RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Journal Name

RSCPublishing

ARTICLE

Folding and Structural Polymorphism of G-quadruplex Formed from a Long Telomeric Sequence Containing Six GGG Tracts

Atsushi Tanaka, Jungkweon Choi*, [a] and Tetsuro Majima*

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2014,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

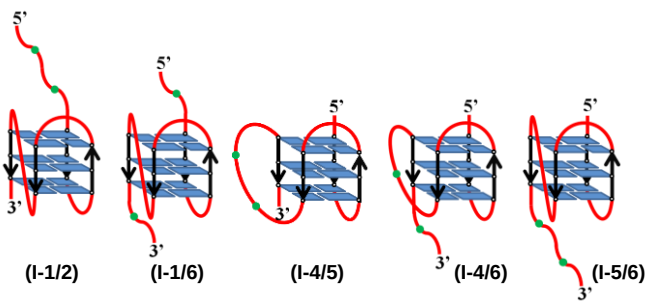
www.rsc.org/

The structure and stability of G-quadruplex formed by a long human telomeric sequence containing five or more TTAGGG repeats are not clear yet. By using the guanine-to-inosine (G-to-I) substitution, thus, we conducted thermodynamic studies on the structural polymorphisms of G-quadruplexes formed by the long telomeric sequences, *37htel* and five G-to-I substituted sequences (I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6), and investigated their folding dynamics in single-molecule level. The thermodynamic study reveals that G-quadruplex formed from I-1/2 has a higher T_m and a larger ΔG_{298K} than those formed by other G-to-I substituted sequences, suggesting that a long telomeric sequence preferentially forms thermodynamically stable G-quadruplex at 3′ end. In addition, from changes in the hydrodynamic radius by the formation of G-quadruplex in single-molecule level, we found that the folding reaction of *37htel* may undergo with two-state mechanism without any detectable intermediate and that the global structure, which leads to the change of a molecular size, is still occurring even after the formation of a secondary structure (G-quadruplex).

Introduction

G-quadruplex is a four-stranded DNA formed by a guanine (G)-rich sequence in the presence of monovalent cations such as Na^+ and K^+ ions.[1-6] Practically, the repetitive $(TTAGGG)_n$ overhang of human telomeric sequence forms G-quadruplex *in vitro*. The structure of G-quadruplex consists of π-π stacking of planar G-tetrads, cyclically bound to each other through eight hydrogen bonds via Hoogsteen base pairs.[1-8] Several research groups have been revealed experimental evidences on presences and functions of G-quadruplex in genomes.[2, 9-11] However, the structure of G-quadruplex in living cells is not clear yet, and its presence in living cells is still the subject of debate.

None the less, G-quadruplex has been received great attention as a target for therapeutic agent in anticancer treatments because of its biological ability to regulate the telomere elongation by telomerase *in vivo*.[5, 12] Furthermore, G-quadruplex is also considered as a promising material for the nanotechnology including a nano-electronic device because of its reversible conformational change and excellent hole trapping ability.[13-15] Especially, the reversible conformational switching between single-stranded DNA (ssDNA) and G-quadruplex can serve as an electronic switch in nano-electronic device. On the other hand, G-quadruplex-forming sequences such as human telomeric sequences have been utilized as a biosensor for K^+ ions *in vitro* and *in vivo* because the G-quadruplex structure cannot be conserved in the absence of a monovalent ion, especially K^+ ion.[16-19] From these perspectives, theoretical and experimental studies on structures and stabilities of human telomeric G-quadruplexes have been extensively conducted *in vivo* as well as *in vitro*, and their structures and biological functions have been highlighted.



37htel	5'-FAM-TA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TT-TAMRA-3'
I-1/2	5'-FAM-TA GIG TTA GIG TTA GGG TTA GGG TTA GGG TTA GGG TT-TAMRA-3'
I-1/6	5'-FAM-TA GIG TTA GGG TTA GGG TTA GGG TTA GGG TTA GIG TT-TAMRA-3'
I-4/5	5'-FAM-TA GGG TTA GGG TTA GGG TTA GIG TTA GIG TTA GGG TT-TAMRA-3'
I-4/6	5'-FAM-TA GGG TTA GGG TTA GGG TTA GIG TTA GGG TTA GIG TT-TAMRA-3'
I-5/6	5'-FAM-TA GGG TTA GGG TTA GGG TTA GGG TTA GIG TTA GIG TT-TAMRA-3'

Scheme 1. Schematic illustration of antiparallel/parallel hybrid (3+1) G-quadruplex structures formed by G-quadruplex-forming sequences (top) and sequences of the oligonucleotide used in this study (bottom).

Practically, numerous human telomeric sequences containing five or more TTAGGG repeats exist in human chromosomes. Nevertheless, most studies on the folding and structure of G-quadruplex in the presence of K^+ ions have been focused on human telomeric sequence containing only four TTAGGG repeats (or four GGG tracts). Although the folding and structure of G-quadruplex formed by long

telomeric sequences with five and more TTAGGG repeats have recently investigated by many research groups,[20, 21] still their folding dynamics and structural polymorphisms are not clear. In addition, the theoretically G-quadruplex can be formed anywhere along the long human telomeric sequence. However, the information for the position where G-quadruplex is formed along the long telomeric sequence has been lacking. Furthermore, little is known about the structural polymorphism of G-quadruplex formed by long telomeric sequences. To address this issue, we thoroughly investigated the folding process of a long G-rich sequence as well as the structural polymorphism of G-quadruplex formed by a long telomeric sequence with six GGG tracts using the thermodynamic analysis and various spectroscopic techniques. In this study, we synthesized and used the dye-labeled human telomeric sequence (5'-FAM-TAGGG(TTAGGG)$_5$-TT-TAMRA-3', 37*htel*) and five guanine-to-inosine (G-to-I) substituted sequences (I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6). (see Scheme 1) FAM (fluorescein) and TAMRA (tetramethylrhodamine) are attached to 5'- and 3'-end, respectively, as an energy donor (D) and an energy acceptor (A) for the fluorescence resonance energy transfer (FRET) experiment. (see Scheme 1) It is known that the replacement by inosine (I) at the middle G of individual GGG tract can selectively induce the exclusion of GIG tract from the formation of G-quartet core, resulting in the reduction of the multiplicity of G-quadruplex conformations.[20] From the thermodynamic studies for six G-rich sequences, we clearly characterized the structures and stabilities of G-quadruplex formed by long G-rich sequences in the presence of K^+ ions. In addition, we demonstrated the folding process of a long telomeric sequence with six GGG tracts in terms of the change of the hydrodynamic radius of DNA in the single-molecule level.

Results and Discussion

Figure 1a shows CD spectra of *37htel* in the absence and presence of 100 mM K^+ ions in 10 mM Tris-HCl buffer solutions (pH 7.4). The formation of G-quadruplex in the presence of 100 mM K^+ ion was confirmed by its characteristic CD spectrum, producing positive band at 290 nm with a shoulder at 265 nm and a negative band at 236 nm (Figure 1a), suggesting that G-quadruplex has a mainly antiparallel/parallel hybrid (3+1) structure. This result, which is consistent with previous reports,[22, 23] also implies that the formation and structure of G-quadruplex upon the addition of K^+ ions are unaffected by the covalent attachment of dyes. To further understand the formation of G-quadruplex structure for *37htel*, we measured the absorption and emission spectra of *37htel* in the absence and presence of K^+ ions under same experimental conditions. (Figure 1b) Although the absorbance of FAM at around 490 nm is slightly enhanced by the increase of the ionic strength upon addition of 100 mM K^+ ions,[24] the absorption spectrum of dyes (D and A) attached to G-quadruplex structure is similar to that in single-stranded DNA (ssDNA). On the other hand, the fluorescence intensity (I_D) of the donor (FAM) attached to *37htel* is significantly quenched by the addition of K^+ ions, whereas the fluorescence intensity (I_A) of the acceptor (TAMRA) is enhanced. (see Figure 1b and S2) The changes in I_D and I_A with adding K^+ ions are due to the conformational change from ssDNA to G-quadruplex structure, resulting in the efficient energy transfer from D to A due to close proximity.

To elucidate the folding reaction of a long telomeric G-quadruplex-forming sequence, we investigated the formation of G-quadruplex structure for *37htel* using the variation of the FRET efficiency ($E_{FRET} = I_A/(I_D + I_A)$) as a function of [$K^+$] and then compared with the formation of G-quadruplex structure for *25htel*, which has a four TTAGGG repeats (5'-FAM-TAGGG(TTAGGG)$_3$-TT-TAMRA-3'). As shown in Figure 1c, the transition curve of *37htel*

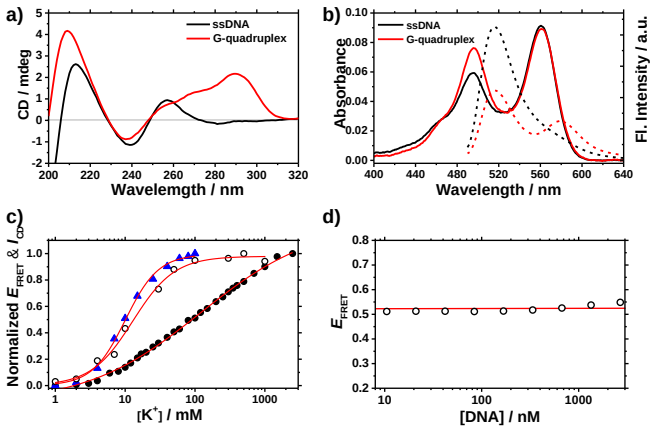


Figure 1. (a) CD spectra of *37htel* in the absence (black) and presence (red) of 100 mM K^+ ions ([*37htel*] = 10 μM). (b) Absorption (straight line) and fluorescence (dotted line) spectra of *37htel* in the absence (ssDNA, black) and presence (G-quadruplex, red) of 100 mM K^+ ions. ([*37htel*] = 168 nM and λ_{Ex} = 485 nm). c) Changes in E_{FRET} for *25htel* (▲) and *37htel* (●) as a function of [K^+] and change in I_{CD} for *37htel* as a function of [K^+] (○). The sequence of *25htel* is 5′-FAM-TAGGG (TTAGGG)$_3$-TT-TAMRA-3′. The I_{CD} is the CD intensity measured at 290 nm. (d) E_{FRET}s for *37htel* at various concentrations.

upon addition of K^+ ions reveals the broad feature compared to that for *25htel*. The transition curves monitored by E_{FRET} (or I_{CD}) are fitted by Eq. (1) (classical Hill equation):

$$E_{FRET} \quad (or \quad I_{CD}) = \frac{\left[K^+\right]^n}{\left[K^+_{1/2}\right]^n + \left[K^+\right]^n} \qquad (1)$$

The n is the Hill coefficient and $K^+_{1/2}$ is the concentration of K^+ ions at which half the G-quadruplex structures were formed. From the transition curve depicted in Figure 1c, $K^+_{1/2}$ and n values for *37htel* were determined to be 117.0 ± 21.7 mM and 0.7, respectively, while$K^+_{1/2}$ and n values for *25htel* were determined as 10.1 ± 0.4 mM and 1.8, respectively. (Figure Table S1) In contrast to the transition curve of *37htel* monitored by E_{FRET}, however, the transition curve of *37htel* monitored by CD intensities is very similar to that of *25htel* ($K^+_{1/2}$ = 14.4 ± 0.8 mM and n = 1.5). (see Figure 2 & 3 and Table S1) The E_{FRET} has been frequently used to measure the distance between domains of a single biomolecule and to obtain the information about the conformational change of a biomolecule. Meanwhile, the CD signal is used to identify the secondary structure of G-quadruplex. Thus, differences in $K^+_{1/2}$ and n values of *37htel* determined by two parameters (I_{CD} and E_{FRET}) suggest that the secondary structure of G-quadruplex is rapidly formed in the low concentration of K^+ ions and then the additional conformational change involving the change in D-A distance takes place in the presence of high concentration of K^+ ions. However, we cannot rule out the possibility that the broad-response of *37htel* for K^+ ions monitored by E_{FRET} is attributed to the G-quadruplex intermolecular interaction in the high concentration of K^+ ions. To test this possibility, we measured the concentration dependence on E_{FRET} of *37htel* in the presence of 1 M K^+ ions. As shown in Figure 1d, no concentration dependence on E_{FRET} of *37htel* was observed, indicating that the broad-response of *37htel* for K^+ ions is probably due to the formation of G-quadruplex and then following

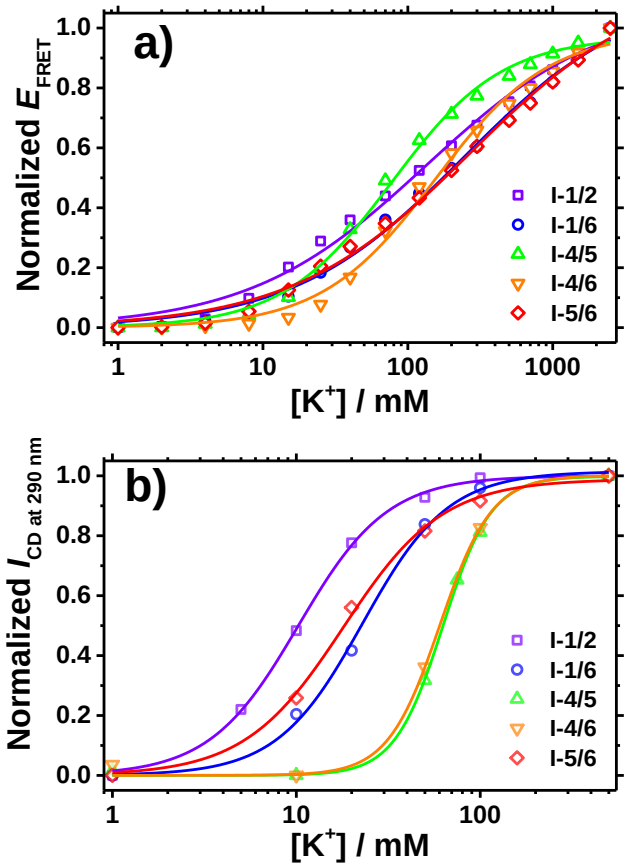


Figure 2. Changes in E_{FRET} (a) and I_{CD} (b) for *37htel*, I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6 as a function of [K+]. Theoretical curves obtained from the fitting analysis using Eq. (1) are shown in solid lines.

an additional conformational change rather than the G-quadruplex intermolecular interaction. The broad-response of *37htel* for K^+ ions will be further discussed later.

As mentioned above, G-quadruplex can be theoretically formed anywhere along the long G-rich sequence such as *37htel*, resulting in the existence of a multi-species with a different G-quadruplex structure as depicted in Scheme 1. The existence of a multi-species can induce the significant difficulty in the study on the structure and stability of G-quadruplex formed by a long human telomeric sequence. To deal with this difficulty, recently Tue et al. used the G-to-I substituted sequence such as I-4/5.[20] As a result, they clearly showed that the GIG tract replaced by inosine (I) at the middle G of individual GGG tract is excluded from the formation of G-quartet core.[20] Substantially, their results showed that the single inosine substitution for d[TA-GGG(TTAGGG)$_n$TT] (n = 5-7), which shows multiple G-quadruplex conformations in the presence of K^+ ions, reduces the multiplicity of its conformation, and revealed that simultaneous inosine substitution in the fourth and fifth G-tracts of *37htel* (I-4/5) led to the emergence of a predominant species (~70%) with a long propeller loop.[20] (see Scheme 1) To elucidate the stabilities and structures of several G-quadruplex structures formed by *37htel* in the presence of K^+ ions, thus, we synthesized five G-to-I substituted sequences (I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6) and investigated the stability and structure of G-quadruplex formed by each oligonucleotide. As depicted in Figure S3, G-quadruplexes formed by five G-to-I substituted sequences show similar CD signals; two positive bands at 270 and 290 nm with a negative band at 240 nm. These results are consistent with that reported by Tue et al.,[20] indicating that all G-quadruplexes formed from five G-to-I substituted sequences have a mainly antiparallel/parallel hybrid (3+1) structure depicted in Scheme 1. However, I-1/2, I-1/6 and I-5/6 form G-quadruplex with the shortest TTA loop, while I-4/6 and I-4/5 form G-quadruplex harboring one or two GIG tracts within a single loop.

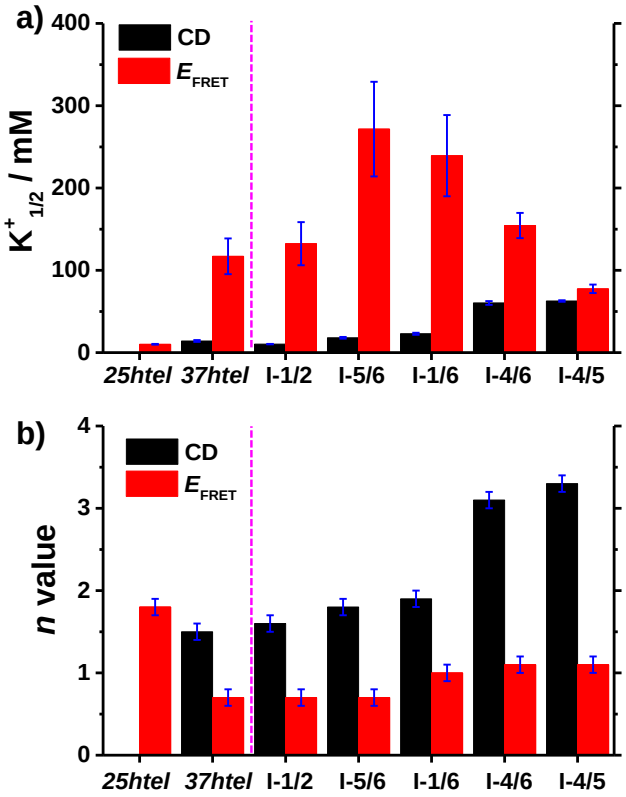


Figure 3. (a) $K^+_{1/2}$ and (b) *n* values for *27htel*, *37htel* and five G-to-I substituted sequences (I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6) determined by E_{FRET} and CD intensities (I_{CD}).

The variation in E_{FRET} and I_{CD} accompanied by the formation of G-quadruplex for five G-to-I substituted sequences with various [K^+] are measured. (see Figure 2) From the quantitative analysis of transition curves obtained from variation of the E_{FRET} as a function of [K^+], $K^+_{1/2}$ values for I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6 were determined to be 132 mM (*n* = 0.7), 239 mM (*n* = 1.0), 77.5 mM (*n* = 1.1), 154 mM (*n* = 1.1), and 272 mM (*n* = 0.7), respectively (Order for $K^+_{1/2}$ value: I-4/5 < I-4/6 < I-1/2 < I-1/6 < I-5/6). (Figure 3 and Table S1) Meanwhile, the transition curves of five G-to-I substituted sequences monitored by CD intensities (I_{CD}) showed contrasting results as shown in Figure 2b; that is, the lower $K^+_{1/2}$ values and the higher *n* values. (Figure 3 and Table S1) It is interesting to note that the $K^+_{1/2}$ values determined from I_{CD} were in the order I-1/2 < I-5/6 < I-1/6 < I-4/6 < I-4/5. These results imply that the formation of G-quadruplex for I-1/2, I-1/6 and I-5/6 is faster than that for I-4/6 and I-4/5. As depicted in Scheme 1, G-quadruplexes formed by I-1/2, I-1/6 and I-5/6 have a short TTA loop compared to those formed by I-4/6

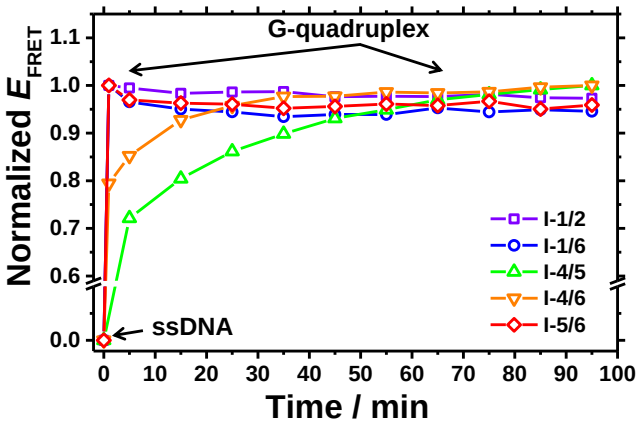


Figure 4. Kinetic aspects for the formation of G-quadruplex triggered by the addition of K^+ ions.

and I-4/5, implying that the loop length affect the formation of G-quadruplex structure. Here, we measured the kinetic aspect for the formation of G-quadruplex by using changes in E_{FRET} monitored at different times after the addition of K^+ ions. As shown in Figure 4, the kinetic aspect for the formation of G-quadruplex triggered by the addition of K^+ ions showed that the formation of G-quadruplex for I-1/2, I-1/6 and I-5/6 is faster than that for I-4/6 and I-4/5. The kinetic traces for I-4/6 and I-4/5 were reproduced with a single-exponential function and the rate constants for G-quadruplex formation for I-4/6 and I-4/5 are determined to be $1.3 \pm 0.1 \times 10^{-3}$ and $0.5 \pm 0.02 \times 10^{-3}$ s^{-1}, respectively. Although we could not determine the rate constant for I-1/2, I-1/6 and I-5/6 with a short TTA loop, the data presented herein clearly reveals that the long G-rich sequence with five or more TTAGGG repeats preferentially forms G-quadruplex with a short TTA loop.

To further elucidate the stability of G-quadruplex formed by each sequence, we measured T_m values of all sequences at various [K^+]. (Figure 5 and S4) As summarized in Table S2 and Figure 5, T_m values of I-1/2, I-1/6 and I-5/6 determined at various [K^+] are higher than those of I-4/5 and I-4/6, indicating that G-quadruplex with a short TTA loop is more thermodynamically stable than that with a long loop. This result coincides with that reported by Koirala et al.[21] Furthermore, the Gibbs free energy change (ΔG_{298K}) for I-1/2, I-1/6 and I-5/6 at 25˚C in the presence of 120 mM K^+ ions are larger than those measured for I-4/5 and I-4/6 (Figure 5b), implying that the structural stability of G-quadruplex is decreased by the presence of a long loop. Indeed, this result agrees with the previous study that suggested the decrease in the stability of G-quadruplex with increasing a total loop length.[25]

Interestingly, as shown in Figure 5b, T_m and ΔG values for I-1/2 determined from the melting curve are greatly higher and larger than those for other G-to-I sequences, implying that G-quadruplex formed from I-1/2 is much more stable compared with those formed from other G-to-I sequences. This means that the thermodynamically most stable G-quadruplex structure formed from *37htel* in the presence of K^+ ions is very similar to that formed from I-1/2. Furthermore, as shown in Figure 5b, the T_m and ΔG values for *37htel* determined from the melting curve are close to those for I-1/2. Considering the G-quadruplex structure formed from I-1/2 (see Scheme 1), we suggest that *in vitro*, a long telomeric sequence containing five or more TTAGGG repeats mainly forms G-quadruplex structure at 3′ end rather than at 5′ end and internal positions. This coincides with the result reported by Tan and coworkers.[26] They suggested that when an open 5′ end is not present, which mimics the telomere G-rich strand *in vivo*, the probability at the 3′ end is higher than at the 5′ end using DMA footprinting and exonuclease hydrolysis.[26] In this study, moreover, we clearly show that from the thermodynamic analysis on G-quadruplex structure, the long telomeric sequence containing five or more TTAGGG repeats with both open 5′ and 3′ ends can form G-quadruplex preferentially at 3′ end rather than at 5′ end and internal positions.

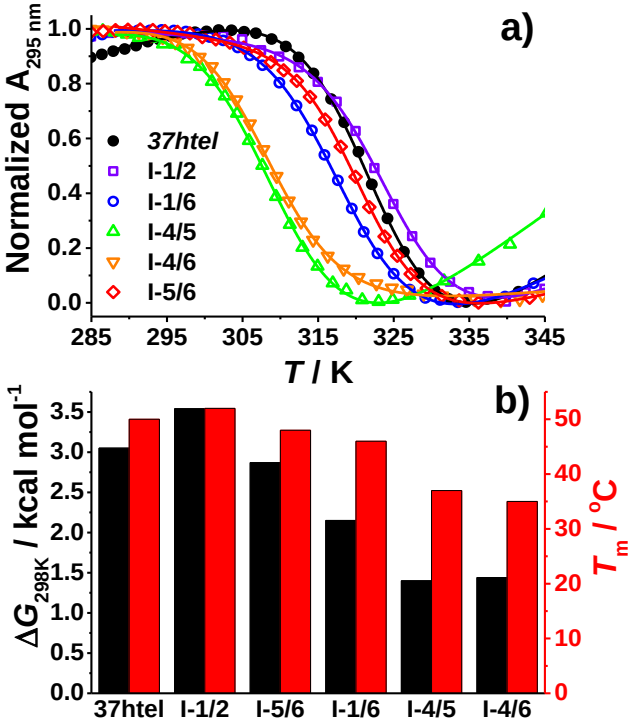


Figure 5. (a) Thermal denaturation curves monitored at 295 nm measured for *37htel*, I-1/2, I-1/6, I-4/5, I-4/6, and I-5/6 in the presence of 120 mM K^+ ions. ([DNA] = 5 μM). Theoretical curves obtained from the fitting analysis using Eq. (S1) are shown in solid lines. (b) Melting temperatures (T_m) and ΔG_{298K} of *37htel*, I-1/6, I-4/5, I-4/6, and I-5/6 determined in the presence of 120 mM K^+ ions.

To further understand the folding dynamics of *37htel* in single-molecule level, on the other hand, we measured the molecular diffusion time of *37htel* as a function of the concentration of K^+ ions using the fluorescence correlation spectroscopy (FCS). FCS is a very useful tool to measure the molecular diffusion time that can provide the information on the change in molecular size associated with the conformational change of biomolecules such as DNA, protein, and so on.[27-30] As depicted in Figure 6a, all FCS curves show two dynamics; one is due to the fast relaxation process corresponding to singlet–triplet (S-T) relaxation of fluorescein and the other is attributed to the translational diffusion of each chemical species. Thus, the autocorrelation function can be expressed by,

$$G(\tau)=\frac{1}{N}\times\left\{\left(1+\frac{\tau}{\tau_d}\right)^{-1}\times\left[1+\left(\frac{s}{u}\right)^2\left(\frac{\tau}{\tau_d}\right)\right]^{-1/2}\right\}\times\left[1+\left(\frac{F}{1-F}\right)\exp\left(-\frac{\tau}{\tau_T}\right)\right] \quad (2)$$

$$\frac{\tau_d'}{\tau_d}=\frac{s^2/4D'}{s^2/4D}=\frac{D}{D'}=\frac{R_h'}{R_h} \qquad \left(D=\frac{kT}{6\pi\eta R_h}\right) \quad (3)$$

where N is the average number of molecules in the observed volume, τ_d is the molecular diffusion time, s/u is the parameter to represent the shape of the observed volume, F is the fraction of the molecules in the triplet state, τ_T is the relaxation time of singlet–triplet relaxation, R_h is a hydrodynamic radius of a molecule, and D is the diffusion coefficient of a molecule.

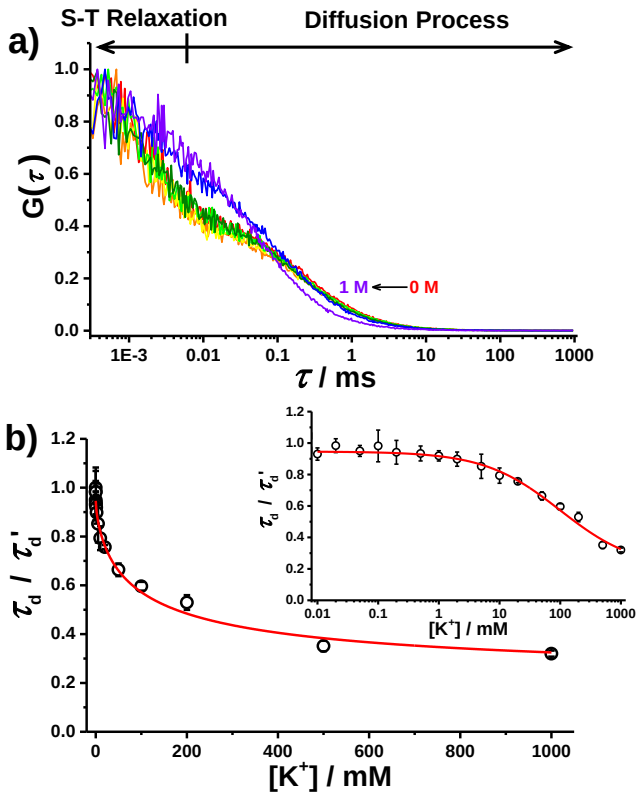


Figure 6. (a) Representative FCS curves of *37htel* measured with 0 (red), 0.01 (orange), 0.1 (yellow), 1 (green), 10 (olive), 100 (blue) and 1000mM (purple) mM K^+ ions. (b) Plot of changes in the molecular diffusion times (τ_d) of *37htel* as a function of [K^+]. The τ_d and τ'_d are the molecular diffusion times in the absence and presence of K^+ ions, respectively. The τ_d values corrected for the increase of the solution viscosity upon addition of K^+ ions. The inset is the log scale plot. Theoretical curves obtained from the fitting analysis using Eq. (2) are shown in solid red lines.

Figure 6b shows changes in molecular diffusion times (τ_d) of *37htel* determined by the quantitative analysis of FCS curves measured at various [K^+] with Eq. (2). It has been suggested that a partially folded G-triplex is an intermediate in the folding reaction of G-quadruplex, and can stably exist in the low concentration of K^+ ions.[21, 31-36] The formation of G-triplex in the low concentration of K^+ ions, thus, should induce the great change in the τ_d of *37htel* because the molecular size of G-triplex is significantly different to that of single-stranded DNA. As shown in Figure 6b, however, *37htel* shows the constant τ_d values in the range of 0 ~ 10 mM, suggesting that G-triplex is not formed in the low concentration of K^+ ions. As shown in Figure 6b, the τ_d of *37htel* is significantly decreased by addition of ≥10 mM K^+ ions and then shows constant τ_d values in the presence of a high concentration of K^+ ions. As expressed in Eq. (3), the τ_d is proportional to the R_h of a molecule. Thus, the decrease of τ_d upon addition of K^+ ions means the decrease of the R_h due to the formation of G-quadruplex. Our result, which is an unexpected result, is in contrast to the result reported by several previous studies, suggesting that a partially folded G-triplex is an intermediate in the folding reaction of G-quadruplex, and can stably exist in the low concentration of K^+ ions.[21, 31-34, 37] Recently, Koirala et al revealed that human telomeric sequences containing four to seven TTAGGG repeats showed the conformational transition into G-quadruplex through a partially folded triplex as well as the direct conformational change from ssDNA to G-quadruplex.[21] However, the folding mechanism of G-quadruplex is still the subject of debate. Indeed, Koirala et al. could not observe the population of a triplex for *hTelo-6*, 5′-(TTAGGG)$_6$TTA-3′,[21] which is consistent with our result. The result reported by Koirala et al. indicates that not all human telomeric sequences containing four or more TTAGGG repeats are folded to G-quadruplex via a G-triplex.[21] In addition, the folding (or unfolding) reaction of G-quadruplex studied by single-molecule techniques did not show the multi-step folding process that occurs with an intermediate such as a hairpin or triplex.[38-40] The presented data herein is consistent with those obtained by several previous single-molecule experiments. Therefore, we suggest that in terms of DNA size change, the folding reaction of *37htel* in single-molecule level occurs without any detectable intermediates, indicating two-state transition.

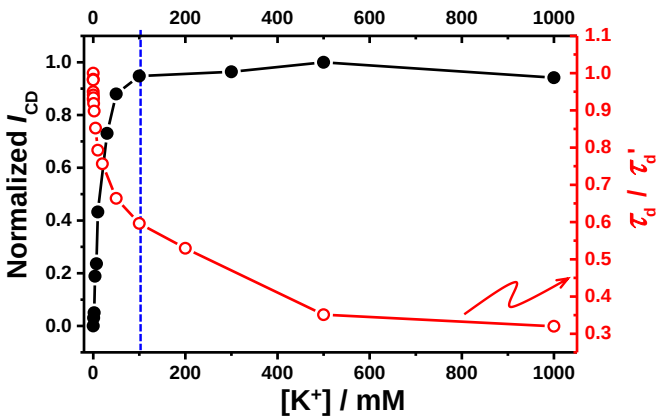


Figure 7. Changes of E_{FRET} and τ_d/τ'_d of *37htel* as a function of the concentration of K^+ ions.

On the other hand, the $K^+_{1/2}$ value of 99 ± 9 mM for *37htel* obtained from changes of τ_d is close to that determined by variation of the E_{FRET}, but is larger than that determined by the variation of I_{CD}. (Figure 7) In addition, the transition region of the transition curve is wider than that obtained from the I_{CD}. (see Figure 7). As mentioned above, the change of τ_d reflects the change in molecular size associated with the global structural of a molecule, whereas the CD signal is sensitive to the secondary structure of G-quadruplex. Thus, the difference in $K^+_{1/2}$ values determined by τ_d and I_{CD} can be

interpreted in terms of the change in the kinetics for the secondary and the global structural change of a DNA sequence upon addition of K^+ ions. In this respect, the smaller $K^+_{1/2}$ value for *37htel* obtained from changes of I_{CD} is probably due to the fast formation of a secondary structure such as G-quadruplex. The unexpected higher $K^+_{1/2}$ values for *37htel* (or five G-to-I substituted sequences) determined by variation of the E_{FRET} may stem from the slow global structural change occurring even after the formation of a secondary structure such as G-quadruplex. The global structural change, which results in the additional change of a molecular size, is due to the existence of a long loop as well as a long overhang at 5′ or 3′ ends.

Conclusions

In contrast to human telomeric sequence containing four TTAGGG repeats, relatively few studies have been done on the structure and stability of G-quadruplex formed from a long human telomeric sequence containing five or more TTAGGG repeats. Substantially, lots of human telomeric sequences containing five or more TTAGGG repeats exist in in human chromosomes, and G-quadruplex can be formed at different positions along the long telomeric sequence. Thus, it is very important to obtain the information on the structure and stability of G-quadruplex formed from a long human telomeric G-rich sequence. From this respect, we have thoroughly investigated the folding process of a long G-rich sequence as well as the structural polymorphism of G-quadruplex formed from long telomeric sequences with six GGG tracts using the thermodynamic analysis and various spectroscopic techniques. The results provided herein revealed that a long telomeric sequence containing five or more TTAGGG repeats forms G-quadruplex dominantly at 3′ end rather than at 5′ end and internal positions. Furthermore, we show that from changes in the hydrodynamic radius associated by the formation of G-quadruplex in single-molecule level, the folding reaction of *37htel* with six GGG tracts can be explained with two-state mechanism without any detectable intermediate. After the formation of a secondary structure (G-quadruplex), a long telomeric sequence containing five or more TTAGGG repeats show a global structural change, resulting in the additional change of a molecular size. We believe that the results provided herein will certainly contribute to understand the structure and stability of G-quadruplex.

Experimental Sections

Full experimental details and characterization of compounds can be found in the Supporting Information.

FAM phosphoramidite, TAMRA CPG, and nucleotide phosphoramidite reagents were purchased from GLEN RESEARCH, Ltd. All oligonucleotides (*25htel*, *37htel*, I-1/2, I-1/6, I-4/5, I-4/6 and I-5/6) studied here were synthesized using Applied Biosystems 3400 DNA synthesizer with standard solid-phase techniques and purified on a JASCO HPLC with a reversed phase C-18 column with an acetonitrile/50 mM ammonium formate gradient. The purified oligonucleotides were lyophilized for three times. Then, all DNA sequences were characterized by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra (Figure S1), and their concentrations were determined by the absorption of FAM labeled to each oligonucleotides.

The steady-state UV−visible absorption, fluorescence, and CD spectra were measured using Shimadzu UV-3100, Horiba FluoroMax-4 and JASCO CD-J720, respectively. To obtain the kinetic aspect for the formation of G-quadruplex at room temperature, the fluorescence spectra were measured at different times following addition of K^+ ions using Horiba FluoroMax-4. The E_{FRET} was determined by I_D and I_A obtained from each spectrum. ($E_{FRET} = I_A/(I_D + I_A)$)

The melting temperature (T_m) was measured using a JASCO V-530. Fluorescence spectra were measured with an excitation wavelength of 485 nm. All sample solutions for the measurement of UV, fluorescence, CD spectrum were prepared with 10 mM Tris-HCl buffer (pH 7.4). The T_m was measured in 10 mM K^+ phosphate buffer (pH 7.4). The thermodynamic parameters and T_m values were determined from the fitting analysis of melting curves using Eq. (S1-S3).

FCS experiments carried out with a time-resolved fluorescence microscope using confocal optics (MicroTime 200; PicoQuant, Berlin-Adlershof, Germany). We used 2 nM sample solutions for measuring FCS curves with various concentrations of K^+ ion.

Acknowledgements

We thank the members of the Research Laboratory for Quantum Beam Science of ISIR, Osaka University for running the linear accelerator. This work has been partly supported by a Grant-in-Aid for Scientific Research (Project 24550188, 25220806, and 25288035) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japanese Government.

Notes and references

The Institute of Scientific and Industrial Research (SANKEN), Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047 (Japan).

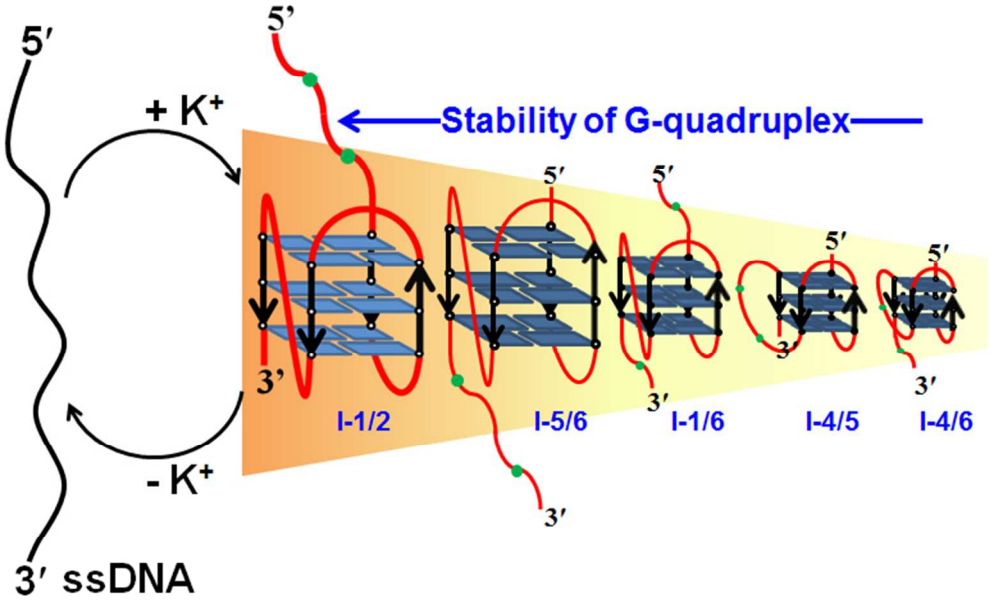
[a] Present address: Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon 305-701, Republic of Korea.

† Electronic Supplementary Information (ESI) available: Experimental details and extra supporting data (Table S1-S2 and Figure S1-S4). See DOI: 10.1039/b000000x/

1. A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, 36, 5482-5515.
2. H. J. Lipps and D. Rhodes, *Trends Cell Biol.*, 2009, 19, 414-422.
3. J. L. Huppert, *FEBS J.*, 2010, 277, 3452-3458.
4. J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, 40, 5893-5909.
5. Y. Xu, *Chem. Soc. Rev.*, 2011, 40, 2719-2740.
6. M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, 13, 770-780.
7. A. Gonnelli, M. G. Ortore, E. J. Baldassarri, G. P. Spada, S. Pieraccini, R. C. Perone, S. S. Funari and P. Mariani, *J. Phys. Chem. B*, 2013, 117, 1095-1103.
8. S. Lena, G. Brancolini, G. Gottarelli, P. Mariani, S. Masiero, A. Venturini, V. Palermo, O. Pandoli, S. Pieraccini, P. Samorì and G. P. Spada, *Chem. Euro. J.*, 2007, 13, 3757-3764.
9. A. Verma, V. K. Yadav, R. Basundra, A. Kumar and S. Chowdhury, *Nucleic Acids Res.*, 2009, 37, 4194-4204.
10. D. Sun, K. Guo and Y. J. Shin, *Nucleic Acids Res.*, 2011, 39, 1256-1265.
11. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, 5, 182-186.
12. G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, 40, 5867-5892.
13. Y. C. Huang and D. Sen, *J. Am. Chem. Soc.*, 2010, 132, 2663-2671.
14. S. Liu, X. Zhang, W. Luo, Z. Wang, X. Guo, M. L. Steigerwald and X. Fang, *Angew. Chem. Int. Ed. Engl.*, 2011, 50, 2496-2502.
15. J. Choi, J. Park, A. Tanaka, M. J. Park, Y. J. Jang, M. Fujitsuka, S. K. Kim and T. Majima, *Angew. Chem. Int. Ed. Engl.*, 2013, 52, 1134-1138.
16. Y. Hong, M. Haussler, J. W. Lam, Z. Li, K. K. Sin, Y. Dong, H. Tong, J. Liu, A. Qin, R. Renneberg and B. Z. Tang, *Chem. Eur. J.*, 2008, 14, 6428-6437.
17. C. K. Kwok, M. E. Sherlock and P. C. Bevilacqua, *Angew. Chem. Int. Ed. Engl.*, 2013, 52, 683-686.
18. C. K. Kwok, M. E. Sherlock and P. C. Bevilacqua, *Biochemistry*, 2013, 52, 3019-3021.

19. Z. Chen, L. Chen, H. Ma, T. Zhou and X. Li, *Biosens. Bioelectron.*, 2013, 48, 108-112.
20. D. J. Yue, K. W. Lim and A. T. Phan, *J. Am. Chem. Soc.*, 2011, 133, 11462-11465.
21. D. Koirala, C. Ghimire, C. Bohrer, Y. Sannohe, H. Sugiyama and H. Mao, *J. Am. Chem. Soc.*, 2013, 135, 2235-2241.
22. H. Q. Yu, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, 128, 15461-15468.
23. A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, 35, 6517-6525.
24. R. Sjöback, J. Nygren and M. Kubista, *Spectrochim. Acta Part A*, 1995, 51, L7-L21.
25. A. Guedin, J. Gros, P. Alberti and J. L. Mergny, *Nucleic Acids Res.*, 2010, 38, 7858-7868.
26. J. Tang, Z. Y. Kan, Y. Yao, Q. Wang, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2008, 36, 1200-1208.
27. H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson and S. Chu, *Proc. Natl. Acad. Sci. USA*, 2002, 99, 4284-4289.
28. J. Choi, S. Kim, T. Tachikawa, M. Fujitsuka and T. Majima, *J. Am. Chem. Soc.*, 2011, 133, 16146-16153.
29. A. P. Fields and A. E. Cohen, *Proc. Natl. Acad. Sci. USA*, 2011, 108, 8937-8942.
30. A. Tanaka, J. Choi, S. K. Kim and T. Majima, *J. Phys. Chem. B*, 2013, 117, 6711-6717.
31. M. Bončina, J. Lah, I. Prislan and G. Vesnaver, *J. Am. Chem. Soc.*, 2012, 134, 9657-9663.
32. V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat and M. Parrinello, *Angew. Chem. Int. Ed. Engl.*, 2013, 52, 2269-2273.
33. W. Li, X. M. Hou, P. Y. Wang, X. G. Xi and M. Li, *J. Am. Chem. Soc.*, 2013, 135, 6423-6426.
34. J. J. Green, L. Ying, D. Klenerman and S. Balasubramanian, *J. Am. Chem. Soc.*, 2003, 125, 3763-3767.
35. M. Boncina, J. Lah, I. Prislan and G. Vesnaver, *J. Am. Chem. Soc.*, 2012, 134, 9657-9663.
36. W. Li, X.-M. Hou, P.-Y. Wang, X.-G. Xi and M. Li, *J. Am. Chem. Soc.*, 2013, 135, 6423-6426.
37. R. D. Gray, R. Buscaglia and J. B. Chaires, *J. Am. Chem. Soc.*, 2012, 134, 16834-16844.
38. R. D. Gray and J. B. Chaires, *Nucleic Acids Res.*, 2008, 36, 4191-4203.
39. M. H. Qureshi, S. Ray, A. L. Sewell, S. Basu and H. Balci, *J. Phys. Chem. B*, 2012, 116, 5588-5594.
40. X. Long, J. W. Parks, C. R. Bagshaw and M. D. Stone, *Nucleic Acids Res.*, 2013, 41, 2746-2755.

A long telomeric sequence preferentially forms thermodynamically stable G-quadruplex at 3′ end rather than at 5′ end and internal positions.



275x177mm (72 x 72 DPI)