J|A|C|S ARTICLES

Published on Web 08/12/2010

Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter

Kah Wai Lim,[†,‡] Laurent Lacroix,*[,§,||] Doris Jia En Yue,[†] Joefina Kim Cheow Lim,[†] Jocelyn Mei Wen Lim,[†] and Anh Tuân Phan*[,†]

School of Physical and Mathematical Sciences, and School of Biological Sciences, Nanyang Technological University, Singapore, and INSERM U565, CNRS UMR 7196, Muséum National d'Histoire Naturelle, Paris, France

Received February 11, 2010; E-mail: phantuan@ntu.edu.sg; laurent.lacroix@curie.fr

Abstract: The catalytic subunit of human telomerase, hTERT, actively elongates the 3′ end of the telomere in most cancer cells. The hTERT promoter, which contains many guanine-rich stretches on the same DNA strand, exhibits an exceptional potential for G-quadruplex formation. Here we show that one particular G-rich sequence in this region coexists in two G-quadruplex conformations in potassium solution: a (3 + 1) and a parallel-stranded G-quadruplexes. We present the NMR solution structures of both conformations, each comprising several robust structural elements, among which include the (3 + 1) and all-parallel G-tetrad cores, single-residue double-chain-reversal loops, and a capping A·T base pair. A combination of NMR and CD techniques, complemented with sequence modifications and variations of experimental condition, allowed us to better understand the coexistence of the two G-quadruplex conformations in equilibrium and how different structural elements conspire to favor a particular form.

Introduction

Telomeres are DNA−protein complexes at the very ends of linear eukaryotic chromosomes.[1] In humans, the telomeres consist of thousands of tandem repeats of the sequence TTAGGG[2] that terminate as single-stranded 3′ G-rich overhangs of 100−200 nucleotides (nt).[3] Telomeres maintain the integrity of the chromosomes by protecting them from erosion and fusion events.[4] In human somatic cells, the telomeres progressively shorten after each cell division, ultimately triggering apoptosis when they reach a critical length.[5] However, this event is averted in most cancer cells,[6] wherein the enzyme telomerase is up-regulated and maintains telomere length by catalyzing the synthesis of telomeric DNA repeats onto 3′ telomeric ends. This process is carried out by the catalytic subunit of telomerase, hTERT, which acts as a reverse transcriptase to extend the 3′ TTAGGG overhang by using the RNA subunit (hTR or hTERC) as a template.[7,8]

Various avenues have been explored for the down-regulation of telomerase activity,[9] thereby suppressing the proliferative capacity of cancer cells. It has been shown that formation of G-quadruplexes,[10−13] a four-stranded helical structure based on the stacking of multiple G·G·G·G tetrads,[14] by telomeric DNA substrate inhibits the activity of telomerase.[15] Subsequent studies have established that human telomeric repeats are highly polymorphic,[16−28] being able to adopt at least five different G-quadruplex folding topologies under different experimental

[†] School of Physical and Mathematical Sciences, Nanyang Technological University.
[‡] School of Biological Sciences, Nanyang Technological University.
[§] Muséum National d'Histoire Naturelle.
[||] Present address: UMR 3244 CNRS/UPMC/Institut Curie, Paris, France.

(1) Zakian, V. A. *Science* **1995**, *270*, 1601–1607.
(2) Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6622–6626.
(3) Makarov, V. L.; Hirose, Y.; Langmore, J. P. *Cell* **1997**, *88*, 657–666.
(4) Sandell, L. L.; Zakian, V. A. *Cell* **1993**, *75*, 729–739.
(5) Harley, C. B.; Futcher, A. B.; Greider, C. W. *Nature* **1990**, *345*, 458–460.
(6) Counter, C. M.; Avilion, A. A.; Lefeuvre, C. E.; Stewart, N. G.; Greider, C. W.; Harley, C. B.; Bacchetti, S. *EMBO J.* **1992**, *11*, 1921–1929.
(7) Morin, G. B. *Cell* **1989**, *59*, 521–529.
(8) Greider, C. W.; Blackburn, E. H. *Nature* **1989**, *337*, 331–337.
(9) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L. *Biochimie* **2008**, *90*, 131–155.
(10) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621–628.
(11) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668–698.
(12) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
(13) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429–7455.
(14) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013–2018.
(15) Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature* **1991**, *350*, 718–720.
(16) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.
(17) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.
(18) Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591.
(19) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
(20) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
(21) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–5719.
(22) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
(23) Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185–11191.
(24) Dai, J.; Carver, M.; Yang, D. *Biochimie* **2008**, *90*, 1172–1183.
(25) 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.

conditions.[16−25,28] This paves the way for the development of specific telomeric quadruplex-stabilizing ligands as anticancer drugs.[29−31]

Others have attempted to inhibit telomerase by the direct targeting of its RNA or catalytic component. Peptide nucleic acids (PNAs) directed against hTR[32] and nucleoside analogues serving as chain-terminators[33] are some examples of telomerase inhibitors applying such strategies.

We and two other groups[34,35] independently worked on the potential G-quadruplex formation in the promoter region of hTERT. This would constitute yet another approach involving selective gene regulation at the transcriptional level.[31,36,37] For instance, oncogenic promoters (*c-myc*, *c-kit*, *bcl-2*, etc.), which often display an overrepresentation of G-quadruplex-forming motifs,[36] have been demonstrated to take up a variety of G-quadruplex structures.[36−47] In particular, G-quadruplex formation, stabilized by the porphyrin TMPyP4, within the *c-myc* promoter nuclease hypersensitivity element III_1 (NHE III_1) has been shown to down-regulate *c-myc* transcription.[37] Recently, a naphthalene diimide derivative has also been shown to down-regulate *c-kit* transcription by inducing quadruplex formation within the *c-kit* promoter sequence.[43] In a similar manner, it is conceivable that the formation of G-quadruplexes within the hTERT promoter could potentially modulate its expression.

Here we show that G-rich sequences in the promoter of hTERT (from 20 to 110 nt upstream of the transcription start site) can form stable G-quadruplex structures. Using NMR and CD spectroscopy, we demonstrated that a particular four-G-tract sequence in the hTERT promoter adopts two distinct G-quadruplex conformations in K^+ solution: a (3 + 1) and a parallel-stranded G-quadruplexes. Together, NMR and CD techniques enabled investigation of the high-resolution structures, as well as the relative populations, of both conformations. A combination of different experimental conditions and sequence modifications facilitated a direct characterization of the interconversion between the two forms.

Methods

DNA Sample Preparation. Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer or purchased from Eurogentec (Belgium). DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species.[48]

Gel Electrophoresis. In-gel migration properties (molecular size, shape, and charge) of the structures formed by DNA oligonucleotides were probed by nondenaturing polyacrylamide gel electrophoresis (PAGE) as previously described.[49] Nonradiolabeled oligonucleotides (30 μM) were prepared in a 20 mM Tris-HCl pH 7.2 buffer supplemented with 100 mM KCl. Samples were heated at 90 °C for 10 min and subsequently quenched in ice (fast-cooling) or cooled to 30 °C over 1 h. The samples were loaded on a 15% polyacrylamide gel supplemented with 20 mM KCl and run at 30 °C; 10% sucrose was added just before loading. The gels were imaged by UV-shadowing.

UV-Melting Experiments and Thermal Difference Spectra. The thermal stability of different oligonucleotides was characterized in heating/cooling experiments by recording the UV absorbance at 295 nm as a function of temperature[50] using a Secomam-Uvikon XL UV−vis spectrophotometer. UV-melting experiments were conducted as previously described[51] in a 20 mM lithium cacodylate pH 7.2 buffer containing 100 mM monovalent cation (KCl, LiCl, or a mixture of both). The heating and cooling rates were 0.2 °C per minute. Experiments were performed with 0.2- or 1-cm path length quartz cuvettes. DNA concentration ranged from 0.5 to 100 μM. 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 process. The thermal difference spectra[52] (TDS) were obtained by subtracting the low temperature (2 °C) UV spectra from the high temperature (94 °C) ones. TDS data are represented in molecular extinction coefficient.

Circular Dichroism. Unless otherwise stated, circular dichroism (CD) spectra were recorded at 25 °C on a JASCO-815 spectropolarimeter using a 1-cm path length quartz cuvette with a reaction volume of 500 μL. DNA concentration was typically 4 μM. The DNA oligonucleotides were prepared in a pH 7.0 buffer containing 70 mM KCl and 20 mM potassium phosphate or in the same buffer as for the UV-melting experiments. For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. For equilibrium CD measurements, the 260-nm and 295-nm peaks were monitored until there were minimal changes over time before the spectra were recorded. For unfolding and refolding kinetics study, a single scan was taken every 21 s. To create molecular crowding condition, PEG

(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) Renčiuk, D.; Kejnovská, I.; Školáková, P.; Bednářová, K.; Motlová, J.; Vorlíčková, M. *Nucleic Acids Res.* **2009**, *37*, 6625–6634.
(28) Phan, A. T. *FEBS J.* **2010**, *277*, 1107–1117.
(29) Sun, D. Y.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.* **1997**, *40*, 2113–2116.
(30) Mergny, J. L.; Hélène, C. *Nat. Med.* **1998**, *4*, 1366–1367.
(31) Balasubramanian, S.; Neidle, S. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345–353.
(32) Norton, J. C.; Piatyszek, M. A.; Wright, W. E.; Shay, J. W.; Corey, D. R. *Nat. Biotechnol.* **1996**, *14*, 615–619.
(33) Strahl, C.; Blackburn, E. H. *Nucleic Acids Res.* **1994**, *22*, 893–900.
(34) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878–10891.
(35) Micheli, E.; Martufi, M.; Savino, M. Putative intramolecular G-quadruplex structures within a regulatory element of the human telomerase (hTERT) promoter. In book of abstracts, *Self-assembly of guanosine derivatives: from biological systems to nanotechnological applications*, Universitätszentrum Obergurgl, Austria, June 20−25, 2009.
(36) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167–1172.
(37) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–11598.
(38) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716.
(39) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048–2058.
(40) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167–173.
(41) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589.
(42) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 4386–4392.
(43) Gunaratnam, M.; Swank, S.; Haider, S. M.; Galesa, K.; Reszka, A. P.; Beltran, M.; Cuenca, F.; Fletcher, J. A.; Neidle, S. *J. Med. Chem.* **2009**, *52*, 3774–3783.
(44) Hsu, S. T. D.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2009**, *131*, 13399–13409.
(45) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 5133–5144.
(46) Xu, Y.; Sugiyama, H. *Nucleic Acids Res.* **2006**, *34*, 949–954.
(47) Cogoi, S.; Xodo, L. E. *Nucleic Acids Res.* **2006**, *34*, 2536–2549.
(48) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059–1077.
(49) Guédin, A.; De Cian, A.; Gros, J.; Lacroix, L.; Mergny, J. L. *Biochimie* **2008**, *90*, 686–696.
(50) Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74–78.
(51) Mergny, J. L.; Lacroix, L. *Current Protocols in Nucleic Acid Chemistry*; 2009; Chapter 17, Unit 17.1.
(52) Mergny, J. L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, e138.

Table 1. Properties of Four-G-Tract hTERT Promoter Sequences Used in This Study

name	sequence[a]	position[b]	length	run of G[c]	%G	$^{K100}T_m$[d]	$^{K10}T_m$[e]
GTERT-110	A GGGG TC GGG AC GGGG C GGGG T	−89	22	4/3/4/4	68.0	82	65
GTERT-105	C GGG A GGGG TC GGG AC GGGG C	−85	21	3/4/3/4	67.0	68	52
GTERT-099	C GGGG ACCC GGG A GGGG TC GGG A	−77	23	4/3/4/3	61.0	−	−
GTERT-094	T GGG CC GGGG ACCC GGG A GGGG T	−72	23	3/4/3/4	61.0	−	−
GTERT-089	A GGGGG CT GGG CC GGGG ACCC GGG A	−65	25	5/3/4/3	60.0	62[f]	−
GTERT-082	A GGG CCCGGA GGGGG CT GGG CC GGGG A	−56	27	3/5/3/4	63.0	61	−
GTERT-076	T GGG A GGG CCCGGA GGGGG CT GGG C	−52	25	3/3/5/3	64.0	58	−
GTERT-071	A GGGG CT GGG A GGG CCCGGA GGGGG C	−46	26	4/3/3/5	65.0	−	−
GTERT-060	A GGGG A GGGG CT GGG A GGG C	−41	20	4/4/3/3	70.0	78	55
GTERT-056	C GGGG CCGCGGAAAGGAA GGGG A GGGG CT GGG A	−24	33	4/4/4/3	60.0	61	56[g]
GTERT-051	A GGG C GGGG CCGCGGAAAGGAA GGGG A GGGG C	−20	32	3/4/4/4	62.5	70	54

[a] Guanines constituting the four G-tracts are underlined. [b] Position of the first nucleotide relative to the transcription start site. [c] Length (number of guanines) of the four successive G-tracts. [d] Melting temperature of the sequence in the presence of 100 mM KCl. [e] Melting temperature of the sequence in the presence of 10 mM KCl and 90 mM LiCl. [f] A second transition with a hysteresis was observed at low temperatures (20−30 °C). [g] This transition presents a hysteresis; the half-transition temperature, $T_{1/2}$(melting) = 59 °C and $T_{1/2}$(annealing) = 52 °C.

200 (10%, 20%, 30%, and 40% v/v) was added to a final DNA concentration of ∼4 μM in 70 mM KCl and 20 mM potassium phosphate. The samples were incubated at 25 °C for 5 days.

NMR Spectroscopy. Samples for NMR study were dialyzed successively against ∼30 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 7.0). 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,[53] site-specific ^{2}H labeling,[54] and through-bond correlations at natural abundance.[55] Spectral assignments were completed by NOESY, COSY, TOCSY, and {^{13}C−^{1}H}-HSQC, as previously described.[56] 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 d[AGGGIAGGGGCTGGGAGGGC] (Form 1) and d[AIGGGAGGGICTGGGAGGGC] (Form 2) G-quadruplexes were classified based on NOESY experiments performed in H_2O (mixing time, 200 ms) and D_2O (mixing times, 100, 200, and 300 or 350 ms). Structure computations were performed using the XPLOR-NIH program[57] in two general steps: (i) distance geometry simulated annealing and (ii) distance-restrained molecular dynamics refinement. Hydrogen bond restraints, interproton distance restraints, dihedral restraints, planarity restraints, and repulsive restraints were imposed during structure calculations. Detailed procedures are described in the Supporting Information Text. Structures were displayed using the PyMOL program.[58]

Data Deposition. The coordinates for the d[AGGGIAGGGGCTGGGAGGGC] (Form 1) and d[AIGGGAGGGICTGGGAGGGC] (Form 2) G-quadruplexes have been deposited in the Protein Data Bank (accession codes 2KZD and 2KZE, respectively).

Results

hTERT Promoter Exhibits Exceptional Potential for G-Quadruplex Formation. The potential of DNA sequences in the telomerase (TERT) promoters from 23 vertebrates (Ensembl release 54, May 2009)[59] to form G-quadruplexes was evaluated with the G4P calculator[60] by analyzing a 240-nt region upstream of the first exon. Eleven species showed high potentials of G-quadruplex formation (at least four stretches of three or more consecutive guanines were identified using a scanning window of 40 nt with a step increment of 1 nt) for either the coding or noncoding strands (Supporting Information Table S1). The high G4P score of 34.17% for the noncoding strand of hTERT promoter suggested a high potential for G-quadruplex formation, especially the segment between −20 and −110 nt upstream of the transcription start site (TSS) (Supporting Information Table S1). This 91 nt long segment contains 14 G-tracts, with linkers ranging in length from 1 to 13 nt. All eleven sequences spanning four successive stretches of three or more consecutive guanines within this region were investigated (Table 1).

Formation of G-Quadruplexes by hTERT Promoter Sequences. Melting experiments were performed for eleven four-G-tract sequences taken from the hTERT promoter (Table 1) by monitoring the UV absorbance at 295 nm.[50] TDS and CD experiments[52,61] were also performed for these oligonucleotides (data not shown). Results from these experiments suggested the formation of G-quadruplexes for eight sequences (*GTERT-089*, *GTERT-082*, *GTERT-076*, *GTERT-110*, *GTERT-105*, *GTERT-060*, *GTERT-056*, and *GTERT-051*) in solution containing 100 mM K^+; the melting temperatures (T_m) ranged from 58 to 82 °C (Table 1). In the presence of only 10 mM K^+, the five latter sequences still formed stable G-quadruplexes with T_m ranging from 52 to 65 °C (Table 1).

The 20-nt Fragment *GTERT-060* Adopts Two Major G-Quadruplex Folds in K^+ Solution. One particular fragment, d[AGGGGAGGGGCTGGGAGGGC] (designated *GTERT-060*; Table 1), positioned between −41 and −60 nt relative to the TSS, produced relatively well-resolved NMR spectra amenable for structural elucidation. The number and intensity of imino proton peaks observed at 10.9−12 ppm (Figure 1G), characteristic of G-tetrad formation, indicated that *GTERT-060* adopts two major G-quadruplex folds in K^+ solution. *GTERT-060* has the highest G-content (70%) among the eleven oligonucleotides, and the structures formed by this sequence in K^+ solution are very stable; the melting temperature (T_m) in the presence of 100 mM KCl is 78 °C (Table 1).

UV-melting profiles (Figure 2A; Supporting Information Table S2), TDS (Supporting Information Figure S1A) and CD spectra (Supporting Information Figure S1B) illustrated the

(53) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
(54) Huang, X.; Yu, P.; LeProust, E.; Gao, X. *Nucleic Acids Res.* **1997**, *25*, 4758–4763.
(55) Phan, A. T. *J. Biomol. NMR* **2000**, *16*, 175–178.
(56) Phan, A. T.; Guéron, M.; Leroy, J. L. *Methods Enzymol.* **2001**, *338*, 341–371.
(57) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.
(58) DeLano, W. L. *The PyMOL User's Manual*; DeLano Scientific: Palo Alto, CA, 2002.
(59) Hubbard, T. J. P. et al. *Nucleic Acids Res.* **2009**, *37*, D690–D697.
(60) Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2006**, *34*, 3887–3896.
(61) Gray, D. M.; Wen, J. D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J. *Chirality* **2008**, *20*, 431–440.

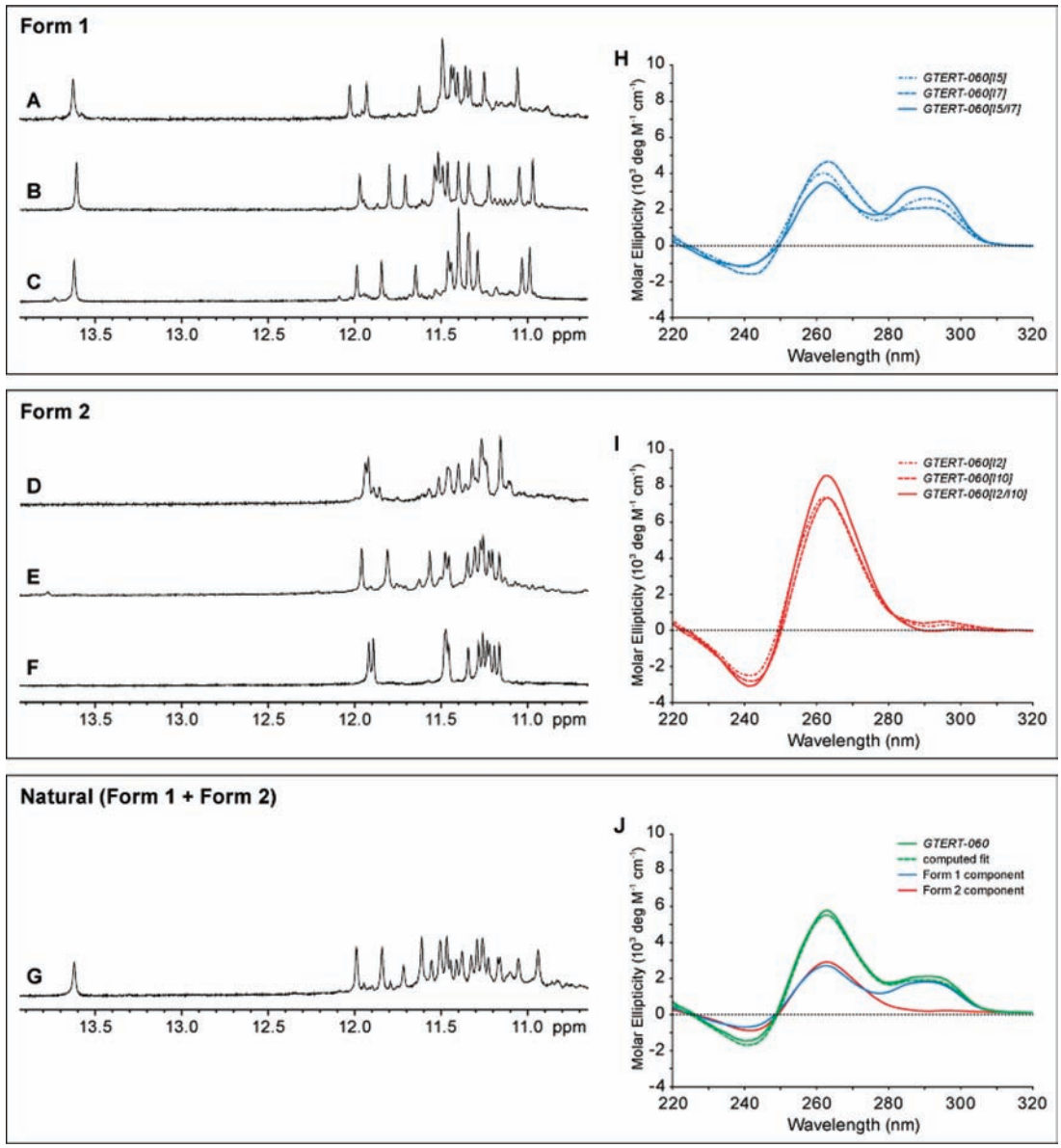

Figure 1. (A−G) Imino proton spectra of the natural and inosine-substituted sequences of the 20-nt hTERT promoter fragment d[AGGGGAGGGGCTGGGAGGGC] in K^+ solution, (A) *GTERT-060[I5]*, (B) *GTERT-060[I7]*, (C) *GTERT-060[I5/I7]*, (D) *GTERT-060[I2]*, (E) *GTERT-060[I10]*, (F) *GTERT-060[I2/I10]* and (G) *GTERT-060*. (H−J) CD spectra recorded at 25 °C for (H) Form 1, (I) Form 2, and (J) the natural hTERT promoter sequence *GTERT-060*. The latter was deconvoluted into its Form 1 and Form 2 component spectra based on a linear fit (see text). Spectra in H and I corresponding to individual oligonucleotides are color-coded on the top right corner, respectively. In J, the blue and red curves represent compositions of Form 1 (64%) and Form 2 (36%), respectively, whose sum (dotted spectrum) gave the best fit to *GTERT-060* (green line).

monovalent cation dependency of the structures formed: both the stability and the folded fraction of G-quadruplexes decreased upon lowering the concentration of K^+. At constant concentration of K^+, the melting temperature of *GTERT-060* was independent of the oligonucleotide concentration (0.5−100 μM; Figure 2B), while nondenaturing gel electrophoresis showed a diffuse fast-migrating band (Supporting Information Figure S2, Lanes 3 and 9). These results argued that *GTERT-060* forms intramolecular G-quadruplex structures.

Favoring a Single Conformation in *GTERT-060* by Inosine Substitutions. We could favor either species by appropriate guanine-to-inosine substitutions (Table 2); single substitution of G5 (*GTERT-060[I5]*) or G7 (*GTERT-060[I7]*) by inosine favored one form (Figure 1A,B), whereas replacement of G2 (*GTERT-060[I2]*) or G10 (*GTERT-060[I10]*) by inosine favored the other (Figure 1D,E). We shall refer to the former and latter conformations as Form 1 and Form 2, respectively. Double inosine substitutions, (G5 and G7; *GTERT-060[I5/I7]*) or (G2 and G10; *GTERT-060[I2/I10]*), retained the same major conformations, respectively, as above (Figure 1C,F). The rationale for these inosine substitutions will be discussed in relation to the structures described below.

The NMR spectrum of Form 1 displayed twelve major peaks at 10.9−12.0 ppm (Figure 1A−C), indicating formation of a three-G-tetrad G-quadruplex. In addition, at 25 °C Form 1 showed a sharp imino proton peak at ∼13.6 ppm, which was not observed in Form 2. At 45 °C, the twelve former peaks remained sharp[25,26] while the latter peak was broadened (Supporting Information Figure S3), consistent with the latter proton belonging to an A·T base pair (see NMR structure below). This signature imino proton was observed for the 22-nt sequence *GTERT-060[+2]* (Table 2; Supporting Information

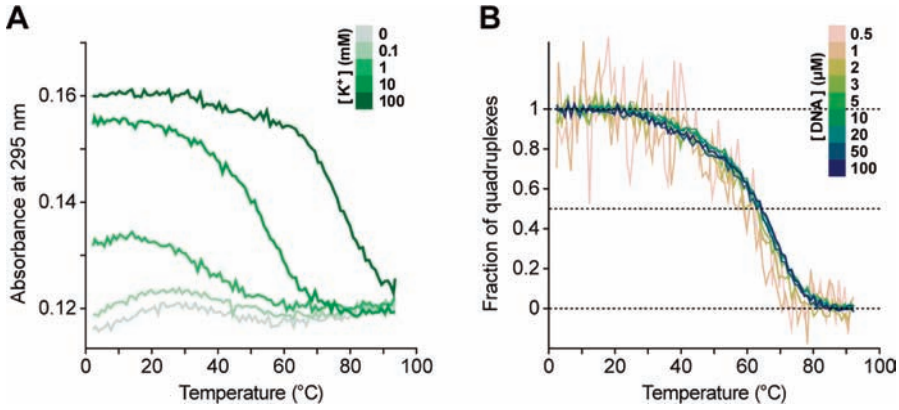

Figure 2. (A) UV-melting profiles at different concentrations of K^+ (0−100 mM; color-coded on the top right corner), illustrating the monovalent cation dependency of *GTERT-060*. For each experiment, DNA concentration was 4 μM; the total concentration of KCl and LiCl was 100 mM; solutions contained 20 mM lithium cacodylate, pH 7.2. (B) Fractions of G-quadruplexes as a function of temperature at different concentrations of *GTERT-060* (0.5−100 μM; color-coded on the top right corner) in the presence of 30 mM KCl and 70 mM LiCl. Solutions contained 20 mM lithium cacodylate, pH 7.2.

Table 2. Natural and Modified Sequences of the hTERT Promoter Fragment *GTERT-060* Used for NMR Analysis, with Estimated Populations for the Two Coexisting G-Quadruplex Conformations at 25 °C

name	sequence[a]	population (%)		
		Form 1	Form 2	others
GTERT-060	A GGGG A GGGG CT GGG A GGG C	>40	<40	<20
GTERT-060[I2]	A **I**GGG A GGGG CT GGG A GGG C	−	>85	<15
GTERT-060[I5]	A GGG**I** A GGGG CT GGG A GGG C	>90	−	<10
GTERT-060[I7]	A GGGG A **I**GGG CT GGG A GGG C	>90	−	<10
GTERT-060[I10]	A GGGG A GGG**I** CT GGG A GGG C	−	>90	<10
GTERT-060[I5/I7]	A GGG**I** A **I**GGG CT GGG A GGG C	>90	−	<10
GTERT-060[I2/I10]	A **I**GGG A GGG**I** CT GGG A GGG C	−	>95	< 5
GTERT-060[+2]	AA GGGG A GGGG CT GGG A GGG CC	>50	<40	<10

[a] Substituted inosine (I) residues are shown in boldface.

Figure S4), an expanded sequence of *GTERT-060*, suggesting that formation of Form 1 in a longer natural sequence context is possible. UV-melting data indicated that *GTERT-060[I5]* adopts an intramolecular structure (T_m was independent of the DNA concentration from 3 to 50 μM; data not shown). Under nondenaturing condition, this sequence migrated as a sharper band as compared to the natural sequence *GTERT-060* and its mobility corresponds to that of the faster-migrating species of the natural sequence (Supporting Information Figure S2, Lane 5).

Form 2, on the other hand, showed twelve major peaks at 11.1−12.0 ppm (Figure 1D−F). The absence of the imino proton at ~13.6 ppm and the distinctive patterns of the twelve imino protons suggested formation of a three-G-tetrad G-quadruplex with a different scaffold. UV-melting data indicated that *GTERT-060[I10]* adopts an intramolecular structure (T_m was independent of the DNA concentration from 3 to 50 μM; data not shown). Under nondenaturing condition, this sequence also migrated as a sharper band as compared to the natural sequence, but its mobility matches that of the slower-migrating species of the natural sequence (Supporting Information Figure S2, Lane 7).

CD Spectra of Two G-Quadruplex Conformations. The CD spectra of Form 1 and Form 2 in K^+ solution at 25 °C differed considerably (Figure 1H,I). Form 1 displayed two positive maxima around 260 and 295 nm, and a negative minimum around 240 nm. In contrast, Form 2 exhibited just a single maximum around 260 nm together with the 240-nm negative minimum. The 295-nm peak is indicative of two adjacent tetrads having opposite hydrogen-bond directionalities, whereas the 260-nm peak is characteristic of a G-quadruplex in which all the quartets have the same polarity.[61] These observations are consistent with Form 1 being a (3 + 1) G-quadruplex and Form 2 adopting a propeller-type parallel-stranded G-quadruplex topology (see NMR data and structures below).

The CD spectrum of the natural sequence *GTERT-060* at 25 °C (solid green curve, Figure 1J) displayed a negative minimum around 240 nm, a single positive maximum around 260 nm, and a plateau stretching from 280 to 300 nm, implying contributions from both Form 1 and Form 2. This was in agreement with the NMR data, which indicated that *GTERT-060* adopts Form 1 and Form 2 as two major conformations in K^+ solution (Figure 1G). With a linear fit $C = \alpha A + (1-\alpha)B$, where α and $(1-\alpha)$ represent the proportions of the two species, A and B represent the respective component spectra, and C represents the composite spectrum, we could estimate the relative populations of both forms in the natural sequence *GTERT-060*, assuming that the proportion of *GTERT-060* adopting other conformation(s) is negligible. Depending on the choice of component spectra, we obtained populations of ~50−80% and ~20−50% for Form 1 and Form 2, respectively (Figure 1J; Supporting Information Text and Figure S5), in accord with the NMR data (Figure 1G and Table 2). Note however that relative populations of the two conformations are sensitive to sample treatment and experimental condition.

Interconversion and Melting of Two G-Quadruplex Conformations. CD spectra of the natural sequence *GTERT-060* at equilibrium in solution containing 70 mM KCl and 20 mM potassium phosphate were recorded over the range of temperatures from 5 to 90 °C (Figure 3A) at 5 °C intervals. At each interval, the CD spectrum was recorded after equilibrium was reached, determined by monitoring the 260- and 295-nm

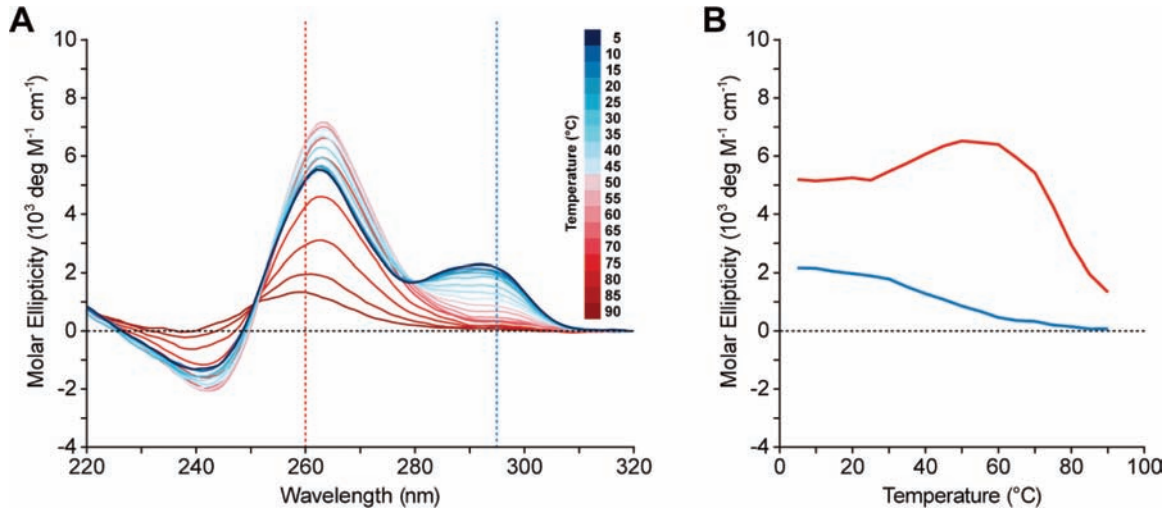

Figure 3. (A) CD spectra of *GTERT-060* at equilibrium from 5 to 90 °C (color-coded on the top right corner). Solutions contained 70 mM KCl and 20 mM potassium phosphate, pH 7.0. (B) The intensity of the peaks at 260 nm (red; Form 1 and Form 2) and 295 nm (blue; Form 1) gave an indication of the interconversion and melting of the two coexisting conformations (see text).

peaks (see below). Form 2 was favored at high temperatures (>25 °C), in agreement with NMR data (5−45 °C; Supporting Information Figure S6). As temperature was increased, Form 1 presumably underwent melting or conversion to Form 2, and this could be directly monitored by the 295-nm peak (Figure 3B, blue curve), which does not contain any contribution from Form 2. On the other hand, the 260-nm peak consists of components from both Form 1 and Form 2 and showed that the population of Form 2 increased with the rise of the temperature from 25 to 50 °C, and, beyond which, decreased with the rise of the temperature from 50 to 90 °C (Figure 3B, red curve). Based on the intensity of the 260-nm peak, the melting temperature of Form 2 should be higher than 75 °C, consistent with results from the UV-melting experiments (Table 1 and Figure 2). This experiment demonstrates that CD could potentially be used to detect the presence of multiple conformations within a particular sequence.

Refolding kinetics of the G-quadruplexes could also be monitored following quick sample cooling from 90 to 20 °C. Formation of the two conformations was observed in the order of minutes, followed by a slow conversion from Form 2 to Form 1 (Supporting Information Figure S7). Note that the NMR spectra of *GTERT-060* (Figure 1G; Supporting Information Figure S6) indicated the possible existence of additional minor conformation(s) (<20%), which could not have been discerned based on CD alone.

Molecular Crowding Condition Favors Form 2. Under molecular crowding condition, simulated by addition of the crowding agent PEG 200 at varying compositions (from 10 to 40% v/v), the equilibrium between two conformations (Form 1 and Form 2) in the natural *GTERT-060* sequence was shifted over time to favor the parallel-stranded G-quadruplex (Form 2), as shown by CD spectra (Figure 4). The conformational conversion is nearly complete at 30% PEG. The results were consistent with previous studies, which showed that molecular crowding condition favors a conversion toward parallel-stranded G-quadruplexes.[23,62]

NMR Spectral Assignments of Form 1 and Form 2. We decided to solve the solution structures of Form 1 and Form 2

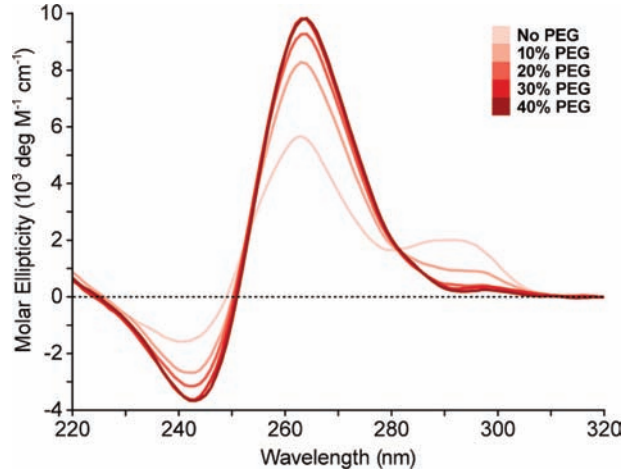

Figure 4. CD spectra showing the conversion of *GTERT-060* from an equilibrium of two conformations (Form 1 and Form 2; light pink) to predominantly a single conformation (Form 2; maroon) induced by the addition of PEG (color-coded on the top right corner) and incubation at 25 °C for 5 days. Solutions contained 70 mM KCl and 20 mM potassium phosphate, pH 7.0.

based on the *GTERT-060[I5]* and *GTERT-060[I2/I10]* sequences, respectively. For both oligonucleotides, guanine imino and H8 protons were unambiguously assigned using site-specific low-enrichment ^{15}N labeling,[53] site-specific ^{2}H labeling,[54] and through-bond correlations at natural abundance[55] (Figures 5 and 6; Supporting Information Table S3). With the help of these unambiguous assignments and other through-bond correlation experiments (COSY, TOCSY, and {^{13}C−^{1}H} HSQC; data not shown), the H8/H6−H1′ NOE sequential connectivity could be traced from A1 through C20 for both sequences (Supporting Information Figures S8 and S9). For *GTERT-060[I5]*, the intensity of intraresidue H8−H1′ NOE cross-peaks (Supporting Information Figure S10) indicated *syn* glycosidic conformation for G2, G8, G9, G13, and G17, and *anti* for all other guanine residues; for *GTERT-060[I2/I10]*, all guanine residues assume *anti* glycosidic conformation (Supporting Information Figure S9).

Form 1 is an Intramolecular (3 + 1) G-Quadruplex. The imino-H8 NOE connectivity patterns (Figure 7A) of *GTERT-*

(62) Miyoshi, D.; Nakao, A.; Sugimoto, N. *Biochemistry* **2002**, *41*, 15017–15024.

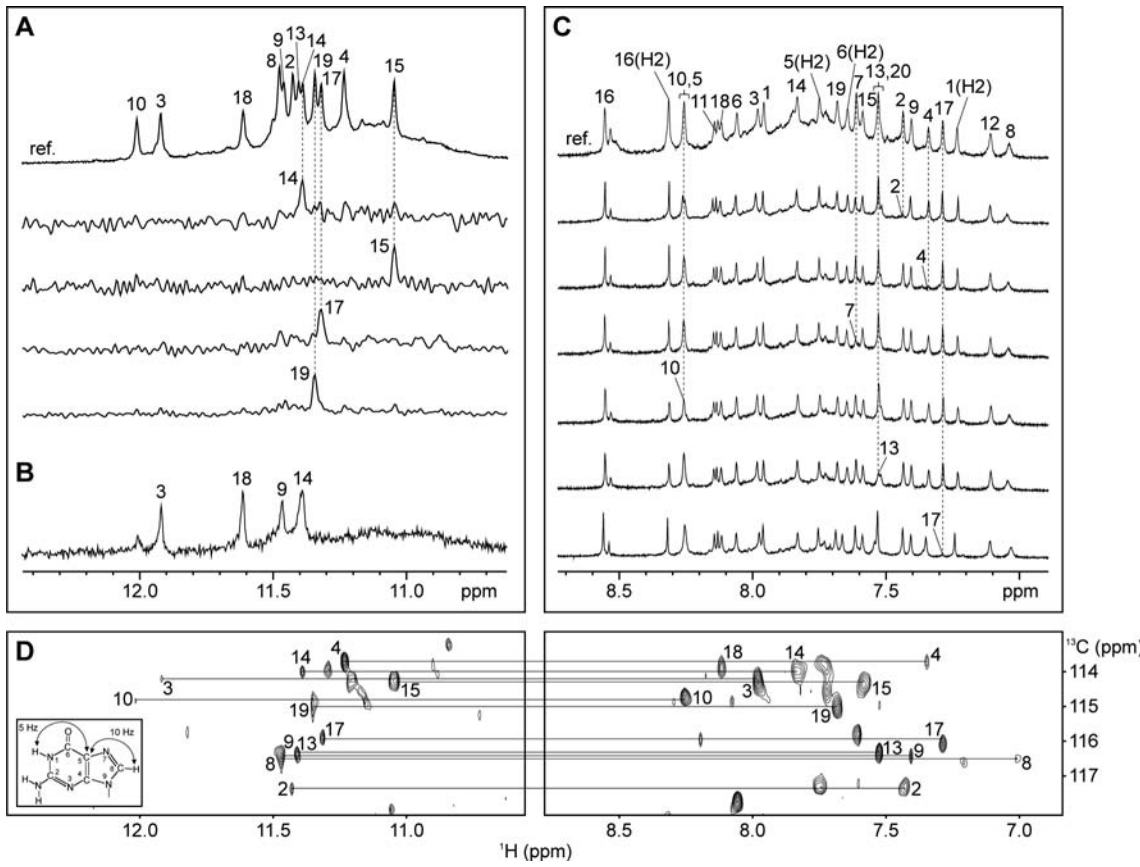

Figure 5. Guanine imino and H8 proton assignments of *GTERT-060[I5]* (Form 1) in K^+ solution. (A) Examples of imino proton assignments in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. The reference spectrum (ref.) is shown at the top. (B) Imino proton spectra after 30 min in D_2O at 25 °C. (C) Examples of H8 proton assignments by site-specific ^{2}H labeling at the indicated positions. The reference spectrum (ref.) is shown at the top. (D) Through-bond correlations between guanine imino and H8 protons via ^{13}C5 at natural abundance, using long-range J-couplings shown in the inset. Assignments are labeled with residue numbers.

060[I5] pointed to an intramolecular (3 + 1) G-quadruplex structure (Figure 7C). The core consists of three G-tetrads, G2·G17·G13·G10, G3·G9·G14·G18 and G4·G8·G15·G19, and the hydrogen-bond directionalities of the G-tetrads are successively anticlockwise-clockwise-clockwise. As viewed from the top, the glycosidic conformations of guanines around the first G-tetrad are *syn*·*syn*·*syn*·*anti* while those for the two other G-tetrads are *anti*·*syn*·*anti*·*anti*. There are one wide, one narrow, and two medium grooves, and the loops are consecutively edgewise-edgewise-propeller. The first loop (I5-A6-G7, at the bottom) spans the wide groove, the second loop (C11-T12, on the top) bridges the narrow groove, and the third loop (A16) connects the two parallel strands across the medium groove.

The structure of the *GTERT-060[I5]* G-quadruplex in K^+ solution (Figure 8) was calculated on the basis of NMR restraints (Table 3; Supporting Information Text). The three-G-tetrad core is comparable to that of human telomeric (3 + 1) G-quadruplexes observed previously (Supporting Information Figure S11).[18−22,24,28] The central placing of the G3·G9·G14·G18 tetrad (between the G2·G17·G13·G10 and G4·G8·G15·G19 tetrads) was consistent with the observation of the imino protons of G3, G9, G14, and G18 being well-protected from the exchange with solvent (Figure 5B).

The top G2·G17·G13·G10 tetrad is capped by a reverse Hoogsteen base pair between A1 and T12, which gives rise to the sharp imino proton peak at ~13.6 ppm (Figure 1A−C,G).[63] In the current structure, T12 stacks over G10 and G13 of the top G-tetrad, supported by numerous NOEs between the base and sugar protons of G10 and T12, as well as several NOEs between the imino proton of G13 and the sugar protons of T12 (data not shown). NOE connectivities between the H8 proton of A1 and the imino protons of G2 and G10 defined the stacking position of A1. The bases of C11 (from the top lateral loop) and A16 (from the double-chain-reversal loop) project outward from their respective loops, consistent with these bases showing very few NOEs to the protons of neighboring residues (data not shown). The alignment of the three-purine loop (I5-A6-G7) and C20 at the bottom of the structure is less well-defined and perhaps reflect the presence of multiple conformations within this region.

The 2D NOESY spectra of *GTERT-060[I7]* corroborated our stance that it adopts the same (3 + 1) G-quadruplex fold as *GTERT-060[I5]*; the excellent correspondence of the H8/H6−H1′ connectivities, coupled with the appearance of five *syn* intraresidue H8−H1′ cross-peaks (data not shown), strongly argued for the same (3 + 1) scaffold.

Form 2 is a Parallel-Stranded Intramolecular G-Quadruplex. The all-parallel arrangement of the four strands constituting the core of *GTERT-060[I2/I10]* quadruplex was deduced from the imino-H8 NOE connectivity patterns around each G-tetrad

(63) Hu, L.; Lim, K. W.; Bouaziz, S.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 16824–16831.

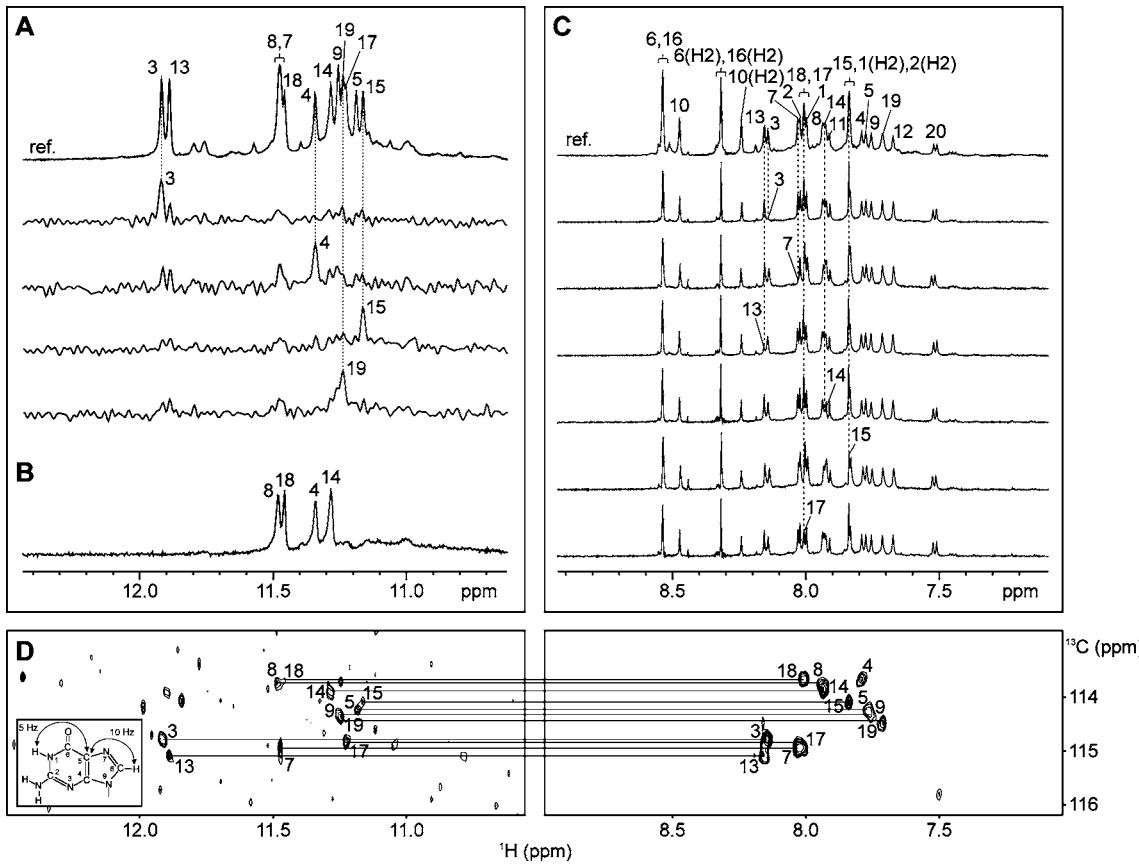

Figure 6. Guanine imino and H8 proton assignments of *GTERT-060[I2/I10]* (Form 2) in K^+ solution. (A) Examples of imino proton assignments in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. The reference spectrum (ref.) is shown at the top. (B) Imino proton spectra after 10 min in D_2O at 25 °C. (C) Examples of H8 proton assignments by site-specific ^{2}H labeling at the indicated positions. The reference spectrum (ref.) is shown at the top. (D) Through-bond correlations between guanine imino and H8 protons via ^{13}C5 at natural abundance, using long-range *J*-couplings shown in the inset. Assignments are labeled with residue numbers.

(G3·G7·G13·G17, G4·G8·G14·G18, and G5·G9·G15·G19; Figure 7A′,C′). All guanines adopt *anti* glycosidic conformation (Supporting Information Figure S9), and the hydrogen-bond directionalities of the G-tetrads are all clockwise (looking down from the top). The three double-chain-reversal loops straddle the three-layered core and the four grooves are of medium widths.

The structure of the *GTERT-060[I2/I10]* G-quadruplex in K^+ solution (Figure 9) was computed on the basis of NMR restraints (Table 4; Supporting Information Text). The three-G-tetrad parallel core is similar to that of human telomeric parallel-stranded G-quadruplex observed previously in a K^+-containing crystal.[17] The central placing of the G4·G8·G14·G18 tetrad (between the G3·G7·G13·G17 and G5·G9·G15·G19 tetrads) was consistent with the observation of the imino protons of G4, G8, G14, and G18 being well-protected from the exchange with solvent (Figure 6B).

In the current structure, the base of I2 folds into the hydrophobic groove, whereas A1 and C20 stack over the outer G-tetrads, with the former possibly forming a Watson−Crick base pair with T12 from the middle 3-nt loop I10-C11-T12. The base I10 interacts with the bases of the core along the groove while C11 projects outward from the loop. The two single-residue double-chain-reversal loops A6 and A16 are situated at opposite sides of the core and adopt almost identical configuration, as shown by many similar spectral characteristics between the segments G3−G9 and G13−G19 including chemical shifts and NOE patterns (Figure 6 and Supporting Information Figure S9).

In the 2D NOESY spectra of *GTERT-060[I10]*, the intensity of intraresidue H8−H1′ cross-peaks indicated that all guanines are *anti* (data not shown), consistent with the formation of a parallel-stranded G-quadruplex structure.

Analysis of Modified Sequences. In order to probe what roles different structural elements might play in stabilizing the two different folds, we modified several loop residues of the natural sequence *GTERT-060*. It has been reported previously that a single-residue double-chain-reversal loop is highly stable, especially so when the base is a thymine.[38,49,64,65] Indeed, double mutations of A6 and A16 to thymines increased the population of Form 2 (see figure legend, Supporting Information Figure S12B). It should be noted that A6 could also be involved in the stabilization of Form 1. In contrast, the contribution of the A1·T12 reverse Hoogsteen base pair toward the stability of Form 1 is apparently more significant, as a single T12C mutation abrogated the population of Form 1 (Supporting Information Figure S12C). Applied together, these three modifications eliminated Form 1 almost entirely (Supporting Information

(64) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415.
(65) Rachwal, P. A.; Brown, T.; Fox, K. R. *FEBS Lett.* **2007**, *581*, 1657–1660.

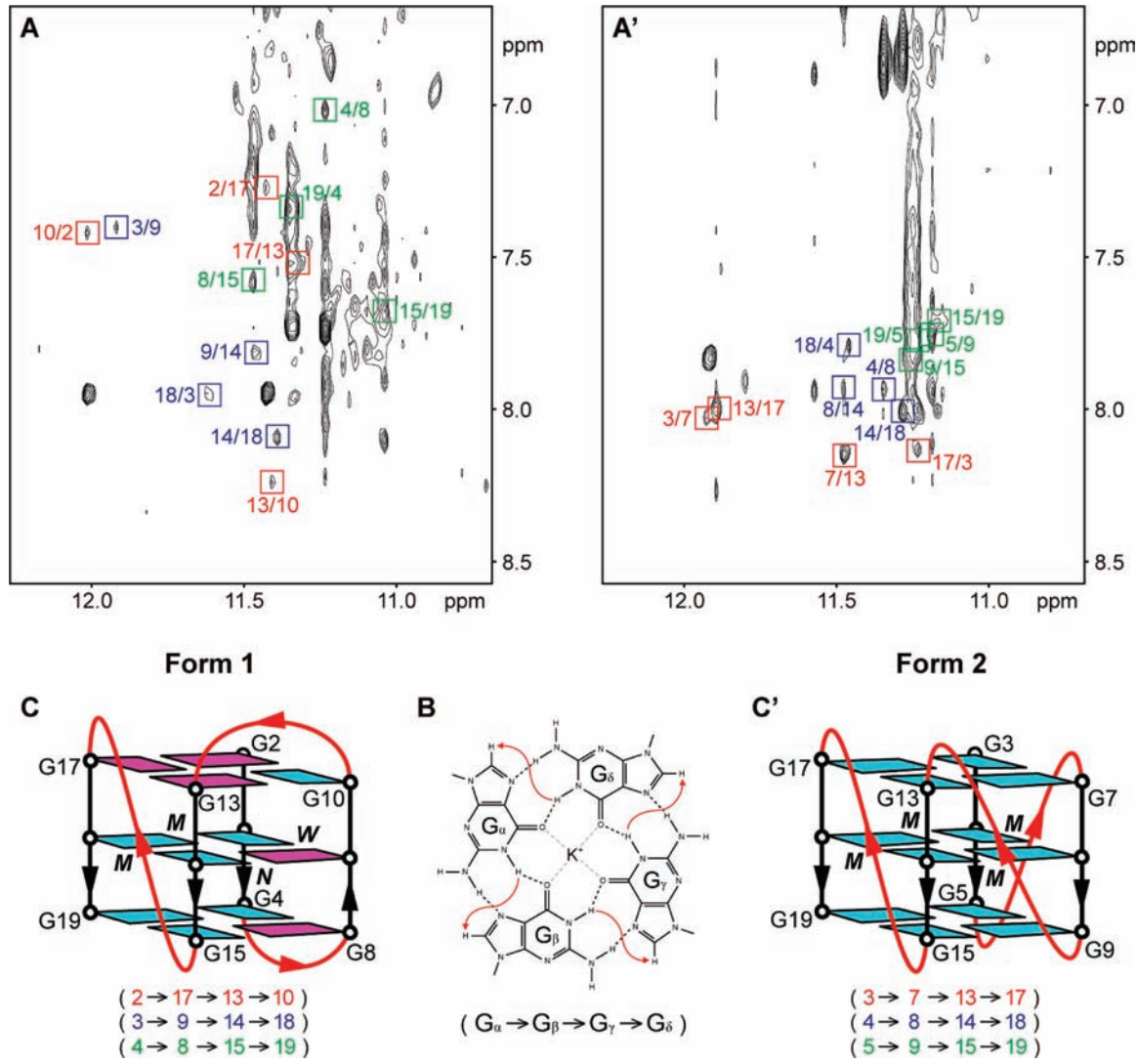

Figure 7. NOESY spectra (mixing time, 200 ms) showing the imino-H8 connectivity of (A) *GTERT-060[I5]* (Form 1) and (A′) *GTERT-060[I2/I10]* (Form 2). Cross-peaks that identify the arrangement of the G-tetrads are framed and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. (B) Characteristic guanine imino-H8 NOE connectivity patterns around a $G_\alpha \cdot G_\beta \cdot G_\gamma \cdot G_\delta$ tetrad as indicated with arrows (connectivity between G_δ and G_α implied). (C and C′) Schematic structures of Form 1 and Form 2 G-quadruplexes adopted by *GTERT-060[I5]* and *GTERT-060[I2/I10]*, respectively, which satisfy the connectivities shown below in parentheses. *anti* and *syn* guanines are colored cyan and magenta, respectively. *W*, *M*, and *N* represent wide, medium, and narrow grooves, respectively. The backbones of the core and loops are colored black and red, respectively.

Figure S12D). Note however the emergence of additional species in the spectra of the modified sequences.

Discussion

Formation of G-Quadruplexes in the hTERT Promoter. We have shown that the G-rich sequence in the hTERT promoter has a high potential to form G-quadruplex structures (Supporting Information Table S1). Eight out of eleven candidate sequences containing four consecutive G-tracts can form stable G-quadruplexes with melting temperatures ranging from 58 to 82 °C in the presence of 100 mM K^+ (Table 1). Note that in a long sequence context multiple G-quadruplexes can be formed by various combinations of G-tracts, including structures that involve nonconsecutive G-tracts. On the other hand, the kinetics of G-quadruplex folding/unfolding should also determine the relative populations of different G-quadruplexes on the folding path[66,67] and one might suppose that the kinetics of intramolecular G-quadruplex folding involving four consecutive G-tracts are among the fastest.

Coexistence and Interconversion between Two G-Quadruplex Conformations. For *GTERT-060*, the four-G-tract sequence with the highest G-content in the hTERT promoter (Table 1), we observed in K^+ solution a mixture of two major intramolecular G-quadruplexes: Form 1 is an intramolecular (3 + 1) G-quadruplex while Form 2 is a propeller-type parallel-stranded G-quadruplex. The observation of the two forms interconverting in solution indicates that they are isoenergetic, with the equilibrium shifting toward Form 2 at high temperatures[67] (Figure 3; Supporting Information Figure S6). Furthermore, Form 2 is also favored under molecular crowding condition (Figure 4), which emulates intracellular environment.[68]

Presence of Robust Structural Elements in Two G-Quadruplex Conformations. The robust (3 + 1) core of Form 1 is reinforced by an A·T reverse Hoogsteen base pair on the top,

(66) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482–5515.

(67) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027.

(68) Zimmerman, S. B.; Trach, S. O. *J. Mol. Biol.* **1991**, *222*, 599–620.

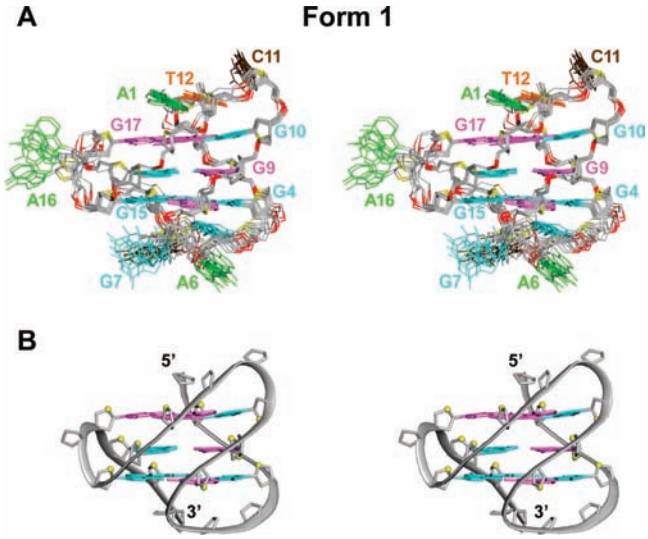

Figure 8. Stereoviews of the *GTERT-060[I5]* (3 + 1) G-quadruplex structure (Form 1) in K^+ solution. (A) Ten superimposed refined structures. (B) Ribbon view of a representative structure. *anti* and *syn* guanines are colored cyan and magenta, respectively; cytosines, brown; adenines, green; thymines, orange; inosines, cyan; backbone and sugar, gray; O4′ atoms, yellow; phosphorus atoms, red.

Table 3. Statistics of the Computed Structures of the 20-nt hTERT Promoter Modified Sequence d[AGGGIAGGGGCTGGGAGGGC] (Form 1)[a]

	D_2O	H_2O
A. NMR Restraints		
distance restraints		
intraresidue distance restraints	329	0
sequential (i, i + 1) distance restraints	106	24
long-range (i, $\geq i$ + 2) distance restraints	14	35
other restraints		
hydrogen bond restraints	52	
dihedral restraints	32	
repulsive restraints[b]	13	
B. Structure Statistics for 10 Molecules Following Distance-Restrained Molecular Dynamics Refinement		
NOE violations		
number (>0.2 Å)	0	
maximum violation (Å)	0.140 ± 0.026	
rmsd of violations (Å)	0.014 ± 0.000	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.692 ± 0.008	
impropers (deg)	0.362 ± 0.005	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except I5, A6, G7, A16, and C20	0.64 ± 0.12	
all heavy atoms	1.24 ± 0.23	

[a] PDB ID: 2KZD. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks.

as well as a single-residue double-chain-reversal loop (Figure 8), a well-known stable structural motif.[38,49,64,65] The (3 + 1) G-quadruplex core topology was first identified in four-repeat *Tetrahymena* sequence in Na^+ solution[69] and subsequently observed under various sequence contexts, including the human telomeric repeats[18−22,24,28] and the *bcl-2* promoter[45] (Supporting Information Figure S11), demonstrating itself to be a robust scaffold. A new feature of the *GTERT-060* (3 + 1) G-quadruplex is the edgewise three-residue loop GAG (or IAG in *GTERT-060[I5]*) that spans the wide groove and involves all purines.

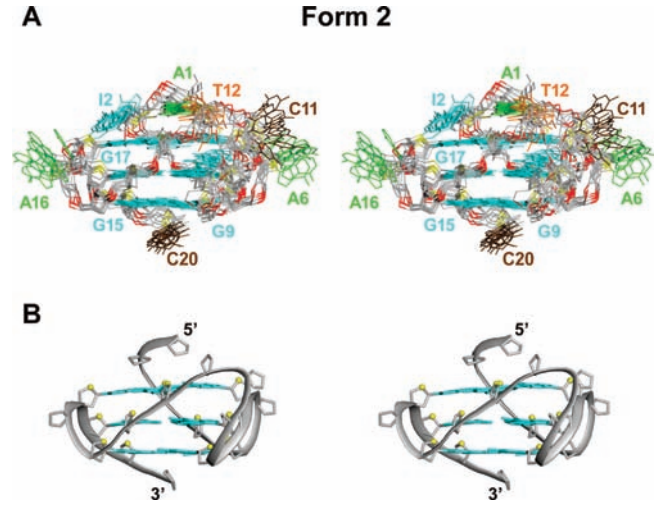

Figure 9. Stereoviews of the *GTERT-060[I2/I10]* parallel-stranded G-quadruplex structure (Form 2) in K^+ solution. (A) Ten superimposed refined structures. (B) Ribbon view of a representative structure. Color-coded as in Figure 8.

Table 4. Statistics of the Computed Structures of the 20-nt hTERT Promoter Modified Sequence d[AIGGGAGGGICTGGGAGGGC] (Form 2)[a]

	D_2O	H_2O
A. NMR Restraints		
distance restraints		
intraresidue distance restraints	310	0
sequential (i, i + 1) distance restraints	99	4
long-range (i, $\geq i$ + 2) distance restraints	10	47
other restraints		
hydrogen bond restraints	48	
dihedral restraints	32	
repulsive restraints[b]	11	
B. Structure Statistics for 10 Molecules Following Distance-Restrained Molecular Dynamics Refinement		
NOE violations		
number (>0.2 Å)	0	
maximum violation (Å)	0.127 ± 0.016	
rmsd of violations (Å)	0.015 ± 0.001	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.003 ± 0.000	
bond angles (deg)	0.674 ± 0.008	
impropers (deg)	0.345 ± 0.007	
pairwise all heavy atom rmsd values (Å)		
all heavy atoms except A6, I10, C11, T12, A16, and C20	0.95 ± 0.15	
all heavy atoms	1.54 ± 0.23	

[a] PDB ID: 2KZE. [b] Distance restraints between pairs of protons that do not exhibit NOE cross-peaks.

On the other hand, the propeller-type parallel-stranded G-quadruplex (Form 2) of *GTERT-060* is stabilized by the incorporation of two robust single-residue double-chain-reversal loops. The two loops fit snugly across the three layers of G-tetrads at opposite sides, giving rise to a compact core. The three-residue double-chain-reversal loop GCT (or ICT in *GTERT-060[I2/I10]*) makes additional contacts with the bases of the core, and possibly the terminal base A1 as well, providing further support to the structure. The configuration of this loop somewhat resembles that of the TTA propeller loop in the parallel G-quadruplex formed by a four-repeat human telomeric sequence in a K^+-containing crystal,[17] albeit the latter loop is more extended possibly due to crystal packing effect.

Many G-rich sequences in the genomes, particularly those in the promoters of oncogenes, contain G-tracts separated by

(69) Wang, Y.; Patel, D. J. *Structure* **1994**, *2*, 1141–1156.

short linkers of 1 or 2 nt, which have a penchant for double-chain-reversal conformation.[38,49,64,65] It is generally thought that this would favor a propeller-type G-quadruplex conformation. Despite the presence of such short loop motifs, there is still a considerable population of *GTERT-060* (>40%) adopting the (3 + 1) G-quadruplex conformation (Form 1). Hence, it seems likely that the predominant form is driven by the establishment of stabilizing interactions within the loop.[25,26,63] Of the two major G-quadruplex conformations observed for this sequence, the (3 + 1) conformer would present more asymmetric surfaces for recognition, making it attractive as a selective target.

Participation of Different Guanines in the G-Tetrad Core: Favoring a Single G-Quadruplex Conformation by Inosine Substitutions. The presence of a G-tract comprising more than three guanines, as seen in the case of *GTERT-060*, leads to variability during the folding of a three-layered G-quadruplex, since two or more sets of three consecutive guanines from this G-tract could potentially take part in G-tetrad formation. Here we employed guanine-to-inosine substitution[40,63,69,70] (which is a softer modification than a G-to-T substitution)[38,39,45] as a strategy to selectively disfavor the guanine at that particular position from forming G-tetrad,[63] thereby driving the equilibrium toward the competing conformation. Specifically, we have demonstrated that inosine substitution at Position 5 and/or Position 7 of *GTERT-060* favored Form 1, whereas inosine replacement at Position 2 and/or Position 10 favored Form 2 (Figure 1 and Table 2).

CD Spectra Deconvolution of a Mixture of G-Quadruplexes. The indistinct shoulder at ~295 nm of the CD spectrum of *GTERT-060* (Figure 1J), together with the fact that this sequence contains G-tracts separated by short linkers of 1 or 2 nt, would have suggested a parallel fold, since there is a strong peak at 260 nm coupled with a negative valley at 240 nm, characteristic of parallel G-quadruplexes.[61] However, we have shown using both NMR and CD that *GTERT-060* adopts the (3 + 1) scaffold (Form 1), in addition to the parallel configuration (Form 2), in K^+ solution, and that we could selectively favor each form over the other by a judicious guanine-to-inosine substitution (Figure 1 and Table 2). We have also demonstrated that quantification of the relative population of each form in the natural sequence based on CD is possible (Figure 1J; Supporting Information Text and Figure S5), on the premise that the contribution arising from individual species could be isolated. The possible coexistence of multiple G-quadruplex conformations within a sequence would demand extra caution when applying only CD to study the topologies.

The populations of the two forms of *GTERT-060* vary as a function of temperature and can be monitored by CD and NMR spectra (Figure 3; Supporting Information Figures S6 and S7). A combination of different experimental conditions (temperature, pH, crowding agent, etc.) and sequence modifications can be applied to vary the relative populations of different forms, aiding structural interpretations. The availability of the CD spectra of the individual species allowed a direct characterization of the interconversion between the two forms.

The CD signatures (a positive peak at 260 nm for Form 2; two positive peaks at 260 and 295 nm for Form 1) were similar to those of other parallel-stranded and (3 + 1) G-quadruplexes, respectively.[18,19,25,27,61] However, for Form 1 the peak at 260 nm is higher than that at 295 nm, while the reverse scenario was previously reported for other (3 + 1) G-quadruplexes.[18,19,25,27] This might reflect the varying degrees of the effects of different loop and terminal base arrangements, e.g. base stacking within the loops and between the loops and the core, in addition to the contribution of the (3 + 1) G-tetrad core, on the CD spectra.

Related Studies on G-Quadruplex Formation in the hTERT Promoter. Putative G-quadruplex formation in the promoter of hTERT has been independently reported by two other groups.[34,35] Using biochemical and spectroscopic approaches, nine sequences from −42 to −109 nt relative to the TSS containing four consecutive G-tracts were examined.[34] This may be compared with the current work, in which eleven sequences from −20 to −110 nt relative to the TSS were examined with a concomitant structural analysis on a particular G-quadruplex-forming fragment. Based on in vitro assays, two DNA polymerase stop sites in the hTERT promoter were reported,[34,35] corresponding to the 3′ ends of the *GTERT-060* and *GTERT-110* sequences, which formed the most stable G-quadruplexes detected in the present work (Table 1). In particular, Palumbo et al.[34] proposed that the 5′-abridged *GTERT-060* sequence forms a parallel-stranded G-quadruplex in a long sequence context, stacked alongside another G-quadruplex formed by remote G-tracts connected by a 26-nt duplex stem. Our data would agree with the formation of a parallel-stranded G-quadruplex (Form 2) by *GTERT-060*, especially since the abridged sequence lacks the terminal base A1 which is essential for the stabilizing effect on Form 1 (Supporting Information Figure S12). In addition, here we could show in detail the coexistence and structural interconversion of G-quadruplexes in the *GTERT-060* fragment. Nevertheless, one should note that even if a particular G-quadruplex is less abundant in the long natural context, it could still be useful so long as it can be stabilized in a pharmacological context.

Biological Significance and Implications for Targeting Telomerase. Numerous quadruplex targets have been identified in the telomeric pathway that could interfere with telomerase activity: (i) quadruplex formation within telomeric G-rich overhang has been shown to prevent its elongation by telomerase[15] or to compete with the binding of telomeric proteins,[71] (ii) quadruplex formation has been shown to negatively impact *c-myc* expression, which in turn affects hTERT expression level,[72] (iii) the splicing pattern of hTERT RNA can be affected by quadruplex formation, which would shift the equilibrium between active and inactive telomerase,[73] and (iv) a quadruplex ligand might bind to a quadruplex-forming motif in the telomerase RNA component (hTR or hTERC)[74] and alter telomerase activity. The present work, together with recent studies,[34,35] highlight potential targets for quadruplex ligand within the hTERT promoter. The 91-nt promoter segment, positioned between −20 and −110 nt relative to the transcription start site, encompasses four Sp1 and one Ets binding sites, representing a potential regulator of hTERT transcription. In particular, our study has focused on the 20-nt fragment (*GTERT-060*) containing an Sp1 binding site and located just upstream of the Ets binding site.

(70) Smith, F. W.; Feigon, J. *Biochemistry* **1993**, *32*, 8682–8692.
(71) Gomez, D.; O'Donohue, M. F.; Wenner, T.; Douarre, C.; Macadré, J.; Koebel, P.; Giraud-Panis, M. J.; Kaplan, H.; Kolkes, A.; Shin-ya, K.; Riou, J. F. *Cancer Res.* **2006**, *66*, 6908–6912.
(72) Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. *Mol. Cancer Ther.* **2002**, *1*, 565–573.
(73) Gomez, D.; Lemarteleur, T.; Lacroix, L.; Mailliet, P.; Mergny, J. L.; Riou, J. F. *Nucleic Acids Res.* **2004**, *32*, 371–379.
(74) Gros, J.; Guédin, A.; Mergny, J. L.; Lacroix, L. *Chembiochem* **2008**, *9*, 2075–2079.

The complementary C-rich strand of *GTERT-060* is also prone to the formation of the i-motif,[75] as indicated by our preliminary results (data not shown). The formation of secondary structures on both strands of this promoter might take part in the inherent regulation of its transcription, and/or it could provide a targeting platform for structure-based transcription-modulating drugs.

Conclusion

The hTERT promoter, which contains many G-tracts on the same DNA strand, exhibits an exceptional potential for G-quadruplex formation. One particular G-rich sequence in this region, *GTERT-060*, coexists in two G-quadruplex conformations in K^+ solution: a (3 + 1) and a parallel-stranded G-quadruplexes. A combination of NMR and CD techniques, complemented with sequence modifications and variations of experimental condition, facilitated a dissection of the equilibrium and interconversion between the two forms.

Acknowledgment. This research was supported by Singapore Biomedical Research Council grant 07/1/22/19/542 to A.T.P., Singapore Ministry of Education grants (ARC30/07 and RG62/07) to A.T.P., INSERM, CNRS, and the Muséum National d'Histoire Naturelle to L.L. We thank Jonathan Aw Zhan Quan and Mahendran Balaji from A.T.P.'s laboratory, and Aurelie Bonfils from L.L.'s laboratory for their participation at the early stage of the project, Jean-Louis Mergny for helpful discussions, and Maurice Guéron for critical reading of the manuscript.

Supporting Information Available: Additional experimental procedures and data (Figures S1−S12 and Tables S1−S3) and complete ref 59. This material is available free of charge via the Internet at http://pubs.acs.org.

JA101252N

(75) Gehring, K.; Leroy, J. L.; Guéron, M. *Nature* **1993**, *363*, 561–565.