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

ABSTRACT

Short contiguous arrays of variant CTAGGG repeats in the human telomere are unstable in the male germline and somatic cells, suggesting formation of unusual structures by this repeat type. Here, we report on the structure of an intramolecular G-quadruplex formed by DNA sequences containing four human telomeric variant CTAGGG repeats in potassium solution. Our results reveal a new robust antiparallel G-quadruplex fold involving two G-tetrads sandwiched between a G·C base pair and a G·C·G·C tetrad, which could represent a new platform for drug design targeted to human telomeric DNA.

INTRODUCTION

Telomeres, the ends of linear eukaryotic chromosomes, consist of tandem repeats of G-rich sequences (1). Acting as the protective caps of the chromosomes, telomeres are critical for chromosomal stability, cell survival and proliferation (2). Human telomeres encompass thousands of canonical (wild-type) TTAGGG repeats (3), which can be interspersed with some sequence-variant repeats (4,5). It has been reported [see accompanying paper (6)] that a particular repeat type, C̲TAGGG (variation is underlined), when present as a short contiguous array within the telomere, causes an extraordinarily high level of localized telomere instability in the male germline and somatic cells. These variant repeats also bind to the telomeric factor POT1 more efficiently than TTAGGG repeats (6). Thus, it is important to understand if unusual structures are formed by these sequence-variant repeats, as they could lead to incongruities during telomere replication, contributing to the instability of the telomere.

Previous studies have established that human telomeric DNA sequences containing four canonical TTAGGG repeats can adopt at least five different G-quadruplex structures under different experimental conditions (7–29). These include three conformations observed in K$^+$ solution (Supplementary Figure S1): the sequences d[TAGGG(TTAGGG)$_3$] and d[TAGGG(TTAGGG)$_3$ TT] form predominantly intramolecular three-G-tetrad (3 + 1) G-quadruplexes Form 1 (21–26,28) and Form 2 (24,27,28), respectively, while the sequence d[GGG(TTA GGG)$_3$T] forms mainly an intramolecular basket-type G-quadruplex involving only two G-tetrads, designated as Form 3 (29). Here, we report on the structure of an intramolecular G-quadruplex formed by DNA sequences containing four human telomeric variant C̲TAGGG repeats in K$^+$ solution. Our results reveal a new robust antiparallel G-quadruplex fold involving two G-tetrads, a G·C base pair and a G·C·G·C tetrad, which could serve as a new platform for drug design targeted to human telomeric DNA.

METHODS

DNA sample preparation

Unlabeled and site-specific labeled DNA oligonucleotides (Table 1; Supplementary Tables S1 and S2) 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 (30).

Gel electrophoresis

The molecular size of the structures formed by DNA oligonucleotides was probed by non-denaturing polyacrylamide gel electrophoresis (PAGE) as previously described (31). DNA samples (^{32}P-labeled) of three different total concentrations (0.05, 4 and 80 μM) were incubated in a 10 mM Tris–HCl pH 7.5 buffer supplemented with 100 mM NaCl or KCl. They were heated at 90°C for 5 min and slowly cooled down (over 2 h) to 20°C. The samples were loaded on a 15% polyacrylamide gel supplemented with 20 mM of the corresponding salt and run at 26°C; 10% sucrose was added just before loading.

UV-melting experiments

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 (32) using a Kontron-Uvikon 940 UV/Vis spectrophotometer. UV-melting experiments were conducted as previously described (33) in a 10 mM lithium cacodylate pH 7.2 buffer containing 100 mM NaCl or KCl. The heating and cooling rates were 0.2°C per minute. Experiments were performed with 1-cm pathlength quartz cuvettes.

Thermal difference spectra

The thermal difference spectra (TDS) of a nucleic acid were obtained by recording the UV absorbance spectra over a range of temperatures (20–90°C) and subsequently taking the difference between each spectrum and that at 90°C. TDS provide specific signatures of different DNA and RNA structural conformations (34). Spectra were recorded between 220 and 320 nm on a Kontron-Uvikon 940 UV/Vis spectrophotometer using quartz cuvettes with an optical pathlength of 1 cm. DNA concentration was 4 μM.

Circular dichroism

Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using a 1-cm pathlength quartz cuvette with a reaction volume of 600 μl. The DNA oligonucleotides (4 μM) were prepared in a 10 mM lithium cacodylate pH 7.2 buffer containing 100 mM NaCl or KCl. The samples were heated at 90°C for 5 min and cooled down to room temperature overnight. 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.

Calorimetry

Microcalorimetry experiments were performed on a Nano DSC-II microcalorimeter as previously described (35). The oligonucleotides were prepared at concentrations ranging from 194 to 223 μM in a 10 mM lithium cacodylate pH 7.2 buffer containing 100 mM KCl. An average of six differential scanning calorimetric (DSC) heating and cooling profiles was taken.

NMR spectroscopy

Samples for NMR study were dialyzed successively against ~50 mM KCl solution and against water. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5–2.0 mM; the solutions contained 70 mM KCl and 20 mM potassium phosphate (pH 7). NMR experiments were performed on 600 MHz 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 (36), site-specific ^{2}H labeling (37), and through-bond correlations at natural abundance (38,39). Resonances for thymine residues were assigned following systematic T-to-U substitutions. Spectral assignments were completed by NOESY, COSY, TOCSY, {^{13}C-^{1}H}-HMBC and {^{13}C-^{1}H}-HSQC, as previously described (29,39). Inter-proton distances were deduced from NOESY experiments at various mixing times. All spectral analyses were performed using the FELIX (Felix NMR, Inc.) program.

Structure calculation

Inter-proton distances for the d[AGGG(CTAGGG)$_3$] quadruplex were deduced from NOESY experiments performed in H$_2$O (mixing time, 200 ms) and D$_2$O (mixing times, 100, 200 and 350 ms). Structure computations were performed using the XPLOR-NIH program (40) in three general steps essentially as previously described (29): (i) distance geometry simulated annealing, (ii) distance-restrained molecular dynamics refinement and (iii) relaxation matrix intensity refinement. Hydrogen bond restraints, inter-proton distance restraints, dihedral restraints, planarity restraints and repulsive restraints were imposed during structure calculations. Structures were displayed using the PyMOL program (41).

Data deposition

The coordinates for the d[AGGG(CTAGGG)$_3$] quadruplex have been deposited in the Protein Data Bank (accession code 2KM3).

RESULTS AND DISCUSSION

Sequences containing four human telomeric variant CTAGGG repeats form a new G-quadruplex fold in K$^+$ solution

The imino proton spectrum of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence, henceforth designated as *22CTA* (Table 1), in K$^+$ solution (Figure 1) was distinct from those observed for the canonical four-repeat TTAGGG human telomeric sequences (22–29) and indicated the formation of a new G-quadruplex structure: eight major peaks at 10.8–11.9 p.p.m. corresponded to the formation of two G-tetrads and three major peaks at 12.8-13.4 p.p.m. corresponded to three Watson–Crick G•C base pairs (see below). The eight former peaks remained sharp at 45°C (Supplementary Figure S2), consistent with their G-tetrad origin (29). Minor G-quadruplex form(s) were also present in *22CTA*, as shown by minor peaks at 10.6–12.2 p.p.m., whose relative
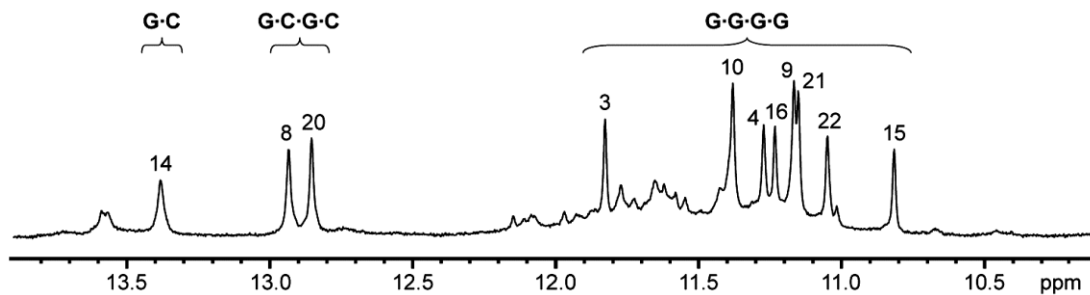
Figure 1. Imino proton spectrum of the 22-nt human telomeric variant d[AGGG(CTAGGG)₃] sequence (*22CTA*) in K⁺ solution with assignments for the major form listed over the spectrum. The imino protons are classified into three categories, corresponding to their involvement in G•G•G•G tetrad, G•C•G•C tetrad, or G•C base pair formation.

Table 1. Representative DNA sequences used in this study[a]

Name	Sequence							
22wt	A	GGG	TTA	GGG	TTA	GGG	TTA	GGG
22CTA	A	GGG	<u>C</u>TA	GGG	<u>C</u>TA	GGG	<u>C</u>TA	GGG
23CTA	A	GGG	<u>C</u>TA	GGG	<u>C</u>TA	GGG	<u>C</u>TA	GGG	C
22TCA	A	GGG	<u>TC</u>A	GGG	<u>TC</u>A	GGG	<u>TC</u>A	GGG

[a]Variations from the canonical repeat are underlined.

population (as compared to the major form) augmented when the temperature was increased (Supplementary Figure S2). Imino proton spectra (Supplementary Figure S3) of many different sequences containing four human telomeric variant <u>C</u>TAGGG repeats with different flanking ends (Supplementary Table S1) showed eight major peaks at 10.8–11.9 p.p.m. along with major peaks at 12.8–13.6 p.p.m., suggesting that they also adopted predominantly a G-quadruplex containing Watson–Crick G•C base pairs in K⁺ solution. A preliminary analysis on a NOESY spectrum of the *23CTA* sequence (Table 1) suggested that this sequence adopted in K⁺ solution the same overall G-quadruplex fold as *22CTA* (see below).

The imino proton spectrum of *22CTA* in Na⁺ solution (Supplementary Figure S4) indicated the presence of a major G-quadruplex form and minor conformation(s), but major sharp peaks around 13 p.p.m. (characteristic of stable Watson–Crick G•C base pairs) were not observed.

Native gel electrophoresis analysis

The molecular size of G-quadruplexes formed by variant human telomeric sequences (Table 1) was probed by native PAGE. In both Na⁺ and K⁺ solution, a single major band was observed for each of the 22-nt human telomeric sequences (Figure 2), *22wt* (containing TTAGG repeats), *22CTA* (containing <u>C</u>TAGGG repeats) and *22TCA* (containing T<u>C</u>AGGG repeats). The rate of migration of the band was independent of oligonucleotide concentrations (from 0.05 to 80 μM), arguing for an intramolecular structure for each of the three sequences. This was further substantiated by the fast migration of the band in relation to those of duplexes (9 and 12 bp), oligothymidylate sequences (15, 21 and 30 nt), and a 22-nt control sequence (*22AgMut4*) incapable of forming a quadruplex.
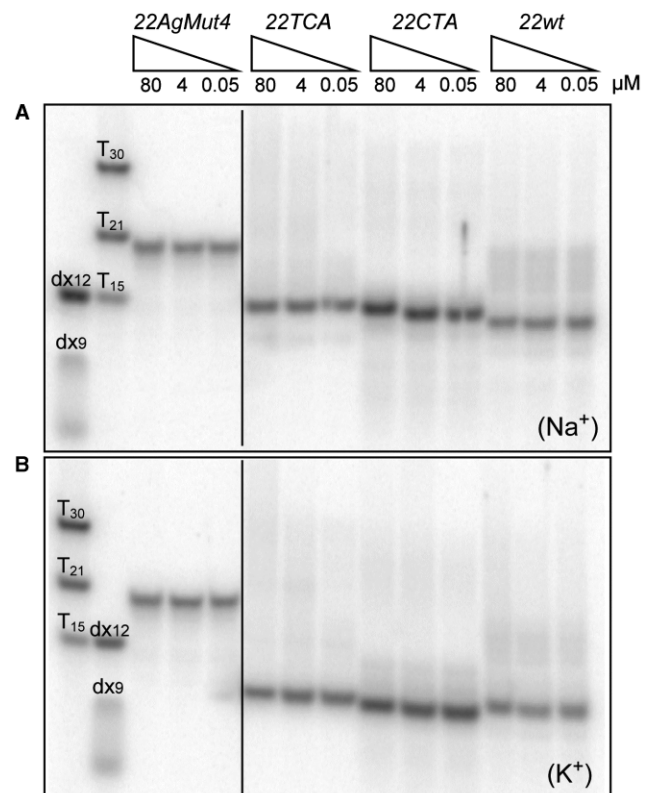

Figure 2. Non-denaturing PAGE analysis of the *22wt*, *22CTA* and *22TCA* human telomeric sequences (Table 1), which were pre-incubated in a 10 mM Tris–HCl pH 7.5 buffer supplemented with (**A**) 100 mM Na⁺ or (**B**) 100 mM K⁺, then loaded on a 15% polyacrylamide gel supplemented with 20 mM of the corresponding salt, and run at 26°C. Migration markers are provided on the left: oligothymidylate sequences (dT₁₅, dT₂₁ and dT₃₀), duplexes (dx9: d[GCGATACGG] + d[CCGTA TCGC] and dx12: d[GCGTGACTTCGG] + d[CCGAAGTCACGC]), and single-stranded *22AgMut4* control sequence d[ATGGTTAGTGT TAGGTTTAGTG] incapable of forming a quadruplex. Note that with respect to the markers, *22wt*, *22CTA* and *22TCA* human telomeric sequences migrate faster in K⁺ solution than in Na⁺ solution.

Thermal stability: UV absorption and calorimetry studies

Melting experiments were conducted for variant human telomeric sequences by monitoring the UV absorbance at 295 nm (32). All transitions were reversible, indicating that the denaturation curves corresponded to a true equilibrium process (Supplementary Figure S5).

Table 2. Melting temperatures (°C) and thermodynamic parameters for quadruplex dissociation of various human telomeric sequences in 100 mM K^+ solution, as measured by 295-nm UV absorbance and DSC

Oligo[a]	UV-melting			DSC: average excess enthalpy/entropy[d]			DSC: deconvolution general model, 1 transition[e]		
	T_m[b] (°C)	ΔH°_{VH} (kcal mol^{-1})	ΔS°_{VH} (cal K^{-1} mol^{-1})	ΔH_{cal} (kcal mol^{-1})	ΔS_{cal} (cal K^{-1} mol^{-1})	T_t(°C)	ΔH (kcal mol^{-1})	ΔCp (kcal K^{-1} mol^{-1})	T_m (°C)
22wt	64	54.8 ± 3.1	163 ± 9	39.6 ± 3.0	116 ± 9	66.9 ± 0.5	51.6 ± 0.9	−0.54 ± 0.18	68.7 ± 0.5
22CTA	62	53.7 ± 0.8	157 ± 4	41.7 ± 3.9	123 ± 12	65.4 ± 0.5	49.8 ± 0.6	−0.45 ± 0.15	66.9 ± 0.4
23CTA	64	53.0 ± 3.2 [c]	157 ± 10[c]	47.4 ± 4.8	139 ± 14	67.1 ± 0.5	55.1 ± 1.0	−0.61 ± 0.21	68.3 ± 0.3

[a]Concentration of *22wt*, *22CTA* and *23CTA* is 223, 194 and 205 µM, respectively.
[b]T_m measured for 5 µM of *22wt*, *22CTA* and *23CTA* is 63.5, 61 and 63°C, respectively.
[c]ΔH°_{VH} and ΔS°_{VH} of 23CTA in KCl are provided for illustration only; in this case, the ln$K$ versus $1/T$ graph significantly deviates from linearity.
[d]$\Delta H_{cal} = \int Cp^{excess}(T)dT$ and $\Delta S_{cal} = \int \frac{Cp^{excess}(T)}{T}dT$ where Cp^{excess} is the excess heat capacity function.
[e]The general transition model directly fits the molar heat capacity Cp (and not the excess heat capacity Cp^{excess}). It is used for transitions with $\Delta Cp \neq 0$. In this model, $\Delta Cp(T)$ is fitted with a second-order polynomial: $\Delta Cp(T) = a + bT + cT^2 = \Delta Cp(T_m) + b(T - T_m) + c(T^2 - T_m^2)$. ΔH and ΔCp are given at $T = T_m$.

The intramolecular nature of G-quadruplex formation was evaluated by varying the oligonucleotide concentration (Table 2). No significant difference (1°C or less) in melting temperature was observed upon about 40-fold increase in concentration (from 5 to ∼200 µM), indicating that the G-quadruplex folding was indeed intramolecular. The variant *22CTA* sequence was found to be marginally less stable than the canonical *22wt* sequence in 20 and 100 mM KCl (Table 2 and data not shown). The stability of both *22CTA* and *22wt* was lower in NaCl than in KCl (data not shown), as usually observed for the formation of G-quadruplexes. Note that the 23-nt human telomeric variant sequence *23CTA* was slightly more stable than *22CTA*. Thermodynamic parameters of the variant human telomeric sequences in K^+ solution, extracted from the UV-melting experiments (32), are listed in Table 2.

Microcalorimetry experiments (35) were performed on the same set of oligonucleotides in a buffer containing 100 mM KCl (Supplementary Figure S6). Analysis of DSC scans confirmed that *22CTA* was slightly less stable than *22wt*. Both transitions gave similar enthalpy values (∼40 kcal/mol, determined using a calorimetry model-independent analysis, as compared to ∼50 kcal/mol, with a van t' Hoff analysis; Table 2). A similar difference in enthalpy has been reported for the wild-type sequence (42).

TDS and CD signatures

The TDS for both *22CTA* and *22wt* in K^+ solution (Figure 3A and B; Supplementary Figure S7) exhibited typical patterns of a G-quadruplex structure, with two positive maxima at 240 and 275 nm and a negative minimum around 295 nm (34). There was an isosbestic point for both sequences around 285 nm, but the shapes of the two TDS differed around 255 nm.

CD spectra for both sequences in K^+ solution (Figure 3C and D; Supplementary Figure S8) were in agreement with the formation of G-quadruplexes. The CD profile of *22CTA* (with a very negative peak around 260 nm) differed significantly from that of the wild-type sequence *22wt*, consistent with the formation of a different fold.

NMR spectral assignments

Guanine imino and H8 protons of *22CTA* were unambiguously assigned using site-specific low-enrichment ^{15}N labeling (36), site-specific ^2H labeling (37) and through-bond correlations at natural abundance (38,39) (Figure 4; Supplementary Table S2). Resonances for thymine residues were unambiguously assigned by T-to-U substitutions (39) (Supplementary Figure S9 and Table S2). With the help of these unambiguous assignments and other through-bond correlation experiments (COSY, TOCSY, {^{13}C-^1H}-HMBC and {^{13}C-^1H}-HSQC) (data not shown), the classical H8/H6-H1' NOE sequential connectivity could be traced from A1 through G22 (Figure 5). The intensity of intraresidue H8-H1' NOE cross-peaks (Supplementary Figure S10) indicated *syn* glycosidic conformation for G3, G9, G15 and G21, in contrast to other residues, which adopted *anti* conformation. These results were supported by NMR spectra of modified sequences in which G3, G9, G15 and G21 were substituted by 8-bromoguanine (^{Br}G) (Supplementary Figure S11 and data not shown).

22CTA forms a chair-type G-quadruplex containing a G·C·G·C tetrad

Analysis of imino-H8 connectivity patterns (Figure 6A) revealed the formation of an intramolecular chair-type G-quadruplex with two G-tetrad layers, G4·G21·G16·G9 and G3·G10·G15·G22 (Figure 6B and D): the hydrogen-bond directionalities of the G-tetrads alternate anticlockwise and clockwise; the glycosidic conformations of guanines around each tetrad are *syn·anti·syn·anti*; each strand is antiparallel to the two adjacent strands; there are two narrow and two wide grooves; the three connecting loops are edgewise; the first and third loops (on the top) span narrow grooves, while the second loop (at the bottom) spans a wide groove. The observation of the imino protons of G8, G20 and G14 at 12.8–13.4 p.p.m. (Figure 1) and their strong NOE cross-peaks to the amino protons of C17, C5 and C11,
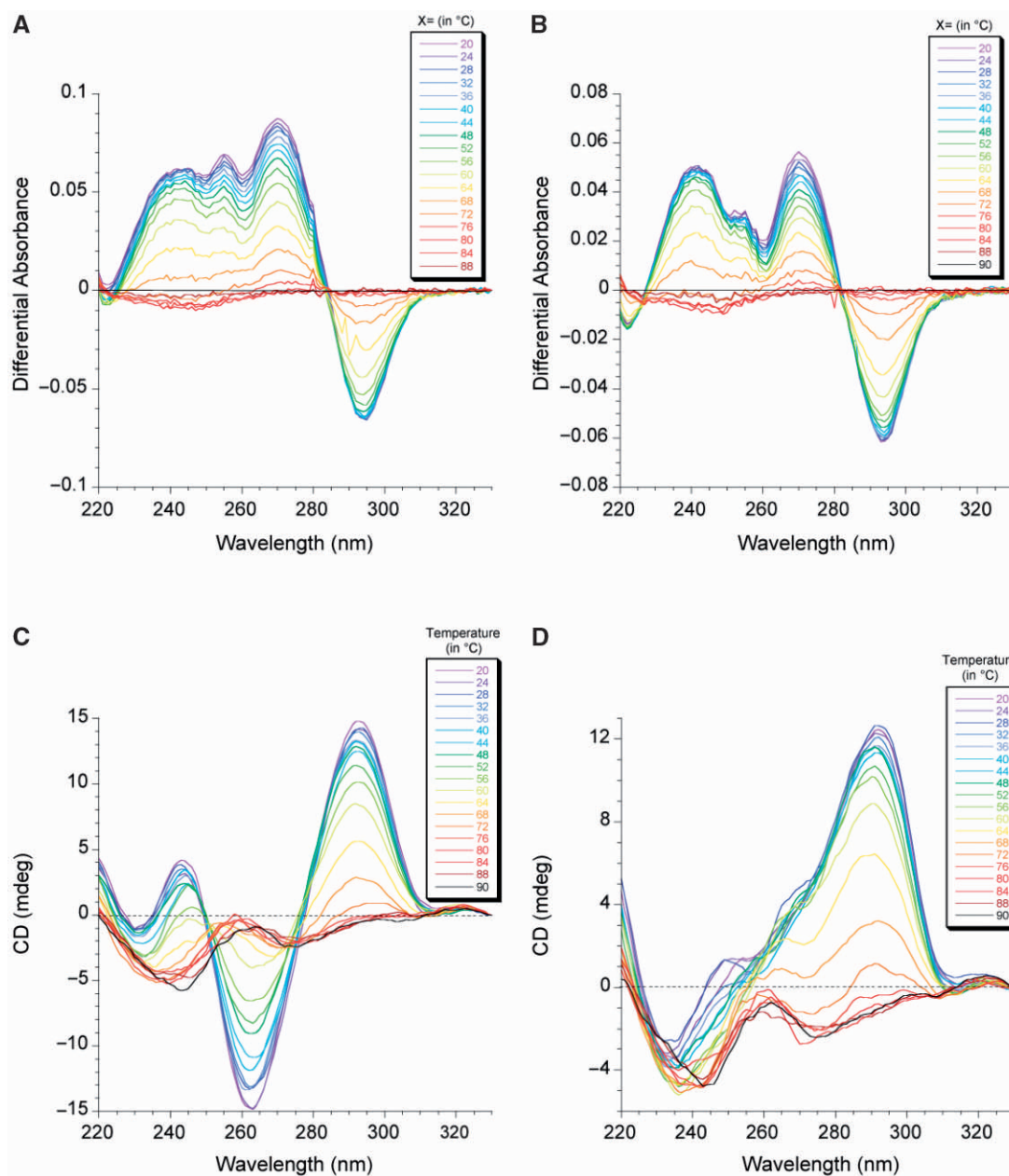
Figure 3. TDS (top panels) and CD spectra (bottom panels) recorded at different temperatures (color coded on the top right corner of each spectrum) for (**A** and **C**) *22CTA* and (**B** and **D**) *22wt*. TDS were obtained by substracting the UV absorbance at X°C from that at 90°C. DNA concentration was 4 μM; solution contained 100 mM KCl and 10 mM lithium cacodylate, pH 7.2.

respectively, (Figure 6A) indicated the formation of three Watson–Crick G•C base pairs: G8•C17 and G20•C5 on the top and G14•C11 at the bottom of the G-tetrad core. The two former base pairs are further aligned to form a G•C•G•C tetrad (Figure 6C). Note that all three Watson–Crick G•C base pairs are formed across the wide grooves of the G-tetrad core. The central position of the G4•G21• G16•G9 tetrad (between the G3•G10•G15•G22 and the G8•C17•G20•C5 tetrads) is consistent with the observation of the imino protons of G4, G21, G16 and G9 being well protected from the exchange with solvent (Supplementary Figure S12).

Solution structure of the *22CTA* G-quadruplex

The structure of the *22CTA* G-quadruplex in K$^+$ solution (Figure 7) was calculated on the basis of NMR restraints (Table 3 and Supplementary Data). In the G-tetrad core, there is extensive stacking between five-membered rings of guanines (Supplementary Figure S13), which belong to the two G-tetrads with opposite hydrogen-bond directionalities (7,29).

Previous studies have demonstrated two different G•C•G•C tetrads that are aligned via the major groove edges of the Watson–Crick G•C base pairs: the direct
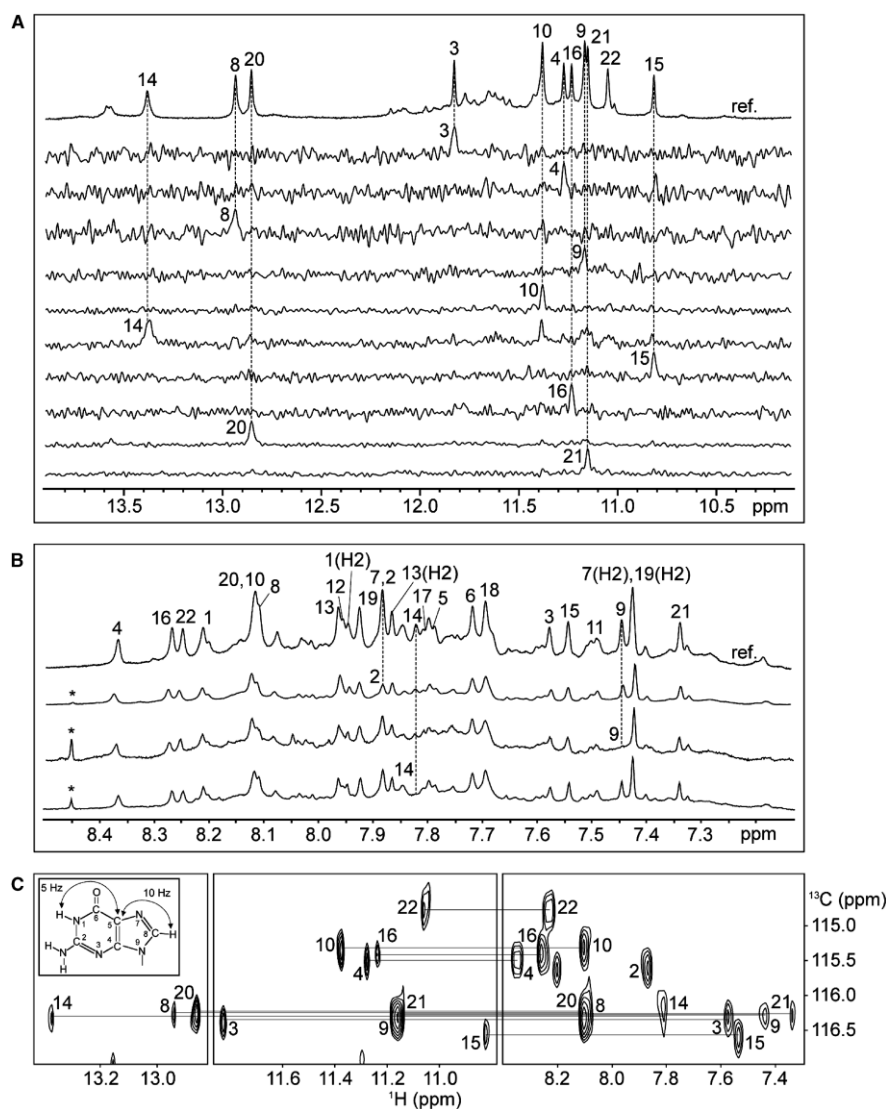
Figure 4. Guanine imino and H8 proton assignments of *22CTA* in K$^+$ solution. (**A**) Imino protons were assigned in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. The reference spectrum (ref.) is shown at the top. (**B**) Examples of H8 proton assignments by site-specific ^{2}H labeling at the indicated positions. The reference spectrum (ref.) is shown at the top. The peak from a contamination is marked by an asterisk. (**C**) 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.

alignment (Supplementary Figure S14A and C) observed in Na$^+$ solution (43–45) and the slipped alignment (Supplementary Figure S14B and D) observed in K$^+$ solution (46). The latter was proposed to provide a K$^+$ coordination site between the two G•C base pairs (Figure 6C; Supplementary Figure S14B) (46). In the direct alignment, the distances from the guanine H8 proton of one Watson–Crick G•C base pair to the cytosine H5 proton and the cytosine amino protons of the other Watson–Crick G•C base pair are ∼3 Å and ∼5 Å, respectively (Supplementary Figure S14C). Accordingly, the guanine H8 proton of one G•C base pair should exhibit a strong NOE to the cytosine H5 proton and weak NOEs to the cytosine amino protons of the other G•C base pair. The reverse scenario should be observed in the slipped alignment: the guanine H8

proton of each Watson–Crick G•C base pair is closer to the cytosine amino protons (∼3 Å) than to the cytosine H5 proton (∼5 Å) of the adjacent Watson–Crick G•C base pair (Supplementary Figure S14D). The G8•C17•G20•C5 tetrad at the top of the current structure adopted the slipped alignment (Figures 6–8). This alignment between the G8•C17 and G20•C5 base pairs was supported by the observations of strong NOEs from the H8 protons of G8 and G20 to the amino protons of C5 and C17 (Figure 6A, see figure legend), respectively, but very weak NOEs from the H8 protons of G8 and G20 to the H5 protons of C5 and C17 (Figure 5), respectively. NOEs from the imino protons of G8 and G20 to the imino protons of G16 and G4, respectively, define the position of the G8•C17•G20•C5 tetrad over the top G-tetrad.
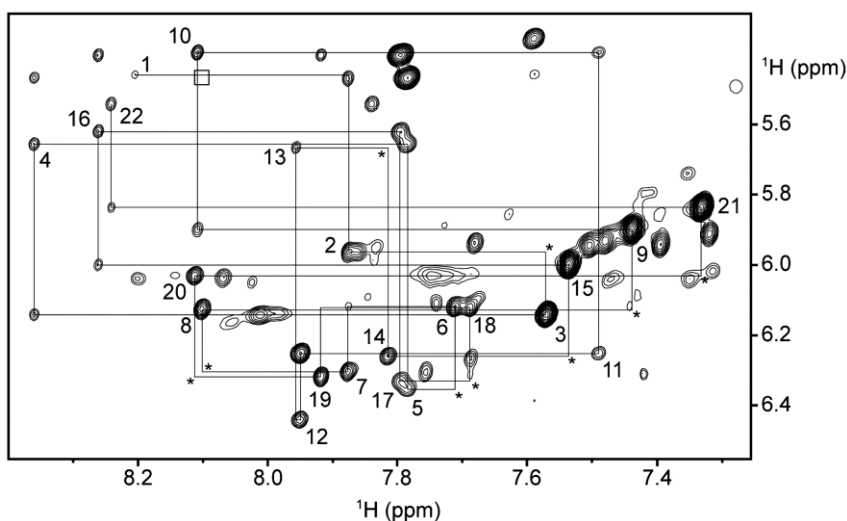
Figure 5. NOESY spectrum (mixing time, 350 ms) showing the H8/H6-H1' connectivity of *22CTA* in K$^+$ solution. Intraresidue H8/H6-H1' NOE cross-peaks are labeled with residue numbers. Weak or missing sequential connectivities are marked with asterisks. The box indicates the position of a very weak cross-peak (not seen at this threshold) corresponding to the NOE from the H8 proton of G8 to the H5 proton of C5, consistent with a slipped alignment for the G8•C17•G20•C5 tetrad (see text).

Two adenine bases, A7 and A19, further stack over the G8•C17•G20•C5 tetrad (Figure 8A–D), supported by the observation of NOEs from the A7 and A19 protons to the C5 and C17 protons, respectively. The two thymine bases, T6 and T18, are projected outward (Figure 8A–D), consistent with the observation of only a few NOEs between the protons of these bases and their neighboring residues (data not shown). Quasi-symmetry (39,47) was observed for the top part of the structure (Figures 7 and 8), consistent with its two halves (segments G4–G9 and G16–G21) displaying many similar spectral characteristics including chemical shifts and NOE patterns (Figures 1, 4–6; Supplementary Figure S15).

The G14•C11 base pair is located at the bottom of the G-tetrad core, directly under G10 and G15 (Figure 8E and F). This placement was supported by the observations of a number of NOE cross-peaks [e.g. a NOE between the imino protons of G14 and G10 (data not shown) and NOEs between the protons of C11 and the bottom G-tetrad (Figure 6A)] and the results from solvent exchange experiment showing that the imino proton of G10 was better protected than that of the three other guanines (G3, G15 and G21) from the same G-tetrad (Supplementary Figure S12). At the bottom loop, the A13 base stacks underneath the G14•C11 base pair, while the T12 base folds into a hydrophobic groove (29).

The *22CTA* G-quadruplex represents a robust fold

The structure of *22CTA* in K$^+$ solution consists of a two-G-tetrad antiparallel core sandwiched between a G•C base pair and a G•C•G•C tetrad. This robust fold appeared to be the preferred one for various four-repeat variant CTAGGG sequences with different flanking ends in K$^+$ solution, contrasting the situation of sequences containing four canonical TTAGGG repeats, which were shown to adopt different G-quadruplex folds (7–29). Our modeling and preliminary NMR spectral

analysis (including NOESY) suggested that sequences containing an additional C at the 3'-end (e.g. *23CTA*) could form a similar G-quadruplex with a second G•C•G•C tetrad at the bottom of the structure. The formation of the G14•C11•G2•C23 tetrad for *23CTA* would be consistent with its higher stability (Table 2) and considerable changes in the chemical shifts of G3 and G15 imino protons of the bottom G-tetrad (Supplementary Figure S3) as compared to *22CTA*. Such a core structure with two G-tetrads sandwiched between two G•C•G•C tetrads, which has already been observed previously (46), would provide a well-aligned channel for K$^+$ ion coordination. Our results indicate that G•C•G•C tetrads may positively contribute to G-quadruplex stability (43–46,48), in contrast to very detrimental effects observed for many tetrads of guanine derivatives (49).

Although the G-quadruplex fold of *22CTA* contains only two G-tetrad layers, its stability is comparable to that of the structures containing three G-tetrads (29). This can be explained by molecular interactions seen in our structure including the formation of a G•C base pair and a G•C•G•C tetrad, and additional base stacking interactions in the loops. The principle of this folding and base arrangement is analogous to that observed for Form 3 human telomeric G-quadruplex in K$^+$ solution (29), and reinforces the view that the overall folding topology of a G-quadruplex is defined not only by maximizing the number of G-tetrads, but also by maximizing all possible base pairing and stacking in the loops (29).

CONCLUSION

We have shown that sequences containing four human telomeric variant CTAGGG repeats adopt a novel antiparallel G-quadruplex involving two G-tetrads
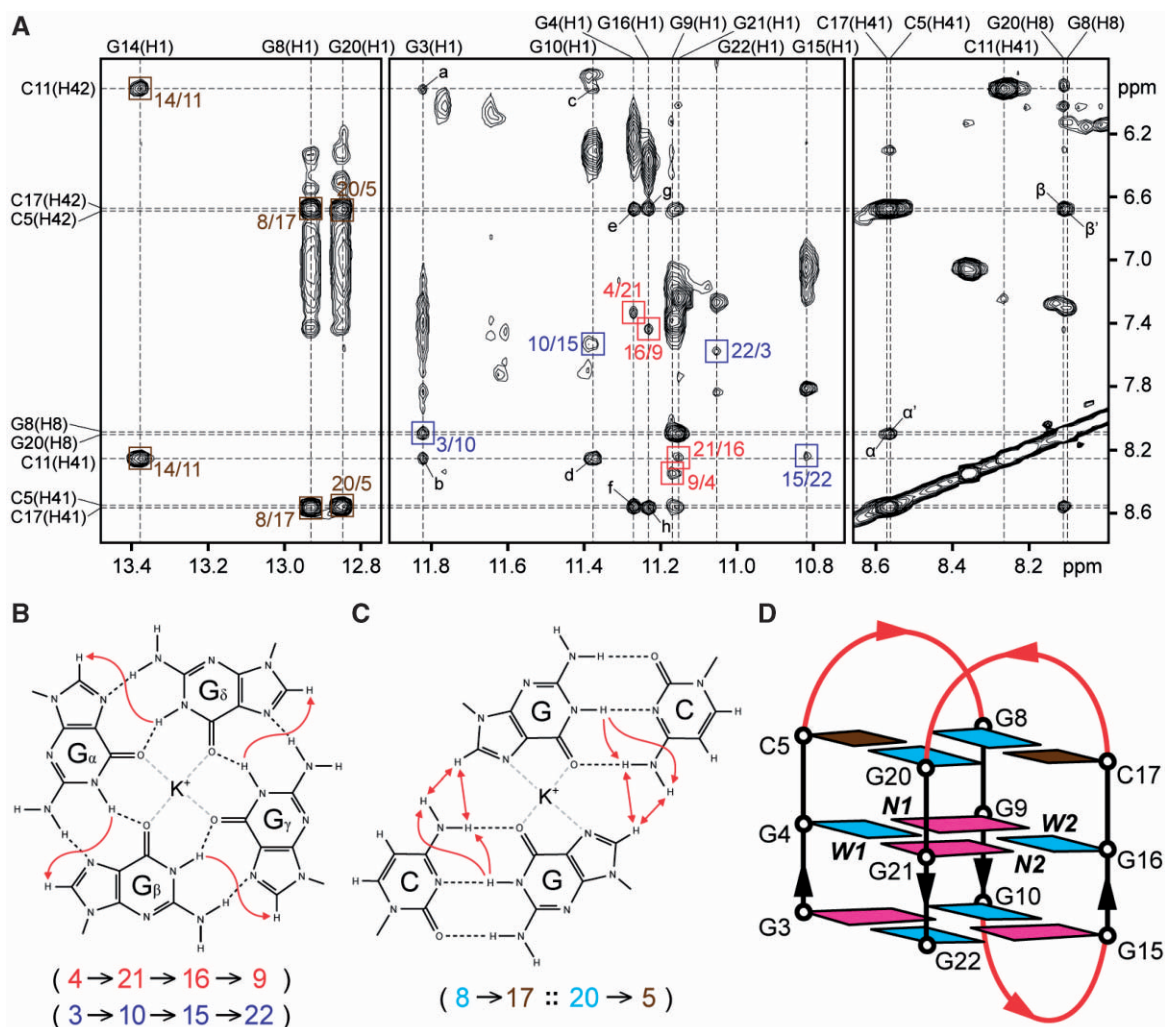
Figure 6. Formation of a chair-type G-quadruplex containing two G•G•G•G tetrads and a G•C•G•C tetrad by *22CTA* in K$^+$ solution. (**A**) NOESY spectrum (mixing time, 200 ms) showing the cross-peaks that identify the arrangement of two G-tetrads and three G•C base pairs. Cross-peaks arising from imino-H8 connectivity around the two G-tetrads are framed in red and blue, and labeled with the residue number of imino protons in the first position and that of H8 protons in the second position. Cross-peaks between guanine imino and cytosine amino protons from the three Watson–Crick G•C base pairs are framed in brown, and labeled with the residue number of guanine imino protons in the first position and that of cytosine amino protons in the second position. The NOE cross-peaks a to h are assigned as follows: a, G3(H1)-C11(H42); b, G3(H1)-C11(H41); c, G10(H1)-C11(H42); d, G10(H1)-C11(H41); e, G4(H1)-C5(H42); f, G4(H1)-C5(H41); g, G16(H1)-C17(H42); h, G16(H1)-C17(H41). The NOE cross-peaks α, α′, β and β′ correspond to G20(H8)-C17(H41), G8(H8)-C5(H41), G20(H8)-C17(H42) and G8(H8)-C5(H42), respectively. (**B**) Characteristic guanine imino-H8 NOE connectivity patterns around a G$_\alpha$•G$_\beta$•G$_\gamma$•G$_\delta$ tetrad as indicated with arrows, with the connectivities observed for the G4•G21•G16•G9 and G3•G10•G15•G22 tetrads shown below. (**C**) The slipped G•C•G•C tetrad arrangement of the two Watson–Crick G8•C17 and G20•C5 base pairs, as supported by NOE cross-peaks α, α′, β and β′ (represented by red double-headed arrows). NOEs from guanine imino protons to cytosine amino protons that establish the G8•C17 and G20•C5 base pairs are represented by single-headed arrows. (**D**) Schematic view of the *22CTA* G-quadruplex. *anti* and *syn* guanines are colored cyan and magenta, respectively, while cytosines are colored brown. *W1*, *W2*, *N1* and *N2* represent wide 1, wide 2, narrow 1 and narrow 2 grooves, respectively. The backbone of the core and loops is colored black and red, respectively.

sandwiched between a G•C base pair and a G•C•G•C tetrad. This structure, with its robust and unique structural features, and its unusual binding properties to the telomeric factor POT1, could be used as a new scaffold for the design of specific G-quadruplex ligands directed against human telomeric DNA.

SUPPLEMENTARY DATA

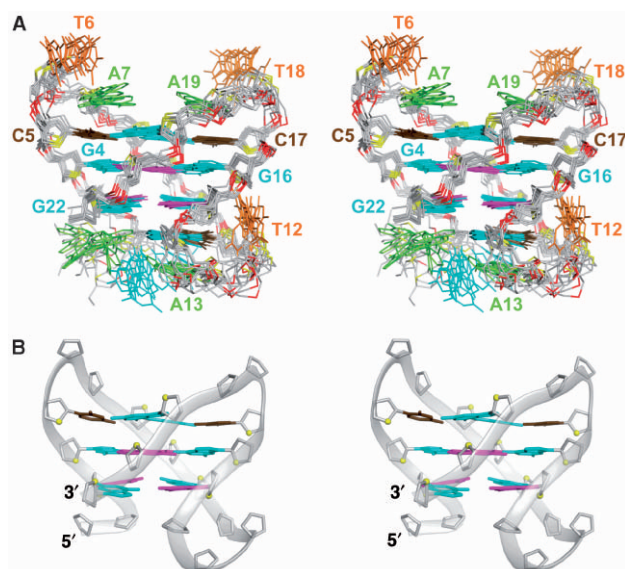Supplementary Data are available at NAR Online.

Figure 7. Stereo views of the *22CTA* G-quadruplex structure 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 are colored brown, adenines, green; thymines, orange; backbone and sugar, gray; O4' atoms, yellow; phosphorus atoms, red.

Table 3. Statistics of the computed structures of the 22-nt human telomeric variant d[AGGG(CTAGGG)$_3$] sequence

NMR restraints		
Distance restraints	D$_2$O	H$_2$O
Intraresidue distance restraints	203	5
Sequential (i, i + 1) distance restraints	179	29
Long-range (i, $\geq i$ + 2) distance restraints	34	72
Other restraints		
Hydrogen bond restraints	50	
Dihedral restraints	34	
Repulsive restraints[a]	6	
Intensity restraints		
Non-exchangeable Protons	108	
(each of three mixing times)		
Structure statistics for 10 molecules following intensity refinement		
NOE violations		
Number (>0.2Å)[b]	0.300 ± 0.458	
Maximum violation (Å)	0.168 ± 0.031	
RMSD of violations (Å)	0.019 ± 0.002	
Deviations from the ideal covalent geometry		
Bond lengths (Å)	0.004 ± 0.000	
Bond angles (deg)	0.746 ± 0.010	
Impropers (deg)	0.413 ± 0.010	
NMR R-factor (R$_{1/6}$)	0.016 ± 0.001	
Pairwise all heavy atom RMSD values (Å)		
All heavy atoms except A1 and G2	1.21 ± 0.14	
All heavy atoms	1.53 ± 0.25	

[a]Distance restraints between pairs of protons that do not exhibit NOE cross-peaks; these restraints are removed during relaxation matrix intensity refinement.
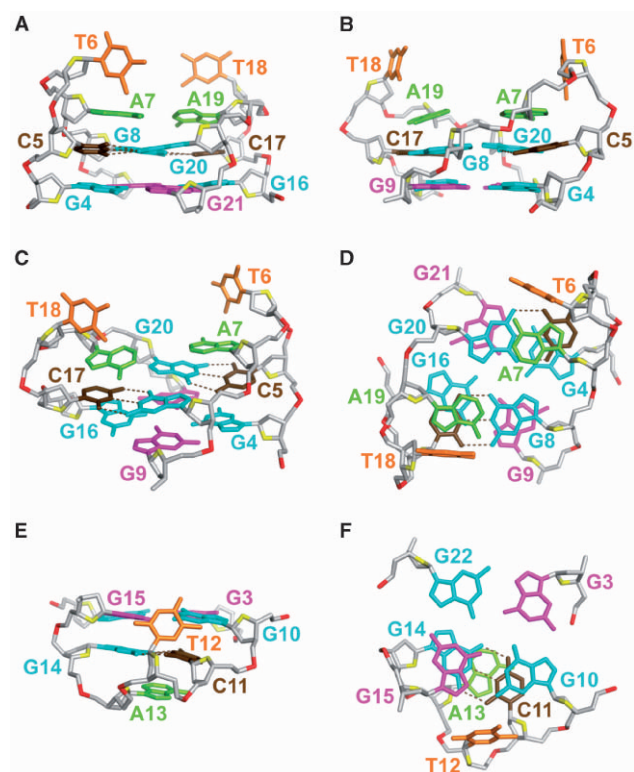[b]The total number of violations divided by the number of structures.

Figure 8. Detailed structure of the *22CTA* G-quadruplex in K$^+$ solution. (**A–D**) The top half: the segments G4-G9 and G16-G21, showing the G4•G21•G16•G9 tetrad, the slipped G8•C17•G20•C5 tetrad, and the two quasi-symmetric loops. (A) View from the side looking into the wide 1 (*W1*) groove. (B) View from the side looking into the narrow 1 (*N1*) groove. (C) Side view from a different angle. (D) Top-down view. (**E** and **F**) The bottom half: the G3•G10•G15•G22 tetrad, the G14•C11 base pair, and the bottom loop, looking from the side (E) and from the top down (F). Color coded as in Figure 7A. Hydrogen-bonds in the Watson–Crick G•C base pairs are shown by dotted lines.

Conflict of interest statement. None declared.

REFERENCES

1. Wellinger,R.J. and Sen,D. (1997) The DNA structures at the ends of eukaryotic chromosomes. *Eur. J. Cancer*, **33**, 735–749.
2. De Cian,A., Lacroix,L., Douarre,C., Temime-Smaali,N., Trentesaux,C., Riou,J.F. and Mergny,J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
3. Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L. and Wu,J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)$_n$, present at the telomeres of human chromosomes. *Proc. Natl Acad. Sci. USA*, **85**, 6622–6626.
4. Allshire,R.C., Dempster,M. and Hastie,N.D. (1989) Human telomeres contain at least 3 types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.*, **17**, 4611–4627.
5. Baird,D.M., Jeffreys,A.J. and Royle,N.J. (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J.*, **14**, 5433–5443.
6. Mendez-Bermudez,A., Hills,M., Pickett,H.A., Phan,A.T., Mergny,J.L., Riou,J.F. and Royle,N.J. (2009) Human telomeres that contain (CTAGGG)$_n$ repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res.*, **37**, 6225–6238.
7. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG$_3$(T$_2$AG$_3$)$_3$] G-tetraplex. *Structure*, **1**, 263–282.

8. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.

9. Ying,L., Green,J.J., Li,H., Klenerman,D. and Balasubramanian,S. (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA*, **100**, 14629–14634.

10. Redon,S., Bombard,S., Elizondo-Riojas,M.A. and Chottard,J.C. (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG₃(T2AG₃)₃ and (T₂AG₃)₄ human telomere sequences in Na$^+$ and K$^+$ solutions. *Nucleic Acids Res.*, **31**, 1605–1613.

11. Risitano,A. and Fox,K.R. (2005) Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg. Med. Chem. Lett.*, **15**, 2047–2050.

12. Rezler,E.M., Seenisamy,J., Bashyam,S., Kim,M.Y., White,E., Wilson,W.D. and Hurley,L.H. (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.*, **127**, 9439–9447.

13. Rujan,I.N., Meleney,J.C. and Bolton,P.H. (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res.*, **33**, 2022–2031.

14. Qi,J. and Shafer,R.H. (2005) Covalent ligation studies on the human telomere quadruplex. *Nucleic Acids Res.*, **33**, 3185–3192.

15. Vorlickova,M., Chladkova,J., Kejnovska,I., Fialova,M. and Kypr,J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.*, **33**, 5851–5860.

16. Li,J., Correia,J.J., Wang,L., Trent,J.O. and Chaires,J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.*, **33**, 4649–4659.

17. Lee,J.Y., Okumus,B., Kim,D.S. and Ha,T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl Acad. Sci. USA*, **102**, 18938–18943.

18. Xue,Y., Kan,Z.Y., Wang,Q., Yao,Y., Liu,J., Hao,Y.H. and Tan,Z. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K$^+$ solution under molecular crowding condition. *J. Am. Chem. Soc.*, **129**, 11185–11191.

19. Gaynutdinov,T.I., Neumann,R.D. and Panyutin,I.G. (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. *Nucleic Acids Res.*, **36**, 4079–4087.

20. Amrane,S., Ang,R.W., Tan,Z.M., Li,C., Lim,J.K., Lim,J.M., Lim,K.W. and Phan,A.T. (2009) A novel chair-type G-quadruplex formed by a *Bombyx mori* telomeric sequence. *Nucleic Acids Res.*, **37**, 931–938.

21. Xu,Y., Noguchi,Y. and Sugiyama,H. (2006) The new models of the human telomere d[AGGG(TTAGGG)₃] in K$^+$ solution. *Bioorg. Med. Chem.*, **14**, 5584–5591.

22. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.A. and Yang,D. (2006) Human telomeric sequence forms a hybrid-type intra-molecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.

23. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K$^+$ solution: an intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.

24. Phan,A.T., Luu,K.N. and Patel,D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K$^+$ solution. *Nucleic Acids Res.*, **34**, 5715–5719.

25. Dai,J., Punchihewa,C., Ambrus,A., Chen,D., Jones,R.A. and Yang,D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple. *Nucleic Acids Res.*, **42**, 481–490.

26. Matsugami,A., Xu,Y., Noguchi,Y., Sugiyama,H. and Katahira,M. (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K$^+$ solution. *FEBS J.*, **274**, 3545–3556.

27. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K$^+$ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.

28. Phan,A.T., Kuryavyi,V., Luu,K.N. and Patel,D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K$^+$ solution. *Nucleic Acids Res.*, **35**, 6517–6525.

29. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K$^+$ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.

30. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.

31. Guédin,A., De Cian,A., Gros,J., Lacroix,L. and Mergny,J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686–696.

32. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.

33. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.

34. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.

35. Amrane,S., Saccà,B., Mills,M., Chauhan,M., Klump,H.H. and Mergny,J.L. (2005) Length-dependent energetics of (CTG)ₙ and (CAG)ₙ trinucleotide repeats. *Nucleic Acids Res.*, **33**, 4065–4077.

36. Phan,A.T. and Patel,D.J. (2002) A site-specific low-enrichment ^{15}N,^{13}C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.

37. Huang,X., Yu,P., LeProust,E. and Gao,X. (1997) An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res.*, **25**, 4758–4763.

38. Phan,A.T. (2000) Long-range imino proton-^{13}C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175–178.

39. Phan,A.T., Guéron,M. and Leroy,J.L. (2001) Investigation of unusual DNA motifs. *Methods Enzymol.*, **338**, 341–371.

40. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.

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

42. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.

43. Kettani,A., Kumar,R.A. and Patel,D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.*, **254**, 638–656.

44. Kettani,A., Bouaziz,S., Gorin,A., Zhao,H., Jones,R.A. and Patel,D.J. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G·G·G·G and G·C·G·C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.*, **282**, 619–636.

45. Zhang,N., Gorin,A., Majumdar,A., Kettani,A., Chernichenko,N., Skripkin,E. and Patel,D.J. (2001) Dimeric DNA quadruplex containing major groove-aligned A·T·A·T and G·C·G·C tetrads stabilized by inter-subunit Watson-Crick A·T and G·C pairs. *J. Mol. Biol.*, **312**, 1073–1088.

46. Bouaziz,S., Kettani,A. and Patel,D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.*, **282**, 637–652.

47. Phan,A.T., Guéron,M. and Leroy,J.L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.

48. Mergny,J.L., De Cian,A., Amrane,S. and Webba da Silva,M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.

49. Gros,J., Rosu,F., Amrane,S., De Cian,A., Gabelica,V., Lacroix,L. and Mergny,J.L. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.*, **35**, 3064–3075.