Peer-reviewed version available at *Molecules* **2019**, *24*, 3074; doi:10.3390/molecules24173074

Review

Structure and function of multimeric G-quadruplexes

Sofia Kolesnikova [1,2] and Edward A. Curtis [1,*]

[1] The Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 166 10, Czech Republic
[2] Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague 166 28, Czech Republic
* Correspondence: curtis@uochb.cas.cz

Received: date; Accepted: date; Published: date

Abstract: G-quadruplexes are noncanonical nucleic acid structures formed from stacked guanine tetrads. They are frequently used as building blocks and functional elements in fields such as synthetic biology and also thought to play widespread biological roles. G-quadruplexes are often studied as monomers but can also form a variety of higher-order structures. This increases the structural and functional diversity of G-quadruplexes, and recent evidence suggests that it could also be biologically important. In this review we describe the types of multimeric topologies adopted by G-quadruplexes and highlight what is known about their sequence requirements. We also summarize the limited information available about potential biological roles of multimeric G-quadruplexes and suggest new approaches that could facilitate future studies of these structures.

Keywords: G-quadruplex; dimer; tetramer; multimer; oligomer; telomere; promoter; R-loop; DNA:RNA hybrid

1. Introduction

The B-form double helix is the most well known nucleic acid structure but it is not the only one. Other examples include double helices with geometries that differ from that of classical B-form DNA, such as Z-DNA [1], triple helices [2], and even four-stranded structures in which canonical A-T and C-G base pairs are absent [3-4]. Among these noncanonical folds, the G-quadruplex (GQ) has received the most attention. This is a four-stranded structure typically stabilized by stacked guanine tetrads connected by short loops [3, 5-6]. Because of their high stability, structural versatility, and functional diversity, GQs have been widely used as building blocks and functional elements in fields such as synthetic biology [7-8]. Sequences with the potential to form GQs are also abundant in the genomes of higher eukaryotes [9-10], and recent studies using GQ-specific antibodies indicate that these structures can form in the context of cells [11-12]. GQs are thought to play a variety of biological roles. These including regulation of transcription, translation, DNA replication, and RNA localization [13-15].

Most biological studies to date have focused on monomeric GQs. However, GQs can also adopt a variety of multimeric forms. These include relatively small structures such as dimers. They also include larger ones like G-wires, which can contain hundreds of GQ monomers. Although the ability of GQs to multimerize has long been recognized, the possibility that such high-order structures form in the context of cells has received less attention. From our perspective, however, two observations suggest that this possibility should be considered. First, the cellular concentrations of GQs are surprisingly high, especially in higher eukaryotes. For example, current estimates suggest that human cells contain at least 716,000 DNA GQs [16] in a volume (for a HeLa cell nucleus) of 0.22 pL [17]. This corresponds to a cellular GQ concentration of 6 μM. Not all of these GQs will be present at the same time in the cell cycle, and not all will be capable of forming dimers. On the other hand, this concentration is orders of magnitude higher than that required for

Peer-reviewed version available at *Molecules* **2019**, *24*, 3074; doi:10.3390/molecules24173074

efficient GQ multimerization. For example, in a recent study we identified GQs with dissociation constants of dimer formation as low as 35 nM [18]. The hypothesis that GQs form multimeric structures in cells is also compelling when the myriad evolutionary and functional advantages of this mechanism are considered [19-22]. These include the ability to regulate biochemical function based on concentration, to detect ligands with enhanced sensitivity by cooperative binding, and to modulate activity in a rapid and reversible manner by exchanging dimerization partners (Figure 1). Inspired by these considerations, in this review we first describe the types of multimeric topologies adopted by GQs and review what is known about their sequence requirements. We then summarize the limited information currently available about the potential biological roles of multimeric GQs in cells and suggest new approaches that could facilitate future studies of these structures, especially in the context of cells.

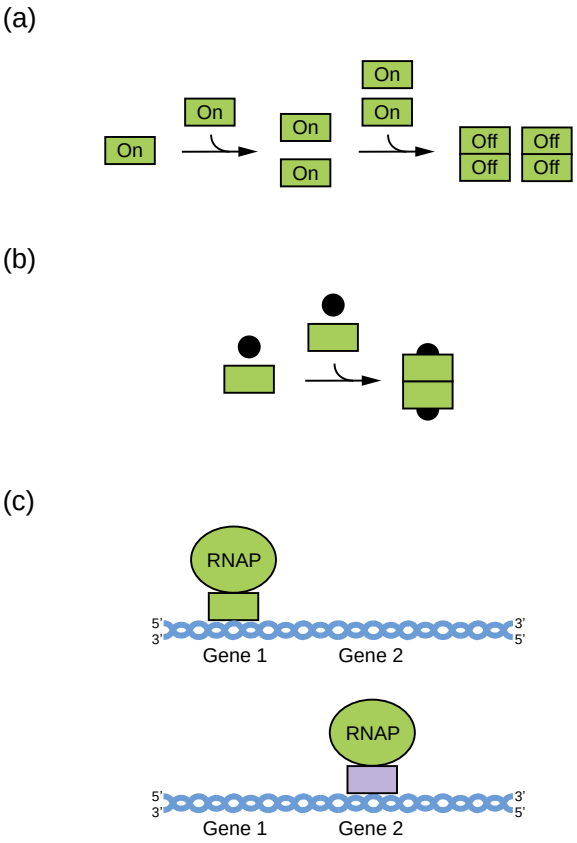


Figure 1. Regulation of biochemical function by multimerization. (**a**) Concentration-based control of biochemical function. In this scheme, monomers are biochemically active and dimers are not. At concentrations below the dissociation constant for dimer formation, most of the population is monomeric and in the active state, while at concentrations above the dissociation constant most of the population is dimeric and in the inactive state. (**b**) Enhanced sensitivity to ligand concentration by cooperative binding. In this scheme, ligand binding is independent when binding sites are monomeric but cooperative when they are linked by multimerization. This leads to all-or-none binding and enhanced sensitivity to ligand concentration. (**c**) Modulation of biochemical activity by the exchange of dimerization partners. In this scheme, the gene transcribed by RNA polymerase is determined by the DNA-binding specificity of its dimerization partner.

2. Definition of multimeric G-quadruplexes and modes of multimerization

A multimer is an aggregate of molecules consisting of multiple monomers. Nucleic acids typically multimerize (hybridize) through duplex formation via the Watson-Crick base pairing of two complementary strands. A double helix provides DNA with structural stability, determined by hydrogen bonding and base-stacking interactions, and facilitates replication [23-24], while

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2019 doi:10.20944/preprints201908.0176.v1

Peer-reviewed version available at Molecules 2019, 24, 3074; doi:10.3390/molecules24173074

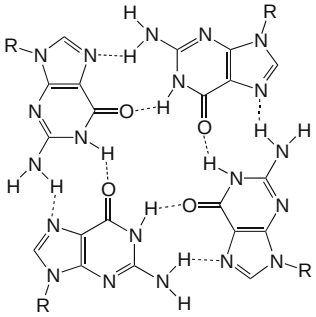


Figure 2. Chemical structure of a GQ tetrad.

double-stranded RNA facilitates genetic interference [25]. Duplex formation is the most common mechanism of multimerization, but not the only one. Multimerization can also occur using a distinct hydrogen-bonding pattern called Hoogsteen base pairing. This enables formation of a G-tetrad, the building block of a unique type of nucleic acid structure called a G-quadruplex (GQ). To form a G-tetrad, four guanines associate via eight hydrogen bonds from both the Watson-Crick and Hoogsteen faces of the base (Figure 2). G-tetrads stack on top of one another giving rise to a GQ (Figure 3). To form a GQ, a sequence typically needs to contain segments of at least two guanines separated by mixed-sequence loop nucleotides. The most widely used models allow loops of 1 to 7 nucleotides [9-10] but in some cases loops can be longer [26-28]. In addition, it is becoming increasingly clear that GQs can accommodate bulges [29], and noncanonical tetrads have also been observed in high-resolution structures [30-36]. Even more complicated topologies are seen in the Spinach, Mango, and Class V GTP aptamers, in which the four clusters of guanines that form the tetrads in the GQ are far apart in the primary sequence [37-39]. Guanines in the G-tetrad can in principle come from one, two, three, or four guanine-rich (G-rich) strands. GQs formed from only one strand are typically defined as intramolecular (unimolecular) (Figure 3A-B), although, as discussed below, multiple GQs on a single strand can also interact to form higher-order structures. GQs that contain more than one strand are termed intermolecular (multimolecular or multimeric) and can be classified according to the number of strands as bimolecular (dimeric) (Figure 3C), trimolecular (trimeric) (Figure 3D), or tetramolecular (tetrameric) (Figure 3E). Multimeric GQs formed from identical strands are called homomultimeric, whereas those composed from non-identical strands are called heteromultimeric.

While the number of molecules in a GQ is the standard way to classify multimers, another approach is to consider the structure of the multimerization interface. GQs utilize two primary modes of multimerization. In the first mode, interfaces are formed by tetrads from two different

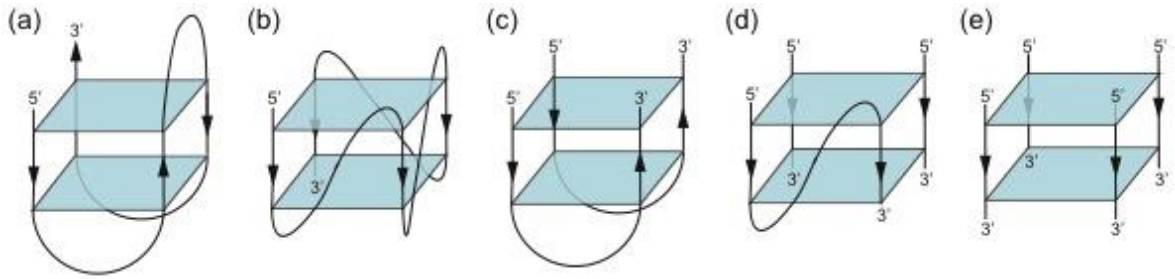


Figure 3. Formation of GQs from different numbers of strands. (**a**) Intramolecular (unimolecular) GQ with antiparallel strands. (**b**) Intramolecular (unimolecular) GQ with parallel strands. (**c**) Bimolecular (dimeric) GQ. (**d**) Trimolecular (trimeric) GQ. (**e**) Tetramolecular (tetrameric) GQ. Note that each of these structures can in principle contain all parallel strands, all antiparallel strands, or a mix of parallel and antiparallel strands.

GQs stacked on top of one another. When intramolecular (unimolecular) GQs multimerize using this mechanism, individual nucleic acid strands first assemble into monomeric GQs such that all four guanines in each G-tetrad come from the same nucleic acid strand. Monomeric GQ subunits then stack on top of one another via π-π stacking interactions of terminal interfaces to form higher-order GQ structures (Figure 4A). Intermolecular (multimeric) GQs can also multimerize in this way if all of the guanines in one of the tetrads at the interface come from one GQ and all of the guanines in the other come from a different GQ (Figure 4A). GQ subunits can stack in three different orientations: 5′ to 3′ (head-to-tail), 5′ to 5′ (head-to-head), and 3′ to 3′ (tail-to-tail). Examples of these orientations can be found in [40] (5′ to 3′), [41] (5′ to 5′), and [42] (3′ to 3′). Experimental and molecular dynamics data suggest that 5′ to 5′ stacking is the most common orientation for GQ structures due to a favorable stacking geometry [41, 43]. The propensity of subunits to stack is also influenced by the topology of intermolecular GQ subunits. Stacking is favorable for parallel GQs (i.e. GQs with all strands oriented in the same direction and with propeller loops on the sides of the tetrads). In antiparallel GQ structures, which have strands in opposite orientations, lateral and diagonal loops are positioned above and below the GQ axis which can impede stacking interactions of terminal G-tetrads [44]. Stacking interactions can also facilitate formation of higher-order structures from tandem GQ subunits folded on a single strand. Such a structure was proposed as one of the models of the human telomere overhang [45-46]. In addition to playing important roles in multimerization, stacking interactions are the most important mode of ligand binding to GQs. End-stacking GQ ligands either bind to terminal GQs tetrads or intercalate between tandem GQs to stabilize the multimeric structure [47-52].

In the second mode of multimerization, guanines from two or more nucleic acid strands hydrogen bond to form tetrads, so that interfaces occur within rather than between tetrads. These intermolecular G-tetrads then stack to form GQs of various lengths (Figure 4B), sometimes using slipped strands [53]. Various G-rich oligonucleotides multimerize via this mode, and a large body of literature has investigated their formation [5-6]. As early as 1988, Sen and Gilbert demonstrated that oligonucleotides containing motifs of four, five or six contiguous guanines fold into tetrameric structures (Sen and Gilbert, 1988) [54]. One year later, two different groups proposed the formation of dimeric GQs from two telomere-derived G-rich sequences [55-56]. Sen and Gilbert [57] observed that short oligonucleotides with three consecutive guanines at the 3′ end assembled

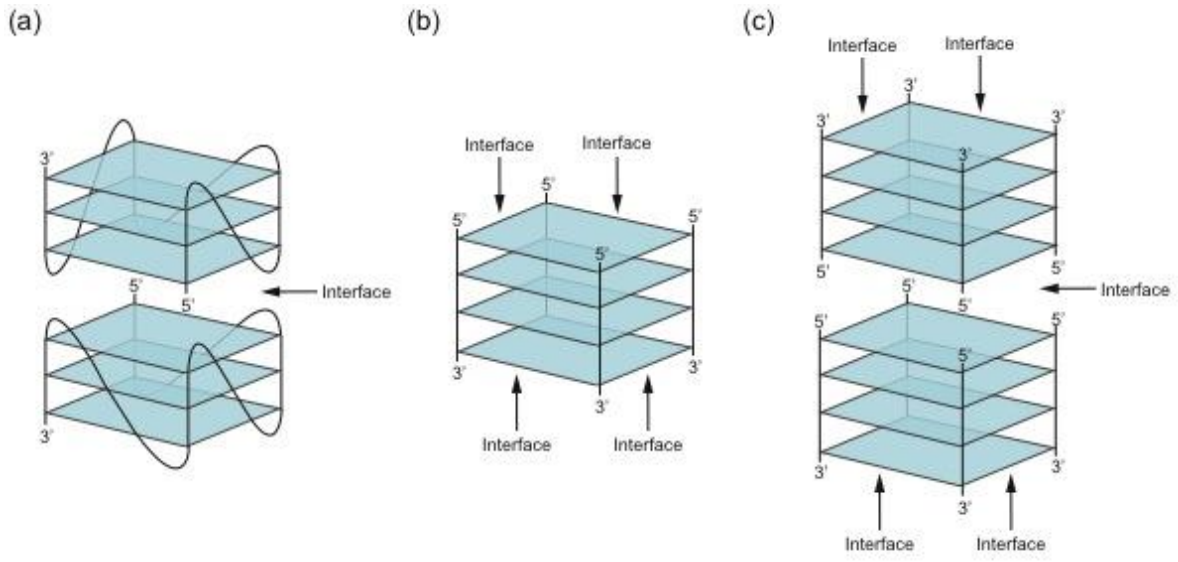


Figure 4. Types of interfaces in multimeric GQs. (**a**) First mode of multimerization. Interfaces are formed between tetrads which stack on top of one another in a 5' to 5' , 3' to 3', or 5' to 3' arrangement. (**b**) Second mode of multimerization. Interfaces are formed within tetrads made up of guanines from multiple DNA strands. (**c**) Structure combining these two modes of multimerization.

Peer-reviewed version available at *Molecules* **2019**, *24*, 3074; doi:10.3390/molecules24173074

into four, eight and twelve-stranded GQ structures. Formation of these structures starts with the self-assembly of slipped tetramers in which strands are not perfectly aligned but contain two overhanging guanines at the 3′end. Tetrameric subunits associate with one another via hydrogen bonding between these extra guanines.

Higher-order GQs which combine these two modes of multimerization have also been described [18, 42, 58-61]. An illustrative example is a GQ structure assembled from eight d(TGGGGT) strands (Figure 4C). The structural subunit is a tetramolecular GQ consisting of four perfectly aligned strands, each in an identical 5′-3′ orientation (second mode of multimerization). Two structural subunits then stack at their 5′ interfaces forming an octamer (first mode of multimerization) [58-59].

Higher-order GQ structures are often initially characterized using low resolution techniques such as circular dichroism, dimethylsulfate footprinting, native PAGE, mass spectrometry, and analytical ultracentrifugation. These methods can be used to establish that the sequence forms a GQ, identify the guanines in tetrads, and determine the number of strands in the structure. Higher-order GQs can be visualized in greater detail using NMR and X-ray crystallography. Examples of high-resolution structures which utilize the first mode of multimerization include 5'-5' stacked dimers with canonical [41, 62-63] or extended [64] tetrads at the interface. Structures which use the second mode of multimerization include interlocked dimers which occur in the promoters and introns of oncogenes [65-67]. A structure which combines these two modes of multimerization is that of a parallel-stranded tetrameric GQ formed from d(TGGGGT) strands [58-59]. Additional examples are discussed in [5, 50, 68] and elsewhere.

3. Sequence requirements of multimeric G-quadruplexes

The propensity of a GQ sequence to fold into a particular multimeric structure depends on a number of factors including the type and concentration of cation in the buffer, the presence of molecular crowding agents, the nucleic acid concentration, the sequence length and the sequence composition [5, 68-70]. Under conditions approximating the intracellular environment, the main factor controlling higher-order structure formation is the sequence. In the context of GQ

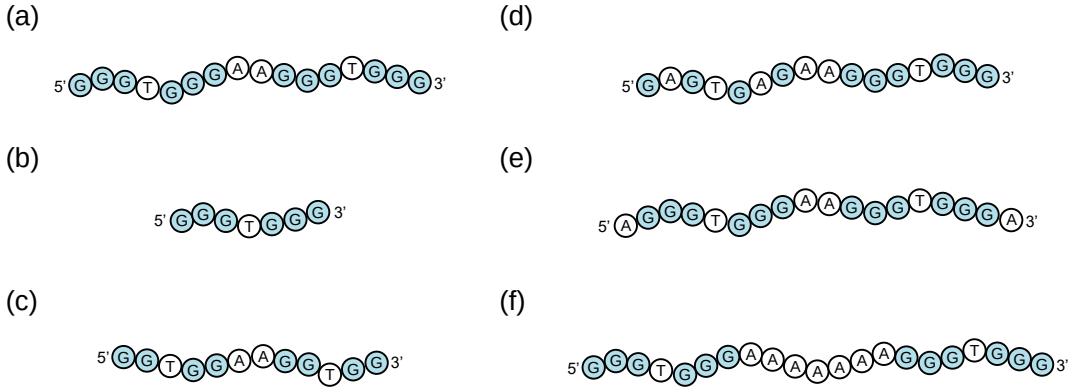


Figure 5. Sequence requirements of multimeric GQs. (**a**) Example of a canonical GQ. (**b**) Variant containing two rather than four G-runs. Such a sequence can form a multimeric GQ but not a monomeric one. (**c**) Variant containing G-runs of two rather than three nucleotides. In some cases such sequences form multimeric rather than monomeric structures. (**d**) Variant containing mutations in tetrads. Such mutations can induce formation of both dimeric and tetrameric structures. (**e**) Variant containing overhanging nucleotides. Such variants typically cannot stack via the first mode of multimerization, but the ability to interact via the second mode of multimerization is unaffected. (**f**) Variant containing an extended central loop. Longer loops favor formation of antiparallel rather than parallel GQs, and such loops can interfere with the ability of GQs to stack via the first mode of multimerization.

Peer-reviewed version available at *Molecules* **2019**, *24*, 3074; doi:10.3390/molecules24173074

formation, a large number of sequences, ranging from short strands with only one G-run to up to 20000 nucleotide long sequences with roughly 3300 G-runs, have been investigated (58-59, 71). The key sequence features affecting multimerization are discussed below and summarized in Figure 5.

The minimum requirement for a sequence to fold into an intramolecular GQ is typically four runs of at least two guanine bases separated by loops ranging from one to seven nucleotides (Figure 5A). Sequences falling into this category often either fold into monomeric GQs that do not interact with one another or form multimeric structures by the first mode of multimerization (stacking of monomeric subunits). Structure prediction is nevertheless complicated by reports of sequences containing four stretches of guanines which form intertwined dimers using the second mode of multimerization [65-67, 72]. Sequences with fewer than four G-runs can typically only adopt GQ structures by interacting to form multimeric structures using the second mode of multimerization (Figure 5B). Sequences with two or three G-runs separated by short loops usually combine both types of multimerization and readily assemble into structures with eight, twelve, or even more strands [57]. At the extreme end of this continuum lie the G-wires [73], long linear ladder-like structures formed from numerous slipped GQ tetrameric subunits (Figure 6). G-wires are longer than any other higher-order GQs with maximum lengths depending on the method of preparation as well as the sequence. For example, a DNA sample of d[$G_4T_2G_4$] formed linear G-wires ranging from 7 to 100 nm in length. A 100 nm G-wire was calculated to contain 75 GQ blocks composed of 140 full and 20 half strands [74]. Positioning of the GQ subunits within the structure can be controlled by adding GC bases to the terminal ends, which enables the formation of G:C:G:C tetrads linking the subunits [75].

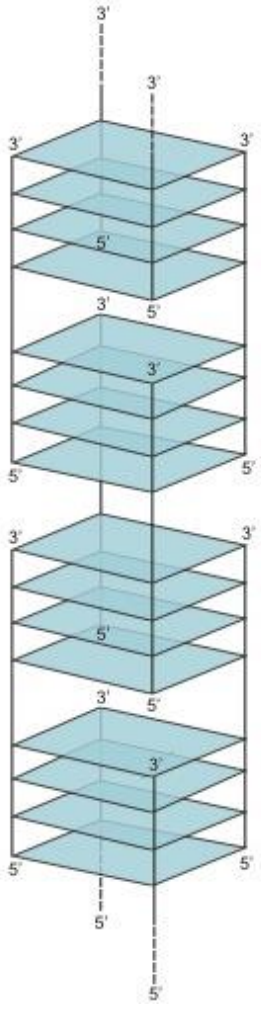


Figure 6. Structure of a G-wire.

In addition to the number of G-runs in the sequence, multimerization also depends on the length of the G-run (Figure 5C). An instructive example is a study of structural transitions caused by truncations of guanine tracts in the telomere-derived DNA strand d($G_4T_4G_4$) [76]. The original sequence and its 5′ truncated analog d($G_3T_4G_4$) formed stable dimeric structures that did not undergo conversion into higher-order structures. In contrast, the 3′ truncated sequences d($G_3T_4G_3$) and d($G_4T_4G_3$) formed a mixture of dimeric, trimeric and tetrameric GQs in a concentration-dependent manner. This structural polymorphism can be explained by varying stabilities of the dimeric structures formed by the reference and the various truncated sequences. The reference d($G_4T_4G_4$)$_2$ dimer contains sixteen guanines which form four G-tetrads, and is therefore more stable than the three-tetrad dimers formed by truncated sequences. A bimolecular structure formed from two d($G_3T_4G_4$) strands is special in that, in contrast to the antiparallel dimers formed by d($G_3T_4G_3$) and d($G_4T_4G_3$), it contains three parallel strands. This specific strand orientation stabilizes the d($G_3T_4G_4$)$_2$ dimer and thereby prevents it from rearranging into higher-order structures.

A third feature of G-runs that can affect multimerization is the presence of bulges (interruptions of G-runs by non-G nucleotides) (Figure 5D). Despite bulges being tolerated by various GQs, they can greatly influence structural stability. The extent to which GQ stability is reduced depends on a number of factors such as the location of the bulges within the GQ sequence, the context of the sequence and the overall GQ topology [29]. The introduction of bulges does not necessarily result in structural changes, and in some cases they do not prevent monomeric GQs from forming. On the other hand, GQs with reduced stability are prone to structural reorganizations, including higher-order structure formation. Bulged nucleotides as drivers of multimerization were systematically investigated in a recent study that analyzed dimerization and tetramerization as a function of guanine substitutions in G-runs [18]. The reference sequence used in this study, d($G_3TG_3AAG_3TG_3A$), folds into a parallel three-tetrad GQ with nucleotides 2, 6, 11 and 15 forming the central tetrad (Figure 7A). Despite containing an exposed 5' tetrad, only monomers were observed when this sequence was analyzed on native gels. Dimerization and tetramerization was induced by introducing substitutions of guanines for other nucleotides at certain positions in the central tetrad. Bulged G-runs prevented these sequences from forming stable intramolecular GQs by themselves and increased their propensity to form multimeric structures. Sequences containing substitutions at position 2, 6 or both were proposed to form

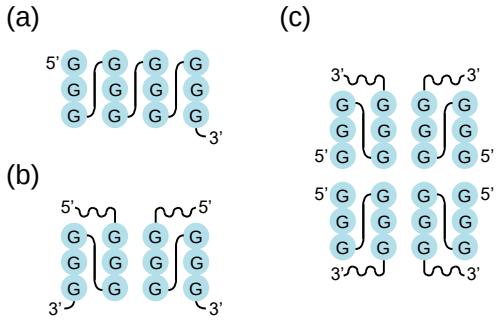


Figure 7. Mutations in tetrads can induce GQ multimerization. (**a**) Secondary structure of a GQ we are studying in our group with the sequence GGGTGGGAAGGGTGGGA. We previously generated a library containing all possible mutations in the central tetrad in this structure (at positions 2, 6, 11 and 15) and tested these variants for a series of biochemical activities associated with GQs [18, 82-83 127]. (**b**) Proposed secondary structure of dimers formed by variants containing mutations at positions 2, 6, or both in the central tetrad of the reference GQ. The 5' part of this structure, which contains the mutated nucleotides, is represented by a wavy black line. (**c**) Proposed secondary structure of tetramers formed by variants containing mutations at positions 11, 15, or both in the central tetrad of the reference GQ. The 3' part of this structure, which contains the mutated nucleotides, is represented by a wavy black line.

intertwined dimers with a core of three canonical G-tetrads at the 3′ end (Figure 7B). Sequences with substitutions at positions 11, 15 or both were instead proposed to form intertwined dimers containing a core of three canonical G-tetrads at the 5′ end (Figure 7C). Such sequences also formed tetramers which, in accord with the most favorable mode of stacking, were proposed to consist of two dimers stacked in a 5′ to 5′ orientation (Figure 7C). Intertwined dimers formed from sequences with other possible combinations of substitutions (2 and 11, 2 and 15, 6 and 11, 6 and 15) would only contain two stacked G-tetrads [18]. This structural arrangement cannot effectively stabilize the structures, which probably explains why these sequences did not also multimerize.

GQ sequences can be designed to begin and/or end with G-runs or they can contain overhanging non-G nucleotides at the 5′ and/or 3′ ends (Figure 5E). Flanking nucleotides often inhibit stacking interactions by sterically hindering formation of interfaces by terminal tetrads [57, 77]. In some cases even the introduction of a 5' phosphate group is sufficient to inhibit 5'-5'stacking [73]. This knowledge can be used when probing multimerization mechanisms: for multimers stacked in a 5′ to 5′ orientation, addition of flanking nucleotides at the 5′ end should interfere with the stacking interaction, while addition of flanking nucleotides to the 3' end should not [18]. Nevertheless, some studies report the stacking of GQ subunits which contain overhanging nucleotides. In these structures, flanking nucleotides either radiate out from the GQ axis or become a part of various unusual structures such as thymine triads [78], guanine:uridine octads [79] or guanine:cytidine octads [60].

The role of loop nucleotides in multimeric GQ assembly has received relatively little attention from researchers, and most studies investigating multimeric GQs have not analyzed the effects of loop length and sequence composition. Despite this, examples in which loop nucleotides affect GQ multimerization have been described. Loop length can indirectly influence multimerization by affecting GQ topology (Figure 5F). Sequences bearing at least one single-nucleotide loop tend to adopt a parallel-stranded orientation, whereas longer loops increase the likelihood that sequences will fold into mixed-strand or antiparallel GQs [80]. As discussed above, this can affect the ability of GQs to interact by the first mode of multimerization: the propeller loops of parallel strand GQs do not generally affect their ability to form multimers through stacking on their terminal tetrads, whereas the lateral and diagonal loops of antiparallel GQs sometimes do. Loop sequence can also play an important role in GQ structure formation. One mechanism by which this can impact the global structure of the GQ is by formation of non-canonical extended G-tetrads. For example, $d(GGA)_4$ strands assemble into dimers by stacking on a guanine:adenine heptad interface formed by two monomeric subunits. Three loop adenines provide hydrogen bonds donors and acceptors in the heptad plane and hence play an indispensable role in dimerization [64]. Moreover, several studies have shown that mutating nucleotides not thought to directly interact with tetrads can lead to major changes in the GQ structure. For example, changing the loops in a human telomere-derived sequence from TTA to AAA abolishes GQ formation [81]. Recent findings in our lab also show that mutations in loops can affect the propensity of sequences to form multimeric GQs [82-83].

In summary, these studies indicate that the main sequence features affecting multimerization are the number and length of the G-runs, the sequence composition of the G-runs, the presence of flanking nucleotides at the 5' and 3' ends of the GQ, and the length and sequence of loop nucleotides (Figure 5). Despite much being understood about the sequence requirements of multimeric GQs, it is generally not possible to predict higher-order GQ topology from sequence. Instead, experimental data from low-resolution methods such as circular dichroism, dimethylsulfate footprinting, native PAGE, mass spectrometry, and analytical ultracentrifugation is used in combination with high-resolution techniques such as NMR and X-ray crystallography.

4. Potential biological roles of multimeric GQs at the tips of telomeres

The first discussions of biologically relevant higher-order GQ structures emerged as the length and sequence composition of telomeres was being characterized [84]. In the majority of eukaryotes, the nucleotide sequence of telomeres consists of a G-rich strand running 5′ to 3′ toward the end of the chromosome and a complementary C-rich strand. For most of their length, telomeres are

double-stranded, except for a single-stranded G-rich 3'overhang. While the telomeric sequence is similar in diverse species, the length of the overhang is not. For example, in *Tetrahymena* the overhang typically consists of tandem repeats of the sequence d(TTGGGG) and is 14 to 21 nucleotides long [85]. On the other hand, in humans the sequence of the overhang is d(TTAGGG) and it ranges in length from 125 to 275 nucleotides [86-87].

While the composition and length of the telomeric overhang has been characterized in many species, its secondary structure is still a matter of debate. The first studies of short telomere-derived synthetic oligonucleotides reported that these sequences folded into several topologically different guanine-rich structures [54-56, 88]. Subsequent studies showed that these correspond to GQ structures, although several different topologies have been reported [6, 89-90]. Examples include parallel-stranded tetrameric and antiparallel dimeric GQs assembled from short telomeric sequences bearing one and two G-runs, respectively as well as intramolecular monomeric GQs formed by sequences containing four G-runs.

Because the human telomeric overhang contains multiple repeats of segments with four G-runs, it can potentially fold into a higher-order structure containing a series of monomeric GQs [91]. A single-stranded DNA with several monomeric GQs can adopt at least two different structural arrangements (Figure 8). The first one is termed "beads-on-a-string", and assumes that monomeric GQ units do not interact with one another [92]. A second model proposes that GQ subunits associate with one another through stacking interactions between the terminal tetrads of consecutive GQs to form high-order structures [45-46]. Despite evidence for both models, a structure in which GQs stack on one another appears to be most likely [91]. Structures consistent with this model have been observed by AFM and EM [71, 93-94], and methods such as FRET have provided additional evidence for interactions between neighboring GQ subunits [94]. Molecular dynamics simulations also support the idea that telomeric GQ monomers can stack to form higher order structures [45-46]. Based on comparisons between calculated and experimentally measured sedimentation coefficients it has been proposed that a structure consisting of hybrid rather than all-parallel GQs is most likely [95]. Despite these advances, a high-resolution structure of the telomeric overhang has not yet been reported.

In vivo evidence for the presence of GQs at telomeres came from studies reporting the binding of GQ-specific antibodies to the tips of telomeres [11-12, 96]. These experiments also showed that telomerase co-localizes with GQ antibodies at chromosomal ends. Telomerase catalyzes telomere elongation, and is directly involved in GQ unfolding [97]. Although these studies indicate that telomeric DNA adopts a GQ structure *in vivo*, they do not distinguish between monomeric and multimeric GQs since the antibody has similar affinities for both types of structures. While

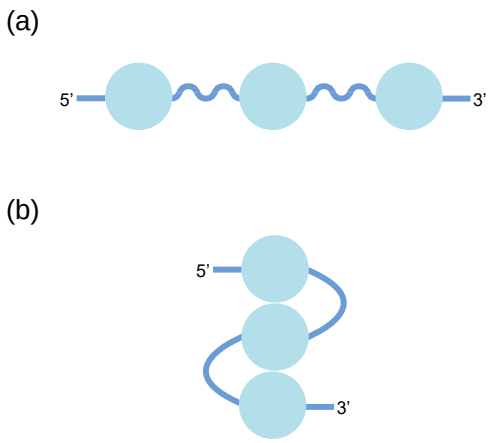


Figure 8. Possible structures of telomeric GQs. (**a**) Beads-on-a-string model in which telomeric GQs do not interact. (**b**) Model in which telomeric GQs stack on one another to form higher-order structures.

Peer-reviewed version available at Molecules 2019, 24, 3074; doi:10.3390/molecules24173074

antibodies specific for multimeric GQs have not yet been developed, several small molecules specific for multimeric GQs have recently been reported [51, 98-100]. One of these compounds, IZNP-1 (triaryl-substituted imidazole derivative), binds and stabilizes multimeric GQs by intercalating into the cavity between two stacked GQ monomers. IZNP-1 treated cells contain a greater number of BG4 foci (human GQ structure-specific antibody foci) at telomeres than untreated cells. They also exhibit signs of DNA damage and telomere shortening, which can lead to cell cycle arrest, apoptosis and senescence. Taken together, these findings suggest that multimeric GQs can form at telomeres under certain conditions and contribute to telomere dysfunction [51].

5. Potential biological roles of multimeric G-quadruplexes in promoters

Visualization of IZNP-1 induced BG4 foci confirmed that telomeric ends have the ability to form multimeric GQs. Surprisingly, however, the percentage of fluorescent foci at telomeres was relatively low (38%) and the majority of antibodies localized to other genomic regions. One implication of this observation is the possibility that multimeric GQs exist outside telomeric DNA. If so, multimeric GQs might have roles beyond chromosomal capping.

A role for monomeric GQ structures in transcriptional regulation was proposed following the discovery that putative GQ sequences are highly enriched in nuclease hypersensitive regions within promoters [13]. In addition, more than 40% of human genes contain at least one GQ motif in their promoter. While formation of monomeric intramolecular GQs has been investigated for a number of these genes including c-MYC, KRAS, and BCL2 [101-104], only a few examples of genes likely to be regulated by multimeric GQs in their promoters have been identified so far. The most well-characterized example is an unusual tetrad:heptad:heptad:tetrad (T:H:H:T) dimer generated from a d(GGA) repeat sequence. An oligonucleotide containing four such repeats folds into an intramolecular structure in which a guanine tetrad and a guanine-adenine heptad are stacked on top of one another. The NMR structure indicates that two monomeric subunits stack 5′ to 5′ on the heptad plane, resulting in a dimeric T:H:H:T structure [64] (Figure 9). A similar structure was adopted by an oligonucleotide containing eight copies of the repeat [105]. The d(GGA) repeat sequence occurs in twelve nearly perfect tandem repeats in the c-MYB promoter. In theory, this sequence can give rise to three independent T:H building subunits (with four tandem repeats per building block), which could then stack in different combinations to form dimers. A study investigating this possibility proposed that a T:H:H:T motif formed by stacking interactions between the first and the third building blocks functions as a negative regulator of c-MYB promoter activity. The authors also identified a transcription factor, MAZ, which binds to the T:H:H:T structure and could play a role in the regulation of c-MYB expression [106-107]. The c-MYB promoter is not the only genomic region rich in GGA repeats, and similar motifs have been identified in the regulatory regions of other genes such as NCAM [108], SPARC [109], KRAS [110],

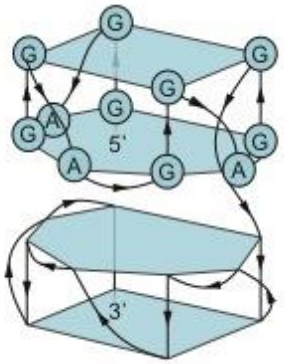


Figure 9. A dimeric GQ formed by GGA repeats. Structure of a dimeric GQ formed by the sequence $(GGA)_8$. See [64] and [105] for more information about this structure.

and CCNB1IP1 [111] (Islam *et al*., 2019). Although it is tempting to assume that the transcription of genes with multiple d(GGA) repeats in the promoter is regulated by T:H:H:T structures, the results presented in [106] and [107] do not completely rule out the possibility that inhibition of transcription is mediated by monomeric forms of these GQs. Testing the effects of antibodies or small molecules specific for the T:H:H:T structure on the extent of transcriptional inhibition could provide further insight into the structural organization of the d(GGA) repeat region in this promoter.

Another multimeric GQ structure that could potentially regulate gene expression occurs in the promoter of hTERT, the gene encoding the reverse transcriptase of telomerase. The core promoter of hTERT contains twelve consecutive G-runs and can therefore form multiple GQs. One proposed structural model for this region consists of two stacked GQs separated by a 26 nucleotide loop which contains another four G-runs and adopts a hairpin structure [112]. An alternative model suggests that all of the G-runs in the sequence participate in GQ folding and that the final structure consists of three parallel GQs tightly stacked on top of one another [40, 113]. Although the available evidence does not indicate which model is correct, it does support the idea that a G-rich structure in this promoter regulates hTERT expression [112-115].

The hypothesis that multimeric GQs in promoters regulate gene expression is also supported by a study investigating the regulatory sequences of muscle-specific genes [116]. The promoter regions of these genes contain a high frequency of G-runs that form both homodimeric and heterodimeric structures *in vitro*. One example is the ITGA7 promoter region, which contains two G-rich sequences, each capable of forming an intramolecular monomeric GQ, separated from one another by 85 nucleotides. These two G-rich sequences formed a heterodimeric structure when analyzed on native gels. This structure is hypothesized to play a role in the regulation of mouse myogenic gene expression. Consistent with this hypothesis, the myogenic determination protein MyoD binds the heterodimeric structure but not a monomeric GQ formed by one of the heterodimer-forming sequences. MyoD exhibited a similar binding preference for a homodimeric GQ formed by a G-rich region in the promoter of the sMtCK gene. Although this G-rich sequence occurs only once in the sMtCK promoter region and therefore cannot assemble into a homodimer *in vivo*, several G-rich clusters are present in adjacent genomic regions. It would therefore be of interest to determine whether a heterodimeric GQ structure can form in this region and is bound by MyoD [116].

Taken together, these data indicate that multimeric GQ structures can form in the promoter regions of several genes and are consistent with the idea that such structures can regulate gene expression. Many additional promoters contain multiple G-runs. For example, genes containing more than eight G-tracts include TRIM13 [67] and c-MYC [102-103]. Although these are candidates for genes regulated by multimeric GQs, our current understanding of GQ folding precludes reliable prediction of higher-order GQ structure from sequence. As discussed above, the development of antibodies or small molecules that selectively stabilize or destabilize specific multimeric GQ structures should greatly facilitate analysis of their potential biological roles in promoters.

6. Potential biological roles of RNA-DNA hybrid G-quadruplexes

In addition to the multimeric DNA GQs discussed so far, multimeric GQs can also form from combinations of DNA and RNA strands [117]. These structures are called DNA:RNA hybrid GQs. They can be readily assembled *in vitro*, and also form during transcription when guanines in the nontemplate DNA strands of genes interact with guanines in RNA molecules encoded by these genes using the second mode of multimerization (Figure 10). DNA:RNA hybrid GQs inhibit transcription, presumably by impeding the progress of RNA polymerase along the DNA template. From the perspective of gene regulation, this is a compelling mechanism because it can act as a negative feedback loop to turn transcription off when mRNA levels are sufficiently high (Figure 1A). The first hint that DNA:RNA hybrid GQs act as regulatory elements came from analysis of a region in the human mitochondrial genome called conserved sequence block II (CSB II) [118]. A G-rich element in CSB II induces transcriptional termination and promotes formation of an unusually stable structure in the template DNA. The stability of this structure suggested that it was not a

Peer-reviewed version available at *Molecules* **2019**, *24*, 3074; doi:10.3390/molecules24173074

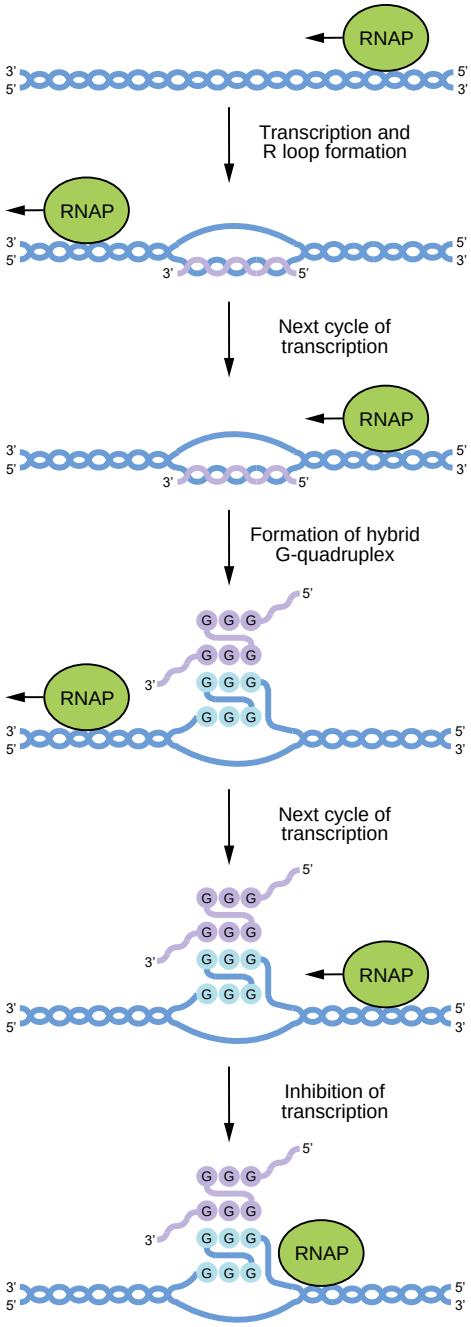
conventional R-loop, in which an RNA transcript and the DNA template encoding it are held together by Watson-Crick base pairing. Instead, it was proposed that this structure was a GQ containing both DNA and RNA strands. Consistent with this hypothesis, a structure was detected by native PAGE that required for its formation the presence of both a DNA oligonucleotide corresponding to the G-rich strand of the template and an RNA oligonucleotide corresponding to the G-rich portion of the RNA transcript. Formation of this structure was inhibited when guanine bases in the DNA strand were replaced by 7-deazaguanine or adenosine, both of which normally prevent GQ formation. The circular dichroism spectrum of this structure was consistent with that of a GQ, and it was resistant to ribonucleases A and H, both of which degrade RNA strands in canonical DNA:RNA duplexes. Additional characterization using DMS fingerprinting, which can distinguish guanines in tetrads from those in single or double-stranded DNA due to their reduced sensitivity to DMS modification, and Zn-TTAPc-mediated photocleavage, which preferentially cleaves GQs, provided additional evidence that the structure is a DNA:RNA hybrid GQ [119]. Similar methods were used to show that a DNA:RNA hybrid GQ also forms in the human NRAS gene [120]. As was the case for the CSB II motif, the structure in the NRAS gene impeded T7 RNA polymerase during *in vitro* transcription reactions. It also inhibited expression of luciferase from plasmids transfected into cultured human cells. Only two stretches of guanines in the non-template strand of the gene are needed to form the DNA:RNA hybrid GQ, with the other two provided by the RNA transcript encoded by the template strand. DNA:RNA hybrid GQs containing two rather than three tetrads can also form, although they are less stable than those containing three tetrads [121]. Bioinformatic studies indicate that sequences with the potential to form DNA:RNA hybrid GQs occur in 97% of human genes [122]. These motifs are enriched in and near promoters, especially downstream of the transcription start site on the nontemplate strand, and this pattern of enrichment is evolutionarily conserved in mammals [122]. Taken together, these studies support the idea that DNA:RNA hybrid GQs represent an important regulatory element in higher eukaryotes.

Additional studies have explored the mechanism of formation of DNA:RNA hybrid GQs during transcription. Zhang and colleagues developed a method to distinguish R-loops from DNA:RNA hybrid GQs based on differences in the sensitivities of these structures to ribonucleases such as RNase A [123]. This was used to show that conventional R-loops form prior to DNA:RNA hybrid GQs during *in vitro* transcription of a template containing the CSB II motif [123]. Formation of DNA:RNA hybrid GQs was also inhibited by the presence of a C-rich oligonucleotide complementary to the G-rich region of both the nontemplate strand and the nascent RNA transcript. Based on these experiments it was proposed that a canonical R-loop is formed in the first round of transcription of the CSB II motif. After the RNA strand in the R-loop is displaced by RNA polymerase in the next round of transcription, it forms a DNA:RNA hybrid GQ with the nontemplate strand and prevents further transcription (Figure 11). Real-time FRET studies indicate that R-loops also stabilize DNA:RNA hybrid GQs [124], probably because they prevent the nontemplate strand from forming a duplex with the template strand. The mechanical strength of DNA:RNA hybrid GQs has also been investigated using optical tweezers [125]. These experiments used a template containing the sequence $(GGGGA)_4$, which occurs downstream of the transcription start site on the nontemplate strand in several hundred human genes [125]. Transcription reactions were performed in the absence of CTP, which caused the polymerase to stall 15 nucleotides downstream of the GQ at the position of the first G in the template. Stalled complexes were tethered between beads and stretched to determine their mechanical stabilities. This revealed that DNA:RNA hybrid GQs are more mechanically stable and form more readily in the nontemplate strand than DNA GQs.

Recent results from our laboratory raise the possibility that formation of DNA:RNA hybrid GQs can be regulated by biologically important small molecules [83]. These experiments used a mutant GQ previously shown by our group to bind GTP [126-127] and to form multimers [18]. When folded in the absence of GTP, this sequence forms multimeric GQs as well as monomers that are probably unfolded. When folded in the presence of physiological concentrations of GTP, however,

Peer-reviewed version available at Molecules 2019, 24, 3074; doi:10.3390/molecules24173074

464 the monomeric form of the GQ is stabilized and formation of multimers is supressed [83]. NMR
465 studies indicate that the GTP ligand is incorporated into a tetrad, and together with other data
466 support a model in which GTP is incorporated into a cavity in the central tetrad of the monomer
467 created by a G to A mutation. Hundreds of examples of sequences with the potential to form
468 GTP-dependent structures were identified in the human genome, some of which were evolutionarily
469 conserved in primates. Ongoing experiments in our group are investigating possible links between
470 GTP-dependent formation of DNA:RNA hybrid GQs and transcriptional regulation.



471 **Figure 10.** Model for the regulation of transcription by DNA:RNA hybrid GQs formed between the
472 noncoding strands of G-rich genes and RNA molecules transcribed from these genes. The newly
473 synthesized RNA transcript is shown in purple. See [123] for more information about this model.

474 **7. Conclusions**

GQs can form a variety of multimeric structures ranging in size from dimers to G-wires. Most contain interfaces formed by either stacked tetrads in adjacent GQ monomers (first mode of multimerization) or guanines from two or more nucleic acid strands that hydrogen bond to form a G-tetrad (second mode of multimerization). Some progress has been made in understanding the sequence requirements of multimeric GQs. For example, overhanging nucleotides often inhibit multimerization by interfering with the stacking of tetrads, while mutations in tetrads can promote multimerization by destabilizing the monomeric form of the structure. Despite this, the existence of unusual folds such as intertwined dimers formed from sequences containing four stretches of guanines highlight our inability to predict the structures of multimeric GQs from primary sequence. The propensity of GQs to multimerize *in vitro*, the high concentrations of GQs in eukaryotic cells, and the advantages of multimerization as a regulatory mechanism raise the possibility that GQ multimerization could be biologically important. Although this hypothesis is only starting to be explored, several lines of evidence suggest that higher-order GQ structures form in both telomeres and promoters. Furthermore, recent studies demonstrate that DNA:RNA hybrid GQs formed between the nontemplate strands of genes and the RNA transcripts they encode inhibit transcription, and suggest that this could be an important regulatory mechanism in higher eukaryotes. We anticipate that future progress in defining the biological roles of multimeric GQs will require the development of tools analogous to those used to characterize monomeric GQs. Examples include bioinformatic models that can identify sequences with the potential to form multimeric GQs in sequenced genomes, antibodies and small molecules specific for different types of multimeric GQs, techniques to visualize multimeric GQs in cells, and high-throughput methods to map multimeric GQs in genomic DNA (including those containing both DNA and RNA strands). The development and application of such tools has the potential to provide a wealth of new infomation about multimeric GQs, especially with respect to their potential biological roles.

Author Contributions: Conceptualization, E.A.C.; Writing – Original Draft Preparation, S.K. and E.A.C; Writing – Review & Editing, S.K. and E.A.C; Visualization, S.K. and E.A.C; Supervision, E.A.C.; Project Administration, E.A.C.; Funding Acquisition, E.A.C.

Funding: This work was supported by an IOCB start-up grant awarded to E.A.C. and "Chemical biology for drugging undruggable targets (ChemBioDrug)" No. CZ.02.1.01/0.0/0.0/16_019/0000729 from the European Regional Development Fund (OP RDE).

Acknowledgments: We thank Václav Veverka and colleagues at the IOCB for helpful discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rich, A. DNA comes in many forms. *Gene* **1993**, *135*, 99-109.
2. Hu, Y.; Cecconello, A.; Idili, A.; Ricci, F.; Willner ,I. Triplex DNA nanostructures: from basic properties to applications. *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 15210-15233.
3. Davis, J.T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 668-698.
4. Day, H.A.; Pavlou, P.; Waller, Z.A. I-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407-4418.
5. Burge, S.; Parkinson, G.N.; Hazel, P.; Todd, A.K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402-5415.
6. Phan, A.T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **2010**, *277*, 1107-1117.
7. Yatsunyk, L.A.; Mendoza, O.; Mergny, J.L. "Nano-oddities": unusual nucleic acid assemblies for DNA-based nanostruc-tures and nanodevices. *Acc. Chem. Res.* **2014**, *47*, 1836-1844.
8. Mergny, J.L.; Sen, D. DNA quadruple helices in nano-technology. *Chem. Rev.* **2019**, *119*, 6290-6325.
9. Huppert, J.L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, 2908-2916.

10. Todd, A.K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*. **2005**, *33*, 2901-2907.
11. Lam, E.Y.N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun*. **2013**, *4*, 1796.
12. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem*. **2013**, *5*, 182-186.
13. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*. **2007**, *35*, 406-413.
14. Kendrick, S.; Hurley, L.H. The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl. Chem*. **2010**, *82*, 1609-1621.
15. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*. **2015**, *43*, 8627-8637.
16. Chambers, V.S.; Marsico, G.; Boutell, J.M.; Di Antonio, M.; Smith, G.P.; Balasubramanian, S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol*. **2015**, *33*, 877-881.
17. Fujioka, A.; Terai, K.; Itoh, R.E.; Aoki, K.; Nakamura, T.; Kuroda, S.; Nishida, E.; Matsuda, M. Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes. *J. Biol. Chem*. **2006**, *281*, 8917-8926.
18. Kolesnikova, S.; Hubálek, M.; Bednárová, L.; Cvačka, J.; Curtis, E.A. Multimerization rules for G-quadruplexes. *Nucleic Acids Res*. **2017**, *45*, 8684-8696.
19. Traut, T.W. Dissociation of enzyme oligomers: a mechanism for allosteric regulation. *Crit. Rev. Biochem. Mol. Biol*. **1994**, *29*, 125-163.
20. Goodsell, D.S.; Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct*. **2000**, *29*, 105-153.
21. Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y.; Oliver, S.G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci*. **2008**, *33*, 220-229.
22. Matthews, J.M.; Sunde, M. Dimers, oligomers, everywhere. *Adv. Exp. Med. Biol*. **2012**, *747*, 1-18.
23. Watson, J.D.; Crick, F.H.C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737-738.
24. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res*. **2006**, *34*, 564-574.
25. Fire, A.; Xu, S.; Montgomery, M.K.; Kostas, S.A.; Driver, S.E.; Mello, C.C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **1998**, *391*, 806-811.
26. Smirnov, I.; Shafer, R.H. Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **2000**, *39*, 1462-1468.
27. Bourdoncle, A.; Estevez Torres, A.; Gosse, C.; Lacroix, L.; Vekhoff, P.; Le Saux, T.; Jullien, L.; Mergny, J.L. Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc*. **2006**, *128*, 11094-11105.
28. Guedin, A.; Gros, J.; Alberti, P.; Mergny, J.L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res*. **2010**, *38*, 7858-7868.
29. Mukundan, V.T.; Phan,A.T. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc*. **2013**, *135*, 5017-5028.
30. Cheong, C.; Moore, P.B. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **1992**, *31*, 8406-8414.
31. Patel, P.K.; Hosur, R.V. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats. *Nucleic Acids Res*. **1999**, *27*, 2457-2464.
32. Patel, P.K.; Bhavesh, N.S.; Hosur, R.V. NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG. *Biochem. Biophys. Res. Commun*. **2000**, *270*, 967-971.
33. Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D.J. Dimeric DNA quadruplex containing major groove-aligned A-T-A-T and G-C-G-C tetrads stabilized by inter-subunit Watson-Crick A-T and G-C pairs. *J. Mol. Biol*. **2001**, *312*, 1073-1088.
34. Pan, B.; Xiong, Y.; Shi, K.; Deng, J.; Sundaralingam, M. Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: implications for specific binding in RNA tetraplexes. *Structure* **2003**, *11*, 815-823.
35. Kimura, T.; Xu, Y.; Komiyama, M. Human telomeric RNA r(UAGGGU) sequence forms parallel tetraplex structure with U-quartet. *Nucleic Acids Symp. Ser. (Oxf.)* **2009**, *53*, 239-240.

36. Virgilio, A.; Esposito, V.; Citarella, G.; Mayol, L.; Galeone, A. Structural investigations on the anti-HIV G-quadruplex-forming oligonucleotide TGGGAG and its analogues: evidence for the presence of an A-tetrad. *Chembiochem*. **2012**, *13*, 2219-2224.
37. Huang, H.; Suslov, N.B.; Li, N.S.; Shelke, S.A.; Evans, M.E.; Koldobskaya, Y.; Rice, P.A.; Piccirilli, J.A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **2014**, *10*, 686-691.
38. Nasiri, A.H.; Wurm, J.P.; Immer, C.; Weickhmann, A.K.; Wöhnert, J. An intermolecular G-quadruplex as the basis for GTP recognition in the class V-GTP aptamer. *RNA* **2016**, *22*, 1750-1759.
39. Trachman III, R.J.; Demeshkina, N.A.; Lau, M.W.L.; Panchapakesan, S.S.S.; Jeng, S.C.Y.; Unrau, P.J.; Ferré-D'Amaré, A.R. Structural basis for high-affinity fluorophore binding and activation by RNA mango. *Nat. Chem. Biol.* **2017**, *13*, 807-813.
40. Chaires, J.B.; Trent, J.O.; Gray, R.D.; Dean, W.L.; Buscaglia, R.; Thomas, S.D.; Miller, D.M. An Improved model for the hTERT promoter quadruplex. *PLoS ONE* **2014**, *9*, e115580.
41. Do, N.Q.; Lim, K.W.; Teo, M.H.; Heddi, B.; Phan, A.T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res*. **2011**, *39*, 9448-9457.
42. Kato, Y.; Ohyama, T.; Mita, H.; Yamamoto, Y. Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAGn) (n = 3–5). *J. Am. Chem. Soc*. **2005**, *127*, 9980-9981.
43. Lech, C.J.; Heddi, B.; Phan, A.T. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res*. **2013**, *41*, 2034-2046.
44. Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E.S.; Bowers, M.T.; De Pauw, E.; Gabelica, V. G-Quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc*. **2008**, *130*, 10208-10216.
45. Haider, S.M.; Parkinson, G.N.; Neidle, S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J*. **2008**, *95*, 296-311.
46. Petraccone, L.; Trent, J.O.; Chaires, J.B. The tail of the telomere. *J. Am. Chem. Soc*. **2008**, *130*, 16530-16532.
47. Monchaud, D.; Teulade-Fichou, M.P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **2008**, *6*, 627-636.
48. Bai, L.P.; Hagihara, M.; Jiang, Z.H.; Nakatani, K. Ligand binding to tandem G quadruplexes from human telomeric DNA. *Chembiochem*. **2008**, *9*, 2583-2587.
49. Haider, S.M.; Neidle, S.; Parkinson, G.N. A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **2011**, *93*, 1239-1251.
50. Zhang, S.; Wu, Y.; Zhang, W. G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem* **2014**, *9*, 899-911.
51. Hu, M.H.; Chen, S.B.; Wang, B.; Ou, T.M.; Gu, L.Q.; Tan, J.H.; Huang, Z.S. Specific targeting of telomeric multimeric G-quadruplexes by a new triaryl-substituted imidazole. *Nucleic Acids Res*. **2017**, *45*, 1606-1618.
52. Funke, A.; Karg, B.; Dickerhoff, J.; Balke, D.; Müller, S.; Weisz, K. Ligand- induced dimerization of a truncated parallel MYC G-Quadruplex. *Chembiochem.* **2018**, *19*, 505-512.
53. Krishnan-Ghosh, Y.; Liu, D.; Balasubramanian, S. Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc*. **2004**, *126*, 11009-11016.
54. Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, 65-70.
55. Sundquist, W.I.; Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989**, *342*, 825-829.
56. Williamson, J.R.; Raghuraman, M.K.; Cech, T.R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **1989**, *59*, 871-880.
57. Sen, D.; Gilbert, W. Novel DNA superstructures formed by telomere-like oligomers. *Biochemistry* **1992**, *31*, 65-70.
58. Laughlan, G.; Murchie, A.I.; Norman, D.G.; Moore, M.H.; Moody, P.C.; Lilley, D.M.; Luisi, B. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **1994**, *265*, 520-524.
59. Phillips, K.; Dauter, Z.; Murchie, A.I.; Lilley, D.M.; Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. *J. Mol. Biol*. **1997**, *273*, 171-182.

60. Borbone, N.; Amato, J.; Oliviero, G.; D'Atri, V.; Gabelica, V.; De Pauw, E.; Piccialli, G.; Mayol, L. d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Res*. **2011**, *39*, 7848-7857.
61. D'Atri, V.; Borbone, N.; Amato, J.; Gabelica, V.; D'Errico, S.; Piccialli, G.; Mayol, L.; Oliviero, G. DNA-based nanostructures: the effect of the base sequence on octamer formation from d(XGGYGGT) tetramolecular G-quadruplexes. *Biochimie* **2014**, *99*, 119-128.
62. Kuryavyi, V.; Cahoon, L.A.; Seifert, H.S.; Patel, D.J. RecA-binding pilE G4 sequence essential for pilin antigenic variation forms monomeric and 5' end-stacked dimeric parallel G-quadruplexes. *Structure* **2012**, *20*, 2090-2102.
63. Adrian, M.; Ang, D.J.; Lech, C.J.; Heddi, B.; Nicolas, A.; Phan, A.T. Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. *J. Am. Chem. Soc.* **2014**, *136*, 6297-6305.
64. Matsugami, A.; Ouhashi, K.; Kanagawa, M.; Liu, H.; Kanagawa, S.; Uesugi, S.; Katahira, M. An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.* **2001**, *313*, 255-269.
65. Kuryavyi, V.; Phan, A.T.; Patel, D.J. Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res*. **2010**, *38*, 6757-6773.
66. Trajkovski, M.; da Silva, M.W.; Plavec, J. Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene. *J. Am. Chem. Soc*. **2012**, *134*, 4132-4141.
67. Wei, D.; Todd, A.K.; Zloh, M.; Gunaratnam, M.; Parkinson, G.N.; Neidle, S. Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc*. **2013**, *135*, 19319-19329.
68. Chen, Y.; Yang, D. Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Protoc. Nucleic Acid Chem*. **2012**, Unit 17.5.
69. Yaku, H.; Murashima, T.; Tateishi-Karimata, H.; Nakano, S.; Miyoshi, D.; Sugimoto, N. Study on effects of molecular crowding on G-quadruplex-ligand binding and ligand-mediated telomerase inhibition. *Methods* **2013**, *64*, 19-27.
70. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal cations in G-quadruplex folding and stability. *Front. Chem*. **2016**, *4*, 38.
71. Kar, A.; Jones, N.; Arat, N.Ö.; Fishel, R.; Griffith, J.D. Long repeating (TTAGGG)n single-stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation. *J. Biol. Chem*. **2018**, *293*, 9473-9485.
72. Podbevsek, P.; Plavec, J. KRAS promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units. *Nucleic Acids Res*. **2016**, *44*, 917-925.
73. Marsh, T.C.; Henderson, E. G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry* **1994**, *33*, 10718-10724.
74. Bose, K.; Lech, C.J.; Heddi, B.; Phan, A.T. High-resolution AFM structure of DNA G-wires in aqueous solution. *Nat. Commun*. **2018**, *9*, 1959.
75. Hessari, N.M.; Spindler, L.; Troha, T.; Lam, W.C.; Drevensek-Olenik, I.; da Silva, M.W. Programmed self-assembly of a quadruplex DNA nanowire. *Chemistry*, **2014**, *20*, 3626-3630.
76. Guo, X.; Liu, S.; Yu, Z. Bimolecular quadruplexes and their transitions to higher-order molecular structures detected by ESI-FTICR-MS. *J. Am. Soc. Mass Spectrom*. **2007**, *18*, 1467-1476.
77. Lu, M.; Guo, Q.; Kallenbach, N.R. Structure and stability of sodium and potassium complexes of dT4G4 and dT4G4T. *Biochemistry* **1992**, *31*, 2455-2459.
78. Creze, C.; Rinaldi, B.; Haser, R.; Bouvet, P.; Gouet, P. Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li+ ions. *Acta Crystallogr. D Biol. Crystallogr*. **2007**, *63*, 682-688.
79. Deng, J.; Xiong, Y.; Sundaralingam, M. X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr(2+) ions at subatomic resolution (0.61 A). *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13665-13670.
80. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc*. **2004**, *126*, 16405-16415.
81. Risitano, A.; Fox, K.R. Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry* **2003**, *42*, 6507-6513.

82. Majerová, T.; Streckerová, T.; Bednárová, L.; Curtis, E.A. Sequence requirements of intrinsically fluorescent G-quadruplexes. *Biochemistry* **2018**, *57*, 4052-4062.
83. Kolesnikova, S.; Srb, P.; Vrzal, L.; Lawrence, M.; Veverka, V.; Curtis, E.A. GTP- dependent formation of multimeric G-quadruplexes. *Submitted* **2019**.
84. Blackburn, E.H.; Gall, J.G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol*. **1978**, *120*, 33-53.
85. Jacob, N.K.; Skopp, R.; Price, C.M. G-overhang dynamics at Tetrahymena telomeres. *EMBO J*. **2001**, *20*, 4299-4308.
86. Moyzis, R.K.; Buckingham, J.M.; Cram, L.S.; Dani, M.; Deaven, L.L.; Jones, M.D.; Meyne, J.; Ratliff, R.L.; Wu, J.R. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci*. USA **1988**, *85*, 6622-6626.
87. Wright, W.E.; Tesmer, V.M.; Huffman, K.E.; Levene, S.D.; Shay, J.W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev*. **1997**, *11*, 2801-2809.
88. Henderson, E.; Hardin, C.C.; Walk, S.K.; Tinoco Jr., I.; Blackburn, E.H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine- guanine base pairs. *Cell* **1987**, *51*, 899-908.
89. Neidle, S.; Parkinson, G.N. The structure of telomeric DNA. *Curr. Opin. Struct. Biol*. **2003**, *13*, 275-283.
90. Li, J.; Correia, J.J.; Wang, L.; Trent, J.O.; Chaires, J.B. Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.* **2005**, *33*, 4649-4659.
91. Petraccone, L. Higher-order quadruplex structures. *Top. Curr. Chem*. **2013**, *330*, 23-46.
92. Yu, H.Q.; Miyoshi, D.; Sugimoto, N. Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.* **2006**, *128*, 15461-15468.
93. Xu, Y.; Komiyama, M. The structural studies of human telomeric DNA using AFM. *Nucleic Acids Symp. Ser. (Oxf).* **2007**, *51*, 241-242.
94. Xu, X.; Ishizuka, T.; Kurabayashi, K.; Komiyama, M. Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem. Int. Ed.* **2009**, *48*, 7833-7836.
95. Petraccone, L.; Spink, C.; Trent, J.O.; Garbett, N.C.; Mekmaysy, C.S.; Giancola, C.; Chaires, J.B. Structure and stability of higher-order human telomeric quadruplexes. *J. Am. Chem. Soc*. **2011**, *133*, 20951-20961.
96. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H.J.; Pluckthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci*. USA **2001**, *98*, 8572-8577.
97. Moye, A.L.; Porter, K.C.; Cohen, S.B.; Phan, T.; Zyner, K.G.; Sasaki, N.; Lovrecz, G.O.; Beck, J.L.; Bryan, T.M. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun*. **2015**, *6*, 7643.
98. Zhao, C.; Wu, L.; Ren, J.; Xu, Y.; Qu, X. Targeting human telomeric higher-order DNA: dimeric G-quadruplex units serve as preferred binding site. *J. Am. Chem. Soc.* **2013**, *135*, 18786-18789.
99. Huang, X.X.; Zhu, L.N.; Wu, B.; Huo, Y.F.; Duan, N.N.; Kong, D.M. Two cationic porphyrin isomers showing different multimeric G-quadruplex recognition specificity against monomeric G-quadruplexes. *Nucleic Acids Res*. **2014**, *42*, 8719-8731.
100. Zhang, Q.; Liu, Y.C.; Kong, D.M.; Guo, D.S. Tetraphenylethene derivatives with different numbers of positively charged side arms have different multimeric G-quadruplex recognition specificity. *Chemistry* **2015**, *21*, 13253-13260.
101. Dai, J.; Dexheimer, T.S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R.A.; Yang, D. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **2006**, *128*, 1096-1098.
102. González, V.; Hurley, L.H. The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol*. **2008**, *50*, 111-129.
103. Mathad, R.I.; Hatzakis, E.; Dai, J.; Yang, D. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res*. **2011**, *39*, 9023-9033.
104. Kerkour, A.; Marquevielle, J.; Ivashchenko, S.; Yatsunyk, L.A.; Mergny, J.L.; Salgado, G.F. High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem*. **2017**, *292*, 8082-8091.

105. Matsugami, A.; Okuizumi, T.; Uesugi, S.; Katahira, M. Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.* **2003**, *278*, 28147-28153.
106. Palumbo, S.L.; Memmott, R.M.; Uribe, D.J.; Krotova-Khan, Y.; Hurley, J.H.; Ebbinghaus, S.W. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. (2008) *Nucleic Acids Res.* **2008**, *36*, 1755-1769.
107. Broxson, C.; Beckett, J.; Tornaletti, S. Transcription arrest by a G-quadruplex forming-trinucleotide repeat sequence from the human c-myb gene. *Biochemistry* **2011**, *50*, 4162-4172.
108. Hirsch, M.R.; Gaugler, L.; Deagostini-Bazin, H.; Bally-Cuif, L.; Goridis, C. Identification of positive and negative regulatory elements governing cell-type-specific expression of the neural cell adhesion molecule gene. *Mol. Cell. Biol.* **1990**, *10*, 1959-1968.
109. Chamboredon, S.; Briggs, J.; Vial, E.; Hurault, J.; Galvagni, F.; Oliviero, S.; Bos, T.; Castellazzi, M. v-Jun downregulates the SPARC target gene by binding to the proximal promoter indirectly through Sp1/3. *Oncogene* **2003**, *22*, 4047-4061.
110. Morgan, R.K.; Batra, H.; Gaerig, V.C.; Hockings, J.; Brooks, T.A. Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator. *Biochim. Biophys. Acta* **2016**, *1859*, 235-245.
111. Islam, I.; Baba, Y.; Witarto, A.B.; Yoshida, W. G-quadruplex-forming GGA repeat region functions as a negative regulator of the Ccnb1ip1 enhancer. *Biosci. Biotechnol. Biochem.* **2019**, *7*, 1-6.
112. Palumbo, S.L.; Ebbinghaus, S.W.; Hurley, L.H. 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.* **2009**, *131*, 10878-10891.
113. Kang, H.J.; Cui, Y.; Yin, H.; Scheid, A.; Hendricks, W.P.D.; Schmidt, J.; Sekulic, A.; Kong, D.; Trent, J.M.; Gokhale, V.; Mao, H.; Hurley, L.H. A pharmacological chaperone molecule induces cancer cell death by restoring tertiary DNA structures in mutant hTERT promoters. *J. Am. Chem. Soc.* **2016**, *138*, 13673-13692.
114. Micheli, E.; Martufi, M.; Cacchione, S.; De Santis, P.; Savino, M. Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives. *Biophys. Chem.* **2010**, *153*, 43-53.
115. Saha, D.; Singh, A.; Hussain, T.; Srivastava, V.; Sengupta, S.; Kar, A.; Dhapola, P.; Dhople, V.; Ummanni, R.; Chowdhury, S. Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J. Biol. Chem.* **2017**, *292*, 15205-15215.
116. Yafe, A.; Etzioni, S.; Weisman-Shomer, P.; Fry, M. Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res.* **2005**, *33*, 2887-2900.
117. Xu, Y.; Kimura, T.; Komiyama, M. Human telomere RNA and DNA form an intermolecular G-quadruplex. *Nucleic Acids Symp. Ser. (Oxf.)* **2008**, *52*, 169-170.
118. Wanrooij, P.H.; Uhler, J.P.; Shi, Y.; Westerlund, F.; Falkenberg, M.; Gustafsson, C.M. A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* **2012**, *40*, 10334-10344.
119. Zheng, K.W.; Wu, R.Y.; He, Y.D.; Xiao, S.; Zhang, J.Y.; Liu, J.Q.; Hao, Y.H.; Tan,Z. A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.* **2014**, *42*, 10832-10844.
120. Zheng, K.W.; Xiao, S.; Liu, J.Q.; Zhang, J.Y.; Hao, Y.H.; Tan, Z. Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* **2013**, *41*, 5533-5541.
121. Xiao, S.; Zhang, J.Y.; Wu, J.; Wu, R.Y.; Xia, Y.; Zheng, K.W.; Hao, Y.H.; Zhou, X.; Tan, Z. Formation of DNA:RNA hybrid G-quadruplexes of two G-quartet layers in transcription: expansion of the prevalence and diversity of G-quadruplexes in genomes. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 13110-13114.
122. Xiao, S.; Zhang, J.Y.; Zheng, K.W.; Hao, Y.H.; Tan, Z. Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals. *Nucleic Acids Res.* **2013**, *41*, 10379-10390.
123. Zhang, J.Y.; Zheng, K.W.; Xiao, S.; Hao, Y.H.; Tan, Z. Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.* **2014**, *136*, 1381-1390.

124. Zhao, Y.; Zhang, J.Y.; Zhang, Z.Y.; Tong, T.J.; Hao, Y.H.; Tan, Z. Real-time detection reveals responsive cotranscriptional formation of persistent intramolecular DNA and intermolecular DNA:RNA hybrid G-quadruplexes stabilized by R-loop. *Anal Chem.* **2017**, *89*, 6036-6042.
125. Shrestha, P.; Xiao, S.; Dhakal, S.; Tan, Z.; Mao, H. Nascent RNA transcripts facilitate the formation of G-quadruplexes. *Nucleic Acids Res.* **2014**, *42*, 7236-7246.
126. Curtis, E.A.; Liu, D.R. Discovery of widespread GTP-binding motifs in genomic RNA and DNA. *Chem. Biol.* **2013**, *20*, 521-532.
127. Švehlová, K.; Lawrence, M.S.; Bednárová, L.; Curtis, E.A. Altered biochemical specificity of G-quadruplexes with mutated tetrads. *Nucleic Acids Res.* **2016**, *44*, 10789-10803.