International Journal of Biological Macromolecules 280 (2024) 136008

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

# The influence of G-tract and loop length on the topological variability of putative five and six G-quartet DNA structures in the human genome

Urša Štefan [a,1], Václav Brázda [b], Janez Plavec [a,c,d], Maja Marušič [c,*]

[a] Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia
[b] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 61265 Brno, Czech Republic
[c] Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia
[d] EN-FIST Center of Excellence, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

*Keywords:*
G-quadruplex
Long G-tracts
NMR spectroscopy

ABSTRACT

Local variation of DNA structure and its dynamic nature play an essential role in the regulation of important biological processes. One of the most prominent noncanonical structures are G-quadruplexes, which form *in vivo* within guanine-rich regions and have been demonstrated to be involved in the regulation of transcription, translation and telomere maintenance. We provide an analysis of G-quadruplex formation in sequences with five and six guanine residues long G-tracts, which have emerged from the investigation of the gapless human genome and are associated with genes related to cancer and neurodegenerative diseases. We systematically explored the effect of G-tract and loop elongations by means of NMR and CD spectroscopy and polyacrylamide electrophoresis. Despite both types of elongation leading up to structural polymorphism, we successfully determined the topologies of four out of eight examined sequences, one of which contributes to a very scarce selection of currently known intramolecular four G-quartet structures in potassium solutions. We demonstrate that examined sequences are incompatible with five or six G-quartet structures with propeller loops, although the compatibility with other loop types cannot be factored out. Lastly, we propose a novel approach towards specific G-quadruplex targeting that could be implemented in structures with more than four G-quartets.

## 1. Introduction

Beside the typical double-stranded right-handed form, DNA molecules are known to adopt various noncanonical structures that influence many fundamental cellular processes through the regulation of gene expression [1]. Among such structures that form *in vivo* are G-quadruplexes (G4) [2–6] that form in guanine-rich regions of both DNA and RNA [7–11]. G4s are intermolecular or intramolecular structures most often stabilised by cations, such as potassium or sodium ions, and π-π stacking interactions [12–16]. Each G4 consists of two or more G-quartets, planar structures of four guanine residues connected by Hoogsteen hydrogen bonds [17]. Sequences with the potential to form G4s predominate at telomeres, DNA replication origins, gene promoters, 5′-UTR regions of mRNA molecules, splice sites and somatic recombination sites in lymphocytes [18], which allows them to control transcription, translation, and influence genome stability [10,18,19]. Noteworthy, G4s often contribute to the regulation of genes associated with cancer [20–22] and the development of several neurodegenerative diseases [23–25].

Various algorithms revealed G4 forming sequences in all groups of living organisms, bacteria [26,27], archaea [28,29], eukaryotes [30,31], and viruses [32,33]. While shorter G-tracts are present across all kingdoms, longer G-tracts that harbour the potential to form G4s with more than three G-quartets are usually associated with eukaryotes - for example, they are highly abundant in humans [34]. Although a G4 is by definition any structure with stacked G-quartets, to the best of our knowledge, there is no structural characterisation of intramolecular structures with more than four G-quartets that occur in nature. However, our recent bioinformatic analyses of the gapless human reference genome GRCh38/hg38 [34] uncovered sequences with five- and six-guanine residues, which consequently have the potential to form G4s with five or six G-quartets. Many of these sequences are located in the promoter regions of genes that are responsible for fundamental cellular processes, such as signalling (WLS) [35], cell proliferation (GRB2) [36],

* Corresponding author.
*E-mail address:* marusic.maja@ki.si (M. Marušič).
[1] University of Copenhagen, Krystalgade 25, 1172 Copenhagen, Denmark.

https://doi.org/10.1016/j.ijbiomac.2024.136008
Received 8 April 2024; Received in revised form 23 September 2024; Accepted 23 September 2024
Available online 24 September 2024
0141-8130/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

regulation of development (HOXC12) [37] and mitophagy (CLEC16A) [38]. The changes in their expression levels can result in the onset and spreading of cancer, neurodegeneration and diabetes [39–41]. Moreover, we have shown that the human genome contains only 190 sequences that have the potential to form a G4 with five G-quartets and 22 that have the potential to form a G4 with six G-quartets [34]. Therefore, such structures would represent an extremely small fraction of all putative G4 forming sequences. Considering that one of the most common drawbacks of G4 specific targeting is the vast number of possible off-targets [42], the rarity of the putative five and six G-quartet structures and their consequent low number in combination with their localisation in the promoter regions of important genes could be extremely valuable assets in the search for novel drug targets. This possibility seems extremely attractive, especially since it is clear that the specific recognition of the large number of G4s in the genome will require the use of additional recognition elements beyond the most commonly used ones, such as planar G-quartets, loops and flanking regions [42,43]. As the number of stacked G-quartets increases, the length of the grooves also changes, providing a new additional binding element that could help discriminate between multi-quartet structures. We therefore undertook a structural study that would reveal the structural constraints and prerequisites for the formation of structures with more than four G-quartets, using a model system related to the sequences with G5 and G6 repeats found in the T2T genome [34].

Given the structural intricacy of G-quadruplexes, it is recommended that a combination of techniques be employed for comprehensive characterisation [44]. NMR spectroscopy provides a direct and clear monitoring system for the formation of a G-quadruplex structure, as the imino protons associated with guanines in G-quartets give rise to characteristic resonances around 10.5–12 ppm, a region completely separated from imino chemical shifts of any other DNA conformations [45]. NMR can offer definitive insights even when other methods yield inconclusive outcomes [46]. Moreover, two-dimensional (2D) NMR experiments can be utilised to ascertain the role of each nucleotide residue in the G-quadruplex structure and to perform detailed topological and structural analysis [47].

We analysed the influence of loop length and G-tract length on G4 formation and polymorphism using NMR spectroscopy in combination with CD spectroscopy and native gel electrophoresis. While we could not definitively disprove the formation of five or six G-quartet structures from the examined sequences, we can unequivocally assert that such structures are incompatible with propeller loops. We structurally characterised four out of eight examined oligonucleotides, while topology determination was not possible for the others due to severe structural polymorphism and oligonucleotide aggregation. Our results thus i) demonstrated the necessity of adjusting the loop length to at least 2 nucleotides ($G_{5/6}X_{2+}$ repeats) for the *in silico* analysis of potential five- and six G-quartet structures, ii) confirmed the preference of G4s for the formation of (3 + 1) or antiparallel structures as soon as the loop length allows it, and iii) highlighted the importance of carefully monitoring the folding conditions of G4s, especially in the case of sequences with a particularly high proportion of guanine residues.

## 2. Materials and methods

### 2.1. Oligonucleotide synthesis and sample preparation

All oligonucleotides were synthesised with DNA/RNA synthesiser H-8 from K&A Laborgeraete GbR using standard solid-phase phosphoramidite chemistry in a DMT-ON mode ($TG_6T$ in DMT-OFF mode) at 1 μmol scale. Phosphoramidites and CPG synthesis columns were purchased from Glen Research and Link Technologies, respectively. Site-specific labelled oligonucleotides were synthesised with an 8 % incorporation of $^{13}C$ and $^{15}N$ isotopically labelled phosphoramidites. Synthesised oligonucleotides were cleaved from the columns and deprotected in a 1:1 solution of $NH_4OH$ and $CH_3NH_2$ at room temperature and 65 °C. Oligonucleotides were purified with Glen-Pak™ Purification protocol by Glen Research and desalted twice on ÄKTA Purifier with HiPrep 26/10 Desalting column (GE Healthcare). Samples were dried with a vacuum centrifuge, diluted in MQ, and pH was adjusted to 7. Their concentration was determined with NanoDrop™ OneC by Thermo Scientific. Extinction coefficients were obtained from Oligo Analyzer™ Tool by Integrated DNA Technologies [48].

The final oligonucleotide concentration in the NMR samples was 0.1 mM for unlabelled and 0.3 mM for site-specific partially labelled oligonucleotides. G4-T1 sample for topology determination was prepared at 1.0 mM concentration. All samples contained 10 % $D_2O$. Different folding conditions were investigated with a gradual five-point increase in the concentration of KPi (pH 7.4) to the final 10 mM concentration, annealing, and quenching of previously titrated samples.

Both annealing and quenching protocol started with a 5-min-long incubation in a water bath at 95 °C. While annealed samples were then left in the water bath to slowly cool overnight, the samples designated for quenching were instead rapidly cooled by submersion in an ice bath.

### 2.2. NMR spectroscopy

All NMR spectra were recorded on Bruker AVANCE Neo 600 or 800 MHz NMR spectrometer at 25 °C unless otherwise specified. 1D proton spectra were recorded with 96 scans, spectral width 21.9429 ppm and interscan delay of 2 s. Excitation sculpting was used for the suppression of water signal. 1D $^{15}N$ HSQC spectra were recorded with 4096 scans, spectral width of 21.9429 ppm and interscan delay of 1 s. 2D $^{1}H—^{1}H$ NOESY spectra were recorded with 32 scans and 1024 increments in indirect dimension, spectral width of 21.9429, interscan delay of 1.8 s and mixing time of 250 ms. Spectra were processed and analysed using TopSpin 4.3.0 [49] and NMRFAM Sparky [50].

### 2.3. Gel electrophoresis

Denaturing gel electrophoresis was conducted on 20 % polyacrylamide gels with 7 M urea in TBE buffer. Before being loaded onto the gel, the samples were mixed with a loading buffer with formamide and incubated at 95 °C for 5 min. Similarly, 20 % polyacrylamide gels in TBE buffer were used for native gel electrophoresis. Gels with samples titrated to 10 mM KPi contained KCl to the final concentration of 10.3 mM. Gels in the denaturing conditions were stained with SYBR Gold (0.5 μL of stock solution for 10 min), while native gels were stained using Stains-all for 30 min and de-stained overnight. All imaging was conducted with Uvidoc HD6 by UVItec Cambridge.

### 2.4. CD spectroscopy

The CD spectra were measured using the CD JASCO J1500 spectrophotometer in a 0.1 mm demountable O-shaped cuvette at 25 °C in the wavelength range between 200 and 320 nm. The scanning rate was set to 20 nm $min^{-1}$, and the signal averaging time was 4 s. 0.1 mM NMR samples were used for CD measurements and all conditions were kept the same for the two techniques.

## 3. Results

### 3.1. Experimental design

Our starting point for the construction of a model system of G4s with more than four G-quartets is the oligonucleotide G3-T1, for which we have systematically extended the G-tracts to create the oligonucleotides G4-T1, G5-T1, G6-T1 (Table 1). G3-T1 represents a very suitable basis for testing stem elongation due to its high stability, its non-polymorphic nature and its regular topology with three G-quartets and three propeller loops with a single thymine residue [51].

**Table 1**
Oligonucleotides with respective names and sequences.

| Name | Sequence (5′ ➔ 3′) | length | G-tract length | loop length |
|---|---|---|---|---|
| $TG_6T$ | T *GGGGGG* T | 8 | 6 | / |
| G3-T1 | T GGG **T** GGG **T** GGG **T** GGG T | 17 | 3 | 1 |
| G4-T1 | T GGGG **T** GGGG **T** GGGG **T** GGGG T | 21 | 4 | 1 |
| G5-T1 | T GGGGG **T** GGGGG **T** GGGGG **T** GGGGG T | 25 | 5 | 1 |
| G6-T1 | T GGGGGG **T** GGGGGG **T** GGGGGG **T** GGGGGG T | 29 | 6 | 1 |
| G6-T2 | T GGGGGG **TT** GGGGGG **TT** GGGGGG **TT** GGGGGG T | 32 | 6 | 2 |
| G6-T3 | T GGGGGG **TTT** GGGGGG **TTT** GGGGGG **TTT** GGGGGG T | 35 | 6 | 3 |
| G6-T4 | T GGGGGG **TTTT** GGGGGG **TTTT** GGGGGG **TTTT** GGGGGG T | 38 | 6 | 4 |

The influence of loop length was furthermore tested on a series of oligonucleotides with six guanine residue tracts and a loop length between one and four thymine residues (G6-T1, G6-T2, G6-T3 and G6-T4). Thymine residues were chosen for the loops as they are the most inert option. While cytosine residues could compete for binding with guanine residues through Watson-Crick base pairing, adenine residues are known to stack strongly with guanine residues and consequently affect folding. Moreover, guanine residues can be both part of loops and G-quartets and therefore cannot be used for controlled extension of the stem. On the other hand, thymine residues do not interfere with G4 topology by binding and exhibit the least steric hindrance due to their size [52,53]. Additionally, single thymine residues were added to both ends of the oligonucleotides to prevent stacking of the structures.

Folding of all oligonucleotides was assessed with stepwise increases in potassium ion concentration and under different sample preparation conditions – annealing or quenching for isolation of thermodynamically or kinetically favoured species.

### *3.2. The effect of the length of G-tract on the topology of the formed G4 structure*

In the presence of potassium cations, G3-T1 is expected to form an intramolecular parallel G4 with three G-quartets and three one-residue propeller loops based on the literature data [51]. CD spectrum of G3-T1 in the presence of 10 mM KPi displays a profile in agreement with a parallel orientation of G-tracts with a minimum at 245 nm and a maximum 265 nm (Fig. 1A). Native gel electrophoresis at the same conditions displays a strong, fast migrating band, indicative of a monomolecular species, while two weak, slowly migrating bands can also be detected that are indicative of multimerisation (Fig. 1B). In full support of the multimerisation process is an observation of broad upfield shifted signals between 10.6 and 11.2 ppm in NMR spectra (Fig. 2A). In agreement with literature data [51], the imino region of the proton 1D NMR spectrum for oligonucleotide G3-T1 displayed 12 distinct and sharp peaks (Fig. 2A), which is characteristic of a G4 with three G-quartets. The spectra do not change significantly with increasing concentration of $K^+$ ions, suggesting that folding is mostly independent of potassium cation concentration (Fig. 2A), which is also evident from a native polyacrylamide gel (Supplementary Fig. 1).

Extending a G-tract length for one from G3-T1 to G4-T1 does not lead to the formation of four G-quartet structure. Instead, we observe twelve resolved peaks of high intensity in the imino region of 1D proton spectra upon the addition of KPi (Fig. 3A). CD spectrum (Fig. 1A) of G4-T1 at 10 mM KPi supports the parallel orientation of DNA strands. The formation of intermolecular structures is observed to a greater extent than in G3-T1, as evidenced by the presence of broad signals between 10.4 and 10.9 ppm in the NMR spectrum (Fig. 3A) and a large number of bands with low mobility on the acrylamide gel, in addition to the high mobility band (Fig. 1B).

The imino regions of the proton 1D NMR spectra of G5-T1 and G6-T1 (Supplementary Fig. 2) show several broad signals that simultaneously indicate the presence of structures with Hoogsteen hydrogen bonded imino protons, but also structural polymorphism. In full support, native acrylamide gel of G5-T1 and G6-T1 displays a large number of bands with different mobilities (Fig. 1B), and a reduction of the intensity of the band with the highest mobility after the addition of potassium cations

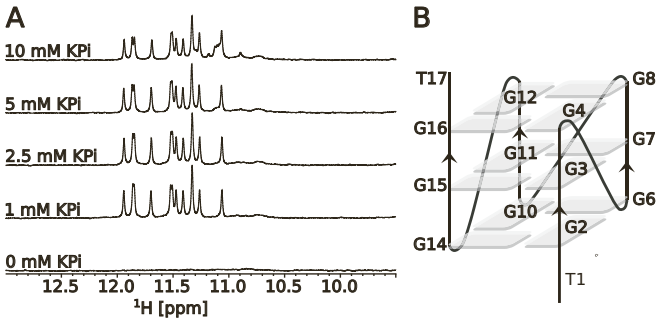


**Fig. 2. A** The imino region of proton 1D NMR spectra of a five-point titration of G3-T1 to the final 10 mM concentration of KPi. Spectra were recorded at 0.1 mM DNA, pH 7, 25 °C on a 600 MHz NMR spectrometer. **B** The topology of the G4 structure adopted by G3-T1 determined using a combination of 1D and 2D NMR methods and literature data [51].

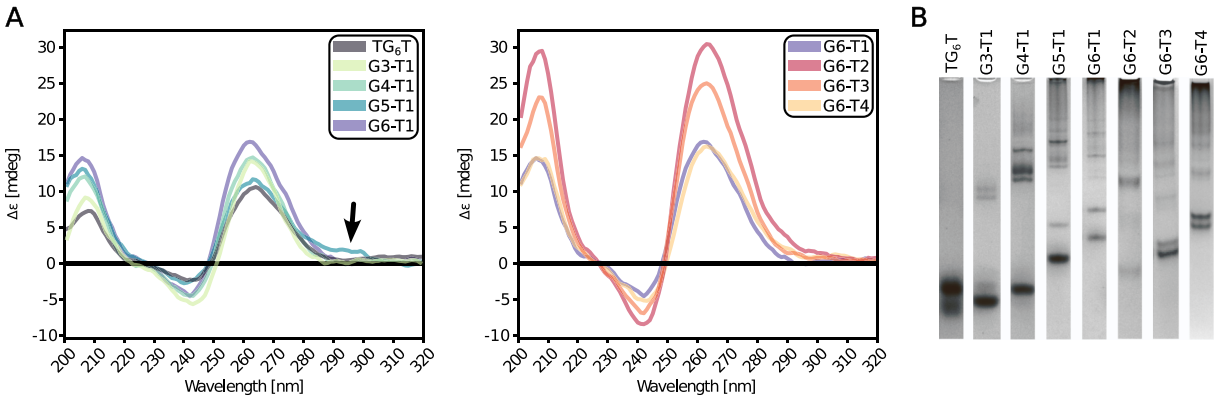


**Fig. 1. A** CD spectra of $G_n$-$T_n$ samples recorded at 0.1 mM oligonucleotide concentration and at 10 mM KPi concentration. G5-T1 was recorded after annealing. The arrow indicates a signal of low intensity at 290 nm for G5-T1. **B** Excerpts of native polyacrylamide gels at 10 mM KPi concentration. The complete gels are provided in Supplementary Fig. 1.

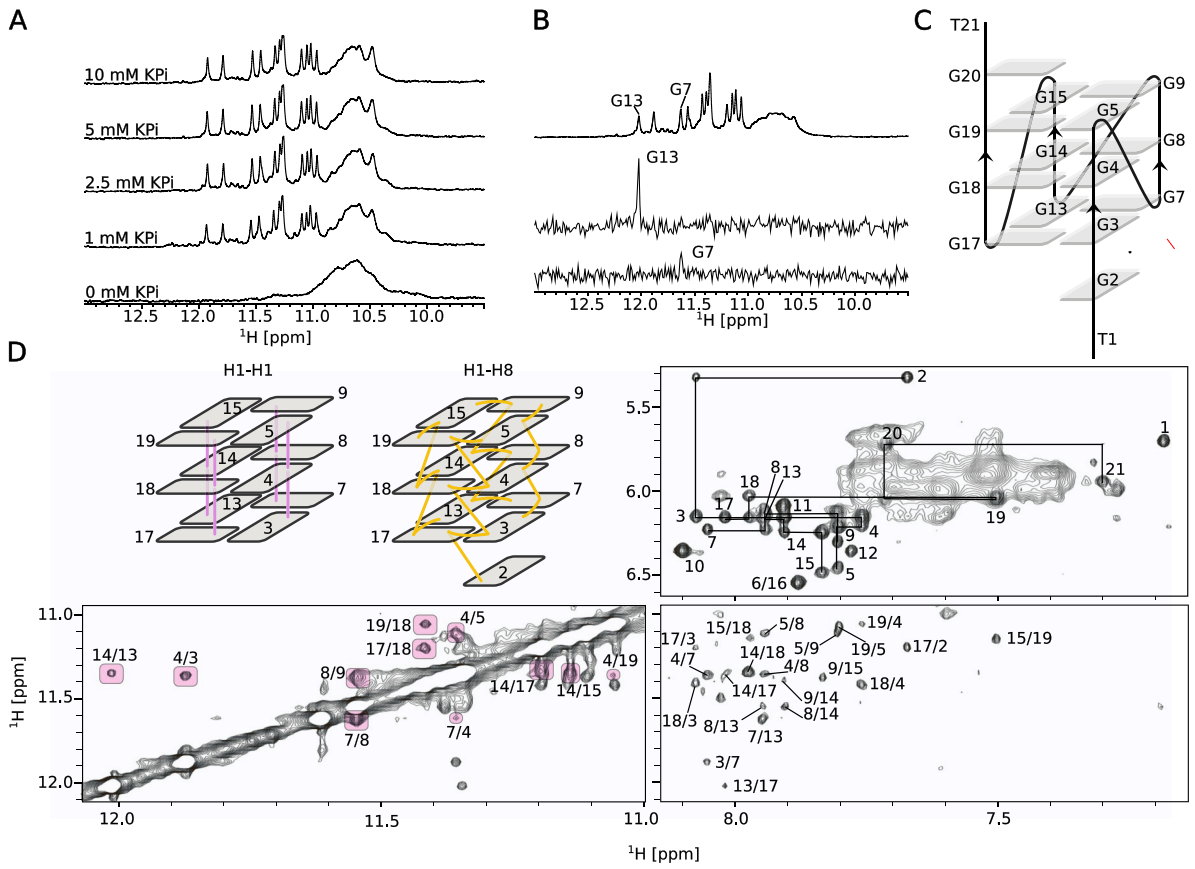


**Fig. 3.** **A** The imino region of proton 1D NMR spectra of five-step titration of G4-T1 to the final 10 mM concentration of KPi. **B** Proton and 1D $^{15}$N-edited spectra of oligonucleotides with site-specifically labelled G7 and G13 residues. **C** The topology of the G4 structure adopted by G4-T1. Residues outside the core are designated with rectangles if continuous stacking outside the G-quadruplex core has been observed. **D** Topology determination of G4-T1 (top left) based on the NOE connectivities in imino-imino (bottom left), imino-aromatic (bottom right) and anomeric-aromatic (top right) region in 2D NOESY spectrum. The anomeric-aromatic region of a 2D NOESY spectrum is displaying sequential connections for G2-G3-G4-G5, G7-G8-G9, G13-G14-G15 and G17-G18-G19-G20. The cross-peaks corresponding to the nucleotides in the loops with no sequential connectivities are labelled with numbers 1, 6, 10, 11, 12, and 16. Spectra were recorded at 0.1 or 1.0 mM oligonucleotide concentration, 10 mM KPi, pH 7.4 and 25 °C on a 600 MHz NMR spectrometer.

(Supplementary Fig. 1). Importantly, the bands with the lowest mobility hardly enter the gel, which suggests the presence of very large intermolecular assemblies that are most likely the result of aggregation. CD spectra of G5-T1 and G6-T1 display minimum at 243 nm and maximum at 263 nm (Fig. 1), which indicates that the molecular species of G5-T1 and G6-T1 are dominated by complexes with parallel DNA chain orientations. In summary, the broad signals in the imino region of the proton 1D NMR spectra for G4-T1, G5-T1 and G6-T1 spectra show that the elongation of the G-tract is associated with the diversification of the G4 topology. The three G-quartet topology of the G4 structure adopted by G4-T1 (Fig. 3A) suggests that some guanine residues in the G-tracts may contribute to the loops, ultimately supporting greater topological diversity, an effect that is likely to be enhanced in oligonucleotides with even longer G-tracts.

### 3.3. Analysis of parallel G3-T1 and G4-T1 structures

To further investigate the predominant topology of G3-T1 we recorded 2D $^{1}$H—$^{1}$H NOESY NMR spectra with $\tau_m$ of 250 ms (Supplementary Fig. 3). The combination of 1D and 2D spectra together with literature data allowed us to confirm the expected three G-quartet parallel topology of G3-T1 with one-nucleotide propeller loops and all guanine residues in *anti* conformation. We followed sequential H2′/″-H8 and H1′-H8 contacts along G-tracts: Ga-Gb-Gc, Gp-Gr-Gs and G14-G15-G16 and G2-G3 (Supplementary Fig. 3). The first and the last G-tracts were unambiguously assigned through sequential connectivities with flanking residues (*i.e.*, T1-G2-G3 and G14-G15-G16-T17). The remaining two G-tracts labelled as a-b-c and p-r-s represent G-tracts G6-G7-G8 and G10-G11-G12, but cannot be unequivocally assigned due to lack of cross-peaks correlating them with other structural elements. The topology of G3-T1 is depicted in Fig. 2B.

The extension of a G-tract from three to four guanine residues in G4-T1 did not allow the formation of a structure with four G quartets. In this way, additional guanine residues are incorporated into loops that allow the formation of many different structures due to the possibility of register shifting. The 2D NOESY spectrum G4-T1 shows that there are no residues in the *syn* conformation along the glycosidic bond (Fig. 3D), indicating the existence of a parallel structure. In addition, we can follow the sequential NOE connectivities from T1 to G5, G7 to G9, and from G17 to T21 (Fig. 3D), which means that there is no propeller loop in this part of the structure, as this would lead to the breaking of the sequential connectivities. From this it can be concluded that the T1-G2 and G20-T21 ends of the molecule are stacked on the final G-quarters and that G2 and G20 therefore do not belong to the core of the structure. Based on all these data, four parallel topologies are possible (Supplementary Fig. 4). In order to determine which of the proposed topologies is adopted by G4-T1, we used $^{15}$N-HSQC spectra with isotopically site-specifically labelled oligonucleotides at positions 7 and 13 (Fig. 3B and Supplementary Fig. 4). The identification of G7 proton signal at 11.63 ppm excludes topologies 2–1-2 and 2–2-1 (Fig. 3B and Supplementary Fig. 4B), where G7 is not part of a G-quartet core. Furthermore, the assignment of the G13 imino proton at 12.02 ppm (Fig. 3B)

confirmed G13 is included in the G4 core. Using the deuterium exchange experiment, we were able to demonstrate that G13 is a constituent of one of the non-central G-quartets since the signal for the G13 imino proton quickly disappeared after the exchange from $H_2O$ to $D_2O$. Using NOE contacts in the imino-imino, imino-aromatic and anomeric-aromatic region of 2D NOESY spectrum (Fig. 3D) together with the unambiguous assignment of G7 and G13 we established that G4-T1 adopts a 1–3-1 parallel topology with residues T6, G10-T11-G12 and T16 constituting three propeller loops (Fig. 3C). Chemical shift of protons of residues in loops matches well with their expected structural position: T6 and T16 in one nucleotide propeller loop are downfield shifted, in agreement with them being exposed to the solvent and not stacked within the structure. T1-G2 and G20-T21 in 5′ and 3′ flanking regions, respectively, are on the other hand upfield shifted, as is expected because they are involved in continuous stacking of the first and the last G-tract. Residues G10-T11-G12 of the central propeller loop match well the with chemical shift pattern observed for G10-C11-T12 central propeller loop in 1–3-1 parallel topology of GTERT-060 G4 [54]. Their chemical shift indicates involvement in stacking interaction at least for T11 and G12. Weak NOE contacts between the methyl group of T11 with both guanine residues of the loop indicate certain interactions within the loop.

Unlike the other two possible topologies (Supplementary Fig. 4C), the final 1–3-1 topology of G4-T1 exhibits two one-residue propeller loops, which are one of the most stabilising elements of G4 structures and which aligns completely with our current knowledge of loop length preference in relationship to structure stability [52].

### 3.4. Annealing partially resolves structural polymorphism for G5-T1

To examine if different sample preparations aid in resolving the observed structural polymorphisms for G5-T1 and G6-T1, the titrated samples were either annealed or quenched. In contrast to G6-T1, where the imino region of the proton 1D NMR spectra before and after annealing or quenching showed similar features, the proton 1D spectra of annealed G5-T1 differed significantly from those before annealing (Fig. 4A). The spectrum for G5-T1 displays two groups of heavily overlapped peaks with the line-shape characteristic of the monomeric species, while the broader signals between 10.30 and 10.75 ppm indicate the presence of intermolecular structures. These results are consistent with native gel electrophoresis, where the fastest moving band of an annealed oligonucleotide exhibited significantly higher mobility than an oligonucleotide prepared under non-annealing conditions (Fig. 4B). Numerous bands seen in native polyacrylamide gel (Fig. 4B), as well as several additional signals of low intensity in the imino region of 1D protons spectra of G5-T1 after annealing indicate that the monomeric species, albeit predominant, is only one of the many possible forms that G5-T1 adopts. Hence, any experimental technique

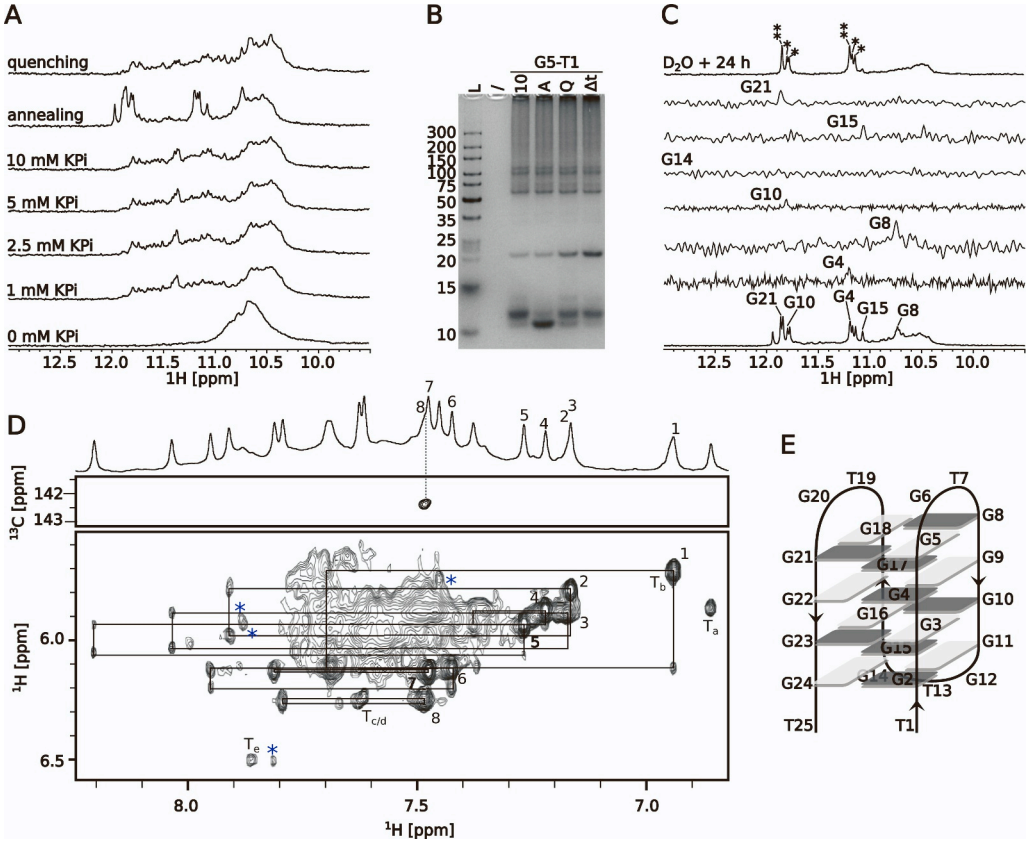


**Fig. 4.** **A** The imino region of proton 1D NMR spectra of G5-T1 at increasing potassium concentration, and after quenching and annealing. **B** Polyacrylamide gel electrophoresis at 10 mM KCl. The structure formed after annealing (A) displays significantly higher mobility than G5-T1 after titration (10), quenching (Q) or two-week long incubation in KPi (Δt). **C** The partial assignment of G5-T1 imino region with the help of site-specifically isotopically labelled oligonucleotides (bottom), and the imino region of 1D proton spectra 24 h after an exchange of solvent from $H_2O$ to $D_2O$ (top). 1D $^{15}$N-edited HSQC and 1D $^{1}$H spectra were recorded at 0.3–0.4 mM DNA concentration, pH 7.4 and 35 °C on an 800 MHz NMR spectrometer. $^{15}$N-edited HSQC spectra for G4 and G10 were recorded at 25 °C. **D** 2D $^{13}$C HSQC spectrum of site-specifically isotopically labelled G5-T1 at position 21 recorded at 35 °C with corresponding $^{1}$H proton spectra (top) and anomeric-aromatic region of 2D NOESY spectrum (bottom) with the labelled sequential connectivities of eight *syn-anti* pairs. Guanine residues in *syn* conformation are labelled with numbers from 1 to 8. Five thymine residues are labelled as Ta-e, while four peaks designating the remaining four guanine residues in the loops are marked with blue asterisks. 2D NOESY spectrum with the mixing time of 250 ms was recorded at 0.44 mM oligonucleotide concentration, 10 mM KPi, pH 7.4 and 35 °C on an 800 MHz NMR spectrometer. **E** The tentative G4 topology of G5-T1. Guanine residues in *syn* and *anti* conformation are presented with dark and light grey rectangles, respectively.

that cannot distinguish the contribution of individual species, will report on average characteristic of an assembly. Nevertheless, the CD spectrum of G5-T1 differs from the CD spectrum of a parallel structure, such as G3-T1 or G4-T1 (Fig. 1), as it displays a shoulder at 290–300 nm (labelled with an arrow in Fig. 1A), which is indicative of a contribution from either antiparallel or (3 + 1) species.

To determine the topology of G5-T1 we used NMR spectroscopy since NMR signals – at least in the regions that are not heavily overlapped – can be attributed to different species based on the signal shape and connectivities. We employed a concerted use of a $D_2O$ exchange experiment, 2D NOESY spectrum, and isotopic labelling of a subset of guanine residues to establish characteristics of G5-T1 topology. After $H_2O$ to $D_2O$ exchange, eight partially overlapping signals were still visible in the imino region after 24 h (Fig. 4C, top), indicating that we observe a four G-quartet structure in which the two central G-quartets are extremely well protected from the solvent. In 2D NOESY spectra we identified the sequential contacts of eight *syn-anti* guanine residue pairs (Fig. 4D), which indicates that the topology of G5-T1 is antiparallel with characteristic *syn-anti-syn-anti* distribution of guanine residues along the G-tract. Moreover, the isotopic labelling unambiguously identified that all labelled guanine residues (G4, G8, G10, G15, G21) except G14 are located in the G-quartets, which allowed us to exclude all antiparallel topologies with four G-quartets where G14 is part of G-quartets and G8 is not (Supplementary Fig. 5). Considering the combined information from the deuterium exchange experiment, 2D NOESY and isotopic labelling suggesting that G21 adopts *syn* conformation along the glycosidic bond (Fig. 4D, top), the remaining topologies can be further narrowed down to seven possible topologies (Supplementary Fig. 6C). Finally, analysis of additional cross-peaks in NOESY suggests that the antiparallel topology with three edgewise loops is the one most likely to be adopted by G5-T1 (Fig. 4E). However, due to the heavy overlap in the imino region and the presence of other species, spectral characteristics within the current setup did not allow us to assign NOE contacts needed for complete topological analysis, which is currently undergoing.

### 3.5. Loop length variation leads to formation of multiple structures

The effect of loop length was examined through gradual loop-length elongation within G6 repeat sequence. We designed four sequences with the G-tract length of six, where the number of thymine residues constituting the loops increased from one to four (Table 1). All samples were subjected to five-point titration to the final KPi concentration of 10 mM, annealing and quenching. All oligonucleotides (G6-T1, G6-T2, G6-T3, and G6-T4) regardless of sample preparation gave rise to multiple broad signals in the imino region of 1D proton NMR spectra (Supplementary fig. 6) indicating the presence of multiple structures. In agreement, several bands of different mobility are observed on native polyacrylamide gels of G6-T1, G6-T2, G6-T3 and G6-T4 (Fig. 1B). CD spectra (Fig. 1A) display profiles characteristic of parallel species with a minimum at 243 nm and a maximum at 263 nm. However, the shoulder in the region from 290 to 300 nm for G6-T2, G6-T3 and G6-T4 suggests that the presence of some non-parallel species cannot be excluded. Because one thymine residue loops are too short to accommodate a structure with six G-quartets, the guanine residues in G-tracts participate in loops instead of G-quartets leading to structural polymorphism. Similarly, the presence of multiple signals in the imino region of proton 1D NMR spectra for G6-T2, G6-T3 and G6-T4 can most likely be attributed to structures with longer loops allowing for more topological diversity.

### 3.6. Six G-quartet structure without structural constraints in loops

We analysed the formation of an intermolecular G4 with six G-quartets to show that loops are the limiting factor for the formation of six G-quartet structures in intramolecular G4s. Before titration and the acquisition of NMR spectra the length of all samples in this study was tested with denaturing gel electrophoresis. Except for $TG_6T$ all oligonucleotides travelled as expected for a denatured DNA strand of a given length (Supplementary Fig. 7). $TG_6T$, however, travelled anomalously slowly, which could be due to its extreme stability and incomplete unfolding under denaturing conditions. After titration with potassium phosphate buffer, the imino region of proton 1D NMR spectra for $TG_6T$ displays six sharp peaks (Fig. 5), demonstrating that with an appropriate G-tract length and in the absence of all structural restraints in the loops, a formation of a parallel G4 structure (Fig. 1A) with six G-quartets is possible.

## 4. Discussion

In the present study, we investigated the effects of loop length and G-tract length on G4 topology in the presence of potassium ions for the oligonucleotides with the potential to form five or six G-quartet structures. Our approach included gradual G-tract elongation from G3-T1 to G6-T1. Altogether, we determined three G-quartet topologies for G3-T1 and G4-T1, and a four G-quartet topology for G5-T1. While one-thymine residue loops easily accommodate the distance and flexibility requirements for the three G-quartet structure of G3-T1, they are not flexible enough to allow G4-T1 to fold into a four G-quartet topology with three propeller loops. Concretely, it has been shown that the sharp turning angle of 136° needed to span four G-quartet layers is incompatible with native linkers with the length of one or two residues [55]. Furthermore, a CD and ESI-MS analysis [56] of $TG_nT$ ($n$ = 14–20) sequences revealed that the oligonucleotides with long G-tracts tend to fold into structures with lower molecularity than expected, indicating that the number of guanine residues in the G-tract does not necessarily equal the number of G-quartets that ultimately form. Our results for G4-T1, G5-T1 and G6-T1 extend this finding to the sequences with four shorter G-tracts. The inability of one-thymine residue loops to accommodate a G4 structure with number of G-quartets equal to G-tract length causes the guanine residues to participate in the loop formation, ultimately resulting in the increase in the number of possible topologies. We uniquely resolved such polymorphism for G5-T1, which adopts a predominant four G-quartet structure after annealing. Importantly, the final topology of G5-T1 adds to less than a handful [57–59] of intramolecular antiparallel four G-quartets structures in the presence of potassium we know of so far. The determination of the precise topology of G5T is beyond the scope of this work, as our results indicate that the large number of guanine residues and the spectral overlap of the signals

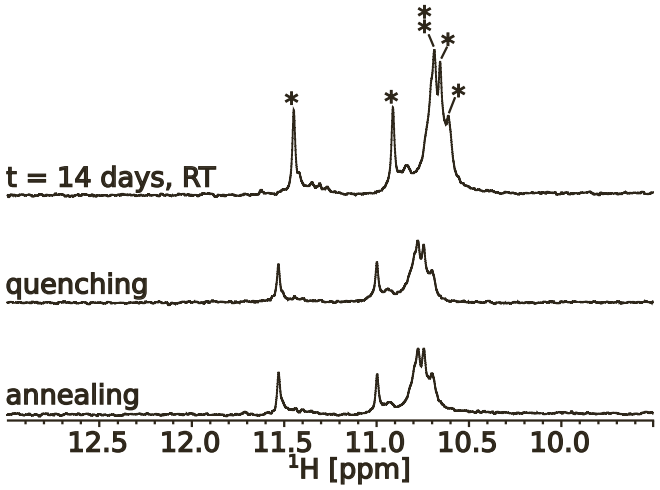


**Fig. 5.** Proton NMR spectra after titration and folding in different conditions for $TG_6T$. All spectra display six intensive peaks representing 24 guanines in six G-quartets. Spectra were recorded at 0.1 mM DNA concentration, pH 7.4 and 25 °C with a 600 MHz NMR spectrometer.

require $^{15}N$ labelling of at least part of the remaining 14 guanine residues. In addition, both the native polyacrylamide gels and NMR spectra show that a significant proportion of the molecules are involved in the formation of higher order structures, which could not be avoided with the current sample preparation and relatively high oligonucleotide concentration needed for the NMR measurements. However, examples from the literature confirm that fine-tuning conditions such as salt concentration, temperature and folding mode can isolate a species to a sufficient extent, and this is a direction in which we will be working in the future.

The effect of loop length on G4 topology was investigated with oligonucleotides G6-T1, G6-T2, G6-T3 and G6-T4. The predominant topologies for these oligonucleotides could not be determined due to the structural polymorphism indicated by the numerous signals in 1D proton NMR spectra. Our data confirms that longer loops allow oligonucleotides to accommodate multiple topologies [52,60,61]. Among those topologies the structures with long total loop lengths most likely fold into non-parallel topologies [52,62], where the increased entropic cost needed to stabilise the loops can be balanced with the interactions within them [52].

Interestingly, both G-tract and loop elongation ultimately resulted in structural polymorphism, which most likely occurred due to the increasing loop length either as a consequence of actual loop elongation or the inclusion of guanine residues in the loops. The presence of multiple topologies adopted by the oligonucleotides with longer loops is represented by numerous signals in the imino regions of proton 1D NMR spectra. On the other hand, the frequently observed broadening of signals is likely connected with aggregation of molecules with long G-tracts. While we unequivocally show that the examined sequences cannot form a five or six G-quartet structures with propeller loops, we cannot indisputably exclude the existence of such structures with non-propeller loops. On the other hand, the presence of multiple G4 structures in equilibrium might point to the potential dynamic nature of such species that actualises through G-strand slipping and base-swapping [63–65]. Such inter-convertibility of G4 structures could represent a type of biological regulation [66] implying the importance of a tight control over transcription and translation of related genes.

In general, due to their position and function in promoter and other regulatory regions of important genes, G4s are deemed as a very desirable therapeutic target [22,67,68]. However, due to their common G-quartet core and a limited number of topologies for thousands of putative G4-forming sequences present in the human genome, specific targeting has proved to be a considerable challenge [43]. Most G4 ligands belong to small molecules, such as acridine derivatives, porphyrins and their derivatives, natural G4 ligands (telomestatin) and certain fluoroquinolone antibiotics [69], however, additional alternative approaches have been explored as well [43,69–72]. No matter the type of the molecule they all exploit the same recognition sites – most commonly the accessible surfaces of the G-quartet core, but also loops and flanking sequences [43] – leaving out an important and possibly highly specific element, which is G4 groove length. The length of groove is rarely considered as a recognition element worth exploring, most likely because there is no evidence of *in vivo* G4 structures with more than four G-quartets that would have a significantly longer stem. Nonetheless, due to the rarity of putative five or six G-quartet G-rich sequences in the human genome and the consequent lack of possible off-targets, this might prove to be a crucial advantage for future specific targeting attempts in the context of novel drug development. With our consequent work we hope to identify and isolate structures with more than four G-quartets that will pave the way for groove-length specific binding, facilitate the identification of these structures within the cell and contribute our understanding of their role in cellular processes.

In conclusion, our systematic study of the effect of loop and G-tract elongation uncovered important effects of loop and G-tract length on topology for potential intramolecular five and six G-quartet structures. This knowledge may contribute to the search for novel G4 forms *in vitro* as well as *in vivo*. Finally, we propose an unexplored approach for specific G4 targeting using the G4 groove length as the recognition element, which might facilitate the future therapeutic application in the case of five and six G-quartet structures.

## CRediT authorship contribution statement

**Urša Štefan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Václav Brázda:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Janez Plavec:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Maja Marušič:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgements*

*Funding sources*

This work was supported by the Slovenian Research and Innovation Agency (ARIS) [grants no. P1-0242 (J.P.) and Z1-2636 (M.M.)], and of the Czech Science Foundation [grant no. 22-21903S (V.B.)]. The authors would moreover like to thank CERIC-ERIC consortium for the access to experimental facilities and financial support [project no. 20227050].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2024.136008.

## References

[1] I. Georgakopoulos-Soares, J. Victorino, G.E. Parada, V. Agarwal, J. Zhao, H. Y. Wong, M.I. Umar, O. Elor, A. Muhwezi, J.Y. An, S.J. Sanders, C.K. Kwok, F. Inoue, M. Hemberg, N. Ahituv, High-throughput characterization of the role of non-B DNA motifs on promoter function, Cell Genomics 2 (2022) 100111.
[2] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, Trends Cell Biol. 19 (2009) 414–422.
[3] C.C. Chang, I.C. Kuo, I.F. Ling, C.T. Chen, H.C. Chen, P.J. Lou, J.J. Lin, T.C. Chang, Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative, Anal. Chem. 76 (2004) 4490–4494.
[4] E.Y.N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, G-quadruplex structures are stable and detectable in human genomic DNA, Nat. Commun. 4 (2013) 1796.
[5] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H.J. Lipps, A. Plückthun, In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei, Proc. Natl. Acad. Sci. USA 98 (2001) 8572–8577.
[6] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, 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. USA 99 (2002) 11593–11598.
[7] J. Kim, C. Cheong, P.B. Moore, Tetramerization of an RNA oligonucleotide containing a GGGG sequence, Nature 351 (1991) 331–332.
[8] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, Nature 334 (1988) 364–366.
[9] W.I. Sundquist, S. Heaphy, Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA, Proc. Natl. Acad. Sci. 90 (1993) 3393–3397.
[10] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, The regulation and functions of DNA and RNA G-quadruplexes, Nat. Rev. Mol. Cell Biol. 21 (2020) 459–474.
[11] L. Dumas, P. Herviou, E. Dassi, A. Cammas, S. Millevoi, G-Quadruplexes in RNA biology: recent advances and future directions, Trends Biochem. Sci. 46 (2021) 270–283.
[12] K. Sato, P. Knipscheer, G-quadruplex resolution: from molecular mechanisms to physiological relevance, DNA Repair (Amst) 130 (2023) 103552.
[13] D. Sen, W. Gilbert, A sodium-potassium switch in the formation of four-stranded G4-DNA, Nature 344 (1990) 410–414.

[14] J.R. Williamson, M.K. Raghuraman, T.R. Cech, Monovalent cation-induced structure of telomeric DNA: the G-quartet model, Cell 59 (1989) 871–880.
[15] J. Spiegel, S. Adhikari, S. Balasubramanian, The structure and function of DNA G-quadruplexes, Trends Chem 2 (2020) 123–136.
[16] W.I. Sundquist, A. Klug, Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, Nature 342 (1989) 825–829.
[17] M. Gellert, M.N. Lipsett, D.R. Davies, Helix formation by guanylic acid, Proc. Natl. Acad. Sci. USA 48 (1962) 2013.
[18] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology, Nucleic Acids Res. 43 (2015) 8627–8637.
[19] J. Robinson, F. Raguseo, S.P. Nuccio, D. Liano, M. Di Antonio, DNA G-quadruplex structures: more than simple roadblocks to transcription? Nucleic Acids Res. 49 (2021) 8419–8431.
[20] S. Balasubramanian, L.H. Hurley, S. Neidle, Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat. Rev. Drug Discov. 10 (2011) 261.
[21] J. Eddy, N. Maizels, Gene function correlates with potential for G4 DNA formation in the human genome, Nucleic Acids Res. 34 (2006) 3887–3896.
[22] N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, G-quadruplexes: a promising target for cancer therapy, Mol. Cancer 2021 20:1 20 (2021) 1–18.
[23] J.W. Cave, D.E. Willis, J.W. Cave, D.E. Willis, G-quadruplex regulation of neural gene expression, FEBS J. 289 (2022) 3284–3303.
[24] E. Wang, R. Thombre, Y. Shah, R. Latanich, J. Wang, G-Quadruplexes as pathogenic drivers in neurodegenerative disorders, Nucleic Acids Res. 49 (2021) 4816–4830.
[25] J. Xu, H. Huang, X. Zhou, G-quadruplexes in neurobiology and virology: functional roles and potential therapeutic approaches, JACS Au 1 (2021) 2146–2161.
[26] M. Bartas, M. Cutová, V. Brázda, P. Kaura, J. ŠtAstný, J. Kolomazník, J. Coufal, P. Goswami, J. Červeň, P. Pečinka, The presence and localization of G-quadruplex forming sequences in the domain of bacteria, Molecules 24 (2019) 1711.
[27] Y. Ding, A.M. Fleming, C.J. Burrows, Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes, Sci. Rep. 8 (2018) 15679.
[28] V. Brázda, Y. Luo, M. Bartas, P. Kaura, O. Porubiaková, J. Šťastný, P. Pečinka, D. Verga, V. Da Cunha, T.S. Takahashi, P. Forterre, H. Myllykallio, M. Fojta, J. L. Mergny, G-quadruplexes in the archaea domain, Biomolecules 10 (2020) 1–23.
[29] G.V. Chashchina, A.K. Shchyolkina, S.V. Kolosov, A.D. Beniaminov, D.N. Kaluzhny, Recurrent Potential G-Quadruplex Sequences in Archaeal Genomes, Front. Microbiol. 12 (2021) 647851.
[30] V.R. Yella, A. Vanaja, Computational analysis on the dissemination of non-B DNA structural motifs in promoter regions of 1180 cellular genomes, Biochimie 214 (2023) 101–111.
[31] A. Vannutelli, L.L.N. Schell, J.P. Perreault, A. Ouangraoua, GAIA: G-quadruplexes in alive creature database, Nucleic Acids Res. 51 (2023) D135–D140.
[32] N. Bohálová, A. Cantara, M. Bartas, P. Kaura, J. Šťastný, P. Pečinka, M. Fojta, J. L. Mergny, V. Brázda, Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection, Biochimie 186 (2021) 13–27.
[33] E. Lavezzo, M. Berselli, I. Frasson, R. Perrone, G. Palù, A.R. Brazzale, S.N. Richter, S. Toppo, G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide, PLoS Comput. Biol. 14 (2018) e1006675.
[34] M. Bartas, V. Brázda, V. Karlický, J. Červeň, P. Pečinka, Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences, Biochimie 150 (2018) 70–75.
[35] C. Bänziger, D. Soldini, C. Schütt, P. Zipperlen, G. Hausmann, K. Basler, Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells, Cell 125 (2006) 509–522.
[36] E.J. Lowenstein, R.J. Daly, A.G. Batzer, W. Li, B. Margolis, R. Lammers, A. Ullrich, E.Y. Skolnik, D. Bar-Sagi, J. Schlessinger, The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling, Cell 70 (1992) 431–442.
[37] R.L. Peterson, T. Papenbrock, M.M. Davda, A. Awgulewitsch, The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions, Mech. Dev. 47 (1994) 253–260.
[38] H. Hakonarson, S.F.A. Grant, J.P. Bradfield, L. Marchand, C.E. Kim, J.T. Glessner, R. Grabs, T. Casalunovo, S.P. Taback, E.C. Frackelton, M.L. Lawson, L.J. Robinson, R. Skraban, Y. Lu, R.M. Chiavacci, C.A. Stanley, S.E. Kirsch, E.F. Rappaport, J. S. Orange, D.S. Monos, M. Devoto, H.Q. Qu, C. Polychronakos, A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene, Nature 448 (2007) 591–594.
[39] G.M.H. Abdel-Salam, H.H. Afifi, M.S. Abdel-Hamid, N.E.B. Ahmed, M.B. Taher, G. El-Kamah, H. Thiele, P.N. Nürnberg, H.J. Bolz, Expanding the phenotypic spectrum and clinical severity associated with WLS gene, J. Hum. Genet. 68 (2023) 607–613.
[40] A. Giubellino, T.R. Burke, D.P. Bottaro, Grb2 signaling in cell motility and cancer, Expert Opin. Ther. Targets 12 (2008) 1021.
[41] I.K. Jang, J. Zhang, Y.J. Chiang, H.K. Kole, D.G. Cronshaw, Y. Zou, H. Gu, Grb2 functions at the top of the T-cell antigen receptor-induced tyrosine kinase cascade to control thymic selection, Proc. Natl. Acad. Sci. USA 107 (2010) 10620–10625.
[42] T. Santos, G.F. Salgado, E.J. Cabrita, C. Cruz, G-Quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions, Pharmaceuticals 14 (2021).
[43] E. Cadoni, L. De Paepe, A. Manicardi, A. Madder, Beyond small molecules: targeting G-quadruplex structures with oligonucleotides and their analogues, Nucleic Acids Res. 49 (2021) 6638–6659.
[44] Y. Luo, A. Granzhan, J. Marquevielle, A. Cucchiarini, L. Lacroix, S. Amrane, D. Verga, J.L. Mergny, Guidelines for G-quadruplexes: I. In vitro characterization, Biochimie 214 (2023) 5–23.
[45] C. Lin, J. Dickerhoff, D. Yang, NMR studies of G-quadruplex structures and G-quadruplex-interactive compounds, Methods Mol. Biol. 2035 (2019) 157–176.
[46] A. Bedrat, L. Lacroix, J.L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759.
[47] M. Adrian, B. Heddi, A.T. Phan, NMR spectroscopy of G-quadruplexes, Methods 57 (2012) 11–24.
[48] OligoAnalyzer Tool - Primer analysis and Tm Calculator | IDT, (n.d.). https://eu.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fanalyzer (accessed October 1, 2023).
[49] TopSpin | NMR Data Analysis | Bruker, (n.d.). https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html (accessed October 1, 2023).
[50] W. Lee, M. Tonelli, J.L. Markley, NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy, Bioinformatics 31 (2015) 1325–1327.
[51] N.Q. Do, A.T. Phan, Monomer–dimer equilibrium for the 5′–5′ stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study, Chem. Eur. J. 18 (2012) 14752–14759.
[52] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, Loop-length-dependent folding of G-quadruplexes, J. Am. Chem. Soc. 126 (2004) 16405–16415.
[53] R. Tippana, W. Xiao, S. Myong, G-quadruplex conformation and dynamics are determined by loop length and sequence, Nucleic Acids Res. 42 (2014) 8106–8114.
[54] K.W. Lim, L. Lacroix, D.J.E. Yue, J.K.C. Lim, J.M.W. Lim, A.T. Phan, Coexistence of two distinct G-quadruplex conformations in the hTERT promoter, J. Am. Chem. Soc. 132 (2010) 12331–12342.
[55] G. Devi, F.R. Winnerdy, J.C.Y. Ang, K.W. Lim, A.T. Phan, Four-layered intramolecular parallel G-quadruplex with non-nucleotide loops: an ultra-stable self-folded DNA nano-scaffold, ACS Nano 16 (2022) 533–540.
[56] L. Joly, F. Rosu, V. Gabelica, d(TGnT) DNA sequences do not necessarily form tetramolecular G-quadruplexes, Chem. Commun. 48 (2012) 8386–8388.
[57] F.W. Smith, P. Schultze, J. Feigon, Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats, Structure 3 (1995) 997–1008.
[58] J. Brcic, J. Plavec, NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism, Nucleic Acids Res. 46 (2018) 11605–11617.
[59] J. Brčić, J. Plavec, Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution, Nucleic Acids Res. 43 (2015) 8590–8600.
[60] A. Bugaut, S. Balasubramanian, A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes, Biochemistry 47 (2008) 689–697.
[61] J. Li, I. Te Chu, T.A. Yeh, D.Y. Chen, C.L. Wang, T.C. Chang, Effects of length and loop composition on structural diversity and similarity of (G3TG3NmG3TG3) G-quadruplexes, Molecules 25 (2020) 1779.
[62] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E.S. Baker, M.T. Bowers, E. De Pauw, V. Gabelica, G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation, J. Am. Chem. Soc. 130 (2008) 10208–10216.
[63] M. Adrian, D.J. Ang, C.J. Lech, B. Heddi, A. Nicolas, A.T. Phan, Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite, J. Am. Chem. Soc. 136 (2014) 6297–6305.
[64] R.W. Harkness, A.K. Mittermaier, G-quadruplex dynamics, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1865 (2017) 1544–1554.
[65] P.A. Rachwal, T. Brown, K.R. Fox, Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes, Biochemistry 46 (2007) 3036–3044.
[66] C. Hennecker, L. Yamout, C. Zhang, C. Zhao, D. Hiraki, N. Moitessier, A. Mittermaier, Structural polymorphism of guanine quadruplex-containing regions in human promoters, Int. J. Mol. Sci. 23 (2022) 16020.
[67] A. Awadasseid, X. Ma, Y. Wu, W. Zhang, G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy, Biomed. Pharmacother. 139 (2021) 111550.
[68] R. Hänsel-Hertsch, M. Di Antonio, S. Balasubramanian, DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential, Nat. Rev. Mol. Cell Biol. 18 (2017) 279–284.
[69] Z.Y. Sun, X.N. Wang, S.Q. Cheng, X.X. Su, T.M. Ou, Developing novel G-quadruplex ligands: from interaction with nucleic acids to interfering with nucleic acid–protein interaction, Molecules 24 (2019) 396.
[70] T.Q.N. Nguyen, K.W. Lim, A.T. Phan, A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs, Sci. Rep. 7 (2017) 11969.
[71] S. Obata, S. Asamitsu, K. Hashiya, T. Bando, H. Sugiyama, G-Quadruplex induction by the hairpin pyrrole-imidazole polyamide dimer, Biochemistry 57 (2018) 498–502.
[72] D.J.Y. Tan, P. Das, F.R. Winnerdy, K.W. Lim, A.T. Phan, Guanine anchoring: a strategy for specific targeting of a G-quadruplex using short PNA, LNA and DNA molecules, Chem. Commun. 56 (2020) 5897–5900.