8684–8696 *Nucleic Acids Research, 2017, Vol. 45, No. 15* *Published online 24 July 2017*
doi: 10.1093/nar/gkx637

Multimerization rules for G-quadruplexes

Sofia Kolesnikova[1,2], Martin Hubálek[1], Lucie Bednárová[1], Josef Cvačka[1] and Edward A. Curtis[1,*]

[1]The Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 166 10, Czech Republic and [2]Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague 166 28, Czech Republic

Received March 17, 2017; Revised July 09, 2017; Editorial Decision July 11, 2017; Accepted July 20, 2017

ABSTRACT

G-quadruplexes can multimerize under certain conditions, but the sequence requirements of such structures are not well understood. In this study, we investigated the ability of all possible variants of the central tetrad in a monomeric, parallel-strand G-quadruplex to form higher-order structures. Although most of these 256 variants existed primarily as monomers under the conditions of our screen, ∼10% formed dimers or tetramers. These structures could form in a wide range of monovalent and divalent metal ions, and folding was highly cooperative in both KCl and $MgCl_2$. As was previously shown for G-quadruplexes that bind GTP and promote peroxidase reactions, G-quadruplexes that form dimers and tetramers have distinct sequence requirements. Some mutants could also form heteromultimers, and a second screen was performed to characterize the sequence requirements of these structures. Taken together, these experiments provide new insights into the sequence requirements and structures of both homomultimeric and heteromultimeric G-quadruplexes.

INTRODUCTION

Multimerization is an important aspect of protein structure and function (1–4). A significant fraction of the proteins in the human genome are thought to dimerize, and trimeric and tetrameric structures are also frequently observed (1–4). Furthermore, aggregates containing tens, hundreds or even thousands of subunits have been reported in the context of structures such as viral capsids (5). In addition to increasing the structural diversity of proteins, multimer formation can also regulate protein function (1–4). For example, a large number of eukaryotic signal transduction pathways are regulated by the GTP-dependent multimerization of heterotrimeric G proteins (6). These proteins cycle between a monomeric 'on' state which is bound to GTP and a trimeric 'off' state which is bound to GDP. Enzymatic conversion of GTP to GDP triggers trimer formation and switches the complex off, while exchange of guanosine-5'-diphosphate (GDP) for guanosine-5'-triphosphate (GTP) promotes trimer dissociation and turns the complex on again.

In the case of nucleic acids, multimerization is most frequently investigated in the context of duplex formation. This involves interactions between complementary strands via A-T and C-G base pairs (in DNA) or A-U and C-G base pairs (in RNA) (7–8). Such interactions play a central role in chromosome structure and function, and determine the substrate specificity of biological processes such as RNA silencing (9). Nucleic acids can also multimerize using mechanisms that do not involve base pairing. For example, non-canonical nucleic acid structures such as G-quadruplexes can form a wide variety of multimeric structures (10–12). The assembly of tetramolecular G-quadruplexes made up of four molecules of short oligonucleotides containing a single stretch of guanosines such as TGGGGGT has been particularly well studied (13–15). G-quadruplexes made from either single or multiple strands can also interact to form higher order structures such as dimers and tetramers (16–27). In some cases, interfaces between subunits are formed by the stacking of canonical GGGG tetrads (18,21,26–27), and stabilization by overhanging nucleotides at the 5′ or 3′ end of the sequence or by nucleotides in loops has also been reported (16–17,20,24,26). When overhanging nucleotides are guanosines, interlocked structures can form in which the tetrad at the interface between subunits is formed by guanosines from two different monomer subunits (17,26). When overhanging nucleotides are not guanosines, they can stabilize the interface by forming unusual structural motifs such as G(:C):G(:C):G(:C):G(:C) octads (20,24) and G(:A):G(:A):G(:A):G heptads (16). Intertwined structures in which tetrads are formed from guanosines from multiple DNA molecules have also been reported (19,22–23,25).

Although the existence of multimeric G-quadruplexes is well established, the types of sequences that can form such structures have not been systematically investigated. Here, we explored this question using a library containing all pos-

*To whom correspondence should be addressed. Tel: +420 733 169 654; Email: curtis@uochb.cas.cz

sible variants of the central tetrad in a parallel-strand G-quadruplex structure. Our experiments indicate that ∼10% of these variants form dimers or tetramers, and that some can also form heteromultimeric structures. As was previously shown for G-quadruplexes that bind GTP and promote peroxidase reactions (28), G-quadruplexes that form dimers and tetramers have distinct sequence requirements. These experiments provide additional evidence in support of the idea that mutations in tetrads can alter the biochemical specificity of G-quadruplexes. They also provide new insights into the sequence requirements of higher-order G-quadruplex structures.

MATERIALS AND METHODS

Reagents

Desalted DNA oligonucleotides were purchased from Sigma. Oligonucleotides were resuspended in Milli-Q water at a concentration of 100 μM and used without additional purification. Stock solutions were stored at −20°C and thawed at room temperature before use. $MgCl_2$, KCl and HEPES buffer were purchased from Sigma. Stock solutions were prepared by dissolving in Milli-Q water and filtered using 0.22 μm filters from VWR. The pH of HEPES solutions were adjusted using KOH purchased from Sigma.

Native gels

In a typical assay, a 100 μM G-quadruplex stock solution (stored at −20°C) was thawed at room temperature. After vortexing, 2 μl was mixed with 2 μl of a 5′ radiolabeled version of the sequence (≤10 nM; prepared as described in Ref. 28) and 6 μl of Milli-Q water. The solution was then heated at 65°C for 5 min, cooled at room temperature for 5 min and mixed with 10 μl of 2 × G-quadruplex buffer (400 mM KCl, 2 mM $MgCl_2$, 40 mM HEPES pH 7.1). Final concentrations were 10 μM G-quadruplex in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1. After incubating at room temperature for 30 min, 4 μl of 6 × gel loading buffer (60% w/v glycerol, 0.15% w/v xylene cyanol and 0.15% w/v bromophenol blue) was added to each sample, and the material was analyzed by native polyacrylamide gel electrophoresis (PAGE) on a 10% gel containing 5 mM KCl in both the gel and buffer. Gels were run at 300 V for 30 min and scanned using a Typhoon phosphorimager. In some experiments (including our initial screen for homomultimer formation), folding reactions were performed using unlabeled DNA, and visualized by staining with GelRed using the protocol recommended by the manufacturer. Pilot experiments indicated that similar results were obtained using unlabeled and radiolabeled material (Supplementary Figure S1). In most experiments we included a 17 nt single-stranded marker (labeled ss) with the sequence GACTGCCTCGTCACGAT as well as a 17 bp double-stranded marker (labeled ds) that contained a mix of this sequence and its reverse complement. In Figure 1B, we also included a 34-bp double-stranded marker that contained a mix of the sequence GGTCATACCAAC CTCTGGTTAGACTTCGATGCAA and its reverse complement.

Circular dichroism spectrometry

In a typical circular dichroism (CD) experiment, a 100 μM G-quadruplex stock solution (stored at −20°C) was thawed at room temperature. After vortexing, 20 μl was mixed with 80 μl of Milli-Q water. This solution was heated at 65°C for 5 min, and cooled at room temperature for 5 min. It was then mixed with 100 μl of 2 × G-quadruplex buffer (400 mM KCl, 2 mM $MgCl_2$, 40 mM HEPES pH 7.1). Final conditions were 10 μM G-quadruplex in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1. After incubating at room temperature for 30 min, electronic circular dichroism (ECD) spectra were measured on a Jasco 815 spectropolarimeter over a spectral range of 200–350 nm. Measurements were made in a quartz cell with a 0.1 cm path length using a scanning speed of 10 nm/min, a response time of 8 s, standard instrument sensitivity and three spectra accumulations. After a baseline correction, spectra were expressed in terms of differential optical density.

Mass spectrometry

In a typical assay, a 100 μM G-quadruplex stock solution (stored at −20°C) was thawed at room temperature. After vortexing, 1 μl was mixed with 4 μl of water. The solution was then heated at 65°C for 5 min, cooled to room temperature and mixed with 5 μl of 400 mM ammonium acetate pH 7. Final concentrations were 10 μM G-quadruplex in a buffer containing 200 mM ammonium acetate pH 7. Control experiments indicated that G-quadruplexes folded in this buffer could not be distinguished from those folded in our standard G-quadruplex buffer (200 mM KCl, 1 mM $MgCl_2$, 20 mM HEPES pH 7.1) on native gels (Supplementary Figure S2). After incubating at room temperature for 30 min, 2 μl was placed in a glass capillary needle and sprayed into the mass spectrometer (Synapt G2, Waters). Data in the range 500–5000 m/z were acquired in negative ion mode at 1.2 kV capillary voltage using a source temperature of 40°C, a 70 V sampling cone, a 2.1 V extraction cone, a 25 V trap collision energy and a 10 V transfer collision energy with the ion mobility separation switched on. Thirty scans were combined, subtracted (polynomial order 15, 3% below curve) and smoothed (Savitzky Golay, 20 channels, two times) to produce the spectra in Figure 2B. Peaks in these spectra were annotated by matching theoretical masses with individual charge states of peaks with the help of the ion mobility spectral profile visualized in DriftScope v2.8 (Waters).

Curve fitting

Titrations were performed by incubating a constant amount of radiolabeled G-quadruplex with increasing concentrations of an unlabeled version of the same sequence and analyzing the resulting complexes on native gels. Although this is typically not an equilibrium technique, it was used in this case because, once formed, dimers and tetramers are extremely stable. Binding curves for duplex and dimer formation were fit using Equation (1), where F is the fraction duplex or dimer, F_{min} is the minimum observed fraction duplex or dimer, F_{max} is the maximum observed fraction duplex or dimer, S is the total DNA concentration and K_d is

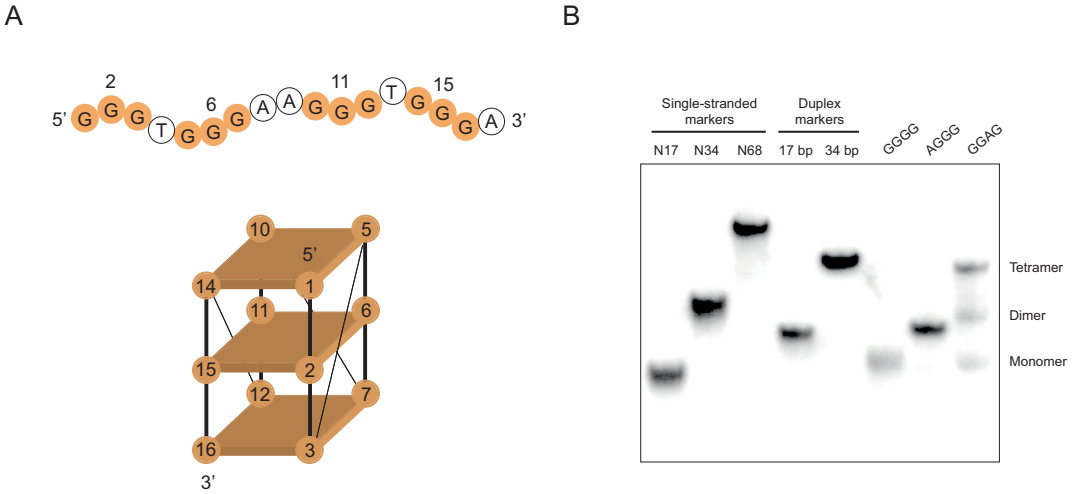


Figure 1. Mutations that induce formation of dimeric and tetrameric G-quadruplex structures. (**A**) Primary sequence and proposed topology of the reference construct used in these experiments. Mutated positions in the central tetrad are numbered. (**B**) Native gel showing different types of multimeric structures formed by mutants of the reference construct. Experiments were performed at 10 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1.

the dissociation constant.

$$F = F_{min} + [(F_{max} - F_{min}) \times (S/K_d + S)] \quad (1)$$

Binding curves for tetramer formation were fit using Equation (2), where F is the fraction tetramer, F_{min} is the minimum observed fraction tetramer, F_{max} is the maximum observed fraction tetramer, S is the total DNA concentration, $K_{0.5}$ is the DNA concentration at which the fraction tetramer is half the maximum value and n is the Hill coefficient (fixed at 1.3 for these fits).

$$F = F_{min} + [(F_{max} - F_{min}) \times ([S]^n/[K_{0.5}]^n + [S]^n)] \quad (2)$$

This equation can be used for tetramer formation when binding of the first subunit is significantly stronger than that of the second and third subunits (29). This appears to be the case for the homotetramers analyzed here because in most titrations only monomers and tetramers were observed, although dimers were sometimes seen when tetramer-forming sequences with different mutations in the central tetrad of the reference construct were mixed.

Binding curves generated in the presence of different $MgCl_2$ and KCl concentrations were fit using Equation (3), where F is the fraction dimer or tetramer, F_{max} is the maximum fraction dimer or tetramer, M is the metal ion concentration, $K_{0.5}$ is the metal ion concentration at which the fraction dimer or tetramer is half the maximum value and n is the Hill coefficient.

$$F = F_{max} \times ([M]^n/[K_{0.5}]^n + [M]^n) \quad (3)$$

RESULTS

Identification of mutations that induce formation of multimeric G-quadruplexes

In a previous study, we investigated the effect of mutating the central tetrad in a parallel-strand DNA G-quadruplex on its ability to bind GTP and to catalyze peroxidase reactions (Figure 1A) (28,30). This sequence was chosen as a model because it exhibited robust GTP-binding and peroxidase activity, and also because the positions in the central tetrad (nucleotides 2, 6, 11 and 15) could be identified (28). Although both gel filtration (28) and native PAGE (Supplementary Figure S3) suggest that this sequence exists primarily as a monomer, some mutants formed multimeric structures when analyzed on native gels (Figure 1B). Two main types of structures were observed. The mobility of one (which we refer to as a dimer) was similar to that of a double-stranded marker oligonucleotide of the same length as these G-quadruplex variants (Figure 1B). The mobility of the other (which we refer to as a tetramer) was similar to a double-stranded marker oligonucleotide of twice the length of these variants (Figure 1B). To obtain additional evidence that the structures we observed on native gels were G-quadruplexes, examples of sequences that form dimers or tetramers were analyzed by CD spectrometry (31–33). This indicated that, for both types of structures, G-quadruplexes made up a significant fraction of the reaction mixture (Figure 2A; Supplementary Figures S4 and 5). To rule out the possibility that these spectra were due to contaminating monomers, we developed a method to obtain pure samples of both dimers and tetramers (Supplementary Figure S6A). After purifying on native gels and eluting into folding buffer, both types of structures could be stored at room temperature for at least 10 days without detectable dissociation (Supplementary Figure S6B). Characterization of purified dimers and tetramers by CD spectrometry indicated that, like that reference construct, they formed parallel-strand G-quadruplexes (Supplementary Figure S6C). Analysis of CD spectra as a function of temperature indicated that monomers were stable even at 90°C, while both dimers and tetramers had melting temperatures of ~70°C (Supplementary Figure S7). To obtain additional evidence for the proposed strand stoichiometries of these structures, unpurified monomers, dimers and tetramers were analyzed by electrospray mass spectrometry (15,17,20,26,34–35). This indicated that dimer-forming se-

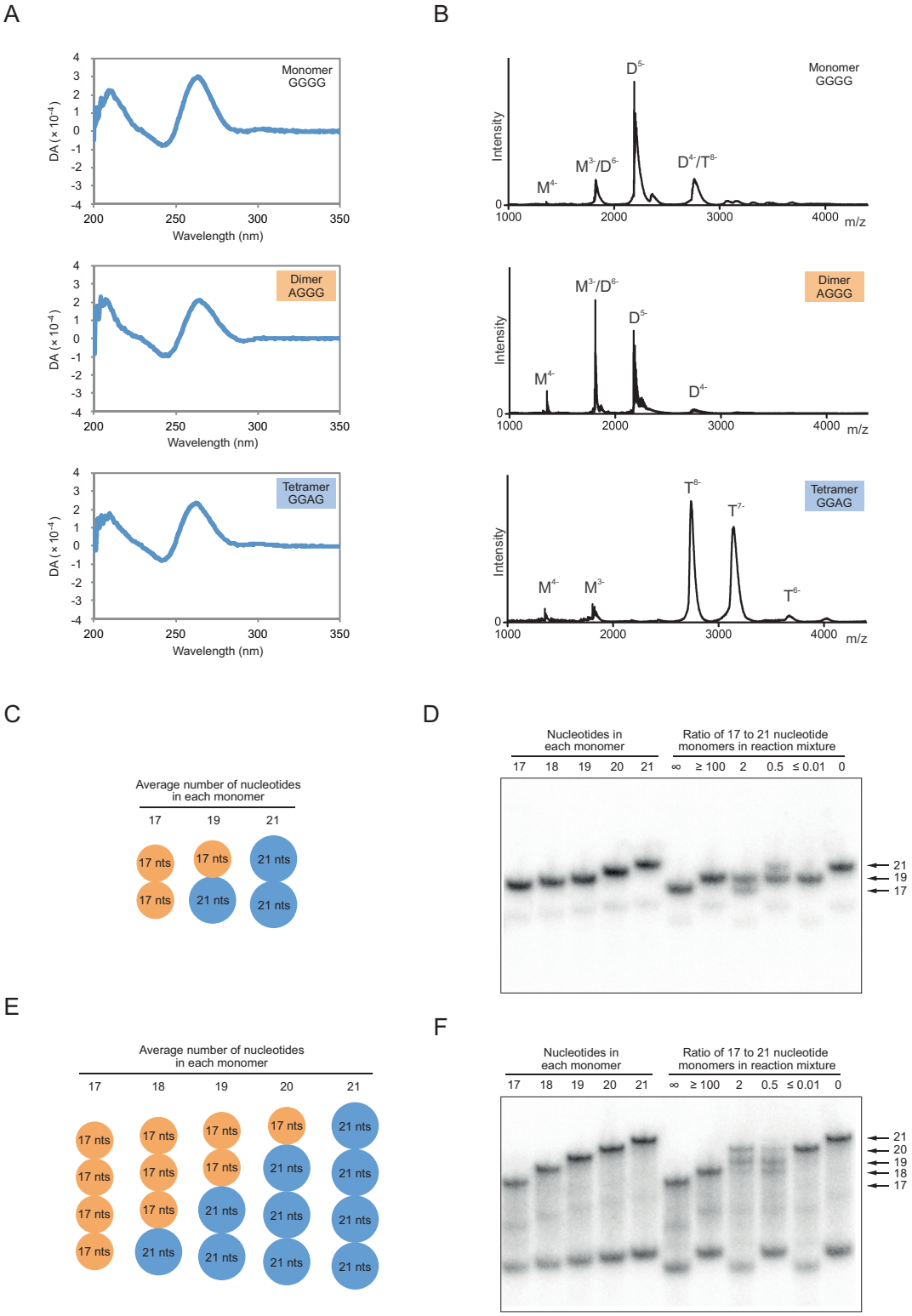


Figure 2. Structural characterization of dimeric and tetrameric G-quadruplexes. (**A**) Circular dichroism (CD) spectra of G-quadruplex variants that exist primarily as monomers, dimers or tetramers under the conditions of our screen. (**B**) Analysis of monomeric, dimeric and tetrameric G-quadruplexes by electrospray mass spectrometry. (**C**) Homodimeric and heterodimeric structures generated from two monomers of different lengths. (**D**) Detection of each of these structures on a native gel. Left part of gel: homodimeric markers generated from 17, 18, 19, 20 and 21 nt dimer-forming G-quadruplexes with the sequence GAGTGGGAAGGGTGGG(A_{1-5}). Right part of gel: homodimeric and heterodimeric structures generated by mixing 17 and 21 nt dimer-forming G-quadruplexes in different ratios. (**E**) Homotetrameric and heterotetrameric structures generated from two monomers of different lengths. (**F**) Detection of each of these structures on a native gel. Left part of gel: homotetrameric markers generated from 17, 18, 19, 20 and 21 nt tetramer-forming G-quadruplexes with the sequence GGGTGGGAAGAGTGGG(A_{1-5}). Right part of gel: homotetrameric and heterotetrameric structures generated by

quences form a mix of structures containing one and two strands and tetramer-forming sequences primarily form a structure containing four strands (Figure 2B). On the other hand, monomers were only detected in small amounts when the reference sequence was analyzed under these conditions (Figure 2B). When interpreted in light of other data (Figure 1B, Supplementary Figure S3 and Ref. 28), this suggests that the relative stabilities of different multimeric forms of the reference sequence are different in the gas phase and in native gels. Similar discrepancies have been reported in other studies. For example, in some cases different results were obtained when the number of ammonium ions bound to G-quadruplexes was determined by mass spectrometry and by nuclear magnetic resonance (NMR) (36). To investigate the stoichiometries of dimers and tetramers using an independent method, mixed-strand experiments were performed in which sequences of different lengths were mixed in various ratios and analyzed on native gels (10). Three products could be detected when two dimer-forming sequences of different lengths were mixed, which is consistent with the pattern expected for a structure containing two strands (Figure 2C and D). Similarly, five products were obtained when two tetramer-forming sequences were mixed, suggesting that this structure contains four strands (Figure 2E and F). Taken together, these experiments indicate that the higher-order structures we observed on native gels are multimeric G-quadruplexes. They also support the proposed stoichiometries of these structures.

Metal ion requirements of dimeric and tetrameric G-quadruplexes

After characterizing the stoichiometries of dimeric and tetrameric G-quadruplexes, we next investigated their metal ion requirements. We initially characterized the role of the monovalent metal ion potassium and the divalent metal ion magnesium on dimer and tetramer formation. Although the buffer used in our screen contained both KCl and $MgCl_2$, dimers and tetramers could form in buffers containing either KCl alone or $MgCl_2$ alone (Figure 3). $MgCl_2$ was a more potent inducer of both dimer and tetramer formation than KCl: although significant levels of such structures were observed at 2 mM $MgCl_2$, 100-fold higher concentrations of KCl were required to achieve a similar effect (Figure 3). Dimers and tetramers could form in a wide variety of other monovalent and divalent cations (Supplementary Figure S8), indicating that such structures simply require positive charge rather than the presence of specific metal ions. Although monovalent metal ions (especially potassium) are known to play important roles in the stabilization of G-quadruplexes, formation of such structures in the absence of monovalent metal ions has been previously demonstrated (37–41). Analysis of the concentration dependence of dimer and tetramer formation revealed that the assembly of such structures is highly cooperative with respect to both KCl and $MgCl_2$ concentration (Figure 3). This effect is especially pronounced for dimer and tetramer formation in the presence of KCl: such structures increase from undetectable levels to the maximum level observed over only a 4-fold range of KCl concentrations. The Hill coefficient of tetramer formation in KCl exceeds 4 (Figure 3), which is considerably higher than that previously reported for the folding of G-quadruplexes in the presence of KCl (42–43). This could reflect the requirement for additional metal ions when assembling a G-quadruplex structure from multiple strands. Taken together, these experiments provide new insights into conditions that promote the formation of dimeric and tetrameric G-quadruplexes. They also show that these structures bind metal ions, especially monovalent metal ions, in a highly cooperative way.

Sequence requirements of multimeric G-quadruplexes

To better characterize the sequence requirements of multimeric G-quadruplexes, we analyzed all 256 possible variants of the central tetrad in our reference G-quadruplex on native gels. Although the majority of these sequences appeared to exist primarily as monomers under these conditions (10 μM DNA, 200 mM KCl, 1 mM $MgCl_2$, pH 7.1), ∼10% formed higher-order structures at detectable levels (Figures 4 and 5). Apparent dissociation constants for dimer formation ranged from 30 to 1000 nM, while concentrations needed to achieve half maximal tetramer formation ranged from 180 to 14 000 nM (Figures 4 and 5; Supplementary Figures S9 and 10). In comparison, the apparent dissociation constant was 2 nM for formation of a duplex from complementary oligonucleotides of the same length as the G-quadruplexes analyzed in this study (Supplementary Figure S11). In the case of dimers, mutations at positions 2 and 6 generally had similar effects, and both G to A and G to T mutations were more potent inducers of dimer formation than G to C mutations (Figure 4 and Supplementary Figure S9). In contrast, the 11A mutation was most important for tetramer formation, and G to A mutations typically had stronger effects than G to T or G to C changes (Figure 5 and Supplementary Figure S10). In addition, multiple mutations were typically required to induce tetramer formation, whereas this was not the case for dimers (Figures 4 and 5; Supplementary Figures S9 and 10). These results indicate that, as is the case for G-quadruplexes that bind GTP and promote peroxidase reactions (28), mutations in tetrads can change the biochemical specificity of G-quadruplex multimer formation.

←

mixing 17 and 21 nt tetramer-forming G-quadruplexes in different ratios. Experiments in panel A were performed at 10 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1, while those in panel B were performed at 10 μM DNA concentration in a buffer containing 200 mM NH_4OAc, pH 7. Experiments in panels D and F were performed at 1–1.5 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1. '∞' lane ≤ 10 nM radiolabeled 17 nt monomer mixed with 1 μM unlabeled 17 nt monomer; '≥100' lane ≤ 10 nM radiolabeled 21 nt monomer mixed with 1 μM unlabeled 17 nt monomer; '2' lane ≤ 10 nM radiolabeled 17 nt monomer mixed with 1 μM unlabeled 17 nt monomer and 0.5 μM unlabeled 21 nt monomer; '0.5' lane ≤ 10 nM radiolabeled 21 nt monomer mixed with 0.5 μM unlabeled 17 nt monomer and 1 μM unlabeled 21 nt monomer; '≤0.01' lane ≤ 10 nM radiolabeled 17 nt monomer mixed with 1 μM unlabeled 21 nt monomer; '0' lane ≤ 10 nM radiolabeled 21 nt monomer mixed with 1 μM unlabeled 21 nt monomer.

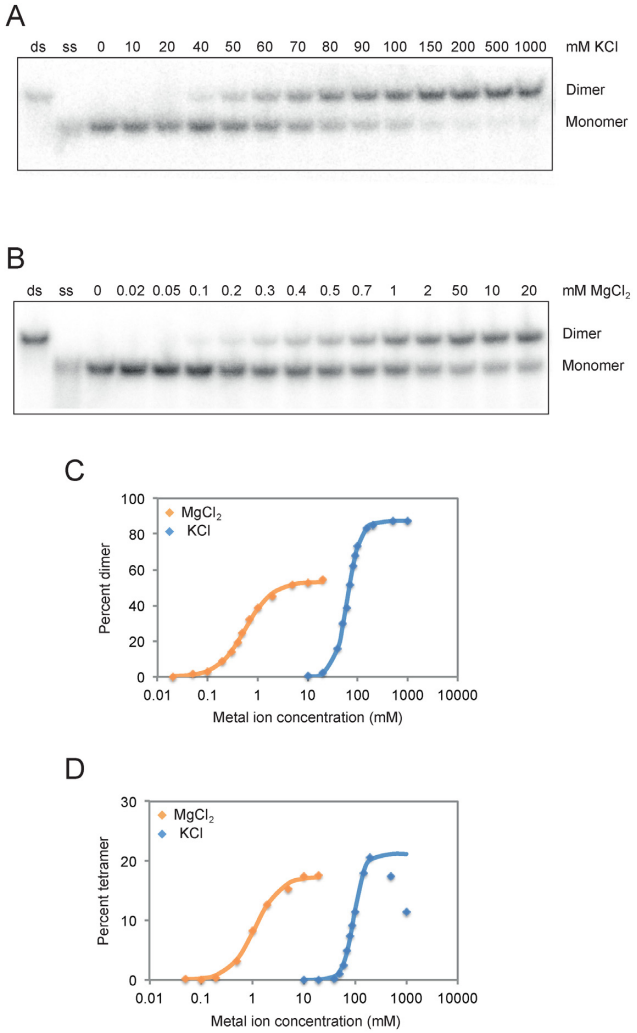


Figure 3. Cooperative folding of dimeric and tetrameric G-quadruplexes. (**A**) Native gel showing dimer formation as a function of KCl concentration. (**B**) Native gel showing dimer formation as a function of $MgCl_2$ concentration. (**C**) Graph showing dimer formation as a function of KCl and $MgCl_2$ concentration. (**D**) Graph showing tetramer formation as a function of KCl and $MgCl_2$ concentration. Dimers were generated using a construct with an AGGG mutation in the central tetrad of the reference construct and the sequence GAGTGGGAAGGGTGGGA. Tetramers were generated using a construct with a GGAG mutation in the central tetrad of the reference construct and the sequence GGGTGGGAAGAGTGGGA. Experiments were performed at 1 μM G-quadruplex concentration in a buffer containing 20 mM HEPES pH 7.1 and either 0–1000 mM KCl or 0–20 mM $MgCl_2$.

To obtain additional information about the sequence requirements of dimer and tetramer formation, we investigated the position requirements of the mutated tetrads identified in our screen. In the case of dimers, most mutations could be transplanted to the 5′ tetrad but not the 3′ tetrad of the reference construct without loss of activity (Figure 6). In the case of tetramer formation, however, this was not the case: such mutations could not typically be transplanted to either the 5′ or 3′ tetrad without inhibiting formation of these structures (Figure 7). We also investigated the loop requirements of dimer and tetramer formation. In both the cases, point mutations did not significantly influence multimerization efficiency, but dimers were considerably more tolerant than tetramers to the effects of multiple mutations in loops (Figures 6 and 7). Moreover, for some tetramer-forming sequences such mutations resulted in the formation of additional products (Figure 7). Because these products were only rarely observed they were not characterized further. The effects of mutations in loops were sometimes different in the presence of different mutations in the central tetrad of the reference construct (Supplementary Figures S12 and 13), which prevented us from developing sequence models that combined information from loop and tetrad mutagenesis experiments. Despite this limitation, comparison to previous models (28) indicated that, with respect to mutations in the central tetrad of the reference construct, the sequence requirements of G-quadruplexes that form tetramers are most similar to those that bind GTP, while the sequence requirements of G-quadruplexes that form dimers are most similar to those that promote peroxidase reactions (Supplementary Figure S14). On the other hand, the loop requirements of G-quadruplexes that form tetramers appear to be most similar to those that promote peroxidase reactions, while the loop requirements of G-quadruplexes that form dimers appear to be most similar to those that bind GTP (Supplementary Figure S14).

Assembly of heteromultimers from G-quadruplex variants with different sequences

Because the G-quadruplex variants analyzed in our initial screen were tested for multimer formation in the absence of other sequences, these experiments only provided information about the sequence requirements of homomultimer formation. To determine whether these mutants could also assemble into heteromultimeric structures, we performed experiments in which variants with different mutations in the central tetrad of the reference construct were mixed and analyzed on native gels. Because testing all $256 \times 256 = 65\,536$ possible pairs of G-quadruplex variants in our library for multimer formation was well beyond the scope of the study, we instead focused on a subset of variants that form stable homodimers or homotetramers. Our analysis included all possible pairwise combinations of seven different dimer-forming sequences and six different tetramer-forming sequences. For each pair analyzed, a trace amount of a radiolabeled version of one sequence was mixed with 10 μM of an unlabeled version of a second sequence, and the resulting complexes were analyzed on native gels. Because only one strand was radiolabeled in this assay, and it was diluted to a concentration at which homomultimers could not form (Supplementary Figure S15), any higher-order structures observed on native gels were likely to contain both radiolabeled and unlabeled strands. These experiments revealed that dimers typically form when two different dimer-forming sequences are mixed (upper left quadrant of Figure 8A), and that tetramers typically form when two different tetramer-forming sequences are mixed (lower right quadrant of Figure 8B). This suggests that, for both types of structures, the formation of interfaces requires mutations at specific positions in the central tetrad but not necessarily specific mutations at these positions. Dimers also formed about half the time when a dimer-forming sequence was

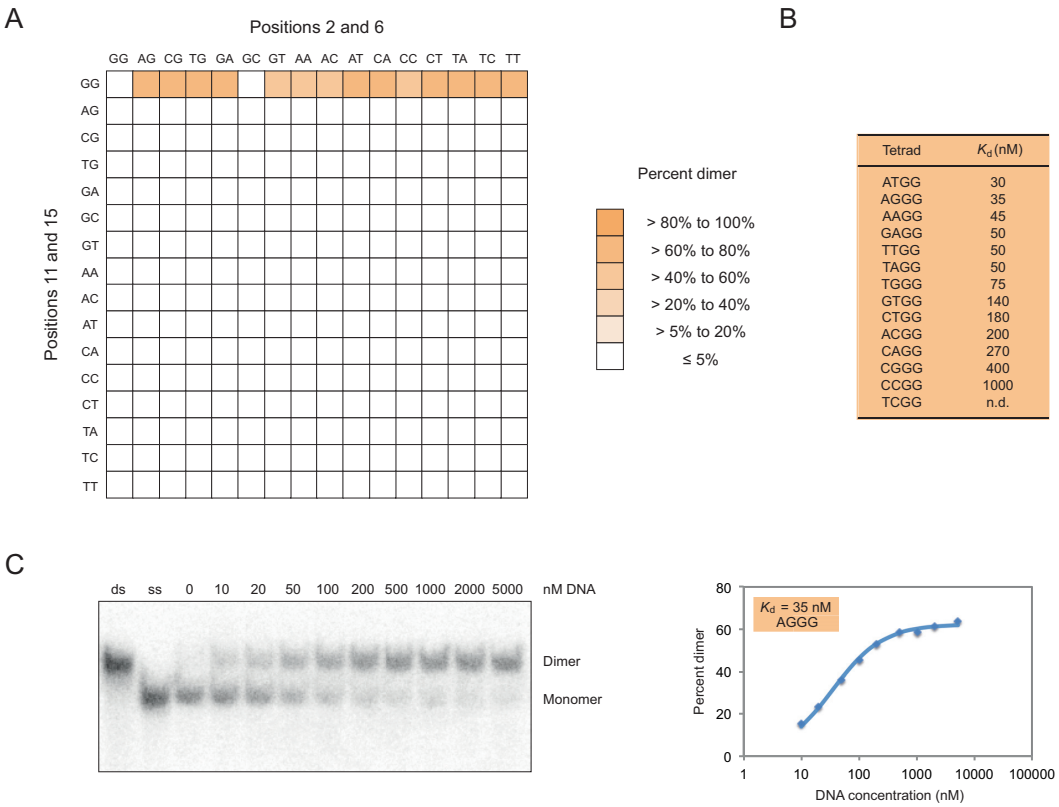


Figure 4. Sequence requirements of G-quadruplex dimer formation. (**A**) Heat map showing the ability of all possible variants of the central tetrad in the reference construct to form dimers at 10 μM DNA concentration. (**B**) Dissociation constants of 13 of the variants identified in our screen. (**C**) Native gel and graph showing dimer formation as a function of DNA concentration. Dimers in panel (C) were generated using a construct with an AGGG mutation in the central tetrad of the reference construct and the sequence GAGTGGGAAGGGTGGGA. Experiments were typically performed between 10 nM and 10 μM G-quadruplex concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1.

mixed with a tetramer-forming sequence (upper right and lower left quadrants of Figure 8A), indicating that mutations that induce dimer formation can be dominant to those that induce tetramer formation (44). In contrast, tetramers never formed when a dimer-forming sequence was mixed with a tetramer-forming sequence (upper right and lower left quadrant of Figure 8B), indicating that each of the four monomers in a tetramer must contain a GGNN mutation in the central tetrad of the reference construct. Finally, in about one third of the cases in which two different tetramer-forming sequences were mixed, both dimers and tetramers were generated (compare the lower right quadrants of Figure 8A and B). This is consistent with the idea that tetramers contain two distinct interfaces: one that can be formed by various combinations of NNGG and GGNN G-quadruplexes, and a second that can only be formed by GGNN G-quadruplexes.

Mapping of multimer interfaces by site-directed mutagenesis

Multimeric G-quadruplexes are often stabilized by stacking interactions between the terminal tetrads of monomeric subunits (16–27), and previous studies have shown that overhanging nucleotides can interfere with interface formation in such structures (10,17–18,21,45). On the other hand, in the case of intertwined structures in which tetrads are formed from guanosines from multiple DNA molecules (19,22–23,25), the addition of overhanging nucleotides does not typically influence multimer formation (19). To probe the interfaces of the G-quadruplexes described here, we investigated the effects of 5′ and 3′ overhanging nucleotides on dimer and tetramer formation. In the case of a representative dimer-forming sequence, overhanging nucleotides had no effect on dimerization efficiency (Figure 9A). This suggests that neither the 5′ nor the 3′ terminus of this sequence is present at the interface, and is consistent with an intertwined structure (for an example of such a structure that maximizes the number of GGGG tetrads see Figure 9B). In the case of a representative tetramer-forming sequence, modifying the 3′ terminus also had little effect on tetramer formation (Figure 9C). Adding a nucleotide to the 5′ terminus, however, resulted in a striking change: such mutants formed dimers rather than tetramers (Figure 9C). These observations are consistent with the idea that tetramers contain two distinct interfaces: one formed by two intertwined monomers, which would not be expected to be disrupted by either 5′ or 3′ overhangs, and a second formed by two 5′-5′ stacked dimers, which would be expected to be destabilized by 5′ but not 3′ overhangs (Figure 9D). Such 5′ to 5′ stacking has been previously observed in multimeric G-quadruplexes, and is thought to be the most favorable mode of stacking for parallel-strand G-quadruplex structures (46).

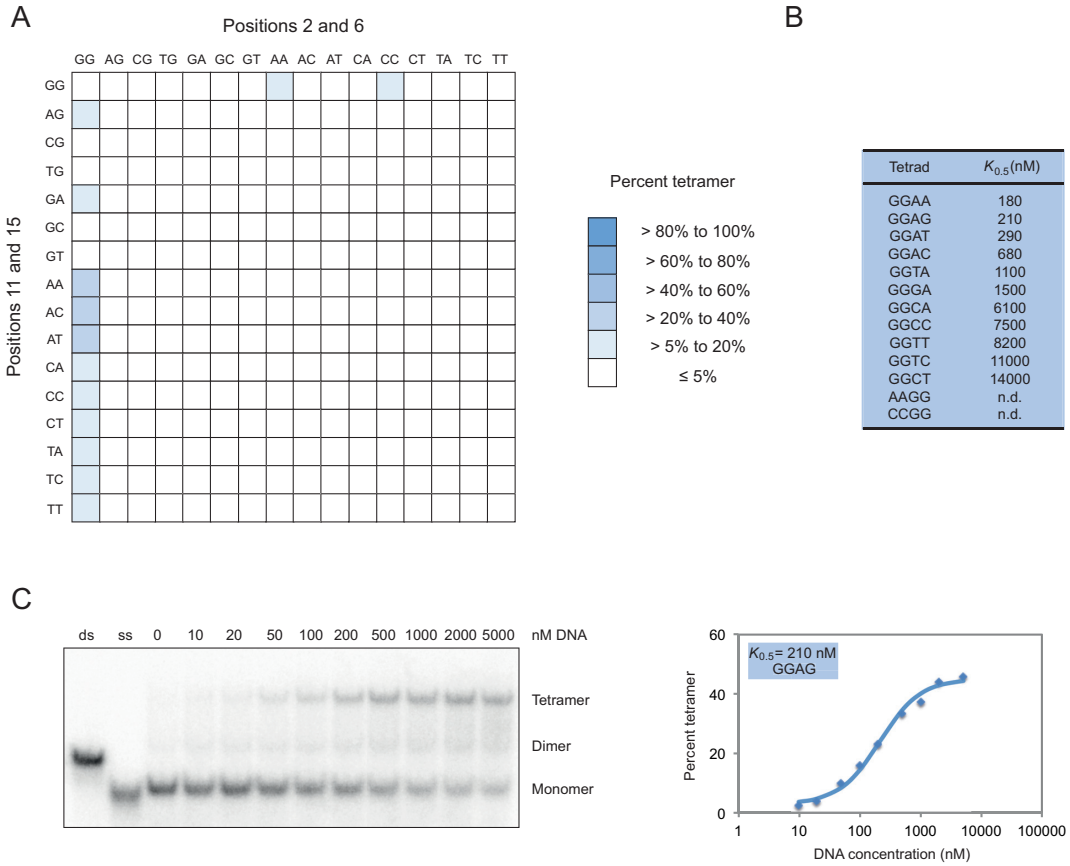


Figure 5. Sequence requirements of G-quadruplex tetramer formation. (**A**) Heat map showing the ability of all possible variants of the central tetrad of the reference construct to form tetramers at 10 μM DNA concentration. (**B**) DNA concentration at which tetramer formation is half the maximum value ($K_{0.5}$) for 11 of the variants identified in our screen. (**C**) Native gel and graph showing tetramer formation as a function of DNA concentration. Experiments in panel (C) were performed using a construct with a GGAG mutation in the central tetrad of the reference construct and the sequence GGGTGGGAAGAGTGGGA. Experiments were typically performed between 10 nM and 10 μM G-quadruplex concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1.

DISCUSSION

In this study, we used a combination of approaches to characterize mutations in the central tetrad of a monomeric, parallel-strand G-quadruplex that induce formation of higher-order structures. Our experiments indicate that some of these mutants form dimeric G-quadruplexes while others form tetramers, and that some variants are also capable of forming heteromultimeric structures. Both dimers and tetramers can form in a wide variety of monovalent and divalent metal ions, and metal binding is highly cooperative, especially in the case of potassium. Perhaps most interestingly, sequences that form dimers typically contain NNGG mutations in the central tetrad of the reference construct, which is similar to the pattern previously observed for G-quadruplexes that promote peroxidase reactions (28). On the other hand, variants that form tetramers typically contain GGNN mutations in the central tetrad of the reference construct, which is more similar to the sequence requirements of G-quadruplexes that bind GTP (28). These observations provide additional evidence in support of our hypothesis that mutations in tetrads play important roles in determining the biochemical specificity of G-quadruplex structures.

Although our experiments do not provide a high-resolution view of the structures of the multimeric G-quadruplexes characterized in this study, they do constrain the possibilities. CD experiments suggest that the building blocks of both types of structures are parallel-strand G-quadruplexes, and both native PAGE and mass spectrometry indicate that dimeric structures contain two DNA molecules while tetrameric structures contain four DNA molecules. These observations are consistent with two fundamentally different types of structures of monomer subunits. In one type of model, monomers form intramolecular G-quadruplexes in which positions 2, 6, 11 and 15 occur in the central tetrad and form either noncanonical tetrads or bulges. In a second type of model, individual strands interact to form intertwined structures containing tetrads made up of guanosines from multiple DNA molecules. An important difference between these models is the positions of 5′ and 3′ nucleotide relative to the interface. In the intramolecular model, interfaces are formed by either 5′ nucleotides (for 5′-5′ stacked structures), 3′ nucleotides (for 3′-3′ stacked structures) or both 5′ and 3′ nucleotides (for 5′-3′ stacked structures). On the other hand, in the intertwined model interfaces are not formed by either 5′ or 3′ nucleotides. Since

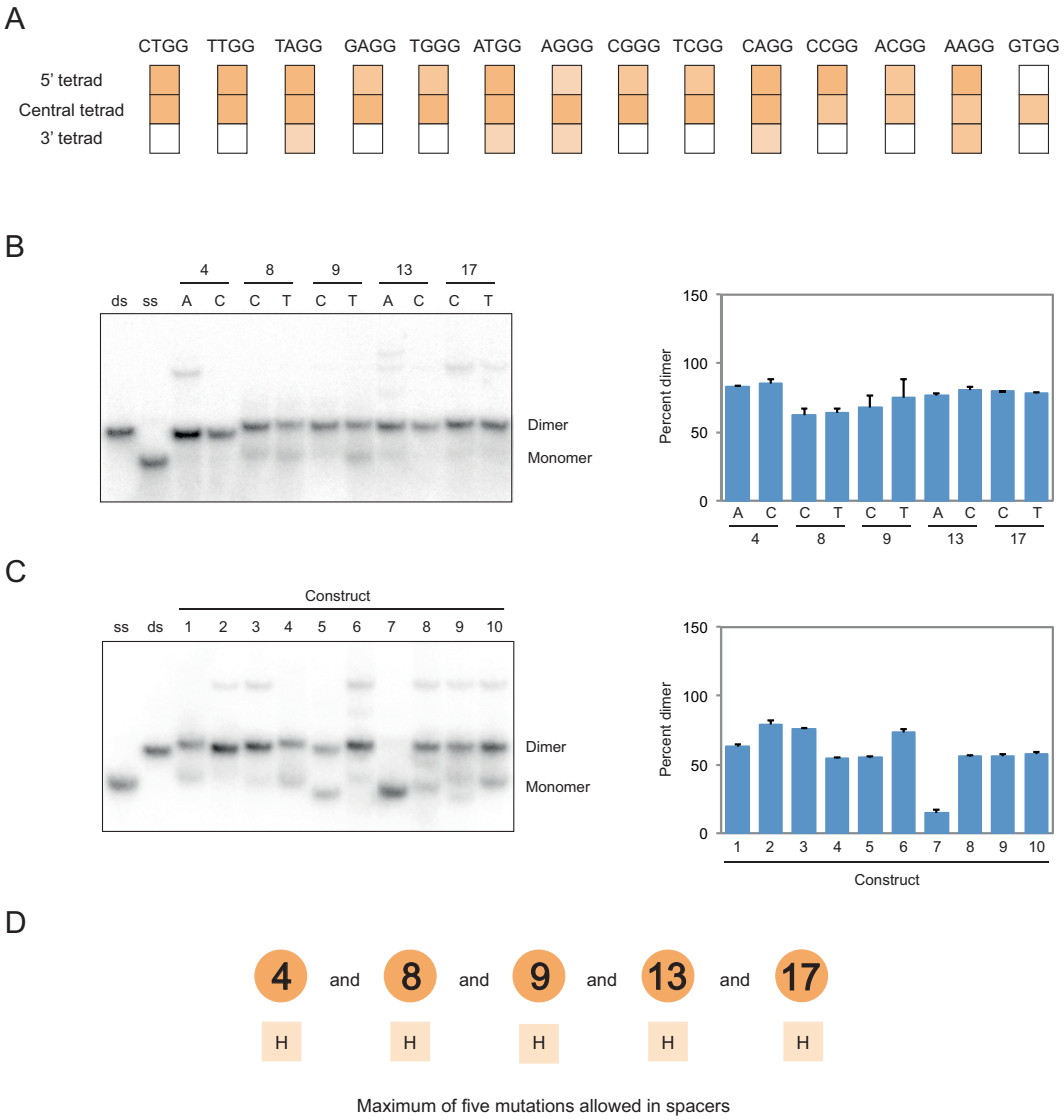


Figure 6. Sequence model for dimeric G-quadruplexes. (**A**) Effect of transplanting mutations that induce dimer formation when present in the central tetrad of the reference construct to the 5′ tetrad (nucleotides 1, 5, 10 and 14) or the 3′ tetrad (nucleotides 3, 7, 12 and 16). Colors match those in Figure 4. (**B**) Native gel and graph showing effects of point mutations in loops on dimer formation. (**C**) Native gel and graph showing the ability of ten randomly chosen sequences with up to five mutations in loops to form dimers. Sequences are given in Supplementary Table S1. (**D**) Loop requirements of G-quadruplex variants that form dimers. Numbers indicate the positions of loops in the reference construct, and letters below indicate the nucleotides that can occur at each position. Experiments were performed at 10 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1. H = A, C or T. Mutations in panels B and C were made in the context of a construct with an AGGG mutation in the central tetrad of the reference construct with the sequence GAGTGGGAAGGGTGGGA.

Table 1. Summary of the sequence requirements of dimeric and tetrameric G-quadruplexes

Experiment	Result for a dimer-forming sequence	Result for a tetramer-forming sequence	Evidence
Mutagenesis of the central tetrad	NNGG required	GGNN required	Figures 4A and 5A
5′ tetrad swap	Possible	Not possible	Figures 6A and 7A
3′ tetrad swap	Not possible	Not possible	Figures 6A and 7A
Point mutations in loops	A, C or T possible	A, C or T possible	Figures 6B and 7B
Total number of mutations in loops	One to five possible	One to two possible	Figures 6C and 7C
Mixing with a different dimer-forming sequence	Dimers formed	Dimers formed	Figure 8A and B
Mixing with a different tetramer-forming sequence	Dimers formed	Dimers and tetramers formed	Figure 8A and B
Addition of a 5′ overhang	Possible	Not possible (dimer formed instead)	Figure 9A and C
Addition of a 3′ overhang	Possible	Possible	Figure 9A and C

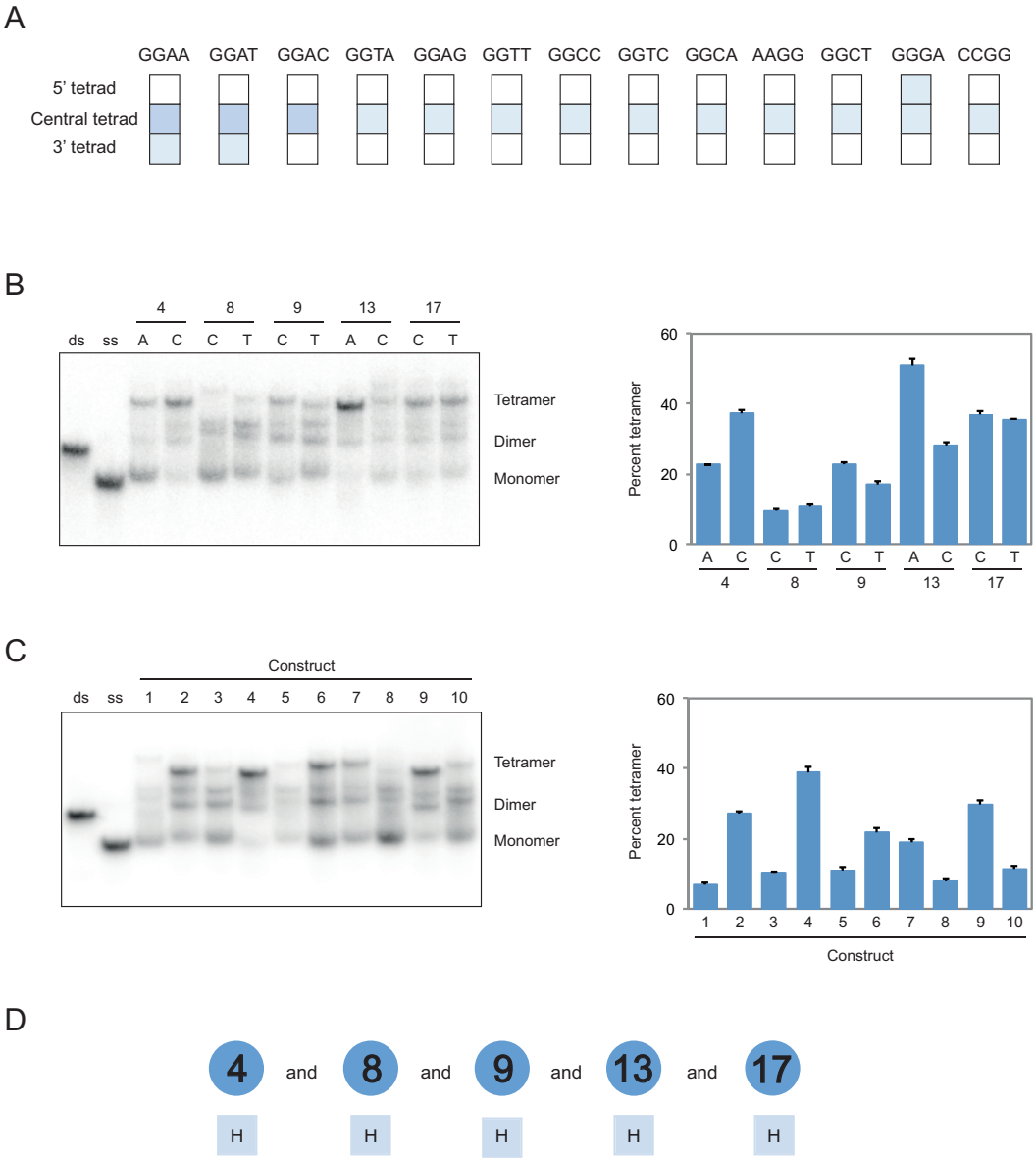
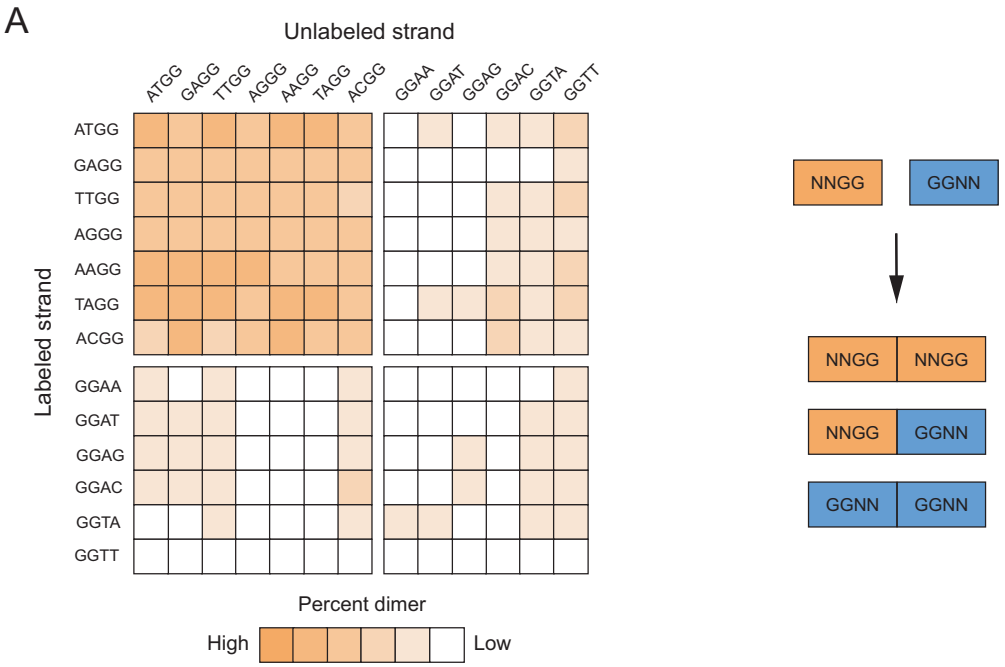


Figure 7. Sequence model for tetrameric G-quadruplexes. (**A**) Effect of transplanting mutations that induce tetramer formation when present in the central tetrad of the reference construct to the 5′ tetrad (nucleotides 1, 5, 10 and 14) or the 3′ tetrad (nucleotides 3, 7, 12 and 16). Colors match those in Figure 5. (**B**) Native gel and graph showing effects of point mutations in loops on tetramer formation. (**C**) Native gel and graph showing the ability of 10 randomly chosen sequences with up to two mutations in loops to form tetramers. Sequences are given in Supplementary Table S1. (**D**) Loop requirements of G-quadruplex variants that form tetramers. Numbers indicate the positions of loops in the reference construct, and letters below indicate the nucleotides that can occur at each position. Experiments were performed at 10 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1. H = A, C or T. Mutations in panels B and C were made in the context of a construct with a GGAG mutation in the central tetrad of the reference construct with the sequence GGGTGGGAAGAGTGGGA.

addition or removal of 5′ or 3′ nucleotides has no effect on dimer formation, we suggest that dimers are most likely to be intertwined structures (Figure 9A and B). On the other hand, the observation that the addition of a 5′ nucleotide to sequences that normally form tetramers causes them to form dimers instead suggests that tetramers are made up of two intertwined dimers stacked in a 5′ to 5′ arrangement (Figure 9C and D). In addition to explaining the effects of flanking nucleotides on dimer and tetramer formation, these models can rationalize several initially puzzling results described in this study. For example, they explain why sequences containing NNGG or GGNN mutations in the central tetrad of the reference construct form multimers but those containing NGNG, NGGN, GNNG or GNGN mutations do not: intertwined dimers containing mutations at positions 2 and 6 or at positions 11 and 15 will contain four consecutive GGGG tetrads while those containing mutations at other pairs of positions will presumably be less



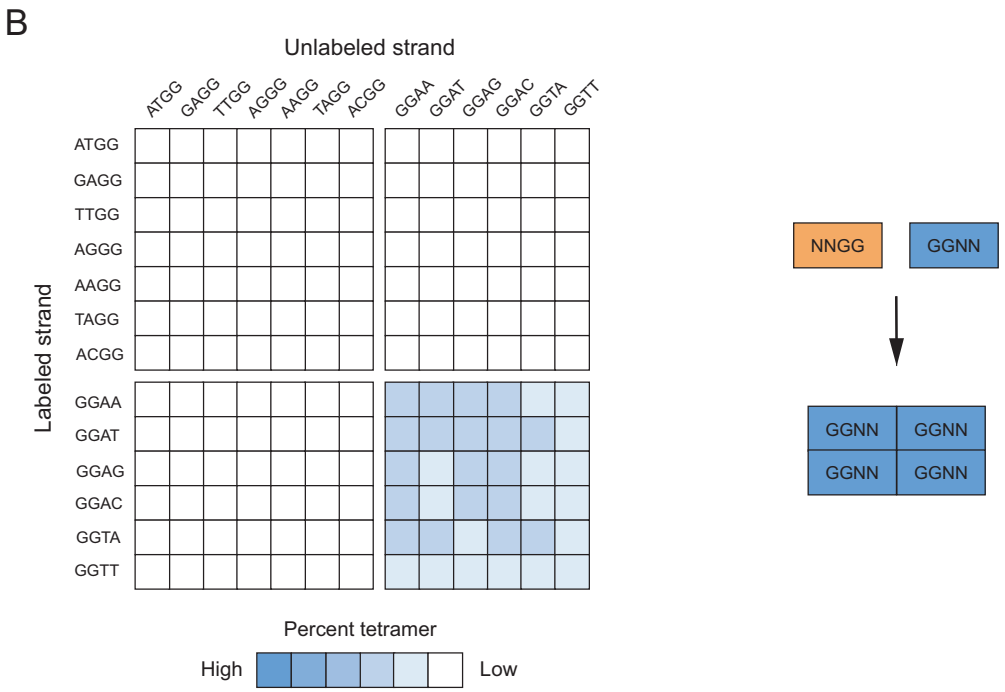


Figure 8. Formation of heteromultimeric G-quadruplexes from pairs of mutants with different sequences. (**A**) Sequence requirements of heterodimer formation. Left: heat map showing the ability of pairs of sequences with different mutations in the central tetrad of the reference construct to form heterodimers. Colors match those in Figure 4. Right: model describing the types of sequences that can form heterodimers. (**B**) Sequence requirements of heterotetramer formation. Left: heat map showing the ability of pairs of sequences with different mutations in the central tetrad of the reference construct to form heterotetramers. Colors match those in Figure 5. Right: model describing the types of sequences that can form heterotetramers. In each heat map, spaces are used to separate the seven dimer-forming sequences (on the left part of the x axis and the top part of the y axis) from the six tetramer-forming sequences (on the right part of the x axis and the bottom part of the y axis). Experiments were performed using 10 μM of the unlabeled G-quadruplex variant indicated on the x axis mixed with ≤10 nM of the radiolabeled G-quadruplex variant indicated on the y axis. All experiments were performed in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1.

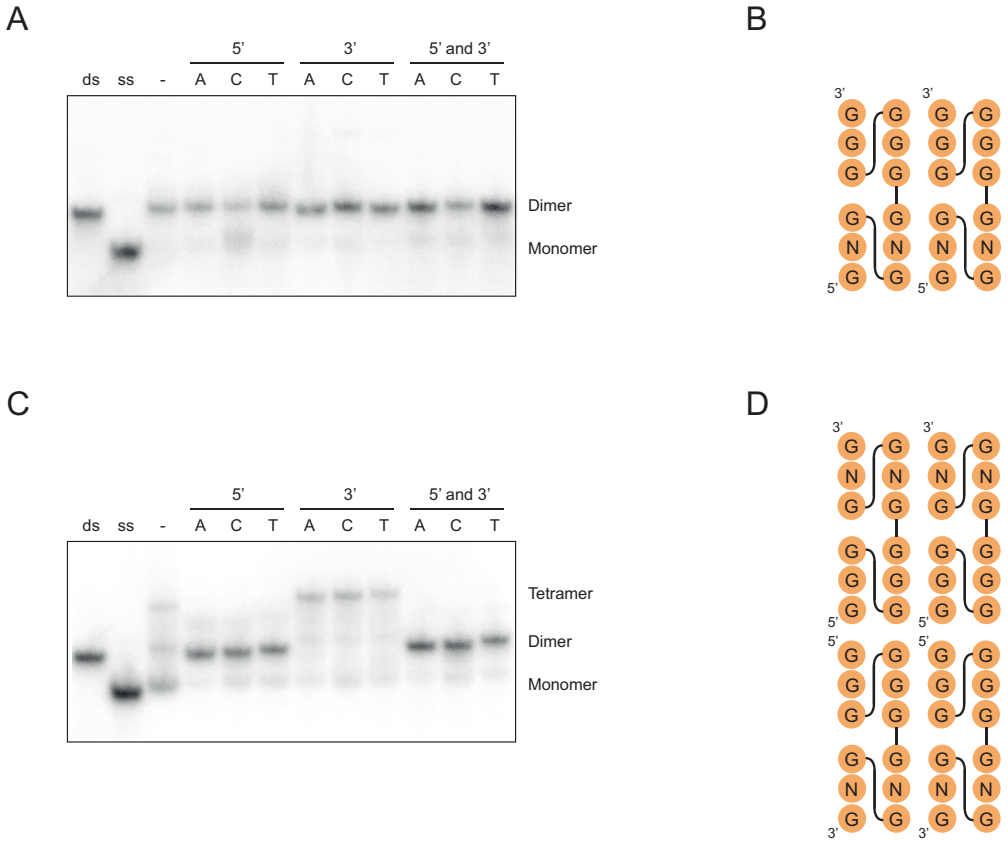


Figure 9. Characterization of multimer interfaces by site-directed mutagenesis. (**A**) Native gel showing the effect of flanking nucleotides on dimer formation. Experiments were performed using the sequences GAGTGGGAAGGGTGGG, xGAGTGGGAAGGGTGGG, GAGTGGGAAGGGTGGGx and xGAGTGGGAAGGGTGGGx (x = A, C or T). (**B**) Hypothetical secondary structure model of dimeric G-quadruplexes consistent with the experiments in panel A. (**C**) Native gel showing the effect of flanking nucleotides on tetramer formation. Experiments were performed using the sequences GGGTGGGAAGAGTGGG, xGGGTGGGAAGAGTGGG, GGGTGGGAAGAGTGGGx and xGGGTGGGAAGAGTGGGx (x = A, C or T). (**D**) Hypothetical secondary structure model of a tetrameric G-quadruplexes consistent with the experiments in panel C. Experiments were performed at 1 μM DNA concentration in a buffer containing 200 mM KCl, 1 mM $MgCl_2$ and 20 mM HEPES pH 7.1.

stable because they only contain two consecutive GGGG tetrads (Supplementary Figure S16). By similar reasoning, these models rationalize why G-quadruplexes containing NNGG mutations in the central tetrad of the reference construct form dimers while those containing GGNN mutations form tetramers: because NNGG mutations generate dimers containing an isolated 5′ tetrad, tetramers formed by such sequences are expected to be less stable than those formed by sequences containing GGNN mutations (Supplementary Figure S17). These models can also explain why, when a G-quadruplex variant containing an NNGG mutation in the central tetrad of the reference construct is mixed with one containing a GGNN mutation in the central tetrad of the reference construct, dimers sometimes form but tetramers do not: the interface in a tetramer formed by the 5′-5′ stacking of two intertwined dimers will be destabilized if it contains even a single molecule with an NNGG mutation (Supplementary Figure S18). Finally, these models explain why the sequence requirements of tetramer formation are in some cases a subset of those of dimer formation, but the converse in never true (Table 1): in addition to requiring the information needed to form an intertwined dimer, tetramers must also be capable of forming a second interface. Although consistent with our data, these models are preliminary, and we anticipate that high-resolution structural methods such as NMR and X-ray crystallography should be useful in testing them further.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Kateřina Švehlová, Filip Teplý, Václav Veverka, Tat'ána Majerová and others at the IOCB for useful discussions.

FUNDING

Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences Start-up Grant (to E.A.C.). Funding for open access charge: Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences.
Conflict of interest statement. None declared.

REFERENCES

1. Powers,E.T. and Powers,D.L. (2003) A perspective on mechanisms of protein tetramer formation. *Biophys. J.*, **85**, 3587–3599.
2. Marianayagam,N.J., Sunde,M. and Matthews,J.M. (2004) The power of two: protein dimerization in biology. *Trends Biochem. Sci.*, **29**, 618–625.
3. Ali,M.H. and Imperiali,B. (2005) Protein oligomerization: how and why. *Bioorg. Med. Chem.*, **13**, 5013–5020.
4. Griffin,M.D. and Gerrard,J.A. (2012) The relationship between oligomeric state and protein function. *Adv. Exp. Med. Biol.*, **747**, 74–90.
5. Mannige,R.V. and Brooks,C.L. III (2010) Periodic table of virus capsids: implications for natural selection and design. *PLoS One*, **5**, e9423.
6. Neves,S.R., Ram,P.T. and Iyengar,R. (2002) G protein pathways. *Science*, **296**, 1636–1639.
7. Watson,J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
8. Turner,D.H. (1996) Thermodynamics of base pairing. *Curr. Opin. Struct. Biol.*, **6**, 299–304.
9. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
10. Sen,D. and Gilbert,W. (1992) Novel DNA superstructures formed by telomere-like oligomers. *Biochemistry*, **31**, 65–70.
11. Marsh,T.C. and Henderson,E. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, **33**, 10718–10724.
12. Protozanova,E. and Macgregor,R.B. Jr (1996) Frayed wires: a thermally stable form of DNA with two distinct structural domains. *Biochemistry*, **35**, 16638–16645.
13. Mergny,J.L., De Cian,A., Amrane,S. and Webba da Silva,M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
14. Bardin,C. and Leroy,J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.
15. Rosu,F., Gabelica,V., Poncelet,H. and De Pauw,E. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids Res.*, **38**, 5217–5225.
16. Matsugami,A., Ouhashi,K., Kanagawa,M., Liu,H., Kanagawa,S., Uesugi,S. and Katahira,M. (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
17. Krishnan-Ghosh,Y., Liu,D. and Balasubramanian,S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.*, **126**, 11009–11016.
18. Kato,Y., Ohyama,T., Mita,H. and Yamamoto,Y. (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAGn) (n = 3-5). *J. Am. Chem. Soc.*, **127**, 9980–9981.
19. Kuryavyi,V., Phan,A.T. and Patel,D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
20. Borbone,N., Amato,J., Oliviero,G., D'Atri,V., Gabelica,V., De Pauw,E., Piccialli,G. and Mayol,L. (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Res.*, **39**, 7848–7857.
21. Do,N.Q., Lim,K.W., Teo,M.H., Heddi,B. and Phan,A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.
22. Trajkovski,M., da Silva,M.W. 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.
23. Wei,D., Todd,A.K., Zloh,M., Gunaratnam,M., Parkinson,G.N. and Neidle,S. (2013) Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc.*, **135**, 19319–19329.
24. D'Atri,V., Borbone,N., Amato,J., Gabelica,V., D'Errico,S., Piccialli,G., Mayol,L. and Oliviero,G. (2014) DNA-based nanostructures: the effect of the base sequence on octamer formation from d(XGGYGGT) tetramolecular G-quadruplexes. *Biochimie*, **99**, 119–128.
25. Podbevšek,P. and Plavec,J. (2015) KRAS promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units. *Nucleic Acids Res.*, **44**, 917–925.
26. Gao,S., Cao,Y., Yan,Y. and Guo,X. (2016) Sequence effect on the topology of 3 + 1 interlocked bimolecular DNA G-quadruplexes. *Biochemistry*, **55**, 2694–2703.
27. Laughlan,G., Murchie,A.I., Norman,D.G., Moore,M.H., Moody,P.C., Lilley,D.M. and Luisi,B. (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, **265**, 520–524.
28. Švehlová,K., Lawrence,M.S., Bednárová,L. and Curtis,E.A. (2016) Altered biochemical specificity of G-quadruplexes with mutated tetrads. *Nucleic Acids Res.*, **44**, 10789–10803.
29. Weiss,J.N. (1997) The Hill equation revisited: uses and misuses. *FASEB J.*, **11**, 835–841.
30. Curtis,E.A. and Liu,D.R. (2013) Discovery of widespread GTP-binding motifs in genomic RNA and DNA. *Chem. Biol.*, **20**, 521–532.
31. Kypr,J., Kejnovská,I., Renčiuk,D. and Vorlíčková,M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
32. Vorlíčková,M., Kejnovská,I., Sagi,J., Renčiuk,D., Bednářová,K., Motlová,J. and Kypr,J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.
33. Tóthová,P., Krafčíková,P. and Víglaský,V. (2014) Formation of highly ordered multimers in G-quadruplexes. *Biochemistry*, **31**, 8112–8119.
34. 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.
35. Yuan,G., Zhang,Q., Zhou,J. and Li,H. (2011) Mass spectrometry of G-quadruplex DNA: formation, recognition, property, conversion, and conformation. *Mass. Spectrom. Rev.*, **30**, 1121–1142.
36. Balthasart,F., Plavec,J. and Gabelica,V. (2013) Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.*, **24**, 1–8.
37. Chen,F.M. (1992) Sr2+ facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry*, **31**, 3769–3776.
38. Smirnov,I. and Shafer,R.H. (2000) Lead is unusually effective in sequence-specific folding of DNA. *J. Mol. Biol.*, **296**, 1–5.
39. Davis,J.T., (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
40. Wei,C., Tang,Q. and Li,C., (2008) Structural transition from the random coil to quadruplex of AG(3)(T(2)AG(3))(3) induced by Zn(2+). *Biophys. Chem.*, **132**, 110–113.
41. Guiset Miserachs,H., Donghi,D., Börner,N., Johannsen,S. and Sigel,R.K. (2016) Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *J. Biol. Inorg. Chem.*, **21**, 975–986.
42. Mullen,M.A., Assmann,S.M. and Bevilacqua,P.C. (2012) Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *J. Am. Chem. Soc.*, **134**, 812–815.
43. Kwok,C.K., Sherlock,M.E. and Bevilacqua,P.C. (2013) Decrease in RNA folding cooperativity by deliberate population of intermediates in RNA G-quadruplexes. *Angew. Chem. Int. Ed. Engl.*, **52**, 683–686.
44. Wilkie,A.O.M. (1994) The molecular basis of genetic dominance. *J. Med. Genet.*, **31**, 89–98.
45. Wang,Y. and Patel,D.J. (1992) Guanine residues in d(T_2AG_3) and d(T_2G_4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry*, **31**, 8112–8119.
46. Lech,C.J., Heddi,B. and Phan,A.T. (2013) Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res.*, **41**, 2034–2046.