



OPEN ACCESS

EDITED BY

Arjun Singh,
Memorial Sloan Kettering Cancer Center,
United States

REVIEWED BY

Soumitra Basu,
Kent State University, United States
Domenica Musumeci,
University of Naples Federico II, Italy
Dmitry N. Kaluzhny,
Engelhardt Institute of Molecular Biology
(RAS), Russia

*CORRESPONDENCE

Robert C. Mosen,
✉ rcmons01@louisville.edu

RECEIVED 24 April 2023

ACCEPTED 30 May 2023

PUBLISHED 07 June 2023

CITATION

Mosen RC (2023), Higher-order G-
quadruplexes in promoters are untapped
drug targets.

Front. Chem. 11:1211512.

doi: 10.3389/fchem.2023.1211512

COPYRIGHT

© 2023 Mosen. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Higher-order G-quadruplexes in promoters are untapped drug targets

Robert C. Mosen*

Department of Medical Oncology and Hematology, UofL Health Brown Cancer Center, University of Louisville, Louisville, KY, United States

G-quadruplexes (G4s) are four-stranded nucleic acid secondary structures that form within guanine-rich regions of chromatin. G4 motifs are abundant in the genome, with a sizable proportion (~40%) existing within gene promoter regions. G4s are proven epigenetic features that decorate the promoter landscape as binding centers for transcription factors. Stabilizing or disrupting promoter G4s can directly influence adjacent gene transcription, making G4s attractive as indirect drug targets for hard-to-target proteins, particularly in cancer. However, no G4 ligands have progressed through clinical trials, mostly owing to off-targeting effects. A major hurdle in G4 drug discovery is the lack of distinctiveness of the small monomeric G4 structures currently used as receptors. This mini review describes and contrasts monomeric and higher-order G-quadruplex structure and function and provides a rationale for switching focus to the higher-order forms as selective molecular targets. The human telomerase reverse transcriptase (hTERT) core promoter G-quadruplex is then used as a case study that highlights the potential for higher-order G4s as selective indirect inhibitors of hard-to-target proteins in cancer.

KEYWORDS

G-quadruplex (G4), higher-order G4, DNA, promoter, drug target

1 G-quadruplex structures in promoters: monomeric vs. higher-order forms

1.1 Intramolecular monomeric promoter G-quadruplexes

G-quadruplexes (G4s) are four-stranded secondary structures created from the stacking of two or more guanine tetrads (“G-tetrads”) (Spiegel et al., 2020). Each G-tetrad is composed of four guanine bases arranged in a square planar configuration, that is, stabilized by Hoogsteen hydrogen bonding (Figure 1A). Monovalent cations are coordinated within the G-tetrad column central channel by the inward facing carbonyl groups, providing stabilization from coordinate bonding and neutralization of the partially negative charges (Lane et al., 2008). A commonly used sequence motif to describe a monomeric G-quadruplex is $G_{3-4}L_{1-7}G_{3-4}L_{1-7}G_{3-4}L_{1-7}G_{3-4}$, where G indicates a guanine tract and L designates any nucleotide in the intervening loop. Historically, the largest loop length has been taken to be seven nucleotides because of the destabilizing effect of large loops *in vitro* (Spiegel et al., 2020; Ravichandran et al., 2021).

G-rich sequences in specific gene promoter regions can adopt G4 structures with a surprisingly diverse array of topological conformations for the intramolecular monomeric G4 *in vitro*, depending on the sequence composition, loop lengths, ionic environment, and hydration (Chen and Yang, 2012; Miller et al., 2010; Bhattacharyya et al., 2016). The most

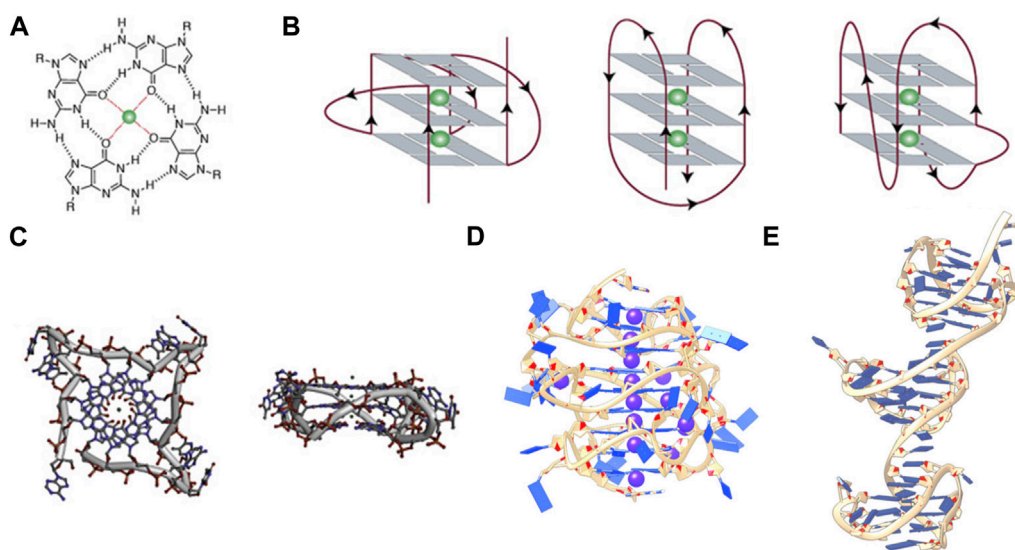


FIGURE 1

(A) Structure of a G-tetrad, (B) From left to right, schematic representations of (A) parallel, antiparallel, and mixed hybrid with bulge, (C) Top and side views of a parallel monomeric G-quadruplex, (D) side view of a stacked intramolecular all-parallel higher-order G-quadruplex, and (E) side view of an intramolecular higher-order all-hybrid G-quadruplex in a beads-on-a-string arrangement. (C–E) are not to scale. Figures 1A–C, are adapted from (Spiegel et al., 2020), (D) is adapted from (Monsen et al., 2020), and (E) is adapted from (Monsen et al., 2021).

frequent topologies are: antiparallel, in which two out of four G-tracts run in the same direction; hybrid, in which three G-tracts run in the same direction; and parallel, in which all G-tracts run in the same direction (Figure 1B) (Spiegel et al., 2020). This review is primarily concerned with promoter G4s that are necessarily intramolecular and flanked on either end by single- or double-stranded regions. In the context of a promoter bubble (Shirude et al., 2007; Monsen et al., 2023), as well as when promoter G4 sequences are flanked by single-stranded regions (this can be realized *in vitro*, for example, by adding two residues on either end of the monomer G4 motif: $L_2G_xL_nG_xL_nG_xL_2$), the parallel topology is preferred (Chen et al., 2021). The biological preference for parallel promoter G4 conformations is supported by the extensive immunofluorescence seen across chromosomes stained with the parallel-favoring anti-G4 antibody “BG4” (Biffi et al., 2013; Javadekar et al., 2020). The first ever near-atomic structural study of a G4 in a near-native duplex bubble has shown that duplex end-stacking at the G-tetrad interface may be a key factor to promoter G4 stability, suggesting that antiparallel and hybrid topologies could be sterically prohibited (Monsen et al., 2023). Overall, *in vitro* monomeric promoter G4 sequences adopt a variety of topological configurations when isolated, but when placed in their biological context prefer the parallel form.

1.2 Intramolecular higher-order promoter G-quadruplexes

Intramolecular higher-order promoter G4s (xG4s) consist of two or more connected monomeric G4 domains. The xG4 motif is degenerate compared to the lower order G4 motif and is of the general form $(G_{2-4}L_{1-12}G_{2-4}L_{1-12}G_{2-4}L_{1-12}G_{2-4}L_{\geq 0})_n$ (Monsen et al.,

2022a), where the length of the 3′ connecting loop can be zero and $n \geq 2$. The xG4 motif becomes substantially more degenerate when accounting for mismatches and/or bulges (i.e., non-guanine nucleotides within a putative G-tract) (Berselli et al., 2020). The xG4 motif is notable for multiple reasons. First, the zero-nucleotide connecting loop presents the possible situation where the last guanine of the 3′ G-tetrad of the first G4 domain and the first guanine of the 5′ G-tetrad of the second G4 domain create a continuous G-tetrad column that spans the entire macromolecule. If both G4s are parallel, this would facilitate formation of a single continuous medium groove spanning the two domains with a favorable locked 3′ to 5′ G-tetrad stacked interface (Kogut et al., 2019). A second, less obvious feature of the xG4 motif is that it accounts for two-tetrad G4s, in which there are only two guanines in the G-tracts. Two-tetrad monomeric G4s cannot exist as stable parallel conformations *in vitro* when isolated (Kejnovska et al., 2021), but have been seen in higher-order G4 assemblies (Monsen et al., 2022a) [we note that stable antiparallel two-tetrad G4s are seen *in vitro* (Lim et al., 2009)]. This is an important feature, as it reveals that biologically relevant xG4s have an expanded structural repertoire compared to the monomeric sequences. Lastly, xG4s seem to tolerate much longer loops (1–12+ nucleotides), both within and between the G4 domains (Berselli et al., 2020; Monsen et al., 2020; Monsen et al., 2022a). This may be the result of tertiary interactions that stabilize either across the G4–G4 domain by loop interactions (Rigo and Sissi, 2017), symmetry in the loop giving rise to a stabilizing hairpin moiety (Berselli et al., 2020), or arise from the stability imparted on the G4 cores from head-to-tail stacking (Kogut et al., 2019).

Structural characterizations of intramolecular promoter xG4s to date have revealed an overwhelming preference for stacked arrangements (Micheli et al., 2010; Rigo and Sissi, 2017; Monsen

et al., 2020; Monsen et al., 2022a), with the highest resolution models showing a preference for an all-parallel stacked globular configuration (Figure 1D) (Monsen et al., 2020; Monsen et al., 2022a). Importantly, all reported putative xG4 sequences that have been investigated form stable G4 structures *in vitro*. In some cases, the presence of long loops (>7 nucleotides) result in hairpins (Monsen et al., 2022a) that contribute to the stability of the individual domains (Ravichandran et al., 2021) while also creating structurally unique interfaces. In contrast to promoter xG4s, other higher-order intramolecular G4s reported, such as the human telomere (Monsen et al., 2021), the insulin-linked polymorphic region (ILPR) minisatellite (Schonhoft et al., 2009), and the CEB25 minisatellite locus (Amrane et al., 2012), show mixed antiparallel and hybrid topologies that are better described as dumbbell shaped or beads-on-a-string configurations. Figure 1 contrasts the structures of a monomer G4 in the parallel conformation, a higher-order all-parallel stacked G4, and a higher-order all-hybrid beads-on-a-string.

The selective binding interface of the xG4s is not imparted by the G-tetrad columns, but rather the specific G4-G4 interaction interfaces, loop sequence, size, and configurations. Figure 1C reveals that there is little targetable real estate associated with the monomeric parallel G4 (Monsen et al., 2023). Conversely, Figures 1D, E reveals multiple putative binding sites that would be large enough and distinct enough for specific protein or drug interaction. Biochemical support for this idea comes from a recent pull-down study conducted by Ceschi et al. (Ceschi et al., 2022). In this study, the authors used a variety of higher-order G4 sequences (such as those in Figures 1D, E) to enrich for tightly interacting proteins from cell lysates that are specific to higher-order G4s over the lower order forms. Surprisingly, the intermediate filament Vimentin was shown to have a selective nanomolar affinity to higher-order G4 structures with no apparent binding to lower order species. While the mechanism of recognition is still unclear, this study supports the hypothesis that xG4s offer unique recognition sites that could be useful in selective targeting.

2 G-quadruplex distribution and function in promoters

2.1 Biological distribution and function of promoter G4s

Promoter G-quadruplexes are prevalent epigenetic regulatory elements. Current estimates show more than 700,000 monomeric G4 motifs across the human genome (Hansel-Hertsch et al., 2017). Huppert and Balasubramanian, using the canonical sequence motif $G_{3+}L_{1-7}G_{3+}L_{1-7}G_{3+}L_{1-7}G_{3+}$, showed that more than 40% of gene promoters had at least one monomeric G4 motif (Huppert and Balasubramanian, 2007). More recently, Hänsel-Hertsch and colleagues have used a chromatin immunoprecipitation (ChIP)-sequencing direct capture approach to show that thousands of G4 structures, not just motifs, are enriched in highly transcribed gene promoters, specifically many involved in cancer (Hansel-Hertsch et al., 2016). Promoter quadruplexes overall appear to be acting as general transcription factor (TF) “binding hubs”,

coinciding with regions of open chromatin and high transcriptional activity (Spiegel et al., 2021).

At a more granular level, promoter G4s act in concert with transcriptional proteins to affect gene transcription in multiple ways. Initially, promoter G4s were thought to only act as simple physical barriers to polymerases, acting as “on/off” switches of transcription (Sarkies et al., 2010). However, studies have now shown that G4s can directly recruit transcription factors with some level of specificity. For instance, two zinc fingers, SP1 and MAZ (Myc-associated zinc finger), show G4 structure-dependent recognition. In the former case, Raiber and colleagues, using pull-down experiments with the transcription factor SP1, showed that 36% of the sequences lacked consensus SP1 binding motifs (Raiber et al., 2012). They went on to show that 77% of those sequences lacking the SP1 consensus motif were putative G-quadruplexes and that, overall, SP1 binding had 87% overlap with G4 sequence motifs. In the latter case, Cogoi and colleagues have shown that MAZ recognizes a G4 formed within the *kRas* promoter and showed that stabilizing the *kRas* G4 with a small molecule could promote MAZ binding and increase transcription, while mutations that destabilized the G4 reduced MAZ binding and transcription (Cogoi et al., 2010).

G4s can also serve as transcriptional repressors. The classical case for this is the G4 formed within the *c-Myc* promoter. In their seminal study, Siddiqui-Jain et al. showed that a G-rich region in the nuclease hypersensitivity element III (NHEIII) of the *c-Myc* P1 promoter forms one or more monomeric G4s (Siddiqui-Jain et al., 2002). Through mutational destabilization, they showed that there is a 3-fold increase in *c-Myc* expression, indicating that the G4 acts as a transcriptional repressor. They showed that stabilization of the G4 with a small molecule could further reduce transcription to below the basal level. Later studies on the protein nucleolin, a multifaceted and abundant protein found in the nucleolus (Tajrishahi et al., 2011), have shown that it can help the folding of promoter G4s like a molecular chaperone (Tosoni et al., 2015). Nucleolin was shown to fold the *c-Myc* NHEIII G4, promoting transcriptional downregulation in cells (Gonzalez and Hurley, 2010). Altogether, these studies show that promoter G4s are important epigenetic regulators of genes, and that their stability and interaction with transcription factors can influence transcription.

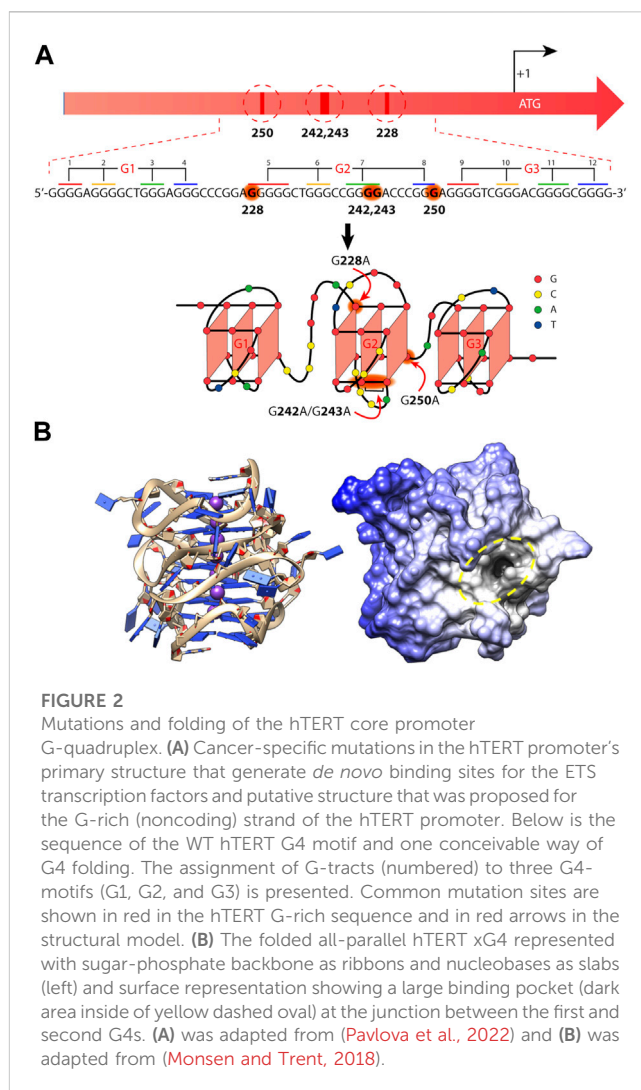
2.2 Biological distribution and function of promoter xG4s

Promoter xG4s are also abundant across the genome. Berselli and colleagues recently developed QPARSE (13), the first algorithm capable of finding xG4s in the genome that accounts for mismatches and bulges in G-tracts (see Figure 1B for a G4 with a bulge). In the study, the authors used their degenerate G4 motif algorithm to find monomeric, dimeric, and trimeric G4 sequences in the range of -200 to +600 of the transcriptional start sites (TSS) across the annotated human genome [GENCODE (Harrow et al., 2012)]. They found that 49%, 15%, and 4% of TSS regions had monomer, dimer, and trimer G4 repeats, respectively, and show that this enrichment cannot entirely be attributed to high GC content. Ceschi et al. have recently used the same algorithm to search just the first 100 bp upstream of the TSS of gene promoters in GENCODE, finding

1,478 dimer and trimer promoter xG4s (Ceschi et al., 2022) (~4% of the 38,404 annotated genes in GENCODE v34).

Parsing out the *in vivo* function of a particular xG4 compared to their monomeric counterparts is a challenging task. At one end you have bioinformatic and G4- or G4-ChIP-sequencing approaches that lack the spatial resolution to distinguish between monomer, dimer, trimer, and other higher-order G4 effects (Park, 2009; Mahony and Pugh, 2015). For instance, the ChIP-sequencing studies mentioned above should, in theory, encompass promoter sites enriched with xG4s. At the other end, there is a dearth of tools that allow direct probing of the effects of monomer versus higher-order G4s. Specifically, there are no reports to date that have convincingly shown selective (de)stabilization of a single G4 domain among a higher-order promoter G4 assembly (Frasson et al., 2022). Mutational reporter assays are the current best approach to parsing out the functionality of xG4s within a cellular context (Siddiqui-Jain et al., 2002; Cogoi et al., 2010), but these come at the risk of altering protein recognition motifs (Bell et al., 2015).

The most thorough investigation of a promoter xG4 to date was conducted in 2019 by Ducani and colleagues on an xG4 found in the promoter of the proto-oncogene *c-Kit* (Ducani et al., 2019). The *c-Kit* gene encodes for a transmembrane tyrosine kinase receptor (*c-Kit* or CD117) that, after activation by stem cell factor (SCF), transduces signals that promote cell proliferation, differentiation, and migration (Liang et al., 2013). Excessive signaling by continued stimulation or mutation has been implicated in a variety of cancers, such as gastrointestinal stromal tumors (GISTs), pancreatic cancer, melanoma, and hematological neoplastic diseases (Gregory-Bryson et al., 2010; Abbaspour Babaei et al., 2016). Since *c-Kit* has a tyrosine kinase domain, these cancers are typically treated with tyrosine kinase inhibitors (TKIs) tailored to the mutation type (Abbaspour Babaei et al., 2016). Unfortunately, treating *c-Kit*-driven cancers with TKIs is often followed with a rapid switch to drug-resistance through mutation (Demetri et al., 2002; Loughrey et al., 2006; Abbaspour Babaei et al., 2016). Therefore, understanding how the *c-Kit* promoter xG4 regulates its transcription is important for drug development efforts. The *c-Kit* xG4 consists of three monomeric G4 motifs, designated K2, SP, and K1 (from 5' to 3') connected by a single dA loop and a hexanucleotide dGCGCAG loop, respectively. Each separate G4 domain has been structurally and/or functionally examined (Rankin et al., 2005; Phan et al., 2007; Hsu et al., 2009; Kuryavii et al., 2010; Raiber et al., 2012; Kotar et al., 2019), and the higher-order structural assembly of domains K2-SP confirmed by integrative structural biology approaches (Rigo and Sissi, 2017; Monsen et al., 2022a). Both structural studies show that the K2-SP regions interact through stacking, although the two studies differ slightly in the size of the sequence examined. Based on circular dichroism of the full-length sequence used by Ducani, however, the full-length K2-SP-K1 sequence adopts an all-parallel conformation, which is consistent with the promoter xG4s structurally verified to date (Monsen et al., 2020; Monsen et al., 2022a). To investigate the biological function of such an arrangement, Ducani and colleagues conducted luciferase assays to test the effect of disrupting all combinations of the G4 regions by mutating guanines essential for structural integrity. They show in the leukemia cell line HEL92.1.7 that the K1 G4 (alone) and K2-SP G4s (together as a higher-order feature) have opposing roles in transcriptional regulation, with the former acting repressively and the latter acting to stimulate transcription. Further, they show that each G4 unit is significantly affected by the formation or disruption of the others, signifying



G4-G4 crosstalk within a cellular context. The readout from each mutational state was more of a continuous distribution, rather than an all or nothing response. Overall, this study suggests that xG4s play a very nuanced role in governing transcriptional activation or repression. It remains to be seen whether a K2-SP disrupting small molecule, or SP-K1 stabilizing small molecule will act as an indirect *c-Kit* inhibitor.

3 Case study: targeting the hTERT core promoter xG4

One of the most studied xG4s to date is found within the *hTERT* core promoter region. *hTERT* encodes the protein catalytic subunit of telomerase, the ribonucleoprotein primarily responsible for maintaining telomere length homeostasis (Bryan and Cech, 1999). Although typically undetectable in somatic cells, *hTERT* is aberrantly over-expressed in more than 90% of aggressive cancers (Shay and Bacchetti, 1997), making it a long sought after cancer-specific target. The *hTERT* gene was first identified as harboring a putative xG4 in its promoter by Palumbo et al. (2009). The hTERT xG4 sequence consists of twelve G-tracts of three or more guanines that enable the maximum formation of three contiguous G4s (Figure 2).

Since its discovery, the tertiary structure of the *hTERT* xG4 has been intensely disputed. In the original work, Palumbo et al. showed by CD that the *hTERT* sequence was entirely parallel (Palumbo et al., 2009), although, based on their DMS footprinting results, they proposed a model consisting of a parallel G4 unit connected to an antiparallel G4 unit connected by a ~26 nucleotide hairpin loop. About a year after, Micheli et al. (2010) independently proposed that the *hTERT* xG4 formed three parallel G4 units that were contiguous and stacked 3' to 5' based on CD and Taq polymerase stop assays. Later studies using an experimental small molecule and DMS footprinting assays reported yet another, slightly different, G4-hairpin arrangement (Song et al., 2019). However, structural studies of the sequence using a combination of CD, hydrodynamics, NMR, modeling, and small-angle X-ray scattering (SAXS) have since confirmed that the most consistent model is one in which three parallel G4s are tightly stacked in a stacked 3' to 5' arrangement (Figure 2) (Monsen et al., 2020). A recent DMS footprinting experiment confirms the all-parallel model as correct (Pavlova et al., 2022).

The *hTERT* xG4 is a potentially selective, transcriptionally repressive structure. The *hTERT* core promoter region has multiple non-coding mutations that are seen across many cancer types, with the two most prominent denoted “G228A” and “G250A”, that coincide with a robust increased telomerase activity (Killela et al., 2013). These mutations exist within G-tracts that form the central G4 unit of the three stacked all-parallel xG4 model (see Figure 2) (Micheli et al., 2010; Chaires et al., 2014; Monsen et al., 2020). Using luciferase expression assays of the wild type (WT) promoter versus mutant promoters of either G228A or G250A, Bell and colleagues showed that a robust increase in promoter activity is gained with either mutant (Bell et al., 2015). Further investigation showed that either G- > A mutation creates a *de novo* ETS (Erythroblast Transformation Specific) transcription factor consensus motif that is recognized by the transcription factor GABP (GA-binding protein). However, the creation of an ETS motif has been shown to be insufficient to fully explain the transcriptional changes seen in mutant cells (Kang et al., 2016).

Early investigations by Micheli et al. showed that the central G4, which encompasses either G- > A mutation, is unstable as an isolated monomeric G4, and only forms stably as a higher-order G4 assembly through stacking with both 5' and 3' G4 regions (Micheli et al., 2010). The destabilizing effects of the mutants on the overall xG4 have since been confirmed (Pavlova et al., 2022). A G4 ligand reported to refold and stabilize the *hTERT* xG4 WT structure (with or without G- > A mutations) was able to restore its repressive effects and showed good selectivity for the *hTERT* promoter over other genes with known promoter G4s (Kang et al., 2016). More recently, Monsen et al., (2022b) used a virtual screening approach to target the loop and G4-junctional regions of the all-parallel *hTERT* xG4 (Figure 2). Using a variety of *in vitro* binding and competition assays, the authors were able to find a drug-like small molecule that stabilized across the second and third G4 regions that showed high selectivity over duplex DNA and moderate selectivity over all other forms of DNA tested. In both cases, repression of *hTERT* expression was confirmed in breast cancer cell lines. Collectively, these studies reveal that the *hTERT* xG4 is an indirect target for down-regulating *hTERT* in cells. Further, these studies show for the first time that a higher-order promoter G4 can be targeted with selectivity using the unique features imparted from G4-G4 domains.

4 Discussion

Higher-order G-quadruplexes have emerged as selective targets in the promoters of thousands of annotated genes across the human genome. While xG4s offer some of the same protein binding recognition that monomeric G4s do, promoter xG4s also appear to encode for transcriptional status, possibly acting as titratable “dimmer switches” of gene activity. Structurally, xG4s offer a much richer drug targeting landscape consisting of G4-G4 interaction junctions adorned with protein-like binding pockets formed among sequence-specific loop features. This is emphasized by the decade-long pursuit targeting the *hTERT* core promoter xG4 which has resulted in both a unique receptor and the first drug-like small molecule targeting an xG4 with selectivity.

While xG4 studies to date have led to exciting new insight into their structural arrangements and potential regulatory mechanisms, there is still much to be understood. One of the major hurdles in studying xG4s is their recalcitrance to the traditional structural biology techniques NMR and X-ray diffraction. To date, all structural models have been derived from medium-to low-resolution integrative structural strategies and/or footprinting methods (Palumbo et al., 2009; Micheli et al., 2010; Rigo and Sissi, 2017; Monsen et al., 2020; Monsen et al., 2021; Monsen et al., 2022a; Pavlova et al., 2022). Cryo-EM is emerging as a possible solution to this problem (Monsen et al., 2023); however, atomic resolution of relatively small, potentially heterogeneous [e.g., G-tract isomers (Harkness and Mittermaier, 2017; Hennecker et al., 2022)] or flexible systems (e.g., inter-domain movements or long flexible loops) (Monsen et al., 2021; Monsen et al., 2022a; Monsen et al., 2023) is still a considerable challenge in the field (Herzik et al., 2019) [although scaffolds might offer a solution (Wu and Rapoport, 2021)]. A second major hurdle, as touched on above, is determining their biological mechanism. Molecular tools, such as xG4-specific fluorescent molecules (Summers et al., 2021) or antibodies (Biffi et al., 2013), should aid in revealing their spatial and temporal formation. Recently, the zinc finger-containing transcription factor Yin Yang-1 (YY1) was shown to bind G4s and bring two G4 domains into proximity through its dimerization (Li et al., 2021), offering a potential tool for investigating biological function of xG4s in a more biologically relevant context.

Author contributions

RM conceptualized the topic, researched, and analyzed the literature, and wrote the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

The author thanks John O. Trent for his suggested use of “xG4” for higher-order promoter G4s, J. Brad Chaires for providing the

inspiration for the topic and proofreading, and Robert D. Gray for proofreading the review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abbaspour Babaei, M., Kamalidehghan, B., Saleem, M., Huri, H. Z., and Ahmadipour, F. (2016). Receptor tyrosine kinase (c-Kit) inhibitors: A potential therapeutic target in cancer cells. *Drug Des. Devel Ther.* 10, 2443–2459. doi:10.2147/dddt.s89114
- Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J. L., et al. (2012). Formation of pearl-necklace monomeric G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.* 134 (13), 5807–5816. doi:10.1021/ja208993r
- Bell, R. J., Rube, H. T., Kreig, A., Mancini, A., Fouse, S. D., Nagarajan, R. P., et al. (2015). The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* 348 (6238), 1036–1039. doi:10.1126/science.aab0015
- Berselli, M., Lavezzo, E., and Toppo, S. (2020). Qparse: Searching for long-looped or multimeric G-quadruplexes potentially distinctive and druggable. *Bioinformatics* 36 (2), 393–399. doi:10.1093/bioinformatics/btz569
- Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Front. Chem.* 4, 38. doi:10.3389/fchem.2016.00038
- Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* 5 (3), 182–186. doi:10.1038/nchem.1548
- Bryan, T. M., and Cech, T. R. (1999). Telomerase and the maintenance of chromosome ends. *Curr. Opin. Cell. Biol.* 11 (3), 318–324. doi:10.1016/s0955-0674(99)80043-x
- Ceschi, S., Berselli, M., Cozzaglio, M., Giantin, M., Toppo, S., Spolaore, B., et al. (2022). Vimentin binds to G-quadruplex repeats found at telomeres and gene promoters. *Nucleic Acids Res.* 50 (3), 1370–1381. doi:10.1093/nar/gkab1274
- Chaires, J. B., Trent, J. O., Gray, R. D., Dean, W. L., Buscaglia, R., Thomas, S. D., et al. (2014). An improved model for the hTERT promoter quadruplex. *PLoS One* 9 (12), e115580. doi:10.1371/journal.pone.0115580
- Chen, J., Cheng, M., Salgado, G. F., Stadlbauer, P., Zhang, X., Amrane, S., et al. (2021). The beginning and the end: Flanking nucleotides induce a parallel G-quadruplex topology. *Nucleic Acids Res.* 49 (16), 9548–9559. doi:10.1093/nar/gkab681
- Chen, Y., and Yang, D. (2012). Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Protoc. Nucleic Acid. Chem.* 50, ncl705s50. doi:10.1002/0471142700.ncl705s50
- Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K., and Xodo, L. E. (2010). The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.* 285 (29), 22003–22016. doi:10.1074/jbc.m110.101923
- Demetri, G. D., von Mehren, M., Blanke, C. D., Van den Abbeele, A. D., Eisenberg, B., Roberts, P. J., et al. (2002). Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N. Engl. J. Med.* 347 (7), 472–480. doi:10.1056/nejmoa020461
- Ducani, C., Bernardinelli, G., Hogberg, B., Keppler, B. K., and Terenzi, A. (2019). Interplay of three G-quadruplex units in the KIT promoter. *J. Am. Chem. Soc.* 141 (26), 10205–10213. doi:10.1021/jacs.8b12753
- Frasson, I., Pirota, V., Richter, S. N., and Doria G-quadruplexes, F. (2022). Multimeric G-quadruplexes: A review on their biological roles and targeting. *Int. J. Biol. Macromol.* 204, 89–102. doi:10.1016/j.ijbiomac.2022.01.197
- Gonzalez, V., and Hurlley, L. H. (2010). The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry* 49 (45), 9706–9714. doi:10.1021/bi100509s
- Gregory-Bryson, E., Bartlett, E., Kiupel, M., Hayes, S., and Yuzbasiyan-Gurkan, V. (2010). Canine and human gastrointestinal stromal tumors display similar mutations in c-KIT exon 11. *BMC Cancer* 10, 559. doi:10.1186/1471-2407-10-559
- Hansel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., et al. (2016). G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48 (10), 1267–1272. doi:10.1038/ng.3662
- Hansel-Hertsch, R., Di Antonio, M., and Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell. Biol.* 18 (5), 279–284. doi:10.1038/nrm.2017.3
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). Gencode: The reference human genome annotation for the ENCODE project. *Genome Res.* 22 (9), 1760–1774. doi:10.1101/gr.135350.111
- Hennecker, C., Yamout, L., Zhang, C., Zhao, C., Hiraki, D., Moitessier, N., et al. (2022). Structural polymorphism of guanine quadruplex-containing regions in human promoters. *Int. J. Mol. Sci.* 23 (24), 16020. doi:10.3390/ijms232416020
- Herzik, M. A., Jr., Wu, M., and Lander, G. C. (2019). High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* 10 (1), 1032. doi:10.1038/s41467-019-08991-8
- Hsu, S. T., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., and Balasubramanian, S. (2009). A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* 131 (37), 13399–13409. doi:10.1021/ja904007p
- Huppert, J. L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35 (2), 406–413. doi:10.1093/nar/gkl1057
- Javadekar, S. M., Nilavar, N. M., Paranjape, A., Das, K., and Raghavan, S. C. (2020). Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms. *DNA Res.* 27 (5), dsaa024. doi:10.1093/dnares/dsaa024
- Kang, H. J., Cui, Y., Yin, H., Scheid, A., Hendricks, W. P. D., Schmidt, J., et al. (2016). A pharmacological chaperone molecule induces cancer cell death by restoring tertiary DNA structures in mutant hTERT promoters. *J. Am. Chem. Soc.* 138 (41), 13673–13692. doi:10.1021/jacs.6b07598
- Kejnovska, I., Stadlbauer, P., Trantirek, L., Renciuik, D., Gajarsky, M., Krafcik, D., et al. (2021). G-quadruplex formation by DNA sequences deficient in guanines: Two tetrad parallel quadruplexes do not fold intramolecularly. *Chemistry* 27 (47), 12115–12125. doi:10.1002/chem.202100895
- Killela, P. J., Reitman, Z. J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L. A., Jr., et al. (2013). TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* 110 (15), 6021–6026. doi:10.1073/pnas.1303607110
- Kogut, M., Kleist, C., and Czub, J. (2019). Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization in atomic detail. *PLoS Comput. Biol.* 15 (9), e1007383. doi:10.1371/journal.pcbi.1007383
- Kotar, A., Rigo, R., Sissi, C., and Plavec, J. (2019). Two-quartet kit^G G-quadruplex is formed via double-stranded pre-folded structure. *Nucleic Acids Res.* 47 (5), 2641–2653. doi:10.1093/nar/gky1269
- 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 (19), 6757–6773. doi:10.1093/nar/gkq558
- Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* 36 (17), 5482–5515. doi:10.1093/nar/gkn517
- Li, L., Williams, P., Ren, W., Wang, M. Y., Gao, Z., Miao, W., et al. (2021). YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol.* 17 (2), 161–168. doi:10.1038/s41589-020-00695-1
- Liang, J., Wu, Y. L., Chen, B. J., Zhang, W., Tanaka, Y., and Sugiyama, H. (2013). The C-kit receptor-mediated signal transduction and tumor-related diseases. *Int. J. Biol. Sci.* 9 (5), 435–443. doi:10.1016/j.ijbs.6087
- Lim, K. W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D. J., et al. (2009). Structure of the human telomere in K⁺ solution: A stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.* 131 (12), 4301–4309. doi:10.1021/ja807503g
- Loughrey, M. B., Waring, P. M., Dobrovic, A., Demetri, G., Kovalenko, S., and McArthur, G. (2006). Polyclonal resistance in gastrointestinal stromal tumor treated with sequential kinase inhibitors. *Clin. Cancer Res.* 12 (20), 6205–6206. doi:10.1158/1078-0432.ccr-06-1079
- Mahony, S., and Pugh, B. F. (2015). Protein-DNA binding in high-resolution. *Crit. Rev. Biochem. Mol. Biol.* 50 (4), 269–283. doi:10.3109/10409238.2015.1051505

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Micheli, E., Martufi, M., Cacchione, S., De Santis, P., and Savino, M. (2010). Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives. *Biophys. Chem.* 153 (1), 43–53. doi:10.1016/j.bpc.2010.10.003
- Miller, M. C., Buscaglia, R., Chaires, J. B., Lane, A. N., and Trent, J. O. (2010). Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.* 132 (48), 17105–17107. doi:10.1021/ja105259m
- Monsen, R. C., Chakravarthy, S., Dean, W. L., Chaires, J. B., and Trent, J. O. (2021). The solution structures of higher-order human telomere G-quadruplex multimers. *Nucleic Acids Res.* 49 (3), 1749–1768. doi:10.1093/nar/gkaa1285
- Monsen, R. C., Chua, E. Y. D., Hopkins, J. B., Chaires, J. B., and Trent, J. O. (2023). Structure of a 28.5 kDa duplex-embedded G-quadruplex system resolved to 7.4 Å resolution with cryo-EM. *Nucleic Acids Res.* 51 (4), 1943–1959. doi:10.1093/nar/gkad014
- Monsen, R. C., DeLeeuw, L., Dean, W. L., Gray, R. D., Sabo, T. M., Chakravarthy, S., et al. (2020). The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res.* 48 (10), 5720–5734. doi:10.1093/nar/gkaa107
- Monsen, R. C., DeLeeuw, L. W., Dean, W. L., Gray, R. D., Chakravarthy, S., Hopkins, J. B., et al. (2022). Long promoter sequences form higher-order G-quadruplexes: An integrative structural biology study of c-myc, k-ras and c-kit promoter sequences. *Nucleic Acids Res.* 50 (7), 4127–4147. doi:10.1093/nar/gkac182
- Monsen, R. C., Maguire, J. M., DeLeeuw, L. W., Chaires, J. B., and Trent, J. O. (2022). Drug discovery of small molecules targeting the higher-order hTERT promoter G-quadruplex. *PLoS One* 17 (6), e0270165. doi:10.1371/journal.pone.0270165
- Monsen, R. C., and Trent, J. O. (2018). G-Quadruplex virtual drug screening: A review. *Biochimie* 152, 134–148. doi:10.1016/j.biochi.2018.06.024
- Palumbo, S. L., Ebbinghaus, S. W., and Hurley, L. H. (2009). Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* 131 (31), 10878–10891. doi:10.1021/ja902281d
- Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10 (10), 669–680. doi:10.1038/nrg2641
- Pavlova, A. V., Savitskaya, V. Y., Dolinnaya, N. G., Monakhova, M. V., Litvinova, A. V., Kubareva, E. A., et al. (2022). G-quadruplex formed by the promoter region of the hTERT gene: Structure-driven effects on DNA mismatch repair functions. *Biomedicine* 10 (8), 1871. doi:10.3390/biomedicine10081871
- Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007). Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* 129 (14), 4386–4392. doi:10.1021/ja068739h
- Raiber, E. A., Kranaster, R., Lam, E., Nikan, M., and Balasubramanian, S. (2012). A non-canonical DNA structure is a binding motif for the transcription factor SP1 *in vitro*. *Nucleic Acids Res.* 40 (4), 1499–1508. doi:10.1093/nar/gkr882
- Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., et al. (2005). Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* 127 (30), 10584–10589. doi:10.1021/ja050823u
- Ravichandran, S., Razzaq, M., Parveen, N., Ghosh, A., and Kim, K. K. (2021). The effect of hairpin loop on the structure and gene expression activity of the long-loop G-quadruplex. *Nucleic Acids Res.* 49 (18), 10689–10706. doi:10.1093/nar/gkab739
- Rigo, R., and Sissi, C. (2017). Characterization of G4-G4 crosstalk in the c-KIT promoter region. *Biochemistry* 56 (33), 4309–4312. doi:10.1021/acs.biochem.7b00660
- Harkness, R. W., and Mittermaier, A. K. (2017). G-quadruplex dynamics. *Biochim. Biophys. Acta Proteins Proteom* 1865 (11), 1544–1554. doi:10.1016/j.bbapap.2017.06.012
- Sarkies, P., Reams, C., Simpson, L. J., and Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. *Mol. Cell.* 40 (5), 703–713. doi:10.1016/j.molcel.2010.11.009
- Schonhoft, J. D., Bajracharya, R., Dhakal, S., Yu, Z., Mao, H., and Basu, S. (2009). Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. *Nucleic Acids Res.* 37 (10), 3310–3320. doi:10.1093/nar/gkp181
- Shay, J. W., and Bacchetti, S. (1997). A survey of telomerase activity in human cancer. *Eur. J. Cancer* 33 (5), 787–791. doi:10.1016/s0959-8049(97)00062-2
- Shirude, P. S., Okumus, B., Ying, L., Ha, T., and Balasubramanian, S. (2007). Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc.* 129 (24), 7484–7485. doi:10.1021/ja070497d
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* 99 (18), 11593–11598. doi:10.1073/pnas.182256799
- Song, J. H., Kang, H. J., Luevano, L. A., Gokhale, V., Wu, K., Pandey, R., et al. (2019). Small-molecule-targeting hairpin loop of hTERT promoter G-quadruplex induces cancer cell death. *Cell. Chem. Biol.* 26 (8), 1110–1121.e4. doi:10.1016/j.chembiol.2019.04.009
- Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends Chem.* 2 (2), 123–136. doi:10.1016/j.trechm.2019.07.002
- Spiegel, J., Cuesta, S. M., Adhikari, S., Hansel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2021). G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* 22 (1), 117. doi:10.1186/s13059-021-02324-z
- Summers, P. A., Lewis, B. W., Gonzalez-Garcia, J., Porreca, R. M., Lim, A. H. M., Cadinu, P., et al. (2021). Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* 12 (1), 162. doi:10.1038/s41467-020-20414-7
- Tajrish, M. M., Tuteja, R., and Tuteja, N. (2011). Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. *Commun. Integr. Biol.* 4 (3), 267–275. doi:10.4161/cib.4.3.14884
- Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., et al. (2015). Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* 43 (18), 8884–8897. doi:10.1093/nar/gkv897
- Wu, X., and Rapoport, T. A. (2021). Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U. S. A.* 118 (41), e2115001118. doi:10.1073/pnas.2115001118