

Structural and thermodynamic studies of G-quadruplexes with unique
structural motifs and their recognition by small molecules

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades

Dr. rer. nat. (Doctor rerum naturalium)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Universität Greifswald

vorgelegt von:

Lie, Yoanes Maria Vianney

Greifswald, Mai 2023

Dekan: Prof. Dr. Gerald Kerth

1. Gutachter: Prof. Dr. Klaus Weisz

2. Gutachter: Prof. Carlos González

3. Gutachter: Assoc. Prof. Lukas Trantirek, Ph.D.

Tag der Promotion: 6 September 2023

Contents

Abbreviations	1
1 Scope of the dissertation	2
2 G-quadruplexes	6
2.1 Structural aspects of G-quadruplexes.....	6
2.2 Non-canonical G-quadruplexes (Article 1)	8
3 Understanding the preferred end-folding of G-quadruplexes through mutational studies on loop residues	12
3.1 G-quadruplex structures derived from naturally occurring sequences in genomes (Articles 2 & 3).....	12
3.2 Rational design of sequences to reveal new G-quadruplex topologies with the help of G analogs (Article 4).....	14
3.3 Towards the explanation of the impact of loop length and composition on the G-quadruplex topologies (Article 5)	18
4 Quadruplex-duplex hybrids	23
4.1 Stacking interactions at the quadruplex-duplex interface (Article 6)	24
5 Ligand binding to quadruplex-duplex hybrids	27
5.1 Quadruplex-duplex junctions as binding hotspots for an indoloquinoline ligand (Article 6).....	27
5.2 Thermodynamic and structural studies on quadruplex-duplex binders (Articles 7 & 8)	29
5.3 Indoloquinolines intercalate at quadruplex-duplex junctions (Article 7)	32
5.4 Factors affecting the selectivity of indoloquinolines upon binding the quadruplex-duplex junction (Article 8).....	34
5.5 Phen-DC3 prefers binding to a quadruplex-duplex junction over an open G-tetrad (Article 8).....	36
5.6 General insights into ligand binding at quadruplex-duplex junctions (Articles 7 & 8)	38
6 Conclusions and perspectives	42
Bibliography	44
Author contributions	51
Article 1	54

Article 2	71
Article 3	87
Article 4	114
Article 5	142
Article 6	202
Article 7	240
Article 8	279
Affirmation	335
Curriculum Vitae	336
Acknowledgments	339

Abbreviations

A	adenine/adenosine
BrG	8-bromoguanosine
C	cytosine/cytidine
CD	circular dichroism
DSC	differential scanning calorimetry
DNA	deoxyribonucleic acid
G	guanine/guanosine
ITC	isothermal titration calorimetry
LNA	locked nucleic acid
NDI	naphthalene diimide
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
nt	nucleotide
PIQ	phenyl-indoloquinoline
QD	quadruplex-duplex
ROESY	rotating frame nuclear Overhauser effect spectroscopy
T	thymine/thymidine
TBA	thrombin binding aptamer
TO	thiazole orange
WSSV	White Spot Syndrome Virus

1. Scope of the dissertation

This dissertation explores and tries to unravel the fundamental basis of G-quadruplex end-folding as well as G-quadruplex interactions with small molecules by thermodynamic and structural approaches. Selective targeting of G-quadruplexes with ligands remains elusive, either because the ligand has considerable binding affinity for other DNA structures or because it fails to discriminate between different G-quadruplex topologies. Unique structural motifs on the G-quadruplex may enhance or inhibit ligand binding to the G-quadruplex. For such aspects, it is necessary to understand the effect of G-quadruplex motifs or elements on the end-folding in order to better tune certain G-quadruplex topologies as model systems. Quadruplex-duplex (QD) junctions as unique structural motifs are shown to be a binding hotspot for various G-quadruplex ligands containing an intercalator motif. Eight articles are to be discussed and summarized in this dissertation.

Article 1: Structural motifs and intramolecular interactions in non-canonical G-quadruplexes

Jana, J., Mohr, S., Vianney, Y. M., & Weisz, K. (2021). *RSC Chemical Biology*, 2 338-353.

This review article discusses various non-canonical G-quadruplexes, particularly from the structural point of view. In many cases, canonical G-quadruplexes contain three layers of G-tetrads. This review focuses on G-quadruplexes containing only two layers of G-tetrads as well as G-quadruplexes containing broken G-columns, i.e., G-quadruplexes in which a G-column has a smaller number of consecutive guanines than the other G-columns. The vacancy created by such a G-column can subsequently be filled with guanosine-based nucleotide(s) via interspersed residues (bulge), or via intramolecular (from a snapback loop for instance), or intermolecular interactions. Aspects related to the thermodynamic stability and a possibility for specific G-quadruplex targeting by ligands following such unique G-quadruplex elements are also discussed.

Article 2: G-Quadruplex formation in a putative coding region of White Spot Syndrome Virus: structural and thermodynamic aspects

Vianney, Y. M., Purwanto, M. G. M., & Weisz, K. (2021). *ChemBioChem*, 22, 1932-1935.

Bioinformatic studies have suggested that the White Spot Syndrome Virus (WSSV) genome contains several putative G-quadruplex forming sequences. One conserved sequence named *WSSV131*, which yields a proton NMR spectrum of good quality, is subjected to structural and thermal stability analyses. There are two mixtures of two parallel G-quadruplexes with different loop length arrangements for the propeller loop. The major species is identified to contain a 1:3:1 loop length arrangement. The 1:3:1 loop isomer is found to be more stable than the 1:2:2 loop isomer, consistent with a classic argument that a

one-nucleotide (1-nt) propeller loop contributes to a more stable G-quadruplex than a longer propeller loop. The presence of multimeric G-quadruplexes was detected by gel electrophoresis with melting temperatures (above 100 °C) far exceeding those of the monomeric G-quadruplex.

Article 3: First tandem repeat of a potassium channel KCNN4 minisatellite folds into a V-loop G-quadruplex structure

Vianney, Y. M., & Weisz, K. (2021). *Biochemistry*, 60, 1337-1346.

The intron in-between the *KCNN4* gene contains minisatellites with several contiguous guanines. The first repeat was subjected to a structural study and found to fold into a conventional V-loop upon truncation of the 5'-flanking sequence. The 3'-overhang seems to give additional stabilizing interactions with the lateral loop preceding the V-loop. On the other hand, the addition of a 5'-overhang greatly diminishes the thermal stability and the quality of the NMR spectrum, which brings into the argument that the addition of an overhang in tandem with a G-vacancy filling strategy is detrimental.

Article 4: Expanding the topological landscape by a G-column flip of a parallel G-quadruplex

Mohr, S., Jana, J., Vianney, Y. M., & Weisz, K. (2021). *Chemistry—A European Journal*, 27, 10437-10447.

A novel hybrid-type G-quadruplex folding with loop progression of -(ppl) (propeller, propeller, and lateral loop together with a counterclockwise progression) was constructed and named hybrid-2R. The formation has to be forced by incorporating 8-bromoguanosines (BrG) at matching positions. In this case, the BrGs were incorporated in the G-column following the lateral loop being the only G-column in an antiparallel direction with respect to the other G-columns. A loop arrangement of 1:1:3 can be employed to create a hybrid-2R topology. On the other hand, the incorporation of a duplex-forming sequence in the lateral loop does not give significant aid to the folding. It is interesting to note that with a loop arrangement containing two 1-nt loops in this design, a (3+1) hybrid G-quadruplex with an exclusive all-homopolar tetrad stacking which usually escapes attention is obtained.

Article 5: Guiding the folding of G-quadruplexes through loop residue interactions

Jana, J., Vianney, Y. M., Schröder, N., & Weisz, K. (2022). *Nucleic Acids Research*, 50(12), 7161-7175.

This article tries to explain the fundamentals of G-quadruplex folding through the combination of structural and thermodynamic approaches. The hybrid-1R/hybrid-1R' topology with +(lpp) loop progression as well as parallel G-quadruplexes were used as model systems in this study. The addition of a lateral snapback loop restricts the possibility of various other topologies and tunes the folding to the desired competing structures. Using various residue mutations (deletion, substitution, addition, using G analogs), various trends were obtained. From the loop itself, the last (third) residue of the lateral loop tends to be ordered

and stacked above the G-tetrad. However, the most crucial one is the importance of the 1-nt loop. Originally, the major species obtained with a loop length arrangement of 3:3:1 was the hybrid-1R', a +(lpp) G-quadruplex with one homopolar and one heteropolar tetrad stacking. Such tetrad stacking patterns are the most commonly perceived stacking for the (3+1) hybrid G-quadruplex. Surprisingly, shortening the central loop into a 3:1:1 loop arrangement also increases the population of both the hybrid-1R topology as well as the parallel topology. Of note, the hybrid-1R topology is a +(lpp) G-quadruplex with all-homopolar tetrad stackings (compare with hybrid-2R in *article 4*). A comparison of free energies was also performed and shortening the loop was found to stabilize both the hybrid-1R and the parallel topology. Thus, the 1-nt loop does not prefer only the parallel G-quadruplex. More precisely, the 1-nt loop prefers to connect G-tetrads of the same polarity, i.e., it prefers to connect two guanosines with the same glycosidic torsion angle. Such progresses answer various questions that have been raised (in *article 4* for instance) and can be used for a better understanding of G-quadruplex folding and for better G-quadruplex engineering.

Article 6: Quadruplex-duplex junction: a high-affinity binding site for indoloquinoline ligands

Vianney, Y. M., Preckwinkel, P., Mohr, S., & Weisz, K. (2020). *Chemistry—A European Journal*, 26, 16910-16922.

The quadruplex-duplex (QD) hybrids have gained interest since their structural determination in several past studies. The QD junction can be created, for example, by incorporating a lateral loop. However, commonly used G-quadruplex binding ligands containing extended aromatic ring systems tend to favor the binding to the open, accessible G-tetrad. The indoloquinoline ligand (PIQ) used in this study was originally developed in this lab and appeared to preferentially target the parallel G-quadruplexes among other G-quadruplex topologies. For the first time, the quadruplex-duplex junction is shown to not represent a blockage, but rather recruits the ligand to bind in the QD junction. Thermodynamic studies with ITC as well as NMR studies were performed to explain this phenomenon.

Article 7: Indoloquinoline ligands favor intercalation at quadruplex-duplex Interfaces

Vianney, Y. M., & Weisz, K. (2022). *Chemistry—A European Journal*, 28, e202103718.

As a continuation of *Article 6*, a QD hybrid containing a G-vacancy which is filled with a terminal 3'-G from the extension of the 3'-overhang by a self-complementary duplex forming sequence is employed. Such a construct is expected to reduce the conformational flexibilities in the QD junction that hinder the structural assignment in *Article 6*. The structural study showed that the duplex 3'-overhang behaves like a lateral snapback loop. In this work, SYUIQ-5, a commercially available ligand, is used for the ligand binding study. SYUIQ-5 is also known to target parallel G-quadruplexes and has the potential for cancer

therapy. NMR studies showed that SYUIQ-5 binds to the QD junction by intercalation. The aliphatic side chain of SYUIQ-5 points towards the direction of the duplex minor groove. A further binding study with cryptolepine, a commercially available ligand with an indoloquinoline ring system without side chains, suggests that cryptolepine also binds to the junction by intercalation. Similar to the duplex intercalator, π - π stacking is the dominant force for the binding. Thus, QD intercalation by indoloquinolines shows higher affinity than stacking at the outer G-tetrad and binding at the duplex. In contrast to what has been proposed, the duplex lateral loop does not appear to block ligand binding, but rather promotes intercalation.

Article 8: High-affinity binding at quadruplex-duplex junctions: rather the rule than the exception

Vianney, Y. M., & Weisz, K. (2022). *Nucleic Acids Research*, 50, 11948-11964.

A similar geometrical arrangement of SYUIQ-5 upon QD intercalation when compared to G-tetrad end-stacking and duplex intercalation with similarly related indoloquinoline-based ligands raises a question about the degree of selectivity towards various DNA binding modes. Various ligands were tested for the affinity and selectivity of the binding at the QD junction. For SYUIQ-5, it was expected that the selectivity is only modest against the duplex compared to both G-quadruplex and the QD junction. In contrast, the PIQ used in *Article 6* exerts a high selectivity against the duplex. A structural study of the PIQ on a QD hybrid was carried out. It was found that upon intercalation at the QD junction, PIQ was found to be subjected to a 180° flip when compared to SYUIQ-5, i.e., the tilted phenyl side chain is directed towards the major groove of the duplex and subsequently covers the rest of the G-tetrad. A similar idea can be deduced from the binding of Phen-DC3 to the QD hybrid. Whereas the phenanthroline ring system of Phen-DC3 intercalates into the QD junction, the two quinoline side arms cover the rest of the G-tetrad. An argument for good QD binders is proposed: increase selectivity by maximizing the interactions with the G-tetrad by entropic contributions and discourage the binding to the duplex. Such an argument is in fact similar to some of the strategies used to design G-quadruplex ligands; however, it should be emphasized that the QD junction is the more affine binding site when compared to the open accessible G-tetrad.

2. G-quadruplex

2.1 Structural aspects of G-quadruplexes

G-quadruplexes are secondary structures of nucleic acids composed of G-tetrads. A G-tetrad is comprised of four guanines, arranged in a square planar pattern and connected via eight Hoogsteen hydrogen bonds (Figure 1A). Two polarities of the tetrad can be categorized according to the direction of the hydrogen bond donor to its acceptor. A G-quadruplex is typically composed of a run of guanines in one G-column/G-tract in which G-tetrads are stacked onto each other in a helical arrangement. Depending on the sugar orientation, the guanosine will have either an *anti*- or *syn*- glycosidic torsion angle to maintain the direction of the polarity within the G-tetrad. A parallel orientation is defined as two strands adjacent to each other displaying the same backbone orientation. The two parallel strands contain two guanines with the same glycosidic torsion angle (*anti/anti* or *syn/syn*). On the other hand, two strands in antiparallel orientation contain two guanines with different glycosidic torsion angles (*anti/syn*) in each of the strands. As a result, three different groove widths, i.e., the distance between two backbones of two adjacent strands, arise from the strand arrangements. A medium groove is obtained from two strands in a parallel orientation while either a wide or narrow groove arises from an antiparallel strand orientation depending on the direction of the hydrogen bond donor to the guanosine acceptor with the different glycosidic torsion angles (Figure 1B). Monocations, usually potassium, stabilize the G-quadruplex and are located in between two G-tetrads which are subsequently coordinated by eight O6 atoms of the guanines (Figure 1C) (1).

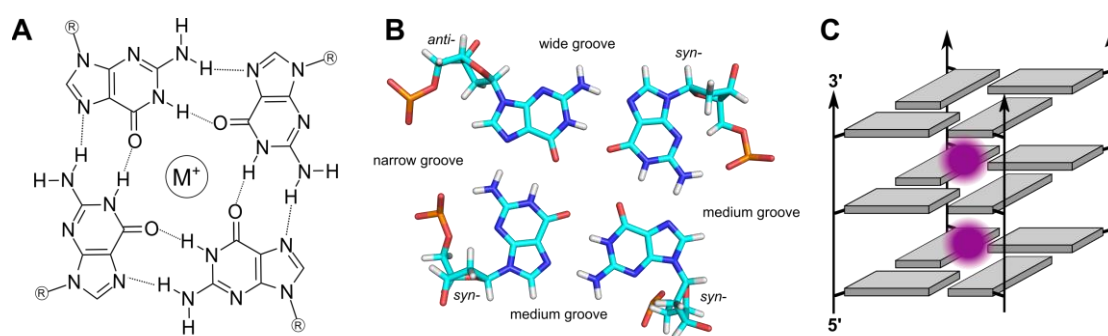


Figure 1. (A) One guanine tetrad arranged and connected by eight Hoogsteen hydrogen bonds with a cation in the center. (B) A guanine tetrad with *syn*→*syn*→*anti*→*syn*-guanosines containing wide, medium, and narrow groove sizes. Cartoon is taken from PDB 2JSM (2). (C) Schematic of a tetrameric parallel G-quadruplex with all *anti*-guanosines (as grey boxes) and alkali metal ions (as a purple sphere) sandwiched between two tetrads.

G-quadruplexes may have a monomolecular, bimolecular, or tetramolecular assembly. For monomolecular and bimolecular G-quadruplexes, loops are needed to connect two G-columns. Three types of loops arise from the arrangement of the strand orientation. A propeller loop connects two G-columns in parallel orientation. A lateral loop joins two adjacent G-columns in antiparallel orientation while the diagonal loop connects two distal antiparallel G-columns. For a right-handed DNA, the G-column helicity already geometrically restricts the direction of the propeller loop due to the twist angle in the range of 20-30° for each base step (Figure 2A) (3). On the other hand, a lateral loop could progress in either a counterclockwise or clockwise direction, subsequently bridging wide or narrow grooves, respectively (Figure 2B). There is an old consensus that short propeller loops (1-2 nucleotides) are favored, whereas a lateral loop is formed from at least two nucleotides in most cases. The diagonal loop is usually formed from more than three nucleotides. A very helpful nomenclature for the propeller and lateral loop directions (- for counterclockwise and + for clockwise) was used in this study, considering the strand orientation and following the starting point of the 5'-guanine in the bottom tetrad in the front-right column (4).

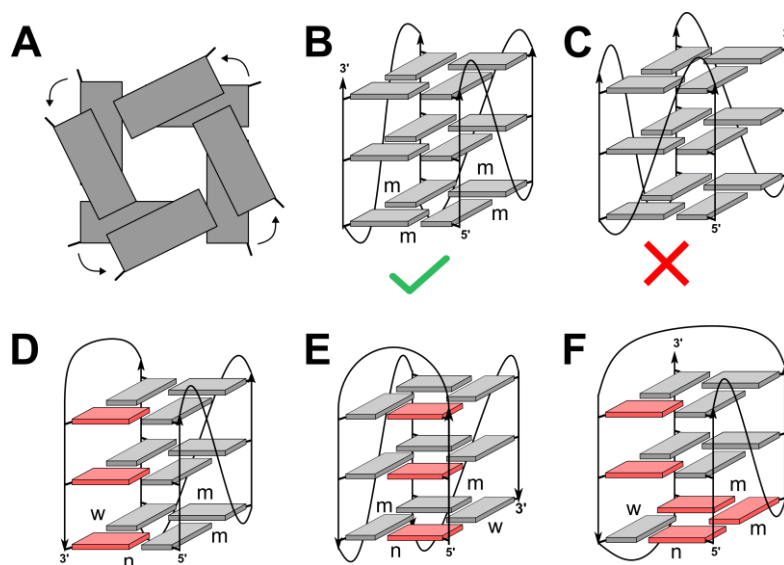


Figure 2. (A) Top view of the two stacked G-tetrads with right-handed helical progression. (B&C) Allowable -(ppp) and non-allowable +(ppp) progression of propeller loops in a right-handed G-quadruplex. (D) Schematic of a G-quadruplex structure containing two propeller loops and a lateral loop bridging a wide groove -(ppl) with all-homopolar tetrad stacking (PDB 7O1H) (5). (E) Schematic of a G-quadruplex structure containing two propeller loops and one lateral loop bridging a narrow groove +(lpp) with all-homopolar tetrad stacking (PDB 6R9K and PDB 7ZEO) (6, 7). (F) Schematic of a G-quadruplex structure (-pd+l) containing propeller, diagonal, and lateral loops bridging a wide groove (PDB 2LOD) (8). The groove width dimensions are denoted as follows: w (wide), m (medium), and n (narrow).

G-quadruplex structures became prominent due to their abundance in the genomic context (9). These structures are known to modulate replication as well as gene expression at both the transcriptional and translational levels. G-quadruplexes could block enzymatic machinery in various nucleic acid transactions or recruit proteins to up/downregulate gene expression (10). Putative G-quadruplex forming sequences are often found in the regulatory region of oncogenes, the best example being the microsatellites of telomeres (TTAGGG)_n. These facts led to the possibility of modulating G-quadruplex formation for cancer therapy (11). On the other hand, G-quadruplexes have found their way into various nanotechnological and analytical applications. A G-quadruplex was found to be able to behave as a DNAzyme system in conjunction with hemin to create a peroxidase-like reaction (12) which allows visualization for various detection purposes (13, 14).

Various small molecules or drugs recognize the G-quadruplex structures by end-stacking owing to the accessible binding site on the terminal G-tetrad. As such, the general design of G-quadruplex ligands involves a polycyclic ring with an extended aromatic ring system. Such designs also prevent intercalation into the duplex DNA due to the unfavorable entropic penalty of exposing large hydrophobic surfaces of the ligand to the solvent or the inability of the ligand to intercalate between the bases in a B-DNA. Additional interactions may come from weak electrostatic forces, hydrogen bonding, or further stacking/capping from the loop or overhang (15).

2.2. Non-canonical G-quadruplexes (Article 1)

The most commonly observed number of stacked G-tetrads is three. Two G-tetrad layers are deemed too unstable and tend to recruit additional stacking interactions, for instance, end-to-end stacking to create a dimeric G-quadruplex (16, 17) or a capping structure from loop residues (18). In fact, the four-layered structures of one left-handed G-quadruplex in one study can be seen as a pair of two-layered G-quadruplexes connected by a linker (19). On the other hand, intact four-layered G-quadruplexes are still rarely observed. It was first observed in antiparallel G-quadruplexes bearing three lateral loops in either clockwise or counterclockwise direction (20, 21). Recently, a propeller loop linking a four-layered G-quadruplex was observed involving non-native nucleotides as its residues. Apparently, the distance between the distal tetrads in the four-layered G-quadruplex is too long, exacerbated by the fact that the 3'-terminal guanine in one G-column already passes the position of the 5'-guanine in the following G-column due to its twist. A hydrophobic penalty in conjunction with steric clash perhaps hinders the formation of the propeller loop using natural nucleotides in a four-layered G-quadruplex (22).

On the other hand, G-quadruplexes can also be composed of interrupted G-tracts. The G-quadruplex is comprised of one or more columns in which the number of consecutive guanines is less

than for other G-columns. Such a design creates a vacancy in the G-quadruplex (23, 24). There are various ways to fill the vacancy. It can be filled from an interspersed G residue in which a bulge is then created (25), from a guanine within an overhang or loop (26, 27), or even from a guanine residue located in another G-quadruplex subsequently creating an interlocked dimeric molecule (Figure 3) (28).

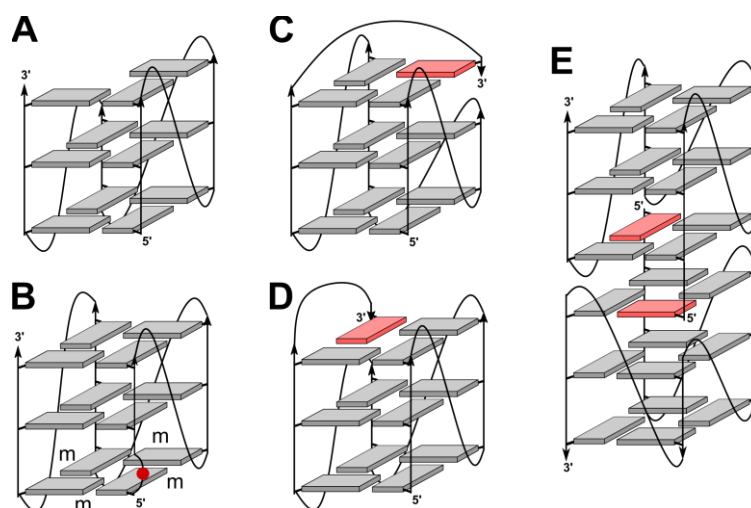


Figure 3. (A) A parallel G-quadruplex bearing one guanine vacancy on its third G-column. (B) A parallel G-quadruplex with a bulge (residue in the bulge is colored in red) on its first G-column. (C) A G-quadruplex with a broken G-tract on the second column subsequently filled with a guanine from the 3'-overhang forming a diagonal snapback loop. (D) A G-quadruplex with a broken G-tract on the third column and subsequently filled with a guanine from the 3'-overhang forming a lateral snapback loop. (E) An interlocked G-quadruplex model that is taken from an HIV integrase aptamer *93del* (PDB 1Y8D) (28).

One possible way to fill a G-vacancy via a loop is by creating a snapback loop. A snapback loop is often formed by the elongation of its overhang sequence, creating either a lateral or diagonal loop. Of note, a propeller snapback loop has never been observed. One best example is taken from a parallel G-quadruplex derived from a *Myc* sequence called *Pu24T* that contains one guanine vacancy on the second column. This vacancy is subsequently filled by a guanine residue from the extension of the 3'-overhang forming a diagonal snapback loop (Figure 3C; PDB 2A5P) (26). On the other hand, a *PDGFR-β*-inspired sequence has a vacancy on the third G-column and is filled by a guanine residue in a lateral-type snapback loop fashion (29). The vacancy is proposed to be filled by a guanosine in *syn*-conformation due to the antiparallel orientation of the backbone of this residue relative to the adjacent nucleotide (Figure 3D).

Other possible non-canonical G-quadruplexes that are peculiar but frequently observed are V-loop G-quadruplexes. These were first reported in 2001 by Patel's lab in a dimeric G-quadruplex (30).

The first observation of a three-layered V-loop G-quadruplex derived from a genomic-inspired sequence was observed in 2010 from the *chl1* intronic sequence (27). This was unprecedented as the *chl1* sequence context does contain four runs of three guanines that can form a canonical G-quadruplex structure. This suggests that the stability of the V-loop G-quadruplex can compete with the canonical G-quadruplexes.

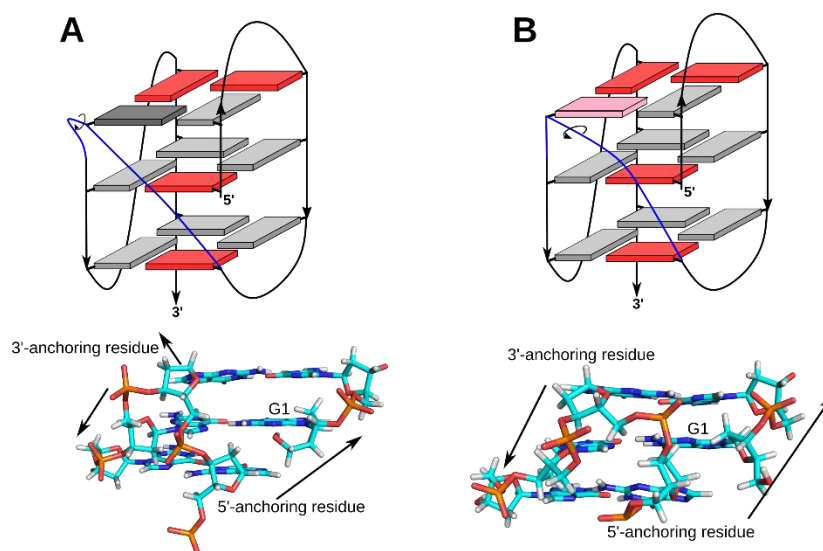


Figure 4. Schematic (top) and high-resolution structures (bottom) focusing on the two G-columns connected by a V-shaped loop of (A) a conventional V-loop G-quadruplex and (B) an alternative V-loop G-quadruplex. For the schematic, the *anti*-, *high-anti*-, *syn*-, and *low-syn*-residues are colored in grey, dark grey, red, and light red, respectively. Note that the arrows in the three-dimensional structures indicate the backbone polarity for residues in the G-columns. The representative three-dimensional structures for the conventional V-loop and alternative V-loop G-quadruplexes are taken from PDB 7ATZ and PDB 6H1K, respectively.

The common motif for the formation of a V-loop G-quadruplex is a vacancy in the first G-column before the 5'-terminus. The vacancy is then subsequently filled by a guanine, which is usually directly linked to the next G-column in the opposite G-tetrad layer. The V-shaped loop is formed from this guanine-filling residue and the 3'-guanine in the adjacent G-column. In contrast to the propeller loop, the orientation of the broken G-column relative to the connected G-column is antiparallel (Figure 4). The terms 5'-anchoring and 3'-anchoring residues are used here to describe the two guanines connected by this V-shaped loop. Two types of V-shaped loops were found after studying various solved structures either derived from naturally occurring sequences or by incorporating G analogs. Two seminal studies were carried out separately and described two kinds of V-shaped loops named conventional (or V_R) and

alternative (or V_s) (31–33). The critical difference is based on the location of the backbone inversion. For the conventional V-shaped loop, the backbone polarity inversion happens after the 3'-anchoring residue. Thereby, it is observed that the 3'-anchoring residue contains a different sugar orientation relative to the 3'-guanosines in the same G-column. On the other hand, the backbone polarity inversion occurs in the V-shaped loop for the alternative one, resulting in the 3'-anchoring residue having the same sugar orientation as the rest of the guanosines in the same G-column (Figure 4A&B). The difference in the backbone orientation for the 3'-anchoring residue results in a unique glycosidic torsion angle for the anchoring residues. Whereas the 5'-anchoring guanosine is typically in a *syn* conformation as can be derived for canonical G-quadruplexes following lateral loop progression preceding the V-loop, the glycosidic torsion angle of the 3'-anchoring residue is not well defined. Often, the 3'-anchoring residue glycosidic torsion angle is found to be in a *high-anti* region for a conventional V-loop G-quadruplex (27, 34, 35), while the glycosidic torsion angle for the 3'-anchoring residue in the alternative one is found in a *low-syn* region (36, 37).

Interesting seminal studies by Haase mentioned the importance of determining the sugar pucker and not only the glycosidic torsion angle for V-loop G-quadruplexes (32). It can be observed that the sugar pucker of both anchoring residues in the conventional V-loop and the alternative V-loop often tend to be in north and south, respectively. In this case, one might consider incorporating G analogs such as ribonucleotides or LNA (locked nucleic acid) that prefer (or lock) the sugar in a north conformation and subsequently favors *anti*-guanosine as the 3'-anchoring residue in a conventional V-loop. Interestingly, incorporation of the LNA in the 5'-anchoring *syn* position still allows formation for the conventional V-loop, suggesting that the sugar conformation can overcome the preference for a nucleoside glycosidic torsion angle when it comes to V-loop G-quadruplex formation (33).

The ability of the G-quadruplex to fold even in the presence of an interrupted G-tract greatly increases the diversity of G-quadruplex folding. However, such motifs can also be used for a specific ligand/drug binding. The G-vacancy can also be filled by small metabolites such as cGMP or oligonucleotides containing guanine motifs which might have importance *in vivo* (24, 38, 39). In addition, the snapback loop motifs, in particular the diagonal snapback loop, are known to block ligand binding to the G-quadruplexes. As mentioned above, G-quadruplex ligands are typically polycyclic compounds with extended aromatic ring systems and prefer to bind via end-stacking on the terminal tetrads. In the presence of the snapback loop, one tetrad is blocked by the loop, leaving the other tetrad as the only accessible site for an affine ligand binding (26, 40, 41). However, there is no study on ligands that specifically target V-shaped loop topologies to date.

3. Understanding the preferred end-folding of G-quadruplexes through mutational studies on loop residues

3.1. G-quadruplex structures derived from naturally occurring sequences in genomes (Articles 2 & 3).

One example of the diversity of G-quadruplex folding can be seen in a study of a gene segment from the White Spot Syndrome Virus (*WSSV*). This simple sequence of 23 residues, named *WSSV131* due to its position on the open reading frame *WSV131* (accession number KY827813) consists of $G_3NG_3N_2G_4NG_3$ with N being any arbitrary nucleotide (Figure 5). NMR spectra suggest a mixture of two species, which arise due to loop isomers from the four runs of guanines in the third segment. A simple $G \rightarrow T$ mutation for each of the guanines of the third G-column (Figure 5) resolves the mixture. Supported by NOESY spectra, the two mutated species are shown to be three-layered parallel G-quadruplexes containing all *anti*-guanosines in the tetrads with exclusive homopolar tetrad stacking. The major species was found to be the G-quadruplex with a loop arrangement of 1:3:1 while the minor species has a loop arrangement of 1:2:2 (Figure 5). The higher melting temperature as well as the slightly higher van't Hoff enthalpy obtained by differential scanning calorimetry (DSC) seem to support two classical arguments:

1. The 1-nt propeller loop is significantly more stable than a propeller loop with longer residues.
2. A favored three-layered parallel G-quadruplex contains two 1-nt loops (first and third) and a middle loop of varying length (42–44).

It is noteworthy that high-melting species are additionally observed in the DSC thermogram. Such species are suggested to be multimeric G-quadruplexes with a higher-order structure (Article 2).

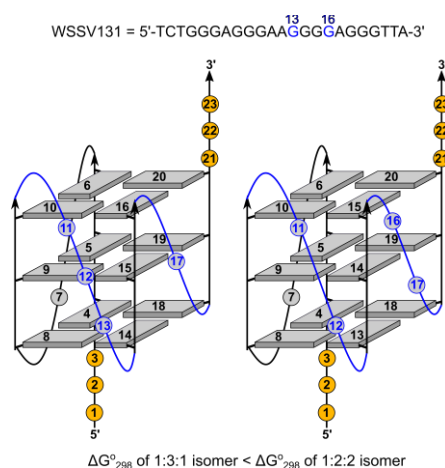


Figure 5. Preferential folding of a *WSSV131*-derived sequence into two parallel G-quadruplexes of loop isomers. Formation of a parallel G-quadruplex with a higher number of 1-nt loops is more favored.

In another case, the first intronic segment of *KCNN4* appears to provide a unique repeat sequence containing several runs of guanines. The first repeat is further studied in terms of G-quadruplex folding in Article 3. The sequence which is 5'-CCTA GG TCTGA GGG A GG A GGGG CT GGGGG TCA-3' forms a G-quadruplex with heterogeneous topologies, as suggested by the 1D proton NMR spectrum. Subsequent mutation and truncation of the flanking sequences resulted in a sequence 5'-GG TCTGA GGG AGA GGGG CT GGG T-3' with one major fold. Such a pattern, carrying only two guanines in the first G-run and four guanines in the third G-run, can potentially form a V-loop G-quadruplex as a way to fill the G-vacancy.

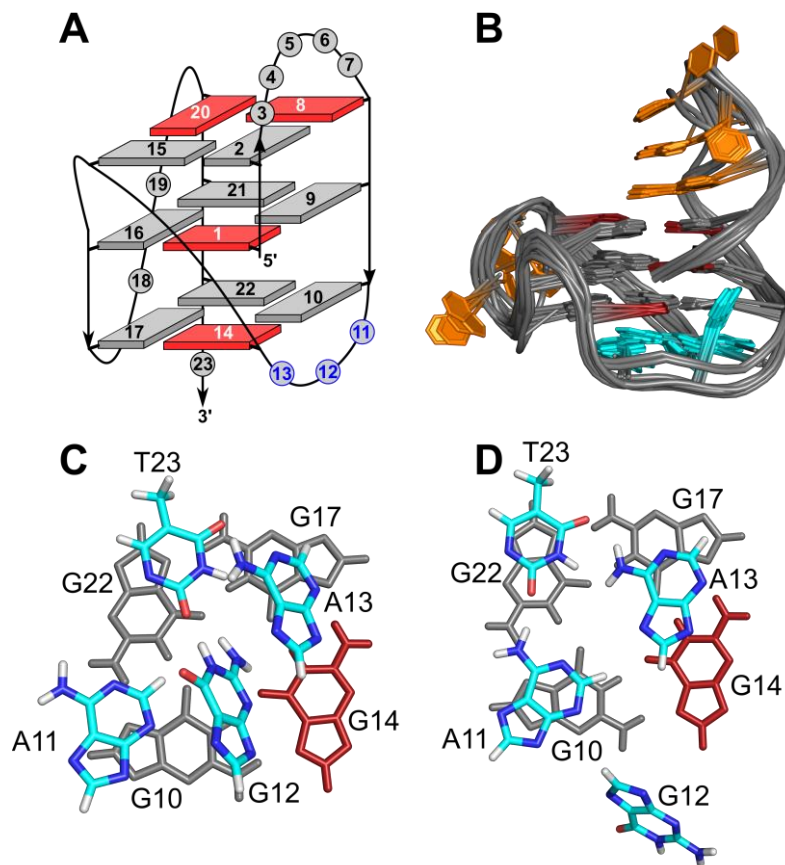


Figure 6. (A) Schematic fold of the *KCNN4* sequence with a conventional V-shaped loop. (B) Superposition of 10 calculated structures of *KCNN4*. (C&D) Top view of the two arrangements of the lateral loop preceding the V-loop and capping the bottom tetrad. The putative G:T hydrogen bond (C) is found in the lowest-energy structure, but only in 2 out of 10 calculated structures. The putative A:T:A triad (D) can be found in 8 out of 10 structures.

After the spectral assignment of this gene segment, the DNA indeed folds into a V-loop G-quadruplex. The vacancy in the first G-column at the 5'-terminus is then filled by the first guanine in the third run of four guanines (Figure 6A). The V-shaped loop connects two G-columns in an antiparallel

fashion, however, there is no strand inversion between the two connected guanines. The strand polarity reversal occurs after the 3'-residue of the V-loop. Such a V-loop topology is defined above as a conventional V-shaped loop. Notably, the sugar pucker of both the 5'- and 3'- anchoring residues of this V-shaped loop adopts a north conformation as expected for the conventional V-loop. Since the 5'-anchoring residue of the V-shaped loop is in *syn* for this (3+1) hybrid-type topology, the 3'-anchoring residue tends to have a glycosidic torsion angle in the *high-anti* range due to the strand polarity inversion following this residue to match the overall upper G-tetrad polarity.

Interestingly, other reported conventional V-loop G-quadruplexes with similar loop progression and without any G analogs also contain an adenine preceding the 5'-anchoring residue (27, 34). In fact, an NAG₄ motif has been proposed as a unique and crucial motif for V-loop G-quadruplex formation (45). For the *KCNN4* sequence, there are two different arrangements for the lateral loop before the V-shaped loop which could be due to the applied force field despite the use of NMR-based distance constraints (Figure 6C&D). Thus, the importance of this lateral loop remains elusive and it can be only speculated whether it is a prerequisite for stable V-loop formation or simply due to the stabilizing behavior of an adenine base as the last residue in a 3-nt lateral loop (7, 46).

It is obvious that the truncation of the flanking sequence is a biased way to obtain such a non-canonical G-quadruplex. In fact, no gene segment in the chromosome will contain two blunt ends. The addition of a 5'-adenine overhang in this design significantly lowered the melting temperature of the G-quadruplex. In addition, the formation of the V-loop topology cannot be unambiguously demonstrated with a 5'-overhang. One study involving a non-canonical parallel G-quadruplex with a snapback loop also suggests that additional overhang residues following the filling guanine significantly lower the free energy of formation. The degree of destabilization is gradual and depends on the number of flanking residues added (47). However, it has also been suggested that the overhang can still be accommodated in the case of the snapback loop design (26) or the V-loop G-quadruplex itself (36). The relevance of a non-canonical G-quadruplex bearing a vacancy in a genomic context remains an interesting question.

3.2 Rational design of sequences to reveal new G-quadruplex topologies with the help of G analogs (Article 4)

The previous section suggests that loop motifs and lengths contribute to the intrinsic stability of the G-quadruplex and ultimately the end-folding. However, mutational studies on sequences derived from genomic sequences suffer from the lack of biological relevance following various residue mutations. As such, one has less freedom to further explore the fundamental basis for the preferred end-folding of G-quadruplexes. It is therefore more inspiring to use rationally designed sequences to engineer G-

quadruplex topologies and then attempt to explain the preferred end-folding by modifying the critical elements observed in the high-resolution structures obtained from NMR studies.

As mentioned above (Section 2.1), a 1-nucleotide propeller loop is most stable. In addition, there was a study mentioning that a 1-nt loop highly favors parallel-stranded G-quadruplexes since various solved G-quadruplex structures forming parallel topologies mostly contain at least one 1-nt propeller loop (42). The propeller loop usually has no direct secondary interaction with the G-core residues and the loop bases are either exposed to the solvent or rather located in the groove to minimize the entropic penalty. However, in a previous study a shorter propeller loop was found to form a more compact G-quadruplex and resulted in a more favorable enthalpy of formation (44).

We initially studied the loop length preferences by constructing new G-quadruplex topologies named hybrid-1R with a +(lpp) loop progression and hybrid-2R with a -(ppl) loop progression. The term “R” comes from the exchange of the lateral and propeller loops in the previously reported hybrid-1 -(pll) and hybrid-2 -(llp) topologies. In the hybrid-1R, the first lateral loop bridges a narrow groove, resulting in a wide groove between the overhang sequences. Previously, it was demonstrated that the hybrid-1R topology can be obtained by adding a duplex through complementary 5'- and 3'-overhang sequences bridging the wide groove (6). Interestingly, the design of a 3:1:1 loop composition still retains the major formation of the parallel G-quadruplex even with the additional energy due to the duplex base-pair formation. Two hybrid-1R topologies can be observed, a major one with exclusive homopolar G-tetrad stacking and one G-column containing all *syn*-guanosines (PDB 6R9L), and a minor species named hybrid-1R' topology with the commonly observed (3+1) hybrid structure presenting a heteropolar as well as homopolar G-tetrad stacking (PDB 6R9K) (Figure 7). The ratio of the coexisting topologies was 2:1:0.4 for parallel:hybrid-1R:hybrid-1R' for the G-quadruplex construct.

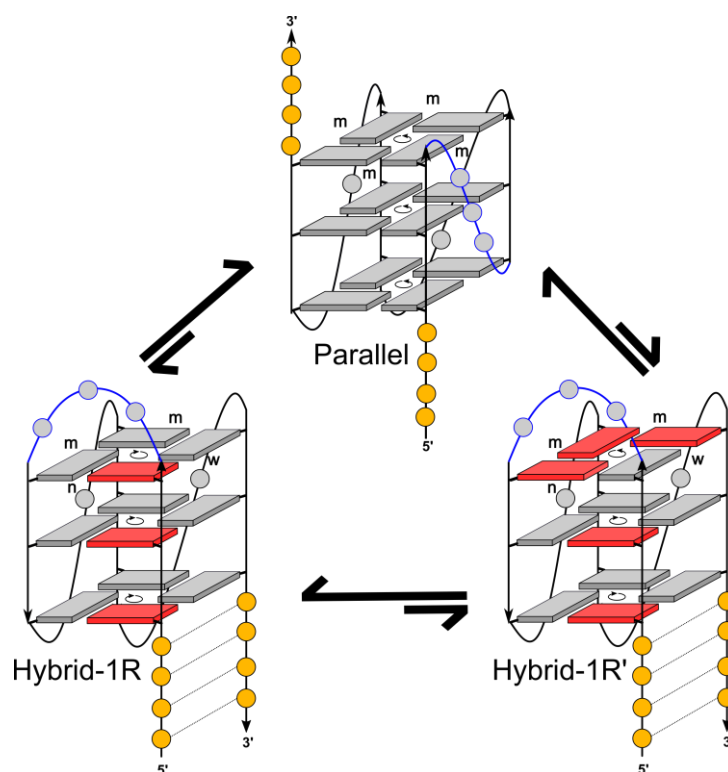


Figure 7. Established equilibrium between three G-quadruplex topologies in a G-quadruplex with a 3:1:1 nucleotide loop progression. The 5'- and 3'-overhangs each contain 4 residues complementary to each other. The parallel G-quadruplex prevailed as the major species, followed by hybrid-1R and hybrid-1R' topologies. The groove dimensions are denoted as follows: w (wide), m (medium), and n (narrow).

Based on this rationale, the new hybrid-2R topology was constructed with an extension of the last loop so that the G-quadruplex contains a loop length progression of 1:1:3 (Article 4). However, only the formation of the parallel G-quadruplex was observed with no hint for the formation of a hybrid-2R topology. We decided to use BrG as a G analog with a *syn*-glycosidic torsion angle preference. Here, the *anti* conformation is disfavored due to the steric clash of the bulky bromine atom with the sugar moiety. The hybrid-2R topology is obtained only with the substitution of BrG in at least two guanosine positions in one G-column. Similar to hybrid-1R, an exclusively homopolar G-tetrad stacking is observed in this topology with an all-*syn*-guanosine fourth G-column (Figure 8A).

The addition of a duplex stem-loop as a lateral loop does not seem to significantly promote a hybrid-2R formation in contrast to the hybrid-1R topology. Of note, a duplex stem-loop can also be formed through a propeller loop lacking only a first base pair and additional base stacking on the G-tetrad when compared to a coaxially stacked lateral loop (Figure 8B; see section 3 on QD hybrids). Such differences in energy contributions, perhaps only provided by the nucleobases at the interface of the G-tetrad and the duplex, are apparently not sufficient to force the (re)folding of the sequence from the preferred parallel

topology to the hybrid-2R topology. This is in contrast to the contribution of the duplex domain for the formation of a hybrid-1R topology in the previously constructed sequence (Figure 7). From the design, it can be seen that residues of the two flanking sequences are located at opposite outer G-tetrad planes when forming a parallel G-quadruplex. In the hybrid-1R case, additional potential energies to enforce folding are given by the interactions (hydrogen bonding, π - π stacking, and London forces) of all nucleobases involved in the formation of the duplex domain itself.

Structure calculations were done on a hybrid-2R topology formed by BrG incorporation at two matching *syn*-guanosine positions and containing a lateral duplex stem-loop as the third loop bridging the wide groove (PDB 7O1H). In this case, the formation of the A:T base pair at the junction between the G-quadruplex and the duplex serves as an indicator for the successful formation of the hybrid-2R topology. It can also be suggested that there are additional capping and hydrogen bond interactions between the two overhang sequences facing each other at the lower G-tetrad (Figure 8B&C).

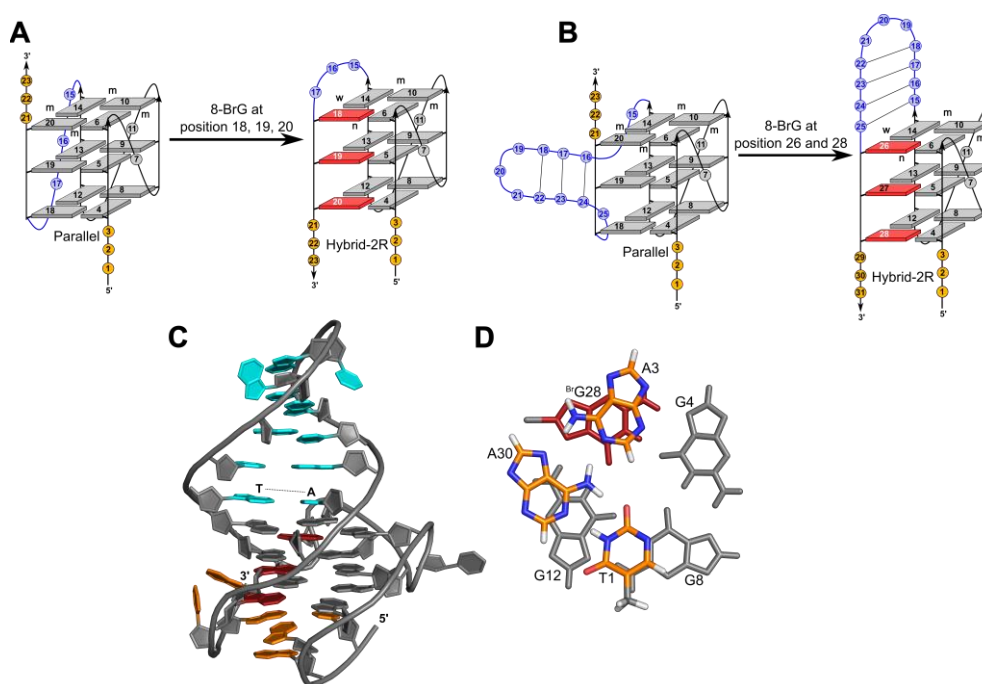


Figure 8. (A) G analogs at matching *syn*-guanosine positions drive refolding from a parallel G-quadruplex into a (3+1) hybrid G-quadruplex named hybrid-2R with exclusively homopolar tetrad stacking and a 3-nt lateral loop. (B) Formation of a hybrid-2R G-quadruplex bearing a lateral duplex stem-loop as its third loop after incorporation of 8-BrG at matching positions to refold from a parallel G-quadruplex. (C) Lowest-energy structure of the hybrid-2R topology (Figure 8B, right) with a lateral duplex stem-loop. An A:T base pair is observed at the interface between the duplex and the G-quadruplex. (D) T:A:A triad formation in the capping structure formed by the 5'- and 3'- flanking sequences bridging the narrow groove. The groove dimensions are denoted as follows: w (wide), m (medium), and n (narrow).

These are initial indications that the 1-nt loop can also form non-parallel G-quadruplexes with exclusively homopolar G-tetrad stacking. Interestingly, no coexisting hybrid-2R' featuring both homopolar and heteropolar tetrad stacking can be observed in contrast to hybrid-1R/hybrid-1R' mixtures. This may indicate an additional loop-position dependence on the formation of a particular G-quadruplex topology. It is also noteworthy that the preference of the 1-nt loop for the parallel G-quadruplex with all *anti*-guanosines outweighs the preference of BrG to adopt a *syn* conformation in a (3+1) hybrid topology.

3.3 Towards the explanation of the impact of loop length and composition on the G-quadruplex topologies (Article 5)

To further investigate the structural basis of the favored G-quadruplex topologies, a model system was used in which the G-quadruplex can only fold into two three-layered G-quadruplexes, i.e., a parallel G-quadruplex with a $-(ppp)$ loop progression and a (3+1) hybrid G-quadruplex with a $+(lpp)$ loop progression. The folding is limited by the use of a vacancy in the first G-column, which can subsequently be filled by a lateral snapback loop starting from the last G-column. It should be noted that a corresponding propeller-type snapback loop has never been observed to date (Figure 9). In addition, two loop arrangements of either 3:1:1 (L311) or 3:3:1 (L331) in this study prevent the formation of a third lateral loop bridging the wide groove in analogy to the hybrid-2R topology above. Careful analysis of the NMR spectra for both L331 and L311 suggested that both sequences form three coexisting topologies with different molar ratios. To resolve the species, a 5'-overhang was added to promote the formation of a parallel G-quadruplex by destabilizing the (3+1) hybrid G-quadruplex with a 5'-terminus at the central tetrad. In addition, the preference of a 5'-G to adopt a *syn* conformation due to a possible hydrogen bond coming from the 5'-OH as the donor is hampered by the addition of this overhang residue (48, 49). On the other hand, the (3+1) hybrid topology was enforced by incorporating a *syn*-favoring BrG in the matching position of the central tetrad whose glycosidic torsion angle pattern is a determinant for different G-quadruplex topologies. To be noted, $+(lpp)$ topologies have been previously observed and explained above (Figure 7) (6).

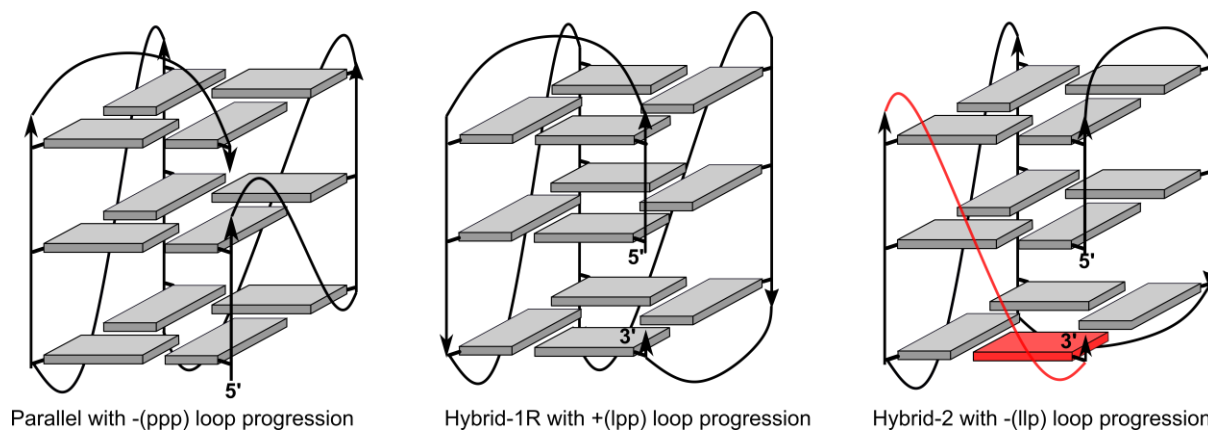


Figure 9. Putative folding of three-layered G-quadruplexes based on the G-quadruplex design with either a 3:1:1 or 3:3:1 nucleotide loop progression and a 3-nt snapback loop after the last G-column filling a vacant site. No information on the tetrad polarity and nucleoside conformers are given in the schematic. The red “propeller” snapback loop on the right has never been observed before, suggesting that the vacancy in such a topology cannot be filled with a 3'-guanine (in red).

Nearly complete NMR assignments were performed on all mutations to confirm the particular fold of interest. This was combined with a van't Hoff analysis to extract thermodynamic parameters for folding. NMR-based structures were determined on the parallel (PDB 7ZEM) and two +(lpp) topologies with either both homopolar and heteropolar tetrad stacking (PDB 7ZEK) or with exclusively homopolar tetrad stacking (PDB 7ZEO) (Figure 10). Several rules emerged regarding the orientation of the loop residues and their position in the G-quadruplex. They can be summarized as follows:

1. For a lateral loop, especially when bridging the narrow groove, the first residue tends to be exposed to the solvent and be disordered while the third residue tends to stack above the G-tetrad. In the present CTA loop, the mutation of the first residue to adenine destabilized the G-quadruplex (Figure 10C&D), in agreement with previous reports that the adenine in the first residue of a loop is disfavored (50, 51). Mutation of the last adenine to thymine or even an abasic residue further destabilized the +(lpp) topology. Such an arrangement can also be seen in several examples of NMR-based calculated structures with a (3+1) hybrid topology derived from telomeric sequences (52).
2. For the propeller loop in the -(ppp) topology, the first residue tends to be exposed to the solvent while the last residue is located in the G-quadruplex groove (Figure 10B). Similarly, adenosine in the first position of the propeller loop destabilizes the parallel G-quadruplex. However, a structural comparison with other structures containing a propeller loop, such as telomeric sequences, lacks a clear trend. Again, this suggests a higher flexibility of the propeller loop residues due to the lack of defined inter-residue interactions.

3. Understanding the preferred end-folding of G-quadruplexes

- Exceptions arise when there are tertiary interactions between loops. Bases can be recruited to form further hydrogen bonds to other loop or overhang residues. This can be seen in the last residue of the first loop in the parallel G-quadruplex structure (PDB 7ZEM) or the last residue of the second loop in the +(lpp) topology (PDB 7ZEK) (Figure 10B&C). Such subtle effects could play a major role in determining the fate of G-quadruplex folding. For instance, a G-quadruplex design based on a single mutation of a loop residue from a telomeric sequence appears to be sufficient to refold the normally observed (3+1) hybrid topology into a +(III) antiparallel topology in potassium ion solution due to an A:T Watson-Crick base pair formation between two lateral loops (53).

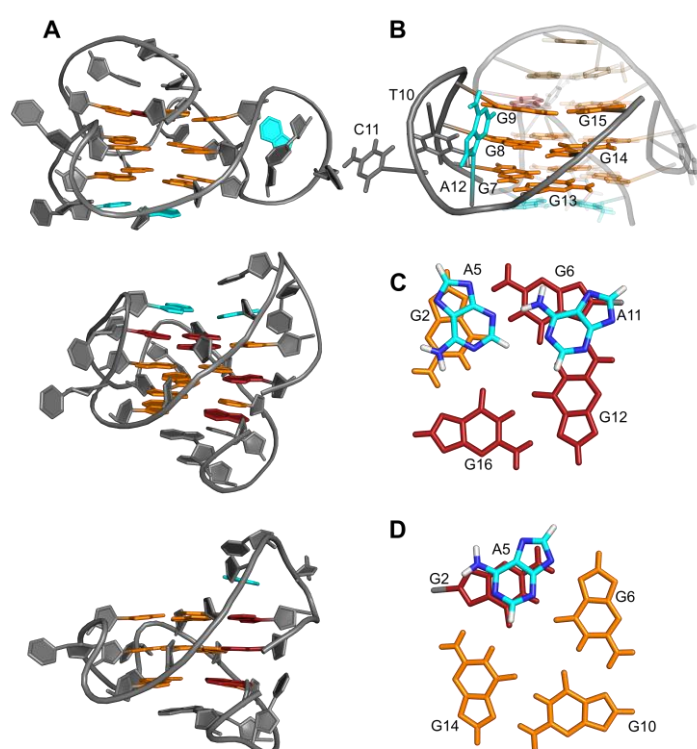


Figure 10. (A) Calculated lowest-energy structures based on L331 and L311 loop arrangement. (Top) Parallel G-quadruplex derived from L331 (PDB 7ZEM); (middle) major hybrid-1R' G-quadruplex from L331 (PDB 7ZEK); (bottom) coexisting hybrid-1R G-quadruplex from L311 (PDB 7ZEO). (B) Side view of the second 3-nt propeller loop for the parallel G-quadruplex. Adenine A12 which is located in the groove is colored cyan. (C) Top view of the A:A base pair capping the G-tetrad in the hybrid-1R' G-quadruplex. A5 is positioned in the 3-nt lateral loop while A11 is positioned in the 3-nt propeller loop. (D) Top view of the outer tetrad highlighting the location of the stacked A5 in the 3-nt lateral loop of the hybrid-1R G-quadruplex.

As mentioned above, the L331 parent sequence is composed of various G-quadruplex topologies. Two of them can be clearly resolved; the +(lpp) hybrid-1R' predominates as the major topology, while the parallel G-quadruplex constitutes the minor species. However, there seems to be another only sparsely populated species. Interestingly, shortening the central 3-nt loop (L331) to a 1-nt loop (L311) resulted in different populations of the G-quadruplexes. Analysis of the three coexisting L311 topologies revealed a population of parallel > hybrid-1R (lpp with exclusively homopolar tetrad stacking) > hybrid-1R' (lpp with both homopolar and heteropolar tetrad stacking). In fact, the hybrid-1R with exclusively homopolar tetrad stacking was found to be the third very low-populated species for the L331 sequence. This +(lpp) topology contains an all-*syn* first G-column, thereby maintaining the (3+1) hybrid topology (Figure 10; PDB 7ZEO). The population ratio of the parallel and hybrid-1R topology was similarly enhanced upon shortening the central loop from three to one nucleotide.

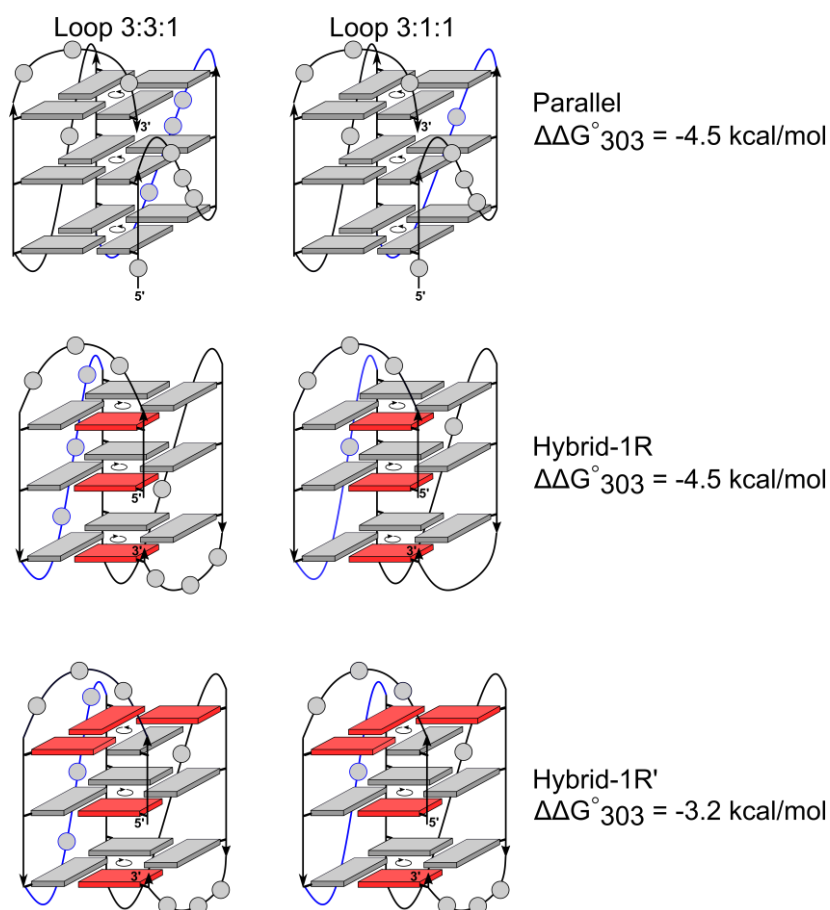


Figure 11. Degree of stabilization upon shortening the central propeller loop from a 3-nt to a 1-nt loop. The $\Delta\Delta G^{\circ}_{303}$ is the difference between the ΔG°_{303} of L331 and that of L311. Free energies were derived from a van't Hoff analysis of the UV melting curves assuming temperature-independent enthalpies and entropies.

3. Understanding the preferred end-folding of G-quadruplexes

It appears that shortening the central 3-nt loop into a 1-nt loop favors not only the parallel G-quadruplex but also the (3+1) hybrid topology containing only homopolar G-tetrad stacking. The differences in ΔG° at 30°C upon shortening the central loop are as follows: -4.5 kcal/mol for the parallel and +(lpp) hybrid-1R G-quadruplex and -3.2 kcal/mol for the +(lpp) hybrid-1R' G-quadruplex (Figure 11). Therefore, the existing rule can be extended as follows:

“a 1 nt-loop favors to connect two guanines in tetrads of the same polarity, i.e., two guanosines with the same glycosidic torsion angle.”

Intriguingly, calculations carried out in 2011 indicated a similar trend (54). Experimental observation of such a behavior of a 1-nt propeller loop was masked in the past by the absence of the corresponding (3+1) hybrid G-quadruplex structure with exclusively homopolar G-tetrad stacking, which has only recently been reported (6). Before, a 1-nt loop was simply described as a loop motif favoring a parallel topology. It should be noted that a parallel and hybrid-1R structure cannot be unambiguously distinguished by circular dichroism spectra.

Nevertheless, it seems difficult to predict the end-folding of a G-quadruplex by considering the sequence context alone, since various tertiary interactions from overhang and loop sequences may change the overall free energy of formation. Of note, the positional dependence of loop length (55) could also result in different energy contributions for the G-quadruplex end-folding, a problem not addressed in this study.

4. Quadruplex-duplex hybrids

It is often observed that G-quadruplexes with long loops tend to have a low degree of thermal stability. This is true for disordered loop residues. Tertiary interactions between loop residues or between loop residues and the G-tetrad often define a certain topology and contribute to the stability of the favored G-quadruplex. Thus, a duplex with its typical Watson-Crick base pairs allows the elongation of a G-quadruplex without its destabilization. Such a construct is defined as a quadruplex-duplex (QD) hybrid. The duplex domain can not only be found in the loop region but also within the overhang sequence. It was first designed upon modifying an antiparallel thrombin binding aptamer (TBA) lateral loop (56) and further explored by Phan *et al.* (57). The duplex domain can be located in all three types of G-quadruplex loops. In general, two orientations of the duplex relative to the G-core can be defined in a QD hybrid. The first orientation, orthogonal to the G-quadruplex tetrad planes, is formed by a double-helical propeller loop. Since the propeller loop bridges two opposite tetrads, a putative first base pair cannot be formed (Figure 12A; PDB 2M93). On the other hand, a coaxially positioned duplex can be observed when located in either a lateral or diagonal loop. As the diagonal loop typically bridges two distal G-columns, two purine bases are usually used for forming a first homopurine base pair (PDB 2M91). On the other hand, the lateral loop bridging a wide groove allows the interfacial first base pair to form typical Watson-Crick hydrogen bonds (57) (Figure 12B; PDB 2M8Z; PDB 2M90). Only the coaxially positioned duplex features the QD interface at the junction with a direct unique interaction, i.e., π - π stacking between the Watson-Crick base pair and the G-tetrad (58). In fact, there is a matching dimension for the wide groove of the G-quadruplex and the minor groove of the duplex which allows a smooth transition and a typical twist for the right-handed helicity to progress continuously (Figure 12C).

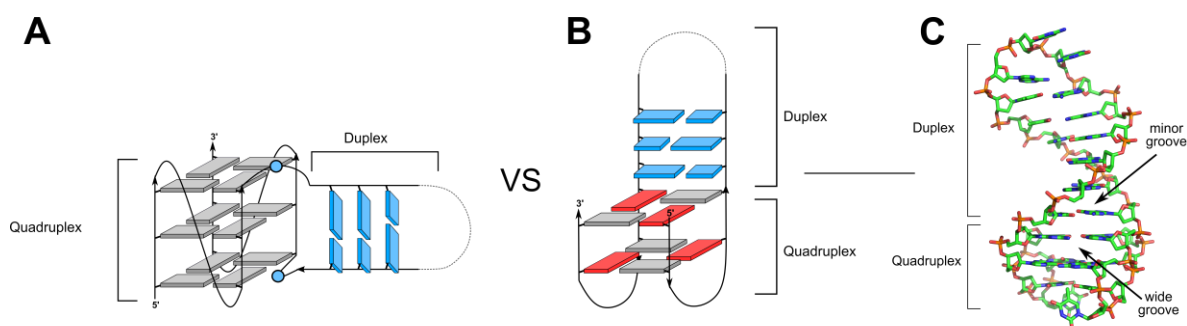


Figure 12. (A) Schematic of a parallel QD hybrid with a duplex-containing propeller loop. (B) Schematic of an antiparallel QD hybrid with a duplex-containing lateral loop bridging a wide groove. (C) Cartoon representation of an antiparallel QD hybrid (PDB 2M8Z) highlighting the smooth progression of the G-quadruplex and duplex backbone due to the matching size of the G-quadruplex wide groove and duplex minor groove.

In contrast to the orthogonally oriented duplex, the presence of a coaxially stacked duplex stem-loop affects the thermal stability of the G-quadruplex depending on its loop length and composition. Notably, there is no change in the melting temperature when a longer double-helical propeller loop is compared to a non-duplex propeller loop. On the other hand, a significant change in the melting temperature is observed when the interfacial base pair directly at the QD junction is varied for the coaxially, but not for the orthogonally oriented duplex stem-loop (59).

The formation of QD hybrids can also be found within the genome. Bioinformatic searches of the human genome revealed putative sequences located in various regulatory regions of many genes capable of forming QD hybrids with the duplex incorporated in a G-quadruplex loop (60). Some of the structures have been confirmed by NMR experiments (61, 62). In fact, the presence of a QD junction and its interface is suggested by the formation of a G-quadruplex within a long double-helical genomic domain, preferring a coaxial stacked duplex domain at the 5'-end of the G-quadruplex (63).

Furthermore, the presence of a duplex-forming sequence in the loop can be used to engineer novel G-quadruplex topologies. It is known that the folding of the duplex stem-loop is kinetically faster than the folding of the G-quadruplex and thus helps to accelerate G-quadruplex formation (64). Correspondingly, the formation of the duplex can subsequently guide the folding of the G-quadruplex (65). A novel topology of G-quadruplexes (Figure 7) can be formed by refolding a parallel G-quadruplex through the incorporation of Watson-Crick pairing residues within the 5'- and 3'-overhang with the two overhang sequences bridging the wide groove (6). The addition of a duplex to an antiparallel G-quadruplex like TBA also increases the binding affinity to its target and its resistance to degradation (56, 66). Despite the lack of detailed structural information, various split G-quadruplex designs first proposed by Willner *et al.* (13) and then widely used for analytical purposes also represent QD hybrids (67–69). A strand capture strategy that subsequently fills a guanine vacancy has also been developed, creating a bimolecular QD hybrid structure (39).

4.1 Stacking interactions at the quadruplex-duplex interface (Article 6)

In this article, QD hybrids were designed with the duplex stem-loop positioned as an overhang at either the 5' or the 3'-end of the G-quadruplex. A parallel G-quadruplex based on *c-Myc* was used to study the stability effects of the duplex stem-loop. Such designs were named *Myc-dup5* and *Myc-dup3* and used the same C:G base pair at the QD interface (Figure 13). These architectures with a non-anchored duplex domain are expected to provide a greater degree of flexibility when compared to a duplex stem-loop incorporated as an intrinsic lateral loop.

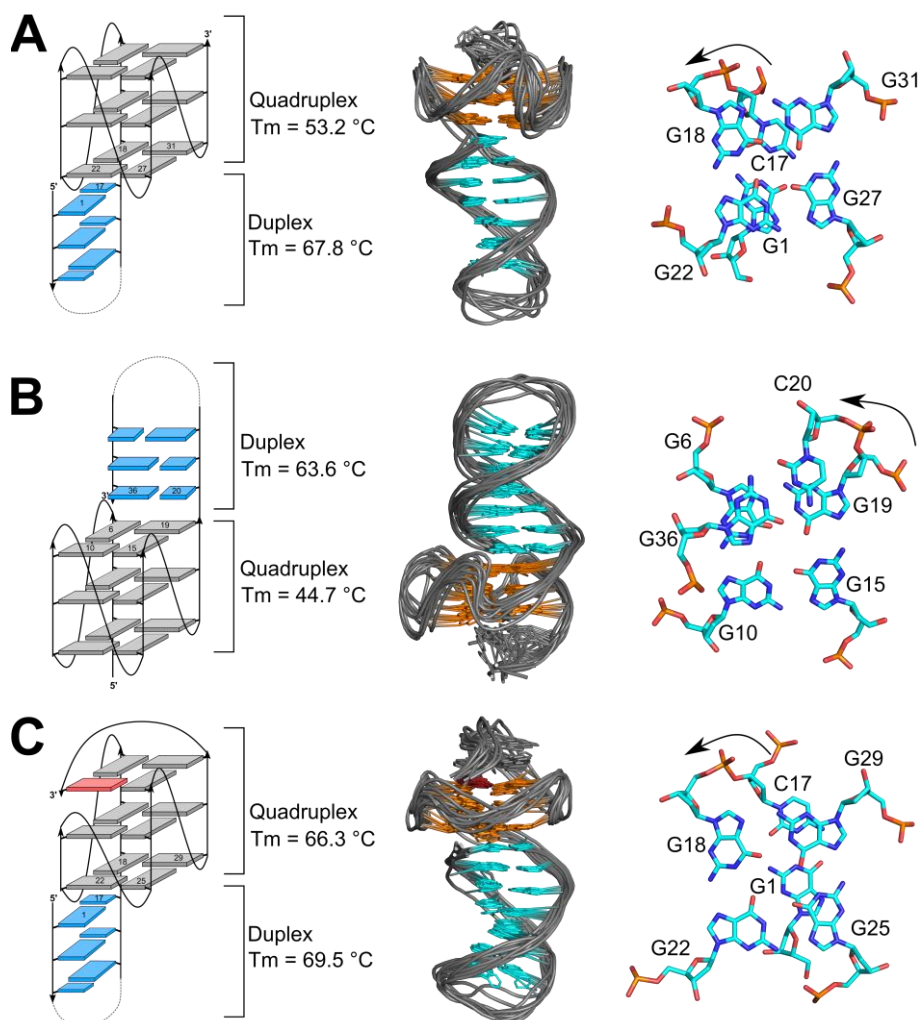


Figure 13. (Left) Schematic QD topologies and their melting temperatures; (center) superposition of corresponding ten lowest-energy structures; (right) QD interface showing stacking interactions between the G-tetrad and the interfacial base pair of the duplex for (A) *Myc-dup5* (PDB 6ZL2), (B) *Myc-dup3* (PDB 6ZL9), and (C) *Myc3l-dup5* (PDB 6ZTE). In the structural ensemble, duplex residues are colored in blue, *anti*-guanosines are colored in orange, and the *syn*-guanosine in the *Myc3l-dup5* is colored in red. Residues in propeller loops and hairpin T₃ residues are omitted for clarity.

NMR-based structural studies were done on the two sequences. Apparently, a defined position for the first Watson-Crick base pair can be observed and is supported by various NOE contacts found between the G-tetrad and the duplex at the QD interface (Figure 13A&B). However, there are different stacking interactions within the two designs. For the *Myc-dup3* design, the first residue (cytosine) of the duplex stem-loop tends to protrude outward and the last residue of the duplex (guanine) shows almost optimal stacking on the G-tetrad (comparable to PDB 2M8Z and 2M90) (57). In contrast, both residues of the first base pair in the *Myc-dup5* design are located above the G-tetrad (Figure 13A). This can be explained by the typical right-handed helical progression from the G-tetrad to the duplex stem-loop and

vice versa. The twist angles for the base step at the QD junction in both designs are smaller than 30°. These smaller twist angles resemble the base step in a G-quadruplex and deviate from the typical twist angle of about 36° for B-DNA. For the *Myc-dup3* design, an average twist angle of about 27° is observed when going from the guanine tetrad to the first residue of the duplex stem-loop, shifting the cytosine away from the G-tetrad in the 10 calculated structures. For the *Mycdup-5* design, a corresponding average twist angle of only 20° is observed.

The thermal stability of the two constructs was studied by UV melting experiments following the hyperchromicity and the hypochromicity at 260 nm for the duplex domain and 295 nm for the G-quadruplex domain, respectively (70). The melting temperatures obtained by the UV spectroscopic technique are in agreement with those obtained by DSC, where two peaks can be observed and deconvoluted in the DSC thermograms. This indicates two independent transitions from two independent melting events for the G-quadruplex and duplex domains. To investigate the effect of the duplex component on the stabilization of the G-quadruplex domain, the salt composition was adjusted to prevent a duplex melting event prior to the G-quadruplex melting. In particular, sodium salt with high ionic strength (~120 mM Na⁺) was used in the buffer system for the melting experiments, replacing the potassium buffer used in the NMR experiments (~10 mM K⁺). Potassium ions are known to stabilize the G-quadruplex better than other alkali metal ions due to their lower desolvation energy and a better match in ionic size (71).

The duplex melting temperatures of *Myc-dup3* and *Myc-dup5* are quite similar. However, there is a large difference in the melting temperatures of the G-quadruplex domains. The melting temperature of the G-quadruplex component in the *Myc-dup5* is significantly higher compared to *Myc-dup3*. This suggests additional stabilizing effects from base stacking of the 5'-duplex on the G-tetrad relative to its 3'-duplex counterpart in *Myc-dup3*. On the other hand, similar melting temperatures of the duplex in both the 5'- and 3'-overhangs can be rationalized by the premature melting of the G-quadruplex with loss of interfacial stacking interactions between the duplex and the G-quadruplex. The small difference in duplex melting is due to the different duplex domains of *Myc-dup5* and *Myc-dup3*, having an inverted 5'- to 3'-polarity.

To study ligand binding, an additional design termed *Myc3I-dup5* derived from *Myc-dup5* by the addition of a diagonal snapback loop was employed (Figure 13C). Due to its different topology, the thermal stability of *Myc3I-dup5* cannot be directly compared with *Myc-dup3* and *Myc-dup5*. Surprisingly, the position of the interfacial duplex base pair is further shifted towards the G-tetrad in *Myc3I-dup5* when compared to *Myc-dup5* as revealed by NMR experiments (Figure 13C). Interestingly, such stacking patterns show a resemblance to a previously designed QD hybrid with a diagonal duplex stem-loop (PDB 2M91) (57).

5. Ligand binding to quadruplex-duplex hybrids

As mentioned above, many G-quadruplex ligands simply stack on an accessible open tetrad. As such, specific discrimination between different G-quadruplex structures and/or different sequences of G-quadruplexes is only poor and leads to unspecific effects when applied inside the cell. The QD hybrids may serve as a unique architecture that can be targeted by typical G-quadruplex binding drugs (72). In a previous study, the failure of ligand binding to a telomeric QD junction may have been caused by the formation of a base triad rather than a base pair at the G-quadruplex interface (58). The first proposal for recognizing a QD interface was made by Phan *et al.* using a G-quadruplex binding ligand with its extended aromatic ring system covalently linked to a duplex minor groove binder, thereby achieving a dual G-quadruplex and duplex recognition (73, 74). Although such an approach seems plausible, no detailed structural study has been made on the actual binding site of the used ligands. It was proposed that the G-quadruplex binding structure binds at the free outer tetrad rather than at the tetrad facing the duplex domain.

5.1. Quadruplex-duplex junctions as binding hotspots for an indoloquinoline ligand (Article 6)

Employing the three designed QD hybrids (Figure 13), binding studies focusing on both structural and thermodynamic aspects were done with a phenyl-indoloquinoline (PIQ)-based ligand previously synthesized in our group (75, 76). The ligand PIQ-4m with a positively charged diethylaminopropyl side chain and a phenyl substituent was shown to be an excellent binder to a parallel G-quadruplex with good selectivities against duplex DNA. It was suggested that the ligand stacks onto the accessible outer G-tetrad (75).

Thermodynamic studies were performed on the binding of PIQ-4m to all three QD hybrids (Figure 14). In fact, the design of *Myc3I-dup5* was expected to alleviate a complex binding behavior as has been observed in the titration of PIQ-4m to *Myc-dup5*. Notably, there are at least three perceivable binding sites for typical DNA binders in *Myc-dup5* and *Mycdup-3*: the QD interface, the opposite open G-tetrad, and the duplex domain. On the other hand, the isothermal titration calorimetry (ITC) thermograms were fitted with only two sets of binding sites, increasing the ambiguity of the fitted parameters due to the interdependence of the different binding processes (77). Therefore, only the parameters of the better-defined most affine binding site(s) were discussed. During the titration of PIQ-4m to *Myc-dup5*, three different binding modes were indicated. The thermogram profile starts with a binding enthalpy of -7.5 kcal/mol, followed by another binding mode with a more exothermic binding enthalpy, and subsequent

5. Ligand binding to quadruplex-duplex hybrids

binding through weak and possibly unspecific interactions such as electrostatic forces. The design of *Myc3l-dup5* simplified analysis by blocking the 3'-open tetrad through a diagonal snapback loop.

Apparently, the binding behavior of PIQ-4m to *Myc-dup3* differs from that of the *Myc-dup5* target. For the former, a clear lower baseline indicating a strong binding event can be observed with a stoichiometry of nearly 1. In addition to a significantly more exothermic binding, the $K_a > 10^7 \text{ M}^{-1}$ is about 8-fold higher than that obtained by titrating the ligand with the *c-Myc* quadruplex (78). Such parameters already suggest that the binding mode to the most affine site of the *Myc-dup3* QD hybrid differs from the simple end-stacking proposed for the parallel G-quadruplex *c-Myc* and the binding site is rather located in the newly introduced QD interface of the 3'-tetrad.

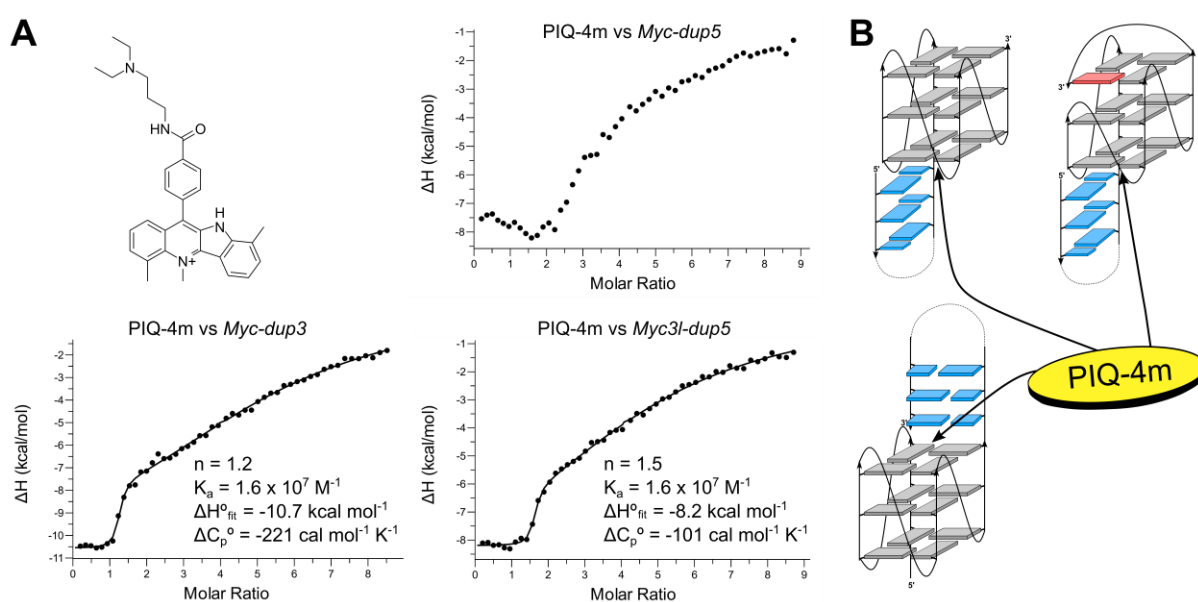


Figure 14. (A) PIQ-4m chemical structure and ITC thermograms of its titration to the three QD hybrids at 313 K in 120 mM K^+ . (B) Proposed location of the PIQ-4m when binding to the three QD hybrids.

On the other hand, the binding behavior of PIQ-4m towards *Myc3l-dup5* also differs from that obtained with the *Myc-3sbl* counterpart without a duplex overhang (41). The obtained affinity constant for PIQ-4m binding to *Myc3l-dup5* is about five times higher and the binding is slightly more exothermic compared to the binding of PIQ-4m to *Myc-3sbl*. This fact also suggests that the binding of the PIQ-4m binding near the QD interface in the 5'-tetrad is not simply an end-stacking. It is also interesting to compare the binding parameters obtained from the ligand titration to *Myc-dup3* and *Myc3l-dup5*. For the high-affinity binding site of *Myc3l-dup5*, a stoichiometry of almost two and significantly less exothermic binding was obtained compared to *Myc-dup3*. There is also a positive entropy for the ligand binding to *Myc3l-dup5*, but not to *Myc-dup3*.

One significant contribution of binding with such a typical non-polar ligand in an aqueous solvent constitutes the hydrophobic effect. This effect is correlated with a molar heat capacity change (ΔC_p°), which is determined by the temperature dependence of the binding enthalpies in temperature-dependent excess-site ITC experiments. As expected, PIQ-4m binding to both *Myc-dup3* and *Myc3I-dup5* resulted in negative ΔC_p° values. Since the obtained ΔC_p° values are larger than the ΔS° values, an enthalpy-entropy compensation is suggested. However, the ΔC_p° value for binding to *Myc-dup3* is more negative than the corresponding value obtained for binding to *Myc3I-dup5*. This contradicts the trend of binding entropies (see above). Obviously, there is a more favorable entropic contribution from the hydrophobic effect when binding to *Myc-dup3* when compared to *Myc3I-dup5*. Thus, significant unfavorable entropic penalties for binding to *Myc-dup3* must result from conformational changes and/or restricted flexibilities in both rotational and translational motions. The more negative change in the molar heat capacity when binding to *Myc-dup3* may indicate different binding geometries of the 3'-QD when compared to the 5'-QD interface as seen in the three-dimensional structures (Figure 13). Notably, both ΔC_p° values obtained for the binding at the QD interface were also found to be more negative than those obtained for the same ligand binding at an outer G-tetrad of the parallel *c-Myc* (78).

Further NMR structural studies were carried out on the binding of PIQ-4m to the three QD hybrids. For *Myc-dup3*, a new set of peaks and the disappearance of the peaks belonging to the free *Myc-dup3* could be observed at the addition of one equivalent of ligand, indicating a saturation at a 1:1 ligand-to-DNA stoichiometry. On the other hand, peaks for free *Mycdup-5* and *Myc3I-dup5* could still be observed at a 1:1 molar ratio, corroborating the stoichiometries as obtained from the ITC. The location of the ligand binding to all three QD hybrids is suggested to be at the QD interface by following the chemical shift changes and various signal broadening effects due to exchange processes. A severe broadening and the loss of NOE contacts from the duplex to the adjacent G-tetrad in *Myc-dup3* suggests a (partial) insertion of the ligand at the QD junction. Further structural details could not be extracted due to the loss of various resonances through their significant broadening. Nevertheless, the combined results from calorimetry and NMR studies on the binding of PIQ-4m to the QD hybrids do not corroborate the previous suggestions (73). Rather, the QD interface is a binding hotspot that provides a more affine binding site when compared to an accessible outer G-tetrad.

5.2 Thermodynamic and structural studies on quadruplex-duplex binders (Articles 7 & 8)

A new construct related to *Myc-dup3*, named *QD3-sbl*, was designed in which the G-quadruplex domain contains an interrupted G-tract in the first G-column with only two consecutive guanine residues. The

5. Ligand binding to quadruplex-duplex hybrids

vacancy is subsequently filled by anchoring a guanine residue from the elongation of the 3'-overhang to form a duplex hairpin structure as a lateral snapback loop (Figure 15). It is expected that the dynamics at the junction is reduced by the filling guanine to provide better resolution for further structural studies. The arrangement of nucleobases at the QD junction of this construct (PDB 7PNE) is identical to *Myc-dup3*.

Five G-quadruplex ligands known as G-quadruplex end-stackers were subsequently titrated to *QD3-sbl* using ITC. Due to the complexity of having different possible binding sites in the *QD3-sbl* design, an additional construct without a duplex domain in the 3'-diagonal snapback loop (*Q3-sbl*) and a free duplex stem-loop (*D3-HP*) were also titrated with these five ligands to better resolve different binding modes on the *QD3-sbl* hybrid (Figure 15).

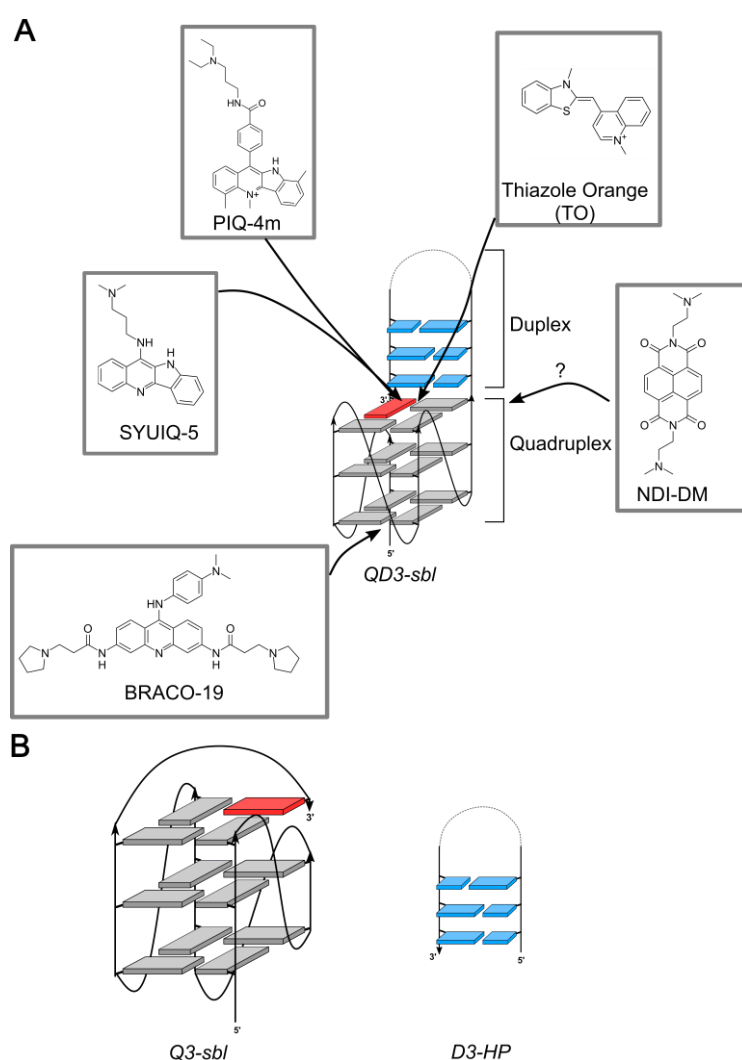


Figure 15. (A) Five ligands and their proposed high affinity binding sites based on ITC and NMR studies. (B) Topologies of *Q3-sbl* and *D3-HP* with seven Watson-Crick base pairing residues used to simplify ligand binding.

The two indoloquinolines PIQ-4m and SYUIQ-5 show similar binding affinities of about 10^7 M^{-1} when titrated against *QD3-sbl*. Expectedly, the binding affinity of the two ligands to *QD3-sbl* is higher than for the truncated sequences (*Q3-sbl* and *D3-HP*). Notably, similar binding enthalpy and binding affinity were obtained for the titration of PIQ-4m to *QD3-sbl* when compared with the titration to *Myc-dup3*, suggesting a comparable binding geometry. A similar binding could also be observed with thiazole orange, although the binding constant to *QD3-sbl* was about 10 times smaller than that of the indoloquinolines. However, with a stoichiometry of 2 the possibility of thiazole orange dimerization or a more complex binding mode cannot be excluded (Figure 16). Supported by ROESY exchange cross peaks, all three ligands were found to bind primarily at the QD junction. Interestingly, thiazole orange known to be a universal DNA binder prefers to bind at the QD junction, again highlighting the QD junction as a binding hotspot. In terms of selectivity, all three ligands have similar discriminatory power when targeting the QD hybrid and the parallel G-quadruplex. However, the trend for the selectivity against the hairpin duplex is notable and follows the order: PIQ-4m \gg TO $>$ SYUIQ-5. Surprisingly, SYUIQ-5 is also known as a G-quadruplex binder and has been tested both biologically and structurally (79, 80), but no detailed thermodynamic study of the binding has been performed to directly compare the selectivity against double-helical DNA. The two ligands PIQ-4m and SYUIQ-5 were subsequently subjected to a further detailed structural characterization upon binding QD hybrids in two separate studies (Articles 7&8).

The ITC thermogram of the naphthalene diimide (NDI) derivative was too complex for a detailed analysis. In addition, its insolubility in DMSO prevented further NMR titration experiments. On the other hand, the trisubstituted acridine-based ligand BRACO-19 (81) showed a stoichiometry of 0.5 when titrated against *QD3-sbl*. This may be due to multiple binding modes reflected in the thermogram, which increases the difficulty in attempts to reliably fit the titration curves. Nevertheless, the binding affinity of BRACO-19 towards the parallel *Q3-sbl* was found to be the highest, consistent with previously reported data (Figure 16) (58, 82). In one of these studies, using a QD hybrid with a triad at the QD interface, a related disubstituted acridine-based ligand was found by molecular docking to rather bind outside of the junction and to end-stack on the open tetrad. Aside from the higher energy required to unstack the triad from the G-tetrad when compared to a base pair, the substituents themselves could result in a steric clash for the binding, preventing interaction at the QD junction (58). Indeed, analysis of the NMR spectra in the present study suggested that BRACO-19 initially binds to the 5'-open tetrad when titrated to *QD3-sbl*.

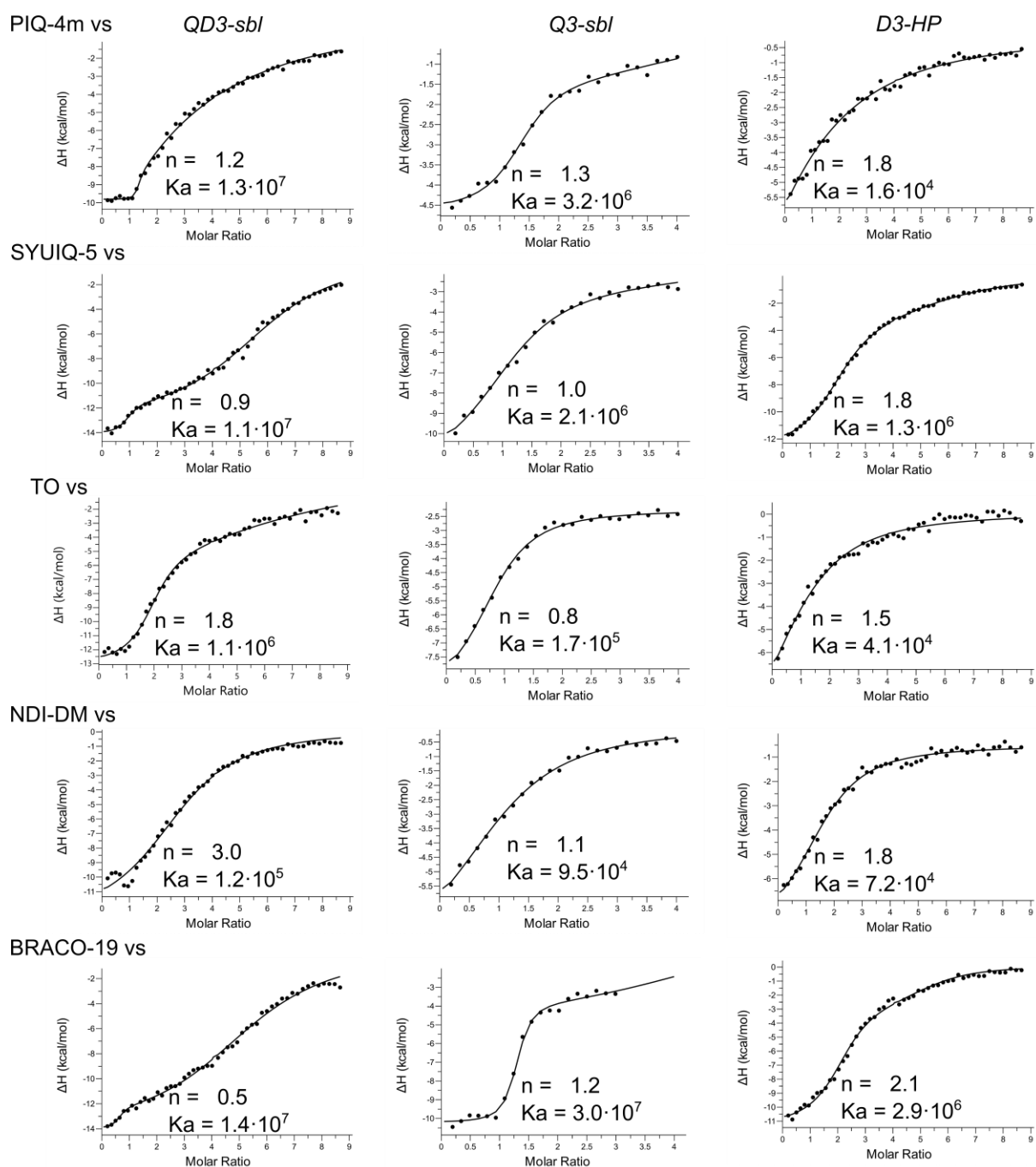


Figure 16. ITC thermograms for five ligands titrated against *QD3-sbl*, *Q3-sbl*, and *D3-HP* in 120 mM K^+ buffer at 313 K. Stoichiometries and binding constants (M^{-1}) for the most affine binding sites are indicated.

5.3 Indoloquinolines intercalate at quadruplex-duplex junctions (Article 7)

An NMR-based structure derived for the complex of the *QD3-sbl* hybrid with bound SYUIQ-5 showed the intercalation of the ligand at the QD junction (Figure 17A). The crescent-shaped ligand inserts into the QD interface with the concave side pointing towards the center of the tetrad, while the convex part as well as the aliphatic side chain is located in the vicinity of the G-quadruplex groove as well as the duplex minor groove. There are no obvious intermolecular interactions between the aliphatic side chain and the

nucleic acid grooves due to the high flexibility of the side chain. Interestingly, the intercalation is followed by a lateral shift of the duplex domain towards the center of the tetrad to maximize π - π stacking interactions with the ligand. On the other hand, the protonated *N5* in the quinoline ring system is located above the central channel of the G-tetrad at the QD interface, featuring a negative potential. Electrostatic interactions towards all carbonyl oxygens of the G-tetrad and also of the guanine in the first base pair of the duplex can be assumed (Figure 17B). Additionally, the binding of SYUIQ-5 was also tested with an antiparallel QD hybrid named *QD2l* as a target (PDB 2M8Z) (57). Following exchange cross peaks of G-quadruplex imino protons between the free and bound states, SYUIQ-5 was again found to intercalate at the QD junction and this was further supported by the observation of a considerably upfield-shifted G-imino proton of the interfacial base pair in the duplex (Figure 17C).

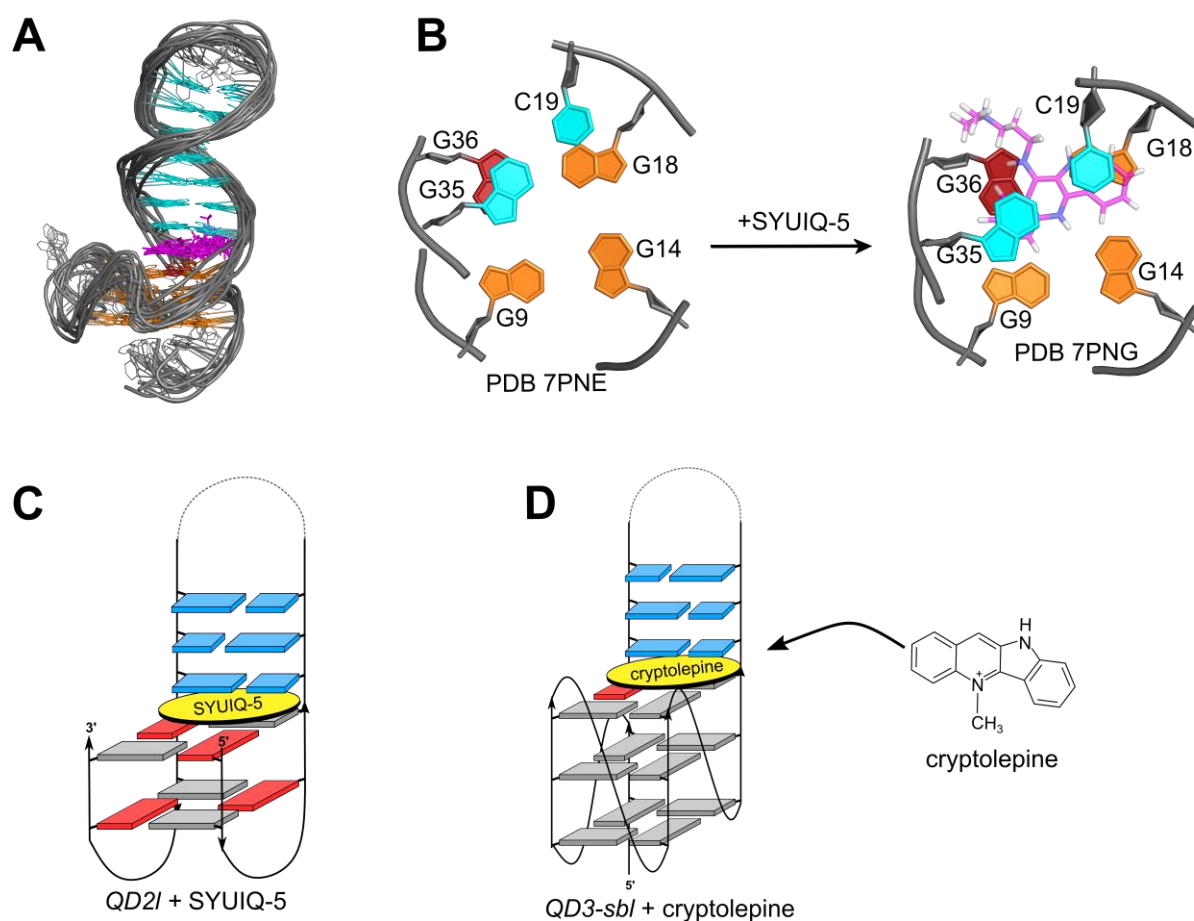


Figure 17. (A) Superposition of 10 lowest-energy structures of *QD3-sbl* complexed with SYUIQ-5. (B) View at the QD interface of the free (PDB 7PNE) and SYUIQ-5 bound *QD3-sbl* highlighting the ligand position and the base pair shift after ligand intercalation (PDB 7PNG). (C&D) Proposed binding of SYUIQ-5 and cryptolepine to *QD2l* and *QD3-sbl*, respectively.

To test the impact of the side chain, a natural product named cryptolepine was used for further binding studies. Cryptolepine is also an indoloquinoline-based ligand lacking any side chain but with a permanent positive charge for the methylated *N5*. This ligand is also known as a duplex intercalator (83). Again, using an exchange experiment with the *QD3-sbl* target, it was found that the ligand primarily binds at the QD interface, most likely intercalating between the G-tetrad and the interfacial base pair (Figure 17D). Because of various exchange cross peaks originating from a single imino proton located at the QD junction, the presence of different complexes with putative flipping of the unsubstituted ligand was suggested. Additionally, the binding constant of cryptolepine to *QD3-sbl* was found to be similar to the binding with the parallel *Q3-sbl* and about five times lower when compared to SYUIQ-5 binding to the *QD3-sbl* hybrid. Consequently, the side chain was suggested to increase both affinity and selectivity. Nevertheless, the results indicated that the recognition of the QD junction is promoted by the indoloquinoline ring system with its matching geometry.

5.4. Factors affecting the selectivity of indoloquinolines upon binding the quadruplex-duplex junctions (Article 8)

As previously reported (section 4.2), different thermodynamic profiles for the binding of PIQ-4m and SYUIQ-5 demonstrate significant differences in QD hybrid selectivity against duplex binding. Obviously, there are additional contributions in the indoloquinoline binding to a QD hybrid. For simplification, NMR binding studies of PIQ-4m were performed with the two-layered G-quadruplex *QD2l* derived from the TBA (84). Of note, ITC studies for PIQ-4m binding to *QD2l* yielded similar thermodynamic profiles of binding when compared to the titration against *QD3-sbl*.

Based on the NMR structural analysis, PIQ-4m also intercalates into the QD junction (PDB 8ABN). In contrast to the SYUIQ-5 binding mode, a 180° flip is observed with the convex part and the out-of-plane phenyl side chain facing the G-tetrad. A less efficient π - π stacking than was observed with SYUIQ-5 could be seen in the final structure. This may explain the less exothermic binding of PIQ-4m to the QD hybrid when compared to SYUIQ-5. A similar shift of the duplex domain towards the center of the G-tetrad can also be seen upon ligand intercalation, particularly for the first cytosine residue of the duplex (Figure 18). Of note, the same arrangement was previously proposed based on an NMR study with a DNA triplex of the homopyrimidine motif. Here, a similar binding geometry of the ligand was observed with the phenyl side chain located in the major groove of the duplex at the triplex-duplex junction (85, 86).

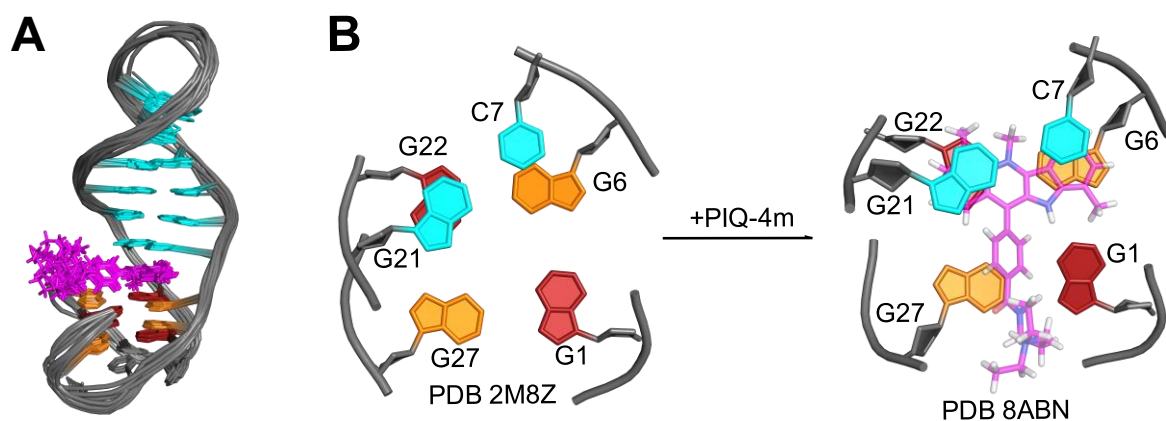
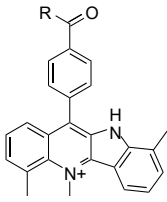
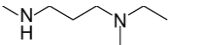
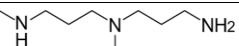
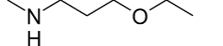


Figure 18. (A) Superposition of 10 lowest-energy structures of PIQ-4m bound to the antiparallel QD hybrid *QD2I*. Residues of the GCA hairpin loop are omitted for clarity. (B) A view of the QD interface highlighting the arrangement of the ligand and the difference between the relative position of the duplex base pair relative to the G-tetrad before (PDB 2M8Z) and after ligand intercalation (PDB 8ABN).

Although there are no obvious interactions of the positively charged amino group in the ligand aliphatic side chain, such additions are usually used to modulate affinity and selectivity towards polyanionic molecules such as DNA. Thus, it has been reported that an increase in the positive charge of the side chain is usually followed by an increase in binding affinity but at the same time also by a decrease in binding selectivity (76, 87). Several PIQ ligands with different aliphatic side chains (76) were therefore studied for their binding to *QD2I* through ITC experiments. It can be expected that the most affine binding site always constitutes a QD junction, while weaker binding sites are likely to involve the duplex domain rather than the opposite tetrad face which is blocked by two lateral loops. These studies revealed trends correlating the side chain structure with the binding affinity for both strong and weak binding sites. Similar affinities were obtained for side chains with one or more positive charges but are reduced for a side chain bearing no additional charge (Table 1). Thus, the addition of more positive charges in the aliphatic side chain does not provide more benefit for the binding affinity towards the QD junction. On the other hand, increasing positive charges in the aliphatic side chain leads to a decrease in selectivity as reflected by smaller difference of binding free energies when binding the QD junction and the duplex domain.

Table 1. ITC-derived standard free energies ΔG° of various PIQ derivatives binding to *QD2I*

					
Name	R	1 st binding site		2 nd binding sites	
		N	ΔG° (kcal/mol)	N	ΔG° (kcal/mol)
PIQ-4m		1.1 ± 0.1	-9.9 ± 0.1	5.0 ± 0.6	-6.2 ± 0.4
PIQ-5m		1.3 ± 0.1	-9.9 ± 0.1	5.0 ± 0.4	-6.7 ± 0.1
PIQ-7m		1.1 ± 0.01	-9.2 ± 0.2	-9.9 ± 0.1	-5.6 ± 0.1

5.5. Phen-DC3 prefers binding to a quadruplex-duplex junction over an open G-tetrad (Article 8)

As indoloquinolines have been shown to prefer intercalation at the QD junction, attempts to target QD hybrids were also made with Phen-DC3 (Figure 19A), being one of the best G-quadruplex binding ligands with good selectivity against binding to a duplex (88, 89). It was suggested that the mirror symmetry of the two quinoline side arms at the phenanthroline ring cover the area of the G-tetrad but are solvent-exposed when phenanthroline intercalates into the duplex. Previously, it was proposed that the Phen-DC3 ligand binds via end-stacking at the open tetrad even in the presence of a QD interface (see above) (40). Such an idea was contested in this study, based on the possibility that a phenanthroline ring may intercalate within the duplex (90).

The low solubility of Phen-DC3 made the ligand not amenable for ITC experiments. Consequently, CD melting experiments were performed on Phen-DC3 binding to QD hybrids, using SYUIQ-5 and PIQ-4m as a reference. However, the melting temperature of the Phen-DC3:*QD3-sbl* complex exceeded the experimental temperature range even in low salt buffer (10 mM K⁺). Therefore, the thermally less stable two-layered *QD2I* system was used as a target (Figure 17). Notably, the binding constant of PIQ-4m to *QD2I* was found to be in the same range (8×10^6 M⁻¹) as that obtained with *QD3-sbl*. An increase of 24 °C for the melting temperature of *QD2I* upon Phen-DC3 complexation was observed, which is about 10 °C higher than that obtained with PIQ-4m. This result suggests a higher binding affinity of Phen-DC3 when compared to PIQ-4m. It is worth noting that the *QD2I* design does not contain an accessible open tetrad, which is commonly considered to be a favored binding site for Phen-DC3.

Subsequently, NMR studies were carried out on *QD3-sbl* featuring a 5'-open tetrad, a 3'-QD interface, and a duplex domain as putative binding sites. Titrations followed by spectral assignments revealed that Phen-DC3 also binds via intercalation at the QD interface (Figure 19B). The phenanthroline ring is sandwiched at the QD junction and the two quinoline side arms extend to cover the rest of the interfacial G-tetrad (Figure 19C). Despite the positive charge of the *N*-methylated quinoline substructure, its participation in electrostatic interaction or hydrogen bonding could not be observed. This suggests that the recognition of the QD interface by Phen-DC3 is mostly derived from the π - π stacking interactions.

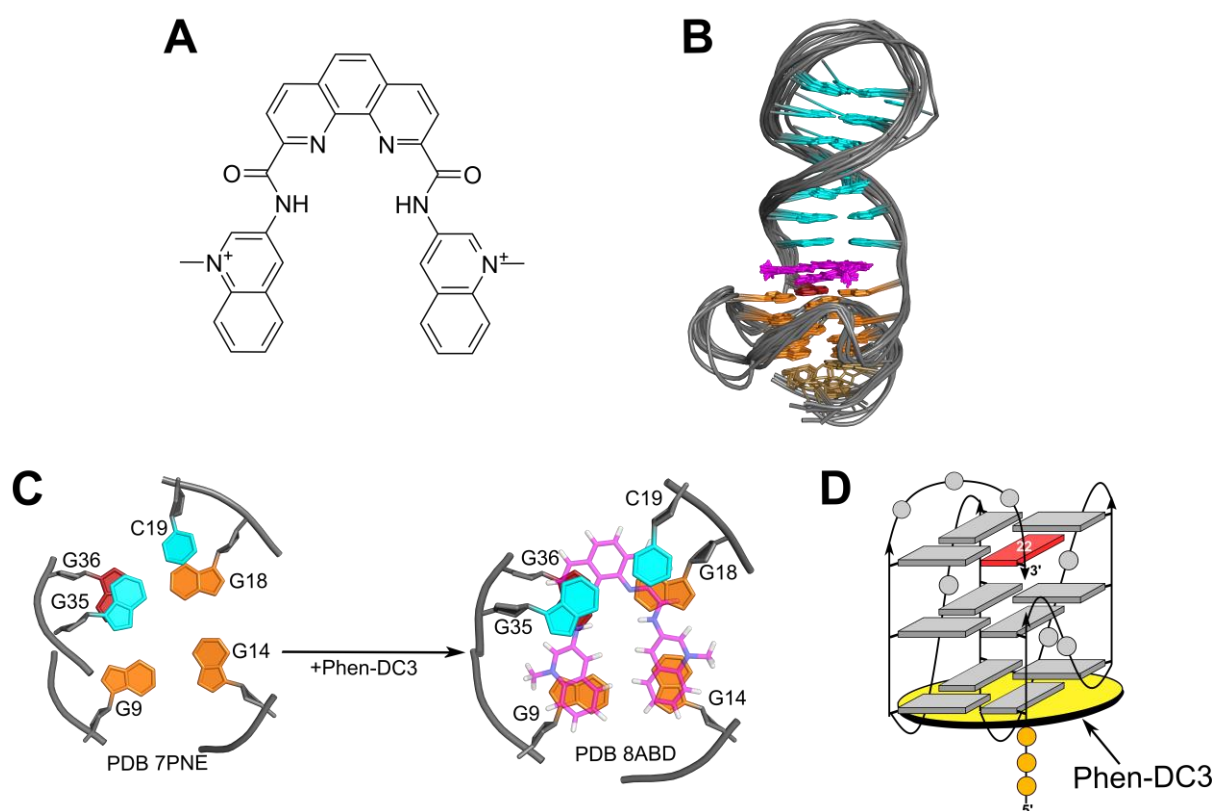


Figure 19. (A) Chemical structure of Phen-DC3. (B) Ten lowest-energy structures of Phen-DC3 bound to *QD3-sbl*. The propeller loop residues and the duplex T₃ stem-loop residues are omitted for clarity. (C) A view of the QD interface highlighting the position of the ligand on the G-tetrad. A shift of the duplex towards the center of the G-tetrad is observed upon ligand intercalation (PDB 8ABD). (D) Proposed binding mode for Phen-DC3 bound to the parallel G-quadruplex *Q3-sbl/2*.

Substitution of the duplex lateral stem loop by a non-duplex forming TGT loop moves the ligand binding to the open opposite outer G-tetrad based on the analysis of NMR experiments with a *Q3-sbl/2* target (Figure 19D). Such binding is in line with the previously reported Phen-DC3 binding mode to a parallel G-quadruplex with a diagonal snapback loop (40). Notably, exchange cross peaks were observed between the symmetry-related quinoline protons of Phen-DC3 when binding the 5'-tetrad of *Q3-sbl/2*. This

is in contrast to Phen-DC3 binding at the QD interface of *QD3-sbl* where no exchange cross peaks could be observed. Exchange cross peaks are likely to result from a release of the ligand, followed by reassociation with a flip orientation. Obviously, such a process is more restricted upon intercalation at the QD interface. Taken together, these results indicate that the QD interface again provides a more affine binding site for Phen-DC3 when compared to binding at the open G-tetrad.

5.6 General insights into ligand binding at quadruplex-duplex junctions (Articles 7 & 8)

A QD interface can be formed in a duplex lateral loop coaxially stacked on a G-tetrad plane. The present studies suggest that a QD interface forms a most favorable binding site when compared to a commonly perceived site for a typical G-quadruplex binder such as an open G-tetrad (15). Instead of blocking ligand access, a QD junction may recruit G-quadruplex ligands with an intercalator motif to intercalate at the QD interface. Such a binding mode provides the most favorable free energy compared to other possible ligand-DNA binding modes, highlighting that the QD junction can be considered a binding hotspot.

Upon ligand intercalation, a conformational readjustment is observed, reflected by a shift of the duplex domain towards the center of the G-tetrad. The most affected shift tends to occur at the 5'-terminal residue of the stacked duplex domain adjacent to the G-column. Interestingly, duplex intercalation was previously reported to be associated with an increase in the backbone torsion angle β and the glycosidic torsion angle χ , the latter being shifted toward the high-anti range for the 3'-position residue of a model dinucleotide duplex (91). The present results suggest a significant increase in the χ angle of the first residue cytosine of the duplex following the G-column. Such a χ angle in the high-anti region is rather rare for pyrimidine nucleosides due to a closer CH6-H2' protons with a potential steric clash.

For a QD intercalation to happen, a typical intercalator motif should be used as core of the ligand. Using ligand intercalation into B-DNA as a general model for ligand intercalation, DNA unwinding followed by base pair unstacking must occur to accommodate subsequent ligand insertion. These processes are usually initiated by ligand binding to the DNA grooves. However, the exact binding mechanisms are dependent on the ligand structure and the type of base pairs (92–94). The energy required is compensated by a more favorable energy as a result of ligand insertion. Provided that the ligand has a suitable geometry for intercalation (95), a typical DNA intercalator can be proposed to readily bind and insert at the QD junction.

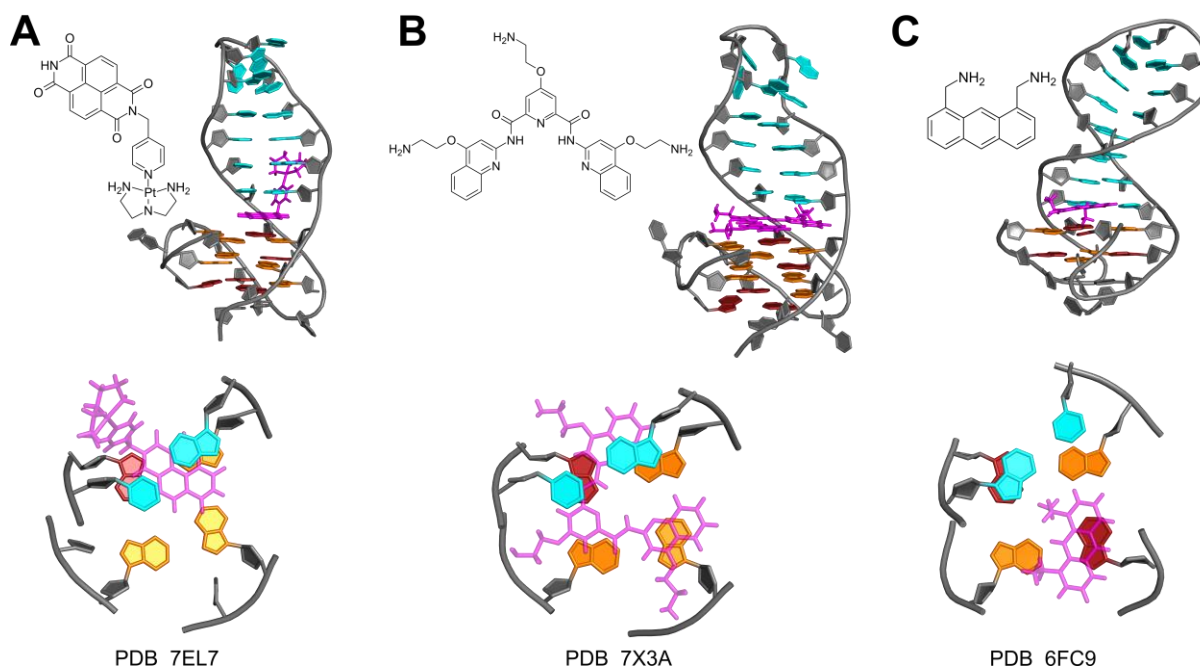


Figure 20. Reported QD hybrids and their interactions with (A) a naphthalene diimide derivative (PDB 7EL7), (B) pyridostatin (PDB 7X3A), and (C) an anthracene derivative (PDB 6FC9). (Top) Chemical structure of the ligands and the representative complexes, and (bottom) view of the QD interface with corresponding ligand arrangements.

Several other ligands have been reported to also exhibit similar binding modes in the presence of a QD hybrid. A naphthalene diimide (NDI) based ligand linked to a positively charged tertiary platinum coordinated complex [Pt-(dien)-(py)] was found to intercalate at the QD junction with a similar binding mode as SYUIQ-5 (PDB 7PNG). The addition of a methyl-containing side chain allowed the formation of additional interactions within the duplex minor groove (Figure 20A) (62). For a pyridostatin ligand (PDS) and one of its derivatives, a behavior similar to Phen-DC3 was found. In contrast to Phen-DC3, it is rather the quinoline ring of PDS that intercalates into the QD junction due to the larger π - π stacking interactions of the quinoline moiety when compared to the pyridine ring of PDS (Figure 20B) (96).

On the other hand, failure of intercalation was observed with an anthracene-based ligand with simple alkylamino substituents (Figure 20C) (97). It can be expected that intercalation of simple aromatic hydrocarbons may prove difficult because the energy provided by π - π stacking interactions may not be sufficient to overcome the activation energy of base unstacking at the QD junction. A simple aromatic hydrocarbon is also more electron-rich with no major polarizing effects, resulting in weaker stacking energies. Structurally, the stacking of the anthracene-based ligand on the accessible open part of the tetrad associated with electrostatic interactions between the positively charged methylamino substituent and the duplex major groove is reminiscent of typical G-quadruplex ligands that stack on an outer G-

tetrad, forming additional interactions with the overhang/loop residues. In yet another study, it is a triad rather than a base pair stacking onto a G-tetrad to prevent ligand intercalation due to a larger energetic penalty for forming an intercalation pocket (58).

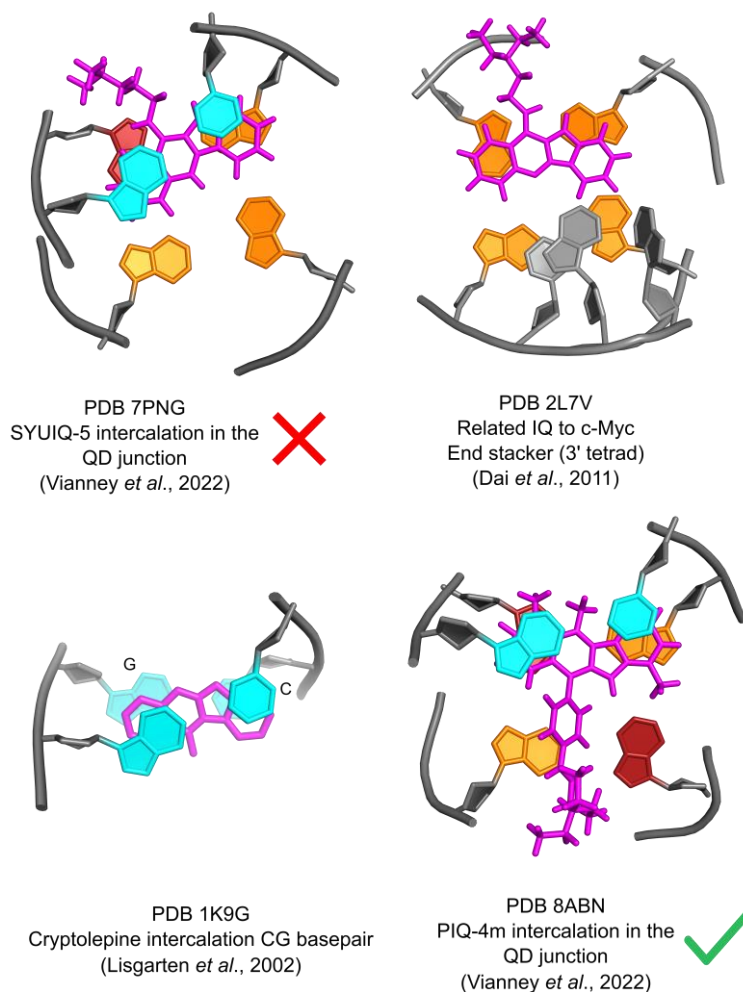


Figure 21. Structures of indoloquinolines bound to QD hybrids, to a parallel G-quadruplex, and to a duplex. SYUIQ-5 is proposed to be a non-selective ligand (marked with the cross). The ligand PIQ with a side chain that is not able to fold back into the duplex minor groove but rather covers the rest of the G-tetrad at the QD interface is proposed to be a selective QD binder (symbolized with the check mark). *Syn*- and *anti*-guanosines are colored in red and orange, respectively, while duplex bases are colored in blue. The ligands are colored magenta.

Although the QD junction can be considered a binding hotspot, the question of selectivity against duplex intercalation must be raised. In fact, the excess number of binding sites for duplex intercalation following nearest-neighbor exclusion makes it important to evaluate the relative binding constants. It was found that the binding constant of SYUIQ-5 towards the QD junction is only about 5 times higher than the

one for duplex binding and almost the same when binding an open tetrad, raising doubt on the use of simple indoloquinolines for G-quadruplex targeting. In contrast, a much more significant selectivity against a duplex was observed for the phenyl-substituted PIQ. Such behavior can be rationalized by structural considerations (Figure 21).

An indoloquinoline related to SYUIQ-5 was previously found to end-stack on a parallel *c-Myc* in both its' accessible 5'- and 3'-tetrad (80). The arrangement of the indoloquinoline ring system on the 3'-tetrad was similar relative to the binding of SYUIQ-5 at the QD interface. The convex side of the indoloquinoline faces outwards from the G-tetrad, positioning the aliphatic side chain towards the groove of the G-quadruplex. On the other hand, an intercalated unsubstituted cryptolepine in a poly-GC DNA duplex was also found to be in a similar orientation with the convex side of the cryptolepine facing the duplex minor groove (83). Intuitively, a similar binding mode and perhaps a nearly similar binding energy of SYUIQ-5 towards the duplex and the QD junction can be perceived and should result in reduced selectivity. In contrast, for PIQ it is rather the convex face that points towards the center of the G-tetrad. The phenyl group with its restricted flexibility cannot efficiently fold back into the duplex minor groove. Thus, one might expect an entropic penalty for duplex intercalation due to an exposed phenyl-containing side chain. In fact, the idea of extending the indoloquinoline by a phenyl substituent to cover the rest of the open QD interface seems similar to the structure of Phen-DC3 with its two quinoline side arms (Figure 19C). The lower binding enthalpy of PIQ-4m when compared to SYUIQ-5 upon QD intercalation (Figure 16) may result from a less efficient stacking of the former. However, a high exothermic binding of SYUIQ-5 is also expected upon duplex intercalation, reducing its selectivity. Apparently, it is not the enthalpic contributions but rather the entropic contributions that promote the binding selectivity for the QD junction over the duplex in these cases.

6. Conclusions and perspectives

G-quadruplexes are nucleic acid secondary structures with a wide variety of folding topologies that are still hard to predict. Clearly, G-rich sequences from the genome can be studied by NMR methods to derive formed G-quadruplex topologies, as has been done for the parallel G-quadruplex of WSSV or the V-loop G-quadruplex of the *KCNN4* sequence. However, a complete overview of all critical factors that determine the fate of the folding is limited due to the impact of flanking sequences, loop lengths, and loop compositions.

For technological applications but also a better understanding of G-quadruplex folding, various canonical and non-canonical G-quadruplexes have been designed in the past. Here, several folding rules, together with the use of unique G-quadruplex motifs, can be employed to engineer a particular G-quadruplex fold. In this regard, engineered model systems can explore the topological space of G-quadruplexes, but more importantly, can deepen the understanding of G-quadruplex end-folding as well as ligand binding in terms of structures and thermodynamics. Therefore, G-quadruplexes based on a “design it yourself” methodology are excellent for extracting fundamental rules of G-quadruplex folding, rather than just solving the structure of a given sequence. For example, a better understanding of the well-known rule that 1-nt propeller loops prefer a parallel topology was extracted by engineering a (3+1) hybrid topology containing all-homopolar tetrad stacking (hybrid-1R). Specifically, a 1-nt propeller loop prefers to connect two guanines in outer tetrads of the same polarity. The positional dependence of the loop length still needs to be considered, based on the fact that more resistance to forming a hybrid-2R topology was observed compared to forming a hybrid-1R topology.

In addition, QD interfaces may be abundant in genomes due to a lateral-type duplex stem-loop stacking on the G-quadruplex core. Such a QD hybrid structure is expected to stabilize the G-quadruplex domain when compared to an unordered long loop segment. Such a QD interface may also provide a unique ligand binding site. The binding mode of the ligands to the QD interface has been shown to consist of a typical intercalation with π - π stacking interactions as the main driving forces. A ligand side chain, on the other hand, not only restricts the dynamics or changes the affinity towards the QD hybrids but may also modify the selectivity against a duplex. Better selectivity can be achieved by incorporating a side chain that prefers to cover the G-tetrad instead of interacting with the duplex groove, as demonstrated by the PIQ phenyl ring or the quinoline side arms of Phen-DC3. Notably, it has been suggested that selectivity can be promoted by more specific intermolecular interactions that are enthalpic in origin. However, binding of the SYUIQ-5 ligand to both a QD junction and to a duplex has been shown to be significantly exothermic, resulting in a poor binding selectivity of the QD hybrid over the duplex. On the

other hand, the more selective PIQ-4m ligand positions its side chain on the open part of the interfacial tetrad, resulting in an entropic gain when compared to the duplex binding.

Higher selectivity for the QD junction over a free G-tetrad is mostly achieved through more favorable π - π stacking interactions. For instance, Phen-DC3 favors intercalation at the QD junction rather than stacking on the oppositely located free tetrad. In case where extended aromatic surface areas disfavor ligand intercalation between duplex base pairs, further gain in QD selectivity but also affinity may be achieved by adding another side chain that further interacts within the duplex groove.

Clearly, much work remains to be done on the targeting of QD junctions with small molecules. In the present studies, two G-quadruplex topologies, i.e., parallel and antiparallel G-quadruplexes were used as model QD hybrids. Typically, both designs showed a favored ligand binding towards the QD junction. It should be mentioned that the exact binding mechanism remains elusive in many cases. In the case of duplex binding, the types of base pairs could influence the degree of intercalation via unique groove interactions that initiate the intercalation. Also, the ligand geometry will also influence the additional stacking energy upon its binding. Different types and numbers of base pairs stacked on a G-tetrad could influence the degree of intercalation at the junction. As the QD junction is a binding hotspot, it will be interesting to see if QD binding ligands can be used to guide the folding of a G-rich sequence into a certain G-quadruplex topology that also includes a QD interface. In addition, the incorporation of a QD junction motif on an aptamer, or the specific targeting of a QD junction in a QD hybrid design by a ligand may be of significant interest for future biological and technological applications.

Bibliography

1. Schultze,P., Hud,N.V., Smith,F.W. and Feigon,J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d(G4T4G4). *Nucleic Acids Res.*, **27**, 3018–3028.
2. 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.
3. Strahan,G.D., Keniry,M.A. and Shafer,R.H. (1998) NMR structure refinement and dynamics of the K⁺-[d(G3T4G3)]₂ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.*, **75**, 968–981.
4. Webba Da Silva,M. (2007) Geometric formalism for DNA quadruplex folding. *Chem. Eur. J.*, **13**, 9738–9745.
5. Mohr,S., Jana,J., Vianney,Y.M. and Weisz,K. (2021) Expanding the topological landscape by a G-column flip of a parallel G-quadruplex. *Chem. Eur. J.*, **27**, 10437–10447.
6. Karg,B., Mohr,S. and Weisz,K. (2019) Duplex-guided refolding into novel G-quadruplex (3+1) hybrid conformations. *Angew. Chem. Int. Ed.*, **58**, 11068–11071.
7. Jana,J., Vianney,Y.M., Schröder,N. and Weisz,K. (2022) Guiding the folding of G-quadruplexes through loop residue interactions. *Nucleic Acids Res.*, **50**, 7161–7175.
8. Marušič,M., Šket,P., Bauer,L., Viglasky,V. and Plavec,J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.*, **40**, 6946–6956.
9. Hänsel-Hertsch,R., di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
10. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
11. Balasubramanian,S. and Neidle,S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
12. Travascio,P., Li,Y. and Sen,D. (1998) DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chem. Biol.*, **5**, 505–517.
13. Xiao,Y., Pavlov,V., Gill,R., Bourenko,T. and Willner,I. (2004) Lighting up biochemiluminescence by the surface self-assembly of DNA - Hemin complexes. *ChemBioChem*, **5**, 374–379.
14. Xiao,Y., Pavlov,V., Niazov,T., Dishon,A., Kotler,M. and Willner,I. (2004) Catalytic beacons for the detection of DNA and telomerase activity. *J. Am. Chem. Soc.*, **126**, 7430–7431.
15. Santos,T., Salgado,G.F., Cabrita,E.J. and Cruz,C. (2021) G-Quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions. *Pharmaceuticals*, **14**, 769.
16. Kettani,A., Gorin,A., Majumdar,A., Hermann,T., Skripkin,E., Zhao,H., Jones,R. and Patel,D.J. (2000) A dimeric DNA interface stabilized by stacked A·(G·G·G·G)·A hexads and coordinated monovalent cations. *J. Mol. Biol.*, **297**, 627–644.
17. Bakalar,B., Heddi,B., Schmitt,E., Mechulam,Y. and Phan,A.T. (2019) A minimal sequence for left-handed G-quadruplex formation. *Angew. Chem. Int. Ed.*, **58**, 2331–2335.
18. 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.

19. Winnerdy, F.R., Bakalar, B., Maity, A., Vandana, J.J., Mechulam, Y., Schmitt, E. and Phan, A.T. (2019) NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res.*, **47**, 8272–8281.
20. Brčić, J. and Plavec, J. (2018) NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism. *Nucleic Acids Res.*, **46**, 11605–11617.
21. Brčić, J. and Plavec, J. (2015) Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Res.*, **43**, 8590–8600.
22. Devi, G., Winnerdy, F.R., Ang, J.C.Y., Lim, K.W. and Phan, A.T. (2022) Four-layered intramolecular parallel G-quadruplex with non-nucleotide loops: an ultra-stable self-folded DNA nano-scaffold. *ACS Nano*, **16**, 533–540.
23. Heddi, B., Martín-Pintado, N., Serimbetov, Z., Kari, T.M.A. and Phan, A.T. (2016) G-quadruplexes with $(4n - 1)$ guanines in the G-tetrad core: Formation of a G-triad·water complex and implication for small-molecule binding. *Nucleic Acids Res.*, **44**, 910–916.
24. Wang, K.B., Dickerhoff, J., Wu, G. and Yang, D. (2020) PDGFR- β promoter forms a vacancy G-quadruplex that can be filled in by dGMP: solution structure and molecular recognition of guanine metabolites and drugs. *J. Am. Chem. Soc.*, **142**, 5204–5211.
25. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.
26. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract g-quadruplex structure from the human *MYC* promoter. *Nat. Chem. Biol.*, **1**, 167–173.
27. Kuryavyi, V. and Patel, D.J. (2010) Solution Structure of a Unique G-Quadruplex Scaffold Adopted by a Guanosine-Rich Human Intronic Sequence. *Structure*, **18**, 73–82.
28. Phan, A.T., Kuryavyi, V., Ma, J.-B., Faure, A., Andreola, M.-L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci.*, **102**, 634–639.
29. Chen, Y., Agrawal, P., Brown, R. v., Hatzakis, E., Hurley, L. and Yang, D. (2012) The major G-quadruplex formed in the human platelet-derived growth factor receptor β promoter adopts a novel broken-strand structure in K^+ solution. *J. Am. Chem. Soc.*, **134**, 13220–13223.
30. Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E. and Patel, D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A·(G·G·G·G) pentad-containing dimeric DNA quadruplex involving stacked G(*anti*)·G(*anti*)·G(*anti*)·G(*syn*) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
31. Maity, A., Winnerdy, F.R., Chang, W.D., Chen, G. and Phan, A.T. (2020) Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.*, **48**, 3315–3327.
32. Haase, L., Dickerhoff, J. and Weisz, K. (2020) Sugar puckering drives G-quadruplex refolding: implications for V-shaped loops. *Chem. Eur. J.*, **26**, 524–533.
33. Haase, L. and Weisz, K. (2020) Locked nucleic acid building blocks as versatile tools for advanced G-quadruplex design. *Nucleic Acids Res.*, **48**, 10555–10566.
34. Marušič, M. and Plavec, J. (2019) Towards understanding of polymorphism of the g-rich region of human papillomavirus type 52. *Molecules*, **24**, 1294.
35. Vianney, Y.M. and Weisz, K. (2021) First tandem repeat of a potassium channel KCNN4 minisatellite folds into a V-loop G-quadruplex structure. *Biochemistry*, **60**, 1337–1346.

36. Butovskaya,E., Heddi,B., Bakalar,B., Richter,S.N. and Phan,A.T. (2018) Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc.*, **140**, 13654–13662.
37. Črnugelj,M., Šket,P. and Plavec,J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.
38. Winnerdy,F.R., Das,P., Heddi,B. and Phan,A.T. (2019) Solution structures of a G-quadruplex bound to linear- and cyclic-dinucleotides. *J. Am. Chem. Soc.*, **141**, 18038–18047.
39. Tan,D.J.Y., Das,P., Winnerdy,F.R., Lim,K.W. and Phan,A.T. (2020) Guanine anchoring: A strategy for specific targeting of a G-quadruplex using short PNA, LNA and DNA molecules. *Chem. Commun*, **56**, 5897–5900.
40. Chung,W.J., Heddi,B., Hamon,F., Teulade-Fichou,M.P. and Phan,A.T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chem. Int. Ed.*, **53**, 999–1002.
41. Schnarr,L., Jana,J., Preckwinkel,P. and Weisz,K. (2020) Impact of a snap-back loop on stability and ligand binding to a parallel G-quadruplex. *J. Phys. Chem. B*, **124**, 2778–2787.
42. Agrawal,P., Hatzakis,E., Guo,K., Carver,M. and Yang,D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.*, **41**, 10584–10592.
43. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
44. Rachwal,P.A., Findlow,I.S., Werner,J.M., Brown,T. and Fox,K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res*, **35**, 4214–4222.
45. Marušič,M. and Plavec,J. (2015) The effect of DNA sequence directionality on G-quadruplex folding. *Angew. Chem. Int. Ed.*, **54**, 11716–11719.
46. Jana,J. and Weisz,K. (2021) Thermodynamic stability of G-quadruplexes: impact of sequence and environment. *ChemBioChem*, **22**, 2848–2856.
47. Jana,J. and Weisz,K. (2020) A thermodynamic perspective on potential G-quadruplex structures as silencer elements in the MYC promoter. *Chem. Eur. J.*, **26**, 17242–17251.
48. Cang,X., Šponer,J. and Cheatham,T.E. (2011) Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.*, **39**, 4499–4512.
49. Chen,J., Cheng,M., Salgado,G.F., Stadlbauer,P., Zhang,X., Amrane,S., Guédin,A., He,F., Šponer,J., Ju,H., *et al.* (2021) The beginning and the end: Flanking nucleotides induce a parallel G-quadruplex topology. *Nucleic Acids Res.*, **49**, 9548–9559.
50. Guédin,A., Alberti,P. and Mergny,J.L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.
51. Sattin,G., Artese,A., Nadai,M., Costa,G., Parrotta,L., Alcaro,S., Palumbo,M. and Richter,S.N. (2013) Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One*, **8**, e84113.

52. Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
53. Liu,C., Zhou,B., Geng,Y., Yan Tam,D., Feng,R., Miao,H., Xu,N., Shi,X., You,Y., Hong,Y., *et al.* (2019) A chair-type G-quadruplex structure formed by a human telomeric variant DNA in K⁺ solution. *Chem. Sci.*, **10**, 218–226.
54. Cang,X., Šponer,J. and Cheatham,T.E. (2011) Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J. Am. Chem. Soc.*, **133**, 14270–14279.
55. Smargiasso,N., Rosu,F., Hsia,W., Colson,P., Baker,E.S., Bowers,M.T., de Pauw,E. and Gabelica,V. (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, **130**, 10208–10216.
56. Macaya,R.F., Waldron,J.A., Beutel,B.A., Gao,H., Joesten,M.E., Yang,M., Patel,R., Bertelsen,A.H. and Cook,A.F. (1995) Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs. *Biochemistry*, **34**, 4478–4492.
57. Lim,K.W. and Phan,A.T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem. Int. Ed.*, **52**, 8566–8569.
58. Russo Krauss,I., Ramaswamy,S., Neidle,S., Haider,S. and Parkinson,G.N. (2016) Structural insights into the quadruplex-duplex 3' interface formed from a telomeric repeat: a potential molecular target. *J. Am. Chem. Soc.*, **138**, 1226–1233.
59. Lim,K.W., Khong,Z.J. and Phan,A.T. (2014) Thermal stability of DNA quadruplex-duplex hybrids. *Biochemistry*, **53**, 247–257.
60. Lim,K.W., Jenjaroenpun,P., Low,Z.J., Khong,Z.J., Ng,Y.S., Kuznetsov,V.A. and Phan,A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630–5646.
61. Tan,D.J.Y., Winnerdy,F.R., Lim,K.W. and Phan,A.T. (2020) Coexistence of two quadruplex – duplex hybrids in the *PIM1* gene. *Nucleic Acids Res.*, **48**, 11162–11171.
62. Liu,L.Y., Wang,K.N., Liu,W., Zeng,Y.L., Hou,M.X., Yang,J. and Mao,Z.W. (2021) Spatial matching selectivity and solution structure of organic–metal hybrid to quadruplex–duplex hybrid. *Angew. Chem. Int. Ed.*, **60**, 20833–20839.
63. Monsen,R.C., Chua,E.Y.D., Hopkins,J.B., Chaires,J.B. and Trent,J.O. (2023) Structure of a 28.5 kDa duplex-embedded G-quadruplex system resolved to 7.4 Å resolution with cryo-EM. *Nucleic Acids Res.*, **51**, 1943–1959.
64. Nguyen,T.Q.N., Lim,K.W. and Phan,A.T. (2020) Folding kinetics of G-quadruplexes: duplex stem loops drive and accelerate G-quadruplex folding. *J Phys. Chem. B*, **124**, 5122–5130.
65. Zhou,J., Bourdoncle,A., Rosu,F., Gabelica,V. and Mergny,J.L. (2012) Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed.*, **51**, 11002–11005.
66. Russo Krauss,I., Spiridonova,V., Pica,A., Napolitano,V. and Sica,F. (2016) Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res*, **44**, 983–991.
67. Deng,M., Zhang,D., Zhou,Y. and Zhou,X. (2008) Highly effective colorimetric and visual detection of nucleic acids using an asymmetrically split peroxidase DNAzyme. *J. Am. Chem. Soc.*, **130**, 13095–13102.

68. Kong,D.M., Wang,N., Guo,X.X. and Shen,H.X. (2010) 'Turn-on' detection of Hg²⁺ ion using a peroxidase-like split G-quadruplex-hemin DNAzyme. *Analyst*, **135**, 545–549.
69. He,H.Z., Chan,D.S.H., Leung,C.H. and Ma,D.L. (2012) A highly selective G-quadruplex-based luminescent switch-on probe for the detection of gene deletion. *Chem. Commun*, **48**, 9462–9464.
70. 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.
71. Zaccaria,F., Paragi,G. and Fonseca Guerra,C. (2016) The role of alkali metal cations in the stabilization of guanine quadruplexes: why K⁺ is the best. *Phys. Chem. Chem. Phys.*, **18**, 20895–20904.
72. Neidle,S. (2016) Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.*, **59**, 5987–6011.
73. Nguyen,T.Q.N., Lim,K.W. and Phan,A.T. (2017) A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs. *Sci. Rep.*, **7**, 11969.
74. Asamitsu,S., Obata,S., Phan,A.T., Hashiya,K., Bando,T. and Sugiyama,H. (2018) Simultaneous binding of hybrid molecules constructed with dual DNA-binding components to a G-quadruplex and its proximal duplex. *Chem. Eur. J.*, **24**, 4428–4435.
75. Funke,A., Dickerhoff,J. and Weisz,K. (2016) Towards the development of structure-selective G-quadruplex-binding indolo[3,2-b]quinolines. *Chem. Eur. J.*, **22**, 3170–3181.
76. Funke,A. and Weisz,K. (2019) Thermodynamic signature of indoloquinolines interacting with G-quadruplexes: Impact of ligand side chain. *Biochimie*, **157**, 142–148.
77. Le,V.H., Buscaglia,R., Chaires,J.B. and Lewis,E.A. (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal. Biochem.*, **434**, 233–241.
78. Funke,A. and Weisz,K. (2017) Comprehensive thermodynamic profiling for the binding of a G-quadruplex selective indoloquinoline. *J. Phys. Chem. B*, **121**, 5735–5743.
79. Ou,T.M., Lin,J., Lu,Y.J., Hou,J.Q., Tan,J.H., Chen,S.H., Li,Z., Li,Y.P., Li,D., Gu,L.Q., *et al.* (2011) Inhibition of cell proliferation by quindoline derivative (SYUIQ-05) through its preferential interaction with c-myc promoter G-quadruplex. *J. Med. Chem.*, **54**, 5671–5679.
80. Dai,J., Carver,M., Hurley,L.H. and Yang,D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.*, **133**, 17673–17680.
81. Read,M., Harrison,R.J., Romagnoli,B., Tanious,F.A., Gowan,S.H., Reszka,A.P., Wilson,W.D., Kelland,L.R. and Neidle,S. (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 4844–4849.
82. Moore,M.J.B., Schultes,C.M., Cuesta,J., Cuenca,F., Gunaratnam,M., Tanious,F.A., Wilson,W.D. and Neidle,S. (2006) Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J. Med. Chem.*, **49**, 582–599.
83. Lisgarten,J.N., Coll,M., Portugal,J., Wright,C.W. and Aymami,J. (2002) The antimalarial and cytotoxic drug cryptolepine intercalates into dna at cytosine-cytosine sites. *Nat. Struct. Biol.*, **9**, 57–60.
84. Macaya,R.F., Schultze,P., Smith,F.W., Roe,J.A. and Feigon,J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 3745–3749.

85. Eick,A., Riechert-Krause,F. and Weisz,K. (2012) Binding and NMR structural studies on indoloquinoline-oligonucleotide conjugates targeting duplex DNA. *Bioconjug. Chem*, **23**, 1127–1137.
86. Dickerhoff,J., Riechert-Krause,F., Seifert,J. and Weisz,K. (2014) Exploring multiple binding sites of an indoloquinoline in triple-helical DNA: A paradigm for DNA triplex-selective intercalators. *Biochimie*, **107**, 327–337.
87. Zuffo,M., Guédin,A., Leriche,E.D., Doria,F., Pirola,V., Gabelica,V., Mergny,J.L. and Freccero,M. (2018) More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.*, **46**, e115.
88. de Cian,A., DeLemos,E., Mergny,J.L., Teulade-Fichou,M.P. and Monchaud,D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.*, **129**, 1856–1857.
89. Tran,P.L.T., Largy,E., Hamon,F., Teulade-Fichou,M.P. and Mergny,J.L. (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie*, **93**, 1288–1296.
90. Gabbay,E.J., Scofield,R.E. and Baxter,C.S. (1973) Topography of nucleic acid helices in solution. XXIX. steric effects on the intercalation of aromatic cations to deoxyribonucleic acid. *J. Am. Chem. Soc.*, **95**, 7850–7857.
91. Shieh,H.S., Berman,H.M., Dabrow,M. and Neidle,S. (1980) The structure of drug-deoxynucleoside phosphate complex: generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.*, **8**, 85–98.
92. Ramstein,J., Dourlent,M. and Leng,M. (1972) Interaction between proflavine and deoxyribonucleic acid influence of DNA base composition. *Biochem. Biophys. Res. Commun.*, **47**, 874–882.
93. Chaires,J.B., Dattagupta,N. and Crothers,D.M. (1985) Kinetics of the daunomycin-DNA interaction. *Biochemistry*, **24**, 260–267.
94. Wilhelm,M., Mukherjee,A., Bouvier,B., Zakrzewska,K., Hynes,J.T. and Lavery,R. (2012) Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.*, **134**, 8588–8596.
95. Wang,K.B., Elsayed,M.S.A., Wu,G., Deng,N., Cushman,M. and Yang,D. (2019) Indenoisoquinoline topoisomerase inhibitors strongly bind and stabilize the MYC promoter G-quadruplex and downregulate MYC. *J. Am. Chem. Soc.*, **141**, 11059–11070.
96. Liu,L.Y., Ma,T.Z., Zeng,Y.L., Liu,W. and Mao,Z.W. (2022) Structural basis of pyridostatin and its derivatives specifically binding to G-quadruplexes. *J. Am. Chem. Soc.*, **144**, 11878–11887.
97. Díaz-Casado,L., Serrano-Chacón,I., Montalvillo-Jiménez,L., Corzana,F., Bastida,A., Santana,A.G., González,C. and Asensio,J.L. (2020) De novo design of selective quadruplex–duplex junction ligands and structural characterisation of their binding mode: targeting the G4 hot-spot. *Chem. Eur. J.*, **27**, 6204–6212.

Author Contributions

Article 1: Structural motifs and intramolecular interactions in non-canonical G-quadruplexes

Jana, J., Mohr, S., Vianney, Y. M., & Weisz, K. (2021). *RSC Chem. Biol.*, 2 338-353.

JJ, SM, YMV, and KW equally contributed to the layout and writing of the review.

Article 2: G-Quadruplex formation in a putative coding region of white spot syndrome virus: structural and thermodynamic aspects

Vianney, Y. M., Purwanto, M. G. M., & Weisz, K. (2021). *ChemBioChem*, 22, 1932-1935.

KW and MGMP initiated the project. YMV performed the experiments and analyzed the data. YMV and KW wrote the manuscript. All authors agreed on the final version of the manuscript.

Article 3: First tandem repeat of a potassium channel KCNN4 minisatellite folds into a V-Loop G-quadruplex structure

Vianney, Y. M., & Weisz, K. (2021). *Biochemistry*, 60, 1337-1346.

KW initiated the project. YMV performed the experiments and analyzed the data. YMV and KW wrote the manuscript.

Article 4: Expanding the topological landscape by a G-column flip of a parallel G-quadruplex

Mohr, S., Jana, J., Vianney, Y. M., & Weisz, K. (2021). *Chem. Eur. J.*, 27, 10437-10447.

KW initiated the project. SM initiated the work. SM, JJ, and YMV performed the experiments. KW and SM wrote the manuscript. All authors agreed on the final version of the manuscript. This article has also been part of the Master Thesis of SM at the Institute of Biochemistry, Universität Greifswald.

Article 5: Guiding the folding of G-quadruplexes through loop residue interactions

Jana, J., Vianney, Y. M., Schröder, N., & Weisz, K. (2022). *Nucleic Acids Res.*, 50(12), 7161-7175.

KW initiated the project. JJ initiated the work. JJ, YMV, and NS performed the experiments. KW, JJ, YMV, and NS analyzed the data. KW, JJ, and YMV wrote the manuscript. All authors agreed on the final version of the manuscript.

Article 6: Quadruplex-duplex junction : a high-affinity binding site for indoloquinoline ligands

Vianney, Y. M., Preckwinkel, P., Mohr, S., & Weisz, K. (2020). *Chem. Eur. J.*, 26, 16910-16922.

KW initiated the project. PP and SM initiated the work. YMV, PP, and SM performed the experiments. YMV and KW wrote the manuscript. All authors agreed with the final version of the manuscript. Part of this manuscript is included in the Bachelor Thesis of PP at the Institute of Biochemistry, Universität Greifswald.

Article 7: Indoloquinoline ligands favor intercalation at quadruplex-duplex interfaces

Vianney, Y. M., & Weisz, K. (2022). *Chem. Eur. J.*, 28, e202103718.

KW initiated the project. YMV performed the experiments and analyzed the data. YMV and KW wrote the manuscript.

Article 8: High-affinity binding at quadruplex-duplex junctions: rather the rule than the exception

Vianney, Y. M., & Weisz, K. (2022). *Nucleic Acids Res.*, 50, 11948-11964.

KW initiated the project. YMV performed the experiments and analyzed the data. YMV and KW wrote the manuscript.

Prof. Dr. Klaus Weisz

Lie, Yoanes Maria Vianney

Article 1



Cite this: *RSC Chem. Biol.*, 2021, 2, 338

Received 18th November 2020,
Accepted 14th January 2021

DOI: 10.1039/d0cb00211a

rsc.li/rsc-chembio

Structural motifs and intramolecular interactions in non-canonical G-quadruplexes

Jagannath Jana, Swantje Mohr, Yoanes Maria Vianney and Klaus Weisz *

Guanine(G)-rich DNA or RNA sequences can assemble or intramolecularly fold into G-quadruplexes formed through the stacking of planar G-G-G-G tetrads in the presence of monovalent cations. These secondary nucleic acid structures have convincingly been shown to also exist within a cellular environment exerting important regulatory functions in physiological processes. For identifying nucleic acid segments prone to quadruplex formation, a putative quadruplex sequence motif encompassing closely spaced tracts of three or more guanines is frequently employed for bioinformatic search algorithms. Depending on the number and type of intervening residues as well as on solution conditions, such sequences may fold into various canonical G4 topologies with continuous G-columns. On the other hand, a growing number of sequences capable of quadruplex formation feature G-deficient guanine tracts, escaping the conservative consensus motif. By folding into non-canonical quadruplex structures, they adopt unique topologies depending on their specific sequence context. These include G-columns with only two guanines, bulges, snapback loops, D- and V-shaped loops as well as interlocked structures. This review focuses on G-quadruplex species carrying such distinct structural motifs. It evaluates characteristic features of their non-conventional scaffold and highlights principles of stabilizing interactions that also allow for their folding into stable G-quadruplex structures.

Introduction

Single-stranded guanine-rich DNA or RNA sequences can fold into intramolecular or intermolecular four-stranded structures called G-quadruplexes (G4s). G4-prone motifs are found in high numbers not only in bacterial and viral, but also in human genomes.

Institute of Biochemistry, Universität Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany. E-mail: weisz@uni-greifswald.de; Fax: +49 3834 420-4427; Tel: +49 3834 420-4426



Jagannath Jana

Jagannath Jana obtained his MSc degree in Chemistry from Vidyasagar University (India). He received his PhD in 2017 from Bose Institute (Calcutta University, Kolkata, India) under the supervision of Dr Subhrangsu Chatterjee where he designed peptides and small molecules as potent stabilizers of G-quadruplex structures. Subsequently, he was a postdoctoral researcher (2018–2019) at Institut Curie (Paris, France) in the laboratory of

Professor Stéphane Vagner. Currently he is working as a postdoc with Prof. Klaus Weisz at the Institute of Biochemistry, Universität Greifswald (Germany). His research interests include thermodynamic and structural studies on G-quadruplexes.



Swantje Mohr

Swantje Mohr received her BSc degree in Biochemistry at the Universität Greifswald (Germany) in 2018. Her thesis in biophysical chemistry focused on the modeling of organic solvents in molecular simulations. She subsequently moved into the field of nucleic acids, working on the structure determination by NMR techniques in the Analytical Biochemistry lab of Prof. Klaus Weisz. She recently finished her MSc thesis project on the refolding of G-quadruplexes.



Thus, G-rich oligonucleotides derived from genomic sequences like those from oncogene promoters and telomeres have been demonstrated to fold into G-quadruplexes. Through their visualization, compelling evidence for the existence of these non-canonical secondary nucleic acid structures has also been found in cellular environments.^{1,2} Our current understanding of the biological roles of quadruplexes suggests that G4s are involved in gene regulation and telomere maintenance, making genomic quadruplexes promising therapeutic targets.³ In this regard, much effort has been devoted during the last decades to searching for G4-stabilizing ligands for pharmaceutical intervention, *e.g.*, for modulating gene expression or telomerase inhibition in cancer cells.⁴ In addition to serving as potential drug targets, synthetic quadruplexes such as the thrombin binding aptamer (TBA) or anti-HIV-1 integrase aptamer constitute an emerging class of therapeutics, binding to various molecules including many pathologically relevant proteins with very high affinity and selectivity.^{5,6} Finally, the increasing use of quadruplexes in supramolecular chemistry as well as in biosensors and nanotechnology as a result of their ability to self-organize into complex two-dimensional networks and long nanowires attests to their enormous potential in medicinal and technological applications.^{7–9}

A typical monomolecular G-quadruplex is formed by sequences harboring four G-tracts of three or more consecutive guanine residues separated by short intervening sequences. Correspondingly, conservative search algorithms are based on a consensus sequence motif $d(G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3)$ for predicting putative G4 structures in genomic DNA.^{10,11} However, a growing number of non-consensus sequences has been reported to actually fold into stable G4 species. The availability of their high-resolution structures has shown a variety of unique conformational features distinct from the 'classical' G4 architecture. Clearly, a better understanding of principles governing quadruplex

folding of such non-standard G-rich sequences will support new algorithms for predicting putative regions within the genome amenable to G4 formation,^{12,13} but may also expand the G4 structural landscape for more effective drug targeting or the engineering of novel G4-based scaffolds.

This review is primarily focusing on the increasing number of G4 structures that do not comply with a consensus sequence motif but rather rely on short G₂-tracts and/or isolated G nucleotides for their architecture. Various strategies to compensate for G-deficiencies within their G-core or for reduced stacking interactions between tetrads are surveyed to give more insight into relevant contributions to G4 stability. Given the large number of deposited G4 structures with unusual sequence motifs, emphasis is placed on the folding behavior of unmodified sequences, with less attention given to quadruplexes featuring several closely spaced tracts of four or more consecutive guanines and non-canonical tetrads, *i.e.*, those composed of additional residues other than Gs.

A short survey on canonical G-quadruplex structures

Upon folding of a sequence composed of four closely spaced GGG triplets, guanine bases from the G-tracts will associate to form planar G-quartets (G-tetrads) through a cyclic hydrogen bond pattern involving both their Hoogsteen and Watson-Crick faces (Fig. 1). In most cases, stacking of three G-tetrads gives a three-layered G-core that is additionally stabilized through monovalent cations with a strength of stabilization in the order $K^+ > Na^+ \geq NH_4^+ > Li^+$.¹⁴ These are coordinated within the central channel of the G-core that is lined by the G-carbonyl oxygens to create a strong negative potential.



Yoanes Maria Vianney

Yoanes Maria Vianney received his MSc degree in Biotechnology at the University of Surabaya (Indonesia) in 2019. His thesis focused on plant tissue cultures and secondary metabolite extraction. He is currently working as a doctoral candidate under the supervision of Prof. Klaus Weisz at the Institute of Biochemistry, Universität Greifswald (Germany), characterizing G-quadruplex structures through calorimetric and, in particular, NMR spectroscopic methods.



Klaus Weisz

Klaus Weisz received his MSc in Organic Chemistry (1983) as a DAAD fellow at the University of Cincinnati, Ohio (USA) and his Diploma in Chemistry (1985) at the University of Stuttgart (Germany). Following his doctoral studies in Physical Chemistry at the University of Stuttgart (1990) he was a post-doctoral fellow at the Department of Pharmaceutical Chemistry at the University of California, San Francisco (1990–1993) and a research fellow at the Free University of Berlin (1993–2001), completing his habilitation in Physical Chemistry in 2000. Since 2001 he has been a Professor of Analytical Biochemistry at the Institute of Biochemistry, Universität Greifswald (Germany). His research focuses on nucleic acids with special emphasis on their tetra-stranded structures.



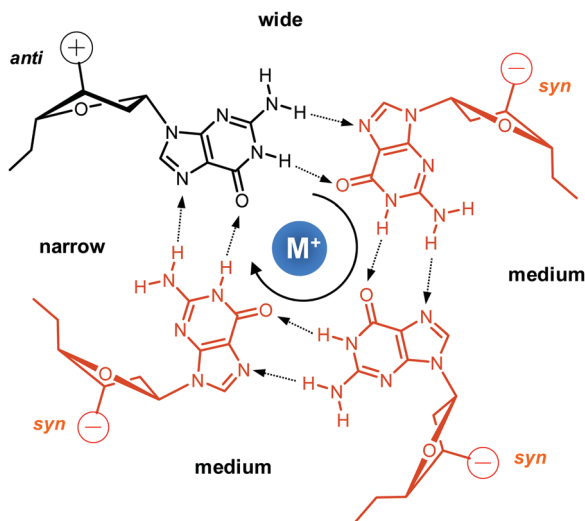


Fig. 1 Guanine tetrad with a centrally located metal ion, residues in *syn* or *anti* conformation, and four grooves of narrow, medium, and wide widths; strand polarities are indicated by + and – with the tetrad polarity running in a clockwise direction.

In case of an intramolecular quadruplex, intervening sequences form loop regions connecting the four G-columns (Fig. 2). A propeller or double-chain-reversal loop links two adjacent G-tracts with parallel orientation whereas lateral (edge-wise) and diagonal loops connect two adjacent or distal anti-parallel G-tracts, respectively. Depending on its topology, a conventional monomolecular quadruplex may be grouped into three

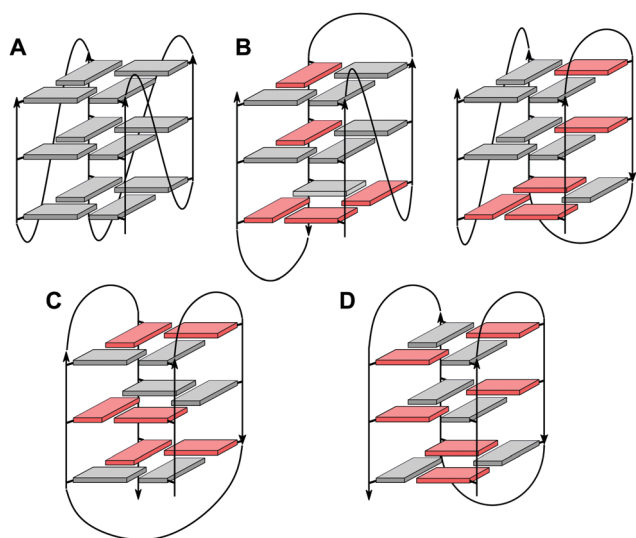


Fig. 2 Topologies of canonical three-layered G-quadruplexes. (A) Parallel quadruplex with strands connected by three propeller loops and all-*anti* G-tetrads; (B) (3+1) hybrid quadruplexes with three parallel and one anti-parallel strands connected by one propeller and two lateral loops; (C) basket-type (2+2) anti-parallel quadruplex with each strand adjacent to a parallel and an anti-parallel strand, two lateral and one diagonal loops, and G(*syn*)–G(*syn*)–G(*anti*)–G(*anti*) tetrads; (D) chair-type anti-parallel quadruplex with each strand adjacent to two anti-parallel strands, three lateral loops, and G(*syn*)–G(*anti*)–G(*syn*)–G(*anti*) tetrads; residues in *anti* and *syn* conformation are colored grey and red, respectively.

major families: a parallel G4 with all four G-tracts being parallel and only containing propeller loops; an anti-parallel G4 with two parallel and two anti-parallel G-runs; and a (3+1) hybrid with three parallel and one anti-parallel G-columns. Because an intramolecular quadruplex is defined by a combination of three different types of loops progressing in either a clockwise or counter-clockwise direction, a large number of topologies is conceivable. For a more systematic nomenclature, a descriptor based on the type of consecutive loops and their progression in relation to a frame of reference has been suggested.¹⁵ In such a system, the parallel topology with three sequential propeller loops progressing in an anti-clockwise direction can be designated as (–p–p–p) (Fig. 2A). Clearly, several of the theoretical loop combinations are forbidden due to geometrical restrictions. In fact, only 14 of these were predicted to be mechanically feasible but four of those have still not been experimentally verified to-date.^{16,17}

Among conformational properties of individual G residues within the quadruplex core, glycosidic torsion angles, *i.e.* *syn* and *anti* conformers, play a critical role for any quadruplex species due to their importance in G-tetrad formation and their close link with relative strand polarities of the four G-columns. In a parallel quadruplex, all residues within a tetrad must adopt the same glycosidic torsion angle for forming a planar G-quartet arrangement held together by the eight Hoogsteen hydrogen bond interactions. Typically, such G4s are composed of an all-*anti* G-core, although exceptions forming a single all-*syn* quartet exist for modified but also unmodified quadruplexes.^{18,19} On the other hand, residues in anti-parallel G-tracts require different glycosidic conformations when participating in the same G-tetrad. This relationship between relative strand polarities and glycosidic torsion angles has frequently been used as a powerful tool to guide folding of a G-quadruplex through the site-specific incorporation of G analogs favoring either *syn* or *anti* glycosidic torsion angles to enforce a particular topology.

Whereas the pattern of glycosidic angles for residues within a G-tetrad is determined by the orientation of the four G-columns, the sequential glycosidic conformation of consecutive G residues within an individual G-run may vary. It should be noted, however, that changing the glycosidic torsion angle within a column will, as a consequence, also change the tetrad polarity, *i.e.*, the clockwise or anti-clockwise direction of Hoogsteen hydrogen bonds within a tetrad plane when going from hydrogen bond donor to hydrogen bond acceptor. Thus, *anti-anti* and *syn-syn* steps will result in homopolar tetrad stacking whereas *syn-anti* and *anti-syn* steps will lead to heteropolar stacking. Computational studies have predicted more favorable interactions for *syn-anti* and *anti-anti* steps with energetic penalties for *anti-syn* and *syn-syn* steps, consistent with conformational properties of most reported G-quadruplex structures.^{20,21}

In addition to their relationship with relative strand orientation and G-tetrad polarity, glycosidic torsion angles will also affect the width of the four grooves featured by the four-stranded quadruplex. Whereas all grooves in parallel quadruplexes are of



medium width, base-paired Gs of different glycosidic conformation as observed in anti-parallel and (3+1) hybrid structures will also form narrow and wide grooves in case of *syn* → *anti* and *anti* → *syn* arrangements within a tetrad, respectively.

From a perspective of intervening sequences, it is their folding into a specific type of loop that defines the topology of the quadruplex architecture. General guidelines have emerged, correlating the length and position of linker sequences to the stability and to favored G4 structures.^{22–24} Due to geometric restraints, formation and stability of particular loops are strongly correlated with the length of the intervening linker sequences. Generally, propeller loops are most stable when composed of only 1–2 residues although even 0 nt propeller loops have been reported in rare cases.^{25,26} Lateral loops often include 2–4 residues depending on bridging a narrow or a wide groove, and diagonal loops require ≥3 residues for linking diagonally positioned G nucleotides. However, even for a conventional sequence, additional complexity may arise due to loop and overhang residues being engaged in subtle tertiary interactions to likewise affect the favored topology. Finally, it is not only the inherent sequence but also the outer conditions like the nature of cations, the ionic strength, and molecular crowding that may significantly impact the topology of a folded quadruplex. Whereas sodium ions have been shown to promote an anti-parallel topology, potassium ions rather tend to destabilize anti-parallel quadruplexes.^{14,27} Folding of the same sequence into either a monomolecular or bimolecular quadruplex may depend on low or high potassium ion concentrations in the buffer solution and is yet another example for an often rather unpredictable folding pathway even when looking at regular G4-forming sequences.^{28,29}

Taken together, intense research during the past years has provided a wealth of information regarding the energetics and structural interdependencies in ‘conventional’ G-quadruplexes. Our present knowledge of folding principles enables us to make a guess as for the most stable quadruplex fold of a given G4 consensus sequence and to tackle the rational design of G4 architectures.^{30,31} However, we are still far from reliably predicting topologies based on primary structure due to more subtle additional interactions involving flanking and intervening residues and also to the impact of specific solution conditions.

Quadruplexes with long loops and quadruplex–duplex hybrids

Longer unstructured loops in G-quadruplex structures tend to be increasingly disfavored because of entropic effects.^{32–34} In fact, only few quadruplexes with long loops of >7 residues, violating the conservative consensus sequence for putative G-quadruplex forming motifs, have been reported to-date. Thus, a G-quadruplex formed by the conserved 26 nt G-rich fragment of the human *CEB25* minisatellite forms a parallel-stranded G-quadruplex with a 9 nt central double-chain-reversal loop (Fig. 3A).³⁵ Within this quadruplex, an A·T Watson–Crick and a potential G·A non-canonical base pair between loop and

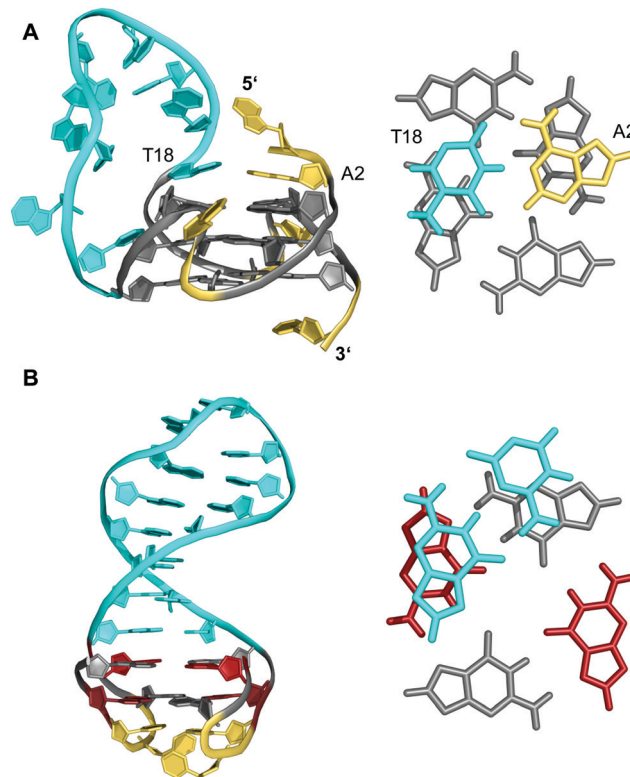


Fig. 3 (A) Solution structure of a human *CEB25* minisatellite sequence with a 9 nt propeller loop formed in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0; $T_m = 76.5$ °C; PDB 2LPW),³⁵ an A·T Watson–Crick base pair between a loop and 5'-overhang residue anchors the 3'-terminus of the loop on top of the 5'-outer G-tetrad (right). (B) Solution structure of a quadruplex–duplex hybrid with a two-layered anti-parallel G-quadruplex and a coaxially stacked duplex hairpin bridging the G4 wide groove formed in K^+ solution (20 mM KP_i + 20 mM KCl, pH 7.0; PDB 2M8Z),³⁸ the duplex GC base pair stacks onto the G-tetrad at the quadruplex–duplex interface (right); *anti*- and *syn*-guanosines of the G-core as well as loop and flanking residues are colored grey, red, and yellow, respectively; the 9 nt propeller loop in (A) and the stem-loop duplex in (B) are colored cyan.

5'-overhang residues fix the 3'-terminal loop domain above the 5'-outer G-tetrad. Another example involves two parallel-stranded G4 conformers from a *KRAS* promoter sequence, that were found to coexist in equilibrium and feature long third propeller loops composed of eleven and twelve nucleotides, respectively. In this case, high-resolution NMR structures determined from single mutants revealed π – π interactions between some bases of the propeller loop as contributors to the overall stability of the structure.³⁶ Also, sequences encompassing five to seven human telomeric (GGGTTA) repeats were shown by NMR to form (3+1) hybrid structures with an up to 21 nt long propeller loop when inner GGG triplets were blocked from participation in G-tetrads through single G → I or G → T substitutions.³⁷ Noticeably, such long loops may constitute new recognition motifs, allowing their targeting by a loop-complementary oligonucleotide to form a double-helical loop region.

Contrary to what is expected assuming most stable 1 nt propeller loops,³⁹ longer loops of ≥5 residues are rather frequently found to be of a propeller type. Apparently, such loops



often allow for stabilizing tertiary interactions with other loop and flanking residues. Following the concept of loop interactions to stabilize longer loop domains, appropriate linker sequences may intrinsically form Watson–Crick paired stem-loop duplexes as part of stable quadruplex–duplex hybrid structures (QDHs). Notably, in contrast to a regular single-stranded linker, quadruplex stabilities of engineered QDHs generally increase with the length of the double-helical hairpin domain.⁴⁰ When forming a lateral-type loop connection, the duplex is favored to bridge a wide groove of the quadruplex G-core to better accommodate distances between the sugar-phosphate backbones of coaxially oriented duplex and quadruplex domains (Fig. 3B).^{31,38} Here, the first base pair at the junction also affects stability due to additional stacking interactions with the quadruplex outer tetrad.⁴⁰ On the other hand, a connecting hairpin element may likewise replace a regular propeller loop, but by connecting G residues at opposite faces of the G-core the first base pair bridging the junction will be invariably disrupted in such an orthogonal arrangement.

Likewise, a duplex-forming diagonal loop with intrinsic Watson–Crick base pairing can be found for a G-rich sequence located in a promoter region of the HIV-1 long terminal repeat (*LTR*).⁴¹ In the major G-quadruplex conformation *LTR-III*, the 12 nt loop contains a stabilizing duplex hairpin element with three base pairs. However, the longer distance across the distal edges of the quadruplex again prevents residues at the quadruplex–duplex interface to be engaged in a stable base pair.

Quadruplexes with a two-tetrad G-core

The stability of G-quadruplexes is mostly derived from the stacking of its planar tetrads with stacking energies estimated to be ~ 80 kJ mol⁻¹ per tetrad.^{42,43} Therefore, the stability increases with an increase of stacked tetrads and only a limited number of monomeric two-layered quadruplex architectures has been reported to date. Among these, the thrombin-binding DNA aptamer (TBA) with its four tracts of only two contiguous Gs is one of the most prominent representatives.^{44–46} Each of its four G-tracts has a favorable 5'-*syn-anti-3'* arrangement resulting in opposite hydrogen bond directionalities of its two stacked G(*syn*)–G(*anti*)–G(*syn*)–G(*anti*) tetrads. The G-runs are connected by two T–T lateral loops on one side and a central T–G–T lateral loop on the other side of the anti-parallel chair-type quadruplex (Fig. 4). Additional stabilization comes from the stacking of a TT base pair from the first and third loop on one of the G-quartets. Other stabilizing contributions may also involve some stacking interactions by bases of the central 3 nt lateral loop on the other face of the G-quadruplex core.

Various TBA modifications have been reported in the past, mostly aiming at an improvement of pharmacological properties. Notably, a TBA analog containing a 5'–5' site of polarity inversion in the first lateral loop resulted in a (3+1) hybrid structure by keeping a 5'-*syn-anti-3'* torsion angle progression along all G-runs.⁴⁷ As a consequence, it differs from the unmodified TBA

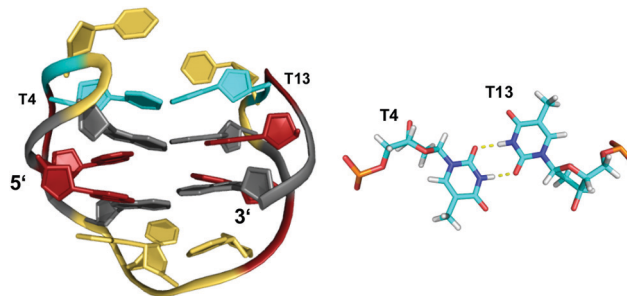


Fig. 4 Chair-type anti-parallel G-quadruplex of the TBA aptamer d(GGTTGGTGTGGTTGG) in K⁺ solution (110 mM KCl, pH 6.1) and hydrogen-bonded base pair formed between two T residues from opposite T–T lateral loops (PDB 148D);⁴⁶ *anti*- and *syn*-Gs of the quadruplex core, loop residues, and the T–T base pair are colored grey, red, yellow, and cyan, respectively.

in having one G(*syn*)–G(*syn*)–G(*syn*)–G(*anti*) and one G(*anti*)–G(*anti*)–G(*anti*)–G(*syn*) tetrad alignment with a parallel 5'–3' strand orientation of the first, second, and fourth strand and a third strand proceeding in the opposite direction. However, stabilizing forces through loop residues are very similar to those found for unmodified TBA.

The TBA quadruplex illustrates a frequently observed principle of stabilization through capping structures formed by base pairing alignments of loop and overhang residues. Such interactions may even be favorable enough in two-layered quadruplexes to successfully compete with three-layered G4s in sequences comprising four GGG-tracts. Thus, the unexpected observation of a G-quadruplex with only two tetrad layers for a human telomeric sequence featuring four G₃-runs emphasizes the potential role of tertiary interactions.^{48,49} Usually, the human telomeric sequence exhibits a (3+1) hybrid form with three stacked G-tetrads in K⁺ solution. However, the 5'-truncated variant d[(GGGTTA)₃GGGT] was shown to favor a two-layered basket-type structure with all G-columns comprising a 5'-*syn-anti-3'* glycosidic bond arrangement.⁴⁸ The conformation is stabilized by A–G–A and G–G–G triples capping the top and bottom faces of the G-core, respectively (Fig. 5). Moreover, two hydrogen-bonded T residues on top of each triad may add further stacking interactions.

Apparently, extensive base pairing and stacking of loop residues can outweigh stabilities of alternate three-layered G4 structures. It should be mentioned, however, that the telomeric two-G-tetrad conformation has been questioned to be a stable form for the extended human telomeric sequence because the addition of a 5'-flanking residue was shown to mostly abolish formation of a two-layered G4 structure.⁴⁹ Instead, the latter was suggested to likely constitute an intermediate in the inter-conversion between different telomeric G-quadruplex topologies.

In close analogy to the human telomeric sequence mentioned above, a sequence with single mutation from the RANKL gene d(G₃TAG₃AGCG₃AGAG₃) adopts a two-layered basket-type topology, again stabilized by a G–G–G and an A–G–A triple on top of the 5'- and 3'-tetrad, respectively.⁵⁰ Here, the critical role of capping base triads and loop residues was uncovered by a structural rearrangement to the anticipated three-layered (3+1)



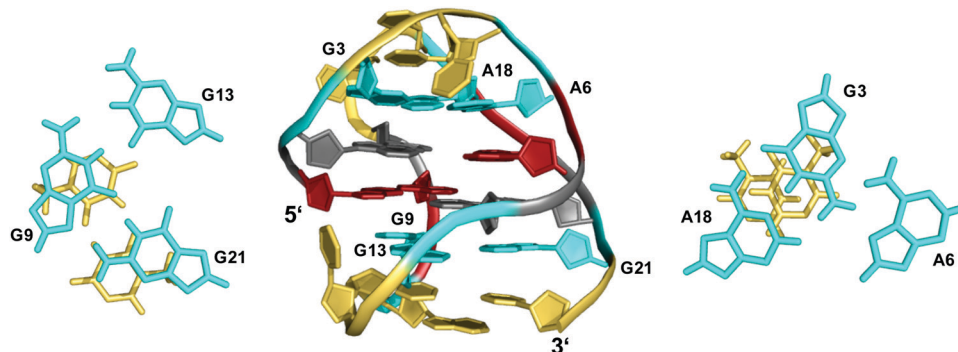


Fig. 5 Structure of a human telomeric G-quadruplex (form 3) in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0; $T_m = 57.0$ °C; PDB 2KF8);⁴⁸ A18-G3-A6 and G21-G9-G13 base triads sandwiched between a G-tetrad and a potential T-T base pair cap the top and bottom of the two-layered G-core; *anti*- and *syn*-residues of the G-quadruplex core, loop and flanking residues, and bases involved in triads are colored grey, red, yellow, and cyan, respectively.

hybrid fold induced by an A5-to-T5 modification. The latter is associated with the destruction of the capping A5-G3-A17 triple, releasing G3 from the A5-G3 base pair. This enables G3 to participate in G-tetrad formation with a concomitant shortening of the 3 nt lateral loop to become a more favorable 2 nt propeller loop.

Other examples exist for the stabilization of a two-layered quadruplex core by additional tiers of planar base pairing arrangements from overhang and loop residues.^{51–54} Thus, a truncated form of the *Bombyx mori* telomeric single repeat sequence d(TAGG) was shown to fold into a four-stranded quadruplex with a two-fold symmetry axis consisting of two G(*syn*)-G(*syn*)-G(*anti*)-G(*anti*) tetrads of different tetrad polarity.⁵¹ The two-layered core is sandwiched between unusual T-A-A triads with one adenosine pairing with the A-T Watson-Crick pair through the minor groove. All three bases and the sugar ring of one adenosine of the triad partially stack over the underlying G residues of the quartet. Inspired by the latter architecture, a sequence d(GGGTTCAGG) was designed and demonstrated to fold into a two-fold symmetric bimolecular

G4 structure with heteropolar stacking of two G(*syn*)-G(*anti*)-G(*syn*)-G(*anti*) tetrads capped by a C-G-A triad on each of the two quadruplex faces.⁵² Emphasizing the important role of additional layers made up by triads, the 12mer sequence d(A₂G₂T₄A₂G₂) with a pair of AAGG repeats folds into a bimolecular structure with 2-fold symmetry and a core of two G(*syn*)-G(*syn*)-G(*anti*)-G(*anti*) tetrads capped on both sides by A-T-A triads.⁵³ The latter, sandwiched between a G-tetrad and an additional outer non-Watson-Crick A-T base pair, contains one adenosine in *syn* conformation that pairs with the thymine through a reverse Hoogsteen alignment (Fig. 6). It should be mentioned that synergistic effects between the unusual base triads and the G4 core result in significant contributions of the stacked triads to the stability of two-layered quadruplexes but also to the promotion of base triad formation through the tetrad platform.

A 12 nt minimal sequence d[GT(GGT)₃G] derived from the anti-proliferating 28 nt DNA aptamer *AGRO100* forms a unique left-handed parallel G-quadruplex with two G-tetrad layers connected by short loops (Fig. 7A).⁵⁵ Lacking additional capping

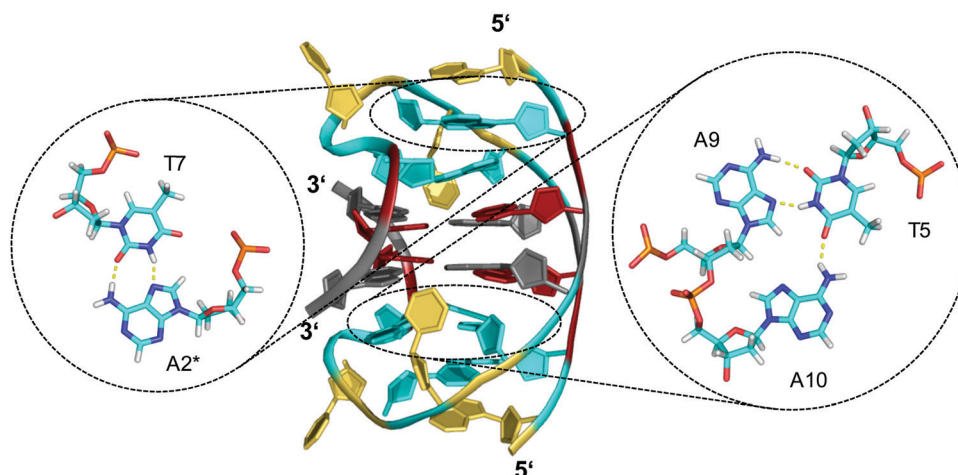


Fig. 6 Diamond-shaped bimolecular G-quadruplex with a two-layered G-core formed in Na^+ solution (5 mM NaP_i + 150 mM NaCl, pH 6.9; PDB 1D6D);⁵³ each of the tetrads forms a platform that is capped by a T(*anti*)-A(*syn*)-A(*anti*) triad (right) and a reversed Hoogsteen A-T base pair (left); an asterisk denotes a residue from the symmetry-related strand; *anti*- and *syn*-residues of the G-quadruplex core, loop and flanking residues, and bases involved in triads and base pairs are colored grey, red, yellow, and cyan, respectively.



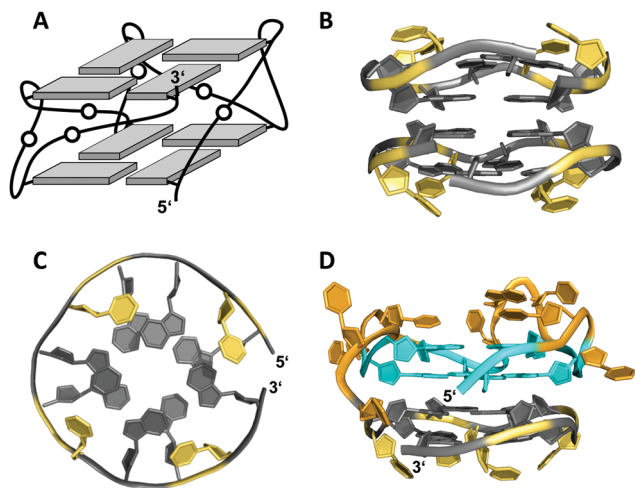


Fig. 7 (A) Schematic representation of a minimal left-handed G4. (B) Crystal structure with two stacked left-handed G4 units (crystals grown from 12 mM spermine and 80 mM KCl, pH 7.0; PDB 6FQ2).⁵⁵ (C) Top view with T loop residues oriented towards the outer tetrad of the left-handed domain; a broken G-column is formed by two split Gs at the 5'- and 3'-ends. (D) Hybrid structure with a right-handed TBA subunit connected to the left-handed motif formed in K⁺ solution (20 mM KPi + 70 mM KCl, pH 7.0; PDB 6JCE).⁵⁶ G residues of the quadruplex core and loop residues are colored grey and yellow for the left-handed G4, and cyan and orange for the TBA subunit in (D).

structures, it dimerizes through 5'-5' stacking for additional stabilization (Fig. 7B). Likewise, two monomers connected by a linker form a four-layered structure with two stacked left-handed subunits of parallel topology. Single-residue loops are clearly favored for the formation of the left-handed G4. In fact, thymine bases of the 1 nt loops collapse toward the terminal G-tetrad and allow for hydrogen bonds between their O4' atoms and amino protons of adjacent tetrad guanines (Fig. 7C). Whereas the TBA sequence features four GG doublets, the 12 nt sequence of the left-handed G4 comprises two single Gs at each terminus. By their stacking upon each other they form an unusual split-guanine tract which is assumed to convey the left-handed twist with its fully circling backbone (Fig. 7C).

Apparently, parallel-stranded two-layered quadruplexes from sequences that encompass closely spaced G-doublets have a strong propensity for additional stacking interactions, either through dimerization or in case of longer sequences through a stacked arrangement of two G4 domains made up of their 5'- and 3'-segments. Several examples for the latter derive from the polymorphic *AGRO100* aptamer that is composed of two domains with four G₂-tracts each. A single G-to-T substitution in the 5'-stretch and addition of thymidine residues at the termini yielded a well-defined sequence that folds into a four-layered G-quadruplex comprising two propeller-type parallel-stranded subunits connected through a central linker.⁵⁷ On the other hand, an alternate G-to-T substitution in the 3'-terminal G-doublet yielded a quadruplex topology termed Z-G4, featuring two stacked G4 domains both with left-handed helicity.⁵⁸ Noticeably, the latter is enforced by the 3'-domain composed of the 12 nt minimal motif mentioned above.⁵⁵

The TBA sequence can also be forced into a parallel topology with its three lateral loops switching into three propeller loops by its linkage to the minimal left-handed G4 sequence. Here, the two G4 units again stack on each other, yet with different helical orientation (Fig. 7D).⁵⁶ Because lateral loops impede stacking, favorable stacking interactions between the two subunits, *i.e.*, the right-handed TBA and the left-handed domain are efficient in driving such refolding into a parallel G4. Also, additional stacking of one base from each propeller loop on the 3'-outer TBA tetrad was observed and may contribute to the stabilization of this TBA topology.

Stabilization can also be provided by bases that are directly linked in-plane to the G-tetrad to form pentads, hexads or heptads. Thus, a dimeric hexad motif with two hexads stacking upon each other was reported for a d(GGAGGAG) sequence in a 150 mM Na⁺ solution.⁵⁹ GGA triplet repeats are abundant in eukaryotic genomes and thought to also be associated with the occurrence of several diseases.^{60,61} In the two tandem GGA triplet repeat sequence, each bimolecular monomer forms a stack of a G·(A)·G·G·(A)·G hexad, a G-tetrad, and an A-A mismatched base pair (Fig. 8). The hexad forms by the in-plane attachment of two adenine bases over their Hoogsteen edge to the G-tetrad through hydrogen bonding with opposite guanine bases. Thus, two out of the four G-tetrad guanines are anchored through a total of six hydrogen bonds. Formation of such hexads is expected to be supported or even driven by extensive π - π stacking interactions between two stacked hexads at the dimer interface.

A corresponding architecture with an intramolecular stack composed of a G·(A)·G·(A)·G·(A)·G heptad and a G-tetrad, additionally stabilized through dimer formation with stacked heptads at the interface, was also found for a four tandem GGA triplet repeat d(GGA)₄. Likewise, an intramolecularly folded d(GGA)₈ extended sequence with two subunits composed of stacked tetrad and heptad arranged in a tail-to-tail orientation with inter-heptad stacking.^{62,63} Notably, although octad formation through the G-tetrad alignment of a fourth adenine base either from the 3'-terminus in d(GGA)₄ or from the adenosine linking the two subunits in d(GGA)₈ is conceivable, it has not been observed. Apparently, the adenosine requires a subsequent 3'-adjacent G residue as part of the G-tetrad to be anchored within the tetrad plane.

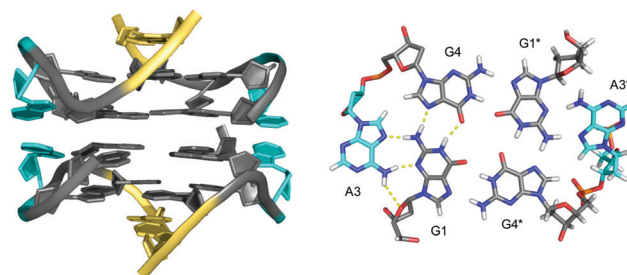


Fig. 8 Dimeric G4 structure formed in Na⁺ solution (2 mM NaPi + 150 mM NaCl, pH 6.6) and composed of four symmetry-related strands with stacked hexads at the dimer interface (PDB 1EEG);⁵⁹ guanines of the all-*anti* quadruplex core, loop and flanking residues, and adenines involved in hexads are colored grey, yellow and cyan, respectively.



In conclusion, the structure of two-layered quadruplexes as presented above emphasize the need for other stabilizing interactions in addition to the stacking of two tetrads in a G-core composed of favorable *syn-anti* or *anti-anti* steps along the four GG-columns. Here, loop residues are of particular importance by forming base pairs and/or base triads as additional stacked layers sandwiching the G-quadruplex core. Also, dimerization or inter-subunit stacking is often observed in case of two-layered quadruplexes with a propeller-type parallel topology, enabling unrestricted stacking with interfacial 5'-outer tetrads generally found to be more favorable.⁶⁴ Stacking interactions can be further optimized by expanding G-tetrads with intervening bases to form hexads or heptads, increasing the stacking interface within dimeric structures.

G-deficient G-quadruplexes and interrupted G-tracts

In the past, an increasing number of G4-forming sequences harboring a shortened G-tract and thus unable to fold into a canonical three-layered quadruplex with four non-interrupted GGG-columns have been reported. Assuming a thermodynamically controlled G4 folding, the final conformer will maximize favorable interactions, primarily striving to fill vacant G-core positions for increased stacking interactions but also through additional interactions involving intervening and flanking segments. Depending on the primary sequence, there are various possibilities for intramolecular G insertions into unoccupied G-core positions, leading to distinct structural features with bulged or interrupted G-columns. These approaches are schematically depicted in Fig. 9. In the following, corresponding G4 folds are reviewed with particular emphasis on non-modified

quadruplexes whose folding pathway is not guided by conformational preferences of incorporated nucleoside analogs (for the latter, see ref. 65).⁶⁵

G-quadruplexes with a guanine vacancy (vG4)

Deviating from the consensus sequence of a canonical G-quadruplex, sequences with three GGG-tracts and one guanine-deficient GG-tract can assemble into a three-layered quadruplex structure with one tetrad bearing a vacant site. Notably, bioinformatics studies have shown that such sequences are abundant in genomes and may be evolutionarily selected in genes with unique distribution patterns in both eukaryotic and prokaryotic organisms.^{66,67} The vacant site can easily accept a guanine base from guanine-containing metabolites such as GTP or GMP to form an intact and strongly stabilized G-core, demonstrated to effectively alter DNA replication *in vitro* at physiological GTP concentration.⁶⁶ Because G-quadruplexes with guanine vacancies (vG4) are distinct from canonical G4 structures in being able to sense intracellular concentrations of guanine derivatives, they have been proposed to play a critical role in gene regulation.

On the other hand, the abundance of vG4 forming sequences in the human genome offers a great potential for therapeutic interventions by more specific, high-affinity targeting. Thus, a bifunctional G4-binding peptide guided through a covalently linked guanine base was shown to feature promising selectivity and affinity toward the G-deficient quadruplex associated with strong suppression of *in vitro* replication.⁶⁸ From an analytical viewpoint, sensors for guanine derivatives based on quadruplexes with a vacant site have been shown to confer exceptional selectivity toward the analyte.⁶⁹

Despite the presence of a destabilizing additional thymine bulge in the short and non-contiguous GG-column at its 5'-end, the sequence d[TTGTG(TGGG)₃T] containing (12-1) guanines was shown by NMR to fold into a G-deficient intramolecular quadruplex with two G-tetrads and one outer G-triad in a parallel-stranded conformation (Fig. 10A).⁷⁰ In fact, molecular dynamics simulations established the formation of a G-triad-water complex with water molecules occupying the vacant site in the G-triad plane. Again, the vacancy being a G-binding hotspot can be specifically recognized by external guanine bases. High-affinity binding was observed for linear and cyclic d(AG) and cGAMP dinucleotides when targeting a T deletion mutant d[TTGG(TGGG)₃T] lacking the bulge.⁶⁷

Another example of a structurally characterized G-deficient quadruplex involves a modified human PDGFR- β gene promoter sequence d(AAG₃AG₃CG₂CG₃ACA) termed Pu19m2.⁷¹ It was shown to adopt two stable G4 structures formed by the G₂-tract shifted toward the 5'- or 3'-terminal quadruplex face with a corresponding vacancy in an outer plane adjacent to either the 3'- or 5'-terminus. The triad layer of the vG4 can again be complemented by the selective and strong external binding of physiologically relevant guanine metabolites such as dGMP, GMP,

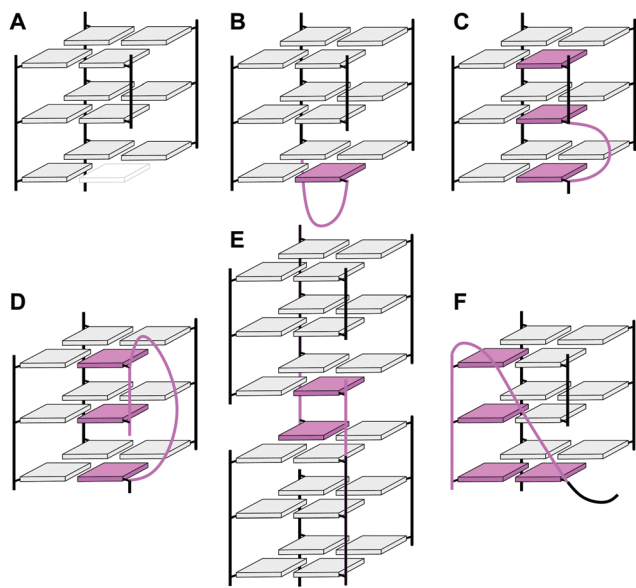


Fig. 9 Strategies to fill vacant positions within a quadruplex G-core. (A) Quadruplex with a vacant site, (B) snapback-loop, (C) bulge, (D) D-shaped loop, (E) interlocked G4, (F) V-shaped loop.



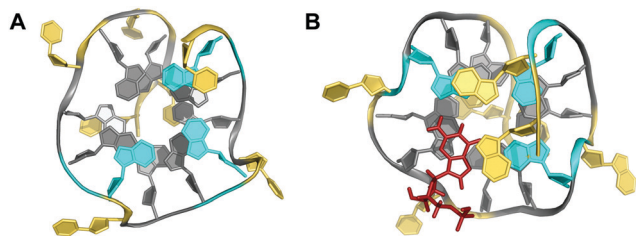


Fig. 10 Top view on (A) the vG4 NMR structure formed in K^+ solution (10 mM KP_i + 35 mM KCl, pH 7.0) from the sequence d(TTGTG(TGGG)₃T) with G-triad (colored cyan) stacked on a G-tetrad (PDB 2N60).⁷⁰ (B) Top view on a dGMP-complexed vG4 structure formed in K^+ solution (12.5 mM KP_i + 37.5 mM KCl, pH 7.0) from the PDGFR- β gene promoter sequence (PDB 6VOL);⁷¹ dGMP (stick model in red) fills the vacant site of the 5'-outer G-layer (colored cyan). Other all-*anti* G-tetrad core residues are colored grey; loop and flanking residues, yellow.

and cGMP but also by guanine-based drugs (Fig. 10B). Interestingly, metabolite binding is able to modulate the equilibrium between the two G₂-shifted isomers, mostly favoring binding to the G-deficient 5'-triad.

Snapback loop G-quadruplexes

If the Pu19m2 sequence of the PDGFR- β promoter with its vG4 fold is extended to also include the wild-type 3'-terminus with another G₃-tract, the resulting sequence d(AAG₃AG₃CG₂CG₃GCAGGG) designated Pu22m1 was found to adopt a parallel-stranded intramolecular quadruplex with three 1 nt propeller loops and an additional 5 nt lateral loop.³⁹ Here, it is a terminal 3'-G in a *syn* conformation that intramolecularly fills the vacant site of the third G₂-run through a snapback loop structure. Interestingly, the sequence itself features four contiguous runs with ≥ 3 guanines, expected to fold into a regular three-layered quadruplex without broken strand but with longer second and third loops. Apparently, the high stability of a parallel quadruplex with 1 nt propeller loops outweighs penalties expected for a fourth snapback lateral loop.

A snapback approach in combination with a 5'-terminal hairpin structure was also shown to fill a single vacancy left by a short G₂-tract. Here, the vacant site acts as an anchor point for the duplex stem-loop in fixing the 5'-terminal G in a *syn* conformation to the tetrad facing the duplex domain.³⁸

Snapback loops can also bridge distal corners as exemplified by a *c-myc* promoter sequence d(TGAG₃TG₄AG₃TG₄AAG₂) containing five guanine tracts. Although able to fold into a regular parallel G4 with 1 and 2 nt propeller loops, it was shown to favor folding into a parallel-stranded fold-back G-quadruplex with the 3'-terminal guanine base filling an empty guanine position within the 3'-tetrad through a diagonal snapback loop.⁷² The three-dimensional NMR structure of a G10I mutant termed Pu24I demonstrates its parallel fold with 1 nt, 3 nt, and 1 nt propeller loops and a fourth diagonal loop bridging two opposite corners of the 3'-G-tetrad with its terminal *syn*-G complementing the second G-column (Fig. 11). Single base substitutions suggest that a G-G-A triad within the diagonal

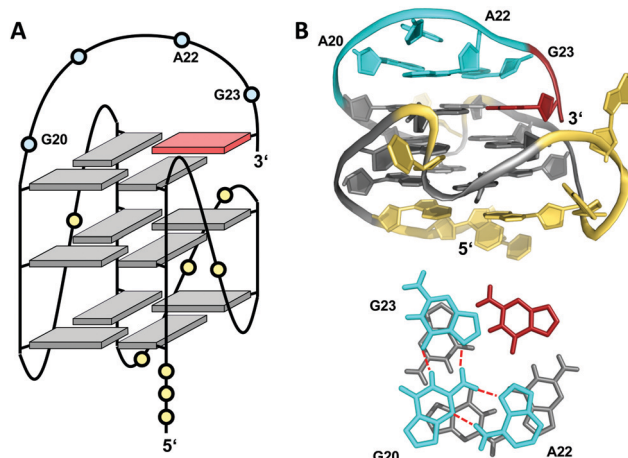


Fig. 11 (A) Schematic representation and (B) three-dimensional structure of Pu24I (PDB 2A5P) formed in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0) with residues of the diagonal snapback loop forming a G-G-A triad stacked on the 3'-outer tetrad (bottom);⁷² *anti*- and *syn*-guanosines of the G-tetrad core are colored grey and red, respectively; loop and flanking residues, yellow; residues forming the snapback loop, cyan.

loop capping the outer G-tetrad seems a critical structural motif for snapback loop formation in Pu24I. Correspondingly, a stacked G-G-A triad from residues of the diagonal snapback loop was likewise found to stabilize one of the two major G4 conformers formed by a G-rich sequence in the *KRAS* nuclease hypersensitive element (NHE) region.³⁶ It should be noted, that the addition of further non-G residues at the 3'-terminus may still allow for a fold-back topology but is expected to compromise the thermodynamic stability as suggested by calorimetric studies on mutated 3'-T extended *c-myc* promoter sequences with five guanine tracts.⁷³

Like the extended PDGFR- β promoter sequence, a G-rich *c-kit* promoter sequence d(AG₃AG₃CGCTG₃AG₂AG₃) encompassing four G₃-tracts and thus capable of forming a regular quadruplex with four continuous G-columns folds into a topology with a snapback loop in K^+ solution (Fig. 12).^{74,75} Again, the pronounced stability of short propeller-type loops in a parallel topology is suggested to drive folding but base pairing alignments in the loops provide for additional stabilization of this structure with several unique features. Here, isolated G10 is recruited to occupy a single outer G-core position and the corresponding G-column is complemented by insertion of the two 3'-terminal *anti*-G residues aligned in a parallel orientation. The two-residue loop directly following G10 links neighboring corners of the same tetrad. By laterally connecting a broken and continuous G-column of the same strand polarity, it shares features of both propeller and edge-wise loops. The 5 nt snapback connection that follows the fourth G₃-column to fill the two vacant sites of the third G-tract with parallel-oriented *anti*-G residues shows base pairing alignments. It is unusual in spanning two G-quartets with a 3'-flanking G being part of the central tetrad, allowing unrestricted DNA sequence extensions at the 3'-terminus. Formally, this rather long 5 nt loop may also be viewed as a propeller-type loop progressing against the right-handed helicity of the G-core. The overall topology is also



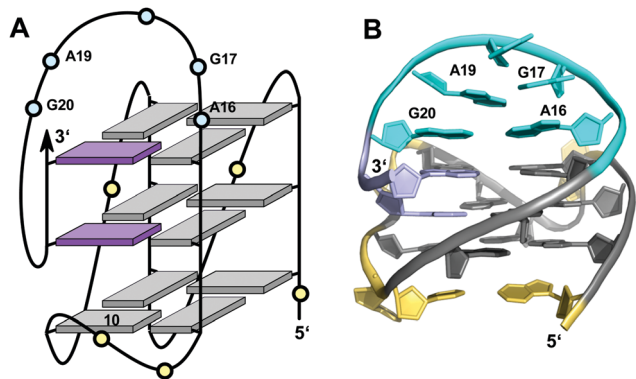


Fig. 12 (A) Schematic representation and (B) solution structure of a quadruplex with a distinct type of snapback loop formed by a *c-kit* promoter sequence in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0; PDB 2O3M);^{74,75} *anti*-guanines of the G-core and two inserted 3'-terminal Gs are colored grey and lilac, respectively; loop and flanking residues, yellow; residues of the snapback loop forming A16-G20 and G17-A19 base pairs, cyan.

conserved when replacing this loop by a hairpin motif within a closely related *c-kit* based sequence to form a unique quadruplex–duplex junction.³⁸ Clearly, it would also be conceivable to fill the two vacant positions through a conventional lateral-type snapback loop with two terminal *syn*-Gs in anti-parallel orientation. However, such a conformer may be disfavored by a less stable *syn-syn* stacking.

Quadruplexes with a bulge

Non-consecutive guanines may assemble into G-quadruplexes that feature a G-column of guanines with interdispersed bases protruding outward to form a bulge. Thus, whereas loop residues connect different columns of the G-tetrad core, a bulge connects adjacent guanines along the same column. Bulges have initially been shown in crystal structures of RNA quadruplexes but more recent reports on several three-dimensional structures of bulged G-quadruplexes in solution attest to their potential prevalence and stability under cellular conditions.^{36,57,76–83} Consequently, participation of isolated guanines within interrupted G-tracts in the formation of a stable ‘bulged’ G-tetrad core will significantly expand the number of genomic sequences with a potential for G-quadruplex formation.

The impact of bulges differing in sequence, size, position, or number on G-quadruplex formation was systematically studied by Mukundan and Phan.⁸⁰ Their results suggest that bulges can be located at any position in a G-quadruplex structure. However, the stability of quadruplexes with a bulge decreases with increasing bulge size in analogy to the length dependence of propeller loops. Also, the G4 stability will depend on their location but also on the sequence context and the G-quadruplex topology. In general, bulges are destabilizing, limiting the number of individual bulges n compatible with formation of three-layered quadruplexes to $n \leq 3$. Destabilization can be attributed to a strained backbone but mostly to the unfavorable

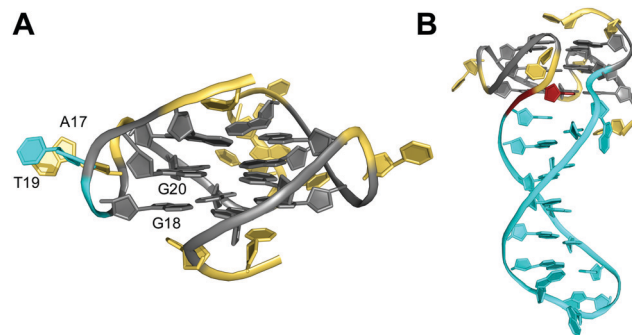


Fig. 13 Solution structure of G-quadruplexes with a bulge. (A) Parallel-stranded *LTR-IV* G-quadruplex from the HIV-1 genome formed in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0; $T_m = 50.5$ °C; PDB 2N4Y);⁸² *anti*-G20 following the bulge adopts a north sugar conformation and the bulged T19 stacks onto A17 of the neighboring propeller loop. (B) Parallel-stranded G4 with a bulge forming a stem-loop duplex in K^+ solution (20 mM KP_i + 30 mM KCl, pH 7.0; $T_m = 46.2$ °C; PDB 7CLS);⁸⁵ the G residue following the hairpin-forming bulge adopts a *syn* conformation. *anti*- and *syn*-guanines of the G-tetrad core are colored grey and red, respectively; loop and flanking residues, yellow; residues in bulges, cyan.

entropy of solvating the protruding residues. Thus, entropic effects likely determine a stacking interaction of a thymine bulge with a single-nucleotide propeller loop adenine to reduce the hydrophobic surface area in the long terminal repeat sequence *LTR-IV* of the proviral HIV-1 genome (Fig. 13A).⁸² Such rather subtle interactions may in fact explain the different impact of bulges on the thermal stability depending on their position in various topologies.

In general, bulges do hardly perturb the G4 core structure which essentially occupies the same conformational space as found for canonical G-quadruplexes. However, guanine residues adjacent to bulges have been reported to frequently populate an additional range of backbone torsion angles.⁷⁸ Also, revisiting available quadruplex structures, G-core residues preceding or following the bulge are often found to adopt sugar conformations in the north rather than in the more typical south domain. It should be noted, however, that in most cases no restraints for sugar dihedral angles were employed for generating the NMR solution structures.

A unique 2 nt GA bulge in a G-quadruplex formed by a G-rich sequence in the regulatory region of a RANKL gene connects *anti*- and *syn*-guanines that occupy G-core positions of a parallel G4.⁸⁴ Consequently, in order to maintain formation and proper stacking of the G-tetrads, the bulge must provide for a turn of the backbone in adopting a pseudo-loop conformation. Remarkably, the corresponding G4-forming sequence encompasses four G₃-tracts to allow for a regular three-layered quadruplex. Assuming the bulge to be destabilizing, the bulge-containing fold seems to be driven by a shorter overall 1-3-1 when compared to a 1-3-3 propeller loop architecture as expected for a bulge-free parallel topology.

Recently, base-paired duplex bulges of different size were incorporated into various positions of a G-quadruplex scaffold, demonstrating their noticeable stabilization when compared to unstructured bulges.⁸⁵ In fact, thermal stabilities of



duplex bulges are slightly increased with increasing bulge sizes, following a similar trend as observed for G4 hairpin loops. The formed quadruplex–duplex junction is reminiscent of an orthogonally aligned propeller-type stem-loop structure with a first disrupted base pair to allow for a progressive transition from the quadruplex to the duplex segment associated with an increase in strand separation (Fig. 13B). However, in contrast to a propeller-type hairpin loop the double-helical foldback bulge continuously stacks onto the 3'-outer G-tetrad and only the first unpaired base projects outward from the groove.

D-Shaped loop

Unlike bulges that link two split G residues within the same G-column in a consecutive way, another distinct type of loop connects residues of a column located at opposite faces of the G-core. Owing to its characteristic progression it has also been termed D-shaped loop.²⁶ This peculiar arrangement positions the 5'- or 3'-terminal G of a d(G₂N_xG) or d(GN_xG₂) tract between the other two G residues when forming a G-quadruplex column (see Fig. 9D). In this regard it is reminiscent of a structural motif reported for a short fragment of telomeric DNA from *S. cerevisiae*. Here, an unusual pseudo-circular G-hairpin with a compact core of three GG base pairs is formed and a chain reversal within a continuous G₃-tract places the 3'-terminal G between the two preceding G residues in the base-paired structure.⁸⁶

In a G-quadruplex, such a structural motif was shown for a G-rich VEGF aptamer carrying three locked nucleic acid modifications.²⁶ Here, a 2 nt D-shaped loop fills a vacant position within the same column by bridging two corners on opposite sides of the G-core (Fig. 14). Notably, with all three tetrads featuring the same polarity and all core guanines adopting an *anti* conformation as demonstrated by NMR data analysis, there seems to be no strand inversion between the flanking outer G-core residues as would be expected for this

structural motif. Interestingly, however, an easy switch to a *syn* conformation was observed for the 3'-flanking G during structure calculations. A 0 nt propeller loop bridging two tetrad planes precedes and another 2 nt loop directly follows the V-shaped loop. The unusual 2 nt loop ties two parallel-oriented G positions at adjacent corners of the same tetrad in analogy to a corresponding loop in the *c-kit* promoter G-quadruplex.^{74,75}

Another example for a D-shaped loop comes from a guanine-rich 36 nt RNA motif named *sc1* capable of quadruplex formation. The solution structure of the *sc1* RNA complexed with an arginine-glycine-rich RGG peptide from the FMRP protein reveals a G-quadruplex domain connected to a flanking duplex stem.⁸⁷ The three-layered all-*anti* G4 is composed of two stacked tetrads of the same polarity and an additional G-tetrad of opposite polarity facing the duplex domain. Here, a strand polarity inversion within one G₃-tract and two 1 nt D-shaped loops associated with a flipped backbone connect the inverted G-tetrad with the other two G-tetrad layers.

Interlocked structures

Interlocked structures are composed of more than one separate G-rich strand and in the past have often been associated with the formation of G-wires. The latter can form if G-rich strands associate out-of-register to present 'sticky' ends. Two such slipped structures may subsequently dimerize through their terminal free G residues to form an extra G-tetrad. Thus, d(GGGT) may align into an octameric complex with five stacked G-tetrads in addition to the tetramolecular d(GGGT)₄ with in-register strand association (Fig. 15A).⁸⁸ If association is enabled at both 5'- and 3'-ends, self-assembly can lead to large nanostructures by the growth of an interlocked G4 in both directions.

A first model of a G-wire formed by the telomeric DNA oligonucleotide d(G₄T₂G₄) was proposed more than 25 years ago^{89,90} but its structural diversity could only be demonstrated by atomic force microscopy in the recent past.⁹¹ Expanding on the self-associative potential of G-rich sequences in a slipped alignment, oligomerization was also shown to be supported by GC overhangs that serve as cohesive 'sticky' ends to form two interfacial GCGC-tetrads by hydrogen bonding through Watson–Crick and Hoogsteen guanine edges.^{92–94} In G-wires, self-recognition and self-assembly relies on G-quartet formation and is expected to be promoted by hydrophobic effects but also by the enthalpic gain of multiple G-tetrad stacking. Correspondingly, these interlocked structures show high thermal stabilities but due to their slow kinetics of formation their population and length strongly depends on concentration, temperature, and cations present. Of note, some of the higher-order G4 structures suggested to coexist in particular with parallel-stranded G-quadruplexes^{23,95,96} may possibly also be traced to the formation of such high-melting interlocked structures.

Narrowing the definition of interlocked G-quadruplexes, G-tetrads at the interface between two G-deficient quadruplex folds may be mutually completed by Gs from the other subunit. This enables the sequence d(G₃AG₂T₃G₃AT), bearing only three

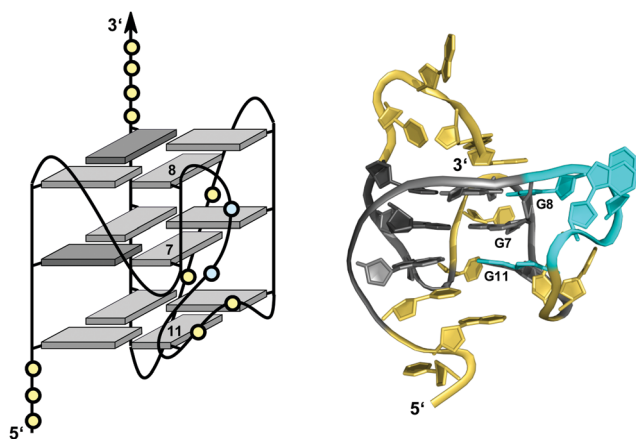


Fig. 14 Schematic representation (left) and three-dimensional solution structure (right) of a G-quadruplex with a D-shaped loop derived from a VEGF aptamer with locked LNA G residues in K⁺ solution (10 mM KP_i + 40 mM KCl, pH 7.0; T_m = 52 °C; PDB 2M53);²⁶ *anti*-guanosines and LNA analogs of the G-core are colored in light and dark grey, respectively; loop and flanking residues, yellow; residues of D-shaped loop, cyan.



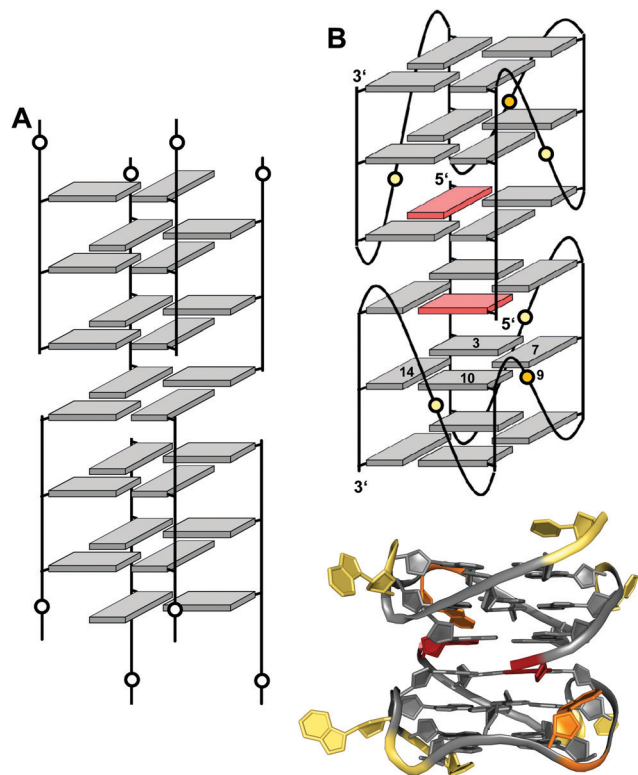


Fig. 15 Interlocked G-quadruplexes. (A) Schematic representation of two interlocked $d(\text{GGGT})_4$ out-of-register quadruplexes forming an octameric species.⁸⁸ (B) Schematic representation (top) and three-dimensional structure (bottom) of an interlocked quadruplex formed in K^+ solution (90 mM KCl, pH 7.0) by an HIV-integrase aptamer (PDB 1Y8D);⁹⁸ the pair of G-tetrads at the dimer interface are mutually filled with a 5'-terminal *syn*-G1 from the other strand; *anti*- and *syn*-guanines of the G-tetrad core are colored grey and red, respectively; loop and flanking residues, yellow; loop adenines A9 aligned in-plane of the tetrad to form a pentad, orange.

G-tracts, to fold into a dimeric four-layered quadruplex.⁹⁷ Each monomer adopts a compact domain with a 1 nt propeller loop, a 3 nt lateral loop, and a 0 nt V-shaped loop (see below). The dimeric interface features a pair of stacked A-(G-G-G-G) pentads through the interaction and in-plane alignment of the propeller loop adenine with a G-quartet. Also, each pentad is completed through the insertion of a 5'-terminal *syn*-G residue of the other monomer. The pentad stacks upon a tetrad in each monomer supported by one broken and three continuous GG-columns.

Likewise, the *93del* aptamer $d(\text{GGGGTGGGAGGAGGGT})$, an HIV-1 integrase inhibitor, forms a very stable six-layered G-quadruplex interlocked dimer in K^+ solution (Fig. 15B).⁹⁸ Each monomeric subunit contains one A-(G-G-G-G) pentad sandwiched between two G-tetrads with the G-tetrad at the interlocking interface complemented by the 5'-terminal *syn*-G from the first G_4 -tract of the other monomer. All G-columns within each monomer are parallel and linked by two 1 nt propeller loops bridging three G-tetrad layers. A second adenosine propeller loop that bridges two G-tetrad layers participates in A-(G-G-G-G) pentad formation.

Based on the *93del* aptamer, sequence variants forming corresponding interlocked G-quadruplex dimers were rationally

designed. These encompass a first long G_4 -tract to compensate with its 5'-G for a G-deficient tetrad of the other monomer, two medium G_3 -tracts, and another short G_2 -tract being positioned as second, third, or fourth G-run.⁹⁹ Indeed, very stable interlocked quadruplexes were demonstrated to form and may constitute robust scaffolds for technological applications.

Expanding on the architecture of locked quadruplexes, a unique intra-locked G4 structure was recently reported for the 28mer G-rich sequence $d[(\text{TGG})_4\text{TTG}(\text{TGG})_3\text{TTGT}]$ harboring multiple G_2 -tracts and two single G residues.¹⁰⁰ This sequence was shown to fold into a structure with two stacked bi-layered subunits formed by its 5'- and 3'-domains. Additionally, intra-molecular locking is achieved by the incorporation of a guanine from the 5'-subunit into the G-deficient interfacial G-layer of the 3'-subunit.

V-loop quadruplexes

V-loops are one of the most prominent non-canonical structural elements in G-quadruplexes and have received growing attention in recent years. V-shaped loops connect two adjacent G-columns by bridging two or three G-tetrad layers (see Fig. 9F). However, in contrast to a propeller-type connection one of the G-columns is broken and the two linked G-tracts are generally oriented anti-parallel to each other. The formation of 0 nt V-shaped loops is most common, but in some cases 1 nt or 2 nt V-loops have also been reported.^{54,100,101}

V-loops can exhibit high intrinsic stability and may even compete with canonical topologies. Thus, modifying all matching *anti*-G positions with strongly *anti*-favoring $^{\text{LNA}}\text{G}$ analogs in the telomeric sequence $d(\text{G}_4\text{T}_4\text{G}_4)$ from *Oxytricha nova* resulted in a rearrangement of the bimolecular anti-parallel quadruplex into a unique scaffold with a topology termed V4 fold.¹⁰² Here, all four G-stretches within two strands fold back in a V-shaped loop with an LNA residue at their 3'-end and interact with each of the other three G-stretches through the formation of four G-tetrads. The V-loop 5'-anchoring position is generally occupied by a *syn*-G being part of a discontinuous G-column. In order to trace favorable and unfavorable contributions to V-loop formation, various sugar-modified G analogs have recently been introduced at specific positions of a (3+1) hybrid quadruplex, triggering rearrangements into a V-loop structure.^{103–105} Detailed analysis of dual-modified V-loop quadruplexes bearing different combinations of G-analogs demonstrated that often overlooked sugar conformational preferences rather than glycosidic conformations were major contributors to V-loop stability. Thus, a stable V-loop structure was even formed when inserting $^{\text{LNA}}\text{G}$ with its fixed C3'-endo conformation (*north*) at the V-loop 5'-anchoring site to enforce a strongly disfavored *syn* conformation when followed by another 3'-flanking *anti*- $^{\text{LNA}}\text{G}$.¹⁰⁵ On the other hand, a sugar pucker in the *north* domain for both 5'- and 3'-flanking residues seems to match backbone conformational requirements of a conventional 0 nt V-shaped loop (Fig. 16A, top). In fact, a corresponding 5'-(*syn,north*)-(anti,north)-3' conformation for V-loop flanking residues is likewise found for other V-shaped



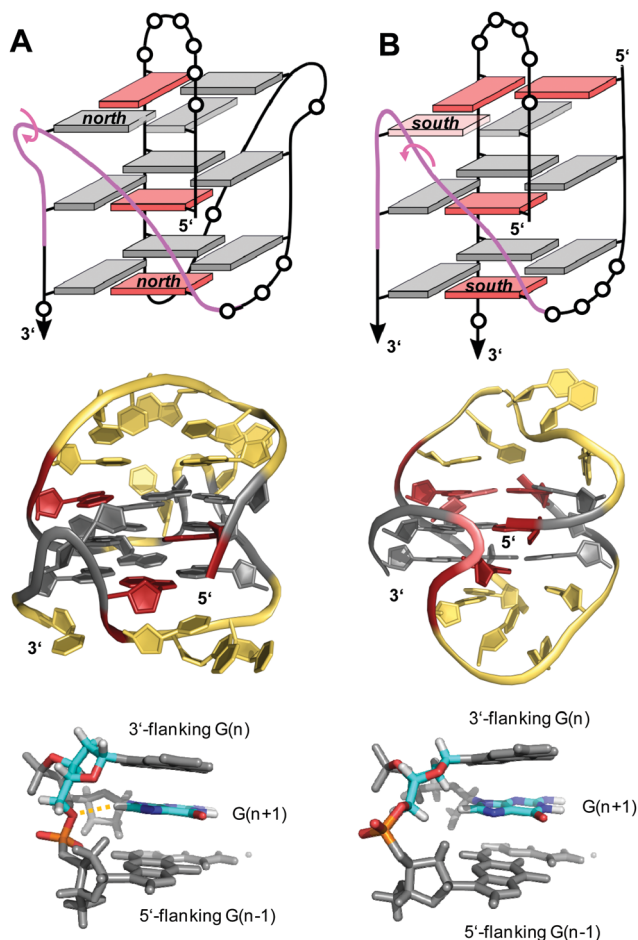


Fig. 16 V-loop topology with polarity inversion sites indicated by circular arrows (top), solution structure (center), and backbone conformation for the V-loop (bottom) comprising a 3 nt stretch that is colored magenta in the topological representations on top. (A) Monomolecular G4 formed in K^+ solution (20 mM KP_i + 80 mM KCl, pH 6.8; $T_m = 54.5$ °C) with conventional V-loop and $O5'(n)$ - $H8(n+1)$ interactions (PDB 5ZEY).¹⁰⁶ (B) Bimolecular G4 in K^+ solution (15 mM KCl, pH 5.5) with alternative V-loop and solvent-exposed phosphate (PDB 1U64).¹⁰⁹ G-core guanines in *anti*, *syn*, and 'low-*syn*' conformation are colored grey, red, and light red, respectively.

loops in unmodified quadruplexes and represents a characteristic feature for such conventional loops.^{106–108} Owing to the *syn* and *anti* anchor residues participating in G-tetrads of reversed polarity, there is no apparent strand polarity inversion inherent to the V-loop but rather between the 3'-flanking G and the following G within the same G-tract. Interestingly, a sharp turn of the sugar-phosphate backbone at the inversion site and a *north*-type sugar pucker of the 3'-anchoring residue places its $O4'$ and $O5'$ oxygen atoms in close vicinity to H8 of the following G to also allow for corresponding C-H...O interactions (Fig. 16A, bottom).

A second type of V-loop conformation has originally been suggested based on a sequence bearing a *south/south-east*-favoring 2'-fluoro-arabinoguanosine analog at the 3'-flanking position.^{104,105} In these alternative V-loop conformers, *south*-puckered residues are mostly found for both V-loop framing

positions, allowing to differentiate V-shaped loops according to two distinct sugar conformational preferences (Fig. 16B, top). Whereas a *syn* conformation at the 5'-anchor site seems mandatory for all regular V-loops, glycosidic torsion angles at the 3'-end of the alternative V-loop are typically outside the *anti/high-anti* range and rather adopt torsion angles in a less defined 'low-*syn*' range.^{41,109,110} As a consequence, a sugar-phosphate backbone inversion can formally be localized within the V-loop in this case. Also, larger inter-atomic distances with a more solvent-exposed phosphate of the 3'-flanking residue prevent C-H...O pseudo-hydrogen bond contacts between the latter and the subsequent G as observed for a conventional type of V-loop, with possible implications for other intermolecular interactions (Fig. 16B, bottom).

Due to a lessening of conformational restraints exerted by the V-shaped loop, few quadruplexes with a less compact architecture, e.g., with more flexible 1 nt or 2 nt V-loops or with neighboring bulges adjacent to the loop, may feature conformers located slightly outside of either of the two characteristic conformational clusters.^{100,101} Of note, foldback bulges in two recently reported parallel-stranded quadruplexes enforce a single *syn* conformer for the 3'-linked G of a central discontinuous G-tract (see above and Fig. 13B).^{84,85} As a result, the loop following this G-core residue with its inverted backbone orientation may be viewed as a 1 nt V-shaped loop rather than a regular propeller-type loop.

Finally, a unique two-layered anti-parallel quadruplex derived from the *AGRO100* aptamer combines a 1 nt V-shaped loop with a 3'-terminal domain forming characteristic base pairing alignments.⁵⁴ The 3'-peripheral motif progresses along two sharp U-turns to form two additional layers composed of a T-T-G triad and a G-T base pair capping the 5'-outer tetrad (Fig. 17). It is attached by a non-terminal guanosine to the G4 core, filling a vacant G4 position in a snapback-type arrangement.

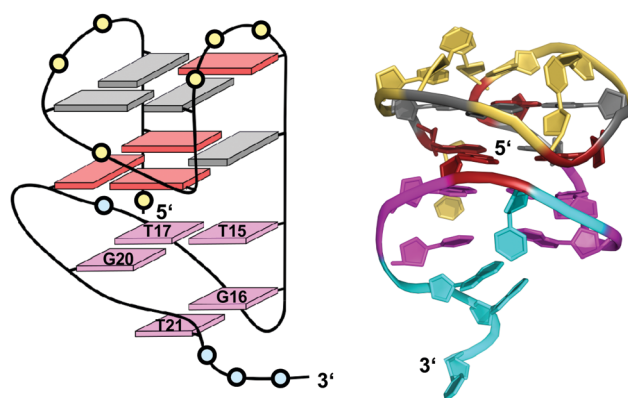


Fig. 17 Schematic representation and three-dimensional solution structure for a V-loop quadruplex of the *AGRO100* derived sequence $d(TG_2TTGTG_2TTTG_2TGTTG_2TG_2T)$ in K^+ solution (20 mM KP_i + 70 mM KCl, pH 7.0; $T_m = 39$ °C; PDB 6JCD);⁵⁴ *anti*- and *syn*-guanines of the G-tetrad core are colored grey and red, respectively; loop and flanking residues of the G4 core structure and of the 3'-peripheral motif are colored yellow and cyan, respectively; T-T-G triad and G-T base pair, magenta.



The compact structural domain of the peripheral sequence was shown to possibly serve as a modular unit, able to replace a diagonal snapback loop in other G4 structures. Notably, in contrast to most V-loop structures the unusual 1 nt V-loop spanning two tetrad planes features a 5'-anchoring guanosine in *anti* conformation.

Summary and perspectives

G-rich sequences covering a wide range of G-runs of different number and length have been demonstrated to fold into four-stranded G-quadruplexes. As surveyed in this review, G-deficiencies as a result of short G-tracts may be compensated by unique structural motifs supplementing vacant sites. These include V- and D-shaped loops, snapback loops, bulges, and interlocked species. On the other hand, additional stacking interactions through G4 association but also through capping base triads or base pairs may considerably stabilize a particular G4 architecture. Formation of corresponding motifs will depend on the particular sequence context and on the superposition of additional interactions that are often difficult to anticipate based on the primary sequence. However, these interactions may even be strong enough to induce folding into a non-conventional structure even if the sequence allowed folding into a canonical topology.

Based on the rapidly growing number of reported crystallographic and NMR G-quadruplex structures, our knowledge on the structural organization and stabilizing inherent interactions of unusual structural motifs has considerably grown. Thus, we are beginning to recognize and understand major contributors to particular topological features. These include tertiary interactions between different domains of the folded quadruplex to form triads, base pairs, or non-canonical tetrads but also more subtle conformational preferences of individual residues with their often decisive impact on equilibria between G4 conformers being close in energy. Sequences that fold into non-canonical G4 structures featuring interrupted G-columns have already been successfully designed.^{38,99,111} However, whereas our ability to decipher the code that relates a G-rich oligonucleotide sequence with four closely spaced G₃-tracts to a preferred G4 topology increases, understanding and predicting the folding of irregular G-rich sequences still poses a challenge.

Detailed insight into folding pathways and into interactions enforcing particular structural motifs will be important for the identification of G-rich fragments prone to G-quadruplex formation and also for a successful engineering of quadruplex architectures for various technology-based G4 applications. On the other hand, non-canonical G4 structures offer additional opportunities in their specific targeting for both pharmaceutical and technological purposes. The majority of known G4 ligands binds through stacking interactions onto a G-tetrad. Also, attempts to increase selectivity with less off-target effects based on different groove dimensions or loop conformations has only brought limited success to-date. Exploiting various non-conventional structural motifs may in fact expand our

toolbox for achieving more selectivity. G-deficient quadruplexes with a vacant site constitute promising targets for a specific and high-affinity binding of bifunctional ligands that are guided by a covalently linked guanine base. Also, interrupted G-tracts with their opening may potentially support insertion of planar ligands between tetrad planes of the quadruplex. Finally, quadruplex–duplex junctions are expected to be widespread in a cellular environment either through hairpin-type loop domains within the G4 architecture or through G-quadruplexes extruding from a B-type duplex as anticipated for oncogenic promoter sequences. Such interfaces between different structural domains have attracted growing interest in recent years and may provide for unique target sites for G4 drugs.^{112–114} In fact, binding to quadruplex–duplex junctions by appropriate ligands has already been shown to be guided by strong interactions, making junctions one of several promising target sites for the future design of more efficient G4-binding drugs.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support (grant no. WE 1933/15-1).

Notes and references

- 1 E. Y. N. Lam, D. Beraldi, D. Tannahill and S. Balasubramanian, *Nat. Commun.*, 2013, **4**, 1796.
- 2 A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J.-M. Perrier-Cornet, J. M. Y. Wong and D. Monchaud, *J. Am. Chem. Soc.*, 2015, **137**, 8521–8525.
- 3 R. Hänsel-Hertsch, M. D. Antonio and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 279–284.
- 4 S. Neidle, *Nat. Rev. Chem.*, 2017, **1**, 0041.
- 5 I. Russo Krauss, A. Merlino, C. Giancola, A. Randazzo, L. Mazzarella and F. Sica, *Nucleic Acids Res.*, 2011, **39**, 7858–7867.
- 6 V. R. de Soultrait, P.-Y. Lozach, R. Altmeyer, L. Tarrago-Litvak, S. Litvak and M. L. Andréola, *J. Mol. Biol.*, 2002, **324**, 195–203.
- 7 A.-M. Chiorcea-Paquim, R. Eritja and A. M. Oliveira-Brett, *J. Nucleic Acids*, 2018, 5307106.
- 8 J. L. Neo, K. Kamaladasan and M. Uttamchandani, *Curr. Pharm. Des.*, 2012, **18**, 2048–2057.
- 9 N. Borovok, N. Iram, D. Zikich, J. Ghabboun, G. I. Livshits, D. Porath and A. B. Kotlyar, *Nucleic Acids Res.*, 2008, **36**, 5050–5060.
- 10 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908–2916.
- 11 A. K. Todd, M. Johnston and S. Neidle, *Nucleic Acids Res.*, 2005, **33**, 2901–2907.



- 12 A. B. Sahakyan, V. S. Chambers, G. Marsico, T. Santner, M. D. Antonio and S. Balasubramanian, *Sci. Rep.*, 2017, **7**, 14535.
- 13 E. Puig Lombardi and A. Londoño-Vallejo, *Nucleic Acids Res.*, 2020, **48**, 1–15.
- 14 D. Bhattacharyya, G. M. Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**, 38.
- 15 M. Webba da Silva, *Chem. – Eur. J.*, 2007, **13**, 9738–9745.
- 16 F. Fogolari, H. Haridas, A. Corazza, P. Viglino, D. Corà, M. Caselle, G. Esposito and L. E. Xodo, *BMC Struct. Biol.*, 2009, **9**, 64.
- 17 S. A. Dvorkin, A. I. Karsisiotis and M. Webba da Silva, *Sci. Adv.*, 2018, **4**, eaat3007.
- 18 B. Karg, L. Haase, A. Funke, J. Dickerhoff and K. Weisz, *Biochemistry*, 2016, **55**, 6949–6955.
- 19 X. Tong, W. Lan, X. Zhang, H. Wu, M. Liu and C. Cao, *Nucleic Acids Res.*, 2011, **39**, 6753–6763.
- 20 X. Cang, J. Šponer and T. E. Cheatham, *Nucleic Acids Res.*, 2011, **39**, 4499–4512.
- 21 J. Šponer, A. Mládek, N. Špačková, X. Cang, T. E. Cheatham and S. Grimme, *J. Am. Chem. Soc.*, 2013, **135**, 9785–9796.
- 22 P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2004, **126**, 16405–16415.
- 23 N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, 2008, **130**, 10208–10216.
- 24 M. Cheng, Y. Cheng, J. Hao, G. Jia, J. Zhou, J. L. Mergny and C. Li, *Nucleic Acids Res.*, 2018, **46**, 9264–9275.
- 25 A. Piazza, X. Cui, M. Adrian, F. Samazan, B. Heddi, A. T. Phan and A. G. Nicolas, *eLife*, 2017, **6**, e26884.
- 26 M. Marušič, R. N. Veedu, J. Wengel and J. Plavec, *Nucleic Acids Res.*, 2013, **41**, 9524–9536.
- 27 T. Fujii, P. Podbevšek, J. Plavec and N. Sugimoto, *J. Inorg. Biochem.*, 2017, **166**, 190–198.
- 28 V. Kuryavyi, A. T. Phan and D. J. Patel, *Nucleic Acids Res.*, 2010, **38**, 6757–6773.
- 29 M. Trajkovski, M. Webba da Silva and J. Plavec, *J. Am. Chem. Soc.*, 2012, **134**, 4132–4141.
- 30 M. Webba da Silva, M. Trajkovski, Y. Sannohe, N. Ma'ani Hessari, H. Sugiyama and J. Plavec, *Angew. Chem., Int. Ed.*, 2009, **48**, 9167–9170.
- 31 B. Karg, S. Mohr and K. Weisz, *Angew. Chem., Int. Ed.*, 2019, **58**, 11068–11071.
- 32 A. Bugaut and S. Balasubramanian, *Biochemistry*, 2008, **47**, 689–697.
- 33 P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2004, **126**, 16405–16415.
- 34 A. Guédin, J. Gros, P. Alberti and J. L. Mergny, *Nucleic Acids Res.*, 2010, **38**, 7858–7868.
- 35 S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J. L. Mergny and A. T. Phan, *J. Am. Chem. Soc.*, 2012, **134**, 5807–5816.
- 36 J. Marquevielle, C. Robert, O. Lagrabette, M. Wahid, A. Bourdoncle, L. E. Xodo, J.-L. Mergny and G. F. Salgado, *Nucleic Acids Res.*, 2020, **48**, 9336–9345.
- 37 D. J. E. Yue, K. W. Lim and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 11462–11465.
- 38 K. W. Lim and A. T. Phan, *Angew. Chem., Int. Ed.*, 2013, **52**, 8566–8569.
- 39 Y. Chen, P. Agrawal, R. V. Brown, E. Hatzakis, L. Hurley and D. Yang, *J. Am. Chem. Soc.*, 2012, **134**, 13220–13223.
- 40 K. W. Lim, Z. J. Khong and A. T. Phan, *Biochemistry*, 2014, **53**, 247–257.
- 41 E. Butovskaya, B. Heddi, B. Bakalar, S. N. Richter and A. T. Phan, *J. Am. Chem. Soc.*, 2018, **140**, 13654–13662.
- 42 R. Jin, B. L. Gaffney, C. Wang, R. A. Jones and K. J. Breslauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 8832–8836.
- 43 L. Petraccone, E. Erra, V. Esposito, A. Randazzo, L. Mayol, L. Nasti, G. Barone and C. Giancola, *Biochemistry*, 2004, **43**, 4877–4884.
- 44 R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 3745–3749.
- 45 I. Smirnov and R. H. Shafer, *Biochemistry*, 2000, **39**, 1462–1468.
- 46 P. Schultze, R. F. Macaya and J. Feigon, *J. Mol. Biol.*, 1994, **235**, 1532–1547.
- 47 L. Martino, A. Virno, A. Randazzo, A. Virgilio, V. Esposito, C. Giancola, M. Bucci, G. Cirino and L. Mayol, *Nucleic Acids Res.*, 2006, **34**, 6653–6662.
- 48 K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 4301–4309.
- 49 Z. Zhang, J. Dai, E. Veliath, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2009, **38**, 1009–1021.
- 50 M. Lenarčič Živković, J. Rozman and J. Plavec, *Angew. Chem., Int. Ed.*, 2018, **57**, 15395–15399.
- 51 A. Kettani, S. Bouaziz, W. Wang, R. A. Jones and D. J. Patel, *Nat. Struct. Biol.*, 1997, **4**, 382–389.
- 52 A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin and D. J. Patel, *J. Mol. Biol.*, 2000, **301**, 129–146.
- 53 V. Kuryavyi, A. Kettani, W. Wang, R. Jones and D. J. Patel, *J. Mol. Biol.*, 2000, **295**, 455–469.
- 54 T. H. A. Truong, F. R. Winnerdy and A. T. Phan, *Angew. Chem., Int. Ed.*, 2019, **58**, 13834–13839.
- 55 B. Bakalar, B. Heddi, E. Schmitt, Y. Mechulam and A. T. Phan, *Angew. Chem., Int. Ed.*, 2019, **58**, 2331–2335.
- 56 F. R. Winnerdy, B. Bakalar, A. Maity, J. J. Vandana, Y. Mechulam, E. Schmitt and A. T. Phan, *Nucleic Acids Res.*, 2019, **47**, 8272–8281.
- 57 N. Q. Do, W. J. Chung, T. H. A. Truong, B. Heddi and A. T. Phan, *Nucleic Acids Res.*, 2017, **45**, 7487–7493.
- 58 W. J. Chung, B. Heddi, E. Schmitt, K. W. Lim, Y. Mechulam and A. T. Phan, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 2729–2733.
- 59 A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones and D. J. Patel, *J. Mol. Biol.*, 2000, **297**, 627–644.
- 60 Y. Mishima, H. Kaizu and R. Kominami, *J. Biol. Chem.*, 1997, **272**, 26578–26584.
- 61 M. Heller, E. Flemington, E. Kieff and P. Deininger, *Mol. Cell. Biol.*, 1985, **5**, 457–465.
- 62 A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi and M. Katahira, *J. Mol. Biol.*, 2001, **313**, 255–269.



- 63 A. Matsugami, T. Okuizumi, S. Uesugi and M. Katahira, *J. Biol. Chem.*, 2003, **278**, 28147–28153.
- 64 N. Q. Do, K. W. Lim, M. H. Teo, B. Heddi and A. T. Phan, *Nucleic Acids Res.*, 2011, **39**, 9448–9457.
- 65 L. Haase, B. Karg and K. Weisz, *ChemBioChem*, 2019, **20**, 985–993.
- 66 X. M. Li, K. W. Zheng, J. Y. Zhang, H. H. Liu, Y. De He, B. F. Yuan, Y. H. Hao and Z. Tan, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 14581–14586.
- 67 F. R. Winnerdy, P. Das, B. Heddi and A. T. Phan, *J. Am. Chem. Soc.*, 2019, **141**, 18038–18047.
- 68 Y. De, He, K. W. Zheng, C. J. Wen, X. M. Li, J. Y. Gong, Y. H. Hao, Y. Zhao and Z. Tan, *J. Am. Chem. Soc.*, 2020, **142**, 11394–11403.
- 69 X. M. Li, K. W. Zheng, Y. H. Hao and Z. Tan, *Angew. Chem., Int. Ed.*, 2016, **55**, 13759–13764.
- 70 B. Heddi, N. Martín-Pintado, Z. Serimbetov, T. M. A. Kari and A. T. Phan, *Nucleic Acids Res.*, 2016, **44**, 910–916.
- 71 K. B. Wang, J. Dickerhoff, G. Wu and D. Yang, *J. Am. Chem. Soc.*, 2020, **142**, 5204–5211.
- 72 A. T. Phan, V. Kuryavyi, H. Y. Gaw and D. J. Patel, *Nat. Chem. Biol.*, 2005, **1**, 167–173.
- 73 J. Jana and K. Weisz, *Chem. – Eur. J.*, 2020, **26**, 17242–17251.
- 74 A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.
- 75 D. Wei, G. N. Parkinson, A. P. Reszka and S. Neidle, *Nucleic Acids Res.*, 2012, **40**, 4691–4700.
- 76 B. Pan, Y. Xiong, K. Shi and M. Sundaralingam, *Structure*, 2003, **11**, 1423–1430.
- 77 B. Pan, K. Shi and M. Sundaralingam, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 3130–3134.
- 78 M. Meier, A. Moya-Torres, N. J. Krahn, M. D. McDougall, G. L. Orriss, E. K. S. McRae, E. P. Booy, K. McEleney, T. R. Patel, S. A. McKenna and J. Stetefeld, *Nucleic Acids Res.*, 2018, **46**, 5319–5331.
- 79 V. T. Mukundan, N. Q. Do and A. T. Phan, *Nucleic Acids Res.*, 2011, **39**, 8984–8991.
- 80 V. T. Mukundan and A. T. Phan, *J. Am. Chem. Soc.*, 2013, **135**, 5017–5028.
- 81 H. Martadinata and A. T. Phan, *Biochemistry*, 2014, **53**, 1595–1600.
- 82 B. De Nicola, C. J. Lech, B. Heddi, S. Regmi, I. Frasson, R. Perrone, S. N. Richter and A. T. Phan, *Nucleic Acids Res.*, 2016, **44**, 6442–6451.
- 83 A. Sengar, J. J. Vandana, V. S. Chambers, M. Di Antonio, F. R. Winnerdy, S. Balasubramanian and A. T. Phan, *Nucleic Acids Res.*, 2019, **47**, 1564–1572.
- 84 M. Lenarčič Živković, J. Rozman and J. Plavec, *Molecules*, 2020, **25**, 4867.
- 85 T. Q. N. Nguyen, K. W. Lim and A. T. Phan, *Nucleic Acids Res.*, 2020, **48**, 10567–10575.
- 86 M. Gajarský, M. Lenarčič Živković, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomáška, J. Šponer, J. Plavec and L. Trantírek, *J. Am. Chem. Soc.*, 2017, **139**, 3591–3594.
- 87 A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R. B. Darnell and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2011, **18**, 796–804.
- 88 Y. Krishnan-Ghosh, D. Liu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2004, **126**, 11009–11016.
- 89 T. C. Marsh and E. Henderson, *Biochemistry*, 1994, **33**, 10718–10724.
- 90 T. C. Marsh, J. Vesenka and E. Henderson, *Nucleic Acids Res.*, 1995, **23**, 696–700.
- 91 K. Bose, C. J. Lech, B. Heddi and A. T. Phan, *Nat. Commun.*, 2018, **9**, 1959.
- 92 M. Webba da Silva, *Biochemistry*, 2003, **42**, 14356–14365.
- 93 M. Webba da Silva, *Biochemistry*, 2005, **44**, 3754–3764.
- 94 N. Ma'Ani Hessari, L. Spindler, T. Troha, W. C. Lam, I. Drevněšek-Olenik and M. Webba Da Silva, *Chem. – Eur. J.*, 2014, **20**, 3626–3630.
- 95 H. T. Le, M. C. Miller, R. Buscaglia, W. L. Dean, P. A. Holt, J. B. Chaires and J. O. Trent, *Org. Biomol. Chem.*, 2012, **10**, 9393–9404.
- 96 V. Rauser and E. Weinhold, *ChemBioChem*, 2020, **21**, 2445–2448.
- 97 N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin and D. J. Patel, *J. Mol. Biol.*, 2001, **311**, 1063–1079.
- 98 A. T. Phan, V. Kuryavyi, J.-B. Ma, F. Aurélie, M.-L. Andréola and D. J. Patel, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 634–639.
- 99 A. T. Phan and N. Q. Do, *Nucleic Acids Res.*, 2013, **41**, 2683–2688.
- 100 A. Maity, F. R. Winnerdy, W. D. Chang, G. Chen and A. T. Phan, *Nucleic Acids Res.*, 2020, **48**, 3315–3327.
- 101 M. Adrian, D. J. Ang, C. J. Lech, B. Heddi, A. Nicolas and A. T. Phan, *J. Am. Chem. Soc.*, 2014, **136**, 6297–6305.
- 102 J. T. Nielsen, K. Arar and M. Petersen, *Angew. Chem., Int. Ed.*, 2009, **48**, 3099–3103.
- 103 L. Haase, J. Dickerhoff and K. Weisz, *Chem. – Eur. J.*, 2020, **26**, 524–533.
- 104 L. Haase and K. Weisz, *Chem. Commun.*, 2020, **56**, 4539–4542.
- 105 L. Haase and K. Weisz, *Nucleic Acids Res.*, 2020, **48**, 10555–10566.
- 106 Y. Liu, W. Lan, C. Wang and C. Cao, *J. Biol. Chem.*, 2018, **293**, 8947–8955.
- 107 M. Marušič and J. Plavec, *Molecules*, 2019, **24**, 1294.
- 108 V. Kuryavyi and D. J. Patel, *Structure*, 2010, **18**, 73–82.
- 109 M. Črnugelj, P. Šket and J. Plavec, *J. Am. Chem. Soc.*, 2003, **125**, 7866–7871.
- 110 C. Wan, W. Fu, H. Jing and N. Zhang, *Nucleic Acids Res.*, 2019, **47**, 1544–1556.
- 111 D. J. Y. Tan, P. Das, F. R. Winnerdy, K. W. Lim and A. T. Phan, *Chem. Commun.*, 2020, **56**, 5897–5900.
- 112 T. Q. N. Nguyen, K. W. Lim and A. T. Phan, *Sci. Rep.*, 2017, **7**, 11969.
- 113 S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando and H. Sugiyama, *Chem. – Eur. J.*, 2018, **24**, 4428–4435.
- 114 Y. M. Vianney, P. Preckwinkel, S. Mohr and K. Weisz, *Chem. – Eur. J.*, 2020, **26**, 16910–16922.



Article 2

G-Quadruplex Formation in a Putative Coding Region of White Spot Syndrome Virus: Structural and Thermodynamic Aspects

Yoanes Maria Vianney,^[a] Maria Goretti M. Purwanto,^[b] and Klaus Weisz*^[a]

White spot disease (WSD) is one of the most devastating viral infections of crustaceans caused by the white spot syndrome virus (WSSV). A conserved sequence *WSSV131* in the DNA genome of WSSV was found to fold into a polymorphic G-quadruplex structure. Supported by two mutant sequences with single G→T substitutions in the third G₄ tract of *WSSV131*, circular dichroism and NMR spectroscopic analyses demonstrate folding of the wild-type sequence into a three-tetrad parallel topology comprising three propeller loops with a major 1:3:1 and a minor 1:2:2 loop length arrangement. A thermodynamic analysis of quadruplex formation by differential scanning calorimetry (DSC) indicates a thermodynamically more stable 1:3:1 loop isomer. DSC also revealed the formation of additional highly stable multimeric species with populations depending on potassium ion concentration.

syndrome virus (accession number KY827813), we found a conserved sequence with putative G4-forming ability termed *WSSV131* (Figure 1). Its first G tract is located three bases downstream the template strand of the open reading frame (ORF) *WSV131*, encoding a putative yet still unknown protein.^[12,13]

The circular dichroism (CD) spectrum of *WSSV131* exhibits positive and negative amplitudes at ~265 and ~240 nm, typical for a quadruplex with parallel topology and exclusive homopolymorphic stacking interactions (Figure 2A). The imino proton

name	sequence
<i>WSSV131</i>	5'-TCT <u>GGGAGGG</u> AA <u>GGG</u> GAGGGTTA-3' 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
<i>WSSV131-G13T</i>	5'-TCT <u>GGGAGGG</u> AA <u>TGG</u> GAGGGTTA-3'
<i>WSSV131-G16T</i>	5'-TCT <u>GGGAGGG</u> AA <u>GGG</u> TAGGGTTA-3'

Figure 1. *WSSV131* wild-type and mutant sequences; G tracts are underlined and positions 13 and 16 marked in red.

G-quadruplexes (G4s) are secondary structures of DNA, formed by repeated runs of contiguous guanosine residues. They are widely found throughout different organisms but also occur in viral genomes.^[1,2] G4s have been shown to be involved in the regulation of gene expression, either acting within the regulatory region or the gene itself.^[3-5] Consequently, formation of G-quadruplexes in the viral functional genome may contribute to viral mortality and G-quadruplex-stabilizing ligands may be employed for viral control.^[6-9] The white spot syndrome virus has emerged as one of the most common and most devastating pathogens for farmed crustaceans such as shrimp.^[10,11] WSSV infects all major shrimp species and has caused huge economic losses in the aquaculture industry worldwide due to the lack of effective treatments. Screening the genome of the white spot

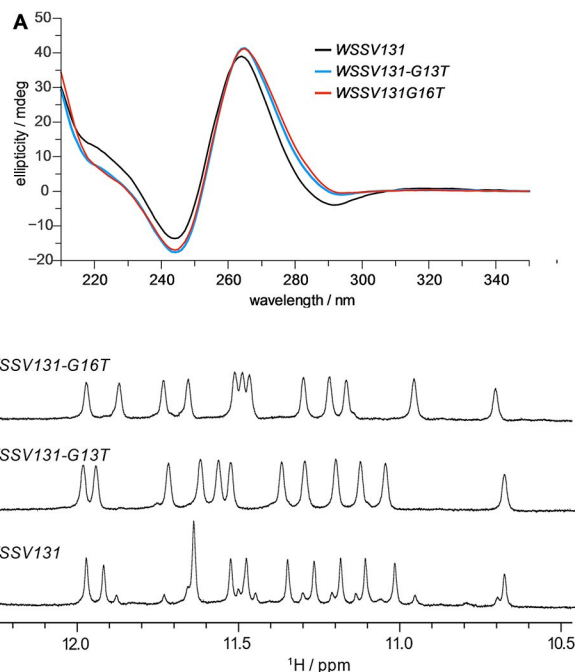


Figure 2. A) CD spectra (20 mM potassium phosphate, pH 7.0, 100 mM KCl, 20 °C) and B) imino proton NMR spectral region (10 mM potassium phosphate, pH 7.0, 25 °C) for *WSSV131* and its G13T and G16T mutants. Note that topologies are conserved under both buffer conditions (Figures S1 and S2 in the Supporting Information).

[a] Y. M. Vianney, Dr. K. Weisz
Institute of Biochemistry
Universität Greifswald
Felix-Hausdorff Str. 4,
17489 Greifswald (Germany)
E-mail: weisz@uni-greifswald.de

[b] Dr. M. G. M. Purwanto
Faculty of Biotechnology
Universitas Surabaya
Kalirungkut Str., Surabaya
60293 (Indonesia)

Supporting information for this article can be found under <https://dx.doi.org/10.1002/cbic.202100064>.

© 2021 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

spectral region of a 1D NMR spectrum for *WSSV131* reveals a set of 12 major imino resonances between 10.5–12.0 ppm characteristic for Hoogsteen hydrogen-bonded guanines arranged in a G-tetrad (Figure 2B). However, additional minor species account for ~30% of the total population. Polymorphism may arise due to the third tract composed of four G residues to potentially form 1:3:1 and 1:2:2 loop isomers. To lock structures to a single loop arrangement, single-site mutants G13T and G16T mimicking 1:3:1 and 1:2:2 loop isomers were evaluated and compared to the wild-type sequence (Figure 1). CD spectra of the mutants featured highly similar signatures typical of a parallel fold (Figure 2A). Notably, imino proton spectral regions for both mutants indicate a single quadruplex comprising twelve imino resonances with their superposition closely matching native *WSSV131* imino signals with the G13T mutant in excess (Figure 2B). These results suggest the presence of a major 1:3:1 and a minor 1:2:2 loop isomer for polymorphic *WSSV131*.

A more detailed NMR structural analysis of the native *WSSV131* sequence employed standard strategies to assign the major loop isomer. Continuous sugar-base NOE connectivities are observed between the 5'-terminal T and the first G-column, between the fourth G-column and the 3' terminus, as well as between guanines constituting the two central G tracts (Figure 3A). Two prominent crosspeaks in the H8/6-H1'/H5 spectral region observed at long but also shorter mixing times are identified as A11 H8-H1' and C2 H6-H5 contacts through a $^1\text{H},^{13}\text{C}$ HSQC experiment (Figure S3). Imino-H8 and imino-imino NOE contacts identified guanines involved in G-tetrad formation (Figure 3B, C). There is no indication of any *syn*-guanine

within the four G tracts in the NOESY spectra in line with the observation of exclusively upfield-shifted guanine ^{13}C resonances in the HSQC spectrum characteristic for *anti* conformers (Figure S3). The all-*anti* G-quadruplex matches a parallel topology with G-columns linked by three propeller-type loops.

Intra-tetrad H8-H1 crosspeaks determine the polarity of the G-tetrads, that is, the direction when going from H-bond donor to H-bond acceptor (Figure 3C). Thus, polarities of 5'-tetrad, central, and 3'-tetrad follow $\text{G}_4 \rightarrow \text{G}_8 \rightarrow \text{G}_{14} \rightarrow \text{G}_{18}$, $\text{G}_5 \rightarrow \text{G}_9 \rightarrow \text{G}_{15} \rightarrow \text{G}_{19}$, and $\text{G}_6 \rightarrow \text{G}_{10} \rightarrow \text{G}_{16} \rightarrow \text{G}_{20}$, respectively. Polarity is additionally confirmed by typical imino-imino NOE contacts (Figure 3B). Strong crosspeaks connect sequential Gs within the same G tract. In addition, intra-tetrad crosspeaks can be seen between neighboring G columns, for example, G10-G6 and G20-G6, and inter-tetrad contacts connect imino protons in adjacent tetrads, such as, G9-G6 and G10-G15.

There is also a weak NOE crosspeak between A7 H8 of the first propeller loop and preceding G6 H2'. Likewise, residues of the second propeller loop can be identified through NOE walks based on H8/H6-H1'/H2'/H3' contacts from A11 through G13. Another weak sequential contact from G16 H2' to A17 H8 of the third propeller loop also allows assignment of G16 as being the last residue of the third column (Figure S4). Taken together, experimental findings clearly confirm a major *WSSV131* G4 with parallel topology and a central 3-nt propeller loop (Figure 4, for a compilation of chemical shifts, see Table S1). A corresponding more detailed 2D NMR spectral analysis on both G13T and G16T mutants confirmed their parallel fold and the same 1:3:1 loop arrangement for *WSSV131-G13T* as identified for the major species of the wild-type sequence (Figure S5 and S6).

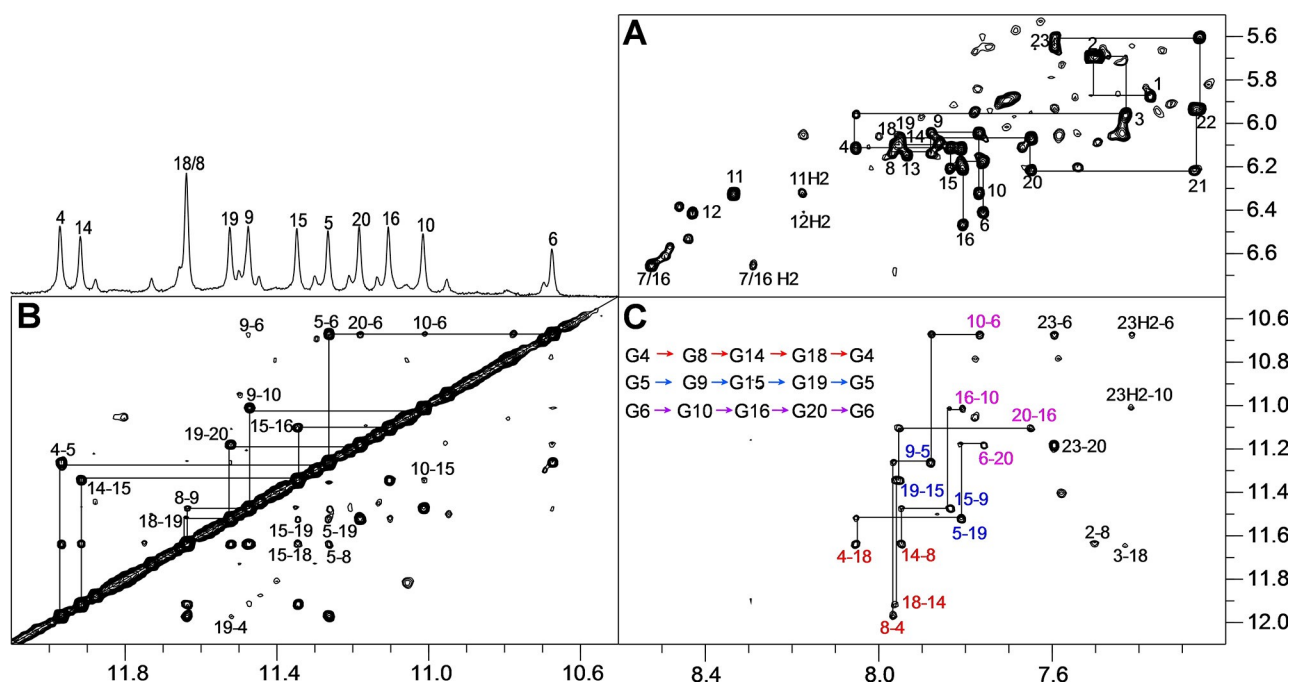


Figure 3. 2D NOE spectral regions of *WSSV131* acquired with a 300 ms mixing time at 25 °C. A) H6/8(ω_2)-H1'/H5(ω_1) spectral region with continuous NOE walks along G tracts and 5'- and 3'-overhang sequences. B) Imino-imino crosspeaks with sequential connectivities traced along the G tracts. C) Intra- and inter-tetrad H8(ω_2)-imino(ω_1) crosspeaks; tetrad polarities as determined from intra-tetrad NOE contacts (marked in red for the 5'-tetrad, in blue for the central tetrad, and in magenta for the 3'-tetrad) are given in the inset.

Initial UV melting studies on the *WSSV131* G13T and G16T mutants, each forming a single loop isomer, revealed slow kinetics of (un)folding at 10 mM K^+ whereas at 120 mM K^+ melting profiles shifted to temperatures too high for the observation of a well-defined high-temperature baseline within the temperature window accessible (not shown). To nevertheless obtain information on their folding thermodynamics, quadruplex formation was analyzed by DSC in a pressurized cell

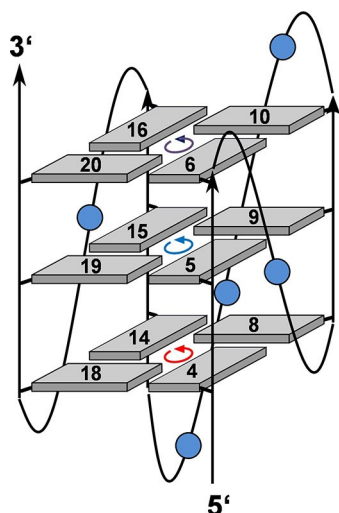


Figure 4. Topology of the *WSSV131* major quadruplex species with tetrad polarities indicated and loop residues represented by circles; residues of the G-core are numbered.

allowing sample heating up to 110 °C without irreversible DNA decomposition. Thermodynamic equilibrium upon heating was verified for both sequences by a match in melting profiles determined with two different heating rates (not shown). After proper baseline correction, DSC data were fitted with a non-two-state model assuming negligible changes in molar heat capacity $\Delta C_p^{[14]}$ to yield T_m as well as calorimetric and van't Hoff molar enthalpies ΔH_{cal}° and ΔH_{vH}° (Figure 5, Table S2).^[15]

In a buffer solution with 120 mM K^+ the major *WSSV131-G13T* quadruplex exhibits a T_m of 68.1 °C, 3.4 °C higher than that of the *WSSV131-G16T* mutant. The higher melting 1:3:1 propeller loop arrangement of *WSSV131-G13T* agrees with systematic studies on the length of propeller loops, reporting a propensity of short first and third loops with a longer central loop.^[16] Interestingly, the less favored *WSSV131-G16T* quadruplex shows another high-temperature transition centered at 103 °C but not fully completed within the experimental temperature range. To shift transitions towards lower temperatures, DSC measurements were also performed in solutions with 90 mM K^+ . Again, in contrast to *WSSV131-G13T* the G16T mutant exhibits an additional high-temperature transition shifted to 99.3 °C but with noticeably reduced height. Based primarily on gel electrophoresis and size exclusion chromatography, high-melting multimeric species have previously been suggested to coexist in particular for parallel-stranded quadruplexes with short loops.^[17–20] However, they have not been observed and characterized by calorimetric methods so far. Taking into account a growing population and faster folding of multimers with increasing K^+ concentration, ΔH_{cal}° for the monomer

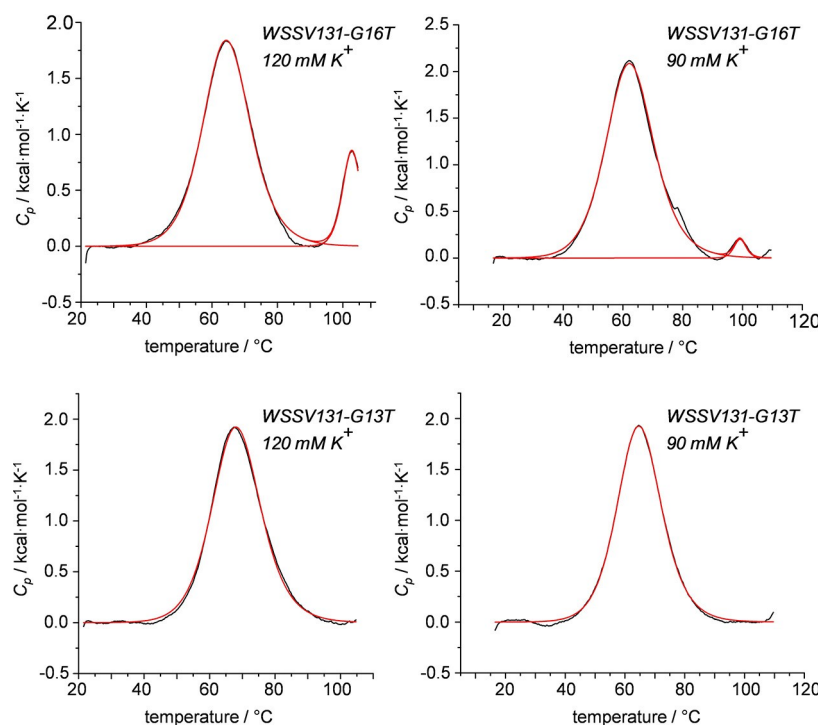


Figure 5. Representative DSC thermograms with a 0.5 °C/min heating rate of *WSSV131* sequences (50 μ M). Melting profiles for *WSSV131-G16T* (top) in the presence of 120 mM K^+ (left) and 90 mM K^+ (right). Melting profiles for *WSSV131-G13T* (bottom) in the presence of 120 mM K^+ (left) and 90 mM K^+ (right). Fitted curves based on a non-two-state model with $\Delta H_{cal}^\circ \neq \Delta H_{vH}^\circ$ are shown in red.

transition at lower temperature is expected to be significantly underestimated in line with a $\Delta H_{\text{vH}}^{\circ}/\Delta H_{\text{cal}}^{\circ} > 1$ in a 120 mM K^{+} buffer. On the other hand, the small population of multimers at 90 mM K^{+} does hardly compromise $\Delta H_{\text{cal}}^{\circ}$ for *WSSV131-G16T*, yielding a $\Delta H_{\text{vH}}^{\circ}/\Delta H_{\text{cal}}^{\circ}$ ratio of about 1 in agreement with a two-state melting transition for the monomer. Because $\Delta H_{\text{vH}}^{\circ}$ is independent of concentration and only depends on the shape of the DSC curve, $\Delta H_{\text{vH}}^{\circ}$ is expected to provide a reliable value for the enthalpy of (un)folding given a two-state transition under equilibrium conditions. With a $\Delta H_{\text{vH}}^{\circ}$ of -47.0 and -46.1 kcal/mol for *WSSV131-G13T* and *WSSV131-G16T* at 120 mM K^{+} as well as -45.9 and -42.9 kcal/mol at 90 mM K^{+} , folding of the more stable G13T mutant seems more exothermic by 1 and 3 kcal/mol. It should also be noted that despite observation of only one DSC transition, $\Delta H_{\text{vH}}^{\circ}/\Delta H_{\text{cal}}^{\circ}$ was found to be > 1 under both salt conditions for *WSSV131-G13T*. This strongly suggests formation of corresponding multimers with even higher thermal stability when compared to those of *WSSV131-G16T*, escaping their detection in the DSC experiment. Indeed, formation of higher-order assemblies for both G16T and G13T mutant sequences are also indicated by native gel electrophoresis experiments in a 90 mM K^{+} buffer (Figure S7).

Taken together, a well-defined parallel quadruplex, highly stable under physiological salt conditions, is formed in a sequence located downstream of a putative coding region in the WSSV viral genome. Such a G4 could possibly be used to regulate viral gene expression and offers the opportunity to ultimately control WSSV infection through the use of G4-binding and G4-stabilizing ligands. Detection and characterization of such G4-prone sequences in the viral genome open new avenues for the target design of antiviral drugs directed against WSSV in crustacean farming.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (INST 292/138-1). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: G-quadruplexes · NMR spectroscopy · thermodynamics · white spot syndrome virus

- [1] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.
- [2] M. Metifiot, S. Amrane, S. Litvak, M. L. Andreola, *Nucleic Acids Res.* **2014**, *42*, 12352–12366.
- [3] K. Derecka, G. D. Balkwill, T. P. Garner, C. Hodgman, A. P. F. Flint, M. S. Searle, *Biochemistry* **2010**, *49*, 7625–7633.
- [4] P. Zizza, C. Cingolani, S. Artuso, E. Salvati, A. Rizzo, C. D'Angelo, M. Porru, B. Pagano, J. Amato, A. Randazzo, E. Novellino, A. Stoppacciaro, E. Gilson, G. Stassi, C. Leonetti, A. Biroccio, *Nucleic Acids Res.* **2016**, *44*, 1579–1590.
- [5] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598.
- [6] R. Perrone, E. Butovskaya, D. Daelemans, G. Palù, C. Pannecouque, S. N. Richter, *J. Antimicrob. Chemother.* **2014**, *69*, 3248–3258.
- [7] S. R. Wang, Q. Y. Zhang, J. Q. Wang, X. Y. Ge, Y. Y. Song, Y. F. Wang, X. D. Li, B. S. Fu, G. H. Xu, B. Shu, P. Gong, B. Zhang, T. Tian, X. Zhou, *Cell Chem. Biol.* **2016**, *23*, 1113–1122.
- [8] S. R. Wang, Y. Q. Min, J. Q. Wang, C. X. Liu, B. S. Fu, F. Wu, L. Y. Wu, Z. X. Qiao, Y. Y. Song, G. H. Xu, Z. G. Wu, G. Huang, N. F. Peng, R. Huang, W. X. Mao, S. Peng, Y. Q. Chen, Y. Zhu, T. Tian, X. L. Zhang, X. Zhou, *Sci. Adv.* **2016**, *2*, e1501535.
- [9] E. Butovskaya, P. Soldà, M. Scalabrin, M. Nadai, S. N. Richter, *ACS Infect. Dis.* **2019**, *5*, 2127–2135.
- [10] A. Sánchez-Paz, *Vet. Res.* **2010**, *41*, 43.
- [11] M. Lillehammer, R. Banger, M. Salazar, S. Vela, E. C. Erazo, A. Suarez, J. Cock, M. Rye, N. A. Robinson, *Sci. Rep.* **2020**, *10*, 20571.
- [12] Y. Han, F. Li, L. Xu, F. Yang, *Vet. Res.* **2017**, *48*, 1–11.
- [13] Z. Li, F. Li, Y. Han, L. Xu, F. Yang, *J. Virol.* **2016**, *90*, 842–850.
- [14] B. Pagano, A. Randazzo, I. Fotticchia, E. Novellino, L. Petraccone, C. Giancola, *Methods* **2013**, *64*, 43–51.
- [15] J. B. Chaires, *Biophys. Chem.* **1997**, *64*, 15–23.
- [16] P. A. Rachwal, I. S. Findlow, J. M. Werner, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2007**, *35*, 4214–4222.
- [17] H. T. Le, M. C. Miller, R. Buscaglia, W. L. Dean, P. A. Holt, J. B. Chaires, J. O. Trent, *Org. Biomol. Chem.* **2012**, *10*, 9393–9404.
- [18] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica, *J. Am. Chem. Soc.* **2008**, *130*, 10208–10216.
- [19] V. Rausser, E. Weinhold, *ChemBioChem* **2020**, *21*, 2445–2448.
- [20] A. Guédin, J. Gros, P. Alberti, J. L. Mergny, *Nucleic Acids Res.* **2010**, *38*, 7858–7868.

Manuscript received: February 11, 2021
 Revised manuscript received: March 4, 2021
 Accepted manuscript online: March 12, 2021
 Version of record online: April 6, 2021

ChemBioChem

Supporting Information

G-Quadruplex Formation in a Putative Coding Region of White Spot Syndrome Virus: Structural and Thermodynamic Aspects

Yoanes Maria Vianney, Maria Goretti M. Purwanto, and Klaus Weisz*

Experimental Section

Sample preparation. Oligonucleotides were purchased from TIB MolBiol (Berlin, Germany), further purified by ethanol precipitation, dried, and dissolved in water. The concentration was measured through their absorbance at 260 nm at 80 °C using the extinction coefficient provided by the manufacturer. After drying, oligonucleotides were resuspended for NMR assignments in a 10 mM potassium phosphate buffer, pH 7.0, with 90% H₂O/10% D₂O. For other experiments, a 20 mM potassium phosphate buffer with 100 mM KCl, pH 7.0, was used unless otherwise stated in the text. Prior to usage, samples were annealed by heating to 90 °C for 5 minutes followed by slow cooling to room temperature.

CD spectroscopy. CD spectra were recorded on samples (5 μM) in 1-cm quartz cuvettes with a Jasco J-810 spectropolarimeter equipped with a Peltier thermostat (Jasco, Tokyo, Japan). Spectra were acquired with a wavelength range of 210-350 nm, a 1 nm bandwidth, a 50 nm/min scanning speed, 4 s response time and 5 accumulations. All spectra were recorded at 20 °C and finally blank-corrected.

NMR spectroscopy. NMR spectra were recorded with a Bruker Avance 600 MHz spectrometer equipped with an inverse ¹H/¹³C/¹⁵N/¹⁹F quadruple resonance cryoprobehead and z-field gradients. Proton chemical shifts were referenced to the water chemical shift at 25 °C while carbon chemical shifts were referenced to DSS by an indirect referencing method. 2D ¹H-¹H NOESY experiments were performed with a 300 ms and 80 ms mixing time. 2D ¹H-¹³C HSQC experiments were recorded with 7500 Hz in the indirect dimension for recording aromatic carbon resonances. All spectra were acquired at 25 °C with oligonucleotide concentrations of about 0.3 mM.

Differential Scanning Calorimetry (DSC). DSC experiments were carried out with a VP-DSC instrument (Malvern Instruments, United Kingdom). The oligonucleotide solution (50 μM) was heated at a heating rate of 0.5 °C·min⁻¹ (three independent measurements) and 0.25 °C·min⁻¹ (two independent measurements) with a pre-equilibration time of 15 min. Buffer vs. buffer scan data were taken by refilling the sample cell in the same running cycle and subtracted from the sample data. After cubic baseline correction, data were fitted with a non-two-state model assuming $\Delta C_p = 0$ kcal·mol⁻¹·K⁻¹ to obtain T_m , ΔH°_{cal} , and ΔH°_{vH} .

Native Gel Electrophoresis. 100 μM of annealed oligonucleotide in a 20 mM potassium phosphate buffer, pH 7.0, supplemented with 70 mM KCl were mixed with glycerol-buffer (4:6) in a 1:1 v/v ratio. Electrophoresis experiments were performed on a 15% polyacrylamide gel (acrylamide:bis-acrylamide 19:1). Samples (1 nmol per lane) were loaded and separation was performed at room temperature in TBE buffer, pH 8.3, supplemented with 90 mM KCl. Gels were stained with 5 μM thiazole orange.

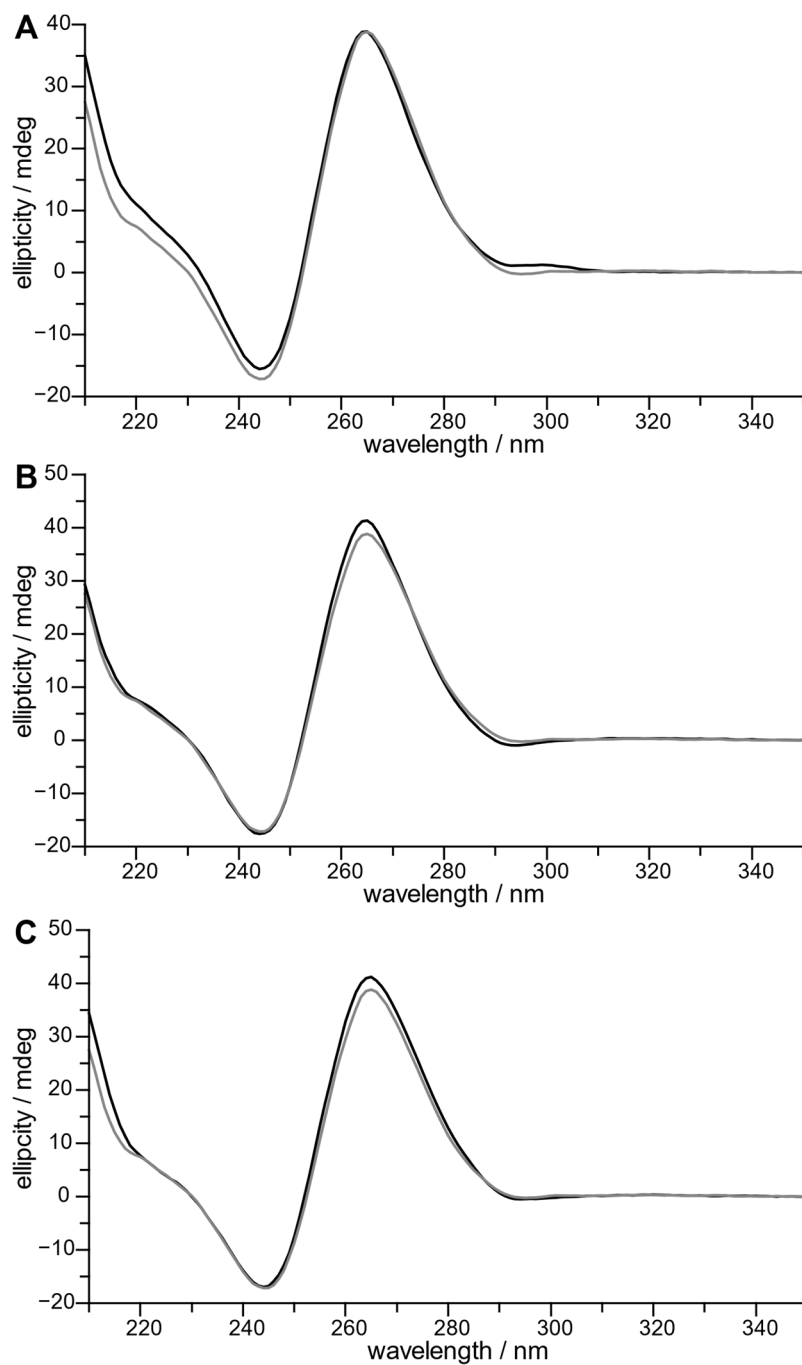


Figure S1. CD spectra for (A) *WSSV131*, (B) *WSSV131-G13T*, and (C) *WSSV131-G16T* in 20 mM potassium phosphate buffer + 100 mM KCl, pH 7.0 (black), and 10 mM potassium phosphate buffer, pH 7.0 (grey). No significant changes are observed for the CD profiles.

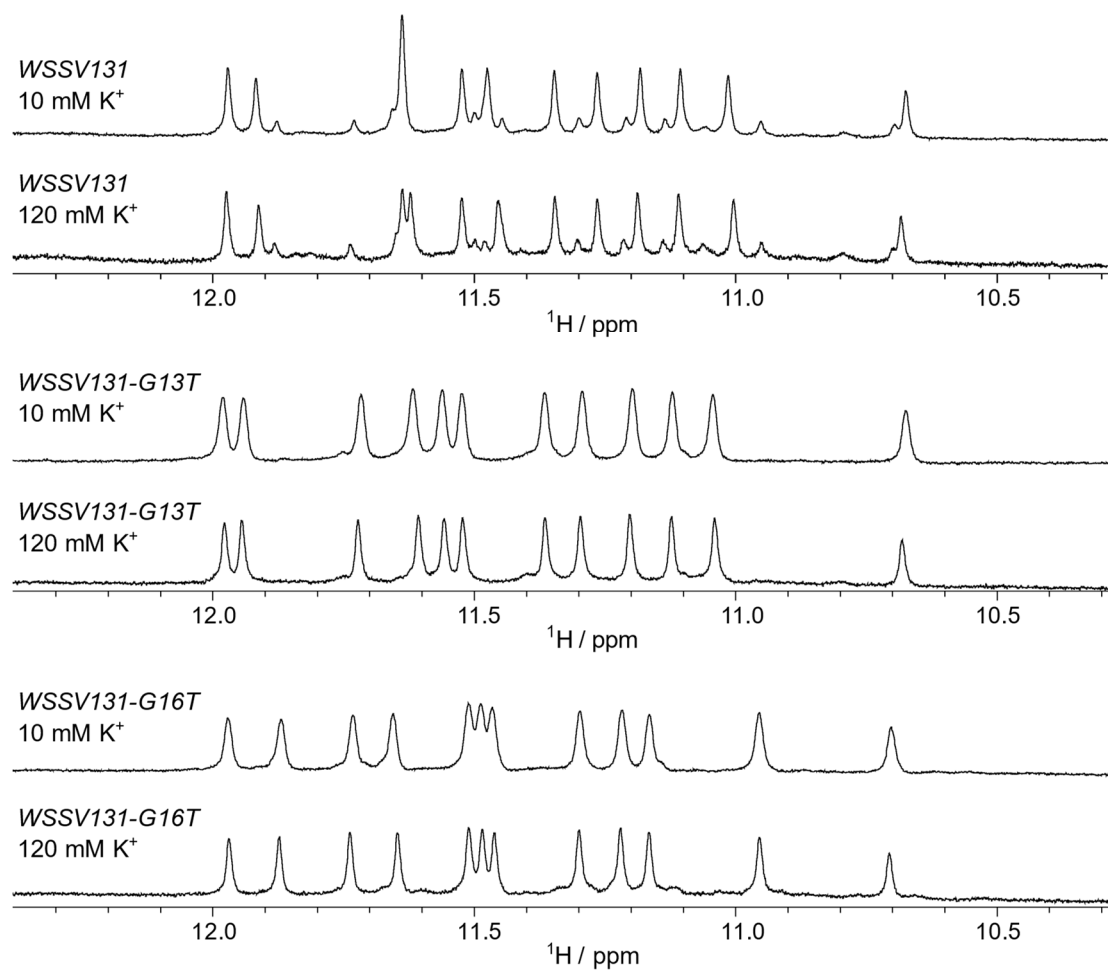


Figure S2. Imino proton spectral region of *WSSV131* (top), *WSSV131-G13T* (center), and *WSSV131-G16T* (bottom) variants (0.3 mM) in 10 mM potassium phosphate buffer, pH 7, and in 20 mM potassium phosphate buffer + 100 mM KCl, pH 7.0.

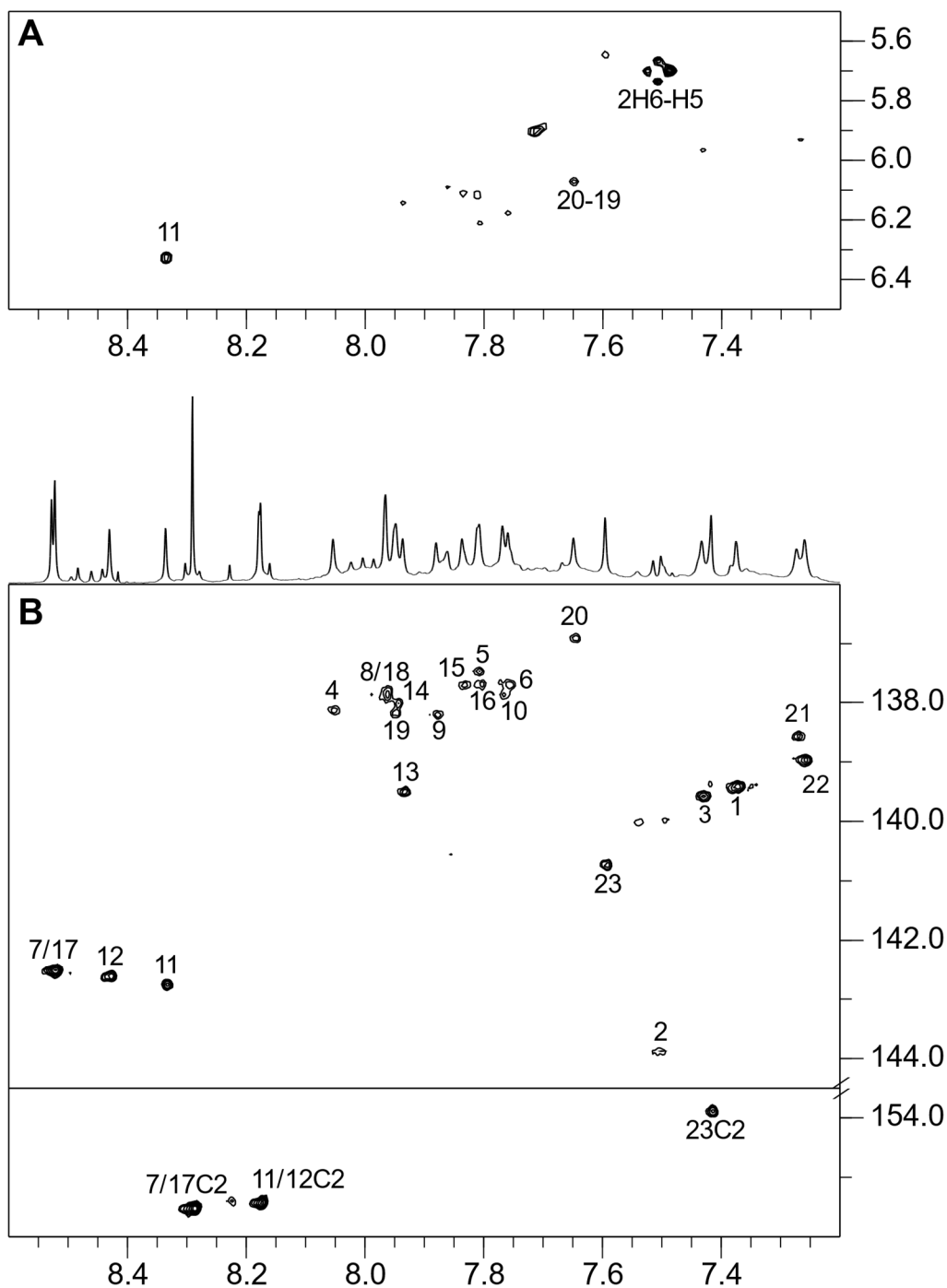


Figure S3. (A) H6/H8 (ω_2) - H1' (ω_1) NOESY spectral region of *WSSV131* at 80 ms mixing time. (B) H6/H8/H2 (ω_2) - C6/C8 (ω_1) spectral region of a ^1H - ^{13}C HSQC spectrum of *WSSV131* (0.3 mM). Both spectra were recorded at 25 °C in 10 mM potassium phosphate buffer, pH 7.

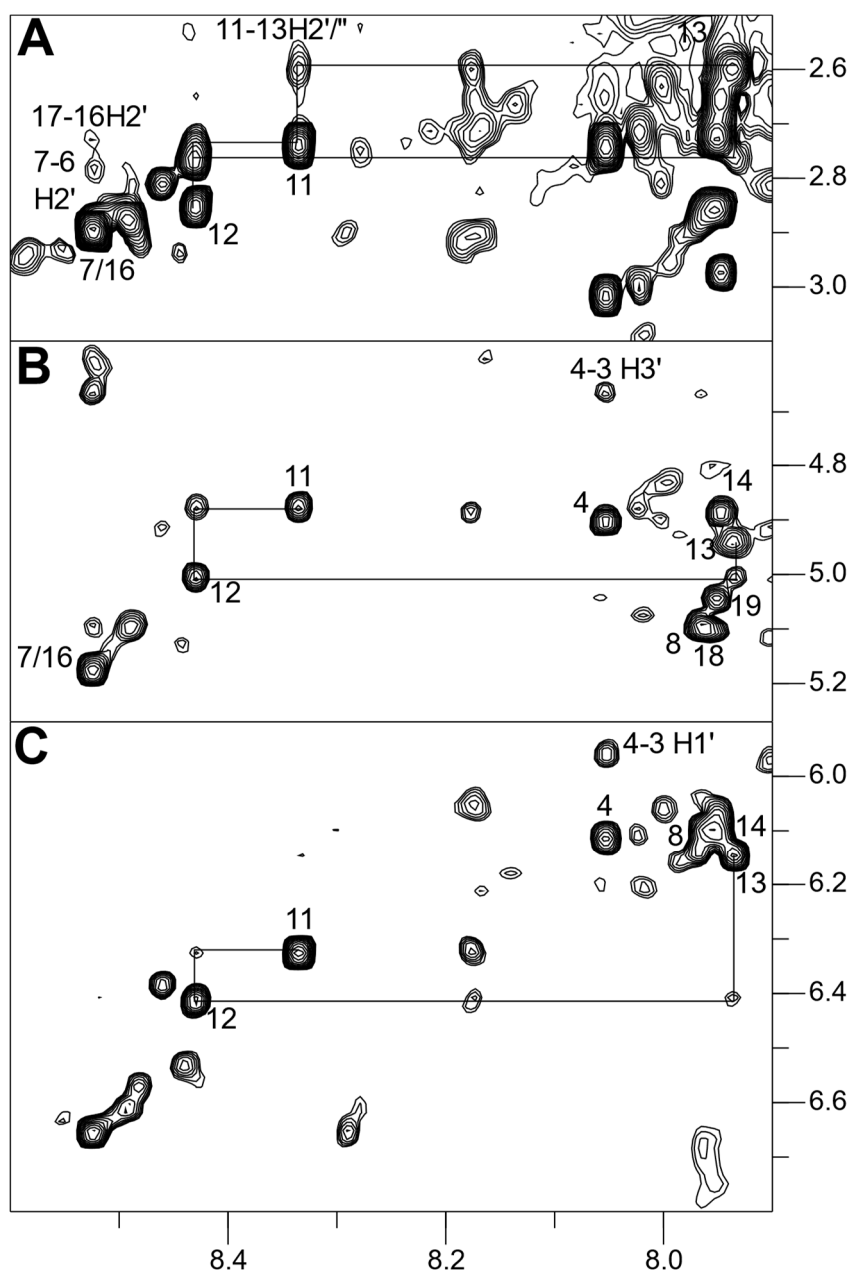


Figure S4. NOESY spectral regions of *WSSV131* showing H8/H6 - sugar contacts involving adenosine in the propeller loop. (A) H8/H6(ω_2) - H2'/H2''(ω_1) region, (B) H8/H6(ω_2) - H3'(ω_1) region, and (C) H8/H6(ω_2) - H1'(ω_1) region. Spectra were recorded at 25 °C in 10 mM potassium phosphate buffer, pH 7.

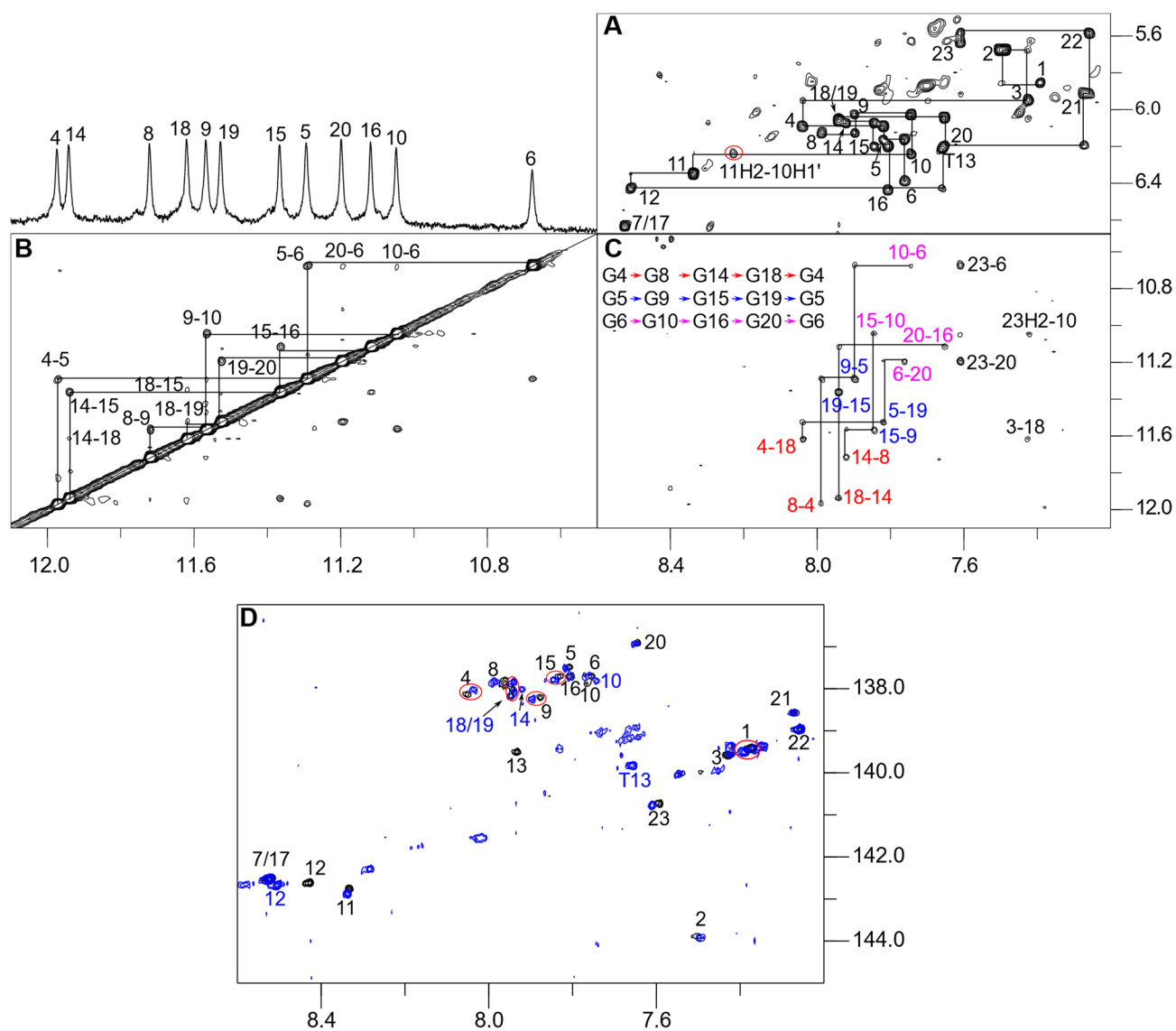


Figure S5. (A-C) 2D NOE spectral regions of *WSSV131-G13T* (0.3 mM) acquired with a 300 ms mixing time. (A) $H6/8(\omega_2)$ - $H1'/H5(\omega_1)$ spectral region with NOE walks along G-tracts and 5'- and 3'-overhang sequences. Loop isomer 1:3:1 is identified through NOE contacts from A11 H2 to G10 H1' (circled in red). (B) Imino - imino cross-peaks with sequential connectivities traced along the G-tracts. (C) Intra- and inter-tetrad $H8(\omega_2)$ - imino(ω_1) cross-peaks; tetrad polarities as determined from intra-tetrad NOE contacts (marked in red for the 5'-tetrad, in blue for the central tetrad, and in magenta for the 3'-tetrad) are given by the inset. (D) $H6/H8(\omega_2)$ - $C6/C8(\omega_1)$ spectral region of a 1H - ^{13}C HSQC spectrum of *WSSV131* (black) and *WSSV131-G13T* (0.3 mM, blue). Notably, most cross-peaks superimpose except for exchanged residue 13, indicating the same fold with a 1:3:1 loop arrangement. Both spectra were recorded at 25 °C in 10 mM potassium phosphate buffer, pH 7.

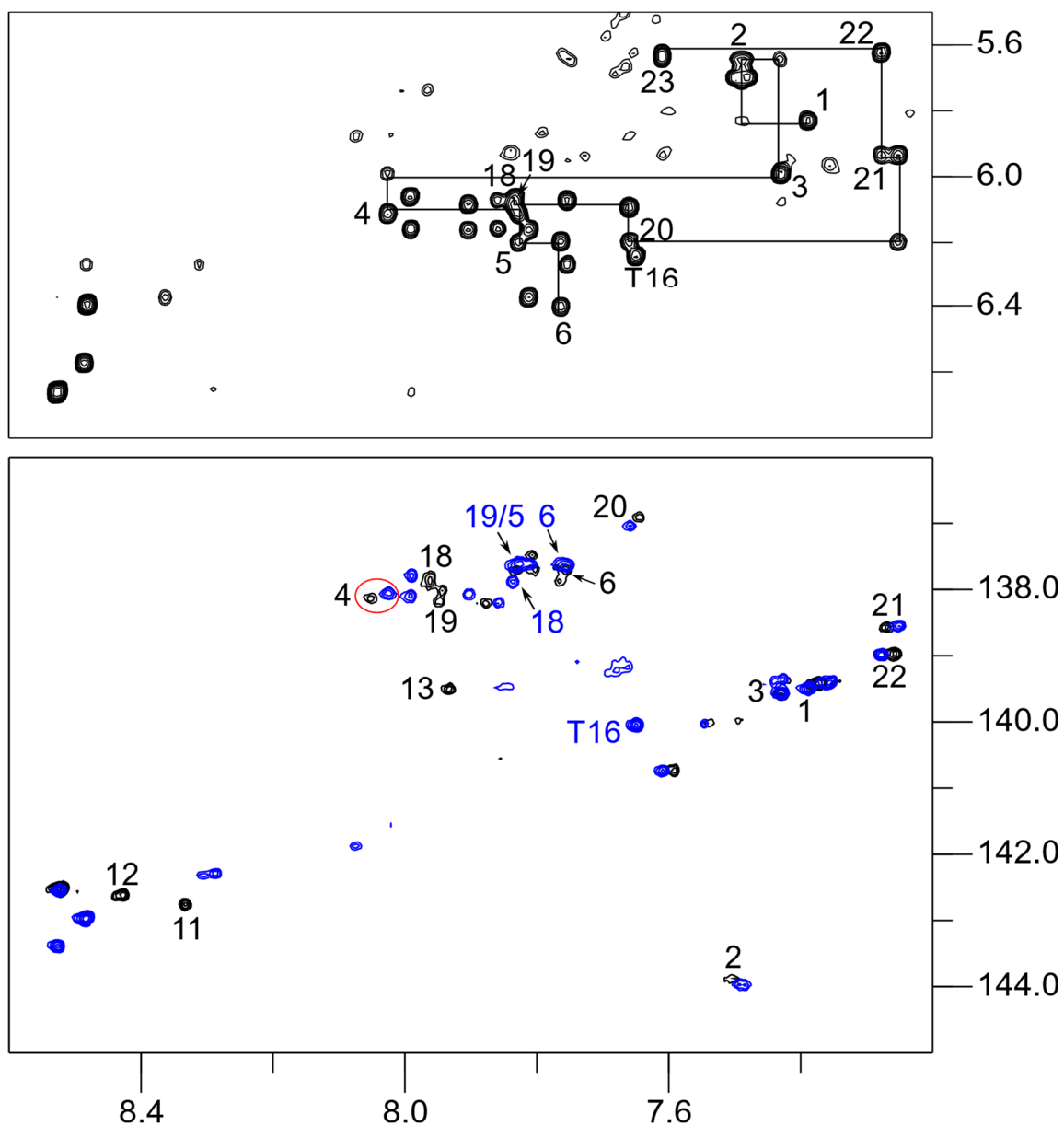


Figure S6. (Top) H6/H8(ω_2) - H1'(ω_1) 2D NOE spectral region of *WSSV131-G16T* (0.3 mM) acquired with a 300 ms mixing time. (Bottom) H6/H8(ω_2) - C6/C8(ω_1) spectral region of a ^1H - ^{13}C HSQC spectrum of *WSSV131* (black) and *WSSV131-G16T* (0.3 mM, blue). More pronounced chemical shift differences are noticeable for several residues. Both spectra were recorded at 25 °C in 10 mM potassium phosphate buffer, pH 7.

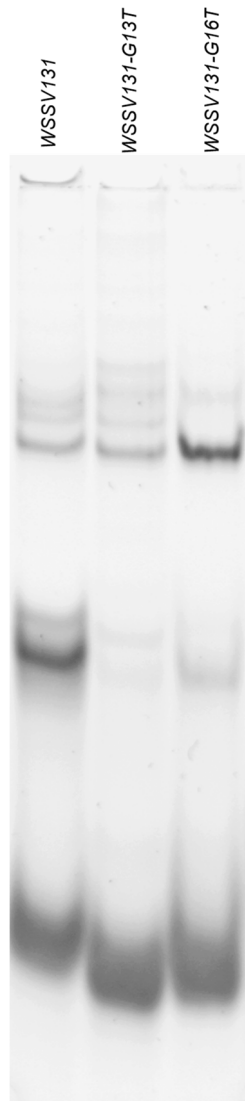


Figure S7. Non-denaturing polyacrylamide gel electrophoresis of *WSSV131* variants in TBE buffer supplemented with 90 mM KCl. Fast, moderate, and slow migrating bands are suggested to be due to monomer, dimers, and multimers.

Table S1. Chemical shifts (ppm) for the major quadruplex of *WSSV131*^a

	Imino	H6/H8	H5/H2/Me	H1'	H2'	H2''	H3'	C6/C8
T1	n.d.	7.37	1.64	5.88	1.96	2.18	4.46	139.42
C2	-	7.50	5.70	5.69	1.89	2.31	4.61	143.89
T3	n.d.	7.43	1.55	5.96	2.13	2.40	4.70	139.58
G4	11.97	8.05	-	6.11	2.74	3.02	4.93	138.14
G5	11.26	7.81	-	6.18	2.65	2.93	5.04	137.48
G6	10.68	7.76	-	6.41	2.78	2.60	5.12	137.72
A7	-	8.52	8.29	6.66	2.90	2.90	5.21	142.52
G8	11.64	7.97	-	6.14	2.46	2.89	5.13	137.89
G9	11.47	7.88	-	6.05	2.61	2.76	4.98	138.21
G10	11.01	7.77	-	6.32	2.61	2.57	4.92	137.88
A11	-	8.33	8.18	6.33	2.74	2.74	4.91	142.78
A12	-	8.43	8.17	6.41	2.85	2.77	5.04	142.64
G13	n.d.	7.94	-	6.12	^b	^b	4.98	139.51
G14	11.92	7.95	-	6.12	2.66	2.98	4.92	138.02
G15	11.35	7.84	-	6.21	2.73	2.95	5.04	137.71
G16	11.10	7.81	-	6.47	2.73	2.61	5.12	137.70
A17	-	8.52	8.29	6.66	2.90	2.90	5.21	142.52
G18	11.64	7.96	-	6.10	2.45	2.87	5.13	137.86
G19	11.52	7.95	-	6.07	^b	2.73	5.07	138.18
G20	11.18	7.65	-	6.22	2.59	2.80	4.95	136.91
T21	n.d.	7.27	1.55	5.94	2.09	2.37	4.74	138.57
T22	n.d.	7.26	1.62	5.61	1.66	2.14	4.44	138.97
A23	-	7.59	7.42	5.65	2.05	2.26	4.28	140.74

^aAt 25 °C in 10 mM potassium phosphate buffer, pH 7. ^bAmbiguous.

Table S2. DSC-derived thermodynamic parameters for the folding of *WSSV131* sequences.^a

sequence	T_m (°C)	$\Delta H_{\text{cal}}^{\circ}$	$\Delta H_{\text{vH}}^{\circ}$	T_m (°C)	$\Delta H_{\text{cal}}^{\circ}$	$\Delta H_{\text{vH}}^{\circ}$
		(kcal/mol) ^b	(kcal/mol)		(kcal/mol) ^b	(kcal/mol)
	20 mM KP _i (pH 7.0) + 70 mM K ⁺			20 mM KP _i (pH 7.0) + 100 mM K ⁺		
<i>WSSV131-G13T</i>	65.0 ± 0.1	-37.6 ± 0.9	-45.9 ± 0.3	68.1 ± 0.2	-37.2 ± 0.8	-47.0 ± 0.4
<i>WSSV131-G16T</i> ^c	61.9 ± 0.5	-42.6 ± 2.1	-42.9 ± 0.9	64.7 ± 0.3	-35.9 ± 0.7	-46.1 ± 0.7
	99.3 ± 0.2	-0.8 ± 0.2	-230.6 ± 30	102.7 ± 0.3	--- ^e	--- ^e

^aAverage value with standard deviation from three independent measurements. ^bApparent calorimetric molar enthalpy for a corresponding transition if $c_{\text{DNA}} = c_{\text{DNA, total}}$. ^cTwo transitions observed. ^dThree transitions observed. ^eNot definable due to incomplete melting.

Article 3

Reprinted with permission from Vianney, Y., Weisz, K. (2021). First tandem repeat of a potassium channel *KCNN4* folds into a V-loop G-quadruplex structure *Biochemistry*, 60, 17, 1337–1346. DOI: 10.1021/acs.biochem.1c00043. Copyright 2021 American Chemical Society

First Tandem Repeat of a Potassium Channel *KCNN4* Minisatellite Folds into a V-Loop G-Quadruplex Structure

Yoanes Maria Vianney and Klaus Weisz*

Cite This: *Biochemistry* 2021, 60, 1337–1346

Read Online

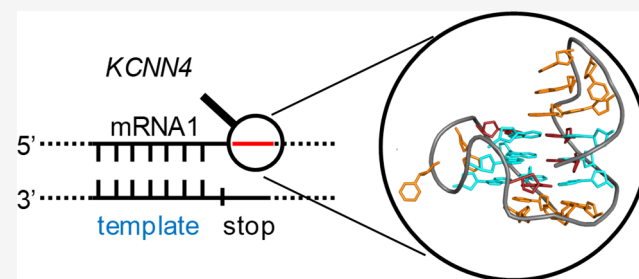
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The *KCNN4* gene encoding a potassium channel protein whose expression has been correlated with tumor progression was found to comprise a guanine-rich minisatellite region with the ability to form a putative G-quadruplex (G4). Given the suggested regulatory role of G4s in gene expression, G-quadruplex formation for the polymorphic first repeat of the minisatellite was studied by nuclear magnetic resonance spectroscopy. A stable G-quadruplex of a truncated mutant sequence was shown to represent one of several coexisting species of the wild-type sequence. The high-resolution structure features a noncanonical G4 with a broken G-column and a V-shaped loop. The presence of a 3'-flanking thymidine interacting with the lateral loop preceding the V loop seems to be critical for the formation of this G4 topology. On the contrary, an additional 5'-flanking residue disfavored but still allowed folding into the V-loop structure. The latter may therefore serve as a putative therapeutic target in strategies for G4-based modulation of *KCNN4* expression.



The alteration of the intracellular ion concentration constitutes a biomarker in cancer development. Pharmacological evidence indicates that membrane proteins responsible for ion transport play crucial roles in tumor progression. They are expressed aberrantly in almost all tumor types and contribute to all hallmarks of cancer.^{1–4} Due to their central role in proliferation, angiogenesis, and migration, ion channels are considered potential therapeutic and diagnostic targets in oncology. A statistical analysis of several cancer types with different patient data sets revealed that modified expression levels occur frequently in potassium ion channels such as *KCNH1*, *KCNE3*, *KCNE4*, and *KCNN4*.⁵ As a result of their overexpression, various families of potassium channels often found in cancer cells cause potassium ion outflux and concomitantly decrease the intracellular potassium ion concentration.

G-quadruplexes (G4s) are noncanonical secondary structures of nucleic acids that can be formed from several repeats of G-rich tracts. G4s have been reported to be involved in the control of various cellular processes, on the DNA or RNA level.⁶ Putative G-quadruplex-forming sequences can be found throughout the genome⁷ and are mostly related to the regulation of replication and the modulation of gene expression. They frequently occur in promoter,^{8–11} exon,^{12,13} intron,^{14–19} and minisatellite loci for homologous recombination in meiosis.^{20–22}

The formation of a G-quadruplex is driven by stabilization through the coordination of cations, mostly potassium, in its central cavity of negative electrostatic potential. Consequently, a change in the concentration and the type of cation may

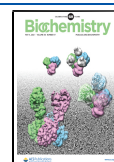
significantly impact the stability of a particular G-quadruplex topology.^{23,24} Because K^+ channels are overexpressed in aggressive cancer cells, intracellular potassium ion concentrations as low as 60 mM are found in transformed cells when compared to normal cells with approximately 100–150 mM K^+ .²⁵ As a result, gene expression may be modulated in response to changes in K^+ concentration during tumor progression through the altered stabilities of G-quadruplex structures within genomic sequences. This was demonstrated by the transcription efficiencies of a luciferase product that was regulated by a G-quadruplex-forming template sequence inserted upstream of the *Firefly* luciferase gene in a pGL3 plasmid. An siRNA-mediated silencing of *KCNH1*, a protein responsible for the outflux of potassium ions, decreased the level of expression of luciferase, demonstrating that the *KCNH1*-controlled K^+ concentration plays an important role in the expression of mRNAs from G-rich sequences.²⁵

In this study, a particularly interesting guanine-rich minisatellite was found 304 bp downstream of the transcription start site (TSS) in the sense strand of the *KCNN4* gene encoding a potassium ion channel protein. In fact, *KCNN4* may represent a potential biomarker for cancer prognosis and

Received: January 14, 2021

Revised: March 15, 2021

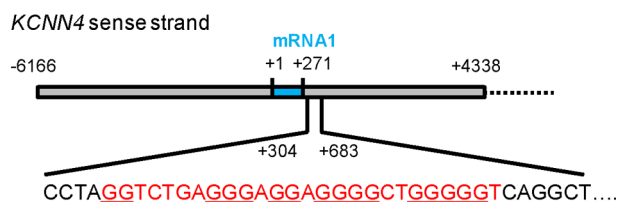
Published: April 12, 2021



is considered a promising therapeutic target.^{26–28} The G-rich minisatellite is located ~30 bp downstream of the first mRNA segment and includes 10 tandem repeats with a repeat sequence of ~37 bp harboring several mutations.

Here, we focused on the first repeat closest to the transcription site, encompassing five G-runs with more than two adjacent Gs (Figure 1). Unlike the other motifs, this first

A Chromosome 19



B

name	sequence
KNA-GGT	5'-GGTCTGAGGGAGGAGGGGCTGGGGGT-3'
KNA-G	5'-GGTCTGAGGGAGGAGGGGCTGGGG-3'
KNA	5'-GGTCTGAGGGAGGAGGGGCTGGG-3'
KNA-G13T	5'-GGTCTGAGGGAGTAGGGGCTGGG-3'
KNA-ΔG13	5'-GGTCTGAGGGAGAGGGGCTGGG-3'
KNA-T	5'-GGTCTGAGGGAGGAGGGGCTGGGT-3'
KNA-ΔG13-T	5'-GGTCTGAGGGAGAGGGGCTGGGT-3'
A-KNA-ΔG13-T	5'-AGGTCTGAGGGAGAGGGGCTGGGT-3'

Figure 1. (A) First repeat of the *KCNN4* minisatellite in the human genome harboring a G → A mutation at position 4. Studies are based on the red-colored 26-nucleotide DNA stretch with its five G-runs. (B) G-rich wild-type sequences and mutants from the *KCNN4* minisatellite. G-tracts comprising more than two contiguous guanines are underlined.

repeat of the minisatellite sequence deviates in featuring a single G → A mutation to leave a run of two guanines within its first G-tract. On the contrary, it also comprises runs of more than three guanines at its 3'-end. Formation of a stable noncanonical quadruplex with a broken G-column is demonstrated for this first repeat of the *KCNN4* minisatellite. This is enabled by a noncanonical backbone progression termed V-shaped loop that links the broken G-tract with an adjacent antiparallel column, bridging three G-tetrad layers.¹⁴ The importance of the 3'-overhang in driving the formation of such a topology is highlighted while discussing the relevance of a V-loop-containing fold within a genomic sequence context.

MATERIALS AND METHODS

G-Rich Sequences and Sample Preparation. Based on a statistical analysis by Biasiotta et al. of significantly modified expression levels for potassium ion channel genes in different tumors,⁵ QGRS mapper was initially employed to screen for G-quadruplex-forming sequence motifs.²⁹ Directed by the search results, we derived wild-type and mutated sequences from the first repeat of the *KCNN4* gene (+304; accession number NG_052672 starting from the 6471st base). Corresponding oligonucleotides were purchased from TIBMOLBIOL (Berlin, Germany). They were further purified by potassium acetate/ethanol precipitation, dried, dissolved in water, and stored at 4 °C until needed. The concentration of oligonucleotides was determined by measuring their absorbance A_{260} at 80 °C using molar extinction coefficients provided by the supplier. Samples

were dried, redissolved in buffer, heated at 90 °C for 5 min, and slowly cooled to room temperature for annealing.

Circular Dichroism (CD). CD spectra were recorded with a Jasco J-810 spectropolarimeter (Jasco, Tokyo, Japan) at 20 °C. Oligonucleotides (5 μM) were prepared in either a low-salt buffer [10 mM potassium phosphate (pH 7.0)] or a high-salt buffer [20 mM potassium phosphate (pH 7.0) supplemented with 100 mM KCl]. Spectra were recorded within a spectral range of 220–330 nm and with a 50 nm/min scanning speed, a 1 nm bandwidth, and five accumulations. All spectra were blank-corrected.

Ultraviolet (UV) Melting Experiments. Oligonucleotides (5 or 50 μM) were dissolved in 1.5 mL of either a low-salt buffer [10 mM potassium phosphate (pH 7.0)] or a high-salt buffer [20 mM potassium phosphate (pH 7.0) supplemented with 100 mM KCl]. UV melting curves were recorded with a Jasco V-650 spectrophotometer (Jasco) equipped with a Peltier thermostat. Temperature-dependent absorbance values were recorded at 295 nm with a bandwidth of 1 nm and a heating rate of 0.2 °C/min. Melting temperatures were determined by the first derivative of the absorbance versus temperature curves.

Nuclear Magnetic Resonance (NMR) Experiments.

Unless otherwise stated, NMR samples were prepared in 10 mM potassium phosphate buffer (pH 7.0) with a 90% H₂O/10% D₂O mixture. NMR spectra were acquired with a Bruker Avance 600 MHz spectrometer equipped with an inverse ¹H/¹³C/¹⁵N/¹⁹F quadruple resonance cryoprobehead and z-field gradients. Data were processed in TopSpin 4.0.7 and analyzed in CcpNmr V2.³⁰ Proton chemical shifts were referenced to the water chemical shift taking into account its temperature dependence at pH 7, while carbon chemical shifts were referenced to DSS through an indirect referencing method. For an initial screening, oligonucleotide concentrations varied between 0.25 and 0.33 mM, while for a complete resonance assignment, a concentration of 0.9 mM was employed. All one-dimensional (1D) ¹H NMR spectra were recorded with 64 transients. Two-dimensional (2D) NOESY experiments with a WATERGATE w5 water suppression scheme were performed with 2K × 1K data points and three different mixing times of 300, 150, and 80 ms with a recycle delay of 2 s. ¹H–¹³C HSQC experiments with a 3–9–19 water suppression scheme were performed with 4K × 500 data points, a 1 s recycle delay, and a spectral width of 7500 Hz to accommodate C6/8 and C2 ¹³C resonances in the indirect dimension. ¹H–¹³C HMBC spectra with a jump-and-return water suppression were acquired with 2K × 192 data points and a 1 s recycle delay and processed with 50% NUS in the indirect dimension. DQF-COSY spectra with water presaturation were recorded in 100% D₂O with 2K × 512 data points and a 2 s recycle delay. Prior to Fourier transformation, FIDs were zero-filled to 4K × 1K data points. Unless otherwise stated, all spectra were recorded at 25 °C.

Structure Calculations. One hundred starting structures of lowest energy were selected from 500 structures generated by a simulated annealing protocol in XPLOR-NIH 3.0.3.³¹ 2D nuclear Overhauser effect (NOE) cross-peak intensities were used for distance restraints. For non-exchangeable protons, distances were fixed to 2.9 ± 1.1 Å for strong cross-peaks, 4.0 ± 1.5 Å for intermediate cross-peaks, 5.5 ± 1.5 Å for weak cross-peaks, and 6.0 ± 1.5 Å for very weak cross-peaks. For exchangeable protons, distances were set to 4.0 ± 1.2 Å for strong cross-peaks, 5.0 ± 1.2 Å for intermediate cross-peaks,

and $6.0 \pm 1.2 \text{ \AA}$ for weak cross-peaks. For overlapped peaks, distances were set to $5.0 \pm 2.0 \text{ \AA}$. Glycosidic torsion angles were fixed in an *anti* ($170\text{--}310^\circ$) or *syn* conformation ($25\text{--}95^\circ$) except for residue G15, which was restrained to an extended torsion angle range ($170\text{--}330^\circ$). Sugar puckers were set to the south domain of pseudorotation ($144\text{--}180^\circ$) for all residues except G1, G6, G14, G15, T19, and T23. G14 and G15 were restrained to a north sugar pucker (pseudorotational angle of $0\text{--}90^\circ$). Additional planarity and hydrogen bond restraints were employed for G-tetrads.

Refinement was performed using AMBER16 with the parmbsc force field and OL15 modifications.³² The 100 starting structures were subjected to simulated annealing *in vacuo* to yield 20 converged structures. The system was initially equilibrated at 300 K for 5 ps followed by heating to 1000 K over 10 ps. After 30 ps, the system was cooled to 100 K within 45 ps and finally to 0 K within 10 ps. Restraint energies were set to $40 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for NOE-based distance restraints, $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for hydrogen bond-based distance restraints, $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ for dihedral angle restraints, and $30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for tetrad planarity restraints.

For a refinement in water, the structures were neutralized with potassium ions with two ions placed in the inner core of the G-quadruplex flanked by two adjacent tetrads. The system was hydrated with TIP3P water in a 10 \AA truncated octahedral box. Initial equilibration was performed with 500 steps of steepest descent minimization followed by another 500 steps of conjugate gradient minimization. The DNA was fixed with a force constant of $25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The system was heated from 100 to 300 K at constant volume during 10 ps, followed by further equilibrations with decreasing energy restraints of 5, 4, 3, 2, 1, and $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ under a constant pressure of 1 atm. The final simulation was performed at 1 atm and 300 K for 4 ns using only NOE and hydrogen bond distance restraints. The trajectory was averaged for the last 500 ps and minimized *in vacuo* to obtain 10 lowest-energy structures.

RESULTS AND DISCUSSION

Sequence Variants and Their Impact on Structural Polymorphism. The minisatellite of the *KCNN4* gene contains various repeats that can potentially fold into a quadruplex structure. Focusing on the first repeating unit, we initially selected a 26-nucleotide guanine-rich sequence termed *KNA-GGT* for structural studies (Figure 1A). Starting with a GG-tract and comprising five runs of two to five consecutive Gs, the G-rich fragment was expected to adopt various structures, thus posing a challenge for structural studies. Indeed, the imino proton spectral region exhibits a large number of overlapping and unresolved resonances typical of extensive structural polymorphism with various coexisting species (Figure 2). To selectively eliminate some of the potential folds through shortening, 3'-truncated sequences *KNA-G* with a GGGG-3' terminus and *KNA* with a GGG-3' terminus were prepared and subjected to NMR analysis. Whereas *KNA-G* was even inferior to the longer *KNA-GGT* wild-type sequence in terms of structural homogeneity, *KNA* showed four high-intensity imino resonances in addition to other imino signals. Therefore, additional mutants with substitutions and/or deletions as summarized in Figure 1B were synthesized and studied for their favored folding topology.

A single-site G13T mutant of *KNA* abolishes the third G-tract with its two neighboring guanosine residues but hardly

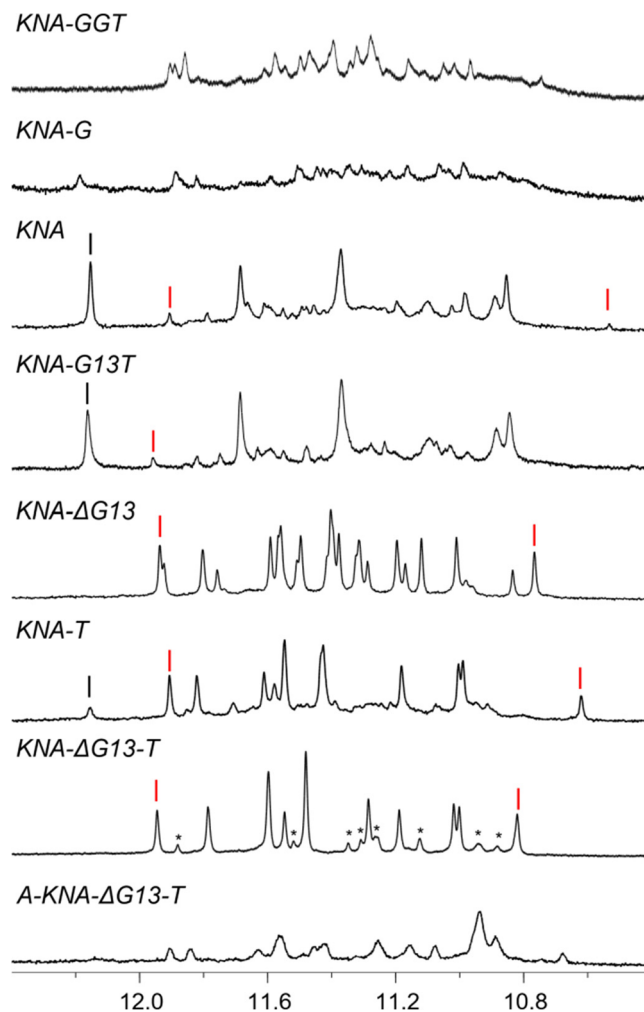


Figure 2. NMR imino proton spectral region of wild-type *KNA* and mutated sequences at $25 \text{ }^\circ\text{C}$. Imino resonances used as markers for two different quadruplex folds are indicated by black and red lines; signals of a minor species for *KNA-ΔG13-T* are marked by asterisks.

impacts folding. In contrast, a G13 deletion to give the 22-nucleotide sequence *KNA-ΔG13* exhibits a better resolved spectrum, indicating the coexistence of two similarly populated species. Remarkably, addition of a 3'-flanking T to the shortened *KNA* variant of the wild-type sequence gives a set of 12 major imino resonances together with weak signals of low-population minor species. Finally, the *KNA-ΔG13-T* sequence that combines a G13 deletion with a 3'-T addition provided the cleanest NMR spectrum appropriate for structural analyses.

Inspection and comparison of the NMR spectra identified several imino protons belonging to the same or to closely related quadruplex structures. Thus, a strong peak at 12.15 ppm in the imino proton spectral region of *KNA* partially and even completely disappears in *KNA-T* and *KNA-ΔG13-T*, respectively (Figure 2). On the contrary, two low-intensity peaks for *KNA* at 11.91 and 10.53 ppm were suggested to most likely match the most downfield- and most upfield-shifted imino signal of the *KNA-ΔG13-T* major species.

No information about coexisting folds can be obtained from spectra of *KNA-G*. However, inspection of 2D NMR spectra provides strong evidence that the major topology of the *KNA-ΔG13-T* mutant is also adopted by G4 species formed by the

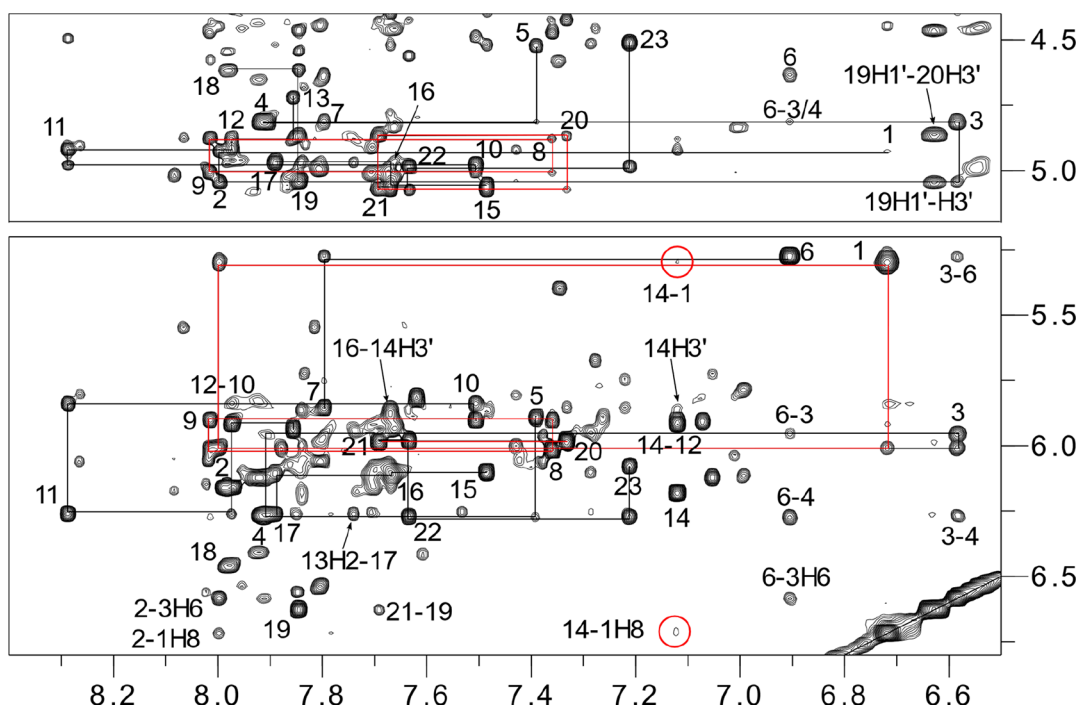


Figure 3. H8/H6(ω_2)–H3'(ω_1) (top) and H8/H6(ω_2)–H1'(ω_1) (bottom) 2D NOESY spectral regions of *KNA-ΔG13-T* (0.9 mM) acquired with a mixing time of 300 ms; NOE walks are traced by vertical and horizontal lines with rectangular cross-peak patterns characteristic of *syn-anti* steps (red). G14 H8–G1 H8 and G14 H8–G1 H1' contacts are marked by circles. Intranucleotide contacts are labeled with single residue numbers; otherwise, the second label refers to the resonance along ω_1 .

other sequences (Figures S1–S3). Thus, alignment of 2D NOESY and HSQC spectral regions revealed several nearly identical cross-peak patterns at the same chemical shift for the *KNA-ΔG13-T* major species and the various mutated sequences. It should be noted that a reevaluation of spectra after completion of the *KNA-ΔG13-T* structural analysis (see below) again corroborated the latter results in showing exact matches with several assigned *syn-* and *anti-G* ^{13}C and/or ^1H resonances. Consequently, mutations in the *KNA-ΔG13-T* variant do not redirect folding into a new topology but rather change populations of coexisting folds. Also, the impact of mutations hints at the importance of a 3'-terminal thymidine and of a putative looped-out intervening stretch between the G_3 - and G_4 -tracts.

CD spectra of all sequence variants in a 10 mM K^+ buffer solution show signatures that demonstrate the presence of both homopolar and heteropolar G-tetrad stacking interactions. These are in line with a (3+1) hybrid topology but also with a mixture of parallel and antiparallel structures in which a parallel topology seems to predominate in the case of *KNA-G* (Figure S4). Nondenaturing gel electrophoresis shows single unresolved monomer bands but also slowly migrating bands of low intensity indicative of multimeric species, especially for *KNA-GGT* and *KNA-G* exhibiting rather featureless NMR spectra (Figure S5). For *KNA-ΔG13-T* as well as for wild-type *KNA-GGT*, both CD and NMR spectra were also acquired at 120 mM K^+ , being compatible with an unaltered topology when compared to a low-salt buffer solution (Figure S6). Melting temperatures (T_m) of the *KNA-ΔG13-T* mutant in 10 and 120 mM K^+ were determined to be 46.8 and 64.4 °C, respectively (Figure S7). No significant hysteresis was observed for any melting curve, indicating fast (un)folding processes. Melting temperatures were also shown to be invariant against oligonucleotide concentration, demon-

strating formation of an intramolecular quadruplex species. Interestingly, addition of a 3'-flanking thymine has only negligible effect on the thermal stability of the major species with a similar T_m for *KNA-ΔG13-T* and *KNA-ΔG13* (Figure S7A,C). This hints at the 3'-T overhang to have only a moderate stabilizing effect on the major species but may disfavor the formation of an additional minor species.

Resonance Assignments. The *KNA-ΔG13-T* sequence folds into a major G-quadruplex amenable to a detailed NMR spectral analysis. However, resonances of a minor species can still be observed in the imino proton spectral region (Figure 2). We nevertheless succeeded in the unambiguous determination of the major fold without resorting to specific isotope labeling strategies. Initially, a ^1H – ^{13}C HSQC spectrum clearly identified H6/H8 resonances of the major species through their high-intensity cross-peaks (Figure S8A). In full agreement with corresponding cross-peak chemical shifts expected for *syn*-guanosines in the HSQC spectrum, H8–H1' cross-peaks with intensities comparable to or higher than those of thymine H6–Me contacts, as observed in a 2D NOESY spectrum acquired with short mixing times, revealed five G residues in a *syn*-glycosidic conformation (Figure S8B).

In the following, H6/H8–sugar proton spectral regions of 2D NOESY spectra recorded with longer mixing times were analyzed in more detail. Continuous NOE walks with concomitant residue assignments were observed between G1 and A7, G8 and A13, G15 and G17, and G20 and the 3'-terminal T23 (Figure 3 and Figure S9). There is a single break in the H8–H1' connectivities between T5 and G6. On the contrary, NOE contacts from G6 H8 to T3 H1', C4 H1', and T3 H6 link all residues between the 5'-terminal G2-tract and the following G3-tract, indicating a lateral-type progression of the five-nucleotide intervening TCTGA sequence. Given the unambiguous assignment of sequential G-tracts through NOE

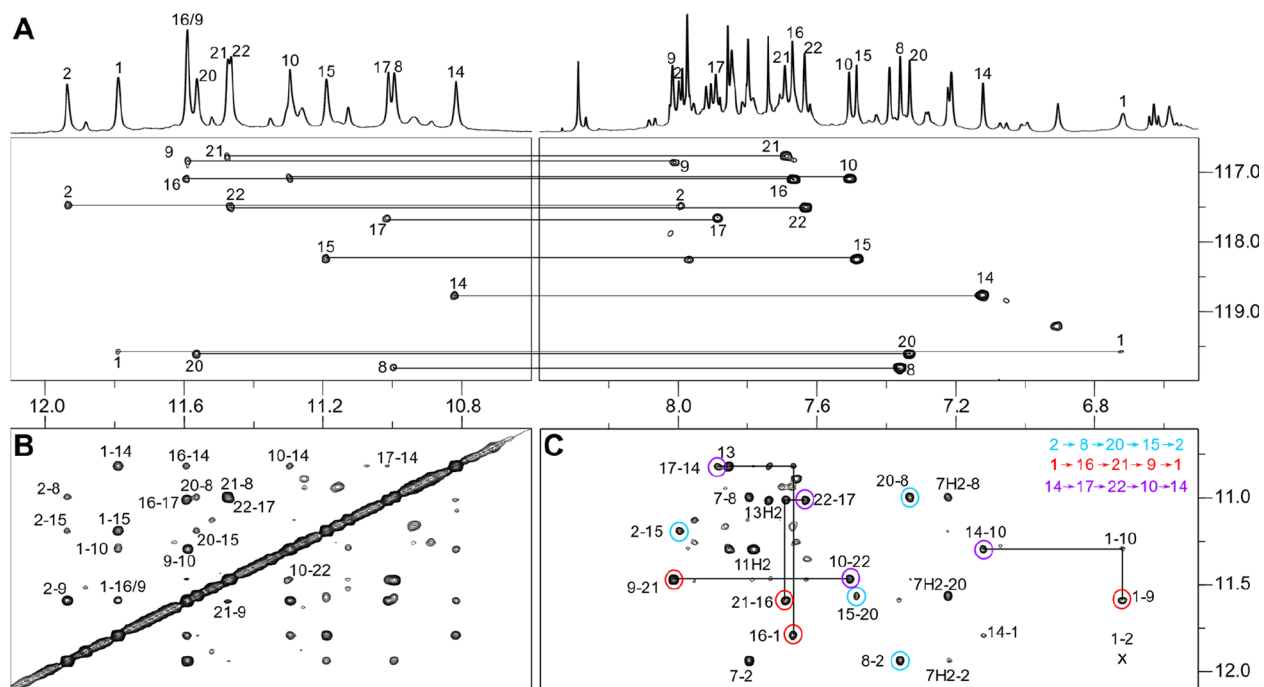


Figure 4. Spectra of KNA-ΔG13-T (0.9 mM). (A) ¹H-¹³C HMBC spectrum showing intrabase H8-H1(ω_2) correlations via long-range couplings to ¹³C5(ω_1) at natural abundance. (B) Imino-imino and (C) H8(ω_2)-H1(ω_1) spectral regions of a 2D NOESY spectrum (mixing time of 300 ms) at 25 °C. In panel C, cross-peaks defining hydrogen bond directionalities in the top, central, and bottom tetrads are circled in blue, red, and purple, respectively. X indicates a weak contact that can be observed only at lower contour levels.

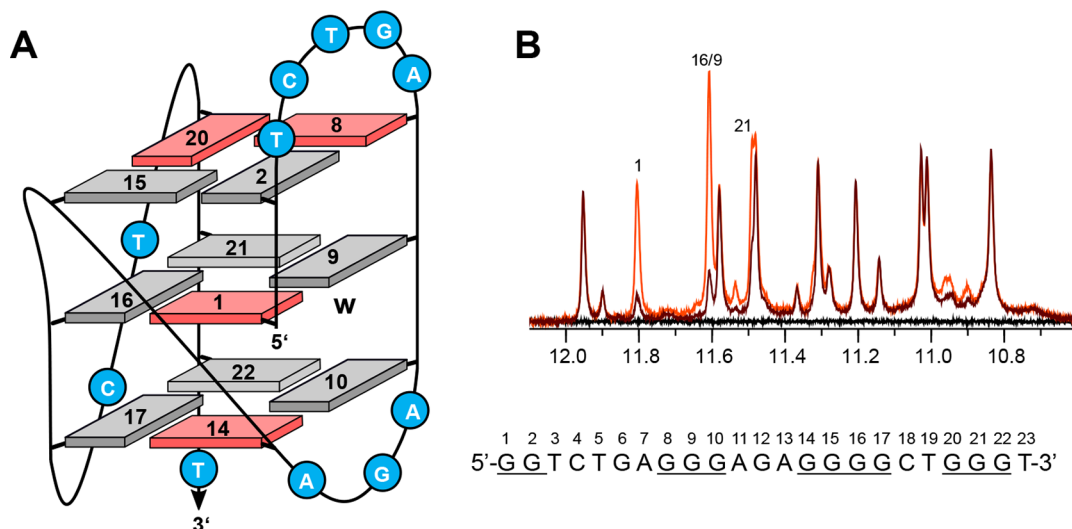


Figure 5. (A) Topology of the KNA-ΔG13-T quadruplex. *Syn* and *anti* residues of the G-core are colored red and gray, respectively; loop and overhang residues are indicated by blue circles, and a w denotes a wide groove. (B) Imino proton spectral region in D₂O (black) and after redissolving in water with waiting times of 10 min (brown) and 240 min (orange). Labeled peaks indicate slowly exchanging imino protons. The sequence with residue numbering is shown at the bottom.

contacts extending into neighboring loop regions, three *syn-anti* arrangements with their characteristic rectangular pattern of intranucleotide and two-way sequential H8-H1' but also H8-H3' cross-peaks are apparent for G1-G2, G8-G9, and G20-G21 steps (Figure 3). A putative G14 H8 resonance lacks sequential H8-H1' NOE contacts but features a strong intranucleotide H8-H1' connectivity typical of a *syn* conformer. A closer look reveals weak additional cross-peaks of the latter to G1 H8 and G1 H1', in agreement with an interrupted *syn*-G14 to *syn*-G1 step (Figure 3). Having assigned most H6/H8 and sugar resonances, we again confirmed via renewed

inspection of the ¹H-¹³C HSQC and 2D NOESY spectrum acquired with a short mixing time a *syn*-glycosidic conformation for G1, G8, G20, and G14 but also for the single G6 (Figure S8). Taken together, the results of the spectral analysis of non-exchangeable protons suggest formation of a three-layer quadruplex composed of two *syn-anti-anti* columns (G8-G9-G10 and G20-G21-G22), one all-*anti* column (G15-G16-G17), and one broken *syn-syn-anti* column (G14-G1-G2).

A subsequent assignment of G imino protons was based on their intranucleotide correlations with H8 protons through ¹H-¹³C scalar couplings with ¹³C5 at natural abundance. A

corresponding ^1H – ^{13}C HMBC experiment shows 12 correlated pairs of guanine H8 and imino protons allowing for the unambiguous assignment of G imino protons engaged in the three-layer quadruplex core (Figure 4A). With the identification of all G imino resonances, tetrad polarities can be traced on the basis of the observed intratetrad H8–H1 NOE contacts. Thus, Hoogsteen hydrogen bonds within the three G-quartets run along G14–G17–G22–G10–G14, G1–G16–G21–G9–G1, and G2–G8–G20–G15–G2. As a consequence, tetrads exhibit both homopolar and heteropolar stacking interactions, in line with the CD spectral signature typical of a (3+1) hybrid structure.

The derived G4 topology features a 5′-terminal G positioned in the central tetrad followed by two lateral loops, a zero-nucleotide V-shaped loop and a propeller loop (Figure 5A). Several imino–imino NOE contacts corroborate the *syn*–*anti* pattern and the resulting tetrad polarities of the quadruplex fold (Figure 4B). These include intra- and intertetrad NOEs between G2 and G8/G15/G9. Other imino–imino contacts connect G16 with G17, G1 with G14/G15, G14 with G10/G17, and G9 with G10. In line with the opposing tetrad polarities, there is no NOE contact between G1 and G2, G8 and G9, G15 and G16, or G20 and G21.

In addition to H1–H8 and H1–H1 contacts, several inter-residue H1–H1′, H1–H2′/H2″, and H1′–H1′ NOE contacts proved to be valuable for corroborating relative strand directionalities (Figure S10). Thus, observation of a conspicuous NOE contact between *syn*-G1 H1 and G14 H2′ protons confirms the position of these residues within a G14–G1–G2 tract but also points to their parallel sugar–phosphate orientation (Figure S10C). Finally, additional evidence for the quadruplex fold comes from solvent exchange experiments that reveal slowly exchanging imino protons of G1, G9, G16, and G21 due to their solvent-protected location within the central G-tetrad (Figure 5B).

After stereospecific assignments of H2′/H2″ proton resonances through the analysis of H6/H8–H2′/H2″ and H1′–H2′/H2″ NOE cross-peaks in 2D NOE spectra with short mixing times, the sugar pucker of individual residues was determined by DQF-COSY experiments in terms of a major south or north conformer. In contrast to south conformers, H1′(ω_2)–H2′(ω_1) unlike H1′(ω_2)–H2″(ω_1) cross-peaks are weak or even absent due to small H1′–H2′ vicinal couplings in a north conformation associated with partial cancellation of the antiphase components of the DQF-COSY cross-peaks. Thus, G14 and G15 were found to adopt a sugar pucker in the north domain of the pseudorotation cycle while most other residues adopt a south conformation (Figure S11). The sugar conformation of residues G1, G6, T19, and T23 is less well-defined and remained ambiguous.

Three-Dimensional Structure of the KNA- Δ G13-T Quadruplex. Structure calculations on KNA- Δ G13-T were performed with experimental restraints that included restraints for the glycosidic torsion angle and sugar pucker (Table 1). On the basis of previous reports on conventional V-shaped loop topologies with both V-loop flanking residues adopting a north-type sugar pucker,³³ the region of glycosidic torsion angle for the 3′-flanking residue G15 was extended beyond *high-anti* to provide for sufficient flexibility. There is a good convergence of the G-tetrad core, including residues T3, G6, and A7 of the first lateral loop in all calculated structures. Only the second lateral and propeller loop show more noticeable fluctuations (Figure 6A).

Table 1. NMR Restraints and Structural Statistics for the Structure Calculations of KNA- Δ G13-T

NMR Restraints	
NOE-based distance restraints	
intraresidual	96
inter-residual	208
exchangeable	85
other restraints	
hydrogen bonds	48
dihedral angles	42
planarity	3
Structural Statistics	
pairwise heavy atom root-mean-square deviation (Å)	
all residues	2.04 ± 0.4
G-tetrad core	0.62 ± 0.1
NOE violations (Å)	
maximum violation	0.165
mean NOE violation	0.004 ± 0.001
deviations from idealized geometry	
bonds (Å)	0.01 ± 0.0001
angles (deg)	2.3 ± 0.04

The quadruplex folds into a V-shaped loop architecture with both V-loop 5′- and 3′-flanking residues adopting a north conformation. Whereas the 5′-flanking G14 located in the lower tetrad features a *syn*-glycosidic torsion angle, the 3′-flanking G15 in the upper tetrad exhibits a conformation in the *high-anti* range in all calculated structures with an average glycosidic angle of 284°. Given different tetrad polarities of the two outer quartets, most of the backbone orientational change occurs within the third G-tract between V-loop 3′-flanking G15 and G16 residues. Such a V-loop conformation with G16 and nearby G14 as well as G15 oriented in opposite directions is also reflected in the observation of unusual long-range NOEs between G16 H8 and G14 H2′/H2″/H3′ sugar protons and of a strong G16 H8–G15 H3′ NOE contact (Figures 3 and 7A). According to a recent classification, the zero-nucleotide loop forms a conventional V-shaped loop.³³ Notably, the unusual downfield-shifted H3′ of the north-puckered *syn*-G14 residue experiences significant deshielding effects through its in-plane orientation with the nucleobase.^{34,35} Also, a short G16 H8–G15 O5′ distance with a putative hydrogen bond interaction is suggested by the three-dimensional structures and matches with corresponding observations in several conventional V-shaped loop quadruplexes (Figure 7A).³³

The first TCTGA lateral loop stacks coaxially onto the G-core with several NOE cross-peaks connecting T3/A7 residues with imino protons of the adjacent G-tetrad (Figure 4C and Figure S10A,C). Given a loop bridging a wide groove, an interfacial Watson–Crick AT base pair should easily be accommodated.^{36–38} However, there is an unusual Hoogsteen-type base pairing alignment between T3 and A7 in all calculated structures (Figure 7B).

The second A11–G12–A13 intervening sequence forms another lateral loop with various NOE contacts of G12 and A13 protons to protons of the neighboring outer tetrad. Interestingly, bases of the latter loop are arranged in two distinct orientations within the calculated structural ensemble. G12 in the lowest-energy structure is found to be positioned below the outer tetrad to form a hydrogen bond between its imino proton and the carbonyl O2 of the 3′-terminal T23 overhang. Whereas A11 and T23 are oriented parallel to the

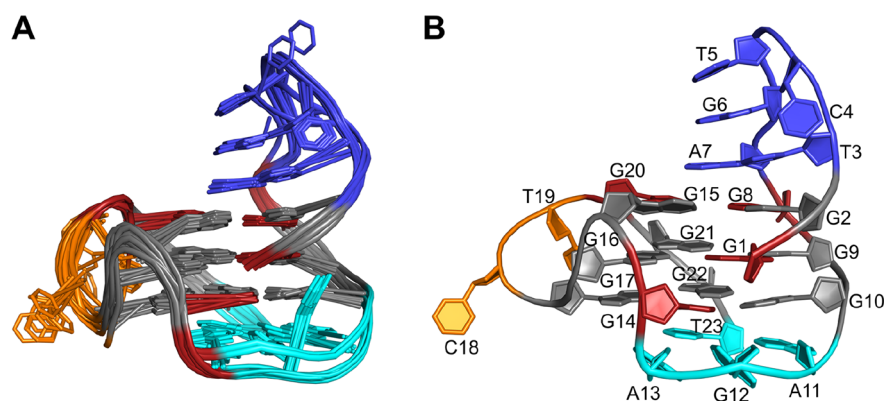


Figure 6. (A) Superposition of the 10 lowest-energy structures and (B) representative structure with labeled residues for the *KNA-ΔG13-T* quadruplex. *syn* and *anti* residues of the tetrads are colored red and gray, respectively. The propeller loop is colored orange, whereas the first and second lateral loops including 3'-T are colored blue and cyan, respectively.

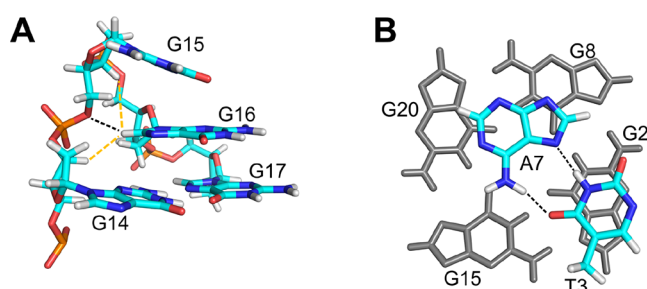


Figure 7. (A) Side view on the G-tract with a conventional V-shaped loop. A putative G16 H8–G15 O5' hydrogen bond interaction and short G16 H8–G14/G15 H3' distances are indicated by black and orange dotted lines, respectively. (B) Top view of the interface between the outer tetrad (gray) and the first lateral loop with a Hoogsteen hydrogen-bonded A·T base pair arrangement.

adjacent G-tetrad plane, G12 and A13 bases are tilted by $\sim 45^\circ$. On the contrary, in eight of 10 calculated structures G12 slides into the wide groove and A11, A13, and T23 bases arrange in a triad stacked below the outer tetrad with the formation of putative A11 amino–T23 O2 and A13 amino–T23 O4 hydrogen bonds (Figure S12A–C). It is noteworthy that both arrangements are only partially consistent with experimental 2D NOESY spectra that lack anticipated cross-peaks for protons in close contact within this AGA-loop conformation. Such a discrepancy may result from the absence of critical experimental restraints but may also reflect dynamic conformational transitions for this lateral loop preceding the V-shaped loop.

The two-nucleotide propeller loop is rather poorly defined by NOE contacts. However, distance restraints through base–sugar NOE contacts between T19 and G20/G21 of the subsequent G-tract suggest another putative hydrogen bond between closely spaced T19 O2 and the G16 amino proton (Figure S12D).

Lacking direct experimental evidence for hydrogen bond interactions involving loop and overhang residues at ambient temperatures, we recorded additional NMR spectra at lower temperatures. In fact, a 2D NOESY spectrum at 278 K reveals several new cross-peaks. Notably, a weak downfield-shifted imino resonance at ~ 13.9 ppm assigned to the T3 residue exhibits a cross-peak to A7 H8 in line with the proposed Hoogsteen AT base pair arrangement (Figure S13D). Also, several amino signals of G-core residues give rise to additional

NOE contacts (Figure S13B,C). Interestingly, cross-peaks involving the G16 amino proton are found to be sharper when compared to other amino protons within the G-tetrads. This may hint at significant restrictions of the corresponding amino group rotation, corroborating an additional hydrogen bond formation between the G16 amino proton and T19 O2 of the propeller loop. In contrast, there is no experimental evidence for hydrogen bonds involving residues of the second lateral loop and the 3'-thymine overhang, possibly due to a dynamic hydrogen bonding network as might be indicated by the two different loop conformations in the structural ensemble.

Mutations and Their Impact on V-Loop Formation.

NMR spectral analysis of the modified *KNA* sequences indicates that their folding into a V-loop topology can be driven by mutations of the V-loop-preceding lateral loop and/or by the addition of thymine at the 3'-end. Apparently, the thymine overhang affects the second lateral loop and more significant H8 chemical shift changes are observed for residues 1, 10, 12, and 22 upon addition of 3'-T to *KNA-ΔG13* (Figure S2). Interactions of 3'-T with bases of the AGA lateral loop are indicated by the structural ensemble of *KNA-ΔG13-T*, but two distinct patterns of base alignments point to weak and/or dynamic interactions. Notably, a thymidine extension restricts formation of another unknown minor species but does not provide for a higher thermal stability as expected for additional stabilizing interactions (Figure 2 and Figure S7). This suggests only small but critical contributions in favor of the V-loop structure or unfavorable interactions for folding into an alternative species.

A comparison with two closely related V-loop structures containing two lateral loops, one V-loop, and one propeller loop reveals some common features.^{14,34,39} Thus, the importance of a 3'-overhang is emphasized, and it is the 3'-T that participates in hydrogen bond formation with bases of the V-loop-preceding lateral loop. An analysis of modified sequences identified 3'-T as being the most effective base for V-loop formation, followed by C and A.¹⁴ Interestingly, 3'-G hardly supports V-loop formation in line with an NMR spectral analysis of *KNA-ΔG13-G*. Contrary to expectations, however, *KNA-ΔG13-GGT* with a 3'-GGT overhang shows formation of a minor V-loop G-quadruplex among various coexisting species, and this can be attributed to its G5-tract and a slipped register isomer with a single T-overhang and the two 5'-Gs participating in a preceding elongated propeller loop.

In addition, all three V-loop structures share the presence of a first lateral loop and of an N_xAG_4 tract that seems to be critical for V-loop formation. On the contrary, a canonical (3+1) hybrid quadruplex⁴⁰ was shown to interconvert into a corresponding V-shaped loop topology by incorporating sugar-modified guanosine analogues with particular sugar pucker propensities at the V-loop 5'- and 3'-flanking positions. In the modified sequences, G3 participates in a lateral loop formed from the native propeller loop whereas the second ACACA loop changes from a diagonal to a second lateral loop preceding the V-shaped loop.^{33,35,41} Clearly, several contributions and mutual interactions of single structural domains add to the favored formation of such V-loop quadruplexes and need to be considered when looking for sequences with a V-loop-forming propensity.

Impact of the 5'-Overhang on the V-Loop Structure.

There is the question of whether a G-quadruplex structure featuring an interrupted G-column with a terminal guanine locked into a central tetrad position or complementing an outer G-tetrad in a snap-back loop arrangement constitutes a physiologically relevant topology if positioned between flanking sequences within the genome. In the V-loop structure presented here, the terminal G occupies a central tetrad position and any additional flanking residue may directly interfere with the broken G-column.

The assigned quadruplex-forming sequence was therefore extended by the addition of a 5'-adenosine for mimicking the wild-type sequence context. Like *KNA-ΔG13-T*, CD profiles of the 5'-A variant *A-KNA-ΔG13-T* exhibit a signature typical of both homopolar and heteropolar G-tetrad stackings (Figure S14A). On the contrary, imino resonances in 1D NMR spectra were considerably broadened, yet a set of ~12 imino signals can still be observed (Figure 2). On the basis of a more detailed comparison of 2D NOESY spectral regions, a population inversion becomes clearly apparent between a major and minor species when compared to the sequence lacking a 5'-overhang (Figure S14C). Thus, a less populated V-loop topology for *A-KNA-ΔG13-T* is indicated by its characteristic 2D NOE cross-peak pattern. Also, the major G4 species for *A-KNA-ΔG13-T* matches a minor fold with only weak signals observed for *KNA-ΔG13-T*. Unfortunately, a more detailed structural assignment of the latter was hampered by weak or broadened signals. A destabilizing effect of a 5'-overhang on thermal quadruplex stability is revealed by additional UV melting experiments with *A-KNA-ΔG13-T*. The latter yielded a melting temperature T_m of 32.8 °C in a low-salt buffer, being 14 °C below the T_m of *KNA-ΔG13-T* under the same conditions (Figures S7 and S14B).

It has been proposed that there is enough space to accommodate a flanking DNA sequence without restricting the V-loop fold based on a simple model of a higher-order RNA structure with V-loop topology and a quadruplex–duplex junction.¹⁴ Likewise, in studies of HIV-1 *LTR* sequences, a truncated *LTRIII* domain was shown to adopt a V-shaped loop topology with its first G participating in the central tetrad. 1D NMR spectra of a longer *LTRIII+IV* sequence with the addition of a 5'-adenosine suggested that the 5'-flanking A residue did not abolish formation of the particular *LTRIII* V-loop structure within this extended sequence context.⁴² Another report involves the proximal promoter sequence of *VEGF* receptor 2 (*VEGFR-2*) that folds into a three-layer V-loop G-quadruplex with its 5'-terminal guanine of the first GGG-tract positioned within the central tetrad and the third G being

part of the following loop region. Extensive mutational studies also included a G1A substitution that would allow for the formation of a V-shaped loop architecture with a 5'-A overhang, a second and third G participating in the central and outer G-tetrad, and a shortened first loop. However, in this case, the results from 1D NMR and CD spectra seem to exclude folding to a corresponding V-shaped loop topology with a 5'-flanking residue.⁴³

In contrast to its positioning within a central tetrad, a terminal G in most snap-back loop structures occupies an empty position of an outer G-tetrad and may possibly better accommodate additional flanking residues. Thus, a snap-back loop quadruplex formed by a *c-myc* promoter sequence was reported to be conserved upon addition of a 3'-T residue to the fold-back guanine with the thymine pointing out of the G-tetrad plane.⁴⁴ However, although a corresponding fold is still retained, recent studies of its thermodynamic stability suggested a destabilizing effect of the 3'-T overhang on the latter topology. As a result, minor species increasingly coexist with addition of a 3'-T and 3'-TT dinucleotide.⁴⁵ On the contrary, a well-defined quadruplex structure has been reported for a snap-back guanine embedded within a unique peripheral motif. Whereas the single guanine fills the outer tetrad, the 3'-flanking overhang sequence folds back to form stacked layers of base pairing alignments for additional stability.⁴⁶

Taken together, addition of a 5'-flanking base within the broken G-tract of the V-loop topology still allows for a corresponding quadruplex fold. However, V-loop formation suffers from destabilizing effects that will likely change folding pathways and equilibria among different folds of similar energy. Although the V-loop quadruplex with its interrupted G-tract may be only poorly populated in a polymorphic mixture derived from the first repeat of the *KCNN4* minisatellite in a longer sequence context, it may well represent a promising drug target for modulating *KCNN4* expression. Thus, a V-loop specific ligand may shift dynamic equilibria upon binding to effectively hit this sequence without more extensive off-target effects.

CONCLUSIONS

The *KCNN4* gene was found to contain a G-rich minisatellite with putative G-quadruplex forming abilities. We have structurally characterized in detail a G4 species with a V-shaped loop proposed to coexist in a polymorphic mixture formed by the first repeat of the minisatellite. Here, a naturally occurring G → A mutation leaves a run of two guanines within the first G-tract that, together with interactions between a 3'-overhang and a lateral loop, drives the formation of such a G-quadruplex. Clearly, sequence requirements as revealed by mutational studies and a detailed analysis of intramolecular interactions increasingly enhance our knowledge of these noncanonical G4 structures, supporting their identification in genomic sequences and also assisting in the rational design of artificial V-loop-forming sequences for technological applications.

Pharmacologically, the V-loop quadruplex formed in the *KCNN4* G-rich minisatellite features specific structural motifs that may promote selective ligand binding and thus a specific targeting through the design and development of new and more potent drugs. It has been demonstrated that quadruplex formation in the proximity of the transcribed region can be related to alterations in gene expression levels. Putative quadruplexes in promoter regions were found to be generally

correlated with a reduced level of transcription.^{47,48} Likewise, genomewide analyses have revealed that quadruplex formation downstream of the transcription start site and even downstream of the transcription termination site may play a regulatory role in gene transcription with effects also depending on the G4-forming sequences located on the sense or template strand.^{49,50} Given the importance of the *KCNN4* gene for the regulation of cell ionic strength and its correlation with cancer development, future research may be directed to quadruplex formation within this minisatellite and its impact on and control of the modulation of *KCNN4* expression within cells.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.1c00043>.

CD spectra, PAGE, UV melting curves, 2D NOEs, ¹H–¹³C HSQC, and DQF-COSY spectral regions of KNA variants as well as molecular models and ¹H and ¹³C chemical shifts of KNA-ΔG13-T (PDF)

Accession Codes

Coordinates for KNA-ΔG13-T have been deposited as Protein Data Bank entry 7ATZ and BMRB entry 34571.

■ AUTHOR INFORMATION

Corresponding Author

Klaus Weisz – Institute of Biochemistry, Universität Greifswald, D-17487 Greifswald, Germany; orcid.org/0000-0002-2736-6606; Phone: (+49) 3834 420-4426; Email: weisz@uni-greifswald.de; Fax: (+49) 3834 420-4427

Author

Yoanes Maria Vianney – Institute of Biochemistry, Universität Greifswald, D-17487 Greifswald, Germany

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.biochem.1c00043>

Funding

This study was funded by the Deutsche Forschungsgemeinschaft (WE 1933/15-1).

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

QGRS, quadruplex-forming G-rich sequence; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; NUS, non-uniform sampling; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single-quantum correlation; HMBC, heteronuclear multiple-bond correlation; DQF-COSY, double-quantum-filtered correlation spectroscopy; TIP3P, three-point water model.

■ REFERENCES

- (1) Maltby, C. J.; Schofield, J. P. R.; Houghton, S. D.; O'Kelly, I.; Vargas-Caballero, M.; Deinhardt, K.; and Coldwell, M. J. (2020) A 5' UTR GGN Repeat Controls Localisation and Translation of a Potassium Leak Channel mRNA through G-Quadruplex Formation. *Nucleic Acids Res.* 48, 9822–9839.
- (2) Bulk, E.; Ay, A. S.; Hammadi, M.; Oquadid-Ahidouch, H.; Schelhaas, S.; Hascher, A.; Rohde, C.; Thoennissen, N. H.; Wiewrodt,

R.; Schmidt, E.; Marra, A.; Hillejan, L.; Jacobs, A. H.; Klein, H. U.; Dugas, M.; Berdel, W. E.; Müller-Tidow, C.; and Schwab, A. (2015) Epigenetic Dysregulation of KCa3.1 Channels Induces Poor Prognosis in Lung Cancer. *Int. J. Cancer* 137, 1306–1317.

(3) Ousingsawat, J.; Spitzner, M.; Puntheeranurak, S.; Terracciano, L.; Tornillo, L.; Bubendorf, L.; Kunzelmann, K.; and Schreiber, R. (2007) Expression of Voltage-Gated Potassium Channels in Human and Mouse Colonic Carcinoma. *Clin. Cancer Res.* 13, 824–831.

(4) Pardo, L. A.; Del Camino, D.; Sánchez, A.; Alves, F.; Brüggemann, A.; Beckh, S.; and Stühmer, W. (1999) Oncogenic Potential of EAG K⁺ Channels. *EMBO J.* 18, 5540–5547.

(5) Biasiotta, A.; D'Arcangelo, D.; Passarelli, F.; Nicodemi, E. M.; and Facchiano, A. (2016) Ion Channels Expression and Function Are Strongly Modified in Solid Tumors and Vascular Malformations. *J. Transl. Med.* 14, 285.

(6) Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; and Balasubramanian, S. (2020) The Regulation and Functions of DNA and RNA G-Quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474.

(7) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; and Balasubramanian, S. (2015) High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* 33, 877–881.

(8) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; and Yang, D. (2005) Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* 44, 2048–2058.

(9) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; and Patel, D. J. (2007) Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* 129, 4386–4392.

(10) Sengar, A.; Vandana, J. J.; Chambers, V. S.; Di Antonio, M.; Winnerdy, F. R.; Balasubramanian, S.; and Phan, A. T. (2019) Structure of a (3 + 1) Hybrid G-Quadruplex in the PARP1 Promoter. *Nucleic Acids Res.* 47, 1564–1572.

(11) 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.

(12) Kettani, A.; Kumar, R. A.; and Patel, D. J. (1995) Solution Structure of a DNA Quadruplex Containing the Fragile X Syndrome Triplet Repeat. *J. Mol. Biol.* 254, 638–656.

(13) Derecka, K.; Balkwill, G. D.; Garner, T. P.; Hodgman, C.; Flint, A. P. F.; and Searle, M. S. (2010) Occurrence of a Quadruplex Motif in a Unique Insert within Exon C of the Bovine Estrogen Receptor α Gene (ESR1). *Biochemistry* 49, 7625–7633.

(14) Kuryavyi, V.; and Patel, D. J. (2010) Solution Structure of a Unique G-Quadruplex Scaffold Adopted by a Guanosine-Rich Human Intronic Sequence. *Structure* 18, 73–82.

(15) Trajkovski, M.; Webba Da Silva, M.; and Plavec, J. (2012) Unique Structural Features of Interconverting Monomeric and Dimeric G-Quadruplexes Adopted by a Sequence from the Intron of the N-Myc Gene. *J. Am. Chem. Soc.* 134, 4132–4141.

(16) Brčić, J.; and Plavec, J. (2015) Solution Structure of a DNA Quadruplex Containing ALS and FTD Related GGGGCC Repeat Stabilized by 8-Bromodeoxyguanosine Substitution. *Nucleic Acids Res.* 43, 8590–8600.

(17) Ribeiro, M. M.; Teixeira, G. S.; Martins, L.; Marques, M. R.; de Souza, A. P.; and Line, S. R. P. (2015) G-Quadruplex Formation Enhances Splicing Efficiency of PAX9 Intron 1. *Hum. Genet.* 134, 37–44.

(18) Huang, H.; Zhang, J.; Harvey, S. E.; Hu, X.; and Cheng, C. (2017) RNA G-Quadruplex Secondary Structure Promotes Alternative Splicing via the RNA-Binding Protein HnRNPF. *Genes Dev.* 31, 2296–2309.

(19) Zizza, P.; Cingolani, C.; Artuso, S.; Salvati, E.; Rizzo, A.; D'Angelo, C.; Porru, M.; Pagano, B.; Amato, J.; Randazzo, A.; Novellino, E.; Stoppacciaro, A.; Gilson, E.; Stassi, G.; Leonetti, C.; and Birocchio, A. (2016) Intragenic G-Quadruplex Structure Formed in the Human CD133 and Its Biological and Translational Relevance. *Nucleic Acids Res.* 44, 1579–1590.

- (20) Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J. L., and Phan, A. T. (2012) Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 minisatellite. *J. Am. Chem. Soc.* *134*, 5807–5816.
- (21) Adrian, M., Ang, D. J., Lech, C. J., Heddi, B., Nicolas, A., and Phan, A. T. (2014) Structure and Conformational Dynamics of a Stacked Dimeric G-Quadruplex Formed by the Human CEB1 minisatellite. *J. Am. Chem. Soc.* *136*, 6297–6305.
- (22) Piazza, A., Cui, X., Adrian, M., Samazan, F., Heddi, B., Phan, A. T., and Nicolas, A. G. (2017) Non-Canonical G-Quadruplexes Cause the HCEB1 minisatellite Instability in *Saccharomyces Cerevisiae*. *eLife* *6*, e26884.
- (23) Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E. S., Bowers, M. T., De Pauw, E., and Gabelica, V. (2008) G-Quadruplex DNA Assemblies: Loop Length, Cation Identity, and Multimer Formation. *J. Am. Chem. Soc.* *130*, 10208–10216.
- (24) Kim, B. G., Evans, H. M., Dubins, D. N., and Chalikian, T. V. (2015) Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter. *Biochemistry* *54*, 3420–3430.
- (25) Tateishi-Karimata, H., Kawachi, K., and Sugimoto, N. (2018) Destabilization of DNA G-Quadruplexes by Chemical Environment Changes during Tumor Progression Facilitates Transcription. *J. Am. Chem. Soc.* *140*, 642–651.
- (26) Jiang, S., Zhu, L., Yang, J., Hu, L., Gu, J., Xing, X., Sun, Y., and Zhang, Z. (2017) Integrated Expression Profiling of Potassium Channels Identifies KCNN4 as a Prognostic Biomarker of Pancreatic Cancer. *Biochem. Biophys. Res. Commun.* *494*, 113–119.
- (27) Lai, W., Liu, L., Zeng, Y., Wu, H., Xu, H., Chen, S., and Chu, Z. (2013) KCNN4 Channels Participate in the EMT Induced by PRL-3 in Colorectal Cancer. *Med. Oncol.* *30*, 566.
- (28) D'Alessandro, G., Catalano, M., Sciacaluga, M., Chece, G., Cipriani, R., Rosito, M., Grimaldi, A., Lauro, C., Cantore, G., Santoro, A., Fioretti, B., Franciolini, F., Wulff, H., and Limatola, C. (2013) KCa3.1 Channels Are Involved in the Infiltrative Behavior of Glioblastoma in Vivo. *Cell Death Dis.* *4*, e773.
- (29) Kikin, O., D'Antonio, L., and Bagga, P. S. (2006) QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* *34*, W676–W682.
- (30) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct., Funct., Genet.* *59*, 687–696.
- (31) Schwieters, C. D., Kuszewski, J. J., and Marius Clore, G. (2006) Using Xplor-NIH for NMR Molecular Structure Determination. *Prog. Nucl. Magn. Reson. Spectrosc.* *48*, 47–62.
- (32) Zgarbová, M., Šponer, J., Otyepka, M., Cheatham, T. E., Galindo-Murillo, R., and Jurečka, P. (2015) Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* *11*, 5723–5736.
- (33) Haase, L., and Weisz, K. (2020) Locked Nucleic Acid Building Blocks as Versatile Tools for Advanced G-Quadruplex Design. *Nucleic Acids Res.* *48*, 10555–10566.
- (34) Marušič, M., and Plavec, J. (2019) Towards Understanding of Polymorphism of the G-Rich Region of Human Papillomavirus Type 52. *Molecules* *24*, 1294.
- (35) Haase, L., Dickerhoff, J., and Weisz, K. (2020) Sugar Puckering Drives G-Quadruplex Refolding: Implications for V-Shaped Loops. *Chem. - Eur. J.* *26*, 524–533.
- (36) Lim, K. W., and Phan, A. T. (2013) Structural Basis of DNA Quadruplex-Duplex Junction Formation. *Angew. Chem., Int. Ed.* *52*, 8566–8569.
- (37) Karg, B., Mohr, S., and Weisz, K. (2019) Duplex-Guided Refolding into Novel G-Quadruplex (3 + 1) Hybrid Conformations. *Angew. Chem., Int. Ed.* *58*, 11068–11071.
- (38) Tan, D. J. Y., Winnerdy, F. R., Lim, K. W., and Phan, A. T. (2020) Coexistence of Two Quadruplex–Duplex Hybrids in the PIM1 Gene. *Nucleic Acids Res.* *48*, 11162–11171.
- (39) Marušič, M., and Plavec, J. (2015) The Effect of DNA Sequence Directionality on G-Quadruplex Folding. *Angew. Chem., Int. Ed.* *54*, 11716–11719.
- (40) Marušič, M., Šket, P., Bauer, L., Viglasky, V., and Plavec, J. (2012) Solution-State Structure of an Intramolecular G-Quadruplex with Propeller, Diagonal and Edgewise Loops. *Nucleic Acids Res.* *40*, 6946–6956.
- (41) Haase, L., and Weisz, K. (2020) Switching the Type of V-Loop in Sugar-Modified G-Quadruplexes through Altered Fluorine Interactions. *Chem. Commun.* *56*, 4539–4542.
- (42) Butovskaya, E., Heddi, B., Bakalar, B., Richter, S. N., and Phan, A. T. (2018) Major G-Quadruplex Form of HIV-1 LTR Reveals a (3 + 1) Folding Topology Containing a Stem-Loop. *J. Am. Chem. Soc.* *140*, 13654–13662.
- (43) Liu, Y., Lan, W., Wang, C., and Cao, C. (2018) A Putative G-Quadruplex Structure in the Proximal Promoter of VEGFR-2 Has Implications for Drug Design to Inhibit Tumor Angiogenesis. *J. Biol. Chem.* *293*, 8947–8955.
- (44) Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human Myc Promoter. *Nat. Chem. Biol.* *1*, 167–173.
- (45) Jana, J., and Weisz, K. (2020) A Thermodynamic Perspective on Potential G-Quadruplex Structures as Silencer Elements in the MYC Promoter. *Chem. - Eur. J.* *26*, 17242–17251.
- (46) Truong, T. H. A., Winnerdy, F. R., and Phan, A. T. (2019) An Unprecedented Knot-like G-Quadruplex Peripheral Motif. *Angew. Chem., Int. Ed.* *58*, 13834–13839.
- (47) Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 11593–11598.
- (48) Cogoi, S., and Xodo, L. E. (2006) G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Res.* *34*, 2536–2549.
- (49) Du, Z., Zhao, Y., and Li, N. (2008) Genome-Wide Analysis Reveals Regulatory Role of G4 DNA in Gene Transcription. *Genome Res.* *18*, 233–241.
- (50) Smestad, J. A., and Maher, L. J. (2015) Relationships between Putative G-Quadruplex-Forming Sequences, RecQ Helicases, and Transcription. *BMC Med. Genet.* *16*, 91.

Supporting Information

First Tandem Repeat of a Potassium Channel *KCNN4* Minisatellite Folds into a V-Loop G-Quadruplex Structure

Yoanes Maria Vianney and Klaus Weisz*

Institute of Biochemistry, Universität Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald,
Germany

*Corresponding author: weisz@uni-greifswald.de

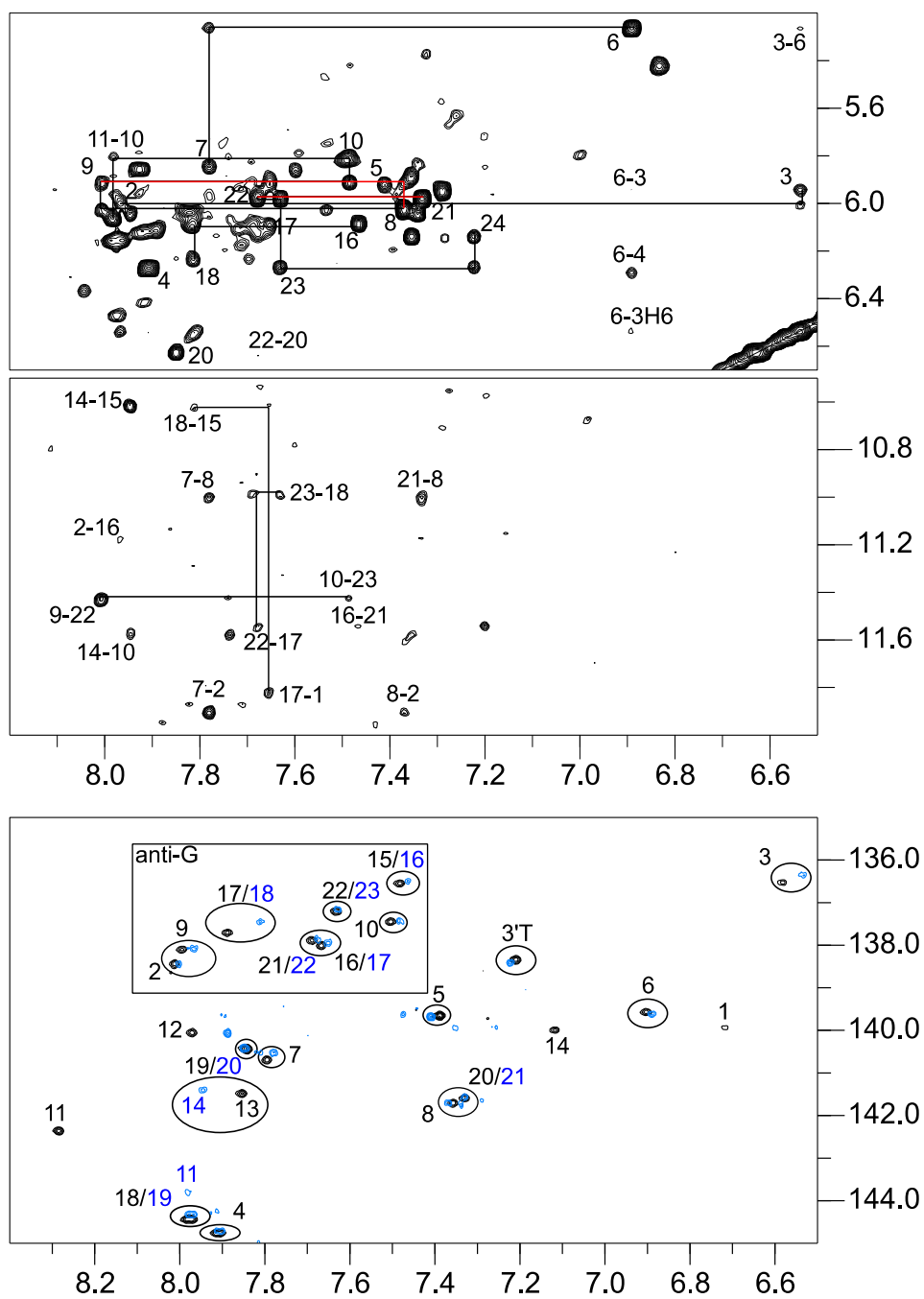


Figure S1. Partial assignment of *KNA-T*; 2D NOESY spectral region showing H6/8(ω_2)-H1'(ω_1) (top) and H6/8(ω_2)-H1(ω_1) cross-peaks (middle). Superimposed ^1H - ^{13}C HSQC spectra with H6/8-C6/8 correlations of *KNA-ΔGI3-T* (black, 0.9 mM) and *KNA-T* (blue) (bottom); overlapping cross-peaks and cross-peak patterns suggest identical G4 topologies.

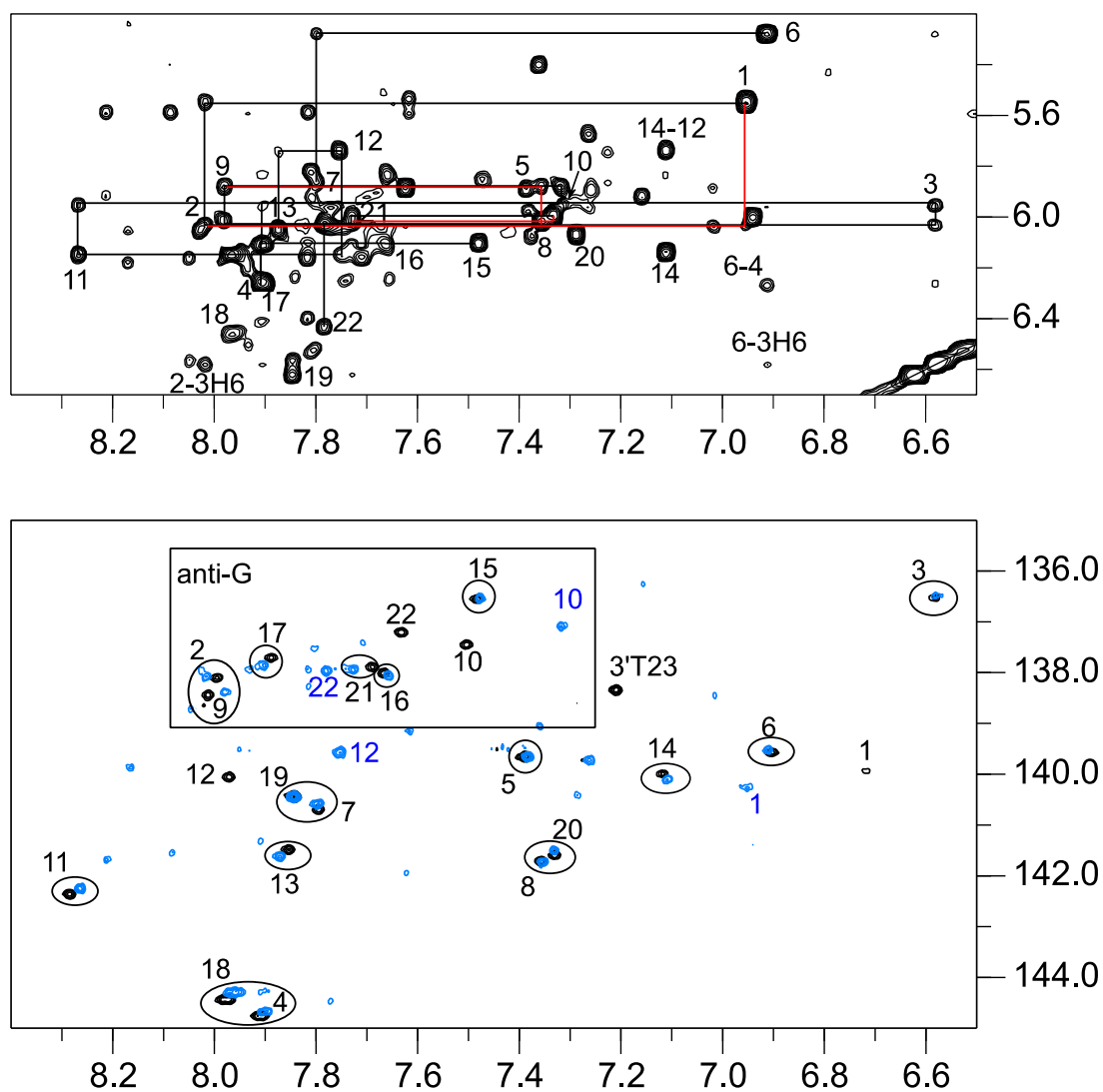


Figure S2. Partial assignment of *KNA-ΔG13*; 2D NOESY spectral region showing H6/8(ω_2)-H1'(ω_1) cross-peaks (top). Superimposed ^1H - ^{13}C HSQC spectra with H6/8-C6/8 correlations of *KNA-ΔG13-T* (black, 0.9 mM) and *KNA-ΔG13* (blue) (bottom). Overlapping cross-peaks and cross-peak patterns suggest identical G4 topologies.

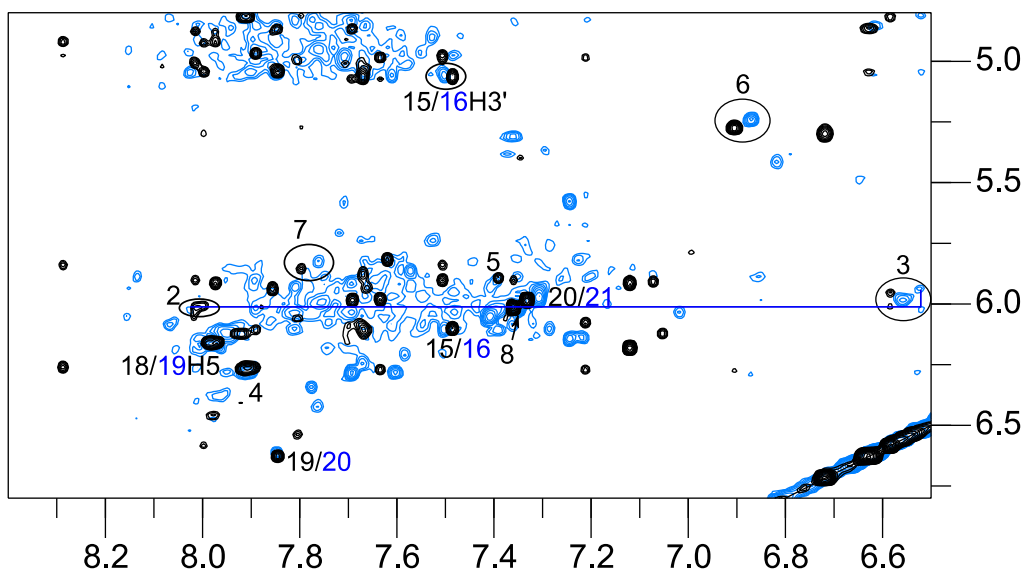


Figure S3. Partial assignment of *KNA-GGT*; superimposed 2D NOESY spectral region showing H6/8(ω_2)-H1'(ω_1) cross-peaks of *KNA- Δ G13-T* (black) and *KNA- Δ G13* (blue); several typical cross-peaks overlap, suggesting identical G4 topologies with assignments based on the *KNA- Δ G13-T* quadruplex.

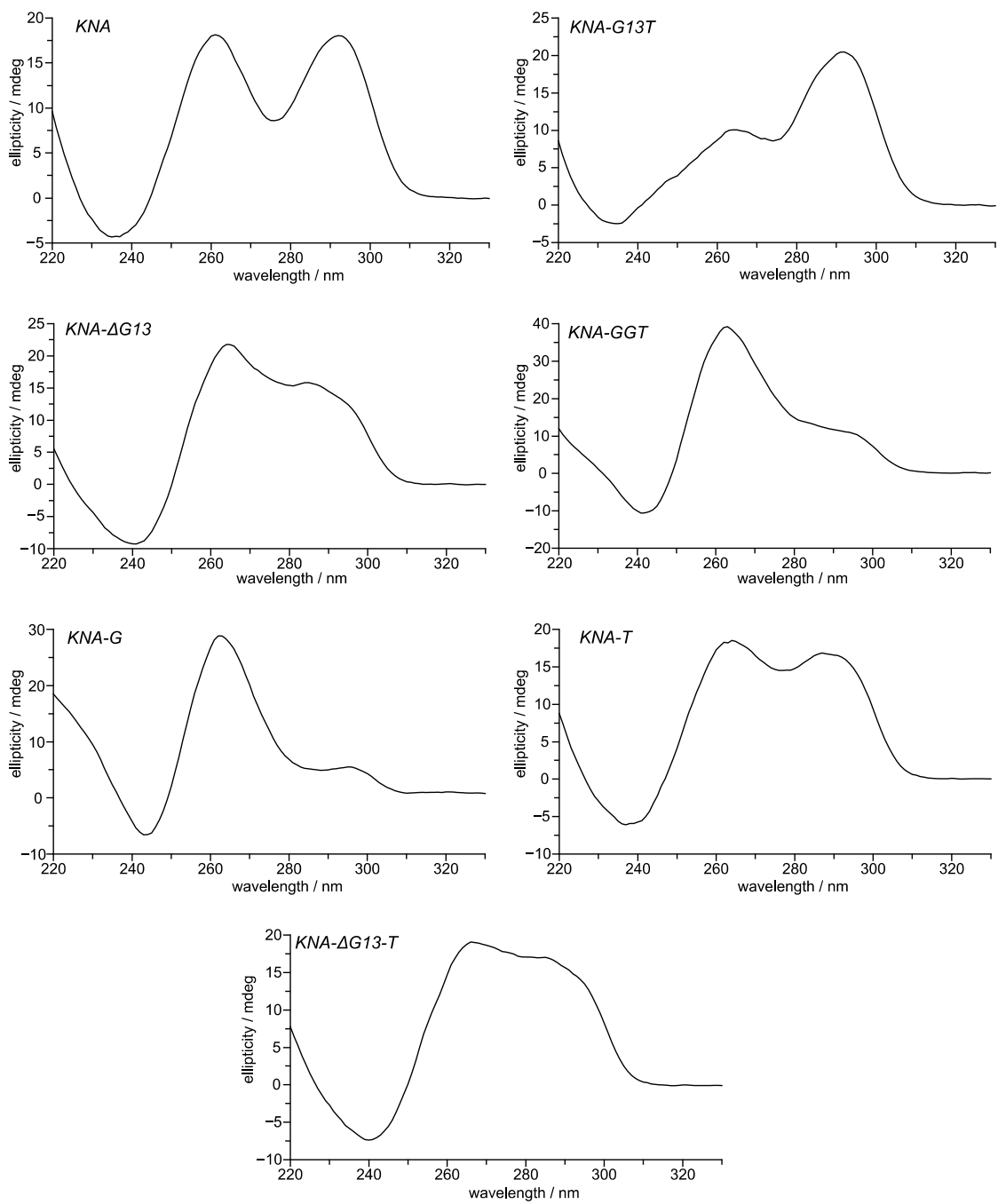


Figure S4. CD spectra of *KNA* variants at 20 °C in 10 mM potassium phosphate buffer, pH 7.

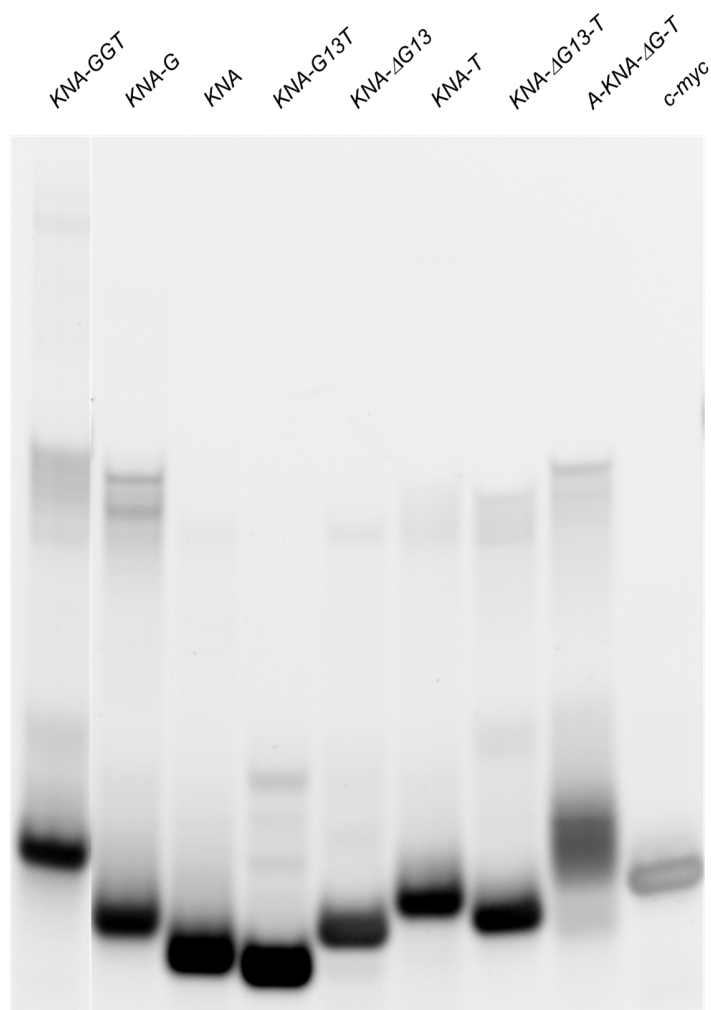


Figure S5. Non-denaturing gel electrophoresis of *KNA* variants and of a parallel 22mer *c-myc* quadruplex as additional reference. Slowly migrating weak bands are indicative of multimeric structures. Experimental procedure: 50 μ M of annealed oligonucleotide samples in a 10 mM potassium phosphate buffer, pH 7.0, were mixed with glycerol-buffer (4:6) in a 1:1 v/v ratio. Samples (250 pmol per lane) were loaded on a 15% polyacrylamide gel (acrylamide:bis-acrylamide 19:1). Separation was performed in TBE buffer, pH 8.3, supplemented with 10 mM KCl. Gels were stained with 5 μ M thiazole orange.

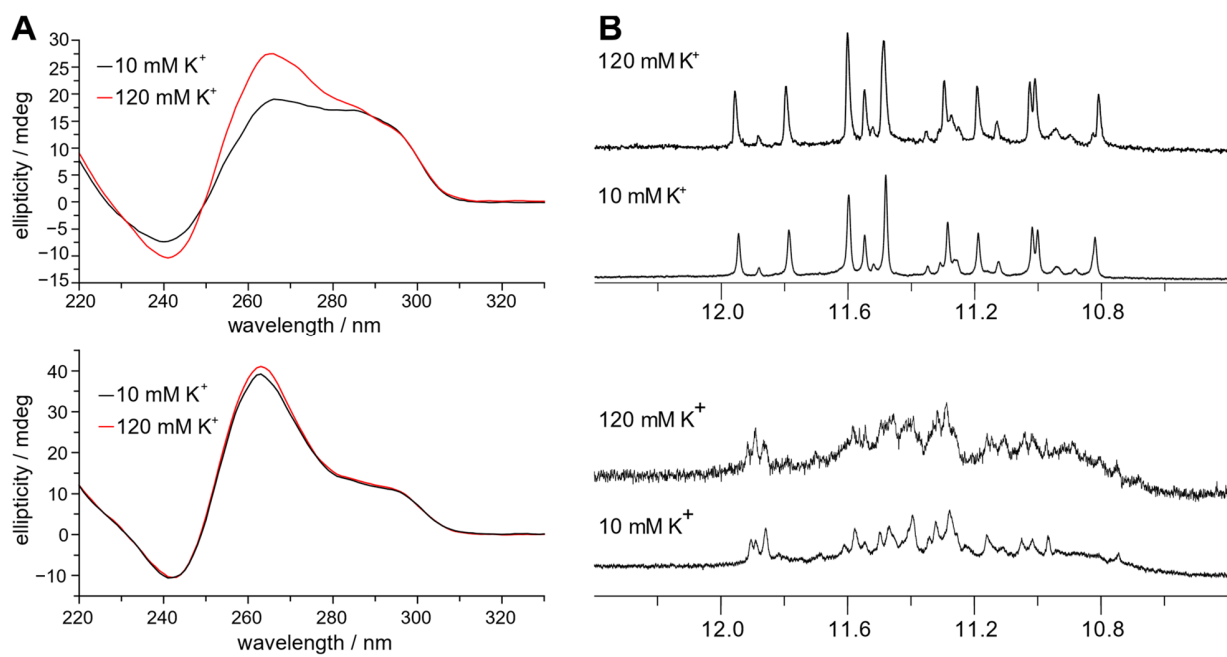


Figure S6. (A) CD spectra and (B) imino proton NMR spectral regions of *KNA-AG13-T* (top) and the *KNA-GGT* wild-type sequence (bottom) in buffer solutions with 10 mM and 120 mM K^+ .

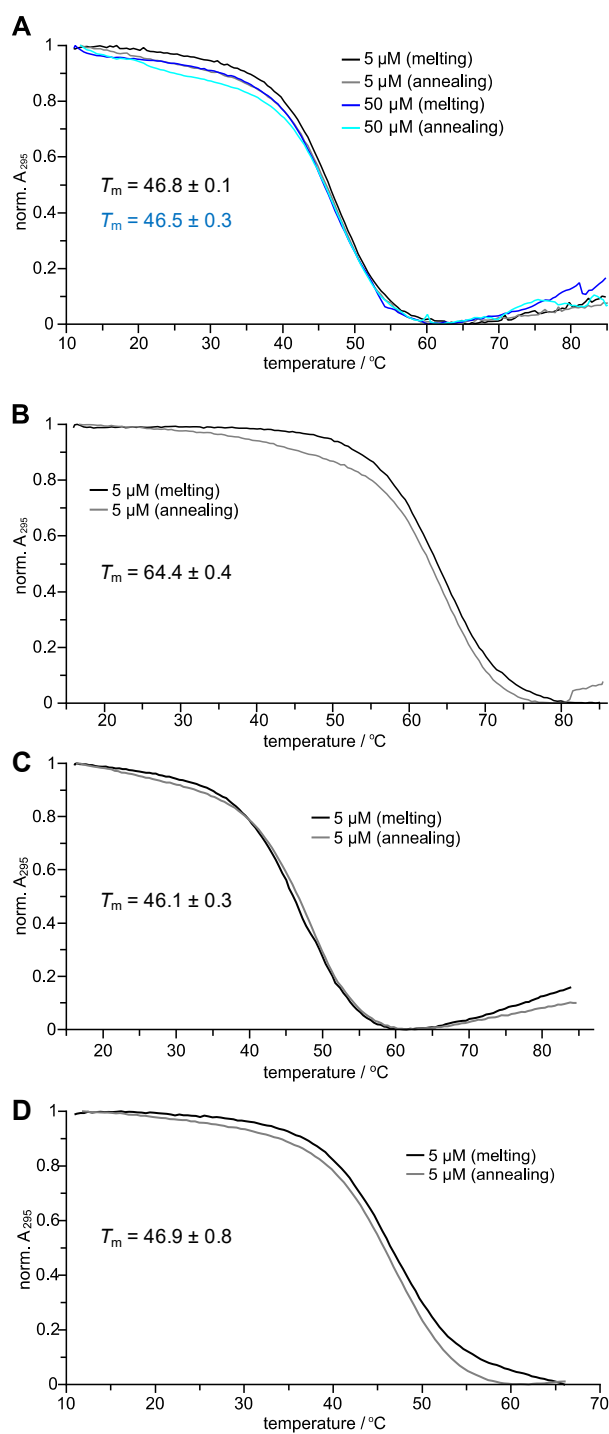


Figure S7. Representative UV melting curves. *KNA-ΔG13-T* in the presence of (A) 10 mM K^+ and (B) 120 mM K^+ ; (C) *KNA-ΔG13* and (D) *KNA-GGT* in the presence of 10 mM K^+ ; melting temperatures T_m are averages from three independent heating curves with standard deviations.

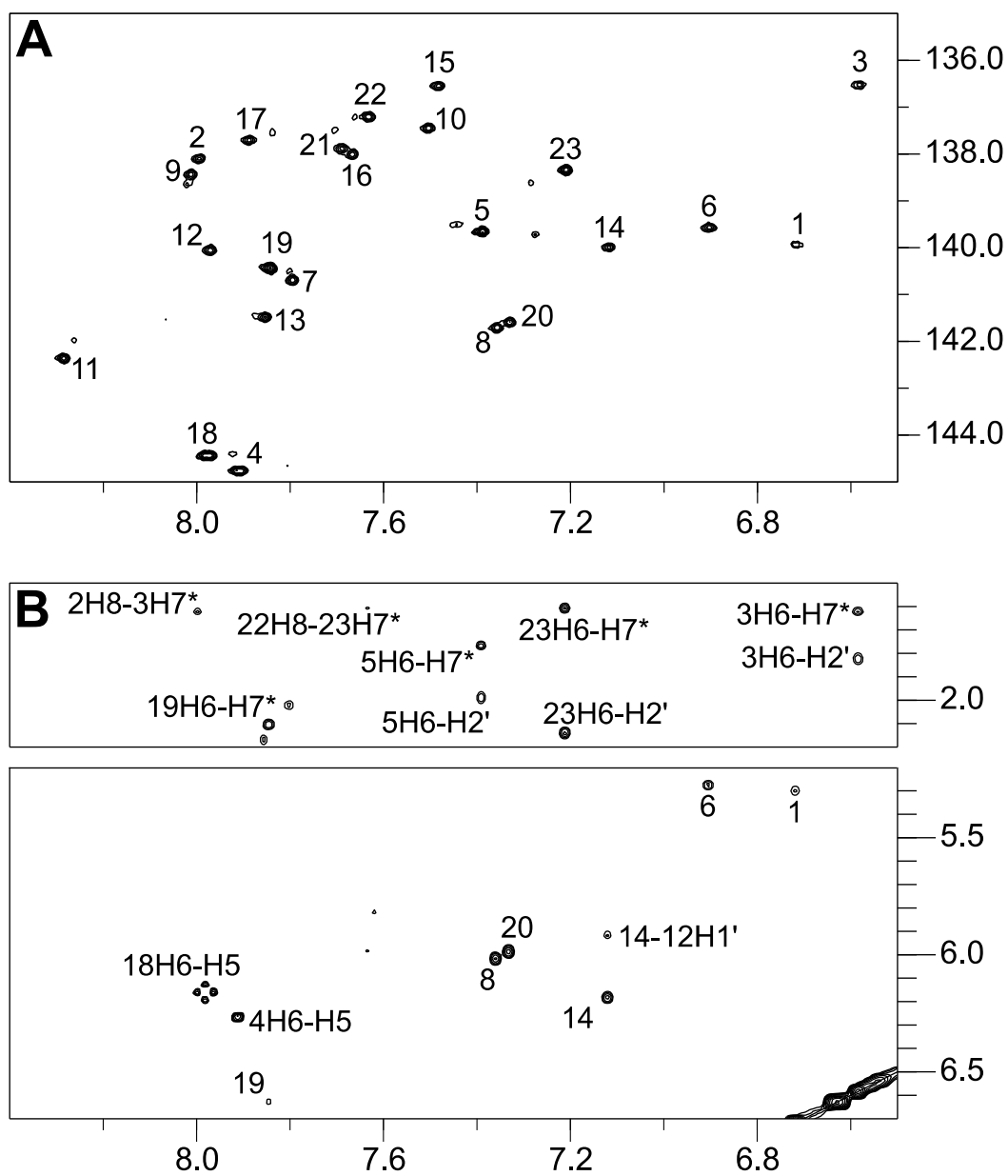


Figure S8. (A) H8/H6(ω_2)-C8/C6(ω_1) HSQC spectral region of *KNA-ΔG13-T* (0.94 mM) in a 10 mM K^+ buffer. (B) H8(ω_2)-H2'/Me(ω_1) (top) and H8(ω_2)-H1'(ω_1) spectral region (bottom) of a 2D NOESY spectrum of *KNA-ΔG13-T* acquired with an 80 ms mixing time at 25 °C in a 10 mM K^+ buffer.

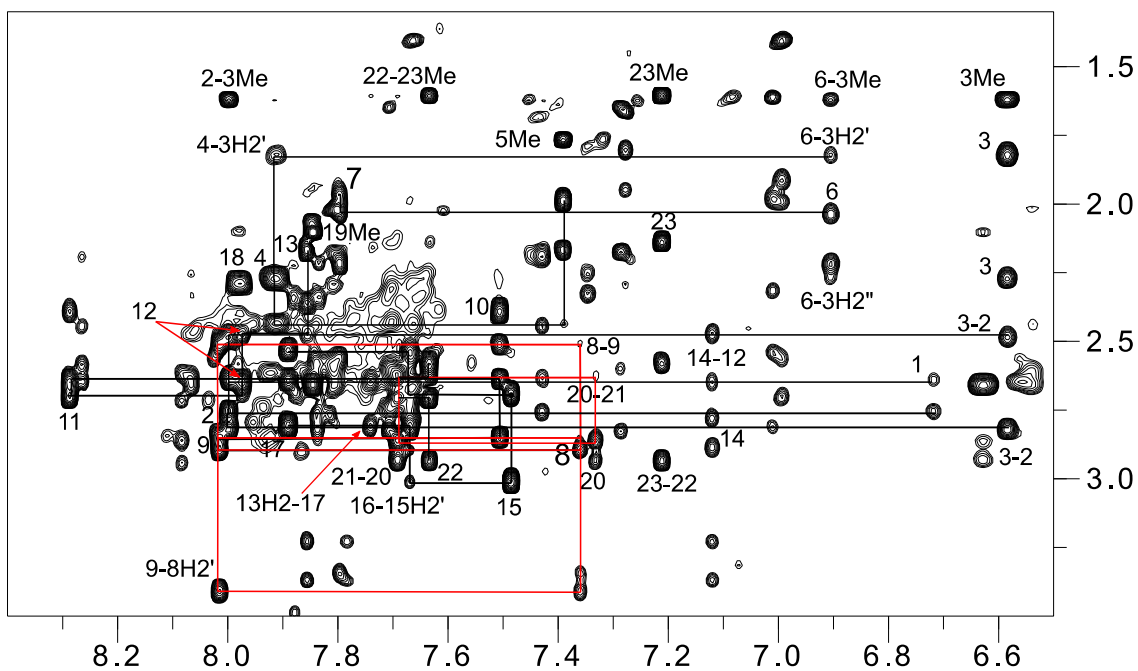


Figure S9. H8/H6(ω_2)-H2'/H2''/Me(ω_1) 2D NOESY spectral region of *KNA-ΔG13-T* (0.9 mM) acquired with a 300 ms mixing time; rectangular cross-peak patterns characteristic for *syn-anti* steps are colored red; the first residue number refers to H8/H6 resonances along ω_2 .

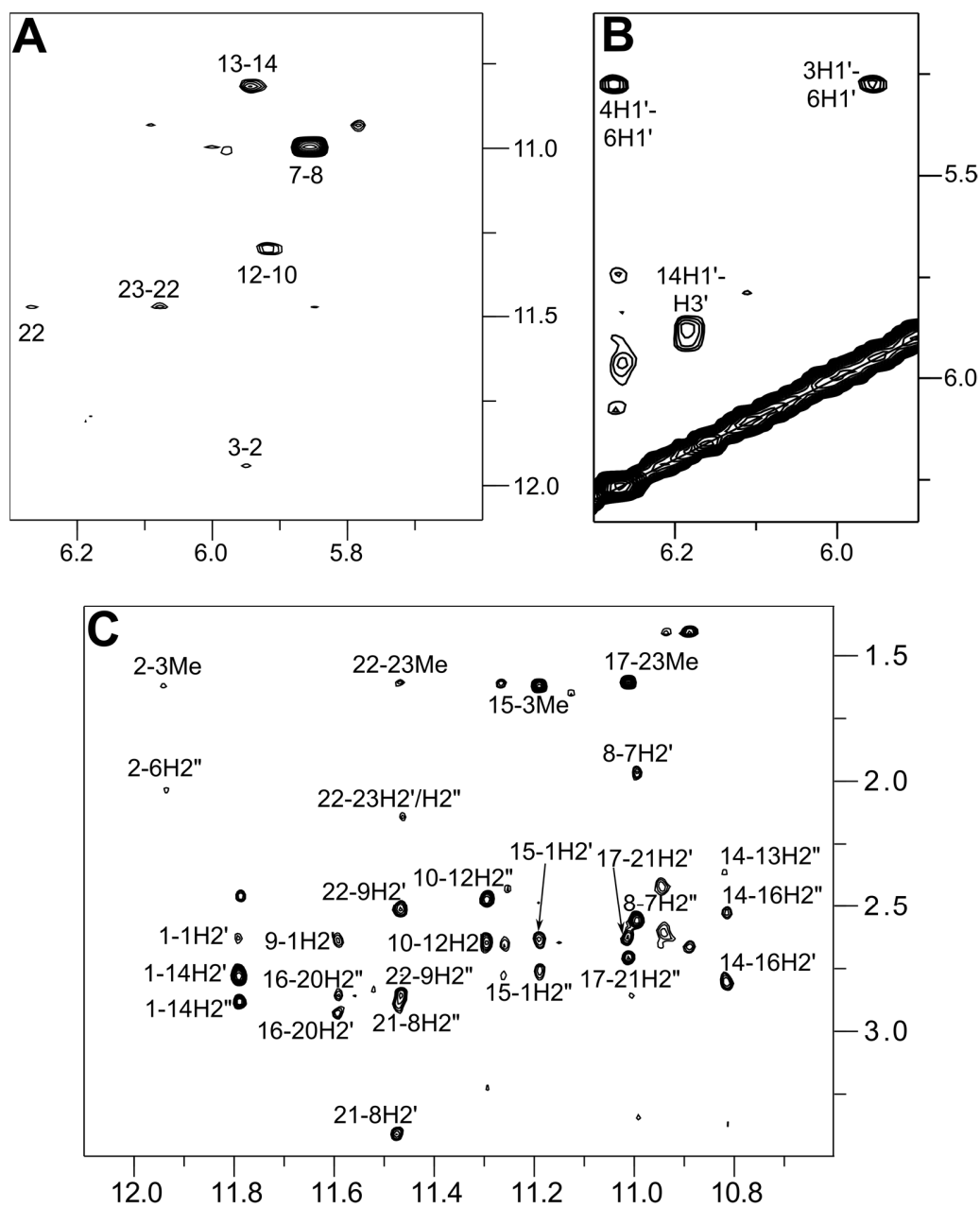


Figure S10. 2D NOESY spectrum of *KCNN4-ΔG13-T* (0.9 mM, 300 ms mixing time). (A) H1'(ω₂)-H1(ω₁) spectral region; (B) H1'(ω₂)-H1'/H3'(ω₁) cross-peaks within the first lateral loop; based on inter-residual H1'-H1' contacts, G6 is positioned close to T3 and C4; the unusually downfield-shifted G14 H3' resonance is also shown; (C) H1(ω₂)-H2'/H2''/Me(ω₁) spectral region; an NOE cross-peak between G9 H1 and G1 H2' indicates their opposite sugar-phosphate orientation.

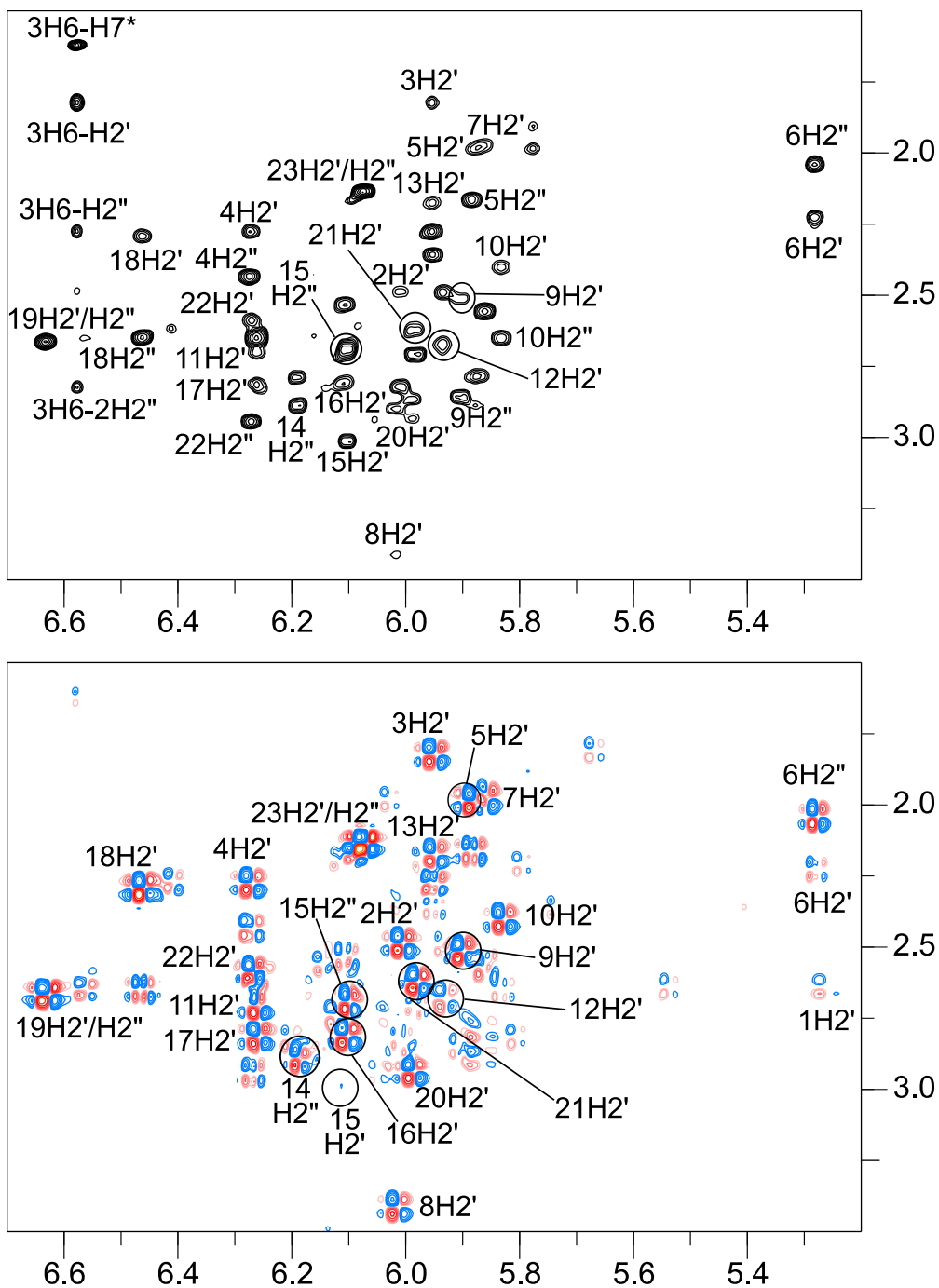


Figure S11. Determination of the sugar conformation. (A) Portion of a 2D NOESY spectrum (mixing time 80 ms) and (B) DQF-COSY spectral region of *KNA-ΔGI3-T* (0.9 mM) in 100% D₂O showing H1'(ω₂)-H2'/H2''(ω₁) cross-peaks.

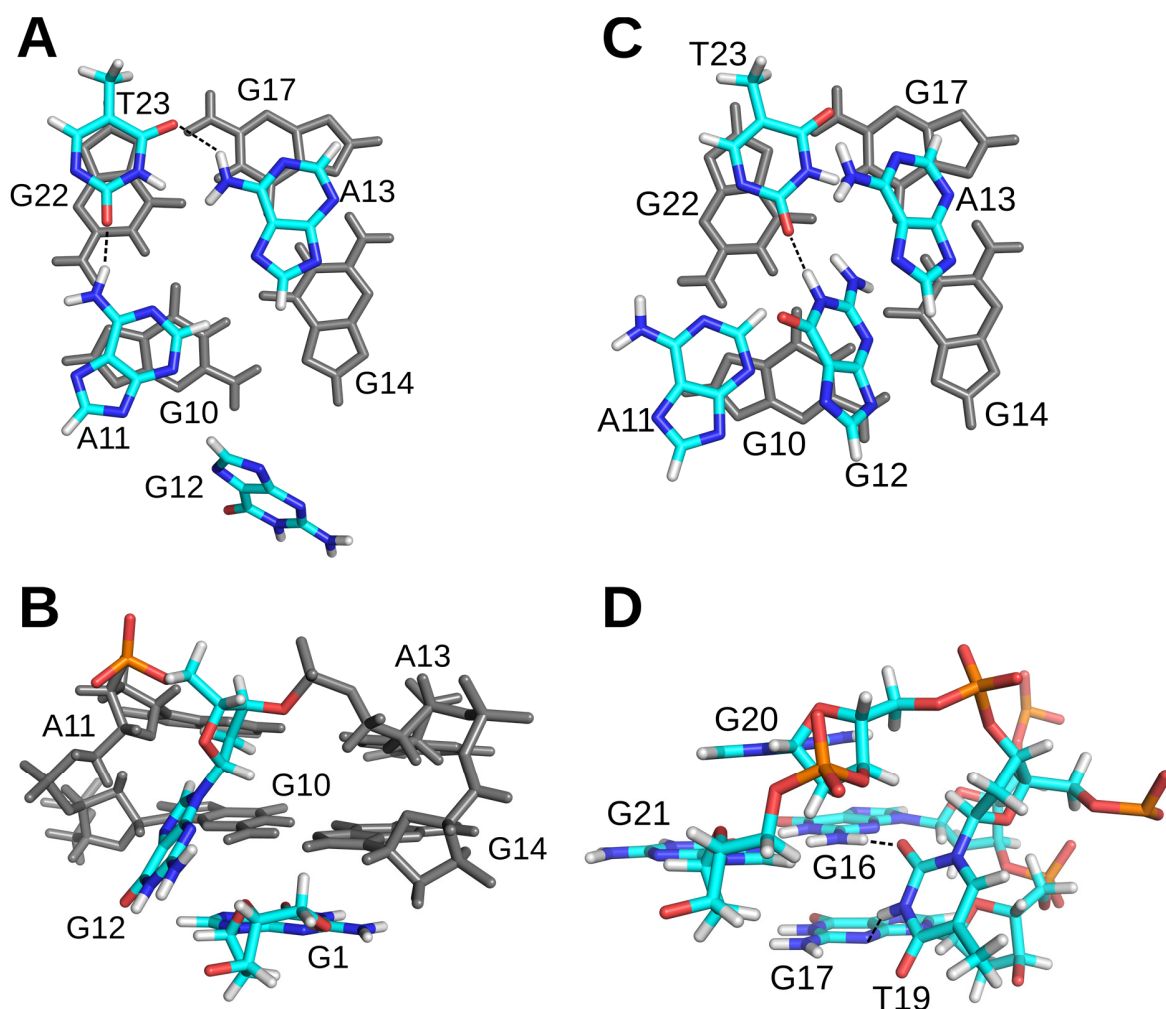


Figure S12. (A,B) Representative model of *KNA-ΔG13-T* with an orientation of the second lateral loop preceding the V-shaped loop as found in 8 out of 10 calculated structures. (A) Top view onto the second lateral loop and adjacent tetrad, highlighting a putative ATA triad between adenine bases of the lateral loop and 3'-terminal thymine T23. (B) Side view, highlighting G12 positioned nearly orthogonally to G1. (C) Top view onto the second lateral loop and adjacent tetrad, highlighting an alternative hydrogen bond formation between guanine G12 in the lateral loop and the 3'-terminal T23; tetrads are colored grey. (D) Putative hydrogen bond formation between the G16 amino proton and T19 O2 as well as between G17 N3 and T19 H3 of the propeller loop; C18 has been omitted for clarity.

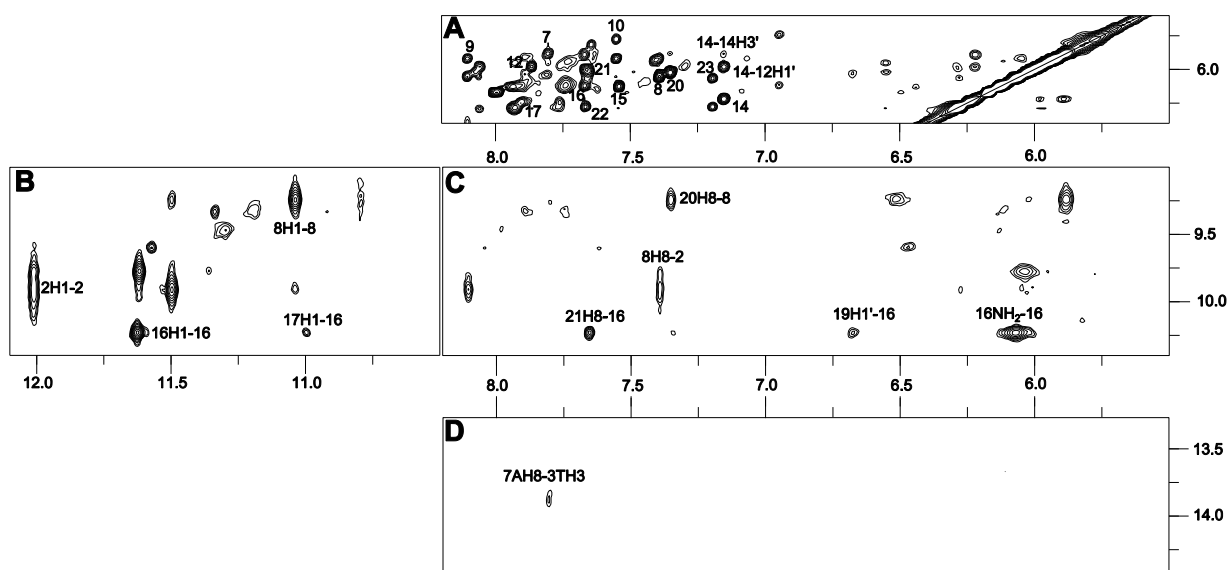


Figure S13. Portions of a 2D NOESY spectrum of *KNA-ΔG13-T* (0.9 mM, 300 ms mixing time, 278 K). (A) H6/8(ω_2)-H1'(ω_1) spectral region, (B) H1(ω_2)-amino proton(ω_1) spectral region, (C) H8/H1'/amino proton(ω_2)-amino proton(ω_1) spectral region. (D) Cross-peak which hints to the formation of an AT Hoogsteen hydrogen bond within the first lateral loop. Cross-peaks observed in this low-temperature spectrum were not included in the NOE-based distance restraints.

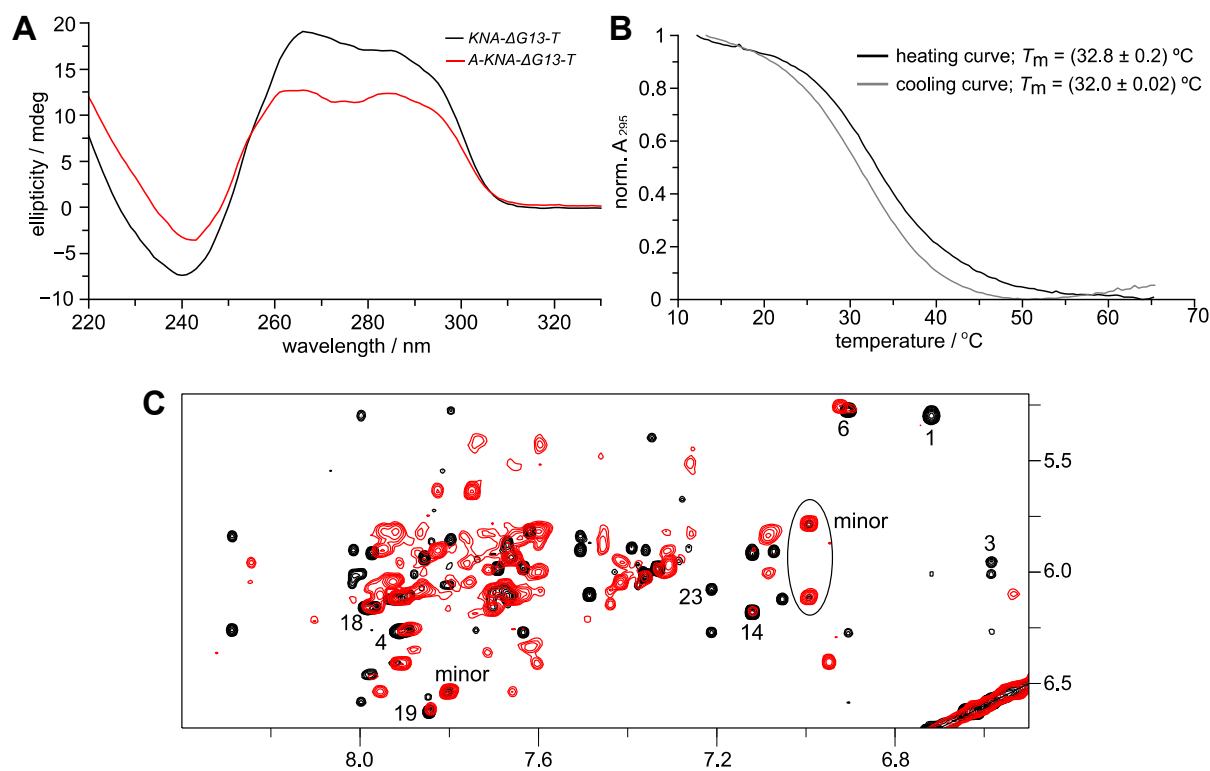


Figure S14. (A) CD spectrum of *A-KNA-ΔG13-T* (red) and *KNA-ΔG13-T* (black) in 10 mM K⁺ buffer. (B) UV melting and annealing curves of *A-KNA-ΔG13-T* in 10 mM K⁺ buffer with very small hysteresis effects. (C) Superposition of 2D NOESY spectra for *A-KNA-ΔG13-T* (red) and *KNA-ΔG13-T* (black) showing H6/8(ω_2)-H1'(ω_1) cross-peaks; labeled cross-peaks identify proton resonances of the V-shaped loop topology. Cross-peaks of the minor species coexisting with the V-shaped loop topology of *KNA-ΔG13-T* become more intense for *A-KNA-ΔG13-T* with a 5'-A overhang (circled); spectra were acquired at 25 °C.

Table S1. ^1H and ^{13}C chemical shifts (in ppm) of *KNA-ΔG13-T* (0.9 mM) at 25 °C in 10 mM potassium phosphate buffer, pH 7.

Residues	H6/H8	H2/5/Me	H1	H1'	H2'	H2''	H3'	C6/8	C5	C2
G1	6.72	-	11.79	5.30	2.64	2.76	4.93	139.93	119.59	-
G2	8.00	-	11.94	6.01	2.49	2.82	5.05	138.10	117.48	-
T3	6.58	1.62	-	5.96	1.82	2.27	4.82	136.53	-	-
C4	7.91	6.27	-	6.27	2.28	2.44	4.81	144.76	-	-
T5	7.39	1.77	-	5.89	2.28	2.44	4.81	139.67	-	-
G6	6.91	-	-	5.28	2.22	2.04	4.64	139.56	-	-
A7	7.80	7.22	-	5.85	1.97	2.56	4.81	140.71	-	153.44
G8	7.36	-	11.00	6.01	3.41	2.90	4.88	141.71	119.82	-
G9	8.01	-	11.59	5.90	2.51	2.85	5.01	138.44	116.84	-
G10	7.51	-	11.30	5.84	2.40	2.64	4.98	137.45	117.09	-
A11	8.29	7.78	-	6.26	2.70	2.65	4.92	142.36	-	154.54
G12	7.97	-	-	5.92	2.65	2.48	4.88	140.07	-	-
A13	7.86	7.74	-	5.94	2.17	2.36	4.72	141.49		154.25
G14	7.12	-	10.82	6.18	2.78	2.89	5.88	140.00	118.77	-
G15	7.49	-	11.19	6.10	3.01	2.70	5.07	136.56	118.24	-
G16	7.67	-	11.59	6.11	2.81	2.54	4.97	138.01	117.10	-
G17	7.89	-	11.01	6.26	2.81	2.65	4.97	137.72	117.66	-
C18	7.98	6.16	-	6.46	2.29	2.65	4.62	144.44	-	-
T19	7.84	2.10	-	6.63	2.66	2.66	5.04	140.44	-	-
G20	7.33	-	11.56	5.99	2.93	2.86	4.87	141.60	119.61	-
G21	7.69	-	11.47	5.98	2.63	2.71	5.08	137.89	116.77	-
G22	7.63	-	11.47	6.27	2.58	2.94	4.98	137.21	117.50	-
T23	7.21	1.61	-	6.08	2.14	2.14	4.51	138.35	-	-

Article 4

Expanding the Topological Landscape by a G-Column Flip of a Parallel G-Quadruplex

Swantje Mohr,^[a] Jagannath Jana,^[a] Yoanes Maria Vianney,^[a] and Klaus Weisz^{*,[a]}

Abstract: Canonical G-quadruplexes can adopt a variety of different topologies depending on the arrangement of propeller, lateral, or diagonal loops connecting the four G-columns. A novel intramolecular G-quadruplex structure is derived through inversion of the last G-tract of a three-layered parallel fold, associated with the transition of a single propeller into a lateral loop. The resulting (3 + 1) hybrid fold features three *syn-anti-anti-anti* G-tetrads with a 3'-terminal all-*syn* G-column. Although the ability of forming a duplex

stem-loop between G-tracts seems beneficial for a propeller-to-lateral loop rearrangement, unmodified G-rich sequences resist folding into the new (3 + 1) topology. However, refolding can be driven by the incorporation of *syn*-favoring guanosine analogues into positions of the fourth G-stretch. The presented hybrid-type G-quadruplex structure as determined by NMR spectroscopy may provide for an additional scaffold in quadruplex-based technologies.

Introduction

G-quadruplexes (G4s) have evoked great interest in the last two decades due to their detection under *in vivo* conditions and their putative regulatory roles in gene expression, making them promising targets for novel therapeutic strategies.^[1,2] In addition, these structures have been increasingly employed as versatile tools in bio- and nanotechnological applications, e.g., in sensor systems, electronic switches, or as DNazymes.^[3,4] In contrast to a B-type genomic DNA duplex, quadruplex functions and properties heavily rely on their tetra-stranded scaffold and may be exquisitely modulated by the remarkable G4 structural diversity. The latter depends on the nucleic acid sequence but also on the outer environment and determines G4 specific interactions with other molecules or ions.

G-quadruplexes can be formed by the intramolecular folding of a single-stranded nucleic acid comprised of four G-runs interrupted by intervening nucleotides. Association of G residues from each of the four G-tracts will form a quadruplex core with stacked G-tetrads of four guanine bases held together in a square-planar arrangement through cyclic hydrogen bonds. The stacked architecture is further stabilized by the coordination of metal ions like K⁺ or Na⁺ in its central cavity. Intervening nucleotides in such an intramolecular assembly will form loops connecting the G-columns of the G4 core structure. Propeller loops link two adjacent columns with the same 5'-3'

backbone orientation, whereas lateral and diagonal loops connect neighboring or distal G-columns of antiparallel orientation, respectively.

With the formation of three loops in a canonical quadruplex with non-interrupted G-columns, there are $3^3 = 27$ theoretical loop combinations giving rise to different topologies. Additional topologies come from a clockwise (+) or anti-clockwise (-) progression of propeller and lateral loops when placed in a common frame of reference as previously proposed.^[5] For characterizing and discriminating the various folds, a simple descriptor composed of the type and progression of consecutive loops has been suggested.^[5,6] Thus, abbreviating lateral, propeller, and diagonal loops with 'l', 'p', and 'd', respectively, the minimalistic designation (-pd+l) identifies a quadruplex topology with a first propeller loop running counter-clockwise, a central diagonal loop, and a third lateral loop running clockwise. Clearly, only a fraction of the theoretical topologies can be realized for mechanical and steric reasons and fragment-based modeling studies have identified 14 topologies that may form under appropriate conditions.^[7,8]

G-quadruplexes of a (3 + 1) hybrid-type with one antiparallel and three parallel G-columns are a recurrent G4 structural motif and typically represent a hybrid-1 (-p-l-l) or a hybrid-2 (-l-l-p) topology. Recently, we have shown refolding of a parallel all-*anti* G4 into a novel (3 + 1) hybrid quadruplex termed hybrid-1R with a (+l+p+p) loop progression, exploiting base complementarity of 5'- and 3'-overhang sequences to form a duplex smoothly extending at one face of the rearranged G4 core.^[9] This topology, associated with a switch from a first propeller into a lateral loop, could be further enforced by the specific incorporation of *syn*-favoring 8-bromo-guanosine (^{Br}G) nucleotides, known to be powerful tools for conformational transitions.^[10] In the present studies, we aimed at the formation of another (3 + 1) hybrid species with a (-p-p-l) topology starting from a parallel fold (Figure 1). Although a corresponding fold is highly ranked in a topology distribution chart of

[a] S. Mohr, Dr. J. Jana, Y. M. Vianney, Prof. Dr. K. Weisz
Institute of Biochemistry, Universität Greifswald
Felix-Hausdorff-Str. 4, 17487 Greifswald (Germany)
E-mail: weisz@uni-greifswald.de

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/chem.202101181>

© 2021 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

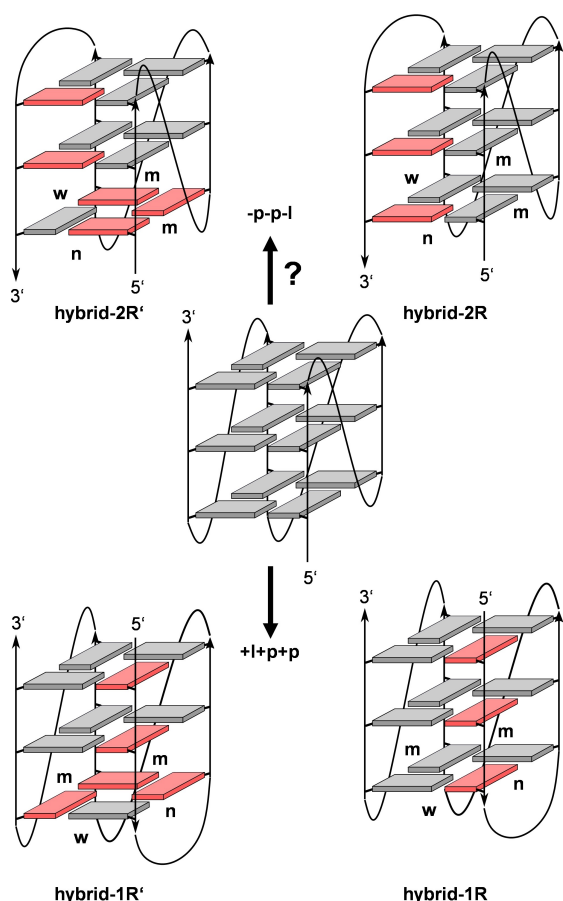


Figure 1. Topology of a parallel quadruplex (center) and possible (3 + 1) hybrid structures formed by flipping the fourth G-tract (top) and the first G-tract (bottom); n, m, and w denote narrow, medium, and wide groove, respectively.

model G-quadruplexes built from reported fragments, it has not been experimentally verified to date.^[7,8] Such a species may thus not only provide insight into critical interactions upon refolding but may also expand the known topological diversity

of G-quadruplex structures for future use in the increasing number of G4-based applications.

Results

The sequence design of a putative -p-p-l G-quadruplex is based on the parallel *c-myc* quadruplex, previously characterized in detail.^[11] Whereas a single nucleotide largely restricts folding of the 1-nt intervening sequence into a propeller-type loop, 3-nt and 4-nt lateral loops spanning the wide groove have most frequently been observed in fragment-based modeling studies for a -p-p-l topology.^[7] We therefore started with sequence *L113* comprising a 1:1:3 nucleotide loop length arrangement. As shown in Figure 1, refolding through a switch from a propeller into a lateral loop results in a flip of the fourth G-tract but is also associated with a relocation of the 3'-overhang to the 5'-face of the quadruplex core. Additionally, *anti*→*syn* conversions are enforced for some G residues to maintain G-quartet formation and these will depend on the glycosidic bond angle pattern along the inverted strand. To probe the impact of differently modified sequences on these structural rearrangements, the fold for a total of 12 oligonucleotides was analyzed in more detail. These include variants with different loop sequences, some of which allow for a stem-loop structure, as well as 8-bromo-2'-deoxyguanosine modified analogues (Table 1).

Circular dichroism (CD)

CD spectra of all sequences exhibit a signature typical of quadruplexes with exclusive homopolar G-quartet stacking, i.e., a positive band at 265 nm and a negative band at 245 nm (Figure 2, for a compilation of all spectra see Figure S1). A noticeable asymmetry towards longer wavelength of the positive band and a slightly red-shifted maximum is generally observed for all sequences bearing third loops with Watson-Crick complementary bases. This points to the formation of a B-type duplex extrusion and a B-DNA signature superimposed on

Table 1. Sequences with UV melting temperatures T_m .^[a,b]

Name	Sequence	T_m [°C]
<i>L113</i>	TGA GGG I GGG I GGG <u>ICA</u> GGG TAA	62.0 ± 0.2
<i>L112-B12</i>	TGA GGG I GGG I GGG <u>IC</u> BBG TAA	59.2 ± 0.6
<i>L114-B12</i>	TGA GGG I GGG I GGG <u>ACTI</u> BBG TAA	50.7 ± 0.6
<i>L113-B12</i>	TGA GGG I GGG I GGG <u>ICA</u> BBG TAA	55.0 ± 0.6
<i>L113-B23</i>	TGA GGG I GGG I GGG <u>ICA</u> GBB TAA	54.3 ± 0.7
<i>L113-B13</i>	TGA GGG I GGG I GGG <u>ICA</u> BGB TAA	54.0 ± 1.1
<i>L113-B123</i>	TGA GGG I GGG I GGG <u>ICA</u> BBB TAA	59.7 ± 0.1
<i>L11GC</i>	TGA GGG I GGG I GGG <u>GCGCGCAGCGC</u> GGG TAA	54.4 ± 0.4
<i>L11AT</i>	TGA GGG I GGG I GGG <u>ACGCGCAGCGT</u> GGG TAA	59.8 ± 0.9
<i>L11AT-B12</i>	TGA GGG I GGG I GGG <u>ACGCGCAGCGT</u> BBG TAA	52.4 ± 0.3
<i>L11AT-B13</i>	TGA GGG I GGG I GGG <u>ACGCGCAGCGT</u> BGB TAA	53.5 ± 0.7
<i>L11AT-B23</i>	TGA GGG I GGG I GGG <u>ACGCGCAGCGT</u> GBB TAA	53.4 ± 0.6

[a] Loop nucleotides are underlined and complementary sequences forming putative stem-loop structures are written in italic; B=8-bromo-2'-deoxyguanosine. [b] Average values with standard deviations from three independent measurements taken at 295 nm in 10 mM potassium phosphate buffer, pH 7.

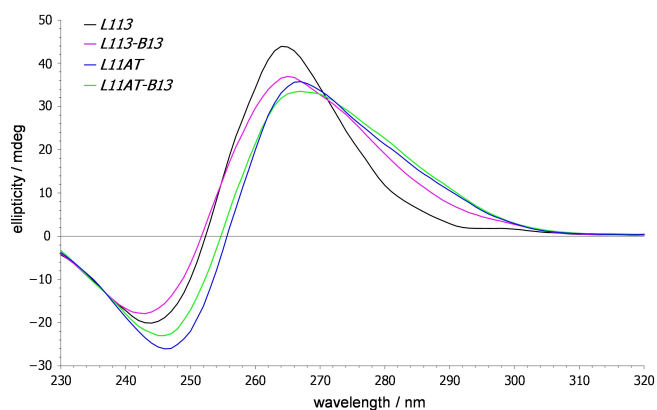


Figure 2. CD spectra of *L113*, *L113-B13*, *L11AT*, and *L11AT-B13*. Spectra were acquired at 20 °C on DNA samples (5 μ M) in 10 mM potassium phosphate buffer, pH 7.

the G4 CD spectral features. Of note, CD spectra are compatible with a parallel as well as a (3 + 1) hybrid topology harboring a single all-*syn* G-tract, but do not allow their mutual discrimination. However, based on the CD experimental results, preferential formation of other hybrid or antiparallel quadruplexes with G-tetrads of different polarity to result in heteropolar stacking interactions can safely be excluded for each of the G-rich sequences.

NMR spectroscopy

Initially, the impact on G4 folding of loop length and of *syn*-favoring G analogues in different positions of the fourth G-column was assessed by a comparison of imino proton NMR spectral regions (Figure 3). Due to their engagement in hydrogen bonds within a G-tetrad, twelve guanine imino resonances in slow exchange with the solvent are expected for a three-layered G-quadruplex. In fact, the observation of about twelve major imino signals for each of the sequences lacking base complementarity in a longer third loop attests to their formation of a G-quadruplex with three stacked tetrads. A closer look over all spectra reveals two different patterns of imino signals, suggesting two different topologies depending on the sequence. One pattern, covering a wider spectral range is shared by *L113* and *L112-B12*, whereas the other pattern is shared by *L113-B23*, *L113-B13*, and *L113-B123*. The presence of a major and minor set of imino resonances for *L114-B12* and *L113-B12* suggests coexistence of both topologies, albeit in different molar ratios.

Sequences *L112-B12* and *L113-B123*, each representing one type of fold, were chosen for an initial determination of topologies. Also, their well-resolved imino proton resonances favor a more detailed NMR structural analysis. For *L112-B12*, a straightforward identification of a standard parallel all-*anti* topology was based on the absence of *syn*-guanosines but also on conspicuous NOE contacts between H8 of 3'-terminal A22 and G6 as well as G19 imino protons located at the G-core face opposite to the 5'-overhang (Figure S2). Thus, all guanine H8-C8 cross-peaks of *L112-B12* are located in a ^1H - ^{13}C HSQC spectral

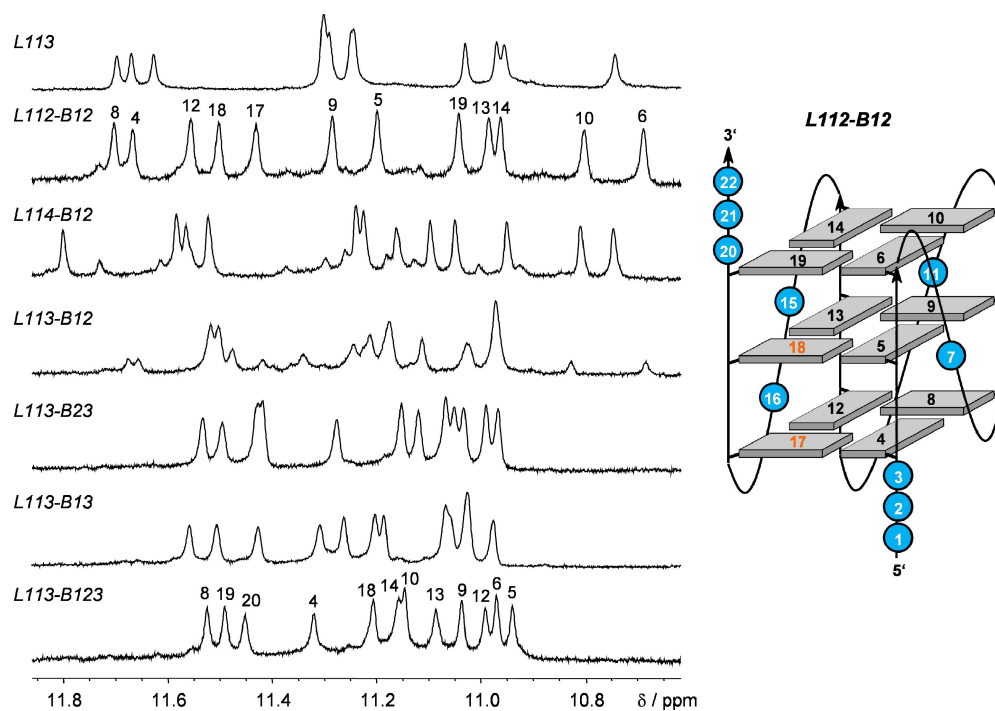


Figure 3. Imino proton spectral region of G4-forming sequences without a putative duplex stem-loop structure; residue assignments are given for *L112-B12* and *L113-B123*. Topology of the *L112-B12* quadruplex with residue numbers (in orange for $^{\text{Br}}\text{G}$ analogues) is shown on the right.

region typical for an *anti* glycosidic conformation (Figure S3A) and this is further supported by the lack of strong intra-nucleotide H8-H1' and associated weak H8-H2'/H2'' contacts in a NOESY spectrum, anticipated for *syn*-residues (for assignments and a compilation of *L112-B12* chemical shifts see Table S1).

The identification of *syn* residues for a topological evaluation of *L113-B123* is restricted due to the absence of G H8 protons in the fully ^{Br}G-modified fourth G-tract. However, various NOE contacts of residues in the third 3-nt loop strongly suggest its lateral progression with a flipped antiparallel fourth G-column in a hybrid-type fold. These include sequential contacts along the loop and cross-peaks between H2/H8 of loop residue A17 with three out of four imino protons in the outer tetrad opposite the 5'-terminus (Figure 4; for assignments and a compilation of *L113-B123* chemical shifts see Table S2).

A close structural similarity of *L113* with *L112-B12* as well as of *L113-B13* and *L113-B23* with *L113-B123*, initially suggested through corresponding signal patterns in the imino proton spectral region, was confirmed by a more detailed analysis of corresponding NOESY and ¹H-¹³C HSQC spectra (not shown). Thus, a single *syn*-G compatible with a ^{Br}G-disubstituted all-*syn* column could be identified in *L113-B13* and *L113-B23*. Also, NOE contacts from loop and 3'-overhang residues to outer tetrad resonances in all sequences could either be traced to a third propeller loop and overhang sequences on opposite sides of the G-core (*L113*) or to a third lateral loop with both termini facing each other (*L113-B13* and *L113-B23*).

Assignments for *L113-B12* and *L114-B12*, exhibiting two coexisting species based on their imino proton spectral region, again relied on spectral comparisons with both the *L112-B12* parallel quadruplex and the *L113-B123* hybrid-type G4. Thus, a

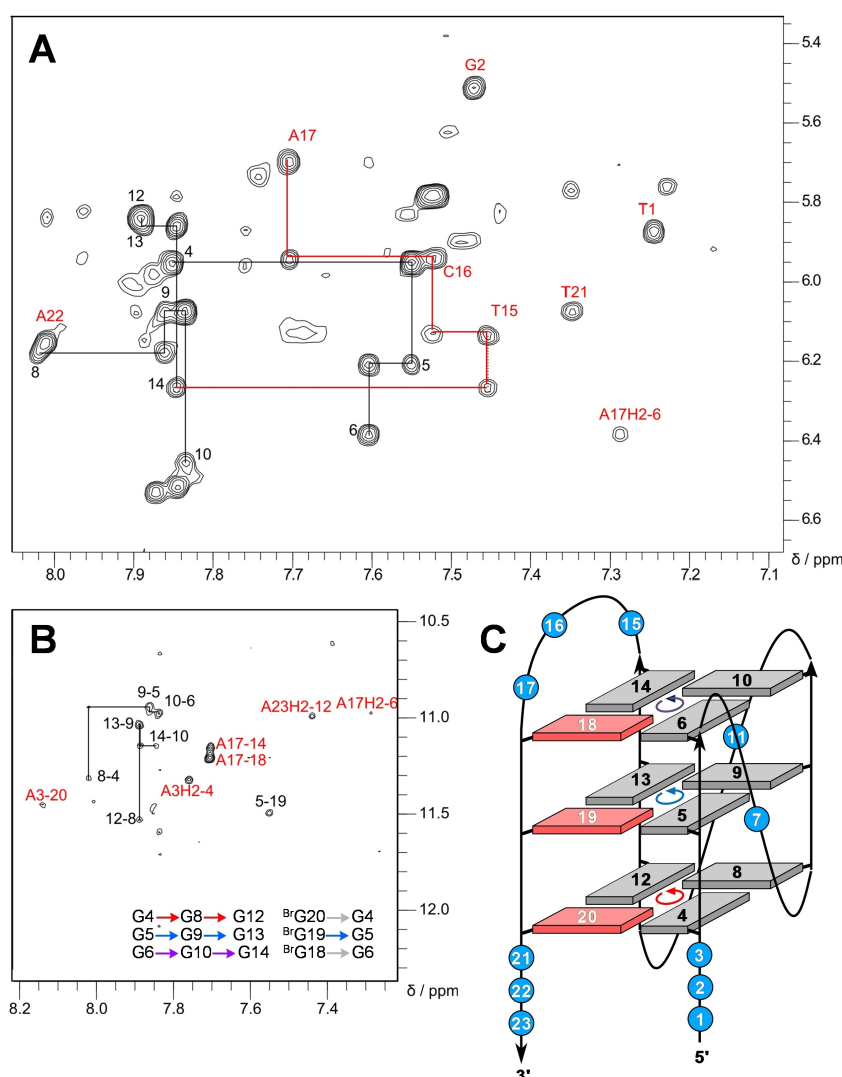


Figure 4. 2D NOE spectral regions of *L113-B123*. (A) H8/6(ω_2)-H1'(ω_1) cross-peaks with sequential contacts traced along the G-core (black) as well as along the 3-nt third loop (red). (B) Intra- and inter-tetrad H8(ω_2)-H1'(ω_1) cross-peaks and additional contacts involving overhang residues and A17 of the 3-nt loop labeled in red. Tetrad polarities as determined from intra-tetrad NOE contacts are summarized at the bottom (grey arrows indicate missing contact). (C) Topology of the *L113-B123* quadruplex with residue numbers and tetrad polarities indicated; *syn*-^{Br}G analogues are shown in red with their residue numbers in white.

superposition of HSQC and NOESY spectral regions of *L113-B12* with those of *L112-B12* and *L113-B123* clearly demonstrated its folding into a major hybrid-type topology with the observation of a single *syn*-G and a minor parallel topology with a population ratio estimated to be about 7:3 (Figure S3). In contrast, the major species of *L114-B12* can likewise be assigned to a parallel fold with cross-peak patterns very similar to *L112-B12* (Figure S4). A decrease in the melting temperature for *L114-B12* by almost 10 °C when compared to *L112-B12* can thus directly be attributed to a destabilizing effect from its longer 4-nt propeller loop (Table 1). With a population of only about 20%, minor peaks of a putative (3 + 1) hybrid fold in *L114-B12* spectra do not allow their unambiguous assignments.

Quadruplexes with a duplex stem-loop

To further promote a switch from the third propeller into a lateral loop with formation of a hybrid-type quadruplex, the loop sequence was expanded to enable hairpin formation through Watson-Crick base pairing. This should allow for a coaxial orientation with contiguous base stacking interactions of a putative hairpin-type lateral loop and the quadruplex helix.^[12] Initially, two sequences *L11AT* and *L11GC* with a 11-nt third loop capable of stem-loop formation and with AT or GC complementary bases adjacent to the G-core were tested for their favored topology.

Inspection of their guanine Hoogsteen imino proton spectral region between 11.8 and 10.6 ppm reveals a closely similar signal distribution to *L113*, indicating a parallel G4 topology for both sequences (Figure 5). Additional downfield-shifted resonances between 12.8 and 13.2 ppm can be attributed to imino protons engaged in GC Watson-Crick base pairs exchanging with the solvent at different rates. Notably, there is no indication of a stable AT Watson-Crick base pair at the quadruplex-duplex junction in *L11AT* because spectra lack an expected imino signal at low field. These observations are consistent with base pairing in a propeller loop, enforcing an orthogonal orientation of duplex and quadruplex helices with unpaired bases at the junction.^[12] Featuring additional base pairs within the loop devoid of additional stacking interactions with outer G-tetrads, G4 stability decreases with the longer 11-nt loop compared to *L113* with its 3-nt loop (Table 1). A more detailed analysis of *L11GC* spectra confirms its parallel fold characterized by (i) the absence of *syn*-Gs in HSQC spectra (not shown), (ii) NOE contacts between residues of the 5'- and 3'-overhang with opposite outer tetrads, and (iii) the absence of sequential contacts between loop residues and the G-core (Figure S5, for assignments and a compilation of *L11GC* chemical shifts see Table S3).

Adopting the same parallel fold as *L11GC*, we selected the *L11AT* quadruplex for additional ^{Br}G substitutions with a putative AT base pair at the quadruplex-duplex junction serving as a convenient probe for refolding into a hybrid-type topology. In fact, the pattern of Hoogsteen imino protons together with the observation of an additional downfield-shifted AT Watson-Crick imino resonance at ~13.5 ppm suggests a non-parallel

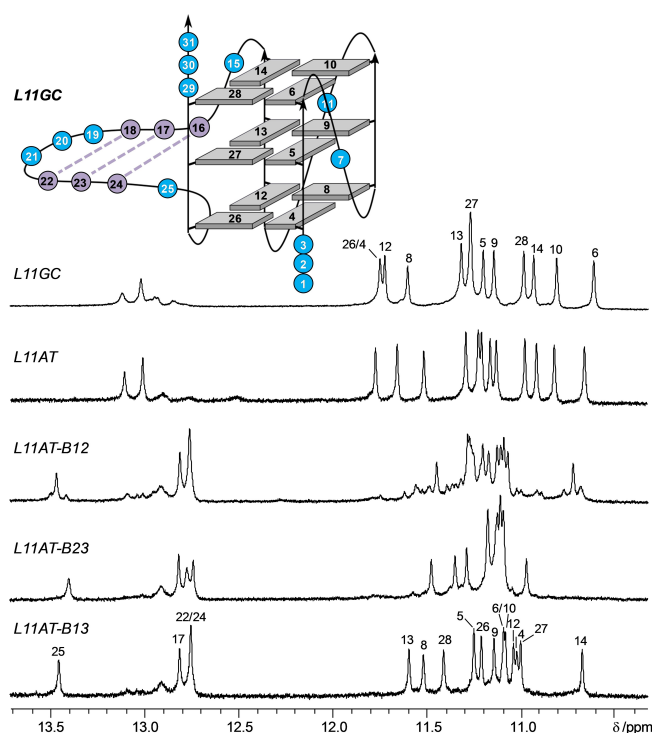


Figure 5. Imino proton spectral region of G4-forming sequences with a duplex stem-loop structure; residue assignments are shown for *L11GC* and *L11AT-B13*. Topology of the *L11GC* quadruplex with residue numbers is shown on top; GC Watson-Crick base pairing within the 11-nt propeller loop is indicated by broken lines.

fold for *L11AT-B12*, *L11AT-B13*, and *L11AT-B23* harboring two ^{Br}G modifications in their 3'-terminal G-tract. Formation of a lateral hairpin structure with an AT base pair at the quadruplex-duplex junction is corroborated by sequential base-sugar contacts which can be traced along the duplex domain into the adjacent G-tracts (Figure 6). Also, folding into a hybrid-type structure for all three dual-modified sequences is suggested by the observation of a single *syn*-G in HSQC and NOESY spectra (see below). Interestingly, *L11AT-B13* and *L11AT-B23* form a single G4 species with about the same melting temperature as those of *L113-B13* and *L113-B23* with short 3-nt lateral loops, excluding significant synergistic effects of duplex and quadruplex domains (Table 1). For *L11AT-B12*, another structure coexists with the major hybrid structure. However, with a population of only about 10%, this minor species escapes a more detailed structural investigation. In the following, we focused on *L11AT-B13* for a three-dimensional structure determination due to its superior signal dispersion with minimal signal overlap.

Structure determination of *L11AT-B13*

Ample evidence from initial NMR spectral analyses points to a (3 + 1) hybrid fold for *L11AT-B13* (Figure 6A). Based on such a topology, complete resonance assignments followed standard procedures. Continuous H8/H6-H1'/H3' and H8/H6-H2'/H2'' NOE walks can be traced along all G-tracts except for the dual-

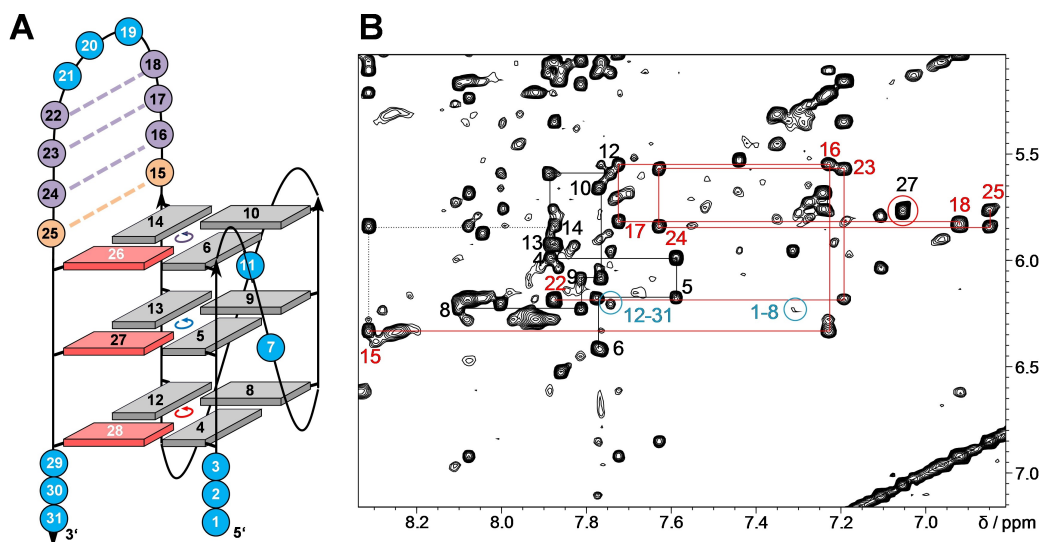


Figure 6. (A) Topology of the *L11AT-B13* quadruplex with residue numbers and tetrad polarities indicated; AT and GC Watson-Crick base pairs within the 11-nt stem-loop are indicated by broken orange and violet lines, respectively; *syn*-BrG analogues are shown in red with white residue numbers. (B) 2D NOESY spectrum of *L11AT-B13*. H8/6(ω_2)-H1'(ω_1) connectivities are traced along the G-columns (black) and the duplex stem (red) with an intra-residual contact of *syn*-G27 circled in red; note a sequential NOE contact between residue 14 and 15 at the quadruplex-duplex interface and cross-peaks (circled in blue) between residues of both 5'- and 3'-overhangs with the 5'-outer tetrad.

modified G-column at the 3'-end. Likewise, corresponding sequential NOE cross-peaks connect residues within the duplex stem-loop as well as A15 and G14 at the duplex-quadruplex interface (Figure 6B). In addition, residues of the duplex stem comprising four Watson-Crick base pairs exhibit well-resolved sequential H8/6-H8/6 cross-peaks (Figure S6A). Non-modified G27 within the 3'-terminal G-tract is identified through its *syn* conformation, exhibiting strong intra-nucleotide H8-H1' and weak H8-H2'/H2'' cross-peaks in NOESY spectra acquired with short mixing times (Figure S7A). A single unmodified *syn*-G27 is further corroborated by its characteristic C8-H8 cross-peak chemical shift in a ^1H - ^{13}C HSQC spectrum (Figure S7B).

Sequential imino-imino NOE contacts link G residues along the same G-column (Figure 7A). Notably, whereas inter-tetrad cross-peaks of imino protons to nearest H2'/H2'' protons located in residues on the 3'-adjacent G-tract are characteristically weak and often unobservable for parallel all-*anti* G-columns, conspicuous sequential H1(n)-H2'/H2''(n-1) cross-peaks, typical for *syn*-residues,^[9,13] were observed along the fourth G-tract and comprise H2'/H2'' of 5'-adjacent T25 of the lateral stem-loop (Figure S6B). Finally, H1-H8 NOE contacts between adjacent Gs within tetrads, only interrupted by missing H8 protons of the two BrG analogues in the outer quartets, identify the direction of Hoogsteen hydrogen bonds and demonstrate the same polarity for all tetrads (Figure 7B).

There are several NOE contacts at the quadruplex-duplex interface. These include NOE cross-peaks of G14 with A15, T25 with BrG26, but also of G6 and G14 with T25, suggesting a well-defined quadruplex-duplex junction (see Figures 7B and S6). This is also corroborated by H₂O-D₂O exchange experiments, showing protection from solvent exchange for G residues of the central tetrad but also slower exchange for G14 and G26 at the quadruplex-duplex junction consistent with continuous stack-

ing (Figure 7C). Few contacts between overhang residues and the adjacent G-tetrad such as G12 H1-A30 H2, G28 H1-A3 H8, G12 H8-A31 H1', and a weak cross-peak between G8 H1' and T1 H6 point to, albeit more flexible, structured 5'- and 3'-flanking sequences (Figure 6B and 7B).

H2' and H2'' resonances were stereospecifically assigned through cross-peak intensities of intra-nucleotide H8-H2'/H2'' and H1'-H2'/H2'' contacts in NOESY spectra acquired at short mixing times. Evaluation of vicinal H1'-H2' and H1'-H2'' scalar couplings in DQF-COSY spectra demonstrated *south* sugar conformations for all residues of the quadruplex and the hairpin domain except for G-core nucleotides G6, G13, G4, and G26 as well as for stem-loop nucleotides C16, G24, and T25. For the latter, sugar conformations remained ambiguous due to heavy signal overlap (Figure S8).

For subsequent structure calculations, NOE-derived distance restraints but no sugar torsion angle restraints were employed (for structural statistics see Table S5). The quadruplex adopts a novel (-p-p-l) topology with two propeller followed by a lateral loop running in a counter-clockwise direction (Figure 8A). We term this quadruplex fold hybrid-2R in analogy to the previously reported hybrid-1R quadruplex with a (+1+p+p) topology derived from hybrid-1 structures by formally substituting propeller for lateral loops and vice versa. Looking at a superposition of the ten lowest-energy structures, the G-core, the 4-bp duplex stem, and especially the quadruplex-duplex interface are well-defined (Figure 8A). In contrast, propeller loops and the loop of the duplex hairpin are less ordered, but also 5'- and 3'-overhang residues seem to experience significant flexibility. With all G residues of the three parallel G-tracts adopting an *anti* glycosidic conformation, the antiparallel fourth G-tract with its two BrG modifications exclusively features *syn*-residues.

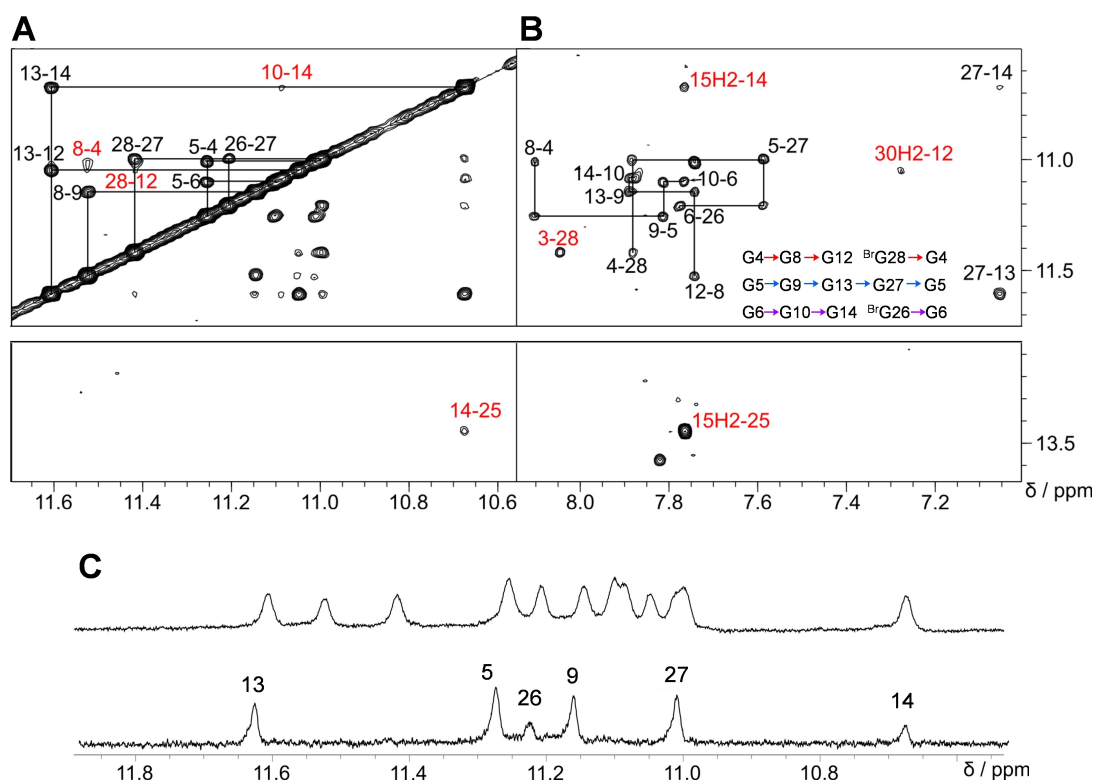


Figure 7. (A) Imino-imino 2D NOE spectral region of *L11AT-B13* with sequential cross-peaks and additional contacts at the junction or within tetrads labeled in red. (B) Intra- and inter-tetrad H8(ω_2)-H1(ω_1) cross-peaks; contacts at the quadruplex-duplex junction, between 5'- and 3'-overhang residues with the 5'-outer tetrad, and within the AT Watson-Crick base pair are labeled in red. Tetrad polarities as determined from intra-tetrad NOE contacts are summarized as inset. (C) H₂O-D₂O exchange experiments. Imino proton spectral region of *L11AT-B13* acquired in H₂O-D₂O (9:1) (top) and shortly after drying and redissolving in 100% D₂O (bottom); imino protons for residues of the central tetrad and, albeit to a smaller extent, of G14 and ^{Br}G26 at the quadruplex-duplex junction are protected from fast solvent exchange.

The stem-loop structure stacks with its terminal AT base pair coaxially on G14 and ^{Br}G26 of the upper quadruplex tetrad (Figure 8B), rationalizing protection of these outer G-core residues from fast solvent exchange as found experimentally (Figure 7C). Because the wide groove of the G-core bridged by the duplex stem-loop and the minor groove of a B-type duplex are of similar width, the quadruplex-duplex junction allows for a smooth transition without major distortions of the sugar-phosphate backbone.

Although less defined and rather flexible, overhang sequences seem to form loose capping structures stacked below the G-tetrad. Thus, corroborated by 2D NOE data, all calculated low-energy structures show stacking of A3 on either G4 or ^{Br}G28 and stacking of T1 on G8 is observed in 7 out of 10 structures. Also, various alignments of flanking bases are observed in the final structures with the lowest energy structure featuring an A-A-T base triad involving 5'-flanking residues T1 and A3 as well as 3'-flanking residue A30 (Figure 8B). Hydrogen bonds in such a base triad lack final confirmation by experimental data, pointing to a high flexibility with fast dissociation rates in rather short-lived base arrangements.

Probing refolding through ligand binding

Refolding of the parallel *L11AT* quadruplex into a hybrid-type quadruplex upon ^{Br}G substitutions in its fourth G-tract results in a coaxial stacking of the stem-loop duplex onto the outer G-tetrad with the formation of a quadruplex-duplex interface. Quadruplex-duplex (Q-D) junctions may serve as specific recognition elements for ligands.^[14,15] In fact, a corresponding Q-D junction has recently been shown to constitute a favorable binding site for a ligand termed PIQ which is based on a phenyl-substituted indoloquinoline heterocyclic ring system.^[16] Isothermal titration calorimetry (ITC) was initially employed to probe PIQ binding to the novel hybrid-2R G-quadruplex with its stacked hairpin lateral loop. To eliminate any putative impact of a bromo substituent at the quadruplex-duplex interface, the *L11AT-B23* analogue was used as receptor for the ligand. Also, a 20 mM potassium phosphate buffer supplemented with 100 mM KCl was used for a better compatibility with previous ITC experiments after verification by NMR spectroscopy that the topology is conserved under the higher salt concentrations (not shown). As an additional benefit of a higher ionic strength, non-specific electrostatic interactions between the cationic ligand and the polyanionic quadruplex may efficiently be suppressed.^[17]

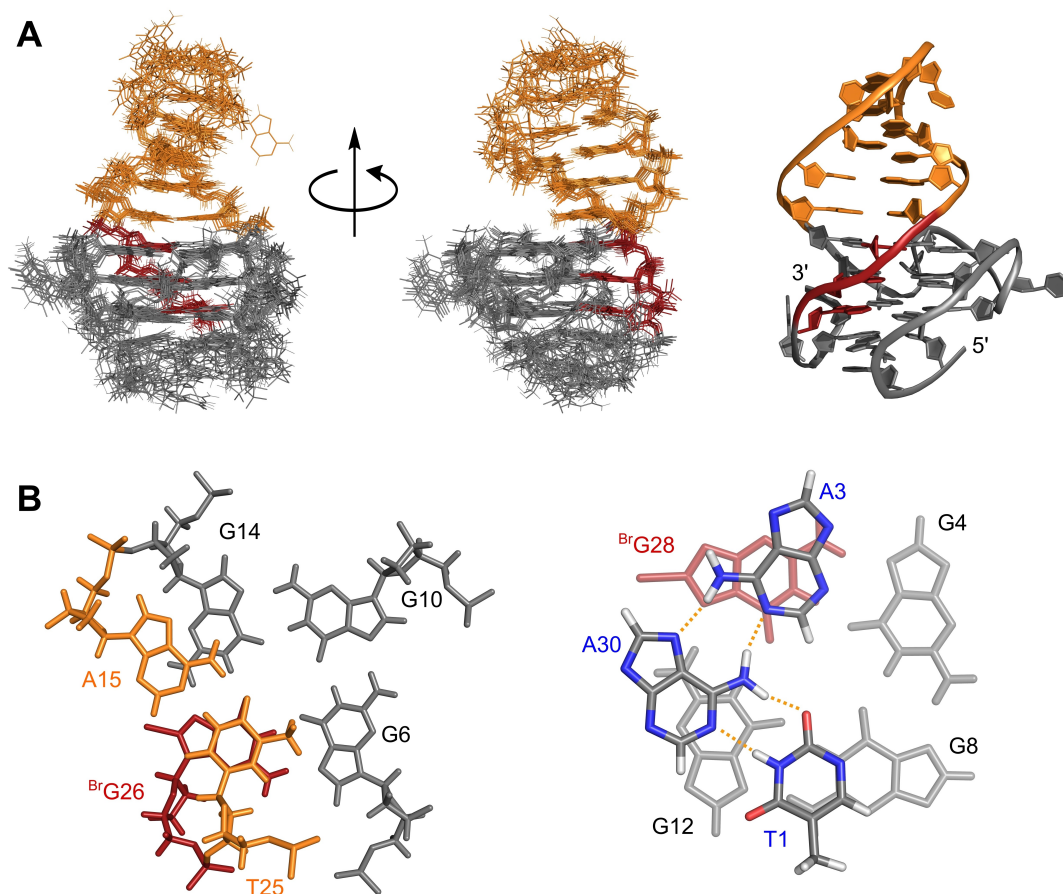


Figure 8. (A) Superposition of 10 lowest-energy structures (left) and representative structure of *L11AT-B13* (right); duplex stem-loop is colored orange, syn-residues of the G-core are colored red. (B) Q–D junction (left) and possible capping structure stacked onto the 5'-tetrad (right); duplex and G-core residues are colored in orange and red/grey, respectively; 5'- and 3'- overhang residues as part of a planar capping structure observed in the lowest energy structure are colored by element with potential hydrogen bonds traced by dotted lines.

A representative ITC isotherm obtained after integration of the power output for each injection and corrections for the heats of dilution is shown in Figure 9A (bottom) for the PIQ titration to the *L11AT-B23* receptor at 40 °C. Strong exothermic binding during initial titration steps followed by only a gradual return to baseline with ligand added in large excess points to at least two non-equivalent binding sites of very different affinity. Fitting the data of three independent experiments with a two-site model yielded an association constant K_a for high-affinity binding of $(4.4 \pm 3.1) \times 10^6 \text{ M}^{-1}$ and a binding molar enthalpy ΔH° of $-8.4 \pm 0.7 \text{ kcal mol}^{-1}$. In contrast, second binding with K_a of $\sim 3 \times 10^4 \text{ M}^{-1}$ is lower by two orders of magnitude and likely suggests PIQ binding to the duplex stem-loop.

Recently, an association constant K_a of $2 \times 10^6 \text{ M}^{-1}$ and a binding enthalpy ΔH° of $-6.3 \text{ kcal mol}^{-1}$ was determined under identical conditions for PIQ binding to a *c-myc*-derived quadruplex with the same flanking residues as in *L11AT-B23*.^[17] Lacking a duplex stem-loop, the latter G4 is forced into a parallel topology through its short 1- and 2-nt loops. Notably, extending one overhang sequence of the *c-myc* G4 to form a regular Watson-Crick duplex stacked on an outer tetrad yielded a stronger and also a more exothermic association, attributable to a favored PIQ

binding at the newly formed Q–D junction.^[16] Based on these results and on the noticeable increase in affinity and exothermicity when going from the non-extended *c-myc* to the hybrid-type *L11AT-B23* quadruplex, high-affinity binding to the latter is proposed to occur at the Q–D junction of *L11AT-B23* as being the most favorable binding site.

Because such a binding preference should promote refolding from a parallel into the hybrid topology in the presence of the ligand, additional NMR titrations of the parallel non-modified *L11AT* quadruplex with PIQ were performed (Figure 9B). With addition of ligand in sub-stoichiometric amounts, a second set of imino resonances emerges. Unfortunately, all G4 imino signals experience significant line-broadening due to dynamic exchange processes upon continued ligand titration. As a result, the newly formed complex resonances could not be analyzed and assigned to a particular topology. On the other hand, a single most downfield shifted AT Watson-Crick imino resonance is expected at the Q–D junction for a hybrid-2R topology. Disregarding severe broadening effects, the absence of a corresponding AT imino seems to argue against a ligand-induced refolding. As a consequence, the gain in binding free energy for the PIQ ligand seems to be insufficient to close the energy gap between the interconverting topologies.

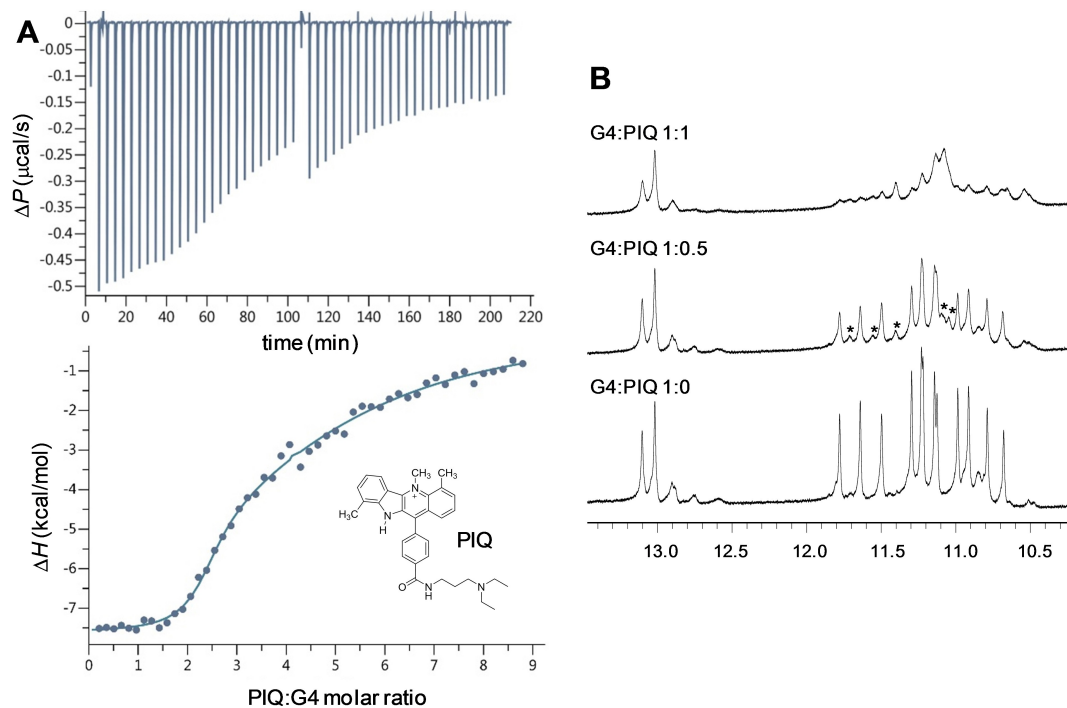


Figure 9. (A) Representative ITC experiment with the PIQ ligand titrated to the *L11AT-B23* quadruplex at 40 °C in 100 mM KCl, 20 mM potassium phosphate buffer, pH 7; the upper panel shows the heat burst for every injection step with a central discontinuity due to reloading the injection syringe for covering a wider concentration range; the lower panel shows the dilution-corrected heat versus the molar ratio with a least square fit. (B) Imino proton spectral region of the *L11AT* quadruplex titrated with the PIQ ligand at 30 °C in 10 mM potassium phosphate buffer, pH 7; some newly observed resonances after addition of 0.5 equivalent of ligand are indicated by an asterisk.

Discussion

Previous studies have succeeded in switching the first propeller loop of a parallel quadruplex to a lateral loop with the formation of a (+l+p+p) topology termed hybrid-1R. To access a novel (-p-p-l) hybrid-type G4 topology, the present work aimed at a corresponding rearrangement of the third loop to invert the last G-tract of a parallel G4. Although a flip of either the 5'- or 3'-terminal G-tract seems largely similar, it imposes different steric constraints. Whereas a 5'-3' backbone inversion of the first G-column in a parallel quadruplex will result in a lateral loop bridging a narrow groove and two overhang sequences facing each other in a wide groove, inversion of the fourth G-column will be accompanied by a lateral loop spanning a wide groove and 5'- and 3'-termini facing each other in a narrow groove. As a result, a 2-nt intervening sequence between the third and fourth G-tract was found to be too short to allow formation of a lateral loop for the present sequences, although 2-nt lateral loops bridging a wide groove have been reported.^[18,19] Here, a 3-nt loop was shown to be favored for the folding into a (-p-p-l) topology whereas an even longer and less structured 4-nt loop proved to be less efficient.

Because the width of a B-DNA minor groove closely matches the width of a quadruplex wide groove, a lateral Watson-Crick stem-loop is expected to smoothly extend coaxially from the quadruplex core with continuous base stacking.^[12] In contrast, short distances will prevent overhang bases next to the G-core from forming a regular Watson-Crick base pair. In line with these

considerations, a more efficient strand flip was attempted by replacing a short lateral loop with a putative stem-loop forming sequence in the present studies. In fact, a mild promotion of refolding into a lateral loop through coaxial stem-loop formation could be observed through a moderate increase in the population of a rearranged hybrid structure in case of *L11AT-B12* when compared to *L113-B12*. However, whereas base pairing between the 5'- and 3'-overhang were previously found to induce a partial flip of the first G-column even without inserting *syn*-favoring G analogues to form a hybrid-1R topology,^[9] base pairing within a lateral loop as in *L11AT* and *L11GC* failed to enforce refolding into a corresponding hybrid-2R fold without support from additional ^{Br}G modifications. The rather small impact of a stem-loop on refolding may be attributed to only partial disruptions of base pairs in a parallel G4. Thus, a propeller loop connecting two faces of the G-core can also accommodate a B-type hairpin, yet an orthogonal orientation between quadruplex core and a duplex stem-loop in such a case excludes base stacking at the interface and also prevents bases proximal to the junction from forming a Watson-Crick base pair.

Having enforced a complete refolding by the incorporation of two ^{Br}G modifications into the fourth G-tract, formed hybrid structures with a 4-bp lateral stem-loop exhibit unaltered thermal stabilities when compared to corresponding hybrid quadruplexes with only a 3-nt lateral loop (Table 1). This is in contrast to previous observations on a two-tetrad antiparallel quadruplex, showing a moderate increase in melting temperature upon

additional stacking of a lateral duplex hairpin onto an outer G-tetrad.^[20] Apparently, there is no significant stability gain for the hybrid-type quadruplex from stacking interactions of an additional base pair with its outer tetrad. On the other hand, the 11-nt hairpin-type propeller loop linked by two unpaired bases to opposite faces of the G-core in *L11AT* and *L11GC* seems to compromise G4 stability. This is suggested by the lowering of melting temperatures when going from *L113* to *L11AT* and *L11GC*, with all three non-modified sequences exclusively folding into a parallel topology. Thus, a favored G-tract inversion in the presence of a base-paired loop as implied by a comparison of *L113-B12* and *L11AT-B12* may rely on a destabilization through the long propeller loop in a parallel G4 rather than on additional G4 stabilization through the stacking of a lateral stem-loop. Clearly, the latter implies that relative melting temperatures also reflect relative thermodynamic stabilities at ambient temperatures. Alternatively, formation of different topologies may also be kinetically controlled and a remarkable increase in folding kinetics has recently been reported for a parallel topology in the presence of a fast forming duplex stem-loop, thought to bring G-tracts closer to each other.^[21] Whereas in the latter studies the double-helical domain was linked by non-complementary unpaired bases to the G-core, initial formation of a hairpin structure with a terminal base pair directly adjoining G-tracts may conceivably promote fast folding into a hybrid structure through its duplex stem-loop mediated preorganization.

Structural analysis of the formed (-p-p-l) hybrid-2R quadruplex demonstrates a fourth G-column exclusively comprising all-*syn* residues to yield four G(*anti*)-G(*anti*)-G(*anti*)-G(*syn*) tetrads of the same polarity. Given more favorable stacking interactions in *syn-anti* and *anti-anti* steps when compared to *syn-syn* and *anti-syn* steps along a G-column,^[22,23] a hybrid-2R' conformation with a G(*syn*)-G(*syn*)-G(*anti*) glycosidic bond pattern of the fourth G-tract associated with a single G(*anti*)-G(*syn*)-G(*syn*)-G(*syn*) tetrad and homopolar as well as heteropolar tetrad stackings would be easily conceivable (Figure 1). This should specifically apply to *L11AT-B12* that comprises two 5'-positioned *syn*-affine ^{Br}G analogues and a natural G residue at the G-tract 3'-position. Although such a hybrid-2R' conformation for a sequence with substitutions at the first two positions in the last G-tract is potentially disfavored by a larger number of non-brominated *syn*-G residues (3 versus 1, see Figure 1), it would harbor only a single unfavorable *syn-syn* step. In fact, the presence of a *L11AT-B12* minor species points to its folding into a coexisting parallel topology, but refolding into an additional hybrid-type conformer with a fourth G(*syn*)-G(*syn*)-G(*anti*) column cannot be excluded. Thus, observation of additional low-intensity AT Watson-Crick imino resonances of *L11AT-B12* are compatible with formation of a hybrid-2R' conformer (see Figure 5). Unfortunately, the low population prevents its unambiguous characterization through CD and NMR spectral analysis. Of note, previously reported hybrid-1 and hybrid-2 topologies, each forming one propeller and two lateral loops, fold into quadruplexes with a glycosidic bond angle pattern featuring one antiparallel G(*syn*)-G(*syn*)-G(*anti*) and three G(*syn*)-G(*anti*)-G(*anti*) columns.^[24–26] Also, in flipping the first G-column of a non-modified parallel quadruplex through complementary overhang sequences to form a (+l+p+p) hybrid-1R topology, quadruplexes with a

first all-*syn* tract were also shown to coexist with a corresponding hybrid-1R' conformation exhibiting a first G(*syn*)-G(*syn*)-G(*anti*) column, albeit with a lower population.

The newly characterized hybrid-2R fold expands the topological toolbox for various G4 technological applications. On the other hand, the present studies suggest that, although sterically feasible, a (-p-p-l) quadruplex topology seems disfavored in sequences lacking nucleotide analogues with specific conformational preferences. However, with an intervening sequence able to form a duplex hairpin when linking the third and fourth G-tract, ligands selectively binding with high affinity to a quadruplex-duplex junction as formed in a hybrid-2R topology may trigger a corresponding rearrangement by shifting equilibria. A ligand-induced refolding of *L11AT* from a parallel into a hybrid-2R topology upon addition of the indoloquinoline PIQ, previously found to favor binding at a quadruplex-duplex junction, remains questionable but highly negative binding free energies of even more specific Q–D binders may overcome energy barriers to effectively drive G4 refolding. As a consequence, hybrid-2R topologies may also occur under *in vivo* conditions in the presence of binding metabolites or proteins without additional modifications.

Experimental Section

Materials and sample preparation

All oligonucleotides were synthesized by TIB MOLBIOL (Berlin, Germany) and additionally purified by precipitation with potassium acetate and ethanol. Concentration of DNA was determined in triplicate by the UV absorbance of its unfolded species at 80 °C, employing molar extinction coefficients as provided by the supplier. For optical and NMR measurements, samples were dissolved in 10 mM potassium phosphate buffer, pH 7. Concentrations of final samples were 5 μM for UV and CD measurements and ranged from 0.2 to 1 mM for NMR studies. Prior to measurements, samples were annealed by heating to 90 °C for 5 min followed by their slow cooling to room temperature. PIQ was prepared as described previously and its concentration determined spectrophotometrically by using a molar extinction coefficient $\epsilon_{376} = 22227 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$.^[27]

UV melting

Melting temperatures for the quadruplexes were determined with a Jasco V-650 spectrophotometer equipped with a Peltier thermostat. Temperature dependent absorbances were recorded with a heating and cooling rate of 0.2 °C per minute at 295 nm. There was no hysteresis, indicating thermodynamic equilibrium conditions. Melting temperatures were determined by the minimum of a first derivative plot. In case of a poor signal-to-noise ratio, the derivative plot was smoothed using the moving window average method. Final melting temperatures are given as averages from the cooling curves of three independent measurements.

CD spectroscopy

CD experiments were performed at 20 °C with a JASCO J-810 spectropolarimeter equipped with a Peltier thermostat. Following stirring of the sample for 2 min to avoid concentration gradients in

the cuvette, spectra were recorded from 210 to 350 nm with a bandwidth of 1 nm, a scanning speed of 50 nm/min and a response time of 4 s using a 1-cm quartz cuvette. Five accumulations were recorded for each sample.

NMR spectroscopy

NMR measurements were performed with a Bruker Avance Neo 600 MHz NMR spectrometer equipped with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{19}\text{F}$ quadruple resonance cryo-probehead and z-field gradients. Unless stated otherwise, experiments were performed at 30 °C. Data were processed in TopSpin 4.0.7 and analyzed using CcpNmr 2.4.2 software.^[28,29] For experimental details see the Supporting Information.

Molecular dynamics simulation

Starting structures calculated with unmodified guanosines replacing 8-bromo-guanosine analogues were generated by simulated annealing in XPLOR-NIH 3.0.3.^[30,31] Partial atomic charges for the modified 8-bromoguanosine residue were calculated using the RED software with DFT approach.^[32] Refinement was performed using AMBER16 with the parmbsc force field and OL15 modifications.^[33] Adding potassium ions in explicit water, ten lowest-energy structures were finally obtained after equilibration for 4 ns and short minimization *in vacuo*. RMSD calculations, alignments, and analyses of the calculated structures were performed with the VMD-1.9.3 software.^[34] Three-dimensional structural representations were prepared with Pymol-1.8.4 [58].^[35] For details of the calculations see the Supporting Information.

Isothermal titration calorimetry

ITC experiments were performed with a Microcal PEAQ ITC micro-calorimeter (Malvern Instruments, United Kingdom) employing a reference power of 4 $\mu\text{cal s}^{-1}$. Oligonucleotides and the PIQ ligand were each dissolved in 100 mM KCl, 20 mM potassium phosphate buffer, pH 7.0, supplemented with 5% DMSO. The PIQ solution (400 μM) was titrated to 20 μM of oligonucleotide with a total of 2 × 26 injections of 1.5 μL each, an injection duration of 3 s, and a spacing between injections of 240 s. The first injection (0.4 μL) was rejected during the fitting process. All experiments were blank- and concentration-corrected. For data analysis, the MicroCal PEAQ-ITC analysis software was used.

Accession Codes

Atomic coordinates and lists of chemical shifts for *L11AT-B13* have been deposited (PDB ID 7O1H, BMRB ID 34615).

Acknowledgements

This research was supported by the Deutsche Forschungsgemeinschaft (grant no. WE 1933/15-1). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: DNA · G-quadruplexes · NMR spectroscopy · quadruplex-duplex junction · topology

- [1] R. Hänsel-Hertsch, M. Di Antonio, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 279–284.
- [2] S. Neidle, *Nat. Chem. Rev.* **2017**, *1*, 0041.
- [3] J. L. Neo, K. Kamaladasan, M. Uttamchandani, *Curr. Pharm. Des.* **2012**, *18*, 2048–2057.
- [4] A.-M. Chiorcea-Paquim, R. Eritja, A. M. Oliveira-Brett, *J. Nucleic Acids* **2018**, *2018*, 5307106.
- [5] M. Webba da Silva, *Chem. Eur. J.* **2007**, *13*, 9738–9745.
- [6] A. I. Karsisiotis, C. O’Kane, M. Webba da Silva, *Methods* **2013**, *64*, 28–35.
- [7] F. Fogolari, H. Haridas, A. Corazza, P. Viglino, D. Corà, M. Caselle, G. Esposito, L. E. Xodo, *BMC Struct. Biol.* **2009**, *9*, 64.
- [8] S. A. Dvorkin, A. I. Karsisiotis, M. Webba da Silva, *Sci. Adv.* **2018**, *4*, eaat3007.
- [9] B. Karg, S. Mohr, K. Weisz, *Angew. Chem. Int. Ed.* **2019**, *58*, 11068–11071; *Angew. Chem.* **2019**, *131*, 11184–11188.
- [10] L. Haase, B. Karg, K. Weisz, *ChemBioChem* **2019**, *20*, 985–993.
- [11] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [12] K. W. Lim, A. T. Phan, *Angew. Chem. Int. Ed.* **2013**, *52*, 8566–8569; *Angew. Chem.* **2013**, *125*, 8728–8731.
- [13] J. Brčić, J. Plavec, *Nucleic Acids Res.* **2018**, *46*, 11605–11617.
- [14] S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Eur. J.* **2018**, *24*, 4428–4435.
- [15] L. Diaz-Casado, I. Serrano-Chacón, L. Montalvillo-Jiménez, F. Corzana, A. Bastida, A. G. Santana, C. González, J. L. Asensio, *Chem. Eur. J.* **2021**, *27*, 6204–6212.
- [16] Y. M. Vianney, P. Preckwinkel, S. Mohr, K. Weisz, *Chem. Eur. J.* **2020**, *26*, 16910–16922.
- [17] A. Funke, K. Weisz, *J. Phys. Chem. B* **2017**, *121*, 5735–5743.
- [18] X. Cang, J. Šponer, T. E. Cheatham III, *J. Am. Chem. Soc.* **2011**, *133*, 14270–14279.
- [19] J. Brčić, J. Plavec, *Nucleic Acids Res.* **2015**, *43*, 8590–8600.
- [20] K. W. Lim, Z. J. Khong, A. T. Phan, *Biochemistry* **2014**, *53*, 247–257.
- [21] T. Q. N. Nguyen, K. W. Lim, A. T. Phan, *J. Phys. Chem. B* **2020**, *124*, 5122–5130.
- [22] X. Cang, J. Šponer, T. E. Cheatham III, *Nucleic Acids Res.* **2011**, *39*, 4499–4512.
- [23] J. Šponer, A. Mládek, N. Špačková, X. Cang, T. E. Cheatham III, S. Grimme, *J. Am. Chem. Soc.* **2013**, *135*, 9785–9796.
- [24] A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
- [25] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
- [26] A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
- [27] A. Funke, J. Dickerhoff, K. Weisz, *Chem. Eur. J.* **2016**, *22*, 3170–3181.
- [28] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins Struct. Funct. Bioinf.* **2005**, *59*, 687–696.
- [29] R. H. Fogh, W. Boucher, W. F. Vranken, A. Pajon, T. J. Stevens, T. N. Bhat, J. Westbrook, J. M. C. Ionides, E. D. Laue, *Bioinformatics* **2005**, *21*, 1678–1684.
- [30] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.
- [31] C. D. Schwieters, J. J. Kuszewski, G. M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
- [32] E. Vanqualef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delapine, P. Cieplak, F. Y. Dupradeau, *Nucleic Acids Res.* **2011**, *39*, W511–W517.
- [33] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.
- [34] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
- [35] The PyMOL Molecular Graphics System Version 1.8 Schrödinger LLC, **2015**.

Manuscript received: April 1, 2021

Accepted manuscript online: May 6, 2021

Version of record online: June 4, 2021

Chemistry–A European Journal

Supporting Information

Expanding the Topological Landscape by a G-Column Flip of a Parallel G-Quadruplex

Swantje Mohr, Jagannath Jana, Yoanes Maria Vianney, and Klaus Weisz*

METHODS

NMR spectroscopy. Experiments were generally performed at 303 K with proton chemical shifts referenced relative to the temperature-dependent water chemical shift at pH 7. Carbon chemical shifts were referenced to DSS using the indirect referencing method. For one-dimensional and two-dimensional NOESY spectra, an optimized WATERGATE with w5 element was used for solvent suppression. For HSQC experiments, the 3-9-19 solvent suppression scheme was employed for samples in 90% H₂O / 10% D₂O, using a spectral width of 6.7 kHz in the ¹³C dimension. NOESY spectra were acquired with mixing times of 80, 150, or 300 ms in 90% H₂O / 10% D₂O. DQF-COSY spectra were recorded in 100% D₂O with solvent suppression through presaturation.

Molecular dynamics simulation. Initially, 100 starting structures of lowest energy were selected out of 400 structures generated by simulated annealing in XPLOR-NIH 3.0.3.^[1,2] Distance restraints were set according to cross-peak intensities in 2D NOESY spectra. For exchangeable protons, peaks were categorized into very strong (2.9 ± 1.1 Å), strong (4.0 ± 1.2 Å), medium (5.0 ± 1.2 Å), and weak (6.0 ± 1.2 Å). For non-exchangeable protons, distances were set to 2.9 ± 1.1 Å for strong, 4.0 ± 1.5 Å for medium, 5.5 ± 1.5 Å for weak, 6.0 ± 1.5 Å for very weak, and 5.0 ± 2.0 Å for overlapping cross-peaks. Distance restraints for H-bonds were included for the tetrads and Watson-Crick base pairs. Glycosidic torsion angle restraints were set to *anti* (170-310°) for all residues except for Gs in the fourth G-tract which were restrained to *syn* (25-95°). Additional planarity restraints were only employed for the tetrads. Starting structures were calculated with unmodified guanosines replacing 8-bromo-guanosine analogs.

Partial atomic charges for the modified 8-bromoguanosine residue were calculated using the RED software with DFT approach.^[3] Refinement was performed using AMBER16 with the parmbsc force field and OL15 modifications.^[4] The 100 starting structures were subjected to simulated annealing *in vacuo* to yield 20 converged structures. Here, the same restraints were used as before with restraint energies set to 40 kcal·mol⁻¹·Å⁻² for NOE distance restraints, 50 kcal·mol⁻¹·Å⁻² for hydrogen bond distance restraints, 200 kcal·mol⁻¹·rad⁻² for dihedral angle restraints, and 30 kcal·mol⁻¹·Å⁻² for planarity restraints in tetrads and base pairs. The system was equilibrated at 300 K for 5 ps, followed by heating to 1000 K for 10 ps. This temperature was maintained for 30 ps, then the system was cooled to 100 K within 45 ps and to 0 K in 10 ps. Ten lowest-energy conformations were selected for further refinement in water.

Next, potassium ions were added to neutralize the system. Two of them were placed in the inner channel of the quadruplex core between tetrad layers. After placing each structure in a 10 Å octahedral box, TIP3P water was added. The simulation began with 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization. Here, the DNA was fixed with a force constant of 25 kcal·mol⁻¹·Å⁻². In the following, the system was heated from 100 to 300 K during 10 ps in a NVT ensemble, continuing with a NPT ensemble equilibration at 1 atm with energy restraints decreasing from 5 to 0.5 kcal·mol⁻¹·Å⁻². Finally, a simulation for 4 ns at 1 atm and 300 K was calculated. The last 500 ps of the trajectory were averaged and shortly minimized *in vacuo*, resulting in 10 lowest-energy structures.

References

- [1] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.
- [2] C. D. Schwieters, J. J. Kuszewski, G. M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
- [3] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, F. Y. Dupradeau, *Nucleic Acids Res.* **2011**, *39*, W511–W517.
- [4] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.

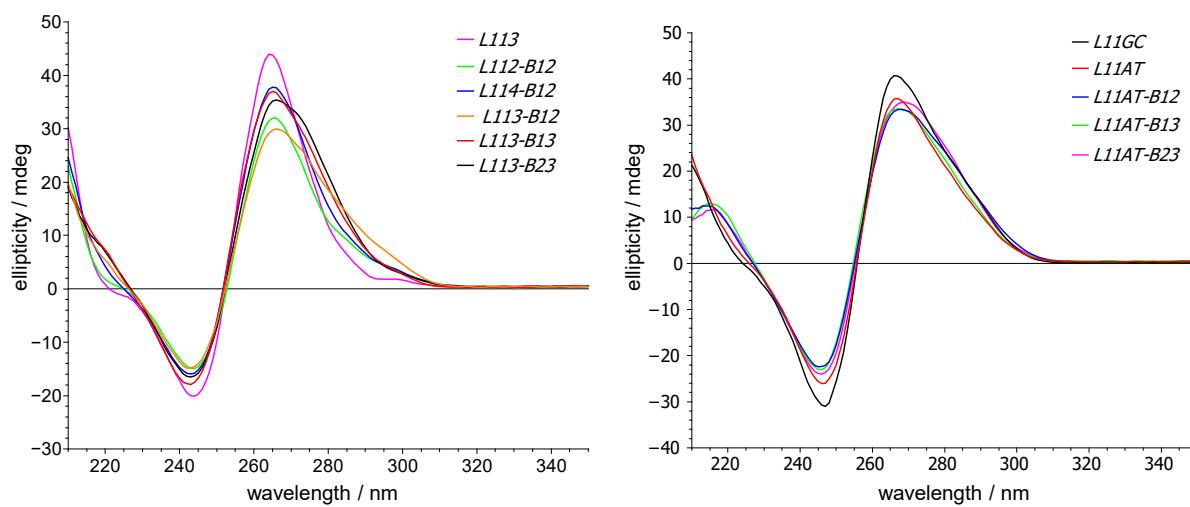


Figure S1. CD spectra of quadruplex-forming sequences without (left) and with a putative stem-loop structure (right) in 10 mM potassium phosphate buffer, pH 7.

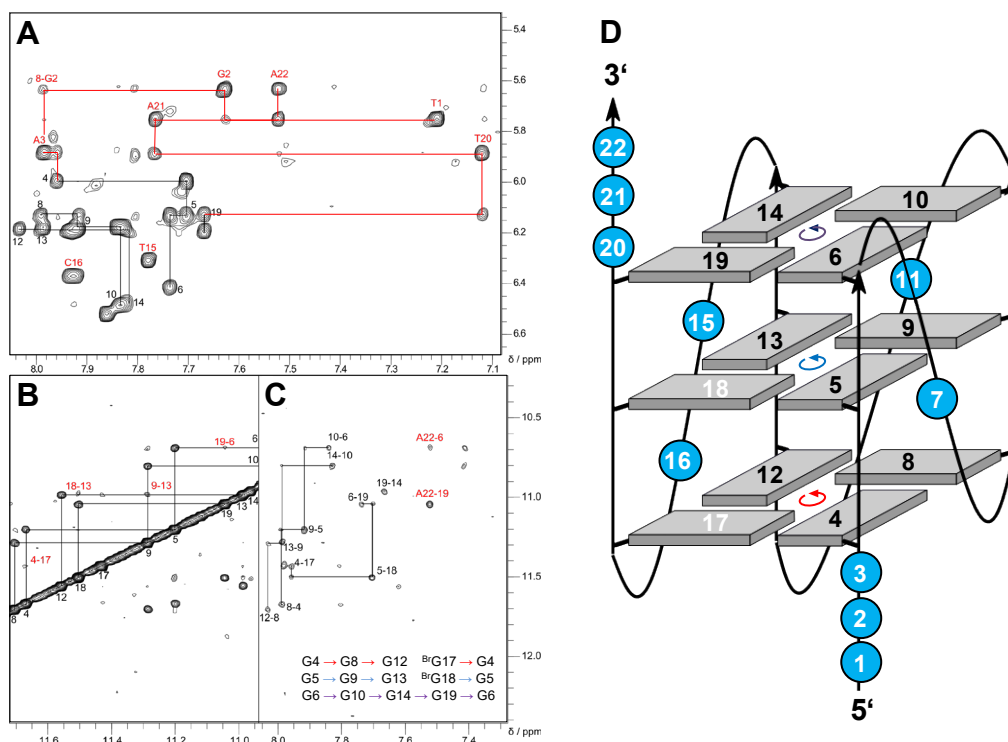


Figure S2. 2D NOE spectral regions of *L112-B12*. (A) H8/6(ω_2)-H1'(ω_1) cross-peaks with sequential contacts traced along the G-core (black) as well as along overhang residues (red). (B) Sequential H1-H1 cross-peaks and additional intra-tetrad contacts labeled in red. (C) Intra- and inter-tetrad H8(ω_2)-H1(ω_1) cross-peaks and additional contacts involving overhang residue A22 labeled in red. Tetrad polarities as determined from intra-tetrad NOE contacts are summarized at the bottom. (D) Topology of the *L112-B12* quadruplex with residue numbers and tetrad polarities indicated; ^{Br}G analogs are labeled with white numbers.

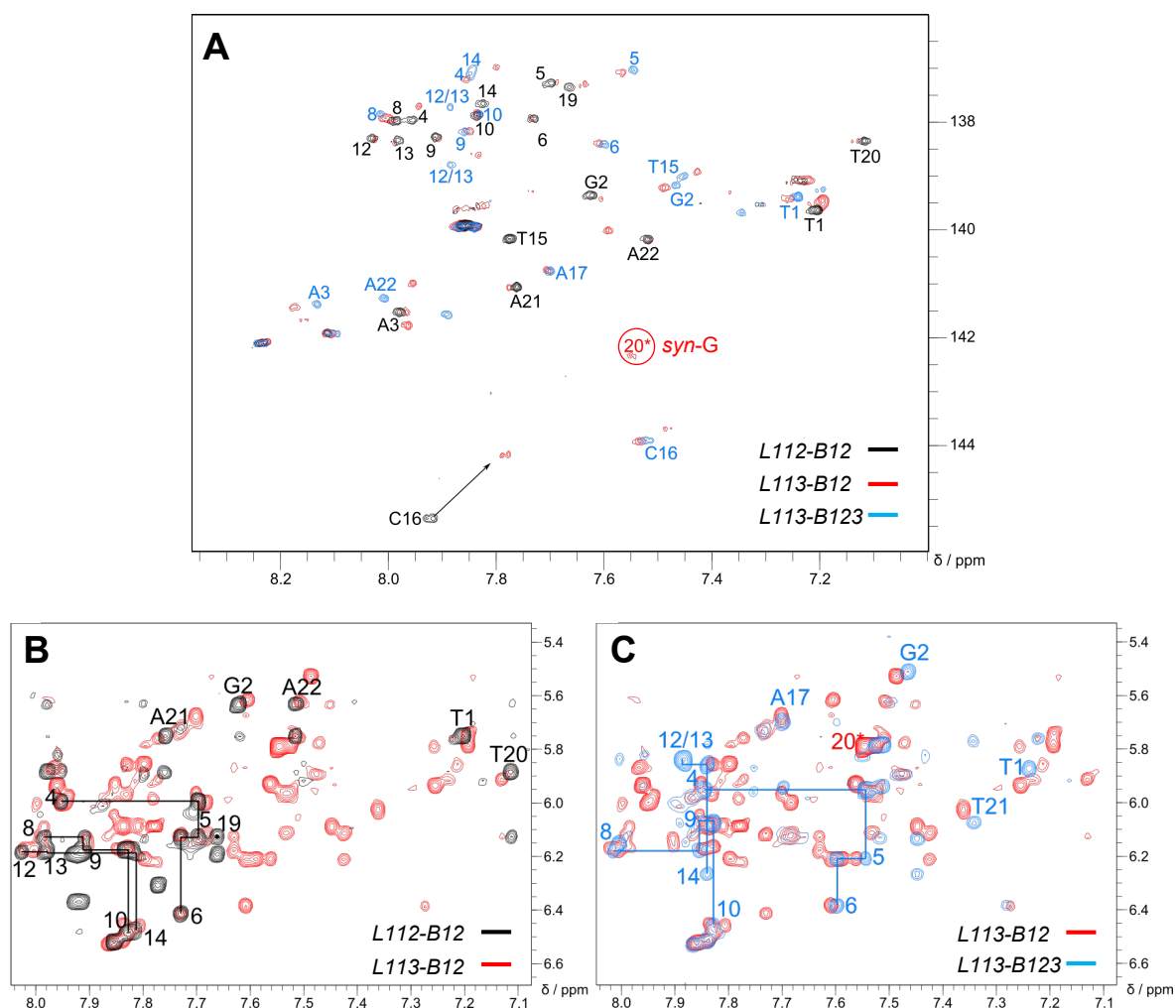


Figure S3. Spectral comparison through superimposed two-dimensional spectra of *L112-B12*, *L113-B12*, and *L113-B123*. (A) H6/8(ω_2)-C6/8(ω_1) correlations in ¹H-¹³C HSQC spectra with corresponding assignments; correlations for *L113-B12* (red) match correlations of both *L112-B12* (black) and *L113-B123* (blue); a single *syn-G* (circled) can be assigned to G20 of the *L113-B12* major species. (B) H8/6(ω_2)-H1'(ω_1) cross-peaks of *L113-B12* (red) and *L112-B12* (black); sequential contacts are traced along the G-columns for *L112-B12*. (C) H8/6(ω_2)-H1'(ω_1) cross-peaks of *L113-B12* (red) and *L113-B123* (blue); sequential contacts are traced along the G-columns for *L113-B123*.

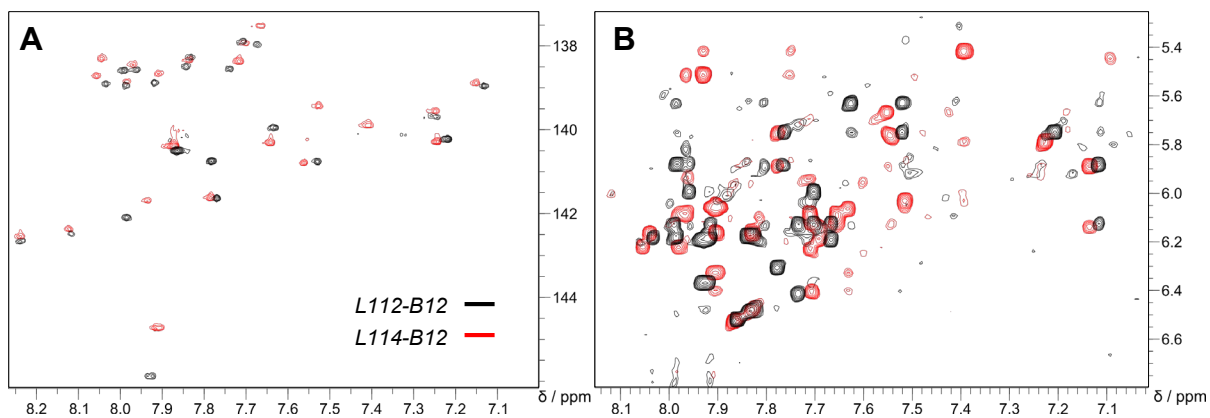


Figure S4. Spectral comparison through superimposed two-dimensional spectra of *L114-B12* (red) and *L112-B12* (black). (A) H6/8(ω_2)-C6/8(ω_1) correlations in ^1H - ^{13}C HSQC spectra and (B) H8/6(ω_2)-H1'(ω_1) correlations in 2D NOESY spectra.

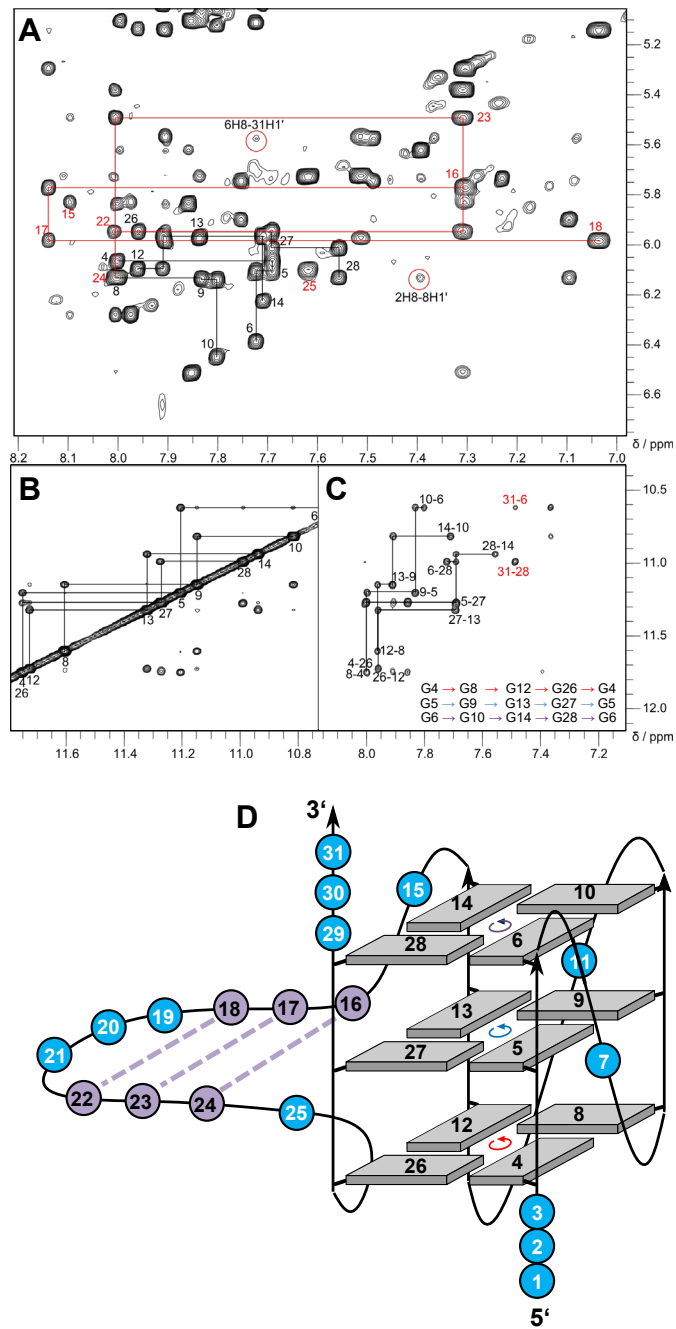


Figure S5. 2D NOE spectral regions of *LIIGC*. (A) H8/6(ω_2)-H1'(ω_1) cross-peaks with sequential contacts traced along the G-core (black) and the duplex stem (red); NOE contacts between overhang residues and outer G-tetrads are marked by red circles. (B) Sequential H1-H1 cross-peaks along the G-columns. (C) Intra- and inter-tetrad H8(ω_2)-H1(ω_1) cross-peaks and additional contacts involving overhang residue A31 labeled in red. Tetrad polarities as determined from intra-tetrad NOE contacts are summarized at the bottom. (D) Topology of the *LIIGC* quadruplex with residue numbers and tetrad polarities indicated; GC Watson-Crick base pairs within the 11-nt propeller loop are indicated by broken violet lines.

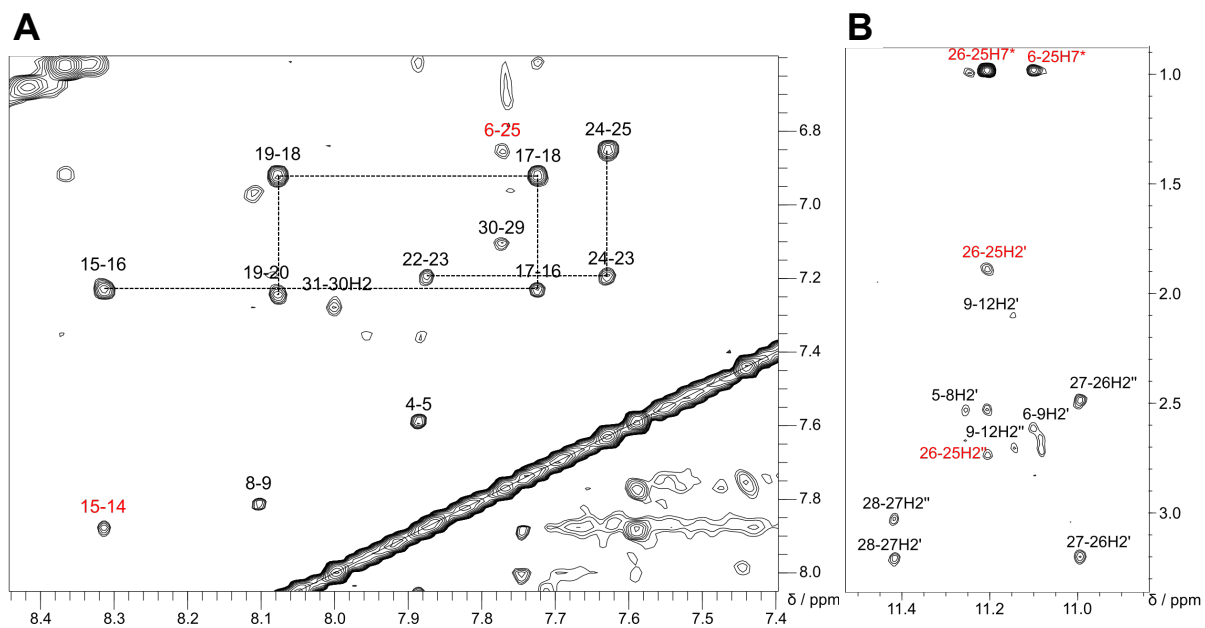


Figure S6. (A) H8/6-H8/6 and (B) H1(ω_2)-H2'/H2''(ω_1) 2D NOE spectral region of *L11AT-B13*; contacts at the quadruplex-duplex interface are labeled in red. Contacts along the duplex stem are traced by horizontal and vertical lines in (A).

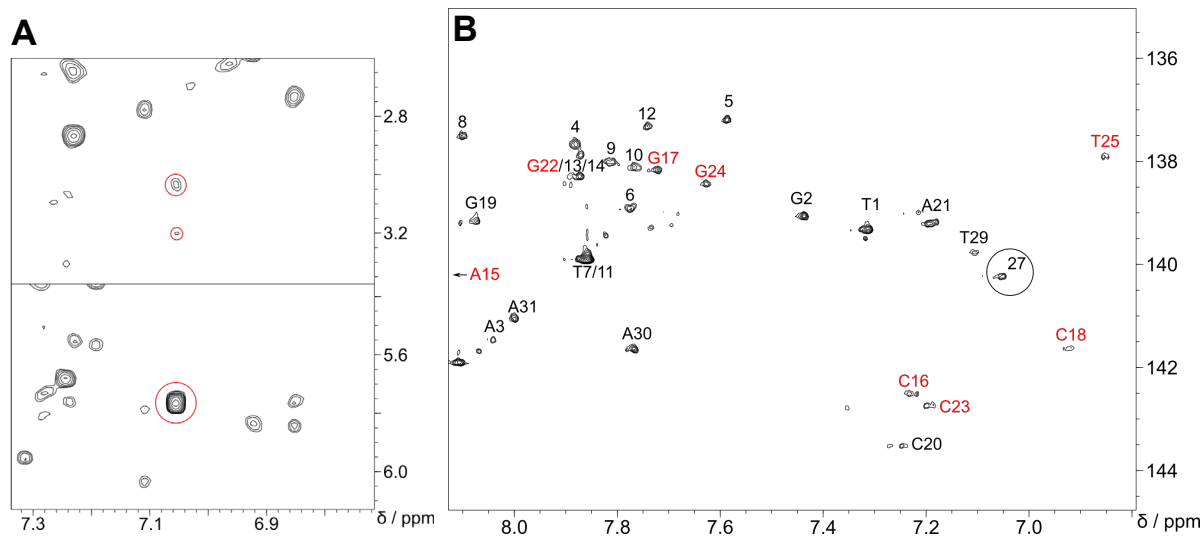


Figure S7. (A) 2D NOESY spectrum of *L11AT-B13*; H8/H6(ω_2)-H2'/H2''(ω_1) (top) and H8/H6(ω_2)-H1'(ω_1) NOE contacts (bottom) with weak and strong intra-residual cross-peaks marked for G27 (80 ms mixing time). (B) H6/8-C6/8 region of a ^1H - ^{13}C HSQC spectrum of *L11AT-B13*; cross-peaks from Watson-Crick base pairs are marked in red; *syn*-G27 is circled.

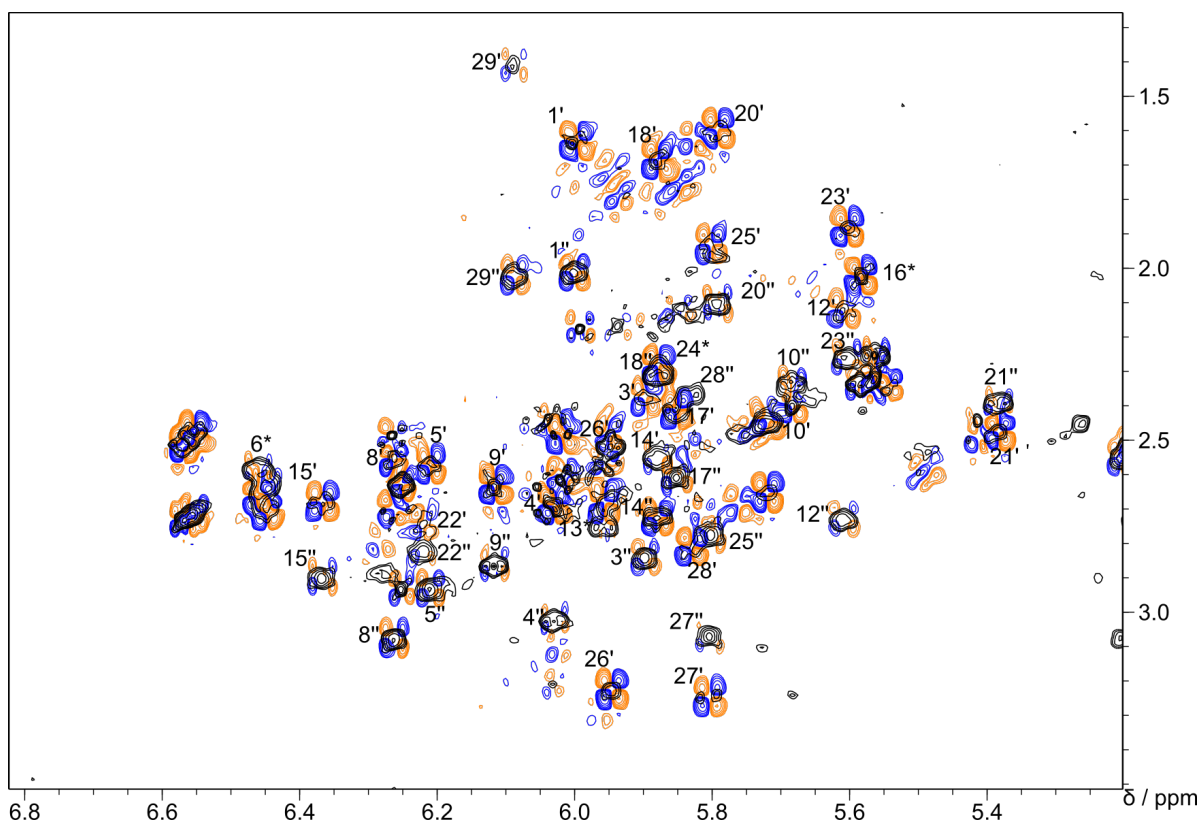


Figure S8. Superimposed H1'-H2'/H2'' spectral region of NOESY (80 ms mixing time, black) and DQF-COSY spectra (blue/orange) for *LIIAT-B13*; H1'-H2' and H1'-H2'' contacts are labeled with residue number; correlations of stereochemically ambiguous resonances are marked with an asterisk.

Table S1. ^1H and ^{13}C chemical shifts δ (ppm) of *L112-B12*.^a

residue	H8/H6	C8/C6	H1'	H2'/H2'' ^b	H1	H5/H2/Me
T1	7.21	139.73	5.75	2.04	n.d.	n.d.
G2	7.63	139.45	5.64	2.38 / 2.45	n.d.	-
A3	7.98	141.63	5.88	2.59	-	n.d.
G4	7.96	138.07	6.00	2.72 / 2.96	11.67	-
G5	7.70	137.41	6.13	n.d.	11.20	-
G6	7.73	138.03	6.42	2.61 / 2.75	10.69	-
T7	7.86	139.94	6.53	n.d.	n.d.	1.98
G8	7.99	138.08	6.13	2.46 / 2.94	11.71	-
G9	7.92	138.37	6.18	2.65 / 2.91	11.29	-
G10	7.84	137.99	6.49	2.64 / 2.77	10.80	-
T11	7.86	139.94	6.53	n.d.	n.d.	1.98
G12	8.03	138.41	6.19	2.46 / 2.97	11.56	-
G13	7.99	138.44	n.d.	n.d.	10.99	-
G14	7.83	137.77	n.d.	n.d.	10.96	-
T15	7.78	140.26	6.31	2.34 / 2.51	n.d.	1.97
C16	7.93	145.46	6.37	2.52 / 2.68	-	n.d.
^{Br} G17	-	-	n.d.	n.d.	11.43	-
^{Br} G18	-	-	6.20	2.63 / 3.14	11.50	-
G19	7.67	137.46	6.13	2.56 / 2.79	11.04	-
T20	7.12	138.45	5.89	1.90 / 2.37	n.d.	1.52
A21	7.76	141.17	5.75	2.00 / 2.46	-	n.d.
A22	7.52	140.28	5.63	2.19	-	n.d.

^a At 303 K in 10 mM potassium phosphate buffer, pH 7.

^b No stereochemical assignment for H2'/H2'' protons.

Table S2. ^1H and ^{13}C chemical shifts δ (ppm) of *L113-B123*.^a

residue	H8/H6	C8/C6	H1'	H2'/H2'' ^b	H1	H5/H2/Me
T1	7.24	139.41	5.88	n.d.	n.d.	n.d.
G2	7.47	139.17	5.51	n.d.	n.d.	-
A3	8.14	141.40	5.96	n.d.	.	7.76
G4	7.85	137.12	5.96	2.66 / 2.92	11.32	-
G5	7.55	137.03	6.21	2.54 / 2.89	10.94	-
G6	7.60	138.42	6.39	2.66	10.97	-
T7	7.87	139.96	6.52	2.46 / 2.70	n.d.	1.98
G8	8.02	137.87	6.18	2.45 / 3.01	11.53	-
G9	7.86	138.18	6.08	2.67 / 2.85	11.04	-
G10	7.84	137.86	6.46	2.57 / 2.67	11.15	-
T11	7.87	139.96	6.52	2.46 / 2.70	n.d.	1.98
G12	7.89	n.d.	5.84	n.d.	10.99	-
G13	7.89	n.d.	5.86	2.66	11.09	-
G14	7.84	137.04	6.27	2.65 / 2.78	11.16	-
T15	7.45	139.01	6.14	2.20 / 2.42	n.d.	1.71
C16	7.52	n.d.	5.94	1.98 / 2.28	-	5.78
A17	7.70	140.78	5.70	1.80 / 2.42	-	7.29
^{Br} G18	-	-	n.d.	n.d.	11.21	-
^{Br} G19	-	-	n.d.	n.d.	11.49	-
^{Br} G20	-	-	n.d.	n.d.	11.45	-
T21	7.34	141.89	6.08	n.d.	n.d.	1.88
A22	8.01	141.28	6.16	n.d.	-	n.d.
A23	n.d.	n.d.	n.d.	n.d.	-	7.44

^a At 303 K in 10 mM potassium phosphate buffer, pH 7.^b No stereochemical assignment for H2'/H2'' protons.

Table S3. ^1H and ^{13}C chemical shifts δ (ppm) of *LIIIGC*.^a

residue	H8/H6	C8/C6	H1'	H2'/H2'' ^b	H1	H5/H2/Me
T1	7.23	139.41	5.74	n.d.	n.d.	1.62
G2	7.39	139.74	5.62	2.33 / 2.83	n.d.	-
A3	7.86	n.d.	5.84	n.d.	-	n.d.
G4	8.00	n.d.	6.07	3.03	11.75	-
G5	7.69	137.40	6.10	2.88	11.21	-
G6	7.72	137.85	6.39	2.59 / 2.74	10.62	-
T7	7.85	139.91	6.51	2.45 / 2.67	n.d.	1.97
G8	8.00	137.75	6.13	2.44 / 2.93	11.60	-
G9	7.83	138.14	6.14	2.60 / 2.87	11.15	-
G10	7.80	137.78	6.45	2.61 / 2.75	10.82	-
T11	7.85	139.91	6.51	2.45 / 2.67	n.d.	1.97
G12	7.96	n.d.	6.09	n.d.	11.73	-
G13	7.91	n.d.	5.97	n.d.	11.32	-
G14	7.71	137.54	6.23	n.d.	10.94	-
G15	7.84	138.97	n.d.	n.d.	n.d.	-
C16	7.31	143.63	5.77	2.48 / 2.69	-	5.30
G17	8.14	139.31	5.98	2.38 / 2.70	n.d.	-
C18	7.04	141.77	5.99	n.d.	-	5.14
G19	7.91	n.d.	5.57	n.d.	n.d.	-
C20	7.51	143.38	5.97	n.d.	-	n.d.
A21	7.84	n.d.	n.d.	n.d.	-	n.d.
G22	8.01	n.d.	5.95	1.65 / 2.22	n.d.	-
C23	7.31	142.63	5.49	2.54 / 2.58	-	5.38
G24	8.01	138.76	6.13	n.d.	n.d.	-
C25	7.62	144.00	6.10	2.36 / 2.46	-	5.73
G26	7.96	n.d.	5.94	2.53 / 2.79	11.75	-
G27	7.69	137.40	6.02	2.62 / 2.68	11.28	-
G28	7.56	137.06	6.13	2.50 / 2.78	10.99	-
T29	7.10	138.25	5.90	n.d.	n.d.	1.43
A30	7.75	141.07	5.75	n.d.	-	7.08
A31	7.49	140.19	5.58	n.d.	-	7.37 / 153.9

^a At 303 K in 10 mM potassium phosphate buffer, pH 7.^b No stereochemical assignment for H2'/H2'' protons.

Table S4. ^1H and ^{13}C chemical shifts δ (ppm) of *LIIAT-B13*.^a

residue	H8/H6	C8/C6	H1'	H2'/H2'' ^b	H3'	H1	H5/H2/C2/Me
T1	7.31	139.34	5.96	1.60 / 1.99	4.39	n.d.	1.74
G2	7.44	139.07	5.53	2.24 / 2.30	4.67	n.d.	-
A3	8.04	141.47	5.87	2.36 / 2.81	4.73	-	7.74 / 154.9
G4	7.88	137.68	5.99	2.67 / 3.00	4.90	11.01	-
G5	7.59	137.20	6.18	2.54 / 2.91	4.80	11.26	-
G6	7.77	138.92	6.42	2.67*	4.65	11.10	-
T7	7.86	139.91	6.52	2.46 / 2.69	n.d.	n.d.	1.99
G8	8.10	137.50	6.23	2.53 / 3.05	5.17	11.52	-
G9	7.81	138.03	6.08	2.61 / 2.84	5.06	11.15	-
G10	7.77	138.11	5.66	2.37 / 2.33	4.65	11.09	-
T11	7.86	139.91	6.52	2.46 / 2.69	n.d.	n.d.	1.99
G12	7.74	137.34	5.59	2.11 / 2.71	5.12	11.05	-
G13	7.89	n.d.	5.93	2.66* / 2.72*	5.07	11.61	-
G14	7.88	n.d.	5.84	2.70 / 2.70	5.07	10.68	-
A15	8.31	141.25	6.33	2.65 / 2.87	5.11	-	7.76 / 155.1
C16	7.23	142.51	5.55	2.00* / 2.31*	4.84	-	5.22
G17	7.72	138.18	5.82	2.39 / 2.58	4.90	12.75	-
C18	6.92	141.65	5.84	1.66 / 2.28	4.74	-	5.03
G19	8.08	139.14	5.68	2.63 / 2.43	4.83	n.d.	-
C20	7.24	143.52	5.76	1.58 / 2.08	4.37	-	5.23
A21	7.19	139.20	5.35	2.45 / 2.37	n.d.	-	n.d.
G22	7.88	n.d.	6.18	2.72 / 2.79	4.77	12.81	-
C23	7.19	142.75	5.57	1.85 / 2.23	4.69	-	5.16
G24	7.63	138.45	5.84	2.25* / 2.52*	4.80	12.76	-
T25	6.85	137.92	5.76	1.90 / 2.74	4.73	13.45	1.00
^{Br} G26	-	-	5.91	3.21 / 2.50	n.d.	11.21	-
G27	7.05	140.25	5.77	3.21 / 3.04	4.88	11.00	-
^{Br} G28	-	-	5.79	2.78 / 2.34	4.89	11.42	-
T29	7.11	139.78	6.04	1.41 / 2.01	4.69	n.d.	1.85
A30	7.77	141.63	5.66	3.24 / 3.24	4.65	-	7.28 / 154.5
A31	8.00	141.06	6.20	2.60 / 2.89	4.60	-	n.d.

^a At 303 K in 10 mM potassium phosphate buffer, pH 7.^b Stereochemically ambiguous H2'/H2'' resonances are marked with an asterisk.

Table S5. NMR restraints and structural statistics for the structure calculations of *L11AT-B13*.

NMR restraints	
NOE-based distance restraints:	
intra-residual	97
inter-residual	153
exchangeable	62
other restraints:	
hydrogen bonds	70
dihedral angles	31
planarity	3
structural statistics	
pairwise heavy atom RMSD value (Å):	
all residues	2.5 ± 0.4
G-tetrad core	0.5 ± 0.1
duplex stem-loop	2.5 ± 0.9
NOE violations (Å):	
maximum violation	0.30
mean NOE violation	0.003 ± 0.001
deviations from idealized geometry:	
bonds (Å)	0.01 ± 0.0001
angles (degree)	2.25 ± 0.02

Article 5

Guiding the folding of G-quadruplexes through loop residue interactions

Jagannath Jana, Yoanes Maria Vianney, Nina Schröder and Klaus Weisz¹*

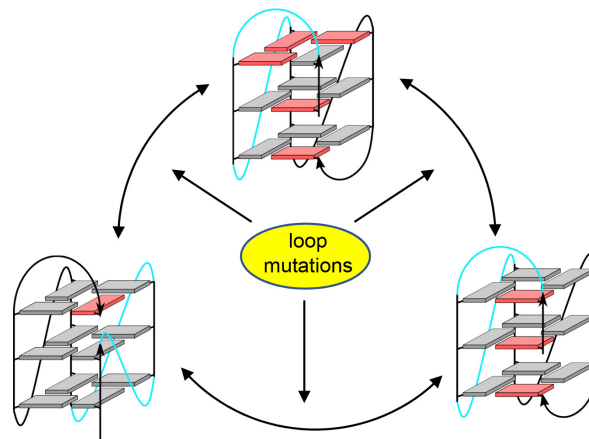
Institute of Biochemistry, Universität Greifswald, D-17489 Greifswald, Germany

Received April 14, 2022; Revised May 25, 2022; Editorial Decision June 08, 2022; Accepted June 14, 2022

ABSTRACT

A G-rich sequence was designed to allow folding into either a stable parallel or hybrid-type topology. With the parent sequence featuring coexisting species, various related sequences with single and double mutations and with a shortened central propeller loop affected the topological equilibrium. Two simple modifications, likewise introduced separately to all sequences, were employed to lock folds into one of the topologies without noticeable structural alterations. The unique combination of sequence mutations, high-resolution NMR structural information, and the thermodynamic stability for both topological competitors identified critical loop residue interactions. In contrast to first loop residues, which are mostly disordered and exposed to solvent in both propeller and lateral loops bridging a narrow groove, the last loop residue in a lateral three-nucleotide loop is engaged in stabilizing stacking interactions. The propensity of single-nucleotide loops to favor all-parallel topologies by enforcing a propeller-like conformation of an additional longer loop is shown to result from their preference in linking two outer tetrads of the same tetrad polarity. Taken together, the present studies contribute to a better structural and thermodynamic understanding of delicate loop interactions in genomic and artificially designed quadruplexes, e.g. when employed as therapeutics or in other biotechnological applications.

GRAPHICAL ABSTRACT



INTRODUCTION

G-Quadruplexes (G4s) are four-stranded structures that are formed by G-rich DNA or RNA sequences. These non-canonical nucleic acids have attracted considerable attention during the last two decades due to their biological role and their potential to serve as targets for medicinal interventions (1–3). In addition, tailored G4s of a particular fold are increasingly engineered, e.g. as aptamers in G4-based therapeutic and diagnostic applications (4). Up to now, a significant number of single G4s formed by truncated G-rich sequences of biological relevance has been determined at high resolution, giving insight into specific structural features and intrinsic interactions (5–8). This also includes non-canonical architectures such as G4s with a V-shaped loop, initially characterized based on a genomic sequence (9). On the other hand, a very recent study employed an iterative integrated structural biology approach to also derive the structures of longer wild-type promoter sequences mostly intractable for high-resolution structure determinations. Notably, results suggested that more complex higher-order G4 assemblies rather than monomeric G4s may be more common and relevant within the cellular environment (10).

For reliably predicting G4 structures within the genome and for a better understanding of folding pathways in the

*To whom correspondence should be addressed. Tel: +49 3834 420 4426; Fax: +49 3834 420 4427; Email: weisz@uni-greifswald.de

design of specific G-rich sequences of a desirable G4 topology, an in-depth understanding of interdependent interactions that determine the overall thermodynamic stability of the system is needed. When trying to derive more general rules that determine G4 folding, studies largely rely on designed sequences that fold into a single G4. Unlike natural sequences of putative physiological relevance, the latter impose fewer restrictions when introducing more extensive mutations to systematically probe sequence dependent effects. Flanking and especially intervening sequences that form different types of loops are critical determinants of major folds and the relationship of their length and composition with favored folding pathways has been scrutinized in the past at different levels (11–16). Based solely on geometric considerations, a particular type of loop may already be excluded by the length of the intervening sequence bridging two G-core positions at varying distances in an intramolecular G4. Thus, whereas a single nucleotide between two G-tracts generally forms a stable one-nucleotide (1-nt) propeller loop upon folding into a three-layered G4, lateral and diagonal loops bridging two adjacent and distal corners of an outer G-tetrad are expected to mostly require a loop length ≥ 2 and ≥ 4 nucleotides (17,18).

In addition to simple geometric restraints set by the minimum length of loops, the collection of data for a large number of canonical G4s has also contributed to several empirical rules relating the length and distribution of loops to the most favored G4 fold (19–21). Notably, G4s with a short central loop seem to favor a parallel topology whereas G4s with a long central loop have a higher propensity of folding into an antiparallel or hybrid-type structure (22). Also, irrespective of the length of a third loop the presence of two 1-nt loops was mostly found to impose a parallel fold (23). Such a propensity of single-nucleotide loops to also enforce propeller-type conformations of additional longer loops seems to be a general phenomenon but its molecular origin remains vague. Apparently, the formation of loops upon G4 folding are interdependent and loop interactions seem to be critical determinants of the overall topology of intramolecular structures. However, relevant forces that drive a specific folding are only poorly understood and their assessment requires both a detailed structural as well as thermodynamic analysis. Lacking a better understanding of loop length and residue dependent G4 folding, a more reliable prediction of favored G4 topologies adopted by a G-rich sequence is severely hampered as is the engineering of a given G4 fold.

Intervening sequences composed of three nucleotides may likewise form propeller and lateral loops depending on the overall sequence context. Such a variability manifests itself for the human telomeric quadruplex comprising three TTA loops. Depending on flanking sequences but also on outer conditions like type of cation in the buffer solution, crystal packing, and molecular crowding effects, several folds with different loop arrangements have been reported (24–26). Thus, being of particular interest, the present study focuses on the propensity of 3-nt intervening sequences to either fold into a favored propeller or lateral loop but also on the still elusive tendency of a central 1-nt propeller loop to promote folding into an all-parallel topology. The strategy is based on the design of a G-rich sequence restricted

to fold into two stable competitive G4 structures with a 3'-terminal snapback loop and on the ability to completely shift the topological equilibrium to either side by employing two different modifications. A combination of thermodynamic and NMR structural studies on various sequence mutants, also including the use of non-natural nucleoside analogs, was employed to unravel loop residue interactions that favor formation of a particular topology. Stabilization of a lateral loop through outer tetrad stacking of a loop 3'-purine but also the importance of steric effects associated with a 1-nt propeller loop bridging three tetrad layers was demonstrated. The present results provide a detailed structural and thermodynamic account for some empirical correlations reported in the recent past and enhances our understanding of G4 folding to support a better prediction but also a more rational design of G4 topologies.

MATERIALS AND METHODS

Materials and sample preparation

All DNA oligonucleotides were purchased from TIB MOL-BIOL (Berlin, Germany) and further purified by ethanol precipitation. Concentrations were determined spectrophotometrically by measuring absorbances at 260 nm at 80°C in salt-free H₂O with molar extinction coefficients as provided by the manufacturer based on a nearest-neighbour model (27,28). All DNA oligonucleotides were dissolved in a low-salt buffer of 10 mM potassium phosphate, pH 7. Samples were annealed with concentrations as used for the subsequent experiments by heating to 85°C for 5 min followed by slow cooling to room temperature and storage in a refrigerator overnight.

Circular dichroism (CD)

CD spectra were acquired with a Jasco J810 spectropolarimeter. Quadruplex sequences ($\sim 5 \mu\text{M}$) were measured in 1-cm quartz cuvettes at 20°C. Spectra were obtained by the accumulation of five scans at a speed of 50 nm/min over a range of 220–330 nm, a bandwidth of 1 nm, and a response time of 4 s.

UV melting and the determination of thermodynamic parameters

Melting experiments were performed in triplicate with a Jasco V-650 spectrophotometer equipped with a Peltier thermostat (Jasco, Tokyo, Japan) using quartz cuvettes of 10 mm path length. Absorbances of oligonucleotides ($\sim 5 \mu\text{M}$) were recorded at 295 nm from 10 to 90°C with a heating and cooling rate of 0.2°C/min. Employing a linear fit with extrapolation of upper and lower baselines, the absorbance vs. temperature curve was converted to a folded fraction α vs. temperature curve. Melting temperatures T_m with a folded fraction $\alpha = 0.5$ are reported as an averaged value from heating curves of three independent measurements. With the equilibrium constant $K(T)$ given by $K(T) = (1-\alpha(T))/\alpha(T)$, thermodynamic parameters ΔH° , ΔS° and ΔG° were derived from a van't Hoff analysis using the standard thermodynamic relationships $\ln K(T) = -\Delta H^\circ/RT +$

$\Delta S^\circ/R$ and $\Delta G^\circ(T) = \Delta H^\circ - T\Delta S^\circ$. The analysis is based on a two-state equilibrium with the assumption of negligible heat capacity effects and temperature independent enthalpies and entropies.

NMR spectroscopy

All NMR spectra were acquired on a Bruker Avance Neo 600 MHz spectrometer equipped with an inverse $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{19}\text{F}$ quadruple resonance cryo-probehead and z-field gradients. Unless stated otherwise, NMR spectra were acquired at 30°C in 10 mM potassium phosphate buffer with 10% D_2O , pH 7.0. For spectral processing and analysis, Topspin 4.0.7 and CcpNmr Analysis 2.4.2 was used (29,30). Proton chemical shifts were indirectly referenced to sodium trimethylsilyl propionate (TSP) through the temperature-dependent water chemical shift at pH 7.0 and carbon chemical shifts were referenced to sodium trimethylsilylpropanesulfonate (DSS) through an indirect referencing method. An optimized WATERGATE with w5 element was employed for solvent suppression in 1D spectra and 2D nuclear Overhauser effect (NOESY) experiments. Two-dimensional NOESY spectra were recorded with mixing times of 300, 150 and 80 ms. ^1H – ^{13}C heteronuclear single quantum correlation (HSQC) experiments were performed with a 3–9–19 water suppression scheme, $4\text{K} \times 500$ data points and a spectral width of 7500 Hz to accommodate $^{13}\text{C}6/\text{C}8/\text{C}2$ resonances in the indirect dimension. ^1H – ^{13}C heteronuclear multiple bond correlation (HMBC) spectra were acquired with a jump-and-return water suppression, $2\text{K} \times 136$ data points, and processed with 50% Non-Uniform Sampling (NUS) in the indirect dimension. Double quantum filtered correlation (DQF-COSY) spectra were either recorded with water suppression through presaturation in 100% D_2O or with a 3–9–19 water suppression in 90% $\text{H}_2\text{O}/10\%$ D_2O employing $2\text{K} \times 512$ data points.

NMR structure calculation

Initially, 100 starting structures of lowest energy were selected from 400 structures calculated by a simulated annealing protocol in XPLOR-NIH 3.0.3 (31,32). Distance restraints were set according to crosspeak intensities in NOESY spectra. For non-exchangeable protons, intensities were categorized as strong ($2.9 \pm 1.1 \text{ \AA}$), medium ($4.0 \pm 1.5 \text{ \AA}$), weak ($5.5 \pm 1.5 \text{ \AA}$) and very weak ($6.0 \pm 1.5 \text{ \AA}$). For exchangeable protons, intensities were assigned as strong ($4.0 \pm 1.2 \text{ \AA}$), medium ($5.0 \pm 1.2 \text{ \AA}$) and weak ($6.0 \pm 1.2 \text{ \AA}$). Glycosidic torsion angles were set to 170 – 310° for *anti* conformers and to 25 – 95° for *syn* conformers. The pseudorotation phase angle was restricted to 144 – 180° for *south*-type sugar puckers. G22 in $^{6\text{Br}}Q$ was restrained to a *north* sugar pucker with a pseudorotational angle of 0 – 90° . Additionally, planarity and hydrogen bond restraints were employed for bases in each G-tetrad.

Refinement was performed *in vacuo* using AMBER18 with the parmbsc force field and OL15 modifications for DNA (33). Partial atomic charges for the modified 8-bromoguanosine residues were calculated with the RED software using a DFT approach (34). The same restraints were used as before. Twenty converged lowest-energy struc-

tures were generated from 100 starting structures with a simulated annealing *in vacuo*.

For a refinement in water, potassium ions were added to neutralize the system and two ions were placed in the inner channel of the G-quadruplex core between two adjacent tetrad layers. The system was hydrated with TIP3P water in a truncated octahedral box of 10 \AA . The final simulation was performed at 1 atm and 300 K for 4 ns using only NMR-derived NOE distance and Hoogsteen hydrogen bond restraints. The trajectory was averaged for the last 500 ps and minimized *in vacuo* to obtain 10 lowest-energy structures. The VMD 1.9.2 software was used for the analysis of calculated structures (35). Three-dimensional structure representations were prepared with Pymol 1.8.4.

RESULTS AND DISCUSSION

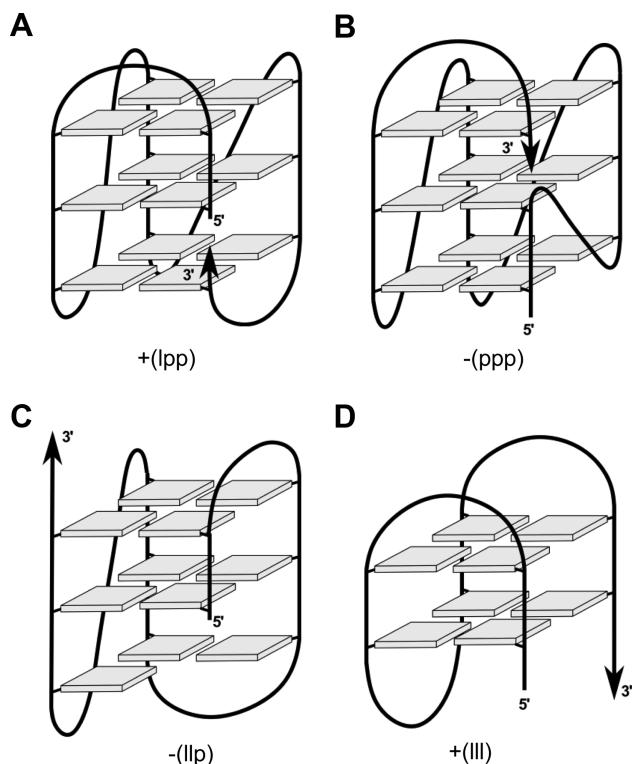
Design of G-quadruplex forming sequences

Initially, a parent G-rich sequence 5'-GGCTAGGGTCAGGGTGGGTCAG-3' with a truncated 5'-terminal GG-tract termed *Qref* was designed to restrict the available topological space and to better control the formation of various putative G4 species for thermodynamic and NMR spectral analyses (for all *Qref* derived oligonucleotides see Table 1). Sequences comprising a guanine-deficient GG-tract in addition to three GGG-tracts may fold into a three-layered quadruplex topology but with one tetrad bearing a vacant site. Interestingly, bioinformatics studies have shown that such sequences are highly abundant in the human genome and may potentially support an environment-responsive regulation in cellular processes (36,37). Thus, the vacant site can be filled with a guanine base from guanine-containing metabolites but also stabilized intramolecularly through snapback loop formation to form an intact G-core (38–40). Accordingly, the 3'-terminal TCAG-sequence of *Qref* was expected to form a snapback loop for filling a putative vacant position as a result of its truncated 5'-GG-tract, allowing a G4 architecture with three intact G-tetrads.

Given a most stable 1-nt propeller loop linking the third and fourth G-column and the absence of any diagonal loop with available loop lengths of ≤ 3 nucleotides, the first two 3-nt intervening segments may either arrange in a lateral or a propeller loop, the latter required to follow the right-handed G4 helicity with a counter-clockwise progression if the 5'-terminal G-column runs towards the viewer (18,41). Also, excluding a highly unlikely propeller-type snapback loop only observed in a rather unique G4 formed by a *c-kit* promoter sequence (42), two stable topologies for a three-layered snapback loop architecture with a 3–3–1 loop length arrangement can be envisaged (Figure 1A, B). For convenience, a systematic G4 notation according to Webba da Silva was employed, classifying different topologies with a simple descriptor that only includes the type (p, l and d for propeller, lateral, and diagonal) and relative direction of loops linking G-tracts (+ and – for clockwise and counter-clockwise progression, respectively) (43,44). Thus, the two three-layered topologies are expected to either comprise a +(lpp) hybrid-type structure with one 3-nt lateral loop followed by a 3-nt and a 1-nt propeller loop (Figure 1A) or a –(ppp) parallel fold with two 3-nt and a single 1-nt propeller

Table 1. Parent sequence *Qref* and *Qref*-derived oligonucleotides with terminal 1-nt additions/deletions as well as single and double mutations^a

Oligonucleotide	Sequence	Major topology ^b
<i>Qref</i>	5'-GG-CTA-GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>3^{'del}Q</i>	5'-GG-CTA-GGG-TCA-GGG-T-GGG-TCA-3'	— ^c
<i>2BrQ</i>	5'-G <u>Br</u> G-CTA-GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>6BrQ</i>	5'-GG-CTA- <u>Br</u> GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>7BrQ</i>	5'-GG-CTA- <u>G</u> BrGG-TCA-GGG-T-GGG-TCA-G-3'	— ^c
<i>8BrQ</i>	5'-GG-CTA-G <u>G</u> BrG-TCA-GGG-T-GGG-TCA-G-3'	— ^c
<i>16BrQ</i>	5'-GG-CTA-GGG-TCA-GGG-T- <u>Br</u> GGG-TCA-G-3'	+(lpp)
<i>5^{'T}Q</i>	5'- <u>T</u> -GG-CTA-GGG-TCA-GGG-T-GGG-TCA-G-3'	-(ppp)
<i>Q-5T</i>	5'-GG-CTT-GGG-TCA-GGG-T-GGG-TCA-G-3'	-(ppp)
<i>Q-5I</i>	5'-GG-CT <u>I</u> -GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-5X^d</i>	5'-GG-CT <u>X</u> -GGG-TCA-GGG-T-GGG-TCA-G-3'	-(ppp)
<i>Q-11T</i>	5'-GG-CTA-GGG-TCT-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-11I</i>	5'-GG-CTA-GGG-T <u>I</u> -GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-11X^d</i>	5'-GG-CTA-GGG-T <u>X</u> -GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-5I-11I</i>	5'-GG-CTI-GGG-T <u>I</u> -GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-3T-10T</i>	5'-GG- <u>T</u> TA-GGG-TTA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-3A</i>	5'-GG- <u>A</u> TA-GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)
<i>Q-3X^d</i>	5'-GG- <u>X</u> TA-GGG-TCA-GGG-T-GGG-TCA-G-3'	+(lpp)

^aMutations are indicated by underlined bold letters.^bBased on imino NMR signal intensities at $T = 30^\circ\text{C}$. ^cSignificant structural heterogeneity. ^dX = abasic 1',2'-dideoxyribose residue.**Figure 1.** Putative G4 topologies of designed sequences given a third 1-nt propeller loop and the absence of diagonal loops with loop lengths ≤ 3 residues. (A) Three-tetrad hybrid-type +(lpp) and (B) parallel -(ppp) topology with snapback loop filling a vacant G-tetrad position; (C) hybrid-2 -(llp) G4 with a G vacancy in its lower tetrad and (D) two-tetrad antiparallel +(lll) G4. Stabilities of the latter two G4s are expected to be compromised by the formation of only two intact G-tetrads. Topologies are characterized by simple loop descriptors according to Webba da Silva (43,44).

loop (Figure 1B). The 3'-terminal snapback loop filling the vacant position of the first G₂-tract was omitted in the descriptors.

Other putative folds suffer from the loss of a third intact G-tetrad. A hybrid-2 topology -(llp) with two lateral loops is expected to be considerably disfavored by leaving an empty outer G-core position with the 3'-terminal snapback G located at the opposite G4 face (Figure 1C). Likewise, the stability of a putative two-tetrad antiparallel G4 +(lll) with three lateral loops is compromised by the presence of only two stacked G-tetrad layers (Figure 1D). Clearly, formation of a three-layered chair-type topology is effectively eliminated by a 1-nt intervening sequence between G₃-tracts which is largely restrained to form a stable propeller-type loop.

It should be mentioned that there are no flanking sequences in the basic snapback loop design of *Qref*. Whereas loop interactions with overhang nucleotides have been shown to influence G4 folding as demonstrated by human telomeric sequences with different flanking residues, eliminating the impact of overhang residues reduces additional complexity to focus on loop interactions inherent to the G4 structure. Also, putative rearrangements of flanking nucleotides, e.g. in genomic DNA upon formation of quadruplex-duplex junctions may easily alter overhang interactions in unpredictable ways.

Topology of G-quadruplex forming sequences

The imino proton spectral region of parent *Qref* suggests formation of a high-populated major and a lower populated minor three-layered G-quadruplex in a molar ratio of about 1:0.4 with additional very low-intensity signals of a third species (Figure 2). Deleting the 3'-terminal G to give the truncated *3^{'del}Q* sequence, imino resonances of both major and minor quadruplexes vanish and a complete re(un)folding is evidenced by downfield-shifted and significantly broadened imino signals. Although structures formed by truncated *3^{'del}Q* are not amenable to a more detailed structural analysis, it can be hypothesized based on their antiparallel CD signature with minima and maxima of

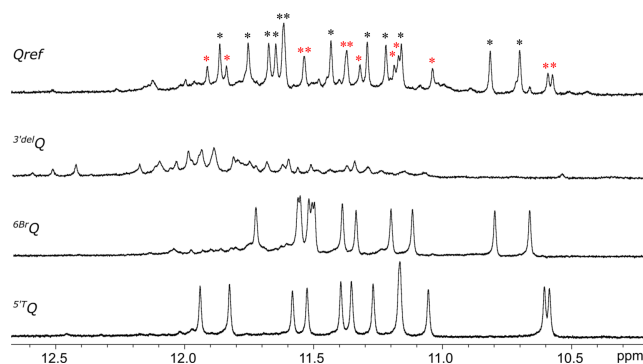


Figure 2. Imino proton spectral region of *Qref*, *3'delQ*, *6BrQ* and *5'TQ*. Black and red asterisks for *Qref* indicate imino resonances of a major and minor fold, respectively. NMR spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30°C.

the amplitude at about 265 and 290 nm (Figure 3) that they include G-register isomers of two-tiered antiparallel folds such as the one shown in Figure 1D or similar to a basket-type telomeric G-quadruplex with only two G-tetrad layers called Form 3 (45). However, formation of multistranded species is also conceivable. Requiring a 3'-terminal G for folding, the *Qref* sequence is expected to form a snapback loop G4 with an interrupted first G-column and with the 3'-G residue participating in a G-tetrad.

To get more insight into the major *Qref* topology, either position 6, 7 or 8 of the second GGG-tract was selectively substituted with a *syn*-favoring 8-bromo-2'-deoxyguanosine residue ^{Br}G. Whereas substitutions at positions 7 and 8 proved deleterious for the original fold with considerable shifts and broadening of imino resonances (not shown), the *6BrQ* sequence with a bromo-modification at position 6 featured a mostly clean imino proton spectral region composed of twelve well resolved resonances (Figure 2).

A detailed NMR spectral analysis of *6BrQ* was based on NOESY, DQF-COSY, ¹H-¹³C HSQC and ¹H-¹³C HMBC experiments (for spectral assignment strategies, spectra, and chemical shifts see Supplementary Figures S1–S5 and Supplementary Table S1 in the Supporting Information). Unambiguous resonance assignments together with H/D isotope exchange experiments demonstrated folding into a +(lpp) hybrid-type topology. It comprises a single first lateral followed by two propeller loops as well as one broken *syn-syn-anti* column (G22–G1–G2) and three *syn-anti-anti* columns (G6–G7–G8, G12–G13–G14 and G16–G17–G18) (Figure 4A).

Comparison of imino proton chemical shifts and in particular observation of closely similar crosspeak patterns in 2D NMR spectra identified the topology of the ^{Br}G-modified sequence to also represent the major fold of *Qref* (Supplementary Figures S6 and S7). In trying to unambiguously identify the lower populated *Qref* G4 species, a single 5'-flanking T was added to the *Qref* sequence to give *5'TQ*. Such an approach follows the expectation that a 5'-T overhang should not impact or rather stabilize a parallel -(ppp) fold with G1 located in the 5'-outer tetrad while preventing a hybrid-type +(lpp) topology with G1 located in the central tetrad due to steric clashes with the fill-in 3'-

terminal G (see Figure 1A, B). Indeed, modified sequence *5'TQ* again features a clean and well resolved imino proton NMR spectral region indicative of a single G4 fold and a CD signature characteristic of a parallel topology (Figures 2 and 3). A detailed NMR spectral analysis using standard strategies demonstrated its parallel topology with three propeller loops and a terminal snapback *syn*-G residue filling the vacant position at the 3'-outer tetrad (Figure 4B; for assignment strategies, spectra, and a compilation of chemical shifts see Supplementary Figure S8–S11 and Supplementary Table S2). Again, a close inspection of *5'TQ* and *Qref* spectra with highly similar signal chemical shifts and crosspeak patterns identified the parallel fold to be the higher populated minor species in the *Qref* mixture (see Figure 2).

Three-dimensional structure of *6BrQ* and *5'TQ*

Following resonance assignments, molecular dynamics simulations were performed for *6BrQ* and *5'TQ* in explicit water using NMR-derived distance and dihedral angle restraints. With rmsd values of 0.7 and 0.9 Å for *6BrQ* and *5'TQ*, respectively, the three-layered G-core is well defined for both G4s (for structural statistics and a superposition of ten lowest-energy structures see Supplementary Table S3 and Supplementary Figure S12). In general, higher rmsd values of 2.2 and 2.5 Å as calculated for all residues can be attributed to significant fluctuations of those loop residues that are mostly exposed to solvent but also partially directed towards a G4 groove. However, all residues of the snapback loop, second and third residues in the (+lpp) lateral loop bridging a narrow groove and the last 3'-residue of the 3-nt propeller loops are well defined.

A representative structure of the *6BrQ* G4 is shown in Figure 5A. The 5'-terminal G1 is positioned in the central tetrad of a *syn-syn-anti* G-tract. A 3-nt lateral loop C3–T4–A5 is followed by a 3-nt propeller loop T9–C10–A11, a 1-nt propeller loop T15, and a 3-nt snapback loop with 3'-terminal G22 filling the vacant position of the first G-column. Residues of the lateral snapback loop T19–C20–A21 bridging a wide groove adopt a well-defined orientation in the structural ensemble. Thus, a T19–A21 Hoogsteen base pair caps the lower G-tetrad with additional outer stacking of C20 onto the base pair (Supplementary Figure S13A). Looking at the first 3-nt lateral and following 3-nt propeller loops, a formed capping structure onto one side of the upper G-tetrad is noticeable and corroborated by several experimental NOE restraints (Figures 5B and Supplementary Figure S14A, B). Thus, 3'-terminal loop residues A5 and A11 stack onto the tetrad, exhibiting a putative hydrogen bond interaction between the A11 amino proton and A5 N3 in most of the structures. T4 stacks over A5 that is sandwiched between the preceding T4 and G2 of the outer G-tetrad. In contrast, A11 only shows poor stacking over *syn*-G6.

A highly similar arrangement with a 3'-adenine of a 3-nt propeller loop recruited through putative base-base interactions by a 3'-adenine of a preceding lateral loop to stack onto the top G-tetrad has also been observed by a human telomeric sequence 5'-(TTAGGG)₄TTA-3' with a single ^{Br}G modification in Na⁺ solution (46). In the formed antiparallel +(lpl) G4, the 3'-A of the lateral loop is addition-

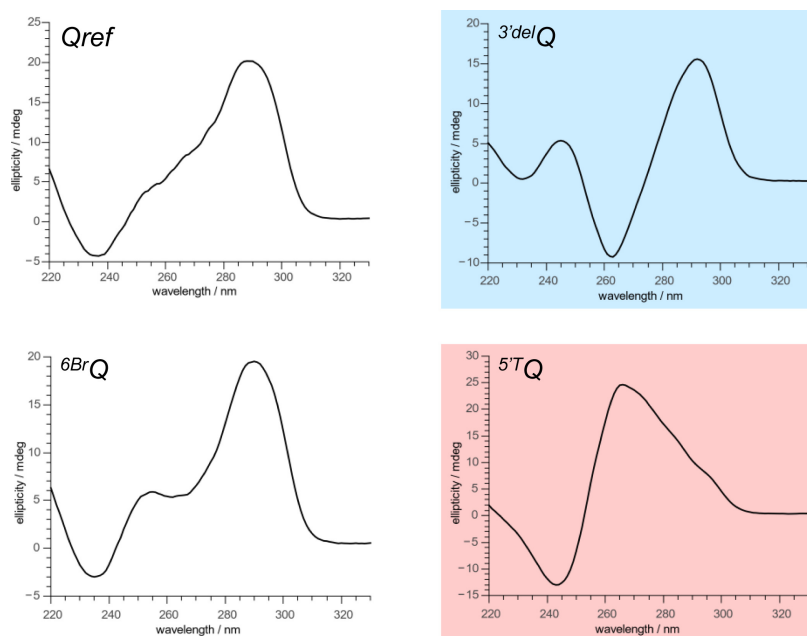


Figure 3. CD spectra of *Qref*, *3'delQ*, *6BrQ* and *5'TQ*. Whereas spectra of *Qref* and *6BrQ* are in line with a mixture of parallel, antiparallel, and/or hybrid-type topologies, spectra of *3'delQ* (blue background) and *5'TQ* (red background) feature a signature typical for an antiparallel and a parallel topology, respectively (20). CD spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 20°C.

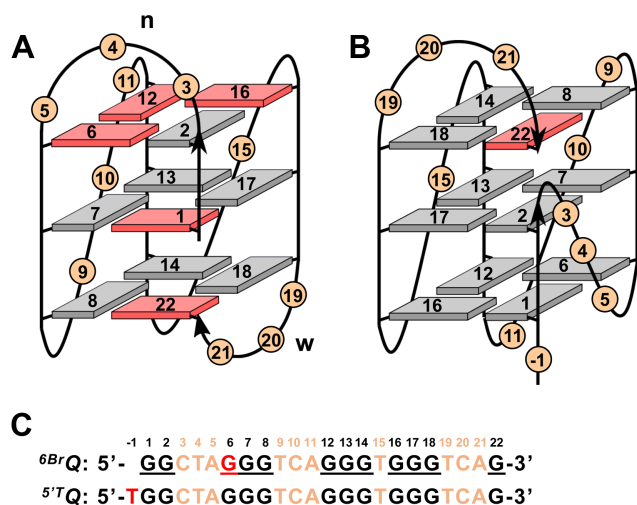


Figure 4. (A) Topologies of *6BrQ* and (B) *5'TQ* quadruplexes; *syn*- and *anti*-G residues are colored red and grey, *n* and *w* denote narrow and wide groove, respectively. (C) Sequence of *6BrQ* and *5'TQ* with residue numbering; added 5'-T and incorporated ^{Br}G residue in red, intervening sequences in orange, and Gs participating in tetrad formation underlined.

ally base-paired to a thymine of the 3'-flanking sequence (Supplementary Figure S14C). Likewise, stacking of a 3'-adenine base of a propeller loop onto an outer tetrad has been found in longer propeller loops if recruited by an adjacent coplanar base in a lateral loop or overhang sequence (47,48).

Calculated structures of *5'TQ* feature a parallel three-layered G4 with a T(-1) 5'-overhang and G1 in an outer tetrad position (Figure 5C). The G1-G2 run of the first G-column is again complemented with 3'-terminal *syn*-G22

through a lateral snapback loop. The latter adopts the same capping structure as found for *6BrQ* with A21 Hoogsteen hydrogen-bonded to T19 and the T-A base pair sandwiched between the lower G-tetrad and C20 (Supplementary Figure S13B). Whereas extensive disorder is observed for the first two residues of both the first and second 3-nt propeller loop, the last loop residues are well defined in their orientation as also demonstrated by various NOE contacts. Thus, A5 stacks onto the bottom tetrad below G6. However, such an arrangement may derive from the first propeller loop bridging only two tetrad layers in this snapback loop architecture. In contrast, bridging three tetrad layers in analogy to the +(lpp) topology of *6BrQ*, the second propeller loop features a 3'-adenine A11 that is found to insert into the G4 groove in all calculated structures while fixed by 10 NOE-based distance restraints (Figure 5D). Consequently, A11 at the 3'-position of the second 3-nt propeller loop has quite distinct orientations/interactions depending on its location within a -(ppp) or a +(lpp) topology.

Impact of mutations on favored topologies

To further assess the participation of loop bases on (de)stabilizing interactions, single and dual base mutations were introduced into intervening sequences and their impact on the favored loop formation and G4 topological equilibria was studied (Table 1). Mutations also involved inosine (I) as a purine nucleoside substitute and 1',2'-dideoxyribose (X) lacking a nucleobase. As shown by the imino proton NMR spectral regions, all mutant sequences generally fold into a major and minor G4 with an additional very low-populated species in most cases (Figure 6). Generally, the low signal intensity of the latter did not allow its detailed assignment without corresponding reference spec-

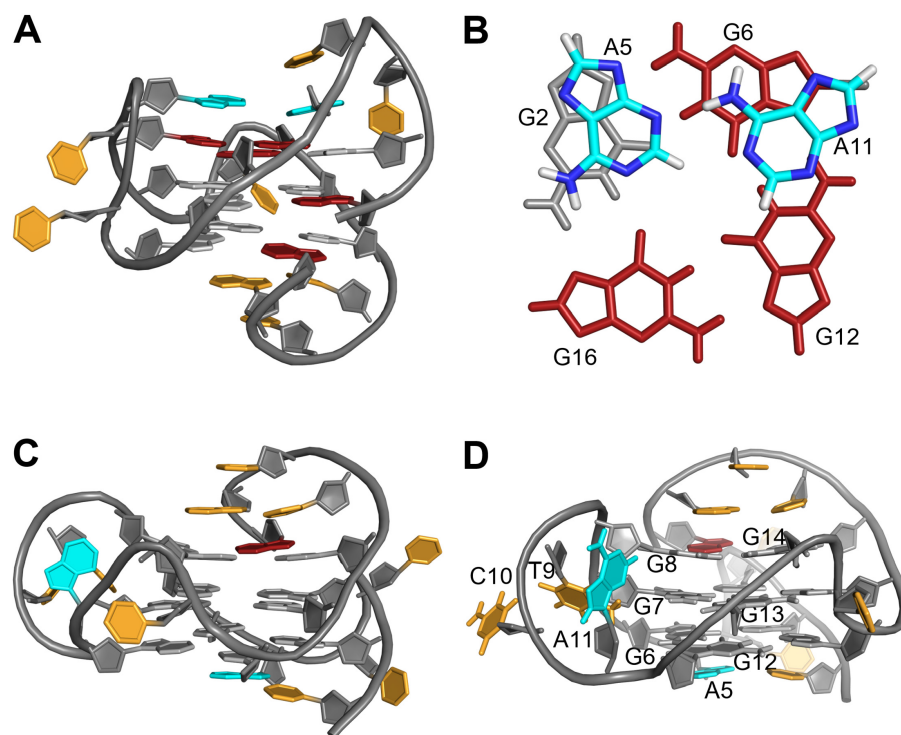


Figure 5. (A) Representative NMR structure of ^{6Br}Q and (B) view onto coplanar A5 and A11 residues stacked onto the upper tetrad with a putative A11 amino – A5 N3 hydrogen bond interaction. (C) Representative NMR structure of ^{5T}Q and (D) side view into a G4 groove with inserted A11. *Anti*- and *syn*-guanosines are colored in grey and red, respectively, A5 and A11 are colored in cyan and other residues are colored in orange.

tra (but see below). On the other hand, the two predominant species, formed in different molar ratios ranging between 0.3:1 up to 1:0.3, were unambiguously identified for each mutant as competing $+(lpp)$ and $-(ppp)$ topologies as formed by ^{6Br}Q and ^{5T}Q , respectively. It should be mentioned that such topological assignments are based on detailed NMR spectral analyses for all mutants through 2D NOESY and 1H - ^{13}C HSQC spectra and were additionally supported by a comparison of crosspeak patterns with ^{6Br}Q and ^{5T}Q reference spectra (see Supplementary Figure S15–S18 for exemplary spectral assignments of Q -5T and Q -11T and Supplementary Figure S19 for CD spectra of all sequences).

The topological ratio was mostly conserved in Q -3A and Q -3X with an adenosine and abasic site at position 3, associated with a $3C \rightarrow 3A$ and $3C \rightarrow 3X$ mutation. Also, replacing cytidine residues C3 and C10 in the 5'-terminal and central position of the first and second loop by T to yield the Q -3T-10T mutant had nearly no effect on the topological equilibrium. Notably, however, the major $+(lpp)$ topology was even maintained when substituting A11 within the second loop for either thymidine, inosine, or an abasic residue. More dramatic effects were induced by mutations at position 5 within the first loop. Whereas a $5A \rightarrow 5I$ purine-to-purine substitution in Q -5I did not noticeably change populations of major and minor fold, populations were inverted in Q -5T and Q -5X to form a major $-(ppp)$ G4. This indicates a critical role of purine bases in the 3-nt lateral loop end position for favorable stacking interactions due to their larger ring system. Although associated with a slight shift in favor of a parallel G4, corresponding dual substitutions with

inosine at loop terminal positions 5 and 11 do not indicate considerable synergistic effects in Q -5I-11I, again conserving the major hybrid topology as observed for the parent sequence.

The high-resolution structure suggests a stabilizing impact of an A5-A11 base pair capping the outer tetrad in the $+(lpp)$ G4 (Figure 5A, B). However, the present mutational studies emphasize the critical role of A5 but not of A11 as contributor for a major $+(lpp)$ fold. Whereas A5 may be replaced by the purine base hypoxanthine without compromising the favored formation of a hybrid against a parallel topology, its substitution by a T5 pyrimidine and in particular by an abasic nucleotide indicates a loss of critical stacking interactions for this lateral loop position, resulting in a significant shift towards a major parallel fold.

Unexpectedly, substituting A11 for inosine, thymidine, and even for an abasic residue did hardly shift equilibria in favor of a parallel structure although A11 was shown to form a capping structure with A5 in the parent G4 (Figure 5B). Apparently, additional A11 stacking and putative hydrogen bond interactions with A5 do not provide for a selective and noticeable stabilization of the hybrid-type species.

Impact of mutations on the thermodynamic profile for G-quadruplex formation

Signal intensities in the NMR spectra yield valuable information on the population of major and minor species formed in solution. However, a substitution-induced shift in molar ratios may result from different combinations of stabilizing and destabilizing effects on the competing topolo-

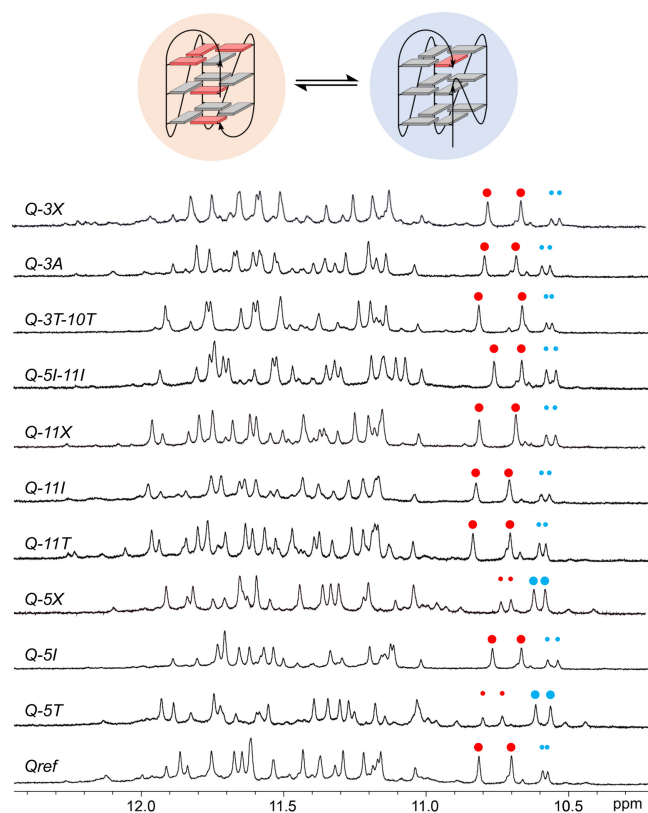


Figure 6. Imino proton spectral region of *Qref* mutants. NMR spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30°C. Red and blue dots of larger and smaller size mark major and minor high-field shifted resonances of G8 and G22 residues in +(lpp) and -(ppp) topologies shown on top.

gies. It would therefore be instructive to determine thermodynamic stabilities for all mutants in each of the two coexisting topologies, i.e. parallel or hybrid-type. Employing a *syn*-affine ^{Br}G6 analog or a 5′T-flanking residue successfully shifted equilibria to exclusively form a +(lpp) or -(ppp) *Qref* quadruplex. Using the same approach, all mutants were likewise modified with the expectation to yield two sets of structures comprising either hybrid or parallel topologies irrespective of the specific mutation.

Initially, the two additional sets of corresponding bromo- and 5′T-modified mutant sequences were subjected to a CD and NMR spectral analysis. Indeed, all 5′T sequences result in CD spectra typical of a parallel fold whereas ^{Br}G-modified sequences exhibit CD signatures consistent with a hybrid-type topology (Supplementary Figure S20 and S21). Accordingly, imino proton spectral regions of the former display clean spectra of a parallel species (Supplementary Figure S22). NMR spectra indicating a single hybrid-type G4 are also observed for most of the ^{Br}G6-modified sequences and additional low-populated species as detected for ^{6Br}*Q-5T* and ^{6Br}*Q-11T* were not expected to significantly compromise the thermodynamic characterization of the major hybrid G4 (Supplementary Figure S23).

Each ^{Br}G- and 5′T-modified mutant was subjected to UV melting experiments. No hysteresis effects were observed between heating and cooling curves, implying a thermody-

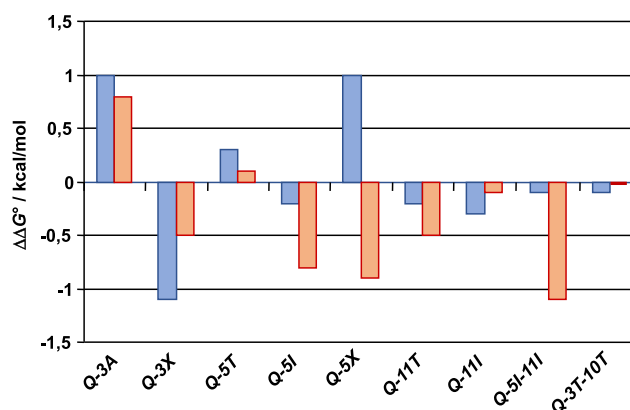


Figure 7. Free energy contribution $\Delta\Delta G^\circ$ of specific mutations upon the formation of a hybrid (blue bars) and a parallel topology (red bars) at 30°C. Uncertainties in the determination of Gibbs free energies averaged over three independent measurements were estimated to be between ± 0.1 to ± 0.2 kcal/mol (for exemplary van't Hoff plots see Supplementary Figure S24). No data were extracted for the *Q-11X* mutant due to a more significant structural heterogeneity.

namically controlled transition. In the following, the melting profiles were evaluated by a van't Hoff analysis in terms of enthalpy ΔH° , entropy ΔS° , and Gibb's free energy ΔG° for G4 formation at 30°C (data are summarized in Supplementary Tables S4 and S5). Favorable or unfavorable free energy contributions for specific mutations in both a hybrid-type fold (^{6Br}*Q*-modified sequences) and a parallel fold (^{5′T}*Q*-modified sequences) are given by changes in ΔG° relative to the free energy of the modified parent sequences ^{6Br}*Q* or ^{5′T}*Q*. Notably, except for the *Q-5I* and *Q-5I-11I* double mutant (see below), a good correlation of topology-dependent (de)stabilizing substitution effects in ^{Br}G- and 5′T-modified sequences with NMR-derived relative populations for unmodified mutants was found. Consequently, the modified analogs are not only good structural mimics of unmodified quadruplexes but also good surrogates for the analysis of mutation-dependent energy profiles.

Free energy contributions of the various mutant sequences when either folded into a hybrid or parallel quadruplex are plotted in Figure 7. Apparently, loop substitutions exert moderate but often noticeable effects by up to 1 kcal/mol on the quadruplex stabilities. Position dependent changes can be summarized as follows:

- (i) Replacing C3 by adenosine has a destabilizing effect for both hybrid and parallel G4s but a 3C→3X substitution stabilizes the G4 and in particular the +(lpp) topology. Destabilizing effects for C/T→A substitutions at the first position of a loop have been reported before (14,49). Because this position seems most disordered in both the lateral and the propeller loop of the hybrid and parallel G4, destabilization can be attributed to more unfavorable hydrophobic effects of a solvent-exposed purine compared to a pyrimidine base. By the same reasoning, an abasic residue should exert a stabilizing effect as observed. Corresponding destabilizing effects for a T→A substitution in 1-nt

- propeller loops have also been well documented in the past (11,50).
- (ii) A 5A→5T and especially a 5A→5X substitution destabilizes the hybrid topology but stabilizes the parallel topology in case of *Q-5X*. Here, destabilization only found for the hybrid fold again confirms critical purine stacking on the outer tetrad for the 3-nt lateral loop end position associated with a shift towards a parallel topology upon replacing A by a pyrimidine or abasic nucleotide.
 - (iii) Substitutions at position 11 but also C→T mutations as in *Q-3T-10T* have only a small impact on stabilities as already suggested by conserved populations of hybrid and parallel species. Apparently, interactions of A11 as indicated in the hybrid fold provide only minor contributions to the G4 stability.
 - (iv) There is a conspicuous stabilizing effect on the parallel topology upon a 5A→5I substitution in *Q-5I* and also on double mutant *Q-5I-11I*. Such a stabilization with a $\Delta\Delta G^\circ$ of about -1 kcal/mol should give an estimated fivefold increase of the population ratio in favor of the parallel species at 30°C. However, NMR signal intensities point to a conserved major hybrid structure. Apparently, inosine here seems to favorably interact with the adjacent 5'-T overhang of the parallel G4, concealing true effects on the non-modified mutant in this particular case.

Interestingly, there is a pronounced enthalpy-entropy compensation on the substitution-induced free energy effects (see Supplementary Tables S4 and S5). However, corresponding changes in G4 stabilities are mostly governed by enthalpic when compared to smaller entropic contributions in line with previous reports on other loop-dependent changes in stability (11,12).

Impact of a central 1-nt loop on the global fold

Two 1-nt loops or even a single but central 1-nt loop are known from empirical studies to strongly promote all-parallel topologies irrespective of the length and composition of other intervening sequences present (22,23). Apparently, formation of a propeller loop affects formation of neighboring loops in a way not yet understood. The snap-back loop G4 architecture allowing for either a stable hybrid or parallel species offers the opportunity to study such an effect in more detail. Consequently, exchanging the second 3-nt loop in the *Qref* sequence by a 1-nt T-loop gives a truncated mutant 5'-GGCTAGGGTGGGTGGGTCAG-3' termed *Q-311-T* with a 3-1-1 loop length arrangement. Of note, a central 1-nt propeller loop seems identical for both topologies in spanning three tetrad layers across a medium groove. Nevertheless, coexisting with two minor species of similar intensity (see below), a predominant G4 populated by about 60% and identified as a parallel (-ppp) conformer was observed for *Q-311-T* in line with expectations (Figure 8). A major parallel fold is also clearly apparent from its CD signature (Supplementary Figure S25).

There have been reports of forming very stable multimeric quadruplexes with sequences comprising short loops even at low concentration (51). To confirm exclusive formation of

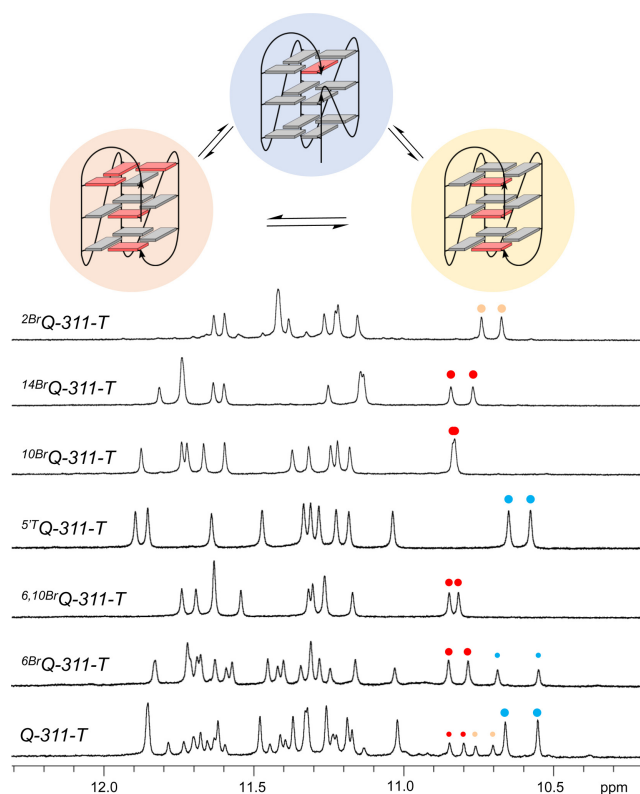


Figure 8. Imino proton spectral region of *Qref* derived quadruplexes with a central 1-nt T-loop. NMR spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30°C. Blue, red, and yellow dots of varying size mark major and minor high-field shifted resonances of G8 and G20 residues in a parallel topology, a +(lpp) topology with a first *syn-syn-anti* column, and a +(lpp) topology with a first *syn-syn-syn* column as shown on top.

monomolecular structures and rule out potential misinterpretations due to aggregation effects, non-denaturing polyacrylamide gel electrophoresis was performed for all G4-forming sequences (Supplementary Figure S26). There was no indication of any noticeable association of the G4s including those with a shortened loop. Apparently, the snap-back loop motif inherent to all structures may effectively prevent multimer formation under the present conditions.

A significant population of a parallel species even for the ^{Br}G-modified ^{6Br}*Q-311-T* analog again demonstrates a strong propensity of the sequence with a second central 1-nt loop to adopt an all-parallel topology. A single hybrid-type species was only enforced through introduction of a second ^{Br}G substitution at *syn*-position 10 to give ^{6,10Br}*Q-311-T*, further shifting topological equilibria (Figure 8). In line with a shift towards the parallel species, there is a strongly stabilizing effect for the parallel G4 upon shortening the central 3-nt propeller loop into a 1-nt T-loop. A van't Hoff analysis of UV melting transitions gives a change in ΔG°_{30} of -4.5 kcal/mol when going from parent ^{5T}*Q* to ^{5T}*Q-311-T*, both exclusively folded into a parallel G4 (Table 2). Such a dramatic stabilization is remarkable and results from a significantly more favorable enthalpy overcompensating for a less favorable entropic change.

Table 2. Thermodynamic parameters of G4 formation for sequences with 3–3–1 and 3–1–1 loop length arrangements at 30°C^a

Sequence	T_m (°C)	ΔH° (kcal/mol) ^b	$-T\Delta S^\circ$ (kcal/mol) ^c	ΔG°_{30} (kcal/mol) ^d	$\Delta\Delta G^\circ_{30}$ (kcal/mol) ^e
5' T Q ^f	44.1 ± 0.3	-53.7 ± 1.9	51.3 ± 1.7	-2.4 ± 0.1	0 ^f
5' T Q-311-T ^f	61.5 ± 0.7	-72.9 ± 2.8	65.9 ± 2.7	-6.9 ± 0.2	-4.5 ± 0.3 ^f
16Br Q ^g	49.3 ± 0.2	-53.2 ± 1.3	50.0 ± 1.2	-3.2 ± 0.1	0 ^g
14Br Q-311-T ^g	59.2 ± 0.2	-72.3 ± 0.6	65.9 ± 0.6	-6.4 ± 0.1	-3.2 ± 0.2 ^g
2Br Q ^h	47.7 ± 0.5	-53.9 ± 0.7	50.9 ± 0.7	-3.0 ± 0.1	0 ^h
2Br Q-311-T ^h	65.5 ± 0.3	-71.4 ± 0.7	63.8 ± 0.6	-7.5 ± 0.1	-4.5 ± 0.2 ^h

^aAverage values with standard deviations derived from the analysis of three independent UV melting experiments. ^bDetermined from a van't Hoff plot.

^c $\Delta S^\circ = \Delta H^\circ / T_m$.

^d $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

^eDifference in Gibb's free energy ΔG°_{30} to the corresponding 3–3–1 quadruplex of the same topology and conformation.

^f-(ppp) topology.

^g+(lpp) topology with one *syn-syn-anti* and three *syn-anti-anti* columns.

^h+(lpp) topology with one all-*syn* and three all-*anti* columns.

Evaluating corresponding effects for a loop shortening in case of a heteropolar stacked hybrid as performed successfully for single residue substitutions through the ^{Br}G modification is hampered by a significant amount of an additional parallel fold found for ^{6Br}Q-311-T (see Figure 8). Although imparting high stability, 1-nt propeller loops also impose steric restraints and higher rigidity. Interestingly, a complete tetrad flip with *anti* → *syn* transitions in an all-*anti* parallel G4 was most effectively induced by the incorporation of a *syn*-affine ^{Br}G at a position following a 1-nt propeller loop. This was attributed to direct steric clashes of the 8-bromine substituent with the 5'-phosphate oxygen atoms in *anti*-^{Br}G but not in *syn*-^{Br}G (52). Consequently, a ^{Br}G substitution at position 10 or 14 following the 1-nt loops in Q-311-T was expected to increasingly shift equilibria towards the hybrid fold. In fact, ^{10Br}Q-311-T but also ^{14Br}Q-311-T features a clean one-component spectrum without any coexisting additional species as observed for mono-substituted ^{6Br}Q-311-T with a bromination site following the 3-nt loop (Figure 8).

To exclude a putative direct impact of the introduced bromo-substituted G on the central loop, the ^{14Br}Q-311-T sequence was selected for a comparative thermodynamic analysis with the corresponding reference sequence ^{16Br}Q featuring a 3-nt central loop. Placed at position 14 or 16, the ^{Br}G analog is far removed from the loop of interest and allows for a straightforward comparison of loop length dependent effects. Initially, partial NMR assignments combined with closely matching crosspeak patterns observed for the already characterized ^{6Br}Q hybrid quadruplex unambiguously identified a +(lpp) hybrid to also constitute the favored conformer for the ^{14Br}Q-311-T and ^{16Br}Q sequence (Supplementary Figure S27). Additional minor resonances observed in the ^{16Br}Q imino proton NMR spectral region were shown by DSC measurements to result from some high-melting species and hardly impact thermodynamic data obtained from a van't Hoff analysis of the lower-melting quadruplex-to-single strand transition (Supplementary Figure S28). With a $\Delta\Delta G^\circ_{30}$ of -3.2 kcal/mol there is also a significant stabilization of the hybrid topology upon shortening the central loop (Table 2). However, in line with empirical observations on the effect of 1-nt loops on G4 folding, the gain in stability of the parallel topology is noticeably higher by more than 1 kcal/mol

when compared to the hybrid structure with its heteropolar stacking.

While the thermodynamic data above constitute an energetics-based explanation for the propensity of 1-nt loops to fold into a parallel topology, the structural basis of enforcing other loops to also adopt a propeller-type conformation remains elusive. As a matter of fact, a propeller loop will always link the two outer tetrads across a medium groove irrespective of the global topology. Also, there should be no direct interactions of the nucleobase within the central 1-nt loop with other residues of the G-core. This was demonstrated by substituting the 1-nt T-loop residue in ^{10Br}Q-311-T with an abasic residue X to give ^{10Br}Q-311-X. In addition to mostly identical thermodynamic profiles obtained from UV melting transitions of the two sequences, a clean hybrid-type spectrum with an imino proton spectral region closely matching imino proton chemical shifts of ^{10Br}Q-311-T was observed for ^{10Br}Q-311-X, corroborating the lack of any noticeable interactions emanating from the 1-nt loop (Supplementary Figure S29 and Supplementary Table S6).

Recently, hybrid-type +(lpp) and -(ppl) topologies with two propeller loops termed hybrid-1R and hybrid-2R have been reported for the first time (Supplementary Figure S30). Formation of the corresponding canonical quadruplexes with four non-interrupted G₃-columns was enforced by additional ^{Br}G modifications and further supported by Watson-Crick base pairing of overhang and loop residues to yield coaxially stacked duplex extensions (53,54). Given their design comprising two 1-nt propeller loops, their resistance against folding into the target hybrid G4 with a single lateral loop can be attributed to the combined effect of two 1-nt loops strongly favoring a parallel topology irrespective of the length of the third loop. It is conspicuous, however, that in both cases a rather unusual (3 + 1) hybrid conformer featuring tetrads of the same polarity with an antiparallel *syn-syn-syn* column and three parallel all-*anti* columns was the major species. A coexisting minor conformer for the +(lpp) G4, termed hybrid-1R', was demonstrated to comprise a *syn-syn-anti* and three *syn-anti-anti* columns (54). On the other hand, it is a corresponding hybrid-1R' G4 with the same pattern of G-core glycosidic conformations that predominates for the present +(lpp) hybrid-type mutants

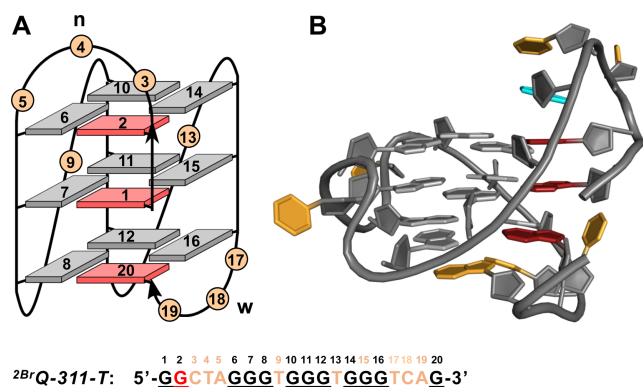


Figure 9. (A) Topology of the ^{2Br}Q-311-T quadruplex with a 3–1–1 loop arrangement; *syn*- and *anti*-G residues are colored red and grey, n and w denote narrow and wide groove, respectively. The sequence of ^{2Br}Q-311-T with residue numbering is shown at the bottom; ^{Br}G residue in red, intervening sequences in orange, and Gs participating in tetrad formation underlined. (B) Three-dimensional structure of ^{2Br}Q-311-T; color scheme as in (A) but with A5 colored in cyan.

with their central 3-nt propeller loop. In fact, based on unfavorable stacking interactions of *syn-syn* steps, the latter should be preferred over the hybrid-1R conformer with its all-*syn* column (55,56). Notably, –(pll) and –(llp) topologies (hybrid-1 and hybrid-2) formed by telomeric sequences in a K⁺ buffer likewise feature three *syn-anti-anti* and a single *syn-syn-anti* column with a heteropolar G-tetrad stacking (Supplementary Figure S30).

To also enforce a hybrid-1R conformer, the ^{Br}G analog was introduced at position 2 of the Q-311-T sequence to match a *syn*-position in the snapback loop G4 in case of an all-*syn* first G-column. The ^{2Br}Q-311-T imino proton spectral region shows a single three-layered G4 species with only very minor additional resonances (Figure 8). In the following, a detailed NMR spectral analysis with complete resonance assignments was performed on the ^{2Br}Q-311-T G4 (for spectral assignment strategies, spectra, and chemical shifts see Supplementary Figure S31–S33 and Supplementary Table S7 in the Supporting Information). In fact, based on the NMR data, formation of a hybrid-1R conformer with a lateral loop followed by two propeller and a snapback loop and with a G-core consisting of a first interrupted all-*syn* tract (G20–G1–G2) and three all-*anti* G-columns (G6–G7–G8, G10–G11–G12 and G14–G15–G16) was unambiguously confirmed (Figure 9A). For a three-dimensional structure determination, restrained molecular dynamics calculations were performed for ^{2Br}Q-311-T in explicit water. With a total rmsd of 1.5 Å and a G-core rmsd of 0.7 Å, the calculated structures are well defined (Supplementary Table S8 and Supplementary Figure S34A). Comprising an all-*syn* column, the G4 exclusively features homopolar tetrad stackings as indicated by its CD signature (Figures 9 and Supplementary Figure S25). The six-membered ring of the A5 purine within the first lateral loop stacks onto the six-membered pyrimidine ring of G2 (Supplementary Figure S34B). This small displacement when compared to the ^{6Br}Q quadruplex can be attributed to the upper tetrad polarity reversal and possibly also to the absence of another interaction with the loop residue from the

second 1-nt propeller loop. Both 1-nt propeller loops are either located in the G4 groove or exposed to solvent as also indicated by the lack of inter-residual contacts.

With the availability of corresponding reference spectra, the two minor species for Q-311-T were unambiguously identified as +(lpp) hybrid-1R' and hybrid-1R G4s with a first *syn-syn-anti* column and a first all-*syn* column through a match of typical crosspeak patterns (Supplementary Figure S35). Likewise, although only poorly populated, a minor third species observable in the parent Qref G4 spectra could finally also be traced to the all-*syn* hybrid-1R topology with three tetrads of the same polarity (Supplementary Figure S36) and this assignment should also apply to additional very low-populated species noticeable in various mutant spectra (see Figure 6).

With the completed assignment of three coexisting G4 conformers for the Qref and Q-311-T sequences, molar ratios were determined based on intensities of non-exchangeable T methyl resonances (Supplementary Figures S35 and S36). For Qref, molar ratios were determined to be 27% for the parallel topology, 11% for the hybrid-1R, and 62% for the major hybrid-1R' conformer. For the Q-311-T sequence with a second 1-nt loop, populations of parallel and hybrid-1R topologies increased by a factor of about two at the expense of the hybrid-1R' conformer, yielding populations of a major parallel, hybrid-1R, and hybrid-1R' species of 54%, 19% and 27%, respectively. Apparently, in addition to supporting a parallel G4, shortening the 3-nt central loop to a 1-nt loop strongly promotes the all-*syn* conformer with respect to the *syn-syn-anti* conformer.

For a more detailed thermodynamic analysis of the impact of loop shortening on the hybrid-1R formation, UV melting transitions were again analyzed in detail for the pair of loop length related G4s with a hybrid-1R fold, namely ^{2Br}Q-311-T and ^{2Br}Q. A corresponding hybrid-1R fold for ^{2Br}Q was again demonstrated through partial NMR assignments and the observation of closely similar crosspeak patterns to ^{2Br}Q-311-T (Supplementary Figure S37). Differences in free energies of G4 formation amount to a significant $\Delta\Delta G^\circ_{30}$ of –4.5 kcal/mol in favor of ^{2Br}Q-311-T with a single-nucleotide central loop (Table 2). Interestingly, such a stabilization is identical to the corresponding loop length mediated stabilization for a parallel topology and exceeds the stabilization for a hybrid-1R' topology by more than –1 kcal/mol.

Taken together, these results suggest that the propensity of a 1-nt propeller loop in also promoting a propeller-type conformation of other loops seems a direct consequence of its preference in linking two outer G-tetrads with the same tetrad polarity, i.e. with two anchoring G residues of the same glycosidic conformation. Such a propensity to drive folding into a parallel structure does not depend on specific interactions between G4 residues but is expected to be based on geometric effects. Thus, accompanied by conformational changes in the sugar-phosphate backbone, guanine bases must rearrange when trying to optimize hydrogen bond interactions within the tetrad upon *anti*→*syn* transitions. Interestingly, explicit solvent molecular dynamics simulations on single-loop G4s also reported a preference for the 1-nt propeller loop traversing three tetrad layers when linking two *anti*-Gs over linking an *anti*-G with a *syn*-G residue

(18). In the same study, a significant preference of a three-nucleotide loop to form a lateral over a propeller loop was reported. However, each 1-nt propeller loop with its propensity to drive folding into a parallel structure will oppose formation of a 3-nt lateral loop, with two 1-nt loops being more efficient than a single 1-nt loop but also with putative loop position dependent effects (21).

Following the considerations above, the parallel and hybrid-1R topology are expected to be equally promoted by 1-nt loops. However, because an all-*syn* column is considerably disfavored due to poor stacking interactions along *syn-syn* steps, an all-*anti* parallel topology will usually predominate unless appropriate modifications are introduced to enforce the hybrid-type G4. Likewise, an energetic penalty associated with an all-*syn* column may favor a hybrid-1R' over a hybrid-1R conformer to still result in a slightly higher population of the former in the *Q-311-T* G4.

It seems counter-intuitive that a *syn*-affine ^{Br}G analog will be most effective in adopting a *syn* conformation associated with a complete tetrad flip when incorporated at a tetrad position directly following a 1-nt propeller loop. However, such a conformational transition results from specific bromine steric clashes in case of ^{Br}G in an *anti*-conformation, outweighing a more favorable homopolarity of bridged outer tetrads. These effects emphasize the often subtle and interrelated loop interactions within a G-quadruplex structure and serve as a reminder that it still remains a challenge to disentangle these rather delicate energetic contributions and to predict their overall impact on putative G4 folds.

Implications for the folding and stability of G-rich sequences

By using a rigorous approach in combining detailed sequence-dependent structural and thermodynamic information, the present studies not only extend some empirical rules on the impact of loops in G4 folding, but also provide for an in-depth molecular understanding of findings put forward in the past. Although results have been derived from a particular G4 architecture lacking overhang sequences to focus on inherent G4 interactions, important generalizations emerge. Thus, whereas first loop residues tend to be rather flexible and mostly exposed to solvent with no long-lived specific interactions, 3'-terminal loop residues are often engaged in stacking interactions with an outer G-tetrad. These can provide for a significant stabilization of lateral loops bridging a narrow groove but may also form in longer propeller loops with ≥ 3 residues. For the latter, stacking of the last residue onto an external tetrad is often promoted by additional base-base interactions with a loop or overhang residue to form a capping structure. This seems to be a more general structural arrangement and can be found in quadruplexes of promoter sequences such as *c-myc* or *PARP1* (48,57). In general, stacking interactions will benefit from larger ring systems, favoring purine over pyrimidine bases. In contrast, if exposed to solvent, a smaller and less extended hydrophobic surface area of a pyrimidine base is expected to be less destabilizing based on hydrophobic effects when located at the loop 5'-end. These effects are reflected in the well-known empirical findings that adenine is destabilizing at first loop positions or in 1-nt pro-

PELLER loops but mostly stabilizing when positioned at the last loop position. Such contributions may also be biologically relevant in selecting for G4 loop isomers from oncogene promoter sequences with short adenine-deficient propeller loops. Also, favorable 5'-TTA-3' intervening tracts in the human telomeric sequence may result from natural selection. On the other hand, such a discrimination between first and last loop residue does not necessarily apply to lateral loops bridging a wide groove. Progressing along the right-handed G-tract helicity (43), the smooth transition to the loop domain in this case allows the first residue to be easily stacked on a tetrad, often forming a capping base pair with the last base of the loop as observed for the lateral snapback loop in the ^{6Br}Q and ^{5T}Q G4s (Supplementary Figure S13).

From a structural point of view, a prominent A·A cap on the upper tetrad formed by end residues of the 3-nt lateral loop bridging the narrow groove and a following 3-nt propeller loop seems an obvious stabilizing element in the present hybrid-type structure. However, point mutations and a more detailed thermodynamic analysis indicates that it is only the stacked 3'-terminal adenine of the lateral loop that promotes a hybrid over a parallel fold whereas the stacked 3'-terminal adenine of the propeller loop hardly exerts any impact on topological equilibria and thermodynamic stabilities. Obviously, structural information on a single competing species can be misleading when it comes to stability effects and only when complemented by thermodynamic profiles for all competing structures allows for the evaluation of critical interactions and ultimately the prediction of most favored folding pathways.

Based on purely geometric restraints it comes as no surprise that a 1-nt intervening sequence will form a propeller-type loop linking two adjacent parallel G-tracts. On the other hand, the observed propensity of 1-nt propeller loops to also enforce a propeller-type conformation on an additional longer loop and to therefore promote an all-parallel topology has been elusive. In the absence of any observable interactions of the single-nucleotide loop residue, kinetic effects may be operative. However, the present studies do not support a kinetic control of folding pathways. Whereas there are no significant hysteresis effects upon melting for each of the topologies examined, changing a second 3-nt to a 1-nt propeller loop does in fact result in a much higher, mostly enthalpy-driven, thermodynamic stabilization of three-layered G4s with exclusively homopolar G-tetrads when compared to a typical hybrid fold with outer G-tetrads of opposite polarity. To put it another way, a 1-nt propeller loop will preferably link two *anti*-G residues in opposite outer tetrads and such a preference will strongly promote formation of all-*anti* parallel quadruplexes in most cases. However, it is reasonable to assume that 1-nt propeller loops will also favor two anchoring *syn*-Gs in outer tetrads of the same polarity, yet corresponding conformers have never been observed in a three-layered G4 until now.

An observed resistance of sequences to folding into +(lpp) and -(ppl) topologies with two 1-nt propeller loops is a direct consequence of the added propensity for those loops to link outer tetrads of the same polarity. The same applies to folding into a -pd+p topology expected to be perfectly feasible but not experimentally verified to date. Thus,

in the presence of single-nucleotide loops a longer third intervening sequence will likewise be forced into a propeller loop to conserve homopolarity as found for several natural promoter sequences like *VEGF*, *HIF-1 α* , *c-kit* or *BCL-2* (58–61). On the other hand, elongating 1-nt to 2-nt propeller loops will result in a delicate balance between a release of the constraints enforcing an all-parallel topology and an enhanced propensity of the propeller loop to rearrange into an easily accessible 2-nt lateral loop.

These loop-dependent equilibria also direct attention to the formation of an alternative homopolar hybrid-type quadruplex with an all-*syn* G-tract, likewise supported by 1-nt propeller loops but unnoticed due to the lack of corresponding high-resolution data until now. Clearly, comprising two unfavorable *syn-syn* steps, such a G4 conformer will be disadvantaged compared to a parallel structure with an all-*anti* G-core. However, a (3 + 1) hybrid fold may be supported by Watson-Crick base pairing of complementary bases in its lateral loop (see also Supplementary Figure S30). In view of reports on the frequent occurrence of duplex stem-loop containing quadruplex sequences within the human genome (62), quadruplex-duplex hybrids with coaxially stacked duplex extensions and a single antiparallel all-*syn* column are conceivable to not only result from a rational sequence design for specific applications, e.g. as aptamers, but to also form more frequently within genomic sequences.

CONCLUSIONS

The artificial design of a non-canonical G-quadruplex with a snapback loop at its 3'-terminus offers the possibility to narrow putative folding pathways for spectral simplifications. It also enables better control on a topological selection while hampering G4 association through the stacking of outer tetrads due to steric hindrance by its snapback loop. With a combination of sequence mutations, high-resolution structural analyses, and thermodynamic profiling, loop residue interactions and their impact on the G4 fold can be characterized in detail to reveal relationships between loop length and loop composition with stabilities and conformational preferences of G4 structures. It also provides for a better structure-based understanding of many empirical findings for a more effective G4 design and sequence-based topological prediction.

The ability of rather weak interactions to drive folding into a particular topology is based on often small differences in stability among competing G4 structures. Even with a growing understanding of loop residue interactions, a reliable prediction of sequence dependent G4 topologies within a biological context where molecular crowding conditions and additional interactions with proteins may easily redirect folding pathways seems challenging. However, the engineering of defined G4 structures under controlled conditions for various medicinal and technological applications is expected to strongly benefit from a detailed understanding of intrinsic forces that stabilize quadruplexes. The availability of an increasing number of high-resolution G4 structures and their detailed thermodynamic profiling is expected to allow for an ever-growing understanding of sequence-specific intrinsic interactions guiding G4 folding.

DATA AVAILABILITY

The atomic coordinates and chemical shifts for ⁶Br Q (PDB 7ZEK; BMRB 34721), ⁵T Q (PDB 7ZEM; BMRB 34722) and ²Br Q-311-T (PDB 7ZEO; BMRB 34723) have been deposited in the Protein Data Bank.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Deutsche Forschungsgemeinschaft [410497337]. Funding for open access charge: Deutsche Forschungsgemeinschaft [410497337] and the Open Access Publication Fund of the Universität Greifswald.

Conflict of interest statement. None declared.

REFERENCES

- Davis, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.*, **43**, 668–698.
- Neidle, S. (2016) Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.*, **59**, 5987–6011.
- Carvalho, J., Mergny, J.L., Salgado, G.F., Queiroz, J.A. and Cruz, C. (2020) G-quadruplex, friend or foe: the role of the G-quartet in anticancer strategies. *Trends Mol. Med.*, **26**, 848–861.
- Roxo, C., Kotkowiak, W. and Pasternak, A. (2019) G-quadruplex-forming aptamers—characteristics, applications, and perspectives. *Molecules*, **24**, 3781.
- Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
- Dai, J., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **128**, 1096–1098.
- Marušič, M. and Plavec, J. (2015) The effect of DNA sequence directionality on G-quadruplex folding. *Angew. Chem. Int. Ed.*, **54**, 11716–11719.
- Lenarčič Živković, M., Rozman, J. and Plavec, J. (2018) Adenine-driven structural switch from a two- to three-quartet DNA G-quadruplex. *Angew. Chem. Int. Ed.*, **57**, 15395–15399.
- Kuryavii, V. and Patel, D.J. (2010) Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure*, **18**, 73–82.
- Monsen, R.C., DeLeeuw, L.W., Dean, W.L., Gray, R.D., Chakravarthy, S., Hopkins, J.B., Chaires, J.B. and Trent, J.O. (2022) Long promoter sequences form higher-order G-quadruplexes: an integrative structural biology study of c-Myc, k-Ras and c-Kit promoter sequences. *Nucleic Acids Res.*, **50**, 4127–4147.
- Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.*, **581**, 1657–1660.
- Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T. and Fox, K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.*, **35**, 4214–4222.
- Kumar, N. and Maiti, S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
- Guédin, A., Alberti, P. and Mergny, J.L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.
- Chen, J., Cheng, M., Salgado, G.F., Stadlbauer, P., Zhang, X., Amrane, S., Guédin, A., He, F., Šponer, J., Ju, H. *et al.* (2021) The beginning and the end: flanking nucleotides induce a parallel G-quadruplex topology. *Nucleic Acids Res.*, **49**, 9548–9559.

16. Pavc,D., Wang,B., Spindler,L., Drevenšek-Olenik,I., Plavec,J. and Šket,P. (2020) GC ends control topology of DNA G-quadruplexes and their cation-dependent assembly. *Nucleic Acids Res.*, **48**, 2749–2761.
17. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
18. Cang,X., Šponer,J. and Cheatham,T.E. III (2011) Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J. Am. Chem. Soc.*, **133**, 14270–14279.
19. Guédin,A., Gros,J., Alberti,P. and Mergny,J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858–7868.
20. Cheng,M., Cheng,Y., Hao,J., Jia,G., Zhou,J., Mergny,J.L. and Li,C. (2018) Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.*, **46**, 9264–9275.
21. Chen,J., Cheng,M., Stadlbauer,P., Šponer,J., Mergny,J.-L., Ju,H. and Zhou,J. (2021) Exploring sequence space to design controllable G-quadruplex topology switches. *CCS Chem.*, **3**, 3232–3246.
22. Arora,A. and Maiti,S. (2009) Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. *J. Phys. Chem. B*, **113**, 8784–8792.
23. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
24. 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.
25. Dai,J., Carver,M. and Yang,D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.
26. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
27. Warshaw,M.M. and Tinoco,I. (1966) Optical properties of sixteen dinucleoside phosphates. *J. Mol. Biol.*, **20**, 29–38.
28. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
29. Fogh,R.H., Boucher,W., Vranken,W.F., Pajon,A., Stevens,T.J., Bhat,T.N., Westbrook,J., Ionides,J.M.C. and Laue,E.D. (2005) A framework for scientific data modeling and automated software development. *Bioinformatics*, **21**, 1678–1684.
30. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Bioinforma.*, **59**, 687–696.
31. Schwieters,C.D., Kuszewski,J.J. and Clore,G.M. (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.*, **48**, 47–62.
32. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
33. Galindo-Murillo,R., Robertson,J.C., Zgarbová,M., Šponer,J., Otyepka,M., Jurečka,P. and Cheatham,T.E. (2016) Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.*, **12**, 4114–4127.
34. Vanqualef,E., Simon,S., Marquant,G., Garcia,E., Klimerak,G., Delepine,J.C., Cieplak,P. and Dupradeau,F.Y. (2011) R.E.D. server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, **39**, W511–W517.
35. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
36. Li,X.M., Zheng,K.W., Zhang,J.Y., Liu,H.H., He,Y.De, Yuan,B.F., Hao,Y.H. and Tan,Z. (2015) Guanine-vacancy-bearing G-quadruplexes responsive to guanine derivatives. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14581–14586.
37. Winnerdy,F.R., Das,P., Heddi,B. and Phan,A.T. (2019) Solution structures of a G-quadruplex bound to linear- and cyclic-dinucleotides. *J. Am. Chem. Soc.*, **141**, 18038–18047.
38. Wang,K.B., Dickerhoff,J., Wu,G. and Yang,D. (2020) PDGFR-β promoter forms a vacancy G-quadruplex that can be filled in by dGMP: solution structure and molecular recognition of guanine metabolites and drugs. *J. Am. Chem. Soc.*, **142**, 5204–5211.
39. Chen,Y., Agrawal,P., Brown,R.V., Hatzakis,E., Hurley,L. and Yang,D. (2012) The major G-quadruplex formed in the human platelet-derived growth factor receptor β promoter adopts a novel broken-strand structure in K⁺ solution. *J. Am. Chem. Soc.*, **134**, 13220–13223.
40. Jana,J., Mohr,S., Vianney,Y.M. and Weisz,K. (2021) Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.*, **2**, 338–353.
41. Fogolari,F., Haridas,H., Corazza,A., Viglino,P., Corà,D., Caselle,M., Esposito,G. and Xodo,L.E. (2009) Molecular models for intrastand DNA G-quadruplexes. *BMC Struct. Biol.*, **9**, 64.
42. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
43. Webba da Silva,M. (2007) Geometric formalism for DNA quadruplex folding. *Chem. Eur. J.*, **13**, 9738–9745.
44. Dvorkin,S.A., Karsisiotis,A.I. and da Silva,M.W. (2018) Encoding canonical DNA quadruplex structure. *Sci. Adv.*, **4**, eaat3007.
45. 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.
46. Lim,K.W., Ng,V.C.M., Martin-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.
47. Marušič,M. and Plavec,J. (2019) Towards understanding of polymorphism of the G-rich region of human papillomavirus type 52. *Molecules*, **24**, 1294.
48. Dickerhoff,J., Onel,B., Chen,L., Chen,Y. and Yang,D. (2019) Solution structure of a MYC promoter G-quadruplex with 1:6:1 loop length. *ACS Omega*, **4**, 2533–2539.
49. Sattin,G., Artese,A., Nadai,M., Costa,G., Parrotta,L., Alcaro,S., Palumbo,M. and Richter,S.N. (2013) Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One*, **8**, e84113.
50. Mathad,R.I., Hatzakis,E., Dai,J. and Yang,D. (2011) c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res.*, **39**, 9023–9033.
51. Smargiasso,N., Rosu,F., Hsia,W., Colson,P., Baker,E.S., Bowers,M.T., De Pauw,E. and Gabelica,V. (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, **130**, 10208–10216.
52. Karg,B. and Weisz,K. (2018) Loop length affects syn - anti conformational rearrangements in parallel G-quadruplexes. *Chem. Eur. J.*, **24**, 10246–10252.
53. Mohr,S., Jana,J., Vianney,Y.M. and Weisz,K. (2021) Expanding the topological landscape by a G-column flip of a parallel G-quadruplex. *Chem. – A Eur. J.*, **27**, 10437–10447.
54. Karg,B., Mohr,S. and Weisz,K. (2019) Duplex-guided refolding into novel G-quadruplex (3+1) hybrid conformations. *Angew. Chem. Int. Ed.*, **58**, 11068–11071.
55. Cang,X., Šponer,J. and Cheatham,T.E. III (2011) Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.*, **39**, 4499–4512.
56. Šponer,J., Mládek,A., Špačková,N., Cang,X., Cheatham,T.E. III and Grimme,S. (2013) Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *J. Am. Chem. Soc.*, **135**, 9785–9796.
57. Sengar,A., Vandana,J.J., Chambers,V.S., Di Antonio,M., Winnerdy,F.R., Balasubramanian,S. and Phan,A.T. (2019) Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. *Nucleic Acids Res.*, **47**, 1564–1572.
58. Agrawal,P., Hatzakis,E., Guo,K., Carver,M. and Yang,D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.*, **41**, 10584–10592.
59. De Armond,R., Wood,S., Sun,D., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming

- region within a polypurine tract of the hypoxia inducible factor 1 α promoter. *Biochemistry*, **44**, 16341–16350.
60. Fernando,H., Reszka,A.P., Huppert,J., Ladame,S., Rankin,S., Venkitaraman,A.R., Neidle,S. and Balasubramanian,S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*, **45**, 7854–7860.
61. Agrawal,P., Lin,C., Mathad,R.I., Carver,M. and Yang,D. (2014) The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K⁺ solution. *J. Am. Chem. Soc.*, **136**, 1750–1753.
62. Lim,K.W., Jenjaroenpun,P., Low,Z.J., Khong,Z.J., Ng,Y.S., Kuznetsov,V.A. and Phan,A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630–5646.

Supplementary Information

for

Guiding the Folding of G-Quadruplexes Through Loop Residue Interactions

Jagannath Jana, Yoanes Maria Vianney, Nina Schröder and Klaus Weisz*
Institute of Biochemistry, Universität Greifswald, 17489 Greifswald, Germany

Contents

NMR spectral analysis and resonance assignments for ⁶ BrQ	S2
NMR spectral analysis of <i>Qref</i>	S8
NMR spectral analysis and resonance assignments for ⁵ TQ	S10
Structural statistics and NMR-derived structures of ⁶ BrQ and ⁵ TQ	S15
NMR spectral analysis for <i>Q-5T</i> and <i>Q-11T</i>	S18
CD spectra of all sequences with a 331 loop length arrangement	S22
Imino proton NMR spectral region of all modified sequences with a 331 loop length arrangement	S25
Thermodynamic analysis for modified sequences with a 331 loop length arrangement	S27
CD spectra of sequences with a 311 loop length arrangement	S29
Non-denaturing polyacrylamide gel electrophoresis of all quadruplexes	S30
Comparative ¹ H- ¹³ C HSQC spectra of ⁶ BrQ, ¹⁶ BrQ, and ¹⁴ BrQ-311- <i>T</i>	S31
Melting analysis of major and minor ¹⁶ BrQ quadruplex	S32
Thermodynamics of ¹⁰ BrQ-311- <i>T</i> and ¹⁰ BrQ-311- <i>X</i> quadruplex formation	S33
Schematic representation of (3+1) hybrid topologies	S34
NMR spectral analysis and resonance assignments for ² BrQ-311- <i>T</i>	S35
Structural statistics and NMR-derived structure of ² BrQ-311- <i>T</i>	S40
Determination of coexisting G4 species with relative populations for <i>Q-311-T</i> and <i>Qref</i>	S42
Comparative ¹ H- ¹³ C HSQC spectra of ² BrQ and ² BrQ-311- <i>T</i>	S44

NMR spectral analysis and resonance assignments for ^{6Br}Q

The ^{6Br}Q sequence exhibits 12 well-resolved imino proton resonances, indicative of a single G-quadruplex conformer with three G-tetrad layers. Guanine imino and H8 protons were unambiguously assigned without specific isotope labeling through inter-residual NOE contacts as well as through ¹H–¹³C HSQC and ¹H–¹³C HMBC experiments at natural abundance. Four *syn*-guanosines were assigned based on their strong H8–H1' cross-peak intensities in 2D NOESY spectra (increasingly pronounced at shorter mixing times) and corroborated by their *syn*-typical chemical shift in a ¹H–¹³C HSQC spectrum (Figures S1B and S2). Another yet non-identified *syn*-G must come from the *syn*-favoring 8-bromoguanosine analog ^{Br}G at position 6, yielding a total of five *syn*-G residues at positions G1, G6, G12, G16, and G22 as well as seven G-core residues with an *anti*-conformation in the G-quadruplex structure. Likewise, H6/H8–H1' and H6/H8–H3' spectral regions of 2D NOESY spectra acquired with longer mixing times (300 ms) showed a rectangular pattern of intra-nucleotide and sequential crosspeaks for G1–G2, G12–G13, and G16–G17, typical for *syn-anti* steps (Figure S1B). Continuous base-sugar NOE walks only interrupted by the 3-nt and 1-nt propeller loop allowed for additional sequential assignments also including lateral and snapback loop residues. Thus, spectral analysis of non-exchangeable protons suggests formation of a three-layered G-quadruplex composed of three *syn-anti-anti* columns (G6-G7-G8, G12-G13-G14, and G16-G17-G18, as well as one broken *syn-syn-anti* column G22-G1-G2).

Following assignments of non-labile protons, guanine imino protons were unambiguously assigned based on their intra-residue correlations to H8 in ¹H–¹³C HMBC spectra (Figure S3). A strong NOE crosspeak observed between G1 H1 and G22 H1 indicates the proximity of 5'- and 3'-terminal G residues within the first broken G-column (Figure S1C). The determination of tetrad polarities was enabled by characteristic H8–H1 intra-tetrad NOE contacts (Figure S1D). Hoogsteen hydrogen bonds within the three G-quartets run along G2-G16-G12-G6, G1-G7-G13-G17, and G22-G8-G14-G18 with one homopolar and one heteropolar tetrad stacking in line with the CD spectral signature typical of a (3+1) hybrid structure (see Figure 3 of the main manuscript). Finally, additional confirmation of the G-quadruplex fold comes from H₂O–D₂O exchange experiments, demonstrating increased exchange rates for imino protons located within the 5'- and 3'-outer tetrads when compared to central tetrad iminos (Figure S4).

Stereospecific assignments for H2'/H2'' sugar protons were accomplished by analyzing NOESY experiments at short mixing times to allow discrimination of the different H1'–H2' and H1'–H2'' NOE crosspeak intensities. Sugar puckers were evaluated through a comparison of H1'–H2' and H1'–H2'' DQF-COSY crosspeaks, making use of cancellation effects of DQF-COSY anti-phase crosspeak components in case of smaller coupling constants (Figure S5).

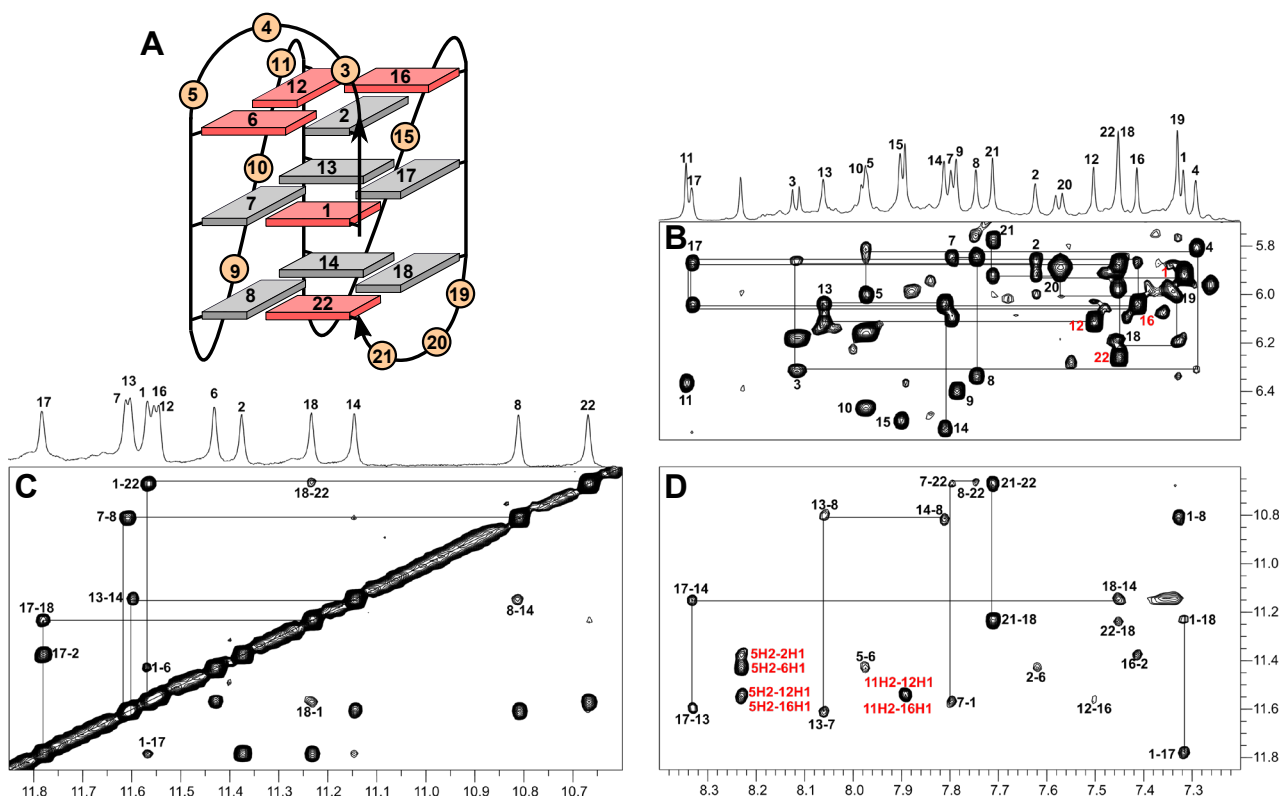


Figure S1. Topology and 2D NOESY spectral regions of ${}^6\text{BrQ}$ in 10 mM K^+ buffer, pH 7.0 (30 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a (3+1) hybrid-type G-quadruplex with a (+lpp) topology and snapback loop adopted by ${}^6\text{BrQ}$; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral region tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeaks of *syn*-guanosines are labelled in red. (C) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (D) H8/H2(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities; connectivities of A5H2 and A11H2 with G imino protons of the upper tetrad (labelled in red) indicate stacking of 3'-terminal loop residues A5 and A11 onto the tetrad.

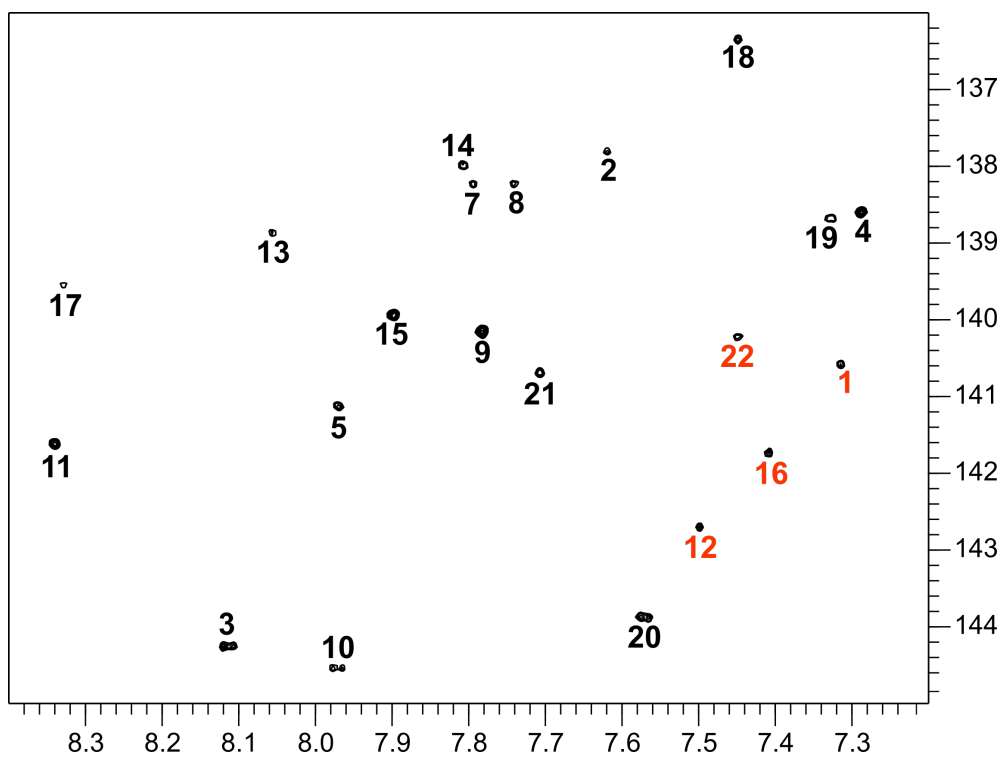


Figure S2. ^1H - ^{13}C HSQC spectrum of ^6BrQ acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. Crosspeaks of *syn*-guanosines are labelled in red.

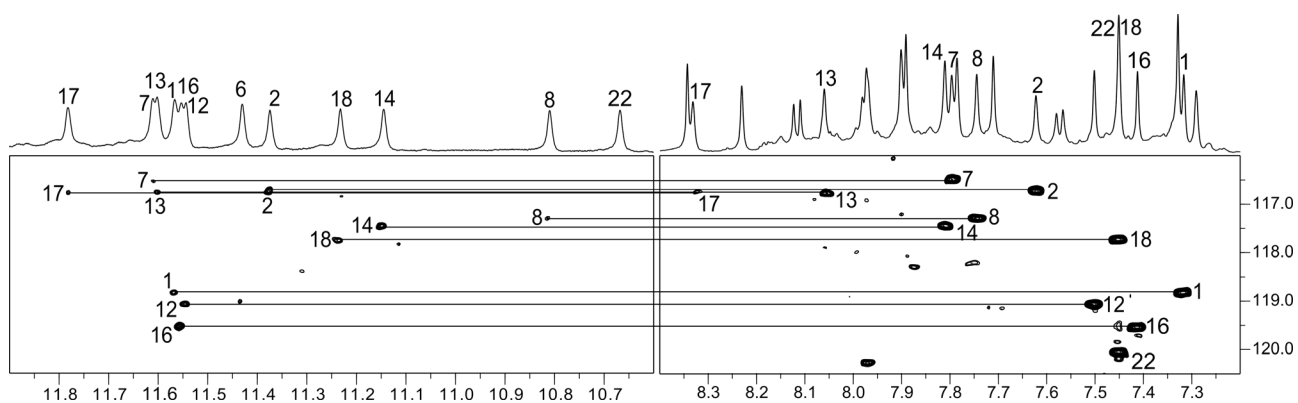


Figure S3. ^1H - ^{13}C HMBC spectrum of ^6BrQ at 30 °C in 10 mM K^+ buffer, pH 7.0, showing through-bond correlations of guanine H1(ω_2) and H8(ω_2) protons via long-range couplings to $^{13}\text{C}5$ (ω_1) at natural abundance.

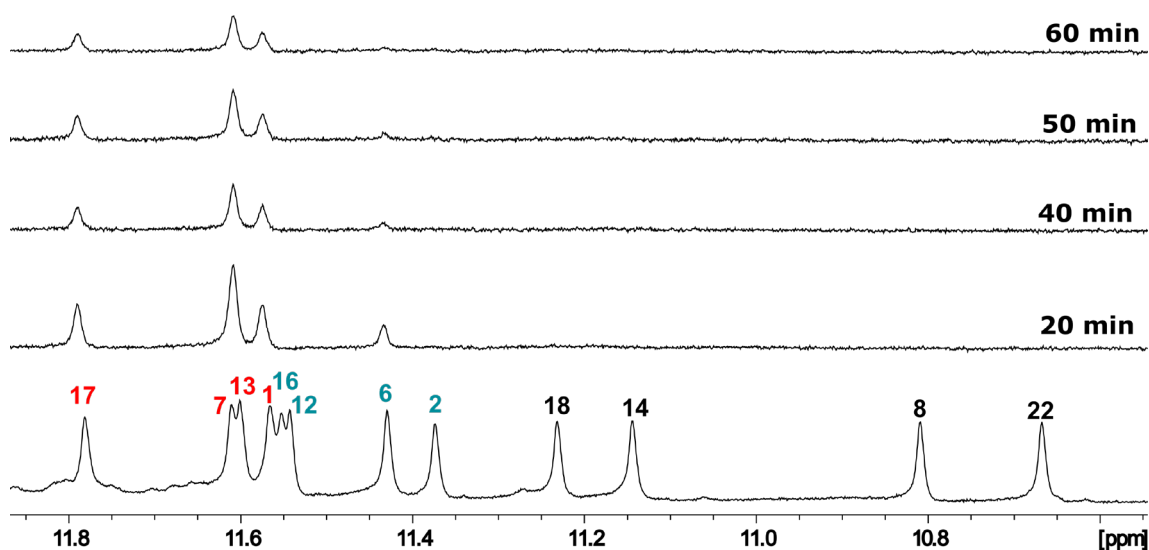


Figure S4. H_2O - D_2O exchange experiments. Imino proton spectral region of ^6BrQ acquired in 10 mM K^+ buffer with 90% H_2O /10% D_2O at 30 °C (bottom) and at increasing time intervals after drying and redissolving the oligonucleotide in 100% D_2O (top). Imino protons of residues located in the central tetrad and, albeit to a smaller extent, of G6 are protected from fast solvent exchange. Numbers of residues in the lower, central, and upper tetrad are labeled in black, red, and blue, respectively.

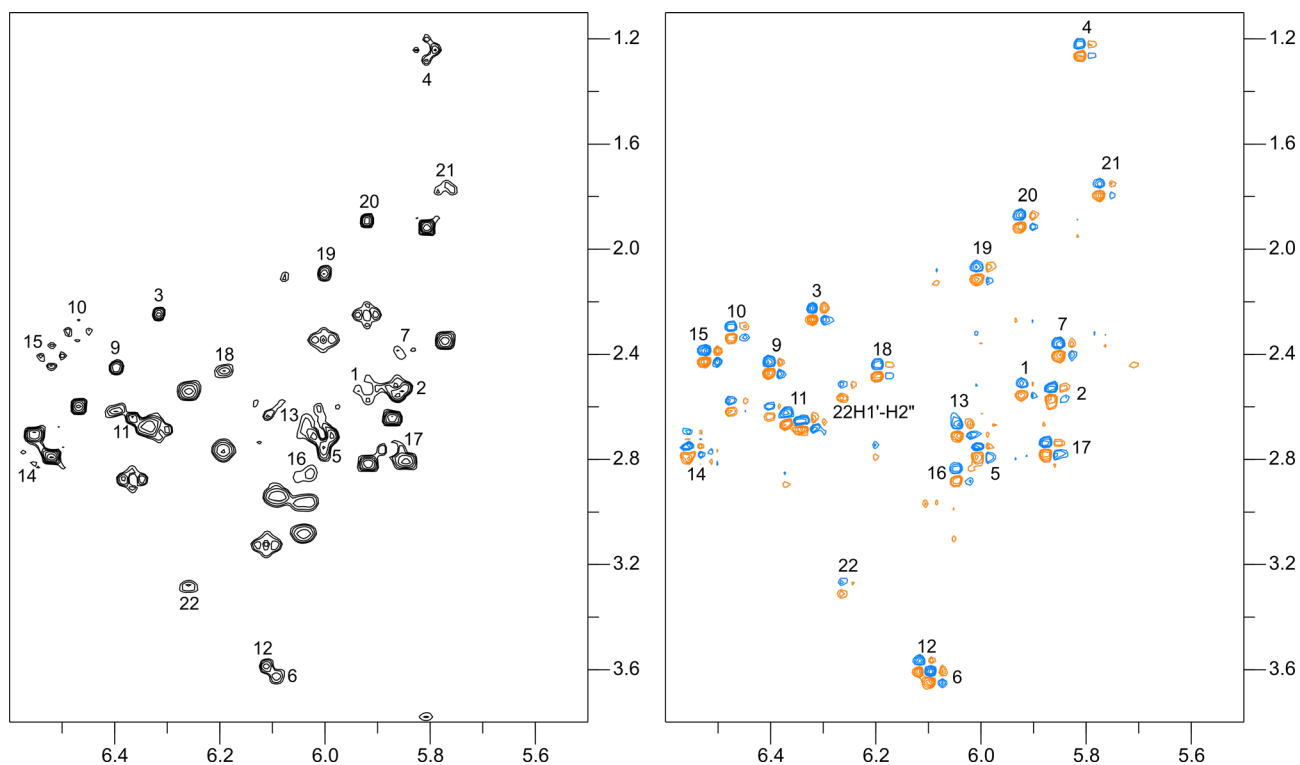


Figure S5. Sugar pucker analysis of ${}^6\text{BrQ}$. (Left) Stereospecific assignments of H2'/H2'' with H1'(ω_2)-H2'/H2''(ω_1) NOESY spectral region at short mixing time (80 ms); crosspeak intensities allow discrimination between H2' and H2''. (Right) DQF-COSY spectral region showing H1'(ω_2)-H2'/H2''(ω_1) crosspeaks; *north*- and *south*-type sugar puckers are associated with different scalar couplings and thus different crosspeak patterns of in-phase and anti-phase components.

Table S1. ^1H and ^{13}C chemical shifts δ of $^{6\text{Br}}\text{Q}$.^a

δ (ppm)	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
G1	7.32	11.57	5.92	2.54/2.82	4.98	-	140.59	-
G2	7.62	11.37	5.86	2.54/2.55	5.13	-	137.81	-
C3	8.11	-	6.31	2.25/2.69	4.90	6.18	144.28	-
T4	7.29	n.d.	5.80	1.24/1.92	4.70	1.71	138.62	-
A5	7.97	-	6.00	2.77/2.72	5.00	8.23	141.15	156.16
^{Br} G6	-	11.43	6.09	3.63/2.95	4.95	-	n.d.	-
G7	7.80	11.61	5.85	2.38/2.81	4.92	-	138.23	-
G8	7.74	10.82	6.34	2.66/2.67	5.03	-	138.23	-
T9	7.78	n.d.	6.40	2.45/2.62	4.83	1.96	140.16	-
C10	7.97	-	6.47	2.31/2.60	4.97	6.16	144.52	-
A11	8.34	-	6.36	2.65/2.88	5.01	7.89	141.61	155.01
G12	7.50	11.54	6.11	3.59/3.13	5.03	-	142.71	-
G13	8.06	11.60	6.04	2.69/2.97	5.12	-	138.88	-
G14	7.81	11.15	6.55	2.77/2.70	5.17	-	137.99	-
T15	7.90	n.d.	6.52	2.41/2.79	5.13	2.00	139.97	-
G16	7.41	11.56	6.04	2.86/3.08	5.12	-	141.73	-
G17	8.33	11.78	5.87	2.76/2.64	5.12	-	139.55	-
G18	7.45	11.23	6.19	2.46/2.77	4.89	-	136.35	-
T19	7.33	n.d.	6.00	2.10/2.34	4.68	1.70	138.68	-
C20	7.57	-	5.92	1.89/2.25	4.56	5.88	143.89	-
A21	7.71	-	5.77	1.78/2.35	4.69	7.33	140.70	153.10
G22	7.45	10.67	6.26	3.29/2.54	4.87	-	140.22	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

NMR spectral analysis of *Qref*

Similar NOE crosspeak patterns of *Qref* and $6BrQ$ in the H6/H8-H1' NOESY spectral region and in the H8/H6-C8/C6 region of 1H - ^{13}C HSQC spectra clearly suggest the formation of a corresponding (+lpp) hybrid-type G-quadruplex with snapback loop as major species formed by unmodified *Qref* (Figures S6 and S7).

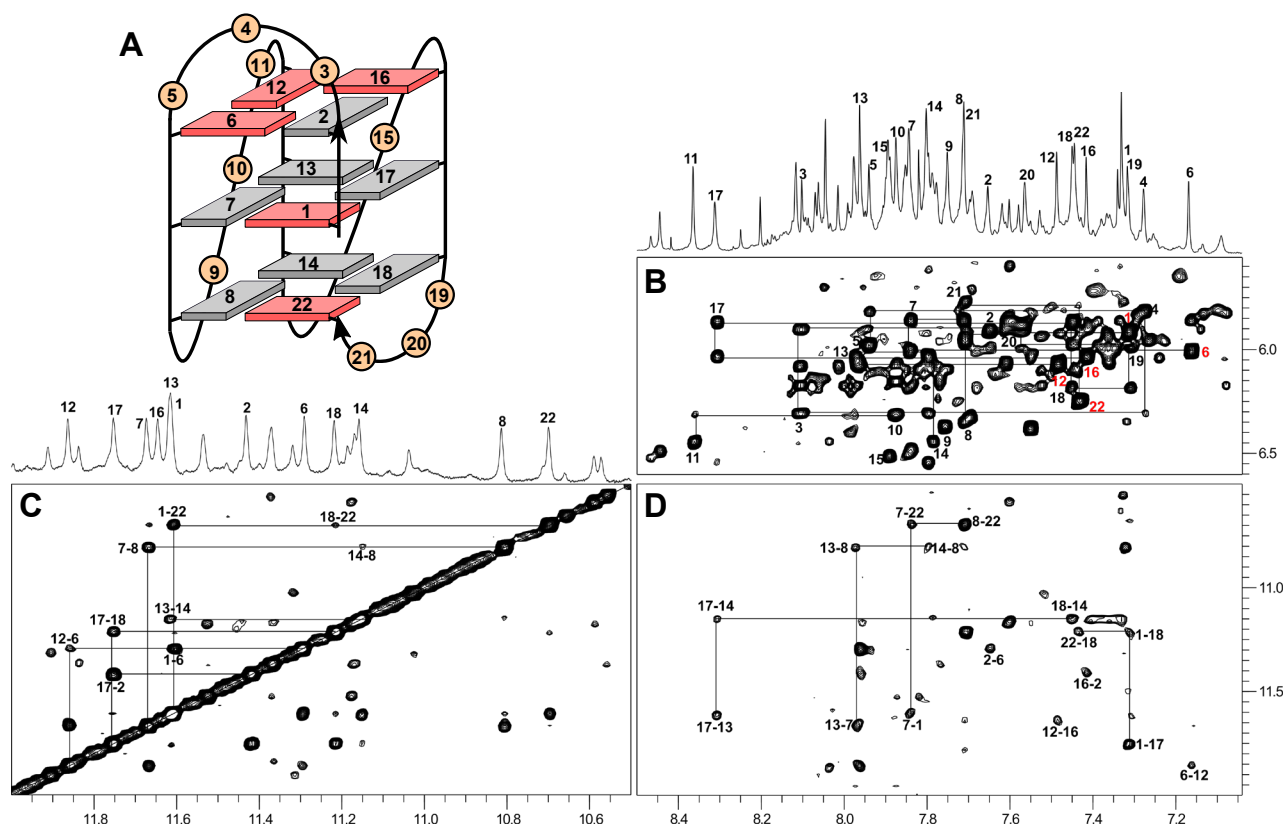


Figure S6. Topology and 2D NOESY spectral regions of *Qref* in 10 mM K^+ buffer, pH 7.0 (30 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a (3+1) hybrid-type G-quadruplex with a (+lpp) topology and snapback loop adopted by *Qref*; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral region tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeaks of *syn*-guanosines are labelled in red. (C) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (D) H8(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities.

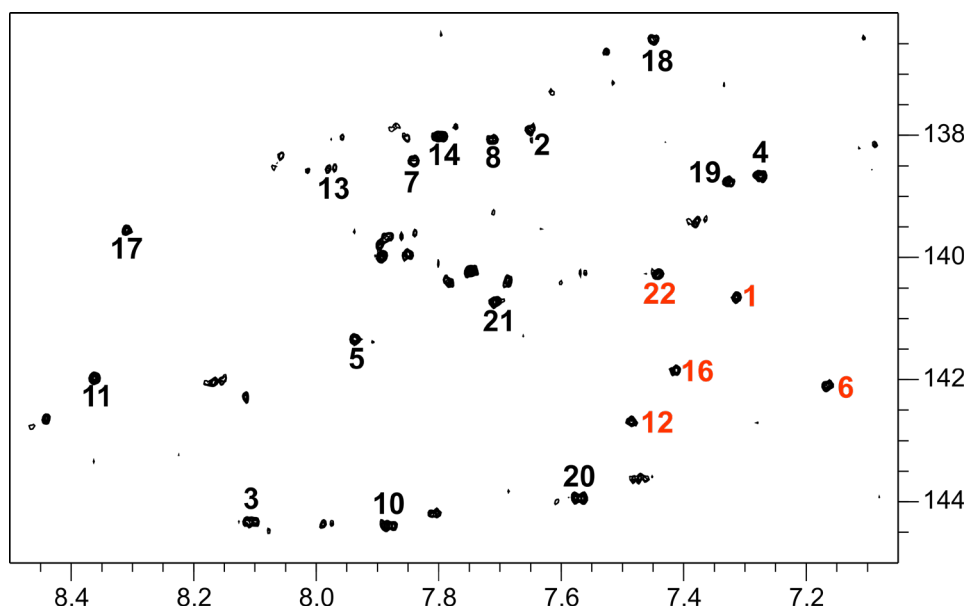


Figure S7. ^1H - ^{13}C HSQC spectrum of *Qref* acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. Crosspeaks of *syn*-guanosines are labelled in red.

NMR spectral analysis and resonance assignments for ⁵TQ

Upon the addition of a 5'-thymidine to *Qref*, a single three-layered quadruplex with a set of 12 G imino proton signals can be observed (Figure S8). The G-core is comprised of a single *syn*- and eleven *anti*-guanosines as suggested by H6/H8-sugar NOE contacts, NOESY crosspeak intensities, and C8 chemical shifts as observed in a ¹H-¹³C HSQC spectrum (Figures S8B and S9). The G-quadruplex has a parallel topology with a vacant site in the first G-tract that is filled by G22 through a lateral snapback loop composed of residues 19-21. An uninterrupted NOE walk can be followed from residue G16 to the 3'-terminal G22, identifying the fourth G-column. The first G-column can be identified through a contact to the T(-1) 5'-overhang and is formed by *anti*-G1 followed by *anti*-G2 and *syn*-G22. The second G-column starting at G6 is discriminated from the third G-tract by sequential contacts between preceding A5 to G6. G-columns G6-G7-G8, G12-G13-G14, and G16-G17-G18 exclusively comprise *anti-anti-anti* steps. Imino protons for all guanines were determined through ¹H-¹³C HMBC spectra, correlating H8 and H1 via ¹³C5 (Figure S10). Finally, homopolarity of all G-tetrads is confirmed by looking at characteristic patterns of GH8-imino and imino-imino NOE contacts (Figure S8C and D). Thus, the direction of Hoogsteen hydrogen bonds within tetrads points along G1-G6-G12-G16, G2-G7-G13-G17, and G22-G8-G14-G18.

The four intervening sequences of the quadruplex form two 3-nt propeller loops, one 1-nt propeller loop and one 3-nt lateral snapback loop. A21 of the lateral snapback loop is expected to stack onto the outer tetrad as indicated by various NOE contacts between adenine aromatic protons to guanine imino protons. The third base A5 of the first 3-nt propeller loop seems to be located below the 5'-outer tetrad with several NOE contacts between adenosine A5 protons and 5'-tetrad guanines including A5H1'-G6H8, A5H8-G1H1, A5H2-G1H1, and A5H2-G6H1. In contrast, 3'-residue A11 of the second 3-nt propeller loop seems to only make contacts with residues of the propeller loop itself (Figure S8) except for a single weak crosspeak observed between A11 H2 and G13 H8 located in the central tetrad (not shown). It should be noted, however, that the first propeller loop only bridges two tetrad layers.

Sugar conformations were determined through an analysis of H1'-H2' and H1'-H2'' DQF-COSY crosspeaks following a stereospecific assignment of H2'/H2'' sugar protons through NOESY experiments at short mixing times (Figure S11). All residues except for G8, G9, A21, and G22 were assigned a *south* conformation. A21 was assigned a *north* conformation due to its strong H1'-H2'' DQF-COSY crosspeak. Whereas G8 and T9 sugar puckers could not be evaluated due to their isochronous H2'/H2'' resonances, puckering of the G22 sugar remained ambiguous.

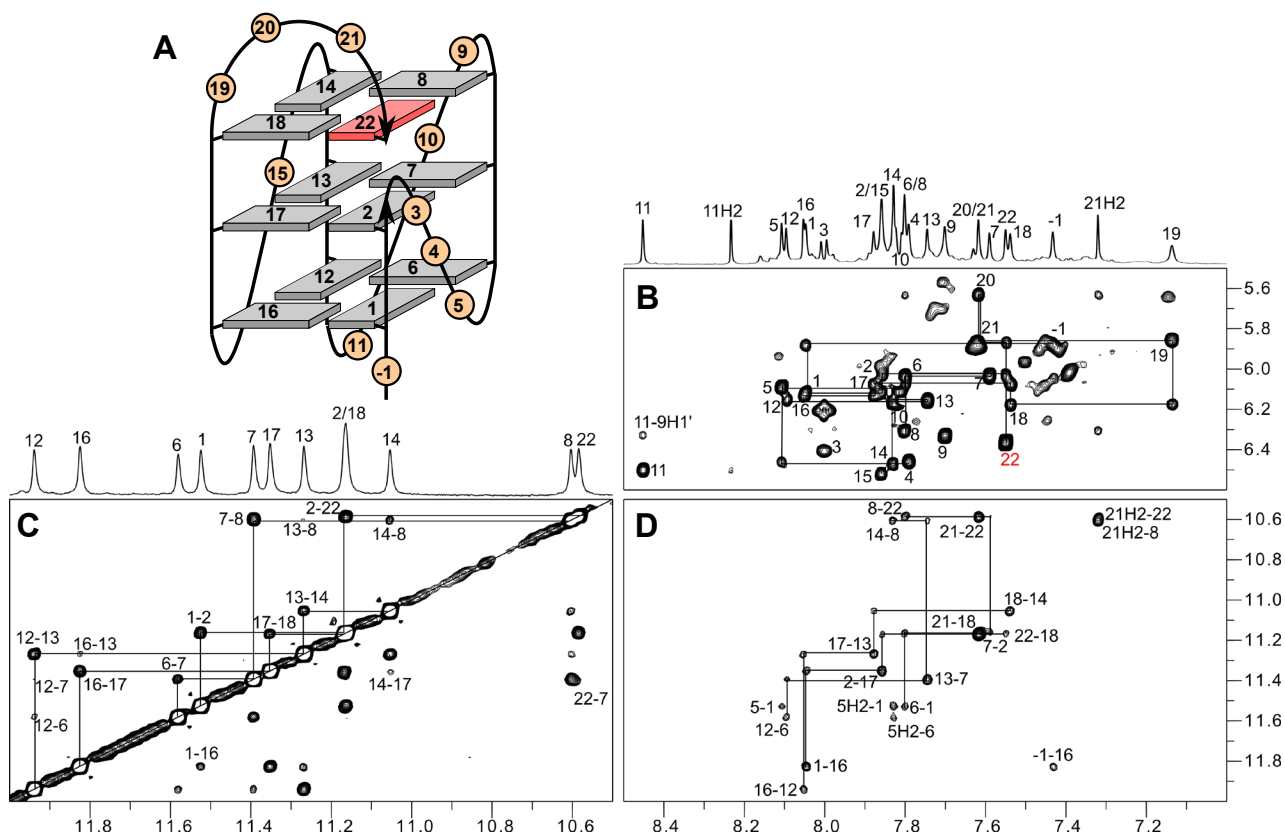


Figure S8. Topology and 2D NOESY spectral regions of $5TQ$ in 10 mM K^+ buffer, pH 7.0 (30 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a parallel G-quadruplex with snapback loop adopted by $5TQ$; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral region tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeak of *syn*-G22 is labelled in red. (C) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (D) H8/H2(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities.

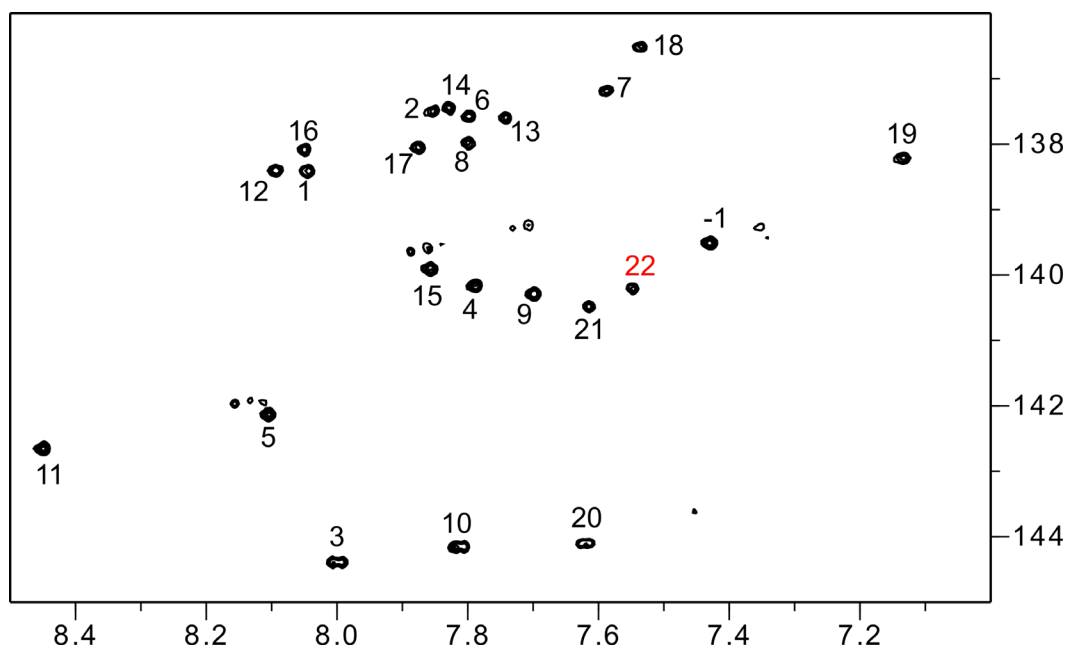


Figure S9. ^1H - ^{13}C HSQC spectrum of ^{57}Q acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. Crosspeak of *syn*-G22 is labelled in red.

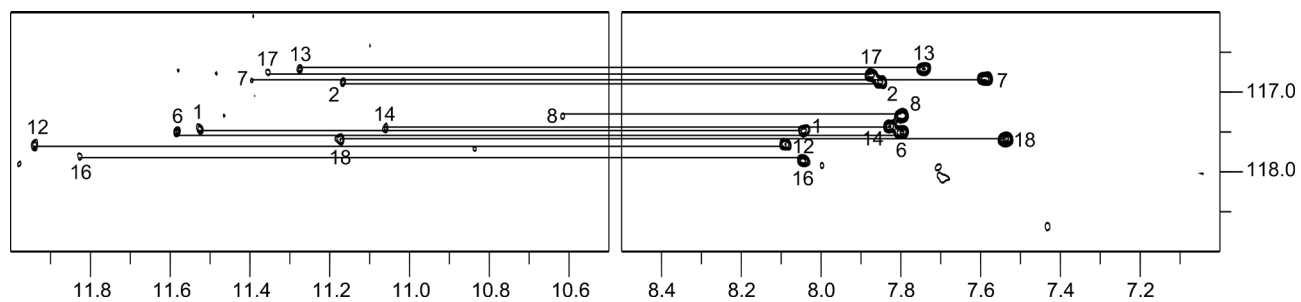


Figure S10. ^1H - ^{13}C HMBC spectrum of ^{57}Q at 30 °C in 10 mM K^+ buffer, pH 7.0, showing through-bond correlations of guanine H1(ω_2) and H8(ω_2) protons via long-range couplings to $^{13}\text{C}5$ (ω_1) at natural abundance.

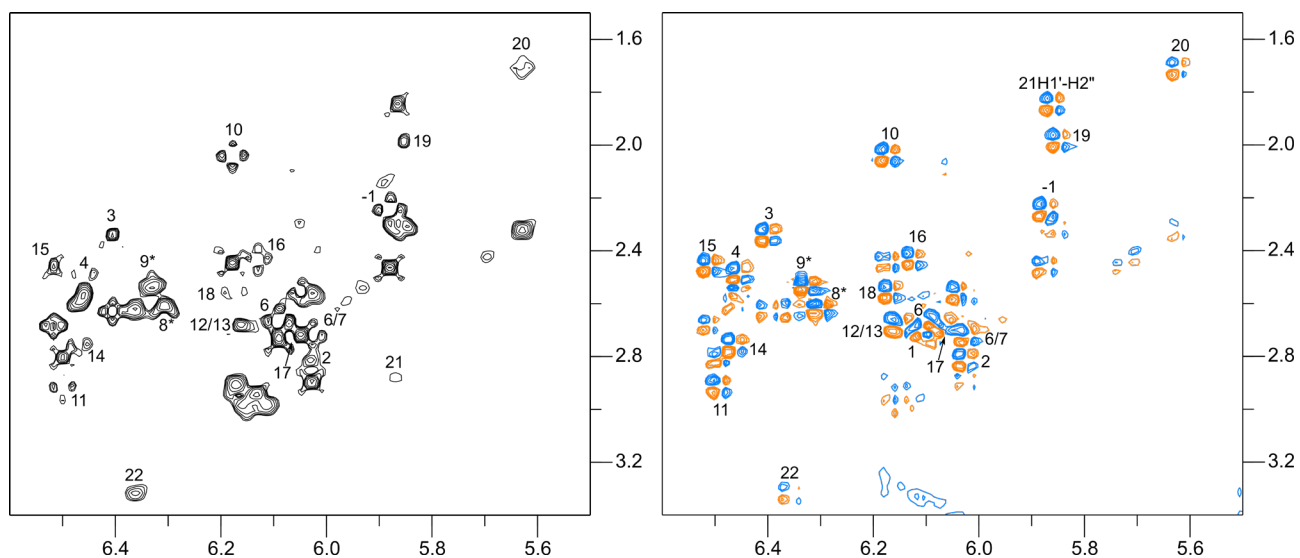


Figure S11. Sugar pucker analysis of 5TQ . (Left) Stereospecific assignments of H2'/H2'' with H1'(ω_2)-H2'/H2''(ω_1) NOESY spectral region at short mixing time (80 ms); crosspeak intensities allow discrimination between H2' and H2''. (Right) DQF-COSY spectral region showing H1'(ω_2)-H2'/H2''(ω_1) crosspeaks; *north*- and *south*-type sugar puckers are associated with different scalar couplings and thus different crosspeak patterns of in-phase and anti-phase components.

Table S2. ¹H and ¹³C chemical shifts δ of ⁵⁷Q.^a

δ (ppm)	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T-1	7.43	n.d.	5.88	2.25/2.47	4.74	1.43	139.52	-
G1	8.05	11.53	6.12	2.70/2.98	5.01	-	138.41	-
G2	7.86	11.16	6.03	2.82/2.57	5.09	-	137.50	-
C3	8.00	-	6.40	2.34/2.63	4.78	6.20	144.40	-
T4	7.79	n.d.	6.46	2.49/2.57	4.95	2.03	140.16	-
A5	8.11	-	6.09	2.66/2.74	4.96	7.83	142.15	154.76
G6	7.80	11.58	6.03	2.72/2.90	5.00	-	137.57	-
G7	7.59	11.40	6.03	2.72/2.56	4.88	-	137.17	-
G8	7.80	10.61	6.31	2.62/2.62	4.99	-	137.98	-
T9	7.70	n.d.	6.33	2.54/2.54	4.91	2.00	140.29	-
C10	7.82	-	6.18	2.05/2.45	4.81	6.11	144.15	-
A11	8.45	-	6.50	2.91/2.81	5.10	8.23	142.66	155.41
G12	8.09	11.94	6.16	2.68/2.99	5.00	-	138.40	-
G13	7.74	11.27	6.17	2.68/2.95	5.02	-	137.60	-
G14	7.83	11.06	6.47	2.76/2.60	5.14	-	137.44	-
T15	7.86	n.d.	6.52	2.46/2.68	5.12	1.98	139.92	-
G16	8.05	11.83	6.13	2.43/2.94	5.15	-	138.08	-
G17	7.88	11.36	6.07	2.71/2.68	5.08	-	138.05	-
G18	7.54	11.17	6.17	2.56/2.91	4.95	-	136.50	-
T19	7.13	n.d.	5.85	1.99/2.32	4.69	1.61	138.21	-
C20	7.62	-	5.63	1.71/2.32	4.68	5.88	144.11	-
A21	7.62	-	5.87	2.89/1.85	4.51	7.32	140.48	152.81
G22	7.55	10.59	6.36	3.32/2.63	4.84	-	140.21	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

Table S3. NMR restraints and structural statistics of calculated structures.

sequence	⁶ BrQ	⁵ TQ
NOE distance restraints		
intra-residual	86	92
inter-residual	122	141
exchangeable	41	53
repulsion	6	0
other restraints:		
hydrogen bonds	48	48
dihedral angles	41	43
planarity	3	3
structural statistics:		
pairwise heavy atom RMSD value (Å)		
all residues	2.0 ± 0.6	2.5 ± 0.6
G-tetrad core	0.7 ± 0.1	0.9 ± 0.2
NOE violations:		
maximum violation (Å)	0.16	0.17
mean NOE violation (Å)	0.002 ± 0.001	0.002 ± 0.001
deviations from idealized geometry:		
bond lengths (Å)	0.01 ± 0.0001	0.01 ± 0.0001
bond angles (degree)	2.3 ± 0.03	2.2 ± 0.04

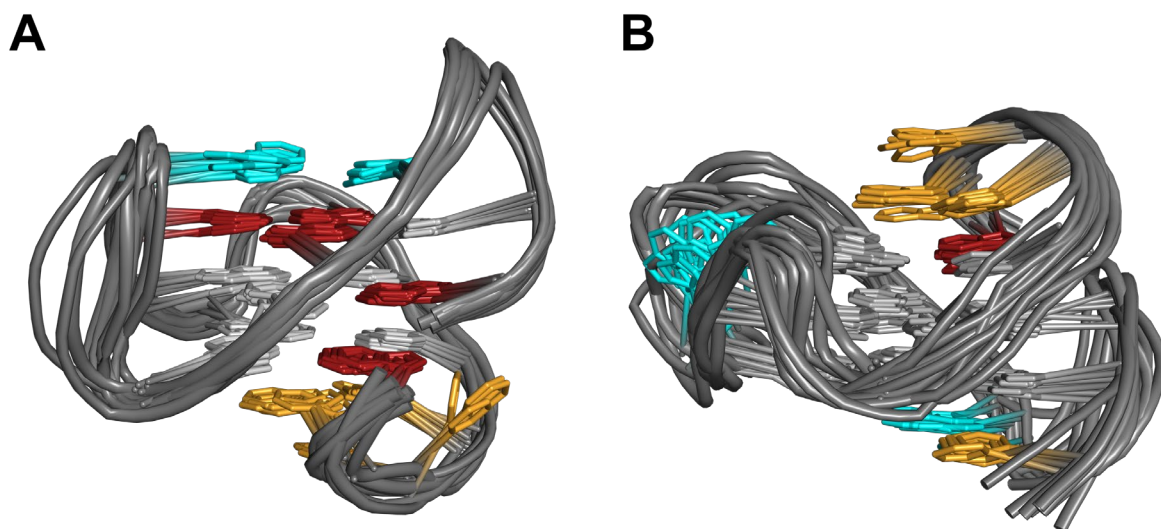


Figure S12. Superposition of ten lowest-energy structures for (A) $6BrQ$ and (B) $5TQ$; loop residues C3, T4, T9, C10, and T15 are omitted for clarity; *anti*- and *syn*-guanosines are colored in grey and red, respectively; A5 and A11 are colored in cyan and T(-1) in $5TQ$ as well as snapback lateral loop residues at the bottom (A) and top (B) are colored in orange.

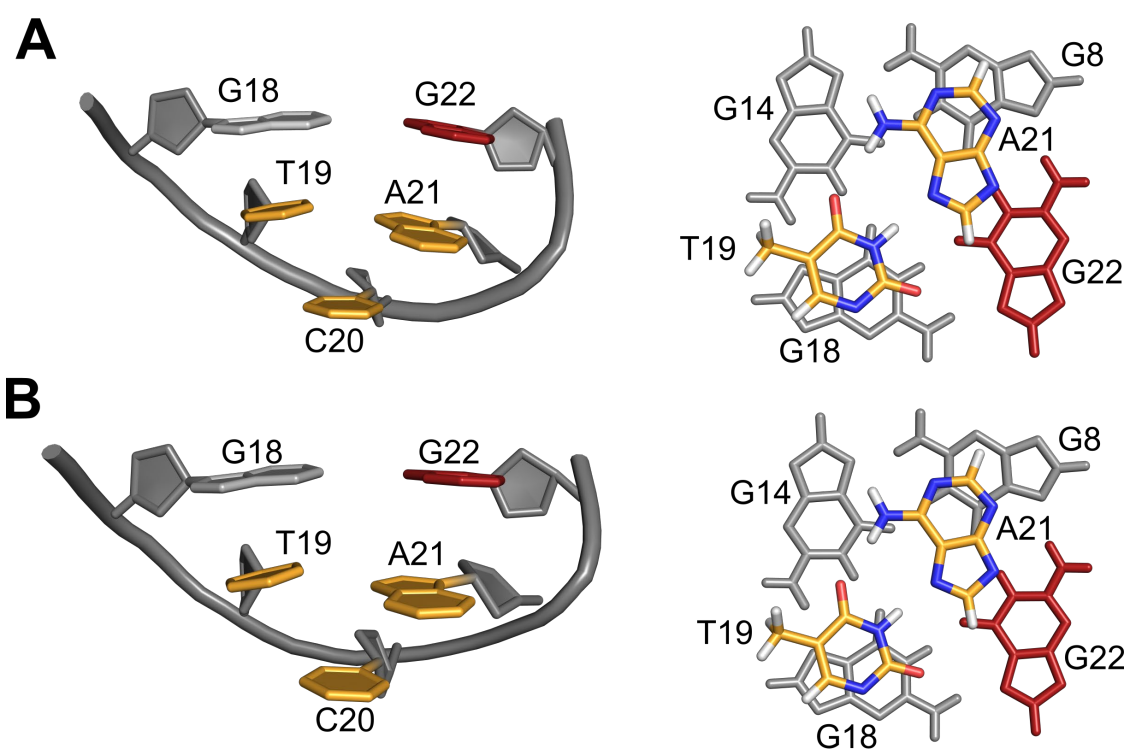


Figure S13. TCA lateral snapback loop of (A) $6BrQ$ and (B) $5TQ$ in side view (left) and top view (right) showing stacking of the Hoogsteen T19-A21 base pair onto the outer G-tetrad. C20 is additionally stacked onto the base pair (not shown in top view).

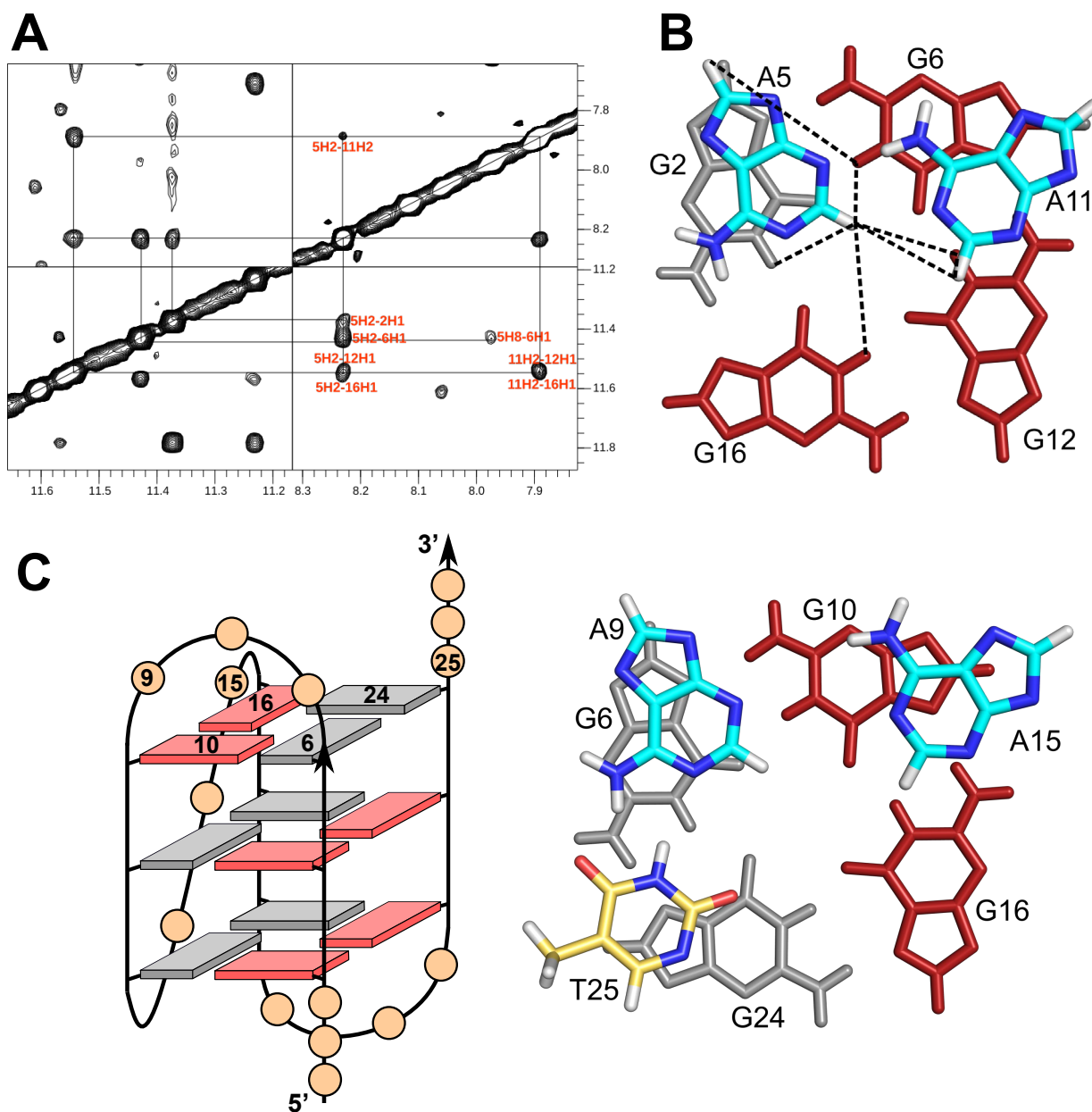


Figure S14. (A) 2D NOESY spectral regions of ${}^6\text{BrQ}$, highlighting some key NOE crosspeaks that define the position of A5 in the first lateral loop and of A11 in the second propeller loop above the outer tetrad. (B) Top view of the ${}^6\text{BrQ}$ outer tetrad with capping loop residues; A5 stacks over G2 and forms a putative hydrogen bond with the A11 amino proton; experimentally observed NOE contacts are indicated by dotted lines. (C) Schematic representation of a $+(lp)$ G-quadruplex formed by a derivative of a 27-nt human telomeric sequence in Na^+ solution (right); its solution structure shows a corresponding positioning of A9 in the first lateral loop and A15 in the second propeller loop, both capping the upper tetrad (left); also shown is T25 of the 3'-flanking sequence forming an additional Watson-Crick base pair with A9 (PDB 2MBJ).

NMR spectral analysis for *Q-5T* and *Q-11T*

Assignments for *Q-5T* revealed its folding into a major parallel species with a snapback loop as evident by eleven *anti*-Gs and one 3'-terminal *syn*-G filling the vacant position of the G-tetrad (Figures S15 and S16). Interestingly, assignments for *Q-11T* clearly demonstrated its alternate folding into a major +(lpp) hybrid fold with crosspeak patterns very similar to ^{6Br}Q and *Qref* (Figures S17 and S18). These results indicate the critical role of A5 but not of A11 as contributor for a major +(lpp) fold.

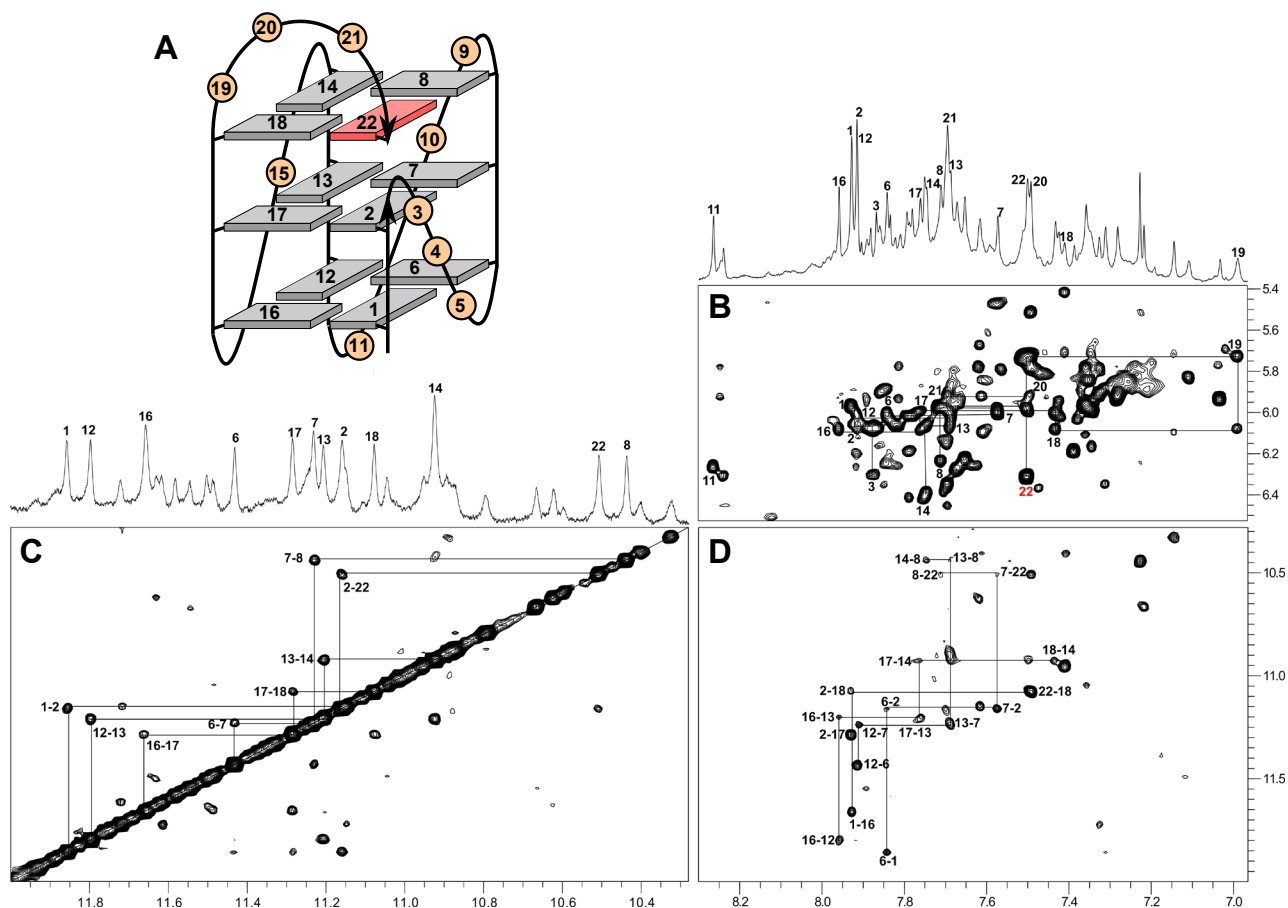


Figure S15. Topology and 2D NOESY spectral regions of *Q-5T* in 10 mM K⁺ buffer, pH 7.0 (20 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a parallel G-quadruplex with snapback loop adopted by *Q-5T*; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral region tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeak of *syn*-G22 is labelled in red. (C) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (D) H8(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities.

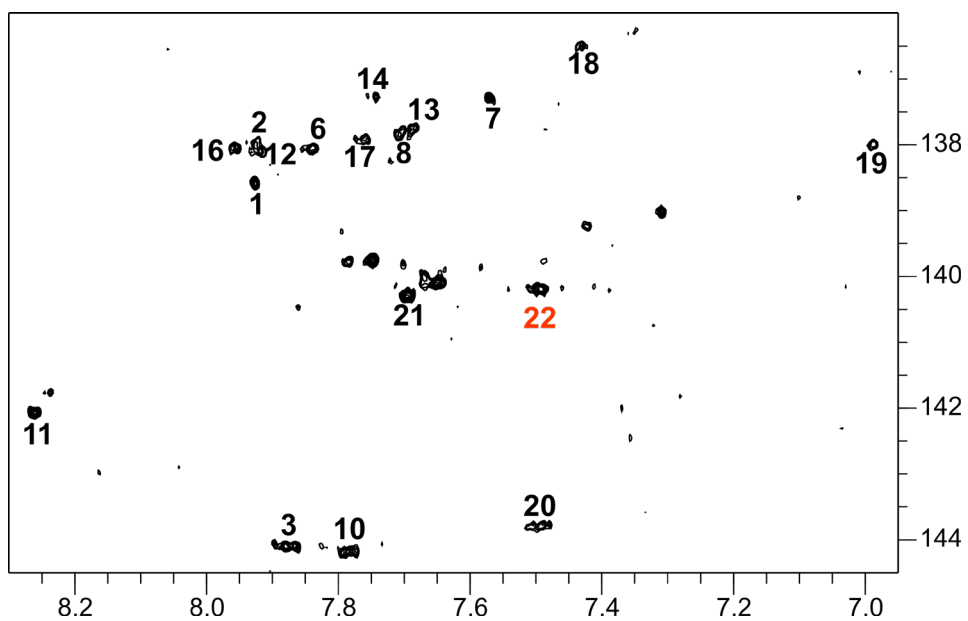


Figure S16. ^1H - ^{13}C HSQC spectrum of Q-57 acquired at 20 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6-C8/C6 correlations. Crosspeak of *syn*-G22 is labelled in red.

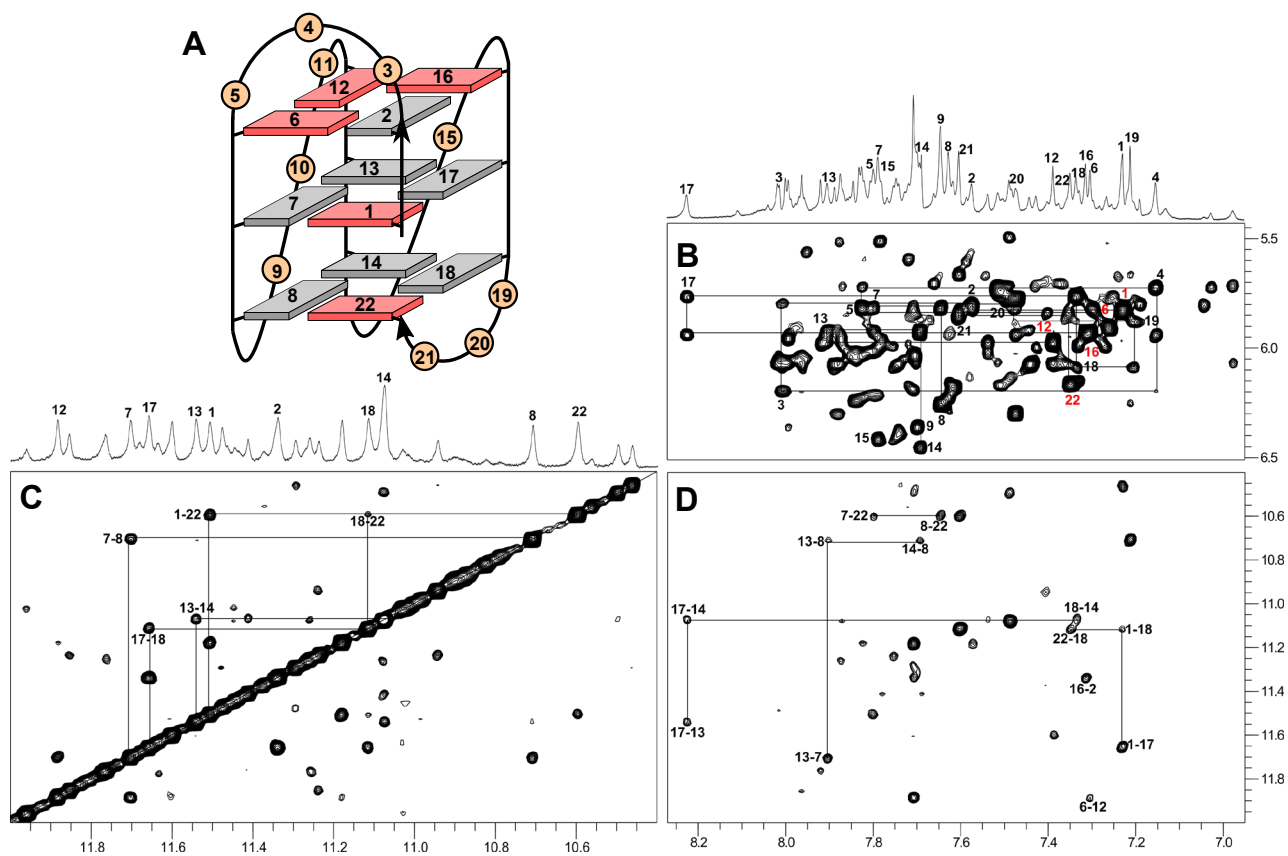


Figure S17. Topology and 2D NOESY spectral regions of *Q-11T* in 10 mM K^+ buffer, pH 7.0 (20 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a (3+1) hybrid-type G-quadruplex with a (+lpp) topology and snapback loop adopted by *Q-11T*; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral region tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeaks of *syn*-guanosines are labelled in red. (C) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (D) H8(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities.

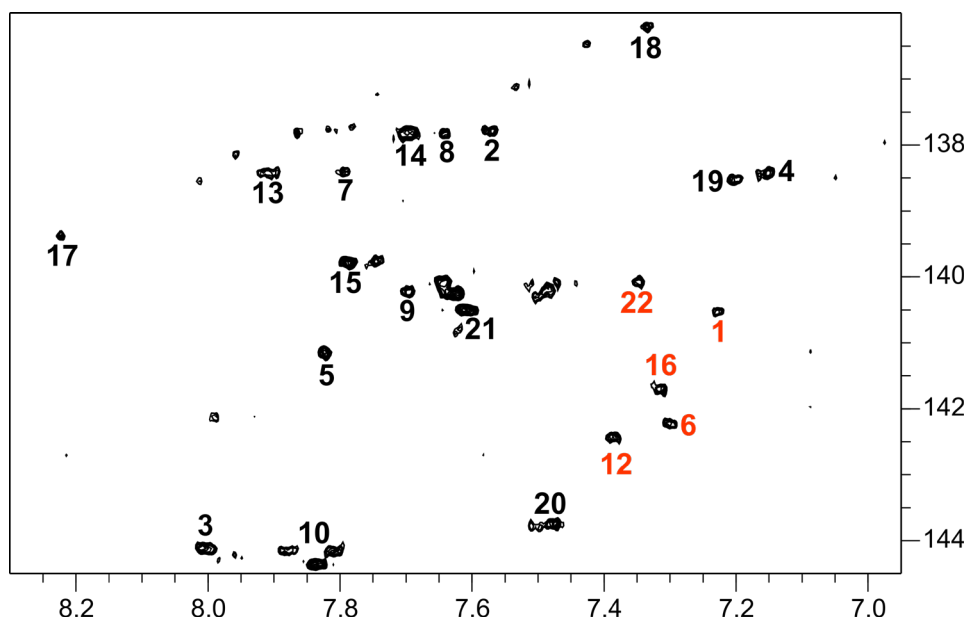


Figure S18. ^1H - ^{13}C HSQC spectrum of Q-11T acquired at 20 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. Crosspeaks of *syn*-guanosines are labelled in red.

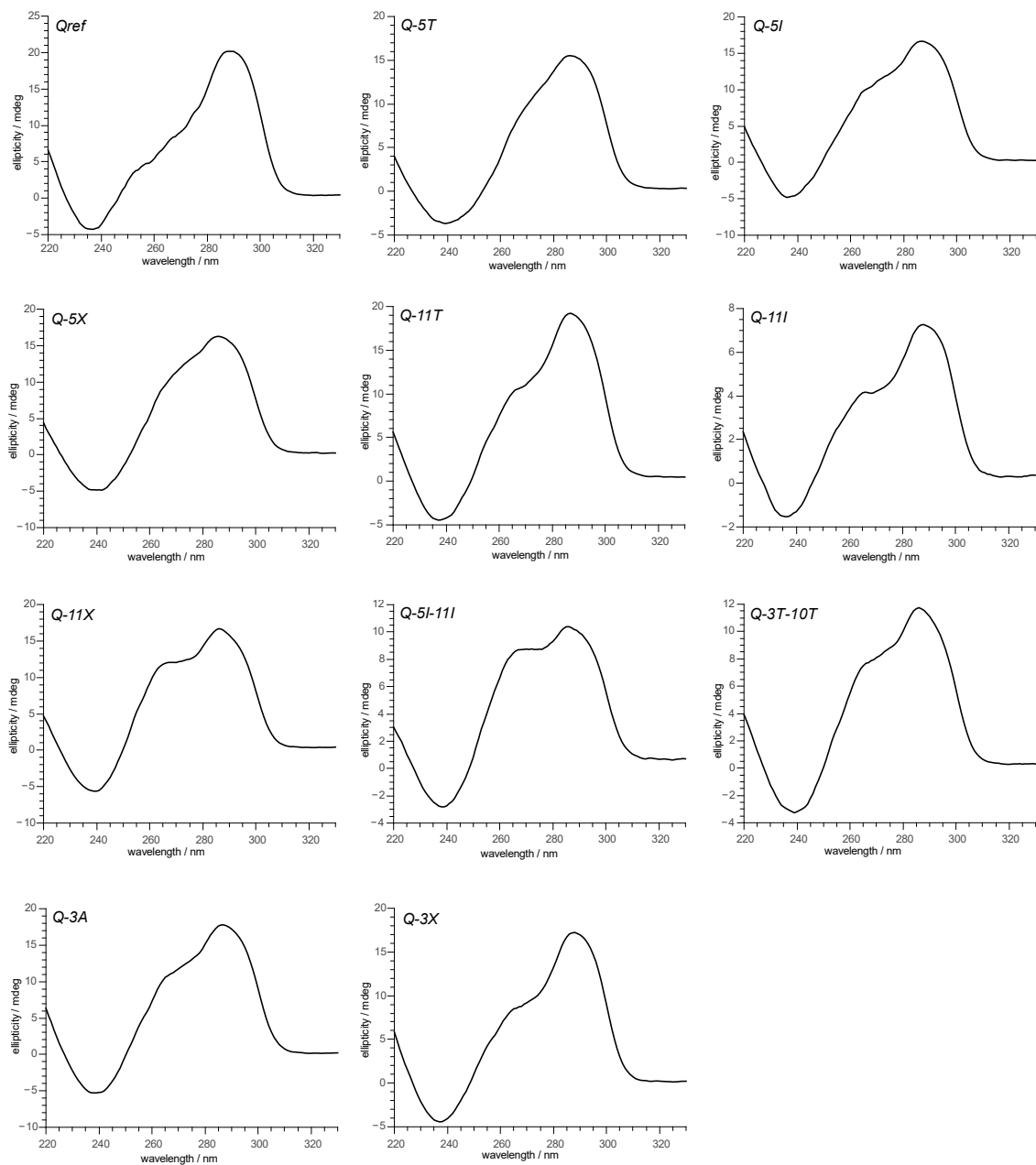


Figure S19. CD spectra of *Qref*-derived sequences with single and double mutations in 10 mM potassium phosphate buffer, pH 7, at 20 °C.

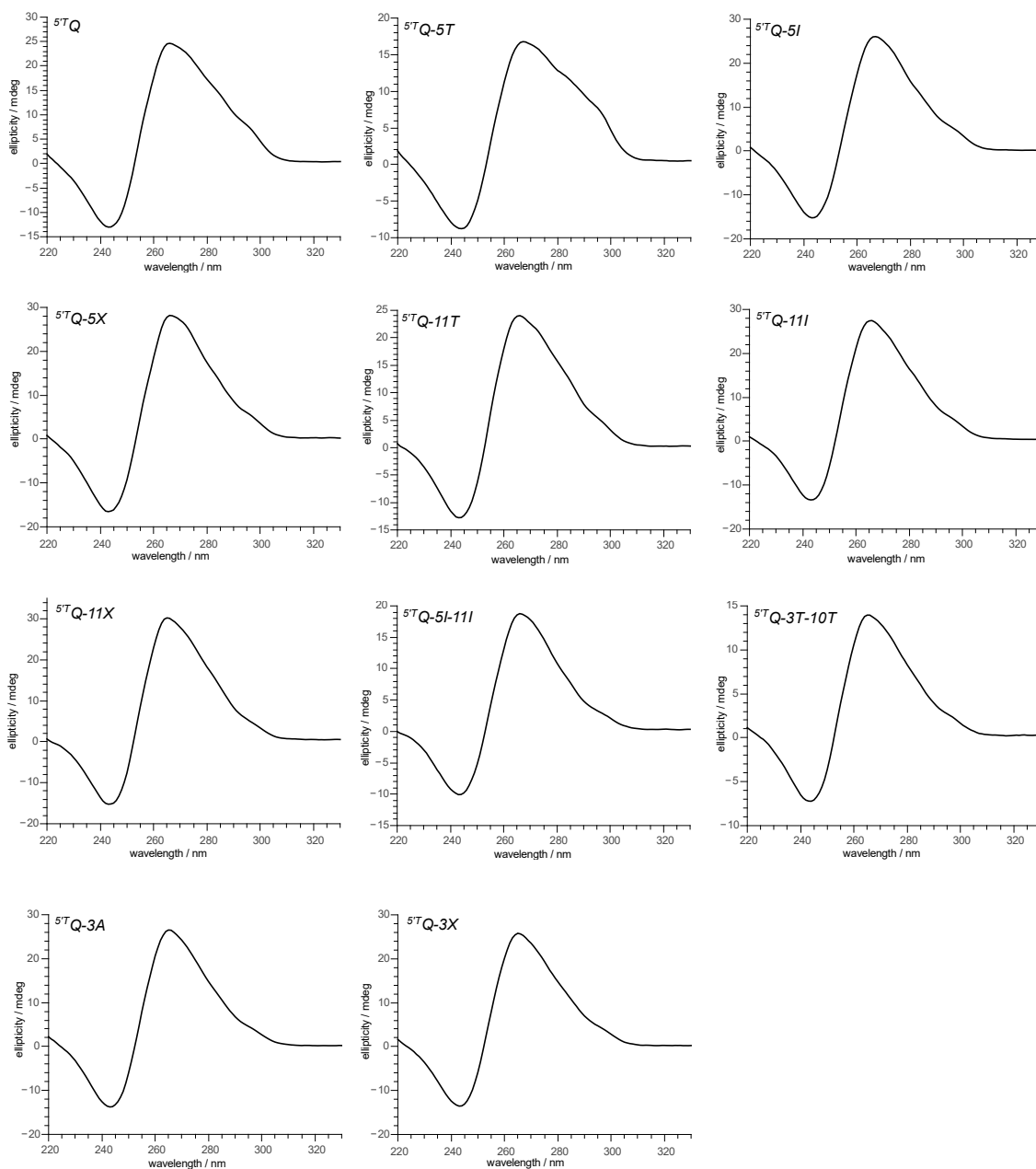


Figure S20. CD spectra of $5TQ$ -modified Q_{ref} -derived sequences in 10 mM potassium phosphate buffer, pH 7, at 20 °C.

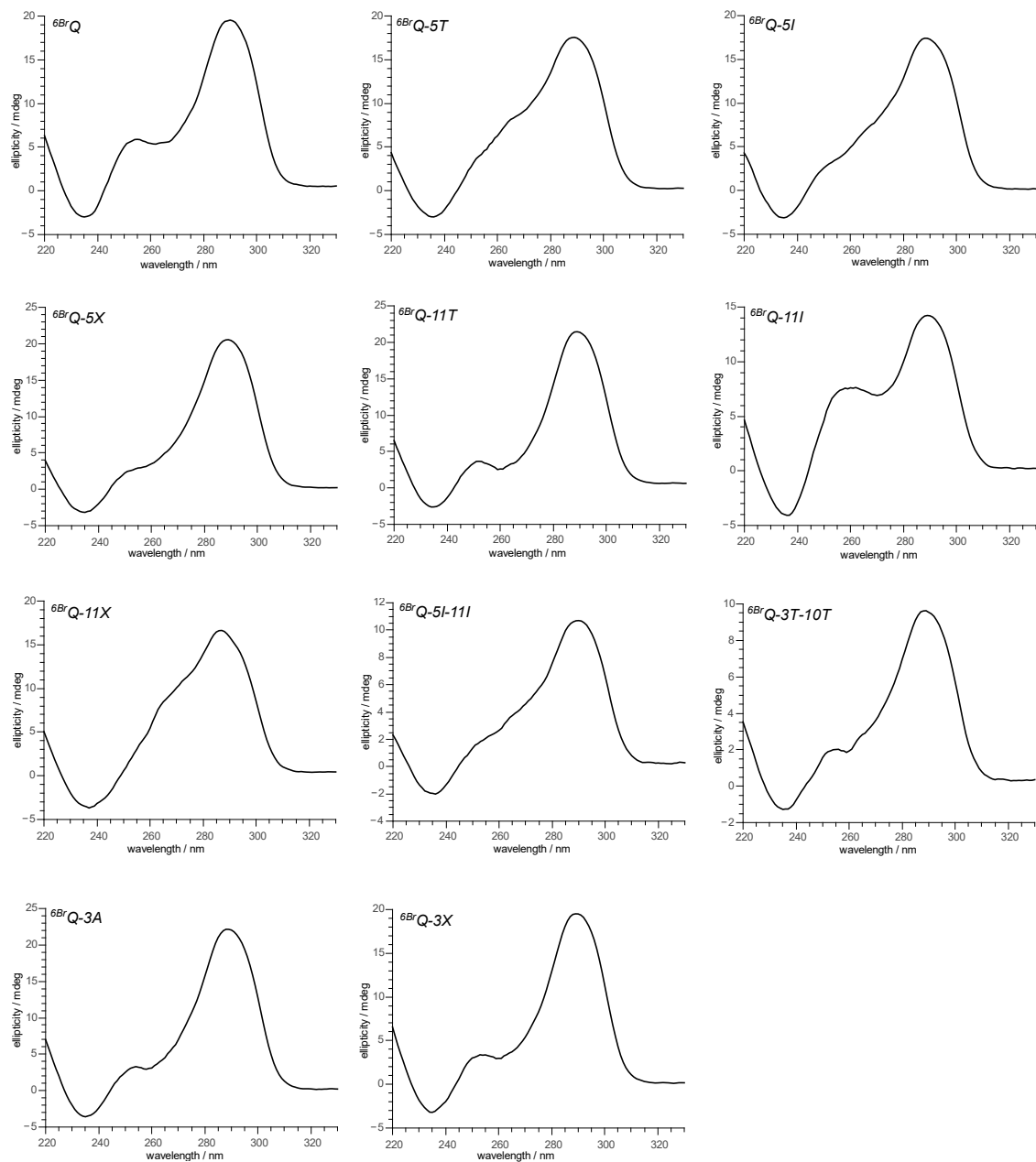


Figure S21. CD spectra of $6BrQ$ -modified Q_{ref} -derived sequences in 10 mM potassium phosphate buffer, pH 7, at 20 °C.

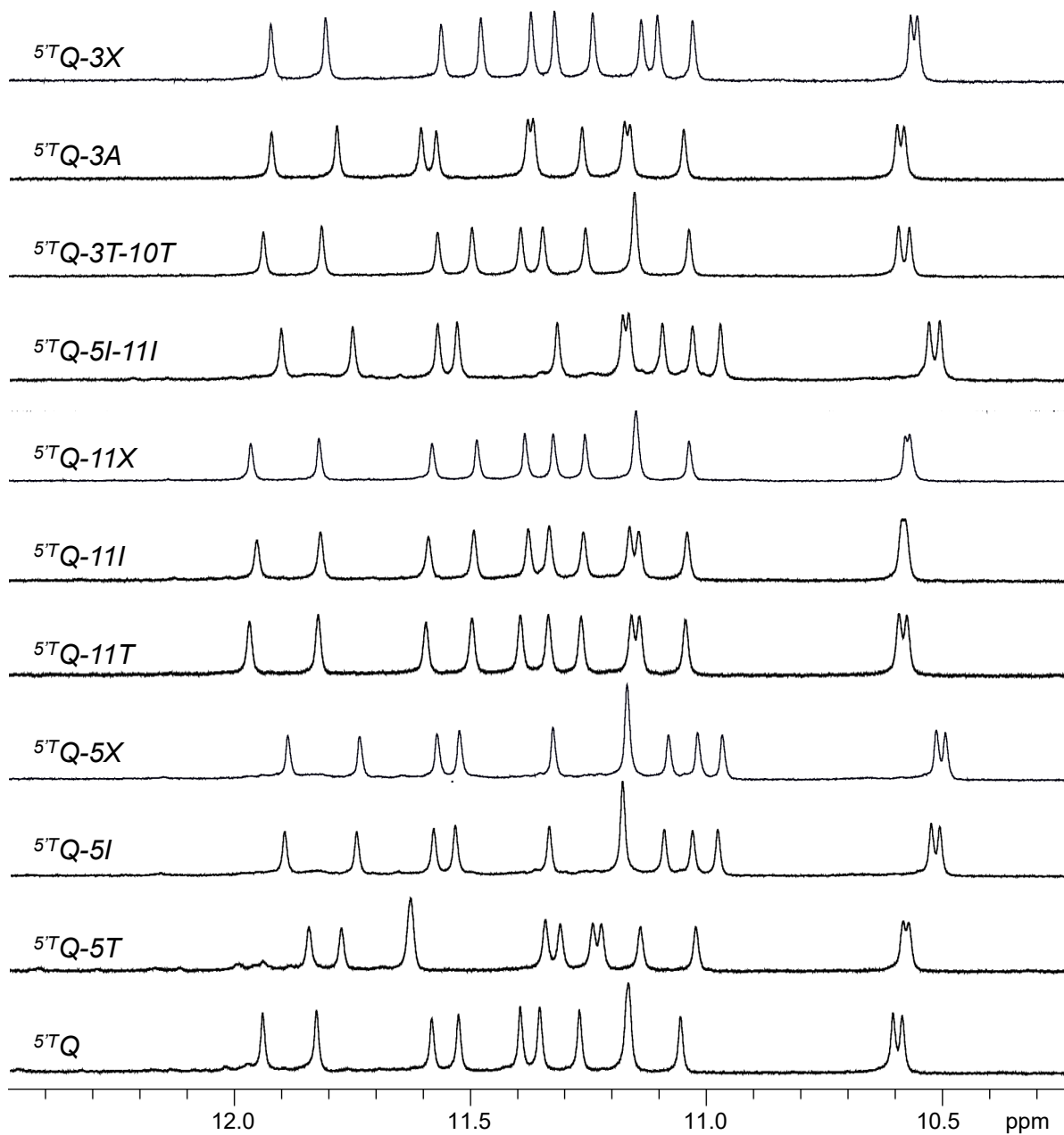


Figure S22. Imino proton spectral region of ^5TQ -modified mutants. NMR spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

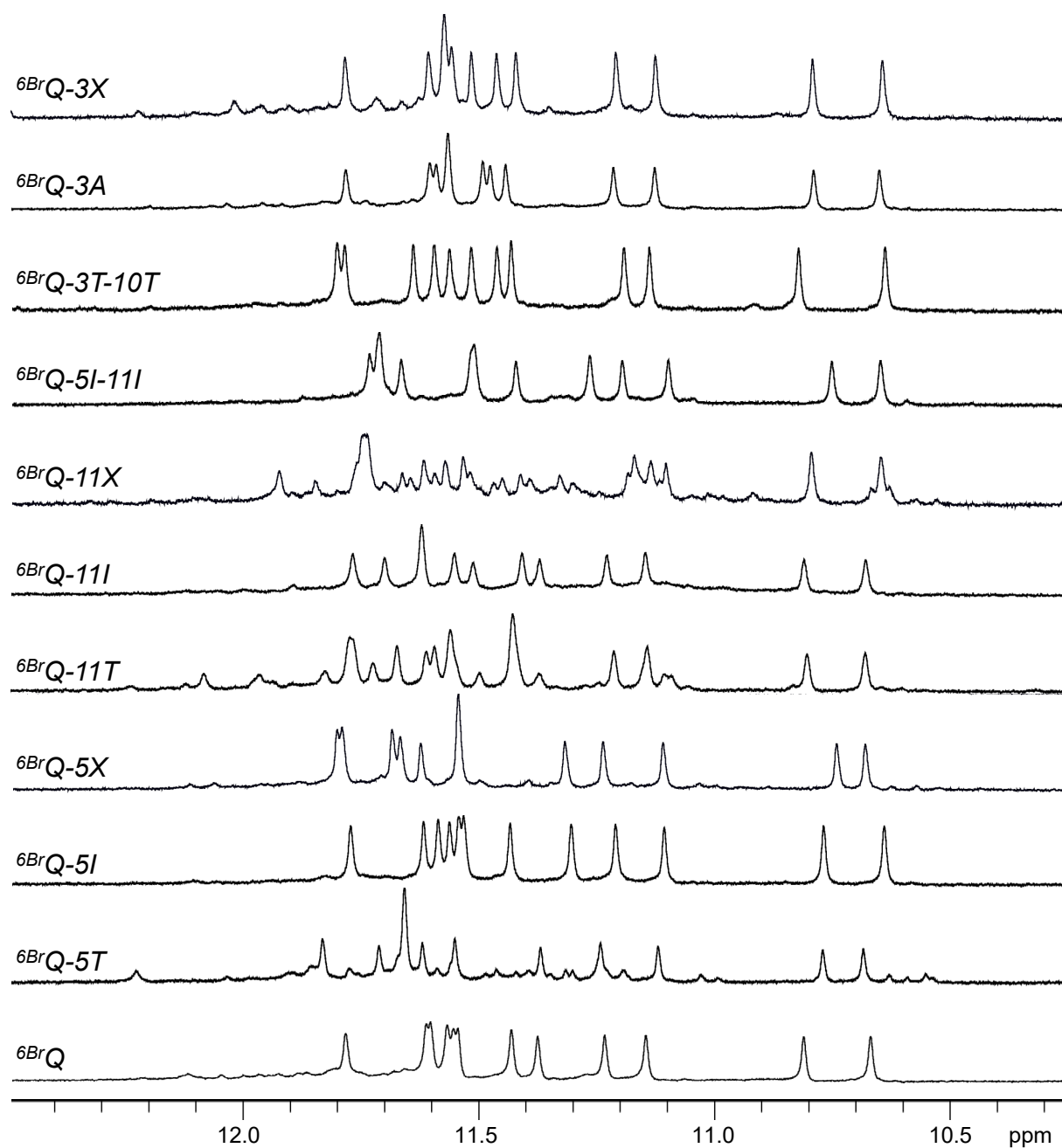


Figure S23. Imino proton spectral region of ⁶BrQ-modified mutants. NMR spectra were acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

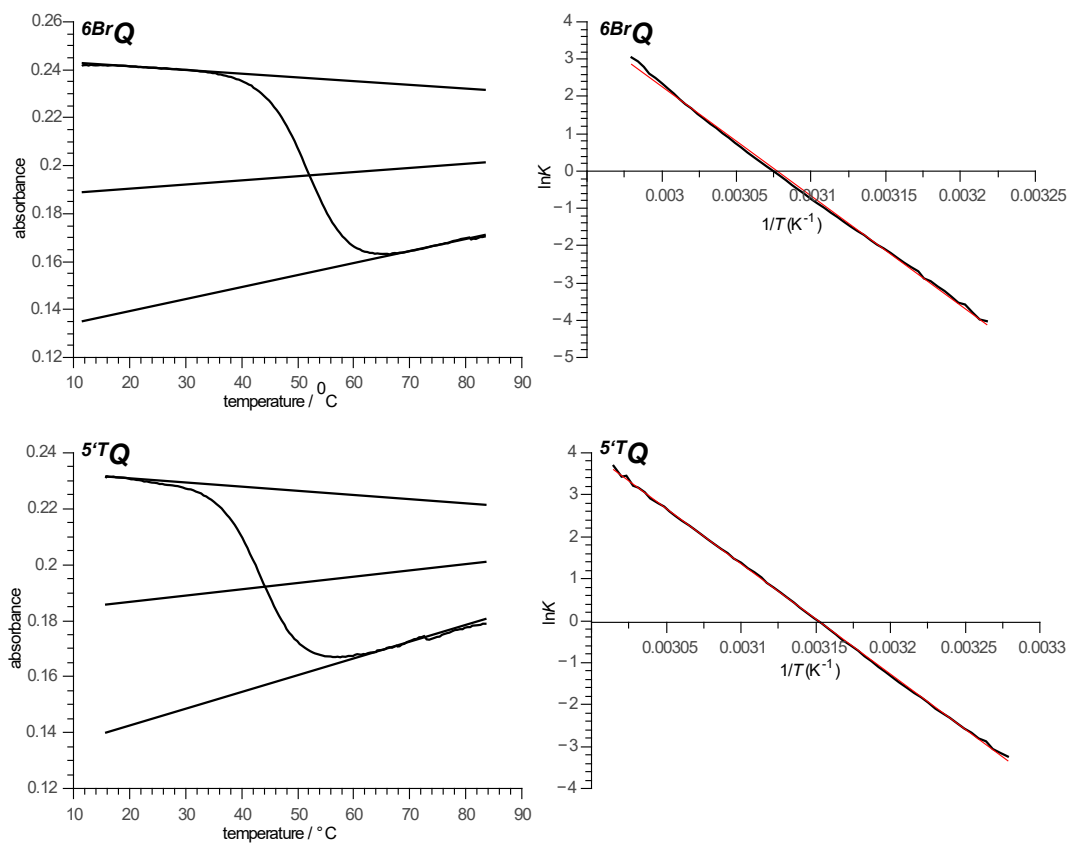


Figure S24. Exemplary van't Hoff analysis of quadruplex melting for the $6BrQ$ (top) and $5TQ$ sequence (bottom). (Left) Melting curves with fits of upper and lower baseline for the determination of temperature dependent populations. (Right) Corresponding van't Hoff plot with a superimposed linear fit curve (red).

Table S4. Thermodynamic parameters for quadruplex formation with +(lpp) topology at 30 °C.^a

oligonucleotide	T_m (°C)	ΔH° (kcal/mol) ^b	$-T\Delta S^\circ$ (kcal/mol) ^c	ΔG°_{30} (kcal/mol) ^d
⁶ BrQ	51.5 ± 1.2	-56.1 ± 0.9	52.4 ± 0.9	-3.7 ± 0.2
⁶ BrQ-3A	46.0 ± 0.9	-52.8 ± 0.9	50.1 ± 1.0	-2.7 ± 0.1
⁶ BrQ-3X ^e	54.2 ± 0.6	-64.7 ± 0.6	59.9 ± 0.7	-4.8 ± 0.1
⁶ BrQ-5T	49.5 ± 0.5	-56.0 ± 2.4	52.6 ± 2.3	-3.4 ± 0.2
⁶ BrQ-5I	51.5 ± 0.3	-58.9 ± 1.2	55.0 ± 1.1	-3.9 ± 0.1
⁶ BrQ-5X ^e	45.2 ± 0.2	-55.2 ± 1.0	52.6 ± 1.0	-2.7 ± 0.1
⁶ BrQ-11T	51.7 ± 0.7	-58.1 ± 1.9	54.2 ± 1.8	-3.9 ± 0.1
⁶ BrQ-11I	51.9 ± 0.6	-58.8 ± 1.7	54.8 ± 1.5	-4.0 ± 0.2
⁶ BrQ-5I-11I	50.3 ± 0.2	-59.6 ± 2.0	55.9 ± 1.8	-3.8 ± 0.2
⁶ BrQ-3T-10T ^e	50.8 ± 0.3	-58.1 ± 0.3	54.3 ± 0.2	-3.8 ± 0.1

^aAverage values with standard deviations derived from the analysis of three independent UV melting experiments.

^bDetermined from a van't Hoff plot. ^c $\Delta S^\circ = \Delta H^\circ / T_m$. ^d $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. ^eX = abasic 1',2'-dideoxyribose residue.

Table S5. Thermodynamic parameters for quadruplex formation with -(ppp) topology at 30 °C.^a

oligonucleotide	T_m (°C)	ΔH° (kcal/mol) ^b	$-T\Delta S^\circ$ (kcal/mol) ^c	ΔG°_{30} (kcal/mol) ^d
⁵ TQ	44.1 ± 0.3	-53.7 ± 1.9	51.3 ± 1.7	-2.4 ± 0.1
⁵ TQ-3A	39.4 ± 0.7	-52.8 ± 1.0	51.2 ± 0.8	-1.6 ± 0.1
⁵ TQ-3X ^e	46.6 ± 0.5	-54.7 ± 0.7	51.8 ± 0.6	-2.9 ± 0.1
⁵ TQ-5T	44.0 ± 0.4	-51.1 ± 1.7	48.8 ± 1.7	-2.3 ± 0.1
⁵ TQ-5I	46.4 ± 0.8	-61.7 ± 2.2	58.5 ± 2.1	-3.2 ± 0.2
⁵ TQ-5X ^e	47.4 ± 0.5	-59.9 ± 1.9	56.7 ± 1.8	-3.3 ± 0.1
⁵ TQ-11T	46.0 ± 0.3	-57.0 ± 1.1	54.1 ± 1.0	-2.9 ± 0.1
⁵ TQ-11I	44.3 ± 0.5	-55.8 ± 1.7	53.3 ± 1.7	-2.5 ± 0.1
⁵ TQ-5I-11I	47.9 ± 0.5	-61.5 ± 0.2	58.0 ± 0.3	-3.5 ± 0.1
⁵ TQ-3T-10T	42.9 ± 0.3	-56.9 ± 0.8	54.5 ± 0.8	-2.4 ± 0.1

^aAverage values with standard deviations derived from the analysis of three independent UV melting experiments.

^bDetermined from a van't Hoff plot. ^c $\Delta S^\circ = \Delta H^\circ / T_m$. ^d $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. ^eX = abasic 1',2'-dideoxyribose residue.

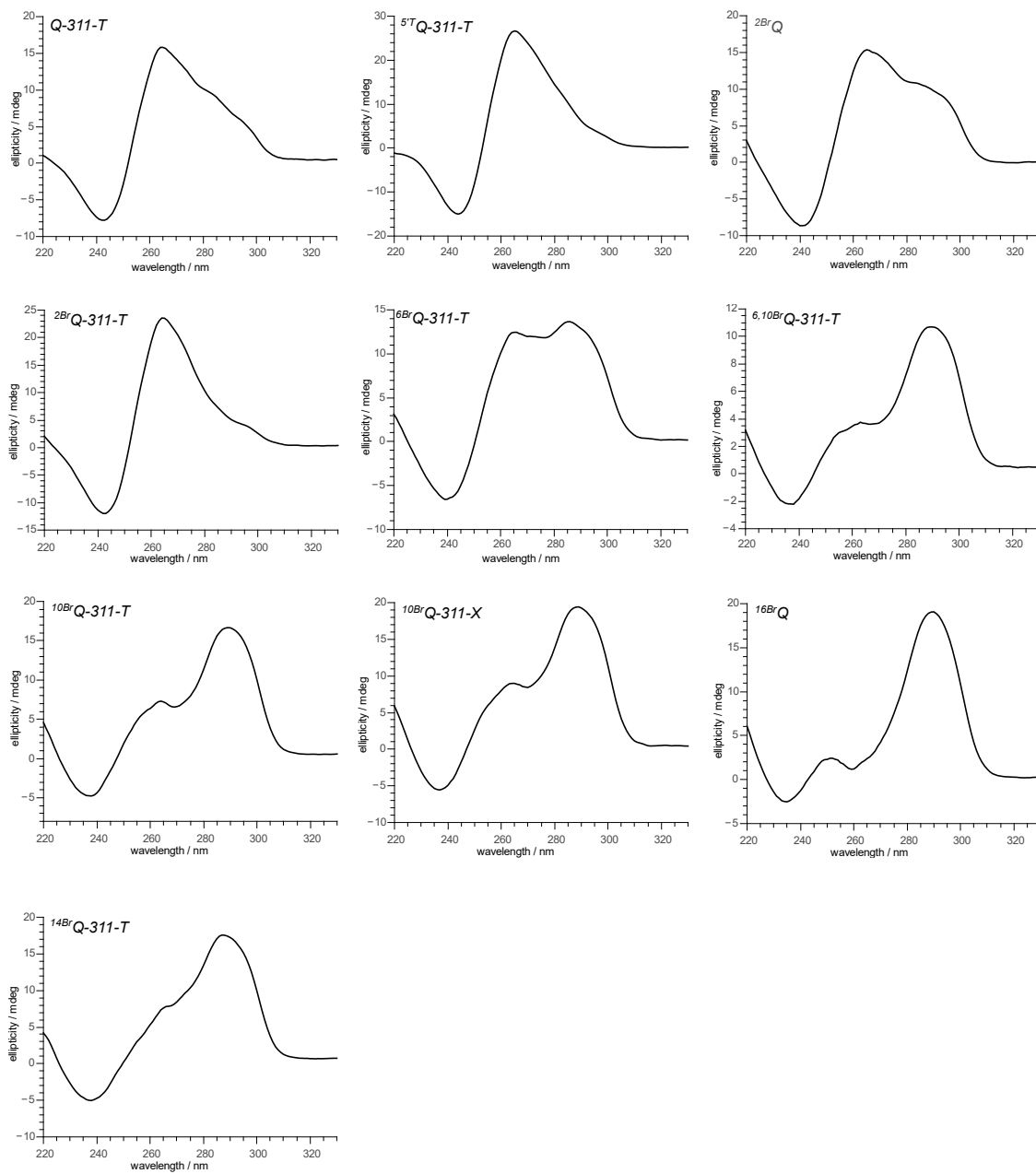


Figure S25. CD spectra of $2BrQ$, $16BrQ$, and sequences with a 311 loop length arrangement in 10 mM potassium phosphate buffer, pH 7, at 20 °C.

Gel electrophoresis. Gel electrophoresis was performed on a 15% polyacrylamide gel (acrylamide:bisacrylamide 19:1). DNA (25 μ M dissolved in 10 mM potassium phosphate buffer, pH 7.0) was loaded and separation was performed at 4 $^{\circ}$ C in 1x TBE buffer supplemented with 10 mM KCl and a voltage of 200 V. Gels were visualized by staining with a 5 μ M thiazole orange solution.

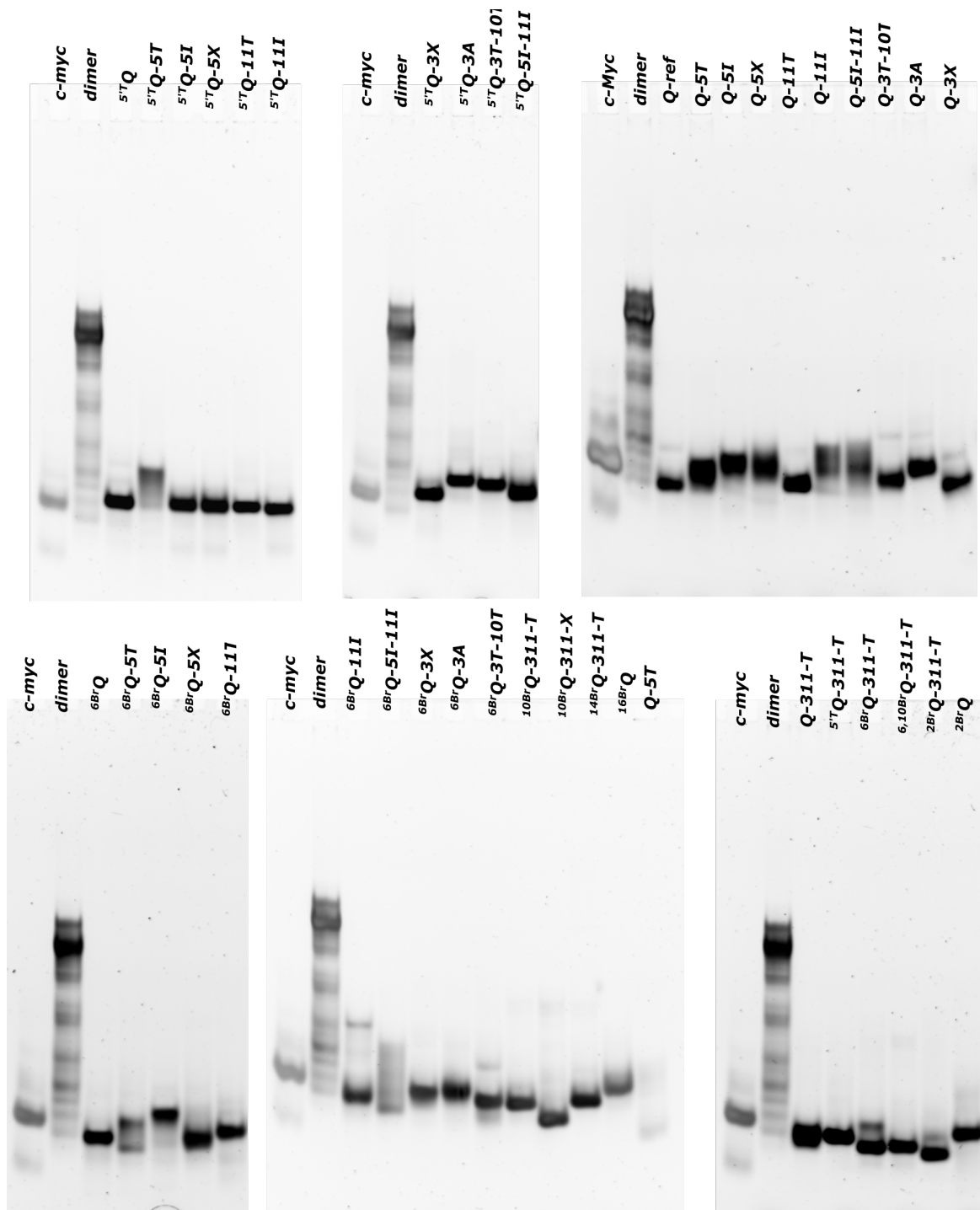


Figure S26. Non-denaturing polyacrylamide gel electrophoresis of quadruplex-forming sequences; a parallel *c-myc* quadruplex and a telomeric dimer were used as a reference.

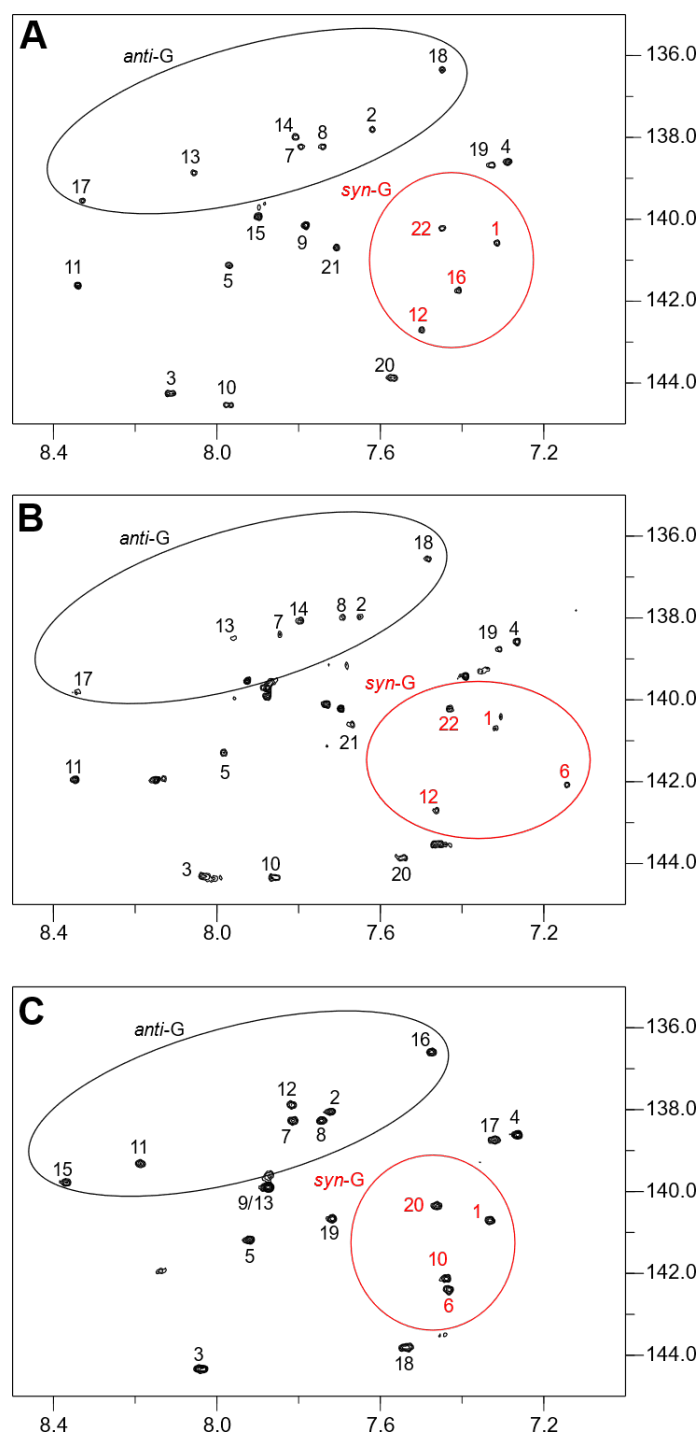


Figure S27. Comparison of ^1H - ^{13}C HSQC spectra of (A) ^6BrQ , (B) ^{16}BrQ , and (C) ^{14}BrQ -311-*T* acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. *Anti*-G and *syn*-G residues of the G-core are circled and labelled in black and red color, respectively. The highly similar pattern of C-H correlations with ^6BrQ comprising four *syn*-G and 7 *anti*-G residues for the G-core demonstrate a hybrid-type fold with one broken *syn*-*syn*-*anti* and three *syn*-*anti*-*anti* G-columns for both ^{16}BrQ and ^{14}BrQ -311-*T*. Note the absence of an observable fifth *syn*-G residue at the $^{\text{Br}}\text{G}$ incorporation site.

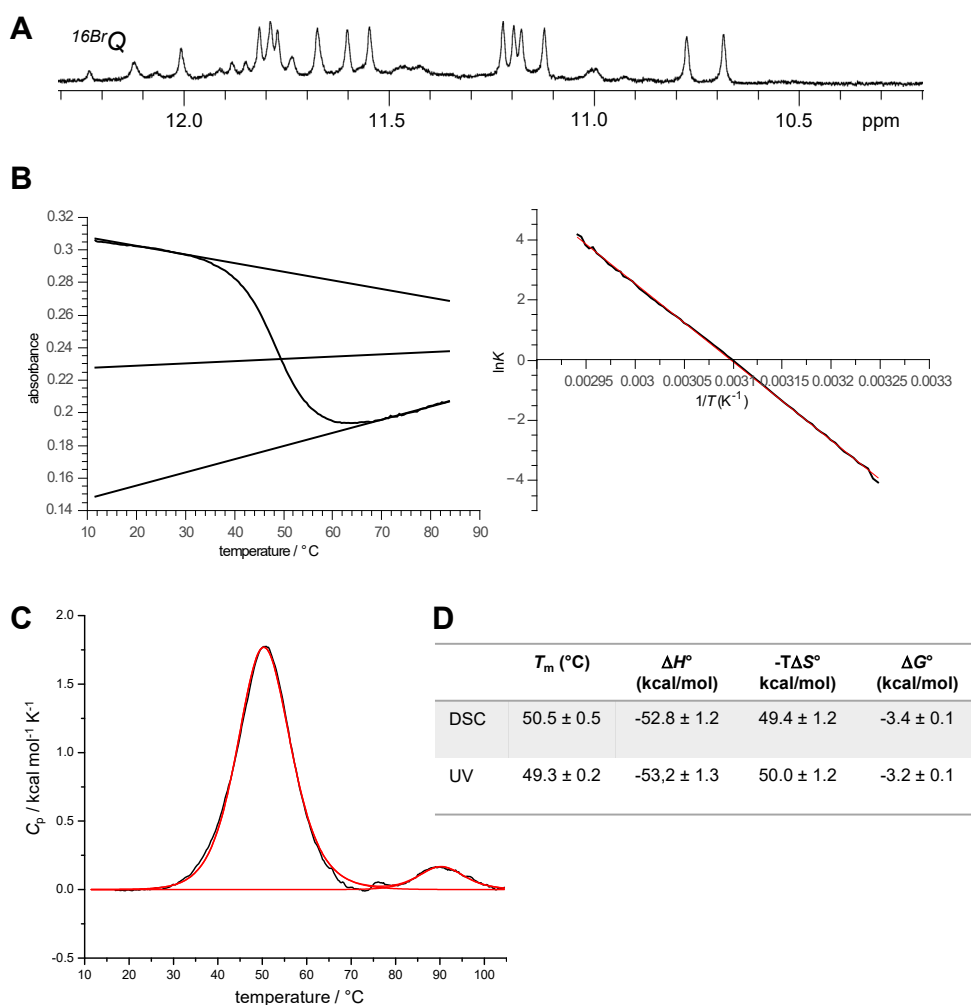


Figure S28. (A) Imino proton NMR spectral region of the ¹⁶BrQ quadruplex acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30 °C. (B) UV melting curve with fits of upper and lower baseline (left) and van't Hoff plot with a superimposed linear fit curve in red (right). (C) DSC thermogram with red fit curve comprising a low- and high-melting transition. (D) Comparison of van't Hoff thermodynamic parameters obtained from UV melting and from the low-temperature DSC transition. Being within experimental uncertainties, the same van't Hoff enthalpies extracted from the UV melting and the low-temperature DSC transition suggests that additional resonances in the NMR spectrum result from some high-melting multimeric associates that do not influence the UV melting transition centered at ~49 °C.

Differential Scanning Calorimetry (DSC) measurements. DSC experiments were carried out with a VP-DSC instrument (Malvern Instruments, United Kingdom). The oligonucleotide solution (50 μM) was heated up to 105 °C with a heating rate of 0.5 °C·min⁻¹. Data from a buffer vs. buffer scan were subtracted from the sample data. After cubic baseline correction, data were fitted assuming $\Delta C_p = 0$ kcal·mol⁻¹·K⁻¹ to obtain T_m and $\Delta H^\circ_{\text{vH}}$. Parameters are averages over three independent experiments.

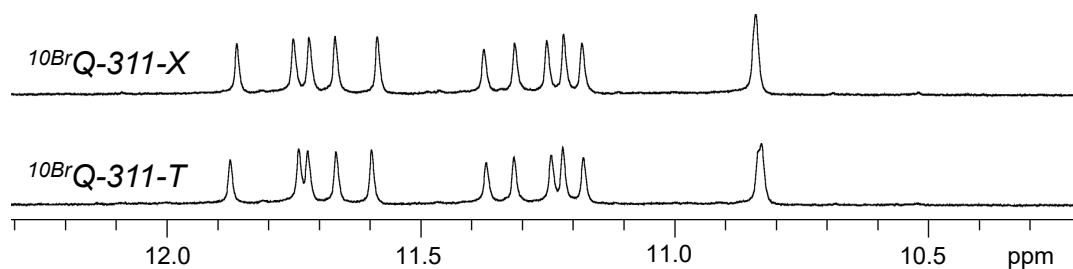


Figure S29. Imino proton NMR spectral region of $^{10}\text{BrQ-311-T}$ and $^{10}\text{BrQ-311-X}$ quadruplexes acquired in 10 mM potassium phosphate buffer, pH 7.0, at 30 °C.

Table S6. Thermodynamic parameters for the formation of $^{10}\text{BrQ-311-T}$ and $^{10}\text{BrQ-311-X}$ quadruplexes at 30 °C.^a

oligonucleotide	T_m (°C)	ΔH° (kcal/mol) ^b	$-T\Delta S^\circ$ (kcal/mol) ^c	ΔG°_{30} (kcal/mol) ^d
$^{10}\text{BrQ-311-T}$	58.5 ± 0.6	-70.1 ± 1.2	64.0 ± 1.2	-6.1 ± 0.1
$^{10}\text{BrQ-311-X}$	59.0 ± 0.4	-72.8 ± 1.7	66.4 ± 1.6	-6.4 ± 0.1

^aAverage values with standard deviations derived from the analysis of three independent UV melting experiments.

^bDetermined from a van't Hoff plot. ^c $\Delta S^\circ = \Delta H^\circ / T_m$. ^d $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

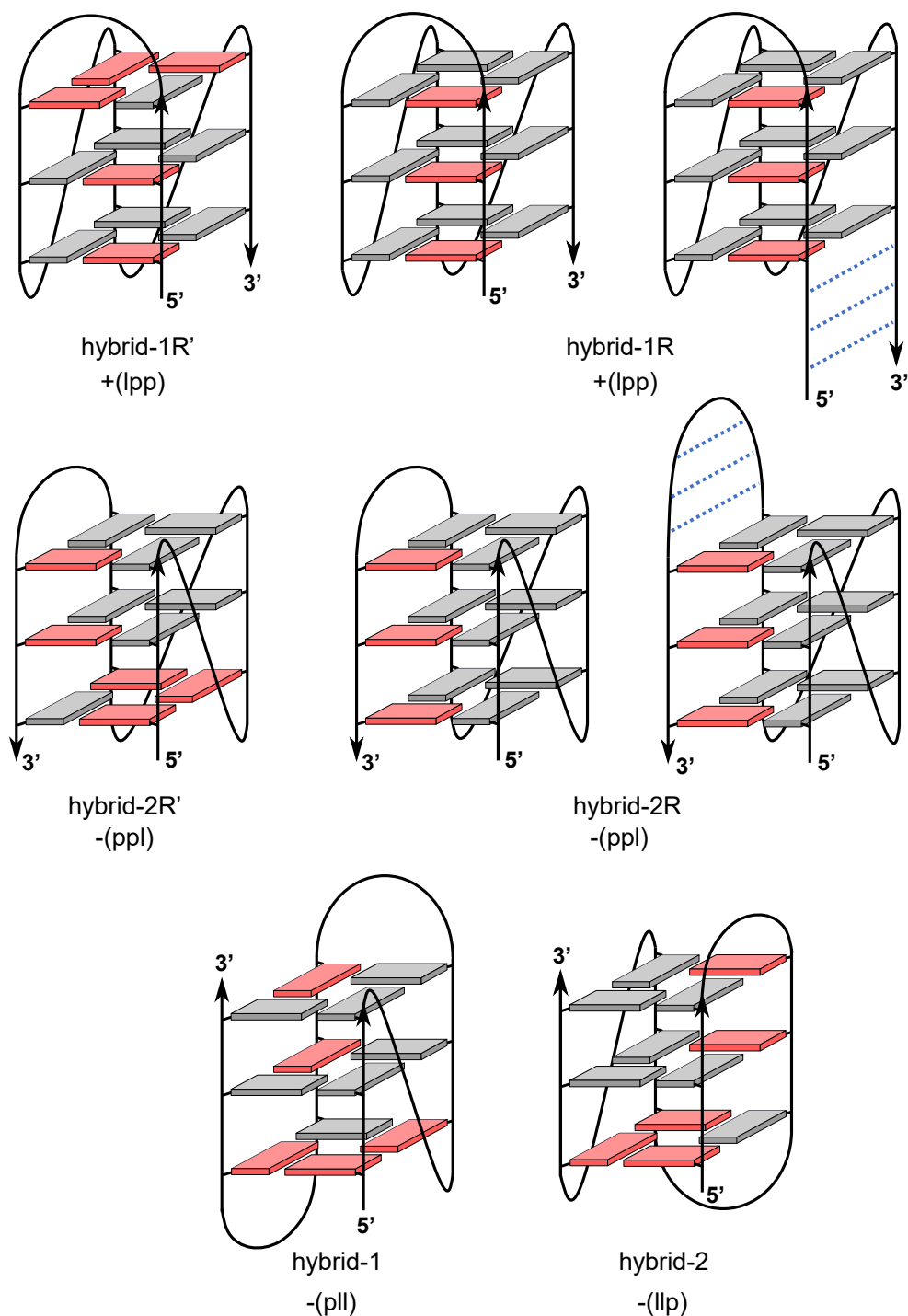


Figure S30. Topologies of (3+1) hybrid quadruplexes with different arrangements of lateral and propeller loops. Due to their all-*syn* antiparallel column, hybrid-1R and hybrid-2R quadruplexes feature only homopolar tetrad stacking as observed in parallel structures; *syn*- and *anti*-G residues are colored red and grey, respectively. For hybrid-1R and hybrid-2R conformations, additional quadruplex-duplex hybrids with Watson-Crick base pairs between flanking sequences or within the lateral loop (dotted blue lines) form a coaxially stacked duplex extension along the quadruplex wide groove, promoting folding into one of the two +(lpp) or –(ppl) topologies (top and middle right).

NMR spectral analysis and resonance assignments for ^{2Br}Q-311-T

Upon shortening the second intervening sequence to a 1-nt propeller loop, three formed species of Q-311-T coexist with the major species adopting a parallel topology like ^{5T}Q (see Figure 8 of the main manuscript). Minor species can be identified through shifting equilibria by deliberate substitutions with ^{Br}G residues at specific positions. The ^{2Br}Q-311-T quadruplex with a corresponding modification at the 2-position yields a set of 12 imino proton resonances which can be attributed to the formation of a three-layered G-quadruplex (Figure S30). Two *syn*- and nine *anti*-guanosines can be identified following NOE connectivities and this is confirmed by ¹H-¹³C HSQC spectra (Figure S31). NOE sequential walks along the G-quadruplex allowed to distinguish three columns comprising all-*anti* steps, namely G7-G8-G9, G10-G11-G12, and G14-G15-G16. One column was found to exclusively include *syn-syn* steps, i.e., G20-G1-G2, providing for the missing ^{Br}G2 in a *syn* conformation. Sequential contacts between aromatic and anomeric protons could be traced between G14 to A19 (Figure S30C). Also, uninterrupted sequential contacts link residues from C3 to G9. Additionally, G2 H1' and G2 H3' show weak and strong sequential contacts to C3 H6, respectively. G20 and G1 feature a reverse sequential NOE crosspeak pattern typical for a *syn-syn* step in the H8-H3' spectral region (Figure S30B). G20 can be distinguished from G1 due to a long-range contact to C18. Following assignments in the aromatic-imino and imino-imino spectral region, the corresponding crosspeak pattern confirms a quadruplex topology with exclusive homopolar tetrad stacking (Figure S30D and E). Hoogsteen hydrogen bonds within individual tetrads run along G2-G6-G10-G14, G1-G7-G11-G15, and G20-G8-G12-G16. Taken together, ^{2Br}Q-311-T folds into a (3+1) hybrid quadruplex with a homopolar tetrad stacking and a single all-*syn* G-tract.

The quadruplex is composed of four intervening sequences, namely a 3-nt lateral loop bridging a minor groove followed by two 1-nt propeller loops and a 3-nt lateral snapback loop bridging a wide groove. The thymidine resonances of the two 1-nt propeller loops cannot be distinguished unambiguously. However, various sequential and long-range contacts define the alignment of both lateral loops. In particular, long-range contacts from A5 to C3 and especially to G2 were found in the aromatic-imino but also aromatic-H1' spectral region (not shown). Long-range contacts between G20 H8 and C18 H1' indicate C18 to be located in the groove formed by the all-*syn* and an all-*anti* column. Various NOE contacts between A19 base and sugar protons as well as T17 methyl protons with G imino protons of the lower tetrad hint to T17 and A19 being stacked onto the plane of the outer tetrad (Figure S30E).

Stereospecific assignments of H2'/H2'' was achieved through NOESY experiments at short mixing times for the discrimination of H1'-H2' and H1'-H2'' crosspeak intensities. Sugar puckers were evaluated by a close inspection of H1'-H2' and H1'-H2'' DQF-COSY crosspeaks. Due to anti-phase signals for the active coupling, small coupling constants are expected to result in partial cancellation and weak signal intensities. Thus, all residues except for A5 and G20 were found to adopt a *south*-type sugar pucker. Isochronous H2' and H2'' resonances hampered the assignment of the A5 sugar pucker. On the other hand, the H1'-H2'' DQF-COSY crosspeak for G20 overlapped with the corresponding crosspeak of C3, yet the H1'-H2' crosspeak intensity suggests at least partially populated *north*-type sugars for G20 (Figure S32).

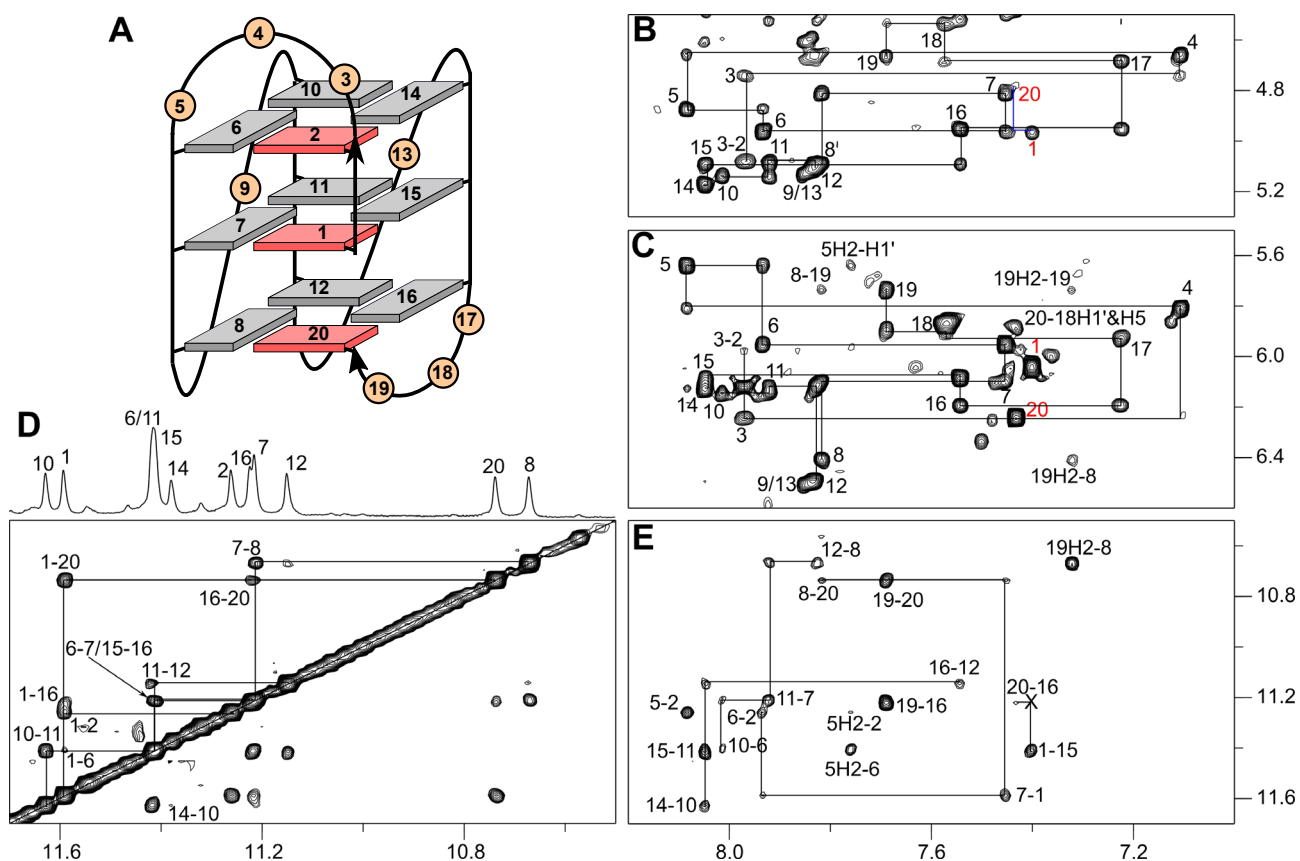


Figure S31. Topology and 2D NOESY spectral regions of $^{2Br}Q-311-T$ in 10 mM K^+ buffer, pH 7.0 (30 °C, mixing time 300 ms). (A) Schematic representation with numbered residues of a (3+1) hybrid-type G-quadruplex with a (+lpp) topology and an all-*syn* G-column adopted by $^{2Br}Q-311-T$; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively. (B) H6/H8(ω_2)-H3'(ω_1) and (C) H6/H8(ω_2)-H1'(ω_1) 2D NOE spectral regions tracing continuous intra-nucleotide and sequential connectivities; intra-nucleotide crosspeaks of *syn*-guanosines are labelled in red; blue lines in (B) indicate a reversed sequential contact typical for *syn-syn* steps. (D) H1(ω_2)-H1(ω_1) crosspeaks with sequential contacts traced along the G tracts. (E) H8/H2(ω_2)-H1(ω_1) NOE contacts with typical intra-tetrad GH8-GH1 connectivities.

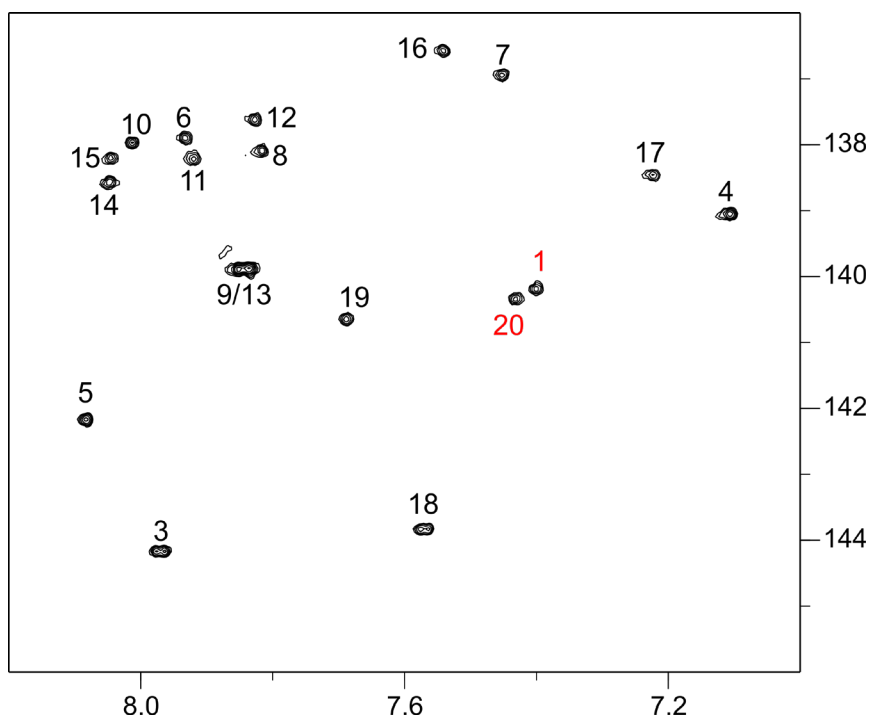


Figure S32. ^1H - ^{13}C HSQC spectrum of $^{2\text{Br}}\text{Q-311-T}$ acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. Crosspeaks of *syn*-G1 and *syn*-G20 are labelled in red.

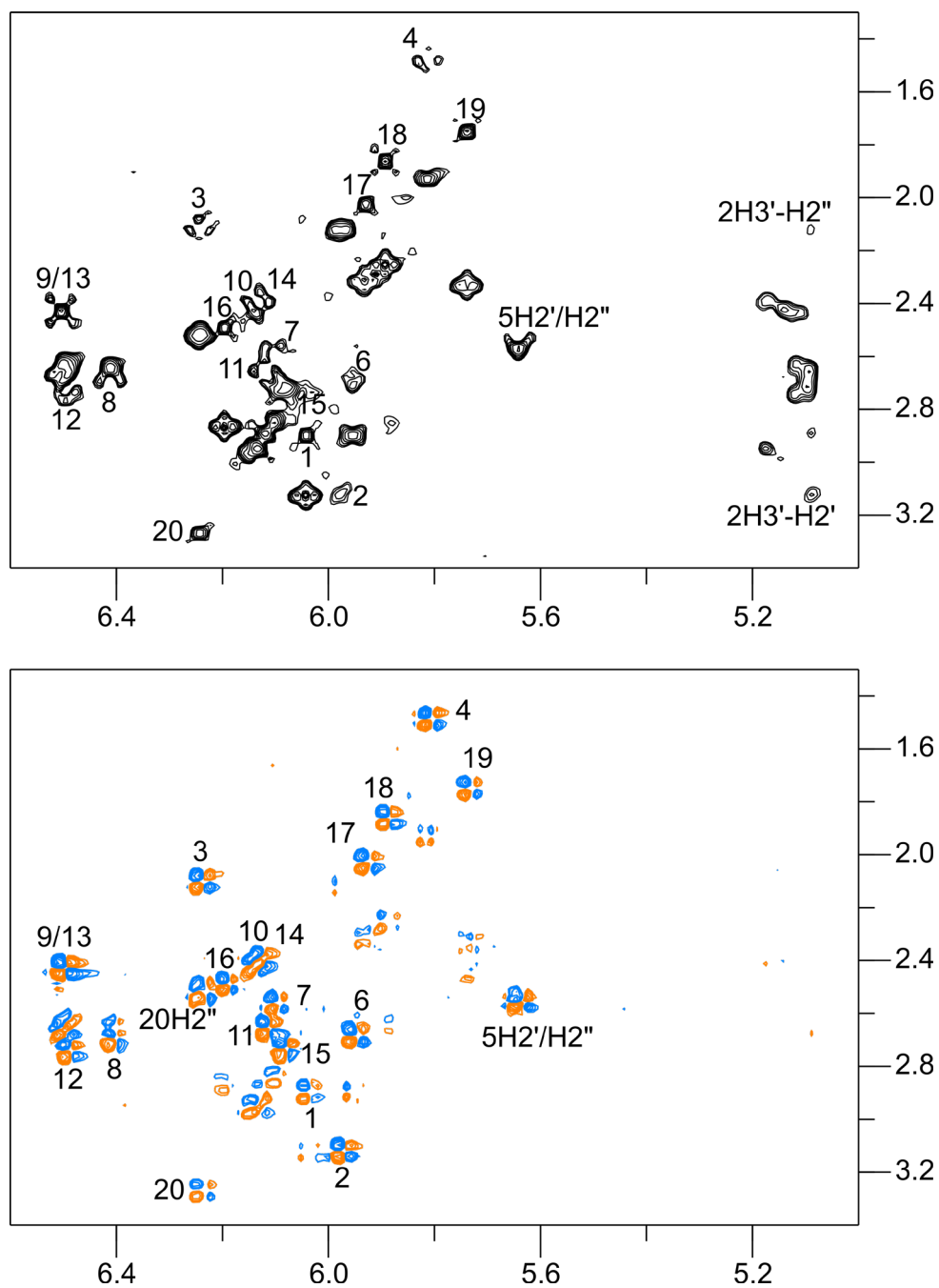


Figure S33. Sugar pucker analysis of $^{2\text{Br}}\text{Q-311-T}$. (Top) Stereospecific assignments of H_2'/H_2'' with $\text{H}_1'(\omega_2)$ - $\text{H}_2'/\text{H}_2''(\omega_1)$ NOESY spectral region at short mixing time (80 ms); crosspeak intensities allow discrimination between H_2' and H_2'' . (Bottom) DQF-COSY spectral region showing $\text{H}_1'(\omega_2)$ - $\text{H}_2'/\text{H}_2''(\omega_1)$ crosspeaks; *north*- and *south*-type sugar pucker conformations are associated with different scalar couplings and thus different crosspeak patterns of in-phase and anti-phase components.

Table S7. ^1H and ^{13}C chemical shifts δ of $^{2\text{Br}}\text{Q-311-T}$.^a

δ (ppm)	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
G1	7.40	11.59	6.04	2.90/3.13	4.97	-	140.19	-
^{Br} G2	-	11.26	5.98	3.12/2.12	5.08	-	n.d.	-
C3	7.97	-	6.24	2.11/2.53	4.74	6.12	144.17	-
T4	7.11	n.d.	5.81	1.49/1.93	4.66	1.52	139.06	-
A5	8.08	-	5.64	2.56/2.56	4.87	7.76	142.19	154.81
G6	7.93	11.41	5.95	2.68/2.90	4.96	-	137.89	-
G7	7.45	11.21	6.10	2.56/2.85	4.81	-	136.95	-
G8	7.82	10.67	6.41	2.69/2.65	5.09	-	138.10	-
T9	7.85	n.d.	6.50	2.43/2.67	5.13	1.96	139.90	-
G10	8.01	11.63	6.15	2.42/2.95	5.14	-	137.98	-
G11	7.92	11.42	6.12	2.65/2.89	5.08	-	138.22	-
G12	7.83	11.15	6.49	2.74/2.61	5.10	-	137.62	-
T13	7.84	n.d.	6.50	2.43/2.67	5.12	1.96	139.88	-
G14	8.05	11.38	6.13	2.40/2.95	5.17	-	138.58	-
G15	8.05	11.41	6.09	2.74/2.72	5.09	-	138.20	-
G16	7.54	11.22	6.19	2.49/2.87	4.95	-	136.58	-
T17	7.22	n.d.	5.93	2.03/2.32	4.69	1.65	138.48	-
C18	7.57	-	5.93	1.86/2.26	4.53	5.87	143.83	-
A19	7.69	-	5.89	1.75/2.33	4.66	7.32	140.65	152.94
G20	7.43	10.74	6.24	3.27/2.51	4.78	-	140.35	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

Table S8. NMR restraints and structural statistics of calculated structures for ²BrQ-311-T

NOE distance restraints	
intra-residual	77
inter-residual	123
exchangeable	42
other restraints:	
hydrogen bonds	48
dihedral angles	38
planarity	3
structural statistics:	
pairwise heavy atom RMSD value (Å)	
all residues	1.5 ± 0.5
G-tetrad core	0.7 ± 0.2
NOE violations:	
maximum violation (Å)	0.14
mean NOE violation (Å)	0.002 ± 0.001
deviations from idealized geometry:	
bond lengths (Å)	0.01 ± 0.0001
bond angles (degree)	2.2 ± 0.02

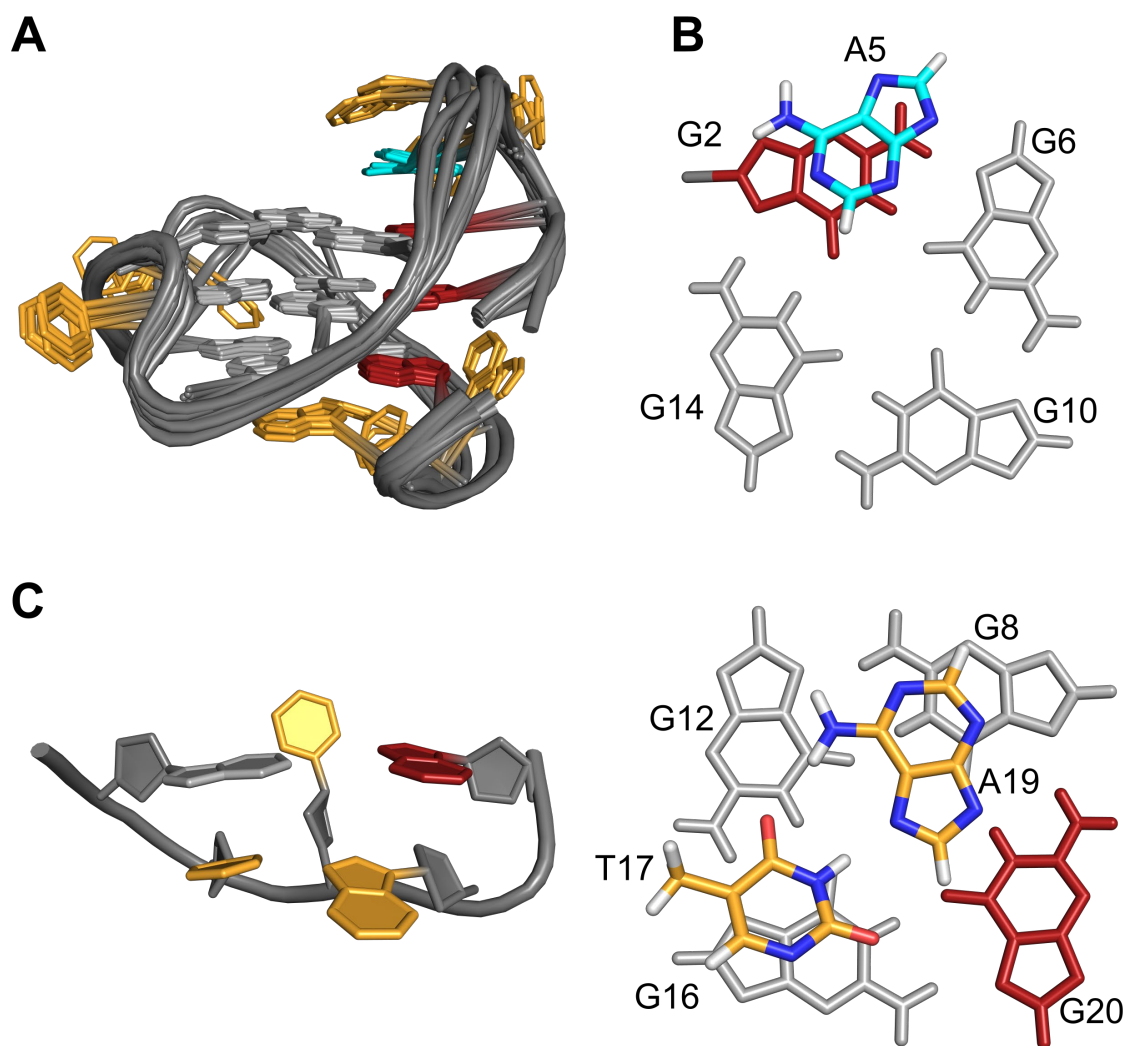


Figure S34. (A) Superposition of ten lowest-energy structures for $^{2Br}Q-311-T$; *anti*- and *syn*-guanosines are colored in grey and red, respectively, A5 is colored in cyan, and other residues are colored in orange. (B) Stacking of A5 onto the upper G-tetrad. (C) TCA lateral snapback loop of $^{2Br}Q-311-T$ in a side view (left) and top view (right) showing stacking of the T17·A19 Hoogsteen base pair onto the outer G-tetrad. C18 is located in the groove (not shown in top view). Such a capping structure of the T17·A19 base pair can be found in 8 out of 10 structures. In two structures, the snapback loop rearranges to have T17 and C18 stacked on each other while being tilted at about 45° whereas A19 remains efficiently stacked onto the tetrad.

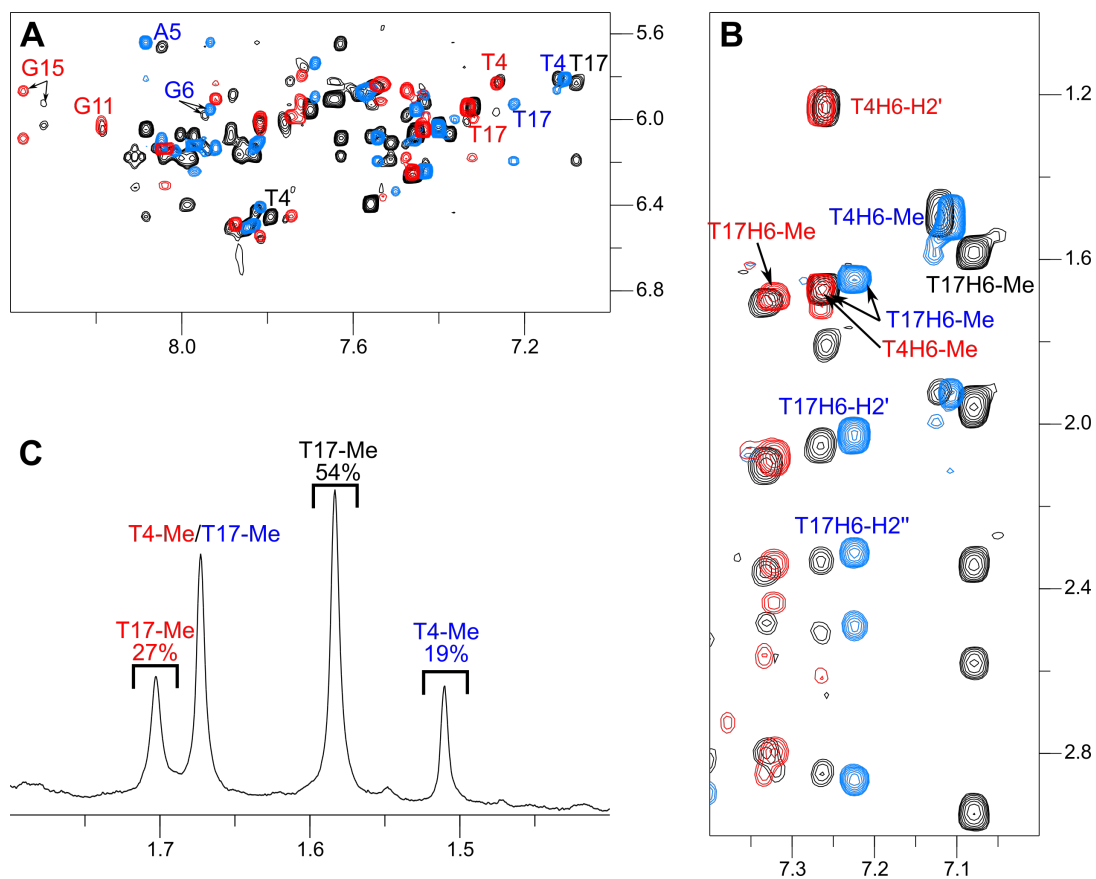


Figure S35. (A) Superimposed H6/8-H1' and (B) H6/8-H2'/Me NOESY spectral regions of Q-311-T (black), ^{14}Br Q311-T (red) and ^{2}Br Q-311-T (blue). (C) Thymine methyl 1D NMR spectral region of Q-311-T with integrals given for the specified peaks. Assigned proton resonances in the parallel, hybrid-1R', and hybrid-1R topologies are colored in black, red, and blue, respectively. The T4 methyl of the major parallel species resonates outside the plotted spectral region.

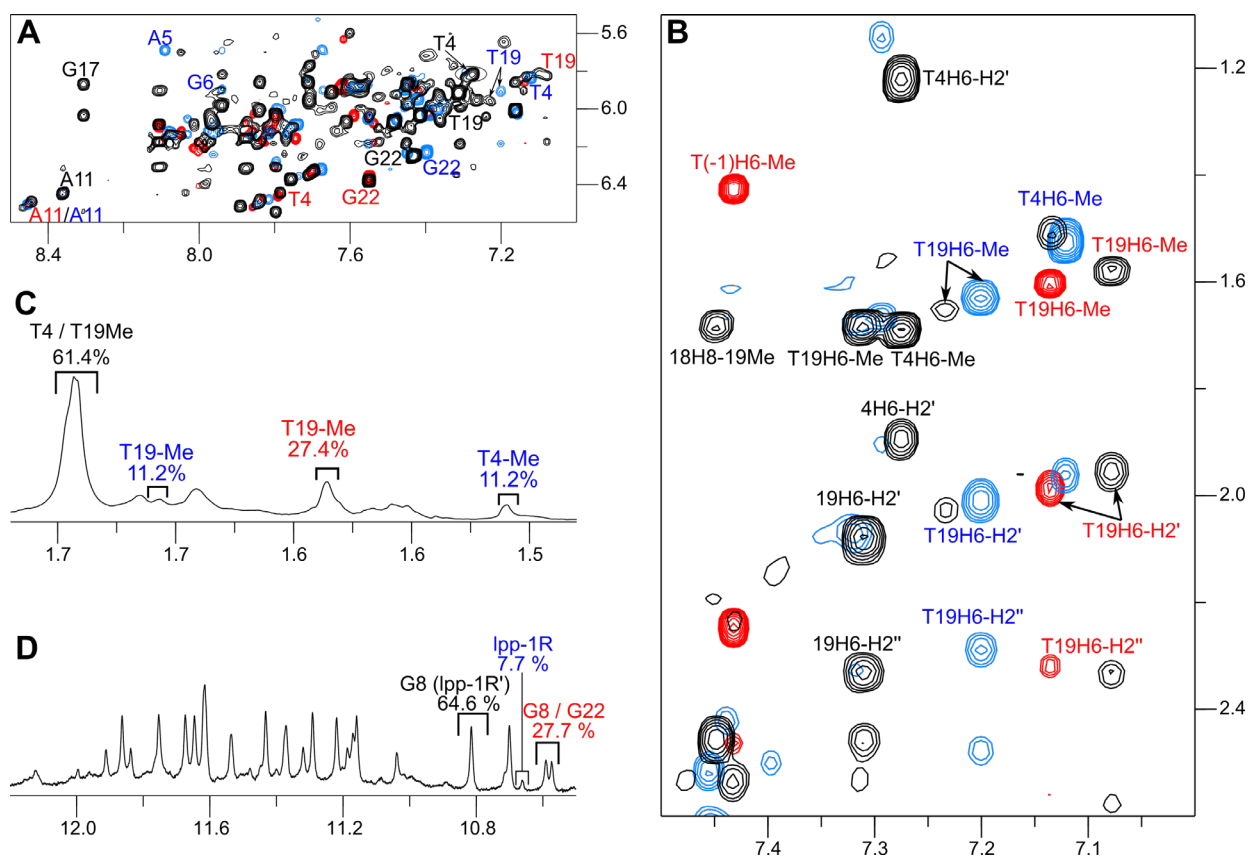


Figure S36. (A) Superimposed H6/8-H1' and (B) H6/8-H2'/Me NOESY spectral regions of *Qref* (black), ^{5T}Q (red) and ^{2Br}Q (blue). (C) Thymine methyl 1D NMR spectral region of *Qref* with integrals given for the specified peaks. Assigned proton resonances in the parallel, hybrid-1R', and hybrid-1R topologies are colored in red, black, and blue, respectively. (D) Imino proton spectral region of *Qref* with topology-specific assignments and integrals of imino resonances in line with the T methyl-based assignments.

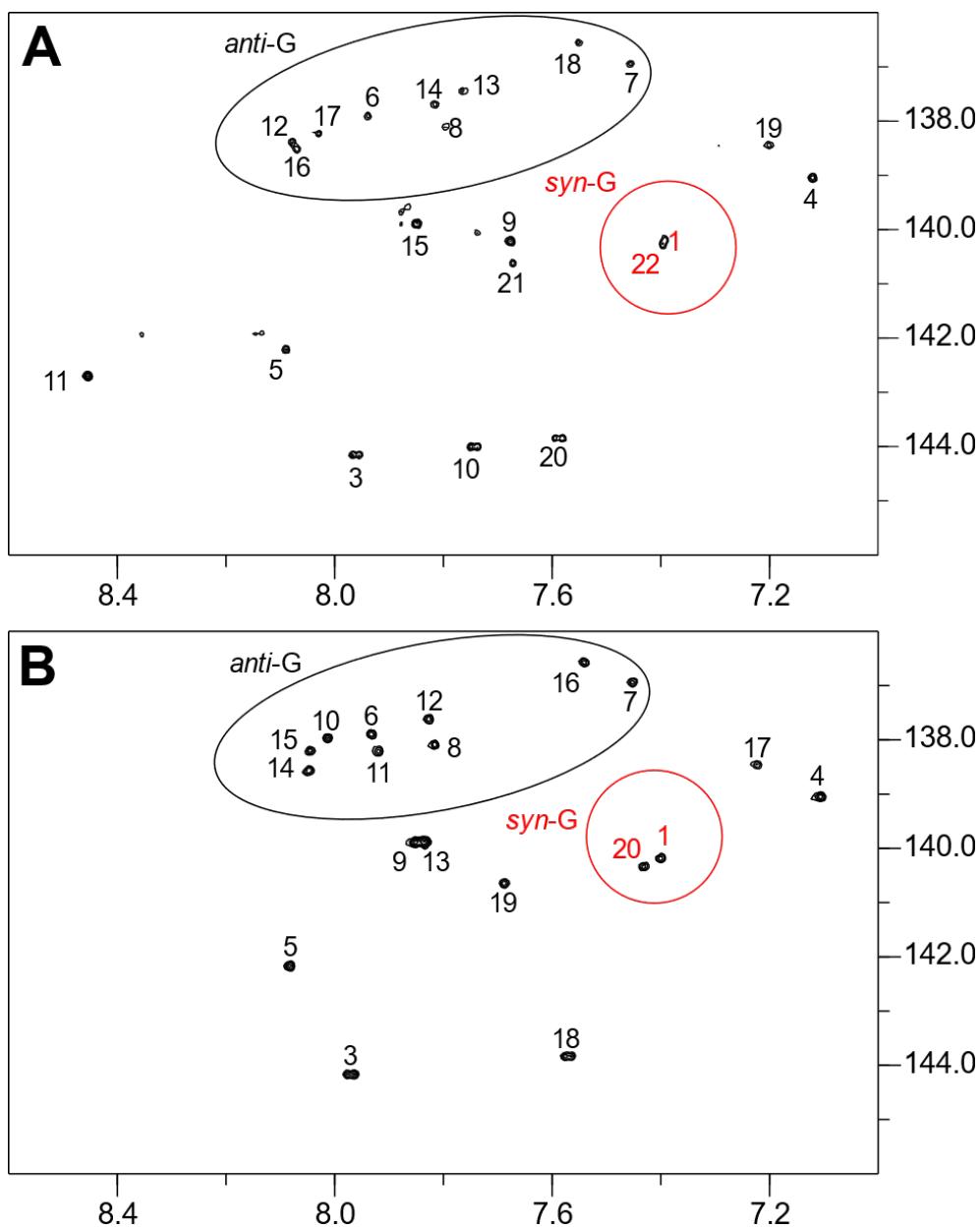


Figure S37. Comparison of ^1H - ^{13}C HSQC spectra of (A) $^{2\text{Br}}\text{Q}$ and (B) $^{2\text{Br}}\text{Q-311-T}$ acquired at 30 °C in 10 mM K^+ buffer, pH 7.0, showing H8/H6–C8/C6 correlations. *Anti-G* and *syn-G* residues of the G-core are circled and labelled in black and red color, respectively. The highly similar pattern of C-H correlations comprising two *syn-G* and 9 *anti-G* residues for the G-core demonstrate the same fold for both sequences. Note the absence of a third *syn-G* residue at position 2 due to $^{\text{Br}}\text{G}$ incorporation.

Article 6

■ G-Quadruplexes

Quadruplex–Duplex Junction: A High-Affinity Binding Site for Indoloquinoline Ligands

Yoanes Maria Vianney, Pit Preckwinkel, Swantje Mohr, and Klaus Weisz*^[a]

Abstract: A parallel quadruplex derived from the *Myc* promoter sequence was extended by a stem-loop duplex at either its 5'- or 3'-terminus to mimic a quadruplex–duplex (Q–D) junction as a potential genomic target. High-resolution structures of the hybrids demonstrate continuous stacking of the duplex on the quadruplex core without significant perturbations. An indoloquinoline ligand carrying an aminoalkyl side chain was shown to bind the Q–D hybrids with a very high affinity in the order $K_a \approx 10^7 \text{ M}^{-1}$ irrespective of the duplex location at the quadruplex 3'- or 5'-end. NMR chemi-

cal shift perturbations identified the tetrad face of the Q–D junction as specific binding site for the ligand. However, calorimetric analyses revealed significant differences in the thermodynamic profiles upon binding to hybrids with either a duplex extension at the quadruplex 3'- or 5'-terminus. A large enthalpic gain and considerable hydrophobic effects are accompanied by the binding of one ligand to the 3'-Q–D junction, whereas non-hydrophobic entropic contributions favor binding with formation of a 2:1 ligand-quadruplex complex in case of the 5'-Q–D hybrid.

Introduction

Sequences with four runs of G-nucleotides can fold into G-quadruplexes (G4s) composed of stacked G-quartets and further stabilized by the coordination of monovalent cations.^[1] G4 forming sequences have been found throughout the genome, with frequent occurrences in telomeres and promoter regions of human oncogenes such as *c-Myc*, *c-Kit*, and *KRAS*.^[2–4] Intramolecular G-quadruplexes with their four G-columns connected by loops show highly diverse topologies. This polymorphism is reflected in different types of loops but may also include discontinuous G-tracts.^[5–10] In general, various G4 topologies in the genome may be specifically targeted by high-affinity ligands for novel pharmaceutical approaches, but quadruplex topologies can also be rationally designed for use in an increasing number of technological applications, for example, as aptamers.^[11–14]

DNA junctions are important elements in cellular maintenance processes.^[15–17] Given that G-quadruplex forming sequences in gene promoter regions originate from duplex DNA,^[18] the presence of quadruplex–duplex (Q–D) motifs seems obvious. Q–D hybrid structures have already been re-

ported 25 year ago^[19] and several variants have since been developed by placing the duplex forming sequence at different internal or external positions of the G4.^[11,20–22] In fact, bioinformatic studies have revealed the abundance of such Q–D hybrid structures in the human genome.^[23] Consequently, Q–D junctions may be considered hotspots of a druggable region in guiding a ligand to bind at the G-quadruplex structure with high selectivity while retaining high affinity. Accordingly, hybrid ligands designed by simply joining known G4 ligands with duplex minor groove binders have been proposed for targeting such type of junctions.^[13,24]

One major challenge for the design of a ligand as specific G4 probe or drug is the selectivity for a quadruplex with respect to other DNA secondary structures but also to a unique quadruplex topology.^[25–27] It is also generally understood that an increase in selectivity, for example, through discrimination of a quadruplex versus duplex species, will often be associated with a decrease in affinity.^[28,29] In previous studies we developed a ligand termed PIQ based on a phenyl-substituted indoloquinoline heterocyclic ring system (Figure 1 A).^[30] This ligand showed high selectivity and affinity towards the parallel *Myc* quadruplex. While stacking on the outer G-tetrad of the *Myc* G4 is favored and associated with the formation of a binding pocket through short overhang sequences, the PIQ ligand was shown to be tightly sandwiched between the 3'-faces of two quadruplexes in a 1:2 complex when binding a 3'-truncated *Myc* sequence.^[31] However, affinity towards duplex DNA is also observed and promoted by its positively charged aminoalkyl substituent, albeit to a small extent.^[30]

In the present paper, a Q–D junction was designed through a duplex stem-loop extension at either the quadruplex 5'- or 3'-terminus (Figure 1). Detailed structural and thermodynamic analysis was employed to characterize the binding behavior of

[a] Y. M. Vianney, P. Preckwinkel, S. Mohr, Prof. Dr. K. Weisz
Institute of Biochemistry, Universität Greifswald
Felix-Hausdorff-Str. 4, 17487 Greifswald (Germany)
E-mail: weisz@uni-greifswald.de

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/chem.202003540>.

© 2020 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

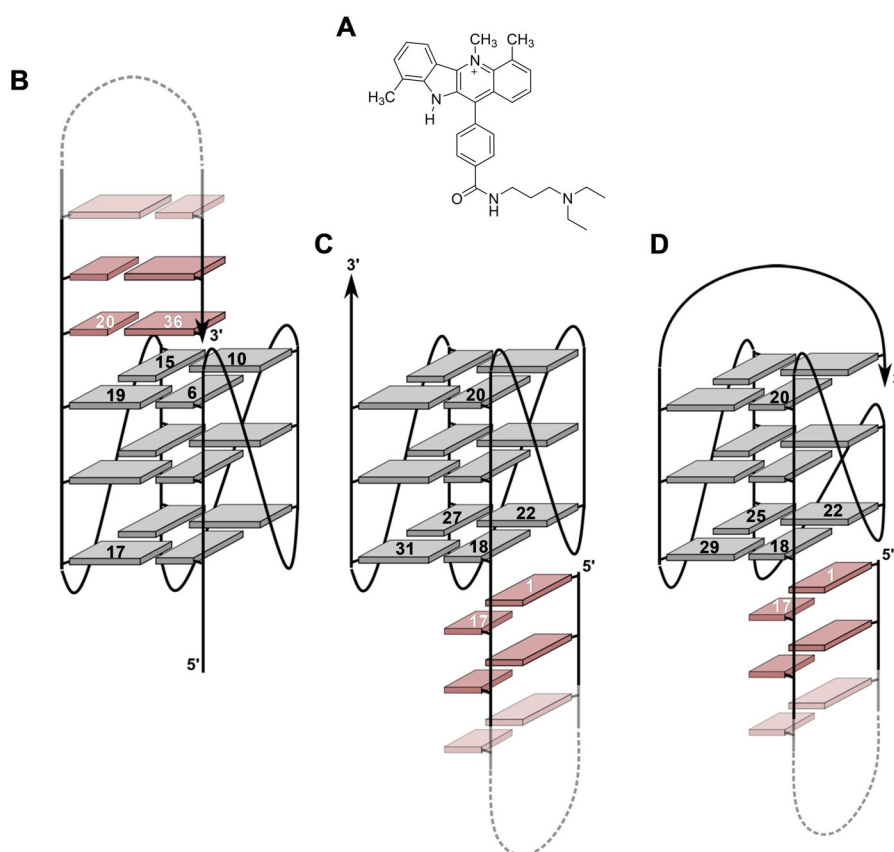


Figure 1. Structure of the PIQ ligand (A). Schematic representation of Q–D hybrids *Myc-dup3* (B), *Myc-dup5* (C), and *Myc3I-dup5* (D) with numbering of residues at the Q–D junction; G-tetrad guanines and base pairs of the duplex hairpin extension are indicated by grey and red rectangles, respectively.

a PIQ ligand as a typical G4 binder. Using the major *Myc* quadruplex but also a modified and extended version with a snap-back loop architecture at its 3'-end as G4 module within the hybrids, the results suggest that PIQ binds selectively and with high affinity to the Q–D junction. However, thermodynamic binding profiles noticeably differ for binding at a junction adjacent to the 5'- or 3'-face of the quadruplex G-core.

Results and Discussion

Circular dichroism spectra

The present design of a quadruplex–duplex hybrid is based on the well-studied parallel G4 derived from the promoter region of the *c-Myc* oncogene.^[32] A long overhang sequence at either the 5'- or 3'-terminus of the G4 core is designed to form a duplex stem-loop structure with a CG base pair adjacent to the 5'- or 3'-outer tetrad of the quadruplex, respectively (Figure 1). In contrast to base pair formation within loop regions of the G-core, the hybrids *Myc-dup5* and *Myc-dup3* are expected to exhibit increased flexibility at the Q–D junction because only a single attachment at one end of the double-helical stem-loop structure links the duplex to the quadruplex motif.

CD spectra of the Q–D hybrids are shown in Figure 2. They exhibit negative and positive maxima at around 240 and 265 nm typical for a parallel quadruplex topology with stacked

G-tetrads of the same polarity and with all-*anti* glycosidic torsion angles for the G-core residues. Because a B-type duplex shares similar CD signatures, the additional presence of a duplex stem-loop structure remains mostly hidden. However, ellipticities in *Myc-dup3* with a double-helical domain at the G4 3'-face are somewhat reduced compared to *Myc-dup5* but also native *Myc* with only short overhangs (Figure S1).^[30] On the other hand, the duplex stem-loop structure at the 5'- or 3'-terminus does not seem to noticeably impact the *Myc* parallel topology.

Upon titrating the PIQ ligand, no major changes were observed in the CD spectra below 300 nm, implying that the topology of both quadruplexes was retained after PIQ binding. Also, the appearance of induced circular dichroism (ICD) effects around the long wavelength absorption maximum of the ligand at 376 nm demonstrates binding of the achiral ligand to the chiral environment of the DNA. It should be noted that small changes in ellipticity at $\lambda < 300$ nm during addition of ligand may indicate small quadruplex conformational adjustments but may also be attributed to ICD effects through short wavelength absorptions of bound PIQ.

ICD effects as a function of increasing ligand concentration give first hints of PIQ binding to the Q–D hybrids. A broad negative ICD band is observed at the start of titration, in line with ligand stacking on an outer tetrad, but is typically also observed for intercalative binding of a planar ligand to duplex DNA.^[33] In subsequent titration steps the amplitude of the ICD

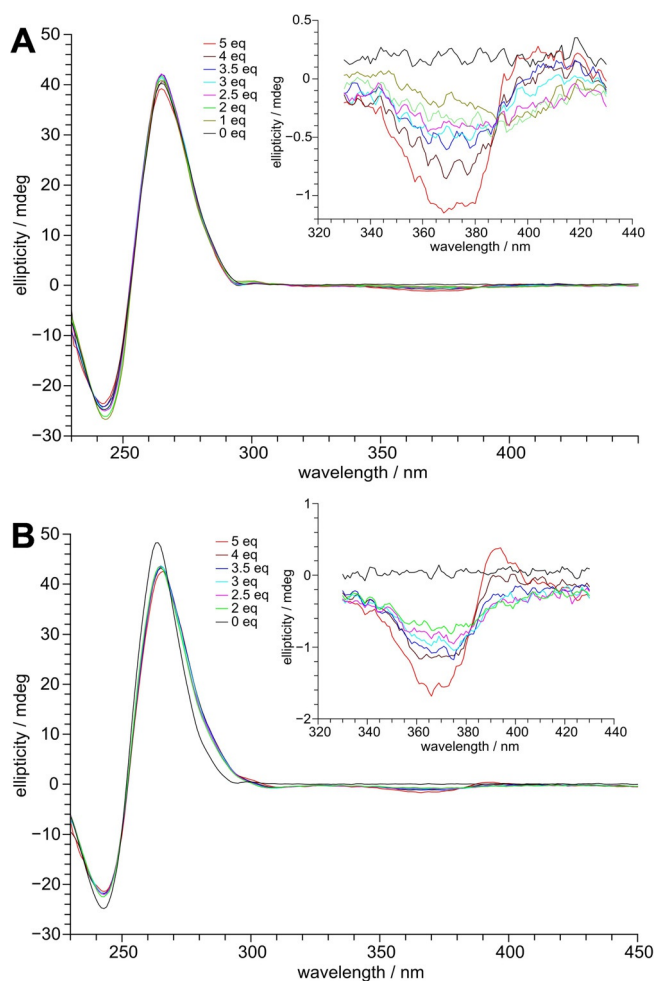


Figure 2. CD spectra of A) *Myc-dup3* and B) *Myc-dup5* following titration with PIQ (0–5 equivalents); the inset shows induced CD effects at the ligand absorption.

band increases at its lower but decreases at its higher wavelength side, becoming even positive for *Myc-dup5* with a five-fold excess of ligand (Figure 2). Such a behavior indicates the gradual onset of an additional binding process characterized by a single positive ICD overlapping the initial band at longer wavelengths. Interestingly, a corresponding ICD is seen for the duplex-free *Myc* G4 known to bind PIQ at its exposed outer tetrads (Figure S1). Alternatively, a developing bisignate ICD due to exciton couplings between bound ligands in close proximity and with negative and positive amplitudes at lower and higher wavelengths, respectively, may superimpose on the initial CD band.

Melting experiments

Bound ligand is expected to affect the stability of DNA as reflected in changes of its melting temperature T_m , which are conveniently determined by temperature-dependent UV and CD experiments. For a separate determination of duplex and quadruplex melting, wavelengths used for the measurements are of particular importance. In UV melting experiments, a hyperchromicity at $\lambda=260$ nm is expected upon the transition

from duplex to single strand whereas the absorbance at $\lambda=295$ nm is most sensitive to quadruplex melting with associated hypochromic effects. However, for quadruplex melting studies in the presence of ligand, temperature-dependent CD experiments were performed to avoid complications due to interferences from ligand absorption below 300 nm but also due to closely similar quadruplex and duplex melting temperatures (see below). With a maximum of its positive band for the parallel G4 topology, a decreased CD signal at 265 nm mostly indicates quadruplex unfolding because ellipticities at this wavelength for the B-type duplex hairpin extensions are less pronounced.

For lowering the high melting temperature of the *Myc* quadruplex, melting experiments were initially performed in a low-salt buffer with 10 mM K^+ . Under these conditions, duplex and quadruplex domains in *Myc-dup5* and *Myc-dup3* melt in two well-separated individual transitions. Melting temperatures for the duplex motifs are 11 and 20 °C below melting of the corresponding quadruplex domains (Table S2 and Figure S2). As expected from the close structural similarity only flanked by a long single-stranded overhang at either the 5'- or 3'-terminus after duplex melting, the UV-derived T_m value for the G4 subunit in *Myc-dup5* was found to be 67 °C and thus only slightly higher by 2 °C when compared to *Myc-dup3*. For a validation of these quadruplex- and duplex-specific results, additional DSC experiments were performed. Deconvolution of the two transitions yielded melting temperatures T_m in full agreement with those obtained from the temperature-dependent absorbances at 295 and 260 nm (Table S2 and Figure S3).

As a consequence of premature duplex melting under the low salt conditions, evaluation of a ligand-induced G4 thermal stabilization suffers from the lack of a defined Q–D junction. We therefore changed the buffer solution from 10 mM K^+ to 120 mM Na^+ , expecting that a higher ionic strength will stabilize the duplex whereas the replacement of potassium by sodium ions will likely destabilize the quadruplex domain without affecting its topology, as demonstrated previously.^[1,30,34] Indeed, as shown by UV and CD melting experiments, relative stabilities of duplex and quadruplex domains change under the new buffer conditions, resulting in a duplex melting more than 10 °C above G4 melting for both hybrids (Table 1 and Figure S2). Also, T_m values for the G4 units in the Na^+ buffer with the flanking duplexes intact differ by 8 °C and suggest more

Table 1. Melting temperatures T_m for the Q–D hybrids without and with addition of 1 equivalent of PIQ in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7.^[a]

Q–D hybrid	T_m (duplex) [°C]		T_m (quadruplex) [°C]	
	w/o PIQ ^[b]	with 1 equiv PIQ ^[b]	w/o PIQ ^[c]	with 1 equiv PIQ ^[c]
<i>Myc-dup3</i>	63.6 ± 0.6	70.8 ± 0.3	44.7 ± 1.8	68.2 ± 1.4
<i>Myc-dup5</i>	67.8 ± 0.6	73.5 ± 1.0	53.2 ± 0.6	69.7 ± 1.2
<i>Myc3l-dup5</i>	69.5 ± 0.5	76.1 ± 0.3	66.3 ± 0.4	73.8 ± 0.4

[a] Averages with standard deviation from three independent measurements except for the quadruplex melting of *Myc3l-dup5* with and without PIQ which was determined in duplicate. [b] T_m data from UV melting experiments. [c] T_m data from CD melting experiments.

stabilizing duplex-G4 interactions at the 5'- when compared to the 3'-outer tetrad.

With the addition of 1 equivalent of ligand in Na⁺ buffer, similar T_m values determined by the temperature dependence of UV absorbances for the duplex and of CD ellipticities for the quadruplex seem to merge at about 70 °C and suggest a single yet rather broad melting transition for both hybrid-ligand complexes given different methods and experimental uncertainties (Table 1). With the duplex domain only moderately stabilized due to ligand binding by $\Delta T_m \leq 7$ °C, the quadruplex shows a significant stabilization by 17 and 23 °C for *Myc-dup5* and *Myc-dup3*, respectively. The higher ligand-induced thermal stabilization of the *Myc-dup3* quadruplex in Na⁺ buffer compensates for its noticeably lower intrinsic stability when

compared to *Myc-dup5*. Also, the increase in melting temperatures for both duplex and quadruplex domains with the addition of ligand in a 1:1 molar ratio points to strong ligand interactions at the Q-D interface.

Thermodynamics of ligand binding to the Q-D hybrids

Isothermal titration calorimetry was used to determine the thermodynamic profile of PIQ binding to the Q-D hybrids. Experiments were performed by titrating the ligand into a 120 mM K⁺ buffer solution of the oligonucleotides. Thermograms obtained after integration of the power output for each injection and corrections for the heats of dilution are shown in Figure 3.

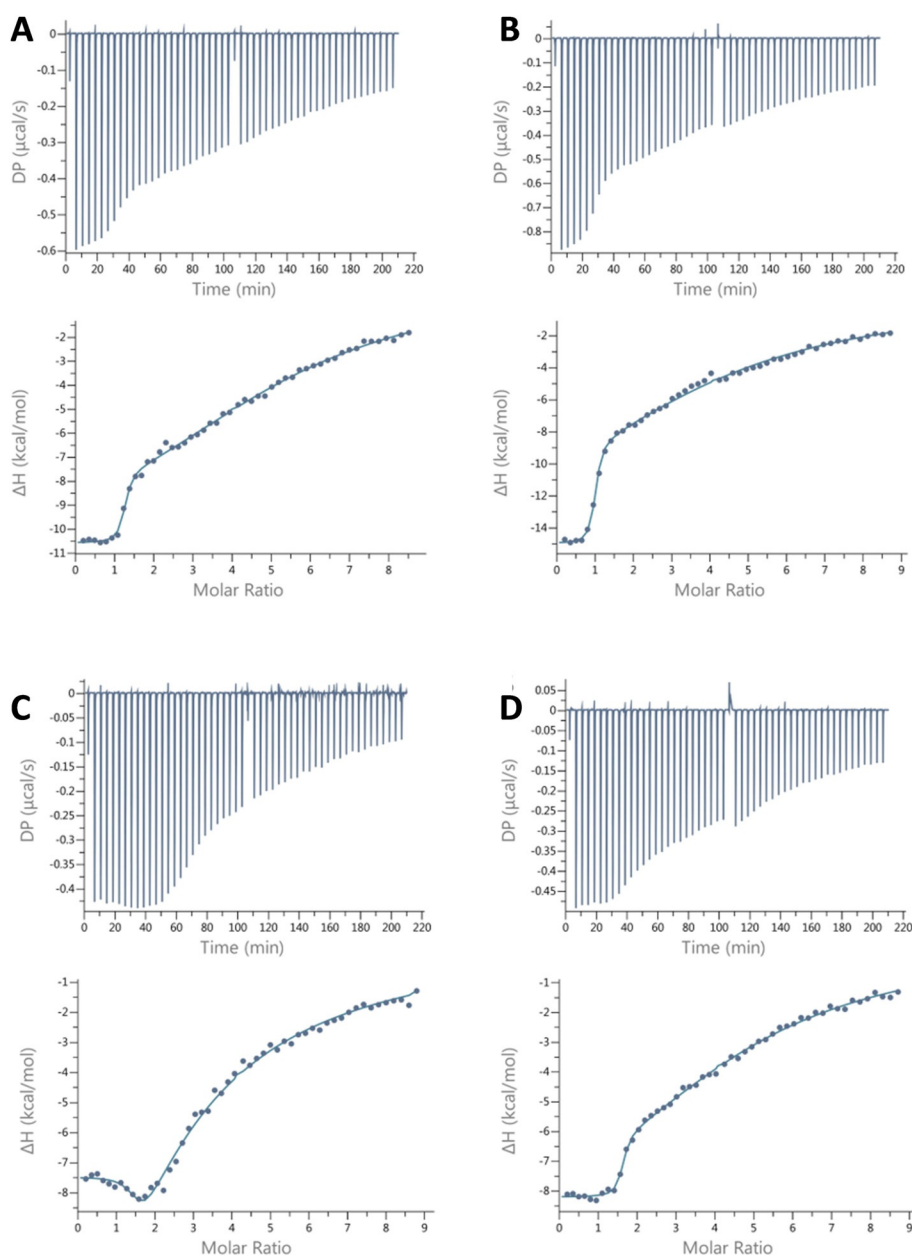


Figure 3. Representative ITC thermograms for the binding of PIQ at 120 mM K⁺ to A) *Myc-dup3* at 40 °C, B) *Myc-dup3* at 50 °C, C) *Myc-dup5* at 40 °C, and D) *Myc3I-dup5* at 40 °C. The upper and lower panels show the heat burst for every injection step and the dilution-corrected heat versus the molar ratio, respectively.

Table 2. ITC-derived thermodynamic parameters for the binding of PIQ to the Q–D hybrids in the presence of 120 mM K⁺ at 40 °C.^[a]

Q–D hybrid	N	K_a [M ⁻¹]	$\Delta H_{\text{fit}}^{\circ}$ [kcal mol ⁻¹] ^[b]	ΔG° [kcal mol ⁻¹] ^[c]	$\Delta H_{\text{es}}^{\circ}$ [kcal mol ⁻¹] ^[b]	$-T\Delta S^{\circ}$ [kcal mol ⁻¹] ^[c]
<i>Myc-dup3</i>	1.2 ± 0.1	(1.6 ± 0.4) × 10 ⁷	-10.7 ± 0.1	-10.3 ± 0.2	-11.9 ± 0.5	1.5 ± 0.5
<i>Myc-dup3</i> ^[d]	1.1 ± 0.2	(1.5 ± 0.5) × 10 ⁷	-15.0 ± 0.8	-10.6 ± 0.2	-14.8 ± 0.2	4.3 ± 0.3
<i>Myc-dup5</i> ^[e]	n.d.	n.d.	n.d.	n.d.	-6.7 ± 0.3	n.d.
<i>Myc3l-dup5</i>	1.5 ± 0.1	(1.6 ± 0.2) × 10 ⁷	-8.2 ± 0.1	-10.3 ± 0.1	-7.3 ± 0.2	-3.0 ± 0.2

[a] Average values with standard deviations from three independent measurements; only values for the high-affinity binding are given. [b] $\Delta H_{\text{fit}}^{\circ}$ and $\Delta H_{\text{es}}^{\circ}$ denote standard molar enthalpy changes determined from curve fitting and from an excess-site method, respectively. [c] From $\Delta G^{\circ} = -RT \ln K_a$ and $-T\Delta S^{\circ} = \Delta G^{\circ} - \Delta H_{\text{es}}^{\circ}$. [d] At 50 °C. [e] Isotherm could not be reliably fitted with a two-site model.

The titration curve for *Myc-dup3* acquired at $T=40$ °C and thus close to physiological temperatures, point to the presence of a single high-affinity binding site with the release of heat upon ligand binding (Figure 3A). A gradual return to baseline at high ligand-to-DNA molar ratios suggests additional weaker binding events and the presence of multiple nonequivalent binding sites. Potential sites include the Q–D junction, the 5'-outer tetrad with its short overhang, and the duplex domain, but unspecific binding, for example, through electrostatic interactions with the negatively charged DNA backbone, should also be considered. Employing a two-site model for curve fitting, excellent and reproducible fits were obtained for the high-affinity binding associated with a very high exothermicity, a stoichiometry of 1, and an association constant of $K_a=1.6 \times 10^7$ M⁻¹ (Table 2). Initial 1:1 complex formation is clearly separated from ensuing binding processes of weaker affinity. Given that more than one subsequent binding event of lower affinity are anticipated to superimpose at later titrations, a second set of thermodynamic parameters for low affinity binding sites through curve fitting was not evaluated in more detail.

To get more insight into potential binding sites, it is instructive to look at closely related DNA-ligand associations. Recently, an association constant K_a of 2×10^6 M⁻¹ and a binding enthalpy ΔH° of -6.3 kcal mol⁻¹ at 40 °C has been determined for PIQ binding to the outer tetrads of the *Myc* quadruplex lacking a duplex extension.^[35] To also assess binding to the duplex domain, we performed ITC titrations of the separate duplex stem-loop structure with the ligand (Figure S4). Here, corresponding binding isotherms indicate a released heat of <5 kcal mol⁻¹ and rather weak binding with estimates of $K_a \leq 10^4$ M⁻¹. These results strongly suggest that the Q–D interface in *Myc-dup3* constitutes the binding site of highest affinity. Binding here is also associated with a very high exothermicity not found in either free quadruplex or hairpin duplex.

Isotherms for *Myc-dup5* at 40 °C exhibit a shallow minimum following a plateau region at initial titration steps and a gradual return to baseline with excess of ligand. It is immediately apparent that such a heat profile indicates the presence of more than two calorimetrically distinct association processes with multiple binding sites at the G4 hybrids. The superposition of more than two binding events during the entire course of titration compromises the extraction of binding parameters through curve fitting for *Myc-dup5* and even restricts the determination of reliable thermodynamic parameters for the site of

highest affinity. To nevertheless obtain a more accurate binding enthalpy for the latter, we employed an excess-site method to yield a ΔH° of about -7 kcal mol⁻¹ at 40 °C (Figure S5). An excess of DNA in this protocol ensures that added ligand is completely bound to high-affinity sites for every injection step and that the area under each power output directly reflects the molar binding enthalpy following normalization.^[36]

Introducing a Q–D hybrid variant with a 3'-snap-back loop

Due to the limitations in analyzing thermograms of *Myc-dup5*, a variant termed *Myc3l-dup5* was designed and tested for ligand binding (Figure 1). *Myc3l-dup5* is based on the quadruplex *Myc3l* that carries an additional 3'-extension together with a central two-nucleotide deletion and was previously shown to fold into a snap-back loop structure. Because the snap-back loop spans the quadruplex 3'-face and effectively prevents the ligand from binding at its 3'-tetrad,^[5,37] elimination of an additional putative binding site should provide for a better resolved thermogram. On the other hand, this variant is expected to closely mimic PIQ binding at the *Myc-dup5* 5'-face with its duplex extension because binding opposite the snap-back loop should be essentially unperturbed (see below). UV and CD melting data of *Myc3l-dup5* are also summarized in Table 1.

When titrating the *Myc3l-dup5* hybrid with the PIQ ligand, a well-defined first binding event with reproducible thermodynamic parameters and an apparent stoichiometry >1 could be extracted (Figure 3D). With a K_a of 1.6×10^7 M⁻¹ it matches with the high-affinity binding of *Myc-dup3* and again suggests binding at the interface of quadruplex and duplex domains. PIQ binding at the 5'-face with no noticeable perturbations by the opposite snap-back loop structure is also strongly suggested by the close binding enthalpies ΔH° of about -7 kcal mol⁻¹ extracted by an excess-site method for the high-affinity binding for both *Myc-dup5* and *Myc3l-dup5*.

Of note, although featuring the same association constant for a proposed binding at the Q–D interface, thermodynamic profiles of PIQ binding are strikingly different for *Myc3l-dup5* and *Myc-dup3*. A significantly more favorable binding enthalpy for *Myc-dup3* is counteracted by a slight loss in entropy whereas a less exothermic binding to *Myc3l-dup5* is associated with a favorable change in entropy at 40 °C to give the same Gibbs free energies (Table 2). Additional ITC titrations for *Myc-dup3* at 50 °C show no significant change in binding constant but a no-

Q–D Hybrid	$\Delta H^\circ_{\text{es},293\text{ K}}$ [kcal mol ⁻¹]	$\Delta H^\circ_{\text{es},303\text{ K}}$ [kcal mol ⁻¹]	$\Delta H^\circ_{\text{es},313\text{ K}}$ [kcal mol ⁻¹]	$\Delta H^\circ_{\text{es},323\text{ K}}$ [kcal mol ⁻¹]	ΔC_p° [cal mol ⁻¹ K ⁻¹]	$\Delta G^\circ_{\text{hyd}}$ ^[b] [kcal mol ⁻¹]
<i>Myc-dup3</i>	-8.1 ± 0.5	-10.1 ± 0.2	-11.9 ± 0.5	-14.8 ± 0.2	-221 ± 19	-17.7
<i>Myc3l-dup5</i>	-5.4 ± 0.1	-6.2 ± 0.2	-7.3 ± 0.2	-8.5 ± 0.2	-101 ± 8	-8.1

[a] Average values for the high-affinity binding with standard deviations from three independent measurements. [b] From the relationship $\Delta G^\circ_{\text{hyd}} = 80 \cdot \Delta C_p^\circ$.

ticeably more negative binding enthalpy of about -15 kcal mol⁻¹ (Figure 3B, Table 2). These results suggest extensive enthalpy-entropy compensation effects with a negative molar heat capacity ΔC_p° exceeding entropic changes ΔS° , as frequently observed for small molecules binding to a biomolecular receptor.^[38]

To determine ΔC_p° for high-affinity binding, binding enthalpies were determined at 20, 30, 40, and 50 °C by an excess-site method and plotted over temperature to give a ΔC_p° of -221 cal mol⁻¹ K⁻¹ and -101 cal mol⁻¹ K⁻¹ for *Myc-dup3* and *Myc3l-dup5*, respectively (Table 3, Figures S6 and S7). Whereas ΔC_p° upon binding *Myc3l-dup5* is close in magnitude to changes in heat capacity effects as obtained for other quadruplex-ligand interactions,^[39,40] ΔC_p° for *Myc-dup3* binding is highly negative, in particular when compared to a $\Delta C_p^\circ = -67$ cal mol⁻¹ K⁻¹ determined for PIQ binding to the *Myc* G4 lacking a duplex extension.^[35] However, a larger change in heat capacity of about -330 cal mol⁻¹ K⁻¹ was reported for minor groove binding of Hoechst 33258 to duplex DNA.^[41]

In general, negative ΔC_p° indicates a reduced solvent-accessible surface area associated with hydrophobic effects by the release of water at the nonpolar solute-solvent interface. Based on solvent-transfer experiments of liquid hydrocarbons, the semi-empirical relationship $\Delta G^\circ_{\text{hyd}} = 80 \cdot \Delta C_p^\circ$ links molar heat capacity changes with the hydrophobic driving force of associ-

ation $\Delta G^\circ_{\text{hyd}}$.^[42] It should be noted that this correlation strictly applies to temperatures near 20 °C and assumes a ΔC_p° that exclusively results from hydrophobic effects. Regardless of such uncertainties, favorable contributions from hydrophobic effects are clearly major contributors for ligand binding to *Myc3l-dup5*. On the other hand, hydrophobic interactions are suggested to play an even more dominant role for ligand binding to *Myc-dup3* with its 3'-duplex extension, exhibiting a $\Delta G^\circ_{\text{hyd}}$ twice as large as determined for *Myc3l-dup5*. With hydrophobic effects being mostly entropic in nature, a less favorable change in total entropy as found for ligand binding to *Myc-dup3* is quite unexpected (Table 2). Apparently, residual entropic contributions differ considerably for these two hybrids and point to more significant reductions in conformational entropy and flexibility in complexes of *Myc-dup3*.

NMR solution structure of Q–D hybrids

Initially all three Q–D hybrids, namely *Myc-dup3*, *Myc-dup5*, and *Myc3l-dup5* were structurally characterized in detail by NMR analysis. ¹H NMR spectra for all three sequences revealed some low-intensity signals of minor species in the imino proton spectral region but these did not hamper proton assignments of the predominant hybrid structure (Figure 4). Thus, 12 major Hoogsteen imino proton resonances in the 10.5–12 ppm

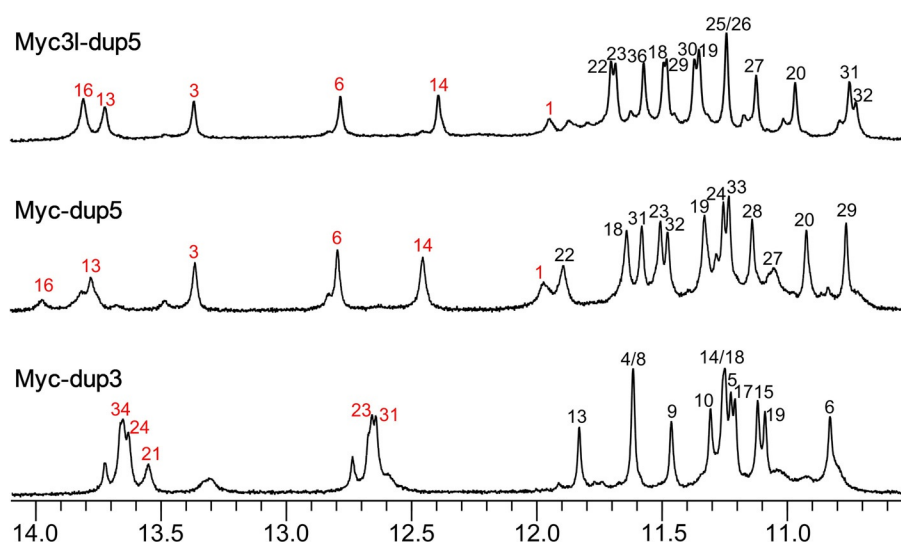


Figure 4. Imino proton spectral region of *Myc-dup3*, *Myc-dup5*, and *Myc3l-dup5*; assigned resonances from the G-core and Watson–Crick stem-loop duplex are indicated by black and red numbers, respectively.

region indicate an intact three-layered G-quadruplex core for *Myc-dup3* and *Myc-dup5*. For *Myc3l-dup5* an additional slow-exchanging imino signal of a guanine base within the snap-back loop is observed, in line with its participation in a capping base triad, as reported previously.^[5,37] The detection of Watson–Crick imino resonances between 12–14 ppm demonstrate complementary base pairing in the stem-loop structure for the 3'- and 5'-flanking sequences. All iminos involved in Watson–Crick hydrogen bonding of the duplex region have been successfully assigned except for the imino signal of the AT base pair adjacent to the T₃ hairpin loop and, in case of *Myc-dup3*, for the imino resonance of the GC base pair at the Q–D junction, likely due to its fast exchange with solvent.

In general, an almost complete resonance assignment following standard strategies based on the analysis of DQF-COSY, ¹H-¹³C HSQC, and 2D NOESY experiments was achieved (Figures S8–S10). Intranucleotide and sequential H6/H8–H1' contacts in NOESY spectra of all three sequences suggest no major structural perturbations of the quadruplex through the formed duplex stem-loop extension. In fact, H6/H8–H1' contacts enable a continuous NOE walk from G17 at the 5'-end of the last G-tract of the quadruplex until the 3'-terminal guanosine G36 of the duplex in *Myc-dup3* and from the 5'-terminal G1 until G20 at the 3'-end of the first G-tract in *Myc-dup5* and *Myc3l-dup5*. Several NOE connectivities involving base and sugar protons between residues in the G-tetrad and the following base pair were observed that define the geometry of the Q–D junction (Table S3). These include a G36 H8–G6 H1 contact in *Myc-dup3*, a G22 H8–G1 H1 contact in *Myc-dup5*, and G1 H8–G22/G25 H1 contacts in *Myc3l-dup5* (Figures S8–S10). Various non-sequential contacts connect C17 of the tetrad-flanking base pair in

Myc-dup5 and *Myc3l-dup5* with G31 and G29 of the adjacent tetrad, respectively, and point to the cytosine being more directed towards the G-core when compared to *Myc-dup3*.

Based on NMR-derived distance and torsion angle restraints, structures were calculated for all three Q–D hybrids (Figure 5, Table S4). Extending from the parallel G-quadruplex domain, the long 3'- and 5'-flanking sequences form a B-type hairpin structure with a T₃ loop. Residues in the short overhangs at the opposite quadruplex face, that is, G2 and A3 in *Myc-dup3* as well as A35 in *Myc-dup5*, cap the tetrad plane. On the other hand, the 3'-overhang in *Myc3l-dup5* forms a snap-back loop that is additionally stabilized by a G32–A34–G35 triad above the 3'-tetrad in seven out of the ten low-energy structures.

It should be noted that the present Q–D hybrids differ from previously reported Q–D architectures that harbor the duplex domain within a G4 loop or form a double helix through complementary 5'- and 3'-overhang sequences.^[11,20,22] With only one strand linked to the G-core, the duplex is expected to exhibit increased flexibility. Nevertheless, with 12–17 NOE-derived distance restraints between base pair and adjacent G-tetrad for each hybrid, the geometry of the Q–D junction is well defined. Interestingly, different stacking patterns between the G–C Watson–Crick base pair and the neighboring quadruplex tetrad are apparent in the hybrid structures (Figure 5). The Q–D junction in *Myc-dup3* is similar to constructs with a double-helical lateral loop (PDB ID 2M8Z and 2M90).^[20] It shows nearly maximum stacking between terminal G36 of the duplex and G6 of the 3'-tetrad with cytosine C20 only poorly stacked on the 5'-sequential guanine G19 (Figure 5A). For the hybrids with the duplex at the 5'-outer tetrad, the CG base pair at the junction is turned towards the G-quadruplex core. As a result, the base-

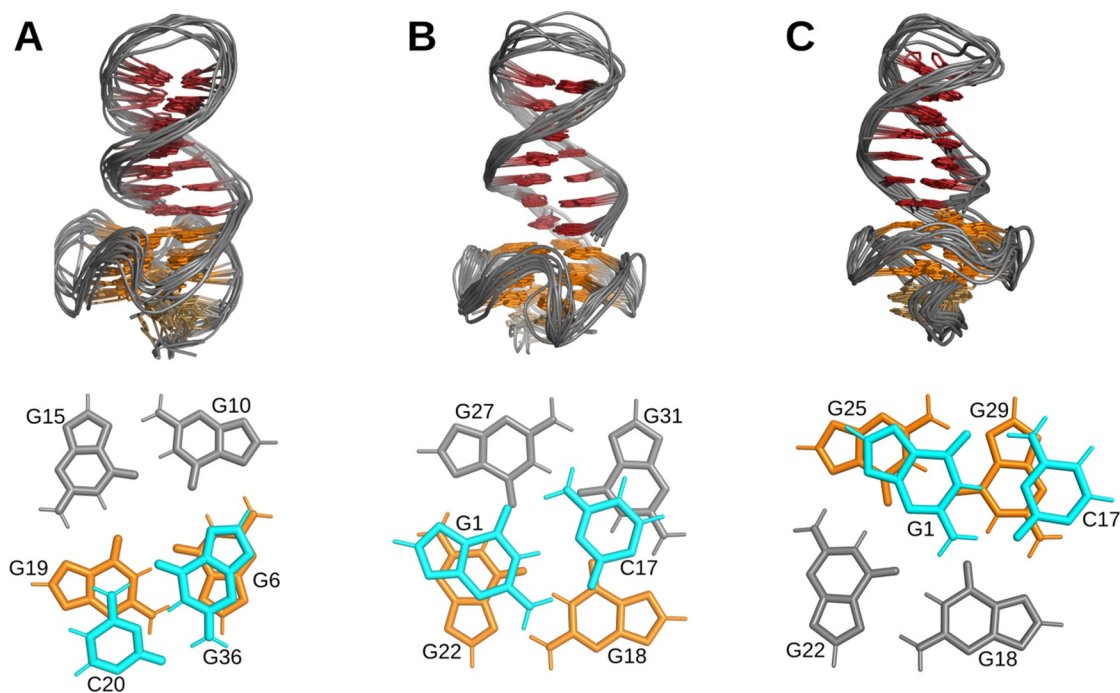


Figure 5. Top: Superposition of ten lowest energy structures of A) *Myc-dup3*, B) *Myc-dup5*, and C) *Myc3l-dup5*. Only the backbone is shown for loop residues. Bottom: Top view of the Q–D junction with stacking interactions between the first duplex CG base pair and the adjacent quadruplex outer tetrad.

paired cytosine exhibits no significant stacking interactions with the 3'-linked guanine of the G-tetrad but is shifted to partially stack on a guanine at the adjacent G-tetrad edge (Figure 5, B and C). Whereas the base pair is centrally stacked on the G-tetrad in *Myc-dup5*, the stacking pattern for the *Myc3I-dup5* hybrid is reminiscent of the previously reported Q-D construct incorporating a duplex with a G-A pair to form a diagonal loop (PDB ID 2M91).^[20]

NMR structural studies on PIQ binding to *Myc-dup3*

Titration of the *Myc-dup3* hybrid with the PIQ ligand is accompanied by the appearance of a new set of G4 imino proton resonances in the 10–12 ppm spectral region (Figure 6A). The

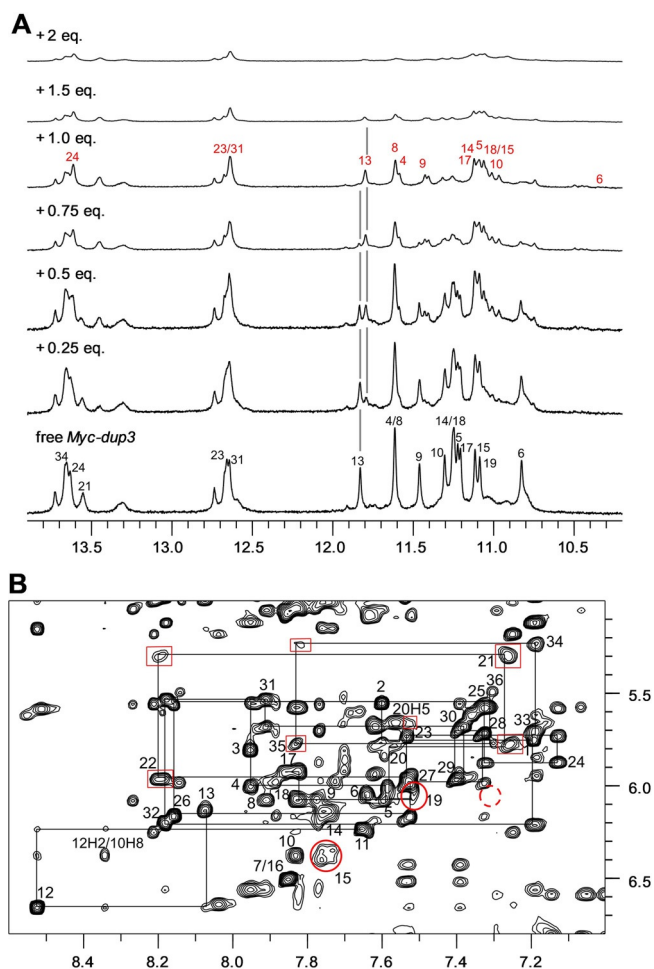


Figure 6. NMR spectra of *Myc-dup3* (0.53 mM) at 20 °C. A) Imino proton spectral region upon addition of the PIQ ligand. Imino proton resonances are assigned to G residues in free *Myc-dup3* and in the ligand-hybrid complex by black and red numbers, respectively. Vertical lines indicate the position of G13 H1 signals with their opposing change in intensity for free and complexed *Myc-dup3*. B) H6/8(ω_2)-H1'(ω_1) spectral region of a 2D NOESY spectrum (300 ms mixing time) in the presence of one equivalent of PIQ. A continuous walk by intranucleotide and sequential NOEs can be traced from G17 of the fourth G-run of the quadruplex until the terminal G36 residue of the duplex. Crosspeaks broadened with respect to the free hybrid spectra are framed by red boxes and red circles for duplex and quadruplex protons, respectively. The dashed red circle indicates the position of a missing contact between residue G36 and G6 at the junction.

latter gradually increase up to a 1:1 molar ratio where free Q-D hybrid resonances have essentially disappeared and imino groups of the newly formed species predominate. Adding ligand in excess, more significant signal broadening indicates exchange processes between complexes at intermediate time-scales.

At a 1:0.5 hybrid-to-ligand molar ratio the spectra point to the coexistence of equally populated free and ligand-bound DNA in slow exchange, which is clearly also demonstrated by two sets of crosspeaks with about equal intensity in a corresponding 2D NOESY spectrum (Figure S11). Unfortunately, ROESY experiments on such samples failed to reveal correlations between the two species, in line with very slow exchange rates (data not shown). Therefore, proton assignments for the complex, making use of standard methodologies for a conventional parallel quadruplex and a B-type duplex, were based on samples with one equivalent of added ligand (Figure 6B). Uninterrupted NOE walks in the H6/8-H1' spectral region can be traced from the 3'-terminal G36 of the duplex extension to G17 in the fourth G-tract of the quadruplex core and also from the 5'-overhang to G6 in the first G-column. Additional NOE connectivities exist between A12 within the propeller loop and guanines in the second and third G-tract. Proton assignments of the complex were completed based on guanine H8-imino and imino-imino NOE contacts of the quadruplex core (Figure S12).

A chemical shift footprint was constructed by plotting ^1H chemical shift changes through ligand binding and these were also mapped on a surface model of the *Myc-dup3* structure by red color of varying intensity (Figure S13, Figure 7). Inspection of the data reveals substantial perturbations of imino and H8 protons for residues located within the 3'-tetrad except for G15. Although being less affected, the H8-H1' intranucleotide NOE crosspeak of the latter seems considerably broadened, possibly due to environments changing at intermediate frequencies (Figure 6B). Likewise, residues of the duplex domain located close to the Q-D junction, that is, C20, T21, and A35, exhibit more significant chemical shift changes. Inspection of

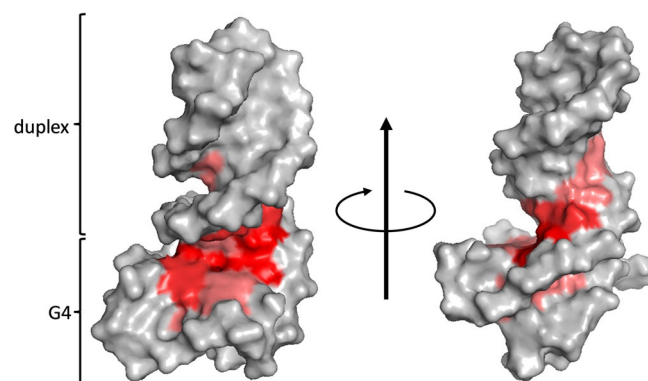


Figure 7. *Myc-dup3* proton chemical shift perturbations after addition of 1 equivalent of PIQ mapped with red color of variable intensity on a surface model of the hybrid. View onto the 3'-tetrad of the quadruplex-duplex junction (left) and rotation around the z-axis with view into the duplex minor groove (right). For a more detailed compilation of chemical shift data see Figure S13 and Table S5 and S6.

the surface model with mapped perturbations immediately identifies the Q–D junction at the 3′-tetrad as major binding site for the ligand with putative additional interactions of the PIQ sidechain within the duplex minor groove in a 1:1 complex (Figure 7). It should be noted, however, that duplex chemical shift changes for non-junction residues may also result from conformational readjustments after ligand binding.

Signal broadening upon the addition of ligand in excess to give a 1:2 hybrid-to-ligand molar ratio severely hampered resonance assignments of the quadruplex domain. However, ¹H resonances of the duplex remained reasonably sharp, which enabled their complete assignment by following walks along H6/H8–H1′ NOE contacts (Figure S14A). Conspicuously, no significant chemical shift differences for assigned duplex proton resonances in *Myc-dup3* spectra were observed when going from a 1:1 to a 1:2 hybrid-to-ligand molar ratio. This also includes residues near the junction and suggests that the ligand neither binds at the duplex nor at the Q–D junction in subsequent titration steps.

Two crosspeaks between G2 and A3 H1′ of the 5′-overhang and imino protons of the 5′-tetrad were observed in the free hybrid and persisted after addition of 1 equivalent of ligand (Figure S14B). However, these two crosspeaks disappeared or considerably shifted after the addition of 2 PIQ equivalents, suggesting signal broadening due to exchange processes and/or conformational readjustments as a result of ligand binding at the 5′-tetrad. Supported by the ITC data and in line with ICD effects in the CD spectra (see Figures 2 and S1), the present NMR data demonstrate a single high-affinity binding site at the Q–D junction followed by the occupation of the 5′-outer tetrad with ligand in excess.

For fixing the bound ligand in a defined orientation, additional assignments of ligand protons and the observation of intermolecular NOE contacts are indispensable. NOESY spectra of the *Myc-dup3* hybrid in the presence of equimolar amounts of ligand exhibit new crosspeaks of a resonance at 10.15 ppm to protons of the 3′-outer tetrad; that is, to G6 H8, G19 H8, and G15 H1 (Figure S12). The deshielded proton at 10.15 ppm likely identifies the indole NH of the indoloquinoline ring system rather than the amide NH of the PIQ sidechain, suggesting stacking of the indoloquinoline on the 3′-tetrad. However, NOE contacts to both G6 H8 and G19 H8 for a single PIQ proton in a unique orientation seems questionable based on interproton distances within the planar tetrad. In fact, various exchange peaks for ligand aromatic signals also including the resonance at 10.15 ppm point to different orientations of bound ligand. Also, no clear NOE contacts are observed between any ligand proton and the duplex domain of *Myc-dup3*. However, some broadening of H8–H1′ crosspeaks for C20, T21, A22, and A35 residues of the stem-loop structure as well as for G15 and G19 of the 3′-tetrad may indicate intermediate exchange between complexes with different ligand orientation, corroborating the absence of a single well-defined complex structure (Figure 6B).

NMR structural studies on PIQ binding to *Myc-dup5* and *Myc3l-dup5*

Upon titrating *Myc-dup5* and *Myc3l-dup5* with ligand, new resonances gradually appeared in analogy to *Myc-dup3* in the G4 imino proton spectral region between 10 and 12 ppm (Figures 8A and S15A). These indicate slowly exchanging species, that is, coexisting free and ligand bound DNA hybrids. Notably, however, signals of free *Myc-dup5* and *Myc3l-dup5* do not vanish but rather persist after the addition of one equivalent of PIQ with roughly equal populations of free and bound species based on imino signal intensities. Such a behavior contrasts sharply with high-affinity binding in a 1:1 stoichiometry to *Myc-dup3* and seems to confirm ITC experiments that consistently indicated higher stoichiometries of $N > 1$ for strong ligand binding to *Myc3l-dup5*.

Other differences in binding to quadruplexes with 5′- and 3′-duplex extensions are apparent when assessing their kinetic behavior. After the addition of 0.5–0.6 equivalent of ligand to *Myc-dup5* or *Myc3l-dup5*, exchange peaks of G4 imino protons can be observed in ROESY experiments as a consequence of faster exchange rates between free and bound species (Figures 8B and S15B). Although resonance assignments for the complex are mostly precluded due to the extensive crowding and overlap of signals in 2D NOESY spectra, exchange peaks allow assignments for most imino protons in the ligand–DNA complex based on assigned resonances of the free hybrid. It should be noted that additional exchange crosspeaks of low intensity in the ROESY spectra point to the presence of minor complex species and therefore only strong exchange peaks indicative of a major complex were used in the construction of chemical shift footprints.

Large upfield shifts are observed for G imino protons at the 5′-outer G4 tetrad in *Myc-dup5* (Figure 8C, Figure S16). Such a pattern clearly demonstrates preferential PIQ binding at the 5′-tetrad and thus at the Q–D junction. On the other hand, minor chemical shift changes for protons in the inner and 3′-tetrad may mostly be due to some conformational adjustments following ligand binding. Additional confirmation of high-affinity binding sites at the quadruplex 5′-face with its duplex extension comes from *Myc3l-dup5* (Figure S15C). As for *Myc-dup5*, chemical shifts of imino protons in the 5′-tetrad experienced most significant perturbations when compared to protons of the inner and 3′-tetrad. Also, a similar chemical shift perturbation pattern was found for *Myc-dup5* and *Myc3l-dup5* upon the addition of less than 1 equivalent of ligand, confirming similar binding characteristics of both hybrids at their 5′-Q–D junction and only a minor impact of the 3′-snap-back loop on binding at the opposite Q–D interface.

Ligand binding and thermodynamic profiles

Association constants for PIQ binding to all Q–D hybrids exceed affinities observed for PIQ binding to the *Myc* quadruplex with only short three-nucleotide flanking sequences or to the corresponding *Myc*-derived G4 with a 3′-snap-back loop by nearly one order of magnitude.^[35,37] Chemical shift perturba-

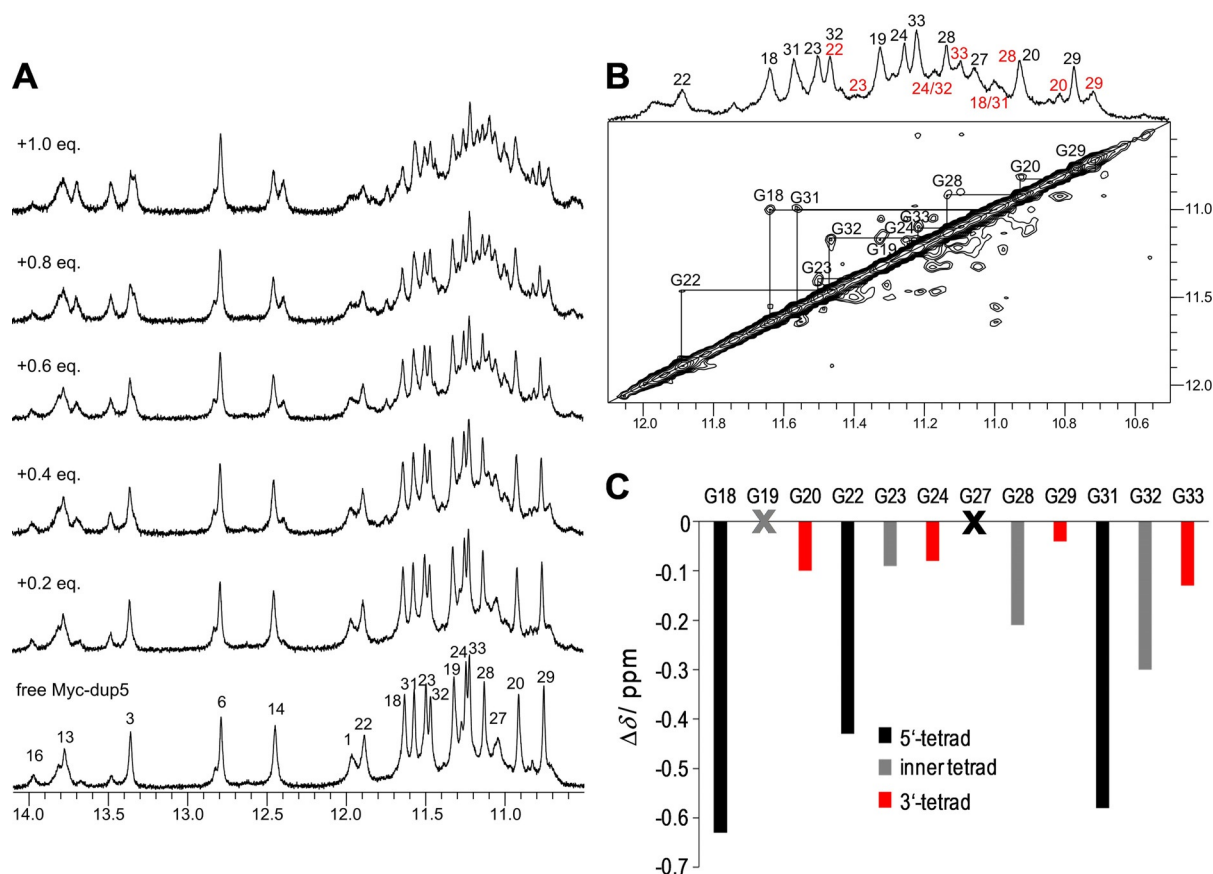


Figure 8. A) Imino proton spectral region of *Myc-dup5* (0.5 mM) titrated with PIQ at 20 °C. B) 1D and ROESY spectrum showing the G4 imino proton spectral region of *Myc-dup5* in the presence of 0.6 equivalent of PIQ at 20 °C. Exchange crosspeaks connect guanine imino resonances of free and complexed species. Imino proton resonances are assigned to G residues in free *Myc-dup5* and in the ligand-hybrid complex by black and red numbers, respectively. C) Imino chemical shift differences of G4 residues between complexed (with 0.6 equiv. of PIQ) and free *Myc-dup5*. Imino resonances of G19 and G27 could not be unambiguously assigned.

tions upon ligand binding clearly identify the Q–D junction as high-affinity binding site. However, Q–D junctions at the 5'- and 3'-face of the quadruplex seem to vary considerably in their stability. Thus, lower melting of the quadruplex domain of free *Myc-dup3* in a sodium-containing buffer, that is, with the duplex extension mostly intact (Table 1), points to higher flexibilities at the junction, as also indicated by a non-observable G36–C20 base pair imino resonance likely due to fraying effects (see above).

In addition to different quadruplex–duplex interactions at the junction of the hybrids, the 5'-outer tetrad of the *Myc* G4 is more hydrophobic and more accessible to additional stacking than the 3'-outer tetrad, often resulting in a noticeable selectivity of ligand binding to *Myc* for one of its two G4 faces.^[43] Likewise, altered binding stoichiometries and binding modes are also suggested by the structural and thermodynamic studies for the 3'- and 5'-junctions. Given the large association constant $K_a > 10^7 \text{ M}^{-1}$ as determined by ITC for the high-affinity binding (Table 2), the presence of equally populated free and ligand bound *Myc-dup5* and *Myc3I-dup5* after the addition of one PIQ equivalent seems only compatible with the cooperative binding of two molecules of ligand or the binding of a ligand dimer to the Q–D hybrids. As a result, only one binding

event is detected by the ITC experiment, although the ligand environment and the strength of ligand-hybrid interactions are anticipated to differ for two PIQ ligands bound to the same G4 receptor. Interestingly, Gibbs free energies for PIQ binding are identical for the 3'- and 5'-duplex hybrids, yet thermodynamic profiles with enthalpic and entropic contributions differ significantly in line with noticeable differences in binding. Although structural details of a 2:1 PIQ-hybrid complex are lacking, binding at or close to the junction is demonstrated by the NMR structural studies. Owing to its planar surface area of only moderate size with the non-fused phenyl substituent rotated out of the indoloquinoline plane, two PIQ molecules can be conceived to stack in a side by side fashion on top of the 5'-outer tetrad of *Myc-dup5* and *Myc3I-dup5* with additional stabilizing interactions provided by the ligand sidechain.

Stacking of the indoloquinoline heterocyclic ring system on the outer G4 tetrad of the Q–D junction is consistent with large upfield shifts of guanine imino protons located within the tetrad plane. Given that high-resolution structures of all the free hybrids show continuous stacking of the duplex extension onto the G-quadruplex, partial insertion of the ligand between base pair and G-tetrad can be assumed. In fact, the loss of a 2D NOESY crosspeak between 3'-terminal G36 H8 and

G6 H1 linking the duplex and quadruplex domain in the *Myc-dup3*–ligand complex corroborates the assumption of a ligand (partially) sandwiched between quadruplex and duplex extension (Figure 6B). To date there is no precedent for a three-dimensional structure with a ligand intercalated between a base pair and G-tetrad at a Q–D junction. Interestingly, however, the G4 ligand 360A has recently been reported to intercalate between GC–GC and GA–GA duplexes within a tetrahelical topology with a flexible central cavity. Binding tightens the structure to create GCGC and GAGA tetrads while increasing the distance between adjacent bases to accommodate the inserted ligand.^[44]

Given high-affinity binding at the Q–D junction for quadruplexes with either 3′- or 5′-duplex extensions, the absence or presence of exchange crosspeaks between free and ligand bound Q–D hybrids in ROESY spectra of *Myc-dup3* and *Myc-dup5* deserves further discussion. In contrast to *Myc-dup5*, exchange rates are slow compared to the mixing time of the ROESY experiment for *Myc-dup3*. Slower ligand dissociation kinetics can be attributed to stronger ligand–DNA interactions, which are also reflected by the more negative binding enthalpy as determined by ITC experiments. Given a more flexible junction in free *Myc-dup3* (see above), the considerable enthalpic gain through ligand binding may be attributed to stronger interactions with the ligand, accompanied by a stiffening of the junction in line with less favorable non-hydrophobic contributions to the binding entropy. It would be tempting to correlate the different thermodynamic binding profiles and associated hydrophobic effects with particular structural characteristics, but different binding stoichiometries and the lack of a well-defined ligand binding mode in *Myc-dup3* and *Myc-dup5*/*Myc3l-dup5* exclude a more detailed assessment of structure–stability relationships.

Conclusions

In addition to the targeting of individual quadruplex architectures, the recognition of a Q–D junction can expand possibilities for a selective DNA targeting by low molecular weight ligands. Such an approach is based on the idea that Q–D junctions are potential hotspots and widely occurring structural elements in the genome. In fact, Q–D junctions may be formed during biological processes associated with the unwinding of a putative G-quadruplex forming sequence either internally as part of a long self-complementary quadruplex loop or externally at the transition from quadruplex to the canonical B-type double-helical structure.

Previous strategies have employed hybrid ligands composed of a quadruplex specific ligand linked with a duplex minor groove binder.^[13,24] However, Q–D junctions may by themselves constitute high-affinity binding sites for known G4 specific ligands, favoring the junction over stacking at exposed outer tetrads of an individual G4 structure as shown here for an indoloquinoline-based ligand. Although not directly clear from their three-dimensional structures, junctions at opposite faces of the G-quadruplex core exhibit pronounced differences in thermal stabilities affected by interactions between the quad-

ruplex and duplex domains. Likewise, thermodynamic profiles for ligand binding are noticeably different for double-helical 3′- and 5′-extensions of the parallel G4. More favorable binding enthalpies, more hydrophobic contributions to the Gibbs free energy through a more negative change in molar heat capacity, and less favorable entropic changes when binding to a Q–D junction at the G4 3′-face are expected to considerably change relative affinities by altering temperatures. It may also hint at a future drug design to selectively address enthalpic and entropic contributions for discriminating between binding at the two quadruplex faces. Together with their high affinity towards appropriate ligands, Q–D junctions may thus constitute targets with a high potential for therapeutic interventions.

Experimental Section

Materials and sample preparation

PIQ was prepared as described previously and its concentration was determined spectrophotometrically by using a molar extinction coefficient $\epsilon_{376} = 22227 \text{ L mol}^{-1} \text{ cm}^{-1}$.^[30] DNA oligonucleotides were purchased from TIB MOLBIOL (Berlin, Germany) and further purified by a potassium acetate–ethanol precipitation. DNA concentrations were determined spectrophotometrically by measuring absorbances A_{260} at 80 °C in water using molar extinction coefficients as supplied by the manufacturer. Prior to usage, oligonucleotides were dried, redissolved in buffer, heated to 90 °C and annealed by slowly cooling to RT. Sequences of oligonucleotides are given in Table S1.

UV/Vis melting experiments

The Q–D hybrid was dissolved in 1.5 mL of either 10 mM potassium phosphate buffer, pH 7, or 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7.0. UV/Vis experiments were performed with a Jasco V-650 spectrophotometer (Jasco, Tokyo, Japan) equipped with a Peltier thermostat. For measurements of the duplex T_m , the final DNA concentration was 2 μM and absorbance was recorded at $\lambda = 260 \text{ nm}$ as a function of temperature (10–90 °C). Data were acquired with a bandwidth of 1 nm and a heating rate of 0.2 °C min⁻¹. For measurements of the quadruplex T_m , the final DNA concentration was 5 μM and absorbance was recorded at $\lambda = 295 \text{ nm}$ between 10 and 90 °C with parameters as given for duplex melting. The melting temperature was determined by the first derivative of the melting curve. For the melting of complexes, the ligand was added up to a 1:1 molar ratio.

CD spectroscopy

CD spectra were measured at 20 °C on the Q–D hybrids (5 μM) in 100 mM KCl, 20 mM potassium phosphate buffer, pH 7.0. A concentrated PIQ solution in DMSO was added up to a 5:1 ligand-to-DNA molar ratio. The DMSO concentration of the mixture was always < 1%. All measurements were performed with a Jasco J-810 spectropolarimeter equipped with a Peltier thermostat (Jasco, Tokyo, Japan). Spectra were recorded for solutions in 1 cm quartz cuvettes from 230 to 450 nm with a bandwidth of 1 nm, a scanning speed of 50 nm min⁻¹, a response time of 4 s, and five accumulations. Prior to measurements, the quadruplex–ligand mixtures were stirred for 10 min to ensure equilibration.

For the determination of melting temperatures, the Q–D hybrid (5 μM) in the absence or presence of 1 equivalent of PIQ was redis-

solved in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7.0. Ellipticities were recorded at $\lambda = 265$ nm between 20 and 95 °C with a bandwidth of 1 nm and a heating rate of 0.2 °C min⁻¹. Melting temperatures were determined by the first derivative of the melting curve.

Isothermal titration calorimetry

ITC experiments were performed with a Microcal PEAQ ITC micro-calorimeter (Malvern Instruments, United Kingdom) employing a reference power of 4 $\mu\text{cal s}^{-1}$. Oligonucleotides and the PIQ ligand were each dissolved in 100 mM KCl, 20 mM potassium phosphate buffer, pH 7.0, supplemented with 5% DMSO. The PIQ solution (400 μM) was titrated to 20 μM of oligonucleotide with a total of 2 \times 26 injections of 1.5 μL each, an injection duration of 3 s, and a spacing between injections of 240 s. The first injection (0.4 μL) was rejected during the fitting process. Excess-site titrations were performed to determine model-independent binding enthalpies directly from averages of peak integrals of the power outputs. For each of the 12 titration steps, 3 μL of ligand solution (200 μM) with a 6 s injection duration were titrated to the oligonucleotide solution (100 μM) with a spacing between injections of 300 s. The first injection volume (0.4 μL) was not included in the calculation of binding enthalpies. Measurements at a temperature range of 20–50 °C were performed to determine the change of heat capacity upon binding. All experiments were blank- and concentration-corrected. For data analysis, the MicroCal PEAQ-ITC analysis software was used.

NMR spectroscopy

NMR spectra were acquired with a Bruker Avance 600 MHz spectrometer equipped with an inverse ¹H/¹³C/¹⁵N/¹⁹F quadruple resonance cryoprobehead and z-field gradients. Data were processed in TopSpin 4.0.7 and assigned in CcpNmr V2.^[45] Oligonucleotides were dissolved in 10 mM potassium phosphate buffer, pH 7.0. Upon titration with a PIQ solution in [D₆]DMSO, the final DMSO concentration with addition of 2 equivalents of ligand was below 5%. Proton chemical shifts were referenced through the water chemical shift taking into account its temperature dependence at pH 7 and carbon chemical shifts were referenced to DSS through an indirect referencing method. Residual HOD in D₂O solutions for DQF-COSY experiments was suppressed by presaturation. A WATERGATE w5 sequence was generally used for water suppression in solutions of 90% H₂O/10% D₂O except for ¹H-¹³C HSQC experiments employing a 3–9–19 pulse sequence. The latter experiments were acquired with a spectral width of 7500 Hz in the F1 dimension, 4 K \times 500 data points, and a 1 s recycle delay. Zero-filling gave a 4 K \times 1 K data matrix that was multiplied with a sine-bell squared window function in both dimensions. Homonuclear 2D spectra were typically recorded with 2 K \times 1 K data points with a relaxation delay of 2 s. Prior to Fourier transformation, FID data were zero-filled to give a final 4 K \times 1 K data matrix and processed with a sine-bell squared window function in both dimensions. 2D NOESY spectra were acquired with mixing times from 80 to 300 ms and ROESY spectra were recorded with a mixing time of 80 ms.

Structure calculation and molecular modeling

Starting structures (100) of lowest energy were selected out of 200 structures generated by a simulated annealing protocol in XPLOR-NIH 2.52.^[46] Distance restraints were obtained from crosspeak intensities in NOESY spectra. Distances were categorized as follows: 2.9 \pm 1.1 Å for strong crosspeaks, 4.0 \pm 1.5 Å for crosspeaks of

medium intensity, 5.5 \pm 1.5 Å for weak crosspeaks, and 6.0 \pm 1.5 Å for very weak crosspeaks. For overlapped peaks, the distance was set to 5.0 \pm 2.0 Å. All χ torsion angles were set as either anti (170–310°) or syn (25–95°) while all sugar puckers were set to the south domain (pseudorotational angle 144–180°). Planarity restraints were employed for tetrads and base pairs.

Restrained simulated annealing was performed using AMBER16 with the parmbsc force field and OL15 modifications.^[47] In vacuo refinement was done for 100 starting structures to yield 20 converged structures by initially equilibrating the system at 300 K for 5 ps followed by heating the system to 1000 K during 10 ps and keeping the temperature for the next 30 ps. The system was cooled to 100 K within 45 ps and finally to 0 K within 10 ps. Restraint energies for simulated annealing in AMBER were 40 kcal mol⁻¹ Å⁻² for NOE based distance restraints, 50 kcal mol⁻¹ Å⁻² for hydrogen-bond based distance restraints, 200 kcal mol⁻¹ rad⁻² for dihedral angle restraints, and 30 kcal mol⁻¹ Å⁻² for tetrad and base pair planarity restraints.

Refinement in water was done by initially neutralizing the system with potassium ions and placing two potassium ions in the inner core of the quadruplex flanked by two tetrad layers. The system was soaked with TIP3P water in a 10 Å truncated octahedral box, initially minimized with 500 steps of steepest descent minimization followed by another 500 steps of conjugate gradient minimization. Both quadruplex and duplex domains were fixed with a force constant of 25 kcal mol⁻¹ Å⁻². The system was heated under constant volume from 100 to 300 K in 10 ps. The system was further equilibrated under a constant pressure of 1 atm with energy restraints decreasing to 5, 4, 3, 2, 1, and 0.5 kcal mol⁻¹ Å⁻². A final simulation was done at 1 atm and 300 K for 4 ns. Snapshots were taken for every 1 ps, the trajectory was averaged for the last 500 ps and shortly minimized in vacuo to obtain 10 lowest-energy structures.

Chemical shift perturbations upon ligand binding were mapped on a surface model of the hybrids by coloring residues depending on chemical shift differences of their imino, base and H1' protons. Each proton was assigned a value based on the maximum change in chemical shift $\Delta\delta_{\text{max}}$ observed for resonances of the same type that was set to 100%. This was followed by averaging and grouping into three classes with decreasing averaged perturbations that were assigned a red color of decreasing intensity.

Accession codes

Atomic coordinates and lists of chemical shifts have been deposited for *Myc-dup3* (PDB ID 6ZL2, BMRB ID 34524), *Myc-dup5* (PDB ID 6ZL9, BMRB ID 34525), and *Myc3l-dup5* (PDB ID 6ZTE, BMRB ID 34533).

Acknowledgements

This research was supported by the Deutsche Forschungsgemeinschaft (INST 292/138-1). Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: calorimetry · DNA structures · G-quadruplexes · NMR spectroscopy

- [1] T. Fujii, P. Podbevšek, J. Plavec, N. Sugimoto, *J. Inorg. Biochem.* **2017**, *166*, 190–198.
- [2] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
- [3] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.
- [4] R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, D. Tannahill, S. Balasubramanian, *Nat. Genet.* **2016**, *48*, 1267–1272.
- [5] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173.
- [6] V. T. Mukundan, A. T. Phan, *J. Am. Chem. Soc.* **2013**, *135*, 5017–5028.
- [7] M. Lenarčič Živkovič, J. Rozman, J. Plavec, *Angew. Chem. Int. Ed.* **2018**, *57*, 15395–15399; *Angew. Chem.* **2018**, *130*, 15621–15625.
- [8] T. H. A. Truong, F. R. Winnerdy, A. T. Phan, *Angew. Chem. Int. Ed.* **2019**, *58*, 13834–13839; *Angew. Chem.* **2019**, *131*, 13972–13977.
- [9] B. Karg, S. Mohr, K. Weisz, *Angew. Chem. Int. Ed.* **2019**, *58*, 11068–11071; *Angew. Chem.* **2019**, *131*, 11184–11188.
- [10] K. B. Wang, J. Dickerhoff, G. Wu, D. Yang, *J. Am. Chem. Soc.* **2020**, *142*, 5204–5211.
- [11] I. Russo Krauss, V. Spiridonova, A. Pica, V. Napolitano, F. Sica, *Nucleic Acids Res.* **2016**, *44*, 983–991.
- [12] T. Bing, W. Zheng, X. Zhang, L. Shen, X. Liu, F. Wang, J. Cui, Z. Cao, D. Shangguan, *Sci. Rep.* **2017**, *7*, 15467.
- [13] T. Q. N. Nguyen, K. W. Lim, A. T. Phan, *Sci. Rep.* **2017**, *7*, 11969.
- [14] M. Kovačič, P. Podbevšek, H. Tateishi-Karimata, S. Takahashi, N. Sugimoto, J. Plavec, *Nucleic Acids Res.* **2020**, *48*, 3975–3986.
- [15] A. L. Brogden, N. H. Hopcroft, M. Searcey, C. J. Cardin, *Angew. Chem. Int. Ed.* **2007**, *46*, 3850–3854; *Angew. Chem.* **2007**, *119*, 3924–3928.
- [16] C. E. Carr, L. A. Marky, *J. Am. Chem. Soc.* **2017**, *139*, 14443–14455.
- [17] K. Duskova, J. Lamarche, S. Amor, C. Caron, N. Queyriaux, M. Gaschard, M. J. Penouilh, G. De Robillard, D. Delmas, C. H. Devillers, A. Granzhan, M. P. Teulade-Fichou, M. Chavarot-Kerlidou, B. Therrien, S. Britton, D. Monchaud, *J. Med. Chem.* **2019**, *62*, 4456–4466.
- [18] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
- [19] R. F. Macaya, J. A. Waldron, B. A. Beutel, H. Gao, M. E. Joesten, M. Yang, R. Patel, A. H. Bertelsen, A. F. Cook, *Biochemistry* **1995**, *34*, 4478–4492.
- [20] K. W. Lim, A. T. Phan, *Angew. Chem. Int. Ed.* **2013**, *52*, 8566–8569; *Angew. Chem.* **2013**, *125*, 8728–8731.
- [21] I. Russo Krauss, S. Ramaswamy, S. Neidle, S. Haider, G. N. Parkinson, *J. Am. Chem. Soc.* **2016**, *138*, 1226–1233.
- [22] K. W. Lim, T. Q. N. Nguyen, A. T. Phan, *J. Am. Chem. Soc.* **2014**, *136*, 17969–17973.
- [23] 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.
- [24] S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Eur. J.* **2018**, *24*, 4428–4435.
- [25] C. C. Chang, I. C. Kuo, I. F. Ling, C. T. Chen, H. C. Chen, P. J. Lou, J. J. Lin, T. C. Chang, *Anal. Chem.* **2004**, *76*, 4490–4494.
- [26] L. Scaglioni, R. Mondelli, R. Artali, F. R. Sirtori, S. Mazzini, *Biochim. Biophys. Acta* **2016**, *1860*, 1129–1138.
- [27] W. Liu, C. Lin, G. Wu, J. Dai, T. C. Chang, D. Yang, *Nucleic Acids Res.* **2019**, *47*, 11931–11942.
- [28] A. Funke, K. Weisz, *Biochimie* **2019**, *157*, 142–148.
- [29] M. Zuffo, A. Guédin, E. D. Leriche, F. Doria, V. Pirota, V. Gabelica, J. L. Mergny, M. Freccero, *Nucleic Acids Res.* **2018**, *46*, e115.
- [30] A. Funke, J. Dickerhoff, K. Weisz, *Chem. Eur. J.* **2016**, *22*, 3170–3181.
- [31] A. Funke, B. Karg, J. Dickerhoff, D. Balke, S. Müller, K. Weisz, *ChemBioChem* **2018**, *19*, 505–512.
- [32] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [33] E. W. White, F. Tanious, M. A. Ismail, A. P. Reszka, S. Neidle, D. W. Boykin, W. D. Wilson, *Biophys. Chem.* **2007**, *126*, 140–153.
- [34] E. Stellwagen, J. M. Muse, N. C. Stellwagen, *Biochemistry* **2011**, *50*, 3084–3094.
- [35] A. Funke, K. Weisz, *J. Phys. Chem. B* **2017**, *121*, 5735–5743.
- [36] I. Haq, T. C. Jenkins, B. Z. Chowdhry, J. Ren, J. B. Chaires, *Meth. Enzymol.* **2000**, *323*, 373–405.
- [37] L. Schnarr, J. Jana, P. Preckwinkel, K. Weisz, *J. Phys. Chem. B* **2020**, *124*, 2778–2787.
- [38] P. Gilli, V. Ferretti, G. Gilli, P. A. Borea, *J. Phys. Chem.* **1994**, *98*, 1515–1518.
- [39] D. S. Pilch, C. M. Barbieri, S. G. Rzuczek, E. J. LaVoie, J. E. Rice, *Biochimie* **2008**, *90*, 1233–1249.
- [40] A. Basu, G. Suresh Kumar, *J. Chem. Thermodyn.* **2016**, *98*, 208–213.
- [41] I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins, J. B. Chaires, *J. Mol. Biol.* **1997**, *271*, 244–257.
- [42] J. H. Ha, R. S. Spolar, M. T. Record, *J. Mol. Biol.* **1989**, *209*, 801–816.
- [43] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [44] A. Kotar, V. Kocman, J. Plavec, *Chem. Eur. J.* **2020**, *26*, 814–817.
- [45] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Protein* **2005**, *59*, 687–696.
- [46] C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
- [47] M. Zgarbová, S. Jir, M. Otyepka, T. E. Cheatham, R. Galindo-murillo, P. Jurec, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.

Manuscript received: July 29, 2020

Revised manuscript received: September 21, 2020

Accepted manuscript online: September 25, 2020

Version of record online: November 16, 2020

Chemistry–A European Journal

Supporting Information

Quadruplex–Duplex Junction: A High-Affinity Binding Site for Indoloquinoline Ligands

Yoanes Maria Vianney, Pit Preckwinkel, Swantje Mohr, and Klaus Weisz^{*[a]}

Table of Contents

Table S1: Sequences of oligonucleotides	S2
Table S2: Melting temperatures of <i>Myc-dup3</i> and <i>Myc-dup5</i>	S2
Table S3: NOE-based distance restraints	S3
Table S4: NMR restraints and structural statistics	S5
Table S5: ¹ H chemical shifts of free <i>Myc-dup3</i>	S6
Table S6: ¹ H chemical shifts of <i>Myc-dup3</i> with one equivalent PIQ	S7
Figure S1: CD spectra of <i>Myc</i> upon PIQ addition	S9
Figure S2: Optical melting curves of <i>Myc-dup3</i> , <i>Myc-dup5</i> , and <i>Myc3l-dup5</i>	S10
Figure S3: DSC melting curves of <i>Myc-dup3</i> and <i>Myc-dup5</i>	S11
Figure S4: ITC thermograms of PIQ binding to duplex hairpins	S12
Figure S5: Excess-site ITC thermogram of PIQ titrated to <i>Myc-dup5</i>	S12
Figure S6: Excess-site ITC titrations of PIQ to <i>Myc-dup3</i> and <i>Myc3l-dup5</i>	S13
Figure S7: Plot of ΔH° over temperature for PIQ binding to the Q-D hybrids	S14
Figure S8: NOESY spectral regions of <i>Myc-dup3</i>	S15
Figure S9: NOESY spectral regions of <i>Myc-dup5</i>	S16
Figure S10: NOESY spectral regions of <i>Myc3l-dup5</i>	S17
Figure S11: NOESY spectral region of <i>Myc-dup3</i> with 0.5 eq. of PIQ	S18
Figure S12: NOESY spectral regions of <i>Myc-dup3</i> with 1 eq. of PIQ	S19
Figure S13: ¹ H chemical shift footprints of <i>Myc-dup3</i>	S20
Figure S14: NOESY spectral regions of <i>Myc-dup3</i> with addition of PIQ	S21
Figure S15: Spectra and chemical shift footprints of <i>Myc3l-dup5</i> upon PIQ titration	S22
Figure S16: Surface model of <i>Myc-dup5</i> with chemical shift perturbations	S23

Table S1. Sequences of oligonucleotides; duplex hairpin domains within the Q-D hybrids are underlined.

name	sequence
<i>Myc</i>	5'-TGA GGG T GGG TA GGG T GGG TAA
<i>Myc-dup3</i>	5'-TGA GGG T GGG TA GGG T GGG <u>CTAGTCA TTT TGACTAG</u> -3'
<i>Myc-dup5</i>	5'- <u>GATCAGT TTT ACTGATC</u> GGG T GGG TA GGG T GGG TA-3'
<i>Myc3l-dup5</i>	5'- <u>GATCAGT TTT ACTGATC</u> GGG T GG T GGG T GGG GAAGG-3'
<i>Dup3</i>	5'-CTAGTCA TTT TGACTAG-3'
<i>Dup5</i>	5'-GATCAGT TTT ACTGATC-3'

Table S2. UV- and DSC-derived melting temperatures T_m of *Myc-dup3* and *Myc-dup5* in 10 mM potassium phosphate buffer, pH 7.^[a]

		T_m from UV-vis (°C)		T_m from DSC (°C)	
<i>Myc-dup3</i>	duplex	45.1 ± 0.3	1 st transition	45.9 ± 0.1	
	quadruplex	65.1 ± 0.2	2 nd transition	64.5 ± 0.4	
<i>Myc-dup5</i>	duplex	56.1 ± 0.3	1 st transition	56.5 ± 0.1	
	quadruplex	66.9 ± 0.5	2 nd transition	66.5 ± 0.1	

[a] Averages with standard deviations from three independent experiments.

Table S3. NOE-based distance restraints for residues at the quadruplex-duplex junction in *Myc-dup3*, *Myc-dup5*, and *Myc3l-dup5*.

<i>Myc-dup3</i>		
duplex	quadruplex	distance (Å)
C20-H5	G19-H8	4.0 ± 1.5
C20-H5	G19-H2'	5.5 ± 1.5
C20-H5	G19-H2''	4.0 ± 1.5
C20-H5	G19-H1'	4.0 ± 1.5
C20-H5	G19-H3'	5.5 ± 1.5
C20-H6	G19-H1'	4.0 ± 1.5
C20-H6	G19-H2'	4.0 ± 1.5
C20-H6	G19-H2''	2.9 ± 1.1
C20-H6	G19-H3'	5.5 ± 1.5
G36-H8	G6-H1'	5.5 ± 1.5
G36-H1'	G6-H1'	4.0 ± 1.5
G36-H2''	G6-H1'	2.9 ± 1.1
G36-H1'	G6-H2'	4.0 ± 1.5
G36-H1'	G6-H2''	4.0 ± 1.5
G36-H8	G6-H1	4.0 ± 1.5
G36-H2'	G6-H1	6.0 ± 1.5
G36-H2''	G6-H1	6.0 ± 1.5
<i>Myc-dup5</i>		
duplex	quadruplex	distance (Å)
C17-H1'	G18-H8	4.0 ± 1.5
C17-H3'	G18-H8	4.0 ± 1.5
C17-H6	G18-H8	5.5 ± 1.5
C17-H5	G31-H8	6.0 ± 1.5
G1-H1'	G22-H8	5.5 ± 1.5
C17-H2'	G31-H1	5.5 ± 1.5

C17-H2''	G31-H1	6.0 ± 1.5
C17-H6	G31-H1	5.5 ± 1.5
G1-H1'	G22-H1	4.0 ± 1.5
G1-H2''	G22-H1	6.0 ± 1.5
G1-H1	G22-H8	5.5 ± 1.5
G1-H1	G31-H1	4.0 ± 1.5
<i>Myc3l-dup5</i>		
duplex	quadruplex	distance (Å)
G1-H8	G25-H8	5.5 ± 1.5
C17-H1'	G18-H8	4.0 ± 1.5
C17-H2'	G18-H8	4.0 ± 1.5
C17-H2''	G18-H8	2.9 ± 1.1
C17-H3'	G18-H8	5.5 ± 1.5
C17-H6	G18-H8	5.5 ± 1.5
C17-H6	G29-H1'	5.5 ± 1.5
G1-H2'	G22-H1	5.5 ± 1.5
G1-H8	G22-H1	6.0 ± 1.5
G1-H1'	G22-H1	4.0 ± 1.5
G1-H2''	G22-H1	5.5 ± 1.5
G1-H8	G25-H1	6.0 ± 1.5
C17-H6	G29-H1	6.0 ± 1.5
C17-H1'	G29-H1	4.0 ± 1.5

Table S4. NMR restraints and structural statistics of calculated structures.

	<i>Myc-dup3</i>	<i>Myc-dup5</i>	<i>Myc3l-dup5</i>
NOE distance restraints:			
intraresidual	161	151	156
interresidual	245	223	270
exchangeable	67	73	99
other restraints			
hydrogen bonds	82	82	82
dihedral angles	72	35	36
planarity	10	10	10
structural statistics:			
pairwise heavy atom			
RMSD value (Å)			
G-tetrad core	0.84 ± 0.18	0.91 ± 0.26	1.01 ± 0.17
all residues	2.70 ± 0.46	2.37 ± 0.36	2.64 ± 0.39
NOE violations (Å)			
maximum violation	0.144	0.197	0.366
mean NOE violation	0.0010 ± 0.0004	0.0016 ± 0.0005	0.0018 ± 0.0011
deviations from idealized			
geometry			
bond lengths (Å)	0.01 ± 0.0001	0.01 ± 0.0001	0.01 ± 0.0001
bond angles (degree)	2.20 ± 0.03	2.17 ± 0.03	2.19 ± 0.03

Table S5. List of ^1H chemical shifts (in ppm) of free *Myc-dup3*; only those protons are listed that were used for a chemical shift footprint.^[a,b]

residue	imino	H6/H8	H2/H5/Me	H1'
T1	n.d.	7.20	1.64	5.76
G2	n.d.	7.61	-	5.57
A3	-	7.99	n.d.	5.82
G4	11.62	7.96	-	6.04
G5	11.22	7.59	-	6.08
G6	10.83	7.55	-	6.19
T7	n.d.	7.85	1.98	6.52
G8	11.62	7.96	-	6.13
G9	11.46	7.78	-	6.17
G10	11.31	7.79	-	6.36
T11	n.d.	7.67	1.94	6.25
A12	-	8.53	8.34	6.66
G13	11.83	8.08	-	6.16
G14	11.25	7.80	-	6.24
G15	11.12	7.79	-	6.41
T16	n.d.	7.84	1.98	6.50
G17	11.21	7.89	-	5.99
G18	11.25	7.88	-	6.10
G19	11.09	7.72	-	6.00
C20	-	7.56	5.43	6.07
T21	13.55	7.33	1.61	5.40
A22	-	8.25	7.31	6.02
G23	12.66	7.59	-	5.79
T24	13.64	7.16	1.24	5.91
C25	-	7.33	5.61	5.57
A26	-	8.16	n.d.	6.17
T27	n.d.	7.52	1.78	5.99
T28	n.d.	7.33	1.58	5.74
T29	n.d.	7.40	1.62	5.97
T30	n.d.	7.38	1.82	5.69
G31	12.65	7.92	-	5.54
A32	-	8.18	7.78	6.21

C33	-	7.18	5.14	5.71
T34	13.66	7.17	1.43	5.23
A35	-	7.85	7.45	5.92
G36	n.d.	7.28	-	5.47

[a] At 20 °C in 10 mM potassium phosphate buffer, pH 7. [b] n.d.: not determined.

Table S6. List of ¹H chemical shifts (in ppm) of *Myc-dup3* after addition of one equivalent PIQ; only those protons are listed that were used for a chemical shift footprint.^[a,b]

residue	imino	H6/H8	H2/H5/Me	H1'
T1	n.d.	7.19	1.63	5.76
G2	n.d.	7.60	-	5.56
A3	-	7.95	-	5.81
G4	11.61	7.95	-	6.01
G5	11.08	7.58	-	6.07
G6	10.32	7.64	-	6.05
T7	n.d.	7.85	1.99	6.51
G8	11.61	7.91	-	6.08
G9	11.42	7.77	-	6.08
G10	11.00	7.83	-	6.38
T11	n.d.	7.65	1.94	6.24
A12	-	8.52	8.32	6.66
G13	11.80	8.07	-	6.14
G14	11.11	7.76	-	6.17
G15	11.06	7.76	-	6.40
T16	n.d.	7.85	1.99	6.50
G17	11.12	7.85	-	5.93
G18	11.06	7.82	-	6.09
G19	n.d.	7.52	-	6.08
C20	-	7.56	5.66	5.78
T21	n.d.	7.26	1.57	5.31
A22	-	8.20	n.d.	5.97
G23	12.65	7.53	-	5.74
T24	13.61	7.13	1.21	5.88
C25	-	7.32	5.59	5.57
A26	-	8.16	n.d.	6.17

T27	n.d.	7.52	1.78	5.99
T28	n.d.	7.32	1.58	5.72
T29	n.d.	7.40	1.63	5.97
T30	n.d.	7.38	1.81	5.68
G31	12.64	7.91	-	5.54
A32	-	8.18	7.77	6.21
C33	-	7.19	5.12	5.71
T34	n.d.	7.19	1.45	5.23
A35	-	7.82	n.d.	5.77
G36	n.d.	7.30	-	5.50

[a] At 20 °C in 10 mM potassium phosphate buffer, pH 7. [b] n.d.: not determined.

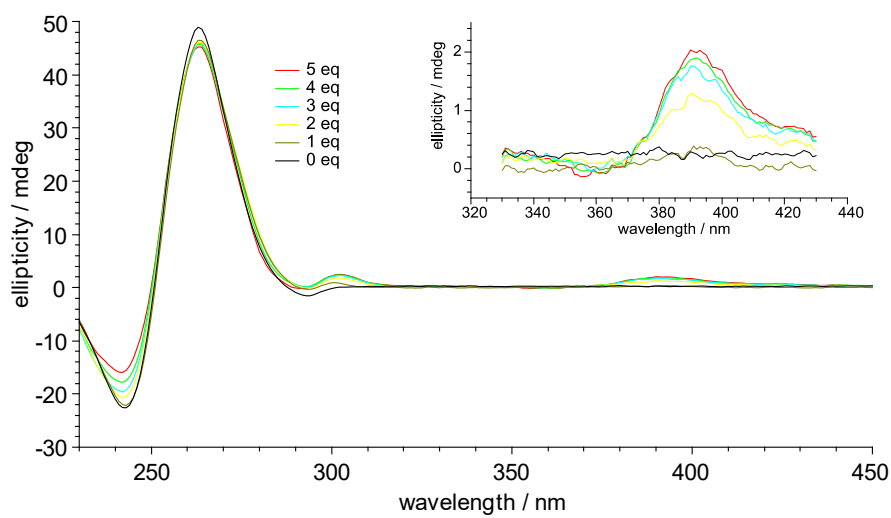


Figure S1. CD spectrum of *Myc* (5 μM) following titration with PIQ (0-5 equivalents) in 100 mM KCl, 20 mM potassium phosphate buffer, pH 7.0; the inset shows induced CD effects at the ligand absorption (from ref. 30).

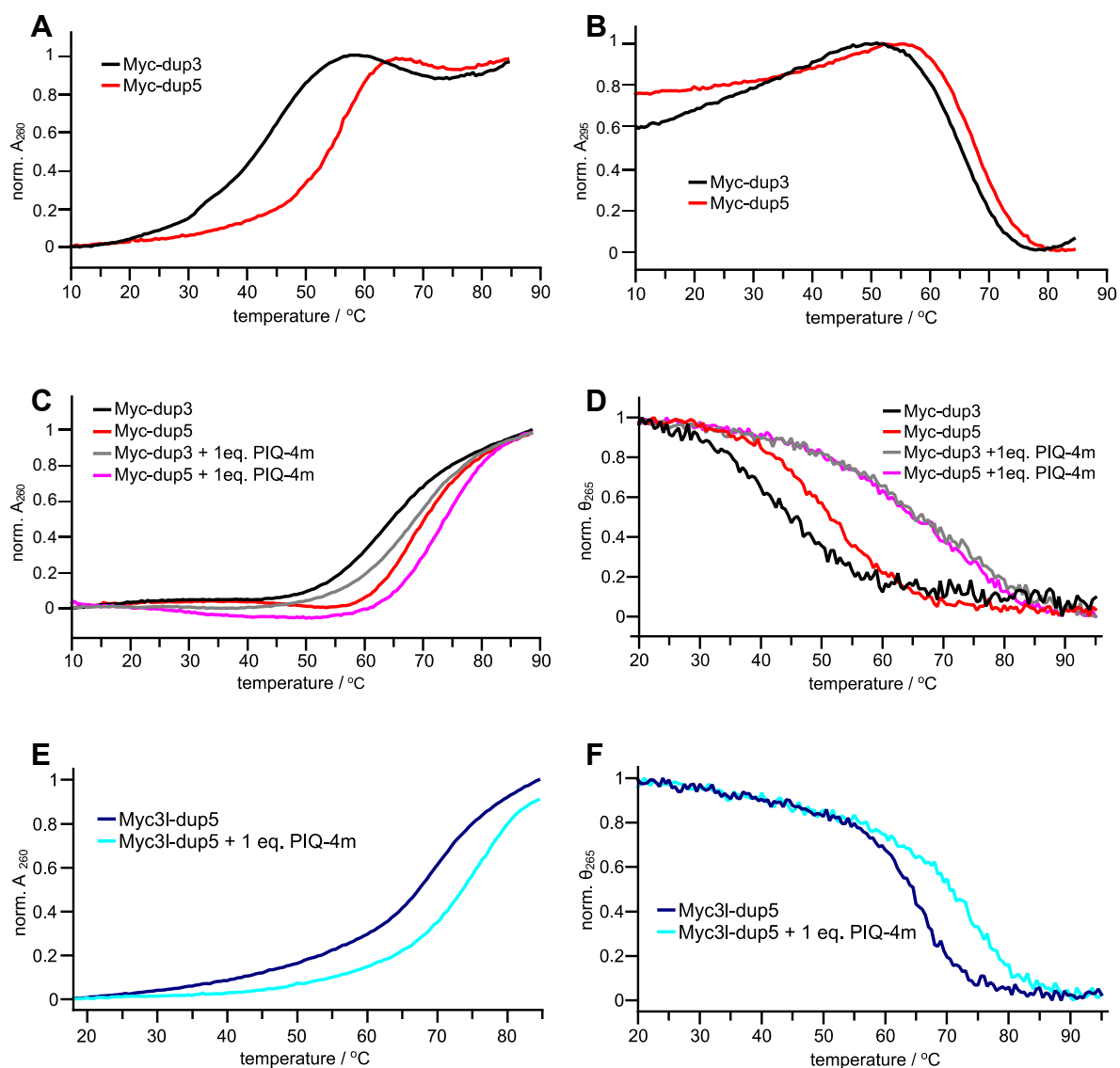


Figure S2. Representative optical melting curves of *Myc-dup3*, *Myc-dup5*, and *Myc3l-dup5*. (A) Normalized absorbances at 260 nm and (B) at 295 nm of *Myc-dup3* and *Myc-dup5* without ligand in 10 mM potassium phosphate buffer, pH 7. (C) Absorbance at 260 nm of *Myc-dup3* and *Myc-dup5* reflecting duplex melting without and with the addition of ligand in a 1:1 molar ratio in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7. (D) Ellipticity at 265 nm of *Myc-dup3* and *Myc-dup5* reflecting quadruplex melting without and with 1 equivalent of added ligand in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7. (E) Absorbance at 260 nm of *Myc3l-dup5* reflecting duplex melting without and with the addition of ligand in a 1:1 molar ratio in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7. (F) Ellipticity at 265 nm of *Myc3l-dup5* reflecting quadruplex melting without and with 1 equivalent of added ligand in 100 mM NaCl, 20 mM sodium phosphate buffer, pH 7.

Differential scanning calorimetry

DSC experiments were performed with a VP-DSC instrument (Malvern Instruments, United Kingdom). The oligonucleotide (50 μM) was dissolved in 10 mM potassium phosphate buffer, pH 7.0. The solution was heated with a heating rate of 0.5 $^{\circ}\text{C}/\text{min}$. Data for a buffer versus buffer scan were subtracted from data obtained for a sample versus buffer scan. A cubic baseline was constructed and melting temperatures were determined from the peaks following deconvolution of the melting transitions. Data were analyzed with the Origin software.

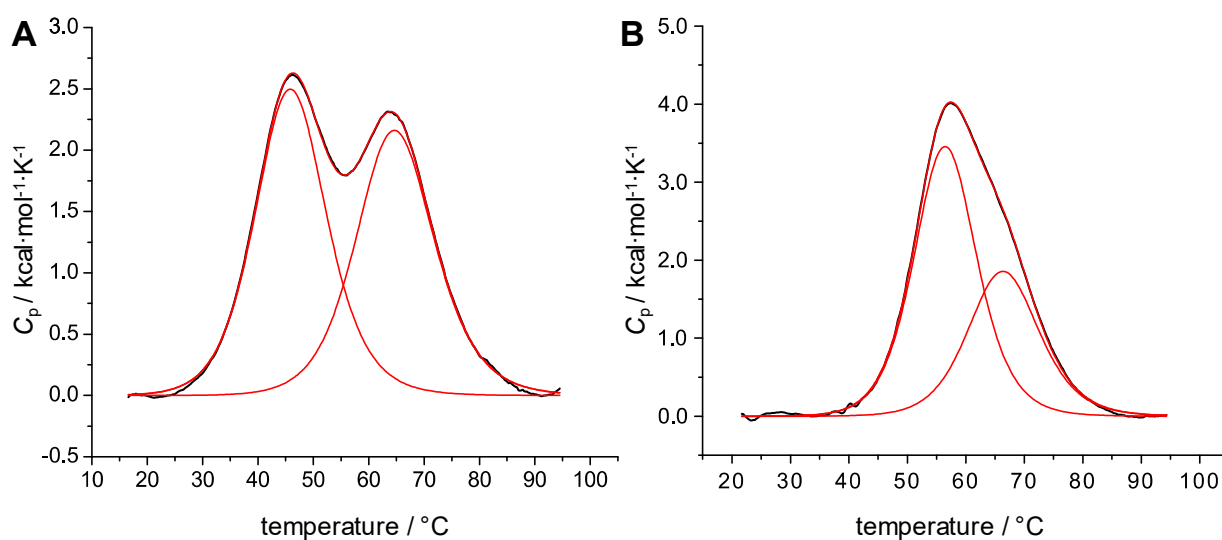


Figure S3. Representative DSC melting curves of (A) *Myc-dup3* and (B) *Myc-dup5* in 10 mM potassium phosphate buffer, pH 7. Thermograms were analyzed and fitted based on two transitions.

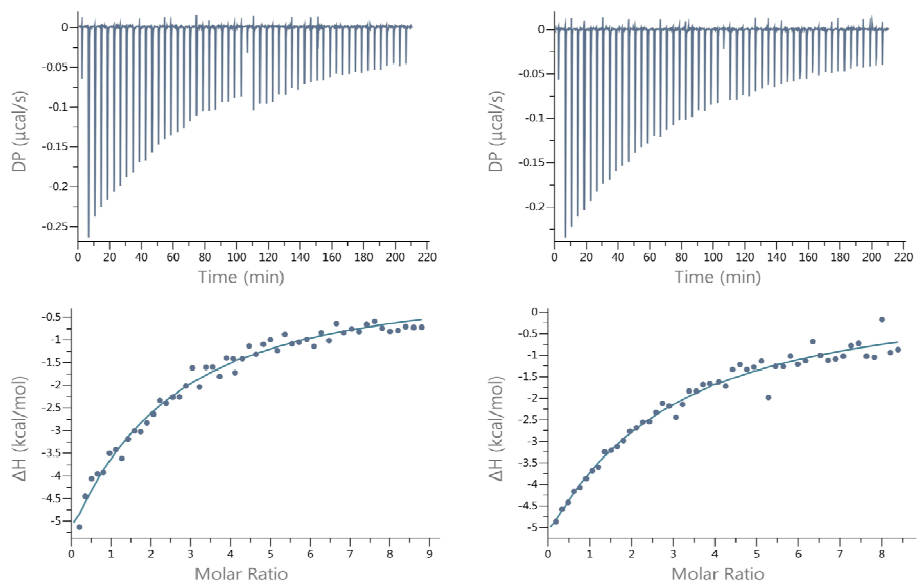


Figure S4. ITC thermograms of PIQ binding to duplex hairpins *Dup5* (left) and *Dup3* (right) at 40 °C. The upper and lower panel shows the heat burst for every injection step and the blank-corrected integrated heat versus molar ratio.

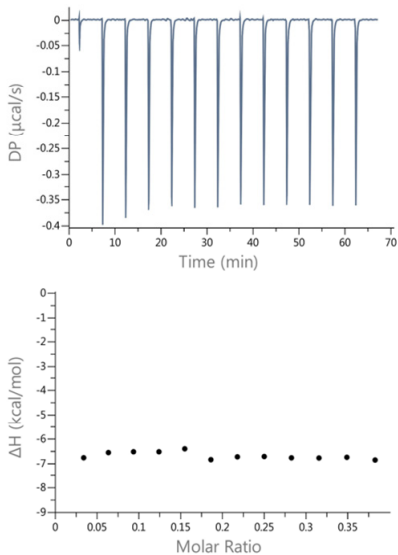


Figure S5. Representative excess-site ITC thermogram of PIQ titrated to *Myc-dup5* at 40 °C. The upper and lower panel shows the heat burst for every injection step and the blank-corrected integrated heat versus molar ratio.

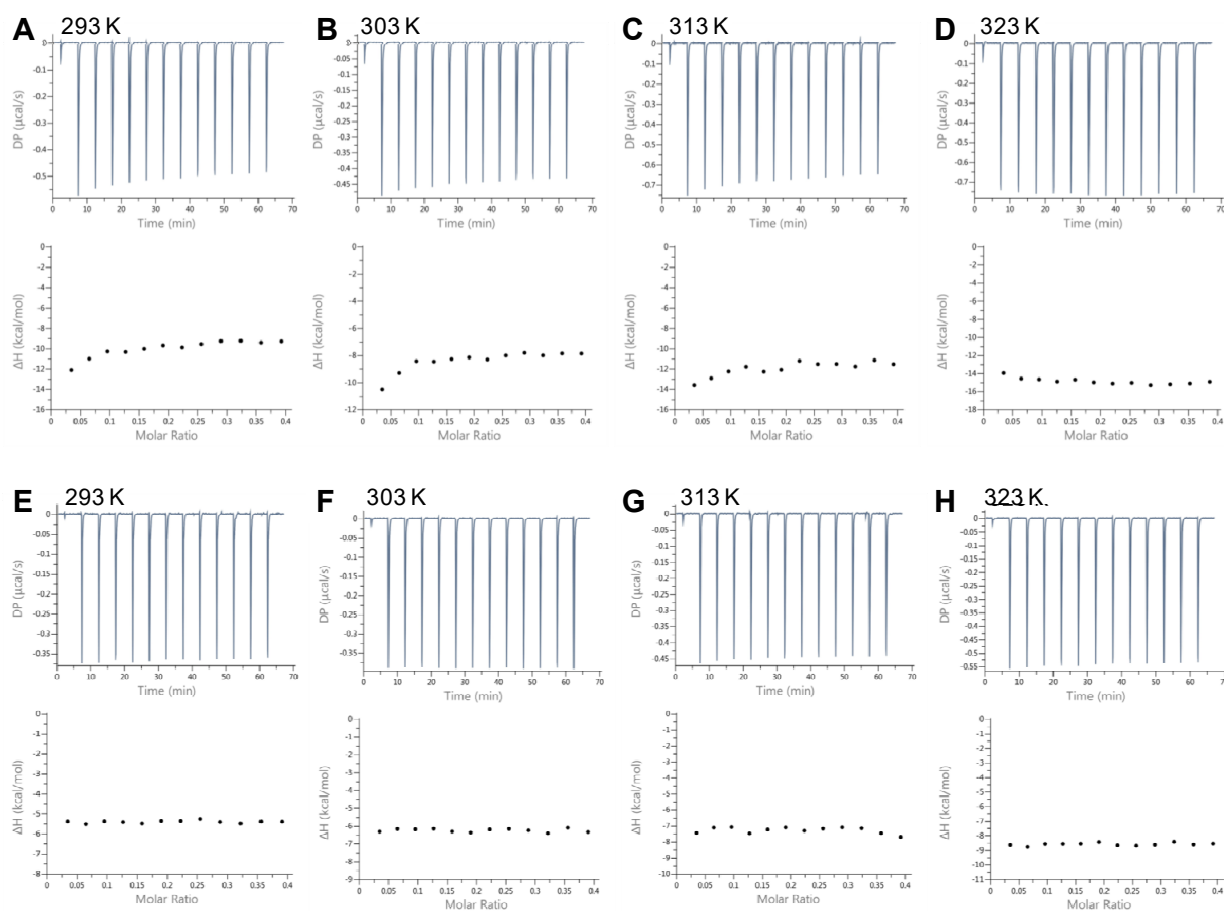


Figure S6. Representative excess-site ITC titrations of PIQ to *Myc-dup3* (A-D) and *Myc3l-dup5* (E-H) at different temperatures. The upper and lower panel shows the heat burst for every injection step and the blank-corrected integrated heat versus molar ratio.

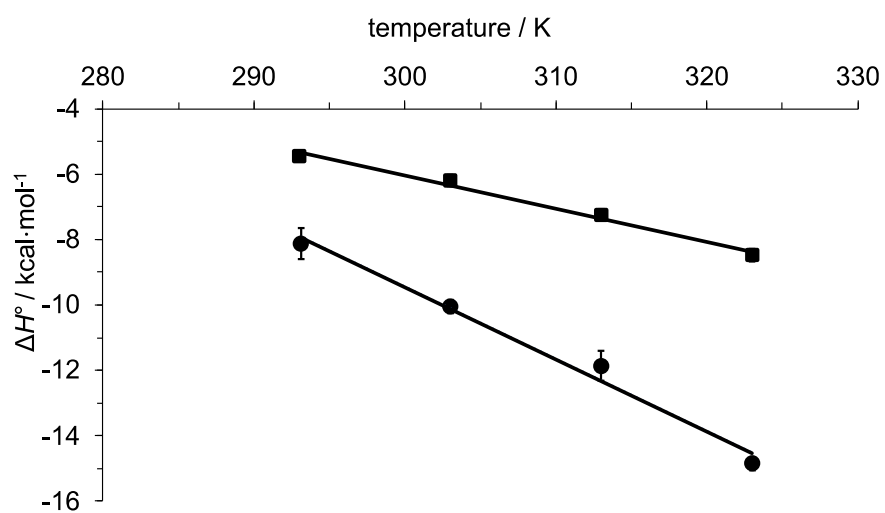


Figure S7. Plot of ΔH° as obtained from excess-site experiments for PIQ binding to *Myc-dup3* (circles) and *Myc3l-dup5* (squares) over temperature. ΔC_p° is given by the slope of the least squares regression line.

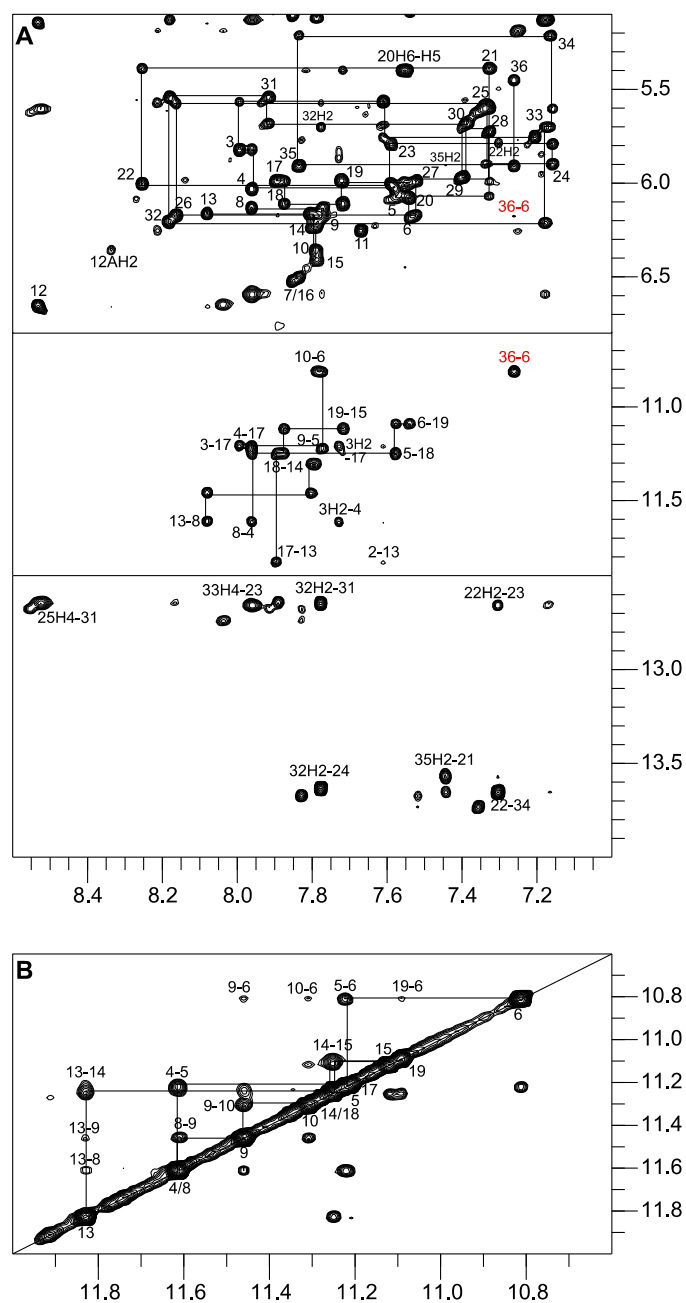


Figure S8. 2D NOESY spectral regions of *Myc-dup3*. (A) H6/H8(ω_2)-H1'(ω_1) (top), H6/H8(ω_2)-Hoogsteen imino(ω_1) (center), and AH2/CNH₂(ω_2)-WC imino(ω_1) connectivities (bottom). An uninterrupted NOE walk through H6/H8-H1' contacts can be followed from G17 along tract IV of the quadruplex to G36 at the duplex 3'-terminus (top); G36 H8 shows weak and strong crosspeaks (labeled in red) to G6 H1' (top) and to the G6 imino at the Q-D junction (center), respectively. (B) Imino-imino connectivities. Spectra were acquired with a 300 ms mixing time at 20 °C in 10 mM potassium phosphate buffer, pH 7.0.

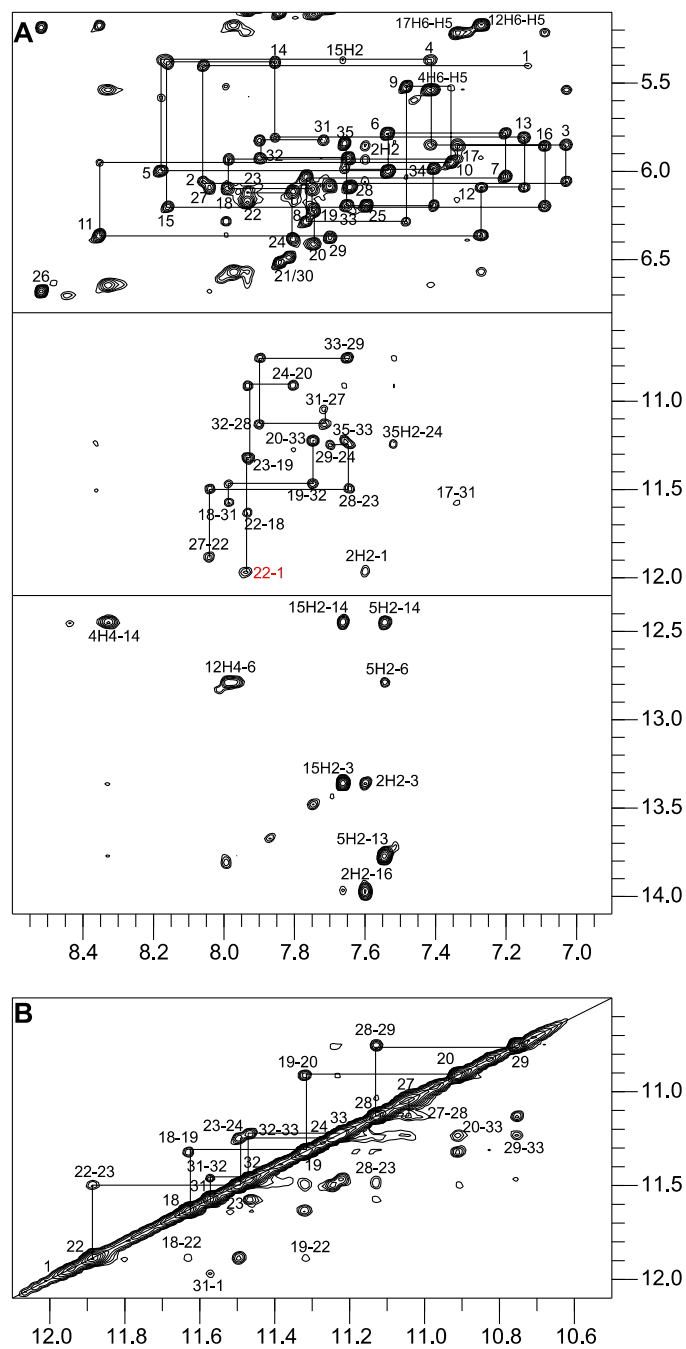


Figure S9. 2D NOESY spectral regions of *Myc-dup5*. (A) H6/H8(ω_2)-H1'(ω_1) (top), H6/H8(ω_2)-Hoogsteen imino(ω_1) (center), and AH2/CNH₂(ω_2)-WC imino(ω_1) connectivities (bottom). An uninterrupted sequential NOE walk through H6/H8-H1' contacts can be followed from G1 of the duplex along tract I of the quadruplex to G20 (top); G22 H8 shows a crosspeak to the G1 imino proton at the Q-D junction (center, labeled in red). (B) Imino-imino connectivities. Spectra were acquired with a 300 ms mixing time at 20 °C in 10 mM potassium phosphate buffer, pH 7.0.

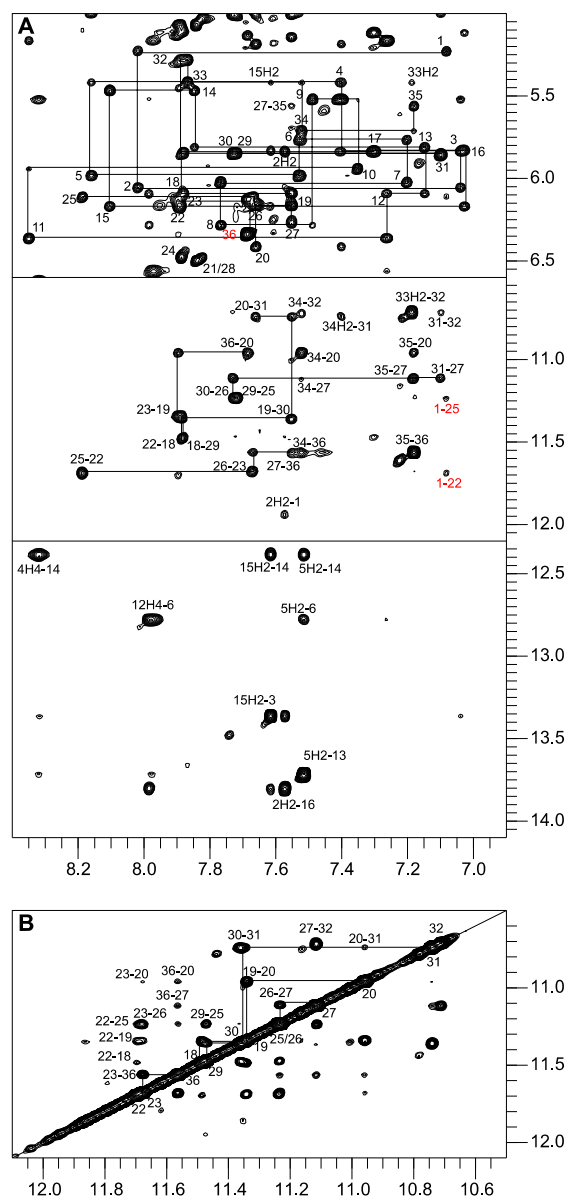


Figure S10. 2D NOESY spectral regions of *Myc31-dup5*. (A) H6/H8(ω_2)-H1'(ω_1) (top), H6/H8(ω_2)-Hoogsteen imino(ω_1) (center), and AH2/CNH₂(ω_2)-WC imino(ω_1) connectivities (bottom). An uninterrupted sequential NOE walk through H6/H8–H1' contacts can be followed from G1 of the duplex along tract I of the quadruplex to G20 (top); a very strong H8-H1' intranucleotide crosspeak indicates a *syn* glycosidic torsion angle for G36 at the 3'-terminus (top, labeled in red); the G1 H8 proton shows crosspeaks to the imino protons of G22 and G25 at the Q-D junction (center, labeled in red). (B) Imino-imino connectivities. Spectra were acquired with a 300 ms mixing time at 20 °C in 10 mM potassium phosphate buffer, pH 7.0.

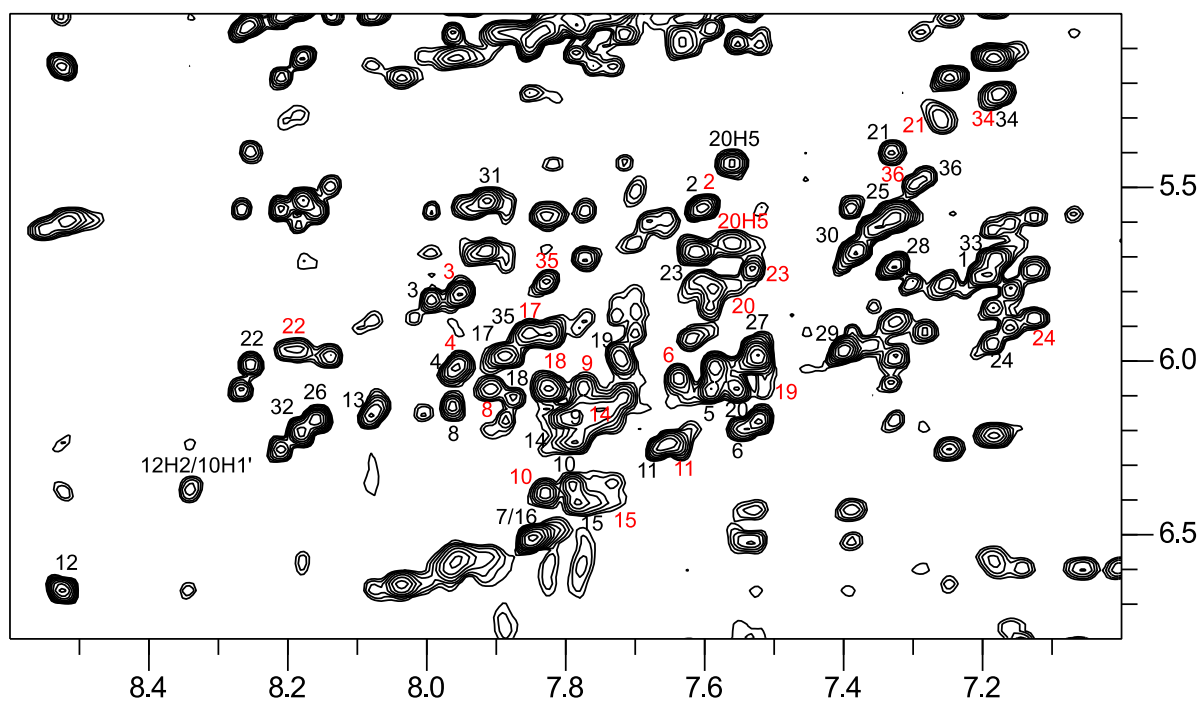


Figure S11. 2D NOESY spectrum of *Myc-dup3* (0.53 mM) with the addition of 0.5 equivalent of PIQ showing H6/8(ω_2)-H1'(ω_1) connectivities. Two sets of crosspeaks are observable with black and red assignments for crosspeaks of the free G4 and of the PIQ-G4 complex, respectively.

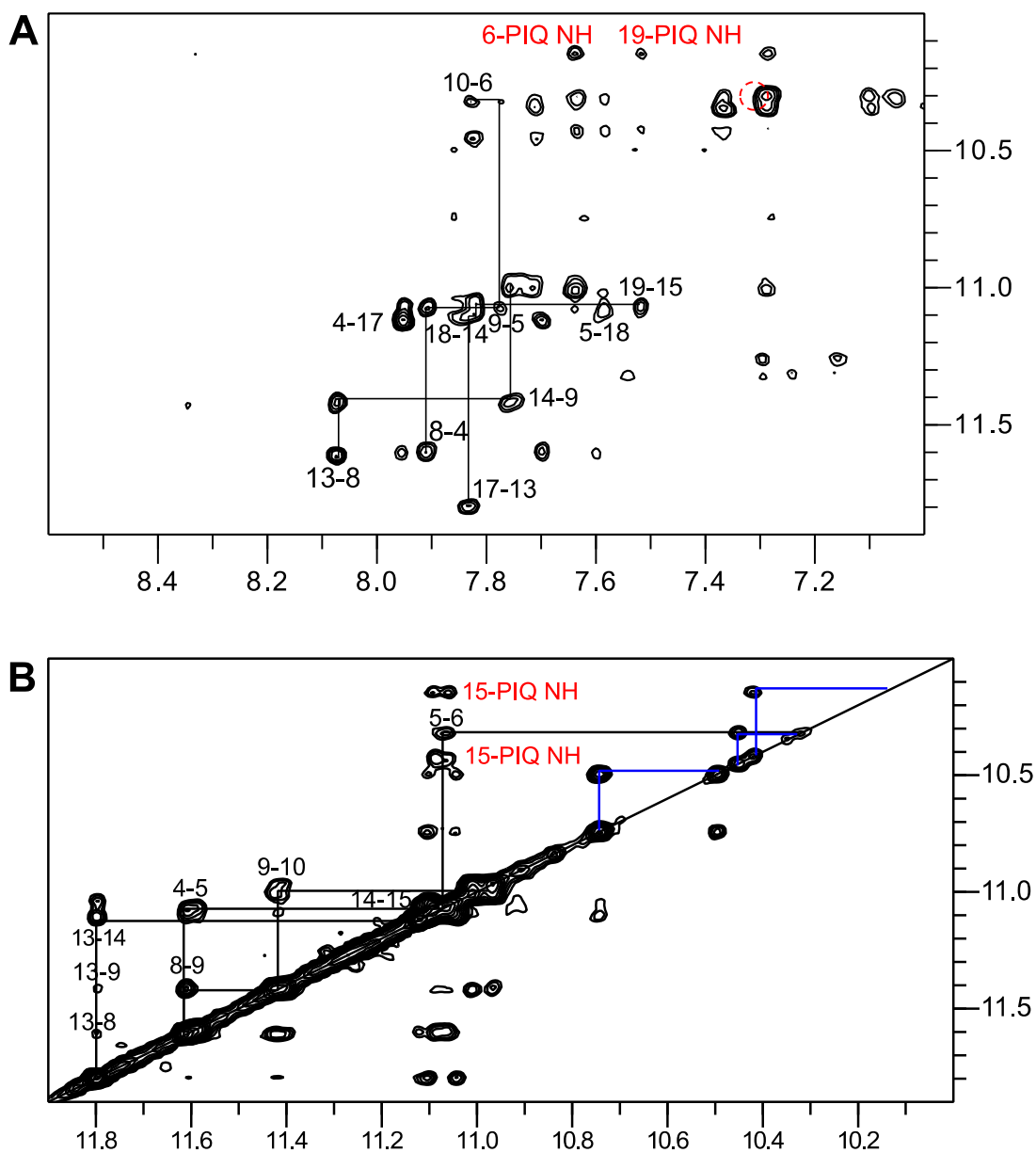


Figure S12. (A) H6/8(ω_2)-imino(ω_1) and (B) imino-imino 2D NOESY spectral region of *Myc-dup3* with addition of 1 equivalent of PIQ. The dashed red circle indicates the position of a missing contact between G36 and G6 at the junction as clearly observed in the free hybrid. Intermolecular crosspeaks with PIQ NH protons are labeled in red and exchange peaks of the ligand as demonstrated by additional ROESY spectra are indicated by blue lines. NOESY spectra were recorded with a 300 ms mixing time.

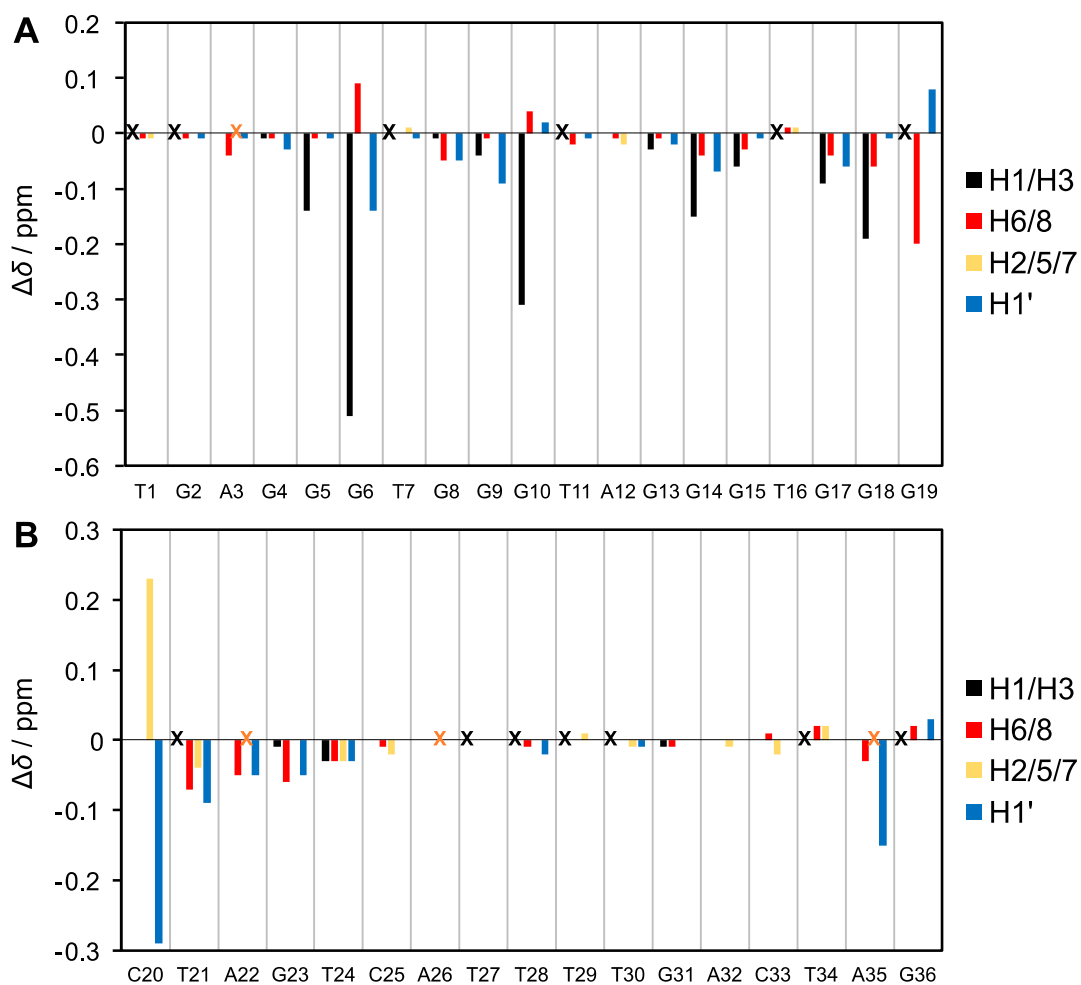


Figure S13. ^1H chemical shift differences for G4 residues in (A) the quadruplex subunit and (B) the duplex extension between complexed (with 1 eq. of PIQ) and free *Myc-dup3*; resonances marked by a cross could not be unambiguously assigned. For a more detailed compilation of chemical shift data see Tables S5 and S6.

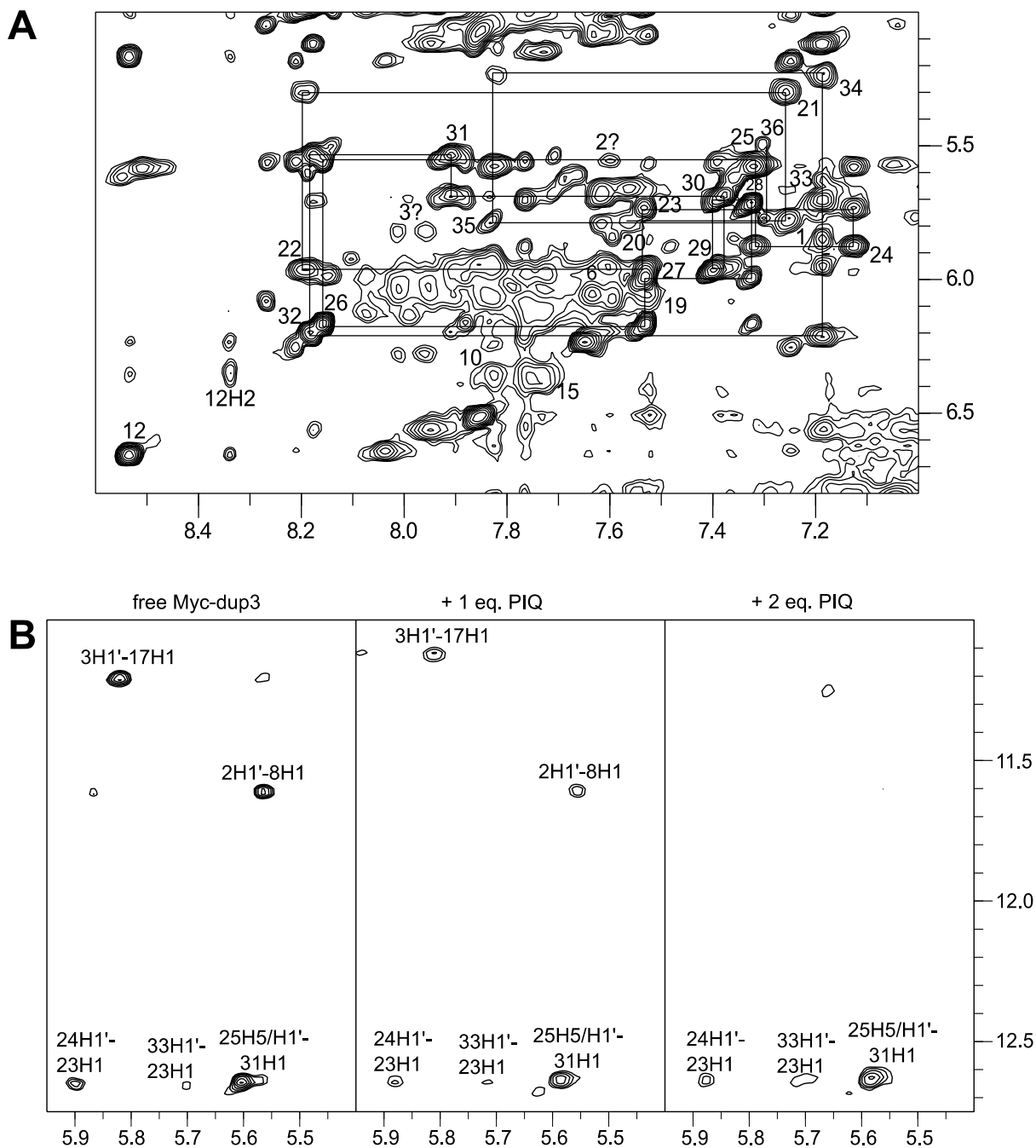


Figure S14. (A) H6/8(ω_2)-H1'(ω_1) 2D NOESY spectral region of *Myc-dup3* (0.53 mM) in the presence of 2 equivalents of PIQ; a continuous NOE walk is traced along the stem-loop duplex. (B) H1'(ω_2)-imino(ω_1) 2D NOESY spectral region of *Myc-dup3* in the absence (left) and in the presence of 1 equivalent (center) and 2 equivalents PIQ (right). All 2D NOESY spectra were acquired with a 300 ms mixing time at 20 °C in 10 mM potassium phosphate buffer, pH 7.0.

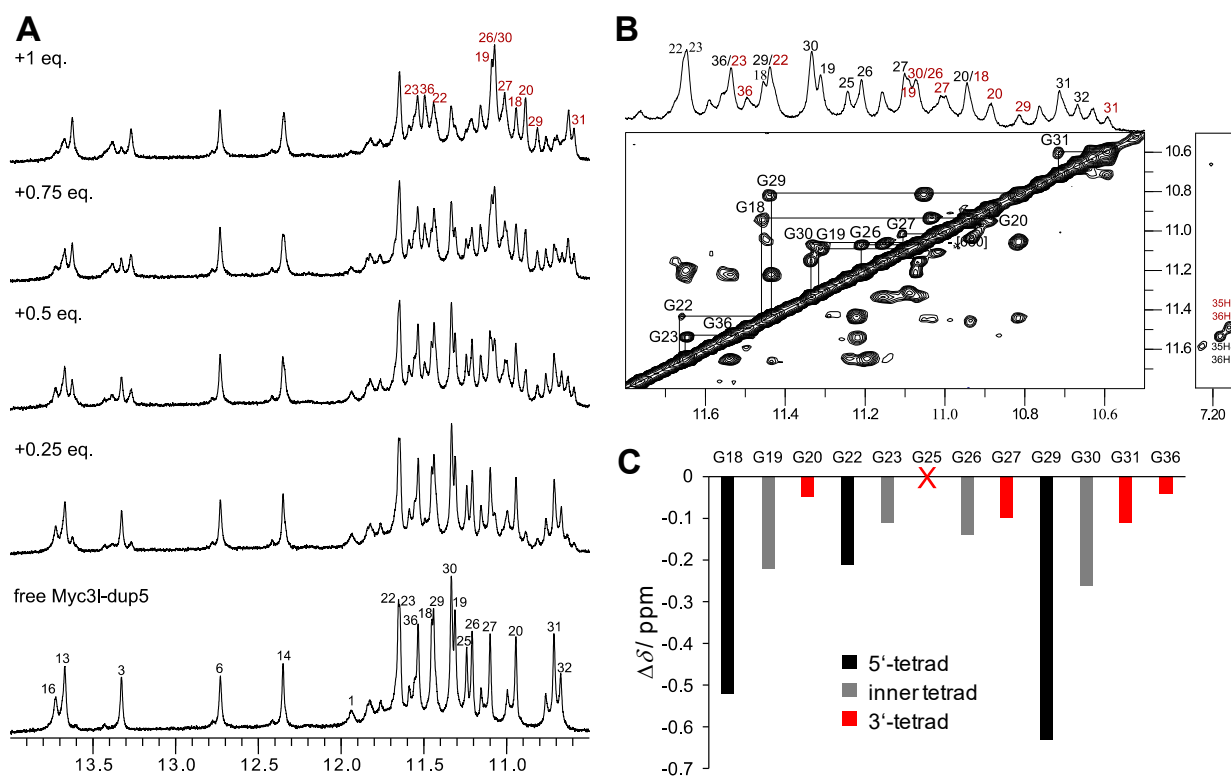


Figure S15. (A) Imino proton spectral region of *Myc31-dup5* (0.62 mM) titrated with PIQ at 30 °C in 10 mM potassium phosphate buffer, pH 7.0. (B) 1D and ROESY spectrum showing the G4 imino proton spectral region of *Myc31-dup5* in the presence of 0.5 equivalent of PIQ at 30 °C. Exchange crosspeaks connect guanine imino resonances of free and complexed species. A small shift of the G36 H1 resonance with a corresponding exchange peak close to the diagonal is confirmed by unambiguous G35 H8-G36 H1 NOE contacts (right). Imino proton resonances are assigned to G residues in free *Myc31-dup5* and in the ligand-hybrid complex by black and red numbers, respectively. (C) Imino chemical shift differences of G4 residues between complexed (with 0.5 eq. of PIQ) and free *Myc31-dup5*. The imino resonance of G25 could not be unambiguously assigned.

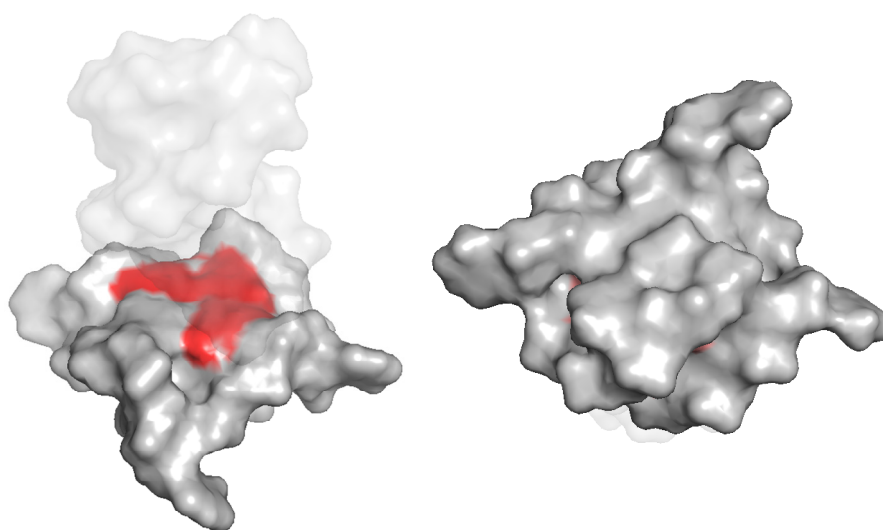


Figure S16. Guanosine imino proton chemical shift perturbations of the quadruplex after addition of 0.6 equivalent of PIQ mapped with red color of variable intensity on a surface model of the *Myc-dup5* hybrid. View onto the 5'-tetrad of the quadruplex-duplex junction (left) and onto the 3'-tetrad (right). The duplex extension is shown in a transparent representation.

Article 7

Indoloquinoline Ligands Favor Intercalation at Quadruplex-Duplex Interfaces

Yoanes Maria Vianney^[a] and Klaus Weisz^{*[a]}

Abstract: Quadruplex-duplex (Q-D) junctions are increasingly considered promising targets for medicinal and technological applications. Here, a Q-D hybrid with a hairpin-type snapback loop coaxially stacked onto the quadruplex 3'-outer tetrad was designed and employed as a target structure for the indoloquinoline ligand SYUIQ-5. NMR spectral analysis demonstrated high-affinity binding of the ligand at the quadruplex-duplex interface with association constants determined by isothermal titration calorimetry of about 10^7 M^{-1} and large exothermicities ΔH° of -14 kcal/mol in a 120 mM K^+ buffer at 40 °C. Determination of the ligand-bound hybrid structure revealed intercalation of SYUIQ-5 between 3'-outer tetrad and the neighboring CG base pair, maximizing π - π

stacking as well as electrostatic interactions with guanine carbonyl groups in close vicinity to the positively charged protonated quinoline nitrogen of the tetracyclic indoloquinoline. Exhibiting considerable flexibility, the SYUIQ-5 sidechain resides in the duplex minor groove. Based on comparative binding studies with the non-substituted *N5*-methylated indoloquinoline cryptolepine, the sidechain is suggested to confer additional affinity and to fix the alignment of the intercalated indoloquinoline aromatic core. However, selectivity for the Q-D junction mostly relies on the geometry and charge distribution of the indoloquinoline ring system. The presented results are expected to provide valuable guidelines for the design of ligands specifically targeting Q-D interfaces.

Introduction

G-rich sequences are able to fold into four-stranded quadruplex structures, exerting important biological roles in the regulation of various physiological processes but also constituting powerful tools for an increasing number of technological applications. It has been pointed out that quadruplex formation in the genome may entail the presence of Q-D junctions through the Watson-Crick pairing within an appropriate loop element or between a flanking sequence with the single-stranded complementary strand.^[1,2] In fact, several natural and designed quadruplex-forming sequences fold to feature Q-D interfaces by having loops, bulges, or flanking sequences able to self-associate into a duplex hairpin.^[3-6] Upon the engineering of quadruplex scaffolds, duplex extensions in quadruplexes were shown to promote quadruplex folding or to drive folding into defined quadruplex topologies.^[7-10] Also, RNA Q-D junctions were reported to be specifically recognized by the human fragile X mental retardation RGG peptide^[11,12] and anti-thrombotic quadruplexes featuring Q-D interfaces have demonstrated

their great potency as biomedical aptamers.^[13,14] Consequently, Q-D junctions have started to become attractive candidates as therapeutic targets but also as novel structural motifs with promising properties. Initial strategies for the design of ligands with a binding propensity for Q-D hybrid structures are based on linking quadruplex-binding ligands composed of extended aromatic ring systems for efficient tetrad stacking with typical duplex minor groove binders for the dual binding of both quadruplex and duplex domains.^[2,15] However, Q-D junctions themselves have not been exploited for a systematic drug targeting to date, suffering from a paucity of detailed structural information.

Natural and artificial compounds that are based on the tetracyclic indoloquinoline scaffold possess a wide range of different biological activities (Figure 1A).^[16] Thus, the natural alkaloid cryptolepine has long been used as an antimalarial agent.^[17] In addition, various indoloquinoline derivatives are known to be potent binders to DNA structures and in particular to G-quadruplexes.^[18,19] These include the anticancer drug SYUIQ-5, shown to exhibit telomerase inhibition activity.^[20,21] The binding of a closely related indoloquinoline to the parallel *c-Myc* quadruplex has been structurally and thermodynamically characterized in detail.^[22,23] As for other typical quadruplex ligands, the planar indoloquinoline ring system binds through end-stacking onto the exposed 5'- and 3'-outer tetrad, additionally fixed through the formation of a binding pocket involving short overhang sequences. Binding is mostly driven by favorable stacking interactions and hydrophobic effects. In trying to optimize ligands for improved biological activities and better quadruplex affinities and selectivities, substitution patterns and ligand sidechains have been modified to provide a large number of indoloquinoline derivatives during the past two

[a] Y. M. Vianney, Prof. Dr. K. Weisz
Institute of Biochemistry
Universität Greifswald
Felix-Hausdorff-Str. 4
17489 Greifswald (Germany)
E-mail: weisz@uni-greifswald.de

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/chem.202103718>

© 2021 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

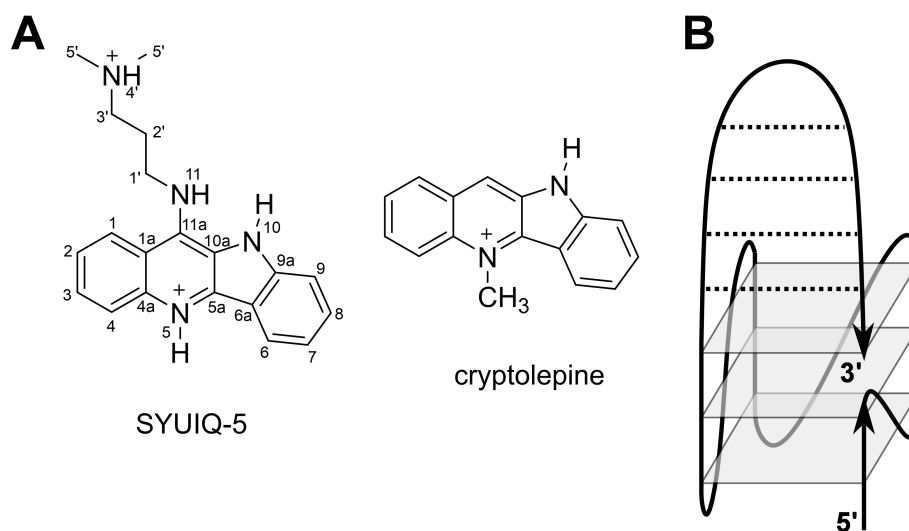


Figure 1. (A) Chemical structure of indoloquinoline derivatives cryptolepine and SYUIQ-5 with atom numbering. (B) Designed quadruplex topology with a Q-D junction.

decades.^[24–26] In fact, indoloquinoline ligands with appropriate sidechains have been shown to not only increase affinities due to additional complex-stabilizing interactions but to also favor quadruplex over duplex binding and to discriminate among different quadruplex folds.^[25,26] However, whereas the impact of sidechains on the binding thermodynamics can easily be evaluated, local sidechain interactions have been difficult to pinpoint in most cases owing to their considerable flexibility.

Building upon the adaptable and promising DNA binding properties of the indoloquinoline scaffold, we recently reported on the binding of an 11-phenyl substituted indoloquinoline derivative to a hybrid structure with a Q-D junction formed by a dangling 3'-hairpin extending from a parallel quadruplex.^[27] A favored enthalpy-driven binding at the Q-D junction could be demonstrated, however, NMR experimental limitations precluded the determination of a high-resolution structure with a well-defined ligand binding site. Therefore, a modified Q-D hybrid was designed to be used as a target for the indoloquinoline SYUIQ-5 in the present study (Figure 1B). The hybrid was constructed from a typical parallel G-quadruplex by extending its 3'-terminus by a self-complementary hairpin-forming sequence. Inspired by various Q-D hybrids originally engineered by Phan,^[28] the duplex stem-loop was additionally fixed to the G-core through a 3'-terminal G, filling a vacant site of the quadruplex outer tetrad. Such a model architecture is expected to decrease flexibilities and to yield a better defined Q-D junction for structure determinations, yet may nevertheless mimic parallel quadruplexes with a coaxially stacked duplex as a potential target in promoter regions of oncogenes. As demonstrated by NMR experiments, SYUIQ-5 with its aminoalkyl sidechain binds the Q-D junction with high affinity. The three-dimensional solution structure of the major 1:1 complex reveals ligand intercalation between the outer G-tetrad and the adjacent duplex base pair. Additional binding studies with

unsubstituted cryptolepine give further insight into the impact of the sidechain on the indoloquinoline binding.

Results

Structure and stability of the *QD3-sbl* hybrid

The 36mer oligonucleotide *QD3-sbl* is based on the parallel-folded *c-Myc* quadruplex^[29] with a 3'-flanking Watson-Crick self-complementary sequence and a 5'-TTA overhang found to exhibit cleaner spectra (Table S1). Also, featuring only a truncated first GG run, a 3'-terminal guanine base appended to the hairpin domain was expected to dock into the last vacant position of the first G column of the parallel fold to result in a duplex stem-loop fixed at both of its ends to the 3'-outer tetrad of the quadruplex (Figure 1B). In the following, NMR and thermal melting experiments were performed in a buffer solution with 10 mM potassium phosphate, pH 7. With no noticeable structural change when compared to an environment with higher K^+ concentrations (not shown), these low-salt conditions allowed for the observation of melting temperatures within a convenient temperature range and also yielded an improved sensitivity in NMR experiments.

The imino proton NMR spectral region of a *QD3-sbl* buffer solution suggests a well-defined structure with Hoogsteen G imino resonances between 10.6 and 12.0 ppm indicative of a three-layered quadruplex and additional more downfield shifted Watson-Crick imino protons through duplex formation between 12.6 and 13.8 ppm (Figure S1A). Resonance assignments of the *QD3-sbl* hybrid were facilitated by strong correspondences of NOE contacts in expected quadruplex and duplex domains with NOE patterns previously found for a closely related quadruplex-duplex hybrid but with non-interrupted G-tracts and a loose duplex 3'-terminus.^[27] Sequential

H8/H6-sugar NOE walks from 5'-terminal T1 to the 3'-penultimate G35 also include T26–T28 of the putative hairpin loop and are only interrupted at propeller loop residues T6, T10, and T15 (Figure 2A, S1B). Noticeably, 3'-terminal G36 lacks sequential contacts to G35 but has cross-peaks to G5 of the first GG-tract as expected when filling the vacant position in the 3'-outer G-tetrad. Also, its *syn* glycosidic torsion angle is shown by a strong intra-nucleotide H8–H1' cross-peak and a rather downfield-shifted ^{13}C in HSQC spectra (Figure S1B, S2). Homo-polarity of all tetrads with hydrogen bonds running into the same direction is demonstrated by typical H8–H1 connectivities within the quadruplex core composed of an *anti-anti-syn* column for G4–G5–G36 and three all-*anti* columns for

G7–G8–G9, G12–G13–G14, and G16–G17–G18 (Figure 2A, S1C). Characteristic imino-imino contacts within the G-core further corroborate the alignment of G residues in the parallel quadruplex (Figure S1A).

Imino protons of the duplex base pairs were assigned according to standard strategies making use of their NOE contacts to cytosine amino and adenine H2 protons. All seven Watson-Crick base-paired imino protons of thymine and guanine bases could be identified. Notably, in addition to the A25–T29 base pair following the flexible T_3 hairpin loop, imino protons of T20 and in particular G35 in the two base pairs bordering the quadruplex-duplex junction are broadened and of low intensity. This suggests enhanced dynamics at the junction with its anchored 3'-terminal *syn*-G36. However, a strong conspicuous contact from G35 H8 to G36 H1 demonstrates continuous stacking of the duplex stem-loop onto the 3'-tetrad (Figure S1C). On the opposite face of the G-core, various cross-peaks connect the short 5'-overhang with the neighboring 5'-outer tetrad. Based on cross-peak patterns in DQF-COSY spectra with large H1'–H2' scalar couplings, 26 residues were unambiguously found to adopt sugar puckers in the pseudorotational south domain (Figure S3).

Structure calculations employed NMR-derived distance and dihedral angle restraints (statistics and a list of chemical shifts are given in Table S2 and S3). Final structures feature a quadruplex-duplex hybrid composed of a parallel three-layered quadruplex with a broken first G-column (Figure 2). A double-helical stem-loop with its seven Watson-Crick base pairs extends from the fourth G-tract with coaxial stacking of the duplex onto the quadruplex domain and is additionally fixed by the appended 3'-terminal *syn*-guanosine that fills an empty G-core position. In fact, the hairpin-type 3'-overhang can be regarded a double-helical lateral snapback loop. Whereas residues of the G-core and the Watson-Crick base pairs are well defined, residues in the 5'-overhang, in the quadruplex propeller loops, and in the hairpin T_3 -loop are more flexible (Figure 2B). With the duplex connecting adjacent edges of the outer tetrad, its minor groove follows the quadruplex groove between the first and fourth G-column while the duplex major groove at the junction faces the center of the G-tetrad. Efficient stacking interactions are observed between G35 and G36 but C19 stacking onto G18 is only poor (Figure 2C). On the other quadruplex face, A3 of the 5'-overhang is found to cap G4 and G16 of the 5'-tetrad in line with corresponding NOESY cross-peaks (Figure S1B,C).

DSC thermograms revealed two distinct melting transitions for the quadruplex and duplex domains at 40.2 °C and at 47.2 °C in a 10 mM K^+ buffer solution (Figure S4). Independent melting of the secondary structures was additionally confirmed by UV melting experiments. By analyzing temperature dependent absorbance changes at 260 nm for the duplex and at 295 nm for the quadruplex, melting of the latter could be assigned to the lower melting transition (Table S4). Apparently, despite their coaxial stacking there is no cooperative melting of quadruplex and duplex domains in line with corresponding observations on a quadruplex carrying a 3'-flanking duplex domain with a dangling terminus.^[27]

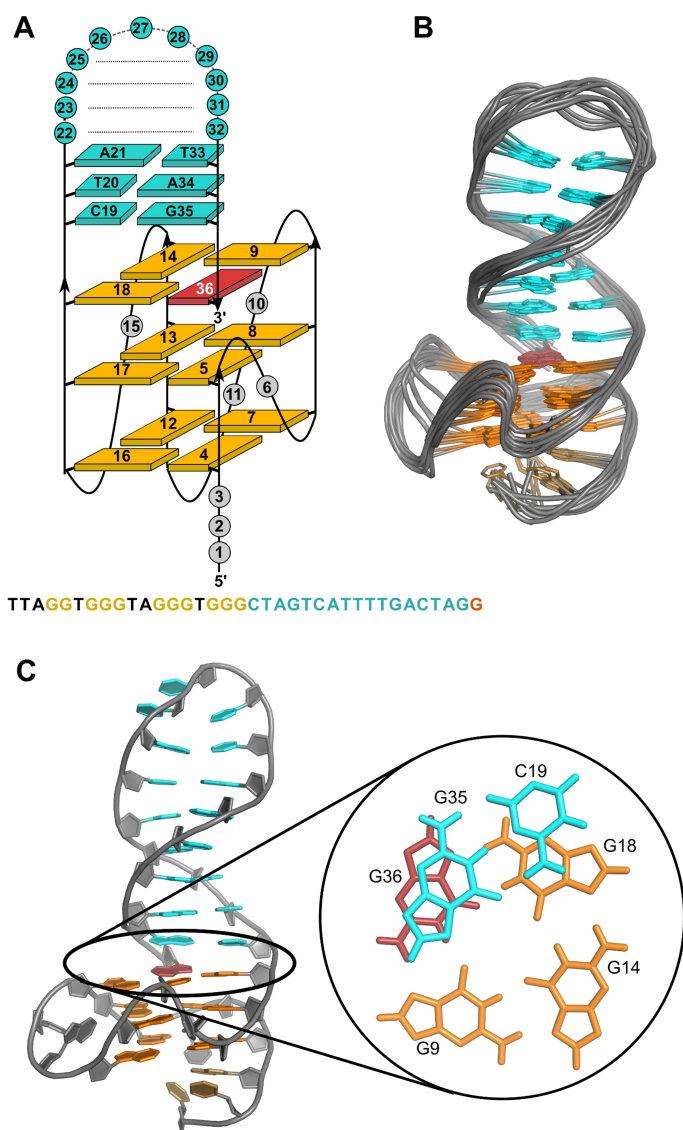


Figure 2. (A) Schematic representation of the *QD3-sbl* hybrid structure with oligonucleotide sequence. (B) Superposition of ten lowest-energy structures of *QD3-sbl*; residues in the quadruplex propeller and the hairpin T_3 -loops are omitted for clarity. (C) Representative structure of *QD3-sbl* with a close-up view of the C19-G35 base pair stacked onto the outer tetrad at the Q-D junction; *anti*-G residues of the tetrad, *syn*-G36, and residues of the stem-loop are colored orange, red, and cyan, respectively.

The Q-D junction constitutes the preferred ligand binding site

Initial CD titrations indicated that addition of SYUIQ-5 to *QD3-sbl* has no significant impact on the Q-D hybrid structure (Figure S5). On the other hand, binding of the ligand is demonstrated by an induced CD effect (ICD) at the ligand absorption centered at 350 nm. Of note, a small-amplitude negative ICD compatible with end-stacking changes to a positive ICD of higher amplitude with ligand in excess. Apparently, at stoichiometries >1 additional ligand interacts with the Q-D hybrid albeit with weaker affinity, overwriting the initial ICD signature.

To yield sharper resonances, subsequent NMR titrations of the ligand to the hybrid were performed at 30 °C. Looking at the imino proton spectral region, additional Hoogsteen G imino signals gradually emerged upon ligand addition with resonances of the free hybrid structure completely lost at a 1:1 molar ratio (Figure 3). Coexisting resonances of free and complexed species observed with 0.5 equivalent of ligand indicate their slow exchange. On the other hand, downfield-shifted Watson-Crick imino resonances seem to only show modest heterogenous and/or homogenous line broadening effects, suggesting smaller perturbations in the duplex stem-loop upon initial ligand binding.

Supported by a close analogy to the already assigned free Q-D hybrid, standard strategies involving NOESY, DQF-COSY, and ^1H - ^{13}C HSQC experiments were again used to identify non-labile protons in the 1:1 complex. Thus, continuous sequential NOE connectivities between H8/H6 and sugar protons, interrupted by the propeller loops, allowed for the assignment of most non-exchangeable proton resonances (Figure 4A,B, S6A).

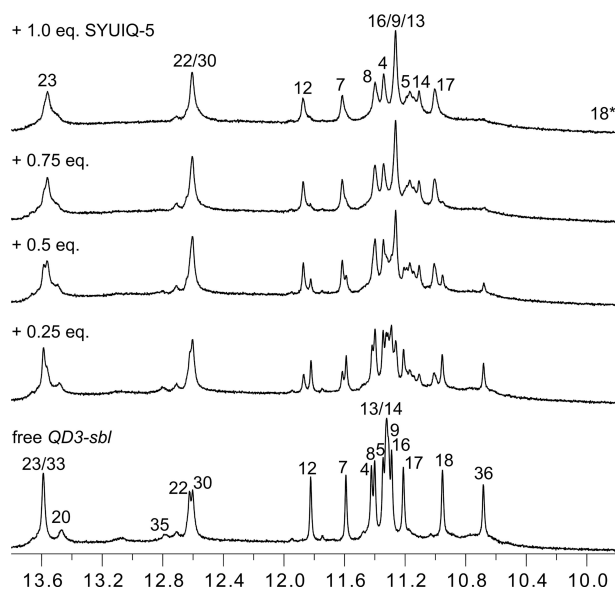


Figure 3. Imino proton spectral region of *QD3-sbl* (1 mM) titrated with SYUIQ-5 at 30 °C. Assigned peaks of the free hybrid and the 1:1 complex are labeled with residue numbers; note that the marked G18 imino signal of the complex is unobservable in the 1D spectrum but unambiguously assigned through exchange cross-peaks in NOESY and ROESY spectra.

Again, 3'-terminal G36 features characteristic NOE contacts to G5 of the first G-tract. However, lost connectivities between G18 and C19 at the Q-D interface adds another interruption to the sequential NOE walk in the ligand-bound hybrid.

Noticeable homogenous and/or heterogenous broadening of resonances at the Q-D junction hampered the unambiguous resonance assignment for affected residues, yet support from sequential H8/6–H8/6 contacts, ^1H - ^{13}C HSQC, and DQF-COSY spectra enabled spectral identification for most protons (Figure S6A, S7, S8). It should be mentioned that some broadening of H8/6–H1' cross-peaks from duplex nucleotides not only depends on their vicinity to the junction but shows an asymmetric behavior along the duplex stem-loop. In contrast to the 5'-terminal hairpin strand extending from the quadruplex core, line broadening effects further continue along the complementary duplex strand beyond interfacial G35 to A34 and up to T33. Finally, if accessible through resolved H1'(ω₂)–H2'(ω₁) DQF-COSY cross-peak patterns and NOESY contacts at short mixing times, the sugar pucker of residues could unambiguously be assigned to a south conformation (Figure S8).

Except for the G36 imino, assignment of all other imino resonances of the G-core was enabled through H8–H1 NOE contacts and additionally supported by characteristic imino-imino connectivities as well as by ROESY exchange cross-peaks observed between the free and complexed Q-D hybrid after the addition of 0.5 equivalent of ligand (Figure 4C, S6B, S9). Imino connectivities also demonstrated formation of a three-layered parallel quadruplex core with a counter-clockwise direction of Hoogsteen hydrogen bonds within G-quartets in line with no major structural rearrangements upon ligand binding. A conspicuous upfield shift of more than 1 ppm for the G18 imino proton after complex formation, confirmed by a corresponding exchange cross-peak in a ROESY experiment (Figure S9), again hints at a ligand binding site at the Q-D junction. Due to a significant ligand-induced broadening of duplex imino resonances especially for residues near the junction and for the base pair following the hairpin loop, only imino protons of centrally located base pairs could unambiguously be assigned through their strong contacts to cytosine H4 or adenine H2 protons (Figure S6C).

Protons of the ligand were assigned based on a combination of DQF-COSY, TOCSY, and NOESY experiments (Figure S10). Amino and aliphatic protons of the SYUIQ-5 sidechain could be traced through their vicinal and long-range couplings as observed in DQF-COSY and TOCSY spectra (Figure S10A). Similarly, correlations in DQF-COSY spectra also enabled assignments of the scalar coupled aromatic protons of quinoline and indole subunits (Figure S10B). A contact from a ligand H2' aliphatic proton to the quinoline moiety of the indoloquinoline discriminates the quinoline and indole ring system. Fast exchange prevented observation of NH protons for the indole N10 and quinoline N5. Given a pK_a of 8.4,^[20] the latter is expected to be protonated even in more hydrophobic environments. Due to the absence of contacts between these exchangeable NH protons to other protons of the indole ring system, unambiguous assignments to H6/H7 and H9/H8 proton

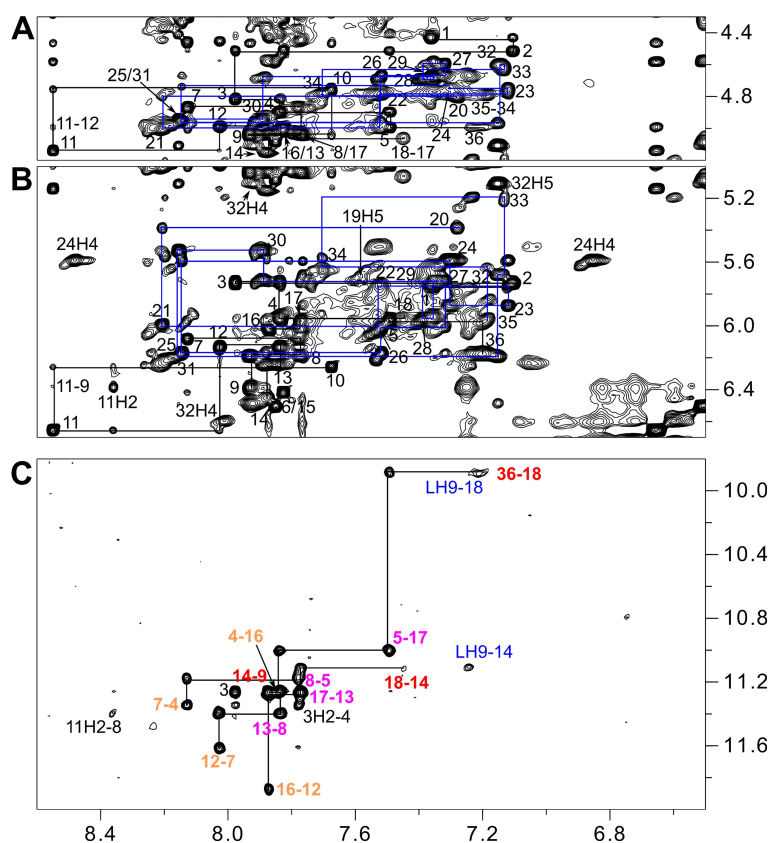


Figure 4. Regions of a 2D NOESY spectrum (300 ms mixing time, 30 °C) of *QD3-sbl* (1 mM) in the presence of 1 equiv. SYUIQ-5. (A) H6/8(ω_2)-H3'(ω_1) and (B) H6/8(ω_2)-H1'(ω_1) spectral region; continuous networks of base-sugar resonances are followed by vertical and horizontal lines with NOE connectivities in the duplex domain traced by blue lines; intra-nucleotide cross-peaks are labeled by residue number. (C) H8/6/2(ω_2)-H1(ω_1) spectral region; intra-tetrad H8(ω_2)-H1(ω_1) cross-peaks are labeled with colors depending on G-tetrad layer; inter-tetrad connectivities along the G-columns are traced by horizontal and vertical lines and intermolecular contacts between quadruplex imino and ligand protons are labeled in blue.

pairs on the two sides of the fused benzene ring were hampered. However, intermolecular NOE connectivities of indole and quinoline with *QD3-sbl* resonances in the complex enabled a discrimination of these non-labile indole protons. A compilation of chemical shifts for *QD3-sbl* and SYUIQ-5 in the complex are given in Table S5 and S6.

A total of 11 intermolecular NOE contacts to the Q-D hybrid position the ligand within the complex (Table S7). These include cross-peaks of ligand quinoline or indole protons to non-exchangeable protons of G18, C19, and G35 at the Q-D interface but also to imino protons of residues G9 and G14 at the exposed edge of the 3'-tetrad (Figure 4C, S11). No contact to G36 could be observed likely due to dynamic processes. Additional non-observable intermolecular contacts between indole H9 and G14 sugar protons with their rather sharp signals were added as repulsion restraints in subsequent structure calculations. While the absence of any intermolecular NOE contact of the dimethylamino group attests to a high flexibility of the ligand sidechain, a single contact was detected from ligand aliphatic H2'a/b protons (restrained as C2') to the G5 H2'' sugar proton in the quadruplex groove just below the Q-D junction. Two unexpected NOE contacts incompatible with the other short intermolecular distances connect aromatic protons

of the ligand with a sugar proton of residue T33 located three bases off the junction and also to A3 H8 located at the opposite face of the G-core (not shown). The latter observations suggest that in addition to a major binding site there may be some minor binding of the ligand within the duplex domain and also onto the 5'-outer tetrad.

Solution structure of the 1:1 complex

A superposition of ten calculated lowest-energy structures shows a good convergence of the 1:1 complex (Figure 5A, Table S3). Despite the limited number of unambiguously assigned intermolecular contacts used as restraints in the structure calculations, the quadruplex-duplex junction with the bound indoloquinoline aromatic ring system is well defined. A single distance violation $> 0.2 \text{ \AA}$ in 1 out of 10 structures applies to contacts between the 5'-overhang and the 5'-tetrad and is thus far removed from the SYUIQ-5 binding site. The ligand intercalates between G18 and G36 of the 3'-outer tetrad and the C19·G35 Watson-Crick base pair at the Q-D junction. As a consequence of the ligand insertion between G-tetrad and base pair, the helical rise at the Q-D junction increases to 7 Å but

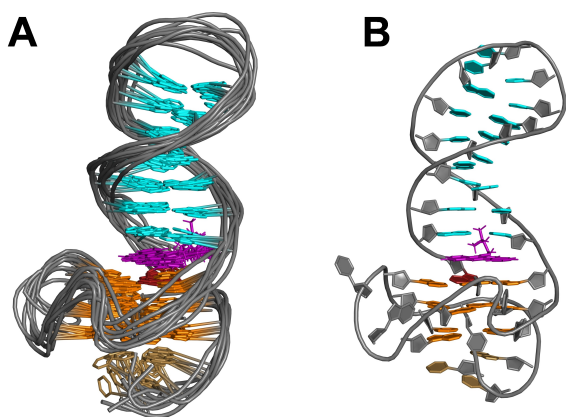


Figure 5. Side view of (A) ten superimposed lowest-energy structures and (B) a representative structure of a 1:1 complex between *QD3-sbl* and SYUIQ-5. Bases in the quadruplex propeller and hairpin T_3 loop are omitted for clarity in (A). *Syn*-guanosines, *anti*-guanosines, duplex bases, and ligand are colored red, orange, cyan, and magenta, respectively.

also results in some conformational adjustments. Compared to the arrangement in free *QD3-sbl*, the stacked duplex stem-loop is shifted towards the center of the G-tetrad with interfacial C19 positioned above G18 (Figure S12). The ligand sidechain faces the minor groove of the duplex as already suggested by the NOESY data. However, poor convergence indicative of high flexibility is observed for the aliphatic substituent and in particular for the protonated dimethylamino group.

The lateral shift of the duplex towards the center of the quadruplex is accompanied by more efficient π - π stacking interactions of the intercalated ligand. Thus, the indoloquinoline tetracyclic ring system is found to insert with the quinoline and indole moieties mostly sandwiched between G36 and G35 and between G18 and C19, respectively. With the protonated and positively charged N5 of the indoloquinoline positioned above the central channel of the G-core at the major groove side of

the duplex stem-loop, additional Coulombic interactions are expected with the four guanine carbonyl oxygen atoms of the 3'-tetrad but also with G35 of the base pair on top, being in close vicinity with distances ~ 4 Å (Figure 6A,B). Potential hydrogen bond interactions with short distances are indicated between indole NH10 and C19 O4' in 3 out of 10 structures and between NH11 of the sidechain and G36 O4' in 5 out of 10 structures (Figure 6C). It should be noted that the latter hydrogen bond is only enabled by the antiparallel orientation of docked G36 relative to the other G-core residues. Given a high flexibility of the ligand sidechain, there is no indication for a hydrogen bond interaction of the terminal dimethylamino group. Yet, electrostatic interactions with the sugar-phosphate backbone can be assumed.

Binding of SYUIQ-5 to an antiparallel quadruplex with a central hairpin lateral loop

To also examine Q-D junctions as a target in a different structural context, an antiparallel quadruplex termed *QD2-I*, derived from the thrombin binding aptamer with the second loop modified by a duplex stem-loop,^[28] was employed in additional binding studies with SYUIQ-5 (Figure 7A). All hydrogen-bonded imino protons of the duplex, of the two-layered G-core, and also of additional T-T base pairs formed between the first and the third lateral loop were observed in a low-salt buffer at 20 °C (Figure S13). With only small shifts for some resonances, further analysis of NOESY spectra confirmed a fold as reported previously under our experimental conditions (PDB 2M8Z).

To identify the major SYUIQ-5 binding site, the ligand was titrated to the *QD2-I* hybrid while monitoring the imino proton spectral region (Figure 7B). Signal broadening but also the appearance of new signals demonstrated slow exchange between a free and ligand-bound quadruplex. Exchange cross-peaks were observed in ROESY spectra on a mixture of the Q-D

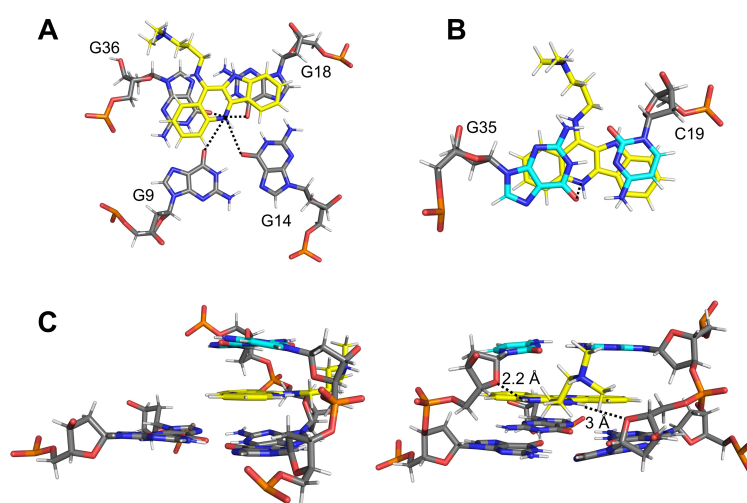


Figure 6. Stacking of the ligand (A) onto the 3'-tetrad and (B) below the G35-C19 Watson-Crick base pair; distances between the protonated ligand N5 and adjacent guanine carbonyl atoms are indicated by dotted lines. (C) Side view (left) and view into the minor groove (right) showing the Q-D junction with intercalated ligand; short NH10-C19 O4' and NH11-G36 O4' distances are indicated by dotted lines. The carbon skeleton of the ligand is colored yellow.

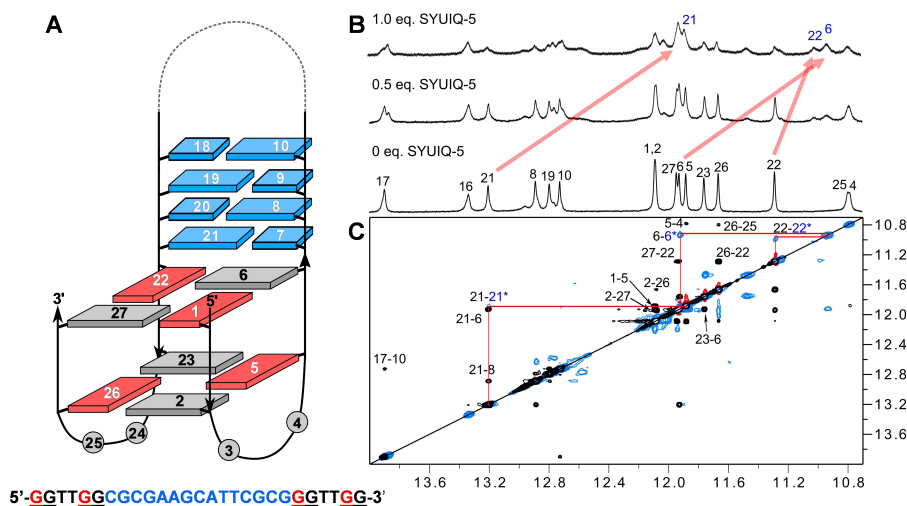


Figure 7. (A) Topology of *QD2-I* with sequence. (B) Imino proton spectral region upon titrating SYUIQ-5 to *QD2-I* (0.5 mM) at 20 °C; most pronounced shifts are indicated by red arrows. (C) Superposition of NOESY spectrum for the free hybrid (black) and ROESY spectrum after the addition of 0.5 equiv. of ligand (blue: positive intensity; red: negative intensity) showing imino(ω_2)-imino(ω_1) correlations; red lines connect most shifted proton resonances at the Q-D junction as verified through exchange cross-peaks between the free and complexed *QD2-I* hybrid. Residue numbers of the free and ligand-bound hybrid are marked in black and blue, respectively.

hybrid with 0.5 equivalent of ligand and allowed the identification of several imino resonances of the complex (Figure 7C). For Hoogsteen imino signals, most prominent exchange cross-peaks positioned far off the diagonal and thus correlating resonances with significant chemical shift differences could be traced to G22 and G6 located at the Q-D interface. The largest chemical shift perturbation was found for the G21 Watson-Crick imino proton within the interfacial base pair. Apparently, these resonances are subjected to significant upfield shifts upon ligand binding, again in line with strong π - π stacking interactions through SYUIQ-5 intercalation at the Q-D junction.

Taken together, Q-D junctions seem to constitute major high-affinity binding sites for the SYUIQ-5 ligand irrespective of the quadruplex topology or external or internal duplex extensions. Intercalation between an outer G-tetrad and a base pair seems to support a selective high-affinity binding with ligands featuring a matched shape for optimal stacking interactions.

Targeting the *QD3-sbl* hybrid with cryptolepine

In the complex structure with bound SYUIQ-5, the ligand sidechain mostly resides in the minor groove of the duplex domain. Because of its high flexibility, no major specific interactions of the aminoalkyl group to the Q-D hybrid could be identified. Nevertheless, van der Waals and electrostatic effects are expected to add to the favorable binding free energy of the indoloquinoline ligand. To examine the impact of the sidechain on the binding selectivity in more detail, the *QD3-sbl* hybrid was also targeted with the natural indoloquinoline alkaloid cryptolepine (Figure 1A). Lacking any additional sidechain, this *N5*-methylated indoloquinoline bears a permanent positive charge but is considered a rather poor quadruplex-binding

ligand both because of its modest discrimination against other nucleic acid secondary structures including duplexes and because of only moderate affinities for G-quadruplexes.^[16]

NMR titrations showed signal broadening and the appearance of new resonances upon cryptolepine addition with slowly exchanging free and bound species at ligand-to-DNA molar ratios < 1 (Figure 8A). Assignments of non-labile protons in the complex are mostly based on NOESY experiments (Figure S14A). In general, cross-peak patterns of the NOESY spectrum resemble *QD3-sbl* when complexed with SYUIQ-5. However, although continuous base-sugar NOE connectivities can be traced along the duplex stem-loop, broadening of cross-peaks increases towards the Q-D junction and only allow to unambiguously follow sequential NOE walks from T20 to A34. For the quadruplex domain, H8–H1' connectivities link all residues along the four G-columns, again featuring more extensive signal broadening for non-labile protons at the Q-D junction when compared to the 5'-tetrad.

Some quadruplex imino proton resonances in the complex were heavily broadened. Nevertheless, almost complete assignments except for G9 H1 were enabled by the observation of ROESY exchange cross-peaks on samples with ligand-to-DNA molar ratios of 0.5 (Figure S15) and further supported by H8–H1 NOESY cross-peaks for the 1:1 complex (Figure S14B). Notably, there are two pairs of prominent exchange cross-peaks of similar intensity in the ROESY spectrum that correlate G18 H1 as well as G36 H1 at the interfacial 3'-tetrad of the free hybrid with two ligand-bound species (Figure S15B). Being most upfield-shifted in both complexes, these observations suggest cryptolepine binding at identical sites but with different ligand orientation, for example, as a result of a 180° flip of the indoloquinoline within the binding pocket. Exchange processes between differently aligned ligand is expected to further broaden resonances at the cryptolepine binding site. Overall, G

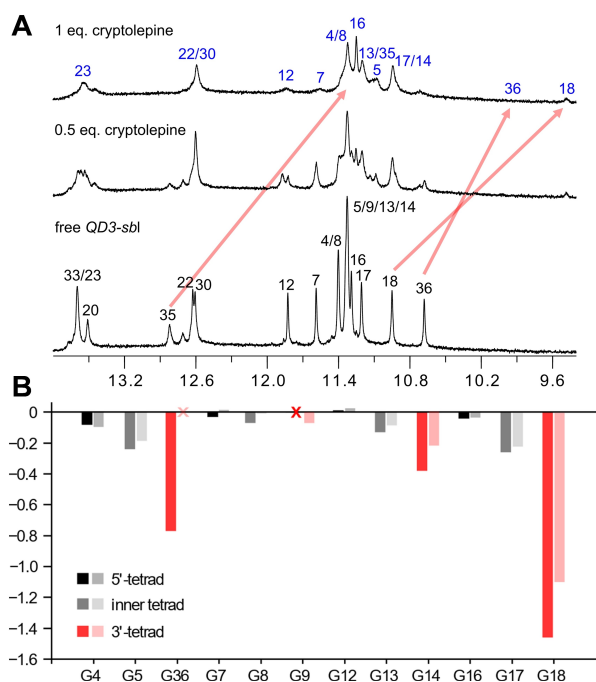


Figure 8. (A) Imino proton spectral region upon titrating cryptolepine to the *QD3-sbl* hybrid structure (0.5 mM) at 30 °C; most pronounced chemical shift changes for interfacial G-tetrad and base pair imino protons, corroborated by exchange cross-peaks in ROESY spectra, are indicated by red arrows. (B) Quadruplex imino proton chemical shift differences between major ligand-bound and free *QD3-sbl*; dark- and light-colored bars represent footprints with cryptolepine and SYUIQ-5, respectively. Red crosses mark non-assigned imino resonances.

imino protons in the 3'-tetrad at the Q-D junction experienced significant chemical shift perturbations when compared to those in the inner and 5'-tetrad. Also, profiles of G imino chemical shift changes on cryptolepine addition closely follow chemical shift footprints on SYUIQ-5 binding, indicating favored cryptolepine binding again through intercalation at the Q-D interface as demonstrated for SYUIQ-5 (Figure 8B). Consequently, it is the indoloquinoline ring system with its particular geometry and electron distribution that seems to favor binding at Q-D junctions.

Thermodynamic profiles for indoloquinoline binding at Q-D interfaces

Isothermal titration calorimetry was employed to evaluate association constants and thermodynamic profiles of SYUIQ-5 and cryptolepine binding to the *QD3-sbl* hybrid structure (Figure S16, Table 1). Being closer to physiological conditions, studies on the binding thermodynamics were performed at 40 °C in a 120 mM potassium phosphate buffer. As has already been suggested by the CD titrations of *QD3-sbl* with SYUIQ-5 (see above), thermograms of both indoloquinolines exhibit a high-affinity binding site followed by additional binding events of lower affinity with only a gradual return to baseline for ligand in excess. Notably, high- and low-affinity binding processes are better resolved for the SYUIQ-5 ligand, indicating its superior binding selectivity. Only focusing on the high-affinity binding, the association constant as determined by a curve fit based on a model with two independent binding sites amounts to $K_a \sim 1 \cdot 10^7 \text{ M}^{-1}$ for SYUIQ-5, more than a factor of three higher when compared to cryptolepine. Also, a strongly exothermic binding for SYUIQ-5 is counteracted by an unfavorable entropic contribution to binding. In contrast, binding of cryptolepine is driven by a considerably smaller enthalpic contribution with no additional entropic penalty. Clearly, such thermodynamic profiles corroborate the presence of significant sidechain interactions at the expense of a reduced SYUIQ-5 conformational freedom.

Low-affinity binding sites likely include the duplex domain and in particular the exposed 5'-face of the quadruplex that has been found to be a favored binding site for indoloquinolines in a regular parallel *c-Myc* quadruplex with its two exposed outer tetrads.^[22,25] For a direct comparison of binding affinities towards the Q-D junction and a more exposed outer G-tetrad, an additional quadruplex *Q3-sbl* was introduced. The sequence of the latter closely resembles a *c-Myc* variant that was reported to fold into a parallel quadruplex with a 4-nt snapback loop and a 3'-terminal G filling a vacant site of its 3'-tetrad.^[30,31] With an additional mutation to match the 5'-overhang in the *QD3-sbl* hybrid, *Q3-sbl* is expected to mimic *QD3-sbl* lacking a Q-D junction at its 3'-outer tetrad. As an additional benefit, the relatively short diagonal snapback loop was previously shown to effectively prevent ligand binding, allowing better defined interactions only at the 5'-face of *Q3-sbl*.^[31]

Initially, the anticipated snapback-driven parallel fold of *Q3-sbl* was demonstrated by NMR experiments (Figure S17). In line with a strong preference for the 5'-tetrad, subsequent ITC

Table 1. Binding thermodynamics of SYUIQ-5 and cryptolepine to target quadruplexes.^[a]

	N	K_a [M^{-1}]	ΔH° [kcal/mol]	ΔG°_{313} [kcal/mol] ^[b]	$-T\Delta S^\circ$ [kcal/mol] ^[b]
SYUIQ-5 to <i>QD3-sbl</i>	0.9 ± 0.1	$(1.1 \pm 0.3) \cdot 10^7$	-14.1 ± 0.2	-10.1 ± 0.2	4.0 ± 0.3
<i>Q3-sbl</i> ^[c]	1.0 ± 0.1	$(2.1 \pm 0.6) \cdot 10^6$	-12.2 ± 0.5	-9.0 ± 0.2	3.2 ± 0.4
cryptolepine to <i>QD3-sbl</i>	1.0 ± 0.1	$(3.3 \pm 0.7) \cdot 10^6$	-8.0 ± 0.1	-9.3 ± 0.1	-1.3 ± 0.2

[a] Average values and standard deviations for the high-affinity binding site obtained from three independent measurements in 120 mM K^+ buffer, pH 7, at 40 °C. [b] $\Delta G^\circ = -RT \ln K_a$ and $-T\Delta S^\circ = \Delta G^\circ - \Delta H^\circ$. [c] *Q3-sbl* sequence: 5'-TTAGGGTGGTAGGGTGGGAAGG-3'.

titrations of *Q3-sbl* with SYUIQ-5 yielded a stoichiometry of 1 for high-affinity binding. On the other hand, a corresponding association constant $K_a \sim 2 \cdot 10^6 \text{ M}^{-1}$ was smaller by a factor of five compared to binding at the Q-D interface in *QD3-sbl*, identifying the Q-D junction as a superior binding site for the indoloquinoline ligand (Table 1). Notably, the heat initially released upon ligand binding at the 5'-outer tetrad of *Q3-sbl* matches the first plateau region that follows the high-affinity binding of SYUIQ-5 to *QD3-sbl*. This suggests a first binding event at the junction with subsequent binding at the 5'-face and possibly additional binding at the duplex domain of the Q-D hybrid.

Discussion

Indoloquinoline ring systems feature a shape that maximizes π - π stacking interactions through intercalation between two guanine bases of the outer tetrad and the adjacent CG base pair in a Q-D hybrid. In addition, electrostatic interactions are promoted by the positive potential at the *N5*-protonated or *N5*-methylated quinoline nitrogen. The specific ligand alignment allows them to be directed towards the central channel of the G-core lined with the G carbonyl oxygen atoms but also towards the carbonyl oxygen of the GC base pair on top. The significance of the latter on binding may only be moderate. Conspicuously, however, cryptolepine was reported to feature a rather peculiar preference for intercalating between two CG base pairs when binding a B-type DNA duplex. In a corresponding crystal structure, stacking interactions were optimized by the excellent geometric fit of cryptolepine with the neighboring CG base pairs.^[32] Notably, in close correspondence with the present SYUIQ-5 complex structure, orientation of cryptolepine in the intercalation pocket positioned the quinoline portion of the ligand between the two guanines, allowing a close contact of the positively charged cryptolepine *N5*-methyl to both of the 6-carbonyl oxygens of the two stacked Watson-Crick paired guanines in the duplex major groove.

Disregarding any significant steric or electronic effects due to *N5*-methylation, the sidechain appended to the tetracyclic ring system in SYUIQ-5 provides for additional binding affinity through its interactions at or within the grooves, considerably increasing affinity constants when compared to cryptolepine. Also, NMR data hint at cryptolepine being subjected to enhanced exchange processes between different ligand orientations. A 180° ring flip of the indoloquinoline possibly followed by minor translational adjustments is easily conceivable for cryptolepine but clearly hampered by the SYUIQ-5 sidechain interacting within a groove. Thus, SYUIQ-5 may be restricted to bind in a more defined orientation.

The only two high-resolution structures reported to date for Q-D hybrids complexed with ligands have revealed two rather divergent binding modes. One study reported on the binding of simple mono- and polyaromatic compounds built on a benzylamine substructure to the *QD2-I* antiparallel quadruplex with its central hairpin-type lateral loop.^[33] A bis-aminomethylated anthracene ligand stacks on the two exposed guanines

of the outer G-tetrad at the junction, being in-plane with the interfacial first GC base pair to form a pseudo-triad. One of the two protonated, positively charged amino substituents on the ligand points towards the central channel of the G-core. In addition to interactions with the central electron-rich guanine oxygen atoms, weaker electrostatic and hydrogen bond interactions with opposite residues at the duplex major groove can also be envisaged. Conspicuously, such a binding geometry is reminiscent of several complexes with polycyclic ligands stacking on a quadruplex outer tetrad. In contrast to macrocycles covering the whole tetrad area, these ligands are often found to bind opposite of an in-plane base recruited from overhang but also loop sequences.^[22,34,35] Possible hydrogen bond interactions within such a pseudo-base pair stacked on top of the outer G-tetrad are often supplemented with a loose cap of another overhang/loop residue.

In the same study, some of the ligands were also used to bind the major G-quadruplex formed in the U3 promoter region of the HIV-1 long terminal repeat (*LTR-III*), being of considerable interest as a novel antiviral target. Notably, the *LTR-III* quadruplex comprises a 12-nt diagonal loop with a duplex-stem but with highly dynamic residues between quadruplex and duplex domains.^[4] Although no three-dimensional complex structure has been reported, biophysical data suggested analogous binding modes for this class of ligands when targeting *QD2-I* and biologically relevant *LTR-III*.^[33]

Another high-resolution structure with binding at the Q-D junction was determined for a conjugated ligand composed of a quadruplex-specific naphthalene diimide (NDI) core linked to a positively charged platinum coordination complex [Pt-(dien)(py)].^[36] In close correspondence to the present indoloquinoline binding, the NDI ring system was found to be sandwiched between interfacial outer tetrad and neighboring base pair of a quadruplex with a lateral duplex stem-loop. Specific binding was further promoted by the platinum coordinated sidechain, interacting within the duplex minor groove through hydrogen bonds and electrostatic interactions.

It should be pointed out that the potential intercalation of a ligand between two G-tetrads should likewise yield favorable binding energies through π - π stacking interactions with two adjacent tetrads. However, such a binding mode has not yet been confirmed on short cation-stabilized quadruplexes and only intercalation between non-conventional GAGA and GCGC quartets of an unusual G-rich tetrahelical structure has been reported for a bis-quinolinium compound.^[37] Also, porphyrin intercalation into long G4 DNA nanowires has only been evidenced in the absence of monovalent cations whereas non-intercalative binding was suggested in a K^+ solution.^[38] Apparently, in addition to the considerable energetic cost when unstacking G-tetrads associated with the unwinding of four strands to provide for an intercalation pocket, a metal ion located within the central channel between tetrads seems to restrict access of a corresponding ligand.

Taken together, planar aromatic ring systems with surface areas only covering part of a G-tetrad as mostly found for quadruplex ligands may bind in two distinctive modes at a Q-D junction (Figure 9). Both involve vertical π - π stacking onto the

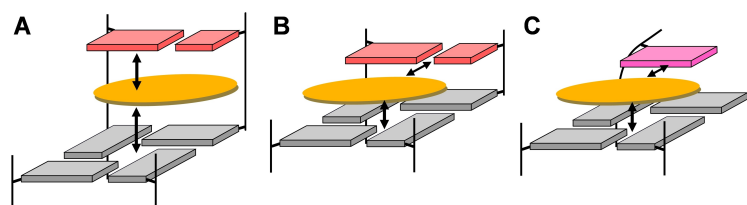


Figure 9. Schematic representation of binding modes for a ligand aromatic moiety to a quadruplex outer tetrad. (A) Intercalation between G-tetrad and base pair at a Q-D junction. (B) Stacking on outer tetrad in-plane with an interfacial base pair at the junction. (C) Stacking on outer tetrad in-plane with an overhang or loop residue of an isolated quadruplex without Q-D interface. Arrows indicate direction of major π - π stacking, electrostatic, and/or hydrogen bond interactions of the orange-colored ligand; bases of the G-tetrad, base pair, and overhang or loop are colored gray, red, and magenta, respectively.

outer tetrad as a major contributor to binding. Additional horizontal electrostatic and hydrogen bond interactions with the interfacial base pair add to the complex stabilization in case of the base pair aligned opposite and in-plane with the ligand. On the other hand, intercalation between outer tetrad and a duplex base pair at a Q-D junction may occur if vertical stacking and electrostatic interactions of the intercalated ligand aromatic moiety overcome the energetic penalty associated with unwinding at the interface to create a binding pocket. It can be assumed that a matched shape and electrostatic potential of ligand and intercalation pocket will strongly favor ligand insertion. In this context it is worth mentioning that a Q-D hybrid structure featuring a base triad platform between quadruplex and duplex motifs in its crystal form was unable to bind a ligand, most likely as a result of a largely occluded G-tetrad surface area in case of an in-plane binding mode or of a larger energy barrier for strand unwinding at the tetrad-triad junction in case of an intercalative binding mode.^[39]

Although being highly flexible in many cases, ligand aliphatic sidechains are important in providing for additional short-lived electrostatic, hydrogen bond, and/or van der Waals interactions. In case of an intercalative binding mode, sidechains seem to favor interactions within the minor groove of the duplex domain and if appropriate may not only increase affinities but also selectivities towards the target hybrid. Finally, as suggested by the NMR analysis of unsubstituted cryptolepine, sidechains may effectively restrict ligand dynamics and exchange between different ligand alignments to fix a major ligand orientation.

Interest in structural details of Q-D interfaces and in their recognition by ligands increasingly grow with the realization that various Q-D junctions can potentially form within G-rich sequences of genomic DNA. It seems obvious to utilize the unique features of quadruplex-duplex interfaces for various technological applications, for example as additional structural motifs in aptamer constructs, but also as hotspots for drug targeting, trying to improve affinities and especially selectivities towards a particular quadruplex-forming site. Up to now, the design of selective and potent ligands to target Q-D junctions is an area still in its very infancy. An obvious approach based on the combination of a large aromatic surface area of a G-selective ligand with typical duplex minor groove binders may suffer from the large size and molecular weight of the conjugates. On the other hand, the specific targeting of Q-D

junctions with small molecules needs more systematic studies. The three-dimensional structure of a Q-D junction complexed with indoloquinolines adds valuable information on the binding selectivity and the ligand binding mode. Critical interactions seem to rely on structural but also electrostatic complementarity that may be strengthened through additional hydrogen bond interactions, for example by appropriate sidechains. The results presented may thus constitute a helpful guide for the future design and development of ligands specifically targeting Q-D interfaces.

Experimental Section

Materials and sample preparation

DNA oligonucleotides were synthesized by TIBMOLBIOL (Berlin, Germany). Samples were additionally purified by ethanol precipitation. Concentration of oligonucleotides was determined by their absorbance A_{260} at 80 °C using a molar extinction coefficient as provided by the manufacturer. The concentration of commercially available SYUIQ-5 and cryptolepine (Sigma-Aldrich Chemie GmbH, Germany) was determined from its weighed mass. Except for the ITC experiments, samples were dissolved in 10 mM potassium phosphate buffer, pH 7.0.

UV melting experiments

UV melting experiments were performed with a Jasco V-650 spectrophotometer (Jasco, Tokyo, Japan) equipped with a Peltier thermostat. For duplex melting, the hyperchromicity of the oligonucleotide solution (2 μ M) was followed at $\lambda = 260$ nm as a function of temperature. For melting of the quadruplex domain, the hypochromicity of the oligonucleotide solution (5 μ M) was observed at $\lambda = 295$ nm. Data were recorded from 10 °C to 90 °C with a heating rate of 0.2 °C min^{-1} and a bandwidth of 1 nm. Melting temperatures were determined by the first derivative of the melting curve. Melting temperatures of the DNA-indoloquinoline complexes were not determined due to their broad, uncooperative melting profiles. All experiments were done in triplicate.

Differential scanning calorimetry (DSC)

To circumvent inaccuracies in UV melting due to mutual perturbations of duplex and quadruplex absorbances, melting temperatures for a *QD3-sbl* solution (50 μ M) were additionally determined by DSC. DSC experiments were performed with a VP-DSC instrument (Malvern Instruments, United Kingdom). The sample was heated

with a heating rate of $0.5^{\circ}\text{C min}^{-1}$ from 10°C to 80°C . Melting temperatures and enthalpy values were determined from a second sample vs. buffer scan after subtracting a buffer vs. buffer scan. A cubic baseline fitting was used and the two peaks associated with duplex and quadruplex melting were deconvoluted. Data were analyzed with the Origin software.

CD spectroscopy

CD spectra were recorded at 20°C with a Jasco J-810 spectropolarimeter equipped with a Peltier thermostat (Jasco, Tokyo, Japan). For recording CD spectra of the Q-D hybrid ($5\ \mu\text{M}$), a bandwidth of $1\ \text{nm}$, a scanning speed of $50\ \text{nm min}^{-1}$, a response time of $4\ \text{s}$, and 5 accumulations were used. A concentrated solution of SYUIQ-5 in DMSO was used for titrations up to a 5:1 ligand-to-DNA molar ratio. All spectra were blank-corrected.

NMR spectroscopy

For NMR experiments, a Bruker Avance 600 MHz NMR spectrometer equipped with an inverse $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{19}\text{F}$ quadruple resonance cryoprobehead and z-field gradients was used. Spectra were processed in TopSpin 4.0.7 and assigned in CcpNMR V2.^[40] Unless indicated otherwise, spectra were acquired on samples in 90% $\text{H}_2\text{O}/10\% \text{D}_2\text{O}$ buffered with $10\ \text{mM}$ potassium phosphate, pH 7.0. SYUIQ-5 was used as a concentrated stock solution in $\text{DMSO-}d_6$. The final concentration of DMSO after the addition of one equivalent of ligand was about 2%. Proton chemical shifts were referenced to TSP through the temperature dependent water chemical shift while ^{13}C chemical shifts were referenced to DSS through an indirect referencing method. For solvent suppression, a WATERGATE w5 pulse scheme was employed in 1D and 2D NOESY experiments whereas a 3–9–19 water suppression scheme was used for DQF-COSY, TOCSY, and $^1\text{H}-^{13}\text{C}$ HSQC experiments. $^1\text{H}-^{13}\text{C}$ HSQC spectra were acquired with $4\ \text{K}\times 500$ data points, a $1\ \text{s}$ recycle delay, and a spectral width of $7500\ \text{Hz}$ in the F1 dimension to cover aromatic C8/C6/C2 carbon resonances of the nucleobases. DQF-COSY and TOCSY spectra with a mixing time of $80\ \text{ms}$ and a DIPSI-2 isotropic mixing scheme were recorded with $4\ \text{K}\times 500$ data points. 2D NOESY spectra with 80 , 150 , and $300\ \text{ms}$ mixing times and EASY-ROESY spectra acquired with a $80\ \text{ms}$ mixing time and a 50° spinlock angle were acquired with $2\ \text{K}\times 1\ \text{K}$ data points. For all 2D homonuclear experiments a $2\ \text{s}$ recycle delay was used. Spectra were zero-filled to $4\ \text{K}\times 1\ \text{K}$ data points and processed with a squared sine-bell window function except for 1D experiments which were multiplied with an exponential function.

ITC experiments

ITC experiments were performed with a Microcal PEAQ-ITC microcalorimeter (Malvern Instruments, United Kingdom) employing a reference power of $4\ \mu\text{cal s}^{-1}$. Oligonucleotides and indoloquinoline ligands were dissolved in $20\ \text{mM}$ potassium phosphate buffer, pH 7.0, supplemented with $100\ \text{mM}$ KCl and 5% DMSO. A ligand solution ($1.5\ \mu\text{L}$, $400\ \mu\text{M}$) was titrated to the oligonucleotide solution ($20\ \mu\text{M}$) with an injection duration of $3\ \text{s}$ and a spacing of $240\ \text{s}$. The first injection ($0.4\ \mu\text{L}$) was discarded before data analysis. Thermograms were subsequently fitted to a model with two sets of binding sites with the Microcal-PEAQ ITC analysis software. All experiments were blank- and concentration-corrected and performed in triplicate.

Structure calculations

Employing NMR-derived distance and dihedral angle restraints as well as H-bond, planarity and repulsion restraints, 100 starting structures were generated for both free and complexed DNA by a simulated annealing protocol in XPLOR-NIH 3.0.3.^[41] Structures were refined using AMBER16 with the parmbsc force field and OL15 modifications for DNA. An additional force field was employed for the ligand and parameterized for AMBER using the R.E.D server.^[42] Geometry optimization for the ligand was done with Hartree-Fock calculations and a 6-31G* basis set and the force field parameters were adapted from parm10 and GAFF. The 100 starting structures were subjected to a simulated annealing protocol to yield 20 lowest-energy structures. Refinement in water was done by neutralizing the DNA with potassium ions, placing two ions in the inner core of the G-quadruplex between two tetrad layers, and soaking the system with TIP3P water in a $10\ \text{\AA}$ truncated octahedral box. The final simulation was done at $1\ \text{atm}$ and $300\ \text{K}$ for $4\ \text{ns}$ using only NOE- and hydrogen bond-based distance restraints. For free QD3-sbl, the trajectory was averaged for the last 500 ps. In contrast, only the last snapshot was used in the complex calculations to prevent distortions of the flexible ligand aliphatic sidechain. Structures were further minimized to obtain ten lowest-energy structures. Structure parameters were extracted with the X3DNA web package.^[43] More details of the structure calculations are given in the Supporting Information.

Acknowledgements

This research was funded by the DFG (project no. 410497337). We thank Dr. Jonathan Dickerhoff for helpful discussions and his support with the structure calculations. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The atomic coordinates and chemical shifts for free QD3-sbl (PDB 7PNE; BMRB 34664) and of its complex with SYUIQ-5 (PDB 7PNG; BMRB 34665) have been deposited in the Protein Data Bank.

Keywords: indoloquinolines · intercalation · isothermal titration calorimetry · NMR spectroscopy · quadruplex-duplex junctions

- [1] 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.
- [2] T. Q. N. Nguyen, K. W. Lim, A. T. Phan, *Sci. Rep.* **2017**, *7*, 11969.
- [3] K. W. Lim, T. Q. N. Nguyen, A. T. Phan, *J. Am. Chem. Soc.* **2014**, *136*, 17969–17973.
- [4] E. Butovskaya, B. Heddi, B. Bakalar, S. N. Richter, A. T. Phan, *J. Am. Chem. Soc.* **2018**, *140*, 13654–13662.
- [5] D. J. Y. Tan, F. R. Winnerdy, K. W. Lim, A. T. Phan, *Nucleic Acids Res.* **2020**, *48*, 11162–11171.
- [6] T. Q. N. Nguyen, K. W. Lim, A. T. Phan, *Nucleic Acids Res.* **2020**, *48*, 10567–10575.

- [7] B. Karg, S. Mohr, K. Weisz, *Angew. Chem. Int. Ed.* **2019**, *58*, 11068–11071; *Angew. Chem.* **2019**, *131*, 11184–11188.
- [8] T. Q. N. Nguyen, K. W. Lim, A. T. Phan, *J. Phys. Chem. B* **2020**, *124*, 5122–5130.
- [9] S. Mohr, J. Jana, Y. M. Vianney, K. Weisz, *Chem. Eur. J.* **2021**, *27*, 10437–10447.
- [10] J. Zhou, A. Bourdoncle, F. Rosu, V. Gabelica, J. L. Mergny, *Angew. Chem. Int. Ed.* **2012**, *51*, 11002–11005; *Angew. Chem.* **2012**, *124*, 11164–11167.
- [11] A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R. B. Darnell, D. J. Patel, *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804.
- [12] N. Vasilyev, A. Polonskaia, J. C. Darnell, R. B. Darnell, D. J. Patel, A. Serganov, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5391–E5400.
- [13] R. F. Macaya, J. A. Waldron, B. A. Beutel, H. Gao, M. E. Joesten, M. Yang, R. Patel, A. H. Bertelsen, A. F. Cook, *Biochemistry* **1995**, *34*, 4478–4492.
- [14] I. Russo Krauss, V. Spiridonova, A. Pica, V. Napolitano, F. Sica, *Nucleic Acids Res.* **2016**, *44*, 983–991.
- [15] S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Eur. J.* **2018**, *24*, 4428–4435.
- [16] F. Riechert-Krause, K. Weisz, *Heterocycl. Commun.* **2013**, *19*, 145–166.
- [17] C. W. Wright, J. Addae-Kyereme, A. G. Breen, J. E. Brown, M. F. Cox, S. L. Croft, Y. Gökçek, H. Kendrick, R. M. Phillips, P. L. Pollet, *J. Med. Chem.* **2001**, *44*, 3187–3194.
- [18] L. Guittat, P. Alberti, F. Rosu, S. Van Miert, E. Thetiot, L. Pieters, V. Gabelica, E. De Pauw, A. Ottaviani, J. F. Riou, J. L. Mergny, *Biochimie* **2003**, *85*, 535–547.
- [19] T. M. Ou, Y. J. Lu, C. Zhang, Z. S. Huang, X. D. Wang, J. H. Tan, Y. Chen, D. L. Ma, K. Y. Wong, J. C. O. Tang, A. S. C. Chan, L. Q. Gu, *J. Med. Chem.* **2007**, *50*, 1465–1474.
- [20] J. L. Zhou, Y. J. Lu, T. M. Ou, J. M. Zhou, Z. S. Huang, X. F. Zhu, C. J. Du, X. Z. Bu, L. Ma, L. Q. Gu, Y. M. Li, A. S. C. Chan, *J. Med. Chem.* **2005**, *48*, 7315–7321.
- [21] J. M. Zhou, X. F. Zhu, Y. J. Lu, R. Deng, Z. S. Huang, Y. P. Mei, Y. Wang, W. L. Huang, Z. C. Liu, L. Q. Gu, Y. X. Zeng, *Oncogene* **2006**, *25*, 503–511.
- [22] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [23] N. Deng, L. Wickstrom, P. Cieplak, C. Lin, D. Yang, *J. Phys. Chem. B* **2017**, *121*, 10484–10497.
- [24] Y. J. Lu, T. M. Ou, J. H. Tan, J. Q. Hou, W. Y. Shao, D. Peng, N. Sun, X. D. Wang, W. Bin Wu, X. Z. Bu, Z. S. Huang, D. L. Ma, K. Y. Wong, L. Q. Gu, *J. Med. Chem.* **2008**, *51*, 6381–6392.
- [25] A. Funke, J. Dickerhoff, K. Weisz, *Chem. Eur. J.* **2016**, *22*, 3170–3181.
- [26] A. Funke, K. Weisz, *Biochimie* **2019**, *157*, 142–148.
- [27] Y. M. Vianney, P. Preckwinkel, S. Mohr, K. Weisz, *Chem. Eur. J.* **2020**, *26*, 16910–16922.
- [28] K. W. Lim, A. T. Phan, *Angew. Chem. Int. Ed.* **2013**, *52*, 8566–8569; *Angew. Chem.* **2013**, *125*, 8728–8731.
- [29] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [30] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173.
- [31] L. Schnarr, J. Jana, P. Preckwinkel, K. Weisz, *J. Phys. Chem. B* **2020**, *124*, 2778–2787.
- [32] J. N. Lisgarten, M. Coll, J. Portugal, C. W. Wright, J. Aymami, *Nat. Struct. Mol. Biol.* **2002**, *9*, 57–60.
- [33] L. Diaz-Casado, I. Serrano-Chacón, L. Montalvillo-Jiménez, F. Corzana, A. Bastida, A. G. Santana, C. González, J. L. Asensio, *Chem. Eur. J.* **2021**, *27*, 6204–6212.
- [34] N. H. Campbell, M. Patel, A. B. Tofa, R. Ghosh, G. N. Parkinson, S. Neidle, *Biochemistry* **2009**, *48*, 1675–1680.
- [35] J. Dickerhoff, J. Dai, D. Yang, *Nucleic Acids Res.* **2021**, *49*, 5905–5915.
- [36] L.-Y. Liu, K.-N. Wang, W. Liu, Y.-L. Zeng, M.-X. Hou, J. Yang, Z. W. Mao, *Angew. Chem. Int. Ed.* **2021**, *60*, 20833–20839.
- [37] A. Kotar, V. Kocman, J. Plavec, *Chem. Eur. J.* **2020**, *26*, 814–817.
- [38] I. Lubitz, N. Borovok, A. Kotlyar, *Biochemistry* **2007**, *46*, 12925–12929.
- [39] I. Russo Krauss, S. Ramaswamy, S. Neidle, S. Haider, G. N. Parkinson, *J. Am. Chem. Soc.* **2016**, *138*, 1226–1233.
- [40] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins Struct. Funct. Genet.* **2005**, *59*, 687–696.
- [41] C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
- [42] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, F. Y. Dupradeau, *Nucleic Acids Res.* **2011**, *39*, W511–W517.
- [43] X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.

Manuscript received: October 14, 2021
Accepted manuscript online: December 14, 2021
Version of record online: January 5, 2022

Chemistry–A European Journal

Supporting Information

Indoloquinoline Ligands Favor Intercalation at Quadruplex-Duplex Interfaces

Yoanes Maria Vianney and Klaus Weisz*

Contents

Experimental details on the structure calculations	S2
Table S1: Sequences of oligonucleotides	S4
Figure S1: NOESY spectral regions of free <i>QD3-sbl</i>	S4
Figure S2: ¹ H- ¹³ C HSQC spectral regions of free <i>QD3-sbl</i>	S5
Figure S3: Determination of the sugar conformations in free <i>QD3-sbl</i>	S6
Table S2: Compilation of ¹ H and ¹³ C chemical shifts in free <i>QD3-sbl</i>	S7
Table S3: NMR restraints and structural statistics	S9
Figure S4: DSC melting curve of <i>QD3-sbl</i>	S10
Table S4: Melting temperatures of <i>QD3-sbl</i>	S10
Figure S5: CD titration of <i>QD3-sbl</i> with SYUIQ-5	S11
Figure S6: NOESY spectral regions of a <i>QD3-sbl</i> - SYUIQ-5 complex	S12
Figure S7: ¹ H- ¹³ C HSQC spectral regions of a <i>QD3-sbl</i> - SYUIQ-5 complex	S13
Figure S8: Determination of sugar conformations in a <i>QD3-sbl</i> - SYUIQ-5 complex	S14
Figure S9: NOESY and ROESY spectra of free and SYUIQ-5-bound <i>QD3-sbl</i>	S15
Figure S10: Proton resonance assignments of bound SYUIQ-5	S16
Table S5: Compilation of ¹ H and ¹³ C chemical shifts in a <i>QD3-sbl</i> - SYUIQ-5 complex	S17
Table S6: Compilation of ¹ H chemical shifts in bound SYUIQ-5	S18
Table S7: List of intermolecular NOE contacts	S19
Figure S11: NOESY spectral regions with intermolecular contacts	S19
Figure S12: Q-D interface before and after ligand binding	S20
Figure S13: NOESY spectral regions and topology of free <i>QD2-l</i>	S21
Figure S14: NOESY spectral regions of a <i>QD3-sbl</i> – cryptolepine complex	S22
Figure S15: NOESY and ROESY spectra of free and cryptolepine-bound <i>QD3-sbl</i>	S23
Figure S16: ITC thermograms for ligand binding	S24
Figure S17: NOESY spectral regions and topology of <i>Q3-sbl</i>	S25

Structure calculations

For both free and complexed DNA, 100 starting structures were selected from 500 structures generated by a simulated annealing protocol in XPLOR-NIH 3.0.3.^[1] Distance restraints were grouped based on cross-peak intensities in NOESY experiments. For free *QD3-sbl*, distances for non-exchangeable protons were fixed to 2.9 ± 1.1 Å for strong cross-peaks, 4.0 ± 1.5 Å for medium intensity cross-peaks, 5.5 ± 1.5 Å for weak cross-peaks, and 6.0 ± 1.5 Å for very weak cross-peaks. For exchangeable protons, distances were set to 2.9 ± 1.1 Å for strong cross-peaks, 4.0 ± 1.2 Å for medium intensity cross-peaks, 5.0 ± 1.2 Å for weak cross-peaks, and 6.0 ± 1.2 Å for very weak cross-peaks. A distance of 5.0 ± 2.0 Å was used for overlapped peaks. Distance ranges for the ligand-DNA complex were fixed according to distances of non-exchangeable protons in free *QD3-sbl* with no further discrimination between exchangeable and non-exchangeable protons due to broadening effects of exchangeable cross-peaks associated with higher uncertainties. Intermolecular distances between ligand and DNA protons were set to $4.0 +1.5/-2.5$ Å for strong cross-peaks, $5.5 +1.5/-2.5$ Å for weak cross-peaks, $6.0 +1.5/-2.5$ Å for very weak cross-peaks, and $5.0 +2.0/-2.5$ Å for overlapped peaks. Additional repulsion restraints were set to 10.0 ± 5.0 Å for non-exchangeable peaks. Torsion angles χ were set to *anti* for residues 1-35 ($170^\circ - 310^\circ$) and to *syn* for G36 ($25^\circ - 95^\circ$). All sugar puckers that could be determined based on DQF-COSY experiments were fixed to a south conformation with a pseudorotation angle of $144^\circ - 180^\circ$. Empirical hydrogen bond-based distance and planarity restraints were added for the three G-tetrads and the seven Watson-Crick base pairs. For structure calculations of the complex, additional α ($260^\circ - 320^\circ$) and β ($190^\circ - 250^\circ$) backbone torsion angle restraints, adopted from experimental data also including dinucleotide-drug model complexes, were employed for residues C19 and G36.^[2,3]

The structure was refined without planarity restraints for the duplex base-pairs using AMBER16 with the parmbsc force field and OL15 modifications for the DNA. For the complex, an additional force field was employed for the ligand and parameterized for AMBER using the R.E.D server.^[4] Geometry optimization for the ligand was done with Hartree-Fock calculations and a 6-31G* basis set and the force field parameters were adapted from parm10 and GAFF. The 100 starting structures were subjected to a simulated annealing protocol to yield 20 lowest-energy structures. Restraint energies were $40 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ for NOE-based distance restraints, $200 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$ for dihedral angle restraints, and $30 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ for planarity restraints. The system was equilibrated at 300 K for 5 ps, followed by heating to 1000 K over 10 ps. Keeping this temperature for the next 30 ps, the system was subsequently cooled to 100 K within 45 ps and to 0 K within 10 ps.

Refinement in water was done by initially neutralizing the DNA with potassium ions, placing two ions in the inner core of the G-quadruplex between two tetrad layers. The system was soaked

with TIP3P water in a 10 Å truncated octahedral box and subsequently minimized with 500 steps of steepest descent minimization followed by another 500 steps of conjugate gradient minimization. The DNA was fixed with a force constant of 25 kcal·mol⁻¹·Å⁻². The system was heated up from 100 K to 300 K under constant volume in 10 ps and further equilibrated under a constant pressure of 1 atm with decreasing energy restraints of 5, 4, 3, 2, 1, and 0.5·kcal·mol⁻¹·Å⁻². The final simulation was done at 1 atm and 300 K for 4 ns using only NOE- and hydrogen bond-based distance restraints. For free *QD3-sbl*, the trajectory was averaged for the last 500 ps. In contrast, only the last snapshot was used in the complex calculations to prevent distortions of the flexible ligand aliphatic sidechain. Structures were further minimized to obtain ten lowest-energy structures. Structure parameters were extracted with the X3DNA web package.^[5]

References

- [1] C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
- [2] H.-S. Shieh, H. M. Berman, M. Dabrow, S. Neidle, *Nucleic Acids Res.* **1980**, *8*, 85–97.
- [3] A. Adams, J. M. Guss, C. A. Collyer, W. A. Denny, L. P. G. Wakelin, *Biochemistry* **1999**, *38*, 9221–9233.
- [4] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, F. Y. Dupradeau, *Nucleic Acids Res.* **2011**, *39*, 511–517.
- [5] X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.

Table S1. Oligonucleotides used in this study.

name	sequence ^a
<i>QD3-sbl</i>	5'-TTAG <u>GTGGGTAGGGTGGG</u> CTAGTCATTTGACTAG <u>G</u> -3'
<i>QD2-l</i>	5'- <u>GGTTGGCGCGAAG</u> CATTCGCGGGT <u>TGG</u> -3'
<i>Q3-sbl</i>	5'-TTAG <u>GGGTAGGTGGGTGGG</u> GAAG <u>G</u> -3'

^aG residues participating in G-tetrad formation are underlined.

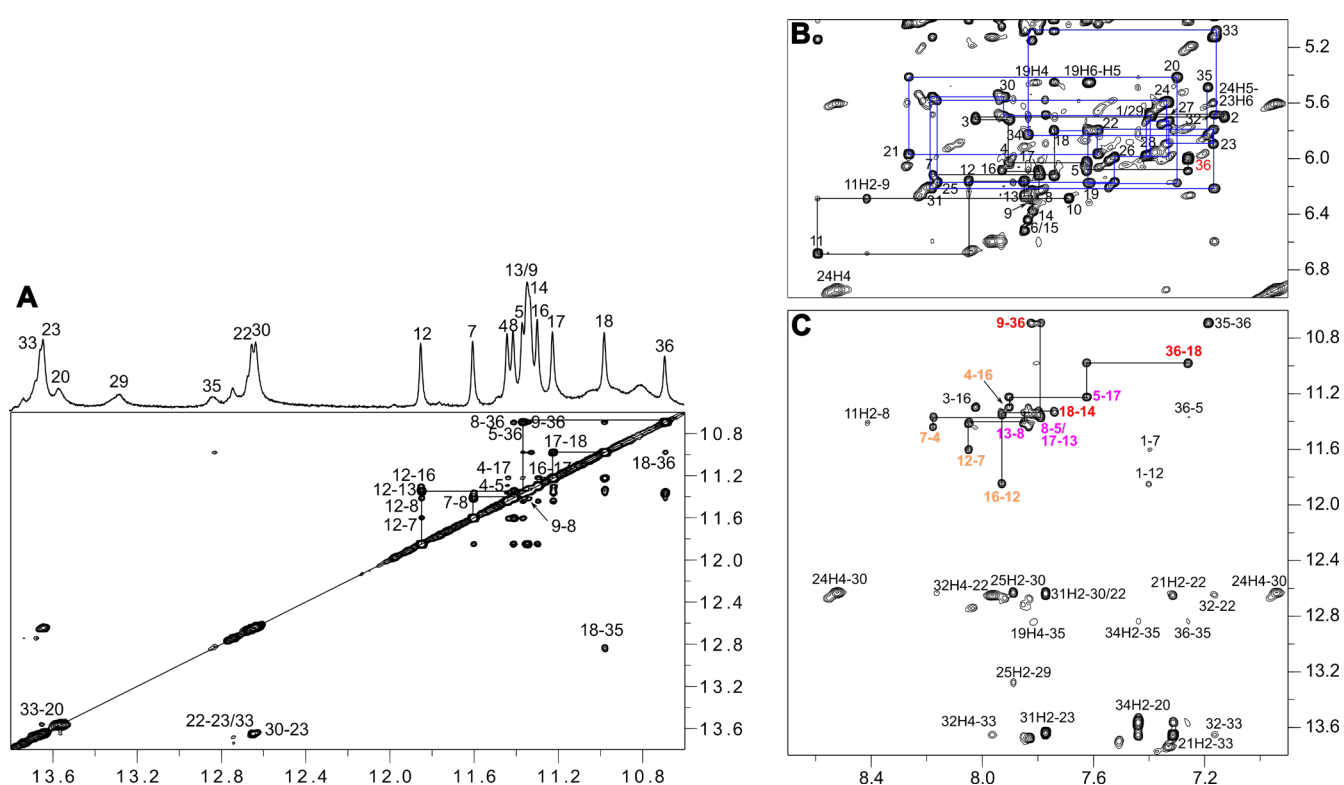


Figure S1. 2D NOESY spectrum (300 ms mixing time, 20 °C) of *QD3-sbl* (1 mM). (A) Imino(ω_2)-imino(ω_1) spectral region; a corresponding 1D spectrum with resonance assignments is given on top. (B) H8/6(ω_2)-H1'(ω_1) spectral region; an intra-nucleotide H8-H1' contact of *syn*-G36 is labeled in red and H8/H6-H1' connectivities within the duplex stem-loop are traced with blue lines. (C) H8/6/2(ω_2)-imino(ω_1) spectral region; intra-tetrad H8(ω_2)-H1(ω_1) cross-peaks are labeled with colors depending on G-tetrad layer; inter-tetrad connectivities along the G-columns are traced by horizontal and vertical lines.

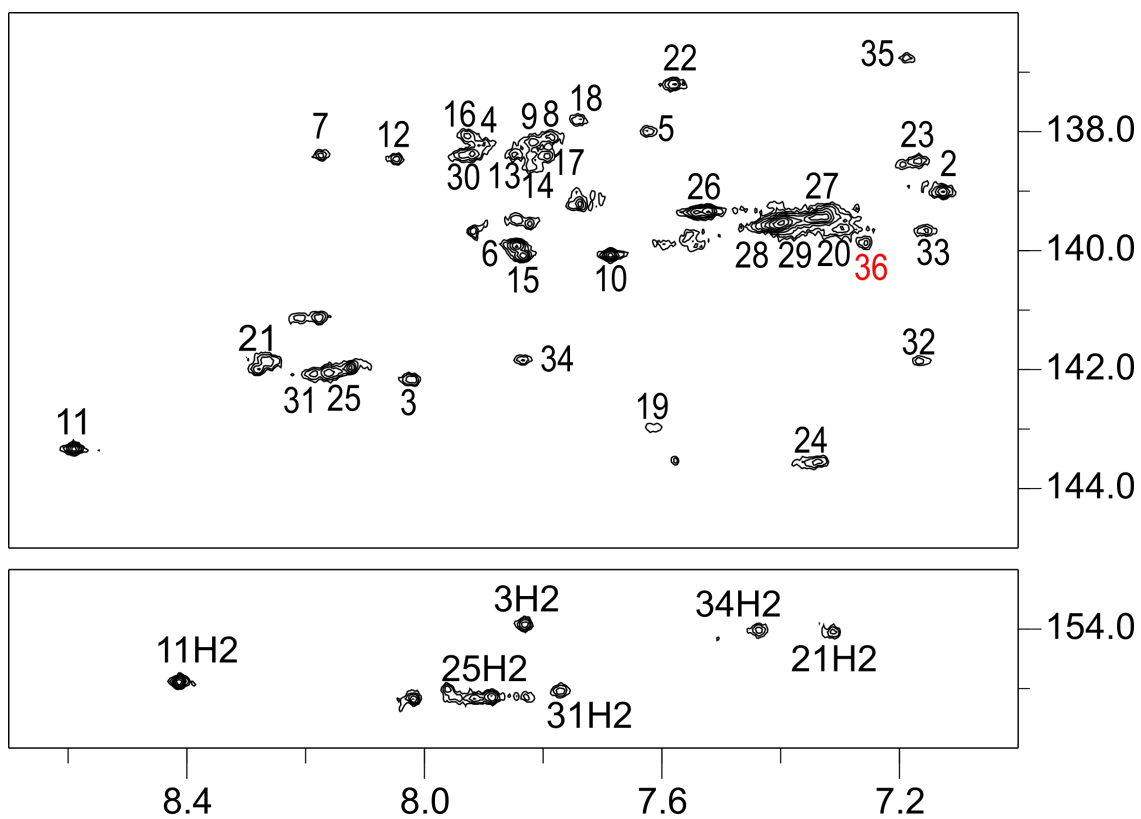


Figure S2. H8/H6(ω_2)-C8/C6(ω_1) (top) and H2(ω_2)-C2(ω_1) (bottom) spectral region of a ^1H - ^{13}C HSQC spectrum of *QD3-sbl* (1 mM) acquired at 20 °C. The *syn*-G36 cross-peak is labeled in red.

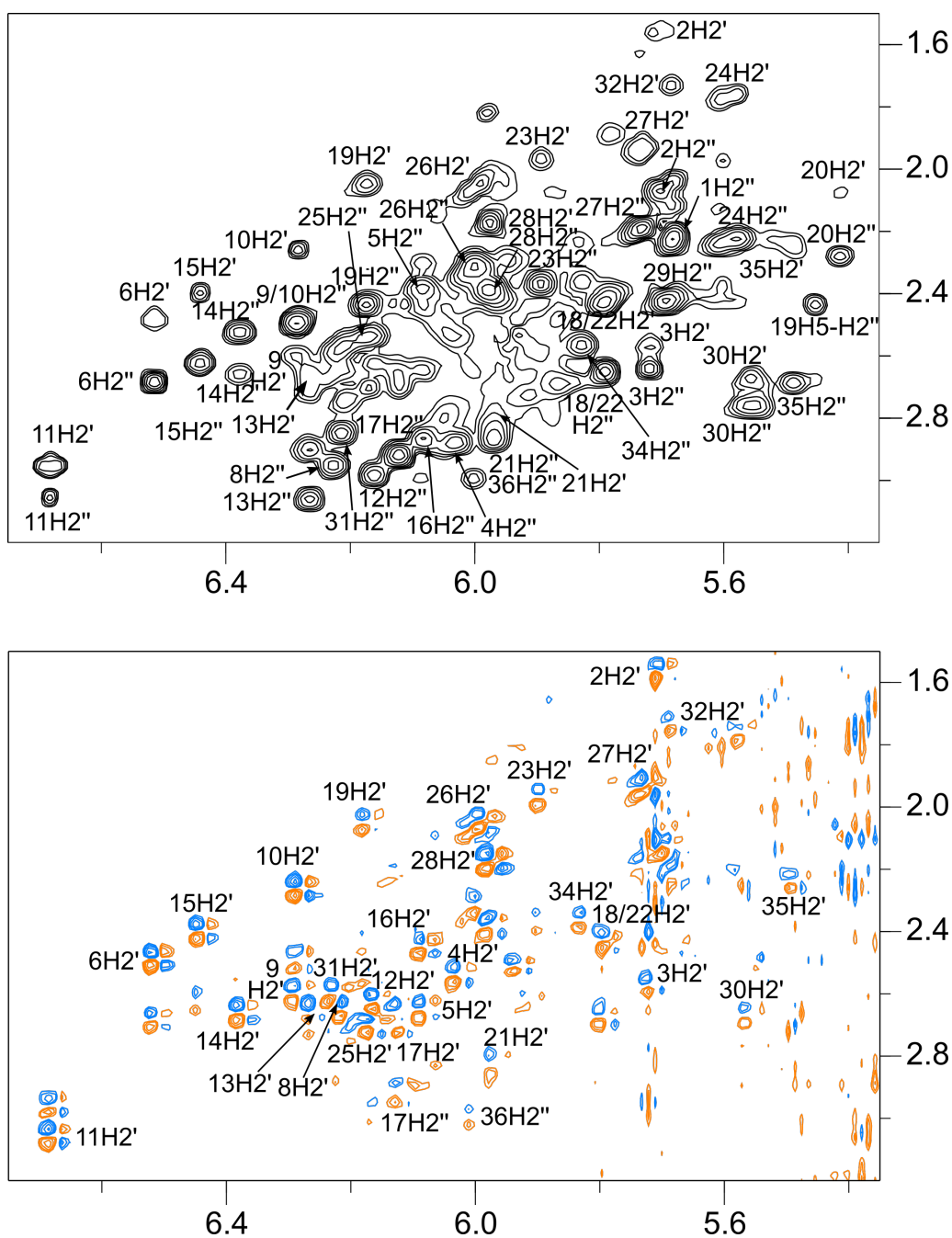


Figure S3. Determination of the sugar conformation. 2D NOESY (top, mixing time 80 ms) and DQF-COSY spectrum (bottom) with H1'(ω₂)-H2'/H2''(ω₁) cross-peaks of *QD3-sbl*. The spectra were acquired at 20 °C. Conformational analysis of DQF-COSY cross-peak patterns and scalar couplings followed stereospecific H2'/H2'' assignments in the NOESY spectrum at short mixing times.

Table S2. ^1H and ^{13}C chemical shifts δ of *QD3-sbl*.^a

δ (ppm)	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T1	7.40	n.d	5.68	2.04/2.22	4.41	1.62	139.54	-
T2	7.13	n.d	5.70	1.57/2.07	4.51	1.50	139.02	-
A3	8.02	-	5.72	2.57/2.64	4.85	7.83	142.17	153.93
G4	7.90	11.44	6.03	2.54/2.88	4.91	-	138.22	-
G5	7.62	11.37	6.08	2.65/2.39	5.01	-	137.99	-
T6	7.85	n.d	6.51	2.48/2.68	5.08	1.99	139.93	-
G7	8.18	11.60	6.12	2.69/2.92	4.95	-	138.41	-
G8	7.79	11.41	6.22	2.60/2.96	5.01	-	138.07	-
G9	7.82	11.34	6.29	2.60/2.51	5.03	-	138.19	-
T10	7.69	n.d.	6.28	2.26/2.49	4.77	1.95	140.08	-
A11	8.59	-	6.68	3.06/2.96	5.15	8.41	143.34	154.91
G12	8.05	11.85	6.16	2.63/2.98	5.00	-	138.46	-
G13	7.85	11.35	6.27	2.65/3.06	5.04	-	138.39	-
G14	7.82	11.33	6.38	2.66/2.53	5.15	-	138.42	-
T15	7.84	n.d.	6.44	2.40/2.62	4.99	1.96	140.09	-
G16	7.93	11.30	6.08	2.45/2.87	5.05	-	138.09	-
G17	7.80	11.22	6.12	2.69/2.92	5.08	-	138.41	-
G18	7.74	10.98	5.79	2.43/2.65	5.00	-	137.82	-
C19	7.62	-	6.17	2.05/2.44	4.90	5.45	142.99	-
T20	7.30	13.56	5.41	2.08/2.28	4.85	1.62	139.63	-
A21	8.27	-	5.97	2.82/2.88	5.03	7.31	141.86	154.05
G22	7.58	12.65	5.79	2.43/2.65	4.79	-	137.21	-
T23	7.17	13.64	5.89	1.97/2.37	4.76	1.23	138.49	-
C24	7.34	-	5.58	1.77/2.23	4.76	5.61	143.57	-
A25	8.16	-	6.17	2.71/2.54	4.97	7.89	142.05	155.15

T26	7.52	n.d.	5.99	2.05/2.32	4.67	1.78	139.35	-
T27	7.33	n.d.	5.73	1.93/2.19	4.60	1.58	139.43	-
T28	7.41	n.d.	5.98	2.17/2.39	4.70	1.63	139.57	-
T29	7.39	13.28	5.69	2.13/2.43	4.70	1.82	139.53	-
G30	7.92	12.63	5.56	2.67/2.76	4.97	-	138.38	-
A31	8.18	-	6.21	2.65/2.85	4.99	7.77	142.06	155.04
C32	7.16	-	5.68	1.73/2.24	4.60	5.13	141.86	-
T33	7.16	13.65	5.08	1.84/1.94	4.60	1.43	139.67	-
A34	7.84	-	5.83	2.37/2.57	4.79	7.44	141.83	154.01
G35	7.19	12.84	5.49	2.24/2.69	4.86	-	136.77	-
G36	7.26	10.69	6.00	2.30/3.00	4.69	-	139.89	-

^aAt 20 °C in 10 mM potassium phosphate buffer, pH 7.0.

Table S3. NMR restraints and structural statistics of calculated structures.

NOE distance restraints	<i>QD3-sbl</i>	<i>QD3-sbl</i> – SYUIQ-5
intra-residual	157	143
inter-residual	265	208
exchangeable	101	64
intermolecular	---	11
repulsion	---	3
other restraints:		
hydrogen bonds	82	82
dihedral angles	62	73
planarity	10	10
structural statistics:		
pairwise heavy atom RMSD value (Å)		
all residues	2.5 ± 0.43	2.75 ± 0.45
G-tetrad core	0.85 ± 0.24	1.00 ± 0.16
duplex stem-loop	1.80 ± 0.62	2.15 ± 0.64
NOE violations:		
maximum violation	0.275 Å	0.204 Å
number of violations (> 0.2)	0.1 ± 0.3	0.1 ± 0.3
mean NOE violation	0.002 ± 0.001	0.002 ± 0.0004
deviations from idealized geometry:		
bond lengths (Å)	0.01 ± 0.0001	0.01 ± 0.0001
bond angles (degree)	2.23 ± 0.03	2.35 ± 0.03

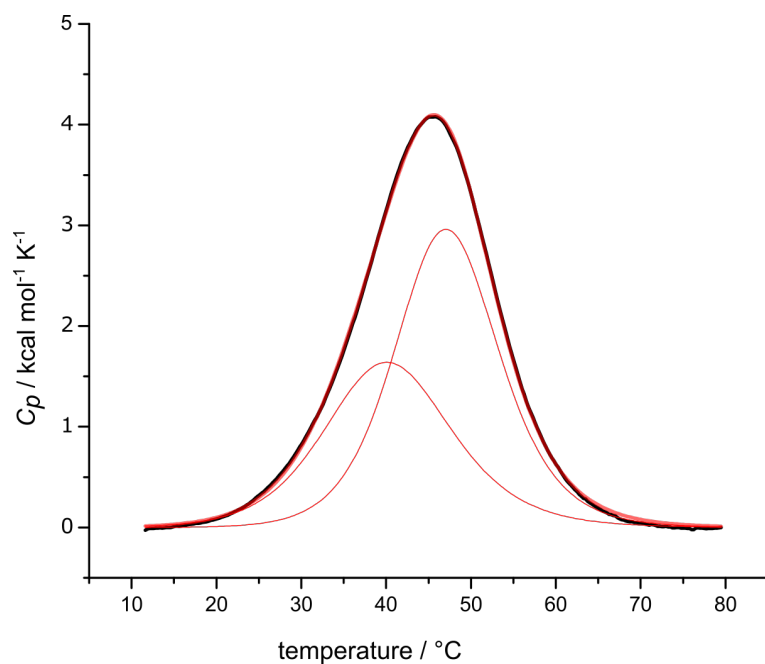


Figure S4. Representative DSC thermogram (black) for *QD3-sbl* in 10 mM KP_i buffer, pH 7.0, and superimposed fit with two melting transitions (red).

Table S4. Melting temperatures T_m of quadruplex and duplex domains of the *QD3-sbl* hybrid.

	T_m ($^{\circ}\text{C}$) ^a	T_m ($^{\circ}\text{C}$) ^b	$\Delta H^{\circ}_{\text{cal}}$ (kcal/mol) ^b	$\Delta H^{\circ}_{\text{vH}}$ (kcal/mol) ^b
quadruplex	40.5 ± 0.9	40.2 ± 0.5	33.3 ± 0.8	38.2 ± 1.2
duplex	47.5 ± 1.1	47.2 ± 0.2	49.1 ± 1.3	49.2 ± 0.6

^aFrom UV melting; average values with standard deviations from three independent measurements.

^bFrom DSC; average values with standard deviations from three independent measurements.

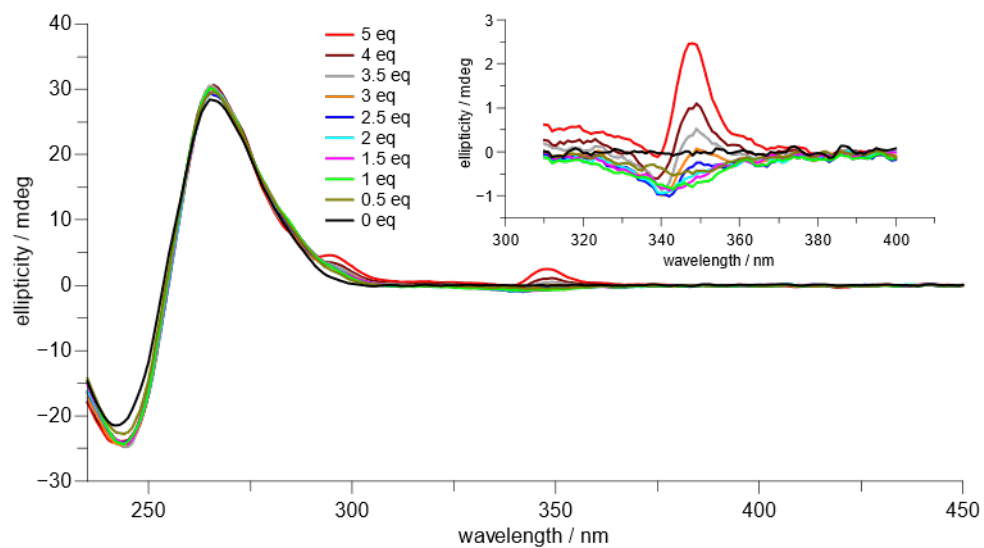


Figure S5. CD titration of *QD3-sbl* with SYUIQ-5 (0-5 equivalents). Spectra were acquired at 20 °C in 10 mM KP_i buffer, pH 7.0.

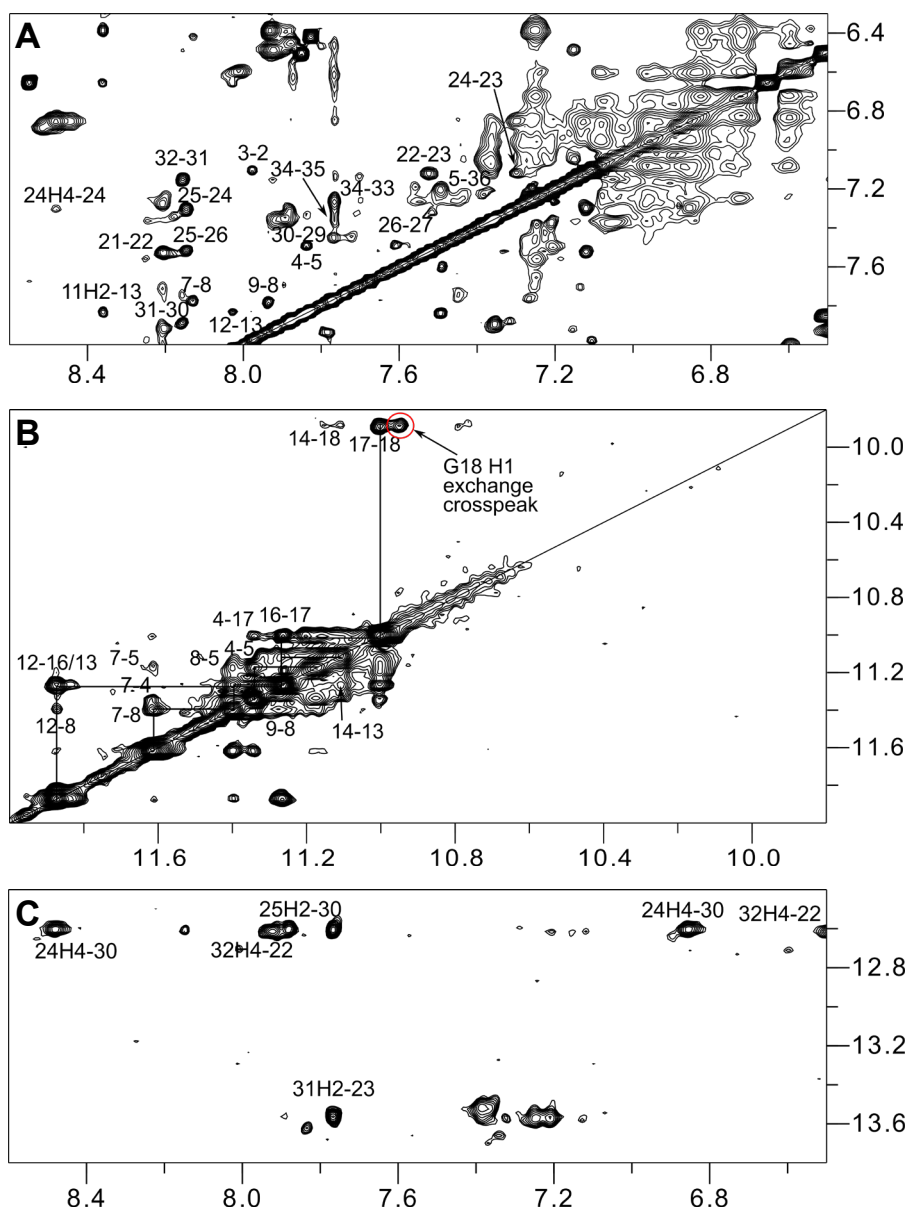


Figure S6. Regions of a 2D NOESY spectrum (300 ms mixing time, 30 °C) with assignments of non-exchangeable and exchangeable protons for a 1:1 mixture of SYUIQ-5 and *QD3-sbl* (1 mM). (A) H6/8(ω_2)-H6/8(ω_1), (B) H1(ω_2)-H1(ω_1), and (C) C H4/A H2(ω_2)-imino(ω_1) spectral region. In (B), an exchange cross-peak for G18 H1 between the free and the complexed structure is circled in red (for a corresponding ROESY spectrum see Figure S9).

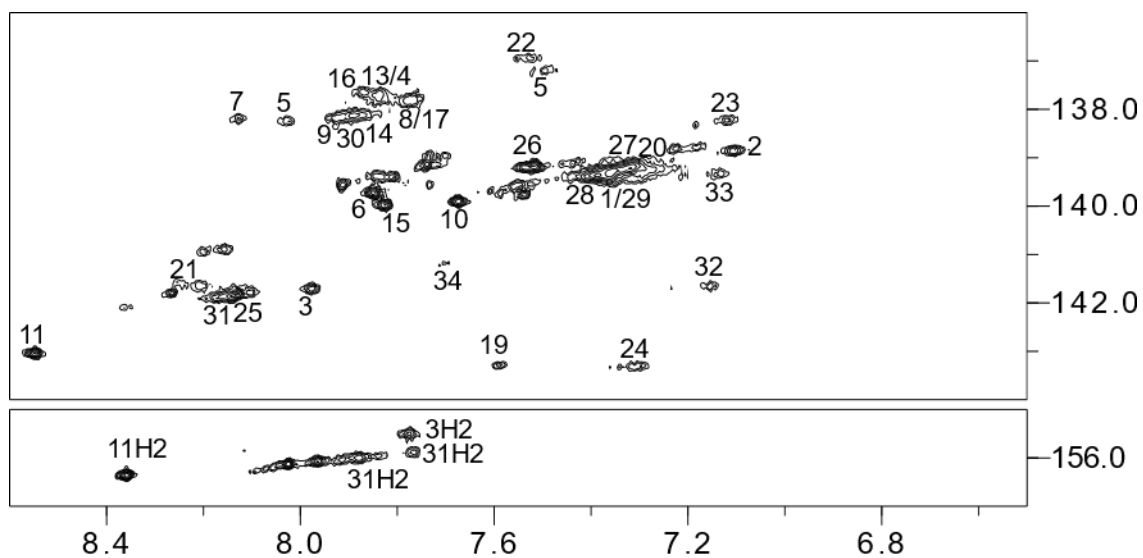


Figure S7. H8/H6(ω_2)-C8/C6(ω_1) (top) and H2(ω_2)-C2(ω_1) (bottom) HSQC spectral region of *QD3-sbl* (1 mM) in the presence of 1 eq. SYUIQ-5. The spectra were acquired at 30 °C.

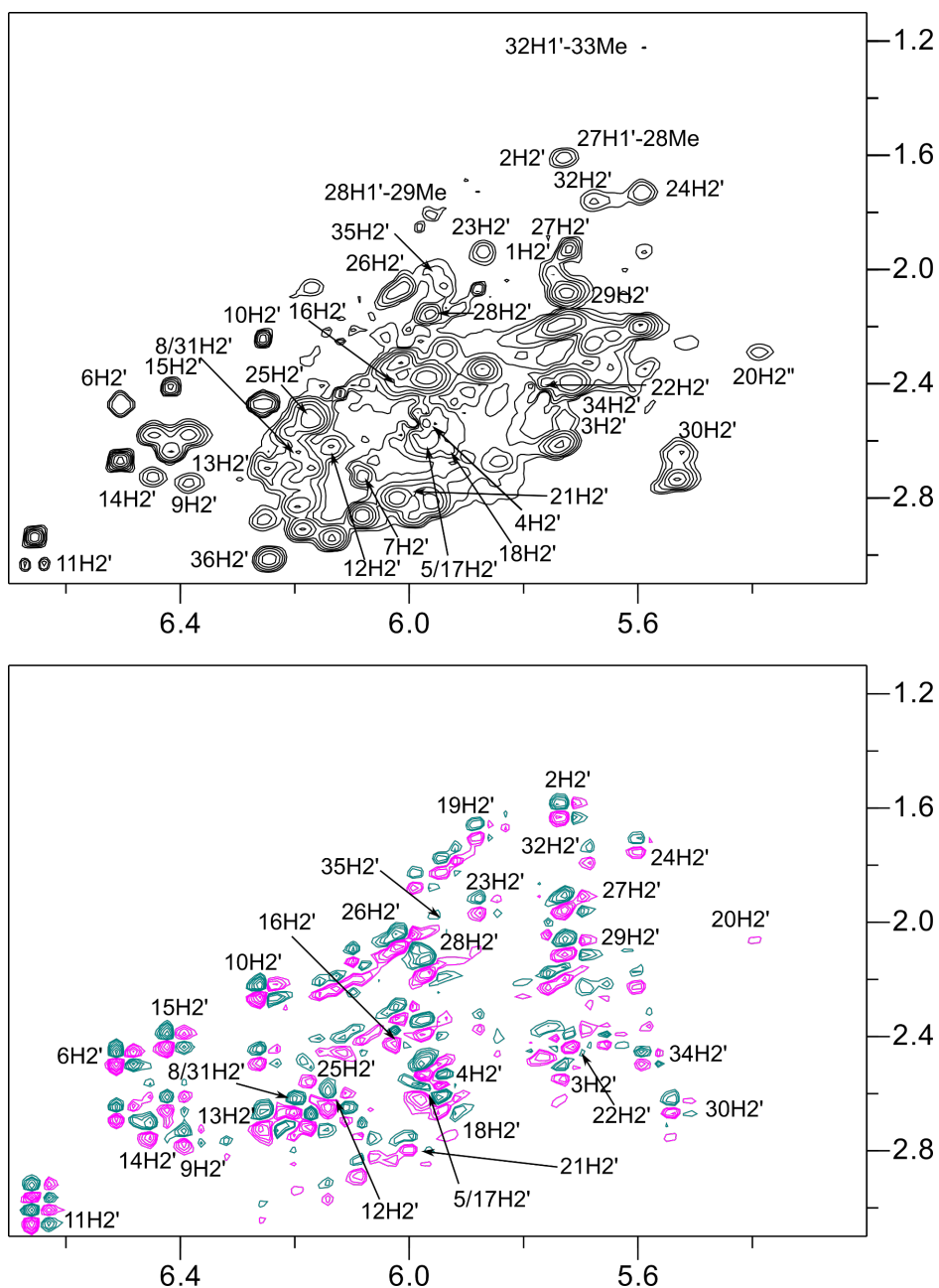


Figure S8. Determination of the sugar conformation in the 1:1 complex of *QD3-sbl* (1 mM) and SYUIQ-5. 2D NOESY (top, mixing time 80 ms) and DQF-COSY spectrum (bottom) with H1'(ω₂)-H2'/H2''(ω₁) cross-peaks. The spectra were acquired at 30 °C. Conformational analysis of DQF-COSY cross-peak patterns and scalar couplings followed stereospecific H2'/H2'' assignments in the NOESY spectrum at short mixing times; correlations for the T33 residue are only observed at lower threshold levels.

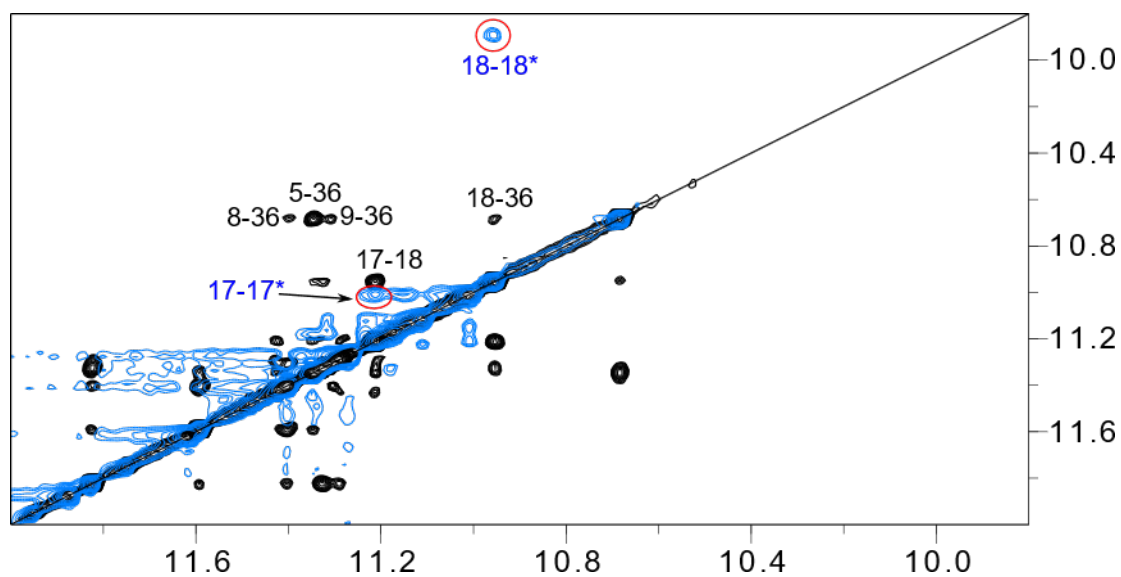


Figure S9. 2D NOESY spectrum of free *QD3-sbl* (300 ms mixing time, black) with superimposed ROESY spectrum of *QD3-sbl* in the presence of 0.5 eq. SYUIQ-5 (blue, only positive contours shown for clarity). Spectra show guanine H1(ω_2)-H1(ω_1) correlations of the G-core and were acquired at 30 °C. Exchange cross-peaks of G17 H1 and G18 H1 between free and complexed *QD3-sbl* are circled in red.

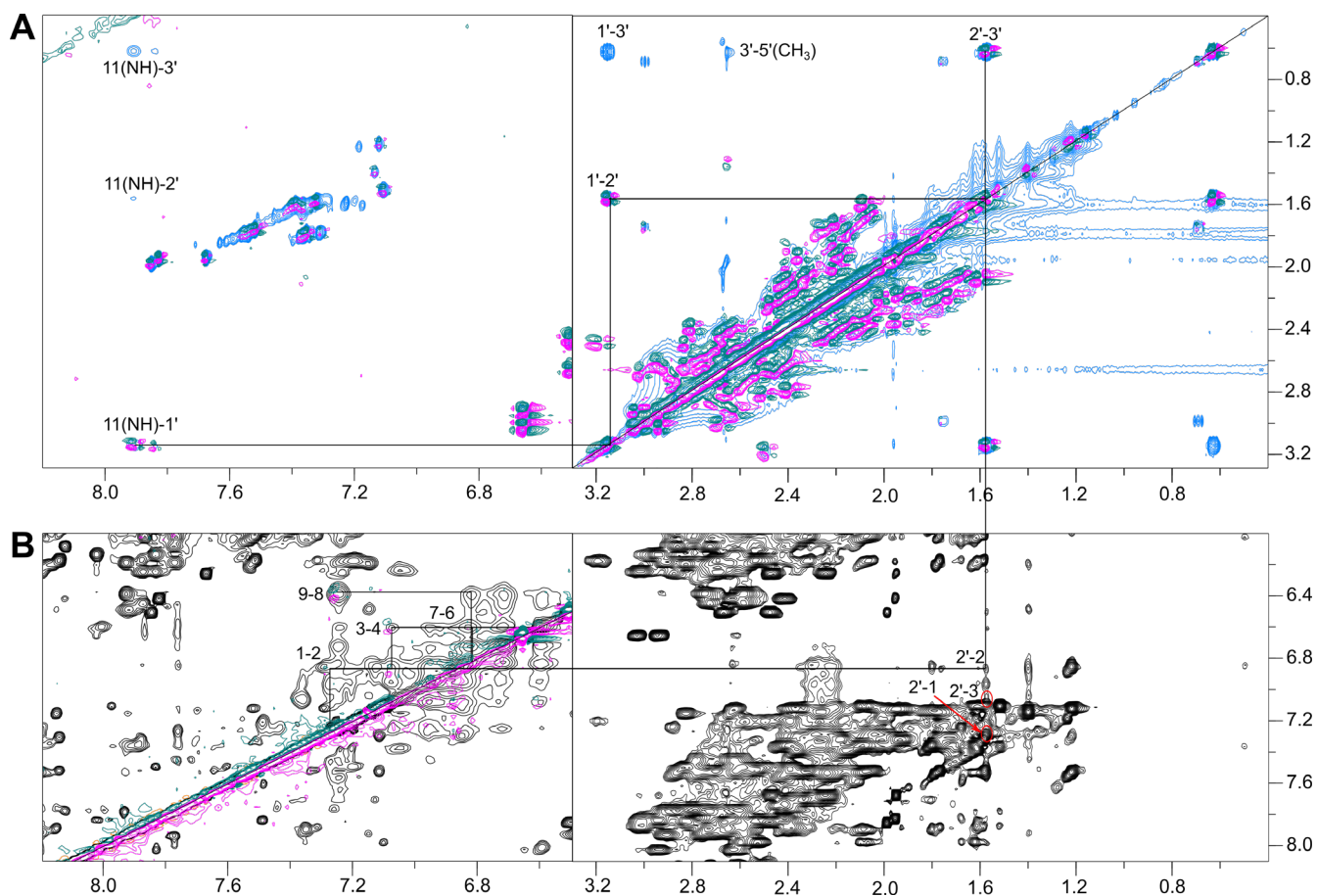


Figure S10. Proton resonance assignments of SYUIQ-5 bound to *QD3-sbl₂*. (A) Superimposed DQF-COSY and TOCSY spectra showing cross-peaks between SYUIQ-5 aliphatic side chain protons. (B) Superimposed DQF-COSY and NOESY spectra (300 ms mixing time) with correlations between aromatic protons (left) and NOESY spectral region with connectivities from SYUIQ-5 aliphatic side chain protons to aromatic protons of the quinoline subunit (right). Spectra were acquired at 30 °C.

Table S5. ^1H and ^{13}C chemical shifts δ of *QD3-sbl* complexed with SYUIQ-5 in a 1:1 molar ratio.^a

δ (ppm)	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T1	7.36	n.d	5.75	2.02/2.21	4.43	1.62	139.33	-
T2	7.11	n.d	5.73	1.61/2.08	4.51	1.52	138.87	-
A3	7.98	-	5.73	2.53/2.60	4.82	7.77	141.71	154.51
G4	7.84	11.34	5.96	2.51/2.83	4.90	-	137.73	-
G5	7.49	11.18	5.98	2.58/2.39	5.00	-	137.20	-
T6	7.85	n.d	6.50	2.48/2.67	5.08	1.99	139.72	-
G7	8.13	11.61	6.08	2.73/2.86	4.87	-	138.21	-
G8	7.78	11.40	6.19	2.64/2.91	5.02	-	137.83	-
G9	7.93	11.27	6.39	2.75/2.58	5.04	-	138.17	-
T10	7.68	n.d.	6.26	2.24/2.47	4.76	1.95	139.91	-
A11	8.55	-	6.65	3.03/2.94	5.14	8.36	143.06	155.36
G12	8.03	11.87	6.13	2.62/2.94	4.99	-	138.25	-
G13	7.84	11.26	6.24	2.69/3.02	5.03	-	137.73	-
G14	7.88	11.11	6.45	2.73/2.58	5.15	-	138.12	-
T15	7.83	n.d.	6.42	2.41/2.63	5.01	1.95	139.98	-
G16	7.87	11.26	6.02	2.40/2.80	5.03	-	137.63	-
G17	7.77	11.00	5.96	2.60/2.81	5.05	-	137.82	-
G18	7.45	9.89	5.93	2.63/2.78	n.d.	-	n.d.	-
C19	7.59	-	5.87	1.68/n.d.	n.d.	5.72	143.31	-
T20	7.28	n.d.	5.39	2.04/2.29	4.80	1.58	139.26	-
A21	8.21	-	5.99	2.77/2.87	5.00	n.d.	141.66	n.d.
G22	7.53	12.61	5.74	2.39/2.62	4.78	-	136.93	-
T23	7.12	13.56	5.87	1.94/2.35	4.75	1.22	138.24	-
C24	7.30	-	5.59	1.73/2.19	4.74	5.59	143.31	-
A25	8.15	-	6.17	2.70/2.53	4.96	7.88	141.86	155.00

T26	7.52	n.d.	6.01	2.07/2.32	4.67	1.77	139.18	-
T27	7.32	n.d.	5.72	1.93/2.18	4.59	1.59	139.22	-
T28	7.39	n.d.	5.96	2.16/2.37	4.69	1.61	139.37	-
T29	7.35	n.d.	5.72	2.08/2.40	4.68	1.80	139.28	-
G30	7.89	12.60	5.53	2.64/2.73	4.95	-	138.12	-
A31	8.15	-	6.19	2.64/2.83	4.96	7.77	141.87	154.89
C32	7.15	-	5.68	1.77/2.26	4.60	5.11	141.65	-
T33	7.14	n.d.	5.22	1.80/1.99	4.63	1.40	139.34	-
A34	7.70	-	5.64	2.41/2.34	4.78	n.d.	141.19	n.d.
G35	7.19	n.d.	5.95	2.01/2.40	n.d.	-	n.d.	-
G36	7.20	n.d.	6.17	2.26./3.19	n.d.	-	n.d.	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

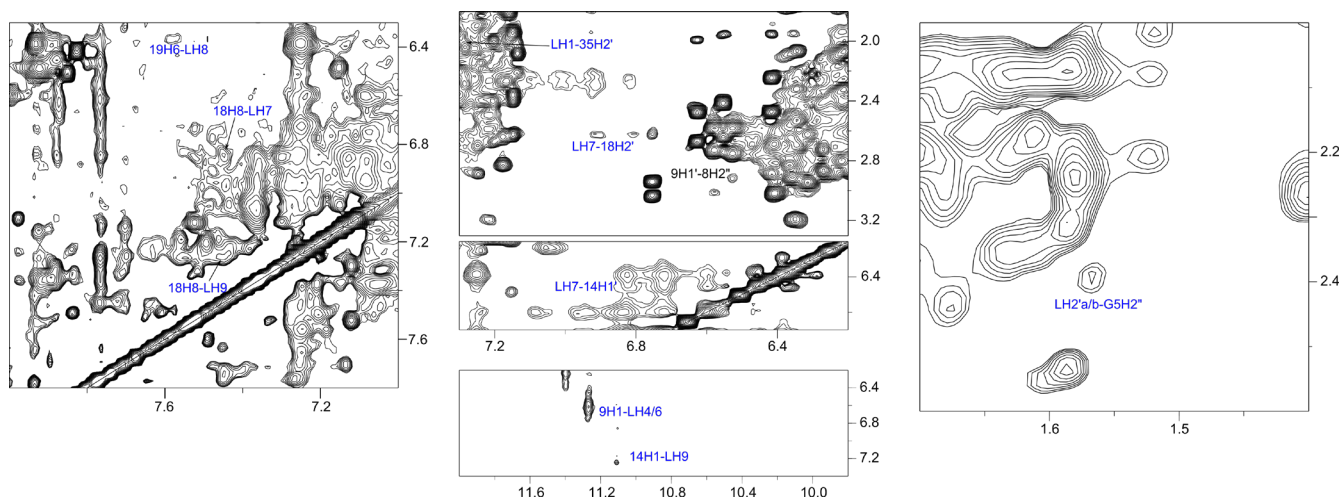
Table S6. ¹H chemical shifts of SYUIQ-5 bound to *QD3-sbl* at a 1:1 molar ratio.^a

δ (ppm)	LH1	LH2	LH3	LH4	LNH5	LH6	LH7	LH8
	7.28	6.88	7.08	6.60	n.d.	6.60	6.83	6.38
δ (ppm)	LH9	LNH10	LNH11	LH1'	LH2'	LH3'	LH5'	
	7.25	n.d.	7.90	3.15	1.57	0.62	2.66	

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

Table S7. Intermolecular NOE contacts observed between *QD3-sbl* and bound SYUIQ-5.

ligand	proton	cross-peak intensity				
		Q-D hybrid	strong	weak	very weak	overlapped
LH1	G35 H2'					
LH4	G9 H1					
LH6	G9 H1					
LH7	G14 H1'					
LH7	G18 H2'					
LH7	G18 H8					
LH8	C19 H6					
LH9	G14 H1					
LH9	G18 H1					
LH9	G18 H8					
LH2'a/b	G5 H2''					

**Figure S11.** Expanded spectral regions of a 2D NOESY spectrum (300 ms mixing time, 30 °C) on a *QD3-sbl* - SYUIQ-5 1:1 mixture with intermolecular NOE contacts between ligand and the Q-D hybrid labeled in blue.

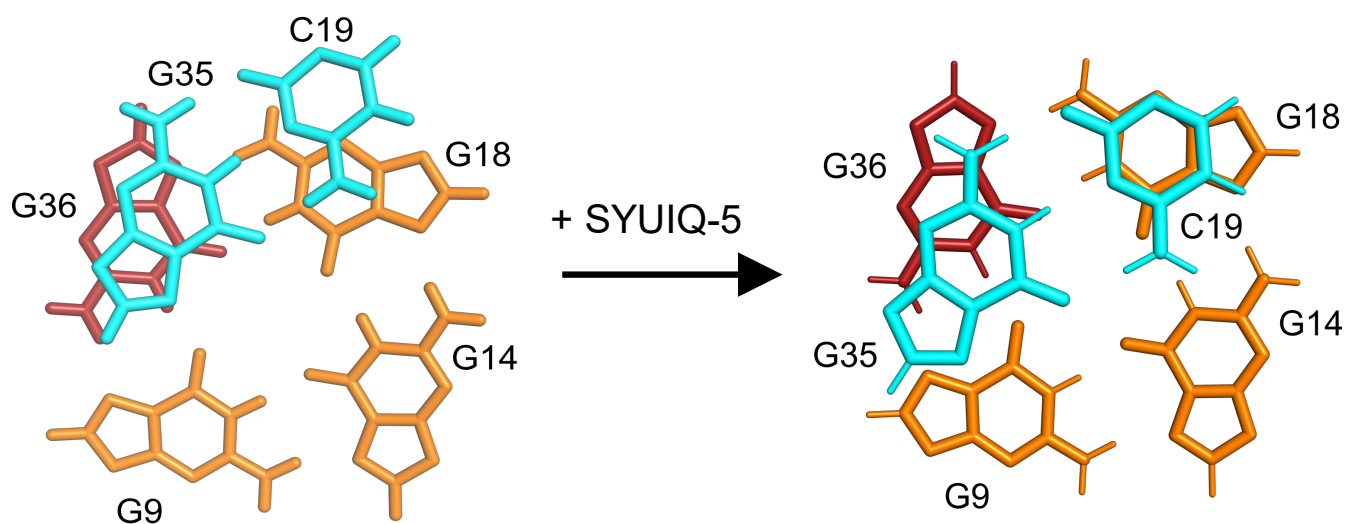


Figure S12. Q-D interface with 3'-outer tetrad and interfacial CG base pair on top. The latter is shifted towards the center of the G-core after ligand binding (ligand has been omitted for clarity); *anti*-G residues of the tetrad, *syn*-G36, and the CG base pair are colored orange, red, and cyan, respectively.

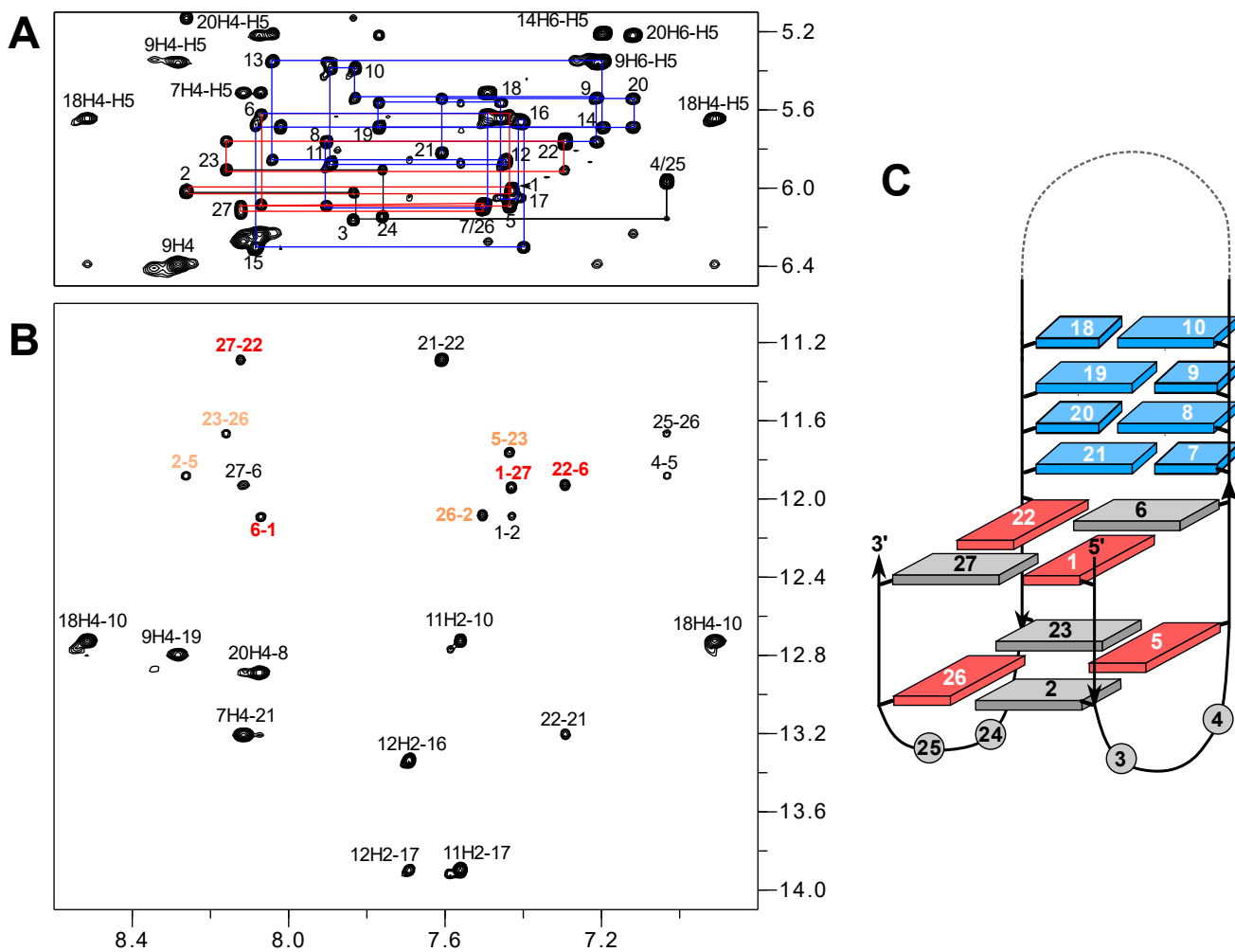


Figure S13. 2D NOESY spectral regions (300 ms mixing time, 20 °C) of *QD2-I* (0.5 mM) showing (A) H8/H6(ω_2)-H1'(ω_1) and (B) H8/6(ω_2)-imino(ω_1) cross-peaks; rectangular NOE patterns of *syn-anti* steps typical for the antiparallel strand alignment are indicated by red lines whereas duplex H8/H6-H1' NOE walks are traced by blue lines; intra-tetrad H8(ω_2)-H1(ω_1) cross-peaks in (B) are labeled with colors depending on G-tetrad layer. (C) Schematic representation of *QD2-I*; *syn*-guanosines, *anti*-guanosines, and duplex residues are colored red, grey, and blue, respectively.

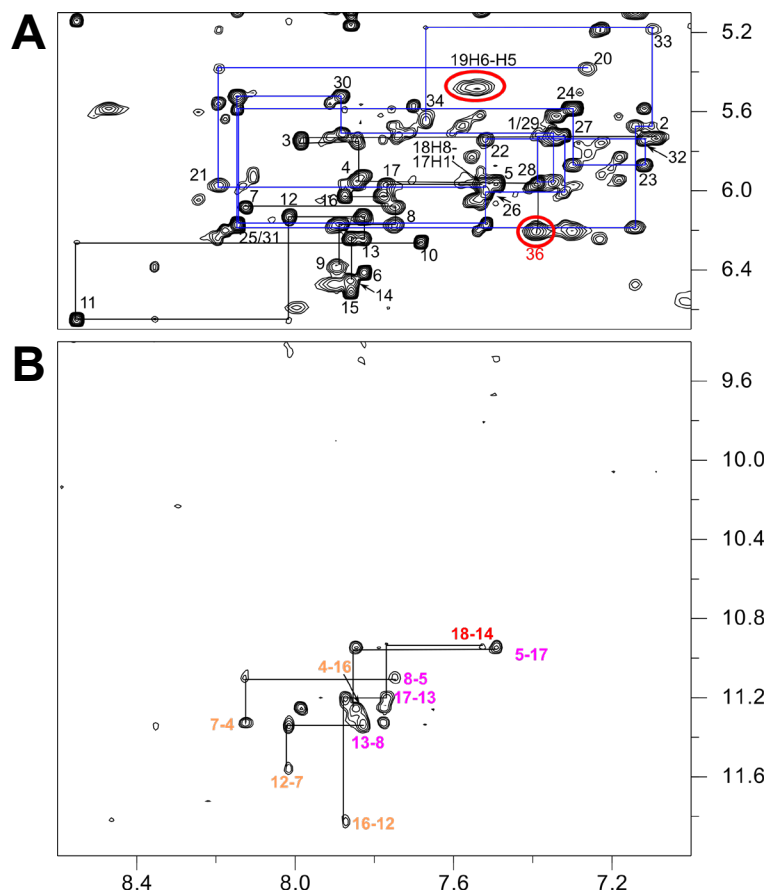


Figure S14. 2D NOESY spectral regions of *QD3-sbl* (0.5 mM) in the presence of 1 eq. cryptolepine. (A) H6/8(ω_2)-H1'(ω_1) connectivities with continuous networks indicated by blue (duplex domain) and black (quadruplex domain) horizontal and vertical lines; broadened intranucleotide H6-H5 and H8-H1' NOE contacts of interfacial C19 and *syn*-G36 are marked by red circles. (B) H6/8(ω_2)-H1(ω_1) connectivities; intra-tetrad H8(ω_2)-H1(ω_1) cross-peaks are labeled with colors depending on G-tetrad layer. Spectra were acquired at 30 °C.

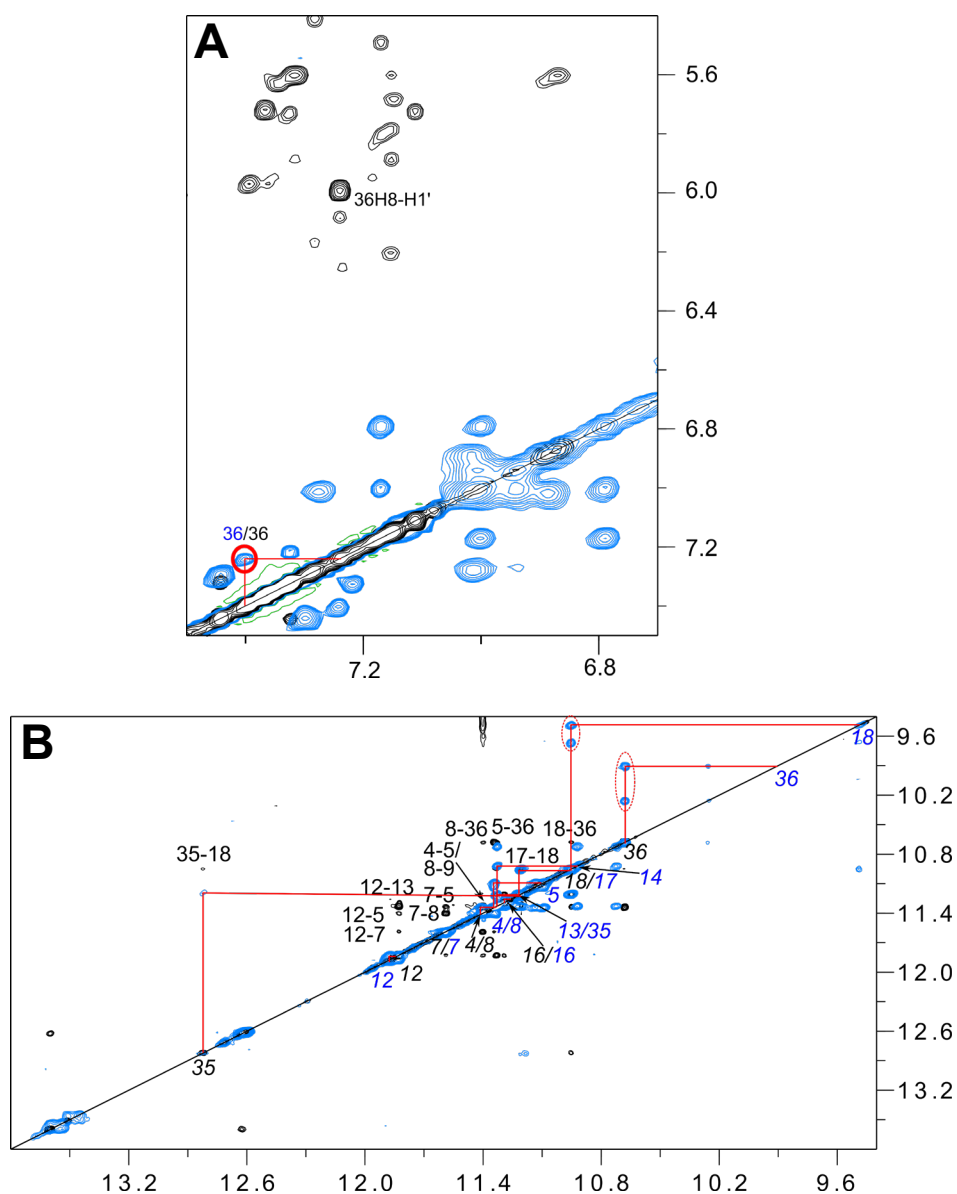


Figure S15. Superimposed NOESY spectral regions of free *QD3-sbl* (300 ms mixing time, black) and corresponding ROESY spectra of *QD3-sbl* in the presence of 0.5 eq. of cryptolepine (blue, only positive contours shown for clarity). (A) H6/8(ω_2)-H1'/H6/H8(ω_1) correlations; G36 H8 of the complex is unambiguously identified through its exchange cross-peak (circled in red) with G36 H8 of the free hybrid. (B) Imino(ω_2)-imino(ω_1) correlations; exchange cross-peaks between free and ligand-bound *QD3-sbl* are traced by red lines with resonances of free and ligand-bound hybrid written in black and blue along the diagonal, respectively; prominent exchange cross-peaks between free and two complexed hybrids are circled in red. Spectra were acquired at 30 °C.

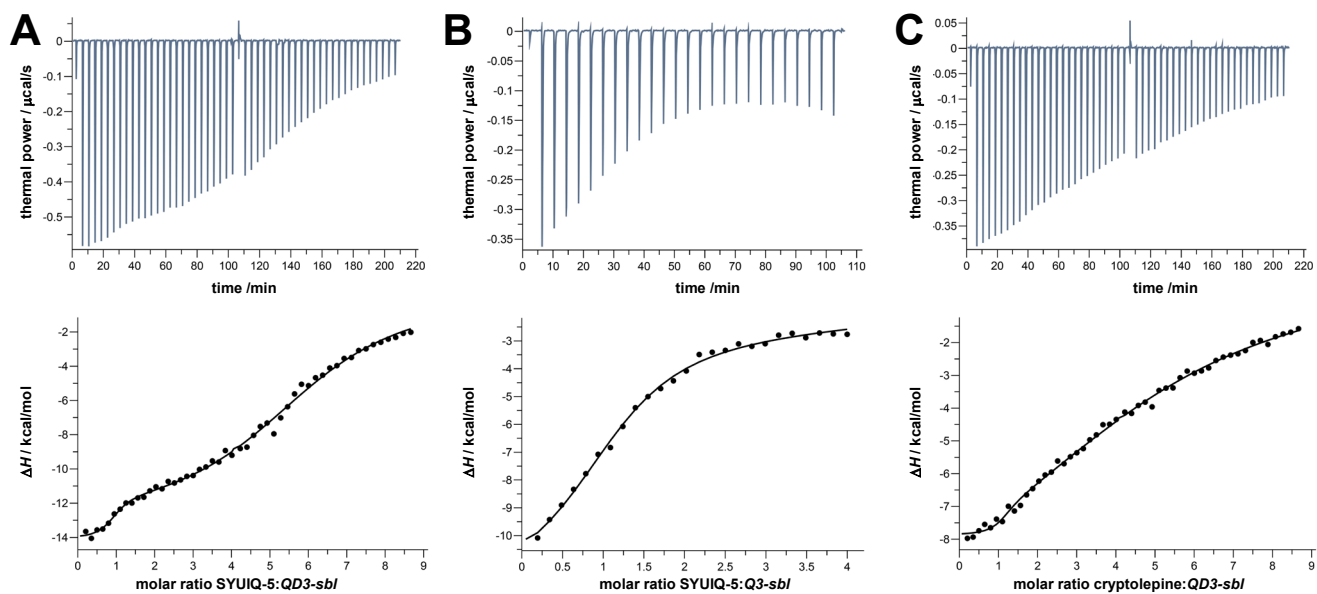


Figure S16. Representative ITC thermograms in 120 mM K^+ buffer at 40 °C for the binding of SYUIQ-5 to *QD3-sbl* (A), SYUIQ-5 to *Q3-sbl* (B), and of cryptolepine to *QD3-sbl* (C). Upper and lower panels show the heat released for each injection step and the blank-corrected normalized heat versus molar ratio, respectively. Curves were fitted with a model of two independent binding sites.

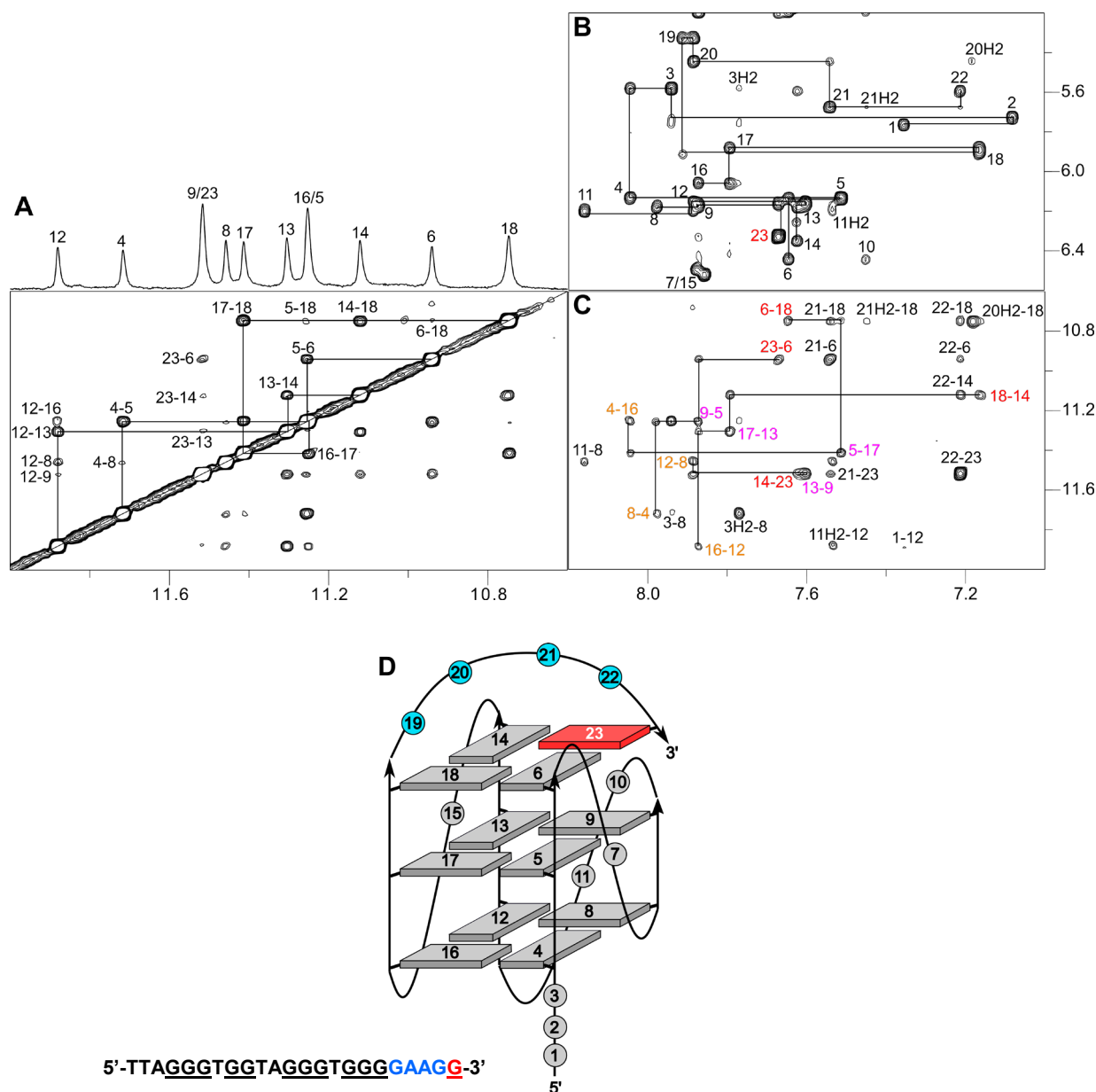


Figure S17. 2D NOESY spectrum (300 ms mixing time, 20 °C) of *Q3-sbl* (0.2 mM). (A) Imino(ω_2)-imino(ω_1) spectral region with corresponding 1D spectrum; resonances are labeled with residue number. (B) H8/H6(ω_2)-H1'(ω_1) spectral region with an intra-nucleotide H8-H1' contact of *syn*-G23 labeled in red. (C) H8(ω_2)-H1(ω_1) spectral region; intra-tetrad H8(ω_2)-H1(ω_1) cross-peaks are labeled with colors depending on G-tetrad layer. Inter-tetrad connectivities along G-columns are traced by horizontal and vertical lines. (D) Topology and sequence of *Q3-sbl*. The sequence folds into a three-layered parallel quadruplex with a 1:2:1 propeller loop and a 3'-terminal snapback loop arrangement. A vacant site in the second G-column is filled by the 3'-terminal *syn*-G residue.

Article 8

High-affinity binding at quadruplex–duplex junctions: rather the rule than the exception

Yoanes Maria Vianney and Klaus Weisz¹*

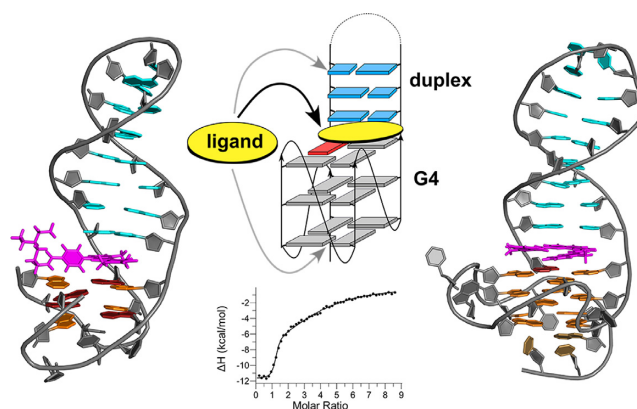
Institute of Biochemistry, Universität Greifswald, Felix-Hausdorff-Str. 4, D-17489 Greifswald, Germany

Received August 31, 2022; Revised October 19, 2022; Editorial Decision October 25, 2022; Accepted October 26, 2022

ABSTRACT

Quadruplex–duplex (Q–D) junctions constitute unique structural motifs in genomic sequences. Through comprehensive calorimetric as well as high-resolution NMR structural studies, Q–D junctions with a hairpin-type snapback loop coaxially stacked onto an outer G-tetrad were identified to be most effective binding sites for various polycyclic quadruplex ligands. The Q–D interface is readily recognized by intercalation of the ligand aromatic core structure between G-tetrad and the neighboring base pair. Based on the thermodynamic and structural data, guidelines for the design of ligands with enhanced selectivity towards a Q–D interface emerge. Whereas intercalation at Q–D junctions mostly outcompete stacking at the quadruplex free outer tetrad or intercalation between duplex base pairs to varying degrees, ligand side chains considerably contribute to the selectivity for a Q–D target over other binding sites. In contrast to common perceptions, an appended side chain that additionally interacts within the duplex minor groove may confer only poor selectivity. Rather, the Q–D selectivity is suggested to benefit from an extension of the side chain towards the exposed part of the G-tetrad at the junction. The presented results will support the design of selective high-affinity binding ligands for targeting Q–D interfaces in medicinal but also technological applications.

GRAPHICAL ABSTRACT



INTRODUCTION

G-rich sequences can fold into non-canonical secondary structures called G-quadruplexes (G4s) (1,2). In these tetra-stranded structures, four guanine bases are positioned in a square planar arrangement and linked by a cyclic array of Hoogsteen hydrogen bonds. In general, G4s comprise a stack of two to four of such G-tetrads with monovalent cations like Na^+ or K^+ coordinated within the central channel of the G-core for additional stabilization. Due to the abundance of G-rich sequences at critical locations within the genome such as in telomeres and oncogenic promoters and also sparked by the observation that G4 formation interferes with cellular processes like telomere maintenance or gene transcription, these alternative nucleic acid structures have been recognized as promising targets for medicinal interventions through selective binding of small drug molecules (3,4). This prompted the design and screening of a plethora of G4 binding ligands, some of them exhibiting a remarkable affinity with dissociation constants in the micromolar and even sub-micromolar range (5,6). On the other hand, the majority of ligands reported to date comprises a polycyclic aromatic ring system and primarily bind through stacking on outer G-tetrads with only weak additional interactions with loop and/or overhang residues (4,7). Consequently, selectivities against other competing G4 topologies but also against genomic B-type duplexes are mostly poor,

*To whom correspondence should be addressed. Tel: +49 3834 420 4426; Fax: +49 3834 420 4427; Email: weisz@uni-greifswald.de

allowing for putative off-target effects in potential therapeutic applications.

In contrast, non-canonical G4 structures with unique structural motifs (8) but especially quadruplex–duplex (Q–D) junctions may provide for a more selective high-affinity targeting through low molecular weight compounds. Additionally, they may also be used as powerful tools to enforce a particular G4 topology or to optimize aptamer affinities (9–13). Q–D junctions are expected to frequently occur within the genome (14–16). They may involve external double-helical overhang sequences or internal duplex stem loops with or without unpaired linker nucleotides between G4 and duplex domains. The orientation of duplex and G4 helices may either be coaxial or orthogonal, depending on the attachment of the duplex to the G4 core. Seminal studies by Phan et al. on the structure and thermodynamics of several engineered Q–D hybrids have revealed continuous stacking at the interface if there is a coaxial orientation of directly linked quadruplex and duplex domains, with different stacking energies depending on the type of base pair at the junction (17,18).

Notably, systematic studies on the ligand binding at Q–D interfaces are still in their infancy. Early reports on the recognition of an RNA Q–D motif involved the positively charged RGG peptide from the human fragile X mental retardation protein (FMRP), binding to the major groove of the duplex domain (19,20). A rational approach for targeting a Q–D motif proposed the design of a two-component hybrid molecule composed of a quadruplex-specific ligand with an extended aromatic surface area and a duplex-specific minor groove binder. Although the simultaneous recognition by both ligand moieties was demonstrated, a more detailed structural characterization has not been provided (21,22). Very recently, the binding of simple aminomethyl-substituted aromatic hydrocarbons as well as indoloquinoline, naphthalene diimide, and pyridostatin derivatives to Q–D junctions resulted in first high-resolution structures of corresponding complexes (23–26). In fact, indoloquinoline-based ligands such as unsubstituted cryptolepine, aminoalkylated SYUIQ-5, and a phenyl-substituted indoloquinoline derivative PIQ-4m recognized the Q–D junction with higher affinity compared to either the free quadruplex or free duplex, suggesting the possibility of a successful ligand design dedicated for selective Q–D targeting (25,27). For SYUIQ-5, the indoloquinoline aromatic ring system intercalates at the Q–D interface of a parallel G4 with a coaxially stacked hairpin-type snap-back loop while the positively charged side chain extends into the minor groove of the duplex (25). A similar binding mode was also reported for the naphthalene diimide derivative (24). Here, the naphthalene diimide plane inserts between an outer G-tetrad of a hybrid G4 and a coaxially stacked lateral duplex stem-loop with the platinum-containing side chain positioned within the duplex minor groove. Likewise, the *N*-methylated indoloquinoline cryptolepine, lacking an additional side chain, preferred to bind through intercalation at the Q–D interface of a G4 with a coaxially stacked duplex domain, yet in a more flexible binding mode with enhanced dynamics (25). This hints at the importance of π - π stacking as a major driving force of Q–D recognition.

To provide a more solid basis for targeting Q–D junctions, we here report on a comprehensive thermodynamic and structural study involving several typical G4 binding ligands (Figure 1A). The combination of comparative calorimetric binding experiments using tailored nucleic acid receptors with NMR structural studies on selected complexes identifies the Q–D junction as primary binding site with the largest association constant for most of the tested ligands. Consequently, the favored recognition of Q–D interfaces constitutes a general rather than a more specific phenomenon. However, discrimination of the Q–D junction against intercalation into a B-type duplex or end stacking on an exposed outer G-tetrad widely varies. Whereas the typical G4 ligand SYUIQ-5 suffers from a poor selectivity in the presence of longer duplex domains, PIQ-4m seems superior in terms of its discriminatory power in favor of the Q–D interface. Based on the collected data, a ligand design for most effective Q–D recognition is proposed, supporting future efforts in targeting G4 structures through existing Q–D motifs.

MATERIALS AND METHODS

Sample preparation

DNA oligonucleotides were purchased from TIBMOBIO (Berlin, Germany) and further purified through ethanol precipitation prior to their use. Concentrations of oligonucleotides were determined by measuring their absorbance A_{260} at 80°C in aqueous solution using extinction coefficients as provided by the manufacturer. PIQ derivatives were prepared as described and their concentration determined spectrophotometrically using a molar extinction coefficient ϵ_{376} of 22 227 M⁻¹·cm⁻¹ in potassium phosphate buffer (28,29). Thiazole orange (TO), SYUIQ-5, BRACO-19 and Phen-DC3 were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). NDI-DM was obtained from ABCR (Karlsruhe, Germany). Thiazole orange concentrations were determined using a molar extinction coefficient for the TO monomer ϵ_{500} of 63 000 M⁻¹·cm⁻¹ in DMSO (30). Concentrations of all other ligands were determined from their weighed mass. All ligands were initially dissolved in a DMSO stock solution except for isothermal titration calorimetry (ITC) measurements. Here, ligands were directly dissolved in the high-salt ITC buffer solution (20 mM potassium phosphate, 100 mM KCl, pH 7.0, supplemented with 5% DMSO). A low-salt buffer with 10 mM potassium phosphate, pH 7.0, was additionally employed for circular dichroism (CD) melting and some of the NMR experiments as described.

UV melting experiments

To confirm complete folding under the ITC experimental conditions, melting of DNA receptors (2–5 μ M depending on duplex or quadruplex melting experiments) was evaluated in the ITC buffer solution with a Jasco V-650 spectrophotometer equipped with a Peltier thermostat. For melting of the duplex and quadruplex domains, temperature-dependent absorbances A_{260} and A_{295} were recorded using a heating rate of 0.2°C·min⁻¹ and a bandwidth of 1 nm. Melting temperatures T_m were determined

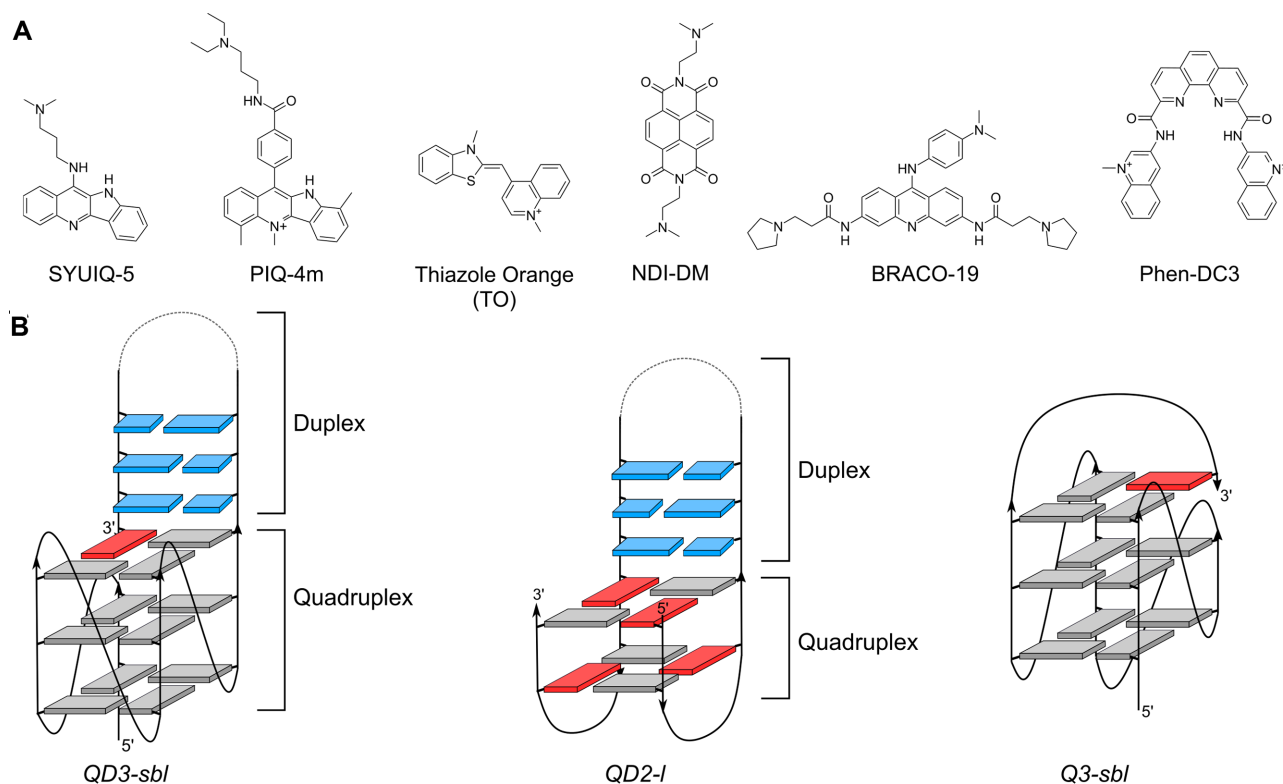


Figure 1. (A) Chemical structure of quadruplex binding ligands. (B) Topologies of quadruplexes with and without a duplex interface; *anti*- and *syn*-guanosines of the G-core are colored grey and red, respectively, Watson–Crick base pairs are colored blue.

in triplicate from the first derivative of the absorbance versus temperature plot.

ITC experiments

ITC experiments were performed at 40°C with a Microcal PEAQ-ITC microcalorimeter using a reference power of 4 $\mu\text{cal}\cdot\text{s}^{-1}$. Oligonucleotides and ligands were dissolved in ITC buffer and the ligand (400 μM) was titrated to the oligonucleotide (20 μM) with a total of 2×26 injections. Titration volumes, duration of injections, and spacing between injection steps were 1.5 μl , 3 s, and 240 s, respectively. The first injection (0.4 μl) was rejected for data analysis through the Microcal PEAQ-ITC analysis software. Generally, data were fitted with a model of two sets of independent binding sites. Additional excess-site titrations were performed for a model-independent determination of binding enthalpies. Here, ligand (3 μl , 200 μM) was titrated in 12 injection steps to an oligonucleotide solution (100 μM) with an injection duration of 6 s and a spacing between injections of 300 s. The first injection (0.4 μl) was discarded for the determination of an average binding enthalpy. All experiments were blank- and concentration-corrected and done in triplicate.

CD spectroscopy

CD spectra were recorded at 30°C with a Jasco J-810 spectropolarimeter equipped with a Peltier thermostat in a high-salt buffer. Ligands in a DMSO stock solution were titrated

to the Q–D hybrid (5 μM) up to a 4:1 ligand-to-DNA molar ratio with final DMSO concentrations $\leq 1.1\%$. Ellipticities were recorded from 230 nm up to 600 nm for the TO ligand using a bandwidth of 1 nm, a scanning speed of 50 $\text{nm}\cdot\text{min}^{-1}$, a response time of 4 s, and five accumulations. All spectra were blank-corrected. For CD melting experiments, the QD2-l hybrid was dissolved in 10 mM potassium phosphate buffer, pH 7.0 (2 ml, 5 μM). For melting of the complex, a concentrated ligand solution in DMSO was titrated to give a 1:1 molar ratio. Ellipticities at 295 nm were recorded from 15°C to 95°C with a heating rate of 0.2°C·min⁻¹ and a bandwidth of 1 nm. Melting temperatures were determined by the first derivative of the melting curve and averaged over two independent experiments.

NMR spectroscopy

NMR spectra were acquired on a Bruker Avance NEO 600 MHz spectrometer equipped with an inverse ¹H/¹³C/¹⁵N/¹⁹F quadruple resonance cryoprobehead and z-field gradients. Spectra were processed in TopSpin 4.0.7 and assigned in CcpNMR V2 (31). Oligonucleotides were dissolved in either a low-salt or high-salt buffer with a 90% H₂O/10% D₂O solvent system. Ligands were added in a DMSO-d₆ stock solution with a final DMSO concentration for a 1:1 mixture of $\leq 4\%$. Proton chemical shifts were referenced to sodium trimethylsilylpropionate (TSP) through the temperature-dependent water chemical shift at pH 7 while ¹³C chemical shifts were referenced to sodium trimethylsilylpropanesulfonate (DSS) through indirect

referencing. For further details on NMR experimental parameters see the Supplementary Information.

Structure calculations

Structures of complexes between *QD3-sbl* and Phen-DC3 as well as between *QD2-l* and PIQ-4m were initially generated with a simulated annealing protocol using XPLOR-NIH 3.0.3 (32). Experimental restraints employed for the calculations included distances as derived from NOESY crosspeak intensities, glycosidic torsion angles χ , sugar puckers from DQF-COSY crosspeak patterns, as well as hydrogen bond and planarity restraints (see the Supplementary Information for more details). Additional chirality restraints were imposed in calculations of the *QD3-sbl*–Phen-DC3 complex. In the following, 100 out of 400 calculated structures were selected for a structural refinement using AMBER18 with the parmbsc0 force field and OL15 modifications according to a protocol described recently (25). In short, having established a force field for the ligand, starting structures were subjected to simulated annealing with experimental and planarity restraint energies to yield 20 lowest-energy structures. For a refinement in explicit water, the DNA was initially neutralized and potassium ions placed within the inner core of the G-quadruplex flanked by the tetrad layers. A final simulation was done at 1 atm and 300 K for 4 ns using only NOE- and hydrogen bond-based distance restraints. The trajectory was averaged for the last 500 ps. It should be mentioned that this averaging process resulted in noticeable distortions of the highly flexible PIQ-4m aliphatic side chain but was eliminated through final minimizations, yielding ten lowest-energy structures. Pymol 2.3.2 was used for visualization and the extraction of conformational parameters.

RESULTS AND DISCUSSION

Ligands and DNA receptors

For getting a deeper insight into the binding affinity and binding mode of G4 ligands when associating with quadruplexes featuring a Q–D junction, a set of six ligands comprising different polycyclic aromatic core structures with up to three side chains were selected for biophysical studies (Figure 1A). With their flat aromatic heterocyclic ring systems prone to π – π stacking interactions, indoloquinolines SYUIQ-5 (33,34) and PIQ-4m (28,29), naphthalene diimide NDI-DM (35,36), trisubstituted acridine BRACO-19 (37,38), and phenanthroline derivative Phen-DC3 (39,40) are all considered typical G4 binding ligands with a preference to stack on exposed outer G-tetrads but also with varying propensities to serve as duplex intercalators. Thus, whereas Phen-DC3 exhibits good selectivity for G4s in comparison with duplexes but shows only poor selectivity against different G4 topologies, BRACO-19 was reported to also show significant binding to duplex structures (41). On the other hand, thiazole orange (TO), frequently applied as fluorescent probe in nucleic acid sensing (42), is regarded a more universal DNA intercalator with a rather poor discriminatory ability for nucleic acid secondary structures.

To probe binding of the ligands at Q–D junctions, the *QD3-sbl* hybrid was employed as DNA receptor. It com-

Table 1. DNA sequences used in the present study; G residues in tetrads are underlined

Name	Sequence (5'-3')
<i>QD3-sbl</i>	TTAGGTGGGTAGGGTGGG-CTAGTCATTTGACTAG-G
<i>Q3-sbl</i>	TTAGGGTGGTAGGGTGGG-GAAG-G
<i>D3-HP</i>	CTAGTCATTTGACTAG
<i>QD2-l</i>	GGTTGG-CGCGAAGCATTGCGG-GGTTGG
<i>D2-HP</i>	CGCGAAGCATTGCGG
<i>QD2-l-2bp</i>	GGTTGG-CGGCACG-GGTTGG
<i>TBA</i>	GGTTGG-TGT-GGTTGG
<i>Q3-sbl2</i>	TTAGGTGGGTAGGGTGGG-TGT-G

prises a three-layered parallel G4 and a 3'-duplex stem loop that is fixed to the interfacial tetrad by the terminal G, filling a vacant tetrad position along the first G-column (Figure 1B, Table 1) (25). Such an architecture showcases different putative binding sites, i.e. the Q–D junction, the exposed 5'-outer tetrad, and the duplex domain. While *QD3-sbl* allows competition among different binding sites, additional truncated constructs were designed to facilitate separation of binding processes in thermodynamic studies. Thus, *Q3-sbl* preserves the 5'-outer tetrad of the *QD3-sbl* hybrid but lacks the coaxially stacked double-helical 3'-extension. Instead, the 3'-outer tetrad is bridged by a diagonal snap-back loop known to effectively block ligand binding to only feature the 5'-tetrad as a putative high-affinity binding site (40,43,44). Likewise, only employing the duplex hairpin of *QD3-sbl* termed *D3-HP* allows exclusive extraction of binding parameters for the double-helical domain.

In addition to the Q–D hybrid based on a parallel G4 topology, another Q–D hybrid *QD2-l* is based on a two-layered chair-type antiparallel quadruplex originally designed by Phan (Figure 1B, Table 1) (17). Here, the second lateral loop of the thrombin binding aptamer *TBA* (45) spanning the G4 wide groove has been replaced by a duplex stem loop. Because the tetrad opposite the Q–D junction is occluded by two TT lateral loops, only binding at the junction or within the duplex domain is expected to compete with each other in this case. Finally, cutting the length of the duplex stem loop in *QD2-l-2bp* or replacing it by a non-base-paired TGT lateral loop as for parent *TBA* or *Q3-sbl2* is expected to give additional information on the structural requirements for ligand binding at Q–D interfaces.

Calorimetry points to the quadruplex–duplex interface of *QD3-sbl* as a binding hotspot for the tested ligands

Although several studies in the past years have reported association constants for some of the ligands upon binding G4 structures, a reasonable comparison of data is often impossible due to the use of different temperatures, buffer conditions, and/or G4 sequences (46–49). Isothermal titration calorimetry (ITC) was therefore employed for extracting information on potential ligand binding sites and for a direct comparison of detailed thermodynamic profiles. To match previously reported conditions (25,27), titrations were performed at a temperature of 40°C in a buffer solution with 120 mM K⁺ ions close to a physiological environment. This high-salt and high-temperature setup was also expected to repress unspecific electrostatic interactions upon binding the cationic ligands. The buffer was supplemented with 5%

DMSO to overcome solubility limitations for some of the compounds. However, even in the presence of the DMSO additive, Phen-DC3 was not amenable to calorimetric measurements due to persisting solubility problems even when trying to set up a reverse titration experiment with the ligand used in a lower-concentrated titrand solution.

Initially, ligands were titrated to the *QD3-sbl* hybrid. More complex binding equilibria can be expected given three different putative binding sites competing for the ligand. ITC curve fitting routines with three or even more sets of binding sites have been employed in the past but in many cases suffer from the large number of free-floating and often interdependent parameters (50). Here, the curve fitting routine was based on a maximum of only two sets of independent binding sites to possibly result in ambiguities of binding parameters in some cases. Also, because lower-affinity binding sites are often less well defined, only parameters for high-affinity binding are discussed in the following (for a compilation of all fit parameters see Supplementary Table S1). Being confronted with these shortcomings when attempting to separate different binding processes, additional titrations were performed on the quadruplex *Q3-sbl* and the hairpin duplex *D3-HP*, expected to closely mimic the exposed tetrad and hairpin duplex of the *QD3-sbl* hybrid, respectively. Before measurements, all receptors including the free duplex hairpin were shown by UV melting to be folded under the ITC experimental conditions (see Supplementary Table S2).

There is a wealth of information gained from the ITC titrations of the ligands to each of the *QD3-sbl*, *Q3-sbl*, and *D3-HP* receptors (Figure 2). Results can be summarized as follows:

- (i) Except for NDI-DM but in particular for BRACO-19 with their comparable or supposedly even higher affinity towards *Q3-sbl*, ligands bind with highest affinity to the *QD3-sbl* hybrid, indicating the Q–D junction to be the favored binding site. Association constants at 40°C vary in a narrow range between $1.1 \cdot 10^7$ and $1.4 \cdot 10^7$ M⁻¹ for PIQ-4m, SYUIQ-5 and BRACO-19 but decrease by one and two orders of magnitude for TO and NDI-DM, respectively.
- (ii) Selectivities towards a particular binding site vary significantly among the ligands. Thus, the binding constant of the indoloquinoline PIQ-4m decreases by a factor of four when going from the *QD3-sbl* hybrid to the *Q3-sbl* parallel G4 and by three orders of magnitude when binding the corresponding free hairpin duplex *D3-HP*. Considering experimental uncertainties, selectivities between G4 hybrid and parallel G4, i.e. between a Q–D junction and an exposed outer tetrad, follow the order PIQ-4m ~ SYUIQ-5 ~ TO > NDI-DM ~ BRACO-19. On the other hand, selectivities of the hybrid against the short hairpin duplex are given by PIQ-4m >> TO > SYUIQ-5 ~ BRACO-19 > NDI-DM. Consequently, PIQ-4m seems superior in differentiating the junction against additional binding sites at both G4 and duplex domains whereas NDI-DM with its two side chains fails to be a selective G4 binder in the absence but also the presence of a Q–D junction. Interestingly, TO, regarded as universal DNA ligand,

exhibits significant discriminating potential with clear preference of the Q–D junction, albeit with only moderate affinity.

- (iii) Given a much better quadruplex–duplex selectivity of the phenyl-substituted PIQ-4m when compared to SYUIQ-5 with its aliphatic side chain, it is conspicuous that G4 binding is considerably more exothermic for SYUIQ-5 irrespective of the DNA receptor (Supplementary Table S1). This is compensated by a much smaller entropic penalty for G4 binding of PIQ-4m. These contrasting profiles point to different interactions of the indoloquinoline side chains in PIQ-4m and SYUIQ-5 with more specific contacts of the flexible aliphatic side chain in SYUIQ-5 within a duplex or quadruplex groove.
- (iv) Whereas high-affinity binding to *QD3-sbl* is associated with a 1:1 stoichiometry for the two indoloquinolines, binding of TO as a dimer is indicated by the determined stoichiometries. However, stoichiometries of 3 and 0.5 for high-affinity binding to *QD3-sbl* in case of NDI-DM and BRACO-19 likely reflect the inability of the binding model to accurately fit and separate competing binding processes of similar affinities at initial titration steps.

High-affinity binding depends on an interfacial base pair

Additional binding studies on a hybrid *QD2-l* featuring a lateral duplex stem loop in a two-layered antiparallel G4 were performed with PIQ-4m as being the most selective Q–D binder (Figure 3). Here, a high affinity with K_a close to 10^7 M⁻¹ slightly lower compared to binding to the *QD3-sbl* hybrid was determined. However, there is a steeper rise in the thermogram after the addition of one equivalent of ligand and the saturation of the first binding site. Because access to the opposite outer tetrad should be effectively blocked through the two TT lateral loops in *QD2-l*, such a behavior is expected if only the duplex domain is left as competing binding site. Again, the latter shows rather weak binding as demonstrated by the ITC titration of the free hairpin duplex *D2-HP* with association constants lower by more than two orders of magnitude when compared to the high-affinity binding for the hybrid receptor.

Next, the duplex domain was truncated to a minimal duplex stem loop comprising only two potential base pairs. With the formation of a stable interfacial base pair as demonstrated by NMR (Supplementary Figure S1), the strength of high-affinity PIQ-4m binding was largely conserved (Figure 3). Clearly, a faster return to baseline at later titration steps results from the reduced number of low-affinity binding sites within the duplex domain. Finally, the parent TBA quadruplex with a non-duplex TGT lateral loop replacing the duplex stem loop was tested as receptor for PIQ-4m binding. Here, binding was very weak and did not even exceed affinities for a free duplex, demonstrating that only an intact base pair at the interface seems to allow strong interactions of the ligand at the Q–D junction.

As mentioned above, no corresponding ITC data could be extracted for Phen-DC3 owing to its poor solubility in the used buffer. To nevertheless obtain an estimate of its binding affinity towards a G4 hybrid, CD melting studies on

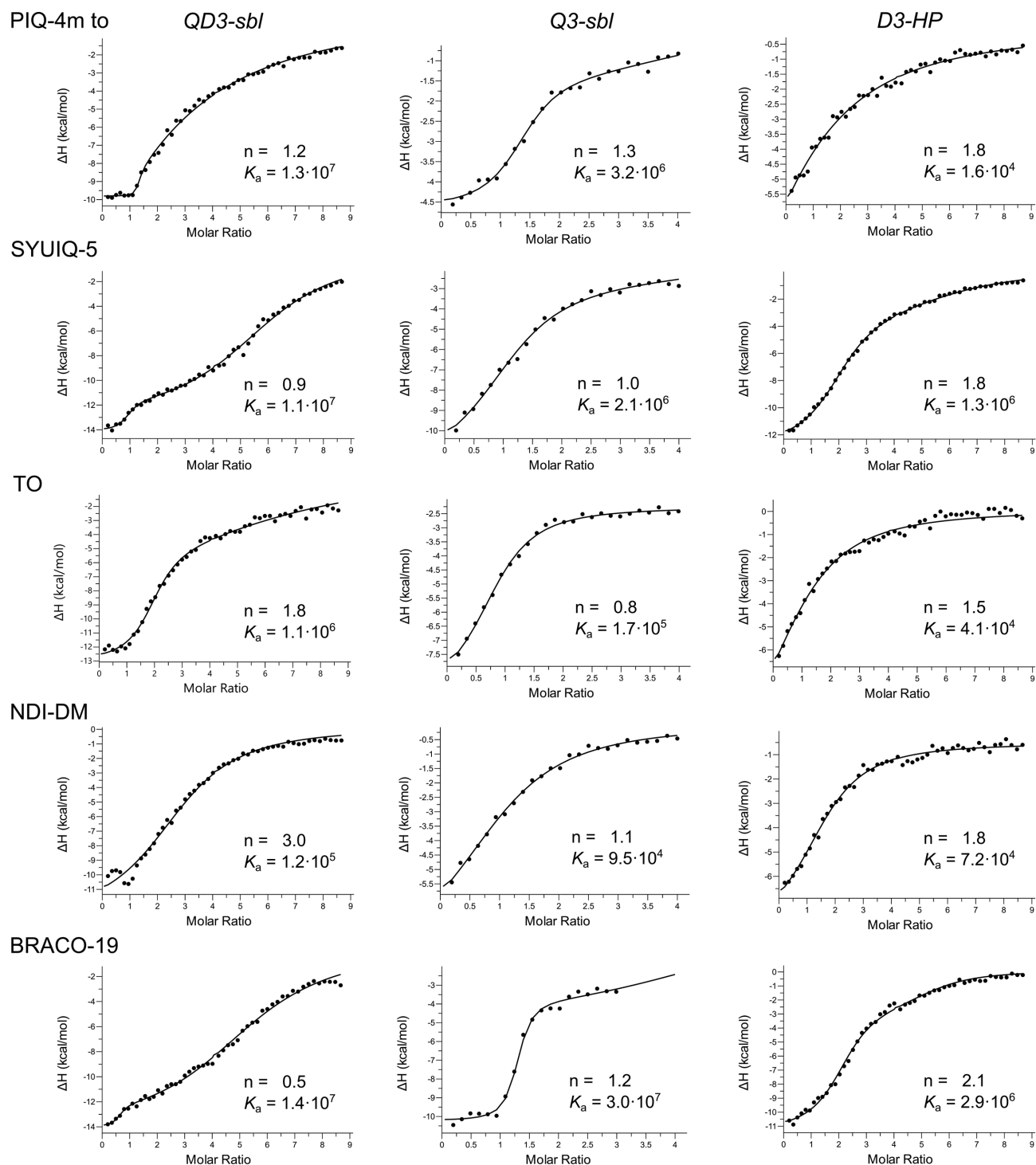


Figure 2. ITC thermograms for G4 ligands titrated to *QD3-sbl*, *Q3-sbl* and *D3-HP* at 40°C in a 120 mM K^+ buffer; stoichiometries n and affinity constants K_a (in M^{-1}) as determined by curve fitting are indicated. For a compilation of all fit parameters with root-mean-square deviations from three independent experiments see Supplementary Table S1.

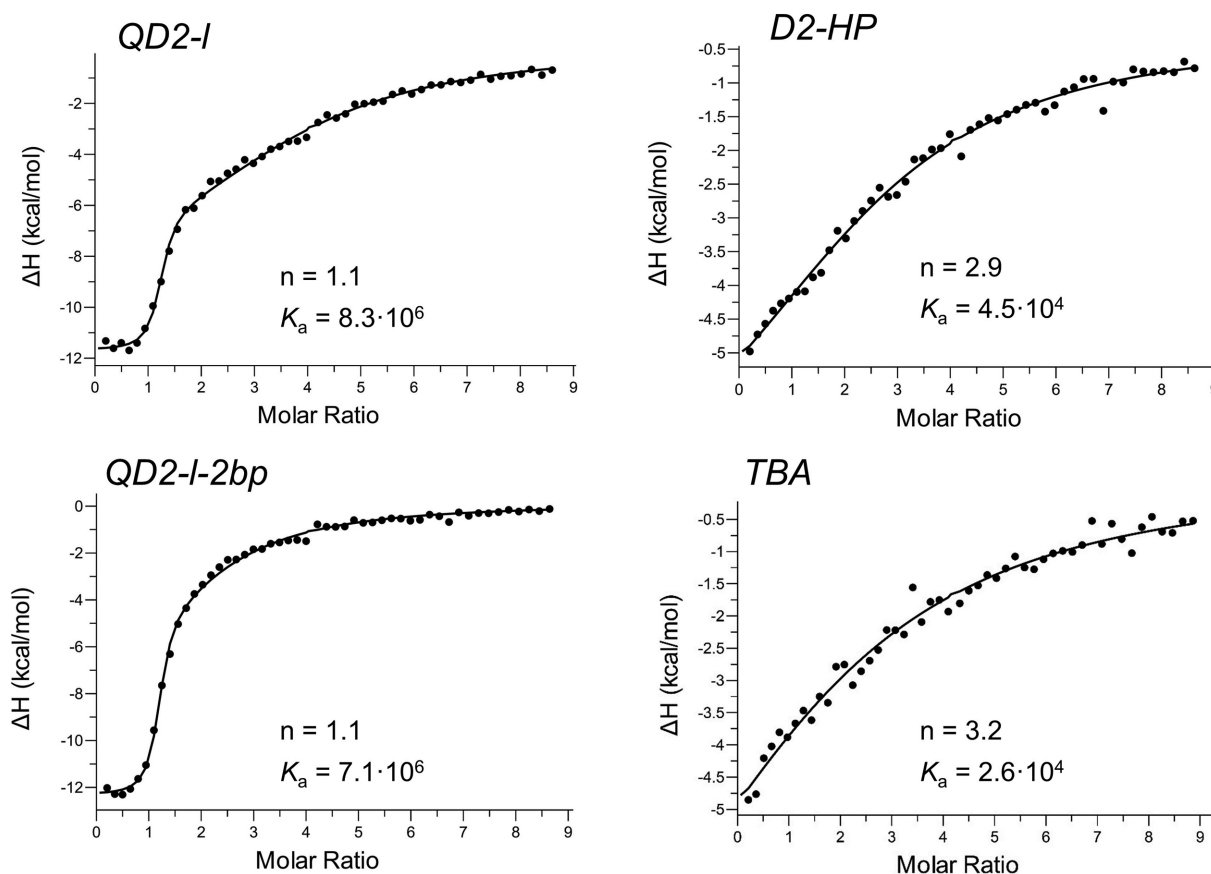


Figure 3. ITC thermograms for the PIQ-4m ligand titrated to *QD2-I*, *QD2-I-2bp*, *D2-HP* and *TBA* at 40°C in a 120 mM K⁺ buffer; stoichiometries n and affinity constants K_a (in M⁻¹) as determined by curve fitting are indicated. For a compilation of all fit parameters with root-mean-square deviations from three independent experiments see Supplementary Table S3.

1:1 mixtures of the hybrid receptor with the ligand were performed and compared to indoloquinolines SYUIQ-5 and PIQ-4m. Because of limitations set by the high melting in case of *QD3-sbl*, the *QD2-I* receptor with a melting temperature T_m of 50.3°C was employed under low-salt conditions, providing for a convenient temperature window for all melting experiments (Supplementary Figure S2). Thus, 1:1 complexes with SYUIQ-5 and PIQ-4m yielded a T_m increase by ~5°C and ~13°C, respectively. The noticeably smaller thermal stability of the SYUIQ-5 compared with the PIQ-4m complex may mostly be derived from the more enthalpy-driven SYUIQ-5 association given the general temperature-dependence of association. Remarkably, with a T_m of 74°C binding of 1 equivalent of Phen-DC3 resulted in a considerable increase of the *QD2-I* melting temperature by almost 24°C. Consequently, Phen-DC3 seems to surpass all other tested ligands in terms of binding affinity towards the hybrid structure, assuming the melting temperature to be a reasonable measure of the binding free energy.

Tracking high-affinity binding sites through NMR experiments

Whereas indoloquinoline ligands and a naphthalene diimide derivative were recently shown to bind at Q–D junctions (24,25,27), there is no structural information on fa-

vored binding sites of the other G4 ligands when targeting the Q–D hybrid. Clearly, different affinities but also exothermicities obtained for initial binding events on the *QD3-sbl* hybrid when compared to the *Q3-sbl* G4 or *D3-HP* duplex fragment suggest that the Q–D junction mostly outcompetes outer G-tetrad stacking or duplex binding. To further validate this conclusion and to exclude putative interfering phenomena such as significant cooperativity effects between binding sites and/or ligand-induced G4 un/refolding, additional structural studies were performed.

Initially, ligand titration to *QD3-sbl* was followed by CD spectroscopy. In addition to induced CD (ICD) effects at the ligand absorption wavelength, only small to moderate changes in the CD amplitudes of minima and maxima at about 240 and 265 nm were observed. This suggests no major G4 conformational rearrangements upon ligand binding (Supplementary Figure S3). ICD effects show ligand-specific shapes and intensities. Although the sign of the ICD is affected by the specific orientation of the bound ligand through its transition dipole moment, ICDs are notoriously difficult to interpret in terms of binding mode or even of a defined binding geometry. A weak negative ICD as observed throughout the titration for NDI-DM or Phen-DC3 has often been associated with duplex intercalation but also tetrad stacking (51,52). On the other hand, bisignate ICDs as seen for most other ligands when added in excess may

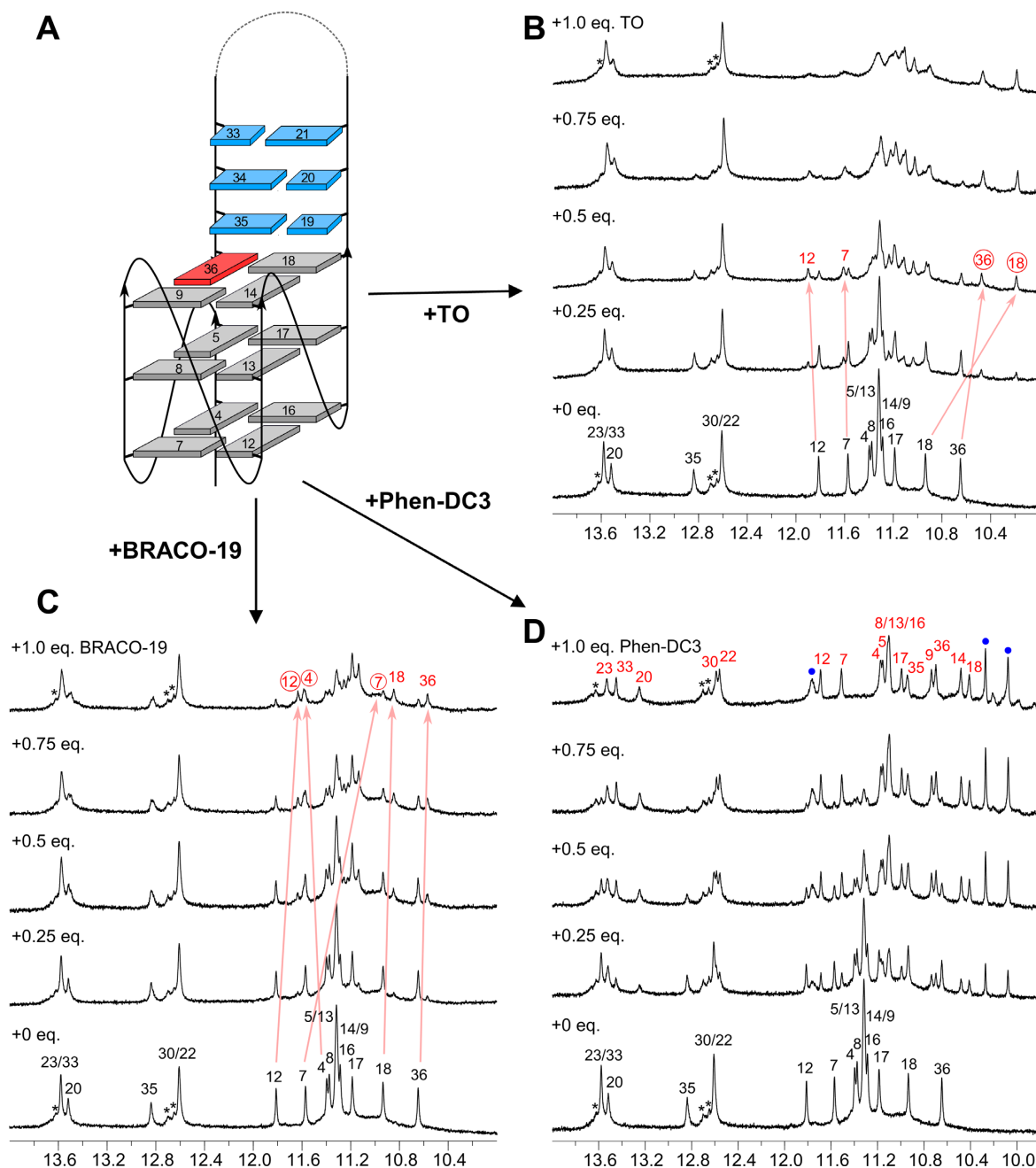


Figure 4. Topology with residue numbers of the *QD3-sbl* hybrid (A) and its titration with TO (B), BRACO-19 (C), and Phen-DC3 (D). Spectra show Watson–Crick and G Hoogsteen imino resonances with full and partial assignments of the free *QD3-sbl* (black) and the *QD3-sbl*–ligand complex (red). Phen-DC3 protons in (D) are marked by blue dots. Representative chemical shift changes of assigned resonances upon TO and BRACO-19 addition are traced by arrows with resonances experiencing more significant shifts circled. Asterisks indicate resonances in the duplex region of a minor species. Spectra were acquired in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0, at 30°C with a *QD3-sbl* concentration of 0.5 mM.

either be attributed to exciton couplings of two or more ligands bound in close proximity or to different binding sites with CD signals of opposite sign and shift. Notably, TO already starts to develop a bisignate ICD before exceeding a 1:1 molar ratio.

In the following, preferred binding sites were probed by NMR titrations of the ligands to *QD3-sbl*. Analysis of 1D and 2D NOESY experiments on the latter at 120 mM K^+ confirmed its folding into a hybrid structure with the du-

plex domain coaxially stacked onto the quadruplex 3'-outer tetrad in analogy to the conformation recently reported for *QD3-sbl* under low-salt conditions (PDB ID: 7PNE) (25). For full resonance assignments and a compilation of chemical shift data see Supplementary Figure S4 and Supplementary Table S4. It should be noted that *QD3-sbl* exhibits additional non-identified signals of very low intensity within the Watson–Crick imino proton spectral region (Figure 4). A putative coexisting minor species may result from different

arrangements of the TTT loop of the duplex hairpin closed by a neighboring AT base pair. In fact, with the hairpin loop remote from the G-core there is no observable impact on the G4 domain nor any shift or change of these weak resonances when binding the ligands (see below).

Adding thiazole orange to *QD3-sbl*, a new set of G imino signals emerges, reaching equal intensity to corresponding resonances of the free hybrid at an 0.5:1 molar ratio (Figure 4B). Clearly, such a behavior suggests binding of the ligand at a distinct binding site. However, imino resonances start to broaden after addition of > 0.5 equivalents of TO in line with the occupation of additional binding sites to form different complexes. Although linebroadening effects associated with the presence of more than a single complex prevented a more detailed structural evaluation of a 1:1 ligand-DNA solution, ROESY experiments on the 0.5:1 mixture revealed various imino proton exchange crosspeaks between free and complexed *QD3-sbl* (Supplementary Figure S5). Based on prior resonance assignments for the free Q-D hybrid, these exchange crosspeaks allowed for the unambiguous assignment of some imino protons in the formed complex. Conspicuously, most significant chemical shift changes were detected for G iminos within the 3'-tetrad such as G18 and G36 located at the Q-D interface. On the other hand, only minor shifts were observed for imino resonances at the 5'-tetrad. These results identify the Q-D junction as hotspot for initial TO binding.

Likewise, addition of BRACO-19 to *QD3-sbl* resulted in the appearance of new G imino resonances, increasing in intensity up to a 1:1 molar ratio (Figure 4C). ROESY experiments on the latter mixture showed various imino exchange crosspeaks, enabling partial imino resonance assignments for a major complex in equilibrium with the free Q-D hybrid (Supplementary Figure S6). Chemical shift perturbations differ significantly from those found upon TO binding with most shifted imino resonances located within the fully exposed 5'-tetrad and hardly affected G iminos at the Q-D interface. However, BRACO-19 seems to display a more complex binding behavior towards *QD3-sbl*, evading a simple description based on a 1:1 high-affinity binding at a single site. Thus, assuming strong binding it is puzzling to observe remaining signals attributable to the free hybrid even after the addition of one equivalent of ligand. Possible reasons such as cooperativity effects in binding or additional groove binding with minimal impact on G-tetrad or duplex imino protons could not be explored in more detail due to limitations in the NMR spectral analysis set by the heterogenous mixture with severe signal overlap. Importantly, however, in the absence of characteristic chemical shift perturbations a preferred binding and intercalation of BRACO-19 at the Q-D junction must be excluded as already suggested by the ITC data.

Titration of Phen-DC3 to *QD3-sbl* resulted in the appearance and gradual increase of a new set of signals for a ligand-DNA complex in slow exchange with the free G4 hybrid (Figure 4D). Also, after the addition of one equivalent of ligand only a single species with well resolved imino signals emerged. Spectra of the 1:1 complex attest to the strong binding of the ligand at a specific binding site on the G4 hybrid, amenable to a more detailed structural analysis.

NMR high-resolution structure of the Phen-DC3 complex

When changing the high-salt buffer (120 mM K⁺) to a low-salt buffer (10 mM K⁺), proton resonances of the *QD3-sbl*-Phen-DC3 1:1 complex experienced some minor shifts whereas crosspeak patterns in the 2D NMR spectra remained virtually unaffected. Therefore, a following spectral analysis was performed on 2D NMR spectra acquired in a low-salt buffer to benefit from a better spectral quality. Established strategies making use of ¹H-¹³C HSQC, DQF-COSY and NOESY spectra with different mixing times allowed for complete resonance assignments of the *QD3-sbl*-Phen-DC3 complex (for details see the Supplementary Information, Figures S7-S9, and Tables S5 and S6).

In general, only small chemical shift changes were observed for the Watson-Crick imino protons upon ligand addition. The interfacial G35 imino represents a striking exception, being significantly upfield-shifted in the complex by nearly 2 ppm to resonate within the Hoogsteen imino proton spectral region (Supplementary Figure S10). For the G-core, most noticeable upfield shifts were experienced by G iminos within the 3'-tetrad at the Q-D interface except for G36 filling the vacant position and showing only small chemical shift perturbations. Overall, a total of 63 intermolecular NOE contacts were identified between ligand protons and various exchangeable, aromatic, and anomeric DNA protons (Supplementary Figures S7 and S9). Contacts were observed between protons in the phenanthroline moiety and protons of G36 and G18 residues of the 3'-outer G-tetrad as well as C19 and G35 of the interfacial base pair. This clearly points to intercalative binding of the phenanthroline at the junction. Additional contacts position the quinoline side arms above the exposed part of the 3'-outer G-quartet to cover the whole tetrad face. Apparently, such a binding mode at the junction deviates from a previously proposed complex structure with Phen-DC3 stacked onto the freely accessible tetrad opposite to the junction of a corresponding Q-D hybrid (21). However, some non-assigned intermolecular contacts of resonances to protons in the 5'-tetrad observed here under high-salt conditions may point to some competing ligand binding, albeit with much lower affinity (not shown).

Structure calculations employing NMR-derived distance and torsion angle restraints yielded a well-defined structural ensemble (for structural statistics see Table 2). The ligand intercalates at the Q-D interface with the phenanthroline located below the C19-G35 interfacial base pair and the two quinoline side arms covering the two exposed G bases of the 3'-tetrad at the junction (Figure 5). There are significant stacking interactions of the ligand with the GC base pair of the duplex domain and guanines of the 3'-tetrad. Based on the ligand orientation, the negligible shift of the *syn*-G36 imino resonance when compared to the other significantly upfield-shifted G iminos within the 3'-tetrad can be attributed to pronounced stacking interactions with G35 in the free hybrid (PDB: 7PNE) combined with its location below the Phen-DC3 amide linkage, associated with less shielding ring current effects in the complex (Figure 5C). On the other hand, there are no obvious electrostatic interactions between the positively charged *N*-methylated quino-

Table 2. NMR restraints and structural statistics for calculated structures

Restraints	<i>QD3-sbl</i> -Phen-DC3	<i>QD2-l</i> -PIQ-4m
NOE distance restraints		
intraresidual	150	114
interresidual	231	175
exchangeable	75	41
intramolecular ligand	0	6
intermolecular	63	41
ligand-DNA		
other restraints		
hydrogen bonds	82	64
dihedral angles	68	53
planarity	3	2
structural statistics after refinement		
pairwise heavy atom RMSD value (Å)		
all residues	2.21 ± 0.39	1.37 ± 0.47
G-tetrad core	0.98 ± 0.2	0.44 ± 0.1
Q-D interface with ligand	0.66 ± 0.15	0.76 ± 0.23
NOE violations		
number of NOE violation (>0.2 Å)	0.4 ± 0.5	0
maximum violation	0.263 Å	0.09 Å
mean NOE violation	0.002 ± 0.0005	0.001 ± 0.0003
deviations from idealized geometry		
bond lengths (Å)	0.01 ± 0.0001	0.01 ± 0.0001
bond angles (degree)	2.25 ± 0.03	2.34 ± 0.02

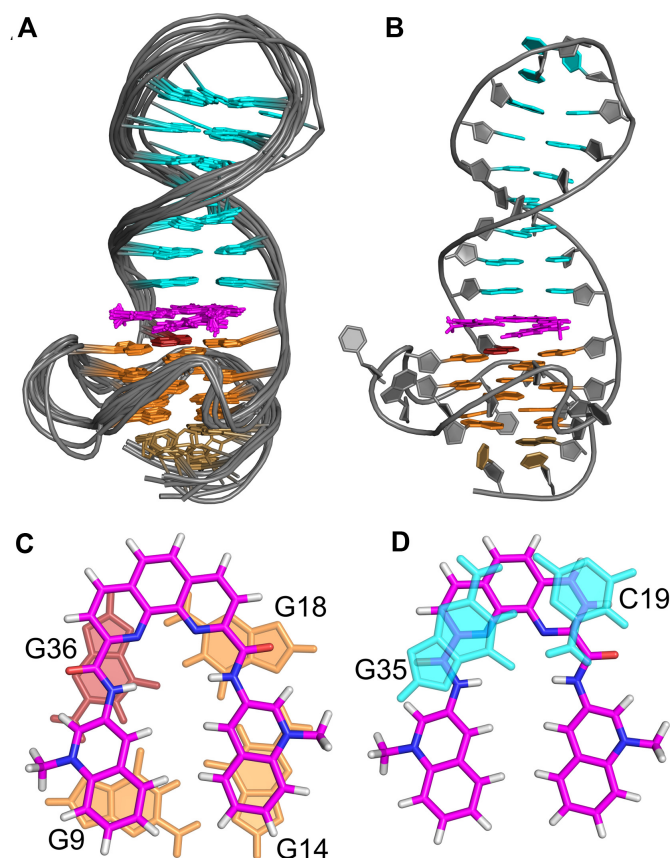


Figure 5. (A) Superposition of ten lowest-energy structures and (B) representative structure of *QD3-sbl* complexed with Phen-DC3. (C) Stacking of Phen-DC3 onto the 3'-tetrad and (D) below the C19-G35 interfacial base pair. In (A), residues of the propeller and duplex hairpin loop have been omitted for clarity.

lines and the DNA hybrid, indicating stacking interactions to be a major driving force for the observed binding mode.

Phen-DC3 changes binding sites upon removal of the duplex stem loop

The *Q3-sbl* sequence employed for the ITC experiments was designed for effectively blocking ligand access to the 3'-tetrad. However, keeping the lateral snapback loop of the original *QD3-sbl* design to only replace the duplex stem loop by a 3-nucleotide non-base-paired TGT lateral loop as for the *TBA* G4 will give additional insight into the importance of regular Q-D junctions for selective Phen-DC3 binding. Folding of the corresponding sequence *Q3-sbl2* was again confirmed by conventional strategies for the assignment of quadruplex structures (for details see the Supporting Information, Supplementary Figure S11 and Supplementary Table S7). Thus, NOESY experiments supported by ^1H - ^{13}C HSQC spectra demonstrated folding of *Q3-sbl2* into a parallel three-layered G4 with exclusive homopolar tetrad stackings and a first broken G-column. In analogy to the *QD3-sbl* hybrid structure, the empty tetrad position as a consequence of the truncated first G-tract is filled by the 3'-G of the snapback loop in a *syn* conformation. The TGT lateral snapback loop seems to effectively cap the 3'-tetrad as suggested by the underlying G22 amino proton shown to be protected from fast solvent exchange.

Initially, titration of Phen-DC3 to *Q3-sbl2* was followed by CD spectroscopy (Supplementary Figure S12A). In analogy to its addition to the *QD3-sbl* hybrid, the parallel CD signature of the G4 is preserved even with ligand in excess and a negative ICD at the ligand absorption develops, albeit only with > 1 equivalent of added ligand. In the following, NMR spectra acquired upon the addition of Phen-DC3 to *Q3-sbl2* showed the emergence of a new set of Hoogsteen imino resonances, fully replacing imino signals of the free G4 at a 1:1 molar ratio (Figure 6A). Full assignment of the 1:1 *Q3-sbl2*-Phen-DC3 complex reveals that most significant chemical shift changes are clearly experienced by imino protons of the 5'-tetrad with iminos of the 3'-tetrad least affected (Figure 6C). Also, NOE contacts involving snapback loop residues are conserved upon ligand binding (Supplementary Figure S12). These data demonstrate that the short non-base-paired lateral snapback loop effectively prevents stacking of Phen-DC3 onto the 3'-tetrad but rather redirects the ligand to the exposed 5'-tetrad at the opposite G4 face. Unlike Phen-DC3 binding to the *QD3-sbl* hybrid but in analogy to a complex with Phen-DC3 stacked on an outer tetrad of a parallel G4 (40), exchange crosspeaks were observed for pairs of protons in the symmetry-related quinoline units of bound Phen-DC3 (Supplementary Figure S13). Such an exchange of the chemical environment for the two quinoline side arms requires the stacked ligand to flip and a corresponding flipping motion is expected to be significantly restricted upon Phen-DC3 intercalation at a Q-D junction in line with present observations.

Taken together, Phen-DC3 known to be a universal G4 binder is shown for the first time to favor a Q-D junction with a coaxially stacked base pair at the interface as most affine binding site over stacking on an exposed outer tetrad. An intercalative binding mode at the junction requires the

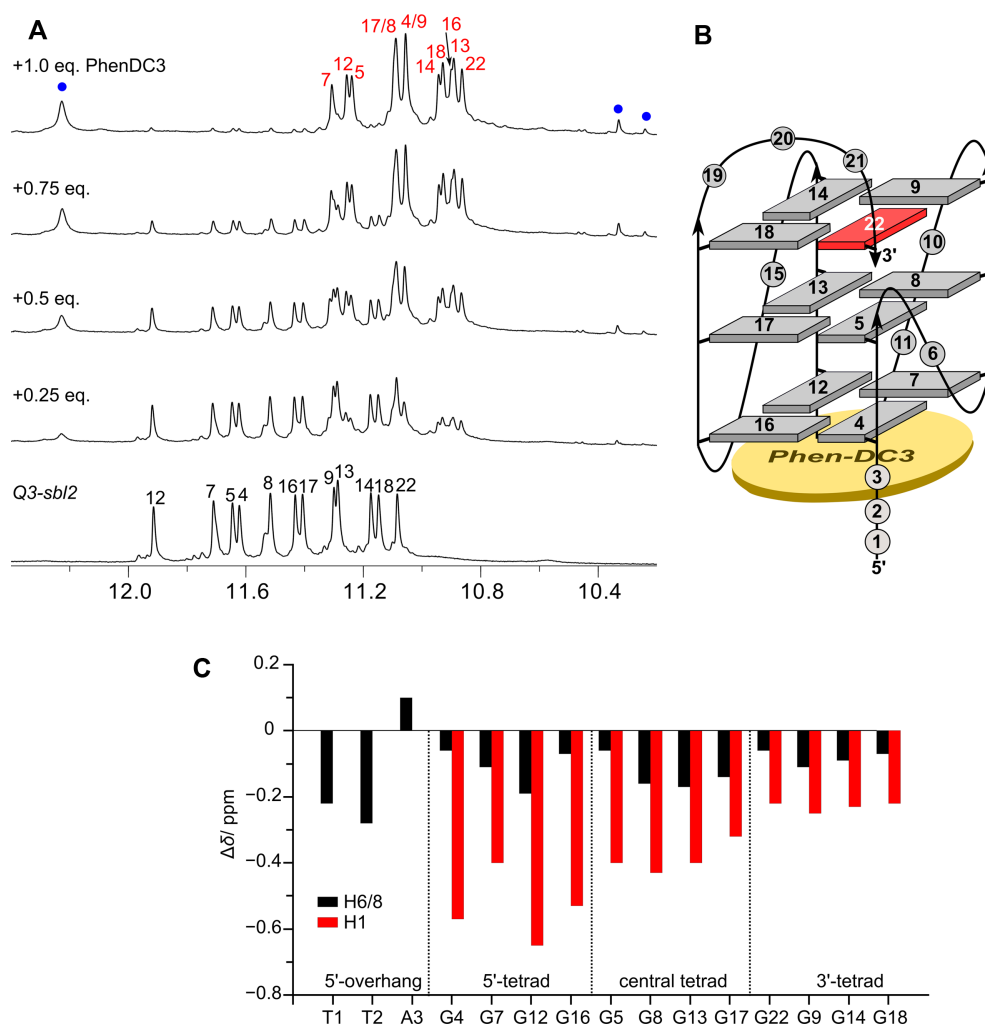


Figure 6. (A) Imino proton spectral region of *Q3-sbl2* (0.75 mM) upon titrating Phen-DC3 up to a 1:1 molar ratio in 10 mM potassium phosphate buffer, pH 7.0, at 30°C; assignments of the free *Q3-sbl2* and the *Q3-sbl2*–ligand complex are shown in black and red; Phen-DC3 resonances are marked by blue dots. (B) Schematic representation of *Q3-sbl2* with residue numbers showing the preferred binding site of Phen-DC3 at the 5'-outer tetrad. (C) Imino and H6/H8 chemical shift perturbations after binding Phen-DC3 to *Q3-sbl2*. Complete resonance assignments of the free and complexed DNA and a compilation of chemical shifts are given in Supplementary Figures S11–S13 and in Tables S7 and S8, respectively.

formation of an intercalation cavity with unwinding and tetrad unstacking of the interfacial base pair. Obviously, these energetically unfavorable processes are overcompensated by the considerable stacking interactions of the bound Phen-DC3 ligand. A lateral-type coaxially stacked duplex domain seems a prerequisite for strong binding at the junction and a non-duplex lateral loop will rather prevent binding due to unfavorable steric and/or compromised stacking interactions.

The phenyl substituent of PIQ-4m extends towards the center of the G-tetrad

ITC experiments have demonstrated an entropically more favored and also a more selective binding at Q–D junctions for the phenyl-substituted indoloquinoline PIQ-4m when compared to its close derivative SYUIQ-5. The latter bears a simple aliphatic aminoalkyl side chain that has recently been shown to be oriented towards the duplex minor groove

upon its intercalation at the junction (25). To get a better understanding of observed selectivities, binding of PIQ-4m was additionally studied on the *QD2-l* hybrid by an NMR structural analysis. The two-layered antiparallel *QD2-l* was employed because of a similar affinity when compared to the parallel *QD3-sbl* hybrid but with the additional benefit of eliminating putative competition with binding at the face opposite the Q–D junction as shown by ITC (see above).

Initially, the topology of *QD2-l* under the present solution conditions was verified to match the structure reported previously for the same sequence (PDB 2M8Z) (17). Again, standard strategies were employed for full resonance assignments of the hybrid (for a detailed description of assignment strategies, spectra, and a chemical shift table see the Supporting Information, Supplementary Figure S14 and Supplementary Table S9). Of note, this antiparallel two-layered G4 hybrid exhibits the same G-tetrad polarity as *QD3-sbl* with respect to the interfacial CG base pair at the junction, making the Q–D interface of both hybrids highly similar.

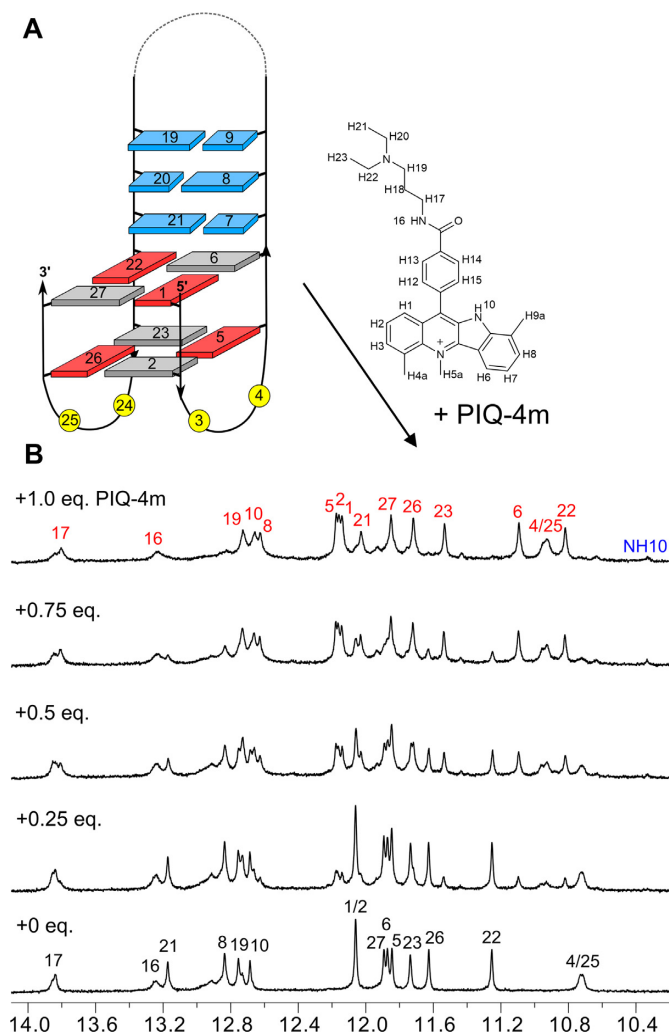


Figure 7. (A) Topology and chemical structure of the *QD2-I* hybrid and PIQ-4m indoloquinoline ligand with residue and atom labeling. (B) Imino proton spectral region of *QD2-I* (0.64 mM) upon titration with PIQ-4m up to a 1:1 molar ratio at 30°C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0. Imino proton resonances in free and ligand-bound *QD2-I* are labeled in black and red, respectively; NH10 represents a resonance from the ligand.

Gradual titration of PIQ-4m to *QD2-I* resulted in the appearance of new imino resonances and the complete disappearance of imino signals from the free hybrid after the addition of 1 equivalent of ligand (Figure 7). ROESY spectra of a 0.5:1 PIQ-4m–*QD2-I* mixture showed exchange cross-peaks between the free and complexed G4 hybrid. Most noticeable upfield shifts upon complex formation were observed for the imino proton of the interfacial C7–G21 base pair as well as for imino protons of G6 and G22 of the tetrad facing the duplex domain (Figure 7, Supplementary Figure S15). With a chemical shift changing by >1 ppm, the C7–G21 Watson–Crick imino signal shifts towards the typical chemical shift range of the G4 Hoogsteen imino protons in analogy to a corresponding shift when binding PhenDC3 to the *QD3-sbl* hybrid. Full resonance assignments of the 1:1 complex by the analysis of NOESY, DQF-COSY, TOCSY and ^1H – ^{13}C HSQC experiments confirmed these

chemical shift perturbations and suggest an intercalative binding mode of the PIQ-4m ligand at the Q–D junction. For more detailed information on the full spectral assignment of the PIQ-4m–*QD2-I* complex with a compilation of proton chemical shifts see the Supporting Information (Supplementary Figures S15–S19, Tables S10 and S11).

A total of 41 intermolecular contacts between ligand and the G4 hybrid could be observed in NOESY spectra. Various contacts position the ligand indole ring system between G6 and C7 and the quinoline moiety between G21 and G22 bases (Supplementary Figure S18). Notably, H14/H15 protons of the PIQ-4m phenyl substituent at about 8 ppm exhibit strong NOE contacts to all four imino protons within the tetrad at the Q–D interface with some corresponding but weaker contacts also observed for H12/H13 protons located on the opposite side of the phenyl ring (Supplementary Figure S18G). Apparently, the para-substituted phenyl side chain points towards the center of the interfacial G4 tetrad rather than towards the exterior and duplex minor groove.

Three-dimensional structures of the complex were calculated using NMR-derived experimental restraints. With average pairwise root mean square deviations of 1.4 Å for all atoms, 0.8 Å for the Q–D junction with intercalated ligand, and 0.4 Å for the G-core, there is a good convergence of final structures (Table 2). As suggested by intermolecular NOE contacts, the indoloquinoline intercalates between G-tetrad and base pair at the Q–D junction with the indole and quinoline subunits sandwiched between G6 and C7 and G21 and G22 residues, respectively (Figure 8). The PIQ-4m phenyl ring located above the center of the G-tetrad is tilted out of plane with respect to the indoloquinoline due to hydrogen-hydrogen steric repulsions in the 11-phenyl-indoloquinoline. Its orientation is well defined whereas the aminoalkyl side chain, amide-linked to the phenyl para-position and extending towards a G4 groove opposite the Q–D interface, seems rather flexible with no apparent long-lived electrostatic or hydrogen bond interactions with the G4. Conspicuously, NH16 of the amide functionality directly points towards the carbonyl oxygen of G1 in the structural ensemble with NHO angles of $157^\circ \pm 5^\circ$. Although this geometry suggests formation of a putative NHO hydrogen bond, H–O distances >3 Å do not support such an interaction to be of any major significance.

Structural determinants of affinities and selectivities for Q–D junctions

Stacking interactions seem to be the main driving force for ligand intercalation at the Q–D interface. Thus, as shown for TO and also recently demonstrated for cryptolepine (25), planar intercalators even without any additional side arms recognize a Q–D junction as binding hotspot to out-compete other available DNA binding sites. A pathway for intercalation suggested initial outside binding of the ligand followed by gradual insertion into a wedge formed at the intercalation site (53). Energy barriers as a result of forming a binding cavity through helical unwinding and separation of stacked bases must be offset by the binding and stacking energies of the intercalator. The energetic penalty of base unstacking is expected to increase with more extended sur-

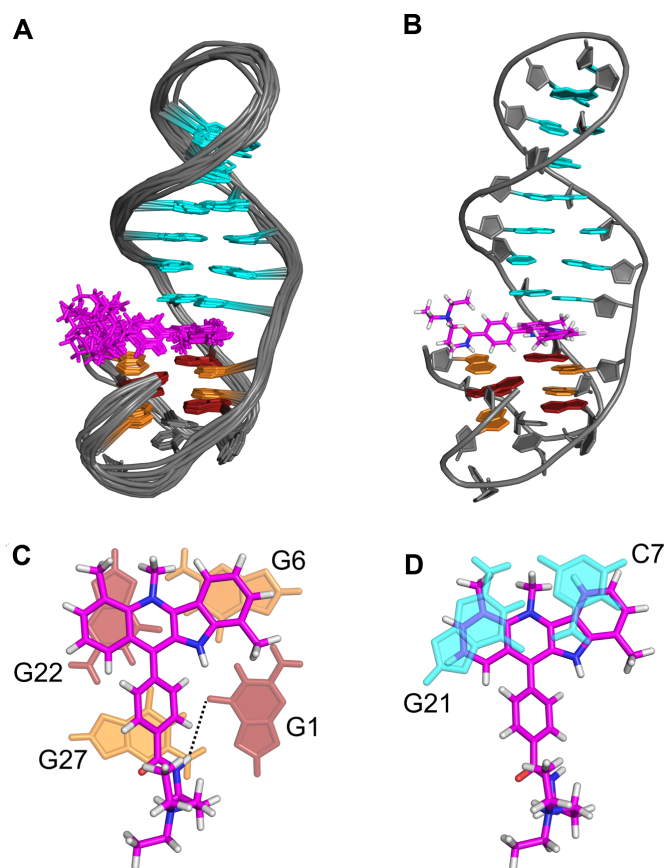


Figure 8. (A) Superposition of ten lowest-energy structures and (B) representative structure of *QD2-l* with bound PIQ-4m. (C) Detailed view of PIQ-4m stacked onto the G-tetrad and (D) below the C7:G21 base pair. A putative NHO hydrogen bond interaction between the PIQ amide and a guanine carbonyl is indicated by the dotted line in (C). In (A), residues of the duplex hairpin loop have been omitted for clarity.

face areas and the previously noted absence of any ligand intercalation at a putative Q–D junction of a particular Q–D hybrid design may derive from a formed base triad rather than a base pair stacked onto the G-tetrad platform (54). On the other hand, a recent study reported on simple aromatic hydrocarbon-based ligands to specifically recognize a Q–D junction by exclusively stacking onto the exposed area of the interfacial G-tetrad with a non-invaded Q–D junction (23). Such a non-intercalative binding mode may again be attributed to the inability of the aromatic hydrocarbons to overcome the energetic cost of forming an intercalation pocket at the junction.

A closer look at the Q–D interface upon ligand intercalation reveals conformational adjustments to optimize stacking interactions. Thus, inspection of available pairs of free and ligand-intercalated high-resolution structures of Q–D hybrids indicates a noticeable shift of the interfacial 5'-base of the duplex stem loop towards the tetrad for maximizing stacking interactions with the sandwiched ligand (Figure 9). Earlier studies on intercalation complexes with a B-type duplex or a dinucleotide base pair model have reported higher glycosidic torsion angles χ for the 3'-residues than for the 5'-residues of the intercalation site (55). In fact, the cyto-

sine residue located at the 5'-end of the duplex stem loop, i.e. at the 3'-side of the junction experiences a significant increase of χ towards the *high-anti* range for complexes with both *QD3-sbl* and *QD2-l* (Supplementary Table S12). As a result, the pyrimidine nucleotide of the base pair at the Q–D junction adopts higher glycosidic torsion angles than the complementary purine nucleotide. This contrasts with the non-complexed hybrids and the general expectation of a higher propensity for purine residues to adopt glycosidic torsion angles in the *high-anti* range. However, the terminal *syn-G* located at the other 3'-end of the intercalation site shows no significant change in its glycosidic torsion angle as expected from its participation and fixation within the outer G-tetrad.

Attaching appropriate side arms to the planar aromatic core structure of G4 ligands is considered highly beneficial for binding affinity but also binding selectivity, allowing for additional stabilizing interactions in grooves or with loop regions. It is instructive to compare binding geometries at the Q–D junction for indoloquinolines PIQ-4m and SYUIQ-5, the latter determined only recently (25). Conspicuously, for intercalated SYUIQ-5 the convex side of the slightly crescent-shaped indoloquinoline is positioned towards the outer surface at the Q–D interface, allowing its flexible side chain to interact within the duplex minor groove. In striking contrast, it is the concave side of intercalated PIQ-4m that is directed towards the exterior, positioning the out-of-plane phenyl substituent in opposite direction at the duplex major groove side and above the exposed part of the tetrad. A corresponding alignment of phenyl-indoloquinoline derivatives has also been reported upon their intercalation at duplex-triplex junctions with the phenyl ring again oriented towards the duplex major groove (56,57). On the other hand, the natural alkaloid cryptolepine lacking any side chain for controlling the ligand alignment was previously shown to intercalate between CG base pairs with its convex side towards the duplex minor groove (58). Apparently, the twisted phenyl substituent disfavors such a 'convex-out' binding mode, expected to be preferred based on a slightly better geometric match of the indoloquinoline sandwiched between a Watson–Crick and a G-tetrad.

More specific interactions can be expected from the SYUIQ-5 side chain if inserted into the duplex minor groove. On the other hand, favorable hydrophobic effects may accompany the positioning of the phenyl substituent of PIQ-4m above the exposed surface area of the tetrad. Indeed, less exothermic but much more favorable entropic contributions are associated with PIQ-4m binding at the junction when compared to SYUIQ-5 (see Supplementary Table S1). Combining enthalpic and entropic contributions, both indoloquinolines bind with similarly high affinity to the Q–D junction, yet selectivities against duplex DNA significantly differ in favor of the PIQ-4m derivative (see Figure 2). Given its less favorable binding enthalpy generally associated with less specific interactions, such an increase in selective binding seems counterintuitive at first. However, side chain interactions in the duplex minor groove of a Q–D hybrid structure are also expected to promote binding to a free duplex if the geometry of the ligand polycyclic ring system also permits intercalation between Watson–Crick base

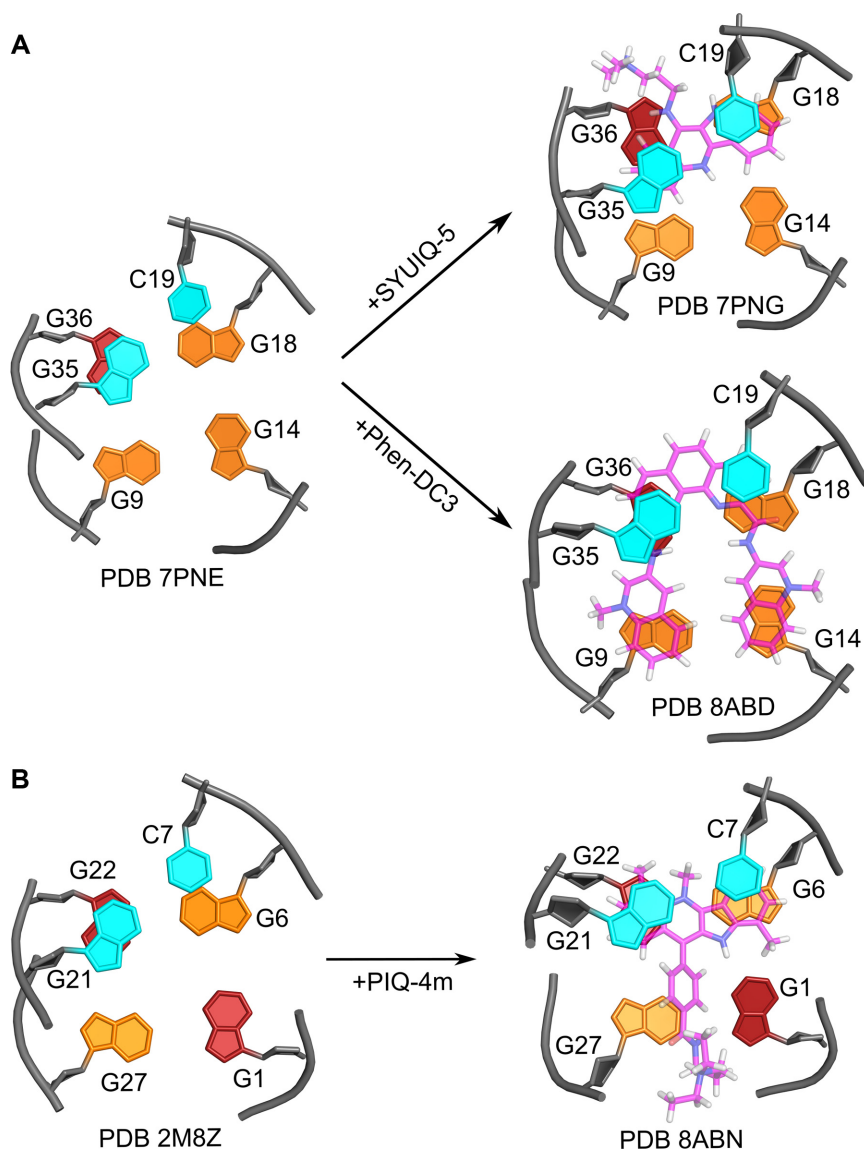
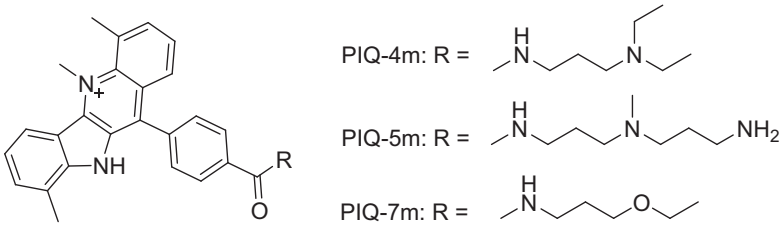


Figure 9. Q–D interface from high-resolution structures of free and ligand-bound Q–D hybrids. (A) Conformational changes of free *QD3-sbl* (PDB 7PNE) upon binding SYUIQ-5 (PDB 7PNG) (25) and Phen-DC3 (this study, PDB 8ABD). (B) Conformational change of free *QD2-l* (PDB 2M8Z) upon binding PIQ-4m (this study, PDB 8ABN).

pairs. Thus, the duplex association constant of SYUIQ-5 is higher by nearly two orders of magnitude when compared to PIQ-4m (Figure 2), raising doubts about the use of SYUIQ-5 as a G4 ligand in the presence of duplex DNA. Although not evident at first, PIQ-4m in fact shares similarities with Phen-DC3 in binding a Q–D junction. With the latter positioning its quinoline arms above the open side of the tetrad, both ligands lack additional interactions with the duplex minor groove but add stabilizing side chain contributions mainly through stacking, van der Waals, and/or hydrophobic interactions.

Side chains are often designed to carry additional positive charges through amine functionalities protonated at neutral pH. Although anticipated to form electrostatic or hydrogen bond interactions with the DNA backbone, the high flexibility of the PIQ-4m aminoalkyl arm did not al-

low observation of any specific contacts. To test the influence of charges on ligand binding to the Q–D hybrid, two additional PIQ derivatives PIQ-5m and PIQ-7m with a doubly charged *N*-aminopropyl-*N*-methyl-propyl and an uncharged ethoxypropyl arm were likewise employed for ITC studies in targeting the *QD2-l* receptor. As expected for ligands of the same family, thermograms strongly resemble corresponding heat profiles observed for PIQ-4m (Supplementary Figure S20). Excellent fits were obtained with a model of two binding sites of different affinity. In the absence of another freely accessible tetrad, these are assumed to be the Q–D junction and the duplex domain. Free energies ΔG° determined by curve fitting for both binding sites indicate a loss in initial high-affinity binding for PIQ-7m with its uncharged arm but no noticeable change when adding a second positively charged amino group in

Table 3. Standard free energies ΔG° for the binding of PIQ derivatives to *QD2-I* at 40°C^a


ligand	n_1	ΔG°_1 (kcal/mol) ^b	n_2	ΔG°_2 (kcal/mol) ^b	$\Delta G^\circ_1 - \Delta G^\circ_2$ (kcal/mol)
PIQ-4m	1.1 ± 0.1	-9.9 ± 0.1	5.0 ± 0.6	-6.2 ± 0.4	-3.7 ± 0.3
PIQ-5m	1.3 ± 0.1	-9.9 ± 0.1	5.0 ± 0.4	-6.7 ± 0.1	-3.1 ± 0.1
PIQ-7m	1.1 ± 0.1	-9.2 ± 0.2	6.0 ± 0.3	-5.6 ± 0.1	-3.5 ± 0.2

^aAverage values from three independent ITC experiments in 20 mM potassium phosphate buffer, pH 7.0, 100 mM KCl, and 5% DMSO at 40 °C; data were fitted with two sets of binding sites. ^b $\Delta G^\circ = -RT \ln K_a$.

PIQ-5m (Table 3; for complete binding profiles see Supplementary Table S13). Consequently, the simply charged side chain of PIQ-4m seems to provide for additional, albeit short-lived and weak electrostatic or hydrogen bond interactions when binding at the junction but the addition of another positive charge hardly exerts any more stabilizing effect. On the other hand, binding to the duplex depends on the charge density of the side chains following the order PIQ-5m > PIQ-4m > PIQ-7m. These results suggest smaller free energy differences and thus less selectivity for PIQ-5m with its doubly charged substituent. Similar selectivities can be expected for PIQ-4m and PIQ-7m, albeit with a higher affinity for the singly charged PIQ-4m ligand. Although uncertainties for extracted parameters have to be taken into account, there is a clear trend in such charge-dependent affinity-selectivity relationships, largely matching previous findings from a comprehensive thermodynamic profiling on the binding of PIQ ligands to the exposed outer tetrad of a parallel *c-myc* quadruplex (29). Based on the above, caution has to be exercised when deciding on charged side arms for G4 recognition in the presence of additional competing nucleic acid structures.

CONCLUSIONS

Q–D junctions are unique structural motifs meeting increasing interest not only for their use in technological applications but also as targets for therapeutics due to their proposed frequent occurrence in genomic sequences. With only few reports on the specific recognition of such junctions through low molecular weight ligands to date, an extended set of detailed thermodynamic and structural data on various G4 binding compounds have been reported here. These demonstrate that Q–D junctions either featuring a hairpin-type snapback loop or a lateral duplex stem-loop with coaxial orientation of duplex and G4 helices represent no structural peculiarities calling for a novel type of

Q–D ligands but are in fact superior binding sites for most G4 ligands usually found to stack on outer G-tetrads. Consequently, the targeting of G4 structures will benefit from higher affinities in the presence of such Q–D junctions due to enhanced stacking interactions upon ligand intercalation.

On the other hand, care must be exercised when following the obvious approach of designing Q–D ligands by combining G4 and duplex binding subunits for optimizing affinities and also selectivities. As suggested by the present studies, moieties designed for duplex minor groove binding, e.g. side chains attached to a planar aromatic core structure, may be detrimental to selectivities against duplex DNA due to an associated strong promotion of duplex binding. To confer more selectivity in the presence of excess double-helical nucleic acids, a strategy is proposed that directs additional interactions towards the G4 core away from the duplex minor groove in contrast to more common perceptions. This will require an extension of the intercalating ligand through side arms to favorably align and interact with the exposed area of the interfacial tetrad. Clearly, such design principles will strictly apply to polycyclic ligands with aromatic surface areas that reasonably match the base pair geometry to also allow for favorable duplex intercalation, as in fact seen for many typical G4 binding compounds. In contrast, macrocyclic or any other G4 ligands with extended surface areas as also represented by U-shaped Phen-DC3 are unable to efficiently intercalate between base pairs. Thus, whereas Phen-DC3 already favors a Q–D junction over any other G4 binding site, Q–D intercalation is expected here to additionally benefit from appropriate minor groove binding side chains, enhancing affinity without compromising selectivity. With studies encompassing different binding partners, the present results provide for valuable general guidelines to support the future design of potent ligands not only for selective Q–D recognition but also for G4 targeting in the presence of predominant duplex structures.

DATA AVAILABILITY

The atomic coordinates and chemical shifts for *QD3-sbl* with bound Phen-DC3 (PDB ID: 8ABD, BMRB ID: 34740) and for *QD2-l* with bound PIQ-4m (PDB ID: 8ABN, BMRB ID: 34741) have been deposited in the Protein Data Bank.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Deutsche Forschungsgemeinschaft [410497337]. Funding for open access charge: Deutsche Forschungsgemeinschaft [410497337]; Universität Greifswald.

Conflict of interest statement. None declared.

REFERENCES

- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Spiegel, J., Adhikari, S. and Balasubramanian, S. (2020) The structure and function of DNA G-quadruplexes. *Trends Chem.*, **2**, 123–136.
- Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
- Neidle, S. (2017) Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.*, **1**, 0041.
- Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627–636.
- Zhang, S., Wu, Y. and Zhang, W. (2014) G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem*, **9**, 899–911.
- Santos, T., Salgado, G.F., Cabrita, E.J. and Cruz, C. (2021) G-Quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions. *Pharmaceuticals*, **14**, 769.
- Jana, J., Mohr, S., Vianney, Y.M. and Weisz, K. (2021) Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.*, **2**, 338–353.
- Mohr, S., Jana, J., Vianney, Y.M. and Weisz, K. (2021) Expanding the topological landscape by a G-column flip of a parallel G-quadruplex. *Chem. Eur. J.*, **27**, 10437–10447.
- Karg, B., Mohr, S. and Weisz, K. (2019) Duplex-guided refolding into novel G-quadruplex (3+1) hybrid conformations. *Angew. Chem. Int. Ed.*, **58**, 11068–11071.
- Zhou, J., Bourdoncle, A., Rosu, F., Gabelica, V. and Mergny, J.L. (2012) Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed.*, **51**, 11002–11005.
- Macaya, R.F., Waldron, J.A., Beutel, B.A., Gao, H., Joesten, M.E., Yang, M., Patel, R., Bertelsen, A.H. and Cook, A.F. (1995) Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs. *Biochemistry*, **34**, 4478–4492.
- Russo Krauss, I., Spiridonova, V., Pica, A., Napolitano, V. and Sica, F. (2016) Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res.*, **44**, 983–991.
- Lim, K.W., Jenjaroenpun, P., Low, Z.J., Khong, Z.J., Ng, Y.S., Kuznetsov, V.A. and Phan, A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630–5646.
- Monsen, R.C., DeLeeuw, L.W., Dean, W.L., Gray, R.D., Chakravarthy, S., Hopkins, J.B., Chaires, J.B. and Trent, J.O. (2022) Long promoter sequences form higher-order G-quadruplexes: an integrative structural biology study of c-Myc, k-Ras and c-Kit promoter sequences. *Nucleic Acids Res.*, **50**, 4127–4147.
- Tan, D.J.Y., Winnerdy, F.R., Lim, K.W. and Phan, A.T. (2020) Coexistence of two quadruplex–duplex hybrids in the PIM1 gene. *Nucleic Acids Res.*, **48**, 11162–11171.
- Lim, K.W. and Phan, A.T. (2013) Structural basis of DNA quadruplex–duplex junction formation. *Angew. Chem. Int. Ed.*, **52**, 8566–8569.
- Lim, K.W., Khong, Z.J. and Phan, A.T. (2014) Thermal stability of DNA quadruplex–duplex hybrids. *Biochemistry*, **53**, 247–257.
- Phan, A.T., Kuryavii, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.*, **18**, 796–804.
- Vasilyev, N., Polonskaia, A., Darnell, J.C., Darnell, R.B., Patel, D.J. and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5391–E5400.
- Nguyen, T.Q.N., Lim, K.W. and Phan, A.T. (2017) A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs. *Sci. Rep.*, **7**, 11969.
- Asamitsu, S., Obata, S., Phan, A.T., Hashiya, K., Bando, T. and Sugiyama, H. (2018) Simultaneous binding of hybrid molecules constructed with dual DNA-binding components to a G-quadruplex and its proximal duplex. *Chem. Eur. J.*, **24**, 4428–4435.
- Díaz-Casado, L., Serrano-Chacón, I., Montalvillo-Jiménez, L., Corzana, F., Bastida, A., Santana, A.G., González, C. and Asensio, J.L. (2020) De novo design of selective quadruplex–duplex junction ligands and structural characterisation of their binding mode: targeting the G4 hot-spot. *Chem. Eur. J.*, **27**, 6204–6212.
- Liu, L.Y., Wang, K.N., Liu, W., Zeng, Y.L., Hou, M.X., Yang, J. and Mao, Z.W. (2021) Spatial matching selectivity and solution structure of organic–metal hybrid to quadruplex–duplex hybrid. *Angew. Chem. Int. Ed.*, **60**, 20833–20839.
- Vianney, Y.M. and Weisz, K. (2022) Indoloquinoline ligands favor intercalation at quadruplex–duplex interfaces. *Chem. Eur. J.*, **28**, e202103718.
- Liu, L.Y., Ma, T.Z., Zeng, Y.L., Liu, W. and Mao, Z.W. (2022) Structural basis of pyridostatin and its derivatives specifically binding to G-quadruplexes. *J. Am. Chem. Soc.*, **144**, 11878–11887.
- Vianney, Y.M., Preckwinkel, P., Mohr, S. and Weisz, K. (2020) Quadruplex–duplex junction: a high-affinity binding site for indoloquinoline ligands. *Chem. Eur. J.*, **26**, 16910–16922.
- Funke, A., Dickerhoff, J. and Weisz, K. (2016) Towards the development of structure-selective G-quadruplex-binding indolo[3,2-*b*]quinolines. *Chem. Eur. J.*, **22**, 3170–3181.
- Funke, A. and Weisz, K. (2019) Thermodynamic signature of indoloquinolines interacting with G-quadruplexes: impact of ligand side chain. *Biochimie*, **157**, 142–148.
- Nygren, J., Svanvik, N. and Kubista, M. (1998) The interactions between the fluorescent dye thiazole and DNA. *Biopolymers*, **46**, 39–51.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Genet.*, **59**, 687–696.
- Schwieters, C.D., Kuszewski, J.J. and Marius Clore, G. (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.*, **48**, 47–62.
- Ou, T.M., Lin, J., Lu, Y.J., Hou, J.Q., Tan, J.H., Chen, S.H., Li, Z., Li, Y.P., Li, D., Gu, L.Q. *et al.* (2011) Inhibition of cell proliferation by quindoline derivative (SYUIQ-05) through its preferential interaction with c-myc promoter G-quadruplex. *J. Med. Chem.*, **54**, 5671–5679.
- Dai, J., Carver, M., Hurley, L.H. and Yang, D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.*, **133**, 17673–17680.
- Tanious, F.A., Wilson, W.D. and Yen, S.F. (1991) Kinetic and equilibrium analysis of a threading intercalation mode: DNA sequence and ion effects. *Biochemistry*, **30**, 1813–1819.
- Parkinson, G.N., Cuenca, F. and Neidle, S. (2008) Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex–drug complexes. *J. Mol. Biol.*, **381**, 1145–1156.
- Read, M., Harrison, R.J., Romagnoli, B., Tanious, F.A., Gowan, S.H., Reszka, A.P., Wilson, W.D., Kelland, L.R. and Neidle, S. (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4844–4849.

38. Campbell, N.H., Parkinson, G.N., Reszka, A.P. and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.*, **130**, 6722–6724.
39. De Cian, A., DeLemos, E., Mergny, J.L., Teulade-Fichou, M.P. and Monchaud, D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.*, **129**, 1856–1857.
40. Chung, W.J., Heddi, B., Hamon, F., Teulade-Fichou, M.P. and Phan, A.T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chem. Int. Ed.*, **53**, 999–1002.
41. Tran, P.L.T., Largy, E., Hamon, F., Teulade-Fichou, M.P. and Mergny, J.L. (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie*, **93**, 1288–1296.
42. Suss, O., Motiei, L. and Margulies, D. (2021) Broad applications of thiazole orange in fluorescent sensing of biomolecules and ions. *Molecules*, **26**, 2828.
43. Phan, A.T., Kuryavii, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract g-quadruplex structure from the human myc promoter. *Nat. Chem. Biol.*, **1**, 167–173.
44. Schnarr, L., Jana, J., Preckwinkel, P. and Weisz, K. (2020) Impact of a snap-back loop on stability and ligand binding to a parallel G-quadruplex. *J. Phys. Chem. B*, **124**, 2778–2787.
45. Macaya, R.F., Schultze, P., Smith, F.W., Roe, J.A. and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 3745–3749.
46. Wang, X.D., Ou, T.M., Lu, Y.J., Li, Z., Xu, Z., Xi, C., Tan, J.H., Huang, S.L., An, L.K., Li, D. *et al.* (2010) Turning off transcription of the bcl-2 gene by stabilizing the bcl-2 promoter quadruplex with quindoline derivatives. *J. Med. Chem.*, **53**, 4390–4398.
47. Monchaud, D., Allain, C. and Teulade-Fichou, M.P. (2007) Thiazole orange: a useful probe for fluorescence sensing of G-quadruplex-ligand interactions. *Nucleosides. Nucleotides Nucleic Acids*, **26**, 1585–1588.
48. Street, S.T.G., Chin, D.N., Hollingworth, G.J., Berry, M., Morales, J.C. and Galan, M.C. (2017) Divalent naphthalene diimide ligands display high selectivity for the human telomeric G-quadruplex in K⁺ buffer. *Chem. Eur. J.*, **23**, 6953–6958.
49. Moore, M.J.B., Schultes, C.M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanius, F.A., Wilson, W.D. and Neidle, S. (2006) Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3,6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *J. Med. Chem.*, **49**, 582–599.
50. Le, V.H., Buscaglia, R., Chaires, J.B. and Lewis, E.A. (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal. Biochem.*, **434**, 233–241.
51. Tuite, E. and Nordén, B. (1994) Sequence-specific interactions of methylene blue with polynucleotides and DNA: a spectroscopic study. *J. Am. Chem. Soc.*, **116**, 7548–7556.
52. White, E.W., Tanius, F., Ismail, M.A., Reszka, A.P., Neidle, S., Boykin, D.W. and Wilson, W.D. (2007) Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophys. Chem.*, **126**, 140–153.
53. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., Hynes, J.T. and Lavery, R. (2012) Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J. Am. Chem. Soc.*, **134**, 8588–8596.
54. Russo Krauss, I., Ramaswamy, S., Neidle, S., Haider, S. and Parkinson, G.N. (2016) Structural insights into the quadruplex–duplex 3' interface formed from a telomeric repeat: a potential molecular target. *J. Am. Chem. Soc.*, **138**, 1226–1233.
55. Shieh, H.-S., Berman, H.M., Dabrow, M. and Neidle, S. (1980) The structure of drug-deoxydinucleoside phosphate complex: generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.*, **8**, 85–97.
56. Eick, A., Riechert-Krause, F. and Weisz, K. (2012) Binding and NMR structural studies on indoloquinoline-oligonucleotide conjugates targeting duplex DNA. *Bioconjug. Chem.*, **23**, 1127–1137.
57. Dickerhoff, J., Riechert-Krause, F., Seifert, J. and Weisz, K. (2014) Exploring multiple binding sites of an indoloquinoline in triple-helical DNA: a paradigm for DNA triplex-selective intercalators. *Biochimie*, **107**, 327–337.
58. Lisgarten, J.N., Coll, M., Portugal, J., Wright, C.W. and Aymami, J. (2002) The antimalarial and cytotoxic drug cryptolepine intercalates into DNA at cytosine-cytosine sites. *Nat. Struct. Biol.*, **9**, 57–60.

Supplementary Information for

High-Affinity Binding at Quadruplex-Duplex Junctions: Rather the Rule than the Exception

Yoanes Maria Vianney and Klaus Weisz*

Institute of Biochemistry, Universität Greifswald, 17489 Greifswald, Germany

Contents

Experimental details	S3
Table S1: Binding parameters of ligands from ITC titrations	S4
Table S2: UV melting temperature of DNA constructs	S5
Table S3: Binding parameters for PIQ-4m from ITC titrations	S6
Figure S1: NMR imino proton spectral region of <i>QD2-l-2bp</i>	S7
Figure S2: CD melting curves of <i>QD2-l</i> with and without ligands	S7
Figure S3: CD spectra of <i>QD3-sbl</i> upon the addition of ligands	S8
Figure S4: NOESY spectral regions of a <i>QD3-sbl</i>	S9
Table S4: Compilation of ¹ H and ¹³ C chemical shifts for free <i>QD3-sbl</i>	S10
Figure S5: ROESY imino exchange crosspeaks for a <i>QD3-sbl</i> – TO complex	S11
Figure S6: ROESY imino exchange crosspeaks for a <i>QD3-sbl</i> – BRACO-19 complex	S12
Resonance assignments of the <i>QD3-sbl</i> – Phen-DC3 complex	S13
Figure S7: ¹ H- ¹³ C HSQC and NOESY spectral regions of the <i>QD3-sbl</i> – Phen-DC3 complex	S15
Figure S8: H1'-H2' NOE and DQF-COSY crosspeaks of the <i>QD3-sbl</i> – Phen-DC3 complex	S16
Figure S9: NOESY spectral regions of the <i>QD3-sbl</i> – Phen-DC3 complex	S17
Table S5: Compilation of ¹ H and ¹³ C chemical shifts in a <i>QD3-sbl</i> – Phen-DC3 complex	S18
Table S6: Compilation of ¹ H chemical shifts of Phen-DC3 bound to <i>QD3-sbl</i>	S19
Figure S10: Chemical shift perturbations of <i>QD3-sbl</i> upon Phen-DC3 binding	S19
Resonance assignments of <i>Q3-sbl/2</i> without and with bound Phen-DC3	S20
Figure S11: ¹ H- ¹³ C HSQC and NOESY spectral regions of <i>Q3-sbl/2</i>	S21

Table S7: Compilation of ^1H and ^{13}C chemical shifts for free <i>Q3-sb/2</i>	S22
Figure S12: ^1H - ^{13}C HSQC and NOESY spectral regions of the <i>Q3-sb/2</i> – Phen-DC3 complex	S23
Figure S13: NOESY and ROESY spectral regions of the <i>Q3-sb/2</i> – Phen-DC3 complex	S24
Table S8: Compilation of ^1H and ^{13}C chemical shifts in a <i>Q3-sb/2</i> – Phen-DC3 complex	S25
Resonance assignments of <i>QD2-I</i> without and with bound PIQ-4m	S26
Figure S14: NOESY spectral regions of free <i>QD2-I</i>	S28
Table S9: Compilation of ^1H chemical shifts for free <i>QD2-I</i>	S29
Figure S15: 1D, ROESY and chemical shift perturbations of <i>QD2-I</i> upon PIQ-4m binding	S30
Figure S16: NOESY and ^1H - ^{13}C HSQC spectral regions of the <i>QD2-I</i> – PIQ-4m complex	S31
Figure S17: H1'-H2' NOE and DQF-COSY crosspeaks of the <i>QD2-I</i> – PIQ-4m complex	S32
Figure S18: NOESY spectral regions of the <i>QD2-I</i> – PIQ-4m complex	S33
Figure S19: DQF-COSY, TOCSY, and ROESY spectra of the <i>QD2-I</i> – PIQ-4m complex	S34
Table S10: Compilation of ^1H and ^{13}C chemical shifts in a <i>QD2-I</i> – PIQ-4m complex	S35
Table S11: Compilation of ^1H chemical shifts of PIQ-4m bound to <i>QD2-I</i>	S36
Table S12: Glycosidic torsion angles for nucleotides at Q-D junctions	S36
Figure S20: ITC thermograms upon the titration of PIQ derivatives to <i>QD2-I</i>	S37
Table S13: Binding parameters for PIQ derivatives from ITC titrations	S37

EXPERIMENTAL SECTION

NMR experiments

A WATERGATE W5 pulse scheme was employed for water suppression in one-dimensional (1D) and two-dimensional (2D) NOESY experiments on a 90% H₂O/10% D₂O phosphate buffer solution. For DQF-COSY, TOCSY, and ¹H-¹³C HSQC experiments, a 3-9-19 W3 binomial-type sequence was used for solvent suppression. ¹H-¹³C HSQC spectra were recorded with 4K x 500 data points, a 1 s recycle delay, and 7500 Hz in the F1 dimension to include aromatic C8/C6/C2 resonances of the nucleobases. TOCSY spectra with a DIPSI-2 isotropic mixing scheme and an 80 ms mixing time as well as DQF-COSY spectra were recorded with 4K x 500 data points. 2D NOESY (80, 150, and 300 ms mixing time) and EASY-ROESY spectra (80 ms mixing time, 50° spinlock angle) were acquired with 2K x 1K data points. For all 2D homonuclear experiments a 2 s recycle delay was used. Data were zero-filled to 4K x 1K data points and processed with squared sine-bell window functions except for time domain data of 1D experiments that were multiplied with an exponential function.

Restraints for structure calculations

For the structure calculations, distance restraints were assigned based on the crosspeak intensities in NOESY spectra. For non-exchangeable protons and intramolecular ligand contacts, these were set according to 2.9 ± 1.1 Å for strong crosspeaks, 4.0 ± 1.2 Å for crosspeaks of medium intensity, 5.5 ± 1.5 Å for weak crosspeaks, 6.0 ± 1.5 Å for very weak crosspeaks, and 5.0 ± 2.0 Å for ambiguous crosspeaks due to signal overlap. For labile protons, distances were restrained according to 2.9 ± 1.1 Å for strong crosspeaks, 4.0 ± 1.2 Å for crosspeaks of medium intensity, 5.0 ± 1.2 Å for weak crosspeaks, 6.0 ± 1.2 Å for very weak crosspeaks, and 5.0 ± 2.0 Å for ambiguous crosspeaks due to signal overlap. Intermolecular ligand-DNA contacts were set to $4.0 +1.5/-2.0$ Å for strong crosspeaks, $5.5 +1.5/-2.0$ Å for weak crosspeaks, $6.0 +1.5/-2.0$ Å for very weak crosspeaks, and 5.0 ± 2.0 Å for ambiguous crosspeaks due to signal overlap. Torsion angles χ were either set to *anti* (170° – 310°) or *syn* (25° – 95°). Sugar pucker information was derived from DQF-COSY crosspeaks, whose in-phase/antiphase pattern is determined by corresponding coupling constants and enables differentiation between a *north* (pseudorotation angle of 0° – 36°) and *south* sugar pucker (pseudorotation angle of 144° – 180°) through the Karplus-relationship. Distance restraints for hydrogen bonds were added for the G-tetrads and all Watson-Crick base pairs and additional planarity restraints were added to the quadruplex core. Force constants for *in vacuo* simulations were set to 40 and 50 kcal·mol⁻¹·Å⁻² for NOE-based distance and hydrogen bond restraints, 200 kcal·mol⁻¹·rad⁻² for glycosidic torsion angle and sugar pucker restraints, and 30 kcal·mol⁻¹·Å⁻² for G-tetrad planarity restraints. Additional chirality restraints with a restraint energy of 10 kcal·mol⁻¹·rad⁻² were added for calculations of the QD3-*sbI* - Phen-DC3 complex. For subsequent simulations in explicit water, only NOE-based distance and hydrogen bond restraints with restraint energies of 15 and 25 kcal·mol⁻¹·Å⁻² were employed.

Table S1. ITC-derived binding parameters of quadruplex ligands^a

PIQ-4m titrated to	n	K_a (M ⁻¹)	ΔH° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol) ^b
<i>QD3-sbl</i>	1.2 ± 0.1	1.3 · 10 ⁷ ± 2.9 · 10 ⁶	-9.8 ± 0.1	-0.4 ± 0.2
	7.3 ± 0.2	7.3 · 10 ³ ± 2.5 · 10 ²	-8.8 ± 0.2	3.2 ± 0.2
<i>Q3-sbl</i>	1.3 ± 0.1	3.2 · 10 ⁶ ± 1.4 · 10 ⁶	-4.9 ± 0.3	-4.4 ± 0.5
	4.7 ± 0.9	8.5 · 10 ⁴ ± 1.4 · 10 ⁴	-1.1 ± 0.4	-6.0 ± 0.3
<i>D3-HP</i>	1.8 ± 0.4	1.6 · 10 ⁴ ± 2.2 · 10 ³	-6.1 ± 0.1 ^c	0.1 ± 0.1 ^c
SYUIQ-5 titrated to				
<i>QD3-sbl</i>	0.9 ± 0.1	1.1 · 10 ⁷ ± 3.3 · 10 ⁶	-14.1 ± 0.2	4.0 ± 0.3
	5.7 ± 0.1	9.2 · 10 ⁴ ± 1.3 · 10 ⁴	-11.8 ± 0.1	4.7 ± 0.2
<i>Q3-sbl</i>	1.0 ± 0.1	2.1 · 10 ⁶ ± 6.1 · 10 ⁵	-12.2 ± 0.5	3.2 ± 0.4
	7.2 ± 1.2	8.4 · 10 ⁴ ± 4.2 · 10 ⁴	-2.8 ± 0.4	-4.2 ± 0.7
<i>D3-HP</i>	1.8 ± 0.3	1.3 · 10 ⁶ ± 1.2 · 10 ⁵	-11.5 ± 0.1 ^c	2.7 ± 0.1 ^c
	4.6 ± 0.7	6.5 · 10 ⁴ ± 5.0 · 10 ³	-3.9 ± 1.1	-3.0 ± 1.1
TO titrated to				
<i>QD3-sbl</i>	1.8 ± 0.1	1.1 · 10 ⁶ ± 2.7 · 10 ⁴	-13 ± 0.6	4.4 ± 0.6
	8.5 ± 0.5	1.4 · 10 ⁴ ± 2.9 · 10 ³	-5.2 ± 0.3	-0.7 ± 0.3
<i>Q3-sbl</i>	0.8 ± 0.1	1.7 · 10 ⁵ ± 6.7 · 10 ⁴	-7.9 ± 1.2	0.4 ± 1.5
<i>D3-HP</i>	1.5 ± 0.2	4.1 · 10 ⁴ ± 2.7 · 10 ³	-11.7 ± 0.4	5.1 ± 0.3
NDI-DM titrated to				
<i>QD3-sbl</i>	3.0 ± 0.2	1.2 · 10 ⁵ ± 2.5 · 10 ⁴	-12.5 ± 0.3	5.2 ± 0.4
<i>Q3-sbl</i>	1.1 ± 0.1	9.5 · 10 ⁴ ± 1.1 · 10 ⁴	-8.7 ± 0.4	1.5 ± 0.4
<i>D3-HP</i>	1.8 ± 0.1	7.2 · 10 ⁴ ± 2.3 · 10 ⁴	-8.7 ± 1.5	1.8 ± 1.7
BRACO-19 titrated to				
<i>QD3-sbl</i>	0.5 ± 0.1	1.4 · 10 ⁷ ± 6.1 · 10 ⁶	-14.6 ± 0.9	4.4 ± 0.6
	5.6 ± 0.6	8.2 · 10 ⁴ ± 1.3 · 10 ⁴	-12.9 ± 1.0	5.8 ± 0.9
<i>Q3-sbl</i>	1.2 ± 0.1	3.0 · 10 ⁷ ± 2.3 · 10 ⁶	-10.3 ± 0.1	-0.3 ± 0.1
	4.2 ± 0.1	1.0 · 10 ⁵ ± 2.5 · 10 ³	-4.0 ± 0.1	-3.2 ± 0.1
<i>D3-HP</i>	2.1 ± 0.1	2.9 · 10 ⁶ ± 6.9 · 10 ⁵	-11.3 ± 0.2	2.1 ± 0.1
	3.9 ± 0.4	1.4 · 10 ⁵ ± 4.8 · 10 ⁴	-2.2 ± 0.2	-5.1 ± 0.1

^aAverage values with root-mean-square deviations obtained from three independent measurements in 20 mM potassium phosphate buffer, pH 7.0, 100 mM KCl, and 5% DMSO at 40 °C; data were fitted with one or two sets of binding sites with fit parameters of a second lower affinity binding site shown on a grey background. ^b $-T\Delta S^\circ = \Delta G^\circ - \Delta H^\circ$ with $\Delta G^\circ = -RT\ln K_a$.

^cDetermined by an excess-site titration.

Table S2. UV melting temperature T_m of DNA constructs

name	T_m (°C) ^a	
	G4	duplex
<i>QD3-sbl</i>	63.3 ± 1.0	63.8 ± 0.9
<i>Q3-sbl</i>	>70 ^b	---
<i>D3-HP</i>	---	55.9 ± 0.1
<i>QD2-I</i>	58.7 ± 0.2	62.7 ± 0.9
<i>D2-HP</i>	---	62.9 ± 0.4
<i>QD2-I-2bp</i>	52.0 ± 1.2	--- ^c
<i>TBA</i>	51.0 ± 0.2	---
<i>Q3-sbl2</i>	n.d. ^d	---

^aDetermined by three independent UV melting experiments in ITC buffer.

^bNo high-temperature baseline reached within temperature range. ^cNo resolved duplex melting. ^dNot determined.

Table S3. ITC-derived binding parameters of PIQ-4m^a

PIQ-4m titrated to	n	K_a (M ⁻¹)	ΔH° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol) ^b
<i>QD2-I</i>	1.1 ± 0.1	8.3·10 ⁶ ± 1.2·10 ⁵	-11.5 ± 0.3	1.6 ± 0.3
	5.0 ± 0.6	2.5·10 ⁴ ± 1.2·10 ⁴	-6.5 ± 0.5	0.2 ± 0.8
<i>D2-HP</i>	2.9 ± 0.2	4.5·10 ⁴ ± 1.4·10 ⁴	-6.6 ± 0.6	0.0 ± 0.8
<i>QD2-I-2bp</i>	1.1 ± 0.1	7.1·10 ⁶ ± 7.7·10 ⁴	-12.6 ± 0.2	2.8 ± 0.2
	2.2 ± 0.5	5.2·10 ⁴ ± 1.0·10 ⁴	-4.7 ± 0.8	-2.0 ± 0.9
<i>TBA</i>	3.2 ± 0.4	2.6·10 ⁴ ± 8.6·10 ³	-8.1 ± 1.7	1.8 ± 1.9

^aAverage values with root-mean-square deviations obtained from three independent measurements in 20 mM potassium phosphate buffer, pH 7.0, 100 mM KCl, and 5% DMSO at 40 °C; data were fitted with one or two sets of binding sites with fit parameters of a second lower affinity binding site shown on a grey background. ^b $-T\Delta S^\circ = \Delta G^\circ - \Delta H^\circ$ with $\Delta G^\circ = -R\ln K_a$.

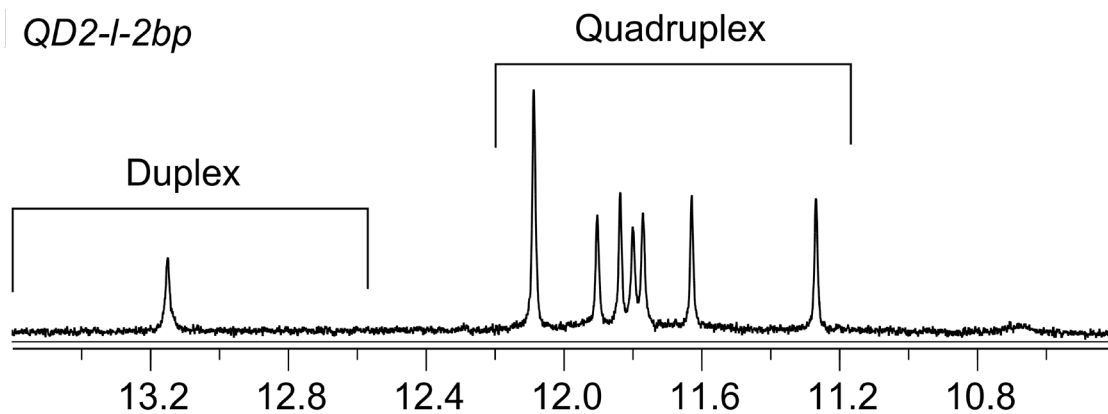


Figure S1. NMR imino proton spectral region of *QD2-I-2bp*. In addition to imino resonances of the quadruplex G-core, a GC Watson-Crick imino signal is also observed. Another potential Watson-Crick imino proton resonance seems to be broadened beyond detection through increased flexibility and associated fast solvent exchange under the present conditions. The spectrum was acquired with a strand concentration of 0.3 mM in 120 mM K^+ buffer, pH 7.0, at 40 °C.

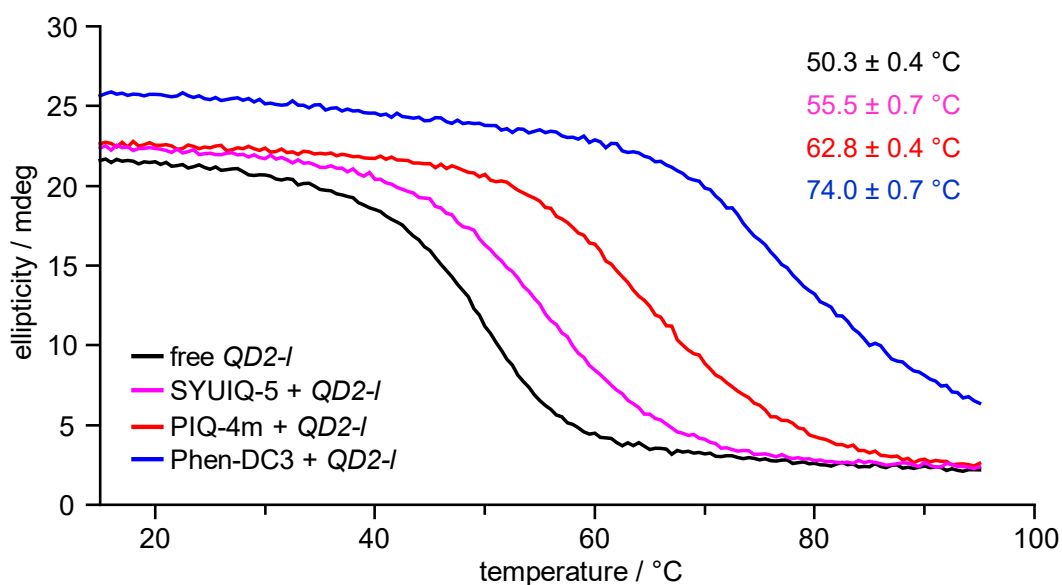


Figure S2. CD melting curves of *QD2-I* without (black) and with addition of 1 equivalent SYUIQ-5 (magenta), PIQ-4m (red), or Phen-DC3 (blue). Corresponding melting temperatures were determined from two independent experiments each in 10 mM potassium phosphate buffer, pH 7.0.

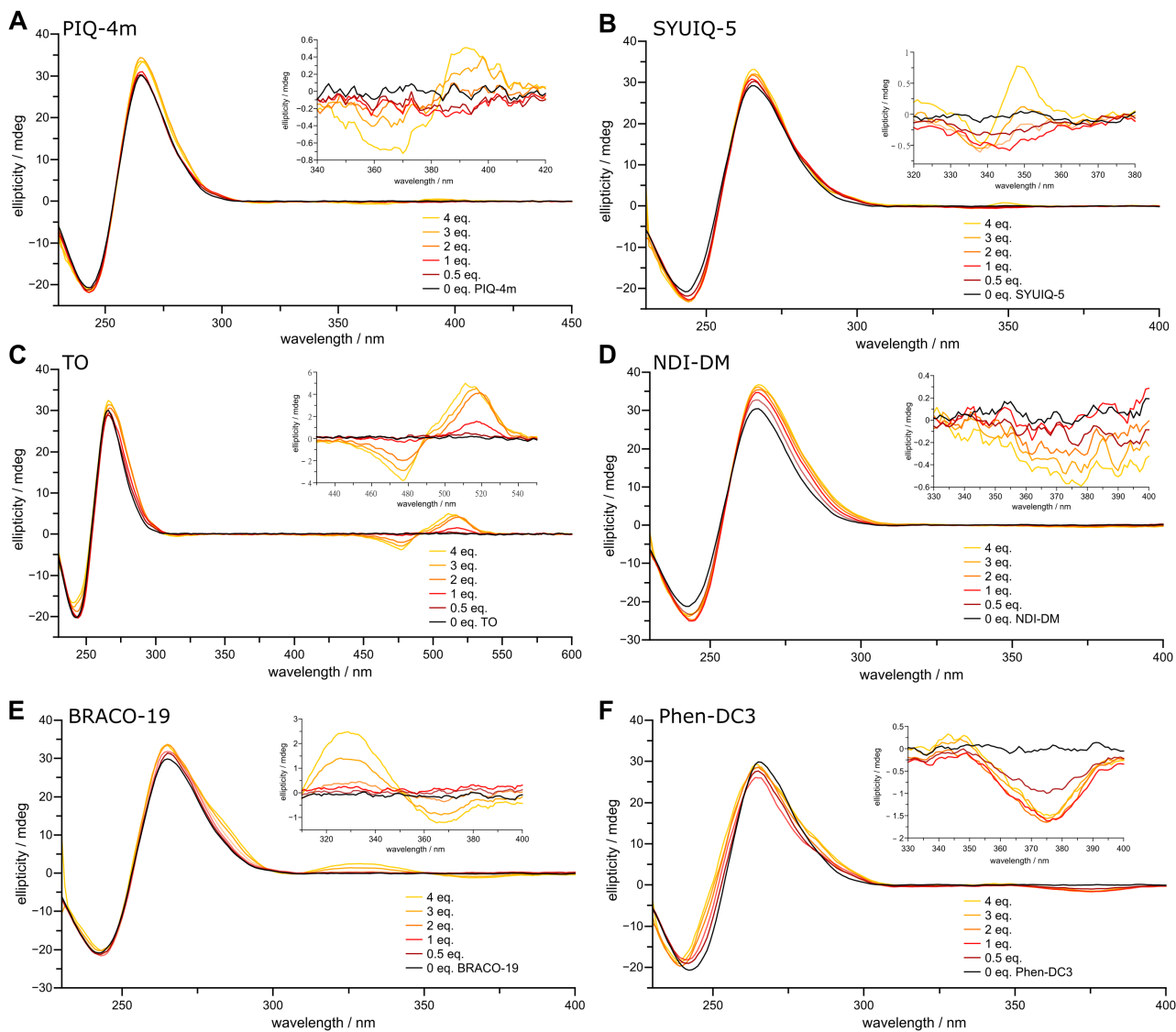


Figure S3. CD spectra of *QD3-sbI* upon addition of (A) PIQ-4m, (B) SYUIQ-5, (C) TO, (D) NDI-DM, (E) BRACO-19, and (F) Phen-DC3. Spectra were recorded at 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7, and are blank-corrected.

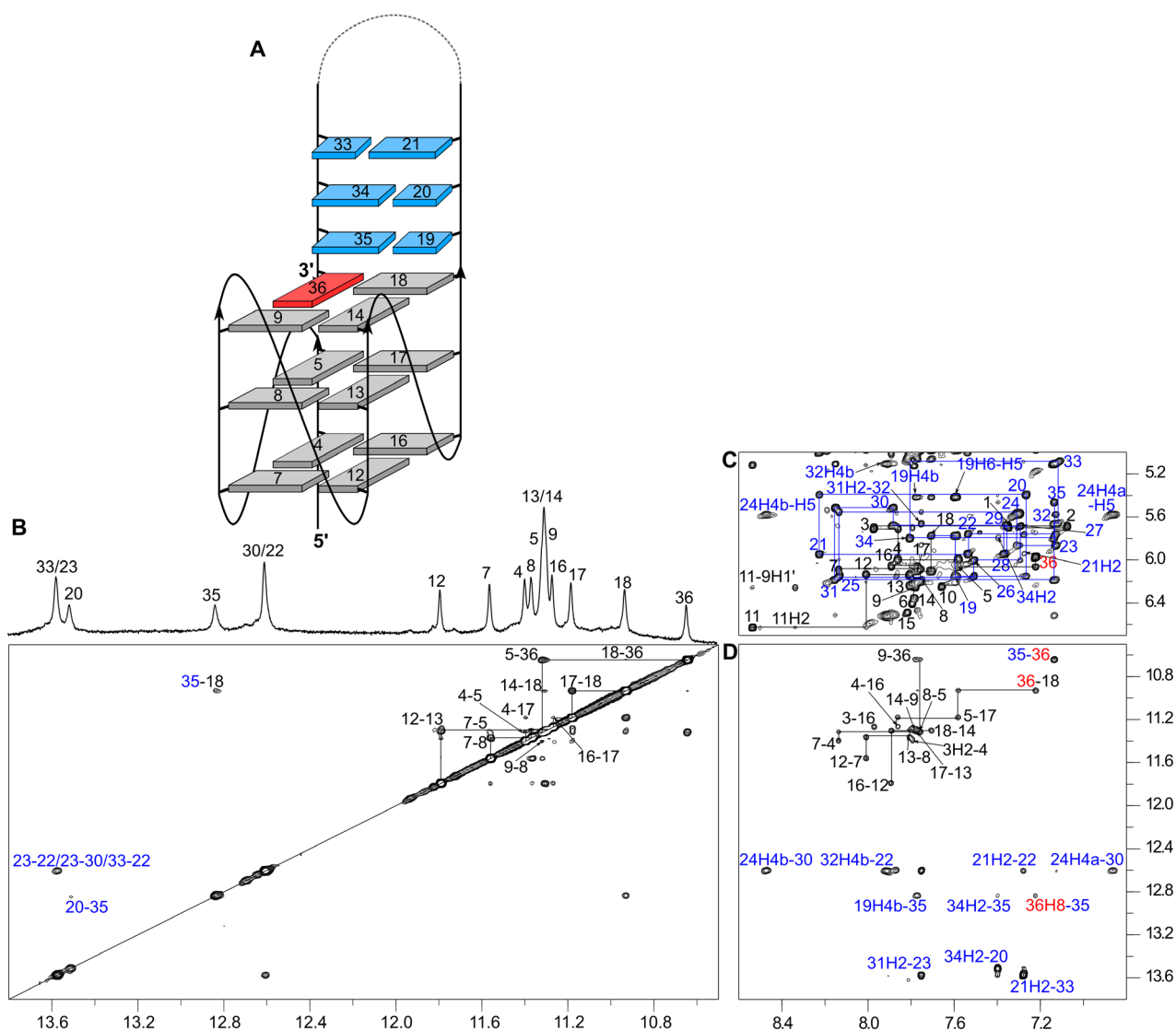


Figure S4. (A) Topology of *QD3-sbl* with residue numbers. (B) Imino(ω₂)-imino(ω₁) NOESY spectral region with corresponding 1D spectrum on top. (C) H6/H8(ω₂)-H1'(ω₁) NOESY spectral region; NOE connectivities and corresponding residue numbers within the quadruplex and duplex domains are given in black and blue color, respectively; *syn*-G36 is labeled in red. (D) H6/H8(ω₂)-imino(ω₁) NOESY spectral region. NOESY spectra (300 ms mixing time) of *QD3-sbl* (0.5 mM) were acquired in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0, at 30 °C.

Note: Resonance assignments and the determination of a three-dimensional NMR structure of the same sequence has been published recently under different conditions (20 °C, 10 mM potassium phosphate buffer, pH 7.0; PDB 7PNE; BMRB 34664).¹ The present NMR analysis demonstrates that different buffer conditions as used here did result in some minor chemical shift changes but did not impact the formed topology.

1. Vianney, Y. M.; Weisz, K. Indoloquinoline Ligands Favor Intercalation at Quadruplex-Duplex Interfaces. *Chem. Eur. J.* **2022**, *28*. e202103718.

Table S4. ¹H and ¹³C NMR chemical shifts δ for free QD3-sb^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2'' ^b	H3'	H5/H2/Me	C8/C6	C2
T1	7.35	n.d.	5.69	1.99/2.18	4.41	1.60	139.52	-
T2	7.08	n.d.	5.69	1.55/2.05	4.48	1.48	139.00	-
A3	7.97	-	5.71	2.55/2.62	4.82	7.80	141.82	154.95
G4	7.86	11.39	6.00	2.50/2.85	4.89	-	138.15	-
G5	7.58	11.31	6.06	2.36/2.61	4.98	-	137.97	-
T6	7.80	n.d.	6.41	2.38/2.61	4.96	1.93	140.04	-
G7	8.14	11.56	6.09	2.67/2.89	4.92	-	138.38	-
G8	7.76	11.37	6.20	2.57/2.93	4.99	-	138.09	-
G9	7.78	11.30	6.26	2.48/2.57	5.02	-	n.d.	-
T10	7.66	n.d.	6.25	2.24/2.46	4.75	1.93	140.11	-
A11	8.54	n.d.	6.63	2.92/3.01	5.12	8.34	143.07	155.73
G12	8.01	11.79	6.14	2.60/2.95	4.98	-	138.40	-
G13	7.81	11.31	6.24	2.63/3.04	5.02	-	138.29	-
G14	7.79	11.30	6.35	2.50/2.63	5.12	-	n.d.	-
T15	7.82	n.d.	6.49	2.46/2.66	5.06	1.96	139.90	-
G16	7.89	11.27	6.06	2.44/2.84	5.01	-	138.01	-
G17	7.77	11.18	6.10	2.62/2.91	5.06	-	138.07	-
G18	7.71	10.93	5.77	2.42/2.64	4.98	-	137.79	-
C19	7.59	-	6.15	2.03/2.42	4.88	5.42	n.d.	-
T20	7.27	13.51	5.39	2.05/2.26	4.82	1.60	139.58	-
A21	8.23	-	5.95	2.79/2.84	5.00	7.28	141.86	154.07
G22	7.54	12.61	5.76	2.39/2.62	4.75	-	137.19	-
T23	7.13	13.58	5.86	1.93/2.34	4.73	1.21	138.45	-
C24	7.31	-	5.56	1.74/2.18	4.73	5.58	143.54	-
A25	8.14	-	6.16	2.51/2.68	4.95	7.87	142.06	155.23
T26	7.51	n.d.	6.00	2.05/2.31	4.65	1.76	139.38	-
T27	7.29	n.d.	5.69	1.91/2.16	4.57	1.56	139.39	-
T28	7.37	n.d.	5.94	2.15/2.35	4.68	1.60	139.54	-
T29	7.36	n.d.	5.68	2.09/2.39	4.67	1.79	139.52	-
G30	7.89	12.61	5.52	2.64/2.73	4.95	-	138.35	-
A31	8.15	-	6.19	2.63/2.82	4.97	7.76	141.14	155.10
C32	7.14	-	5.67	1.71/2.22	4.58	5.11	n.d.	-
T33	7.12	13.58	5.09	1.79/1.92	4.57	1.40	139.61	-
A34	7.81	-	5.80	2.33/2.54	4.75	7.40	141.81	154.02
G35	7.14	12.84	5.46	2.20/2.65	4.83	-	136.65	-
G36	7.22	10.64	5.97	2.28/2.96	4.73	-	139.89	-

^aAt 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0. ^bNo stereospecific assignment.

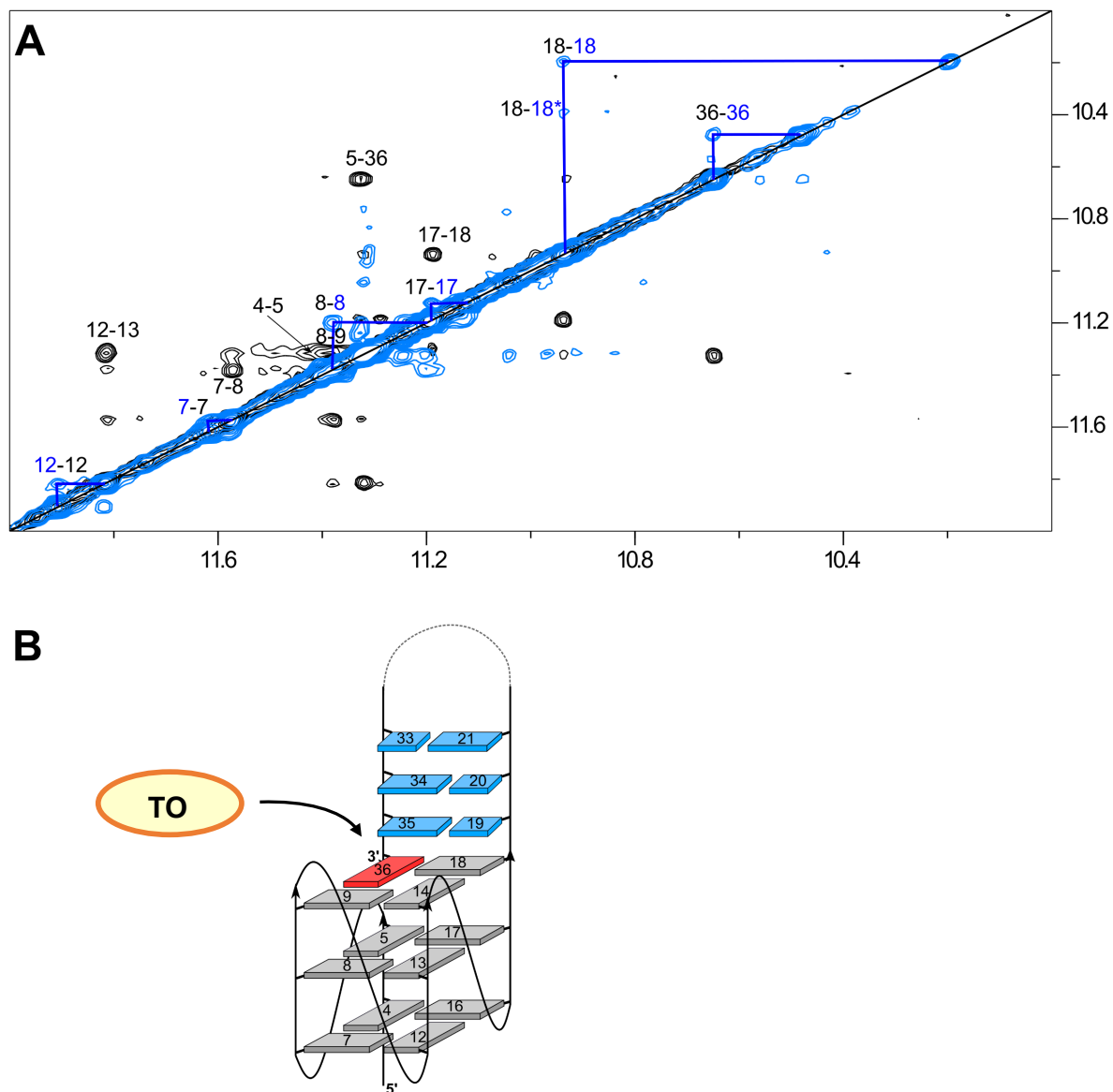


Figure S5. (A) Superposition of the imino(ω_2)-imino(ω_1) region in a NOESY spectrum (300 ms mixing time, black) on the free *QD3-sbl* hybrid and in a ROESY spectrum (blue) on a mixture of *QD3-sbl* and 0.5 equivalent of TO ligand. For simplicity, only positive exchange crosspeaks are shown in the ROESY spectrum with negative peaks suppressed. Blue lines connect imino protons exchanging between the free (labeled in black) and the complexed DNA (labeled in blue). Spectra were acquired in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0, at 30 °C. (B) Schematic representation of high-affinity TO binding to *QD3-sbl*.

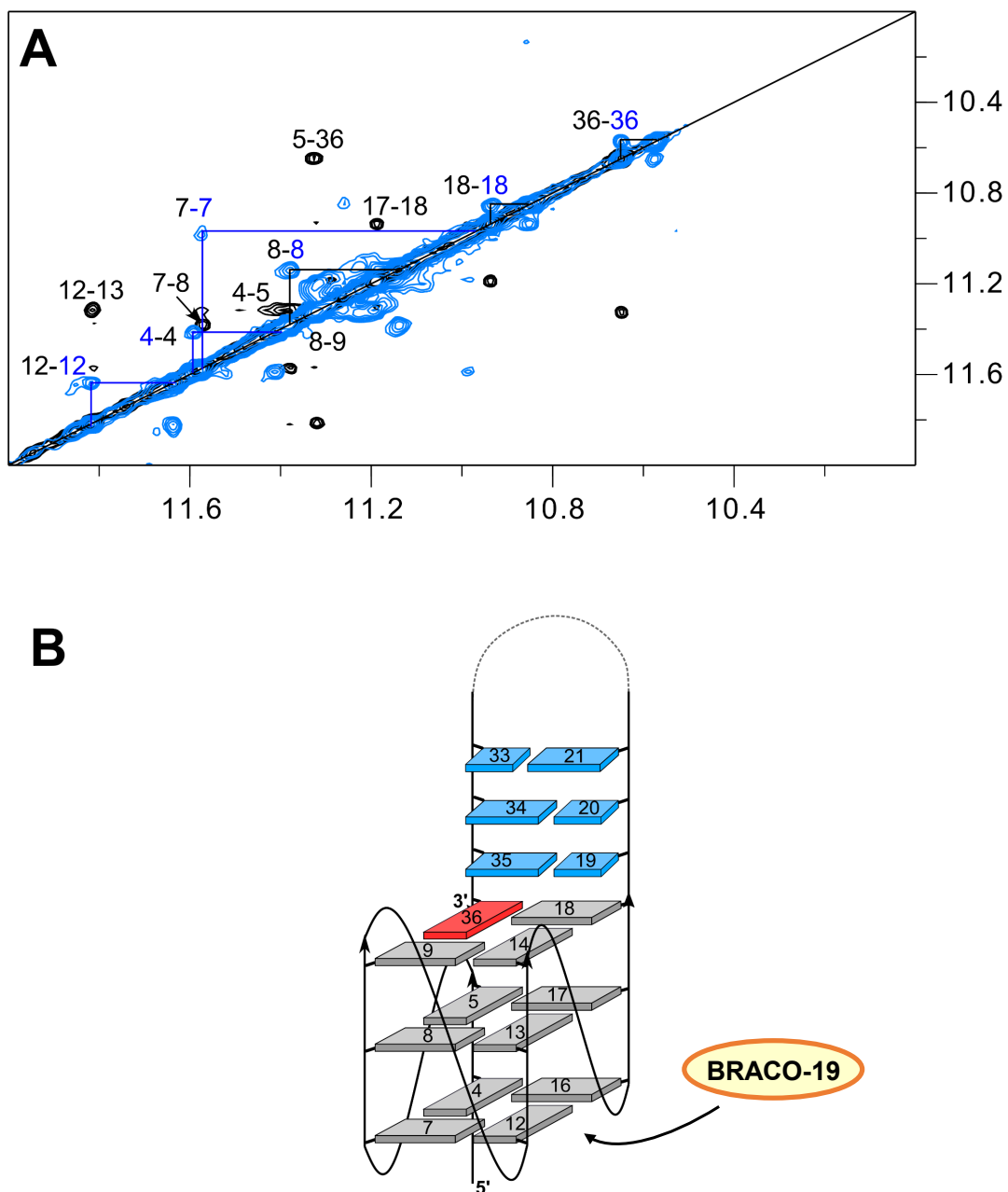


Figure S6. (A) Superposition of the imino(ω_2)-imino(ω_1) region in a NOESY spectrum (300 ms mixing time, black) on the free *QD3-sbl* hybrid and in a ROESY spectrum (blue) on a mixture of *QD3-sbl* and 1 equivalent of the BRACO-19 ligand. For simplicity, only positive exchange crosspeaks are shown in the ROESY spectrum with negative peaks suppressed. Blue lines connect imino protons exchanging between the free (labeled in black) and the complexed DNA (labeled in blue). Spectra were acquired in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0, at 30 °C. (B) Schematic representation of high-affinity BRACO-19 binding to *QD3-sbl*.

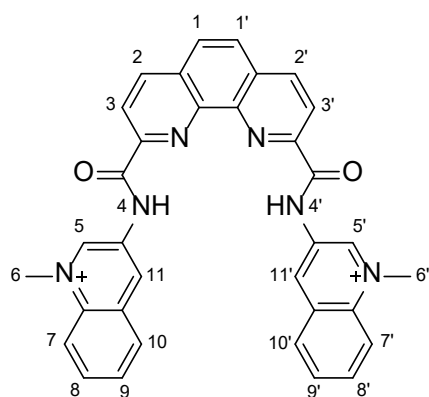
Resonance assignments of a *QD3-sbl* complex with Phen-DC3

Continuous sugar-base NOE contacts can be followed from T1 through G5 and up to *syn*-G36 as well as along the other three G-columns with interruptions by the propeller-type loops (Figure S7B,C). The second and the third G-column can be differentiated by NOE contacts of A11 in the second propeller loop, showing sequential contacts to G12 but also additional contacts in particular through its H2 proton to G9 H1'. Sequential NOEs also connect all residues from C19 to G35 of the duplex stem. However, in contrast to the free *QD3-sbl*, NOE connectivities from the 3'-tetrad to the duplex are interrupted. The presence of G36 adopting a *syn* conformation is demonstrated by its strong H8-H1' intra-nucleotide crosspeak and its typical downfield-shifted G ¹³C8 resonance observed in a ¹H-¹³C HSQC experiment (Figure S7A). Stereospecific assignments of the H2'/H2'' sugar protons are based on H1'-H2' crosspeak intensities in NOESY experiments acquired at short mixing times (Figure S8A). Sugar puckers were grouped into *south*-type and *north*-type conformations depending on the pattern and intensity of DQF-COSY H1'-H2' and H1'-H2'' crosspeaks (Figure S8B). Most of the residues were found to be in a *south* conformation with some sugars remaining ambiguous and left unassigned. In contrast, T1 and G36 could be restrained to a *north* sugar conformation. Guanine imino (H1) protons of the G-core were assigned without specific isotope labeling by following the characteristic pattern of intra- and inter-tetrad H8-H1 contacts. Their assignment was confirmed by imino-imino connectivities typical for an all-homopolar stack of G-tetrads with polarities of the three tetrads when going from hydrogen bond donor to acceptor along G4→G7→G12→G16, G5→G8→G13→G17, and G36→G9→G14→G18 (Figure S9D,E). Contacts from H1' sugar protons of 5'-overhang residues to four imino resonances in the 5'-tetrad further confirmed their assignment (Figure S9F). Overall, the topology of *QD3-sbl* was retained after the addition of 1 equivalent Phen-DC3. Unlike free *QD3-sbl*, a prominent crosspeak correlating G35 H8 with G36 H1 is missing in the complex (Figure S9E). Imino protons of the Watson-Crick base pairs were identified through their crosspeaks with cytosine amino protons (H4a/H4b for non-hydrogen and hydrogen-bonded) and adenine H2 (Figure S9G,H). Notably, a contact between the hydrogen-bonded C19 amino with the considerably upfield-shifted G35 imino resonance in a NOESY experiment with 80 ms mixing time confirms an intact Watson-Crick CG base pair at the junction (Figure S9H).

Two sets of different resonances were observed for the symmetry-related protons of Phen-DC3 upon binding. These exhibit no exchange crosspeaks in a ROESY experiment on a mixture with a 1:1 *QD3-sbl* – Phen-DC3 molar ratio, indicating only slow ring flips of the ligand when bound to the hybrid (not shown). Proton resonances of the ligand were fully assigned by standard strategies using COSY correlations to identify groups of scalar coupled protons assisted by NOESY experiments to link the different spin systems (Figure S7, S9). The two most downfield-shifted isochronous H4 amide protons at about 11.8 ppm serve as a convenient starting point by exhibiting NOE contacts to the quinoline moiety including *N*-methyl protons (Figure S9A).

Various contacts of phenanthroline protons H1, H2, and H3 but also of the H4 amide and quinoline protons closest to the phenanthroline moiety with residues G18, C19, G35, and G36 point to their location at the Q-D junction (Figure S7 and S9). On the other hand, quinoline protons distant from the phenanthroline ring system exhibit NOE contacts to the exposed interfacial G-tetrad residues G9 and G14. These connectivities suggest

the phenanthroline to be sandwiched between base pair and G-tetrad at the Q-D junction whereas both quinoline side arms project towards the open part of the G-tetrad.



Chemical structure of Phen-DC3 with atom numbering as used in this study.

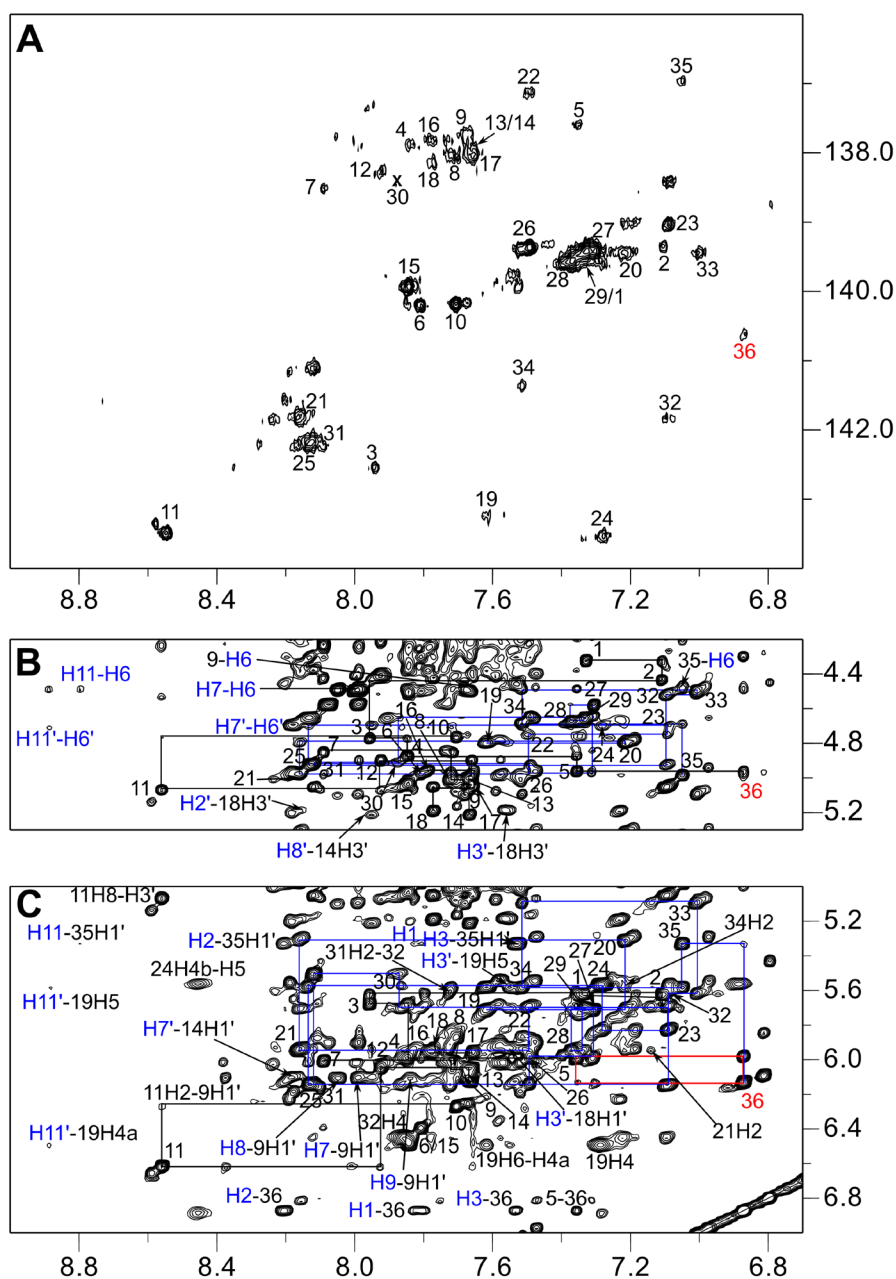


Figure S7. NMR spectra with assignments for a 1:1 complex of QD3-sbl (0.5 mM) with Phen-DC3 acquired at 30 °C in 10 mM potassium phosphate buffer, pH 7. (A) ^1H - ^{13}C HSQC spectral region with H6/H8(ω_2)-C6/C8(ω_1) correlations; the crosspeak with downfield-shifted ^{13}C 8 of *syn*-G36 is labeled in red; a G30 H8-C8 correlation only observed at lower threshold levels is marked by a cross. (B) H6/H8(ω_2)-H3'(ω_1) and (C) H6/H8(ω_2)-H1'(ω_1) spectral region of a NOESY spectrum (300 ms mixing time). NOE sequential connectivities for the quadruplex and the duplex domain are traced by black and blue lines, respectively. *Syn*-G36 with its strong intra-nucleotide H8-H1' crosspeak shows a weak NOE contact of its H1' proton to G5 H8 across the *anti-syn* step with antiparallel strand orientation (highlighted by the red rectangular pattern). Intra- and intermolecular ligand contacts are labeled with the ligand proton written in blue.

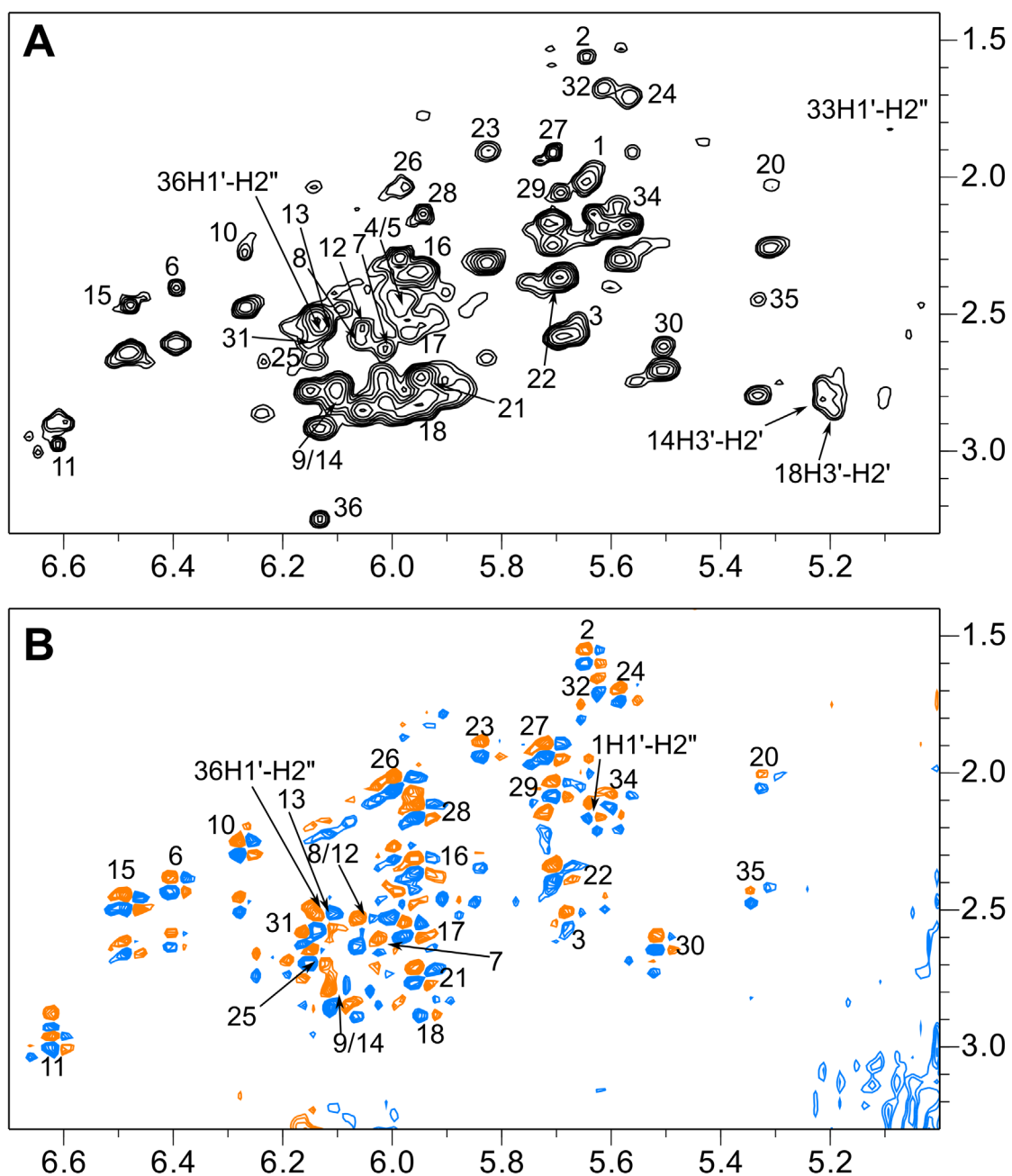


Figure S8. (A) NOESY spectrum (80 ms mixing time) and (B) DQF-COSY spectrum for a 1:1 complex of QD3-*sbl* with Phen-DC3, showing the H1'(ω_2)-H2'/H2''(ω_1) spectral region (30 °C, 10 mM potassium phosphate buffer, pH 7). Following a stereospecific H2'/H2'' assignment based on NOE crosspeak intensities, sugar conformations can be assessed through inspection of DQF-COSY crosspeak patterns, demonstrating a *north*-type sugar pucker for residues T1 and G36.

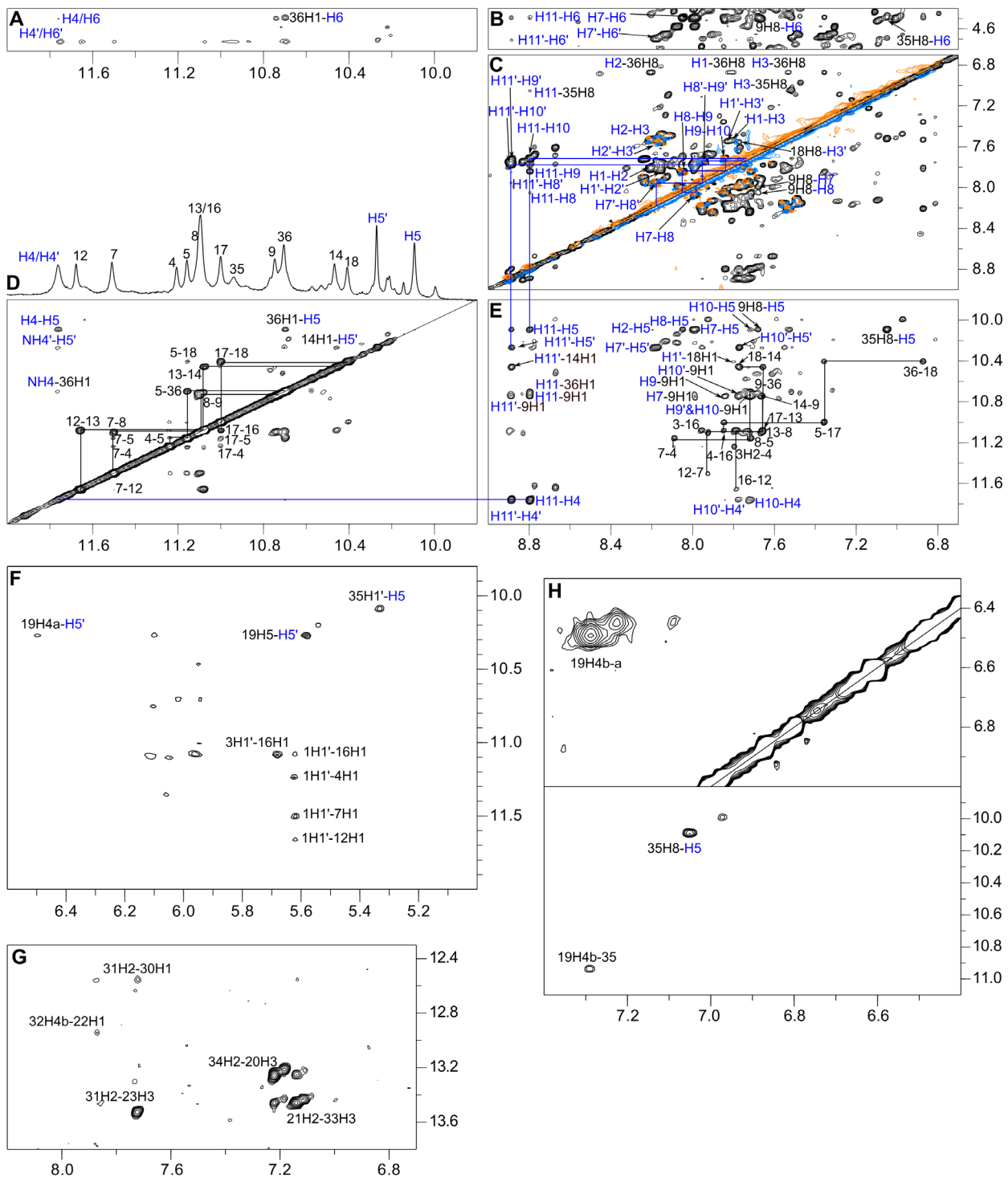


Figure S9. Resonance assignments of a 1:1 Phen-DC3 - QD3-*sbI* complex and intermolecular DNA-ligand contacts. (A,B) NOESY spectral regions showing inter- and intramolecular NOE contacts to the *N*-methyl H6/H6'(ω_1) protons of Phen-DC3. (C) Superposition of a NOESY (black) and DQF-COSY spectrum with assignments of ligand resonances through both scalar couplings and NOE connectivities. (D) NOESY spectral region with imino(ω_2)-imino(ω_1) crosspeaks and corresponding 1D spectrum with Hoogsteen G imino resonances (top). (E) H6/8(ω_2)-imino(ω_1) NOESY spectral region also comprising intra- and intermolecular crosspeaks of Phen-DC3 protons. (F) H1'(ω_2)-imino(ω_1) NOE crosspeaks with additional ligand H5/H5' contacts. (G) C amino-G imino and A H2-T H3 NOE contacts for the duplex domain with Watson-Crick hydrogen bonds. (H) Intra-base C19 amino(ω_2)-amino(ω_1) contact (top) and intra-base pair C19 amino(ω_2)-C35 imino(ω_1) contact (bottom). DNA and ligand protons are labeled with black and blue color, respectively. NOESY spectra were acquired at 30 °C in 10 mM potassium phosphate buffer, pH 7, with a 300 ms mixing time except for the spectral region in (H) that derives from a NOESY experiment with an 80 ms mixing time.

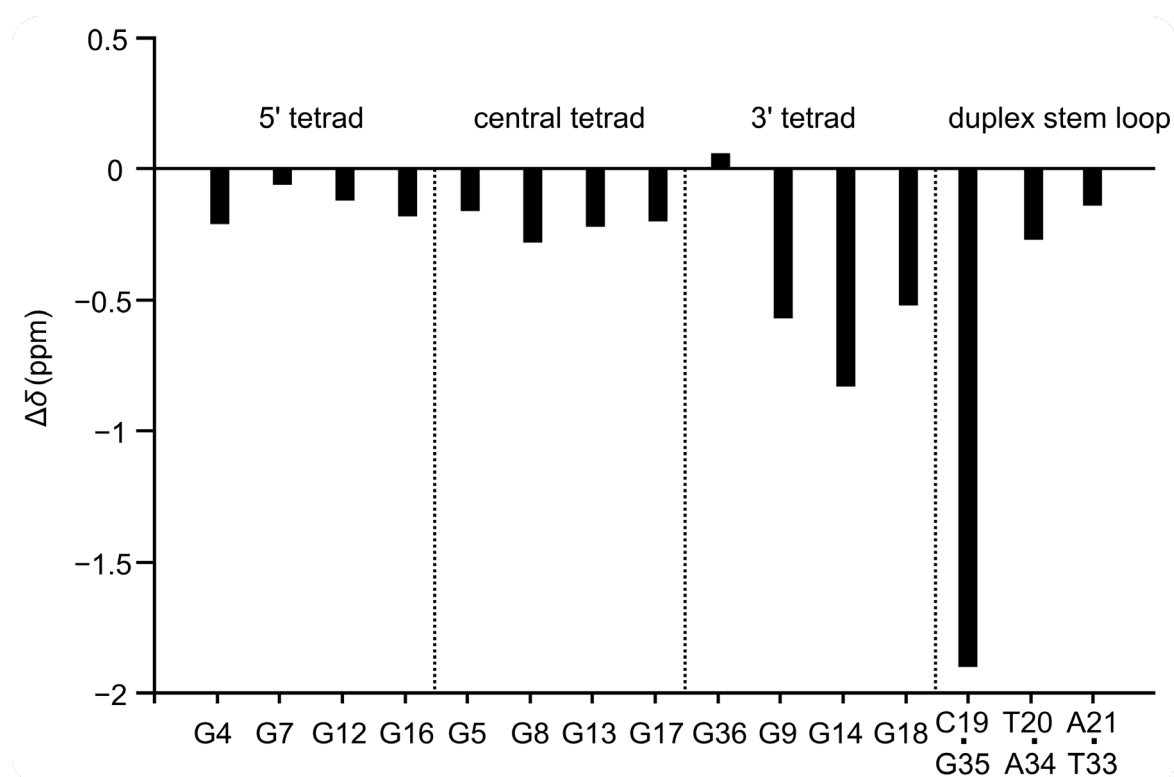
Table S5. ¹H and ¹³C NMR chemical shifts δ of a QD3-sbl - Phen-DC3 1:1 complex^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T1	7.33	n.d.	5.62	1.95/2.14	4.33	1.55	139.43	-
T2	7.11	n.d.	5.63	1.58/2.01	4.44	1.44	139.35	-
A3	7.96	-	5.68	2.54/2.58	4.77	7.79	142.54	n.d.
G4	7.85	11.24	5.99	2.49/2.87	4.87	-	137.87	-
G5	7.36	11.16	5.98	2.45/2.37	4.97	-	137.60	-
T6	7.81	n.d.	6.39	2.40/2.61	4.98	1.94	140.21	-
G7	8.09	11.50	6.01	2.62/2.83	4.86	-	138.52	-
G8	7.72	11.10	6.05	2.59/2.84	5.01	-	138.04	-
G9	7.68	10.74	6.10	2.82/2.79	5.10	-	137.77	-
T10	7.70	n.d.	6.27	2.27/2.48	4.77	1.98	140.17	-
A11	8.56	-	6.62	2.99/2.90	5.07	8.37	143.50	154.37
G12	7.92	11.66	6.05	2.55/2.85	4.90	-	138.27	-
G13	7.66	11.08	6.12	2.55/2.91	5.00	-	138.01	-
G14	7.67	10.46	6.10	2.81/2.76	5.21	-	138.01	-
T15	7.84	n.d.	6.48	2.47/2.64	5.03	2.00	139.94	-
G16	7.79	11.08	5.95	2.32/2.73	4.96	-	137.80	-
G17	7.66	11.00	5.96	2.57/2.85	5.05	-	138.02	-
G18	7.77	10.41	5.94	2.85/2.80	5.19	-	138.13	-
C19	7.62	-	5.71	2.20/2.25	4.80	5.58	143.24	-
T20	7.21	13.25	5.31	2.03/2.25	4.80	1.52	139.46	-
A21	8.16	-	5.94	2.74/2.82	4.98	7.14	141.81	153.84
G22	7.50	12.94	5.69	2.36/2.58	4.75	-	137.11	-
T23	7.09	13.52	5.83	1.91/2.31	4.69	1.19	139.04	-
C24	7.28	-	5.58	1.71/2.18	4.70	5.57	143.55	-
A25	8.13	-	6.14	2.66/2.51	4.93	7.86	142.21	154.89
T26	7.49	n.d.	5.99	2.03/2.29	4.97	1.74	139.36	-
T27	7.31	n.d.	5.71	1.91/2.17	4.58	1.57	139.42	-
T28	7.37	n.d.	5.95	2.14/2.35	4.67	1.59	139.57	-
T29	7.33	n.d.	5.70	2.05/2.37	4.66	1.77	139.48	-
G30	7.87	12.55	5.50	2.61/2.70	4.91	-	138.37	-
A31	8.12	-	6.14	2.59/2.77	4.92	7.72	142.16	155.03
C32	7.09	-	5.62	1.67/2.18	4.52	5.05	141.83	-
T33	7.01	13.46	5.09	1.66/1.83	4.49	1.28	139.45	-
A34	7.51	-	5.58	2.10/2.30	4.69	7.22	141.37	153.93
G35	7.05	10.94	5.33	2.44/2.80	4.99	-	136.98	-
G36	6.87	10.70	6.13	3.25/2.55	5.00	-	140.61	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

Table S6. ^1H NMR chemical shifts δ (ppm) of Phen-DC3 bound to *QD3-sb*^a

H1	H2	H3	H4	H5	6-CH ₃	H7	H8	H9	H10	H11
7.81	8.20	7.53	11.76	10.09	4.49	7.99	8.05	7.84	7.71	8.80
H1'	H2'	H3'	H4'	H5'	6-CH ₃ '	H7'	H8'	H9'	H10'	H11'
7.82	8.16	7.55	11.76	10.27	4.69	8.18	7.95	7.73	7.77	8.89

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.**Figure S10.** Chemical shift perturbations $\Delta\delta$ of imino protons in *QD3-sb* upon binding the Phen-DC3 ligand (120 mM K^+).

Resonance assignments of free *Q3-sbl2* and of its complex with Phen-DC3

Non-interrupted base-sugar NOEs can be traced from T1 to G5 and *syn*-G22, demonstrating a first truncated G-column with a broken G-tract and the open position filled by the 3'-terminal G22 (S11C). In analogy to *QD3-sbl*, second and third G-columns can be distinguished by following contacts of A11 in the second propeller loop. While a continuous NOE walk can be traced from G16 to T19, a long-range contact between G20 and G18 identifies the fourth G-column and the TGT lateral snapback loop. The presence of a single *syn*-guanosine at position 22 is additionally corroborated by its typical ¹³C8 chemical shift in ¹H-¹³C HSQC spectra (Figure S11B). Based on a parallel quadruplex, imino protons were unambiguously assigned without specific isotope labeling by intra- and inter-tetrad H8-imino connectivities and further confirmed by their sequential imino-imino contacts (Figure S11D,E). Also, a NOE crosspeak between the G22 imino and the T21 H6 proton demonstrates positioning of the snapback loop above the 3'-tetrad, effectively protecting an observable G22 amino proton from solvent exchange.

Assignments for the quadruplex in the 1:1 complex with Phen-DC3 closely follows the assignments of the free *Q3-sbl2*, demonstrating a conserved parallel topology, a lateral snapback loop, homopolar tetrad stacking, and a G22 amino proton protected from solvent exchange (Figure S12). Various contacts from Phen-DC3 quinoline protons to the quadruplex can be observed, including contacts to all G imino protons in the 5'-tetrad but also to T1 H1' in the 5'-overhang. These unambiguously show that the Phen-DC3 ligand stacks onto the 5'-face of the quadruplex (Figure S12C-F). However, some unidentified crosspeaks to protons of the 3'-tetrad also suggest small amounts of a minor complex with a putative Phen-DC3 binding at the 3'-tetrad (Figure S12D). Interestingly, in contrast to Phen-DC3 intercalated at the Q-D junction of *QD3-sbl*, symmetry-related ligand protons are subject to chemical exchange through a flip of the ligand as shown by exchange crosspeak in a ROESY experiment (Figure S13).

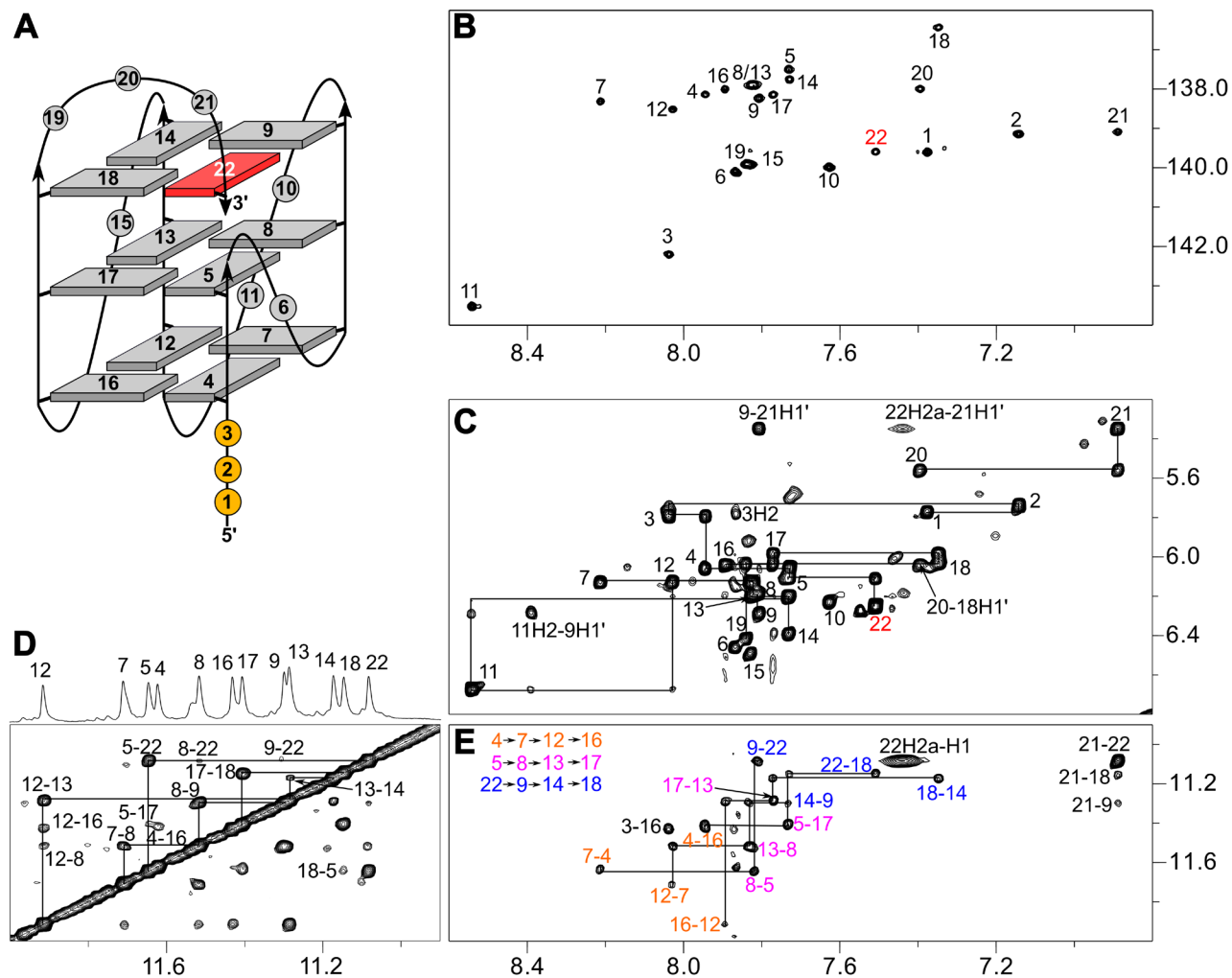


Figure S11. (A) Topology of Q3-sbI2. (B) ^1H - ^{13}C HSQC spectrum of Q3-sbI2 with H6/H8(ω_2)-C6/C8(ω_1) correlations; a single *syn*-G22 is labeled in red. (C) H6/H8(ω_2)-H1'(ω_1) NOESY spectral region of Q3-sbI2 with the strong intra-nucleotide crosspeak of *syn*-G22 labeled in red. (D) Imino(ω_2)-imino(ω_1) NOESY spectral region and corresponding 1D spectrum with assignments (top). (E) H6/H8(ω_2)-imino(ω_1) NOESY spectral region with intra-tetrad contacts labeled in different color for each tetrad; the latter determine tetrad polarities and are indicative of exclusive homopolar stacking for Q3-sbI2; lines trace sequential connectivities between tetrads. NOESY (300 ms mixing time) and HSQC spectra were acquired at 30 °C in 10 mM potassium phosphate buffer, pH 7.0, with a strand concentration of 0.75 mM.

Table S7. ^1H and ^{13}C NMR chemical shifts δ for free Q3-sb/2^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T1	7.38	n.d.	5.77	2.04/2.24	4.45	1.64	139.61	-
T2	7.14	n.d.	5.74	1.66/2.12	4.54	1.54	139.15	-
A3	8.04	-	5.79	2.57/2.66	4.85	7.87	142.22	153.93
G4	7.94	11.63	6.06	2.56/2.89	5.02	-	138.14	-
G5	7.73	11.64	6.11	2.76/2.45	5.07	-	137.50	-
T6	7.87	n.d.	6.46	2.43/2.67	5.07	1.97	140.10	-
G7	8.21	11.71	6.13	2.87/2.87	4.89	-	138.34	-
G8	7.82	11.52	6.19	2.62/2.84	5.00	-	137.91	-
G9	7.81	11.30	6.29	2.66/2.50	4.98	-	138.23	-
T10	7.63	n.d.	6.23	2.20/2.44	4.68	1.90	139.99	-
A11	8.54	-	6.67	3.07/2.92	5.15	8.39	143.53	155.12
G12	8.03	11.91	6.13	2.57/2.98	4.99	-	138.52	-
G13	7.83	11.29	6.20	2.66/3.00	5.00	-	137.91	-
G14	7.73	11.17	6.39	2.68/2.53	5.12	-	137.75	-
T15	7.83	n.d.	6.49	2.45/2.66	5.06	1.96	139.92	-
G16	7.89	11.43	6.04	2.39/2.80	5.04	-	138.00	-
G17	7.77	11.41	5.98	2.56/2.75	5.04	-	138.15	-
G18	7.35	11.15	6.04	2.68/2.67	5.10	-	136.44	-
T19	7.84	n.d.	6.42	2.37/2.63	4.75	1.96	139.91	-
G20	7.40	n.d.	5.56	1.89/2.16	4.59	-	138.00	-
T21	6.89	n.d.	5.35	1.76/2.31	4.48	1.63	139.10	-
G22	7.51	11.08	6.25	3.12/2.59	4.95	-	139.60	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

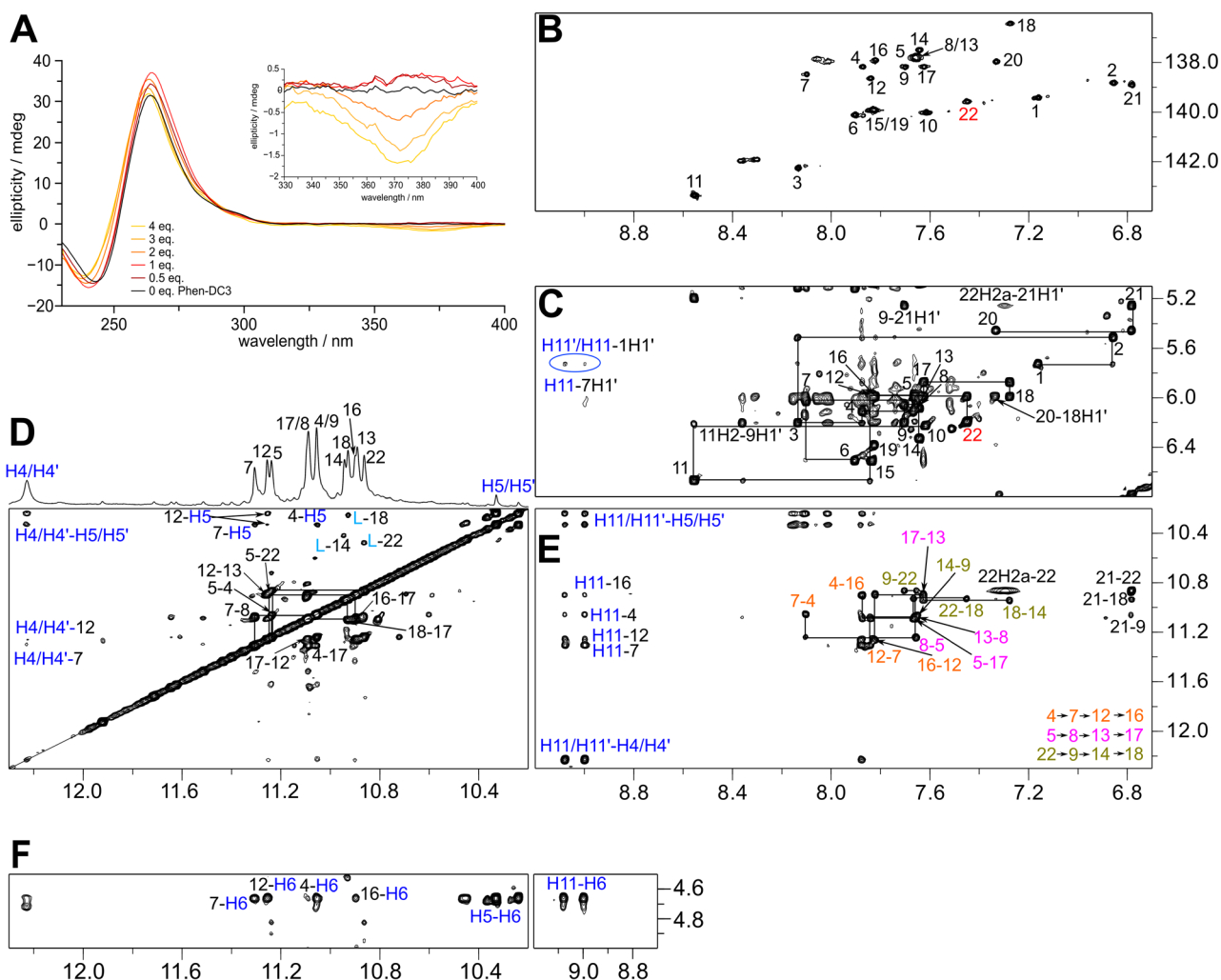


Figure S12. (A) CD spectra of Q3-*sbl2* upon its titration with Phen-DC3; the inset shows the y-expanded ligand absorption region. (B) ^1H - ^{13}C HSQC spectrum of the 1:1 Q3-*sbl2* - Phen-DC3 complex with H6/H8(ω_2)-C6/C8(ω_1) correlations; *syn*-G22 is labeled in red. (C-F) NOESY spectrum (300 ms mixing time) of a 1:1 Q3-*sbl2* - Phen-DC3 complex with ligand resonances labeled in blue. (C) H6/H8(ω_2)-H1'(ω_1) spectral region with the intra-nucleotide crosspeak of *syn*-G22 labeled in red; two representative crosspeaks between Phen-DC3 protons and 5'-overhang residues are circled. (D) Imino(ω_2)-imino(ω_1) spectral region; unassigned ligand resonances L with NOE contacts to the 3'-tetrad, suggesting the presence of a minor complex, are labeled in light blue; a corresponding 1D spectrum with assignments is shown on top. (E) H6/H8(ω_2)-imino(ω_1) spectral region with intra-tetrad contacts labeled in different color for each tetrad; lines trace sequential connectivities between tetrads. (F) Crosspeaks of Phen-DC3 *N*-methyl protons H6(ω_1) with G4 imino and Phen-DC3 protons. Spectra were acquired at 30 °C in 10 mM potassium phosphate buffer, pH 7.0, and a strand concentration of 0.75 mM.

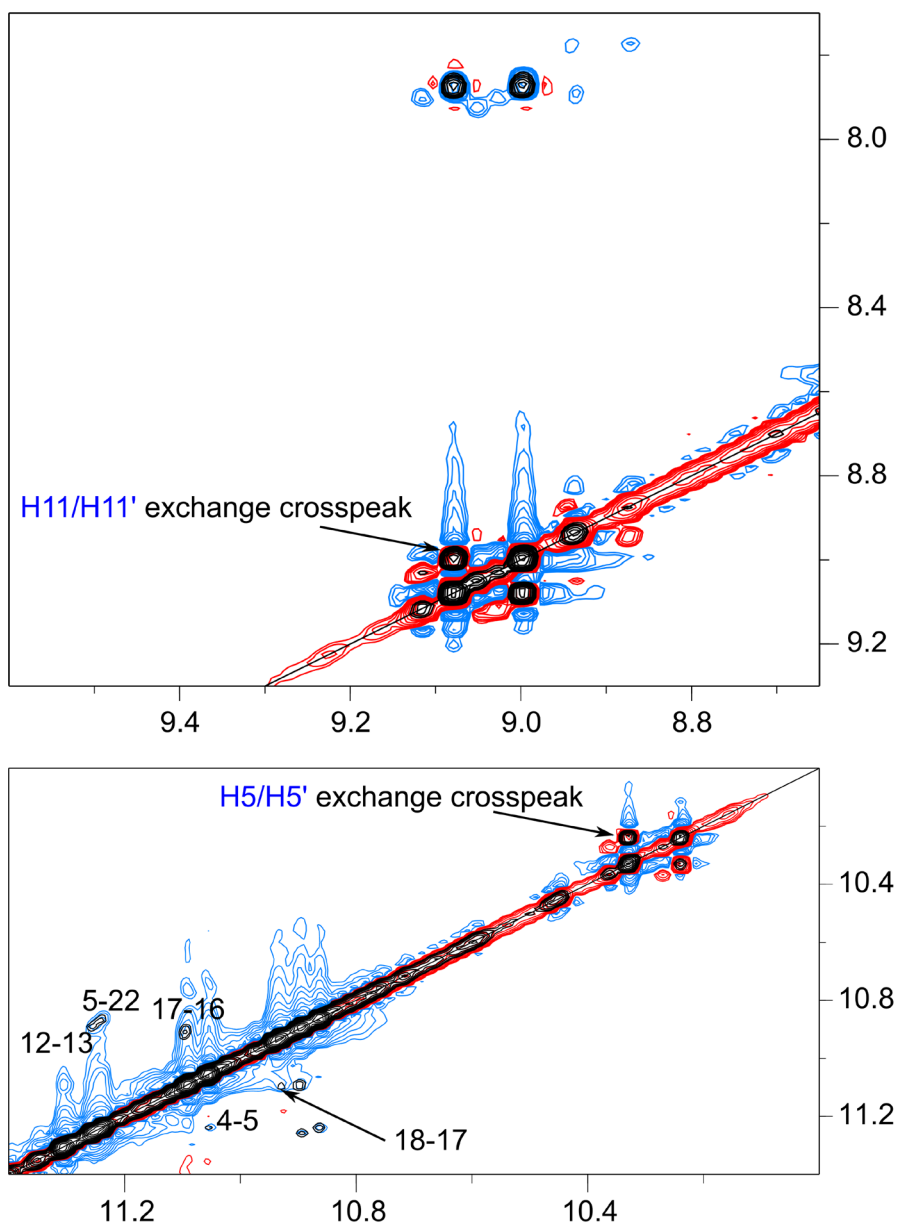


Figure S13. Superposition of NOESY (300 ms mixing time) and ROESY spectral regions for the 1:1 complex of Q3-*sbl2* (0.75 mM) with Phen-DC3. ROESY exchange crosspeaks of the ligand with their positive sign are indicated. Spectra were acquired at 30 °C in 10 mM potassium phosphate buffer, pH 7.0. Signals in the NOESY spectrum are colored black whereas positive and negative signals in the ROESY spectrum are given in red and blue color, respectively.

Table S8. ^1H and ^{13}C NMR chemical shifts δ for a Q3-sb/2 - Phen-DC3 1:1 complex^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
T1	7.16	n.d.	5.73	1.84/2.11	4.41	1.52	139.43	-
T2	6.86	n.d.	5.52	1.47/1.92	4.50	1.36	138.83	-
A3	8.14	-	6.20	2.88/2.96	5.11	7.75	142.25	155.03
G4	7.87	11.06	6.11	2.51/2.89	5.02	-	138.19	-
G5	7.67	11.24	5.99	2.76/2.43	5.07	-	137.81	-
T6	7.90	n.d.	6.51	2.47/2.70	5.12	2.00	140.13	-
G7	11.31	8.10	6.03	2.60/2.87	4.91	-	138.49	-
G8	11.09	7.66	6.06	2.51/2.78	4.95	-	137.79	-
G9	11.06	7.70	6.21	2.60/2.45	4.95	-	138.19	-
T10	7.62	n.d.	6.23	2.18/2.43	4.69	1.89	140.01	-
A11	8.56	-	6.67	3.07/2.95	5.20	8.36	143.39	155.32
G12	7.84	11.26	5.99	2.48/2.91	5.03	-	138.67	-
G13	7.66	10.89	6.08	2.57/2.93	4.98	-	137.81	-
G14	7.64	10.94	6.33	2.63/2.50	5.08	-	137.50	-
T15	7.84	n.d.	6.51	2.46/2.68	5.07	1.96	139.92	-
G16	7.82	10.90	5.99	2.25/2.77	5.08	-	137.93	-
G17	7.63	11.09	5.87	2.47/2.66	5.00	-	138.18	-
G18	7.28	10.93	5.99	2.62/2.62	5.06	-	136.44	-
T19	7.83	n.d.	6.39	2.34/2.59	4.34	1.95	139.91	-
G20	7.33	n.d.	5.46	1.82/2.09	4.52	-	137.98	-
T21	6.78	n.d.	5.26	1.69/2.24	4.43	1.51	138.92	-
G22	7.45	10.86	6.20	3.04/2.51	4.83	-	139.57	-

^aAt 30 °C in 10 mM potassium phosphate buffer, pH 7.0.

Resonance assignments of free *QD2-I* and of its complex with PIQ-4m

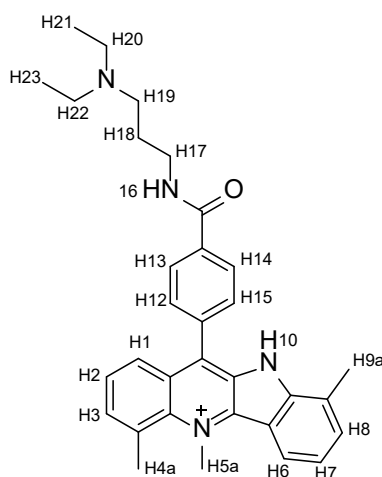
The imino proton spectral region of the *QD2-I* hybrid shows six and eight imino resonances with chemical shifts typical of Watson-Crick base pairs and of more upfield shifted G iminos involved in Hoogsteen hydrogen bonds, respectively. Identifying four *syn-anti* steps along the G-columns, a two-layered antiparallel G-quadruplex can be established. The first and third G-columns were assigned based on continuous sugar-base NOE contacts to the following TT lateral loops. The second G-column is identified due to non-interrupted NOE connectivities from *syn*-G5 along the duplex stem loop up to G21 (Figure S14C,D). Various contacts at the Q-D interface including G21 H8 to G22 H1 and C7 H4 to G6 H1 positions the duplex stem loop coaxially with the G-core. A typical heteropolar stacking pattern can be observed through strong non-sequential imino-imino contacts such as between G26 H1 and G22 H1. G-tetrad polarity following hydrogen bond donor to acceptor runs along G1→G6→G22→G27 and G2→G26→G23→G5. Intra-tetrad NOE contacts between G imino and G H8 protons of either *syn*- or *anti*-Gs determine the quadruplex groove width. Thus, the two TT lateral loops bridge a narrow groove while the duplex stem loop bridge the wide groove of the quadruplex.

For the 1:1 complex of *QD2-I* with the PIQ-4m ligand, similar sugar-base NOE connectivities as found for the free hybrid identify the first and third G-column followed by the two TT lateral loops and the duplex hairpin loop with uninterrupted NOE connectivities from C7 to G21. However, interruption of sequential contacts at the junction from G6 to C7 indicates ligand intercalation. *Syn*-guanines were assigned by their downfield-shifted ¹³C8 resonance while three adenine H2 resonances were identified by their H2-C2 correlations in a ¹H-¹³C HSQC spectrum (Figure S16C). Stereospecific assignments of H2'/H2'' protons were based on a NOESY experiment with short mixing times (80 ms) and the following determination of sugar conformations made use of the pattern and intensity of H1'-H2' and H1'-H2'' crosspeaks in a DQF-COSY spectrum (Figure S17). Except for T3 and T24, located in lateral loops bridging the narrow groove, all assigned sugar puckers are in the *south* domain of the pseudorotational cycle. Imino protons were assigned by following exchange crosspeaks between the free and complexed *QD2-I* hybrid (Figure S15) and heteropolar tetrad stacking was confirmed by characteristic intra- and inter-tetrad H8-imino NOE contacts. Taken together, the two-layered antiparallel topology of free *QD2-I* with exclusive *syn-anti* steps along the G-columns, two TT lateral loops bridging a narrow groove, and one duplex stem loop bridging a wide groove was retained after ligand addition. However, sequential contacts bridging the quadruplex-duplex interface were lost.

Ligand proton resonances were assigned by a combination of COSY and NOESY experiments. Spin systems with corresponding COSY correlations derive from protons in the phenyl and the fused indole and quinoline ring systems (Figure S18F). Discrimination of quinoline and indole resonances was enabled by a strong NOE crosspeak from a phenyl proton to quinoline H1 and H2 protons. Additional NOE crosspeaks of methyl substituents of the quinoline and indole moiety were observed to ring protons in their proximity, with the quinoline *N*-methyl proton resonating close to the water signal (Figure S18D,E). A strong NOE contact connects the indole NH with a resonance at about 8 ppm (Figure S18G). TOCSY and ROESY experiments identified the latter as being two isochronous ortho-positioned phenyl protons, explaining the observation of only a single scalar coupled proton pair of the phenyl ring in a DQF-COSY spectrum. The NH16 amide is assigned by following NOE contacts from adjacent phenyl protons with connectivities continuing to the aliphatic

side chain (Figures S18F, S19A). Methylene protons H17 adjacent to the amide were found to be non-equivalent, indicating a restricted C-C bond rotation upon binding. Other resonances of the aliphatic side chain were identified through their mutual scalar couplings observed in COSY and TOCSY experiments (Figure S19B). It should be noted that NOE crosspeaks for aliphatic side chain protons are rather weak and broadened due to a high flexibility towards the terminus with changing signs for NOE crosspeaks of terminal ethyl protons (not shown).

Various intermolecular ligand-DNA contacts were observed in representative NOESY spectral regions, yielding a total of 41 NOESY-derived distance restraints (Figures S16 and S18). In fact, indole H1/H2/H3 protons primarily show NOE contacts to G6 and C7 while quinoline H6/H7/H8 protons show contacts to G21 and G22 residues. Quinoline methyl substituents H4a and H5a are connected to G6 and G21 iminos (Figure S18A,B). Interestingly, indole methyl protons H9a located on the opposite side of the indoloquinoline together with its extended side chain feature a contact to the exposed G1 imino not covered by the duplex stem loop (Figure S18A). Additionally, phenyl protons, in particular H14 and H15, show NOE contacts to all imino protons within the G-tetrad at the Q-D interface. Such a NOE pattern suggests the indoloquinoline to be sandwiched at the Q-D junction with its side chain projected towards the exposed part of the interfacial G-tetrad.



Chemical structure of PIQ-4m with atom numbering as used in this study.

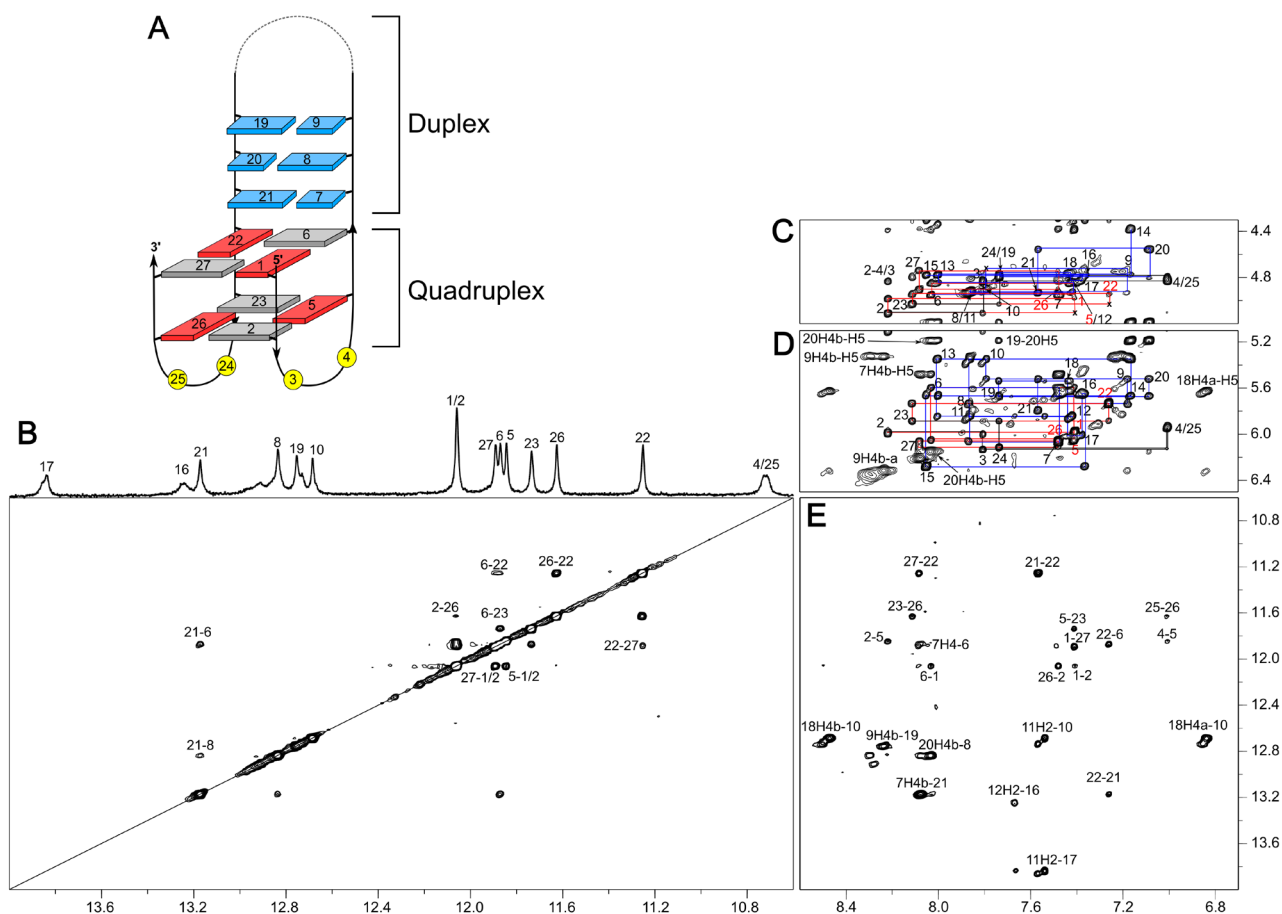


Figure S14. (A) Topology and (B-E) NOESY spectral regions of the *QD2-I* hybrid. (B) Imino(ω_2)-imino(ω_1) crosspeaks and corresponding 1D spectrum with assignments on top. (C) H6/H8(ω_2)-H3'(ω_1) and (D) H6/H8(ω_2)-H1'(ω_1) spectral region; continuous NOE connectivities in the quadruplex and duplex domain are traced by black and blue lines, respectively. Characteristic rectangular patterns for *syn-anti* steps and labels of *syn*-residues G1, G5, G22, and G26 with their strong intra-nucleotide H8-H1' crosspeak are shown in red. Two NOE contacts in (C) only observed at lower threshold levels are marked by a cross. (E) H6/H8(ω_2)-imino(ω_1) crosspeaks of the G-core and adenine H2/cytosine amino(ω_2)-imino(ω_1) crosspeaks of the duplex domain. NOESY spectra (300 ms mixing time) were acquired with a *QD2-I* concentration of 0.64 mM at 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0.

Note: A three-dimensional NMR structure of the same sequence under different temperature and buffer conditions has been published previously (PDB 2M8Z).² It matches the Q-D hybrid topology as derived here from analysis of the NMR spectral data collected under the present conditions. Assignments of T4/T25 imino protons lacking NOE crosspeaks in the present study is based on a comparison with the previously published structural analysis.

2. Lim, K. W.; Phan, A. T. Structural Basis of DNA Quadruplex-Duplex Junction Formation. *Angew. Chem. Int. Ed.* 2013, 52, 8566–8569.

Table S9. ¹H NMR chemical shifts δ for free QD2-^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2'' ^b	H3'	H5/H2/Me
G1	7.41	12.06	5.98	2.90/2.90	4.98	-
G2	8.22	12.06	6.00	2.35/3.00	5.11	-
T3	7.81	n.d.	6.13	2.15/2.51	4.83	1.94
T4	7.01	n.d.	5.94	1.98/2.58	4.83	0.89
G5	7.41	11.85	6.05	2.98/3.39	4.85	-
G6	8.03	11.87	5.60	2.40/2.64	4.95	-
C7	7.48	-	6.06	1.88/2.40	4.95	5.48
G8	7.87	12.84	5.74	2.58/2.64	4.93	-
C9	7.18	-	5.52	1.76/2.22	4.72	5.33
G10	7.79	12.69	5.34	2.56/2.65	4.91	-
A11	7.86	-	5.84	2.25/2.57	4.93	7.54
A12	7.42	-	5.85	2.02/2.38	4.84	7.67
G13	8.00	n.d.	5.34	2.31/2.60	4.78	-
C14	7.17	-	5.66	1.56/2.02	4.38	5.19
A15	8.05	-	6.28	2.91/2.6	4.78	n.d.
T16	7.36	13.25	5.65	2.08/2.47	4.79	1.80
T17	7.38	13.84	6.01	2.15/2.44	4.83	1.59
C18	7.43	-	5.54	2.00/2.29	4.76	5.63
G19	7.74	12.75	5.67	2.43/2.48	4.80	-
C20	7.09	-	5.52	1.66/2.12	4.56	5.19
G21	7.57	13.17	5.79	2.55/2.94	4.93	-
G22	7.26	11.26	5.73	2.60/3.03	4.94	-
G23	8.11	11.74	5.89	2.26/2.87	5.03	-
T24	7.74	n.d.	6.11	2.08/2.46	4.80	1.88
T25	7.01	n.d.	5.94	1.97/2.54	4.80	0.83
G26	7.48	11.63	6.06	2.99/3.51	4.91	-
G27	8.08	11.89	6.11	2.38/2.62	4.74	-

^aAt 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0. ^bNo stereospecific assignment.

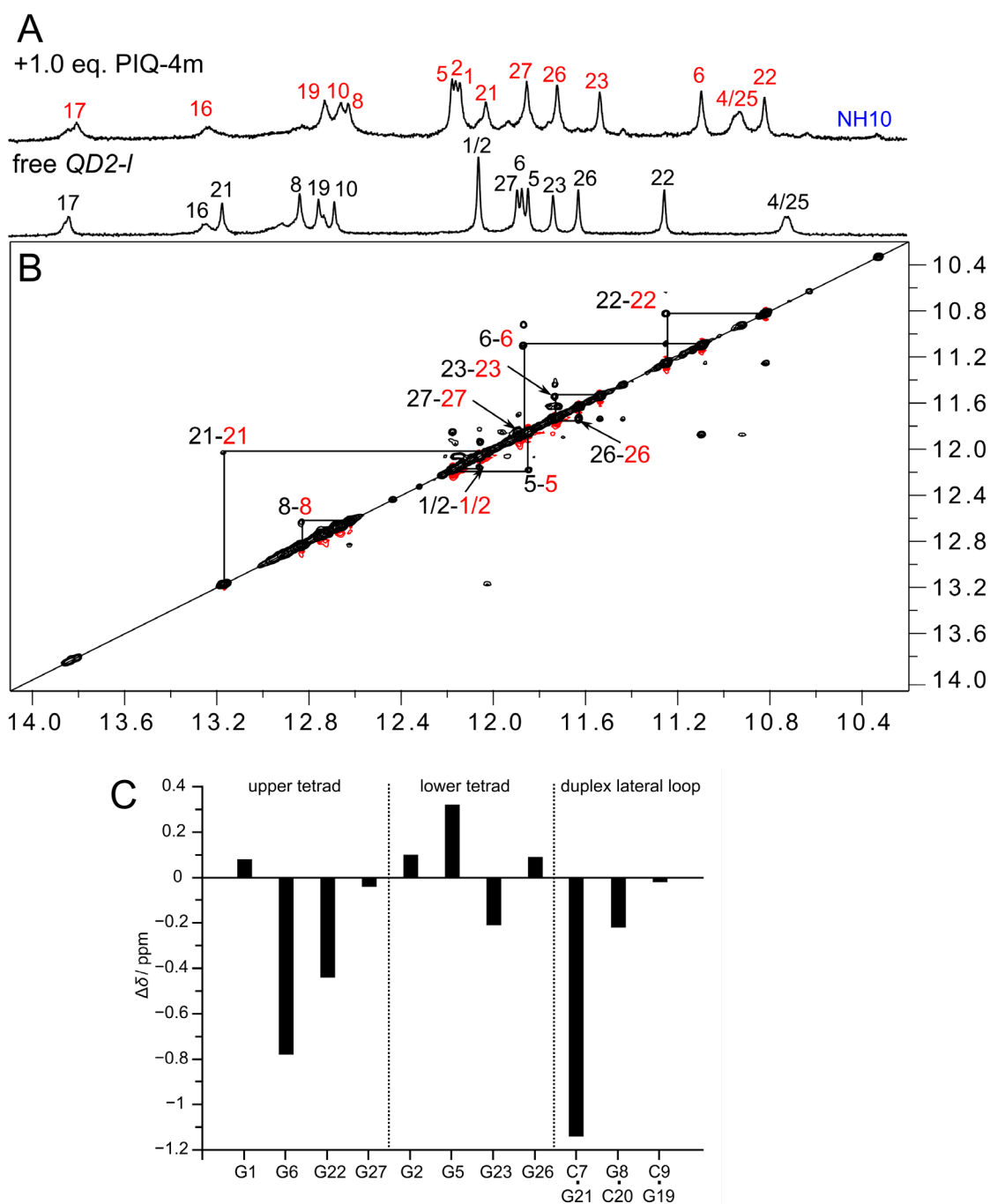


Figure S15. (A) 1D imino proton spectral region with resonance assignments of *QD2-I* before (bottom) and after addition of 1 equivalent of PIQ-4m (top). (B) Imino(ω_2)-imino(ω_1) spectral region of a ROESY spectrum for *QD2-I* with 0.5 equivalent of PIQ-4m; exchange crosspeaks of positive sign are colored black and signals of opposite sign colored red; imino proton resonances in free and ligand-bound *QD2-I* are labeled in black and red, respectively; NH10 represents a ligand resonance. Spectra were acquired at 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0, with a *QD2-I* concentration of 0.64 mM. (C) Imino proton chemical shift perturbations upon PIQ-4m binding to *QD2-I*. For a compilation of chemical shifts for the free Q-D hybrid and its complex with ligand see Tables S8-S10.

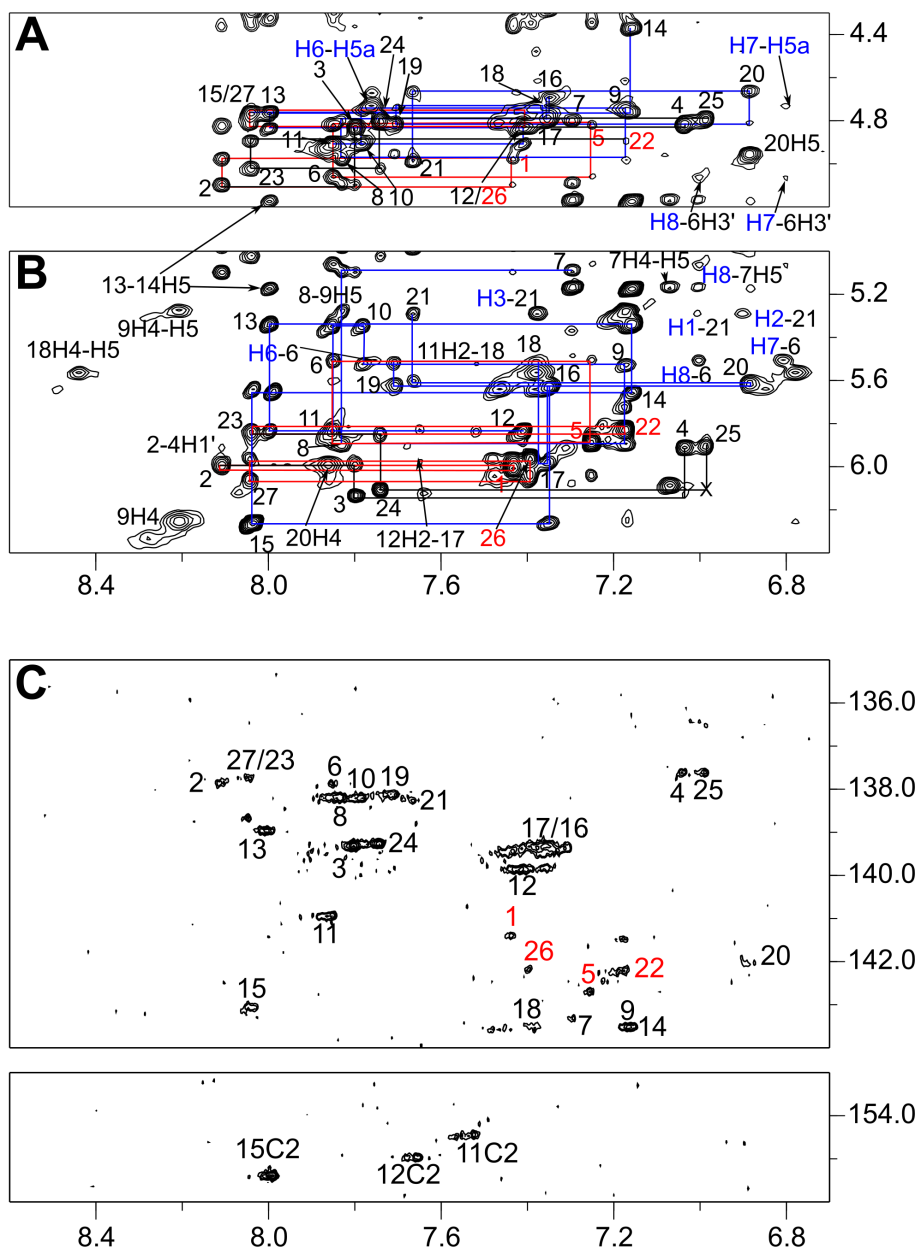


Figure S16. Resonance assignments for the 1:1 complex between QD2-I (0.64 mM) and PIQ-4m. (A) NOESY spectral region with H6/H8(ω_2)-H3'(ω_1) and (B) H6/H8(ω_2)-H1'(ω_1) crosspeaks. Sequential NOE connectivities for the quadruplex and duplex domain are traced by black and blue lines, respectively. Also shown are some ligand-DNA intermolecular contacts with ligand resonances labeled in blue. A T25 H6-T24 H1' crosspeak only observed at lower threshold levels is marked by a cross. (C) ^1H - ^{13}C HSQC spectrum showing H6/H8(ω_2)-C6/C8(ω_1) (top) and adenine H2(ω_2)-C2(ω_1) correlations (bottom). *Syn*-residues G1, G5, G22, and G26 exhibiting strong intra-nucleotide H8-H1' NOE crosspeaks and characteristic rectangular patterns for *syn-anti* steps (A,B) as well as more downfield shifted guanine $^{13}\text{C}8$ resonances (C) are labeled in red. Experiments were performed at 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0. NOESY spectra were acquired with a 300 ms mixing time.

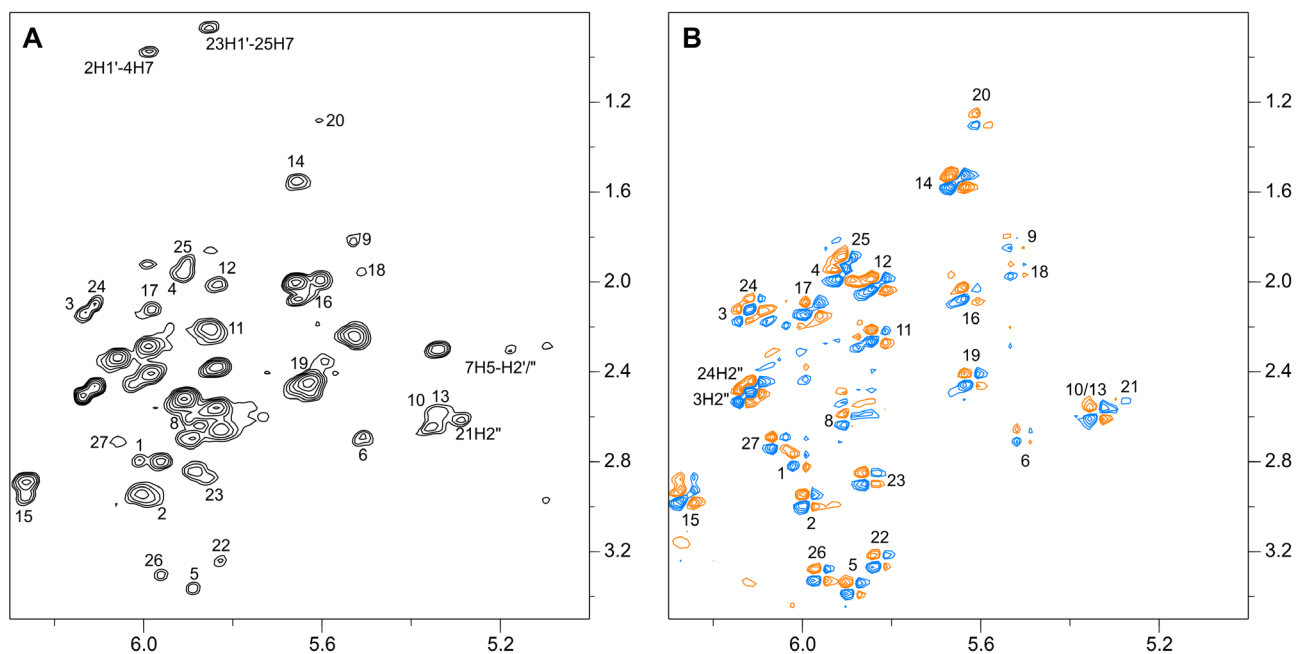


Figure S17. (A) NOESY spectrum (80 ms mixing time) and (B) DQF-COSY spectrum for a 1:1 complex of QD2-I (0.64 mM) with PIQ-4m, showing the H1'(ω_2)-H2'/H2''(ω_1) spectral region (30 °C, 20 mM potassium phosphate buffer, 100 mM KCl, pH 7). Following stereospecific H2'/H2'' assignments based on NOE crosspeak intensities, sugar conformations were evaluated through inspection of DQF-COSY crosspeak patterns, demonstrating a *north*-type sugar conformation for T3 and T24 with all other residues adopting a *south* sugar pucker.

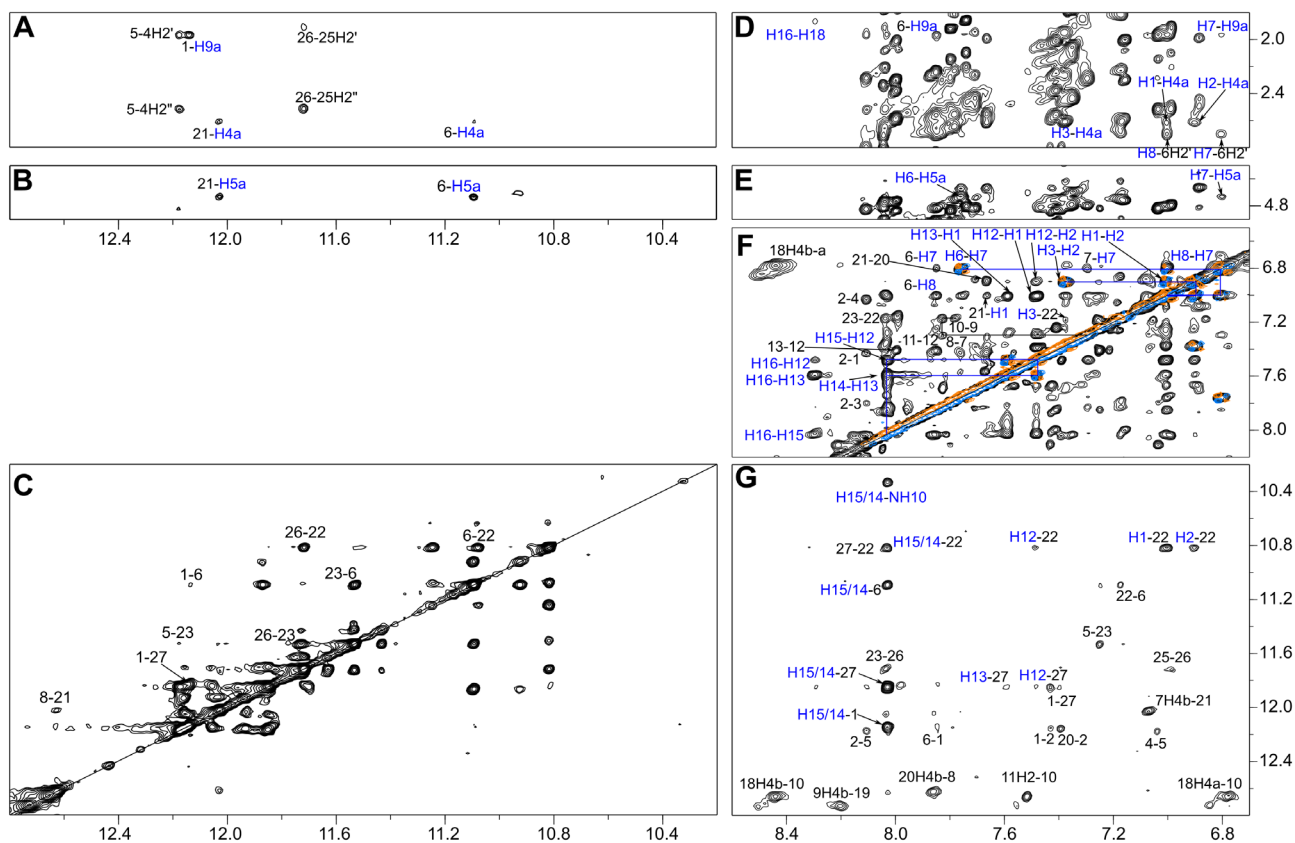


Figure S18. NOESY spectral regions for the 1:1 complex between *QD2-I* (0.64 mM) and PIQ-4m with assignments of ligand protons and intermolecular DNA-ligand contacts. (A,B) NOE contacts from imino(ω_2) to H2'/H2'' and to ligand methyl protons, also including more deshielded PIQ-4m *N*-methyl (H5a) protons. (C) Imino(ω_2)-imino(ω_1) spectral region. (D,E) NOE contacts between aromatic protons(ω_2) and H2'/H2''/methyl protons(ω_1). (F) 2D NOE spectral region with intra- and intermolecular crosspeaks between aromatic protons. For the unambiguous assignment of the PIQ-4m aromatic ring protons, a corresponding DQF-COSY spectrum (orange-blue for positive-negative signals) is superimposed on the NOESY spectral region (black). (G) NOE contacts between aromatic protons(ω_2) and imino protons(ω_1). DNA and ligand protons are labeled in black and blue, respectively. NOESY spectra (300 ms mixing time) were acquired at 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0.

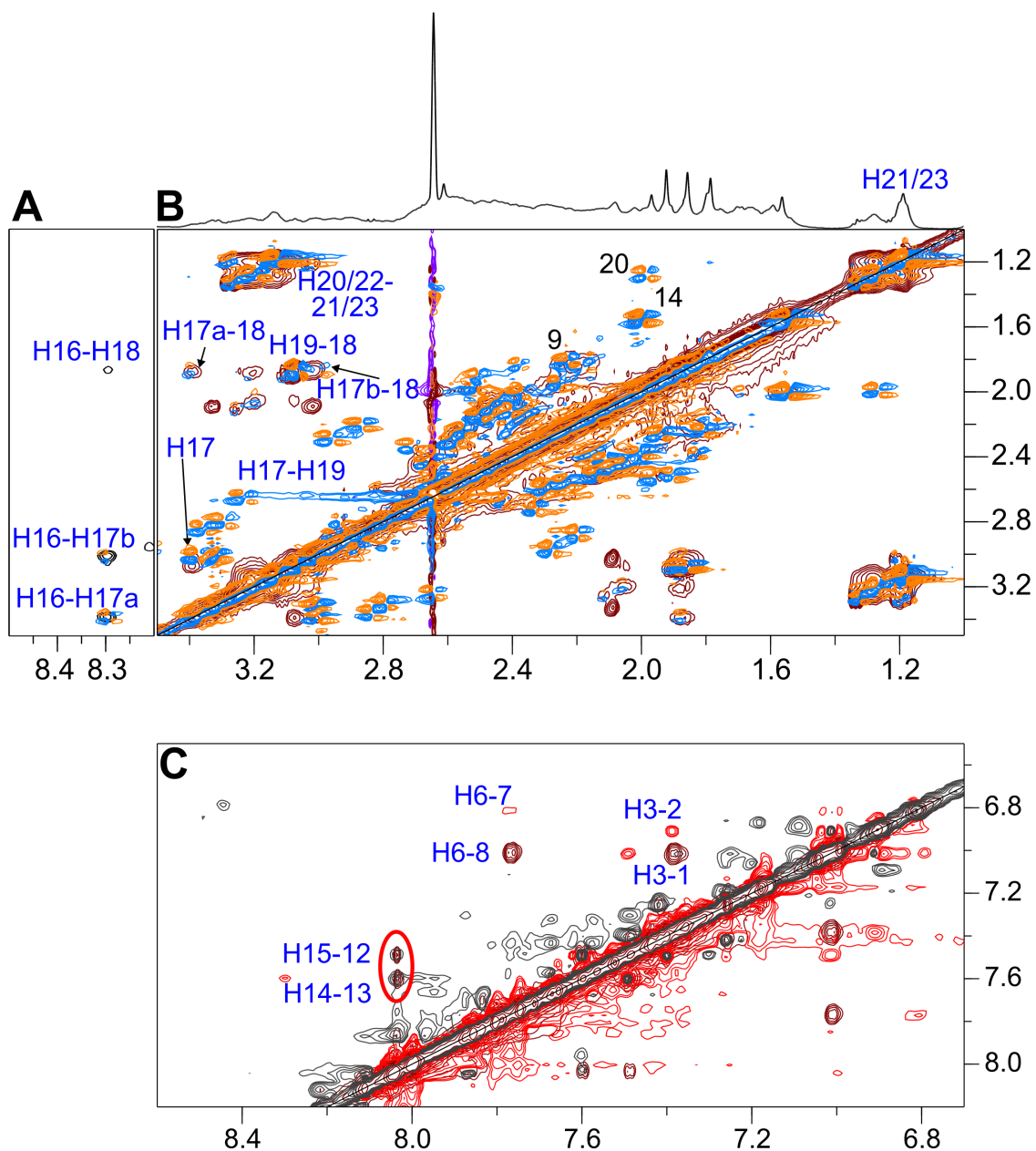


Figure S19. Proton assignments of the PIQ-4m side chain in a 1:1 complex with QD2-I. (A) Superposition of a NOESY (300 ms mixing time, black) and DQF-COSY spectrum (orange-blue for positive-negative signals) with correlations between amide NH16 and adjacent methylene protons. (B) Superposition of a DQF-COSY (orange-blue for positive-negative signals) and a TOCSY spectrum (brown), showing DNA intra-nucleotide H2'-H2'' crosspeaks and correlations between resonances of the ligand aliphatic side chain including terminal H21/H23 with their triplet fine structure (see 1D spectrum on top). (C) Superposition of TOCSY (brown) and ROESY spectrum (black-red for positive-negative signals) with the assignment of aromatic protons of PIQ-4m. Two exchange crosspeaks of isochronous phenyl protons H14/H15 with their symmetry-related H13/H12 protons are circled in red.

Table S10. ¹H and ¹³C NMR chemical shifts δ for a QD2-I - PIQ-4m 1:1 complex^a

δ / ppm	H8/H6	H1/H3	H1'	H2'/H2''	H3'	H5/H2/Me	C8/C6	C2
G1	7.43	12.14	6.01	2.80/2.94	4.98	-	141.40	-
G2	8.11	12.16	5.99	2.97/2.29	5.10	-	137.84	-
T3	7.80	n.d.	6.13	2.14/2.50	4.83	1.92	139.32	-
T4	7.04	n.d.	5.91	1.97/2.52	4.82	0.98	137.62	-
G5	7.25	12.17	5.89	3.36/2.84	4.82	-	142.70	-
G6	7.85	11.09	5.51	2.69/2.25	5.06	-	137.86	-
C7	7.30	-	5.09	2.29/2.29	4.80	5.17	143.31	-
G8	7.83	12.62	5.90	2.60/2.69	4.98	-	138.19	-
C9	7.17	-	5.53	1.82/2.23	4.75	5.28	143.51	-
G10	7.78	12.66	5.35	2.56/2.64	4.91	-	138.20	-
A11	7.85	-	5.84	2.24/2.56	4.91	7.52	140.93	154.46
A12	7.41	-	5.84	2.01/2.38	4.83	7.65	139.88	154.93
G13	8.00	n.d.	5.34	2.58/2.30	4.76	-	138.97	-
C14	7.16	-	5.66	1.55/2.00	4.37	5.17	143.50	-
A15	8.04	-	6.26	2.95/2.89	4.77	8.00	143.05	155.39
T16	7.35	13.23	5.63	2.06/2.44	4.69	1.78	139.35	-
T17	7.35	13.81	5.98	2.12/2.41	4.80	1.56	139.35	-
C18	7.38	-	5.52	1.95/2.26	4.74	5.56	143.50	-
G19	7.71	12.73	5.62	2.43/2.49	4.81	-	138.13	-
C20	6.89	-	5.61	1.28/1.99	4.67	4.96	142.04	-
G21	7.66	12.03	5.29	2.56/2.61	4.99	-	138.28	-
G22	7.17	10.82	5.83	3.24/2.66	4.89	-	142.21	-
G23	8.04	11.53	5.85	2.87/2.21	5.02	-	137.74	-
T24	7.74	n.d.	6.11	2.10/2.47	4.80	1.86	139.27	-
T25	6.99	n.d.	5.90	1.91/2.51	4.79	0.86	137.62	-
G26	7.39	11.72	5.96	3.30/2.80	4.82	-	142.19	-
G27	8.04	11.85	6.06	2.71/2.34	4.76	-	137.74	-

^aAt 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0.

Table S11. ^1H NMR chemical shifts δ (ppm) of PIQ-4m bound to QD2-*I*^a

H1	H2	H3	H4a	H5a	H6	H7	H8	H9a	10-NH
7.01	6.90	7.38	2.61	4.73	7.76	6.8	7.00	1.96	10.33
H12	H13	H14	H15	16-NH	H17	H18	H19	H20/H22	H21/H23
7.48	7.59	8.03	8.03	8.30	3.38/3.01	1.87	3.06	3.13	1.19

^aAt 30 °C in 20 mM potassium phosphate buffer, 100 mM KCl, pH 7.0.**Table S12.** Glycosidic torsion angles χ (°) at Q-D interfaces without and with an intercalated ligand

Q-D hybrid (PDB ID)	Q-D interface			
	χ 5'-G (G4)	χ C-3'	χ 5'-G	χ G-3' (G4)
free QD3- <i>sbl</i> (7PNE)	(G18) 246.1 ± 7.9	(C19) 248.5 ± 4.6	(G35) 251.1 ± 5.7	(<i>syn</i> -G36) 62.6 ± 4.3
QD3- <i>sbl</i> with SYUIQ-5 (7PNG)	(G18) 238.7 ± 11.4	(C19) 266.5 ± 6.6	(G35) 254.1 ± 14.8	(<i>syn</i> -G36) 66.1 ± 5.3
QD3- <i>sbl</i> with Phen-DC3 (8ABD)	(G18) 251.7 ± 6.5	(C19) 271.6 ± 6.0	(G35) 261.3 ± 7.5	(<i>syn</i> -G36) 78.2 ± 12.4
free QD2- <i>I</i> (2M8Z)	(G6) 241.2 ± 3.4	(C7) 237.1 ± 2.6	(G21) 241.4 ± 1.6	(<i>syn</i> -G22) 54.4 ± 1.2
QD2- <i>I</i> with PIQ-4m (8ABN)	(G6) 254.6 ± 2.8	(C7) 275.0 ± 1.8	(G21) 266.9 ± 0.7	(<i>syn</i> -G22) 68.1 ± 2.0

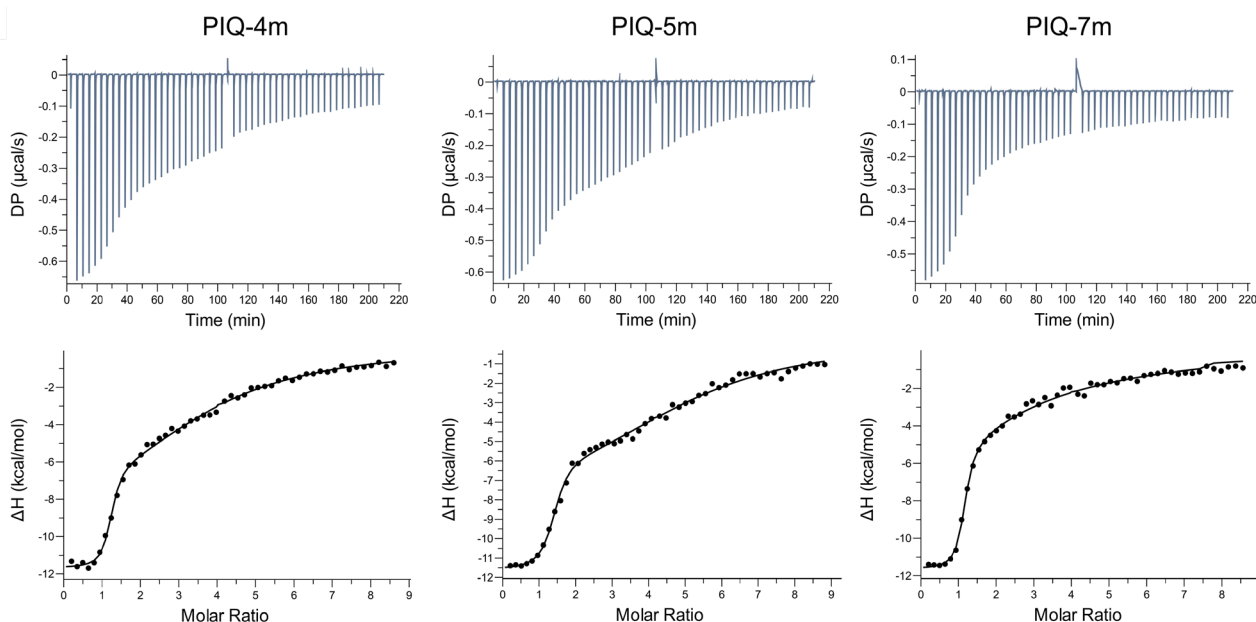
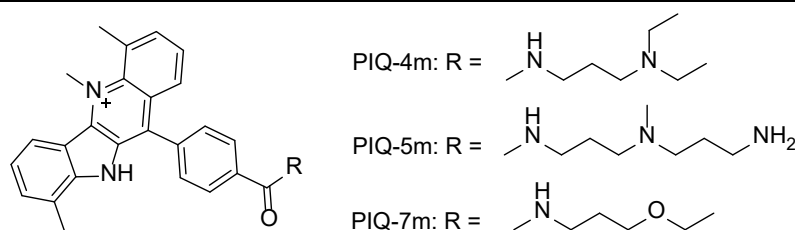


Figure S20. Representative ITC thermograms for the titration of PIQ-derivatives PIQ-4m, PIQ-5m, and PIQ-7m to the *QD2-I* hybrid at 40 °C (20 mM potassium phosphate buffer, pH 7.0, 100 mM KCl, 5 % DMSO). Upper and lower panels show the heat released for each injection step and the blank-corrected normalized heat versus molar ratio. Curves were fitted based on a model with two sets of binding sites.

Table S13. ITC-derived binding parameters of PIQ derivatives with different aliphatic side chains when binding to *QD2-I* at 40 °C^a



ligand	n	K_a (M^{-1})	ΔH° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol) ^b
PIQ-4m	1.1 ± 0.1	$8.3 \cdot 10^6 \pm 1.4 \cdot 10^6$	-11.5 ± 0.3	1.6 ± 0.3
	5.0 ± 0.6	$2.5 \cdot 10^4 \pm 1.2 \cdot 10^4$	-6.5 ± 0.5	0.2 ± 0.8
PIQ-5m	1.3 ± 0.1	$7.9 \cdot 10^6 \pm 0.9 \cdot 10^6$	-12.0 ± 0.4	2.2 ± 0.4
	5.0 ± 0.4	$5.2 \cdot 10^4 \pm 1.1 \cdot 10^4$	-6.2 ± 0.2	-0.5 ± 0.3
PIQ-7m	1.1 ± 0.1	$2.6 \cdot 10^6 \pm 0.7 \cdot 10^6$	-12.1 ± 0.3	2.9 ± 0.4
	6.0 ± 0.3	$8.7 \cdot 10^3 \pm 0.5 \cdot 10^3$	-5.3 ± 0.2	-0.3 ± 0.2

^aAverage values with root-mean-square deviations obtained from three independent measurements in 20 mM potassium phosphate buffer, pH 7.0, 100 mM KCl, 5% DMSO; data were fitted with two sets of binding sites with fit parameters of the second lower-affinity binding shown on a grey background. ^b $-T\Delta S^\circ = \Delta G^\circ - \Delta H^\circ$ with $\Delta G^\circ = -RT \ln K_a$.

Eigenständigkeitserklärung

Hiermit erkläre ich, dass diese Arbeit bisher von mir weder an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Greifswald noch einer anderen wissenschaftlichen Einrichtung zum Zwecke der Promotion eingereicht wurde.

Ferner erkläre ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die darin angegebenen Hilfsmittel und Hilfen benutzt und keine Textabschnitte eines Dritten ohne Kennzeichnung übernommen habe.

Lie, Yoanes Maria Vianney

Curriculum Vitae

Personal Data

Name: Lie, Yoanes Maria Vianney

Date of Birth: 07 April 1995

Place of Birth: Semarang, Indonesia

Education

June 2013 Graduated from high school at Karangturi Senior Highschool, Semarang, Indonesia

2013-2017 **S.Si. in Biology**, Universitas Surabaya, Surabaya, Indonesia

2017-2019 **M.Biotek. in Biotechnology**, Universitas Surabaya, Surabaya, Indonesia

From September 2019 **PhD student in Analytical Biochemistry**, Universität Greifswald, Greifswald, Germany under the guidance of Prof. Klaus Weisz

Selected Publications

Articles

1. Vianney, Y. M., & Weisz, K. (2022) High-affinity binding at quadruplex–duplex junctions: rather the rule than the exception. *Nucleic Acids Res.*, **50**, 11948-11964.
2. Jana, J., Vianney, Y. M., Schröder, N., & Weisz, K (2022). Guiding the folding of G-quadruplexes through loop residue interactions. *Nucleic Acids Res.*, **50**, 7161-7175.
3. Vianney, Y. M., & Weisz, K. (2022) Indoloquinoline Ligands Favor Intercalation at Quadruplex-Duplex Interfaces. *Chem. Eur. J.*, **28**, e202103718.
4. Mohr, S., Jana, J., Vianney, Y. M., & Weisz, K. (2021) Expanding the Topological Landscape by a G-Column Flip of a Parallel G-Quadruplex. *Chem. Eur. J.*, **27**, 10437-10447.
5. Jana, J., Mohr, S., Vianney, Y. M., & Weisz, K. (2021) Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.*, **2**, 338-353.
6. Vianney, Y. M., Purwanto, M. G. M., & Weisz, K. (2021) G-Quadruplex Formation in a Putative Coding Region of White Spot Syndrome Virus: Structural and Thermodynamic Aspects. *ChemBioChem*, **22**, 1932-1935.
7. Vianney, Y. M., & Weisz, K. (2021) First Tandem Repeat of a Potassium Channel KCNN4 Minisatellite Folds into a V-Loop G-Quadruplex Structure. *Biochemistry*, **60**, 1337-1346.

8. Vianney, Y. M., Preckwinkel, P., Mohr, S., & Weisz, K. (2020) Quadruplex–Duplex Junction: A High-Affinity Binding Site for Indoloquinoline Ligands. *Chem. Eur. J.*, **26**, 16910-16922.
9. Tang, Q., Vianney, Y. M., Weisz, K., Grathwol, C. W., Link, A., Bornscheuer, U. T., & Pavlidis, I. V. (2020). Influence of Substrate Binding Residues on the Substrate Scope and Regioselectivity of a Plant O-Methyltransferase against Flavonoids. *ChemCatChem*, **12**, 3721-3727.

Book Chapter

Vianney, Y. M., Jana, J., Schröder, N., & Weisz, K. (2023). Nucleic Acid Structure and Biology. In *Nucleic Acid Biology and its Application in Human Diseases* (pp. 1-34). Singapore: Springer Nature Singapore.

Deposited Structures

1. Vianney, Y.M., & Weisz, K. (2020). Structure of a parallel c-Myc modified with 3' duplex stem-loop overhang. PDB 6ZL2. <https://doi.org/10.2210/pdb6ZL2/pdb>
2. Vianney, Y.M., & Weisz, K. (2020). Structure of a parallel c-Myc modified with 5' duplex stem-loop overhang. PDB 6ZL9. <https://doi.org/10.2210/pdb6ZL2/pdb>
3. Vianney, Y.M., & Weisz, K. (2020). Structure of a parallel c-Myc modified with 5' duplex stem-loop and 3' diagonal snap-back loop. PDB 6ZTE. <https://doi.org/10.2210/pdb6ZL2/pdb>
4. Vianney, Y.M., & Weisz, K. (2021). G-quadruplex with V-shaped loop from the first repeat of KCNN4 minisatellite. PDB 7ATZ. <https://doi.org/10.2210/pdb6ZL2/pdb>
5. Mohr, S., Vianney, Y.M., & Weisz, K. (2021). Hybrid-2R quadruplex-duplex with (-p-p-l) topology and 3 syn residues. PDB 7O1H. <https://doi.org/10.2210/pdb6ZL2/pdb>
6. Vianney, Y.M., & Weisz, K. (2022). Parallel Q-D hybrid with 3' duplex stem-loop as a lateral snapback loop. PDB 7PNE. <https://doi.org/10.2210/pdb6ZL2/pdb>
7. Vianney, Y.M., & Weisz, K. (2022). Solution structure of 1:1 complex of an indoloquinoline derivative SYUIQ-5 to parallel quadruplex-duplex (Q-D) hybrid. PDB 7PNG. <https://doi.org/10.2210/pdb6ZL2/pdb>
8. Jana, J., Vianney, Y.M., Schroeder, N., & Weisz, K. (2022). Structure of a parallel G-quadruplex with a snapback loop. PDB 7ZEM. <https://doi.org/10.2210/pdb6ZL2/pdb>
9. Jana, J., Vianney, Y.M., Schroeder, N., & Weisz, K. (2022). Structure of a hybrid-type G-quadruplex with a snapback loop and an all-syn G-column (hybrid-1R). PDB 7ZEO. <https://doi.org/10.2210/pdb6ZL2/pdb>

10. Jana, J., Vianney, Y.M., Schroeder, N., & Weisz, K. (2022). Structure of a hybrid-type G-quadruplex with a snapback loop (hybrid-1R'). PDB 7ZEK. <https://doi.org/10.2210/pdb6ZL2/pdb>
11. Vianney, Y.M., & Weisz, K. (2022). Solution structure of Phen-DC3 intercalating into a quadruplex-duplex hybrid. PDB 8ABD. <https://doi.org/10.2210/pdb6ZL2/pdb>
12. Vianney, Y.M., & Weisz, K. (2022). Solution structure of a phenyl-indoloquinoline intercalating into a quadruplex-duplex hybrid. PDB 8ABN. <https://doi.org/10.2210/pdb6ZL2/pdb>

Conferences

1. Vianney, Y.M., & Weisz, K.: Quadruplex-duplex junctions as targets for small molecules. Nucleic Acid Secondary Structures: G4s and Beyond- G4 webinar series, 9th February 2023.
2. Vianney, Y.M., & Weisz, K.: Quadruplex-duplex junctions as binding hotspots: intercalation drives small-molecule recognition. G4thering, 8th International Meeting on Quadruplex Nucleic Acids, Marienbad, Czech Republic.
3. Vianney, Y.M., Kandinata, P., Weisz, K., & Goretti, M.G.M. (2023). Identification of DNA G-quadruplex forming sequence in shrimp white spot syndrome virus (WSSV). In E3S Web of Conf. Volume 374, 2023. The 3rd International Conference on Natural Resources and Life Sciences (NRLS) 2020, September, 23rd-24th, 2020, Surabaya.

Lie, Yoanes Maria Vianney

Acknowledgments

I greatly thank Prof. Dr. Klaus Weisz as my doctoral supervisor. You have given me a lot of support, both on a professional and personal level. You gave me the opportunity to enjoy not only the vibrant research environment but also a lot of guidance on a fundamental level. In addition, you were very open to new topics and spent a lot of time on constructive discussions. You were always available when I was looking for a discussion. In addition, you spent a lot of time correcting my draft written in English and also this thesis. I am grateful for the last four years of my stay in this working group with you.

Many thanks to the members of AKW during my stay here; Petra: for always providing loving care for this working group; Jagannath: for staying all through the 4 years of my study here and always bearing with me. Thank you also for the guidance and inputs you gave me for the completion of my work, study, and this thesis; Linn: for introducing me to the methodologies, especially NMR and molecular dynamics simulation sections, which are important for my work and study. It was also a nice experience to go to the NMR school with Linn, as well as the NMR workshop and G4 conference with Jagannath. In addition, many thanks to Nina, Dorothea, as well as Gregor for the many constructive sessions we had during my time here, which expanded our topics altogether. Additionally, I would like to thank the former members of AKW whom I had the opportunity to interact with during my visit here in 2018: Trixi, Andrea (thank you for the synthesized PIQ ligands), Lena for all of your guidance during my work back then which introduced me to the working topics of this group, Jonathan who also helps me to solve a computational problem in one of my studies here, and especially Maria who was my supervisor during my studies in Surabaya for the support and further introduction to Klaus.

I would especially like to thank Norman for the time you gave me during my problems in handling the computational study, providing options and suggestions, in addition to helping with my computer (OS) related problems. I also thank Theo for the discussion and help with gel electrophoresis and visualization.

Finally, I would like to thank my family; father, mother, and little brother who always support me in my life choice and provide me very caring atmosphere. A special thanks also to my friend, girlfriend, fiancé, and now wife, Ivana Rosalyn for putting up with me all the time as we have to stay at a very long distance for three years, not to mention the difficulty in the waiting period prior to your arrival in Germany. Thank you for always being there for me, and for the love and patience all of you have given to me.