J|A|C|S ARTICLES

Published on Web 10/29/2009

Giardia Telomeric Sequence d(TAGGG)$_4$ Forms Two Intramolecular G-Quadruplexes in K$^+$ Solution: Effect of Loop Length and Sequence on the Folding Topology

Lanying Hu,[†] Kah Wai Lim,[†] Serge Bouaziz,[‡] and Anh Tuân Phan*,[†]

Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371 and Unité de Pharmacologie Chimique et Génétique, Université Paris Descartes−INSERM U640−CNRS UMR 8151, 4 avenue de l'Observatoire, 75006 Paris, France

Received July 14, 2009; E-mail: phantuan@ntu.edu.sg

Abstract: Recently, it has been shown that in K$^+$ solution the human telomeric sequence d[TAGGG(TTAGGG)$_3$] forms a (3 + 1) intramolecular G-quadruplex, while the *Bombyx mori* telomeric sequence d[TAGG(TTAGG)$_3$], which differs from the human counterpart only by one G deletion in each repeat, forms a chair-type intramolecular G-quadruplex, indicating an effect of G-tract length on the folding topology of G-quadruplexes. To explore the effect of loop length and sequence on the folding topology of G-quadruplexes, here we examine the structure of the four-repeat *Giardia* telomeric sequence d[TAGGG(TAGGG)$_3$], which differs from the human counterpart only by one T deletion within the non-G linker in each repeat. We show by NMR that this sequence forms two different intramolecular G-quadruplexes in K$^+$ solution. The first one is a novel basket-type antiparallel-stranded G-quadruplex containing two G-tetrads, a G•(A-G) triad, and two A•T base pairs; the three loops are consecutively edgewise−diagonal−edgewise. The second one is a propeller-type parallel-stranded G-quadruplex involving three G-tetrads; the three loops are all double-chain-reversal. Recurrence of several structural elements in the observed structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies, for which different elements could be extracted from one G-quadruplex and inserted into another.

Introduction

Guanine-rich DNA and RNA sequences can form G-quadruplex structures[1−8] through stacking of multiple G-tetrads.[9] There has been increasing interest in these structures, as G-rich sequences have widespread prevalence in the human genome and have been identified in biologically important regions such as telomeres, immunoglobulin switch regions, promoter regions of oncogenes, and recombination hot spots.

G-quadruplexes can adopt a range of folding topologies regarding strand directionalities, loop connectivities, and glycosidic conformations of guanines around G-tetrads.[1–8] There are four possibilities for relative strand orientations in the G-tetrad core (all four oriented in one direction; three in one direction and the fourth in the opposite direction; two adjacent in one direction and the two remaining in the opposite direction; two across a diagonal in one direction and the two remaining in the opposite direction) and three main possibilities for loops (edgewise; diagonal; double-chain-reversal).

It is important to be able to predict the G-quadruplex structure based on sequence, as well as to design a sequence that will adopt a desired G-quadruplex topology. Systematic studies on the effect of loop length/sequence and G-tract length on the topology and stability of G-quadruplexes have been conducted by several groups.[10−13] However, structural interpretations of G-quadruplexes in these works were limited by using mainly CD spectra.

As for high-resolution structural studies, five different intramolecular G-quadruplexes have been solved for DNA sequences containing human telomeric TTAGGG repeats under

[†] Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University.
[‡] Unité de Pharmacologie Chimique et Génétique, Université Paris Descartes.

(1) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429–7455.
(2) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; Royal Society of Chemistry: Cambridge, UK, 2006; pp 81−99.
(3) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5015.
(4) Phan, A. T.; Kuryavyi, V.; Patel, D. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 288–298.
(5) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668–698.
(6) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621–628.
(7) Gilbert, D. E.; Feigon, J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 305–314.
(8) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730.
(9) Gellert, M. N.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013–2018.
(10) Rachwal, P. A.; Brown, T.; Fox, K. R. *Biochemistry* **2007**, *46*, 3036–3044.
(11) Rachwal, P. A.; Findlow, I. S.; Werner, J. M.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2007**, *35*, 4214–4222.
(12) Bugaut, A; Balasubramanian, S. *Biochemistry* **2008**, *47*, 689–697.
(13) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J. L. *Biochimie* **2008**, *90*, 686–696.

 10.1021/ja905611c CCC: $40.75

different experimental conditions.[14−24] In particular, the four-repeat human telomeric sequence d[TAGGG(TTAGGG)$_3$] has been shown to form a (3 + 1) intramolecular G-quadruplex structure in K^+ solution, in which the G-tetrad core contains three strands oriented in one direction and the fourth in the opposite direction.[18] Also in K^+ solution, the *Bombyx mori* telomeric sequence d[TAGG(TTAGG)$_3$], which differs from the human counterpart only by one G deletion in each repeat, forms a chair-type intramolecular G-quadruplex, indicating an effect of G-tract length on the folding topology of G-quadruplexes.[25] A small variation within the non-G linkers might result in a dramatic change in G-quadruplex topology: the four-repeat variant human telomeric sequence d[AGGG(CTAGGG)$_3$] (variation is underlined) forms a chair-type intramolecular G-quadruplex involving two G-tetrads and a G•C•G•C tetrad in K^+ solution.[26]

To further explore the effect of loop length and sequence on the folding topology of G-quadruplexes, here we examine the structure of the four-repeat *Giardia* telomeric sequence d[TAGGG-(TAGGG)$_3$], i.e., d(TAGGG)$_4$, in which the non-G linkers are shorter than the human counterpart only by one T deletion in each repeat. The TAGGG repeats have also been detected as a potential variant that are interspersed within the human telomeres,[27] among the canonical TTAGGG repeats. We found that this sequence forms two different intramolecular G-quadruplexes in K^+ solution. The first one is a novel basket-type antiparallel-stranded G-quadruplex containing two G-tetrads, a G•(A-G) triad, and two A•T base pairs; the second one is a propeller-type parallel-stranded G-quadruplex involving three G-tetrads. Recurrence of several structural elements in the observed structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies.

Methods

DNA Sample Preparation. Unlabeled and site-specific labeled DNA oligonucleotides (Table 1; Table S1 of the Supporting Information) were chemically prepared using an ABI 394 DNA/RNA synthesizer, as previously described.[25] DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.[28]

Table 1. Natural and Modified *Giardia* Telomeric DNA Sequences[a]

name	sequence							
natural sequence	TA	GGG	TA	GGG	TA	GGG	TA	GGG
I18-Form 1	TA	GGG	TA	GGG	TA	GGG	TA	**I**GG
Δ*A12-Form 2*	TA	GGG	TA	GGG	T-	GGG	TA	GGG

[a] Modifications from the natural sequence are shown in boldface.

Gel Electrophoresis. The molecular size of the structures formed by DNA oligonucleotides was probed by nondenaturing polyacrylamide gel electrophoresis (PAGE). The electrophoresis experiment was performed with 10 × 7 cm native gel containing 20% acrylamide (Acrylamide:Bis-acrylamide = 37.5:1) supplemented with 10 mM KCl in TBE buffer pH 8.3 at 120 mV, in 140 minutes. Gel was viewed by UV shadowing.

Circular Dichroism (CD). CD spectra were recorded on a Jasco J-810 spectropolarimeter using a 1-cm path-length quartz cuvette as previously described.[25] The buffer contained 10 mM KCl, 40 mM LiCl and 20 mM lithium phosphate (pH 7). For each spectrum, an average of 3 scans was taken. DNA concentration was typically around 5 μM.

UV Spectroscopy. The UV melting experiments were performed on a Varian CARY-300 spectrophotometer by monitoring the UV absorption at 295 nm as a function of temperature.[29] The concentration of DNA varied from 3 to 300 μM. Samples were covered with approximately 100 μL of mineral oil to prevent evaporation. They were equilibrated at 90 °C for 10 min, then cooled to 20 °C and heated to 90 °C twice consecutively at a rate of 0.15 °C per minute. Data were collected every 1 °C during both cooling and heating processes.

NMR Spectroscopy. Samples for NMR study were dialyzed successively against ∼50 mM KCl solution and against water. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5−2.0 mM; the solutions contained 70 mM KCl and 20 mM potassium phosphate (pH 7). NMR experiments were performed on 600 and 700 MHz Bruker spectrometers at 25 °C, unless otherwise specified. Resonances for guanine residues were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling,[30] site-specific ^{2}H labeling,[31] and JRHMBC through-bond correlations at natural abundance.[32,33] Resonances for thymine residues were assigned following systematic T-to-U substitutions. Spectral assignments were completed by NOESY, COSY, TOCSY, and {^{13}C−^{1}H}-HSQC, as described previously.[33] Interproton distances were deduced from NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure Calculation. Interproton distances for the *I18-Form 1* G-quadruplex were deduced from NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 150, 200, and 350 ms). Structure computations were performed using the XPLOR-NIH program[34] in three general steps essentially as previously described:[24] (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement, and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, interproton distance restraints, dihedral restraints, and planarity

(14) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.
(15) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.
(16) Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591.
(17) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
(18) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
(19) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–5719.
(20) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
(21) Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. *FEBS J.* **2007**, *274*, 3545–3556.
(22) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
(23) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
(24) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
(25) Amrane, S.; Ang, R. W.; Tan, Z. M.; Li, C.; Lim, J. K.; Lim, J. M.; Lim, K. W.; Phan, A. T. *Nucleic Acids Res.* **2009**, *37*, 931–938.
(26) Lim, K. W.; Alberti, P.; Guédin, A.; Lacroix, L.; Riou, J. F.; Royle, N. J.; Mergny, J. L.; Phan, A. T. *Nucleic Acids Res.* **2009**, *37*, 6239–6248.
(27) Allshire, R. C.; Dempster, M.; Hastie, N. D. *Nucleic Acids Res.* **1989**, *17*, 4611–4627.
(28) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059–1077.
(29) Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74–78.
(30) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
(31) Huang, X.; Yu, P.; LeProust, E.; Gao, X. *Nucleic Acids Res.* **1997**, *25*, 4758–4763.
(32) Phan, A. T. *J. Biomol. NMR* **2000**, *16*, 175–178.
(33) Phan, A. T.; Guéron, M.; Leroy, J. L. *Methods Enzymol.* **2001**, *338*, 341–371.
(34) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.

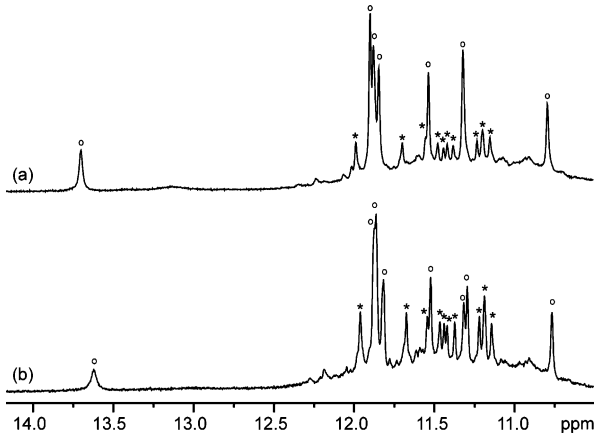

Figure 1. Imino proton spectra of the 20-nt *Giardia* telomeric d(TAGGG)$_4$ sequence in K^+ solution at (a) 25 °C and (b) 35 °C. Two sets of peaks, corresponding to two different conformations of G-quadruplexes, are labeled with circles and asterisks, respectively.

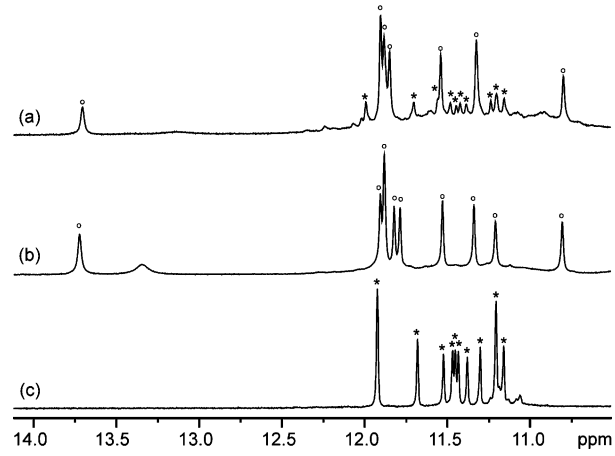

Figure 2. Imino proton spectra of natural and modified *Giardia* telomeric sequences in K^+ solution. (a) Natural d(TAGGG)$_4$ sequence; (b) *I18-Form 1*; and (c) Δ*A12-Form 2*. Peaks from Form 1 and Form 2 are labeled with circles and asterisks, respectively.

restraints were imposed during structure calculations. Structures were displayed using the PyMOL program.[35]

Data Deposition. The coordinates for the *I18-Form 1* G-quadruplex have been deposited in the Protein Data Bank (accession code 2KOW).

Results

Four-Repeat *Giardia* Telomeric Sequence d(TAGGG)$_4$ Forms Two G-Quadruplex Structures in K^+ Solution. Imino proton spectra of the four-repeat *Giardia* telomeric sequence d(TAGGG)$_4$ in K^+ solution (Figure 1; Figure S1 of the Supporting Information) show two sets of peaks (differentiated by circles and asterisks, respectively) at 10.8−12.0 ppm, which could be distinguished by their respective intensities at different temperatures, indicating the formation of two different G-quadruplexes. At 25 °C (Figure 1a), the major conformation, designated Form 1, represented about 80% of the population (peaks with circles), while the minor conformation, designated Form 2, represented about 20% of the population (peaks with asterisks). When the temperature increased, the relative population of Form 2 with respect to Form 1 increased (Figure 1b). At temperatures higher than 40 °C, Form 2 was favored over Form 1 (Figure S1 of the Supporting Information). Note that Form 1 was also characterized by an imino proton at 13.6 ppm.

Favoring a Single Conformation by Sequence Modifications. We identified small sequence modifications (Table 1) that favored a single G-quadruplex conformation, thereby allowing detailed structural analysis of both conformations: substitution of G18 by an inosine favored Form 1 (Figure 2b), whereas the deletion of A12 favored Form 2 (Figure 2c). The rationale of these modifications will be discussed in view of the structures described below. Comparison of the NMR spectra of these modified sequences (designated *I18-Form 1* and Δ*A12-Form 2*, respectively) with those of the natural sequence indicated that the modified sequences adopted the same G-quadruplex folds as Form 1 and Form 2 of the natural sequence, respectively (Figure 2 and data not shown).

Monitoring Two G-Quadruplexes by Circular Dichroism. At 25 °C, the CD profile of *I18-Form 1* was very different from that of Δ*A12-Form 2* (Figure 3). The former showed positive peaks at 245 and 295 nm and a negative peak at 265 nm, typical

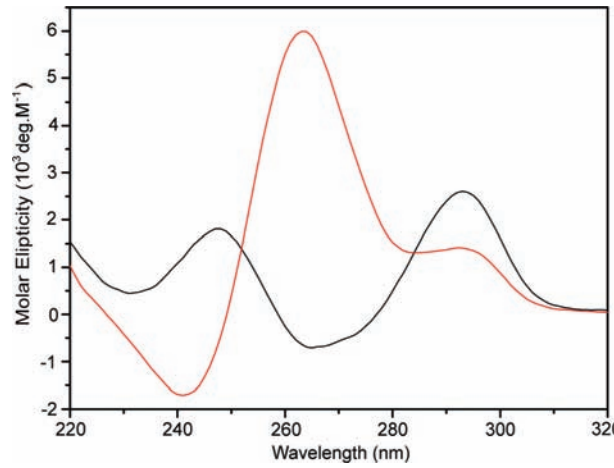

Figure 3. Normalized CD spectra of *I18-Form 1* (black line) and Δ*A12-Form 2* (red line) at 25 °C. The former displays the CD signature of antiparallel G-quadruplexes, whereas the latter shows the CD signature of parallel G-quadruplexes.

of an antiparallel-stranded G-quadruplex, while the latter showed a positive peak at 260 nm, characteristic of a parallel-stranded G-quadruplex.[36] Note that the normalized intensity of the 260-nm peak of Δ*A12-Form 2* (red curve, putative parallel G-quadruplex) is roughly three times that of the 295-nm peak of *I18-Form 1* (black curve, putative antiparallel G-quadruplex).

CD spectra of the natural sequence d(TAGGG)$_4$ recorded at various temperatures (Figure S2 of the Supporting Information) were consistent with this sequence forming two G-quadruplex structures (as monitored by peaks at 295 and 260 nm, respectively) and that the parallel-stranded form was favored at high temperatures as observed by NMR (see above). We remark that these CD data show that an interpretation of G-quadruplex topology based only on a single CD spectrum might be misleading due to the coexistence of multiple G-quadruplex conformations. A combination of different experimental conditions (temperature, pH, etc.) and sequence modifications can be used to vary the relative populations of different forms, aiding structural interpretations.

(35) DeLano, W. L. *The PyMOL User's Manual*; DeLano Scientific: Palo Alto, CA, 2002.

(36) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.* **1992**, *20*, 4061–4067.

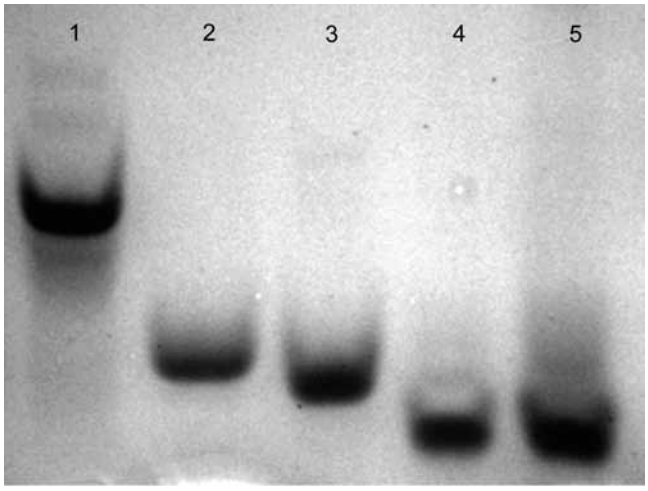

Figure 4. Nondenaturing PAGE analysis of the *Giardia* telomeric sequences. Migration markers are provided on the left. Lane 1: a migration marker of an interlocked dimeric G-quadruplex;[37] Lane 2: a migration marker of a monomeric propeller-type parallel-stranded G-quadruplex (unpublished); Lane 3: Δ*A12-Form 2*; Lane 4: *I18-Form 1*; Lane 5: natural *Giardia* telomeric sequence $d(TAGGG)_4$.

Formation of Form 1 and Form 2 G-quadruplexes by the natural sequence $d(TAGGG)_4$ and their interconversion could be monitored by CD (Figure S3 of the Supporting Information). After a quick sample cooling from 90 to 10 °C, formation of the two conformations was observed (as monitored by peaks at 295 and 260 nm, respectively) at similar rates (in the order of minutes), which was followed by their interconversion toward the equilibrium with Form 1 being the major conformation.

Stoichiometry of Two *Giardia* Telomeric G-Quadruplexes. The molecular size of Form 1 and Form 2 G-quadruplexes formed by the natural and modified *Giardia* telomeric sequences (Table 1) was probed by native polyacrylamide gel electrophoresis (PAGE) (Figure 4). A single major band was observed for each sequence. The migration rate of the natural sequence, *I18-Form 1* and Δ*A12-Form 2* was comparable with that of a reference monomeric propeller-type parallel-stranded G-quadruplex containing three G-tetrads (unpublished data) and significantly faster than that of a dimeric interlocked G-quadruplex containing totally six G-tetrads,[37] arguing for an intramolecular structure for both Form 1 and Form 2.

Melting experiments were conducted for *I18-Form 1* and Δ*A12-Form 2* by monitoring the UV absorbance at 295 nm. All transitions were reversible, indicating that the denaturation curves corresponded to a true equilibrium process (Figure S4 of the Supporting Information). No significant difference in melting temperature (1 °C or less for *I18-Form 1*; 4 °C or less for Δ*A12-Form 2*) was observed upon 100-fold increase in concentration (from 3 to 300 μM), consistent with intramolecular G-quadruplex formation. A slightly larger concentration-dependent behavior observed for Form 2 could be explained by the stacking of two G-quadruplex blocks (see below) in the presence of high DNA and/or K^+ concentration. Such a high-order structure was also observed in a gel electrophoresis experiment (Figure S5 of the Supporting Information).

NMR Spectral Assignments. Rigorous approaches[30–33] were used for NMR spectral assignments. Guanine imino and H8 protons of *I18-Form 1* and Δ*A12-Form 2* sequences were unambiguously assigned (Figures S6−S10 and Table S1 of the Supporting Information) using site-specific low-enrichment ^{15}N labeling[30] (Figures S6 and S8), site-specific ^2H labeling[31] (Figure S9), and JRHMBC through-bond correlations between imino and H8 protons via $^{13}C5$ at natural abundance[32,33] (Figures S7 and S10 of the Supporting Information). Resonances for thymine residues were unambiguously assigned by T-to-U substitutions[33] (Table S1 of the Supporting Information). NMR spectral assignments were completed with other through-bond correlation experiments (COSY, TOCSY, and $\{^{13}C-^1H\}$-HSQC) (data not shown) and through-space correlation NOESY experiments.[33]

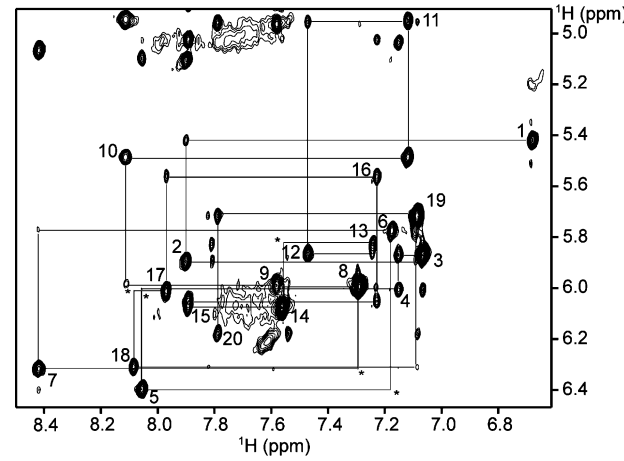

Figure 5. NOESY spectrum (mixing time, 350 ms) showing the H8/H6−H1′ connectivity of *I18-Form 1* in K^+ solution. Intraresidue H8/H6−H1′ NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks.

Structure of Form 1: Novel Basket-Type G-Quadruplex with Two G-Tetrads and a G•(A-G) Triad. With the help of the unambiguous assignments (described above), the classical H8/H6−H1′ NOE sequential connectivity of *I18-Form 1* could be traced (Figure 5). The intensity of intraresidue H8−H1′ NOE cross-peaks (Figure 5 and Figure S11 of the Supporting Information) indicated *syn* glycosidic conformation for G3, G8, G14, and G19, and *anti* conformation for other residues. Analysis of imino-H8 connectivity patterns revealed the formation of a novel intramolecular basket-type G-quadruplex with two G-tetrad layers, G3•G9•G20•G14 and G4•G15•G19•G8 (Figure 6). The structure of the *I18-Form 1* G-quadruplex in K^+ solution (Figure 7) was calculated on the basis of NMR restraints (Table 2). Two edgewise loops, formed by the G5-T6-A7 and T16-A17-I18 segments, bridge across a narrow and a wide groove, respectively, at the bottom of the G-tetrad core. The diagonal loop, formed by the G10-T11-A12-G13 segment, connects two opposite corners of the top G-tetrad. In this diagonal loop, formation of the G10•(A12-G13) triad (Figure 8c) was supported by the observation of the imino proton of G10 at 10.8 ppm and the NOEs between A12(H8) and G13(H8). Formation of the Hoogsteen A2•T11 (Figure 8b) and Watson−Crick A7•T16 base pairs (Figure 8d) at the top and bottom of the structure, respectively, was supported by the observation of the imino protons of T11 at 13.6 ppm and T16 at 13.3 ppm (Figure 2 and Figure S6 of the Supporting Information), as well as a number of NOEs (data not shown). A continuous stacking between the G-tetrad core and the base triad/pairs at the top and the bottom was observed (Figure 8a).

Structure of Form 2: Propeller-Type G-Quadruplex. The H8/H6−H1′ NOE sequential connectivity of Δ*A12-Form 2* is

(37) Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andréola, M. L.; Patel, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 634–639.

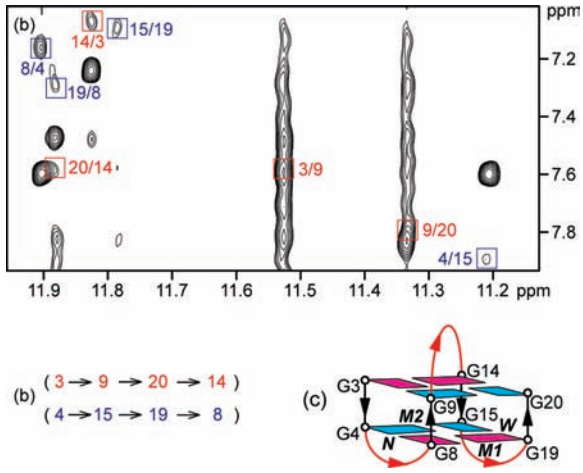

Figure 6. Determination of G-quadruplex folding topology. (a) NOESY spectrum (mixing time, 200 ms) showing imino-H8 connectivity of *I18-Form 1*. Cross-peaks that identify the two G-tetrads are framed and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. (b) Guanine imino-H8 NOE connectivities observed for *I18-Form 1* with G3•G9•G20•G14 and G4•G15•G19•G8 tetrads. (c) Schematic view of the *I18-Form 1* G-quadruplex. *anti* guanines are colored cyan; *syn* guanines are colored magenta. *W*, *M1*, *M2*, and *N* represent wide, medium 1, medium 2, and narrow grooves, respectively. The backbone of the core and loops is colored black and red, respectively.

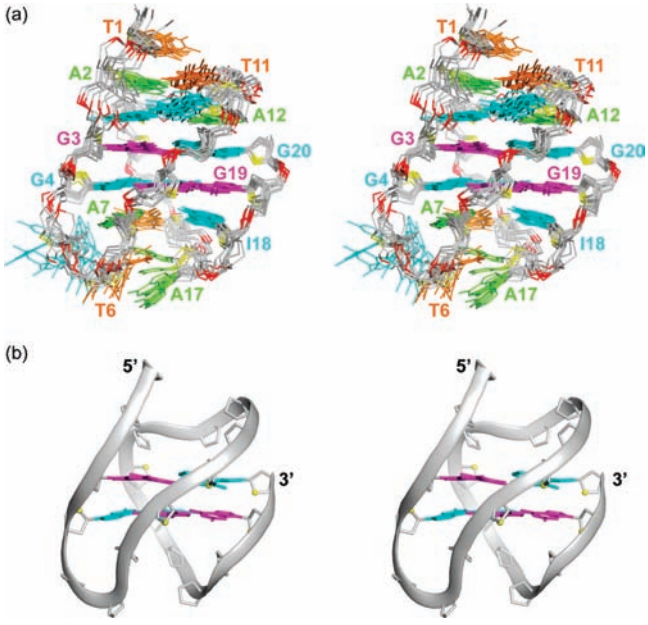

Figure 7. Stereo views of the *I18-Form 1* G-quadruplex structure in K^+ solution. (a) Ten superimposed refined structures of *I18-Form 1*. (b) Ribbon view of a representative structure. The *anti* and *syn* guanines are colored cyan and magenta, respectively; adenines are colored green; thymines, orange; inosines, cyan; backbone and sugar, gray; O4' atoms, yellow; phosphorus atoms, red.

displayed in Figure 9. All residues were observed to adopt *anti* glycosidic conformation (Figure 9; Figure S12 of the Supporting Information). Analysis of imino-H8 connectivity patterns revealed the formation of an intramolecular propeller-type parallel-stranded G-quadruplex with three G-tetrad layers, G3•G8•G12•G17, G4•G9•G13•G18, and G5•G10•G14•G19 (Figure 10). There are three double-chain-reversal loops formed by single residue or two residues. At high concentration of K^+ and/or DNA, end stacking can occur between two such G-quadruplex blocks to

Table 2. Statistics of the Computed Structures of the 20-nt d[(TAGGG)$_3$TAIGG] Sequence

A. NMR restraints		
distance restraints	D_2O	H_2O
intraresidue distance restraints	252	0
sequential (i, $i+1$) distance restraints	144	21
long-range (i, $\geq i+2$) distance restraints	45	37
other restraints		
hydrogen bond restraints	48	
torsion angle restraints	20	
intensity restraints		
nonexchangeable protons (each of four mixing times)	196	
B. Structure statistics for 10 molecules following intensity refinement.		
NOE violations		
number (>0.2 Å)	0.200 ± 0.600	
maximum violation (Å)	0.159 ± 0.035	
rmsd of violations (Å)	0.020 ± 0.002	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.005 ± 0.000	
bond angles (deg)	0.796 ± 0.015	
impropers (deg)	0.452 ± 0.017	
NMR R-factor ($R_{1/6}$)	0.014 ± 0.001	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except G5, T6, and A17	0.96 ± 0.16	
all heavy atoms	1.37 ± 0.21	

form a higher-order structure[38] consistent with our gel-shift and spectroscopic data (see above).

Analysis of Modified Sequences: Structure-Based Rationale for Favoring a Single Conformation in Solution. Imino proton spectra of sequences modified from the natural sequence, in which a guanine located in the loops of Form 1 (namely, G5, G10, G13 and G18) was substituted by an inosine, are plotted in Figure S13 of the Supporting Information. Except for the G10-substituted sequence, the three other sequences favored Form 1. In view of the structures of Form 1 and Form 2, these modifications do not affect Form 1 but destabilize Form 2. In contrast, the G-to-I substitution at position 10 removes a critical amino group for the G10•(A12-G13) in Form 1, and destabilizes this structure. This observation emphasized the importance of the G10•(A12-G13) in the stabilization of Form 1.

Deletion of A12 from the natural sequence (resulting in Δ*A12-Form 2*) favored Form 2 because the removal of this residue (i) abolishes the G10•(A12-G13) triad, destabilizing Form 1, and (ii)

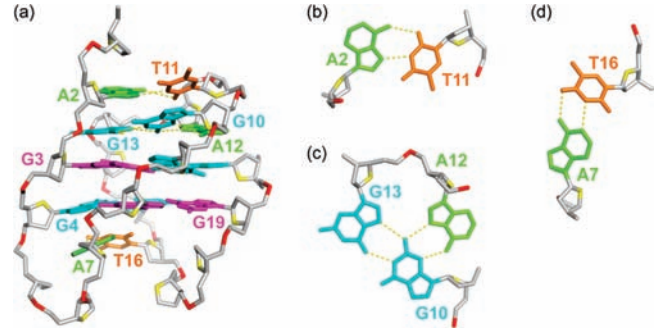

Figure 8. Base pairing and stacking in the *I18-Form 1* G-quadruplex structure. (a) Side view of the structure. (b) A2•T11 Hoogsteen base pair. (c) G10•(A12-G13) triad. (d) A7•T16 Watson−Crick base pair. Color coded as in Figure 7a. Hydrogen bonds between the base triad/pairs are shown by yellow dotted lines.

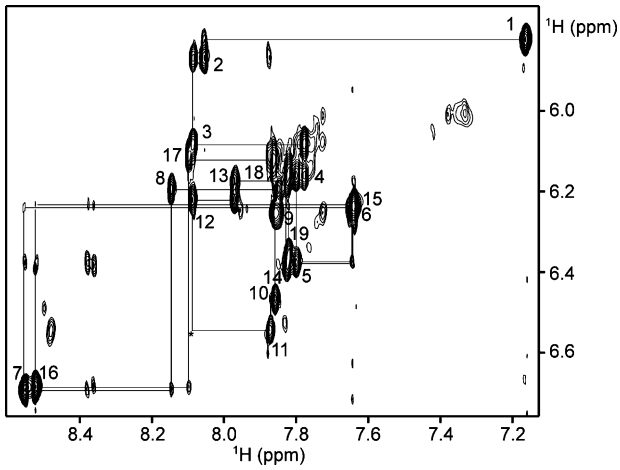

Figure 9. NOESY spectrum (mixing time, 350 ms) showing the H8/H6−H1′ connectivity of Δ*A12-Form 2* in K^+ solution. Intraresidue H8/H6−H1′ NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks.

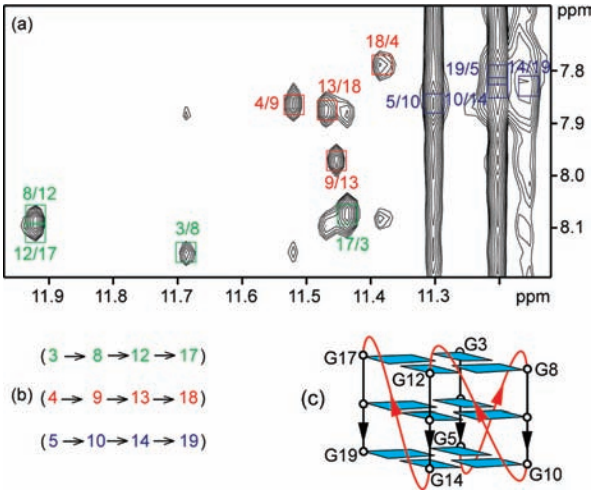

Figure 10. Determination of G-quadruplex folding topology. (a) NOESY spectrum (mixing time, 200 ms) showing imino-H8 connectivity of Δ*A12-Form 2*. Cross-peaks that identify the three G-tetrads are framed and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. (b) Guanine imino-H8 NOE connectivities observed for Δ*A12-Form 2* with G3•G8•G12•G17, G4•G9•G13•G18, and G5•G10•G14•G19 tetrads. (c) Schematic view of the Δ*A12-Form 2* G-quadruplex. The *anti* guanines are colored cyan. The backbone of the core and loops is colored black and red, respectively.

stabilizes Form 2 (a double-chain-reversal loop with one residue is more stable than that with two residues[39,40,10–13]).

Discussion

***Giardia* Telomeric Sequence Forms Two G-Quadruplexes: Loop Sequence Effect.** We have shown that the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ forms two different intramolecular G-quadruplex structures in K^+ solution: the first one is a basket-type antiparallel-stranded G-quadruplex containing two G-tetrads, a G•(A-G) triad, and two A•T base pairs, while the second one is a propeller-type parallel-stranded G-quadruplex containing three G-tetrads. This result shows a dramatic effect of linker length and sequence on the folding topology of G-quadruplexes. Human telomeric sequence, which has the same terminal residues and differs by an extra T in each repeat, forms a (3 + 1) G-quadruplex structure.[18] This is consistent with previous findings that a small change to a G-rich sequence might result in a dramatic change in G-quadruplex topology.[41,42]

Technically, this work demonstrates again that small sequence modifications can be used to manipulate the relative populations of different G-quadruplex conformations and to favor a single conformation for structural analysis.[16–24,32,43]

Interconversion between G-Quadruplexes. In this work, we observed the interconversion between two distinct G-quadruplexes for the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution, indicating that these different G-quadruplex conformations are isoenergetic. The interconversion between different G-quadruplexes has been previously observed for *Tetrahymena* and human telomeric sequences.[43–46] Formation of a mixture of various G-quadruplex forms might be a general property of telomeric sequences,[16–26,43–46] as well as other G-rich genomic sequences.[39,47]

The kinetics of G-quadruplex formation and unfolding might also be an important factor to determine the relevance of different G-quadruplex forms in different biological processes. Different folding/unfolding kinetics could result in different G-quadruplex populations, when the system is not at equilibrium. Previously, distinct folding and unfolding rates have been observed for two human telomeric G-quadruplexes;[43] relative populations of these two forms could vary during the folding/unfolding processes.[43] For the four-repeat *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution, although Form 1 is the major form at equilibrium at low temperatures, the two forms might have comparable populations after a few minutes of folding time (Figure S3 of the Supporting Information).

In nature, different proteins or other cellular factors can recognize a particular G-quadruplex form and selectively promote or abolish this conformation as a step within a regulation pathway. It could be interesting to specifically target this structure by small-molecule ligands with high affinity toward only one form. In these cases, proteins or small-molecule ligands should recognize and distinguish between different structural elements of G-quadruplexes, such as G-tetrad cores and loops.

Propeller-Type Parallel-Stranded G-Quadruplexes. Intramolecular propeller-type parallel-stranded G-quadruplexes have been observed for human telomeric sequences in a K^+-containing crystal[15] but not in dilute solution.[14,16–24] In contrast, the propeller-type G-quadruplex is observed for a *Giardia* telomeric sequence, in which each linker is shorter by one T. This finding is consistent with previous observations that shorter

(38) Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570–2578.
(39) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716.
(40) Hazel, P; Huppert, J; Balasubramanian, S; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415.
(41) Crnugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2003**, *125*, 7866–7871.
(42) Crnugelj, M; Hud, N. V.; Plavec, J. *J. Mol. Biol.* **2002**, *320*, 911–924.
(43) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027.
(44) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Mol. Biol.* **2004**, *338*, 93–102.
(45) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14629–14634.
(46) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938–18943.
(47) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167–173.

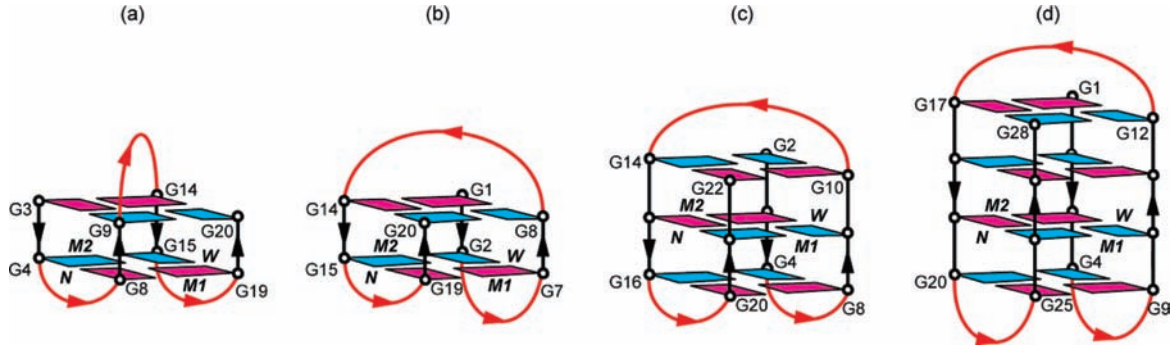

Figure 11. Schematic structures of intramolecular antiparallel basket-type G-quadruplexes formed by (a) the sequence $d(TAGGG)_4$ in K^+ solution (this work), (b) the sequence $d[(GGGTTA)_3GGGT]$ in K^+ solution,[24] (c) the sequence $d[AGGG(TTAGGG)_3]$ in Na^+ solution,[14] and (d) the sequence $d[(GGGGTTTT)_3GGGG]$ in Na^+ solution.[48,49] *anti* guanines are colored cyan; *syn* guanines are colored magenta. The backbone of the core and loops is colored black and red, respectively.

double-chain-reversal loops are more stable than the longer ones.[39,40,10–13]

For the *Giardia* telomeric sequence $d(TAGGG)_4$, high temperatures favor the propeller-type parallel-stranded G-quadruplex over the basket-type antiparallel-stranded G-quadruplex. Similar temperature-dependent behavior was observed for the equilibrium between dimeric parallel-stranded and antiparallel-stranded G-quadruplexes formed by a two-repeat human telomeric sequence in K^+ solution.[43]

Our experimental data on the *Giardia* telomeric sequence $d(TAGGG)_4$ in K^+ solution show that the propeller-type parallel-stranded G-quadruplex form, in which the G-tetrad core's termini are exposed, is more favorable for end-stacking than the antiparallel-stranded G-quadruplex form, in which the G-tetrad core is capped at two ends with loops.

Basket-Type Antiparallel-Stranded G-Quadruplexes. Different basket-type intramolecular G-quadruplexes have been reported previously, containing two G-tetrads (formed by human telomeric sequence in K^+ solution[24] (Figure 11b)), three G-tetrads (formed by human telomeric sequence in Na^+ solution[14] (Figure 11c)) or four G-tetrads (formed by *Oxytricha* telomeric sequence in Na^+ solution[48,49] (Figure 11d)). These three basket-type G-quadruplexes have similar loop arrangements, which are different from the novel basket-type of the Form 1 *Giardia* telomeric G-quadruplex (Figure 11a). When the G-tetrad cores of these four G-quadruplexes are oriented similarly according to the groove widths, the 5′-end of these two different basket types starts from different corners (Figure 11). An alternative way to distinguish these basket-type G-quadruplexes is to consider consecutive loop arrangements. In all basket-type G-qadruplexes, the loops are consecutively edgewise-diagonal-edgewise, but in Form 1 *Giardia* telomeric G-quadruplex the first (edgewise) loop spans across a narrow groove and the third (edgewise) loop spans across a wide groove, while in the other basket-type G-quadruplexes the reverse was observed: the first and third loops span across a wide and narrow groove, respectively.

G-Quadruplexes with Two G-Tetrad Layers. G-quadruplex structures with only two G-tetrad layers have been observed when they are further stabilized by other base pairing and stacking.[25,50] Recently, it has been observed that sequences containing four tracts of three consecutive Gs, can fold into

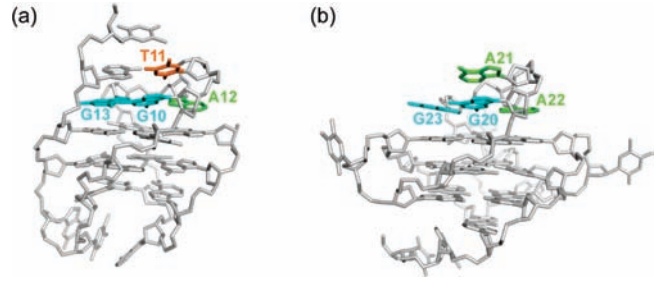

Figure 12. Recurrence of G-quadruplex structural motifs. G•(A-G) triad observed in (a) *I18-Form 1* (this work) and (b) *Pu24I* (pdb code: 2A5P). Only the bases from the diagonal loops are highlighted, all other atoms are colored gray.

G-quadruplex structures involving only two G-tetrad layers.[24,26] To compensate for the loss of a potential additional G-tetrad, again, base pairing and stacking in the loops contribute to the stability of the structure. The observation of Form 1 *Giardia* telomeric G-quadruplex with only two G-tetrad layer reinforces the view that the overall folding topology of a G-quadruplex is defined not only by maximizing the number of G-tetrads, but also by maximizing all possible base pairing and stacking in the loops.[24,26]

Recurrence of Structural Elements in G-Quadruplexes: A "Cut and Paste" Principle for Structure Design and Prediction. Recurrence of several structural elements was found in the structures of Form 1 and Form 2 *Giardia* telomeric G-quadruplexes. (i) Single nucleotide has been reported to form a very stable double-chain-reversal loop;[39,40,10–13] such a loop was found to stabilize Form 2 in Δ*A12-Form 2*. (ii) Multiple base pairs and base triads have been observed in the loops of G-quadruplexes and stabilize these structures by stacking continuously on the G-tetrad cores;[18–24,26] a Hoogsteen A•T base pair, a G•(A-G) triad, and a Watson−Crick A•T base pair were observed to stack at the top and bottom of the G-tetrad core and stabilize Form 1. (iii) The G•(A-G) triad in Form 1, formed in the diagonal loop G10-T11-A12-G13 (single-nucleotide turn is underlined) (Figure 12a), matches well to a G•(A-G) triad previously observed in the *MYC* promoter,[47] formed in a diagonal loop G-A-A-G (Figure 12b). Although the G-tetrad cores of the two structures are quite different, the configurations of the triad-containing diagonal loops are very similar: the first (G), third (A) and fourth (G) bases in the loop participate in a G•(A-G) triad, while the second base (T or A) forms a single-base turn and stacks on top of the triad. The adjacent G-tetrads in the two structures are different (with glycosidic conformations

(48) Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1995**, *251*, 76–94.
(49) Smith, F. W.; Schultze, P.; Feigon, J. *Structure* **1995**, *3*, 997–1008.
(50) Kelly, J. A.; Feigon, J.; Yeates, T. O. *J. Mol. Biol.* **1996**, *256*, 417–422.

being *syn*•*syn*•*anti*•*anti* and *syn*•*anti*•*anti*•*anti*, respectively), but the configurations of the connection points are similar: both loops connect an *anti* guanine to a *syn* guanine and span the same diagonal distance across the corresponding G-tetrad. The recurrence of different structural elements in these structures suggests a "cut and paste" principle for the design and prediction of G-quadruplex topologies, for which different elements could be extracted from one G-quadruplex and inserted in another one. However, one should be cautious that the formation of a particular G-quadruplex "structural element" by a "sequence element" might also be context-dependent due to the presence of different competing conformations.

Acknowledgment. This research was supported by Singapore Ministry of Education grants (Nos. ARC30/07 and RG62/07) and Nanyang Technological University start-up grants (Nos. SUG5/06 and RG138/06) to A.T.P. We thank Chun Li and Samir Amrane for their participation in the early stage of the project, Ngoc Quang Do for his assistance with the gel electrophoresis experiments, and Jean-Louis Mergny, Laurent Lacroix, and Patrizia Alberti from the Museum National d'Histoire Naturelle de Paris for stimulating discussions.

Supporting Information Available: Table S1 (list of all unlabeled and site-specific labeled DNA oligonucleotides used in this study) and Figures S1−S13 (additional gel electrophoresis, CD, UV, and NMR data). This material is available free of charge via the Internet at http://pubs.acs.org.

JA905611C