Bioorganic & Medicinal Chemistry 16 (2008) 6873–6879

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution

Kenji Okamoto †, Yuta Sannohe, Tomoko Mashimo, Hiroshi Sugiyama, Masahide Terazima *

Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

ARTICLE INFO

Article history:
Received 23 April 2008
Revised 23 May 2008
Accepted 24 May 2008
Available online 29 May 2008

Keywords:
Cancer therapeutic target
Triple-strand-core model
Confocal microscopy
Bin size-independent distribution analysis

ABSTRACT

The human telomere is known to form the G-quadruplex structure to inhibit the activity of telomerase. Its detailed structure has been of great interest. Recently, two kinds of the parallel–antiparallel hybrid structures have been specified in K^+ solution. However, the G-quadruplex structure is generally thought to be in equilibrium among different structures. Here, we describe the single-pair fluorescence resonance energy transfer (sp-FRET) experiments on telomere samples with bromoguanine (BrG)-substitutions, which control the G-quadruplex structures, at different positions and one without any substitution. The observed FRET distributions were decomposed into five components and the relative population of these components depended on the BrG-substitution positions. In order to consistently explain the variety of conformations, we proposed a novel structural model, the so-called triple-strand-core model. On the basis of this model, the components of the FRET distributions were attributed to the mixed-chair hybrid structures, which were reported recently, and chair-type antiparallel structures, which can be predicted from this model. The FRET efficiencies of these structures were explained in terms of partially broken structures due to steric hindrance and inappropriate capping. This basic model also consistently explains experimental results reported previously. Furthermore, using this model, the folding pathway of the hybrid structures and T-loop formation can be predicted.

1. Introduction

Telomeres at the ends of eukaryotic chromosomes are thought to play important roles in biological processes, including cell aging and death. Telomere commonly contains tandem repeats of Guanine-rich sequences with no genetic information. Telomeres are shortened on DNA replication and can be elongated by the enzyme, telomerase. The telomerase's activity is detected in >80% of cancer cells[1] and hence thought to be related to proliferation of cancer cells. Telomere inhibits its activity by forming a three-dimensional structure, the so-called G-quadruplex, and hence it has become an attractive therapeutic target for the development of novel anticancer agents.[2–9] Human telomere has a G-rich sequence $(GGGTTA)_n$ and its G-quadruplex consists of stacked G-tetrad planes, which are stabilized by monovalent cations (K^+, Na^+) and connected by TTA linkers. The detailed molecular structure, including the glycosidic *syn*/*anti* conformation of guanine residues, the relative orientation of the G-tetrad planes and the conformations of linking loops, has been of great research interest. In 1993, a basket-type antiparallel structure was first determined for four-repeat telomere with the 22nt sequence of $d[AGGG(TTAGGG)_3]$ in Na^+ solution by NMR analysis.[10] In 2002, propeller-type parallel structure was obtained for the same 22nt sequence by X-ray diffraction in a crystal grown in the presence of K^+ ions.[11] However, the structure in K^+ solution had been unknown for years, although it was an important structure under the physiological condition. In 2006, Xu et al. proposed a detailed structure of $d[AGGG(TTAGGG)_3]$ based on CD and NMR experiments.[12,13] They investigated the thermal stability of a series of systematic derivatives of telomeric DNA, in which guanines were substituted with 8-bromoguanines. This result indicated that the G-quadruplex consisted of three parallel G-tracts and an antiparallel tract connected by two lateral and a double-chain-reversal loops. The so-called hybrid-1 mixed-chair-type structure is shown in Figure 1(a). The glycosidic conformations were also determined identically as indicated by color: *anti* as blue and *syn* as red. At almost the same time, two other groups reported the same structure with the similar sequences obtained by NMR measurements.[14–16] Within a year, another mixed-chair structure, the so-called hybrid-2, shown in Figure 1(b), was identified by NMR.[17–19] Hybrid-2 structure has different loop and glycosidic conformations from those of the hybrid-1, but has similar structure consisting of three parallel and an antiparallel G-tracts connected by two lateral and a double-chain-reversal loops. While some solution structures were determined, the structural dynamics among them is not understood well. The folding process to

* Corresponding author. Tel.: +81 75 753 4023; fax: +81 75 753 4000.
E-mail address: mterazima@kuchem.kyoto-u.ac.jp (M. Terazima).
† Present address: Cellular Informatics Laboratory, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.05.053

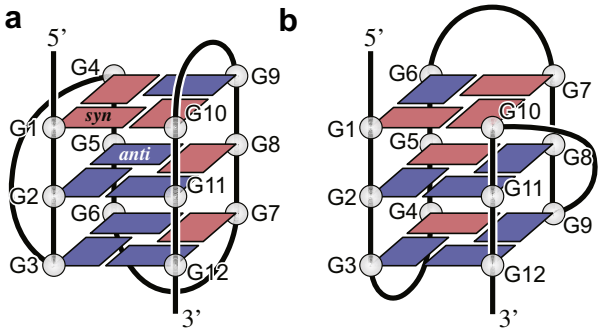

Figure 1. Experimentally determined G-quadruplex structures of human telomere DNA. (a) Hybrid-1 mixed-chair.[12–16] (b) Hybrid-2 mixed-chair.[17,18] Glycosidic conformations are indicated by color: *anti* as blue, *syn* as red.

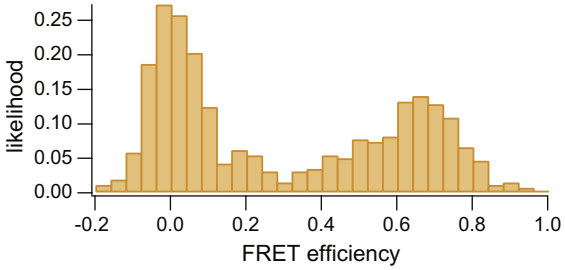

Figure 2. FRET histogram for native species obtained by sp-FRET measurement.

solution structures is another puzzling problem although there are some proposed models.[20]

In this paper, we investigated the G-quadruplex structure of human telomere DNA using the single molecule measurement of the fluorescence resonance energy transfer (FRET) and bromoguanine(BrG)-substitution. The single molecule measurement is free from ensemble-averaging and enables us to measure individual molecules distinctively. This technique is advantageous to the investigation of G-quadruplex structures, since it has been considered that the telomere DNA has some structures in K^+ solution and may even be in equilibrium among them. The existence of these multiple conformations is one of the problems, which had prevented us from identifying the G-quadruplex structure. The FRET measurement was realized by labeling double-strand DNA (dsDNA) with two kinds of fluorescent dye molecules. The difference in molecular conformation can be detected as the ratio of the fluorescence intensities and resolved in dimensions of angstroms. The single-pair FRET (sp-FRET) measurement has shown its ability to distinctively detect coexisting molecular species[21] and has been successfully employed to investigate the DNA structures.[22,23] Especially, Lee et al. made a series of sp-FRET experiments with human telomere G-quadruplex.[23] However, their discussion was based on contemporary knowledge in 2005, when the stable structures in K^+ solution were unknown, and focused on the low K^+ concentration case presumably because the FRET distribution with higher K^+ concentration seemed broad and unresolvable. Here we introduced the technique of BrG-substitution at specific positions, which can control the G-quadruplex folding topology.[12,13,19,24] We prepared four kinds of oligonucleotides with BrG-substitution in the G-quadruplex sequence as well as unmodified (native) one and examined the dependence of the FRET efficiency distribution on the BrG-substitution positions. We propose a novel triple-strand-core model and show that it consistently explains our experimental results as well as those reported previously.

2. Results

The distribution of the FRET efficiency for native species was obtained by the imaging sp-FRET experiment and plotted as a histogram in Figure 2. The result composed of two broad distributions is similar to that obtained by Lee et al. in the presence of 100 mM K^+.[23] The distribution at $E_{FRET} = 0$ corresponds to the signals from acceptor-inactive species, typically consisting of the molecules, on which the acceptor is photobleached. Another broad peak is formed at high FRET region $E_{FRET} > \sim 0.4$. Since the donor position of our sample is further from quadruplex part for 4 bases than that of Ref. 23, the broad peak's position is shifted toward the lower E_{FRET}. Since the peak at $E_{FRET} = 0$ contained no information on the molecular structure, we focused our attention on the other distributions in the following.

The imaging sp-FRET measurements were conducted for four other BrG-substituted species and the FRET histograms of all five species in the region $E_{FRET} > 0.3$ are shown in Figure 3. The histogram shapes are dependent on the samples, which have different BrG-substitution positions.

In order to further analyze those distributions, we first composed a data set by taking a summation of all data points in Figure 3 and plotted it as a histogram in Figure 4(a). The components in the distribution were analyzed using a recently proposed method.[21] Briefly, the FRET efficiency distribution, which was composed of 1,129 points in the range $E_{FRET} > 0.35$, was plotted in a style of a cumulative distribution function (CDF) (Fig. 4(b)) and the plot was fitted by a sum of error functions. The number of the error functions was varied and fitting residuals were calculated for these fittings (Fig. 4(c)). The number of the components in the distribution was determined based on the following criterion. When the number of the function was smaller than that for the best fitting of the distribution, the residuals should significantly decrease with increasing the number of the error function. On the other hand, when the number was larger than the optimal number, the residual may further decrease but the change should not be significant with increasing the number further. In this case, the number of components was determined to be five (Fig. 4(c)) and the peak positions of these components were determined to be E_{FRET} ~0.44, ~0.56, ~0.64, ~0.73, and ~0.82. Here, we named those peaks as low, mid-low, middle, mid-high, and high, respectively, in the order of the FRET efficiency. The best-fitted error functions were converted to the Gaussian functions and are shown on the histogram in Figure 4(a).

Next, we applied the CDF-based fitting analysis to the five distributions of single DNA species. The molecules in the range $E_{FRET} > 0.35$, the numbers of which were 267, 227, 205, 185, and 245 for native, BrG8, BrG2,8, BrG5, and BrG5,11 species, respectively, were dedicated to the analyses. The positions of five peaks were fixed at those determined in the above analysis under the assumption that the distributions consisted of common components and they varied their populations depending on the BrG-substitutions. The populations of five components were plotted in Figure 5 for the species we investigated. It should be noted that the populations were normalized so that the sum of five components is unity.

3. Discussion

In our results, the broad distributions were observed in all five species and decomposed into five components, namely, low, mid-low, middle, mid-high, and high peaks in the order from low to high E_{FRET}. This result indicates that several structures coexist in each species possibly under equilibrium. At the same time, it

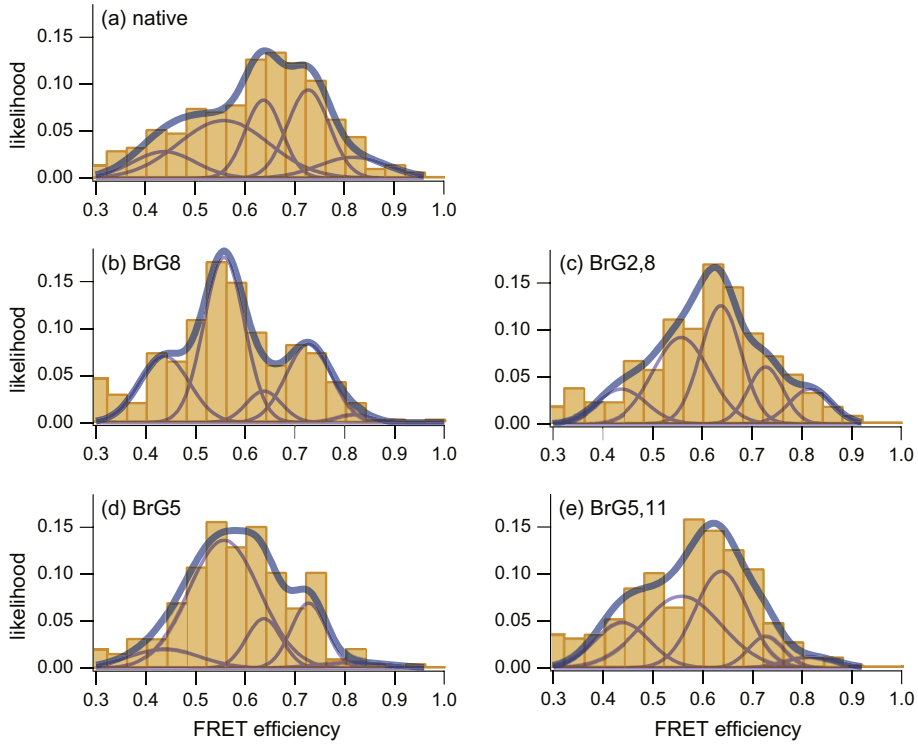

Figure 3. FRET histograms of native and four BrG-substituted DNA species in the range $E_{FRET} > 0.35$. Gaussian functions show the components obtained by the distribution analyses.

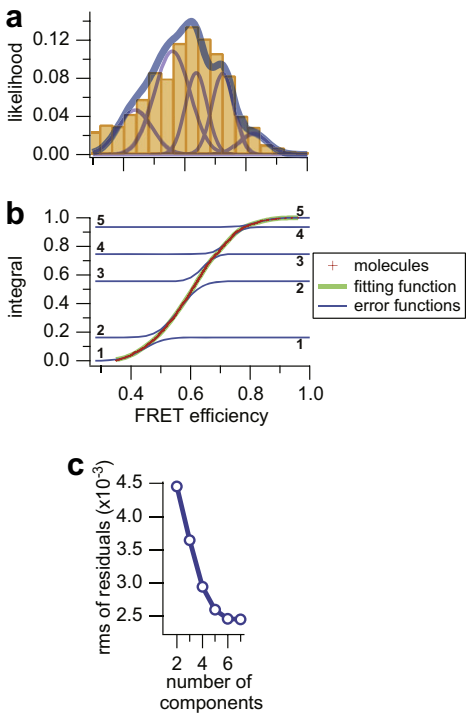

Figure 4. (a) FRET histogram including all data points from five species in the range $E_{FRET} > 0.35$. (b) The same data set is plotted in a CDF-style and analyzed by fitting error functions. (c) The dependence of the rms of the fitting residual on the number of fitting functions indicates the number of components is five in this case. The resolved components are shown as Gaussian functions in (a) and as error functions in (b).

should be noted that the fractions of these components are dependent on the species.

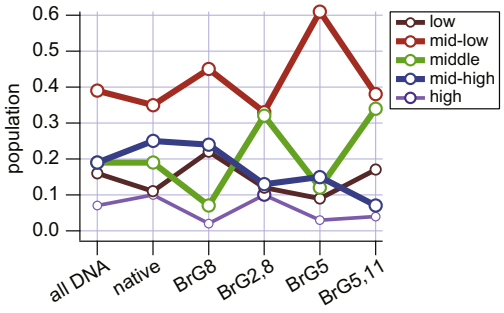

Figure 5. The dependence of populations of five components on BrG-substitution positions. Data labeled as 'all DNA' are obtained from the combined data from all five species and these correspond to the distribution shown in Figure 4(a). Other data correspond to specific species and correspond to the distributions shown in Figure 3.

For the assignment of DNA species to resolved components, we paid attention to the dependence of the population on species (Fig. 5), particularly the effect of the BrG-substitution. First, we noted that the population of the mid-high peak of the native form and BrG8 is relatively large compared with those of the other species. Considering that the hybrid-1 mixed-chair-type quadruplex is expected to be stabilized by BrG8, we attributed the mid-high peak to the hybrid-1 mixed-chair structure. Furthermore, we should note that this mid-high peak is the highest distribution with a narrow width for the native species (Fig. 3). Since the hybrid-1 structure is the first experimentally determined structure, it is reasonable that this structure is highly populated under the native condition. Second, the population of the mid-low peak is characteristically large for BrG5. It is expected that the BrG-substitution at G5 stabilizes the *syn*-type of G5, that is, the hybrid-2 structure. Hence, we attributed the mid-low peak to the hybrid-2 structure. Third, the middle peaks of BrG2,8 and BrG5,11 are relatively large

compared with those of the other species. Therefore, these peaks correspond to a structure that is commonly preferable for these two species. Fourth, the high peaks are small and almost constant for all species. However, examining the distribution in Figure 3, we noticed that the peak is distinct only for BrG2,8 and the other ones are rather broad and not distinct. Hence, we consider that the high peak is related to a preferable structure of BrG2,8. Finally, the low peak is larger for BrG8 and BrG5,11 than for others and hence is related to a preferable structure of those species.

In addition to the mid-high and the mid-low peaks, which were assigned to the previously determined hybrid-1 and -2 structures as discussed above, other peaks may be related to specific structures. We would like to further consider the detailed structures of unspecified peaks, too, based on the above assignments. The possible structures may be similar to stable hybrid structures and be modifiable from them by small perturbations like a few BrG-substitutions. If this is the case, it may be reasonable to consider that the structures consist of a rather rigid core part and a modifiable peripheral part. Here, we should notice that both the hybrid structures contain the common structure including *anti*/*syn* glycosidic arrangements. We considered that this common part could be the core and propose a core structure model of the G-quadruplex, the 'triple-strand-core'. Figure 6(a) shows the 'triple-strand-core' consisting of three repeats of the human telomere DNA sequence. The G-bases coordinate two cations, which may be K^+ under the physiological condition. Three G-tracts are connected by two lateral loops of TTA linkers. A T and an A bases on a lateral loop coordinate the incomplete G-quartet surface at the top and the bottom and may stabilize the core by capping as Dai et al. suggested.[16,18] This molecular structure core (Fig. 6(a)) is included in both the hybrid-1 and -2 mixed-chair structures, which are all of experimentally determined structures at this moment. If there is another repeat of telomere sequence on the 5′- or the 3′-end as shown with dimmed color in Figure 6, it may be bound via a double-chain-reversal loop to form the hybrid-1 or -2 structure, respectively. Considering this core model, we can predict other possible inter-convertible structures from the hybrid structures. For example, since the binding of the extra G-tract, which is connected via a double-chain-reversal loop, may be weaker than other three G-tracts forming the core, the extra G-tract may be relatively easily unbound from the core and flipped to form the chair-type antiparallel structures, which were suggested previously[12,25] (Fig. 6(c) and (e)). Hence, the quadruplex structure may be equilibrated between the mixed-chair hybrid and the chair-type antiparallel structures in two systems as shown in Figure 6. Since the BrG-substitutions in BrG2,8 and BrG5,11 adapt to the chair-1 and -2, respectively, those structures may be preferable for these species. If this consideration is correct, the middle peak is assigned to the chair structures. The high peak is also assigned to the chair structure but only to the chair-1 since BrG5,11 does not form a distinct peak at high position as shown in Figure 3(e). Moreover, this 'triple-strand-core' model can explain other previously reported structures as will be discussed later.

Comparing the hybrid and the chair structures shown in Figure 6, we can explain some of the differences in the FRET efficiency among the structures. In our system, the distance and the directions of the 5′- and the 3′-ends of the quadruplex part are important since the acceptor is attached on the 5′-end and the donor is labeled on the double-strand part, which is connected to the 3′-end. In the chair structures, the G-tracts at the 5′- and 3′-ends are almost antiparallel and the donor and the acceptor are placed on the same side of the quadruplex. On the other hand, the dyes

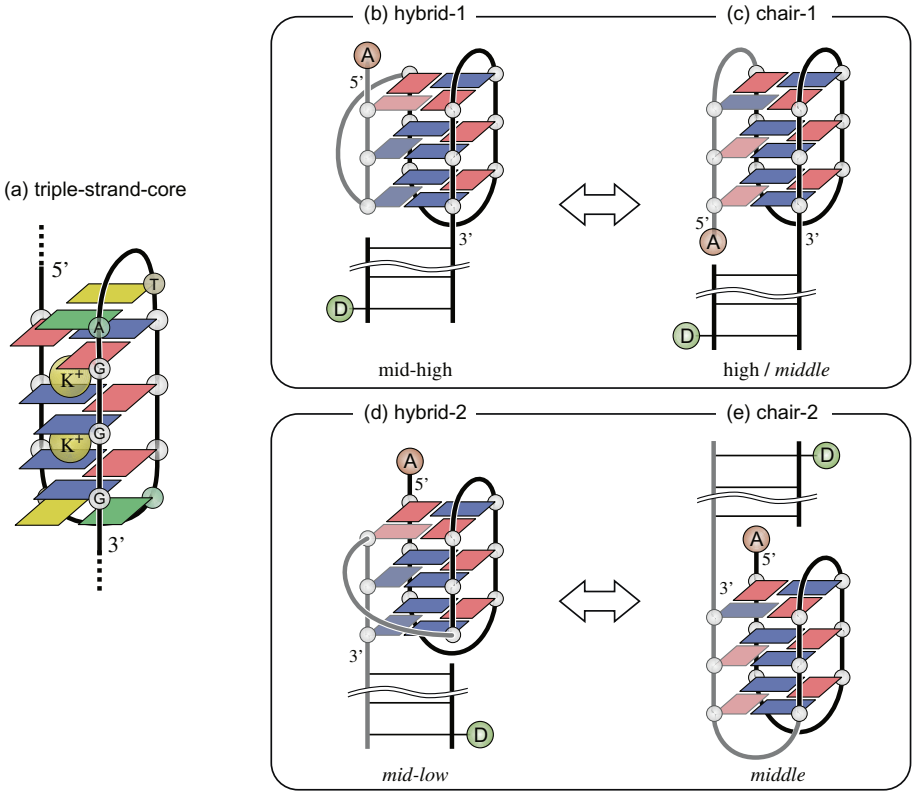

Figure 6. The proposed core structure and equilibrating states. (a) The triple-strand-core is commonly included both in (b) the hybrid-1 and (d) the hybrid-2 structures. The quadruplex with an extra G-tract on the 5′-end of the core forms (b) the hybrid-1 or (c) the chair-1 structures. Addition of an extra G-tract on the 3′-end results in (d) the hybrid-2 or (e) the chair-2 structures. The hybrid-1(2) and the chair-1(2) structures may be equilibrated in solution. For each structure, our assignment to FRET components is shown under the model (italic indicates broken structures).

are separated over the quadruplex and extended to the opposite direction in the hybrid structures. Therefore, it seems to be reasonable that the high peak is assigned to the chair-1 showing higher E_{FRET} than other hybrid structures.

However, there are some questions on a relation between the assigned structures and the experimentally observed FRET efficiencies. (1) Our assignment indicates that the FRET efficiency of the hybrid-2 (mid-low peak) is lower than that of the hybrid-1 (mid-high peak). Why is there this difference in the FRET efficiency between two hybrid structures? (2) Why is the FRET efficiency of the chair structure (middle peak) lower than that of the hybrid-1 structure? (3) Why is the chair-1 populated both in the high and the middle peaks while the chair-2 is populated only in the middle peak? These questions cannot be solved as long as we consider only the ideal structures (Fig. 6). However, if the G-quadruplex structures are partially broken, for example, the extra G-tracts is unbound due to relatively weak bonds in the triple-strand-core model, the distance between the donor and acceptor dyes may become longer and such structures may show a lower FRET efficiency. Below, we consider instabilities of the G-quadruplex structures of these isomers caused by some factors arising from coordination on the G-quartet surfaces to explain these questions.

First, we consider steric repulsion near the 3′-end of the quadruplex between the lateral loop and the complementary stem strand. Although the quadruplex structures are often illustrated as a parallelepiped as shown in Figures 1 and 6, it is actually a twisted hexahedron and hence not rotationally symmetrical. The conformations of the DNA backbone and bases near the G-quartet surface at the 3′-end are illustrated in Figure 7. There is a lateral loop to provide bases to cap the G-quartet surface.[16,18] In the hybrid-1 and the chair-1, the lateral loop is connected to the next two G-tracts on the clockwise side of the 3′-end, while the double-strand part extends to the counterclockwise side (Fig. 7(a)). There is sufficient space for the complementary strand to penetrate to couple with the last base in the double-strand part. On the other hand, the lateral loop in the hybrid-2 and the chair-2 is connected to the next two G-tracts on the counterclockwise side and facing to the double-strand part (Fig. 7(b)). It seems to sterically disturb the end part of the complementary strand (shown as broken lines in the figure). This instability may partially destroy the G-quadruplex structures of the hybrid-2 and the chair-2 so that the FRET efficiencies of these isomers are expected to be smaller than those of the hybrid-1 and the chair-1, respectively.

Secondly, the difference in the capping of the G-quartet surface near the 3′-end may affect the stability of the hybrid structures. It was reported that an A base from the flanking segment was coupled with a T from a lateral loop by hydrogen bonding and capped the G-quartet surface in the hybrid-1,[16] presumably to stabilize the G-quadruplex. On the other hand, a T base on the flanking segment caps the tetrad surface with a T and an A bases from a lateral loop in the hybrid-2.[18] In our experiments, the base nearest to the quadruplex part in the double-strand part, which corresponds to the flanking segment, is the A. Hence, it is plausible that the hybrid-1 is stabilized while the hybrid-2 is unstable. This difference may lead to the smaller FRET efficiency of the hybrid-2 structure compared with that of the hybrid-1.

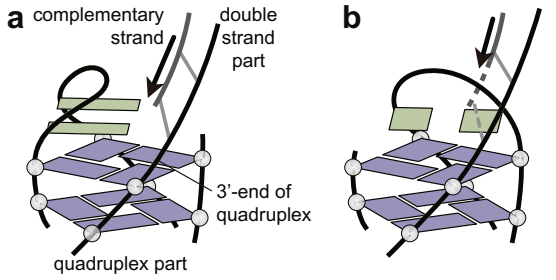

Figure 7. Illustrations of the base and backbone conformations near the 3′-end of the quadruplex part. (a) Hybrid-1quadruplex allows the complementary strand of the double-strand part to penetrate. (b) Hybrid-2 quadruplex seems to interfere with the complementary strand.

Thirdly, as a general problem in the human telomere, two lateral loops facing each other may destabilize the quadruplex structure. A lateral loop is thought to provide two bases to coordinate a G-quartet surface.[16,18] However, in the chair structures, two lateral loops have to face each other on one of the G-quartet surfaces. Coexistence of four bases may be too dense for a G-quartet surface and make the coordination from one of lateral loops incomplete whereas the coordination may stabilize the lateral loop. Xu et al. reported the experiments using BrG-substituted DNA, which was expected to adapt to the chair-1 structure.[12] However, they could not detect the existence of the chair structures, but mainly the hybrid structures. It was shown that the chair structure is difficult to be formed both by bulk and single-molecule experiments. The results indicate the energetic disadvantage of the chair-type structures. This energetic disadvantage will shorten the lifetime of the complete chair structure. Most part of BrG2,8, which should be conformed to the chair, may not be able to form the complete chair structure, which populates in the high peak, but the partially broken chair structure, which populates in the middle peak. BrG5,11 is affected by the first and the second factors in addition to this instability, and therefore it cannot contribute to a detectable distribution as the high peak.

The above considerations still do not explain the low peak's specificity to BrG8 and BrG5,11. Probably, the low FRET efficiency indicates some broken structures and it is difficult to discuss the detailed structures of such broken structures.

In our experimental results, even BrG8 species, which was thought to adopt the most stable structure, did not form a dominant distribution. It is generally thought that the G-quadruplex structure of four-repeat telomere is sensitive to various conditions, such as the BrG-substitutions and the sequence in the flanking segment, in K^+ solution. Even the hybrid structures are stabilized only under specific conditions. Although our experimental condition was not optimized to stabilize specific structures, sp-FRET technique enabled us to resolve subpopulations and observe the competition among them. We particularly paid attention to the changes in equilibrium dependent on the BrG-substitution.

Since the fluorescence photons were averaged over several seconds, the components of the observed FRET efficiency distribution represent the mean of fluctuation during this period which could be longer than typical conformational dynamics of biomolecules in solution. However, it was shown that some of the human telomere structures had a longer lifetime, even more than 200 s.[23] Such stable structures must have contributed to the peak formation in the FRET distributions.

It would be important to mention that the novel triple-strand-core model we proposed consistently explains the experimental results obtained in the previous researches as well as ours. As mentioned before, the core is included in both of the hybrid structures, which are all of the structures experimentally determined with the four-repeat human telomere in K^+ solution. Zhang et al. reported that three repeats of telomeric DNA also tended to form dimers with the (3+1) quadruplex structure, not the (2+2) complex.[26] This report suggested the stability of the triple-strand-core structure. Although their results showed the different glycosidic conformations, it should be noted that their experiments were conducted in Na^+ solution. The triple-strand-core model consistently explains the selectivity of the hybrid structures and the linker conformations. In the experiments using four-repeat sequences, the telo-

mere may be folded into the G-quadruplex stepwise, that is, the triple-strand-core is first formed and then the fourth G-tract is bound. Which hybrid is formed depends on where the triple-strand-core is formed in four repeats, that is, the hybrid-1 is formed when the core is made of three G-tracts from G1 to G9, whereas the core composed of G4–G12 yields the hybrid-2. In four-repeat molecules, the TTA linker connecting the fourth G-tract may lead to a double-chain-reversal loop to avoid the competition between lateral loops. The triple-strand-core model also explains the formation of T-loop structure,[27] which was previously proposed by Xu et al.[12] In this structure, the one of four G-tracts is separated by numbers of telomeric repeats while other three tracts are successive and form the triple-strand-core. Once the core is formed, it may be able to keep its structure for a while without the fourth tract bound. Then another 3G sequence, which happens to approach the core, may be bound. If the fourth tract comes from a separate site, the T-loop structure predicted by Xu et al. is formed. The formation of such kind of T-loop was successfully observed in another series of experiment by Xu et al. by using the oligonucleotides having three successive repeats and one separated telomere sequence.[28]

There is basically no telomere DNA only with four repeats in the nature, that is, the human body. That means that the stable structure composed of successive four repeats of telomere sequence does not necessarily exist. In this paper, we introduced the triple-strand-core model, which can consistently explain the experimental results in this paper as well as those of previous researches and the folding process of the G-quadruplex.

4. Materials and methods

4.1. Oligonucleotides

The reversed-phase HPLC-purified DNA oligonucleotides, which are labeled with the fluorescence dyes for FRET measurements, were purchased from Japan Bio Services Co., Ltd (Saitama, Japan). The Cy-5-labeled oligonucleotides were confirmed by PAGE analysis and the tetramethylrhodamine (TMR)-attached oligonucleotide was confirmed by TOF-MASS spectroscopy (calcd 9312.2, found 9317.6). The homogeneity of all oligonucleotides was confirmed by the HPLC.

The G-quadruplex strand consists of the quadruplex part and the double-strand part and has the sequence 5′-Cy5-$(GGG\ TTA)_3$ GGG AGA GGT AAA AGA ATA ATG GCC ACG GTG CG-3′-biotin. A Cy5 dye is attached on the 5′-end and a biotin for immobilization is attached on the 3′-end. The double-strand part acts as a spacer from the glass surface. The complementary stem strand has the sequence 5′-CGC ACC GTG GCC ATT AT(amino-C6 dT) CTT TTA CCT CT-3′, which is complementary to the double-strand part of the G-quadruplex strand. A tetramethylrhodamine (TMR) dye is attached via amino-C6 dT. The almost same sequences were used by Lee et al. and it was confirmed that the G-quadruplex showed the diverse structures and dynamics without the influence from the surface.[23]

In addition to the native sequence, we prepared four kinds of G-quadruplex strands with the identical sequences except for BrG-substitutions at different positions in the quadruplex part. Since BrG adopts a *syn* conformation, the BrG-substitution stabilizes a structure that has *syn* guanines at the substituted positions, whereas substituting *anti* guanines causes the energetic penalty.[12,13,19,24,29] Comparing hybrid-1 and -2 structures shown in Figure 1, one can find that the glycosidic conformations are identical except for G5 and G8, that is, only hybrid-1 has the *syn*-type G8 and hybrid-2 has the *syn*-type G5. Therefore, we prepared BrG8 with a BrG-substitution only at G8, which may stabilize the hybrid-1 mixed-chair-type quadruplex, and BrG5, which may preferably form the hybrid-2 mixed-chair. We prepared two other strands BrG2,8 and BrG5,11, in which the additional BrG-substitution at G2 and G11 might destabilize the hybrid structures since none of known hybrid structures had *syn* at those positions. Sequences in the quadruplex part of all species are summarized in Table 1.

Table 1
DNA sequences with BrG-substitutions in the quadruplex part of the G-quadruplex strands

Name	Sequence
Native	5′-Cy5-(GGG TTA GGG TTA GGG TTA GGG) ···
BrG8	5′-Cy5-(GGG TTA GGG TTA $GG^{Br}G$ TTA GGG) ···
BrG2,8	5′-Cy5-($GG^{Br}G$ TTA GGG TTA $GG^{Br}G$ TTA GGG) ···
BrG5	5′-Cy5-(GGG TTA $GG^{Br}G$ TTA GGG TTA GGG) ···
BrG5,11	5′-Cy5-(GGG TTA $GG^{Br}G$ TTA GGG TTA $GG^{Br}G$) ···

Experimental solvent was 5 mM cacodylic acid + 100 mM KCl solution. In imaging measurements, DNA was immobilized to glass surface by a tethering method based on biotin–avidin coupling process,[30] in which the final concentration of DNA solution was ~30 pM. Specificity of biotin–avidin immobilization was confirmed by comparison with control experiments without streptavidin (data not shown).

4.2. Imaging sp-FRET measurements

The imaging sp-FRET measurement was conducted on a confocal microscope equipped with two photon-counting detectors. The details of apparatus are described elsewhere.[21] In this paper, we obtained the fluorescence images by sample scanning. An *x*–*y* piezoelectric translator scans 125 × 125 points in the scan area of ~8 × 8 μm in a few minutes. The fluorescence signal was accumulated for 5 ms at each point. An image typically contained several fluorescence spots, which corresponded to single dsDNA, in a weak background signal. A single molecule was scanned in several seconds and typically yielded hundreds of photon counts. Fluorescence signal was integrated on each detector in a rectangular area surrounding a single molecule and the background was subtracted. The FRET efficiency calculated from the integrated fluorescence intensities as $E_{FRET} = (I_A - \beta I_D)/(I_A + \gamma I_D)$ was counted as a data point, where I_A and I_D were the fluorescence intensity on the acceptor and the donor detectors, respectively, β was the coefficient for the leakage of the donor fluorescence onto the acceptor channel, and γ was a factor correcting differences in the dyes' quantum yields, the detection efficiencies between dyes and detectors, and the fluorescence signal leakage. β and γ were experimentally determined from acceptor's photobleach events[31,32] of at least several immobilized molecules for each sample. In our system, the typical values for β and γ were ~0.26 and ~0.82, respectively. Since the FRET efficiency is related to the donor–acceptor distance by $E_{FRET} = \{1 + (r/R_0)^6\}^{-1}$, where R_0 is Förster distance and is typically 50–100 Å, the molecular structure can be inferred from the FRET efficiency, that is, the high E_{FRET} indicates the shorter donor–acceptor distance and the low E_{FRET} suggests the extended molecular structure.

Acknowledgments

K.O. was supported by Research Fellowship of Graduate School of Science, Kyoto University. This work was supported by Grants-in-Aid for Scientific research in Priority Areas from the Japan MEXT.

References and notes

1. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* **1994**, *266*, 2011.

2. Mergny, J.-L.; Hélène, C. *Nat. Med.* **1998**, *4*, 1366.
3. Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441.
4. Perry, P. J.; Jenkins, T. C. *Mini-Rev. Med. Chem.* **2001**, *1*, 31.
5. Hurley, L. H. *Nat. Rev. Cancer* **2002**, *2*, 188.
6. Neidle, S.; Parkinson, G. *Nat. Rev. Drug Disc.* **2002**, *1*, 383.
7. Riou, J.-F. *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 439.
8. Seenisamy, J.; Bashyam, S.; Gokhale, V.; Vankayalapati, H.; Sun, D.; Siddiqui-Jain, A.; Streiner, N.; Shin-ya, K.; White, E.; Wilson, W. D.; Hurley, L. H. *J. Am. Chem. Soc.* **2005**, *127*, 2944.
9. Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167.
10. Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.
11. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876.
12. Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.* **2006**, *14*, 5584.
13. Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. *FEBS J.* **2007**, *274*, 3545.
14. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963.
15. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723.
16. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 2440.
17. Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715.
18. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927.
19. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517.
20. Mashimo, T.; Sugiyama, H. *Nucleic Acids Symp. Ser.* **2007**, *51*, 239.
21. Okamoto, K.; Terazima, M. *J. Phys. Chem. B* **2008**, doi:10.1021/jp712104h.
22. Morgan, M. A.; Okamoto, K.; Kahn, J. D.; English, D. S. *Biophys. J.* **2005**, *89*, 2588.
23. Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938.
24. Xu, Y.; Sugiyama, H. *Nucleic Acids Res.* **2006**, *34*, 949.
25. He, Y.; Neumann, R. D.; Panyutin, I. G. *Nucleic Acids Res.* **2004**, *32*, 5359.
26. Zhang, N.; Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 17277.
27. Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T. *Cell* **1999**, *97*, 503.
28. Xu, Y.; Sato, H.; Shinohara, K.; Komiyama, M.; Sugiyama, H. *Nucleic Acids Symp. Ser.* **2007**, *51*, 243.
29. Dias, E.; Battiste, J. L.; Williamson, J. R. *J. Am. Chem. Soc.* **1994**, *116*, 4479.
30. Ha, T.; Zhuang, X.; Kim, H. D.; Orr, J. W.; Williamson, J. R.; Chu, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9077.
31. Sabanayagam, C. R.; Eid, J. S.; Meller, A. *Appl. Phys. Lett.* **2004**, *84*, 1216.
32. Ha, T.; Ting, A. Y.; Liang, J.; Caldwell, W. B.; Deniz, A. A.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 893.