Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202401190
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

Hot Paper

Solution-State Structure of a Long-Loop G-Quadruplex Formed Within Promoters of *Plasmodium falciparum* B *var* Genes

Marina Juribašić Kulcsár,[a, b] Valérie Gabelica,[c] and Janez Plavec*[a, d, e]

We report the high-resolution NMR solution-state structure of an intramolecular G-quadruplex with a diagonal loop of ten nucleotides. The G-quadruplex is formed by a 34-nt DNA sequence, *d*[$CAG_3T_2A_2G_3TATA_2CT_3AG_4T_2AG_3T_2$], named UpsB-Q-1. This sequence is found within promoters of the *var* genes of *Plasmodium falciparum*, which play a key role in malaria pathogenesis and evasion of the immune system. The [3 + 1]-hybrid G-quadruplex formed under physiologically relevant conditions exhibits a unique equilibrium between two structures, both stabilized by base stacking and non-canonical hydrogen bonding. Unique equilibrium of the two closely related 3D structures originates from a North-South repuckering of deoxyribose moiety of residue T27 in the lateral loop. Besides the 12 guanines involved in three G-quartets, most residues in loop regions are involved in interactions at both G-quartet-loop interfaces.

Introduction

G-rich nucleic acid sequences form secondary structures called G-quadruplexes (G4).[1] G4s consist of Hoogsteen-type hydrogen-bonded (H-bonded) guanines, forming G-quartets (Figure 1a). Consecutive G-runs are connected by loops. G4 observation *in vivo* and extensive studies *in vitro* provide compelling evidence for their involvement in 'normal' biology and genomic pathologies, and support the quest for ligands that will selectively target G4s.[2–6] Putative G4-forming sequences are non-randomly distributed within genomes.[2–5] They are usually conserved within telomeric and promoter regions of different species, from bacteria to mammals.[2,3] In the last decade, the formation of G4s has been related to regulation of biological processes in live cells, including DNA replication, transcription, translation and telomere maintenance.[2–4] Their participation in the initiation of replication may facilitate or

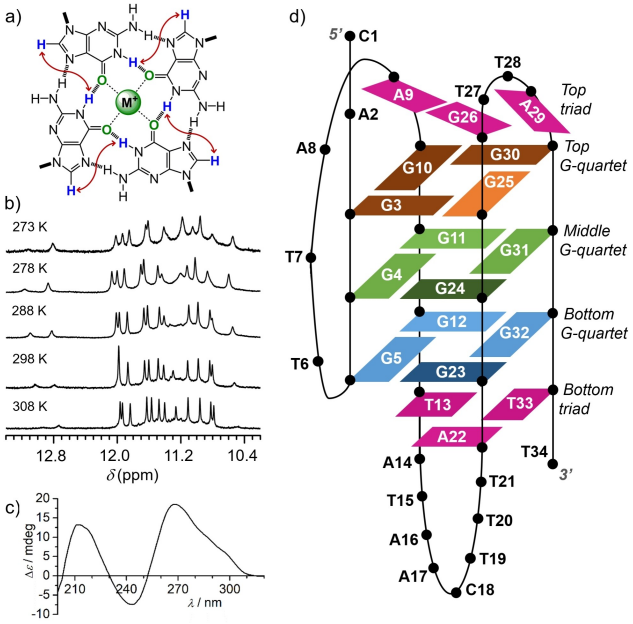

Figure 1. a) G-quartet showing inter-residual imino-aromatic (H1-H8) contacts supportive of its formation. b) Imino region of ^{1}H NMR spectra in 273–308 K temperature range. Spectra were recorded using 1.6 mM oligonucleotide concentration *per* strand in 10% (*v/v*) $^2H_2O/H_2O$ in the presence of 150 mM KCl in a K-phosphate buffer (pH 7.0). c) CD spectrum of the UpsB-Q-1 recorded at 298 K. d) Topology of the G4 adopted by the UpsB-Q-1. Darker shades of color in each of the three G-quartets indicate *syn* orientation across glycosidic bonds of G3, G10, G23, G24 and G30.

[a] Dr. M. Juribašić Kulcsár, Prof. Dr. J. Plavec
Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana (Slovenia)
Tel: +386 1 4760353
Fax: +386 1 4760300
E-mail: janez.plavec@ki.si

[b] Dr. M. Juribašić Kulcsár
Division of Physical Chemistry, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000 Zagreb (Croatia)

[c] Prof. Dr. V. Gabelica
School of Pharmaceutical Sciences, University of Geneva, 1 rue Michel-Servet, CH-1211 Geneva 4 (Switzerland)

[d] Prof. Dr. J. Plavec
Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana (Slovenia)
Tel: +386 1 4760353
Fax: +386 1 4760300
E-mail: janez.plavec@ki.si

[e] Prof. Dr. J. Plavec
EN-FIST Centre of Excellence, Trg OF 13, SI-1000 Ljubljana (Slovenia)
Tel: +386 1 4760353
Fax: +386 1 4760300
E-mail: janez.plavec@ki.si

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202401190

repress the process by interacting with the replication-relevant proteins.[4]

Plasmodium parasites are still annually responsible for over 245 million malaria cases and about 620 000 malaria deaths worldwide.[7] When the unusually A/T-biased genome of *P. falciparum* (overall 81 % A/T) was searched for putative G4 motifs,[8] a few non-randomly located G-rich sequences in non-telomeric regions were found, only after allowing extended 11-nt long loop regions. Four distinct 34-nt G4-forming sequences, named UpsB-Q-1 to UpsB-Q-4, were found within promoters of subtype B *var* genes. Importantly, these sequences do not exist in the human genome.[8] *P. falciparum* avoids immune response by expressing about 60 polymorphic *var* genes that have been divided into five subtypes named A to E. Expression of *var* genes is mutually exclusive and regulated mainly at the transcription initiation level.[9,10] Identifying specific *var* gene regulatory factors and pathways will deepen our understanding of the malaria pathogenesis and could lead to improved treatment. A recent study has indicated that G4s are involved in recombination and diversification of *var* gene families in *Plasmodium*.[11,12] G4s are consistently present in the nuclei during *P. falciparum* intraerythrocytic development and influence the growth of the blood-stage parasites.[13] In addition, RNA G4s influence the gene expression and can affect *in vitro* translation in the parasite.[14] Besides, *P. falciparum* genome revealed genetic codes for two RecQ helicases,[6,11] which facilitate G4 unwinding.[15]

UpsB-Q-1, *d*[CAG GG^{5}T TAA G^{10}GG TAT^{15} AAC TT^{20}T AGG G^{25}GT TAG^{30} GGT T] (detailed residue numbering is available in Figure 1d), is the most frequent of the G-rich sequences in *var* genes. Its guanine tract distribution follows known rules for the formation of standard G4, with four G-tracts of at least three contiguous guanines that are separated by AT-rich loops of 4, 10 and 4-nt. The number of residues in the loops affects the structure of G4s, with loop elongation usually reducing the stability of the structure. Such long loops without Watson-Crick pairing within the loop are thus uncommon,[16–20] although some studies have suggested that prolongation of the central loop region even up to 30 residues does not prevent G4 formation at physiological temperatures.[21–23] Putative long-loop G4-forming sequences are usually found within promoter regions[18–20,24–27] and thus these G4s could have a prominent influence on biological processes such as transcription and gene regulation.[1–6,14,21]

Here we report high-resolution structural studies of the UpsB-Q-1 G4 by NMR spectroscopy in solution. We identified structural factors contributing to the overall thermodynamic stability of two closely related structures that are involved in a slow exchange process on NMR chemical shift timescale. Both structures are well-defined and exhibit G4s comprising a three-quartet core and a [3 + 1]-hybrid topology with propeller, diagonal and lateral loops exhibiting unique structural features. Unique equilibrium of the two fully characterized and closely related 3D structures originates from a repuckering of deoxyribose moiety of residue T27 in the third lateral loop. High-resolution structures provide insights into H-bonding and base stacking interactions within loops that stabilize these biologically-relevant G4 structures. Described structural insights open a door towards using novel structural motifs as recognition elements for biomolecules and drug targets in search of a pharmaceutical lead.

Results

NMR assignment and topology. UpsB-Q-1 comprising a 34-nt DNA sequence has been folded into a three-quartet G4 structure according to the observed signals of imino protons of guanine residues forming a G-quartet (Figure 1). The NMR sample exhibited long-term stability and endured repeated changes in temperature (273–308 K), pH (5.5–8.0) or potassium ion concentration (70–150 mM). The melting temperature of 320 K for UpsB-Q-1 in 150 mM aqueous KCl[8] suggests that its G4 structure can exist in the temperature range relevant to organisms involved in the *P. falciparum* life cycle.

Twelve imino signals in δ 10.7–12 ppm region of ^{1}H NMR spectrum of UpsB-Q-1 (Figure 1b) suggest the formation of three G-quartets (Figure 1d). We unambiguously assigned the resonances using partial (8 %) residue-specific isotopic labelling and single-residue mutations of thymine to uracil accompanied by multiple homo- and heteronuclear NMR experiments. ^{15}N-filtered HSQC spectra of isotopically guanine-labelled oligonucleotides (Figure 2a) reveal that G26 in the third G-tract is not involved in a G-quartet, but is rather a part of the third loop (Figure 1d).

Three G-quartets, G3→G25→G30→G10 (top), G4→G11→G31→G24 (middle) and G5→G12→G32→G23 (bottom), have been identified following the assignment of the guanine imino and aromatic protons. Relative positions of the three G-quartets are supported by the imino-aromatic and imino-imino connectivities in 2D ^{1}H-^{1}H NOESY spectrum (τ_m 250 ms, Figure 2c). Spectral data suggest the same directionality of H-bonds within the middle and the bottom G-quartets and a directionality change between the top and the middle G-quartets (Figure 2d).

In addition to 12 guanine imino signals, ^{1}H NMR spectrum of UpsB-Q-1 shows other broad(er) imino signals at δ 13.04, 12.80 and 10.52 ppm at 298 K (Figure 1b), which have been assigned to imino protons of T27, G26 and T7, respectively, using ^{15}N-filtered HSQC spectra of residue-specifically labelled samples. Chemical shifts and intensity of signals assigned to imino protons from residues T7, G26 and T27 not involved in G-quartets are temperature-dependent, suggesting dynamics within the loops. However, NOESY connectivities are conserved and the same sequential walk is found in the 273–313 K temperature range, which suggests that the overall topology of the G4 is unchanged.

UpsB-Q-1 adopts a [3 + 1]-hybrid topology (Figure 1d). Strong intra-residual H1′-H8 cross-peaks in 2D ^{1}H-^{1}H NOESY spectrum have been observed for the five guanine residues in *syn* conformation *i.e.*, G3, G10, G23, G24 and G30 (Figure 2). All other residues adopt an *anti* conformation across glycosidic bonds. All three loops connect *anti* guanines at the 3′-ends of G-tracts (G5, G12 and G25) with *syn* guanines at the 5′-ends of G-tracts (G10, G23 and G30). The first two G-tracts are in parallel

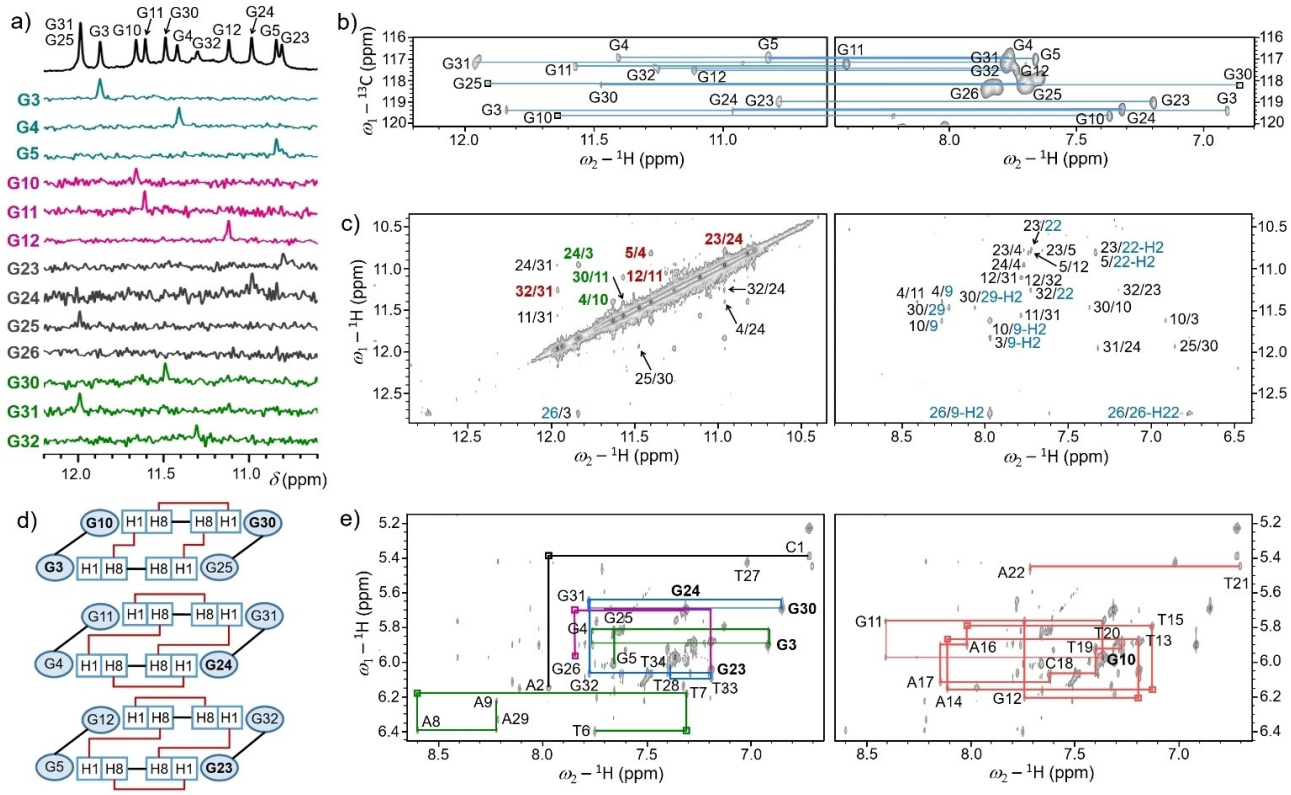

Figure 2. NMR spectral assignment of UpsB-Q-1. a) 1D ^{15}N-filtered HSQC spectra of residue-specifically ^{15}N-labelled (8%) UpsB-Q-1 at the indicated positions of guanine, at 298 K. b) Imino and aromatic regions of 2D ^{1}H-^{13}C jrHMBC spectrum at 308 K showing correlations between H8 and the corresponding H1 proton via guanine C5 atom in 10% (*v/v*) $^{2}H_2O/H_2O$. c) H1-H1 and H1-H8 connectivities in 2D ^{1}H-^{1}H NOESY spectrum (τ_m 250 ms) at 308 K in 10% (*v/v*) $^{2}H_2O/H_2O$. d) Topology of G4 core of three G-quartets displaying the observed H1-H8 contacts shown in red within individual G-quartets. e) Aromatic-anomeric region of 2D ^{1}H-^{1}H NOESY spectra (τ_m 250 ms) at 308 K in 100% $^{2}H_2O$. Sequential walk is marked for C1-A2 (black), G3-A9 (green), G23-A29 (magenta) and G30-T34 (blue) in the left panel. For clarity, the right panel highlights sequential walk within G10-A22 segment (red). Spectra in b–e were recorded using 1.6 mM (in 10% (*v/v*) $^{2}H_2O/H_2O$) or 1.8 mM (in 100% $^{2}H_2O$) oligonucleotide concentration *per* strand in the presence of 150 mM KCl in a K-phosphate buffer (pH 7.0).

orientation and are connected by a four-residue (T6-T7-A8-A9) propeller-type loop (Figure 1d). G12 and G23 from the second and the third G-tracts are connected by a diagonal-type loop comprising ten residues (T13···A22). A lateral loop comprising four nucleotides (G26-T27-T28-A29) links the third and the fourth G-tract. The third G-tract, G23-G24-G25, is antiparallel to the other three G-tracts.

Sequential aromatic-anomeric proton connectivities are observed along the diagonal loop except at the T20-T21 step (Figure 2e and Figures S2–S14). However, sequential $(H1')_n$-$(H6/8)_{n+1}$ connectivities for the 5′-overhang C1-A2 as well as T6-T7-A8 and G26-T27-T28-A29 loop segments are weak (Figure 2e). These data, along with the observed long-range NOE contacts, suggest a well-defined structure of G-quadruplex adopted by UpsB-Q-1.

Interactions of loops and overhang residues with the neighboring G-quartets. NOE contacts between guanine protons in the G4 core with the loop and overhang residues (Supplementary Figures S8–S11) uncover a unique structure comprising two base triads located above the top and below the bottom G-quartets.

Interactions above the top G-quartet. NOE contacts between A29H8 and G30H1 (Figure 2c) as well as A29H2 and G30H8 (both of medium intensity), G26H22-A29H8 (medium), G30H1′-A29H8 (weak) and G26H22-A29H2 (strong, Figure 3a) support stacking of the purine moiety of A29 on the G30 nucleobase and its involvement in a sheared G26·A29 base pair (Figure 3). The signal of G26H1 at δ 12.74 ppm in 1D ^{1}H spectrum of UpsB-Q-1 at 308 K is shifted considerably downfield compared to the unpaired guanine imino proton of a typical sheared GA base pair, which usually resonates in the δ 9.5–10.5 ppm range.[28,29] Thus, the observed chemical shift suggests that G26H1 is involved in H-bonding interactions as part of an A9-G26·A29 base triad. Accordingly, long-range NOE contacts between the proton A9H2 from the propeller loop and the protons G3H1, G3H8 and G10H1 in the top G-quartet suggest a well-defined positioning of the A9 nucleobase moiety above the top G-quartet (Figures 2c and 3a). Together with NOE connectivities of the amino proton G26H22 with all four imino protons of the top G-quartet and with A9H2 and A29H8, the NMR data confirm a 'face-to-face' A9-G26 interaction. This type of GA pairing has been reported in DNA repair processes.[29] In addition, signals of adenine amino protons in 1D ^{15}N-filtered HSQC spectra recorded at 308 K using samples with residue-specifically labelled adenines A9 (δ 6.5 ppm) and A29 (δ 6.8 ppm) fully support the formation of an A9-G26·A29 base triad (Figure S12).

A careful examination of the NMR spectra reveals two sets of resonances for T27 sugar protons (H1′, H2′, H2″, H3′ and H4′,

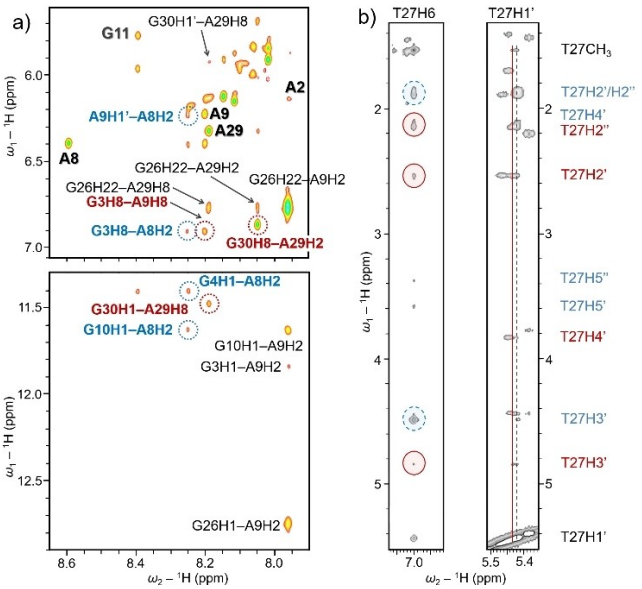

Figure 3. a) Selected NOE connectivities of the aromatic protons of A8, A9 and A29 with nearby protons observed in the 2D ^{1}H-^{1}H NOESY spectrum (τ_m 150 ms). b) Selected parts of the 2D ^{1}H-^{1}H NOESY spectrum (τ_m 250 ms) showing the doubling of the NOE cross-peaks for T27 together with the respective assignments related to structure I (red) and structure II (blue). The spectra were recorded at 308 K using the sample ($c = 1.6$ mM *per* strand) in 10% (*v/v*) $^{2}H_2O/H_2O$, pH 7.0.

Figure 3b) as well as two sets of aromatic (H2 and H8) and sugar protons (H1′, H2′, H2″, H3′ and H4′) for A2 from the 5′-overhang. Moreover, the assigned signals of the nucleobases of A2, G25 and G26 appear weaker than signals of other residues. However, no doubling of signals for residues other than A2 and T27 has been detected. These data indicate two closely related G4 structures involved in a slow exchange on the NMR chemical shift time-scale. The two structures exhibit the same topology and differ only in the structure's part above the top G-quartet. Therefore, we collected additional NMR data in order to characterize them in more detail.

The cross-peaks corresponding to the sugar protons of T27 in the 2D ^{1}H-^{1}H DQF-COSY spectrum (Figures S16–S18) show a strong bias towards the C3′-*endo* sugar pucker in one structure (structure I) and towards the C2′-*endo* in the other (structure II). The preference of the C3′-*endo* pucker for T27 in structure I together with the strong intra-residual T27H3′-H6 NOE contact and the chemical shift of T27H4′ at δ 2.10 ppm are reminiscent of an inverted 3- or 4-nt lateral GN(M)A loop, where N and M are any nucleotide.[29–32] In the edgewise loop of the G4 structure I, a sheared GA base pair is formed, resulting in orientation of A29H8 towards the center of the top G-quartet. In addition, nucleobase of the unpaired nucleotide N in the GN(M)A loop, T27 here, usually stacks upon its 5′-neighbour, *i.e.*, G26 here. A2H2 from structure I shows long-range NOE contacts with H8 and H1′, H4′ and H5′/H5″ protons of A9 suggesting positioning of the A2 nucleobase in proximity of A9 in the propeller loop. No NOE cross-peaks between the protons of A2 and T27 were observed for structure I. However, chemical shifts obtained by 1D ^{15}N-filtered HSQC spectra of samples with residue-specifically labelled A2 (δ 6.7 ppm at 308 K) and T27 (δ 13.04 ppm at 298 K) suggest the formation of the A2·T27 base pair over the A9-G26·A29 base triad in structure I.

Structure II contains the same structural features as structure I except for residues A2 and T27. The NMR data for structure II shows NOE cross-peaks of A2H8 with H1′ (weak) and H2′/H2″ (medium) of T27 indicates that the 5′-overhang is closer to the lateral loop than in structure I. The NOE data for T27 are consistent with the C2′-*endo* sugar pucker (Figure 3, cross-peaks marked in red). This sugar conformation pushes the pyrimidine base of the T27 residue further away from the backbone (Figure 4c), which in turn disrupts the A2·T27 interaction in structure I. Consequently, the A2 in the 5′-overhang of structure II is more mobile and moves away from the pyrimidine base of T27.

Interactions involving the middle G-quartet. The correlations of the T7 imino proton with the amino and imino protons of residue G4 observed in the 2D ^{1}H-^{1}H NOESY spectrum (τ_m 150 ms) below 280 K (Supplementary Figures S2 and S8) indicate that T7 is turned toward the G4. We find that, in addition to T7, A8 is also oriented with its nucleobase toward the G-quartet in the core of G4 structure and interacts with the top and middle G-quartets according to the characteristic NOE

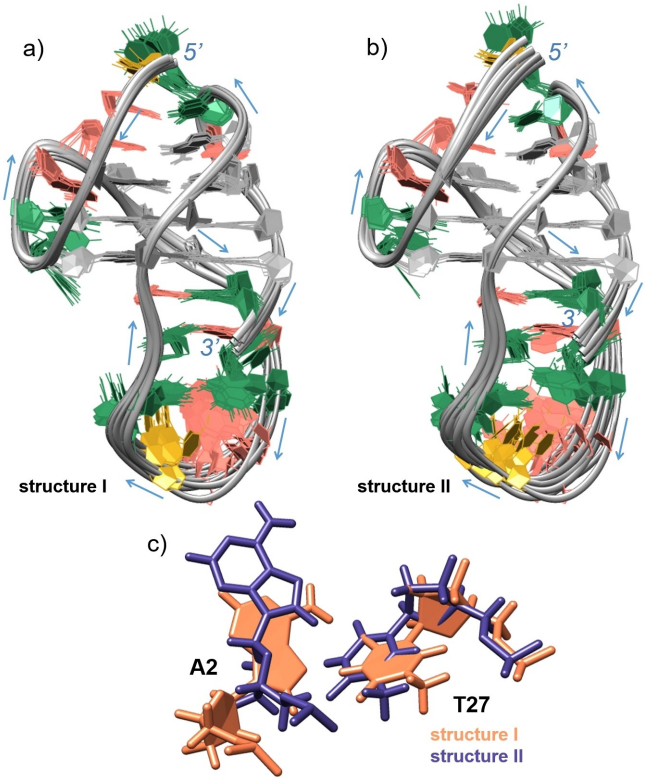

Figure 4. Two structural ensembles of G4s adopted by UpsB-Q-1. Superposition of the family of 10 structures with the lowest energies for a) structure I and b) structure II. Blue arrows indicate the 5′→3′ direction of the oligonucleotide chain. Guanines and the backbone are presented in grey, adenines in pink, cytosines in yellow, and thymines in green. c) The residues A2 and T27 in the structure I (pink) and in the structure II (violet). The structures I and II have been overlapped using the least-square superposition of guanine residues involved in the top G-quartet. Hydrogen atoms are omitted for clarity.

contacts of A8H2 (see Figure 3) with G3H8, G10H1 and G10 sugar protons (top quartet) and G4H1 and G11H8 (middle quartet).

Interactions below the bottom G-quartet. The NOE contacts of the aromatic protons of A22, A22H2 with G5H1 and G23H1 and A22H8 with G23H1 and G32H1 (Figure 2c), show that the nucleobase of A22 is positioned below the bottom G-quartet. Long-range NOE cross-peaks of T33CH3 protons and G12H1, T13H1′, T13H4′, A14H1′ and A14H4′, accompanied by sequential contacts of T33 with G32, show that T33 is stacked on the bottom G-quartet in addition to A22. A significant, albeit weak, NOE contact between G5H1 and T13CH3 positions the T13 nucleobase near the bottom G-quartet.

High-resolution UpsB-Q-1 G4 structure. The solution-state structure of [3 + 1]-hybrid G-quadruplex adopted by UpsB-Q-1 was calculated using the NOE-derived distance, as well as glycosidic torsion angle (*syn* or *anti*), chirality and sugar type (C2′-*endo* or C3′-*endo* pucker) restraints according to the experimental NMR data. Eight H-bonding restraints *per* G-quartet were used during structure calculations. Additional H-bonding restraints were used for base triads A9-G26·A29 and T13-A22·T33 (for both structures) and base pairs (A2·T27 only for structure I and A14-T21 for both structures). No other H-bonding restraints were employed. Planarity restraints were only used for the middle and bottom G-quartets throughout the calculations.

NMR-derived restraints allowed determination of high-resolution structures I and II adopted by the UpsB-Q-1 with overall pairwise heavy atom RMSD values of 1.1 Å and 1.3 Å, respectively (Figure 4, Table 1). The G-quartet core is the best-defined part of both structures with negligible structural variability (calculated RMSD of 0.6 Å for both structures; Table 1). The stacking of the top and middle G-quartet, which have opposite polarities of the H-bonds, involves an overlap of the imidazole units (Figure 5a), consistent with a twist angle of 87 ± 3°. The directional change of H-bonding between the middle and top G-quartet is clearly supported by medium intensity of imino-imino cross-peaks between guanines in the neighboring G-quartets, *i.e.*, G3/G24, G4/G10 and G11/G30 (cross-peaks marked in green in Figure 2c). In addition, the NOE contacts G3H1-G25H8 and G25H1-G31H1 were not observed. This was rationalized by broad and weak signals of the residue

Table 1. NMR restraints and structure statistics for UpsB-Q-1 G-quadruplex structures.

NMR restraints	Structure I	Structure II
NOE-derived distance restraints:		
Total	989	962
Intra-residue exchangeable	4	4
Intra-residue non-exchangeable	612	602
Inter-residue exchangeable	61	61
Inter-residue non-exchangeable	312	295
Hydrogen bonding restraints	36	34
Torsion angle restraints	63	62
G-quartet planarity restraints	24	24
Structure statistics	Structure I	Structure II
Violations:		
Mean NOE restraint violation (Å)	0.052 ± 0.010	0.050 ± 0.008
Max. NOE restraint violation (Å)	0.069	0.072
Max. torsion angle restraint violation (°)	0	0
Deviations from idealized geometry:		
Bonds (Å)	0.011 ± 0.000	0.011 ± 0.000
Angles (°)	2.285 ± 0.015	2.295 ± 0.020
Pairwise heavy atom RMSD (Å):		
Overall	1.14 ± 0.38	1.27 ± 0.31
G-quartets	0.61 ± 0.24	0.55 ± 0.25
G-quartets and propeller loop	0.80 ± 0.23	0.67 ± 0.20
G-quartets and diagonal loop	1.1 ± 0.5	1.3 ± 0.4
G-quartets and lateral loop	0.67 ± 0.21	0.62 ± 0.20
G-quartets and 5′-overhang	0.69 ± 0.22	0.62 ± 0.22
G-quartets and 3′-overhang	0.65 ± 0.23	0.63 ± 0.23

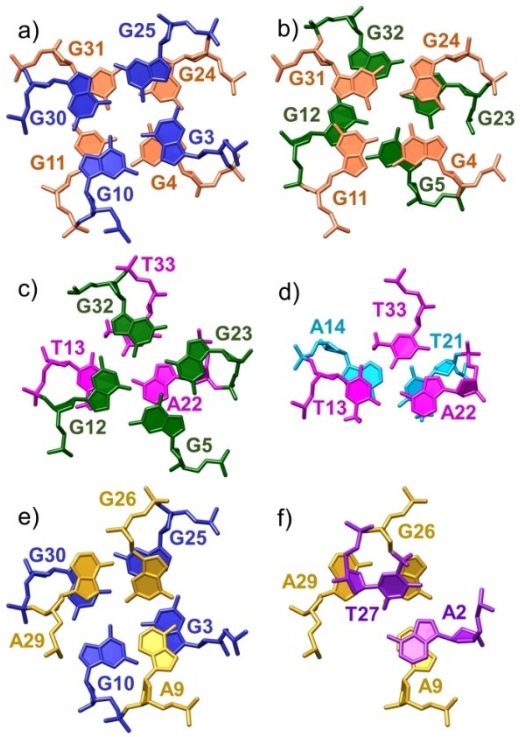

Figure 5. Base-base stacking of adjacent G-quartets and extra quartet residues for structure I. a) Top (blue) and middle (orange) G-quartets. b) Middle (orange) and bottom (green) G-quartets. c) Bottom G-quartet (green) and T13-A22·T33 base triad (magenta). d) T13-A22·T33 base triad (magenta) and A14-T21 base pair (light blue). e) Top G-quartet (blue) and A9-G26·A29 base triad (yellow). f) A9-G26·A29 base triad (yellow) and A2·T27 base pair (purple). The hydrogen atoms are omitted for reasons of clarity.

G25. Same H-bonding directionality in the middle and bottom G-quartets (Figure 5b) agrees with a twist angle of $62 \pm 7^{\circ}$ and is evidenced by medium-intensity imino-imino peaks of G4/G5, G11/G12, G23/G24 and G31/G32 (cross-peaks marked in red in Figure 2c). Well-defined structures of 4-nt propeller and 4-nt lateral loops are evident from RMSD values calculated for residues in G-quartets and respective loop regions that are below 0.8 Å (Table 1 and Figure 4). The long AT-rich diagonal loop is slightly less well-defined with RMSD values of 1.1 Å (structure I) and 1.3 Å (structure II) calculated for G-quartets and the T13-T21 region (Table 1).

Two of the three G-quartets exhibit three guanine residues in *anti* and one in *syn* conformation. The third G-quartet has one *anti* and three *syn* guanines. This arrangement of the G-quartet core with propeller, diagonal and lateral loops defines the groove dimensions as medium sized between the guanines of the same conformation, while the other two grooves are narrow and wide.[1a] The grooves of the G4 structure adopted by UpsB-Q-1 agree to this definition except in the size of the groove between G3 and G25. Indeed, the narrow groove between G4-G24 and G5-G23 widens in the top G-quartet to medium size which is attributed to the change in the glycosidic conformation of G3 and G25. Furthermore, P–P distance of 20.5 Å in G25-G30 of the top quartet is reduced to 17.3 Å for the P–P distance in the H-bonded sheared G26·A29 base pair (Supplementary Table S1). The groove between G23-G32, G24-G31 and G25-G30 is wide. Interestingly, the P–P distance is narrowed by 8.4 Å going from the bottom G-quartet to the bottom triad due to formation of the A22·T33 interaction (P–P distance for A22·T33 is only 11.6 Å). The particular groove arrangement and groove sizes are the same in structures I and II (see Supplementary Table S1 for detailed data), and offer opportunities for ligand design.

Propeller loop. Weak sequential $(H1')_n$-$(H6/8)_{n+1}$ connectivities within the T6-T7-A8-A9 loop (Figure 2e) coincide with a gradual counterclockwise twist, ending with A9 folded over the top G-quartet (Figure 6a). T6 is the most flexible residue in this loop and its nucleobase points away from the G-quartets. The positioning of the nucleobase of T7 towards the middle G-quartet is consistent with the observed NOE contacts between the imino proton of T7 and the imino and amino protons of G4. The nucleobase of A8 is almost co-planar with T7, which enables stacking within the propeller loop, an additional stabilization element that is uncommon in G4s (Figure 6). The A8H2 proton is buried inside the groove and oriented towards the G-core close to G10 (Figure 6b). A9 is stacked between A2 and G3, which is involved in the top G-quartet (Figure 6a). This structural arrangement induces a stretching of the backbone between A2 and G3 as well as a sharp turn between A8 and A9. The unusual conformation of the backbone is confirmed by the downfield shifts of $\Delta\delta$ 2.3 and 0.7 ppm for ^{31}P NMR signals of G3 and A9 compared to other phosphorous resonances.

Diagonal loop. To our knowledge, the 10-nt region T13···A22 is one of the longest G-quadruplex loops that have been resolved in high-resolution structures to date. It lacks the obvious sequence complementarity between opposing strands that normally allows stabilization of the structure by Watson-Crick base pairing in quadruplex-duplex hybrids.[21] The structure of this 10-nt loop is governed by H-bonding interactions of residues close to the G-quartet. The diagonal loop can be formally divided into two regions, (*i*) six structurally-defined residues involved in H-bonding (T13, A14, T15, T20, T21 and A22), and (*ii*) four less rigid residues A16-T19 with bases aligned to maximize base stacking within the loop (Figure 7). The purine base of A22 is located below the bottom G-quartet (between

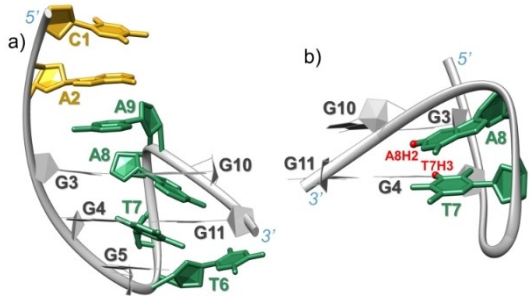

Figure 6. Structure of the propeller loop. a) Base-base stacking of the propeller loop residues allows b) T7 and A8 to be buried inside the groove with T7H3 pointing toward nucleobase G4, which is involved in the middle G-quartet, and A8H2 pointing toward the G-quadruplex core near residue G10. The guanines involved in the G-quartets and the backbone are gray. The hydrogen atoms except T7H3 and A8H2 in b) are omitted for clarity.

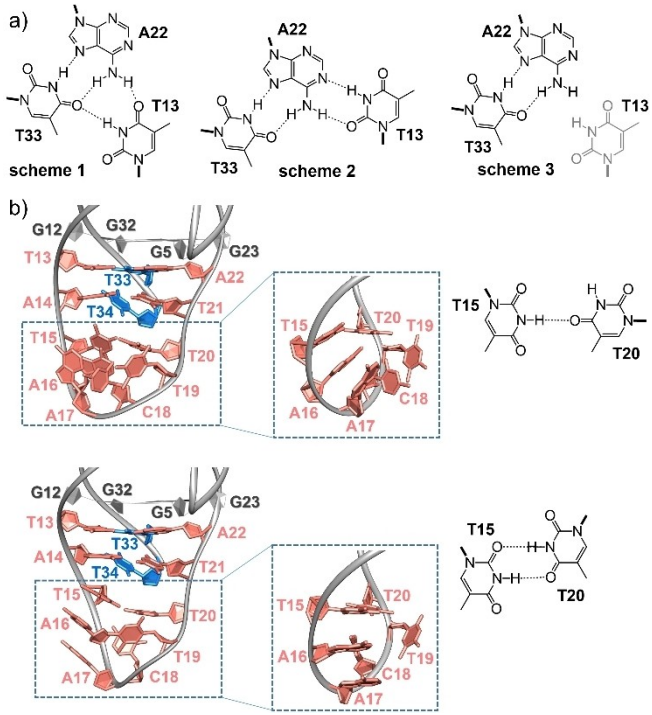

Figure 7. Diagonal loop structure. a) H-bonding schemes for T13, A22 and T33. b) Two conformations of the diagonal loop (pink), in interaction with the 3′-overhang (blue), that differ mainly in T15-T20 section. Enlarged and rotated views highlighting T15–T20 sections are in blue squares. Guanines involved in the bottom G-quartet and the backbone are grey. Hydrogen atoms are omitted for clarity.

G5 and G23, Figure 5c). The sugar moiety of T21 is turned towards A22, which influences the sequential connectivity between T20 and T21. This unusual backbone twist between T21 and A22 is supported by $\Delta\delta$ 0.7 ppm upfield shift of the A22 phosphorus signal with respect to the remaining phosphate resonances. The 3′-overhang residues are engaged in weak H-bond with the diagonal loop residues: T33 interacts with A22, while T34 interacts with A14 or T15.

The loop residues present unusual H-bonding arrangements. A22 forms a TAT base triad with T13 and T33. In agreement with the NOE data, T33 is placed below G32 and is H-bonded to the Hoogsteen face of A22. Three interaction schemes that can be envisaged have A22·T33 base pair and differ in T13 interactions (Figure 7a): T13 bridges A22·T33 base pair *via* A22H21-T13O4 and T13H3-T33O4 interactions (Scheme 1 in Figure 7a), T13 and A22 form a reversed Watson-Crick base pair in a TAT base triad (Scheme 2 in Figure 7a), or T13 is turned away from H-bonded A22 and T33 bases and interacts with T21, thus disrupting the A14-T21 base pair (Scheme 3 in Figure 7a). Molecular modelling calculations starting with the arrangements in Schemes 1, 2 or 3 showed that all three schemes fit the experimental NOE data and can interconvert readily. However, HSQC data (Supplementary Figure S9) suggests that one A14 amino proton and both A22 amino protons are H-bonded, in line with Schemes 1 and 2. However, the TAT base triad schematically presented in Scheme 2 converts into arrangements shown in Schemes 1 or 3 during calculations with no H-bonding restraints for the diagonal loop and the 3′-overhang residues. Therefore, the TAT base triad shown in Scheme 1 seems the best structural solution, and was used for calculations of structures I and II. We note that partial interconversion between structures comprising TAT base triads shown in Schemes 1 and 3 in solution could explain the unequal intensities of A22 amino proton signals in 2D ^{15}N-filtered HSQC spectra. In order to further investigate this, we have performed additional calculations without H-bonding restraints for the diagonal loop and 3′-overhang using structure I, *i.e.*, starting with G-quadruplex comprising residues T13, A22 and T33 as in Scheme 1 and the A14-T21 base pair. The family of 10 structures with the lowest energies comprises 8 G-quadruplexes adopting structure I and II comprising the T13-A22·T33 base triad and A14·T34 base pair formed *via* T34H3-A14N3 interaction that leaves amino protons of A14 unbound. In an extended family of 20 structures with the lowest energies from the above-described calculations, we detected only one structure exhibiting interaction of T13, A22 and T33 residues as in Scheme 2 and one structure as in Scheme 3, both without A14-T21 base pairing (Figure 7).

Apart from the TAT base triad and the AT base pair described above, a T15·T20 H-bonded base pair is formed below the A14-T21 base pair in the diagonal loops of structures I and II. Depending on the position of the T15 pyrimidine, either T15O2 or T15O4 is H-bonded to T20H3, leaving T15H3 either H-bonded to T20O4 or free (Figure 7b). If two conformations arising from two T15·T20 H-bonding modes are compared (Figure 7b, blue rectangles), positions of the T15 residue and T20 pyrimidine differ. Shift of the T15 residue has an influence on backbone conformation of the A14-T19 part. Because of the lack of strong interactions, A16-T19 region of the diagonal loop is the most flexible part of the structure (Figure 4) and base stacking determines its final conformation. Base stacking of T15, A16 and A17 is conserved for both T15·T20 H-bonding interactions.

The structural dynamics of the diagonal loop and exposure of individual residues to solvent agree with the absence of thymine imino signals assigned to AT- or TT-base paired residues below the bottom quartet, even at lower temperature and pH values (down to 273 K and pH 5.5).

Lateral loop. Taking NMR data into account, the H-bonded A9-G26·A29 base triad is calculated above the top quartet (Figure 8a). The purine of A29 is stacked above G30 (Figure 5e) and forms a sheared G26·A29 base pair, according to the observed NOE data. G26 and A29 residues are not coplanar (Figure 8b), as previously described for inverted 3-nt or 4-nt GN(M)A loops.[29–32] T28 base is flipped away from the G-quadruplex core (Figure 8b), in agreement with the absence of its sequential NOE connectivities with T27 and A29. In structure I (Figure 4a), T27 is H-bonded to A2 and has its sugar moiety in the C3′-*endo* conformation, which agrees with the strong intra-residual NOE cross-peak T27H3′-H6. The A2·T27 base pair is stacked above A9-G26 bases (Figure 5f). In structure II, the 5′-overhang comes closer to the lateral loop than in structure I. The position of the A2 base on top of the A9 base is more flexible in structure II than in structure I which

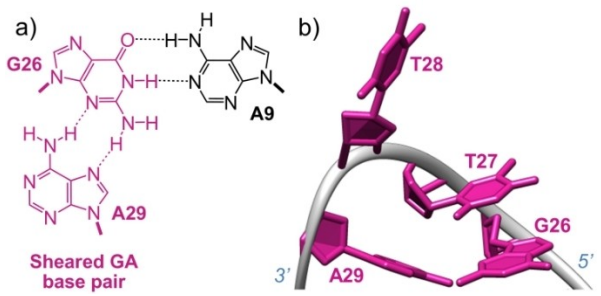

Figure 8. Structure of the lateral loop. a) H-bond in the base triad A9-G26-A29 above the top G-quartet. b) Inverted GTTA lateral loop (violet) in structure I in which nucleobase of T27 is stacked on nucleobase of G26. Backbone is grey. The hydrogen atoms are omitted for clarity.

agrees with the low number of inter-residual NOE interactions involving A2 protons and absence of H-bonding with T27.

NOE cross-peaks with water. Apart from imino signals of guanines constituting the middle quartet, $^{1}H/^{2}H$ exchange experiments reveal long lived imino signals for guanines involved in the bottom G-quartet (Supplementary Figure S14), indicating that these protons are protected from exchange with bulk water due to stacking with the neighboring base triad. Imino protons of G30 and G26 exhibit cross-peaks, albeit of very weak intensities, with water in 2D NOESY spectrum (τ_m 150 ms) at 308 K indicating that water molecules might be localized between the top G-quartet and the A9-G26·A29 base triad. In addition, two strong broad cross-peaks of water protons with protons at ca. δ 6.7 ppm and δ 6.5 ppm, corresponding to the assigned adenine A2, A9, A14, A22 and A29 amino proton chemical shifts and two weak cross-peaks between water and imino protons of G23 and G32 from the bottom G-quartet are observed (Supplementary Figure S14). We note that chemical shifts of the assigned amino protons from adenine residues involved in the base triads and base pairs overlap, hindering assignment of the NOE contacts with water.

Discussion

Instead of the parallel propeller-type G4 structure previously suggested for this sequence,[11] UpsB-Q-1 adopts a unique [3 + 1]-hybrid G-quadruplex structure with propeller, diagonal and lateral loops. This topology, first reported in 2009 on an artificial design,[33] and orientations of the three loops are reminiscent of some previously described G4 structures.[34–37] The UpsB-Q-1 G4 is nevertheless unique, as it comprises longer loops as well as 5′- and 3′-overhangs that contribute to the stabilization of its structure. Most G4 structures with long loops reported so far have a parallel topology with a long loop in the middle,[17,19,20,38–40] or [3 + 1]-hybrid topologies with the long loop folded into a hairpin stabilized by several canonical base pairs.[24–26]

In addition, UpsB-Q-1 displays unique interactions between the G4 core, the loops and the flanking regions. These structural elements are governing G4 stability and other properties, including potential drug targeting.[1,2] Moreover, they might be involved in the folding of the G4.[41] The G4 core, with its three stacked G-quartets, is not the only or most important contributor of overall stability. All three loops are stabilized by H-bonding and base stacking at the loop-quadruplex interface and within the loop regions. The loop sequences determine their ability to form stable structural motifs and thus modulate the relative stability of the accessible topologies. Most excitingly, we observed an equilibrium between two structures (named I and II). These structures differ only in lateral loop and 5′-overhang. The G**T**TA lateral loop includes a sheared GA base pair that is part of the AGA base triad (Figure 8), and can also accommodate a change in the sugar conformation of a thymine residue (marked **bold**) while retaining its own unique inverted structure. This conformational change of one-residue influences the 5′-overhang position and does not affect other parts of the G4, which are adopted by UpsB-Q-1. Such localized structural diversity induced by only one residue reveals that including overhang residues can be essential for the complete structural characterization of G4s.

In many G4s formed within promoter regions, the first loop is a 1-nt propeller loop.[1,18–20,26,27] UpsB-Q-1's first propeller loop is 4-nt long allowing for three major interactions. T7 base stacks with A8 and interacts with the middle quartet (Figure 6). A9 participates in A9-G26·A29 base triad above the top quartet (Figure 8). The AT-rich diagonal loop demonstrates that long loops often found in G4s occurring in promoter regions of biologically-relevant systems,[18–20] even without significant base complementarity, can nevertheless be structured themselves or adjacent to the core of G4.

The long second loop of UpsB-Q-1 forms an unusual H-bonding motif with the 3′-overhang, which contributes to the stability of the structure. Yet, the loop remains considerably flexible, allowing interconversions among several interaction schemes. Along with the ability to accommodate various interactions within the loop, these features could be beneficial as recognition elements in rational design of ligands. Our results suggest that various local energy minima are separated with low barriers of interconversion thus allowing for structural accommodation in interactions with binding partners. The disordered part of the loop can also bring entropic stabilization to the structural ensemble. These observations along with the unique lateral loop structural features are indeed thought provoking. It is possible that such conformational flexibility, which is unprecedented in G4s, is attributed to the potential of loop residues to associate with neighboring H-bonding partners in various geometries.

The unique features of UpsB-Q-1 structure may stimulate the community to consider G4s from an alternative perspective, focusing on the loops rather than on the G-quartet core alone. Perusal of UpsB-Q-1 structure also shows that the 5′- and 3′-flanking regions are fully involved in defining and stabilization of the overall fold through their engagement in specific interactions with the remainder of the molecule. The unique interactions in the loops of the UpsB-Q-1 G-quadruplex are exquisitely sensitive to the loop sequence. Understanding the structuring of loop regions in relation to their primary sequence

is of importance for further improving of G-quadruplex prediction tools.

Ten additional residues (here marked **bold** in the UpsB-Q-1 sequence C**A***GGG*TTA**A***GGG***TA**TAACTT**TA***GGG***GTTA***GGG***TT**) aside from the four *GGG*-tracts (marked *italic*), are essential for the well-defined G4. Remarkably, all ten significant residues identified for the UpsB-Q-1 are conserved in two other G-rich sequences UpsB-Q-3 and UpsB-Q-4 also found in promoters of B *var* genes in *P. falciparum*.[8]

Conclusions

A high-resolution solution structure of the [3 + 1]-hybrid G-quadruplex adopted by the UpsB-Q-1, a 34-nt oligonucleotide, was determined. All three loop types, 4-nt propeller, 10-nt diagonal and 4-nt lateral loop, were found to exhibit H-bonding and base stacking interactions at the G-core/loop interface and within the loop. The structure of the 10-nt long diagonal loop is ordered, although its sequence lacks complementarity that would allow stabilization by Watson-Crick base pairing. Furthermore, two base triads located above and below the G-quartets, as well as the structural diversity introduced by the single-residue conformational change (*i.e.*, the T27 sugar pucker), show that loops can have unexpected interactions and lead to unique conformations. Therefore, G4 loops should not only be considered as flexible links of G-tracts. Their specific sequence and structure show the highest potential for specific G4 recognition.

Our detailed structural study lays the ground for understanding the role of G4s in *var* gene regulatory factors and pathways relevant in malaria pathogenesis. G-rich sequences in promoter regions are potential regulators of gene expression. The new G4 structure is therefore an attractive way to interfere with replication, recombination and diversification of *var* genes, which are important for *P. falciparum* virulence and maintenance of chronic infections.[11,14] Moreover, putative *P. falciparum* RecQ helicase(s) might play an important role in G4 processing[6,11] especially if they could be related to previously reported bacterial RecQ helicase which binds to a sheared GTAA loop within a duplex environment.[42]

UpsB-Q-1 and its variants could be of great importance as a biological or pharmaceutical target within the *P. falciparum* genome. Loop and groove regions are increasingly recognized as the most promising binding sites to achieve ligand specificity among different G-quadruplex topologies.[1,2] However, so far loops have not been used for drug design, possibly because of the lack of structural information at atomic level of resolution. Most well-resolved structures contain short loops that rarely display specific structural features. The structure of the UpsB-Q-1 G-quadruplex is unique because it has long loops with well-defined, unique structures. Our study also prompts revising the typical search parameters for putative G4-forming sequences. Many queries are still limited to only 7-nt long loop regions, and the effect of the loop sequence is thus ignored. Developing better prediction tools will require further documenting the relationship between loop sequence, topology, structure and stability. Broadening of the typical G4 definition is particularly important because long(er) loops could be flexible enough to form various types of interactions within the G4 structure, as well as with other molecules in solution, making them an attractive target site for drug design.

Experimental Section

Synthesis. Oligonucleotides with natural-abundant or partially (8%) ^{15}N,^{13}C site-specifically labeled residues were synthesized on a K&A Laborgeraete GbR DNA/RNA Synthesizer H-8 using standard phosphoramidite solid-phase chemistry. Cleavage of the protecting groups was carried out in concentrated aqueous ammonia at 328 K overnight. Samples were purified by extensive dialysis against water and concentrated using 3000 Da cut-off ultrafiltration membrane (regenerated cellulose, Millipore).

Sample preparation. Folding of the oligonucleotides was done in a K-phosphate buffer (pH 7.0). After adding KCl solution to achieve final 150 mM K^+ ion concentration, ca. 2 mM oligonucleotide solution was heated to 363 K and cooled over 16 hours to 298 K using linear gradient. The obtained sample was diluted to about 1.2 mM concentration of the oligonucleotide, K^+ ion concentration was adjusted to 150 mM, and another cycle of 16 hour annealing from 363 to 298 K was conducted. The sample in 100% ^2H_2O was prepared by two consecutive 12 h lyophilisations using 1 mL of 100% ^2H_2O and final dissolution in 100% ^2H_2O.

UV spectroscopy. The concentration of the samples was determined on a Varian CARY-100 BIO UV-vis spectrophotometer using absorption at 260 nm in 1.0 cm cuvettes. Extinction coefficient for UpsB-Q-1 sequence was determined by the nearest-neighbor method.

NMR spectroscopy. NMR spectra were recorded on a Varian V NMRS 600 MHz and 800 MHz spectrometers at 273–313 K. Chemical shifts are referenced to the DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). DPFGSE pulse sequence was used to suppress the water signal. Spectra were processed and analyzed using V NMRJ program (Agilent Technologies). Cross-peak assignment and integration were done in the SPARKY software (UCSF).[43]

Resonance assignment. Identification of aromatic, imino and amino protons in partially (8%) ^{15}N,^{13}C residue-specifically labelled samples was performed by 1D and 2D ^{15}N-filtered and ^{13}C-filtered HSQC experiments. H1-H8 intra-residual connectivities through the C5 atom were established with the use of a 2D ^1H-^{13}C jrHMBC experiment (Figure 2b). Standard 2D NMR experiments recorded in 100% ^2H_2O including 2D ^1H-^1H DQF-COSY, ^1H-^{31}P HP-COSY, ^1H-^1H TOCSY (τ_m 80 ms) and ^1H-^1H NOESY (τ_m 80, 150 and 250 ms) were used to assign non-exchangeable protons. Resonances of exchangeable protons were assigned using 2D ^1H-^1H NOESY (τ_m 80, 150 and 250 ms) recorded in 10% (*v/v*) $^2H_2O/H_2O$. Analysis of 2D ^1H-^{13}C jrHMBC data and ^1H-^1H NOESY sequential connectivities extended the H6/H8 assignment to other residues. Additional support for assignments of T6, T7, T21, T27, T28 and T34, for which partial breaking of the NOESY walk or overlap obstructed the definitive assignment, was obtained by residue-specific ^{13}C-labelling (8%) or residue-specific mutation of thymine to an uracil residue. Sugar and some H5′/H5″ protons were assigned using the combination of 2D ^1H-^1H NOESY, ^1H-^1H TOCSY, ^1H-^1H DQF-COSY and ^1H-^{31}P HP-COSY data.

NOE (Nuclear Overhauser Effect) distance restraints for non-exchangeable protons were obtained from a 2D ^1H-^1H NOESY spectrum recorded at 308 K in 100% ^2H_2O (τ_m 150 ms). The volume of H5-H6 cross-peak of C1 residue was used as the distance

reference of 2.45 Å. Cross-peaks were classified as strong (1.8–3.6 Å), medium (2.6–5.0 Å) and weak (3.5–6.5 Å). After such NOE classification, class distance ranges were adjusted for thymine methyl groups to strong (2.3–4.4 Å), medium (3.1–5.5 Å) and weak (4.0–7.0 Å), to account for restraining of the distances to methyl carbon C7 instead of methyl proton. Classified NOE distance restraints for exchangeable protons were obtained from 2D ^{1}H-^{1}H NOESY spectra recorded at 308 K in 10% (*v/v*) $^{2}H_2O/H_2O$ (τ_m 80, 150 and 250 ms). Cross-peaks of medium and weak intensity in 2D ^{1}H-^{1}H NOESY spectrum (τ_m 80 ms) were classified as strong (1.8–3.6 Å) and medium (2.6–5.0 Å), respectively. Cross-peaks that appeared in 2D ^{1}H-^{1}H NOESY spectrum (τ_m 250 ms) were classified as weak (3.5–6.5 Å).

Structure calculations. Structure calculations were performed using AMBER14 software[44] and experimental restraints. A total of 1000 structures were calculated in 200 ps of NMR-restrained simulated annealing (SA) using the generalized Born implicit model to account for solvent effects. The cut-off for non-bonded interactions was 20 Å and the SHAKE algorithm for hydrogen atoms was used with the tolerance of 0.0004 Å. For each SA simulation, a different starting velocity was used. The temperature program was as follows: the temperature was raised from 300 K to 1000 K in 5 ps and held constant at 1000 K for 95 ps. The temperature was scaled down first to 500 K in the next 60 ps and then to 100 K in the next 20 ps and finally reduced to 0 K in the last 20 ps. Restraints used in the calculation were G-quartet and base pairing hydrogen bond (force constant 200 and 20 kcal mol^{-1}Å^{-2}, respectively) and NOE-derived distance restraints (force constant 60 kcal mol^{-1}Å^{-2}), torsion angle ε, χ and ν2 restraints (force constant 200 kcal mol^{-1}rad^{-2}) and planarity restraints for G-quartets (force constant 200 kcal mol^{-1} rad^{-2}). All residues except A14, A16, A17, C18 and T27 were restrained to C2′-*endo* sugar conformation. For the structure I with A2·T27 H-bonding T27 sugar was restrained to C3′-*endo* conformation, whereas for the structure II was left unrestrained. All structures were minimized with a maximum of 10 000 steps of energy minimization and a family of 10 structures was selected based on the smallest restraints' violations and lowest energy. Twist angles for two stacked G-quartets were calculated according to the literature.[45] Average of four values for each twist angle is reported. Visualization and structural analysis were performed in UCSF Chimera Software.[46]

Supporting Information

Supporting Information includes plots of NMR data (1D and 2D heteronuclear spectra) and tables with structural analysis.

Author Contributions

The manuscript was written through contributions of all authors.

Acknowledgements

This work was supported by the Slovenian Research and Innovation Agency [ARIS, grant P1-0242]. The authors acknowledge the CERIC-ERIC consortium for the access to experimental facilities and financial support. Financial support by Wallonia-Brussels region in Belgium and Slovenian Research Agency (grant no. BI-BE/11-12-V-003) under Bilateral Joint Research Program is gratefully acknowledged. Jaka Zavasnik is kindly acknowledged for performing some initial experiments.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

Structural data are available under PDB IDs: 5MTA and 5MTG, and BMRB IDs 34083 and 34084. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Keywords: G-quadruplex · nuclear magnetic resonance · long loop · loop interactions · *Plasmodium falciparum*

[1] a) A. I. Karsisiotis, C. O'Kane, M. W. da Silva, *Methods* **2013**, *64*, 28–35; b) P. Šket, J. Plavec, in *Biological Relevance & Therapeutic Applications of DNA- & RNA-Quadruplexes*, Future Science Ltd, **2015**, pp. 22–36; c) S. Takahashi, A. Kotar, H. Tateishi-Karimata, S. Bhowmik, Z.-F. Wang, T.-C. Chang, S. Sato, S. Takenaka, J. Plavec, N. Sugimoto, *J. Am. Chem. Soc.* **2021**, *143*, 16458–16469; d) A. Ghosh, M. Trajkovski, M.-P. Teulade-Fichou, V. Gabelica, J. Plavec, *Angew. Chem. Int. Ed.* **2022**, *61*, e202207384.
[2] D. Rhodes, H. J. Lipps, *Nucleic Acids Res.* **2015**, *43*, 8627–8637.
[3] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
[4] M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770–780.
[5] G. W. Collie, G. N. Parkinson, *Chem. Soc. Rev.* **2011**, *40*, 5867–5892.
[6] L. M. Harris, C. J. Merrick, *PLoS Pathog.* **2015**, *11*, e1004562.
[7] *World Malaria Report*, World Health Organization, ISBN 978-92-4-006489-8, **2022**.
[8] N. Smargiasso, V. Gabelica, C. Damblon, F. Rosu, E. De Pauw, M.-P. Teulade-Fichou, J. A. Rowe, A. Claessens, *BMC Genomics* **2009**, *10*, 362–373.
[9] G. A. Josling, M. Llinás, *Nat. Rev. Microbiol.* **2015**, *13*, 573–587.
[10] A. Scherf, J. J. Lopez-Rubio, L. Riviere, *Annu. Rev. Microbiol.* **2008**, *62*, 445–470.
[11] A. Stanton, L. M. Harris, G. Graham, C. J. Merrick, *BMC Genomics* **2016**, *17*, 859–874.
[12] H. L. Gage, C. J. Merrick, *BMC Genomics* **2020**, *21*, 236.
[13] L. M. Harris, K. R. Monsell, F. Noulin, M. T. Famodimu, N. Smargiasso, C. Damblon, P. Horrocks, C. J. Merrick, *Antimicrob. Agents Chemother.* **2018**, *62*, e01828–17.
[14] F. Dumetz, E. Y.-C. Chow, L. M. Harris, S. W. Liew, A. Jensen, M. I. Umar, B. Chung, T. F. Chan, C. J. Merrick, C. K. Kwok, *Nucleic Acids Res.* **2021**, *49*, 12486–12501.
[15] J. A. Smestad, L. J. Maher, *BMC Med. Genetics* **2015**, *16*, 91–104.
[16] D. J. E. Yue, K. W. Lim, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 11462–11465.
[17] a) S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J.-L. Mergny, A. T. Phan, *J. Am. Chem. Soc.* **2012**, *134*, 5807–5816; b) Y. M. Vianney, K. Weisz, *Nucleic Acids Res.* **2022**, *50*, 11948–11964.
[18] J. Marquevielle, C. Robert, O. Lagrabette, M. Wahid, A. Bourdoncle, L. E. Xodo, J.-L. Mergny, G. F. Salgado, *Nucleic Acids Res.* **2020**, *48*, 9336–9345.
[19] P. Agrawal, C. Lin, R. I. Mathad, M. Carver, D. Yang, *J. Am. Chem. Soc.* **2014**, *136*, 1750–1753.
[20] B. Onel, M. Carver, G. Wu, D. Timonina, S. Kalarn, M. Larriva, D. Yang, *J. Am. Chem. Soc.* **2016**, *138*, 2563–2570.
[21] Y. Qin, L. H. Hurley, *Biochimie* **2008**, *90*, 1149–1171.
[22] N. Kumar, S. Maiti, *Nucleic Acids Res.* **2008**, *36*, 5610–5622.

[23] a) K. W. Lim, P. Jenjaroenpun, Z. J. Low, Z. J. Khong, Y. S. Ng, V. A. Kuznetsov, A. T. Phan, *Nucleic Acids Res.* **2015**, *43*, 5630–5646; b) C. Papp, V. T. Mukundan, P. Jenjaroenpun, F. R. Winnerdy, G. S. Ow, A. T. Phan, V. A. Kuznetsov, *Nucleic Acids Res.* **2023**, *51*, 4148–4177.
[24] S. L. Palumbo, S. W. Ebbinghaus, L. H. Hurley, *J. Am. Chem. Soc.* **2009**, *131*, 10878–10891.
[25] E. Butovskaya, B. Heidi, B. Bakalar, S. N. Richter, A. T. Phan, *J. Am. Chem. Soc.* **2018**, *140*, 13654–13662.
[26] D. J. Y. Tan, F. R. Winnerdy, K. W. Lim, A. T. Phan, *Nucleic Acids Res.* **2020**, *48*, 11162–11171.
[27] Y. M. Vianney, P. Preckwinkel, S. Mohr, K. Weisz, *Chem. Eur. J.* **2020**, *26*, 16910–16922.
[28] S. L. Lam, *Nucleic Acids Res.* **2007**, *35*, W713–W717.
[29] S.-H. Chou, L. Zhu, B. R. Reid, *J. Mol. Biol.* **1997**, *267*, 1055–1067.
[30] M. J. van Dongen, M. M. Mooren, E. F. Willems, G. A. van der Marel, J. H. van Boom, S. S. Wijmenga, C. W. Hilbers, *Nucleic Acids Res.* **1997**, *25*, 1537–1547.
[31] G. D. Balkwill, T. P. Garner, M. S. Searle, *Mol. BioSyst.* **2009**, *5*, 542–547.
[32] G. D. Balkwill, T. P. Garner, H. E. L. Williams, M. S. Searle, *J. Mol. Biol.* **2009**, *385*, 1600–1615.
[33] M. Webba da Silva, M. Trajkovski, Y. Sannohe, N. M. Hessari, H. Sugiyama, J. Plavec, *Angew. Chem. Int. Ed.* **2009**, *48*, 9167–9170.
[34] M. Marušič, P. Šket, L. Bauer, V. Viglasky, J. Plavec, *Nucleic Acids Res.* **2012**, *40*, 6946–6956.
[35] S. Lago, M. Nadai, E. Ruggiero, M. Tassinari, M. Marušič, B. Tosoni, I. Frasson, F. M. Cernilogar, V. Pirota, F. Doria, J. Plavec, G. Schotta, S. N. Richter, *Nucleic Acids Res.* **2021**, *49*, 847–863.
[36] M. L. Živković, M. Gajarský, K. Beková, P. Stadlbauer, L. Vicherek, M. Petrová, R. Fiala, I. Rosenberg, J. Šponer, J. Plavec, L. Trantírek, *Nucleic Acids Res.* **2021**, *49*, 2317–2332.
[37] a) J. Dickerhoff, K. Weisz, *J. Phys. Chem. Lett.* **2017**, *8*, 5148–5152; b) J. Dickerhoff, L. Haase, W. Langel, K. Weisz, *ACS Chem. Biol.* **2017**, *12*, 1308–1315; c) Y. M. Vianney, N. Schröder, J. Jana, G. Chojetzki, K. Weisz, *J. Am. Chem. Soc.* **2023**, *145*, 22194–22205.
[38] K. W. Lim, T. Q. N. Nguyen, A. T. Phan, *J. Am. Chem. Soc.* **2014**, *136*, 17969–17973.
[39] K. W. Lim, P. Jenjaroenpun, Z. J. Low, Z. J. Khong, Y. S. Ng, V. A. Kuznetsov, A. T. Phan, *Nucleic Acids Res.* **2015**, *43*, 5630–5646.
[40] K. W. Lim, A. T. Phan, *Angew. Chem. Int. Ed.* **2013**, *52*, 8566–8569.
[41] J. Jana, Y. M. Vianney, N. Schröder, K. Weisz, *Nucleic Acids Res.* **2022**, *50*, 7161–7175.
[42] K. A. Manthei, M. C. Hill, J. E. Burke, S. E. Butcher, J. L. Keck, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4292–4297.
[43] T. D. Goddard, D. G. Kneller, Sparky 3; University of California: San Francisco, CA, USA, **2004**.
[44] A. Perez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham 3rd, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817–3829.
[45] C. J. Lech, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 2034–2046.
[46] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
[47] S. Čeru, P. Šket, I. Prislan, J. Lah, J. Plavec, *Angew. Chem. Int. Ed.* **2014**, *53*, 4881–4884.
[48] S. Takahashi, A. Kotar, H. Tateishi-Karimata, S. Bhowmik, Z.-F. Wang, T.-C. Chang, S. Sato, S. Takenaka, J. Plavec, N. Sugimoto, *J. Am. Chem. Soc.* **2021**, *143*, 16458–16469.
[49] T. Frelih, B. Wang, J. Plavec, P. Šket, *Nucleic Acids Res.* **2020**, *48*, 2189–2197.
[50] A. Kotar, R. Rigo, C. Sissi, J. Plavec, *Nucleic Acids Res.* **2019**, *47*, 2641–2653.
[51] S. Hadži, V. Kocman, D. Oblak, J. Plavec, J. Lah, *Angew. Chem. Int. Ed.* **2019**, *58*, 2387–2391.
[52] M. Lenarčič Živković, J. Rozman, J. Plavec, *Angew. Chem. Int. Ed.* **2018**, *57*, 15395–15399.

Manuscript received: March 25, 2024
Accepted manuscript online: April 22, 2024
Version of record online: June 3, 2024