J|A|C|S ARTICLES

Published on Web 01/30/2009

Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K^+ Solution

Herry Martadinata and Anh Tuân Phan*

Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371

Received August 19, 2008; E-mail: phantuan@ntu.edu.sg

Abstract: Very recent studies showed that mammalian telomeres were transcribed into telomeric-repeat-containing RNAs and suggested that these RNA molecules were biologically important. Here we report on a structural study of RNA G-quadruplexes formed by human telomeric RNA sequences in K^+ solution. Our data indicated that these sequences formed propeller-type parallel-stranded RNA G-quadruplexes. We have determined the NMR-based solution structure of a dimeric propeller-type RNA G-quadruplex formed by the 12-nt human telomeric RNA sequence r(UAGGGUUAGGGU). We also observed the stacking of two such propeller-type G-quadruplex blocks for the 10-nt human telomeric RNA sequence r(GGGUUAGGGU) and a higher-order G-quadruplex structure for the 9-nt human telomeric RNA sequence r(GGGUUAGGG). Based on these findings we proposed how higher-order structures might be formed by long telomeric RNA.

1. Introduction

The ends of eukaryotic chromosomes are called telomeres. Human telomeric DNA contains thousands of repeats of the sequence d(TTAGGG). This DNA G-rich strand can form four-stranded G-quadruplex structures[1] based on stacking of G•G•G•G tetrads.[2] Such DNA structures are potential anticancer targets.[3] Human telomeric DNA sequences have been shown to form a variety of G-quadruplex topologies in different experimental conditions.[4] The four-repeat human telomeric d[AGGG(TTAGGG)$_3$] sequence forms an intramolecular basket-type antiparallel-stranded G-quadruplex in Na^+ solution.[4b] In this structure, four GGG segments form the G-tetrad core involving three stacked G-tetrads with *syn•syn•anti•anti* glycosidic conformations; each strand has both parallel and antiparallel adjacent strands; three connecting TTA linkers form successively edgewise, diagonal and edgewise loops. The same DNA sequence forms in a K^+-containing crystal a completely different intramolecular propeller-type parallel-stranded G-quadruplex.[4c] In this structure, all four strands are parallel, the connecting TTA loops are double-chain-reversal, and all guanines adopt *anti* glycosidic conformations. In K^+ solution, two-repeat human telomeric sequences can adopt both parallel- and antiparallel-stranded G-quadruplexes;[4d] the parallel-stranded form[4d] is the same as the propeller-type G-quadruplex topology observed in the crystalline state.[4c] It has been reported that molecular crowding conditions can favor parallel-stranded G-quadruplex conformation(s).[4l] Recently, four-repeat human telomeric sequences have been observed to form in K^+ solution two different intramolecular (3 + 1) G-quadruplexes,[4h−k] whose populations depend on the flanking sequences. In these structures, the (3 + 1) core contains three strands oriented in one direction and the fourth in the opposite direction; the glycosidic conformations of G-tetrads are *syn•anti•anti•anti* and *anti•syn•syn•syn*. Both structures have one double-chain-reversal and two edgewise TTA loops, but they differ in the successive order of loop arrangements.

Very recent studies[5] showed that mammalian telomeres were transcribed by DNA-dependent RNA polymerase II. RNA molecules, ranging from 100 to 9000 nt and containing telomeric

(1) (a) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730. (b) Gilbert, D. E.; Feigon, J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 305–314. (c) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621–628. (d) Davis, J. T. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 668–698. (e) Phan, A. T.; Kuryavyi, V.; Patel, D. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 288–298. (f) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415. (g) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429–7455.

(2) Gellert, M. N.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.

(3) (a) Sun, D.; 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. (b) Mergny, J. L.; Hélène, C. *Nat. Med.* **1998**, *4*, 1366–1367.

(4) (a) Wang, Y.; Patel, D. J. *Biochemistry* **1992**, *31*, 8112–8119. (b) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282. (c) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880. (d) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027. (e) He, Y.; Neumann, R. D.; Panyutin, I. G. *Nucleic Acids Res.* **2004**, *32*, 5359–5367. (f) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649–4659. (g) Zhang, N.; Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 17277–17285. (h) Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591. (i) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735. (j) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. (k) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–5719. (l) Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185–11191.

(5) (a) Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. *Science* **2007**, *318*, 798–801. (b) Schoeftner, S.; Blasco, M. A. *Nat. Cell Biol.* **2008**, *10*, 228–236. (c) Horard, B.; Gilson, E. *Nat. Cell Biol.* **2008**, *10*, 113–115.

10.1021/ja806592z CCC: $40.75

Table 1. Natural Human Telomeric RNA Sequences Used in This Study

	Residue number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sequences	12-nt	U	A	G	G	G	U	U	A	G	G	G	U											
	10-nt			G	G	G	U	U	A	G	G	G	U											
	9-nt			G	G	G	U	U	A	G	G	G												
	23-nt	U	A	G	G	G	U	U	A	G	G	G	U	U	A	G	G	G	U	U	A	G	G	G

repeats r(UUAGGG), were found in nuclear fractions. Data suggested that telomeric RNAs were associated with the chromosome ends; their levels were regulated by different factors and dependent on cellular developmental stages.[5] It appears that much of the telomeric RNA biology remains to be uncovered.[5c] As these RNA sequences are rich in guanines, they can form G-quadruplex structures, which might be important for telomere biology and good targets for anticancer drug design. The biological relevance of RNA G-quadruplexes has been demonstrated for G-rich sequences in different mRNAs.[6] However, so far in contrast to the situation for DNA G-quadruplexes,[1] very few structural studies on RNA G-quadruplexes have been performed[7] and knowledge about structural diversity or conservation of RNA G-quadruplexes is limited.

Here we report on a structural study of RNA G-quadruplexes formed by human telomeric RNA sequences in the physiological K^+-containing solution. Our data indicated that these sequences formed propeller-type parallel-stranded RNA G-quadruplexes. We have determined an NMR-based solution structure of a propeller-type RNA G-quadruplex formed by the 12-nt human telomeric RNA sequence r(UAGGGUUAGGGU). We have also established the stacking of two such propeller-type G-quadruplex units for the 10-nt human telomeric RNA sequence r(GGGUUAGGGU) and observed a higher-order G-quadruplex structure for the 9-nt human telomeric RNA sequence r(GGGUUAGGG). Based on these findings we proposed how higher-order structures might be formed by long telomeric RNA.

2. Results and Discussion

2.1. G-Quadruplex Formation by Two- And Four-Repeat Human Telomeric RNA Sequences. In this work we examined two-repeat and four-repeat human telomeric RNA sequences (Table 1). Imino proton NMR spectra of the two-repeat (Figure 1 and Figure S1, Supporting Information) and four-repeat (Figure S2) sequences in K^+ solution showed peaks at 10.5−11.8 ppm, indicating formation of G-quadruplexes. CD spectra of these sequences showed a positive band at 260 nm and a negative band at 240 nm (Figure 2), characteristic of parallel-stranded G-quadruplex conformation(s).[8]

We further focused our efforts on the two-repeat (12-nt, 10-nt, and 9-nt) human telomeric RNA sequences (Table 1) due to their higher-quality NMR spectra. To probe their molecular sizes, we performed gel electrophoresis experiments using DNA oligonucleotides dT_{12} and dT_{24} as references (Figure 3). All the tested RNAs migrated slower than dT_{12}, indicating formation of multimeric structures. The band corresponding to the 12-nt r(UAGGGUUAGGGU) sequence migrated faster than those of the 10-nt r(GGGUUAGGGU) and 9-nt r(GGGUUAGGG) sequences, indicating that the structure of the 12-nt RNA has a lower stoichiometry than that of the 10-nt and 9-nt RNAs. The 9-nt RNA migrated as two bands, which were respectively faster and slower than the 10-nt RNA. The faster-migrating one might correspond to the same stoichiometry as that of the 10-nt RNA, while the slower-migrating one must correspond to a higher stoichiometry. Taken all together, gel electrophoresis data were consistent with the 12-nt and 10-nt RNAs forming a dimeric and a tetrameric G-quadruplex, respectively, while the 9-nt RNA formed a tetrameric G-quadruplex and a higher-order structure. This conclusion was supported by further NMR structural analysis (see below).

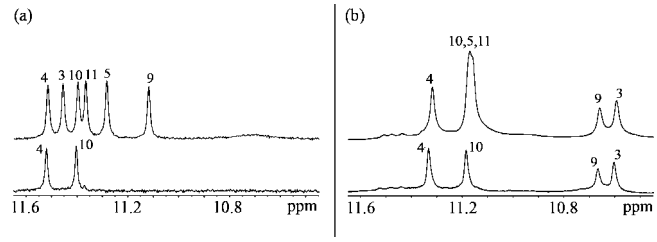

Figure 1. Imino proton NMR spectra of two-repeat human telomeric RNA sequences in K^+ solution: (a) 12-nt r(UAGGGUUAGGGU) and (b) 10-nt r(GGGUUAGGGU). Top spectra were recoded in H_2O; bottom spectra were recorded after 3 h in D_2O. Experimental conditions: temperature, 25 °C; RNA concentration, 1.25 mM.

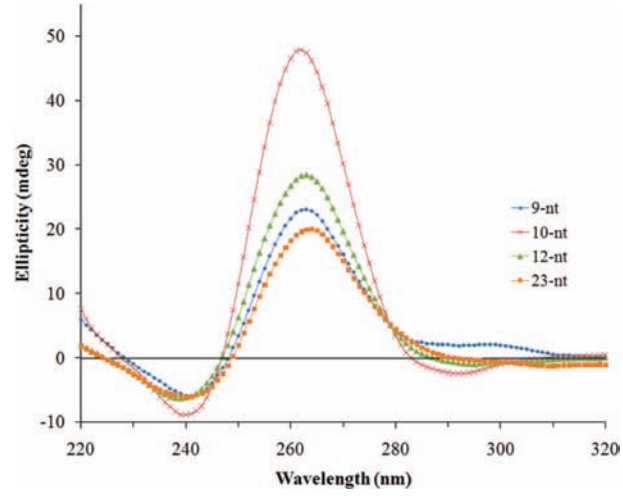

Figure 2. CD spectra of the 9-nt, 10-nt, 12-nt, and 23-nt human telomeric RNA sequences (Table 1) in K^+ solution. Experimental conditions: temperature, 25 °C; RNA concentrations, 5−8 μM.

(6) (a) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nat. Chem. Biol.* **2007**, *3*, 218–221. (b) Wieland, M.; Hartig, J. S. *Chem. Biol.* **2007**, *14*, 757–763. (c) Arora, A.; Dutkiewicz, M.; Scaria, V.; Hariharan, M.; Maiti, S.; Kurreck, J. *RNA* **2008**, *14*, 1290–1296.
(7) (a) Cheong, C.; Moore, P. B. *Biochemistry* **1992**, *31*, 8406–8414. (b) Deng, J.; Xiong, Y.; Sundaralingam, M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13665–13670. (c) Liu, H.; Matsugami, A.; Katahira, M.; Uesugi, S. *J. Mol. Biol.* **2002**, *322*, 955–970. (d) Pan, B.; Shi, K.; Sundaralingam, M. *J. Mol. Biol.* **2006**, *363*, 451–459.
(8) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.* **1992**, *20*, 4061–4067.

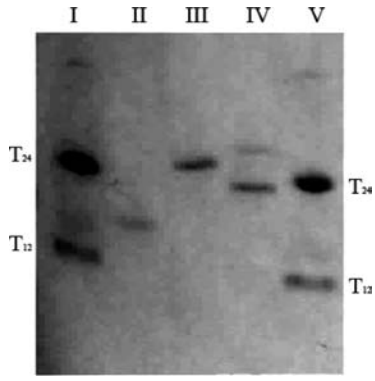

Figure 3. Gel electrophoresis monitored by UV shadowing method. Lanes I and V, markers dT_{12} and dT_{24}; Lanes II, III, and IV are for the 12-nt, 10-nt, and 9-nt human telomeric RNA sequences, respectively. RNA concentrations: 12-nt, 50 μM; 10-nt, 80 μM; 9-nt, 50 μM.

2.2. NMR Spectral Assignment of 12-nt and 10-nt Human Telomeric RNA Sequences. The number and intensity of NMR peaks observed for the 12-nt and 10-nt human telomeric RNA sequences in K^+ solution (Figure 1 and Figure S3) indicated formation of symmetric structures. For each sequence, six guanine imino protons were observed. As expected for G-quadruplexes, they were protected from the exchange with solvent, and some of them remained after the samples were dissolved in D_2O (Figure 1). To facilitate the comparison between the two sequences, residues in the 10-nt RNA were numbered according to the 12-nt RNA (Table 1).

Nonexchangeable base and sugar protons of the 12-nt and 10-nt RNA sequences could be classified and sorted by residues using a combination of homonuclear NOESY, COSY and TOCSY experiments, as well as heteronuclear [$^{13}C-^{1}H$] HSQC experiments at natural abundance.[9] Sequential connectivities for the H6/8-H1′ region could be traced in NOESY spectra (Figure 4). Many similarities in spectral NOE patterns were observed for the 12-nt and 10-nt RNAs (Figure 4). Spectral analysis of sequences containing single site-specific dG-for-rG and dT-for-rU substitutions (Table 2 and Figure S4) independently confirmed the assignments. NOE patterns indicated that these single deoxyribose-for-ribose substitutions (Table 2) did not perturb the general G-quadruplex fold; H2′ and H2″ protons of the substituted residue were upfield-shifted and could be easily recognized (at ~2 ppm) (Figure 5). We note that this site-specific deoxyribose-for-ribose substitution approach is an attractive method for unambiguous NMR spectral assignments of RNA given that isotopic labeled ribo-phosphoramides are not yet available at moderate cost.[10]

Within each individual guanine, through-bond correlation between H8 and imino protons via $^{13}C5$ at natural abundance[9] provided the imino proton assignments (Figure 6).

2.3. G-Quadruplex Folding and Stacking. The G-tetrad alignments in the 12-nt and 10-nt human telomeric RNAs were identified from NOESY spectra (Figure 7a and b) on the basis of the characteristic imino-H8 connectivity pattern around a G-tetrad (Figure 7c). In both cases, data revealed the formation of three G-tetrads: G3•G9•G3•G9, G4•G10•G4•G10, and G5•G11•G5•G11 (Figure 7d).

(9) (a) Phan, A. T. *J. Biomol. NMR* **2000**, *16*, 175–178. (b) Phan, A. T.; Guéron, M.; Leroy, J. L. *Methods Enzymol.* **2001**, *338*, 341–371.
(10) (a) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161. (b) Phan, A. T.; Patel, D. J. *J. Biomol. NMR* **2002**, *23*, 257–262.

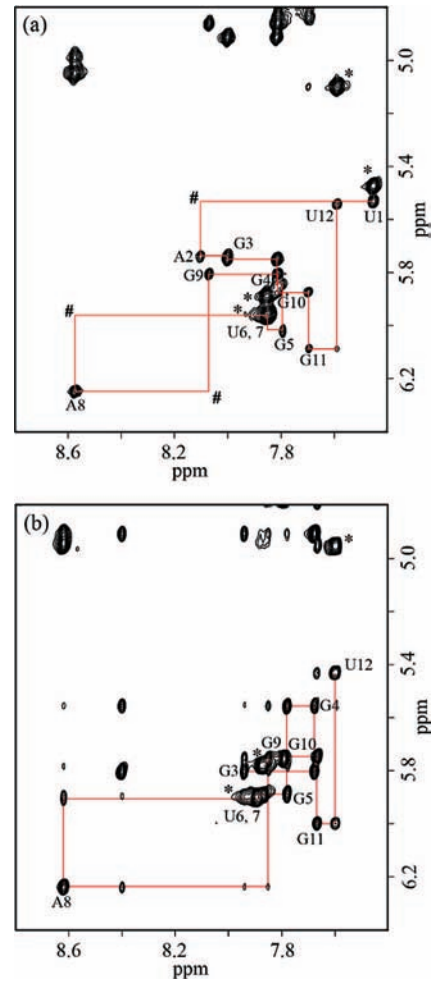

Figure 4. NOESY spectra in D_2O (mixing time, 300 ms), showing the H8/H6−H1′ connectivities of (a) the 12-nt and (b) 10-nt human telomeric RNA sequences in K^+ solution. Intraresidue H6/H8−H1′ NOE cross-peaks are labeled with residue numbers. Intraresidue H5−H6 peaks of U are labeled with a star. Peaks labeled (#) were detected at a lower contour threshold.

Table 2. Modified RNA Sequences Containing a Single Deoxyribose Substitution Used in This Study

	Residue number	1	2	3	4	5	6	7	8	9	10	11	12
Sequences	12-nt(d5)	U	A	G	G	**dG**	U	U	A	G	G	G	U
	12-nt(d6)	U	A	G	G	G	**dT**	U	A	G	G	G	U
	12-nt(d7)	U	A	G	G	G	U	**dT**	A	G	G	G	U
	12-nt(d9)	U	A	G	G	G	U	U	A	**dG**	G	G	U
	12-nt(d10)	U	A	G	G	G	U	U	A	G	**dG**	G	U

Data for the 12-nt human telomeric RNA were consistent with formation of a dimeric propeller-type G-quadruplex involving three G-tetrad layers (Figure 8a). The G-tetrad core is parallel-stranded with all guanines adopting an *anti* glycosidic conformation, consistent with the H8−H1′ NOE cross-peak intensity observed for all guanines (Figure 4a). Two UUA loops adopt the double-chain-reversal configuration, located at the groove sides of the G-quadruplex, connecting two adjacent parallel strands and bridging three G-tetrad layers. This G-quadruplex fold (Figure 8a) was supported by many other proton−proton distances measured in NOESY experiments (see solution structure below). This structure was also supported by the deuterium exchange experiment which showed that only the imino protons from the central G-tetrad (G4 and G10) were

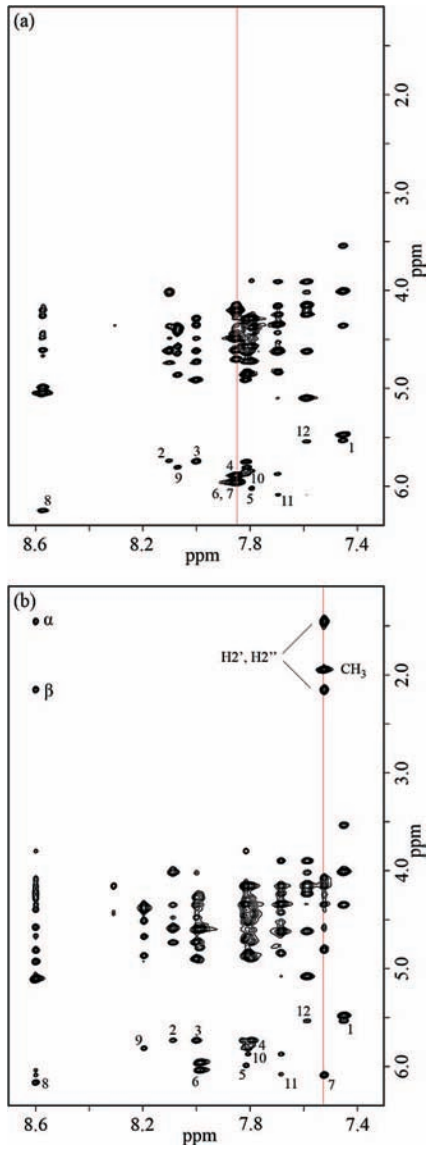

Figure 5. Example for the identification of the deoxyribose-substituted residue by upfield-shifted H2′ and H2″ protons. NOESY spectra in D_2O of (a) the 12-nt human telomeric RNA sequence and (b) 12-nt modified sequence containing dT-for-rU substitution at position 7. Vertical lines were traced on the frequency of U7(H6) and T7(H6), respectively. α/β correspond to cross-peaks between A8(H8) and T7(H2′/H2″).

well protected from the exchange with solvent: after 3 h in D_2O solvent, these imino protons were still detected, while those from the two remaining G-tetrads disappeared (Figure 1a).

Data for the 10-nt human telomeric RNA were consistent with the 5′-end stacking (Figure 8b) of two G-quadruplex blocks, each of which was similar to the dimeric propeller-type G-quadruplex described above for the 12-nt RNA (Figure 8a). The formation of this tetrameric G-quadruplex was supported by (i) the gel shift data showing that the 10-nt RNA migrated slower than the 12-nt RNA (Figure 3), (ii) the deuterium exchange data showing that the imino protons of G3 and G9 were now highly protected from the exchange with solvent (Figure 1b) but still less well protected than the imino protons of G4 and G10 (Figure S5), (iii) the observation of the upfield shifts for protons of G3 and G9 (Figure 1) at the proposed stacking interface (Figure 8b), and (iv) the observation of several NOE cross-peaks across the interface consistent with the stacking of G9 and G3 on top of each other (Figures 7, 9, and S6; see figure legends).

At low concentrations (~0.07 mM), NMR spectra of the 9-nt human telomeric RNA sequence showed a set of major peaks and a set of minor peaks (Figure S1). The spectral features of the former, including the chemical shift and solvent exchange properties, were similar to those of the 10-nt RNA (at 0.07−1.25 mM) (Figures 1, S1, S3, and S7), suggesting formation of the same major G-quadruplex structure. However, at higher concentrations of the 9-nt RNA (0.36−1.25 mM) the intensity of the latter became comparable to that of the former (Figures S7 and S8). These observations were consistent with formation of higher-order structure(s) for the 9-nt RNA as seen in the gel electrophoresis experiments (Figure 3). We note that the resonances of the 10-nt and 12-nt RNAs were broadened when the sample concentration exceeded 2.0 mM, also suggesting formation of aggregations or higher-order structures.

2.4. Sugar Puckering Conformations. Both C2′-*endo* and C3′-*endo* sugar puckering conformations were observed in these RNA G-quadruplexes. For the 12-nt RNA, the COSY spectrum (Figure 10a) showed large *J*-couplings between H1′ and H2′ for loop residues (U6, U7, and A8), terminal residues (U1 and U12), and some core residues (G3 and G5), indicating their C2′-*endo* sugar conformations.[11] Except for G9 (showing a weak H1′−H2′ COSY peak), the absence of H1′−H2′ COSY peaks for the remaining core residues (Figure 10a) indicated their C3′-*endo* conformations.[11] For the 10-nt RNA, strong H1′−H2′ COSY peaks (corresponding to C2′-*endo* conformations) were also observed for all loop (U6, U7 and A8) and terminal (U12) residues, but only for G3 from the core (Figure 10b). The observation of both C2′-*endo* and C3′-*endo* sugar puckering was previously also reported for NMR and X-ray structures of different RNA quadruplexes.[7]

2.5. K^+ vs Na^+: Structural Conservation of RNA G-Quadruplexes. When this manuscript was in preparation, a paper of Xu et al.[12] appeared online and reported on the same 12-nt RNA sequence forming a propeller-type parallel-stranded G-quadruplex in Na^+ solution. The conservation of G-quadruplex folding topology for this human telomeric RNA sequence in Na^+ solution[12] and in K^+ solution (this work) contrasts with the results reported for human telomeric DNA sequences, for which various G-quadruplex topologies[4] were observed in the presence of Na^+ and K^+. This structural conservation might be related to the preference of RNA residues to adopt an *anti* glycosidic conformation.[13,14]

The stability of the 12-nt RNA G-quadruplex in K^+ and Na^+ solution was characterized by UV melting experiments (Figure S9). Cations, such as K^+ and Na^+, stabilize G-quadruplexes by coordinating to the channel in the middle of the G-tetrad core.[15] Although the same RNA G-quadruplex topology was formed in both K^+ and Na^+ solution, the structure in K^+ solution was significantly more stable (with T_m being 10 to 20 °C higher) than that in Na^+ solution (Figure S9). Note that the structure in

(11) de Leeuw, F. A. A. M.; Altona, C. *J. Chem. Soc., Perkin Trans.* **1982**, *2*, 375–384.
(12) Xu, Y.; Kaminaga, K.; Komiyama, M. *J. Am. Chem. Soc.* **2008**, *130*, 11179–11184.
(13) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, NY, 1984.
(14) Qi, J.; Shafer, R. H. *Biochemistry* **2007**, *46*, 7599–7606.
(15) Hud, N. V.; Plavec, J. , In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; Royal Society of Chemistry: Cambridge, UK, 2006; pp 100−126.

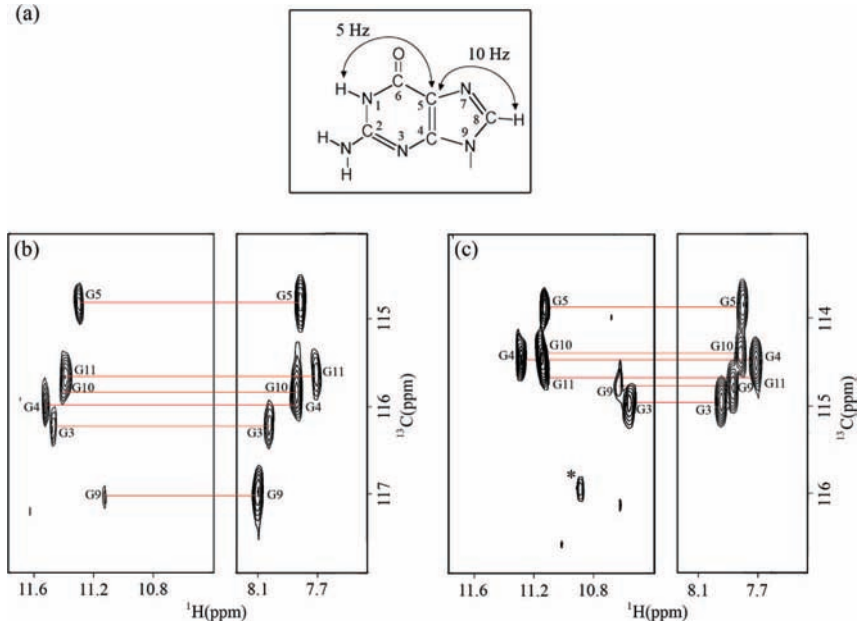

Figure 6. (a) Long-range *J*-couplings between imino and H8 protons via ^{13}C5. (b, c) Through-bond correlations between imino and H8 protons via ^{13}C5 at natural abundance for the (b) 12-nt and (c) 10-nt human telomeric RNA sequences, using long-range *J*-couplings as shown in (a). A peak from a minor species is labeled with a star.

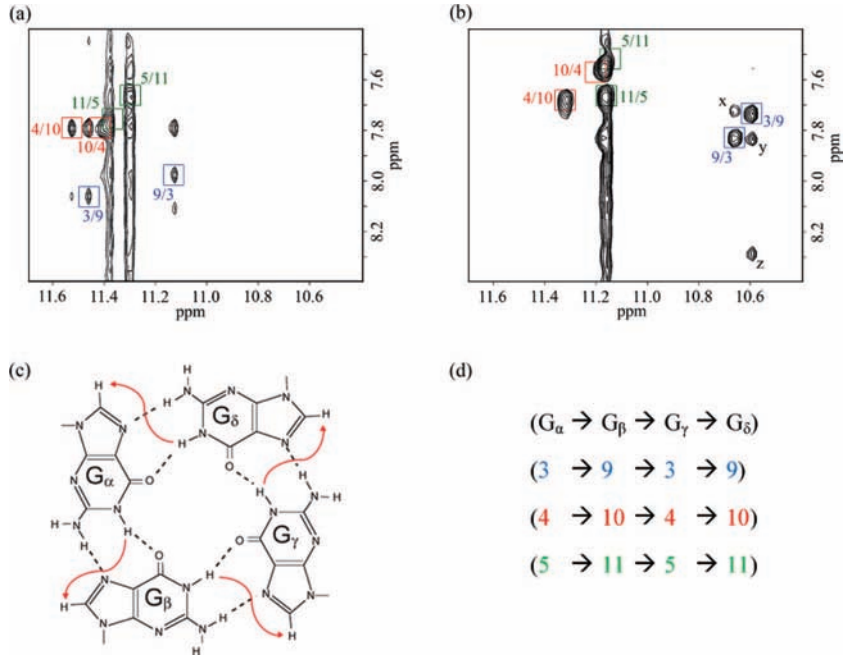

Figure 7. Determination of G-quadruplex folding topology for the human telomeric RNA sequence in K^+ solution. (a, b) NOESY spectra in H_2O (mixing time, 200 ms) of the (a) 12-nt and (b) 10-nt human telomeric RNA sequences. Cross-peaks that identify the three G-tetrads are framed and labeled with the residue number of imino proton in the first position and that of H8 in the second position. (c) Characteristic guanine imino-H8 NOE connectivity patterns around a G_α•G_β•G_γ•G_δ tetrad as indicated with arrows (connectivity between G_δ and Gα implied). (d) Guanine imino-H8 NOE connectivities observed for G3•G9•G3•G9 (blue), G4•G10•G4•G10 (red), and G5•G11•G5•G11 (green) tetrads. Letters correspond to the following peaks: x, G9(H1)−G9(H8); y, G3(H1)−G3(H8); *z*, G3(H1)−A8(H2).

K^+ solution also exhibited stronger cation concentration dependence than that in Na^+ solution (Figure S9).

2.6. Solution Structure of a Human Telomeric RNA G-Quadruplex. The structure of the G-quadruplex formed by the 12-nt human telomeric RNA sequence (Figure 11) was calculated on the basis of NMR restraints (Table 3) by using the X-PLOR program.[16] Figure 11a displayed 10 superpositioned lowest-energy refined structures. The structure of the core is better defined than that of the loop and terminal residues. The parallel-stranded G-tetrad core is defined with four medium-size grooves, two of which are occupied by the U6U7A8 loops. Several arrangements of the residues in the UUA double-chain-reversal loops were observed, reflecting the dynamics of this region. In general, U7 and A8 were partially stacked and pointed toward one direction, while U6 pointed to another direction (Figure 11). A8 was located closer to the G-tetrad core while U6 and U7 pointed outward, consistent with the observation of many NOEs between A8 and G4/G5. Previously, various

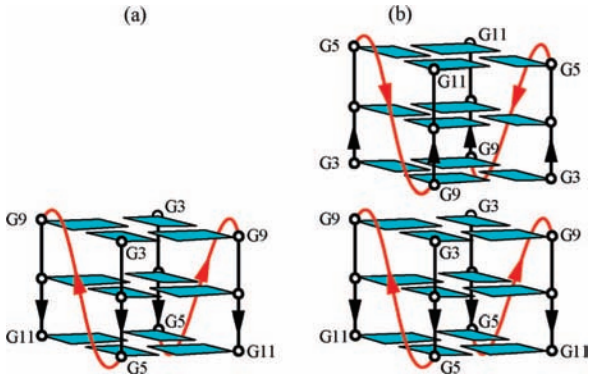

Figure 8. Schematic structure of the (a) 12-nt and (b) 10-nt human telomeric RNA sequences. *Anti* guanines are colored cyan; UUA linkers are colored red.

possible configurations[17,4c,j] have been reported for the double-chain-reversal TTA loop in G-quadruplexes, including a configuration similar to the one observed here.[4j] Terminal bases U1 and U12 stack on the top and the bottom of the G-tetrad core, respectively, consistent with the upfield shifts of U1(H5/6) and U12(H5/6) protons (Figure 4) and the observation of many NOEs between these residues and the adjacent G-tetrads. A2 is projected outward, consistent with the lack of NOEs between this base's protons and imino protons of the adjacent G-tetrad.

2.7. Structural Model for Long Telomeric RNA. We have shown that the 10-nt human telomeric RNA sequence r(GGGUUAGGGU) formed in K^+ solution a structure involving 5′-end stacking of two propeller-type three-layer G-quadruplex blocks. The lack of two residues UA at the 5′-end of the 10-nt RNA (as compared to the 12-nt RNA) might favor this stacking structure. Our data also indicated that a higher-order structure was formed by the 9-nt human telomeric RNA sequence r(GGGUUAGGG). The lack of U at the 3′-end might favor further stacking of G-quadruplexes at this end.

Based on these findings we proposed a higher-order structure (rod-like) formed by a long human telomeric RNA sequence involving both 5′-end and 3′-end stacking of multiple three-layer G-quadruplex blocks (Figure 12a). Each block is a propeller-type parallel-stranded three-layer G-quadruplex formed by the four-repeat human telomeric RNA sequence, which is similar to the crystal structure previously observed for the four-repeat human telomeric DNA sequence.[4c] Two consecutive blocks are connected by a UUA linker that spans four G-tetrad layers (shown in red, Figure 12a). Each block has one 5′-end/5′-end and one 3′-end/3′-end stacking interfaces with its two neighboring blocks. Four strands within each block are parallel and define the direction of the block, but the directions of different blocks alternate up−down along the rod (Figure 12a). We name this model for the human telomeric RNA "alternate-direction stacking" (or "up−down stacking") because it involves the alternation of directions of G-quadruplex blocks. This model can be compared with two other models previously proposed for human telomeric DNA, the "beads-on-a-string"[18] and the "same-direction stacking" (or "up−up stacking").[4c] Both models involve formation of three-layer G-quadruplex blocks. In the former model, the G-quadruplex blocks can move relatively independently from each other and are constrained only by the connecting TTA linkers.[18] In the latter model, G-quadruplex blocks stack continuously head-to-tail; blocks are oriented in the same direction.[4c] Our model is closely related to the latter model, but differs in three mutually related properties: (i) relative orientations of blocks (up−down vs up−up), (ii) stacking interfaces between blocks (5′−5′ and 3′−3′ vs 5′−3′), and (iii) configuration of connecting linkers between blocks (bridging four G-tetrad layers vs two G-tetrad layers). Regarding the last factor, it seems that a three-nucleotide linker can be better packed against the groove of the G-quadruplex when it bridges four G-tetrad layers then when it bridges only two layers. Previously, a linker spanning three G-tetrad layers to connect two G-quadruplex blocks was observed.[19]

To test the feasibility of our model, a structure of the twelve-repeat human telomeric RNA r[UAGGG(UUAGGG)$_{11}$] sequence (Figure 12b) was built by using the X-PLOR program[16] with artificial constraints derived from the 12-nt and 10-nt RNAs (see Methods). In this structure, the length of the UUA linkers connecting consecutive blocks (shown in red, Figure 12b) matched well with the connecting distances, thereby resulting in these linkers being nicely packed in the grooves (Figures 12b and S10). In addition, the arrangement of G-quadruplex blocks in our model resulted in a unique pattern of continuous groove progression (Figure 12). We note that in an attempt to compute 100 structures of the eight-repeat human telomeric RNA r[UAGGG(UUAGGG)$_7$] sequence (see Methods), a structure with the 5′−5′ stacking interface (Figure S11) was observed among the 20 lowest-energy structures, even when no constraints for the stacking between G-quadruplex blocks were imposed.

It is possible that all the three types of G-quadruplex arrangements described above ("beads-on-a-string",[18] "up−up stacking"[4c] and our "up−down stacking" model) (co)exist in the context of long telomeric RNA (or DNA), and their formation may depend on different interacting partners in the cell.

3. Conclusion

RNA human telomeric repeats form propeller-type parallel-stranded G-quadruplex structure in K^+ solution. Several blocks of three-layer propeller-type G-quadruplexes can stack on each other to form higher-order structures.

4. Methods

4.1. Sample Preparation. RNA and modified oligonucleotides were purchased from Research Instruments Pte Ltd. (Singapore). RNA sequences used in this study are shown in Tables 1 and 2. The samples were treated according to the prescribed protocol and dialyzed against water, 70 mM KCl solution, and water, successively. Unless otherwise stated, experiments were carried out in a buffer containing 20 mM potassium phosphate (pH 7.0) and 70 mM KCl. Samples for NMR experiments in H_2O contained 90%

(16) Brünger, A. T. *X-PLOR: A system for X-ray crystallography and NMR*, Yale University Press: New Haven, CT, 1992.
(17) (a) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517–6525. (b) Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. *FEBS J.* **2007**, *274*, 3545–3556. (c) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927–4940. (d) Haider, S.; Parkinson, G. N.; Neidle, S. *Biophys. J.* **2008**, *95*, 296–311. (e) Parkinson, G. N.; Cuenca, F.; Neidle, S. *J. Mol. Biol.* **2008**, *381*, 1145–1156.
(18) (a) Vorlícková, M.; Chládková, J.; Kejnovská, I.; Fialová, M.; Kypr, J. *Nucleic Acids Res.* **2005**, *33*, 5851–5860. (b) Yu, H. Q.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2006**, *128*, 15461–15468.
(19) Matsugami, A.; Okuizumi, T.; Uesugi, S.; Katahira, M. *J. Biol. Chem.* **2003**, *278*, 28147–28153.

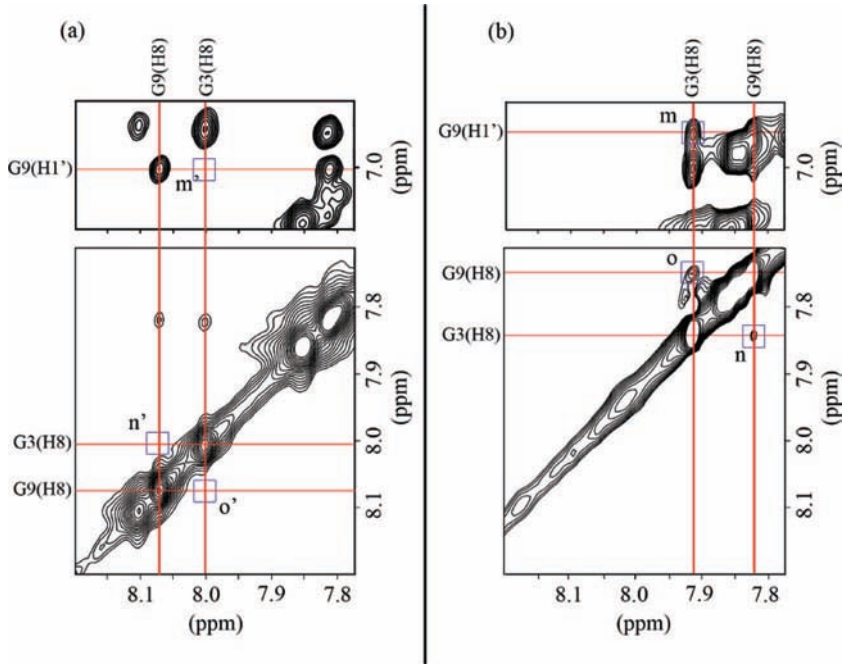

Figure 9. NOESY in D_2O pattern differences between the (a) 12-nt and (b) 10-nt human telomeric RNA sequences. NOE peaks for the 10-nt RNA indicating the stacking between two G-quadruplex blocks are framed and labeled: m, G3(H8)−G9(H1′); n and o, G9(H8)−G3(H8). Equivalent peaks for the 12-nt RNA were absent. The corresponding positions are framed and labeled with letters m′, n′, and o′.

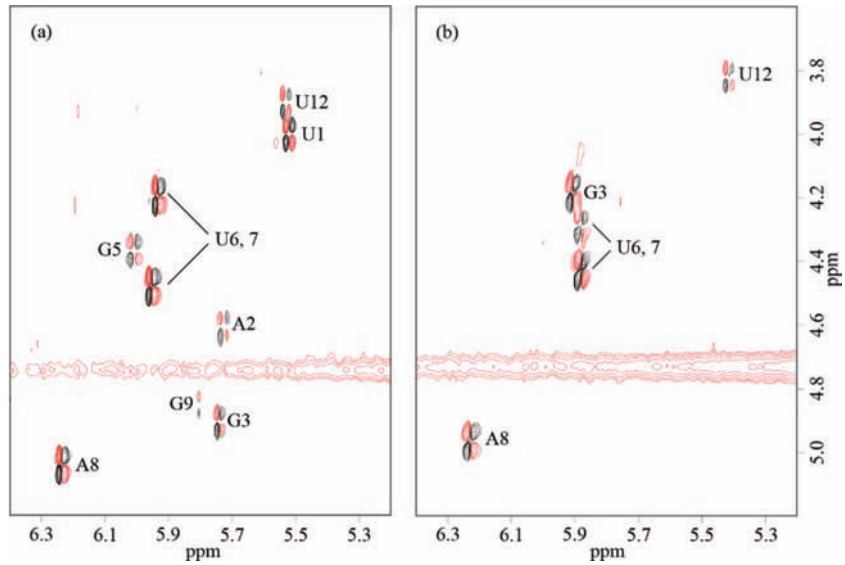

Figure 10. COSY spectra, showing H1′−H2′ couplings in the (a) 12-nt and (b) 10-nt human telomeric RNA sequences. Positive and negative levels are colored black and red, respectively.

H_2O and 10% D_2O; samples for experiments in D_2O contained 99.96% D_2O.

4.2. Gel Electrophoresis. Gel electrophoresis experiments were performed in 20 × 20 cm native gel containing 18% acrylamide (Acrylamide/Bis-acrylamide = 37.5:1) in TBE buffer (pH 8.3). 20 mM KCl was added into the buffer and the gel. Electrophoresis experiments were run at 150 V for 8 h, and the gel was viewed by the UV shadowing method or SYBR Gold staining. The concentrations of loaded RNA samples were 0.05−0.1 mM. DNA oligonucleotides dT_{12} and dT_{24} were used as molecular markers.

4.3. Circular Dichroism. CD spectra at 25 °C were recorded on a Jasco J-810 spectropolarimeter using a 1-cm path-length quartz cuvette in a reaction volume of 500 μL. The concentration of RNA samples was 5−8 μM. Scans from 220 to 320 nm were performed with a 200 nm/min scanning speed. For each spectrum, an average of three scans was taken, and the spectral contribution from the buffer was subtracted.

4.4. Melting Experiments. The thermal stability of the 12-nt RNA G-quadruplex under different salt conditions was measured in UV-monitored melting experiments.[20] Absorbance at 295 nm was recorded as a function of temperature using a Varian Cary 300 Bio UV−vis spectrophotometer. The temperature ranged from 20 to 90 °C. The heating and cooling rates were 0.1 °C/min. The melting temperature (T_m) was defined as the temperature of the midtransition point. Experiments were performed with quartz cuvettes, with a 1-cm path length. RNA concentration was 75 μM. The buffer contained 18 mM lithium phosphate (pH 7.0), 102 mM LiCl, and different NaCl or KCl concentrations (Figure S8).

(20) Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74–78.

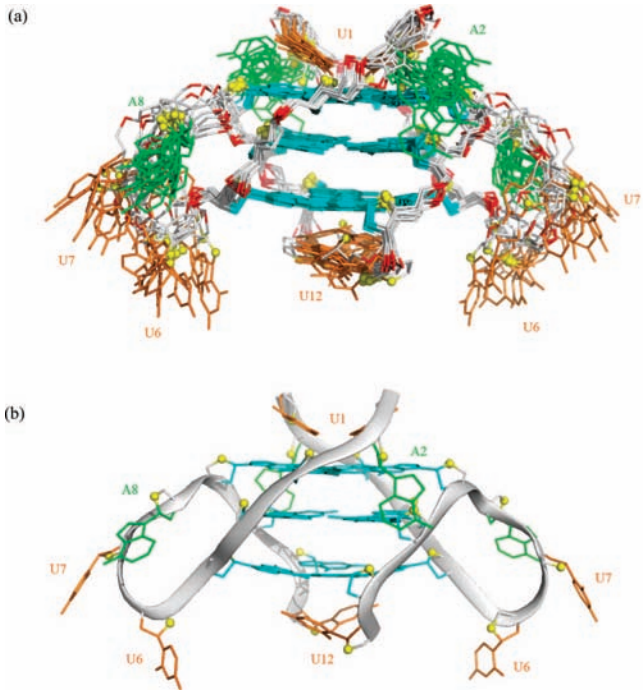

Figure 11. (a) Superimposed 10 computed structures of the 12-nt human telomeric RNA G-quadruplex in K^+ solution. (b) Ribbon view of the lowest-energy structure. Bases of guanines are colored cyan; adenines green; uracils orange; O4′ and P are colored yellow and red, respectively.

Table 3. Statistics of the Computed Structures

A. NMR restraints	nonexchangeable	exchangeable
Distance restraints		
intraresidue distance restraints	126	0
sequential (i, i+1) distance restraints	64	4
long-range (i, $\geq i$+1) distance restraints	48	34
ambiguous distance restraints	20	0
repulsive restraints	4	0
Other restraints		
hydrogen bonding restraints	48	
torsion angle restraints	184	

B. Statistics for 10 structures following refinement	
NOE violations	
number[a] (>0.2 Å)	0.60 ± 0.97
maximum violation (Å)	0.19 ± 0.03
rmsd of violations (Å)	0.04 ± 0.00
Deviations from the ideal covalent geometry	
bond lengths (Å)	0.00 ± 0.00
bond angles (deg)	0.58 ± 0.01
impropers (deg)	0.24 ± 0.00
Pairwise all heavy atom rmsd values (Å)	
all heavy atoms except U1, A2, U6, U7, A8, U12	1.11 ± 0.20
all heavy atoms	2.12 ± 0.41
R-factor	0.06 ± 0.01

[a] Total number of violations divided by the number of structures.

4.5. Nuclear Magnetic Resonance. NMR experiments were performed on 600 and 700 MHz Bruker spectrometers, both equipped with a CryoProbe. Spectra were recorded at 25 °C, unless otherwise specified. The strand concentration of NMR samples was typically 0.1−1.3 mM. NMR experiments in H_2O used JR-type pulse sequences[21,9b] for water signal suppression. Site-specific deoxyribose substitutions were used to assist resonance assignments. The substituted residues were recognized thanks to their upfield-shifted H2′/H2″ peaks. Resonances were assigned using a combination of through-bond correlations at natural abundance[9] (JRHMBC, HSQC, COSY, and TOCSY) and NOESY experiments. Sugar puckering conformations were characterized using *J*-couplings between H1′ and H2′ protons estimated from COSY experiments. Interproton distances were measured using NOESY experiments (mixing times: 50, 120, 200, 300, and 450 ms).

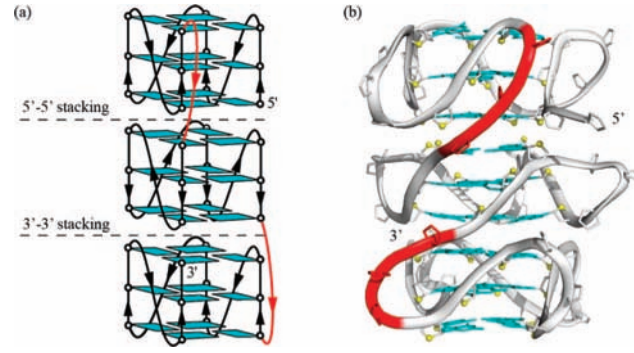

Figure 12. A model for a higher-order structure of the long human telomeric RNA based on alternate-direction stacking of three-layer G-quadruplex blocks. (a) Schematic structure; (b) Ribbon view of a computed structure. UUA linkers connecting consecutive blocks are colored red.

4.6. Structure Calculation. The structure of the dimeric G-quadruplex formed by the 12-nt human telomeric RNA sequence was calculated using the NIH version[22] of the X-PLOR program.[16] NMR-restrained computations were performed as described previously.[23] Structures were displayed using PyMOL[24] and analyzed using MOLMOL.[25]

4.6.1. Topology and Parameter Files. The topology (dna-rna-allatom.top) and parameter (dna-rna-allatom.param) files were obtained from http://cns-online.org/v1.21/.

4.6.2. NOE Distance Restraints. The distances between non-exchangeable protons of the 12-nt RNA were obtained from NOESY cross-peaks at various mixing times. NOESY at various mixing times (50, 120, 300, and 450 ms) were used to classify the distances. The peaks were classified as strong (strong peaks at 50 ms), medium (peaks which were weak at 50 ms, but strong at 120 ms), weak (peaks which were weak at 120 ms, but strong at 300 ms), and very weak (only observed at 450 ms). The distances restraints corresponding to these classifications are 2 ± 1, 3.75 ± 1.25, 5.5 ± 1.5, and 7.5 ± 1.5 Å, respectively. The peaks from exchangeable protons were also classified manually with a NOESY with mixing time 200 ms. The peaks were classified as medium, weak, or very weak, with ranges of 3 ± 2 Å for medium and 4.5 ± 2 Å for weak and 5.75 ± 2.25 Å for very weak.

4.6.3. Dihedral Restraints. All the glycosidic bonds were restrained according to their *anti* conformation. The sugar puckering conformation C2′-*endo* was restrained for U1, A2, G3, G5, U6, U7, A8, and U12 with 40° flexibility.

4.6.4. Hydrogen-Bond and Planarity Restraints. The G3•G9•G3•G9, G4•G10•G4•G10, and G5•G11•G5•G11 tetrads within the dimeric RNA G-quadruplex were restrained with O6−N1 and N7−N2 distances, which were set to 2.95 ± 0.1 and 2.90 ± 0.1 Å, respectively. Restraints were applied to the bases of these G-tetrads to enforce their planarity.

4.6.5. Distance Geometry and Simulated Annealing. The calculation started with generating two extended RNA strands. Initial distance geometry calculation was performed by incorporat-

(21) Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310–7311.
(22) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 66–74.
(23) Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andréola, M. L.; Patel, D. J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 634–639.
(24) DeLano, W. L. *The PyMOL User's Manual*; DeLano Scientific: Palo Alto; CA, 2002.
(25) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51–55.

ing hydrogen-bond, dihedral angle, planarity, and NOE restraints. This step gave the overall folding of the RNA G-quadruplex. Subsequent simulated annealing was performed involving restraints mentioned above. As NMR spectra indicated the symmetry between the two strands, noncrystallographic symmetry restraints were also applied in the calculation with a force constant of 500 kcal·mol^{-1}·Å^{-2}.

4.6.6. Refinement. The 100 structures calculated from the simulated annealing step were further refined. The refinement process followed the molecular dynamics protocol. The molecule was initially heated until 1000 K from 300 K in 5 ps. Equilibration for 1 ps was performed with the force constant for nonexchangeable and exchangeable NOE restraints being kept at 2 kcal·mol^{-1}·Å^{-2}. The force constants for both types of NOE restraints were then increased to 50 kcal·mol^{-1}·Å^{-2} in 26 ps. The system was then cooled down until 300 K in 14 ps, followed by 10 ps of equilibration. Coordinates of the molecule were saved every 0.5 ps during the last 4.0 ps and then averaged. The structure produced was then subjected to minimization until the gradient of energy was less than 0.1 kcal·mol^{-1}. The energy terms used for the refinement were van der Waals, electrostatic, bond angle, bond length, dihedral angle, improper angle, NOE, noncrystallographic symmetry, and planarity. Dihedral (20 kcal·mol^{-1}·Å^{-2}) and planarity (1 kcal·mol^{-1}·Å^{-2}) restraints were maintained throughout the refinement process. The 10 lowest-energy structures were selected. Rmsd calculation was performed using MOLMOL.[24]

4.6.7. R-Factor. R-factor, a measure of the quality of the structure, was calculated based on the peak intensities from NOESY spectra (mixing time: 120 and 300 ms). The R-factor was calculated using X-PLOR-NIH,[22] by comparing the structure-based calculated intensities with the experimentally derived intensities.

4.7. Modeling of Long Telomeric RNA. Structural models for long telomeric RNA were built by using the X-PLOR program with the same protocols as those described above. 12-repeat and 8-repeat human telomeric RNA sequences, r[UAGGG(UUAGGG)$_{11}$] and r[UAGGG(UUAGGG)$_7$], were assumed to form a three-block and a two-block G-quadruplex structure, respectively (see above text). Artificial hydrogen-bond, planarity, and dihedral angle restraints were used to model the three-layer G-tetrad core; NOEs derived from the 12-nt RNA were used to model the double-chain-reversal UUA loops, except for the UUA linkers that connect consecutive blocks; NOEs (H8−H8, H8−H1′ and H8−H1) derived from the 10-nt RNA were used for the 5′−5′ stacking interface between blocks; artificial distance restraints (H8−H8 and H8−H1) were used for the 3′−3′ stacking interface between blocks. The model of the three-block RNA G-quadruplex was built using constraints for both 5′−5′ and 3′−3′ stacking interfaces. For the two-block RNA G-quadruplex, the computation was performed with no restraints for the stacking interfaces. Out of 100 computed structures, the 16th lowest-energy structure showed a 5′−5′ stacking conformation (Figure S11).

4.8. Data Deposition. The coordinates for the human telomeric RNA G-quadruplex formed by the r(UAGGGUUAGGGU) sequence have been deposited in the Protein Data Bank (accession code 2KBP).

Acknowledgment. This research was supported by Singapore Ministry of Education Grant ARC30/07 to A.T.P.

Supporting Information Available: This material is available free of charge via the Internet at http://pubs.acs.org.

JA806592Z