Electronic Supplementary Material (ESI) for Chemical Communications.

Supporting Information

Impact of loop length and duplex extensions on the design of hybrid-type G-quadruplexes

Jagannath Jana, Yoanes Maria Vianney and Klaus Weisz*

Institut für Biochemie, Universität Greifswald, Felix-Hausdorff Str. 4, D-17489 Greifswald, Germany

* Corresponding Author; email: weisz@uni-greifswald.de.

Table of Contents

Materials and methods

Sample preparation. DNA oligonucleotides were purchased from TIB MOLBIOL (Berlin, Germany) and further purified by ethanol precipitation. DNA concentrations were determined spectrophotometrically in triplicate by measuring the UV absorbance at 260 nm at 80°C in H_2O using molar extinction coefficients based on a nearest-neighbor model as provided by the supplier. DNA oligonucleotides were dissolved in a low-salt buffer of 10 mM potassium phosphate, pH 7. Prior to measurements, all samples were annealed by heating to 85°C for 5 minutes followed by slow cooling to room temperature and storage in a refrigerator overnight.

Circular dichroism (CD). CD experiments were performed with a JASCO J-810 spectropolarimeter equipped with a thermoelectrically controlled cell holder at 20°C. Spectra were acquired by accumulating five scans at a rate of 50 nm/min over a range of 220-320 nm, a bandwidth of 1 nm, and a response time of 2 s and subsequently blank-corrected. All CD spectra were measured in 10 mM potassium phosphate buffer, pH 7.0, using 1-cm quartz cuvettes.

NMR experiments. All NMR spectra were acquired on a Bruker Avance NEO 600 MHz NMR spectrometer equipped with an inverse $^{1}H/^{13}C/^{15}N/^{19}F$ quadruple resonance cryoprobehead and z-field gradients. Topspin 4.0.7 and CcpNmr Analysis 2.4.2 were used for spectral processing and analysis.[1] Proton chemical shifts were referenced indirectly to sodium trimethylsilylpropionate (TSP) via the temperature dependent water chemical shift at pH 7.0 and carbon chemical shifts were referenced to sodium trimethylsilylpropanesulfonate (DSS) using an indirect referencing method. Unless otherwise stated, NMR spectra were acquired at 30 °C in a 10 mM potassium phosphate buffer with 90% H_2O/10% D_2O, pH 7.0.

A WATERGATE with w5 element was used for solvent suppression in 1D and 2D nuclear Overhauser enhancement (NOESY) experiments. Two-dimensional NOESY spectra were recorded with mixing times of 300, 150, and 80 ms. Phase-sensitive ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) experiments were recorded with a 3-9-19 water suppression scheme, 4K×500 data points and a typical spectral width of 7.5 kHz in the indirect dimension to accommodate $^{13}C6/C8/C2$ resonances of the nucleobase. Double-quantum-filtered correlation (DQF-COSY) spectra were recorded with a 3-9-19 water suppression scheme in 90% H_2O/10% D_2O using 4K×512 data points. ^{1}H-^{13}C heteronuclear multiple bond coherence (HMBC) spectra were acquired with a jump-and-return water suppression, 2K×144 data points, and processed with 50% non-uniform sampling (NUS) in the indirect dimension.

^{15}N-editing was achieved through 1D ^{1}H-^{15}N HMQC experiments on selectively ^{15}N-labeled oligonucleotides with 10% ^{15}N enrichment. For the guanine H1 detection, a selective 90° pulse

was used for imino proton excitation followed by a standard HMQC pulse sequence with ^{15}N decoupling during acquisition. The experiment was optimized for a N-H heteronuclear coupling constant of 90 Hz.

Structure calculations. Initially, 100 lowest-energy starting structures were selected from 400 structures calculated by using a simulated annealing protocol in XPLOR-NIH 3.0.3.[2] The distance constraints were set according to the intensities of the cross-peaks in the NOESY spectra. For non-exchangeable protons, intensities were categorized as strong (2.9 ± 1.1 Å), medium (4.0 ± 1.2 Å), weak (5.5 ± 1.5 Å), and very weak (6.0 ± 1.5 Å). In case of ambiguous cross-peaks due to signal overlap, distances were set to 5.0 ± 2.0 Å. Intensities from exchangeable protons were assigned strong (2.9.0 ± 1.1 Å), medium (4.0 ± 1.2 Å), weak (5.0 ± 1.2 Å), and very weak (6.0 ± 1.2 Å). Glycosidic torsion angles were restrained to the range 170-310° for *anti*-conformers and to 25-95° for *syn*-conformers. The pseudorotation phase angle P was restricted to 144-180° for south-type sugar conformers based on experimentally determined scalar couplings as assessed from DQF-COSY cross-peak patterns. Distance restraints for hydrogen bonds and planarity restraints were added for G-tetrads. In addition, chirality restraints were employed for all residues.

Twenty converged structures were obtained by simulated annealing of the 100 starting structures. The refinement was carried out in vacuum using AMBER18 with the parmbsc force field and OL15 modifications for DNA.[3] Ten conformations of lowest energy were selected for further refinement in water. Potassium ions were added to neutralise the system and two ions were placed in the inner channel of the G-quadruplex core between the eight O6 atoms of two adjacent G-tetrad layers. The system was hydrated with TIP3P water in a truncated octahedral box of 10 Å.[4] The final simulation was performed at 1 atm and 300 K for 4 ns, only using NMR-derived distance, Hoogsteen hydrogen bond, and planarity restraints. To obtain the final ten lowest-energy structures, the trajectory was averaged over the last 500 ps and minimised in vacuum. Atomic coordinates of the final structures and NMR chemical shifts have been deposited in the Protein Data Bank and in the Biological Magnetic Resonance Bank (accession code 8R4W and 34877 for L322 as well as 8R4E and 34876 for dsL322). For the analysis of the calculated structures, the VMD 1.9.2 software was used. Pymol 1.8.4 was employed for the three-dimensional structure representations.

References

1 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J L. Markley, J. Ionides and E. D. Laue, *Proteins Struct. Funct. Bioinforma.*, 2005, **59**, 687–696.

2 C. D. Schwieters, J. J. Kuszewski and G. M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2006, **48**, 47–62.

3 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *J. Chem. Theory Comput.*, 2015, **11**, 5723–5736.

4 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.,* 1983, **79**, 926-935.

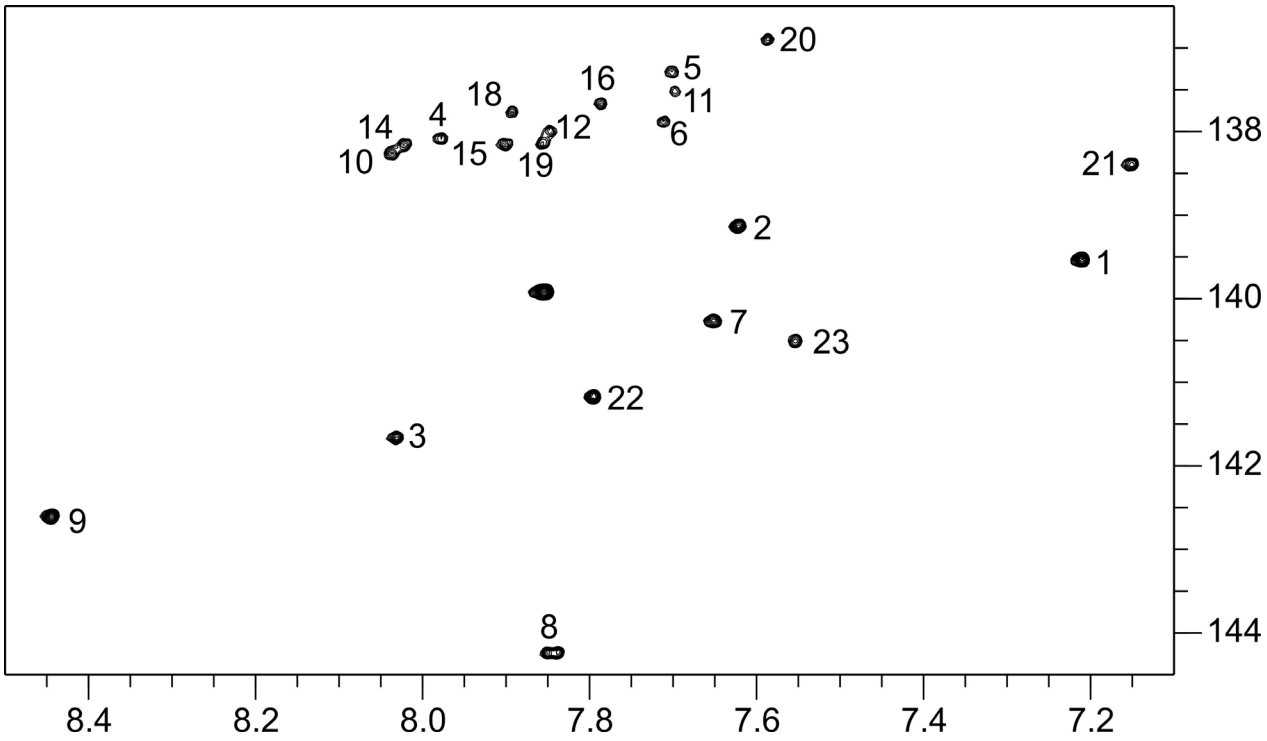

Figure S1. ^{1}H-^{13}C HSQC spectrum of L311 showing H8/H6(ω_2)-C8/C6(ω_1) correlations with cross-peaks assigned to individual residues. Chemical shifts of correlations for all G residues suggest an *anti* glycosidic conformation in agreement with an all-parallel G4 topology.

Spectral assignment of L322. The L322 sequence yields a set of 12 imino proton resonances between10.5 and 12 ppm, which can be attributed to the formation of a three-layered G-quadruplex (Figure S2). Also, an additional imino proton was observed at 13.35 ppm. The presence of five *syn*-G residues in addition to seven *anti*-guanosines can be identified from their strong intra-nucleotide H8-H1' cross-peaks in NOESY spectra and their downfield shifted ^{13}C8 resonances in ^{1}H-^{13}C HSQC spectra (Figure S2 and S3). 2D NOESY spectra acquired with longer mixing times (300 ms) showed a rectangular pattern of intra-nucleotide and sequential H8-H1' cross-peaks for G4-G5, G11-G12, G15-G16, and G20-G21 steps, indicative of *syn-anti*-steps to allow identification of G4, G10, G11, G15, and G20 residues as *syn* conformers. Continuous base-sugar sequential NOE walks were observed from T1 to A3. A continuous NOE walk in the H8-H1' region could be followed along the first G-column G4-G5-G6 with *syn-anti-anti* steps to the first loop T7-C8-A9. The second G-column G10-G11-G12 is found to be comprised of *syn-syn-anti* steps. Here, a sequential NOE walk can be followed from G11 to T14 in the second loop. The fourth G-column G20-G21-G22 with a *syn-anti-anti* glycosidic torsion angle pattern can be identified by uninterrupted NOE connectivities from G20 up to 3'-flanking A25. Finally, the third G-column G15-G16-G17 is again found to comprise *syn-anti-anti* steps. It can be identified by following the continuous NOE contacts from G15 to G17 by a process of elimination. Sequential connectivities are interrupted by a T18-T19 propeller loop.

Guanine imino protons were unambiguously assigned based on inter-residual H1-H1 NOE contacts (Figure S2C) and on their intra-residual H1-H8 correlations, observed for all G-core residues via long-range couplings to ^{13}C5 at natural abundance in ^{1}H-^{13}C HMBC spectra (Figure S4). Additionally, assignments for G4, G11, and G15 imino protons were confirmed by ^{15}N edited experiments employing residue-specific ^{15}N labelled oligonucleotides with 10% ^{15}N enrichment (Figure S5). Tetrad polarities were determined using characteristic intra-tetrad H8-H1 NOE contacts (Figure S2D). Hoogsteen hydrogen bonds from donor to acceptor within the three G-tetrads run along G4-G20-G15-G12 (5'-tetrad), G5-G11-G16-G21 (central tetrad) and G6-G10-G17-G22 (3'-tetrad), featuring heteropolar as well as homopolar tetrad stacking.

Based on full resonance assignments, strong NOE contacts were observed between G4 and G10 imino protons in the outer tetrads with H8 base protons of A3 and A9, respectively (Figure S2). Also, a strong NOE contact with A3 H8 of the slowly exchanging and downfield-shifted T14 imino proton, unambiguously identified by a T14U exchange (Figure S6), is compatible with an A·T reverse Hoogsteen base pair.

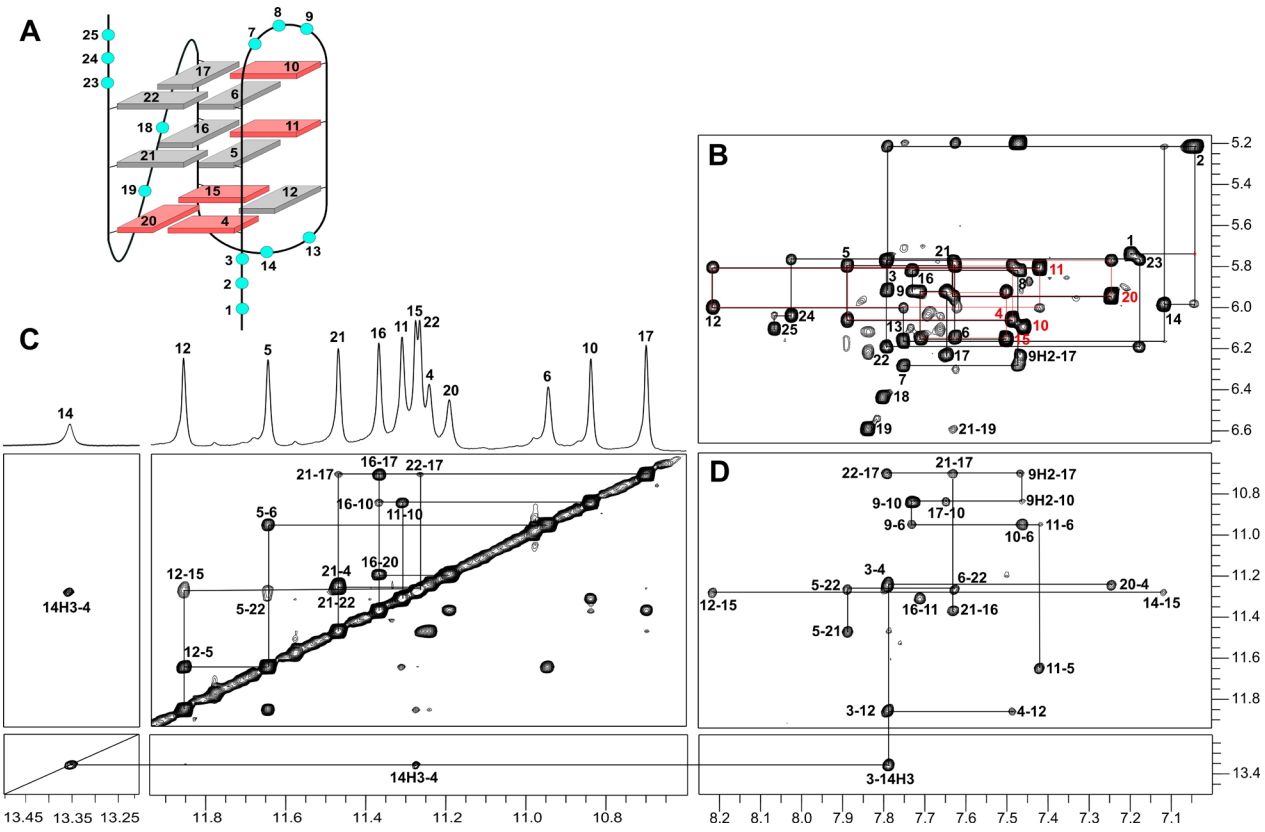

Figure S2. (A) (3+1) Hybrid-2 topology of L322 with residue numbers; *anti*- and *syn*-G residues of the G-core are colored grey and red, respectively; loop and overhang residues are represented by circles colored cyan. (B)-(D) 2D NOESY spectral regions (mixing time 300 ms) of L322. (B) H8/H6(ω_2)-H1'(ω_1), (C) H1-H1, and (D) H8/H6(ω_2)-H1(ω_1) NOE spectral region; in (B), *syn*-guanosines and a typical rectangular connectivity pattern for *syn-anti* steps are shown in red; in (C), the 1D imino proton spectral region with assignments is shown on top. NMR spectra were recorded in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

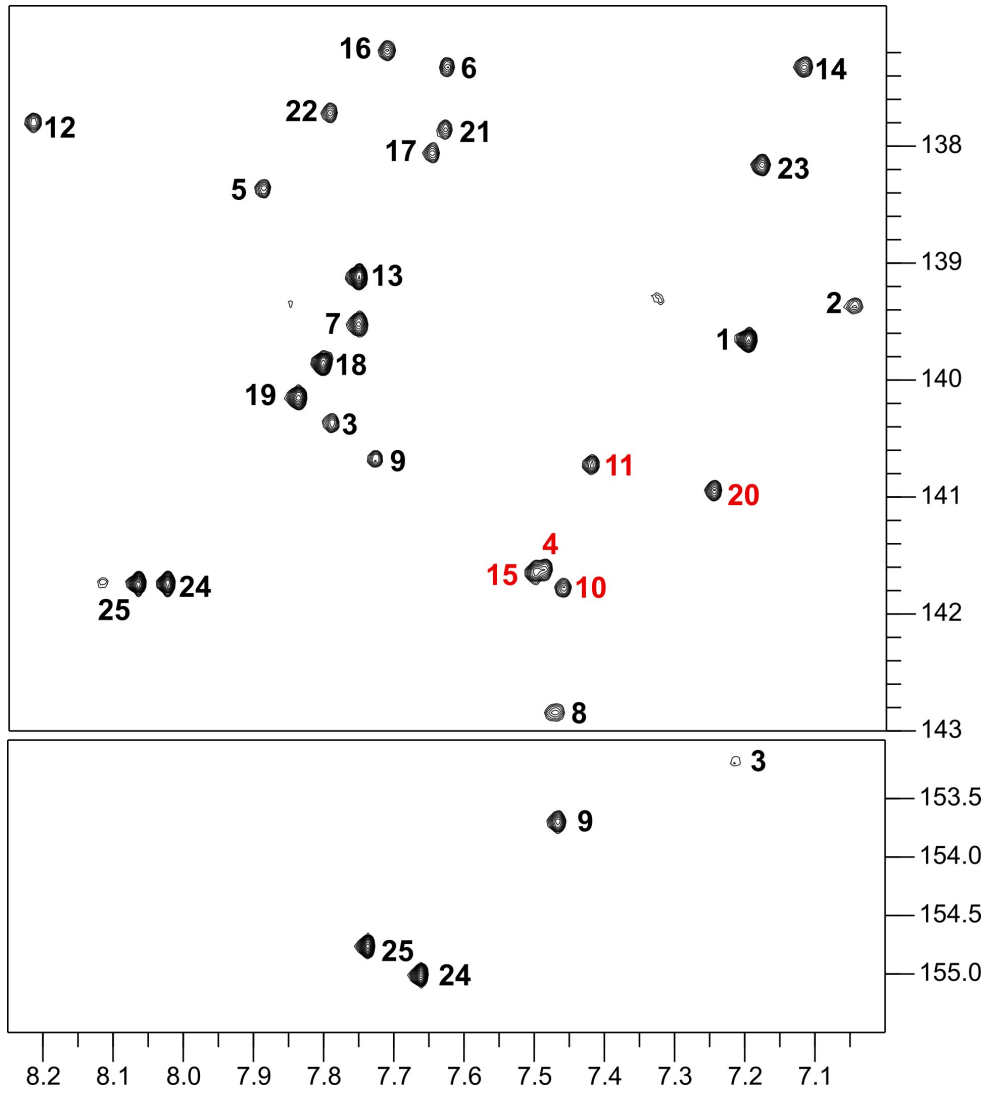

Figure S3. ^{1}H-^{13}C HSQC spectrum of L322 showing H8/H6(ω_2)-C8/C6(ω_1) (top) and adenine H2-C2 correlations (bottom). Upfield-shifted H8 and downfield-shifted ^{13}C8 of *syn*-guanosines are marked in red. Spectra were acquired in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

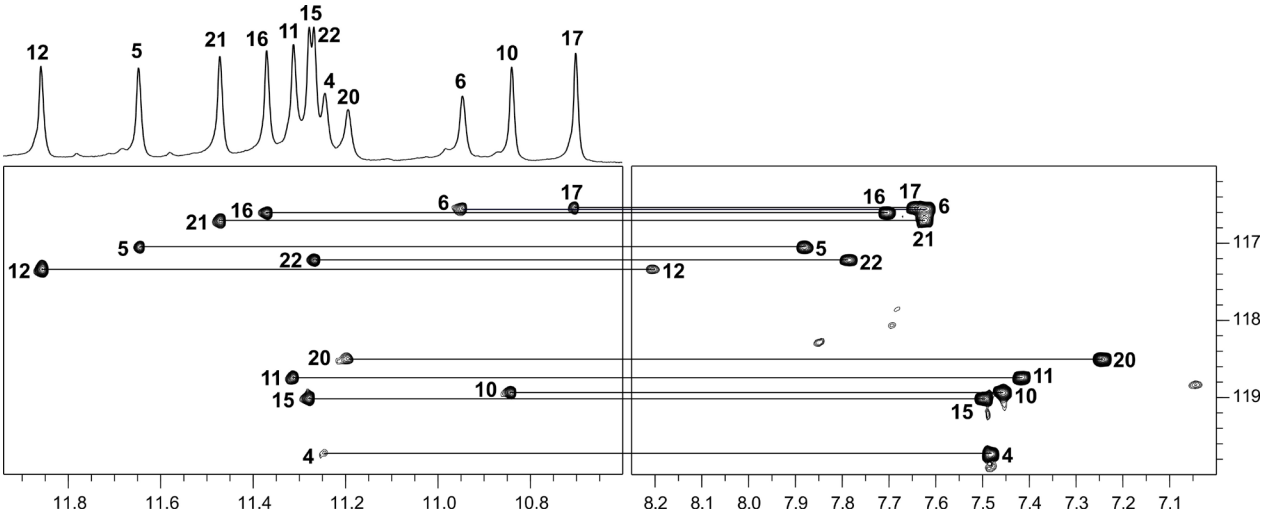

Figure S4. ^{1}H-^{13}C HMBC spectrum of L322 showing correlations between guanine H1 and H8 resonances (ω_2) via long-range scalar couplings to ^{13}C5 (ω_1) at natural abundance. Spectra were acquired in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

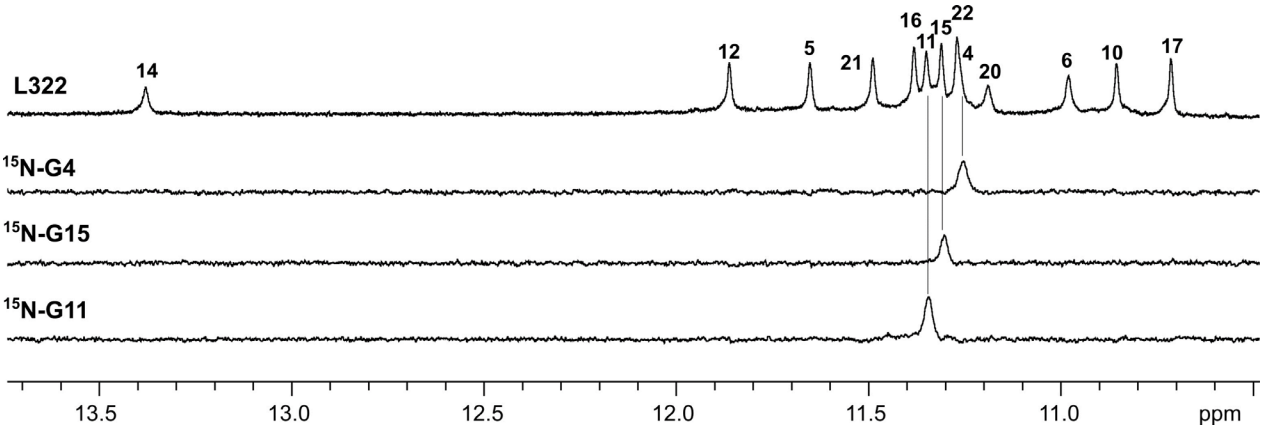

Figure S5. (A) Imino proton spectral region of non-labeled L322 (top) and ^{15}N-edited spectra for L322 sequences site-specifically labelled at G4, G15, and G11 (10% ^{15}N enrichment).

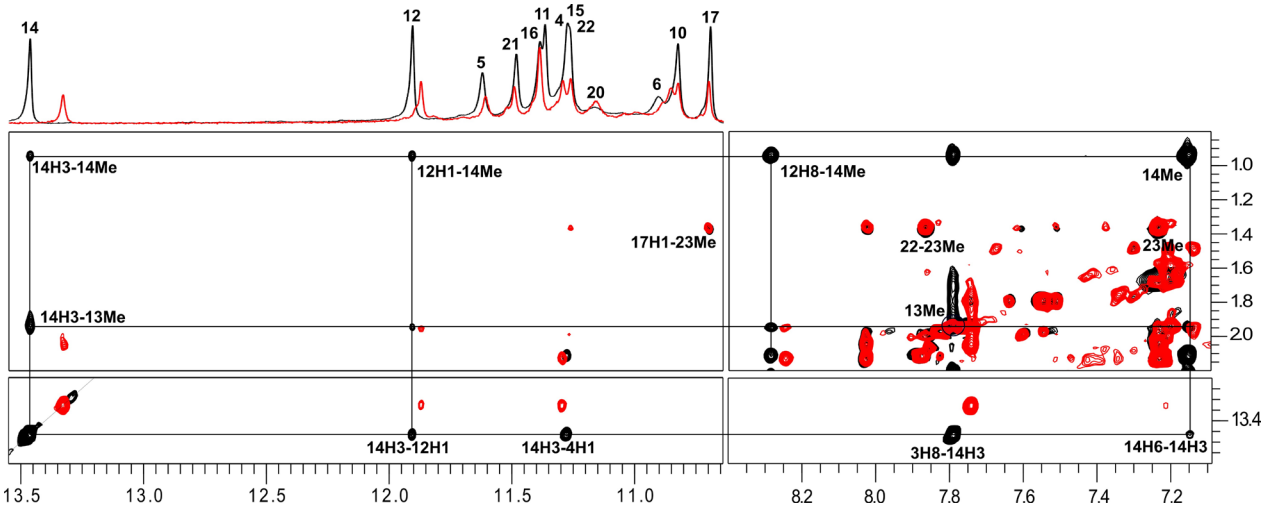

Figure S6. Superimposed 1D imino proton and 2D NOESY spectral regions of L322 (black) and L322-T14U (red) showing cross-peaks of the imino proton of residue 14 to H1/H8/H6(ω_2) protons (bottom) as well as H1/H3(ω_2)-H2'/H2"/Me(ω_1) (top, left) and H6/H8(ω_2)-H2'/H2"/Me(ω_1) (top, right) cross-peaks. Of note, the intra-residue NOE cross-peak (14H3-14Me) is missing for the L322-T14U sequence, confirming assignment of the downfield-shifted 14H3 resonance. NMR spectra were recorded in a 10 mM potassium phosphate buffer, pH 7.0, at 5°C.

Table S1. NMR restraints and structural statistics of L322.

sequence	L322
NOE distance restraints	
intra-residual	104
inter-residual	194
exchangeable	58
other restraints:	
hydrogen bonds	48
dihedral angles	41
planarity	3
chirality	125
structural statistics:	
pairwise heavy atom RMSD value (Å)	
all residues	1.9 ± 0.4
G-tetrad core	0.6 ± 0.1
NOE violations:	
maximum violation (Å)	0.21
mean NOE violation (Å)	0.0027 ± 0.0007
deviations from idealized geometry:	
bond lengths (Å)	0.01 ± 0.0001
bond angles (degree)	2.3 ± 0.03

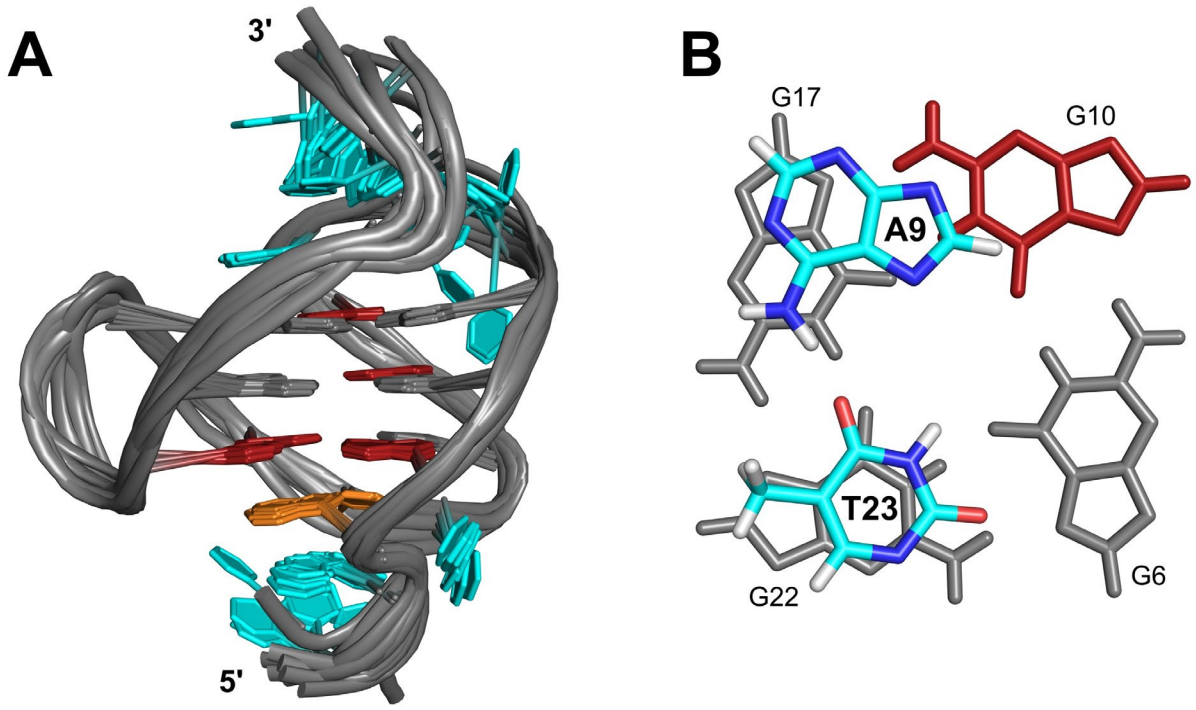

Figure S7. (A) Superposition of 10 lowest-energy structures of L322; *anti*- and *syn*-guanosines are colored grey and red, respectively; lateral loop and overhang residues are colored cyan; an A3·T14 reverse Hoogsteen base pair stacked on the 5'-tetrad is shown in orange; propeller loop residues have been omitted for clarity. (B) View on the 3'-tetrad with A9 in the first lateral loop and 3'-flanking T23 stacked on top.

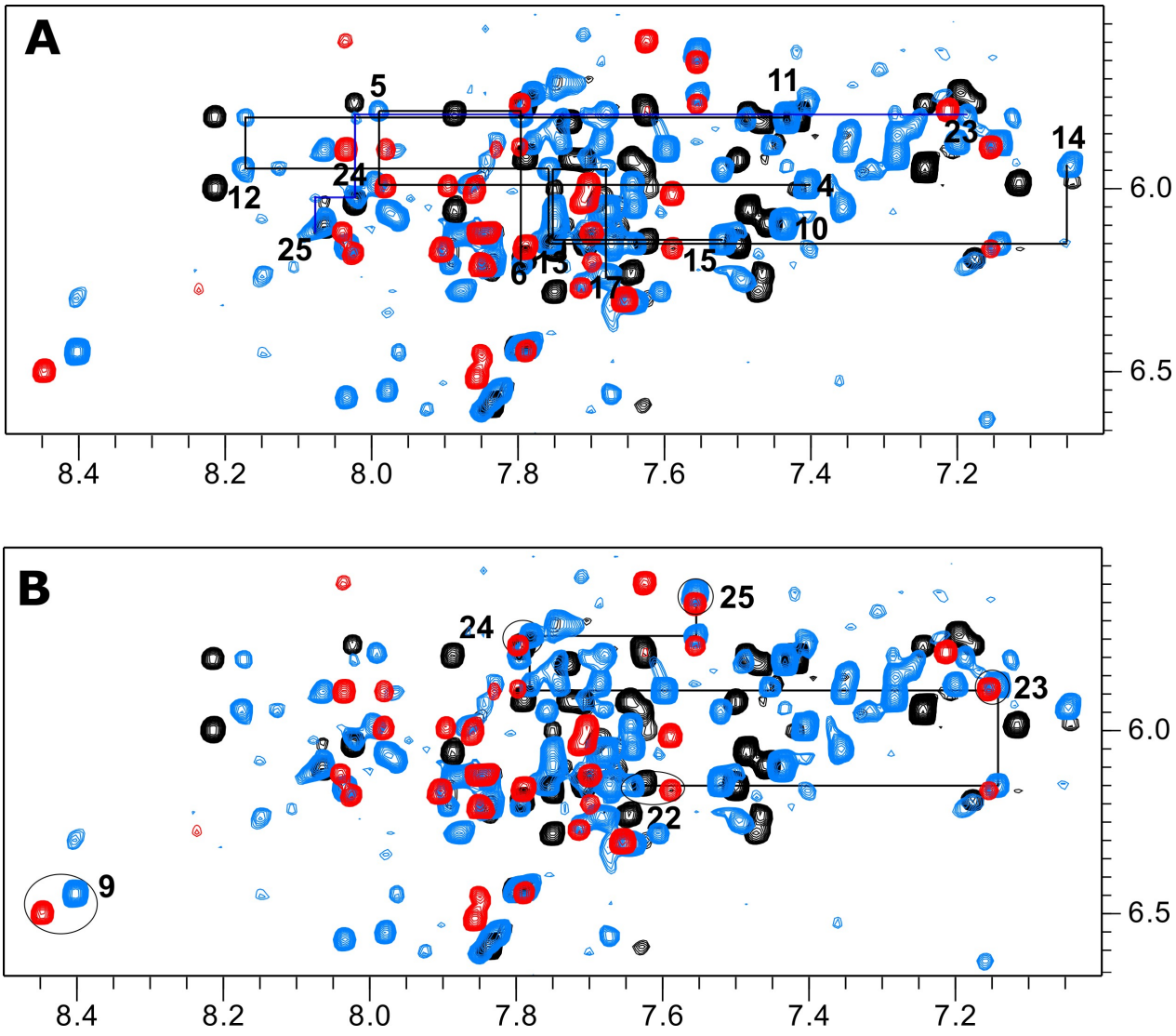

Figure S8. Superimposed H6/H8-H1' NOESY spectral regions of L311(red), L322 (black), and L322-A3T (blue) highlighting some representative NOE cross-peaks of L322-A3T with (A) L322-matching connectivity patterns and with (B) L311-matching connectivity patterns.

Spectral assignment of dsL322. The dsL322 sequence shows twelve major G Hoogsteen imino proton resonances in the 10.5-12.0 ppm region, indicating formation of a three-layered G-quadruplex core. In addition, the presence of Watson-Crick imino resonances between 12-14 ppm strongly suggests the formation of a duplex by base pairing of complementary 4-nt 5'- and 3'-flanking overhangs (Figure S9). Resonance assignments for dsL322, in particular for the duplex domain as well as for the first and the second G-column of the G-quadruplex domain, were strongly supported by a comparison with NOE patterns previously found for a closely related structure (PDB ID 6R9L).[1] All imino protons engaged in Watson-Crick hydrogen bonding of the duplex were assigned except for the G1 imino signal that was broadened beyond detection due to fraying effects and fast solvent exchange. On the basis of strong intra-residual H8-H1′ cross-peak intensities in 2D NOESY spectra, five *syn*-guanosines could be identified and further confirmed by typical chemical shifts of C8-H8 correlation peaks in ^{1}H-^{13}C HSQC spectra (Figure S10). Four out of the five *syn*-residues feature a rectangular H8-H1' NOE connectivity pattern typical for *syn-anti* steps with their 3'-neighboring G (Figure S9B).

Sequential H6/H8-H1' NOE contacts could be followed from G1 to T4. The first G column G5-G6-G7 comprises a G(*syn*)-G(*syn*)-G(*anti*) alignment, confirmed by tracing NOE connectivities to TCA residues of the first loop. Likewise, the fourth G-column could be identified following a NOE walk along *syn*-G21, *anti*-G22, and *anti*-G23 that continues up to residues in the 3'-flanking sequence. Notably, discrimination of the second and the third G-tract was supported by the comparison with a related construct (PDB ID 6R9L) as mentioned above. Both of these latter two G-columns also feature *syn-anti-anti* steps. This points to a G-quadruplex with a +(lpp) loop progression. Notably, first and also second residues of the two TT propeller loops exhibit isochronicity for both their aromatic and anomeric protons.

Guanine imino protons were assigned through correlations with their H8 in ^{1}H-^{13}C HMBC experiments. Tetrad polarities were subsequently determined by characteristic intra-tetrad H8-H1 NOE contacts and further verified by H1-H1 NOE contacts. Strong NOE connectivities of the T4 methyl protons were observed to imino protons of G5 and G13 whereas weak NOE contacts were observed to imino protons of G18 and G23, confirming assignments of residues in the 5'-tetrad (Figure S11). Additionally, A10 H2 shows NOE contacts of moderate intensity to all G imino protons located in the 3'-tetrad, i.e., G7, G11, G21, and G16 (Figure S9D).

Taken together, the G-quadruplex comprises three G-tetrads with homopolar and heteropolar tetrad stackings. Hoogsteen hydrogen bonds from donor to acceptor within the three G-quartets run along G5-G13-G18-G23 (5'-tetrad), G6-G12-G17-G22 (central tetrad), and G7-G21-G16-

G11 (3'tetrad) forming a +(lpp) topology (hybrid-1R) with a first 3-nt lateral loop followed by two 2-nt propeller loops.

References

1 B. Karg, S. Mohr and K. Weisz, *Angew. Chemie Int. Ed.*, 2019, **58**, 11068–11071.

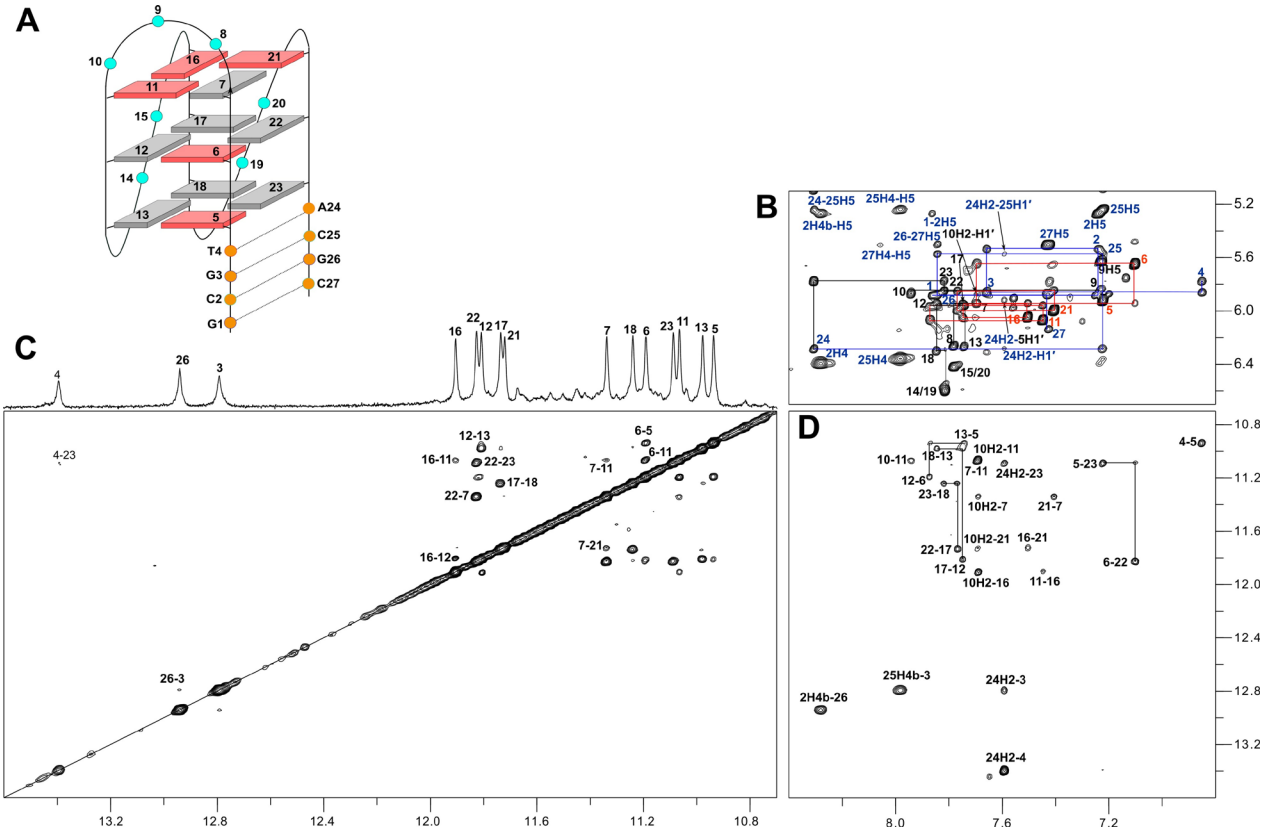

Figure S9. (A) Hybrid-1R topology +(lpp) of dsL322 with residue numbers; *anti*- and *syn*-G residues of the G-core are colored grey and red, respectively; loop and duplex residues are represented by circles colored cyan and orange, respectively. (B)-(D) 2D NOESY spectral regions (mixing time 300 ms) of dsL322. (B) H8/H6(ω_2)-H1'(ω_1), (C) H1/H3-H1/H3, and (D) H8/H6(ω_2)-H1/H3(ω_1) NOE spectral region; in (B), *syn*-guanosines and a typical rectangular connectivity pattern for *syn-anti* steps are shown in red whereas labels and NOE walks in the duplex domain are shown in blue; in (C), the 1D imino proton spectral region with assignments is shown on top. NMR spectra were recorded in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

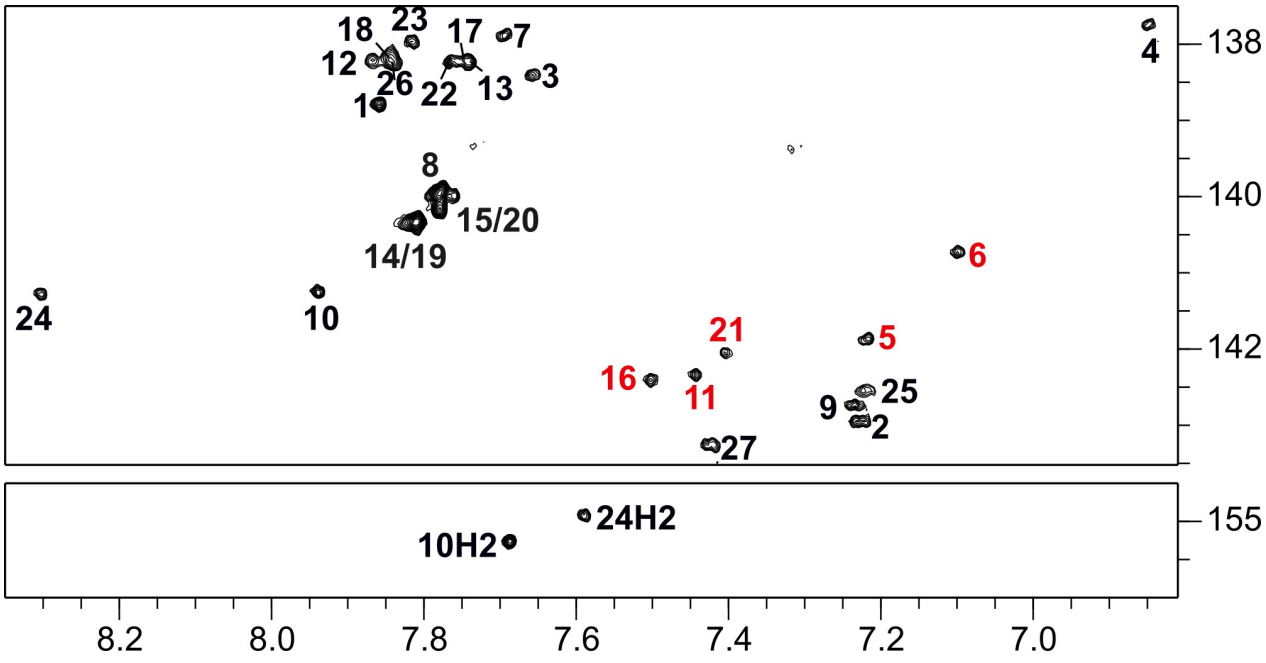

Figure S10. ^{1}H-^{13}C HSQC spectrum of dsL322 showing H8/H6(ω_2)-C8/C6(ω_1) (top) and adenine H2-C2 correlations (bottom). *Syn*-guanosines with their upfield-shifted H8 and downfield-shifted ^{13}C8 are marked in red. Spectra were acquired in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

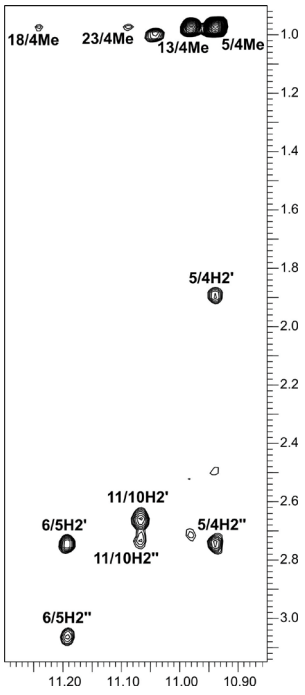

Figure S11. 2D NOE spectral region of dsL322 showing H1(ω_2)-H2'/H2''/Me(ω_1) connectivities. Spectra were acquired in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

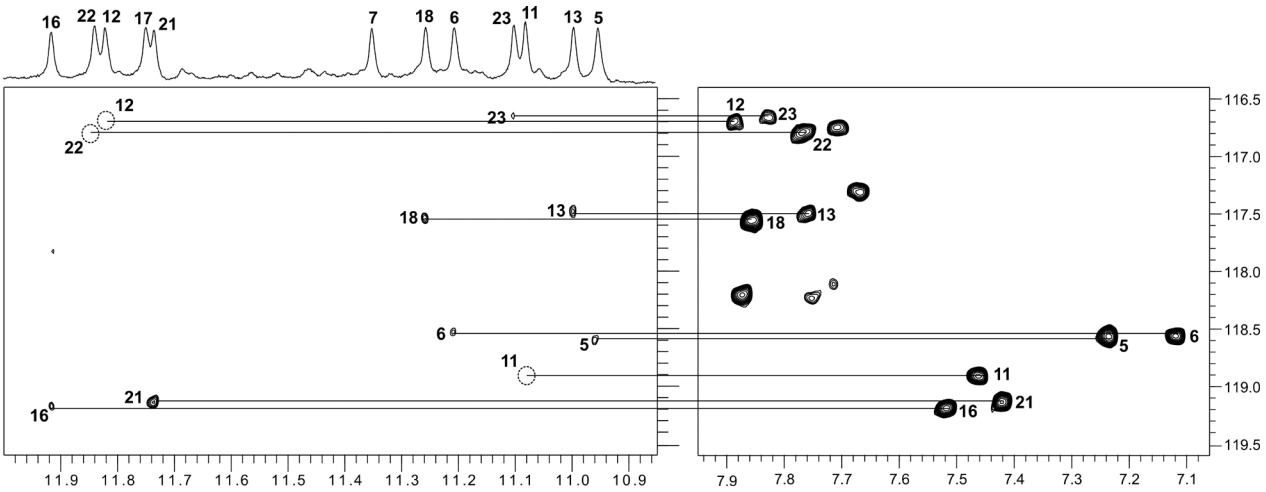

Figure S12. ^{1}H-^{13}C HMBC spectrum of dsL322 showing G H1(ω_2)-G H8(ω_2) correlations through long-range couplings to their ^{13}C5(ω_1); dotted circles indicate correlation peaks only seen at lower threshold levels. Spectra were acquired in a 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

Table S2. NMR restraints and structural statistics of dsL322.

sequence	dsL322
NOE distance restraints	
intra-residual	112
inter-residual	164
exchangeable	53
other restraints:	
hydrogen bonds	70
dihedral angles	50
planarity	7
chirality	135
structural statistics:	
pairwise heavy atom RMSD value (Å)	
all residues	2.2 ± 0.4
G-tetrad core	0.9 ± 0.2
NOE violations:	
maximum violation (Å)	0.10
mean NOE violation (Å)	0.0008 ± 0.0003
deviations from idealized geometry:	
bond lengths (Å)	0.01 ± 0.0001
bond angles (degree)	2.3 ± 0.03

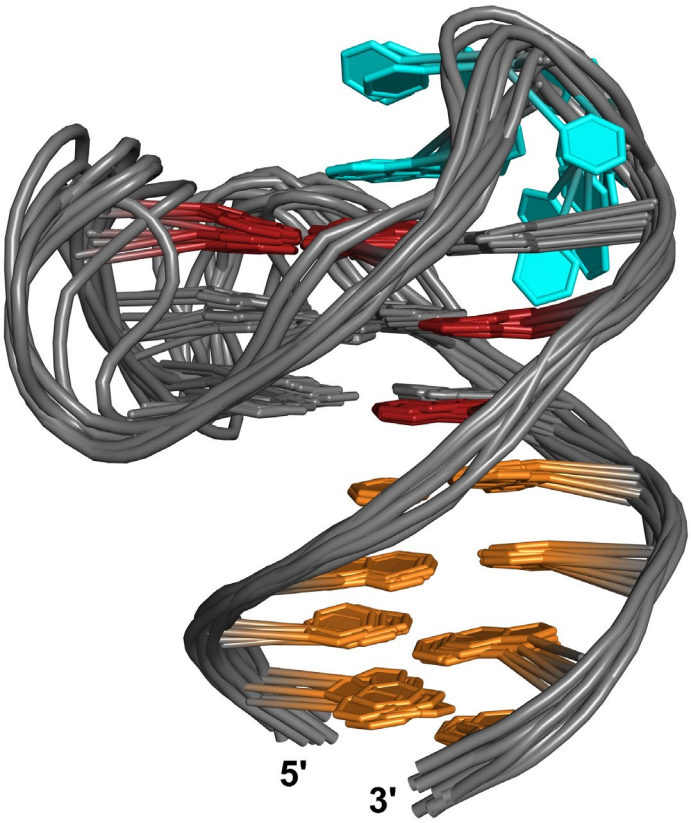

Figure S13. (A) Superposition of 10 lowest-energy structures of dsL322; *anti*-guanosines, *syn*-guanosines and lateral loop residues are colored grey, red, and cyan, respectively; overhang residues forming a coaxially stacked duplex domain are colored orange; propeller loop residues have been omitted for clarity.