Published in final edited form as:
J Am Chem Soc. 2009 April 1; 131(12): 4301–4309. doi:10.1021/ja807503g.

Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers

Kah Wai Lim[1], **Samir Amrane**[1], **Serge Bouaziz**[3], **Weixin Xu**[2], **Yuguang Mu**[2], **Dinshaw J. Patel**[4], **Kim Ngoc Luu**[4], and **Anh Tuân Phan**[1,*]

1*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*

2*School of Biological Sciences, Nanyang Technological University, Singapore*

3*Unité de Pharmacologie Chimique et Génétique, INSERM U640 — CNRS UMR 8151, Université Paris Descartes, France*

4*Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, USA*

Abstract

Previously, it has been reported that human telomeric DNA sequences could adopt in different experimental conditions four different intramolecular G-quadruplexes each involving three G-tetrad layers, namely Na^+-solution antiparallel-stranded basket form, K^+-crystal parallel-stranded propeller form, K^+-solution (3+1) Form 1 and K^+-solution (3+1) Form 2. Here we present a new intramolecular G-quadruplex adopted by a four-repeat human telomeric sequence in K^+ solution (Form 3). This structure is a basket-type G-quadruplex with only two G-tetrad layers: loops are successively edgewise, diagonal and edgewise; glycosidic conformations of guanines are *syn*•*syn*•*anti*•*anti* around each tetrad; each strand of the core has both a parallel and an antiparallel adjacent strands; there are one narrow, one wide and two medium grooves. Despite the presence of only two G-tetrads in the core, this structure is more stable than the three-G-tetrad intramolecular G-quadruplexes previously observed for human telomeric sequences in K^+ solution. Detailed structural elucidation of Form 3 revealed extensive base pairing and stacking in the loops capping both ends of the G-tetrad core, which might explain the high stability of the structure. This novel structure highlights the conformational heterogeneity of human telomeric DNA. It revealed a new folding principle for G-quadruplexes and suggests new loop sequences and structures for targeting in human telomeric DNA.

Keywords

anticancer targets; human telomere; intramolecular G-quadruplexes

INTRODUCTION

Telomeres, the ends of linear eukaryotic chromosomes, are important for the chromosomal stability[1]. Human telomeric DNA comprises thousands of tandem repeats of the G-rich sequence $(GGGTTA)_n$, with a 3′-end overhang of 100-200 nt long[2]. Under physiological ionic conditions, these sequences are capable of forming a four-stranded helical structure, known as the G-quadruplex[3-9], based on the stacking of multiple G•G•G•G tetrads[10]. It has been shown that formation of such quadruplex structures by telomeric DNA inhibits the activity of telomerase[11], an enzyme that is critical for the proliferation of most cancer cells[12]. Therefore,

*Corresponding author: phantuan@ntu.edu.sg.

stabilization of telomeric G-quadruplexes by ligands is a promising strategy for the development of anticancer drugs[13,14].

Previous studies on various four-repeat human telomeric G-rich sequences have revealed a diverse range of intramolecular G-quadruplex structures under different experimental conditions[15-44]. The human telomeric d[AGGG(TTAGGG)$_3$] sequence adopts in Na^+ solution an antiparallel-stranded G-quadruplex structure[15], in which the core is made up of three stacked G-tetrads adjoined by one diagonal and two edgewise loops, giving it a basket-like appearance (Figure 1a). The glycosidic conformations of guanines around each tetrad are *syn•syn•anti•anti*; each strand constituting the core is parallel to one adjacent strand and antiparallel to the other adjacent strand; there are one narrow-, one wide- and two medium-width grooves. In contrast, the same sequence assumes a parallel-stranded propeller-type G-quadruplex fold (Figure 1b) in a K^+-containing crystal[16]. All four strands of the core are oriented in the same direction; all guanines adopt *anti* glycosidic conformation and the four grooves are of medium width; the three connecting TTA loops are of the double-chain-reversal type.

Structure of the human telomere in K^+ solution has attracted intense research[17-44], as K^+ is much more abundant than Na^+ in intracellular environments. However, the presence of multiple G-quadruplex conformations in K^+ solution makes structure elucidation difficult[17]. This conformational heterogeneity was overcome by altering the flanking nucleotides of the four-repeat human telomeric sequence[18-20,25-27] and/or by substituting core guanine(s) with guanine analogues[17,20,22-24] such as 8-bromoguanine (^{Br}G), thus favoring only one predominant form. Based on this approach, two different human telomeric G-quadruplex structures (Figure 1c, d) with the (3+1) G-tetrad core topology were discovered[18-27]. The (3+1) core consists of three G-tracts oriented in one direction and the fourth in the opposite direction; one *anti•syn•syn•syn* and two *syn•anti•anti•anti* G-tetrads; one wide, one narrow and two medium grooves. Both structures have one double-chain-reversal and two edgewise loops, but they differ in the order of loop arrangements[19] (Figure 1c, d).

Here, we present the structure of a new intramolecular G-quadruplex conformation formed by four-repeat human telomeric sequences in K^+ solution. This is an antiparallel-stranded basket-type G-quadruplex involving only two layers of G-tetrads. Interestingly, we found that this G-quadruplex is more stable than its three-tetrad counterparts reported previously for human telomeric sequences in K^+ solution. Detailed structural elucidation of this G-quadruplex form revealed extensive base pairing and stacking in the loops capping both ends of the G-tetrad core, which might explain the high stability of the structure. This novel structure highlights the conformational heterogeneity of human telomeric DNA. It indicates a new folding principle for G-quadruplexes and suggests new loop sequences and structures for targeting in human telomeric DNA.

RESULTS

Human telomeric sequences adopt a novel G-quadruplex fold (Form 3) in K^+ solution

Previously, we showed that the natural human telomeric sequences d[TAGGG(TTAGGG)$_3$] and d[TAGGG(TTAGGG)$_3$TT] adopted in K^+ solution predominantly intramolecular (3+1) G-quadruplexes Form 1 (Figure 1c) and Form 2 (Figure 1d), respectively. Both these structures contained three G-tetrad layers, leading to the observation of twelve guanine imino protons at 10.6-11.8 ppm, which were well protected from the exchange with solvent[18-20].

In contrast, imino proton spectra of the 22-nt natural human telomeric d[(GGGTTA)$_3$GGGT] sequence (Table 1) in K^+ solution displayed a different pattern with two major upfield-shifted peaks at 10.5-10.8 ppm and eight major peaks at 11.4-11.9 ppm (Figure 2a). At 45 °C, the

upfield-shifted peaks were broadened due to the exchange with solvent while the eight other peaks remained sharp (Figure S1, Supporting Information), suggesting that the former did not belong to the G-tetrad core. These data suggested that the major form (about 60%) of d[$(GGGTTA)_3GGGT$] was a novel G-quadruplex conformation containing only two G-tetrads (see below). Here we will call the d[$(GGGTTA)_3GGGT$] sequence natural Form 3. A single ^{Br}G-for-G substitution at position G7 of the sequence, resulting in d[GGGTTA(^{Br}G)GGTTAGGGTTAGGGT] (designated as ^{Br}G7-Form 3, Table 1), further increased the population of this G-quadruplex fold to about 90% in K^+ solution (Figure 2b). Characteristic patterns of imino proton spectra observed for several other natural four-repeat human telomeric sequences in K^+ solution (Figure S2) also suggested formation of Form 3 in these sequences.

UV melting and CD studies of human telomeric G-quadruplex Form 3

We recorded CD spectra of three natural human telomeric sequences and their ^{Br}G-substituted versions (Table 1) which form predominantly G-quadruplexes Form 3, Form 2 and Form 1 in K^+ solution, respectively. Their CD signatures are presented in Figure 3a-c. All the CD spectra showed a positive peak at 290 nm, characteristic of antiparallel G-quadruplexes[45]. However, these CD profiles with positive contribution at 270 nm were different from that of the antiparallel basket-type human telomeric G-quadruplex in Na^+ solution, which showed negative peak at 270 nm.[34] The CD profile of Form 3 (Figure 3a) showed positive peaks at 250 nm and 270 nm with the intensity of the 270-nm peak lower than that of the (3+1) G-quadruplex Form 1 and Form 2 (Figure 3b, c).

Melting of the human telomeric G-quadruplexes was monitored by UV absorption at 295 nm.[46] The melting profiles of the three G-quadruplex forms adopted by natural human telomeric sequences and ^{Br}G-modified counterparts (Table 1) are plotted in Figures 3d and e, respectively. The melting temperatures (T_m) are listed in Table 2. The T_m value of natural Form 3 was higher than those of natural Form 1 and Form 2 by 3 °C and 10 °C, respectively. An appropriate ^{Br}G-for-G substitution in each case induced an increase in T_m by 4 to 7 °C and the same trend was conserved with T_m of Form 3 being 5 to 9 °C higher than the values for Form 1 and Form 2 (Table 2). These results were unexpected as Form 3 contained only two G-tetrad layers, while Form 1 and Form 2 contained three G-tetrad layers.

UV melting experiments were performed for different concentrations of ^{Br}G7-Form 3. The fractions of folded and unfolded species were determined using two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states. The normalized melting curves of ^{Br}G7-Form 3 at different DNA concentrations, ranging from 10 μM to 250 μM, showed that the melting temperature was independent of DNA concentrations (Figure 3f). This indicated that Form 3 was a monomeric intramolecular G-quadruplex, consistent with narrow linewidths of the observed NMR peaks.

NMR spectral assignments

NMR spectral patterns indicated formation of the same general G-quadruplex fold for the major form of the natural human telomeric d[$(GGGTTA)_3GGGT$] sequence and its ^{Br}G7-substituted counterpart (Figure 2). The spectral correspondence between the two sequences was helpful for the assignments in both cases. The assignments of many resonances in the ^{Br}G-substituted sequence could be obtained by comparing spectral patterns of the natural and modified sequences. Conversely, the assignments of several peaks for the natural sequence were obtained or confirmed with help of the corresponding cleaner spectra of the ^{Br}G-substituted sequence. This spectral comparison also facilitated the identification of peaks belonging to the major form of the natural sequence.

Guanine imino and H8 protons were unambiguously assigned using the site-specific low-enrichment labeling approach and through-bond correlations at natural abundance[47-50] (Figures 4, 5 and S3; Table S1). For example, Figure 4 shows unambiguous assignments of guanine imino protons in d[$(GGGTTA)_3GGGT$] using samples containing site-specific 2% ^{15}N-labeled G. Resonances for thymine residues were unambiguously assigned using samples containing site-specific 2% ^{15}N-labeled T and systematic T-to-U substitutions[47-50] (Figures S4-S6; Table S1). With help of these unambiguous resonance assignments and other through-bond correlation experiments (COSY, TOCSY and {^{13}C-^{1}H}-HSQC), the classical H8/H6-H1′ NOE sequential connectivities could be traced from G1 through T22 (Figure 6). The assignment correspondence between the natural and modified sequences could be seen in Figures 2, 5, 6, and S7-10. The intensity of intra-residue H8-H1′ NOE cross-peaks (Figure S6) indicated *syn* glycosidic conformation for G1, G7, G14 and G19, in contrast to other residues, which adopted *anti* conformation.

A basket-type G-quadruplex with only two G-tetrad layers

The arrangement of G-tetrads within the core of Form 3 (for both natural and modified sequences) was deduced from characteristic guanine imino-H8 connectivity patterns around individual tetrads (Figure 7). Connecting corners of this G-tetrad core with linking sequences established a basket-type G-quadruplex fold (Figure 8c). The core consists of two tetrads, G1•G14•G20•G8 and G2•G7•G19•G15; the hydrogen-bond directionalities of the two G-tetrads alternate clockwise and anti-clockwise; the glycosidic conformations of guanines around the tetrads are *syn*•*syn*•*anti*•*anti*; each strand constituting the core is parallel to one adjacent strand and antiparallel to the other adjacent strand. The loops are successively edgewise-diagonal-edgewise.

Overall solution structure of human telomeric G-quadruplex Form 3

The structure of Form 3, an antiparallel-stranded basket-type G-quadruplex (Figure 8), adopted by the natural and ^{Br}G-substituted human telomeric sequences was calculated on the basis of NMR restraints (Table 3) (see Methods). The core displays much greater convergence as compared to the loops (Table 3). There is extensive stacking between the five-membered rings of guanines from two G-tetrad layers (Figure 9). This stacking pattern is specific for adjacent G-tetrads with alternate clockwise and anti-clockwise hydrogen-bond directionalities[15].

The strand orientations in the G-tetrad core result in one wide, one narrow, and two medium grooves (Figures 8 and S11). The wide groove is formed between two adjacent antiparallel segments G1-G2 and G7-G8, while the narrow groove is formed between the two other adjacent antiparallel segments G14-G15 and G19-G20. On the other hand, two pairs of adjacent parallel strands, namely G1-G2 and G14-G15, and G7-G8 and G19-G20, give rise to the medium grooves (Figures 8 and S11).

Structure of loops and caps: formation of base triples and base pairs

The two-tetrad G-quadruplex core is protected at both ends by base triples (Figures 8 and 9). The base triple at the top is made up of G9, G13, and G21 (Figure 9b, c). Formation of this base triple is consistent with the observation of imino protons of G21 and G9 at ~10.5 ppm (Figure 2), characteristic for imino protons hydrogen-bonded to oxygen. In the current structure, imino proton of G21 is hydrogen-bonded to O6 of G9, while imino proton of G9 is hydrogen-bonded to O6 of G13 (Figure 9b, c). The base triple consisting of G3, A6, and A18 (Figure 9e, f) was observed at the bottom of the G-tetrad core and stacked well under the G2•G7•G19•G15 tetrad (Figure 9e). Formation of this base alignment was supported by the observation of G3 imino proton (Figures 2 and S3) and NOEs between these bases and the bottom G-tetrad (not shown). This base triple is reminiscent of the G•(C-A) triad observed previously by Kettani et al.[51] (Figure S12).

Structure computation suggested formation of the T11•T22 base pair at the top of the structure (Figure 9) and possibly the T5•T17 base pair at the bottom of the structure. Formation of such base pairs would be consistent with the observation of the imino protons of these thymines at ~10.5 ppm (Figure S5). The base of A12 further stacks on top of the T11•T22 base pair (Figure 8).

The stacking of thymine bases (T4, T10, and T16) into the hydrophobic grooves (Figure 10) might also contribute to the stability of Form 3. At low temperatures and pH 5, the imino protons of these thymines were observed at ~11 ppm (Figure S5). Such conformations of bases seem to occur often in G-quadruplex loops, as they have also been found in other structures[15,20].

Analysis of modified sequences

We modified some residues (Table 1) in the top G9•G21•G13 and bottom A18•G3•A6 base triples, in order to probe the importance of these structural motifs. For the top base triple, modification at position G13, either with a G-to-T substitution or with only a replacement of oxygen 6 by a sulfur, seemed to disrupt formation of the G9•G21•G13 base triple but maintain the general fold of Form 3 (Figure S13). For the bottom base triple, substitution of G3 by a T maintained the general fold of Form 3, while replacing A18 by a C favored other G-quadruplex folds (Figure S14).

DISCUSSION

A new stable basket-type G-quadruplex with only two G-tetrad layers

Intense research over fifteen years on the structure of four-repeat human telomeric G-rich DNA sequences[15,16,18-20,22,23,25-27] has established formation of four different intramolecular G-quadruplex folds, all involving three G-tetrad layers and TTA loops (Figure 1). Here we have shown that four-repeat human telomeric sequences can adopt a new basket-type intramolecular G-quadruplex structure (Form 3 in K^+ solution) involving only two G-tetrad layers (Figure 8). Base pairing and stacking in the loops, whose sequences are GTTA, GTTAG and TTA, respectively, contribute to the stability of this form. This new folding principle contrasts the common thought that a DNA sequence containing four tracts of GGG would be most favorable to form G-quadruplexes with three G-tetrad layers. Interestingly, despite the presence of only two G-tetrad layers, Form 3 is more stable than the previously reported three-tetrad human telomeric intramolecular G-quadruplexes in K^+ solution: for sequences listed in Table 1, the melting temperature of Form 3 is higher than that of Form 1 and Form 2 by 3 to 10 °C. However, it should be noted that the most stable G-quadruplex form for different sequences might be different.

G-quadruplex stability: extensive base pairing and stacking of loop residues

Base pairing and stacking in loops are often found to stabilize G-quadruplexes[20]. In the current structure of the human telomeric G-quadruplex Form 3, extensive base pairing alignments (G21•G9•G13 base triple and possible T22•T11 base pair at the top; A6•G3•A18 base triple and possible T5•T17 base pair at the bottom) were observed in the loops capping both the top and bottom of the G-tetrad core. Combining base pairing layers from each end of the G-tetrad core and the two G-tetrad layers in the core, the structure of Form 3 is made up of totally four to six base pair/triple/tetrad layers (Figure 9), which might explain the high stability of this two-tetrad G-quadruplex. This structure can be compared with the dimeric G-quadruplex of d($A_2G_2T_4A_2G_2$) which involves two tetrads, two triads and two base pairs[52]. The new folding of Form 3 indicates that the overall G-quadruplex topology of a G-rich sequence is defined not only by maximizing the number of G-tetrads, but also by maximizing all possible base pairing and stacking in the loops.

Basket-type G-quadruplexes in the human telomere

The human telomeric G-quadruplex Form 3 in K^+ solution (this work) can be directly compared with the basket-type G-quadruplex observed for the human telomeric d[AGGG(TTAGGG)$_3$] sequence in Na^+ solution[15]. Although the latter contains three G-tetrad layers, the cores of the two structures are quite similar (Figure 11), due to the same strand directionality. In both structures, loop arrangements are edgewise-diagonal-edgewise. However, sequences and sizes of these loops are quite different, so are the base pairing and stacking interactions in the loops.

Based on ^{125}I radio probing, Gaynutdinov et al.[32] proposed a basket-type G-quadruplex fold for human telomeric DNA in K^+ solution. However, the proposed structure contained three G-tetrad layers, in contrast to the two-tetrad structure observed here.

New loops and implications for targeting human telomeric G-quadruplexes

Intramolecular G-quadruplexes formed by human telomeric DNA sequences are potential anticancer targets. Four different G-quadruplex structures have been solved previously and the basket-type G-quadruplex presented here is the fifth one. The previously reported structures show different configurations for TTA loops in various contexts[16,18,20,23,26,27,53-55]. It has been suggested that these loops are dynamic and can be good targets for specific ligand recognitions[20,53-55]. In the present structure, for the first time different loop sequences, GTTA and GTTAG, have been observed in the context of human telomeric G-quadruplexes. The concept of loop dynamics and targeting should be therefore extended to these loop sequence variants.

CONCLUSION

Four-repeat human telomeric sequences can form in K^+ solution a new stable basket-type intramolecular G-quadruplex structure involving only two G-tetrad layers. The new fold revealed new loops (GTTA and GTTAG) in the contexts of human telomeric G-quadruplexes. Extensive base pairing and stacking in the loops contribute to the high stability of this structure.

METHODS

Sample preparation

Unlabeled and site-specific low-enrichment (2% ^{15}N-labeled or ^{15}N,^{13}C-labeled) DNA oligonucleotides (Tables 1 and S1) were chemically prepared using products from Glen Research, Spectra Gases and Cambridge Isotope Laboratories. Samples were dialyzed successively against ~50 mM KCl solution and against water. DNA concentration was expressed in strand molarity using a nearest-neighbour approximation for the absorption coefficients of the unfolded species[56]. The same extinction coefficient was used for the natural and the ^{Br}G-substituted oligonugleotides. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5-2.5 mM; the solutions contained 70 mM KCl and 20mM potassium phosphate (pH 7).

UV melting experiments

The thermal stability of different DNA oligonucleotides was characterized by recording the UV absorbance at 295 nm[46] as a function of temperature (20 to 90 °C) using a Varian Cary 300 Bio UV-Vis spectrophotometer. The heating and cooling rates were 0.25 °C per minute. Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting/folding transition. The melting temperature (T_m) was defined as the temperature of the mid-transition point. Experiments were performed

with quartz cuvettes, 1-cm path-length for low DNA concentrations and 0.2-cm path-length for high DNA concentrations.

Circular dichroism

CD spectra were recorded at 20 °C on a JASCO-810 spectropolarimeter using a 1-cm path-length quartz cuvette in a reaction volume of 800 μl. The DNA oligonucleotides (8 μM) were prepared in 20 mM potassium-phosphate buffer (pH 7.0) containing 70 mM KCl. The samples were heated at 95°C for 5 minutes and cooled down to room temperature overnight. For each sample, an average of 3 scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. The spectra were finally normalized to the concentration of the DNA samples.

NMR spectroscopy

Experiments were performed on 600 MHz and 700 MHz Varian and Bruker spectrometers at 25°C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment labeling and through-bond correlations at natural abundance[47-50]. Imino protons of thymine residues were assigned unambiguously by using site-specific low-enrichment labeling[47]. Other resonances for thymine residues were assigned following systematic T-to-U substitutions[49]. Spectral assignments were completed by NOESY, COSY, TOCSY and {^{13}C-^{1}H}-HSQC, as described previously[47-50]. Spectral comparison between the natural and ^{Br}G7-substitutted sequences facilitated the identification of peaks belonging to the major form of the natural sequence. ROESY spectra were used to confirm the dipolar interaction nature of peaks observed in NOESY. Inter-proton distances were measured using NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure Calculations

Inter-proton distances for ^{Br}G7-Form 3 were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 150, 200 and 300 ms). Structure computations were performed using the X-PLOR program[57] in three general steps: (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, as well as planarity restraints were imposed during structure calculations. The detailed procedure is described in the Supplementary Text. Structures were displayed using PyMOL.[58]

The structure of natural Form 3 was calculated from the structure of ^{Br}G7-Form 3. The bromine atom of a representative structure of ^{Br}G7-Form 3 was replaced by a hydrogen atom and the molecule was then subjected to refinement with inter-proton distances deduced from NOESY experiments performed in D_2O (mixing times, 100 and 300 ms).

Data Deposition

The coordinates for the d[$(GGGTTA)_3$GGGT] G-quadruplex and for the ^{Br}G-substitute d [GGGTTA(^{Br}G)GGTTAGGGTTAGGGT] G-quadruplex have been deposited in the Protein Data Bank (accession code XXX and YYY).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This research was supported by Singapore Ministry of Education grant ARC30/07 and Nanyang Technological University (NTU) start-up grants (SUG5/06, RG138/06 and M58110032) to A.T.P, Singapore Ministry of Education AcRF Tier 2 grant (T206B3210RS) and the NTU University Research Committee grant (RG65/06) to Y.M., and US National Institutes of Health grant GM34504 to D.J.P. We thank Dr. Vitaly Kuryavyi (Memorial Sloan-Kettering Cancer Center) for helpful discussions and Prof. Lars Nordenskiöld (NTU School of Biological Sciences) for allowing us to use the UV spectrophotometer of his laboratory.

REFERENCES

1. De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF, Mergny JL. Biochimie 2008;90:131–155. [PubMed: 17822826]
2. Makarov VL, Hirose Y, Langmore JP. Cell 1997;88:657–666. [PubMed: 9054505]
3. Patel DJ, Phan AT, Kuryavyi V. Nucleic Acids Res 2007;35:7429–7455. [PubMed: 17913750]
4. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Nucleic Acids Res 2006;34:5402–5015. [PubMed: 17012276]
5. Phan AT, Kuryavyi V, Patel DJ. Curr. Opin. Struct. Biol 2006;16:288–298. [PubMed: 16714104]
6. Davis JT. Angew. Chem. Int. Ed. Engl 2004;43:668–698. [PubMed: 14755695]
7. Simonsson T. Biol. Chem 2001;382:621–628. [PubMed: 11405224]
8. Gilbert DE, Feigon J. Curr. Opin. Struct. Biol 1999;9:305–314. [PubMed: 10361092]
9. Williamson JR. Annu. Rev. Biophys. Biomol. Struct 1994;23:703–730. [PubMed: 7919797]
10. Gellert MN, Lipsett MN, Davies DR. Proc. Natl. Acad. Sci. USA 1962;48:2013–2018. [PubMed: 13947099]
11. Zahler AM, Williamson JR, Cech TR, Prescott DM. Nature 1991;350:718–720. [PubMed: 2023635]
12. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. Science 1994;266:2011–2015. [PubMed: 7605428]
13. Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH. J. Med. Chem 1997;40:2113–2116. [PubMed: 9216827]
14. Mergny JL, Hélène C. Nat. Med 1998;4:1366–1367. [PubMed: 9846570]
15. Wang Y, Patel DJ. Structure 1993;1:263–282. [PubMed: 8081740]
16. Parkinson GN, Lee MPH, Neidle S. Nature 2002;417:876–880. [PubMed: 12050675]
17. Phan AT, Patel DJ. J. Am. Chem. Soc 2003;125:15021–15027. [PubMed: 14653736]
18. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. J. Am. Chem. Soc 2006;128:9963–9970. [PubMed: 16866556]
19. Phan AT, Luu KN, Patel DJ. Nucleic Acids Res 2006;34:5715–5719. [PubMed: 17040899]
20. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Nucleic Acids Res 2007;35:6517–6525. [PubMed: 17895279]
21. Xu Y, Sugiyama H. J. Am. Chem. Soc 2004;126:6274–6279. [PubMed: 15149224]
22. Xu Y, Noguchi Y, Sugiyama H. Bioorg. Med. Chem 2006;14:5584–5591. [PubMed: 16682210]
23. Matsugami A, Xu Y, Noguchi Y, Sugiyama H, Katahira M. FEBS J 2007;274:3545–3556. [PubMed: 17561958]
24. Okamoto K, Sannohe Y, Mashimo T, Sugiyama H, Terazima M. Bioorg. Med. Chem 2008;16:6873–6879. [PubMed: 18555689]
25. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Nucleic Acids Res 2006;34:2723–2735. [PubMed: 16714449]
26. Dai J, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang D. Nucleic Acids Res 2007;42:481–490.
27. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Nucleic Acids Res 2007;35:4927–4940. [PubMed: 17626043]
28. Ying L, Green JJ, Li H, Klenerman D, Balasubramanian S. Proc. Natl. Acad. Sci. USA 2003;100:14629–14634. [PubMed: 14645716]
29. Redon S, Bombard S, Elizondo-Riojas MA, Chottard JC. Nucleic Acids Res 2003;31:1605–1613. [PubMed: 12626701]

30. Ourliac-Garnier I, Elizondo-Riojas MA, Redon S, Farrell NP, Bombard S. Biochemistry 2005;44:10620–10634. [PubMed: 16060671]
31. He Y, Neumann RD, Panyutin IG. Nucleic Acids Res 2004;32:5359–5367. [PubMed: 15475390]
32. Gaynutdinov TI, Neumann RD, Panyutin IG. Nucleic Acids Res 2008;36:4079–4087. [PubMed: 18535007]
33. D'Isa G, Galeone A, Oliviero G, Piccialli G, Varra M, Mayol L. Bioorg. Med. Chem. Lett 2004;14:5417–5421. [PubMed: 15454237]
34. Hazel P, Huppert J, Balasubramanian S, Neidle S. J. Am. Chem. Soc 2004;126:16405–16415. [PubMed: 15600342]
35. Risitano A, Fox KR. Bioorg. Med. Chem. Lett 2005;15:2047–2050. [PubMed: 15808465]
36. Rezler EM, Seenisamy J, Bashyam S, Kim MY, White E, Wilson WD, Hurley LH. J. Am. Chem. Soc 2005;127:9439–9447. [PubMed: 15984871]
37. Rujan IN, Meleney JC, Bolton PH. Nucleic Acids Res 2005;33:2022–2031. [PubMed: 15817566]
38. Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A. J. Phys. Chem. B 2005;109:3594–3605. [PubMed: 16851398]
39. Qi J, Shafer RH. Nucleic Acids Res 2005;33:3185–3192. [PubMed: 15933211]
40. Vorlickova M, Chladkova J, Kejnovska I, Fialova M, Kypr J. Nucleic Acids Res 2005;33:5851–5860. [PubMed: 16221978]
41. Li J, Correia JJ, Wang L, Trent JO, Chaires JB. Nucleic Acids Res 2005;33:4649–4659. [PubMed: 16106044]
42. Lee JY, Okumus B, Kim DS, Ha T. Proc. Natl. Acad. Sci. USA 2005;102:18938–18943. [PubMed: 16365301]
43. Yu HQ, Miyoshi D, Sugimoto N. J. Am. Chem. Soc 2006;128:15461–15468. [PubMed: 17132013]
44. Xue Y, Kan ZY, Wang Q, Yao Y, Liu J, Hao YH, Tan Z. J. Am. Chem. Soc 2007;129:11185–11191. [PubMed: 17705383]
45. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V. Nucleic Acids Res 1992;20:4061–4067. [PubMed: 1508691]
46. Mergny JL, Phan AT, Lacroix L. FEBS Lett 1998;435:74–78. [PubMed: 9755862]
47. Phan AT, Patel DJ. J. Am. Chem. Soc 2002;124:1160–1161. [PubMed: 11841271]
48. Phan AT. J. Biomol. NMR 2000;16:175–178. [PubMed: 10723997]
49. Phan AT, Guéron M, Leroy JL. Methods Enzymol 2001;338:341–371. [PubMed: 11460557]
50. Phan AT, Modi YS, Patel DJ. J. Mol. Biol 2004;338:93–102. [PubMed: 15050825]
51. Kettani A, Basu G, Gorin A, Majumdar A, Skripkin E, Patel DJ. J. Mol. Biol 2000;301:129–146. [PubMed: 10926497]
52. Kuryavyi V, Kettani A, Wang W, Jones R, Patel DJ. J. Mol. Biol 2000;295:455–469. [PubMed: 10623538]
53. Parkinson GN, Ghosh R, Neidle S. Biochemistry 2007;46:2390–2397. [PubMed: 17274602]
54. Haider S, Parkinson GN, Neidle S. Biophys J 2008;95:296–311. [PubMed: 18375510]
55. Parkinson GN, Cuenca F, Neidle S. J. Mol. Biol 2008;381:1145–1156. [PubMed: 18619463]
56. Cantor CR, Warshaw MM, Shapiro H. Biopolymers 1970;9:1059–1077. [PubMed: 5449435]
57. Brünger, AT. X-PLOR: A system for X-ray crystallography and NMR. Yale University Press; New Haven, CT: 1992.
58. DeLano, WL. The PyMOL User's Manual. DeLano Scientific; Palo Alto, CA: 2002.

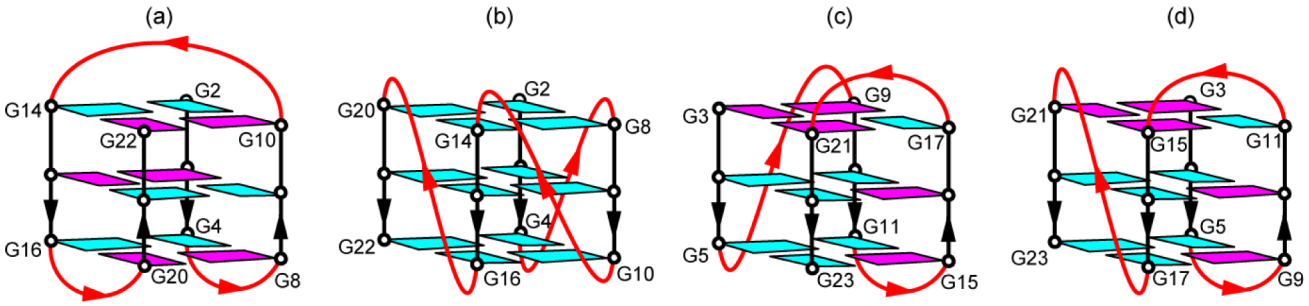

Figure 1.

Schematic structures of intramolecular G-quadruplexes formed by the human telomeric sequences: (a) basket-type form observed for d[A(GGGTTA)$_3$GGG] in Na^+ solution[15], (b) propeller-type form observed for d[A(GGGTTA)$_3$GGG] in a K^+-containing crystal[16], (c) (3 +1) Form 1 observed for d[TA(GGGTTA)$_3$GGG] in K^+ solution[18], and (d) (3+1) Form 2 observed for d[TA(GGGTTA)$_3$GGGTT] in K^+ solution[19]. *anti* guanines are colored cyan; *syn* guanines are colored magenta.

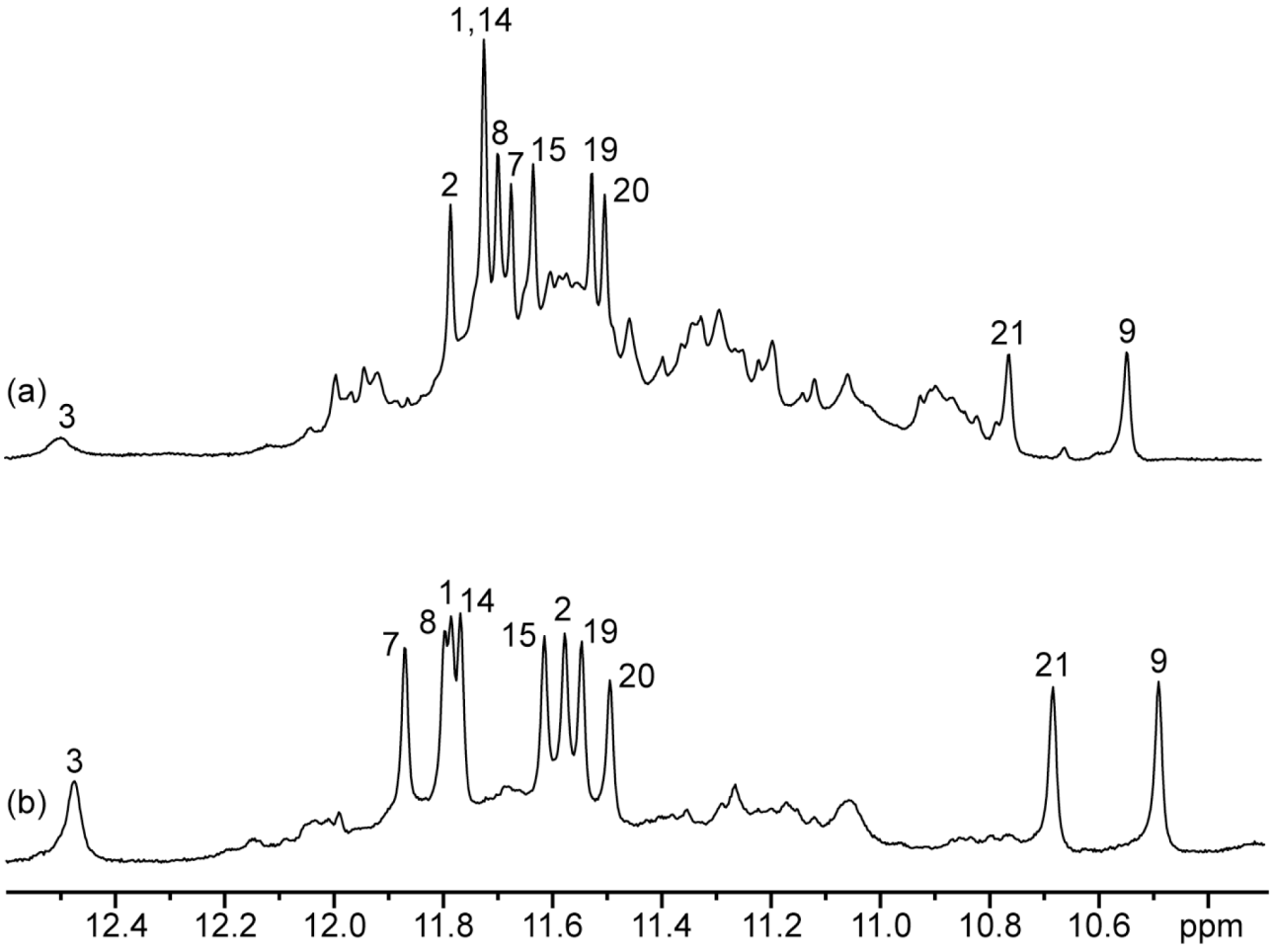

Figure 2.

Imino proton spectra of (a) the natural 22-nt human telomeric d[$(GGGTTA)_3GGGT$] sequence and (b) the modified d[GGGTTA(^{Br}G)GGTTAGGGTTAGGGT] sequence in K^+ solution with assignments for the major forms listed over the spectra.

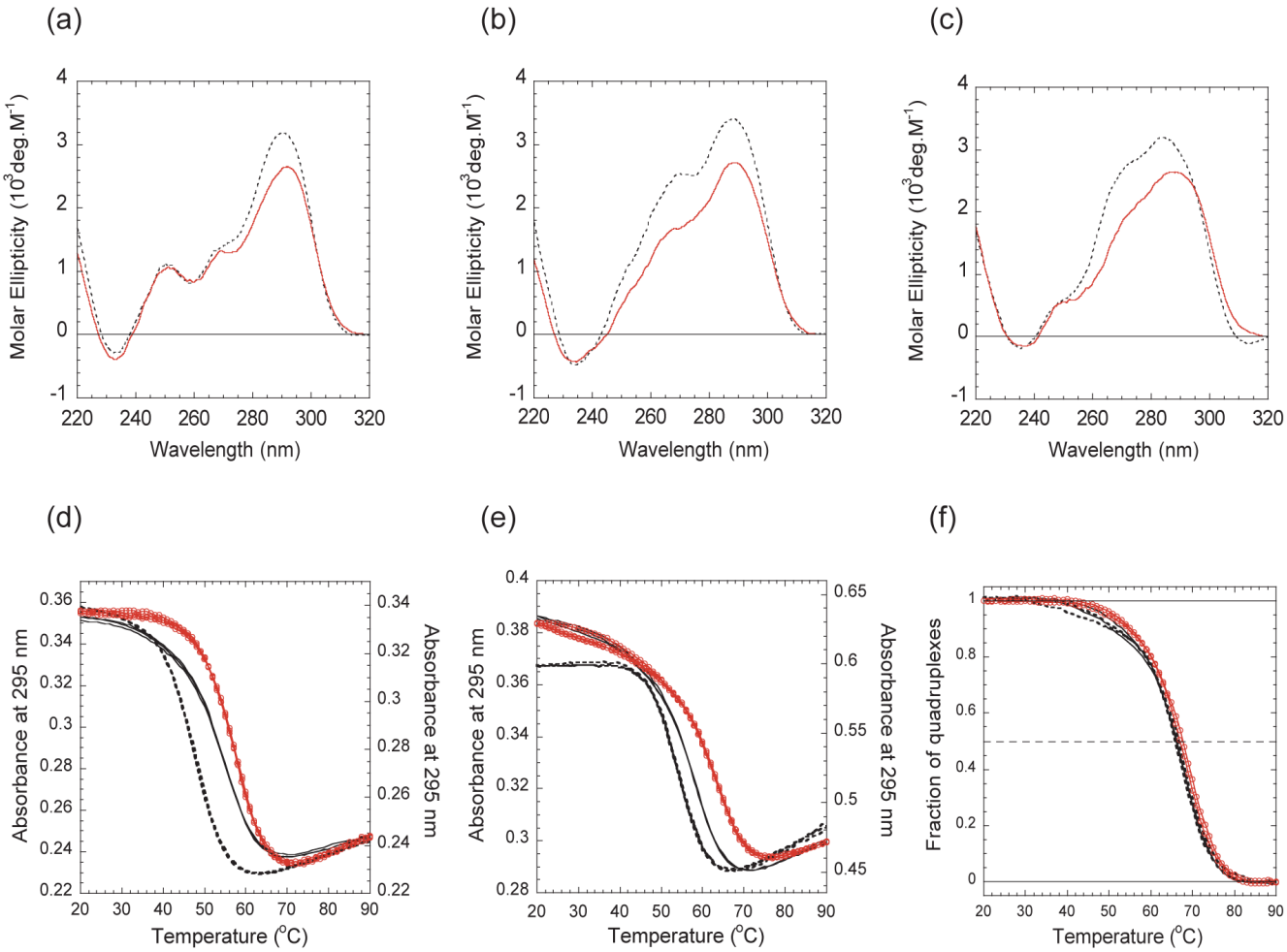

Figure 3.

(a-c) CD spectra of (a) Form 3 (this work), (b) Form 2 and (c) Form 1 intramolecular human telomeric G-quadruplexes[20] (strand concentration, 8 μM; potassium phosphate, 20 mM; KCl, 70 mM; pH 7). Dotted black and continuous red curves are for the natural and ^{Br}G-modified human telomeric sequences, respectively. (d,e) Melting curves of (d) the natural and (e) ^{Br}G-modified sequences of Form 1 (continuous line, left axis), Form 2 (dotted line, left axis) and Form 3 (red circles, right axis) monitored by UV absorption at 295 nm (DNA strand concentration, 8 μM; potassium phosphate, 20 mM; pH 7). (f) The fractions of G-quadruplexes as a function of temperature for three different DNA concentrations of ^{Br}G7-Form 3: 10 μM (continuous line), 50 μM (dotted line) and 250 μM (red circles), in a buffer containing 20 mM potassium phosphate pH7 and 20 mM KCl. In each case, both heating and cooling curves are plotted.

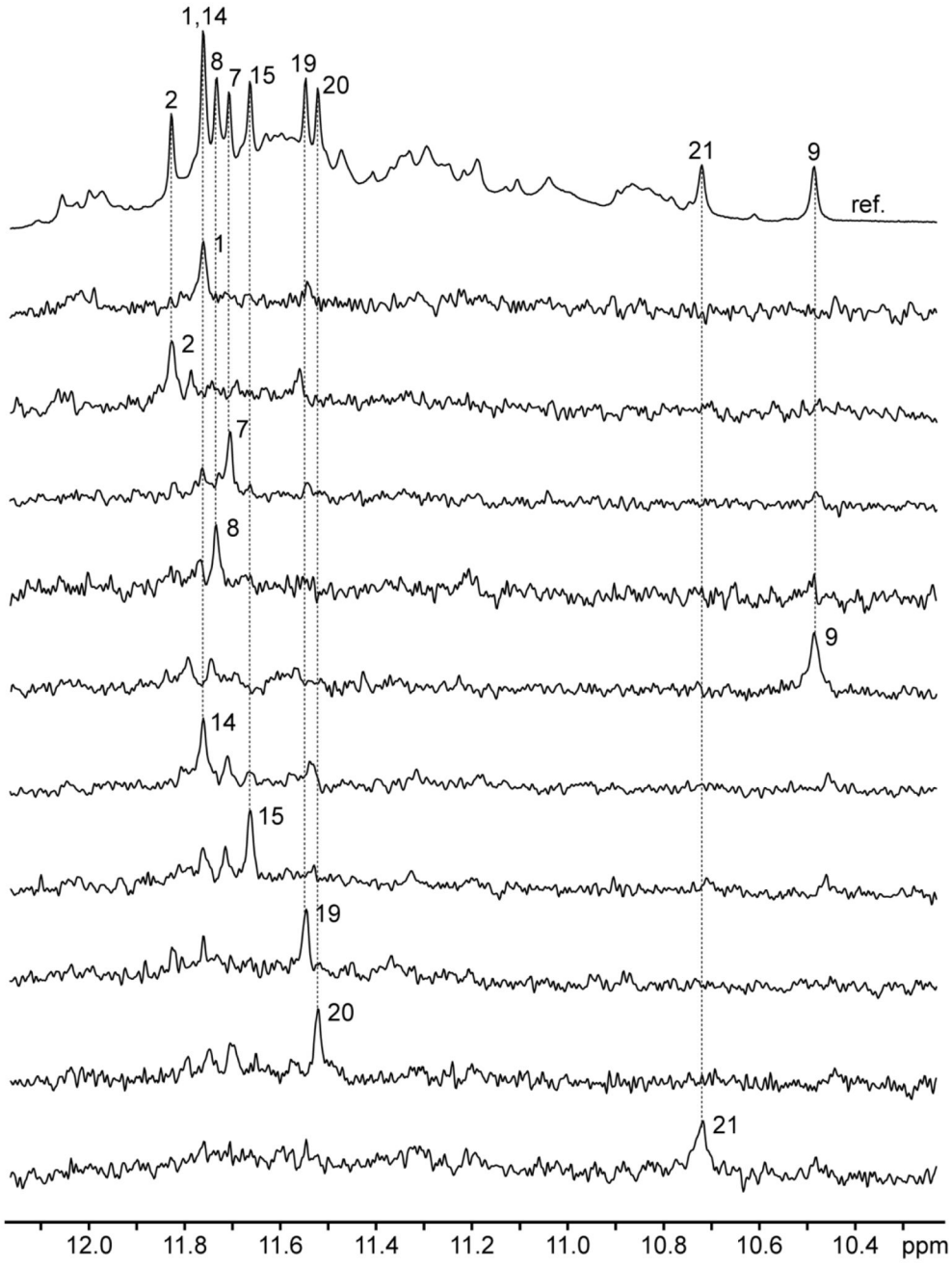

Figure 4.

Imino proton spectra and assignments of the natural human telomeric d[$(GGGTTA)_3GGGT$] sequence in K^+ solution. Imino protons were assigned in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated position. The reference spectrum (ref.) is shown at the top.

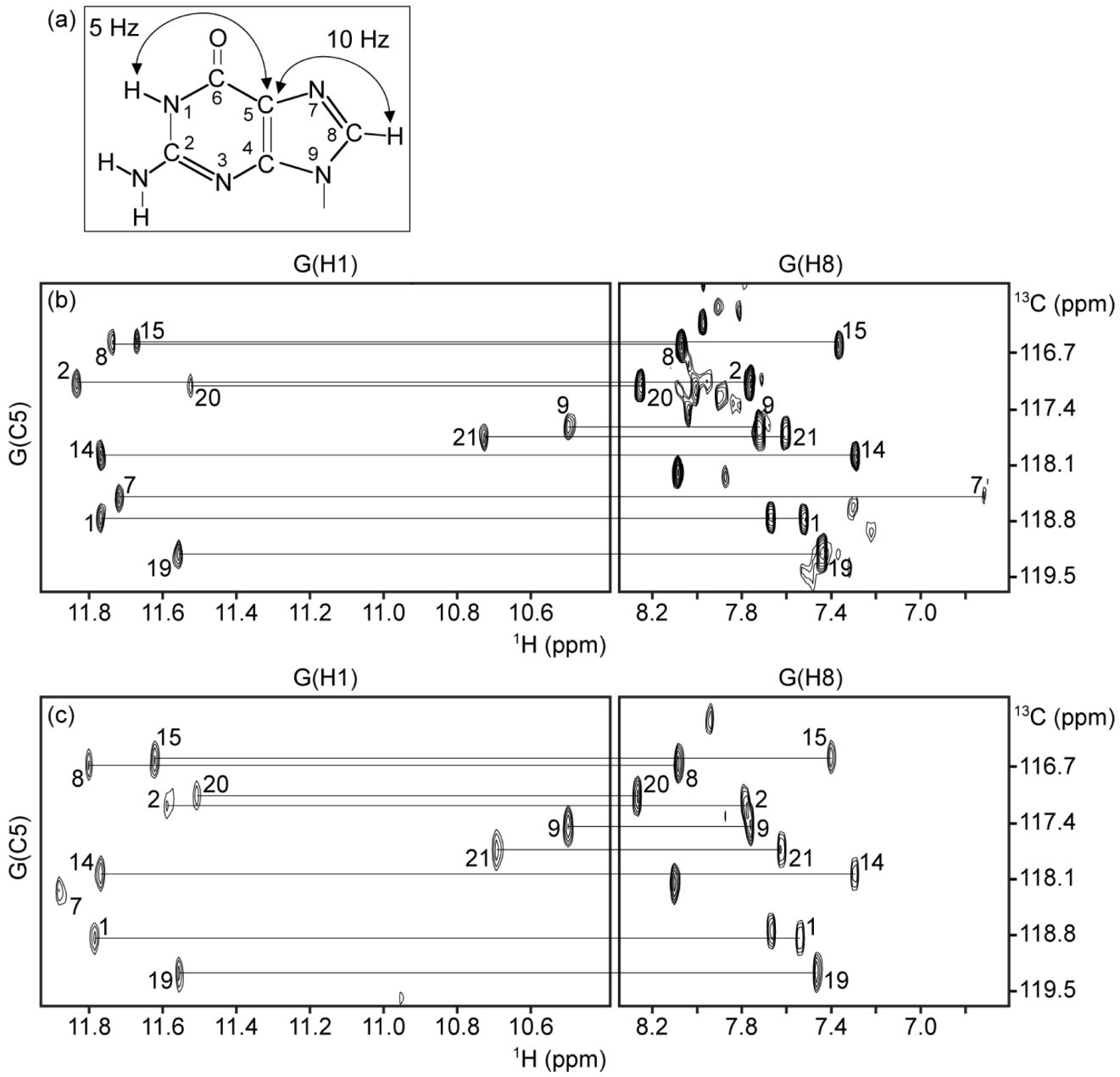

Figure 5.

H8 proton assignments. (a) Long range J-couplings in a guanine. (b, c) Through-bond correlations between guanine imino and H8 protons via ^{13}C5 at natural abundance for (b) the natural d[(GGGTTA)$_3$GGGT] sequence and (c) the modified sequence, using long range J-couplings shown in (a).

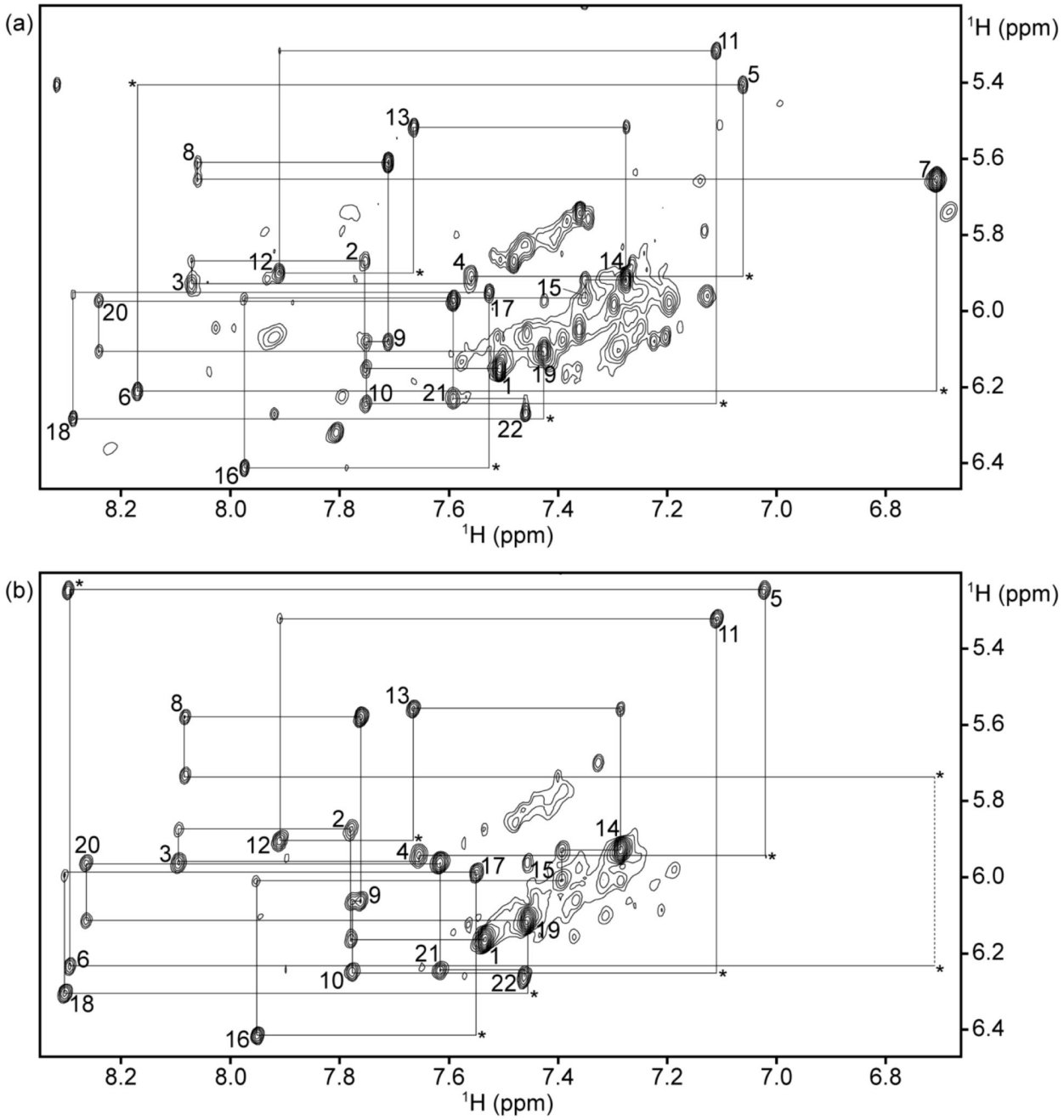

Figure 6.

NOESY spectra (mixing time, 300 ms), showing the H8/H6-H1′ connectivity of (a) the natural d[(GGGTTA)$_3$GGGT] sequence and (b) the modified sequence in K^+ solution. Intra-residue H8/H6-H1′ NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks. For the modified sequence (b), the connectivity is broken at position 7 (shown by dotted line) because the proton H8 of G7 was replaced by bromine.

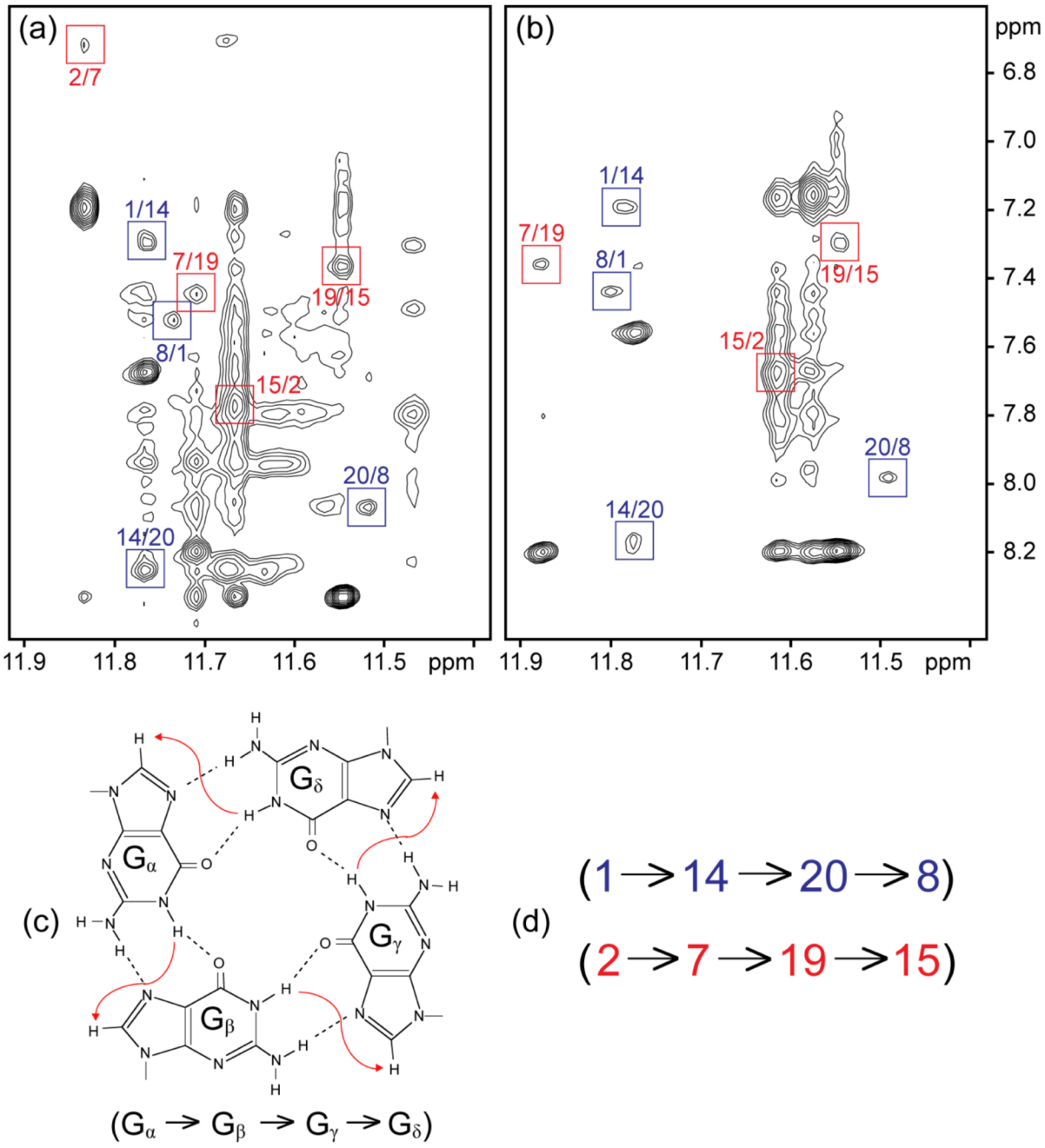

Figure 7.

Determination of G-quadruplex folding topology. (a, b) NOESY spectra (mixing time, 200 ms), showing imino-H8 connectivity of (a) the natural d[$(GGGTTA)_3GGGT$] and (b) the modified sequence. Cross-peaks that identify the two G-tetrads are framed and labeled with the residue number of imino proton in the first position and that of H8 in the second position. (c) Characteristic guanine imino-H8 NOE connectivity patterns around a G_α•G_β•G_γ•G_δ tetrad as indicated with arrows (connectivity between G_δ and G_α implied). (d) Guanine imino-H8 NOE connectivities observed for G1•G14•G20•G8 and G2•G7•G19•G15 tetrads.

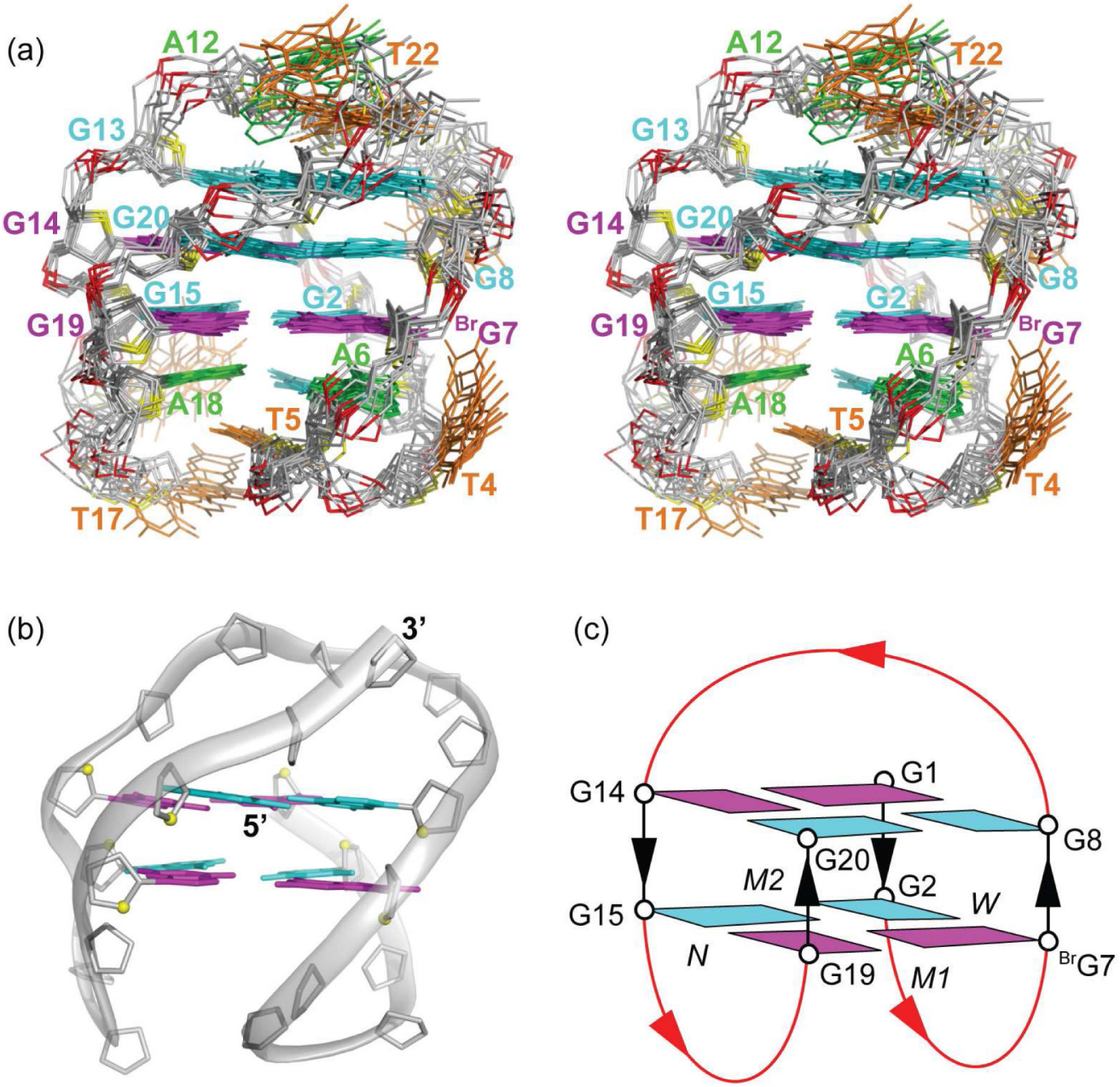

Figure 8.

Structure of the human telomeric G-quadruplex Form 3 in K^+ solution. (a) Stereo view of eight superpositioned refined structures of ^{Br}G7-Form 3. (b) Ribbon and (c) schematic view of a representative structure of natural Form 3. *anti* and *syn* guanines are colored cyan and magenta, respectively; adenines are colored green; thymines, orange; backbone, gray; O4′ atoms, yellow; phosphorus atoms, red. *W*, *M1*, *M2* and *N* represent wide, medium 1, medium 2 and narrow grooves, respectively.

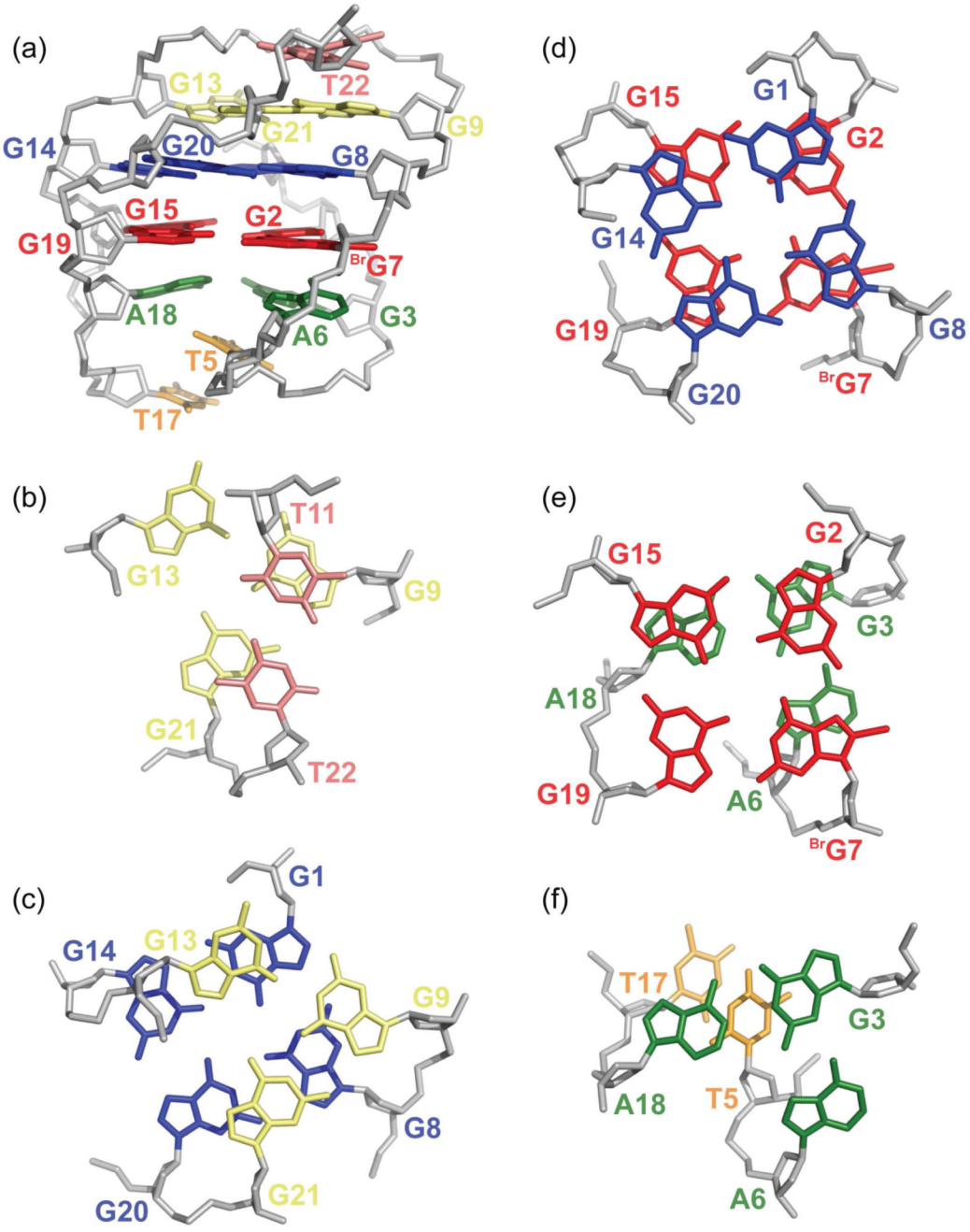

Figure 9.

Stacking between different layers in the human telomeric G-quadruplex Form 3 in K^+ solution. (a) View from the side looking into the medium groove 1. (b) Top T•T base pair and top triple. (c) Top triple and top tetrad. (d) The two tetrads. (e) Bottom tetrad and bottom triple. (f) Bottom triple and bottom two T bases.

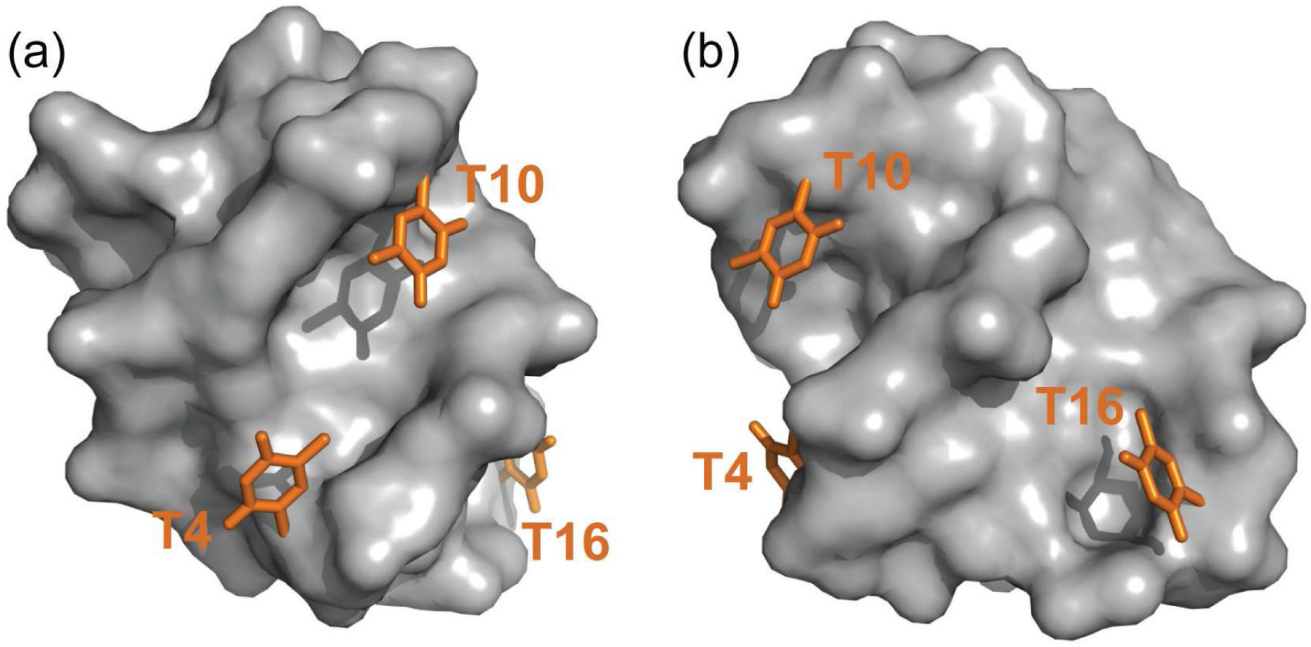

Figure 10.
Folding of the base of T4, T10, and T16 into the hydrophobic groove of the G-quadruplex, viewing from the side looking into (a) the wide groove (*W*) and (b) the medium groove 2 (*M2*).

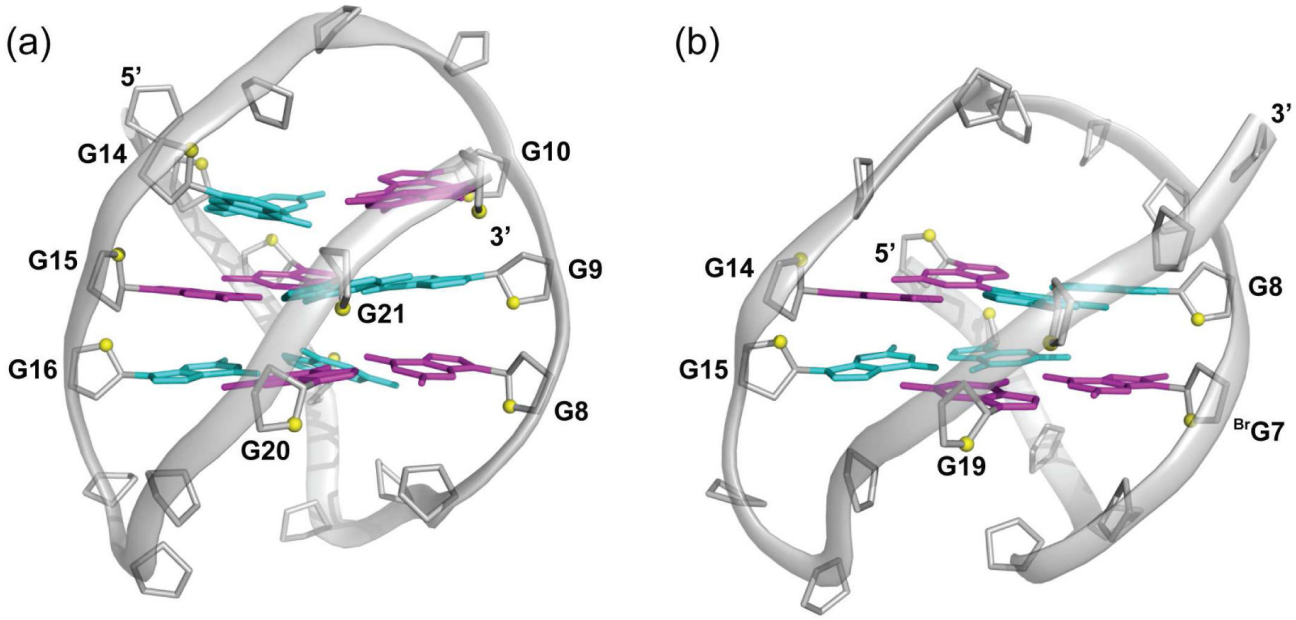

Figure 11.
Comparison of (a) the three-tetrad basket-type human telomeric G-quadruplex observed in Na^+ solution[15] and (b) the Form 3 two-tetrad basket-type human telomeric G-quadruplex observed in K^+ solution (this work).

Table 1

Natural and modified human telomeric DNA sequences [(a)].

Name	Sequence									
Natural Form 3		GGG	TTA	GGG	TTA	GGG	TTA	GGG	T
^{Br}G7-Form 3		GGG	TTA	^{Br}GGG	TTA	GGG	TTA	GGG	T
[S13]-^{Br}G7-Form 3		GGG	TTA	^{Br}GGG	TTA	SGGG	TTA	GGG	T
[T13]-Form 3		GGG	TTA	GGG	TTA	TGG	TTA	GGG	T
[T3]-Form 3		GGT	TTA	GGG	TTA	GGG	TTA	GGG	T
[C18]-Form 3		GGG	TTA	GGG	TTA	GGG	TTC	GGG	T
Natural Form 1 [(b)]	TA	GGG	TTA	GGG	TTA	GGG	TTA	GGG	
^{Br}G16-Form 1 [(b)]	TA	GGG	TTA	GGG	TTA	G^{Br}GG	TTA	GGG	
Natural Form 2 [(b)]	TA	GGG	TTA	GGG	TTA	GGG	TTA	GGG	TT
^{Br}G15-Form 2 [(b)]	TA	GGG	TTA	GGG	TTA	^{Br}GGG	TTA	GGG	TT

[(a)] Natural sequences are shown in red; modified residues are in boldface

[(b)] These sequences were studied in ref. [20]. ^{Br}G is 8-bromoguanine; ^{S}G is 6-thioguanine.

Table 2

Melting temperatures (°C) of G-quadruplexes adopted by different human telomeric sequences in 20 mM K^+ solution, as measured by 295-nm UV absorbance.

	Natural	^{Br}G-substituted
Form 1	53.6	57.5
Form 2	47.2	54.0
Form 3	57.0	62.8

Table 3

Statistics of the computed structures of the modified d[GGGTTA(^{Br}G)GGTTAGGGTTAGGGT] sequence.

A. NMR restraints		
Distance restraints	D_2O	H_2O
Intra-residue distance restraints	269	3
Sequential distance restraints (i, i + 1)	199	20
Long-range distance restraints (i, ≥ i + 2)	44	45
Other restraints		
Hydrogen bonding restraints	42	
Torsion angle restraints	20	
Intensity restraints		
Non-exchangeable Protons (each of four mixing times)	140	
B. Structure statistics for 10 molecules following intensity refinement.		
NOE violations		
Number [(a)] (>0.2Å)	0.400 ± 0.490	
Maximum violation (Å)	0.193 ± 0.029	
RMSD of violations (Å)	0.018 ± 0.002	
Deviations from the ideal covalent geometry		
Bond lengths (Å)	0.005 ± 0.000	
Bond angles (deg)	0.752 ± 0.028	
Impropers (deg)	0.433 ± 0.015	
NMR R-factor ($R_{1/6}$)	0.016 ± 0.001	
Pairwise all heavy atom RMSD values (Å)		
All heavy atoms except T4, T5, T10, T11, A12, T16, T17, T22	1.06±0.13	
All heavy atoms	1.54±0.21	

[(a)] the total number of violations divided by the number of structures.