Sequence variant $(CTAGGG)_n$ in the human telomere favors a G-quadruplex structure containing a G•C•G•C tetrad

Kah Wai Lim[1], Patrizia Alberti[2], Aurore Guédin[2], Laurent Lacroix[2], Jean-François Riou[2], Nicola J. Royle[3], Jean-Louis Mergny[2,*] and Anh Tuân Phan[1,*]

[1]Division of Physics and Applied Physics,
School of Physical and Mathematical Sciences,
Nanyang Technological University, Singapore

[2]INSERM U565, CNRS UMR 7196, USM 503,
Muséum National d'Histoire Naturelle, Paris, France

[3]Department of Genetics,
University of Leicester, Leicester, LE1 7RH, UK

SUPPORTING INFORMATION

SUPPLEMENTARY TEXT

Methods for structure calculation

Hydrogen Bond Restraints. Each hydrogen bond was restrained by two distances (donor atom to acceptor atom and proton to acceptor atom) which correspond to ideal hydrogen bond geometry. The force constant for hydrogen bond restraints was kept at 100 $kcal.mol^{-1}.Å^{-2}$ throughout the computation.

Non-exchangeable Proton Distance Restraints. Distances between non-exchangeable protons were deduced from NOESY spectra in D_2O (mixing times, 100, 200, and 350 ms) and implemented as distance restraints during structure calculations.

For well-resolved cross-peaks, the initial slopes of the buildup curves were obtained by performing a second-order fit on the volumes measured at 100- and 200-ms mixing times. The slopes were then converted to inter-proton distances using the cytosine H5-H6 distance of 2.4 Å as reference, and given bounds of 20% or 30%, depending on the resolution of the peak.

Cross-peaks that are not well-resolved were manually classified as strong or weak, and assigned to inter-proton distances of 1.8 - 4.2 Å or 3.6 - 7.8Å, respectively. Heavily overlapped cross-peaks were given the distances of 1.8 to 7.8 Å.

Cross-peaks involving a methyl group were interpreted as distances to the methyl carbon. The upper bounds for these peaks were increased by 0.5 Å to account for the radius of the methyl group.

Exchangeable Proton Distance Restraints. Cross-peaks involving exchangeable protons were obtained from the NOESY spectrum in H_2O at 200-ms mixing time and classified manually.

NMR Dihedral Restraints. The glycosidic χ torsion angle for experimentally-determined *syn* guanine residues G3, G9, G15, and G21 were restrained to (60±30)°, while that for residues G4, C5, G8, G10, G16, C17, G20, and G22 were fixed at (240±40)°, characteristic of *anti* conformation. In addition, the H1'-C1'-N9/N1-C4/C2 torsion angle for G3, G9, G15, and G21 were set to (180±30)°, whereas that for G4, C5, G8, G10, G16, C17, G20, and G22 were set to (0±40)°. The ε dihedral angles between residue pairs G3-G4, G4-C5, G8-G9, G9-G10, G15-G16, G16-C17, G20-G21, and G21-G22 were set to be (225±75)°, consistent with the stereochemically allowed range for this angle. Prior NOE-restrained computations produced structures in which G2 adopted a near-*syn* conformation, in disagreement with the intensity of the intraresidue H8-H1' cross-peak for this residue. As such, the glycosidic χ torsion angle and the H1'-C1'-N9-C4 angle of G2 were restrained to (240±70)° and (0±70)°, respectively, for all subsequent calculations.

Planarity Restraints. Planarity restraints were enforced on the base atoms for the G3•G10•G15•G22 and G4•G21•G16•G9 tetrads, and the G8•C17, G20•C5, and G14•C11 base-pairs.

Repulsive Restraints. Repulsive restraints (4.5 - 80 Å) were applied on pairs of protons that do not exhibit cross-peaks in NOESY.

Distance Geometry Simulated Annealing. An extended conformation of the d[AGGG(CTAGGG)$_3$] sequence with ideal geometry was generated randomly using XPLOR-NIH. The extended DNA was then subjected to distance geometry simulated annealing (DGSA) by supplying the full set of hydrogen bond, distance, dihedral, planarity, and repulsive restraints. 100 structures were generated and subjected to further refinements.

Distance-restrained Molecular Dynamics Refinement. All 100 DGSA structures were refined with distance-restrained molecular dynamics. The system was heated from 300 K to 1000 K in 5 ps and allowed to equilibrate for 1 ps, wherein the force constants for the distance restraints were kept at 2 kcal.mol^{-1}.Å^{-2}. The force constants for non-exchangeable and exchangeable proton distance restraints were then scaled to final values of 50 and 40 kcal.mol^{-1}.Å^{-2}, respectively, in a 26-ps interval. Subsequently, the system was slowly cooled down to 300 K in 14 ps, after which equilibration was performed for another 10 ps. Coordinates of the molecule were saved every 0.5 ps during the last 4.0 ps and averaged. In the final step, the average structure was subjected to minimization until the gradient of energy was less than 0.1 kcal.mol^{-1}. Dihedral (200 kcal.mol^{-1}.rad^{-2}) and planarity (5 kcal.mol^{-1}.Å^{-2} for both tetrads and G•C base-pairs) restraints were maintained throughout the course of refinement. 10 best structures were selected, based on the overall energy term, to undergo relaxation matrix intensity refinement.

Relaxation Matrix Intensity Refinement. Relaxation matrix intensity refinement was next performed to account for spin diffusion effects. The relaxation matrix was set up for the non-exchangeable protons. NOE intensity volumes of 108 cross-peaks for each of three mixing times (100, 200, and 350 ms) were employed. The exchangeable imino and amino protons were all replaced by deuterons. Heating was done from 5 K to 300 K. The force constant for NOE intensities was increased from 2 to 300 kcal.mol^{-1}.Å^{-2}, while the force constant for non-exchangeable proton distance restraints were decreased from 50 to 30 kcal.mol^{-1}.Å^{-2}. The force constant for exchangeable proton distance restraints was maintained at 40 kcal.mol^{-1}.Å^{-2}. After that, the system was equilibrated at 300 K for 3 ps and energy-minimized. The planarity restraints were set to very low values of 1 kcal.mol^{-1}.Å^{-2} for the tetrads and 0.5 kcal.mol^{-1}.Å^{-2} for the G•C base-pairs.

SUPPLEMENTARY TABLES

Table S1. Natural and modified human telomeric DNA sequences used in this study[a,b]

Repeat Type	Sequence									Name
Canonical	A	GGG	TTA	GGG	TTA	GGG	TTA	GGG		*22wt*
CTA variant										
natural		GG	CTA	GGG	CTA	GGG	CTA	GGG	C	
		GGG	CTA	GGG	CTA	GGG	CTA	GGG		
		GGG	CTA	GGG	CTA	GGG	CTA	GGG	C	
		GGG	CTA	GGG	CTA	GGG	CTA	GGG	CT	
		GGG	CTA	GGG	CTA	GGG	CTA	GGG	CTA	
	A	GGG	CTA	GGG	CTA	GGG	CTA	GGG		*22CTA*
	A	GGG	CTA	GGG	CTA	GGG	CTA	GGG	C	*23CTA*
	A	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CT	
	A	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CTA	
	TA	GGG	CTA	GGG	CTA	GGG	CTA	GGG		
	TA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	C	
	TA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CT	
	TA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CTA	
	CTA	GGG	CTA	GGG	CTA	GGG	CTA	GGG		
	CTA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	C	
	CTA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CT	
	CTA	GGG	CTA	GGG	CTA	GGG	CTA	GGG	CTA	
modified	A	GGG	CTA	G^{Br}**G**G	CTA	GGG	CTA	GGG		
	A	GGG	CTA	GGG	CTA	GGG	CTA	G^{Br}**G**G		
	A	GGG	CTA	G^{Br}**G**G	CTA	GGG	CTA	G^{Br}**G**G		
	A	G^{Br}**G**G	CTA	GGG	CTA	GGG	CTA	GGG		
	A	GGG	CTA	GGG	CTA	G^{Br}**G**G	CTA	GGG		
TCA variant										
natural	A	GGG	TCA	GGG	TCA	GGG	TCA	GGG		*22TCA*

[a] Variations from the canonical sequence are underlined.
[b] Modified residues are shown in boldface; ^{Br}**G** is 8-bromoguanine.

Table S2. Site-specific labeled DNA sequences used in this study[a-d]

Type	Sequence							
^{P15}N-labeled	A	*GGG	CTA	GGG	CTA	GGG	CTA	GGG
	A	G*GG	CTA	GGG	CTA	GGG	CTA	GGG
	A	GG*G	CTA	GGG	CTA	GGG	CTA	GGG
	A	GGG	CTA	*GGG	CTA	GGG	CTA	GGG
	A	GGG	CTA	G*GG	CTA	GGG	CTA	GGG
	A	GGG	CTA	GG*G	CTA	GGG	CTA	GGG
	A	GGG	CTA	GGG	CTA	*GGG	CTA	GGG
	A	GGG	CTA	GGG	CTA	G*GG	CTA	GGG
	A	GGG	CTA	GGG	CTA	GG*G	CTA	GGG
	A	GGG	CTA	GGG	CTA	GGG	CTA	*GGG
	A	GGG	CTA	GGG	CTA	GGG	CTA	G*GG
^{2}H-labeled	A	#GGG	CTA	GGG	CTA	GGG	CTA	GGG
	A	GGG	CTA	G#GG	CTA	GGG	CTA	GGG
	A	GGG	CTA	GGG	CTA	#GGG	CTA	GGG
	A	GGG	CTA	GGG	CTA	GGG	CTA	#GGG
U-substituted	A	GGG	C**U**A	GGG	CTA	GGG	CTA	GGG
	A	GGG	CTA	GGG	C**U**A	GGG	CTA	GGG
	A	GGG	CTA	GGG	CTA	GGG	C**U**A	GGG

[a] Variations from the canonical sequence are underlined.
[b] Individual guanines 2%-^{15}N-labeled are marked by asterisks and colored red.
[c] Individual guanines ^{2}H-labeled at the H8 position are marked by hash signs and colored blue.
[d] Modified residues are shown in boldface and colored green; **U** is uracil.

SUPPLEMENTARY FIGURES

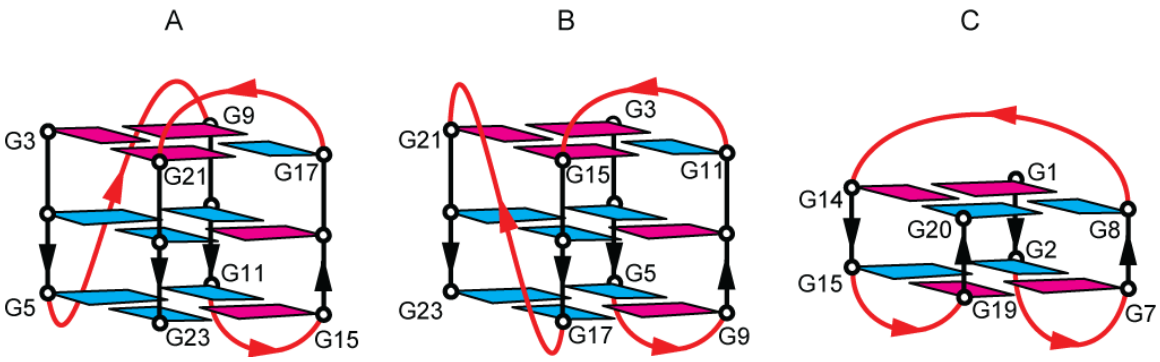

Figure S1. Schematic structures of intramolecular G-quadruplexes formed by four-repeat TTAGGG human telomeric sequences in K^+ solution: (**A**) three-G-tetrad (3+1) Form 1 observed for d[TAGGG(TTAGGG)$_3$], (**B**) three-G-tetrad (3+1) Form 2 observed for d[TAGGG(TTAGGG)$_3$TT], and (**C**) two-G-tetrad basket-type Form 3 observed for d[GGG(TTAGGG)$_3$T]. *anti* guanines are colored cyan; *syn* guanines are colored magenta. The backbone of the core and loops is colored black and red, respectively.

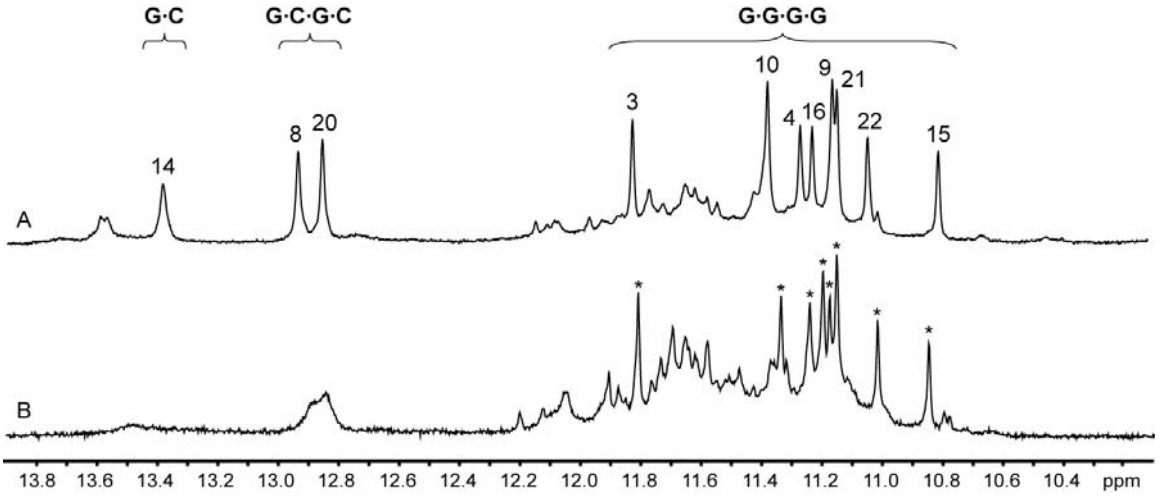

Figure S2. Imino proton spectra of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence (**A**) at 25°C and (**B**) at 45°C. The imino proton peaks of G14, G8, and G20 are broadened at high temperature, suggesting that they are not involved in G-tetrad formation, while the eight other tetrad-forming imino peaks (indicated by asterisks) remain sharp.

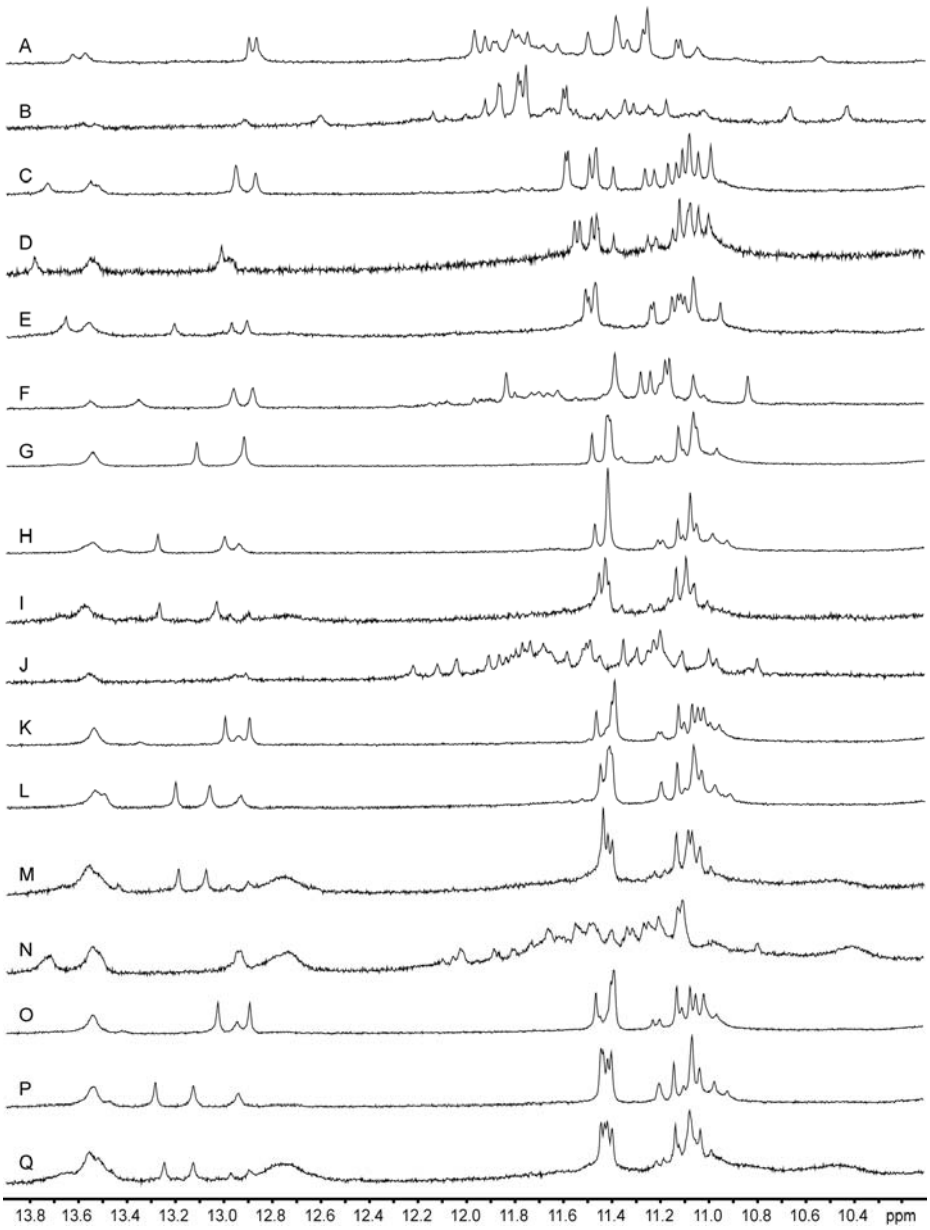

Figure S3. Imino proton spectra of four-repeat human telomeric variant sequences containing CTAGGG repeats in K^+ solution 25°C. (**A**) d[GG(CTAGGG)$_3$C], (**B**) d[GGG(CTAGGG)$_3$], (**C**) d[GGG(CTAGGG)$_3$C], (**D**) d[GGG(CTAGGG)$_3$CT], (**E**) d[GGG(CTAGGG)$_3$CTA], (**F**) d[AGGG(CTAGGG)$_3$], (**G**) d[AGGG(CTAGGG)$_3$C], (**H**) d[AGGG(CTAGGG)$_3$CT], (**I**) d[AGGG(CTAGGG)$_3$CTA], (**J**) d[TAGGG(CTAGGG)$_3$], (**K**) d[TAGGG(CTAGGG)$_3$C], (**L**) d[TAGGG(CTAGGG)$_3$CT], (**M**) d[TAGGG(CTAGGG)$_3$CTA], (**N**) d[(CTAGGG)$_4$], (**O**) d[(CTAGGG)$_4$C], (**P**) d[(CTAGGG)$_4$CT], (**Q**) d[(CTAGGG)$_4$CTA]. Most of them adopt predominantly the same chair-type intramolecular G-quadruplex, as suggested by the number of major tetrad-forming imino protons at 10.8-11.8 ppm and the two major downfield-shifted sharp imino signature peaks.

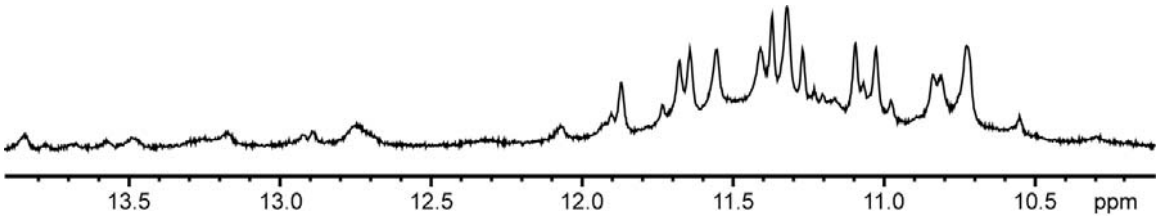

Figure S4. Imino proton spectrum of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence in Na^+ solution at 25°C.

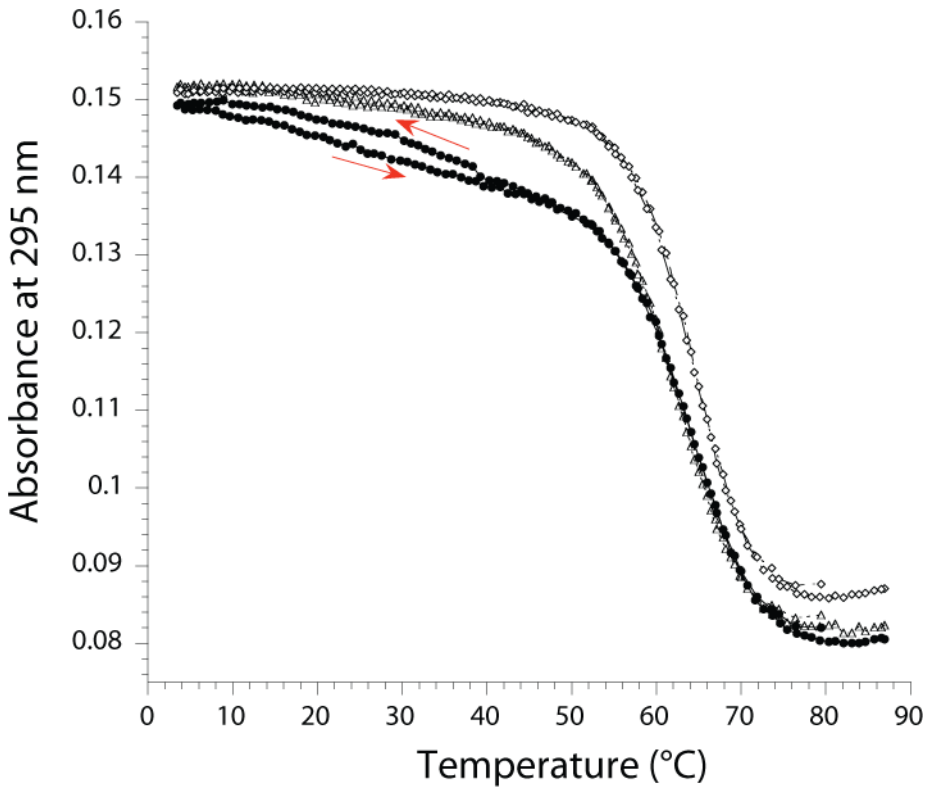

Figure S5. Examples of UV-melting profiles obtained for the human telomeric *22CTA* (triangles), *23CTA* (diamonds), and *22wt* (filled circles) sequences in a 10 mM lithium cacodylate, 100 mM KCl pH 7.2 buffer. Both heating and cooling profiles are shown. Note the slight but reproducible hysteresis obtained at low temperature for the *22wt* sequence. In contrast, the profiles of the *22CTA* and *23CTA* variants are fully reversible.

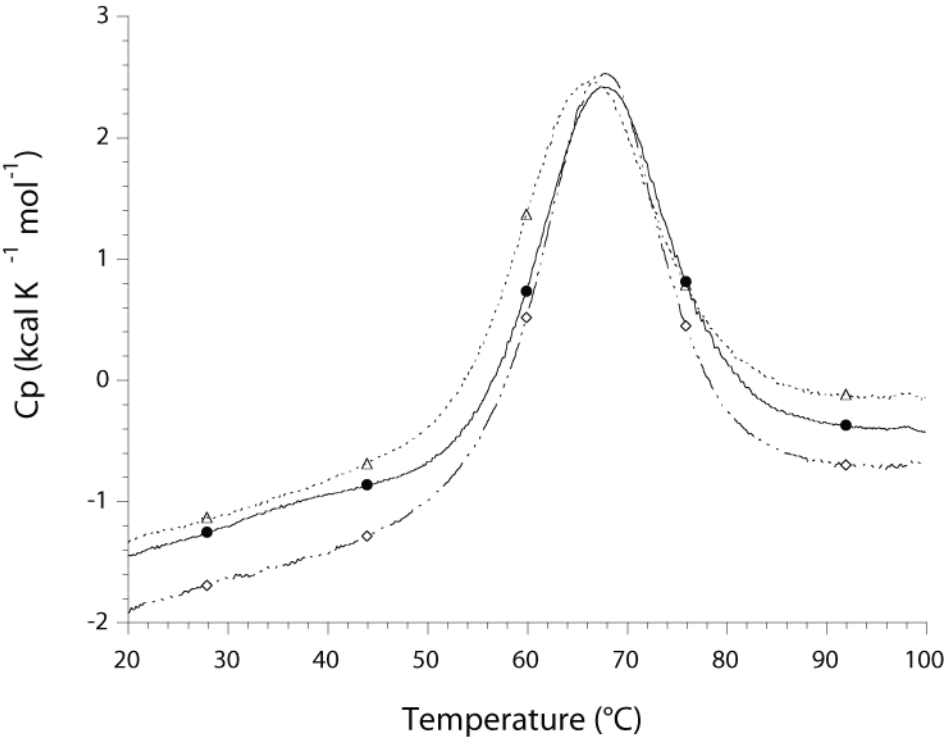

Figure S6. DSC profiles obtained for the human telomeric *22CTA* (triangles), *23CTA* (diamonds), and *22wt* (filled circles, full lines) sequences in a 10 mM lithium cacodylate, 100 mM KCl pH 7.2 buffer. Only a few experimental points are shown for clarity.

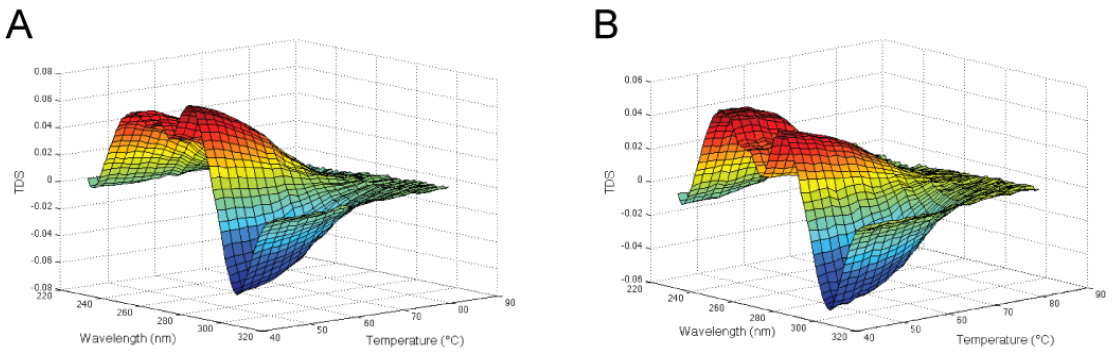

Figure S7. 3D TDS of the 22-nt human telomeric (**A**) variant *22CTA* sequence and (**B**) canonical *22wt* sequence. DNA concentration was 4 μM; solution contained 100 mM KCl and 10 mM lithium cacodylate, pH 7.2.

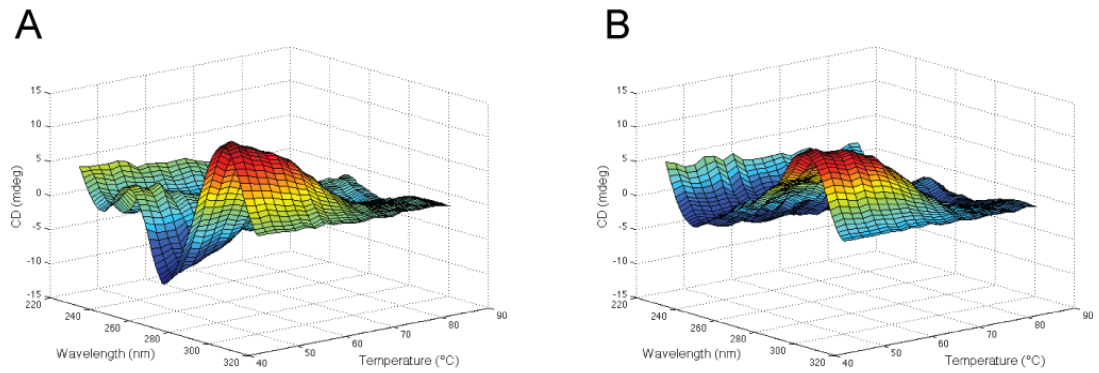

Figure S8. 3D CD of the 22-nt human telomeric (**A**) variant *22CTA* sequence and (**B**) canonical *22wt* sequence. DNA concentration was 4 μM; solution contained 100 mM KCl and 10 mM lithium cacodylate, pH 7.2.

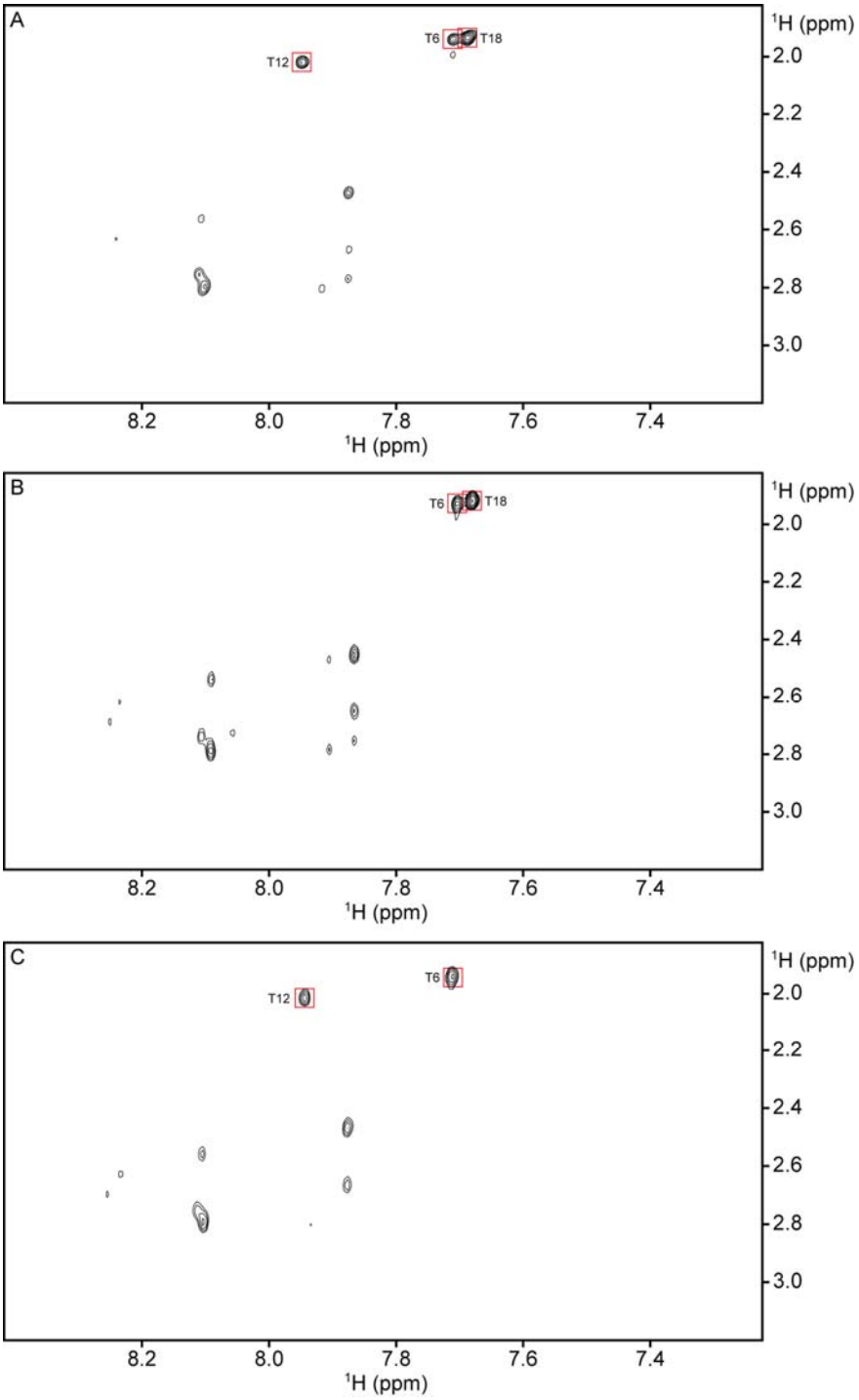

Figure S9. Examples of thymine H6 and methyl proton assignments by T-to-U substitution. (**A**) NOESY spectrum (mixing time, 350 ms) of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence showing the three thymine intraresidue H6-methyl cross-peaks. The H6-methyl cross-peak for (**B**) residue 12 and (**C**) residue 18 disappeared following T-to-U substitution at position 12 and position 18, respectively.

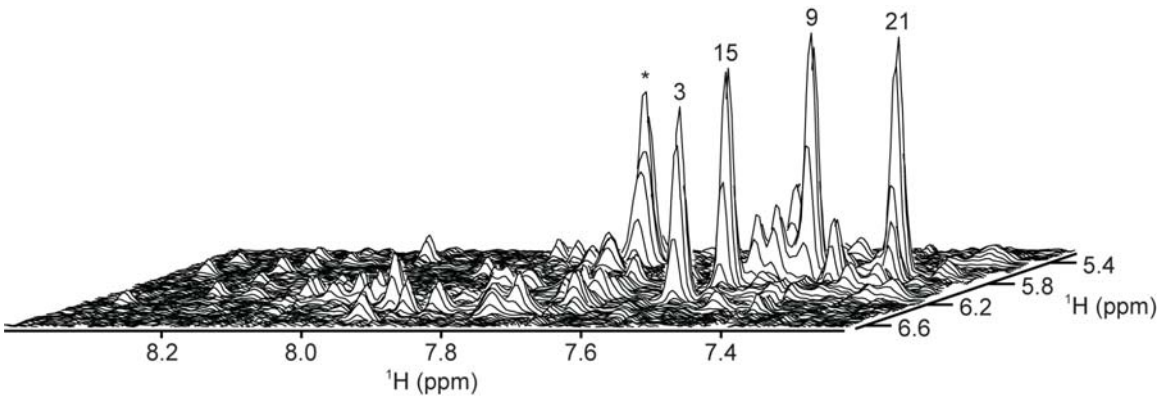

Figure S10. Stacked plot of NOESY spectrum (mixing time, 100 ms) of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence that shows four strong intraresidue H8-H1' cross-peaks for G3, G9, G15, and G21, indicating *syn* glycosidic conformations. The strong cross-peak marked by an asterisk corresponds to the intraresidue NOE correlation between H5 and H6 protons of Cytosine 5.

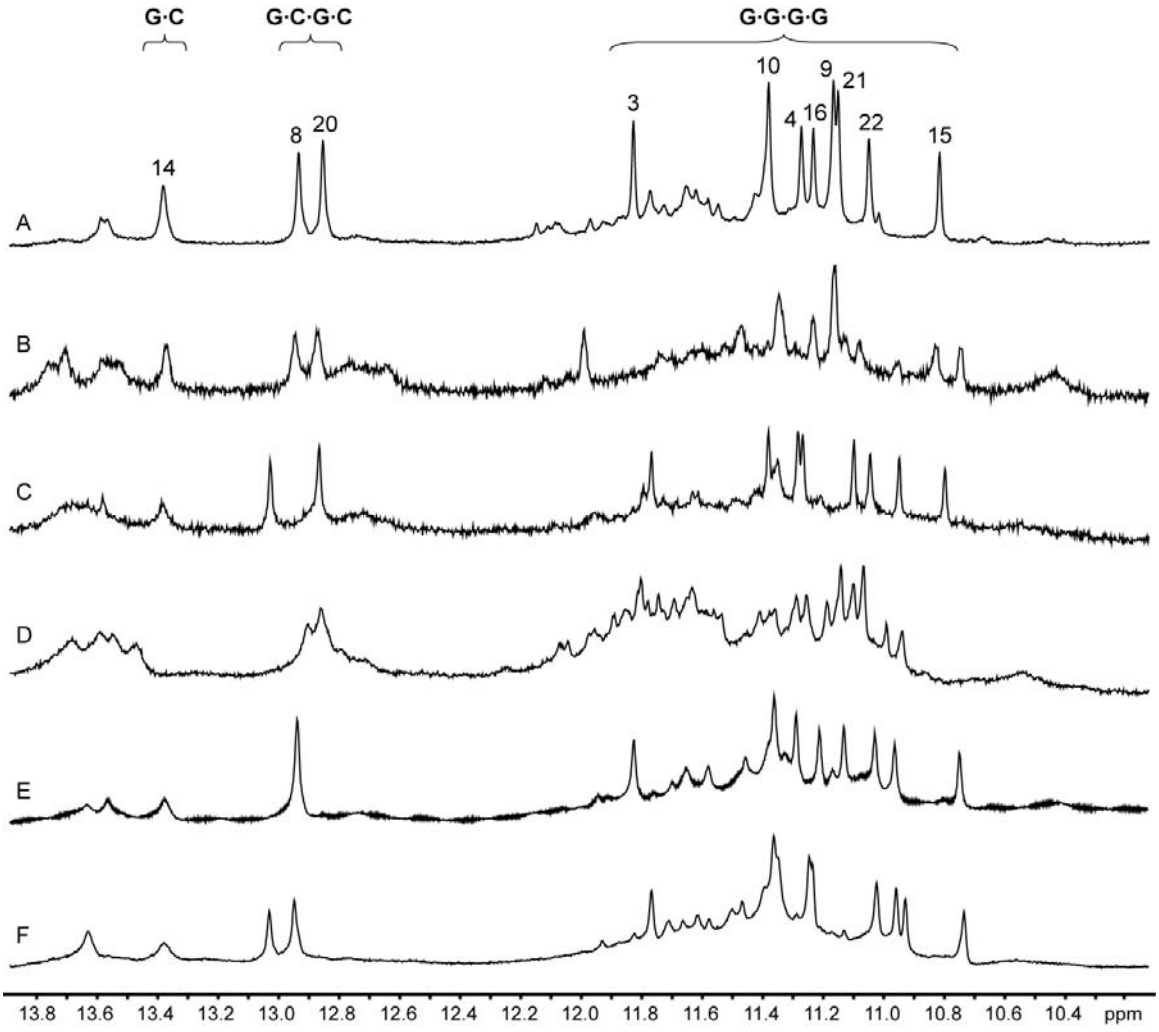

Figure S11. Imino proton spectra of (**A**) d[AGGG(CTAGGG)$_3$], (**B**) d[AG(^{Br}G)G(CTAGGG)$_3$], (**C**) d[AGGGCTAG(^{Br}G)G(CTAGGG)$_2$], (**D**) d[AGGGCTAGGGCTAG(^{Br}G)GCTAGGG], (**E**) d[AGGG(CTAGGG)$_2$CTAG(^{Br}G)G], and (**F**) d[AGGGCTAG(^{Br}G)GCTAGGGCTAG(^{Br}G)G]. ^{Br}G substitution of *syn*-guanine(s) for the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence maintains the same general fold as the unmodified sequence, as suggested by the number of major tetrad-forming imino protons and the presence of the two downfield-shifted sharp imino signature peaks.

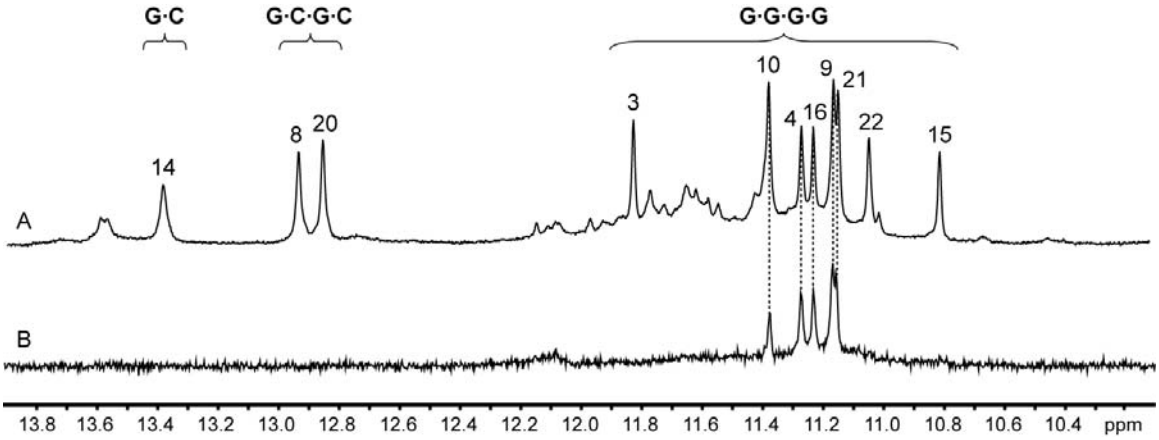

Figure S12. Imino proton spectra of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence (**A**) in H_2O and (**B**) in D_2O for 1 hour.

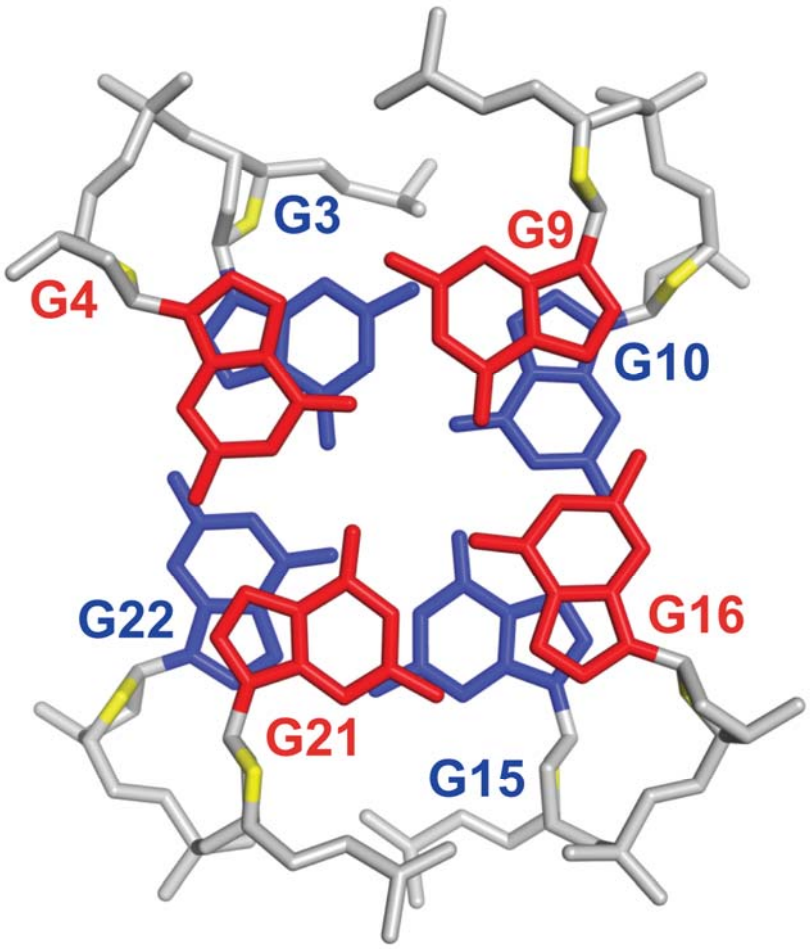

Figure S13. The two G-tetrad layers of the central core, G4•G21•G16•G9 (red) and G3•G10•G15•G22 (blue), which alternate anticlockwise and clockwise. There is extensive overlap between the five-membered rings of the guanines. Sugar and backbone atoms are colored gray while O4' atoms are colored yellow.

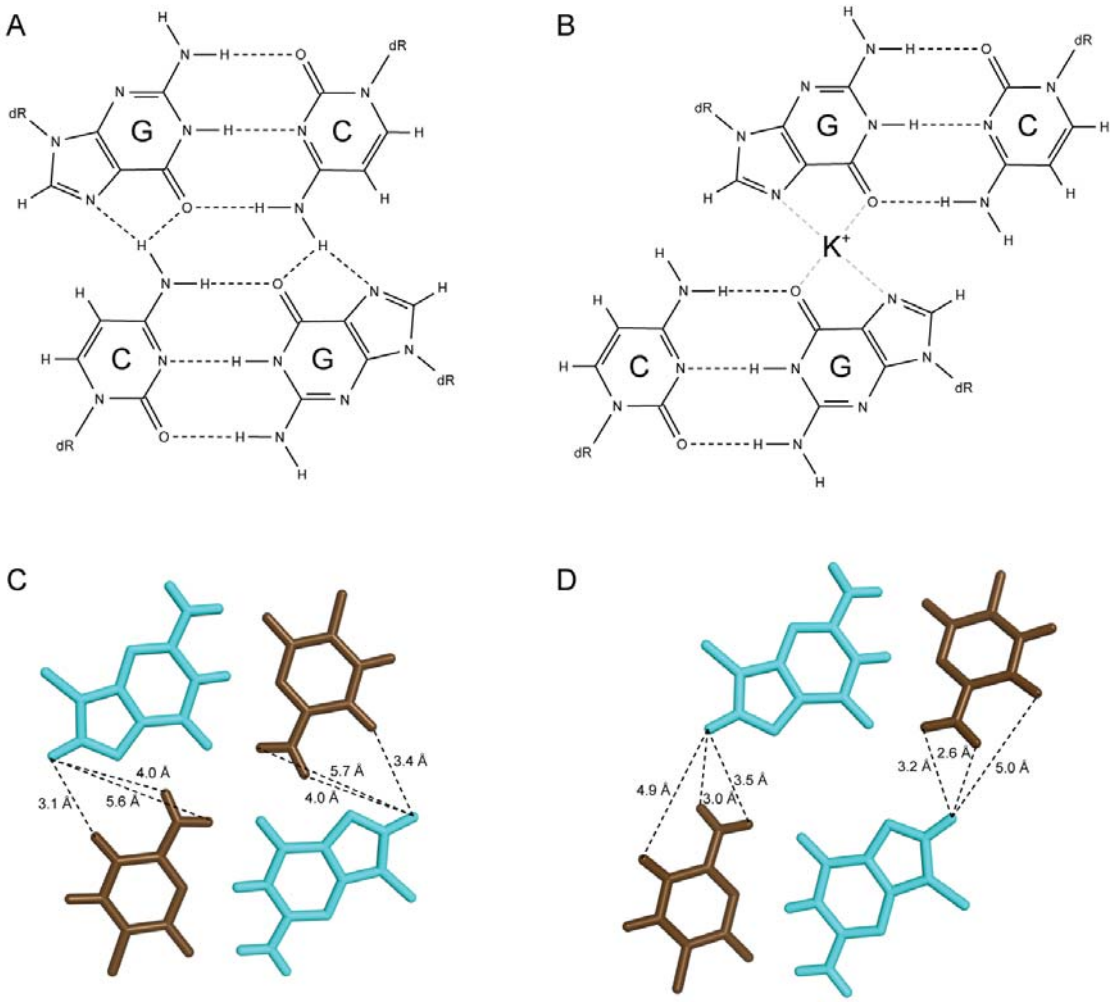

Figure S14. Two possible arrangements of a G•C•G•C tetrad. The two G•C base-pairs are paired up across the major groove edges, either in (**A**, **C**) a direct alignment or (**B**, **D**) a slipped alignment. Inter-proton distances that can be used to distinguish between the two alignments are shown by black dotted lines in (C) and (D). The PDB ID for the structures in (C) and (D) are 1A8N and 1A8W, respectively.

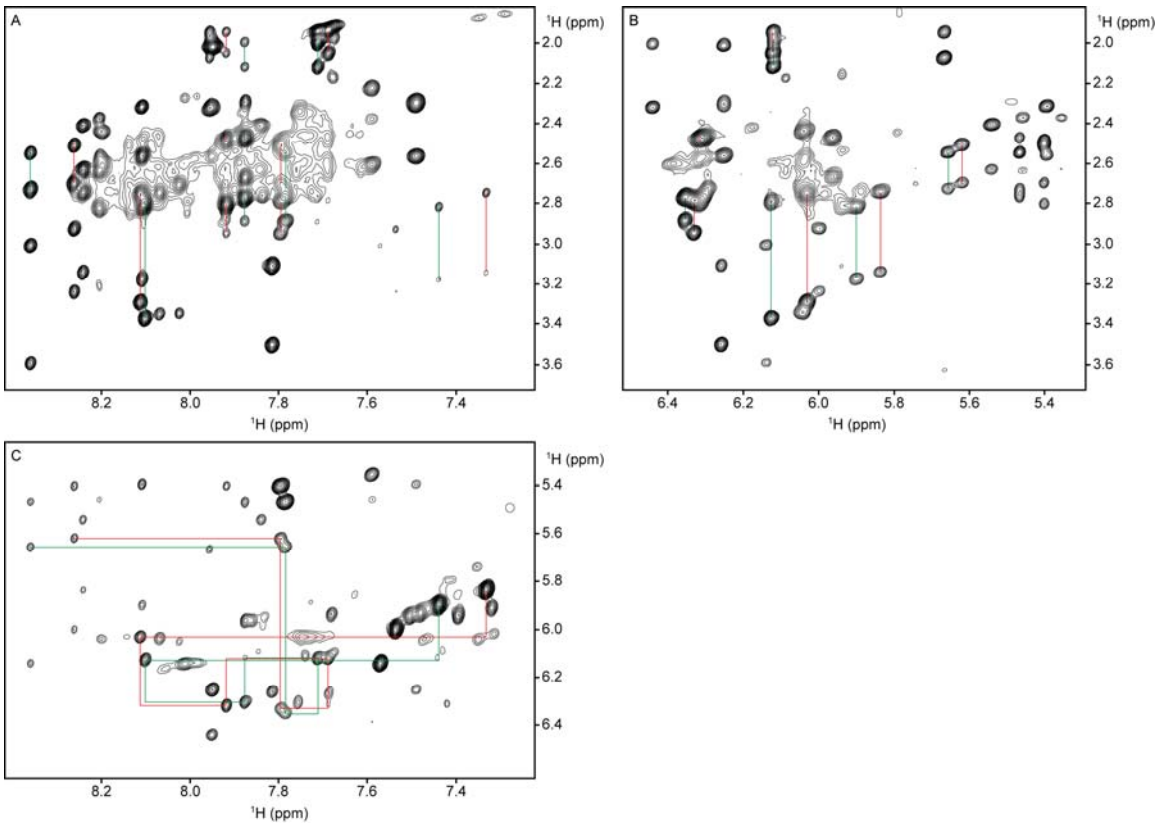

Figure S15. NOESY spectra (mixing time, 350 ms) showing the similar NOE patterns displayed by the two quasi-symmetric segments, G4-G9 (green) and G16-G21 (red), of the d[AGGG(CTAGGG)$_3$] quadruplex. (**A**) H8/H6-H2'/H2'' region, (**B**) H1'-H2'/H2'' region, and (**C**) H8/H6-H1' region.