G-quadruplex structure of the *C. elegans* telomeric repeat: a two tetrads basket type conformation stabilized by a non-canonical C–T base-pair

Julien Marquevielle[1,2,†], Aurore De Rache[1,2,3,†], Brune Vialet[1], Estelle Morvan[2], Jean-Louis Mergny ⬤[1,2,4,*] **and Samir Amrane** ⬤[1,2,*]

[1]Univ. Bordeaux, Inserm U1212, CNRS UMR 5320, ARNA laboratory, 146 rue Léo Saignat F-33000 Bordeaux, France, [2]Institut Européen de Chimie et Biologie, UMS 3033 US001, CNRS-Université de Bordeaux, 2 rue Robert Escarpit, F-33600 Pessac, France, [3]Department of Chemistry, UNamur, 61 rue de Bruxelles, B-5000 Namur, Belgium and [4]Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, 91128 Palaiseau, France

ABSTRACT

The *Caenorhabditis elegans* model has greatly contributed to the understanding of the role of G-quadruplexes in genomic instability. The GGCTTA repeats of the *C. elegans* telomeres resemble the GGGTTA repeats of the human telomeres. However, the comparison of telomeric sequences (*Homo sapiens*, *Tetrahymena*, *Oxytricha*, *Bombyx mori* and *Giardia*) revealed that small changes in these repeats can drastically change the topology of the folded G-quadruplex. In the present work we determined the structure adopted by the *C. elegans* telomeric sequence d[GG(CTTAGG)₃]. The investigated *C. elegans* telomeric sequence is shown to fold into an intramolecular two G-tetrads basket type G-quadruplex structure that includes a C–T base pair in the diagonal loop. This work sheds light on the telomeric structure of the widely used *C. elegans* animal model.

GRAPHICAL ABSTRACT

C. Elegans Telomeric repeat

5' GGCTTAGGCTTAGGCTTAGG 3'
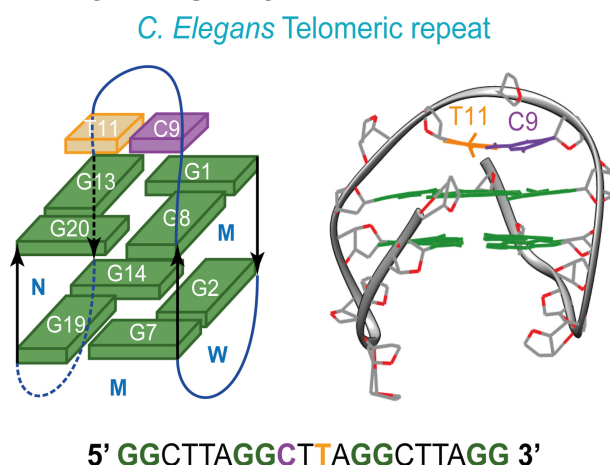
INTRODUCTION

Telomeric sequences constitute the nucleic part of the telomere, a nucleoprotein complex capping and protecting the extremities of linear chromosomes in eukaryotes. Most telomeric sequences contain a G-rich repetitive motif (1) and are able to form G-quadruplexes (2). These nucleic acid structures, associated to a wide variety of biological processes and pathologies (3), are formed by the stacking of a minimum of two G-tetrads. Each G-tetrad consists of four guanines interacting through a Hoogsteen type H-bonds network. Thanks to the development of novel biological probes, direct evidence for the *in vivo* existence of G-quadruplexes at telomeres has been obtained (reviewed here (4)). For instance, antibodies raised against telomeric G-quadruplexes revealed the presence of G-quadruplexes

---

*To whom correspondence should be addressed. Tel: +33 5 40 00 22 24; Email: samir.amrane@u-bordeaux.fr
Correspondence may also be addressed to Jean-Louis Mergny. Tel: + 33 1 69 33 50 01; Email: jean-louis.mergny@inserm.fr
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

in *Stylonychia lemnae* cells ([5]). In another strategy, G-quadruplexes were detected in the telomeres of human T98G cells by using a specific G-quadruplex tritium-labeled ligand ([3]H-360A ([6])). Furthermore, electron microscopic experiments showed G-quadruplex formation in a long single-stranded telomeric fragment indicating the formation of higher order bead-like structures ([7]). This condensation into G-quadruplex beaded filaments has also been confirmed by other biophysical approaches ([8,9]). The ability of telomeric sequences to form G-quadruplexes *in vitro* is conserved in highly divergent organisms. Indeed, in addition to the G_3T_2A motif (reviewed in ([10])) which is present in many organisms (from fungi ([11]) to human ([12])) many other G-quadruplex structures formed by other telomeric sequences have been investigated: *Tetrahymena* ([13]), *Oxytricha* ([14]), *Bombyx* ([15]), *Giardia* ([16]).

All these telomeric G-quadruplex forming motifs are closely related; nevertheless, the structural diversity in this set of sequences is quite high (Figure 1). Along the years, various modulations of the human telomeric sequence have been studied and a high diversity of G-quadruplex topologies were reported ([17,18]) (Figure 1 A-H): One 2+2 anti-parallel (Figure 1 A) ([19,20]), one 3 tetrads basket type anti-parallel (Figure 1 B) ([19,20]), two 3+1 (hybrid) forms ([21–23]) with three G-tetrads (Figure 1 C,D), one parallel ([24,25]) (Figure 1 E), two anti-parallel basket type with 2 G-tetrads ([26,27]) (Figure 1F and G). A variant human telomeric sequence (G_3CTA repeats) was also reported to fold into a chair type antiparallel G-quadruplex with two G-tetrads and one G:C:G:C tetrad ([28]) (Figure 1H). In the case of the non-human telomeric repeats (Figure 1I–L), the *Oxytricha* G_4T_4 telomeric sequence folds into a basket type G-quadruplex with 4 G-tetrads ([14]) (Figure 1I). Differing from the human telomeric G_3T_2A sequence only by removal of one guanine of each G-track, the *Bombyx mori* telomeric G_2T_2A sequence folds into a completely different topology, namely a chair-type anti-parallel G-quadruplex ([15]) (Figure 1J). Finally, the *Giardia* G_3TA telomeric sequence forms two intramolecular G-quadruplexes: a basket type with 2 G-tetrads and a parallel-stranded structure with 3 G-tetrads ([16]) (Figure 1K and E). The *Tetrahymena* G_4T_2 telomeric sequence, which differs from the human one by a single guanine to adenine substitution, adopts a 3+1 G-quadruplex structure with 3 G-quartets ([13]) (Figure 1L). When looking closely at these structures, the topological diversity is even more pronounced. For instance the 2-tetrads basket type conformations from the Human telomeric motifs (Figure 1F and G) form very distinctive structures. In Figure 1F, the first and last lateral loops span a wide and narrow groove respectively and the loop progression is anti-clockwise. In Figure 1G, it is exactly the reverse while it adopts the same topology as the telomeric repeat of *Giardia* (Figure 1K). Likewise, the human variant chair type structure (Figure 1H) differs from the chair type structure adopted by *Bombyx mori* (Figure 1G). They both have three lateral loops, however in the human conformation, the first and last lateral loops span a narrow groove and the middle one a wide groove, with a clockwise loop progression, while it is exactly the reverse for *Bombyx mori*. The two hybrid structures (Figure 1C, D) also differ in the position of

the propeller loop and the sizes of the grooves. Depending on the number of tetrads and the specific conformation, almost all these motifs form unique G-quadruplex topologies highlighting the tremendous polymorphism of these structures.

C. elegans is a widely used biological model system both in general terms and in the G-quadruplex field. Indeed not only *Caenorhabditis elegans* allowed the first demonstration of replication-associated instability at G-quadruplex motifs in 2008 ([29]), but the understanding of the implication of polymerase θ in a repair pathway associated with this instability was also obtained using this model ([30]). While *C. elegans* is not a pathogen, a number of nematodes are parasites for plants and animals, and several species infect humans, including ascarids (e.g. *Ascaris lumbricoides*), filarias, hookworms, pinworms and whipworms (e.g. *Trichuris trichiura*). Most, if not all of these nematodes share the same telomeric motif, composed of TTAGGC repeats.

Surprisingly, the d[GG(CTTAGG)$_3$] sequence of nematodes is still missing in the overview of the G-quadruplex structures formed by telomeric sequences provided above. The *C. elegans* telomeric sequence corresponds to the replacement in the human telomeric GGGTTA repeats of the first guanine of each track by a cytosine. In terms of G-quadruplex folding propensity, an interesting feature of this sequence is its relative poorness in G residues (with only 33% of Gs) as compared to other G-quadruplexes, which translates into a relatively low G4Hunter score ([31]): d[GG(CTTAGG)$_3$] has a score of only 0.65, generally considered to be too low to allow stable G4 formation. In this work, we investigate the structure of the G-quadruplex formed by this sequence.

MATERIALS AND METHODS

Oligonucleotide synthesis and samples preparation

All oligonucleotides were ordered from Eurogentec or synthesized by the phosphoramidite methodology using an automated Expedite 8909 DNA synthesizer at one micromole scale. The SynBase CPG 1000 Å primer support, the $^{8-Br}$dG^{dmf} amidite precursor, the phosphoramidites reagents (dA^{Bz}, dT, dG^{dmf} and dC^{Ac}) and solvents were purchased from Glen Reasearch (VA, USA) and LGC LINK (Scotland). The ^{15}N labelled sequences were synthesized according to previously published methodology ([31]). Incubation in ammonium hydroxide at room temperature for 24 h was used to cleave and deprotect the oligonucleotides. All oligonucleotide solutions were prepared in 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. The sample concentrations were determined from the absorbance at 260 nm and their absorption coefficients determined from ([32]) (see Supplementary Table S-1).

UV melting and thermal difference spectra (TDS)

A SAFAS UVmc2 double-beam spectrophotometer (Monte Carlo, Monaco) equipped with a 10-cells holder regulated by a Peltier controller was used to perform the UV melting experiments. The oligonucleotide strand concentration ranged between 3.3 and 7.2 μM. The melting
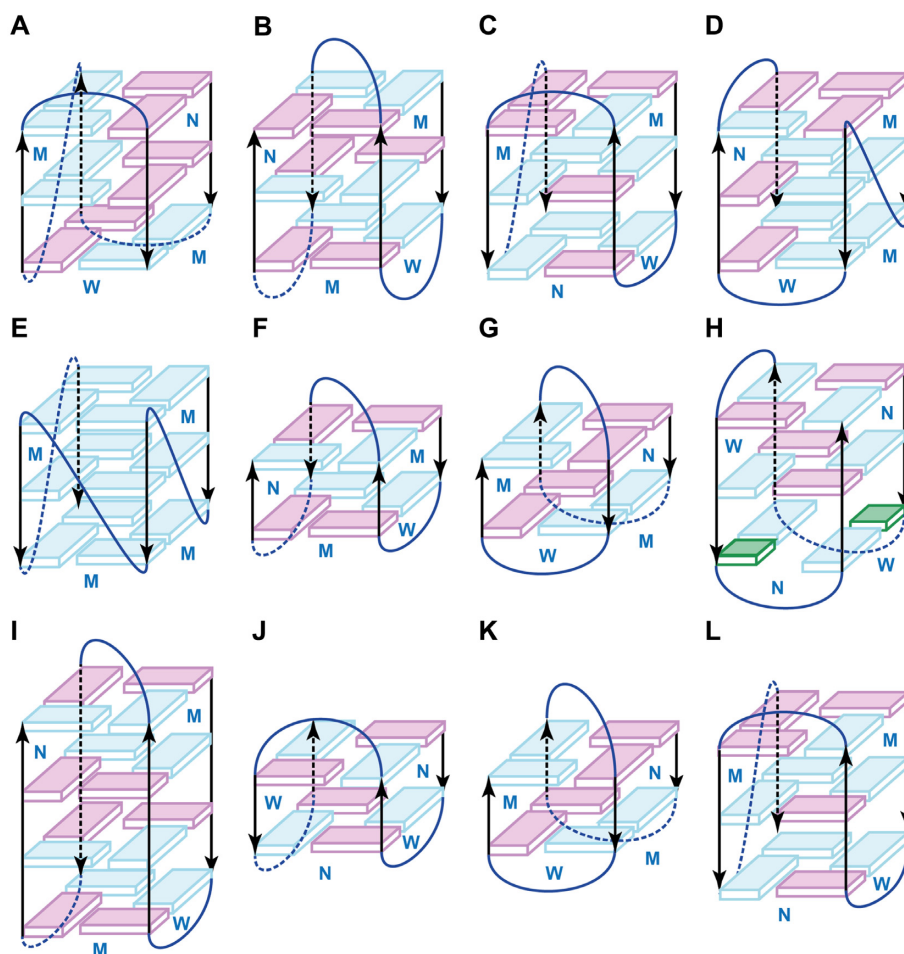
Figure 1. Reported G4 topologies for telomeric sequences from (A–H) human (GGGTTA repeat or GGGCTA repeat for (H)) and non-human (I–L) organisms. *Syn* and *anti* guanines are represented in magenta and cyan, respectively, while cytosines are shown in green. W, N and M refer to wide, narrow and medium grooves, respectively. Each topology is described using the single descriptor nomenclature (SDN) [51] containing the number of guanines in the stem (n = 2, 3 or 4) along with the type of loops (l_w: lateral spanning a wide groove, l_n: lateral spanning a narrow groove, p: propeller and d for diagonal) and relative direction (+ or −) of loops linking G-tracts of the stem: n(L1, L2,L3). (**A**) 2+2 conformation in Na$^+$ (pdb: 2MBJ) [19,20]. **SDN:** 3(+Ln-P-Lw). (**B**) 2+2 basket type conformation in Na$^+$ (pdb: 143D) [19,20]. **SDN:** 3(−LwD+Ln). (**C**) 3+1 conformation in K$^+$ (pdf: 2JPZ) [21–26]. **SDN:** 3(−Lw−Ln−P). (**D**) 3+1 conformation in K$^+$ (pdb: 2GKU) [21–26]. **SDN:** 3(−P−Ln−Lw). (**E**) All parallel conformation (pdb: 2LD8) [21–26] **SDN:** 3(-P-P-P). (**F**) 2+2 basket type conformation in K$^+$ (pdb: 2KF8 & 2KKA [55]) [21–26]. **SDN:** 2(−LwD+Ln). (**G**) 2+2 basket type conformation in K$^+$ (pdb: 5LQG) with only 3 telomeric repeats [27]. **SDN:** 2(+LnD−Lw). (**H**) Chair type Anti-parallel conformation for the human variant in K$^+$ (pdb: 2KM3) [28]. **SDN:** 2(+Ln+Lw+Ln). (**I**) 2+2 basket type conformation for *Oxytricha* ($G_4T_4)_n$ in Na$^+$ (pdb: 201D) [13]. **SDN:** 4(−LwD+Ln). (**J**) Chair type Anti-parallel conformation for *Bombyx mori* ($G_2TTA)_n$ in K$^+$ [15]. **SDN:** 2(-Lw-Ln-Lw). (**K**) 2+2 basket type conformation for *Giardia* ($G_3TA)_n$ in K$^+$ (pdb: 2KOW) [16]. **SDN:** 2(+LnD−Lw). (**L**) 3+1 conformation for *Tetrahymena* ($G_4T_2)_n$ in Na$^+$ (pdb: 186D) [13]. **SDN:** 3(−Lw−Ln−P).

curves were recorded both ways (heating and cooling) between 95°C and 2°C at 0.2°C min^{-1}. TDS were calculated by subtracting the UV absorbance spectrum at 5°C from the one at 95°C.

Circular dichroism (CD) spectroscopy

CD-spectra were measured between 215 and 330 nm (0.5 nm data pitch) using a Jasco J-815 spectropolarimeter in 1-cm path-length quartz cells. The temperature was set at 20°C with a Peltier temperature controller. Each spectrum is the average of four scans recorded at a speed of 50 nm min^{-1} (2-nm bandwidth and 1-s data integration time). The oligonucleotide concentration ranged between 3.2 and 6.5 μM.

NMR experiments

A 700 MHz Bruker spectrometer was used to perform the NMR experiments. Sample strand concentrations ranged from 100 μM to 3 mM. Except otherwise stated, all measurements were carried out at 15°C (288 K). Resonance assignments were based on bromo-guanine labelling and through bond correlations at natural abundance (2D JRHMBC, {^{13}C−^{1}H}-HMBC and {1H−1H}-TOCSY) [33,34]. Several {1H-1H} NOESY experiments have been recorded at 50, 150, 250 and 350 ms mixing times. Diffusion ordered spectroscopy (DOSY) NMR was used to measure translational diffusion coefficients (D) using ^{1}H NMR bipolar pulsed gradients with the following parameters at 298 K: intergradient delay Δ = 150 ms; gradient pulse duration δ = 2 ms. The pulse gradients (G) were in-

cremented from 2 to 95% of the maximum gradient strength in a linear ramp in 16 steps.

Structure calculation

Distance restraints between protons were obtained from ^1H–^1H NOESY spectrum (250 ms) at 288 K using SPARKY software for spectral assignment (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco) and CCPN Analysis for volume calculation. NOESY spectra were acquired at 50, 150, 250 and 350 ms mixing times (Supplementary Figure S-1). Distance restraints were derived from the NOESY spectrum recorded at 250 ms mixing time. Interproton distances were calibrated using the average volume of H7–H6 isolated cross-peaks of T5 and T16 corresponding to a distance of 3.0 Å. All restraints files were then included in CNS1.2 within ARIA2.3 software for structure calculation. The distance bounds have been set with a minimum distance of 1.8 Å and a maximum of 7.0 Å. For each distance a fluctuation of ±25% was allowed in order to prevent errors from peak integration, overlaps or spin diffusion (35). Hoogsteen hydrogen bonds and planarity restraints were set considering the two tetrads G1→G13→G20→G8 and G2→G7→G19→G14; and the C9-T11 base pair. Eight iterations were performed from 100 to 750 calculated structures with mixed Cartesian and torsion angle dynamics during the simulated annealing runs. For distances and hydrogen bonds, 50 kcal mol^{-1} Å^{-2} was applied during the initial stage of dynamics and was increased up to 100 kcal mol^{-1} Å^{-2} for the remaining steps of the dynamics. Twenty structures were then extracted and analysed in order to select the ten lowest energy structures from calculation before going through structure refinement using AMBER20 molecular dynamics (see below). A more detailed protocol has been described in a previous work (36).

MD simulation

Structure refinement in explicit solvent was performed using AMBER20 (37) molecular dynamics. For each structure selected from ARIA calculation, a K^+ cation was inserted within the G-core. The system was then neutralized with potassium cations and solvated with water molecules using a truncated octahedral TIP3P box. The first step was a minimization using harmonic potential position restraints at 50 kcal mol^{-1}.Å^{-2} over 1000 steps of steepest descent minimization. A second step consisted in heating the system from 100 to 288 K for 25 ps with position restraints at 50 kcal mol^{-1} Å^{-2}. Then a series of equilibrations was performed by gradually reducing positional restraints: 25, 20, 15, 10, 5 and finally 2.5 kcal mol^{-1} Å^{-2}. The last step was the equilibration of the system without positional restraints for 2.5 ns at 288 K. The ten lowest-energy structures were extracted and went through a series of minimizations to eliminate clashes and to correct bonds lengths and torsions. Distance and planarity restraints were imposed during molecular dynamics refinement.

Table 1. List of the DNA sequences studied here

Name	Sequence (5′ → 3′)	T_m (°C)
Ce20	GGC TTA GGC TTA GGC TTA GG	40.0
	Br-substituted sequences	
BrG-1	**Br****G**GC TTA GGC TTA GGC TTA GG	46.7
BrG-7	GGC TTA **Br****G**GC TTA GGC TTA GG	49.5
BrG-13	GGC TTA GGC TTA **Br****G**GC TTA GG	45.4
BrG-19	GGC TTA GGC TTA GGC TTA **Br****G**G	50.2
	G4 forming mutants	
C3T	GG**T** TTA GGC TTA GGC TTA GG	44.0
T5C	GGC T**C**A GGC TTA GGC TTA GG	42.0
C9T	GGC TTA GG**T** TTA GGC TTA GG	38.0
T10A	GGC TTA GGC **A**TA GGC TTA GG	40.0
T11C	GGC TTA GGC T**C**A GGC TTA GG	<30.0
A12T	GGC TTA GGC TT**T** GGC TTA GG	42.0
C15T	GGC TTA GGC TTA GG**T** TTA GG	43.0
T17C	GGC TTA GGC TTA GGC T**C**A GG	37.0
	Defective G4 mutants	
T4C	GGC **C**TA GGC TTA GGC TTA GG	-
T10C	GGC TTA GGC **C**TA GGC TTA GG	-
T11A	GGC TTA GGC T**A**A GGC TTA GG	-
T16C	GGC TTA GGC TTA GGC **C**TA GG	-

Melting temperatures of the G-quadruplexes determined by UV-melting at 295 nm in 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. The melting profile of Ce20 is provided in Figure 2B and Supplementary Figure S2A; Bromo derivatives are shown in Supplementary Figure S5 and other mutants in Supplementary Figures S8D, S9D; S10D and S11D. (–) This sequence does not fold into a G-quadruplex under these conditions.

BMRB and PDB deposition

The NMR chemical shifts have been deposited in the Biological Magnetic Resonance Bank (accession code, 34631) and coordinates have been deposited in the Protein Data Bank (accession code, 7OQT)

RESULTS

An intramolecular two-tetrad G-quadruplex

In K^+ solution, the thermal difference absorbance spectrum (TDS) of Ce20 (Table 1) presents a negative minimum at 297 nm and two positive maxima at 246 and 272 nm (Figure 2A); this pattern is characteristic of a G-quadruplex structure (38). A thermal denaturation experiment in which absorbance is recorded at 295 nm allowed us to determine a melting temperature of 40°C, well above the typical temperature at which *C. elegans* typically lives (Figure 2B). The corresponding CD-spectrum (Figure 2C) displays a negative peak at 262 nm and two positive peaks at 244 and 290 nm. This signature is typical for anti-parallel G-quadruplexes. Both results are consistent with previous reports including this sequence (2,39). The proton NMR spectrum of the Ce20 sequence is composed of eight major well-resolved peaks in the 10-11 ppm imino protons region (Figure 2D). Each G-tetrad involves four imino protons; the presence of those eight peaks indicates the formation of a single G-quadruplex structure composed of two G-tetrads. Notably, the presence of an unusually sharp peak at 10.1 ppm suggests that at least one additional hydrogen-
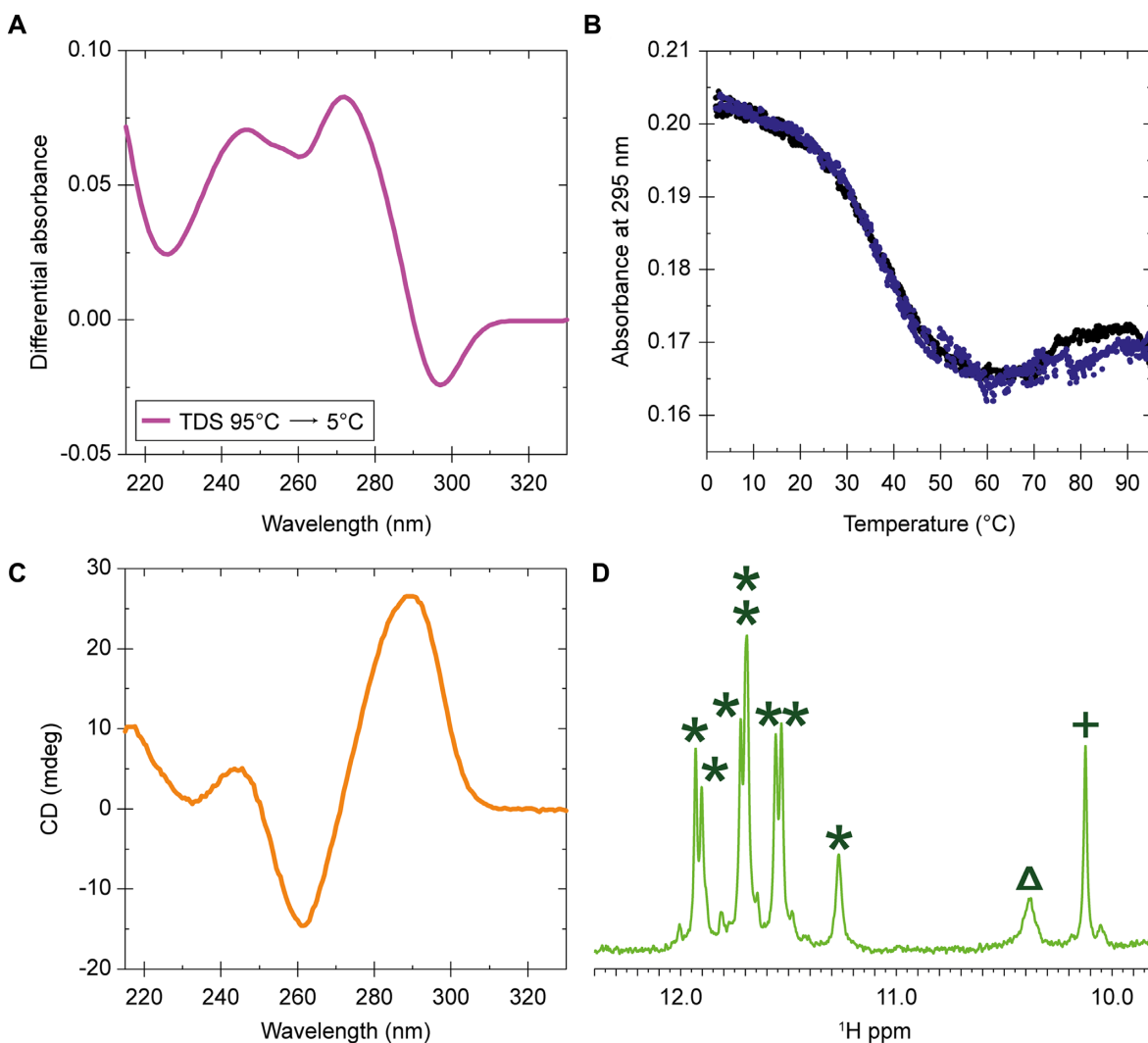
Figure 2. (A–C) Optical spectroscopy on a 5 μM Ce20 sample in 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. (**A**) Thermal difference spectrum (TDS) (**B**) UV-melting (both heating and cooling profiles are shown; $T_{\mathrm{m}} = 40°$C) (**C**) CD-spectrum. (**D**) ^{1}H NMR spectrum at 288 K in 90% H_2O/10% D_2O 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. (*) Guanine imino protons from tetrads, (+) Thymin imino proton, (Δ) guanine amino proton.

bond forms between loop bases and contributes to the overall folding.

UV-melting and NMR DOSY experiments (40) were used to determine the molecularity of the folded structure. The UV-melting temperature is independent of the sample concentration between 1 and 100 μM (Supplementary Figure S-2A). In addition, using NMR DOSY experiments, we estimated the translational diffusion coefficient (D) of Ce20 at 1.78×10^{-10} m^{2}s^{-1} (25°C) for a sample at 3 mM oligonucleotide concentration. Using the Stokes–Einstein equation and assuming a spherical shape of the molecule, this diffusion coefficient (D) corresponds to a spherical diameter of 27×10^{-10} m, in agreement with other G-quadruplex monomeric structures of the same length (Supplementary Figure S-2B). Along with the already reported fast electrophoretic mobility on native PAGE (2), these results imply that the folded structure is monomeric.

NMR spectral assignments

In the aromatic/anomeric region, the NOESY experiment at short (50 ms) mixing time for Ce20 (Figure 3A) shows several short distance H–H correlation peaks. Three peaks correspond to the aromatic H5–H6 correlations from the cytosines of the loops, as confirmed by TOCSY experiments (Supplementary Figure S-3A). The four most intense peaks are intra-residue H8–H1' correlations between the aromatic and the anomeric protons of the guanine residues adopting *syn* glycosidic torsion angles. Their presence confirms the anti-parallel nature of the core of this G-quadruplex.

We introduced ^{8Br}G substitutions in which a Bromo atom replaces the H8 hydrogen atom. Because of the presence of the bulky bromo atom on position 8, ^{8Br}G substitutions are known to stabilize G-quadruplexes when replacing a guanine in *syn* glycosidic conformation, while they destabilize the structure in the case of an *anti* conformation (21,41,42). Several studies have already taken advantage of Br-guanine
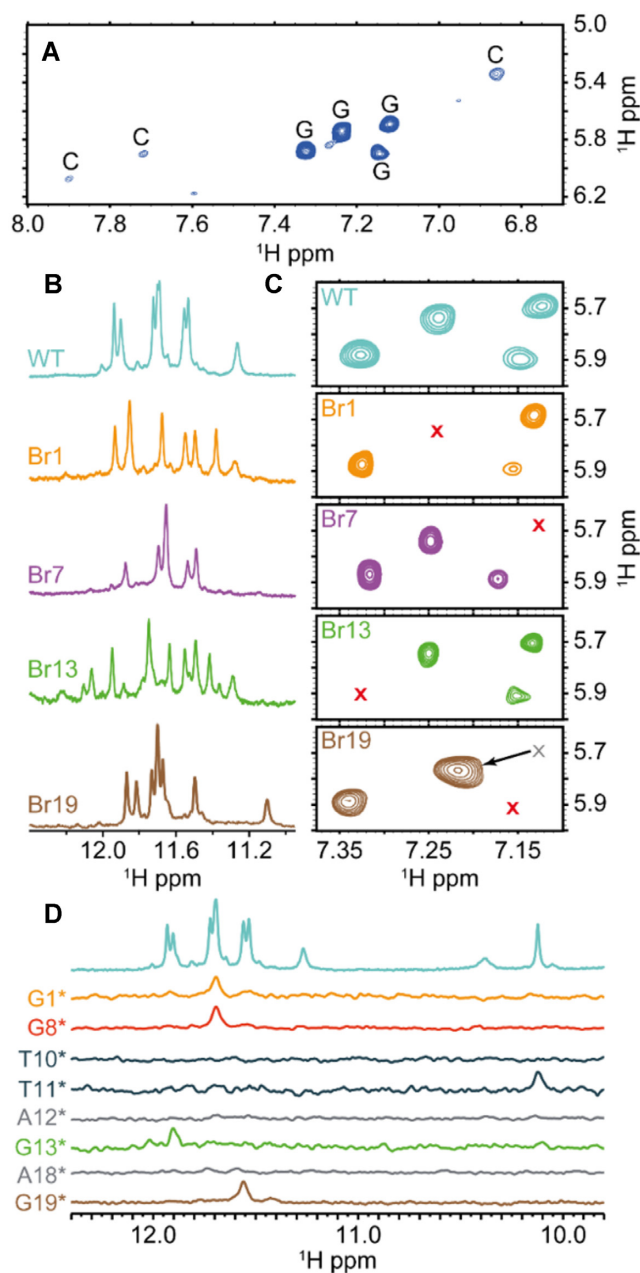
Figure 3. Unambiguous assignment of Imino/aromatic protons of guanines based on ^{15}N-labelling and Br-guanine substitutions. Experiments were recorded in 90% H_2O /10% D_2O 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. (**A**) {^{1}H–^{1}H} NOESY spectrum of Ce20 with a 50 ms mixing time. (**B, C**) ^{1}H 1D NMR spectra of Ce20, Br1, Br7, Br13 and Br19 at 288 K with in (**B**) 1D spectra and in (**C**) NOESY experiments with a mixing time of 50 ms). (**D**) ^{15}N-filtrated spectra using samples containing 5% of ^{15}N-enriched isotope in order determine or confirm residues implicated in Ce20 G4 formation.

modifications to overcome the difficulty arising from conformational heterogeneity either by displacement of the conformational equilibrium towards one major form (42) or by sorting out the contributions of minor conformation(s) (21). The Br-guanine substitution has also been used in combination with inosine substitutions in the early 2000s to facilitate the assignments in NOESY spectra (43–45). The

Table 2. Statistics of the computed structures of Ce20 after structure calculation and refinement by molecular dynamics

NMR restraints	H$_2$O
distance restraints	510
intraresidue distance restraints	325
sequential (I, i+1) distance restraints	125
long-range (I, $\geq i$+2)	38
short-range non sequential distance restraints	22
dihedral restraints	8
H bonds restraints	18
Structural statistics	
structure calculation	
total calculated structures	750
NOE violations	
number (>0.3 Å)	0.06
RMSD of violations (Å)	0.193 ± 0.009
molecular dynamics	
simulation time (ns)	2.5
extracted structures (lowest energy)	10
RMSD	
all heavy atoms (Å)	0.73

anti-parallel core of the G-quadruplex could present either *5'syn-anti3'* or *5'anti-syn3'* steps between two stacked guanines. According to the G-quadruplex structures already published, the *5'syn-anti3'* steps are more predominant than the *5'anti-syn3'* steps which have been shown to provide the least energetically favourable stacking mode (46). Hypothesizing the presence of the more favourable *5'syn-anti3'* stacking step, we decided to modify the Gs at positions 1, 7, 13 and 19 (Table 1). Prior to the NMR analysis, the TDS, CD and thermal stability of the Br-substituted sequences were compared to those of the unmodified sequence. For all four Br-substituted sequences, the TDS (Supplementary Figure S-4A) were typical of G-quadruplexes (38) and similar to Ce20. The UV-melting curves (Table 2 and Supplementary Figure S-5) show that the stability of the modified sequences is higher than that of the unmodified one. Along with the CD-spectra of the ^{Br}G sequences showing similar conformational signatures (Supplementary Figure S-4A), these data suggest that the anti-parallel structure is maintained in all four cases. We also ensured that the initial topology was maintained in the presence of the Br-modified residues by carefully comparing the ^{1}H–^{1}H NOESY spectra of both modified and unmodified oligonucleotides (Supplementary Figure S-4B). Unsurprisingly, despite the topology being preserved upon ^{Br}G substitutions, the 1D imino proton NMR signature significantly changed (Figure 3B). As shown by *Lech et al.*, this variation in imino proton chemical shift is mainly due to slight changes in hydrogen-bond geometry within the modified tetrads (47). In this case, the most affected imino proton was the one that is directly hydrogen-bonded to the Br-guanine. The other imino protons are also affected which prevent the straightforward identification of the signals corresponding to the same protons in the unmodified and modified structures. However, when comparing the intra-residue H8-H1' NOE cross-peaks recorded at a 50 ms mixing time for each G1, G7, G13 Br-labelled sequence with those of the unmodified one, only minor changes were observed for the three remaining H8-H1' correlations from the *syn* guanines. De-
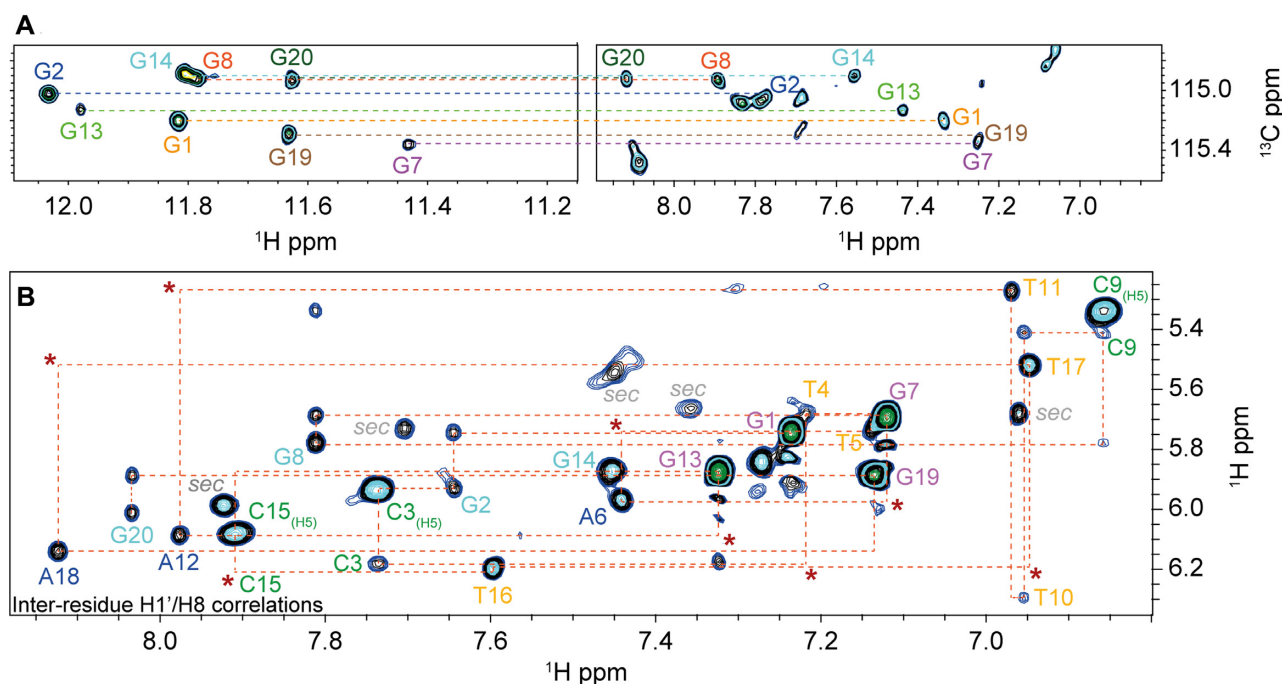
Figure 4. Assignment of aromatic and imino protons of guanines in 90% H_2O /10% D_2O 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. (**A**) {$^{13}C-^1H$}-HMBC at 302 K showing intra-residue H1/H8 correlations used to identify H8 of guanines implicated in Ce20 tetrads. (**B**) Ce20 inter-residue H1'/H8 correlations ("walk") between H8 from a residue and H1' from the previous residue {$^1H-^1H$} NOESY experiment with a 250 ms mixing time. Red stars represent weak correlations.

spite the slightly more important shifts in the case of ^{Br}G19, these data again confirm that the global topology adopted by Ce20 sequence is not altered by the Bromo-substitutions. Finally, we took advantage of these substitutions to unambiguously assign the chemical shifts of the H8 and H1' of all the guanines in *syn* conformation (Figure 3C). In each 50 ms mixing time NOESY, the missing H8-H1' NOE cross-peaks was attributed to the replaced proton for G1, G7, G13 and G19.

Imino (H1) protons were assigned in two steps. Imino protons of guanines G1, G8, G13 and G19 were unambiguously assigned using site-specific low-enrichment ^{15}N labelling (Figure 3-D). The four other imino protons of guanines were assigned from through-bond correlations {$^{13}C-^1H$}-HMBC at 288 and 302 K (Figure 4-A and Supplementary Figure S-3B). Some of the HMBC correlations did not appear at 288 K likely due to unfavourable T2 relaxation times for some of the imino peaks, but we managed to assign all the imino peaks at 302 K. The assignments of the H8 of G13 and G19 deduced by combining ^{15}N labelling for H1 and HMBC were consistent with those previously made from Bromo-substitution ruling out any doubt about the assignments of the H8 and H1' protons for G13 (due to the difference in 1D pattern in Figure 3-B) or for G19 (with the difference in 2D pattern in Figure 3C). Site-specific low-enrichment ^{15}N labelling of residues T10, T11, A12 and A18 was used to assign the sharp peak at 10.1 ppm (Figure 3D) to the H3 of T11. On the NOESY experiment at 250 ms mixing time (Figure 4B); 5'-*syn-anti*-3' steps between G1-G2, G7-G8, G13-G14 and G19-G20 are evidenced by specific rectangular NOE patterns (48) composed of double sequential H1'/H8 correlations. The assignments of the H8 of G2,

G8, G14 and G20 were made according to those patterns and to already assigned protons.

Structure of the Ce20 G-quadruplex core

The characteristic H1–H8 proton cyclic NOE connectivity patterns (Figure 5B) define two G-tetrads: G1–G13–G20–G8 and G2–G7–G19–G14 (Figure 5C). The stacking of these two quartets corresponds to a G-quadruplex with C3–T4–T5–A6 and C15–T16–T17–A18 as lateral loops spanning a wide and a narrow groove, respectively while the central C9–T10–T11–A12 is a diagonal loop (Figure 5D). Inter-tetrad H1–H1 correlations (Figure 5A) between two guanines spanning the same groove in opposite directions (G13–G19, G2–G8 and G7–G20) confirm the anti-parallel nature of the structure (the G14–G1 correlation is hidden in the diagonal). Furthermore, the H8–H8 correlations (Supplementary Figure S-6) between the G1–G2, G7–G8 and G19–G20 confirm the 5'*syn-anti* 3' steps between two stacked guanines (the G13–G14 correlation is hidden in the diagonal) (48). Ce20 G-quadruplex structures were calculated with a total of 510 NOE-derived restraints (Table 2) extracted from {$^1H-^1H$} NOESY experiments with a mixing time of 250 ms in H_2O and D_2O. The 10 lowest-energy structures obtained after structure calculation and refinement using Molecular Dynamics in explicit solvent in the presence of counter-ions (K^+) are shown in Figure 6A and B. The final heavy-atom RMSD of the converged structures is 0.73 Å. Variation of the RMSD is mainly due to C and T residues within the two lateral loops. T11 H3 NOE correlations (Supplementary Figure S-7) with C9 H6 and C9 sugar protons (H1', H2', H2", H4' and H5') evidence the forma-
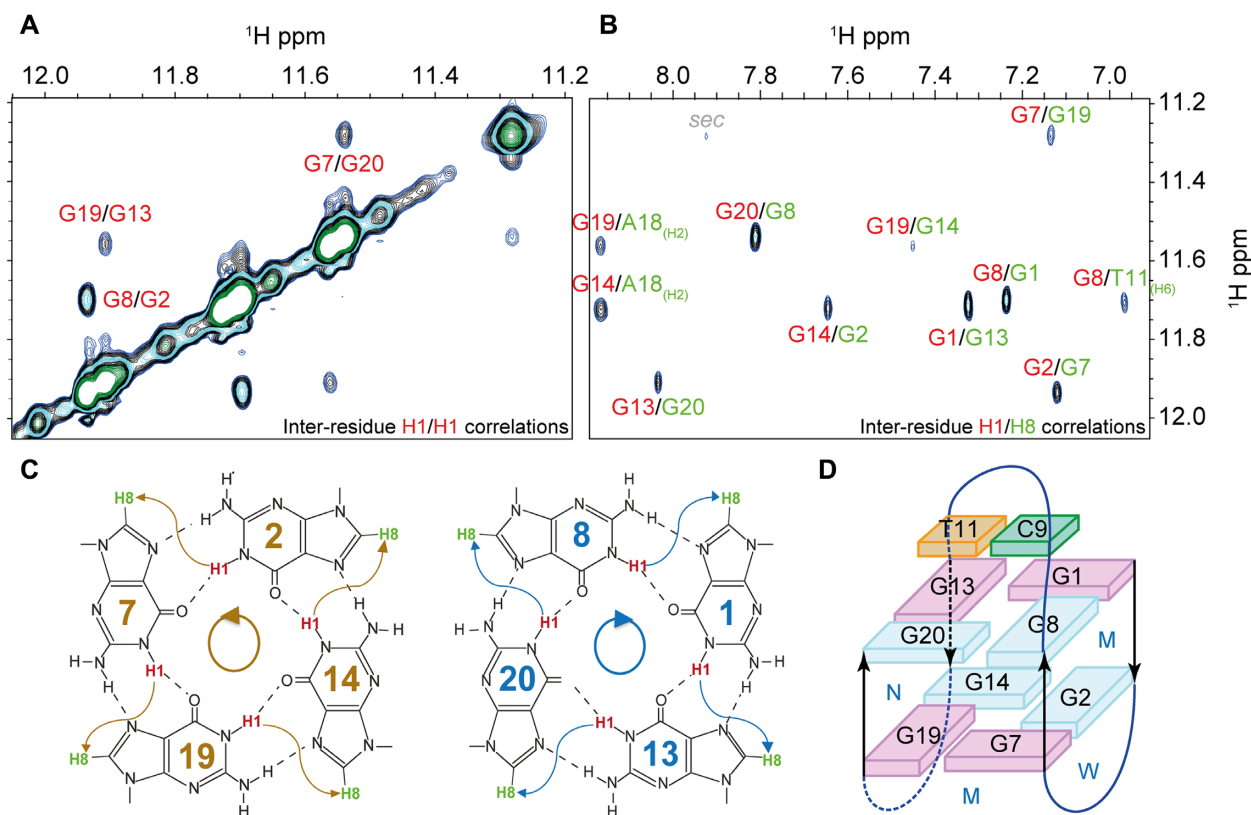
Figure 5. Determination of the G-quadruplex topology in 90% H$_2$O /10% D$_2$O 20 mM potassium phosphate buffer pH 6.9 supplemented with 70 mM KCl. (**A**) and (**B**) Inter-residue H1/H1 and H1/H8 correlations used to determine tetrads patterns in {^{1}H–^{1}H} NOESY experiment with a 250 ms mixing time. Tetrads patterns and Ce20 topology are respectively shown in (**C**) and (**D**). *Syn* and *anti* guanines are represented in magenta and cyan, respectively, while cytosines are shown in green. W, N and M refer to wide, narrow and medium grooves, respectively.

tion of a base pair between T11 and C9. This non canonical C9-T11 base pair involves hydrogen bonds between the O2 and H3 of T11 and N3 and H4 of C9 (Figure 6C). Notably, H3 and O4 of Ts, and not O2, are usually involved in the canonical Watson-crick AT base pairs. The C9-T11 base pair is stacking on top of G1 and G8, involving π–π interactions between the aromatics rings of the bases (Figure 6D). A partial overlap of the 6-member ring of T11 with the 6-member ring of G1 is observed, while an almost complete overlap is observed between the 6-member ring of C9 and the 6-member ring of G8. On the other side, the second tetrad is stabilized by π–π interactions with A18 (Figure 6C).

Mutations within the loops

We investigated a set of 12 mutated sequences (see Table 1). Mutations of any of the three CTTA loops to a CCTA loop (T4C, T10C and T16C) led to TDS that are positive on the all-spectral range with a minimum around 225 nm, a maximum around 270 nm and no shoulder at higher wavelengths (Supplementary Figure S-8B). This shape is intermediate between the signatures reported for 100% AT and 100% GC self-complementary duplexes (38). The CD-spectra present a negative peak around 245 nm and a broad positive band centred around 275 nm (Supplementary Figure S-8A). This signature could correspond to the formation of a hairpin

duplex. This hypothesis is further confirmed by the presence of proton signals between 12.5 and 14 ppm in the ^{1}H-NMR spectra of those mutants (Supplementary Figure S-8C), which reflects the formation of canonical base pairs. For T4C and T10C, the presence of peaks in the 10–12 ppm region argues for the presence of additional G:G mismatch base pairs. Notably these T to C mutations that introduce two consecutive cytosines in the sequence of the loop (CCTA) are detrimental for G-quadruplex formation and favour the duplex structures. Conversely CTTA to CTCA mutations in the lateral loops (T5C and T17C) did not affect the overall TDS and CD-signatures (Supplementary Figures S-9B and S9A) and therefore their antiparallel G-quadruplex formation. Nevertheless, while the ^{1}H-NMR spectrum of T5C is very similar to the Ce20 one, the T17C spectrum evidences the presence of a second G-quadruplex population (Supplementary Figure S-9C). In both cases the sharp T11 signal is maintained. In the case of T5C, the T_m value slightly increased to 42°C while it decreased to 37°C in the case of T17C. In the middle loop, the CTTA to CTCA mutation (T11C) induced an equilibrium between a hairpin duplex and two G-quadruplex populations (see TDS, CD and ^{1}H NMR spectra of Supplementary Figure S-9) and the sharp signal of T11 disappeared. In the absence of the C9/T11 base pair, the T_m value dropped below 30°C. The role of the bases in the middle diagonal loop seems to differ from their role in the lateral loops. This is confirmed by the
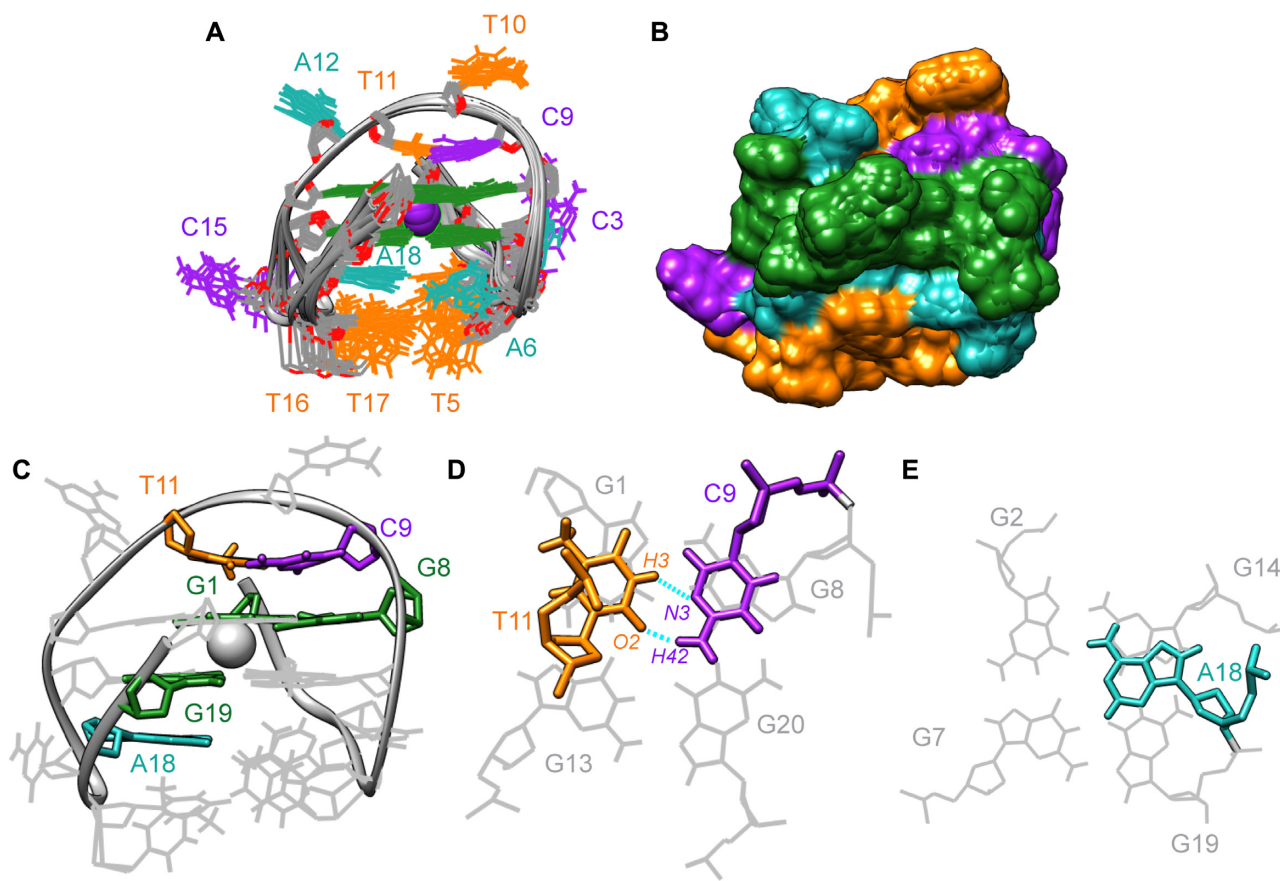
Figure 6. Structure of the Ce20 G-quadruplex. (**A**) Ensemble of the ten best structures obtained after structure calculation and refinement. Guanines are coloured in green, Adenines in cyan, cytosines in purple and thymines in orange (**B**) Surface view of the ensemble. Specific features of Ce20 G-quadruplex structure are shown in (**C**) with the stacking of A18 (cyan) on top of G18 and the C9–T11 base pair (purple and orange) at the bottom of the bottom tetrad. (**D**) Top view of the C9–T11 base pair that stacks on top of G1 and G8. (**E**) Top view of the A18 that stacks on top of the G2–G7–G19–G14 tetrad.

CTTA to TTTA mutations (C3T, C9T and C15T) which all three preserved the G-quadruplex TDS-signature, the anti-parallel CD-signature (Supplementary Figures S-10B and S-10A) and imino protons between 11 and 12.5 ppm (Supplementary Figure S-10C). The C15T ^{1}H-NMR spectrum evidences the presence of two G-quadruplex populations. For CTTA to TTTA mutations in the lateral loops, the melting temperatures slightly increased to 43°C (C3T) and 44°C (C15T). The mutation in the diagonal loop (C9T) was the only one inducing the disappearance of the T11 imino peak at 10.1 ppm, the overall ^{1}H NMR signature of Ce20 being otherwise preserved. This observation is consistent with the presence of the C9-T11 base pair in the central loop. This mutation also caused a drop of 2°C in the T_m as compared to the wild type sequence highlighting the contribution of the C9-T11 base pair to the stability of the structure. The CTTA to CTAA mutation (T11A) did not preserve the G-quadruplex folding (for TDS, CD and ^{1}H NMR spectra see Supplementary Figure S-11). Conversely, T10A and A12T both fold into the anti-parallel G-quadruplex (see TDS and CD of Supplementary Figure S-11). But in both ^{1}H-NMR spectra, at least one sharp amino peak was observed together with the eight imino protons, the C9-T11 base pair being preserved. In the case of T10A, the melting temper-

ature was maintained at 40°C while it slightly increased to 42°C in the case of A12T.

DISCUSSION

Structures of the two-quartets G-quadruplexes

We have shown that the nematode d[GG(CTTAGG)$_3$] telomeric sequence adopts a single conformation. It is a basket-type G-quadruplex that involves a two quartets core, two lateral loops and a central diagonal loop. This latter loop includes a non-canonical C–T base pair. This type of folding differs from those previously reported for two-quartets monomorphic G-quadruplexes. Most of them fold into chair-type G-quadruplexes. The *Bombyx mori* telomeric d[TAGG(TTAGG)$_3$] sequence adopts a chair type anti-parallel topology where the central loop spans across a narrow groove and the two other span wide grooves (15) (Figure 7A). The thrombin binding aptamer (TBA) d[GGTTGGTGTGGTTGG] topology differs from the *Bombyx mori* one by having the central loop spanning a wide groove and the others a narrow one (49) (Figure 7B). Other reported topologies like the HIV-PRO1 (a sequence located in the promoter of the HIV virus) (50) and the variant human telomeric d[AGGG(CTAGGG)$_3$] are built on
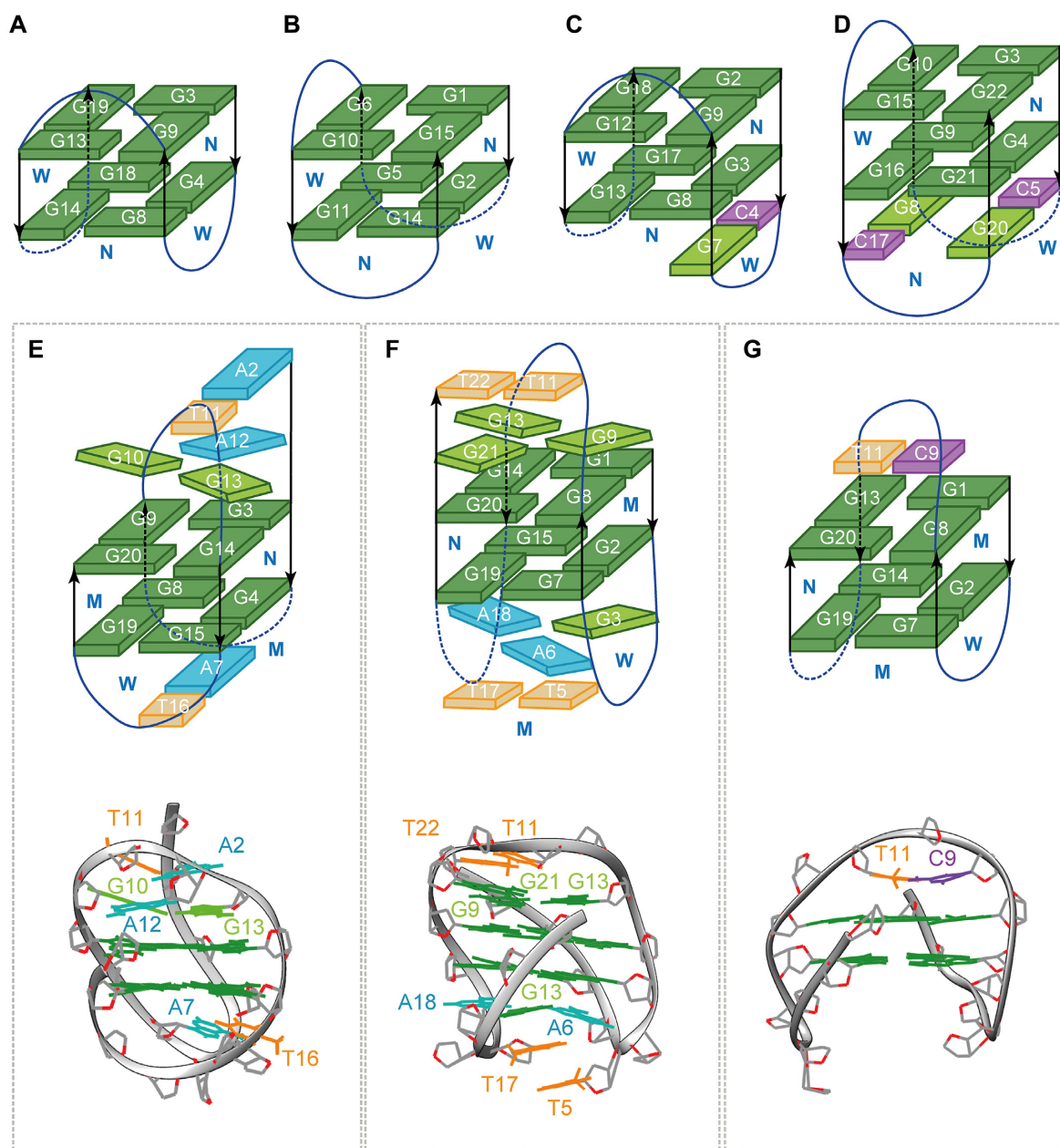
Figure 7. Topologies of 2-quartet intramolecular G-quadruplexes. Guanine tetrads are represented in dark green, while guanines involved in base-pairs or triads are shown in light green. Cytosines, Thymines and Adenines are represented in violet, yellow and blue. W, N and M refer to wide, narrow and medium grooves, respectively. Each topology is described using the single descriptor nomenclature (51) containing the number of guanines in the stem (n = 2) along with the type of loops (l_w: Lateral spanning a wide groove, l_n: Lateral spanning a narrow groove and d for diagonal) and relative direction (+ or −) of loops linking G-tracts of the stem: n(L1, L2,L3). (**A–D**) Chair type anti-parallel conformation for: (**A**) *Bombyx mori (G_2TTA)$_n$* in K$^+$ (15). **SDN:** 2(-Lw-Ln-Lw). (**B**) The thrombin binding aptamer d[GGTTGGTGTGGTTGG] (PDB: 148D). **SDN:** 2(+Ln+Lw+Ln). (**C**) The HIV-PRO1 sequence d[TGGCCTGGGCGGGACTGGG] (56). **SDN:** 2(-Lw-Ln-Lw). (**D**) The telomeric human variant in K$^+$ containing a G:C:G:C tetrad (pdb: 2KM3) (28). **SDN:** 2(+Ln+Lw+Ln). (**E, F**) Basket type anti-parallel conformation for: (**E**) The telomeric motif of *Giardia (G_3TA)$_n$* in K$^+$ (pdb: 2KOW) (16). **SDN:** 2(+LnD−Lw). (**F**) The telomeric motif of *Human* in K$^+$ (pdb = 2KF8 & 2KKA(55)) (21–26). **SDN:** 2(−LwD+Ln). (**G**) The telomeric motif of *C. elegans* (this work; PDB: 7OQT). **SDN:** 2(−LwD+Ln).

the same cores. The HIV-PRO1 adds a C–G base pair in the first lateral loop of the *Bombyx mori* core (Figure 7C) and the variant human telomeric includes a C–G–C–G tetrad on top of the TBA core (Figure 7D).

Although never reported as single conformation of natural sequences, basket-type two quartets G-quadruplexes have been previously described. Like *Bombyx mori* and TBA form the pair of the possible cores for two quartets chair-type G-quadruplexes, some of the reported forms of the *Giardia* and the human telomeric sequences describe all possible two quartets basket-type cores. Indeed, in the structure adopted by the d[(TAGGG)₃TAGGG] *Giardia* telomeric sequence, the first loop (GTA) spans the narrow groove while the third loop (TAG) spans the wide groove (16) (Figure 7E). In the Form 3 of the human telomeric sequence (reported in K$^+$), the first loop spans the wide groove and the third one the narrow groove (26) (Figure 7F). The structure of the *C. elegans* d[GG(CTTAGG)₃] telomeric sequence reported here is built on the same core as the human telomeric Form 3. Differences between the two structures lie in the presence of the C-T base pair in the diagonal loop, the size and composition of the three loops (GTTA, GTTAG and TTA for the human telomere and CTTA in all three loops in the *C. elegans* sequence). The same topology has also been reported for a sequence containing only three human telomeric repeats (27) (pdb: 5LQG) and three other synthetic constructs (51) (pdb: 2M6V, 5J4P, 5J4W). Interestingly, the global conformation of the central diagonal loop of a basket type G-quadruplex seems to favour the formation of non-canonical base-pairs. Indeed, a GGA triad is present within the central diagonal loop of Giardia G-quadruplex (Figure 7E), while a GG and a CT base-pair are present in the Form 3 of the human telomeric sequence (Figure 7F) and the *C. elegans* telomeric sequence (Figure 7G).

G-quadruplex versus (hairpin-)duplex formation

C. elegans d[GG(CTTAGG)₃] telomeric sequence was reported to be involved in an equilibrium between a G-quadruplex and a non-G-quadruplex structure where the non-G-quadruplex structure is likely a hairpin-duplex with two out of three nucleotides engaged in Watson–Crick base pairs (2). Two factors could drive a displacement of this equilibrium: either the destabilization of the G-quadruplex or the stabilization of the hairpin-duplex. It should be noted that the primary sequence of *C. elegans* telomeric repeats does not look very prone to G-quadruplex formation: its G4Hunter score (31) of 0.65 makes it one of the lowest genuine G4 structures solved so far and its thermal stability is relatively modest (40°C—still far above its physiological temperature range, typically between 15 and 25°C). It should therefore be relatively easy to tip the equilibrium and disfavour G4 formation and/or promote alternative structures. Examples of both situations can be found in the literature. The TDS recorded in 100 mM LiCl (Supplementary Data of (2)) illustrates a displacement of the equilibrium driven by the destabilization of the G-quadruplex (as compared to KCl containing solutions). Another report showing that longer *C. elegans* telomeric sequences (21, 44 and 45 bases) do fold in a hairpin-duplex structure (39) reflects

the stabilization of the hairpin-duplex by the possibility to form more Watson-Crick base pairs.

In our experimental conditions, this equilibrium is fully displaced towards the G-quadruplex structure. However, the influence of point mutations on the TDS, CD and ^{1}H NMR spectra of the sequence suggests that some mutations can partially or totally displace the equilibrium towards the formation of the hairpin-duplex. Independently of their position in the sequence, CTTA to CCTA mutations (Supplementary Figures S-8 to S-10) strongly displace the equilibrium towards the hairpin-duplex. As this means that a mismatch would be replaced by a Watson–Crick base pair, these mutations clearly stabilize the hairpin-duplex. This stabilization is certainly the predominant driving force of the displacement of the equilibrium, although a concomitant destabilization of the G-quadruplex by incorporation of a cytosine cannot be ruled out. The CTTA to CTAA mutation (tested only in the central diagonal loop) also displaced the equilibrium towards the hairpin-duplex by introduction of a supplementary base pair and a concomitant destabilization of the G-quadruplex by disruption of the C–T base pair. Conversely, CTTA to TTTA, CATA or CTTT did not displace the equilibrium and maintained the G-quadruplex structure. This is explained by the absence of either an additional base pair in the corresponding hairpin-duplex (no stabilization of this structure) or a predictable decrease in G-quadruplex stability. Finally, the effect of CTTA to CTCA mutations was position-dependent. In the lateral loops, the G-quadruplex structure was maintained but a careful look at the ^{1}H NMR spectra (Supplementary Figure S-8 to S-10) reveals the presence of a small fraction of hairpin-duplex. In the absence of any stabilization of this latter structure, this effect can be attributed to a slight destabilization of the G-quadruplex. The higher proportion of hairpin-duplex in T11C than in m6C and T17C suggests that this destabilization is more significant when the CTTA to CTCA mutation occurs in the diagonal loop, consistently with the disruption of the C–T base pair in this loop.

Recently, it has been shown that the telomeric G-overhangs in *Saccharomyces cerevisiae* (52) are able to form both G-hairpins (53) and antiparallel G-quadruplexes depending on the length of the sequence. It has also been proposed that the folding kinetics and equilibrium between these secondary structures may play a role in the regulation of telomere maintenance. *C. elegans* is usually grown at temperatures comprised between 15°C and 25°C and the longest lifespan have been observed at 16°C (54). This temperature range is therefore favourable for the formation of the Ce20 G-quadruplex studied here despite its relatively low thermal stability ($T_m = 40$°C). We also believe that this low stability could actually be an advantage. This feature, which implies easy transition between folded and unfolded states, would enable the G4 to easily participate in dynamic biological processes without being an obstacle as can be a highly stable and rigid G-quadruplex structure. Indeed, *in vivo* proper functioning of biological processes always requires a high degree of structural plasticity based on easy conformational transitions. Therefore, having a "fragile" G-quadruplex structure in equilibrium with a single strand or hairpin-duplex could be used as a struc-

tural switch in order to regulate functions related to the telomeres. This balance may be affected by the nature of the cation present in the environment (Ce20 G4 is unfolded in Na$^+$ conditions (Supplementary Figure S12)) or regulated by the presence of G-quadruplex, hairpin or single strand binding proteins. Such G-quadruplex/Hairpin equilibrium observed in *Saccharomyces cerevisiae* could also play a role in the telomere dynamics of *C. elegans*. Furthermore, these telomeric repeats are shared by most, if not all, nematodes, including parasitic species infecting humans such as ascarids (e.g. *Ascaris lumbricoides*), filarias, hookworms, pinworms and whipworms (e.g. *Trichuris trichiura*). We have recently discovered that *Ascaris lumbricoides* has an unusual genomic distribution of quadruplex motifs and that small molecules able to selectively recognize G-quadruplex exhibit potent anti-helminth activity (57). These results emphasize the need to better understand quadruplex formation in nematodes and platyhelminths.

DATA AVAILABILITY

The NMR chemical shifts have been deposited in the Biological Magnetic Resonance Bank (accession code, 34631) and coordinates have been deposited in the Protein Data Bank (accession code 7OQT).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Teixeira,M.T. (2013) *Saccharomyces cerevisiae* as a model to study replicative senescence triggered by telomere shortening. *Front. Oncol.*, **3**, 101–101.
2. Tran,P.L.T., Mergny,J.-L. and Alberti,P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res.*, **39**, 3282–3294.
3. Maizels,N. (2015) G4-associated human diseases. *EMBO Rep.*, **16**, 910–922.
4. Bryan,T.M. (2020) G-Quadruplexes at telomeres: friend or foe? *Mol.*, **25**, 3686.
5. Postberg,J., Tsytlonok,M., Sparvoli,D., Rhodes,D. and Lipps,H.J. (2012) A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene.*, **497**, 147–154.
6. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.-F., Mergny,J.-L. *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.
7. Kar,A., Jones,N., Arat,N.Ö., Fishel,R. and Griffith,J.D. (2018) Long repeating (TTAGGG)n single-stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation. *J. Biol. Chem.*, **293**, 9473–9485.
8. Abraham Punnoose,J., Ma,Y., Hoque,M.E., Cui,Y., Sasaki,S., Guo,A.H., Nagasawa,K. and Mao,H. (2018) Random formation of G-quadruplexes in the full-length human telomere overhangs leads to a kinetic folding pattern with targetable vacant G-tracts. *Biochem.*, **57**, 6946–6955.
9. Saintomé,C., Amrane,S., Mergny,J.-L. and Alberti,P. (2016) The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res.*, **44**, 2926–2935.
10. Phan,A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.
11. Schechtman,M.G. (1990) Characterization of telomere DNA from *Neurospora crassa*. *Gene.*, **88**, 159–165.
12. 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. U.S.A.*, **85**, 6622–6626.
13. Wang,Y. and Patel,D.J. (1994) Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure*, **2**, 1141–1156.
14. Wang,Y. and Patel,D.J. (1995) Solution Structure of theOxytrichaTelomeric Repeat d[G4(T4G4)3] G-tetraplex. *J. Mol. Biol.*, **251**, 76–94.
15. Amrane,S., Ang,R.W.L., Tan,Z.M., Li,C., Lim,J.K.C., Lim,J.M.W., 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.
16. Hu,L., Lim,K.W., Bouaziz,S. and Phan,A.T. (2009) Giardia telomeric sequence d(TAGGG)4 forms two intramolecular G-quadruplexes in K+ solution: effect of loop length and sequence on the folding topology. *J. Am. Chem. Soc.*, **131**, 16824–16831.
17. Karsisiotis,A.I., O'Kane,C. and Webba da Silva,M. (2013) DNA quadruplex folding formalism – a tutorial on quadruplex topologies. *Methods*, **64**, 28–35.
18. Webba da Silva,M., Trajkovski,M., Sannohe,Y., Ma'ani Hessari,N., Sugiyama,H. and Plavec,J. (2009) Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem. Int. Ed. Engl.*, **48**, 9167–9170.
19. 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.
20. Lim,K.W., Ng,V.C.M., Martín-Pintado,N., Heddi,B. and Phan,A.T. (2013) Structure of the human telomere in Na$^+$ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.*, **41**, 10556–10562.
21. 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.
22. 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.
23. 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 formation. *Nucleic Acids Res.*, **35**, 2440–2450.
24. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
25. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
26. 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.

27. Galer,P., Wang,B., Sket,P. and Plavec,J. (2016) Reversible pH switch of two-quartet G-quadruplexes formed by human telomere. *Angew. Chem. Int. Ed. Engl.*, **55**,1993–1997.

28. Lim,K.W., Alberti,P., Guédin,A., Lacroix,L., Riou,J.-F., Royle,N.J., Mergny,J.-L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.

29. Kruisselbrink,E., Guryev,V., Brouwer,K., Pontier,D.B., Cuppen,E. and Tijsterman,M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. *Curr. Biol.*, **18**, 900–905.

30. Koole,W., van Schendel,R., Karambelas,A.E., van Heteren,J.T., Okihara,K.L. and Tijsterman,M. (2014) A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.*, **5**, 3216.

31. Bedrat,A., Lacroix,L. and Mergny,J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.

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

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

34. Phan,A.-T., Guéron,M. and Leroy,J.-L. (2002) In: Thomas,L., James,V.D. and Uli,S. (eds). *Methods Enzymol.* Academic Press, Vol. **338**, pp. 341–371.

35. Webba da Silva,M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.

36. Marquevielle,J., Robert,C., Lagrabette,O., Wahid,M., Bourdoncle,A., Xodo,L.E., Mergny,J.-L. and Salgado,G.F. (2020) Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Res.*, **48**, 9336–9345.

37. Case,D.A., Cheatham,T.E. 3rd, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

38. 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.

39. Školáková,P., Foldynová-Trantírková,S., Bednářová,K., Fiala,R., Vorlíčková,M. and Trantírek,L. (2015) Unique C. elegans telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res.*, **43**, 4733–4745.

40. Ambrus,A. and Yang,D. (2007) Diffusion-ordered nuclear magnetic resonance spectroscopy for analysis of DNA secondary structural elements. *Anal. Biochem.*, **367**, 56–67.

41. Dias,E., Battiste,J.L. and Williamson,J.R. (1994) Chemical probe for glycosidic conformation in telomeric DNAs. *J. Am. Chem. Soc.*, **116**, 4479–4480.

42. 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.

43. Webba da Silva,M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.

44. Kuryavyi,V., Kettani,A., Wang,W., Jones,R. and Patel,D.J. (2000) A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads11Edited by P. E. Wright. *J. Mol. Biol.*, **295**, 455–469.

45. Kuryavyi,V., Majumdar,A., Shallop,A., Chernichenko,N., Skripkin,E., Jones,R. and Patel,D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of unimolecular DNA quadruplex containing a pair of stacked G(syn)·G(syn)·G(anti)·G(anti) tetrads flanked by a G·(T-T) triad and a T·T·T triple. *J. Mol. Biol.*, **310**, 181–194.

46. Lech,C.J., Heddi,B. and Phan,A.T. (2013) Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res.*, **41**, 2034–2046.

47. Lech,C.J., Lim,C., Joefina,K., Lim,W., Jocelyn,M., Amrane,S., Heddi,B. and Phan,Anh T. (2011) Effects of site-specific guanine C8-modifications on an intramolecular DNA G-quadruplex. *Biophys. J.*, **101**, 1987–1998.

48. Adrian,M., Heddi,B. and Phan,A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11–24.

49. Kelly,J.A., Feigon,J. and Yeates,T.O. (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.*, **256**, 417–422.

50. Amrane,S., Kerkour,A., Bedrat,A., Vialet,B., Andreola,M.-L. and Mergny,J.-L. (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **136**, 5249–5252.

51. Dvorkin,S.A., Karsisiotis,A.I. and Webba da Silva,M. (2018) Encoding canonical DNA quadruplex structure. *Sci. advances*, **4**, eaat3007.

52. Jurikova,K., Gajarsky,M., Hajikazemi,M., Nosek,J., Prochazkova,K., Paeschke,K., Trantirek,L. and Tomaska,L. (2020) Role of folding kinetics of secondary structures in telomeric G-overhangs in the regulation of telomere maintenance in Saccharomyces cerevisiae. *J. Biol. Chem.*, **295**, 8958–8971.

53. Gajarský,M., Živković,M.L., Stadlbauer,P., Pagano,B., Fiala,R., Amato,J., Tomáška,L.u., Šponer,J., Plavec,J. and Trantírek,L. (2017) Structure of a stable G-hairpin. *J. Am. Chem. Soc.*, **139**, 3591–3594.

54. Zhang,B., Xiao,R., Ronan,Elizabeth A., He,Y., Hsu,A.-L., Liu,J. and Xu,X.Z.S. (2015) Environmental temperature differentially modulates longevity through a thermosensitive TRP channel. *Cell Rep.*, **11**, 1414–1424.

55. Zhang,Z., Dai,J., Veliath,E., Jones,R.A. and Yang,D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*, **38**, 1009–1021.

56. Amrane,S., Kerkour,A., Bedrat,A., Vialet,B., Andreola,M.L. and Mergny,J.L. (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **136**, 5249–5252.

57. Cantara,A., Luo,Y., Dobrovolná,M., Bohalova,N., Fojta,M., Verga,D., Guittat,L., Cucchiarini,A., Savrimoutou,S., Häberli,C., Guillon,J. et al. (2022) G-quadruplexes in helminth parasites. *Nucleic Acids Res.*, **50**, 2719–2735.