HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2018 July 07.

Published in final edited form as:

J Mol Biol. 2017 July 07; 429(14): 2127–2147. doi:10.1016/j.jmb.2017.05.017.

RNA G-quadruplexes in biology: principles and molecular mechanisms

Marta M. Fay[1,2], **Shawn M. Lyons**[1,2], and **Pavel Ivanov**[1,2,3,*]

[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA, 02115

[2]Department of Medicine, Harvard Medical School, Boston, MA, 02115

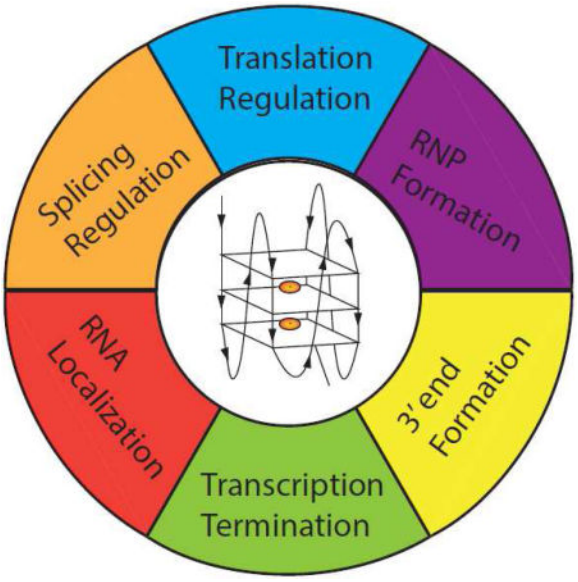
[3]The Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA

Abstract

G-quadruplexes are extremely stable DNA or RNA secondary structures formed by sequences rich in guanine. These structures are implicated in many essential cellular processes and the number of biological functions attributed to them continues to grow. While DNA G-quadruplexes are well understood on structural and to some extent on functional levels, RNA G-quadruplexes and their functions have received less attention. The presence of *bona fide* RNA G-quadruplexes in cells has long been a matter of debate. The development of G-quadruplex-specific antibodies and ligands hinted on their presence *in vivo* but recent advances in RNA sequencing coupled with chemical footprinting suggested the opposite. In this review, we will critically discuss the biology of RNA G-quadruplexes focusing on the molecular mechanisms underlying their proposed functions.

Graphical abstract

[*]To whom correspondence should be addressed. Pavel Ivanov: Tel: 1-617-525-1233; Fax: 1-617-525-1310; pivanov@rics.bwh.harvard.edu.



Introduction

It was first reported in 1910 that high concentrations of guanylic acid form a gel [1]. 50 years later, the G-tetrad (also known as G-quartet) structure, derived from X-ray fiber diffraction studies, was postulated to be the basis for the aggregation of 3′- or 5′-guanosine monophosphate into such gelatinous substance [2]. Similar aggregation was also observed with the first chemically synthesized deoxyguanosine homopolymers[3]. In comparison to well-known Watson-Crick base pairing in nucleic acids (Figure 1A), G-tetrads are formed when guanines are organized into planar quartets where each base is connected to two other bases via Hoogsteen base pairing (Figure 1A–B). Hydrogen bonds between each pair of guanines involves four donor/acceptor atoms, the N1, N7, N2 and O6 atoms, such that a G-quartet has eight total hydrogen bonds (four N2-H…N7 and four N1-H…O6 bonds). Four carbonyl oxygen (O6) atoms form a negatively charged core in the center of G-quartet. When three or more G-quartets stack onto one another, they form a stable right-handed helical structure known as a G-quadruplex (G4) (Figure 1B). In such vertical stacking, individual G-quartets are separated by ≈3.3 Å [4]. Metal ions such as monovalent cations can intercalate into the central anionic core of a G-tetrad (Figure 1B) or bind between two G-quartets to coordinate, stabilize hydrogen-bonded tetrads and enhance base-stacking interactions [5; 6] (Figure 1B). As the central channel has defined geometry and size, only cations with an adequate charge, size and dehydration energy are expected to coordinate a G4 [7; 8]. As a general principle, when the ionic radius fits in the middle of a G-quartet, it stabilizes G4 structures [7; 9; 10; 11; 12; 13; 14; 15; 16]. In particular, physiologically relevant monovalent cations such as K^+, Na^+ and NH_4^+ play important roles in G4 stabilization and to different extents (Figure 1C). While relatively small Na^+ can be embedded into the middle of a single G-tetrad, larger K^+ and NH_4^+ cations can fit into the space between two tetrads. As a consequence, larger cations coordinate eight oxygen atoms in G4s while smaller ones coordinate only four, thus contributing differently to G4 stability. In contrast, cations with very small ionic radius such as Li^+ do not favor G4 formation [8]. Within G4s, G-quartets may

be connected by loops of divergent sequence, structure and length (reviewed in [12; 17]). Potential G4 motif is commonly described as $\mathbf{G_X\text{-}N_{1–7}\text{-}G_X\text{-}N_{1–7}\text{-}G_XN_{1–7}\text{-}G_X}$, where *x* is 3–6 and *N* corresponds to any nucleotide (A, G, T, C or U) (Figure 1D).

Similarities and differences between RNA and DNA G-quadruplexes

The central building block of any quadruplex nucleic acids (DNA, RNA) or their analogs (locked nucleic acid, peptide nucleic acid) is the G-quartet (Figure 1B) [12; 17; 18; 19; 20; 21]. However, an assumption that RNA G4s are DNA counterparts is oversimplified. While DNA mostly exists in a double-stranded conformation, RNA is single-stranded or adopts various secondary structures (hairpins, loops, bulges, pseudoknots, etc.). Moreover, at least in eukaryotes, DNA is located in the nucleus bound to histones and auxiliary factors, while RNA is found in both nuclear and cytoplasmic compartments with a great diversity in protein-binding partners.

Of course, the most fundamental differences between RNA and DNA G4s are the same differences as between RNA and DNA, the presence of uracil instead of thymine and of a ribose sugar instead of a deoxyribose sugar (Figure 1E). The presence of a 2′-hydroxyl group in the ribose sugar has several consequences. First, it allows more intramolecular interactions within RNA G4s, leading to enhanced stability. Second, the ordered 2′-hydroxyls within RNA G4 grooves are favored to bring water molecules, leading to a more stable structure compared to DNA G4s [22]. Third, it has additional steric constraints on the G4 topology where 2′-hydroxyl groups prevent orientation of the base in the *syn*-conformation, instead strongly favoring the *anti*-conformation (via restrains on the glycosidic torsion angle), and imposition of additional constraints on sugar puckering (the ribose having a preference for C3′-*endo* puckering). As a consequence, topology of RNA G4s is limited to the parallel conformation where all four strands are oriented in the same direction (Figure 1B). In contrast, DNA G4s can adapt parallel, antiparallel or mixed conformations (Figure 1B). Although all known naturally occurring RNA G4s adopt a parallel topology, an artificial RNA aptamer *Spinach* demonstrated an unexpected antiparallel conformation[23], suggesting that the repertoire of G4 topologies may potentially be expanded. The extensive overview of the various RNA G4 structures based on high-resolution studies and structural modeling as well as current challenges in their structural investigations is beyond the scope of our manuscript, but is available elsewhere [12].

The impact of uracil versus thymine residues on G4 stability is not well studied. Substitution of thymine with uracil (thus omission/removal of methyl groups) in the loop of G4 oligos makes G4s more stable and less hydrated within the loop around the grooves [24; 25]. Indeed, the composition and length of loops strongly affects the stability of both DNA and RNA G4s where shorter loops provide greater stability [26] (Figure 1E). In one study, the importance of flanking versus the central G-quartet on the stability of DNA or RNA G4 derived from telomeric region was addressed [27]. Biophysical and mutagenesis studies indicated that each G-quartet contributes equally to RNA G4 stability while the central quartet is of greater importance for DNA G4 stability [27], in agreement with previous observations [28; 29].

Another important parameter is that DNA and RNA G4s show distinct differences in cation binding specificity[7]. Extensive analysis of G4 formation and stability in the presence of various monovalent and divalent cations (Li^+, K^+, Na^+, NH_4^+, Rb^+, Cs^+, Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}) was recently performed on two of the most studies model RNA G4s - *TERRA* (telomeric repeat-containing RNA) and *NRAS* oligo (an 18 nt sequence derived from the *NRas* mRNA) and their DNA G4 counterparts [7]. While K^+ dramatically stabilizes both DNA and RNA G4s, Na^+ only had a strong effect on DNA but not RNA G4s. In agreement with loops having a stabilizing role on G4 structures, the *NRAS* G4 oligo with the shorter loops was more stable than the *TERRA* G4 with the longer loops in the presence of K^+ [7]. For divalent ions, only Sr^{2+} increases stability of the RNA G4s, although Sr^{2+} ions have no known biological function. In contrast, the biologically relevant divalent cations Mg^{2+} and Ca^{2+} had no effect on RNA G4 stability. However, the data also may suggest that other divalent cations, such as harmful heavy metals (Pb^{2+}, Cd^{2+} or Hg^{2+}), may bind and disturb G4 structures contributing to their toxicity; such binding is reported for guanine-rich oligos [30; 31].

Thus despite numerous similarities, RNA G4s are more compact, less hydrated and often more thermodynamically stable than DNA G4s [11; 32] (Figure 1E). It is important to keep in mind when discussing the relevance of RNA G4 formation/presence on biology: the majority of studies were done *in vitro* with chemically synthesized oligonucleotides.

How to identify RNA G4: bioinformatic predictions

Despite many years of *in vitro* research, experimental evidences for G4s in cells are limited or even controversial. Biologically relevant G4s were first postulated to be present in eukaryotic chromosomal telomeric DNA, specific DNA sequences at chromosome ends [33; 34] and then it was observed that telomeric regions of eukaryotic chromosomes could adapt a compact four-stranded structural arrangement[35; 36]. In vertebrates, telomeric sequences vary in length (up to 70kb) and are composed of repetitive hexanucleotide repeats (5′-TTAGGG) that largely form canonical DNA duplexes with the complimentary sequence repeats (3′-AATCCC). However, the 3'-terminal region remains single stranded for approximately 15 – 200 nucleotiddes due to the end replication problem. Similar findings were also recorded for guanine-rich sequences in immunoglobulin switch regions[37]. Authors assumed that these structures are held together by guanines bonded by Hoogsteen pairing; this assumption was experimentally proven *in vitro* using short DNA oligonucleotides derived from these regions [37].

Subsequent to the discovery of the first putative G4 (pG4) motifs in telomeric DNA sequences, several reports noted that the promoter regions of many genes have G-rich sequences that are potentially capable of forming G4s. Such pG4s have since been validated *in vitro* using DNA oligonucleotides derived from these sequences [38; 39; 40; 41; 42]. The first validated DNA G4 was found in the promoter of the oncogene *c*-MYC [41], which was followed by attempts to transcriptionally downregulate *c*-MYC expression using small molecules to modulate the *c*-MYC G4 [43]. Subsequent structure/functional analyses have assessed G4 structures in promoters of other oncogenes including *c-KIT* and *KRAS* [44; 45; 46; 47]. Later, multiple rep

Upon sequencing of the human genome, a systematic computational approach using the simple algorithm (*Quadparser*) mined the genome for the potential G4 motif $G_XN_{1–7}$ $G_XN_{1–7}$ $G_XN_{1–7}$ G_X (Figure 1D). This and similar analyses identified pG4s in the human genome (over 375,000 potential G4 sites) as well as other genomes [38; 42]. pG4s were found in promoters, coding regions, introns and untranslated regions (UTRs) of genes as well as intergenic regions. Various genome-wide validation strategies including G4-specific probes and sensors (e.g. G4 antibodies or small molecules as pyridostatin (PDS) [48]) coupled to *in vitro* biophysical analysis have indicated the existence of numerous pG4s structures in the human genome[49; 50; 51; 52; 53].

An important finding from the DNA G4 genome-wide studies is the presence of pG4s in the protein-coding (open reading frames (ORFs)) and non-coding (e.g. introns, 5′- and 3′-UTRs) regions of the transcriptome. Given that transcribed mRNAs are single-stranded, and RNA G4s are generally more stable than their DNA counterparts (*vide supra*), it was proposed that mRNAs contain G4s [39; 46]. Computational analysis suggested that pG4s are overrepresented in 5′- and 3′-UTRs [39], regions critical for post-transcriptional regulation. Several 5′-UTR- and 3′-UTR-derived G4 sequences have been validated *in vitro*, and multiple potential functions proposed (*vide infra* and reviewed in [54; 55]).

While bioinformatics predictions were instrumental in identifying pG4s in cellular RNAs, both false positive and false negative results have been reported. For example, RNAs that included loops greater than 7 nucleotides (up to 15 nucleotides) have been reported to fold into stable G4s *in vitro* [26]. Similarly, several pG4s in 5′-UTRs matching all algorithm prerequisites were unable to fold into G4s [56]. In this case, pG4 flanking sequences (within 15 nucleotides either 5′- or 3′- of pG4) such as C-rich tracks impaired G4 folding by interacting with key guanines via Watson-Crick base pairing to promote alternative secondary structures. Substitutions of cytosines for adenines in such non-folding pG4s were able to promote G4 formation [56]. Thus, caution should be used with G4 computational analysis predictions.

In vitro approaches to identify RNA G4

Assessing synthetic RNA oligonucleotides derived from pG4s by structural and biophysical techniques remains the leading methodology to characterize RNA G4s (discussed in details in [57]). Circular dichroism (CD) [58], ultraviolet (UV) melting curves [59] and nuclear magnetic resonance (NMR) spectroscopy [60] are most commonly used to compare biophysical signals of a pG4 in the presence of G4-stabilizing and destabilizing ions. RNA G4s adopt parallel G4 topology and display a characteristic CD spectrum of a negative signal at 240 nm and a positive signal at 265 nm [58; 61]. Although convenient for the initial characterization, the CD signature of RNA G4s is very similar to that of an A-form helix RNA (often found in stem-loop structures) and may be misleading [58]. UV measurements rely on a unique G4 hypochromic signature at 295 nm (contrasting to the absorbance peak for nucleic acids at 260 nm) and used to monitor pG4 stability by determination of the G4 melting temperature in G4-stabilizing or destabilizing buffer conditions [59]. Although a time-consuming technique, one-dimensional NMR is often used for determining G4 structures in short oligonucleotides [61; 62]. The NMR spectra of RNA molecules (signature that appears from

the imino protons of guanine and uracil) display typical signals at 10–15 ppm (parts per million). When these protons are engaged in hydrogen bonding (via Watson-Crick or Hoogsteen base pairing), they do not exchange with the solvent, become protected and are visible in the NMR spectra. Since the resonance frequency of the iminos in Watson-Crick base pairing at 12–15 ppm is clearly different from the iminos in G4s (10–12 ppm), NMR is able to distinguish G4s from helix structures, especially by comparing spectra in the presence of G4-stabilizing and destabilizing cations [60; 62; 63]. In addition to biophysical approaches, other methods are often used to study G4s. They include functional approaches such as G4-specific ligand binding, site-specific mutagenesis of G4 sequences, usage of RNA reporters and biochemical approaches to determine secondary structure.

Hundreds of small molecules have been described to bind DNA and RNA G4s [64]. Such ligands are diverse chemically and structurally, and their specificity towards G4s is often low, and should be used with precautions. Nonetheless, several ligands are commonly used in functional assays where their addition may affect a functional outcome via stabilization or destabilization of G4s (such as splicing or translation efficiency of a reporter mRNA). Reporter assays are based on the inclusion of a candidate G4 into a functional cassette that monitors splicing (inclusion of alternatively spliced exons) [65; 66; 67; 68] or translation (luciferase production) [56; 69; 70; 71] reactions *in vitro* and/or *in vivo*. Such approaches are often paralleled by site-directed mutagenesis of key guanines involved in the G4 and then wild type and mutated forms are compared in functional assays. Although the reporter results may suggest that the G4 is functional, the results should be analyzed with caution since introduction of sequences into a functional cassette may change structure of mRNA, affect binding of protein factors or introduce cryptic regulatory elements.

Biochemical methods of RNA G4 characterization employ chemical molecules or enzymes to determine the secondary structure of RNA in the presence or the absence of G4-permissive or non-permissive buffer conditions. Chemical probing uses small molecules that report RNA structural footprints to accurately map secondary structures. Nucleobase-specific (such as dimethyl sulfate (DMS) and 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate (CMCT)) and ribose-specific (selective 2′-hydroxyl acylation) probes are commonly used *in vitro* and *in vivo* to characterize RNA secondary structures [72; 73; 74; 75; 76; 77; 78]. Chemical footprinting has been used to determine the presence of G4s in both single RNAs and transcriptome-wide studies [73; 79; 80]. Enzymatic probing is often used for RNA secondary structure determination; it uses ribonucleases that selectively target single-stranded or double-stranded regions and has been used in RNA G4 studies [68; 81]. Unfortunately, ribonucleases do not efficiently cross cell membranes an therefore are not used for *in vivo* studies.

Finally, a new strategy was recently developed to study RNA G4s *in vitro* [57; 68]. This method is based on the different hydrogen-bonding patterns of stem-loop structures and G4s, as the former requires only Watson-Crick base pairing and the latter relies on Hoogsteen base pairing. Nitrogen at position 7 of guanine (N7, see Figure 1A–B) is absolutely required for Hoogsteen base pairing, and its substitution to carbon (as in 7-deazaguanine) prevents G4 formation [82; 83]. By assessing the differences between native RNA and RNA with guanines substituted with 7-deazaguanines, it is possible to conclude whether a pG4 actually forms a

bona fide G4. One advantage of this method is that 7-deazaguanine-substituted RNAs can be used under functional conditions (e.g., in cellular extracts or in transfection experiments) and does not rely on artificial buffer conditions (e.g., in the presence of non-physiological levels of Li^+) [68].

A major problem with most of the above-described methods is that they rely on *in silico* analysis and *in vitro* experimental approaches in quasi-physiological conditions. Are there evidences for RNA G4 formation in cells?

What is the evidence for *in vivo* formation of RNA G-quadruplexes?

RNA G4s are extremely stable with *in vitro* melting temperatures exceeding the physiological range. The typical intracellular concentration of free metal cations is sufficient to support G4 formation as assessed under *in vitro* conditions, 5–15 mM Na^+, 140 mM K^+, 0.5–2 mM Mg^{2+}, pH 7.2 [84]. However, the evidence to support RNA G4s *in vivo* are weak. Earlier studies that addressed this question included immunofluorescence using G4-specific antibodies which revealed detectable ribonuclease-sensitive cytoplasmic and nuclear staining [49]. Upon treatment of cells with G4-stabilizing ligands, the intensity of the G4 signal increased in accordance with the presence of RNA G4s in cells [49]. While immuno-based approaches are useful, they have significant limitations, including fixation and permeabilization, allowing for the possibility that G4 folding could occur during these steps. Additionally, G4-stabilizing ligands and antibodies may shift the balance from the naturally existing RNA populations of mixed forms toward otherwise unstable G4-enriched forms. Moreover, the specificity of some antibodies (e.g. monoclonal antibody 1H6) towards G4 structures is questionable due to cross-reactivity to other sequence motifs [85].

With the development of high-throughput RNA sequencing (RNA-seq) and chemical probing methods, experimental approaches to systematically map RNA G4s throughout the entire transcriptome have become available [79; 80]. These methods exploit reverse transcriptase stalling or stops (RTS) caused by the presence of strong secondary structures such as G4s. The Balasubramanian lab developed RNA G-quadruplex sequencing or rG4-seq which monitors RTS sites coupled with next generation sequencing in the presence of K^+ cations or the G4-specific ligand PDS and K^+ (K^+/PDS) [80]. Parallel analysis was done in the presence of G4 destabilizing Li^+ cations and serves as a control. By applying rG4-seq *in vitro* to profile RNA G4s in the polyadenylated (poly(A)) fraction of RNA (covering over 17,000 transcripts), it was determined that about 3,500 (distributed in approximately 2,500 genes) and 11,500 RTS sites (distributed in approximately 6,000 genes) in K^+ and K^+/PDS conditions can be assigned as RNA G4s, respectively. Selected transcript candidates were validated *in vitro* by chemical probing using 2′-hydroxyl acylation [73]. Importantly, as this transcriptome-wide analysis used purified cellular RNA, it does not truly convey the structural footprint of transcripts in living cells.

The Bartel lab developed a method that combines next generation sequencing, sensitivity to selected cations (K^+, Na^+ versus Li^+) and chemical probing techniques [79] including DMS modification (base modification) [86; 87] and 2′-hydroxyl acylation (ribose modification) [88; 89]. Poly(A) containing transcripts were analyzed *in vitro* and approximately

6,000–12,000 RTS sites were assigned as intramolecular RNA G4s. Most importantly, as chemical reagents used to probe secondary structures can penetrate living cells, similar analysis was also done *in vivo*. Unexpectedly, such analysis revealed that RNA G4s are overwhelmingly depleted in eukaryotic cells suggesting that machinery prevents G4 folding and/or actively monitors and unfolds existing G4s in eukaryotic cells. In support of this hypothesis, when model RNA G4s were overexpressed in eukaryotic cells, these G4s were also found unfolded. In contrast, when the same model RNA G4s were ectopically expressed in *Escherichia coli*, they were found in the G4 conformation. Further transcriptome analysis in three bacterial species revealed global depletion/absence of pG4s. Moreover, bacteria expressing model G4s demonstrated growth defects compared to wild type bacteria or strains expressing mutant variants of the same model G4s [79]. Together this study suggests that RNAs with the propensity to adopt a G4 conformation are largely absent in the bacterial transcriptome. In the eukaryotic genome, mRNAs with this propensity exist, but are actively disassembled or prevented from forming by an unidentified machinery.

While instrumental in our understanding of RNA G4s, transcriptome-wide studies also have technical limitations and can be misleading. For example, although high-throughput sequencing should generally be sensitive to detect RNAs that are expressed at extremely low levels, RNA-seq/chemical probing analyses require relatively high expression of transcripts to detect G4s. Further, in a particular mRNA, a G4 may only fold during a fraction of the mRNAs lifetime (e.g. during splicing or transport), yet still be biologically relevant. In these instances, the unfolded G4 mRNA will dilute the signal of the folded G4. Smaller G4-containing transcripts (such as microRNAs), which are shorter than the length of fragments selected for library construction and sequencing, would presumably be lost in such analysis. In addition, these methods were designed to preferentially detect intramolecular G4s thus missing potential intermolecular G4s. Both labs selected polyadenylated RNAs, missing other classes of RNAs. Taking into consideration that only ~2% of the human genome is protein coding [90], we expect G4s to be within introns of pre-mRNAs as well as non-protein coding transcripts (both nuclear and cytoplasmic RNAs) and to have biological functions.

Despite limitations, these studies have revealed some important findings. First, they have proven the existence of RNA G4s *in vitro*. Second, in both studies, a large fraction of identified G4s do not follow the predicted motif (Figure 1D) and therefore were missed by *in silico* discovery approaches. The identified RNA G4s demonstrate diverse variations in their structures especially in the loop regions, greatly expanding the possible RNA G4 structures. Analysis of the regions surrounding a RNA G4 suggest that flanking regions such as the presence of proximal C-rich tracks may compete with G4 folding consistent with previous *in vitro* analysis [56; 91]. Third, in accordance with bioinformatics predictions, pG4s are enriched within UTRs relative to the ORF regions. This may suggest important regulatory roles for G4s in UTRs and possible selection against their placement in ORFs where they can interfere with mRNA translation. Fourth, although only poly(A) transcripts were analyzed in these studies, the similar approaches could be used for non-poly(A) fraction of cellular RNAs.

Most importantly, it was recognized that in eukaryotic cells RNA G4s are globally unfolded *in vivo*, despite their ability to readily assemble *in vitro*. This is presumably due to cellular

machinery that actively monitors and unwinds G4s [79]. The identity of the RNA G4-bound proteome remains elusive; although many RNA-binding proteins (RBPs) bind G-rich RNA sequences (see below, and Table 1) and possibly function to prevent G4 folding. A recent paper by Benhalevy *et al*. suggests that CNBP/ZNF9 is one of such proteins that binds G-rich elements in mRNAs, most of which form G4s *in vitro* [92]. Functionally, CNBP increases target mRNA translation efficiency by resolving G4s and other stable structures on mRNAs. It is also tempting to speculate that other RBPs (such as heterogeneous nuclear ribonucleoproteins (hnRNPs)) may keep G4 unwound by binding single-stranded G-rich sequences co-transcriptionally in the nucleus, prior to the export of mRNA to the cytoplasm.

RNA G-quadruplex binding proteins

In general, many proteins have affinities for G4s, whether these proteins recognize G4 structural determinants rather than G-rich sequences remains largely uncharacterized and requires further experimentation. Additionally, many of these proteins have been characterized *in vitro* with little to no confirmation that these interactions occur in cells (Table 1). Yet an *in vivo* interaction between a RNA G4 and a binding protein is dependent on RNA folding into G4s in cells, and since there is currently no consensus whether, and to what extent, RNA G4s fold in cells this question remains largely unanswered. Thus, as the field progresses we expect to identify two groups of proteins – (1) those that bind and prevent RNA G4s from forming and (2) those that bind and stabilize RNA G4s. Additionally, we expect some degree of specificity between G4s and binding proteins as loops and surrounding sequences will impact binding. We expect that G4 structures can serve as topological marks that act as specific binding sites for regulatory or structural proteins (Figure 2A). G4s as extremely stable structures can act as barriers or kinetic traps for movement of proteins or protein machines (e.g. ribosomes) along RNA (Figure 2B). Finally, G4s can also disrupt adjacent or overlapping alternative secondary structures or sequence determinants thus serving to inhibit binding of other proteins (Figure 2C). Whatever their roles, G4s must either be folded or unfolded, and both processes likely require assistance of proteins, especially in cells (Figure 2D).

Moving forward it is important to employ stringent criteria to define G4 verses G-rich interactions. Transcriptome-wide approaches such as PAR-CLIP that identify the binding sites of cellular RBPs with nucleotide resolution can be used to identify *bona fide* targets of candidate G4-binding factors. Biochemical approaches using 7-deazaguanine-subsituted RNAs to compare with native RNAs can be used as a control for G4 binding specificity. Finally, with a few exceptions, structural information on the protein-RNA G4 interaction is lacking; such studies are required for in depth insight into the biological relevance of the interaction.

Proposed functions of RNA G-quadruplexes

Putative RNA G4s are widely distributed in coding and non-coding regions of pre-mRNAs and mRNAs such as introns, 5'- and 3'-UTRs. Their enrichment in mRNA regions with regulatory functions (5'- and 3'-UTRs) hints that RNA G4s exist to regulate RNA metabolism. Below we discuss proposed biological functions of RNA G4s, and it should be

noted that many of these functions are based on *in vitro* studies and the usage of G4-specific ligands.

Role of G4s in mRNA transcription and processing

The initial finding that up to 50% of human genes may contain pG4s in their promoter regions suggested a role for G4s in regulating gene expression. After the first evidence that promoter-associated G4s in the oncogene *c*-MYC affect transcription *in vivo* [41; 43; 119], genome-wide studies in yeast and human cells identified transcriptional changes in numerous pG4-containing genes upon treatment with the G4-specific ligand TMPyP4 [120; 121].

During transcription, the newly synthesized RNA can base pair with the complimentary template DNA strand to form a RNA:DNA hybrid. Together with displaced single-stranded DNA, this RNA:DNA hybrid forms a structure known as an R-loop (Figure 3A). When sequences involved in R-loop formation contain two or more neighboring guanines, they can potentially fold into an intermolecular RNA:DNA G4 (Figure 3B). Formation of such hybrid G4 requires as few as two tandem G-tracks (guanine-rich sequences) on the nascent RNA transcript and a non-template DNA strand instead of four or more G-tracks required for formation of canonical intramolecular G4s (Figure 3B). Bioinformatic analysis identified that putative hybrid G4s are enriched within regions downstream of transcription start sites and are found in >97% of human genes, with an average of 73 putative hybrid G4s per gene [122; 123]. Formation of hybrid G4s was confirmed using T7 RNA polymerase *in vitro* transcription [123; 124; 125], *in vitro* crosslinking, site-specific mutagenesis and transfection studies using luciferase reporters [123]. Results from these studies suggest that hybrid G4s inhibit transcription *in vitro* and represent *cis*–elements that are built into a gene and activated co-transcriptionally.

Hybrid G4s have also been implicated in transcription termination (Figure 3C). Several observations have highlighted pG4s as terminator sequences that cause RNA Polymerase II (Pol II) transcription to pause [126; 127]. R-loops formed behind elongating Pol II are especially prevalent over G-rich pause sites positioned downstream of poly(A) signals and are capable of G4 formation. This transcriptional termination mechanism relies on the helicase Senataxin (SETX) to resolve R-loops. SETX-mediated R-loop resolution allows the 5'-3' exonuclease Xrn2 access to the 3' cleavage poly(A) sites causing nascent RNA release, 3' cleavage product degradation and finally Pol II termination (Figure 3D). Upon SETX depletion, R-loops are stabilized downstream of poly(A) signals, preventing efficient pause-mediated termination [126; 127].

Transcriptional termination is also coupled with 3'-end polyadenylation (Figure 4). Addition of the poly(A) tail requires recognition of a defined poly(A) signal and downstream signaling motifs on the pre-mRNA by cleavage/polyadenylation machinery (Figure 4A). These include the hexanucleotide AAUAAA signal upstream of a CA dinucleotide where 3'-end cleavage occurs and a G/U rich sequence predicted to fold into G4s [128]. In mammals, canonical polyadenylation involves binding of multisubunit complexes (CFI, CFII, CstF and CPSF) to these signals (Figure 4B, **Left**) [129]. In instances where a G4 forms at polyadenylation signals, such as *TP53* mRNA, G4s interact with the splicing/

polyadenylation factor hnRNP H/F to regulate polyadenylation (Figure 4B, **Right**) [130; 131]. Under optimal conditions, in the absence of DNA damage, mRNAs lacking G4s at poly(A) signals are efficiently processed whereas efficient 3'-end processing is inhibited for *TP53* mRNA presumably due to the G4 thereby decreasing its expression. In response to DNA damage/genotoxic stress, there is global repression of mRNA 3'-end formation rates which contributes to decreased mRNA synthesis. Mechanistically, the essential polyadenylation factor CstF is sequestered in a non-active protein complex composed of Pol II, PARN, BARD1 and BRCA1 [132] (Figure 4C, **Left**). In parallel, 3'-end processing of *TP53* mRNA is up-regulated to increase expression of TP53 (also known as p53) [133]. This mechanism requires recognition of the G4 in the *TP53* pre-mRNA by hnRNP H/F causing efficient recruitment of CstF, 3'-end formation, and ultimately increased p53 expression under DNA damage (Figure 4C, **Right**). In turn, p53 regulates expression of downstream target genes encoding multiple stress-related and apoptosis-related proteins.

RNA G4s are implicated in splicing of pre-mRNAs (Figure 5). Genome-wide analysis of alternatively spliced transcripts found over 3 million pG4s that map to approximately 30,000 mammalian genes [134; 135]. Alternative splicing is mechanistically regulated by combinatory action of many RBPs with RNA elements that impact spliceosome assembly at neighboring splice sites, therefore G4s assembled in the vicinity of splice sites may directly impact binding of regulatory RBPs. This can be achieved by masking or steric overlap with regulatory elements, or recruitment of specific G4-binding proteins (Figure 2). RNA G4s are found associated with both exon splicing enhancers and intron splicing enhancers and silencers (Figure 5A–B). For example, two G4s are found within the FMRP-binding site (FBS) on its own pre-mRNA (*FMR1*), which give rise to different FMRP isoforms (major longer and minor shorter isoforms) (Figure 5B) [136]. The FBS was shown to be a potent exonic splicing enhancer in a minigene system and acts as a control element that regulates alternative splicing in response to intracellular levels of FMRP. Binding of the major FMRP isoform to FBS results in decreased synthesis of the major FMRP isoforms (carrying a complete exon 15) concomitant with an increase of minor isoforms. Mutations in the FBS that affect its ability to form a G4 (although maintaining a G-rich sequence) decrease FMRP binding, ablate exonic splicing enhancer activity and change the splicing pattern of *FMR1* pre-mRNA [136]. While G4s in the *FMR1* pre-mRNA act as exonic splicing enhancers, RNA G4s found in intron 6 of the human telomerase (hTERT) are proposed to serve as an intronic splicing silencer (Figure 5A). Targeting this structure with a G4-interacting agent (ligand 12459) stabilizes G4s and impairs *hTERT* splicing [137]. In *TP53* pre-mRNA, a G4 located in intron 3 stimulates splicing of intron 2, working as an intronic splicing enhancer and leading to differential expression of transcripts encoding distinct p53 isoforms (Figure 5A). Site-directed mutagenesis of residues involved in G4 formation decreased the excision of intron 2 (by ~30%) in a reporter splicing assay [66].

Although our understanding of how G4s impact pre-mRNA processing and maturation is limited, the above-discussed examples suggest that G4s may have roles both in global and transcript-specific regulation of mRNA synthesis.

Role of G4s in mRNA localization

Transport of mRNAs to subcellular compartments is a key post-transcriptional mechanism that allows localized protein synthesis. This mechanism is especially important in asymmetric cells such as neurons where transcribed mRNAs travel large distances to their sites of translation such as the tip of a growing axon or synapse (Figure 6). G4s belong to a list of emerging *cis*-elements participating in subcellular sorting and are most commonly found in 3'-UTRs (aka “zipcodes”). Analysis of known dendritic mRNAs identified G4s in approximately 30% of such mRNAs (11 out of 34 candidate transcripts). Using RNA structure probing and mutagenesis, Subramanian *et al.* showed that G4s are present in the 3'-UTRs of two mRNAs encoding important postsynaptic proteins PDS-95 (post-synaptic density protein 95; contains three G4s) and CaMKIIa (Ca^{2+}/calmodulin-dependent protein kinase II; contains one G4) [138]. Transfection of neurons with mRNA reporters containing these 3'-UTRs was sufficient for localization to neuronal processes while mutating the predicted G4 region in these reporters failed to localize. This supports the idea that G4s are neurite-targeting elements that contribute to the transport of *CaMKIIa* and *PDS-95* mRNAs in neurites *in vivo* [138]. Following export from the nucleus (Figure 6A), these mRNAs may bind accessory proteins as previously discussed. It was hypothesized that these G4 binding proteins may link mRNAs to molecular motors such as kinesins (e.g. KIF5, kinesin heavy chain isoform 5) (Figure 6B). A proteomic screen identified a number of RBPs (including selected hnRNPs (e.g. hnRNP U), FUS/TLS and FMRP that binds G4 structures) as KIF5-binding partners [139]. Bound to kinesins, G4-associated mRNPs may be transported along neuronal microtubules to neurites for local mRNA translation (Figure 6C).

Role of G4s in mRNA translation

mRNA translation is a complex process that involves coordination between dozens of translation factors and the ribosome to produce proteins. Translation initiation is the rate-limiting and most regulated step of protein synthesis. It involves 5'-cap recognition by the eIF4F complex that consists of the cap-binding protein eIF4E, the scaffolding protein eIF4G and an RNA helicase eIF4A. Cap-bound eIF4F then associates with the multi-subunit eIF3 translation initiation complex and the 40S ribosomal subunit to form the pre-initiation complex (43S). The eIF2–GTP–$tRNA_i^{Met}$ ternary complex binds the 43S to scan through the 5'-UTR to identify the initiator AUG codon. Upon start codon recognition, the 60S ribosomal subunit joins the 40S to form translationally competent 80S ribosomes that proceed to the elongation step of translation.

The 5'-UTR is the key element for translation initiation and translational control. G4s are overrepresented in the 5'-UTRs of mRNAs implying important regulatory functions. As other thermodynamically stable RNA structures (such as hairpins) are known to compromise translation initiation [140], it is intuitive that extremely stable G4s would also inhibit translation. Indeed, when various naturally occurring G4s (e.g. found in *FMR1*, *Zic*-1, *NRas*, *MT3-MMP*, *Bcl-2*, *TRF2*, *ESR1* mRNAs) are placed in the 5'-UTRs of mRNA reporters, potent inhibition of translation is observed both *in vitro* and *in vivo* (reviewed in details in [141]) (Figure 7A). The efficiency of translation inhibition by G4s is not simply explained by steric constrains of the scanning pre-initiation complexes but may also strongly depends on the G4 location relative to the cap structure and/or initiation AUG codon. In two reports,

G4s located closer to these elements caused more robust translation inhibition [69; 142]. Whether this is universal for all G4s is unknown but there is a clear positive correlation between G4 stability and translation inhibition *in vitro* [143]. Moreover, different G4s and elements surrounding G4s may also influence translation, presumably by recruitment of different RNPs and/or by adapting alternative secondary structures. This also emphasizes the importance of studying G4s in their natural mRNA context to truly understand their biological consequences.

In addition to inhibiting translation, selected 5'-UTR-derived G4s have been shown to stimulate translation (Figure 7A). Two studies demonstrated that GC-rich regions within 5'-UTRs of *FGF2* (fibroblast growth factor 2) and *VEGF* (vascular endothelial growth factor) mRNAs contain structures that stimulate translation [144; 145]. Chemical and enzymatic footprinting using *in vitro* transcribed RNAs showed that these regions contain G4s. Deletion analysis using plasmid reporters coupled with transfection studies confirmed that translation stimulation required G4 sequences. Based on further site-directed mutagenesis, the authors concluded that these G4s were part of IRESes (Internal Ribosome Entry Site), structures that initiate protein synthesis by a non-canonical 5'-cap-independent manner. While the presence of *bona fide* IRESes in viral genomes is well characterized, whether cellular IRESes exist is a matter of debate [146]. Under oxidative stress, the 5'-UTR of *NRF2* mRNA (encoding the transcription factor NRF2, which regulates expression of antioxidant and detoxification genes) stimulates its translation in a G4-dependent manner [147]. Similarly, the G4 in the 5'-UTR of the *SNCA* (α-synuclein) mRNA enhances SNCA production under conditions when cap-dependent translation is attenuated [148]. The ability of specific RNA G4s to stimulate protein synthesis is not understood (Figure 7A).

Several reports suggest that RNA G4s within 3'-UTRs and ORFs also modulate translation (Figure 7A). In one study, several G4 motifs with loops greater than seven nucleotides located in 3'-UTRs both stimulated or inhibited translation without affecting mRNA transcription [149]. Canonical G4s in the 3'-UTRs of *PIM1* (encoding oncogene PIM1) [150] and *APP* (encoding Amyloid Precursor Protein)[151] mRNAs negatively regulate translation yet have no effect on mRNA stability. The molecular mechanisms that underlie this inhibition are unknown but may rely on G4-specific protein(s) that act(s) as translational repressors. For example, FMRP binds to a G4 in the *APP* mRNA ORF to inhibit its translation[152]. The relationship/relative contribution of 3'-UTR- and ORF-located G4s in regulating *APP* translation is unclear. Similarly, G4s within the ORF of the virally encoded EBNA1 transcript (the Epstein-Barr virus-encoded nuclear antigen 1) may hinder translation elongation by either promoting ribosomal pausing or ribosomal dissociation [70]. ORF-situated G4s are reported to act as roadblock for elongating ribosomes [153; 154]. RNA G4s also stimulate ribosomal frameshifting [155], the recoding event when ribosomes are forced to move backward or forward leading to alternative ORF on the same mRNA [156]. In contrast, poorly studied RGG domain-containing RBP Aven recognizes ORF-situated G4s in its mRNA targets to stimulate translation[157]. Aven cooperates with DHX36, a helicase that is required for the optimal translation of Aven-regulated mRNAs and is known to unwind RNA G4s in cells [110]. It is tempting to speculate that DHX36 stimulates translation of Aven mRNA targets through its G4 resolvase activities, removing roadblocks for elongating ribosomes [157] (Figure 7A).

Finally, RNA G4s may be directly involved in RAN translation, a non-canonical and poorly understood translation mechanism observed in some repeat expansion disorders. In this mechanism, expansion of short repetitive sequences (3–6 nucleotides) can provoke expression of proteins in all three reading frames without an initiating AUG codon [158]. RAN translation is reported in amyotrophic lateral sclerosis (ALS)/frontotemporal dementia (FTD) patients with intronic expansions of the hexameric repeats $(GGGGCC)_n$ in the *C9ORF72* gene. RAN translation of these intronic repeats produces di-peptide repeat proteins in all three frames *in vitro* and these same proteins have been detected in tissues and cells of ALS/FTD patients [159; 160]. RNA $(GGGGCC)_n$ repeats are capable of forming G4s [94; 114; 161; 162] and bioactive small molecules targeting these repeats significantly inhibit RAN translation and the production of dipeptide repeat proteins [162]. As different RNA G4s are able to interact with ribosomal proteins[93] and/or bind the 40S ribosomal subunit [163], we propose that *C9ORF72* hexameric RNA G4 repeats may directly recruit translationally-competent 48S-like pre-initiation complexes to initiate RAN translation generating di-peptides. We are currently testing this hypothesis (Figure 7B).

Emerging roles of RNA G4s in biology of non-coding RNAs

Recent bioinformatics analysis of human ncRNAs found a significant number of pG4s (approximately 750 unique transcripts) [164]. A subset of these pG4s were analyzed biophysically and found to fold into G4s *in vitro* [164]. While long ncRNAs (lncRNAs) are implicated in different biological processes and found to be dysregulated in many pathological conditions [165; 166], the biological significance of G4s in lncRNAs is still unclear. Recent work from Matsumura *et al.* demonstrated the biological relevance of a G4 in the lncRNA *GSEC* (G-quadruplex-forming sequence containing lncRNA) [167]. This cytoplasmic lncRNA is implicated in the development of colorectal cancer where it stimulates migration of colon cancer cells, and depletion of *GSEC* results in the reduction of cancer cell mobility. Immunoprecipitation and pull-down assays using *GSEC* RNA identified DHX36 as a binding partner. DHX36 recognizes a G4 in *GSEC*, and mutations of key guanines that affect G4 folding also reduce their interaction. Functionally, DHX36 inhibits mobility of colon cancer cells and *GSEC* antagonizes its effect by acting as a molecular decoy to suppress DHX36 function [167].

MicroRNAs (miRs) are 20–22 nucleotide small ncRNAs with roles in RNA metabolism, cellular homeostasis, development, growth and differentiation. Briefly, mature miRs are formed as a consequence of tightly regulated processing events in both the nucleus and cytoplasm. First, RNA Pol II transcribes primary transcripts (pri-miRs) in the nucleus that are recognized by the endoribonuclease Drosha, which coverts pri-miRs into smaller precursor RNAs (pre-miRs) that assemble into a hairpin structure. Upon nuclear export by the nuclear factor Exportin-5, pre-miR hairpins are further processed by the ribonuclease Dicer into miR duplexes. One of the duplex strands as a mature miR is effectively loaded into the RISC complex (RNA-induced silencing complex) where, in the complex with the protein Argonaute (Ago), regulates mRNA stability and translation. About 130 miR genes (~ 5% of all miR genes) are predicted to contain G4s in their promoter regions [168] and some G4 are found within miR transcripts [169; 170]. Since Dicer effectively recognizes canonical hairpin structures in pre-miRs, alternative structures such as G4s in pre-miRs should inhibit

their maturation into miRs. Indeed, two studies have suggested that this is the case. In the first study, pre-miR-92b was shown to exist in two conformations [169]. Using an *in vitro* Dicer-mediated cleavage assay, pre-miR-92b was processed into mature miR-92b from a canonical hairpin structure under conditions that disfavor G4 formation (e.g. presence of Li^+ ions). In contrast, conditions that favor G4 formation significantly inhibited pre-miR-92b processing *in vitro*. In another study, authors compared processing of the naturally occurring (pre-let-7e) and artificial synthetic (quad-pre-miR-27a) pre-miRs containing G4 structures into mature miRs [170]. In agreement with the previous work, G4s inhibit Dicer-mediated conversion of pre-miRs into mature forms. Both studies also demonstrated that pre-miRs can exist as G4s in living cells [169; 170], and treatment of cells with a G4-destabilizing ligand promotes production of mature miRs *in vivo* (up to 4-fold) [170]. These studies have uncovered biologically relevant RNA G4s that regulate the biogenesis and functions of miRs.

PERSPECTIVES, KEY QUESTIONS AND CHALLENGES

We have only begun to understand the biological significance and molecular mechanisms that underlie RNA G4 structures. Recent transcriptome-wide studies have demonstrated that RNA G4s are mostly unfolded *in vivo*. While this is generally unexpected in the light of extensive *in vitro* data, this may also suggest that RNA G4s can exist under certain conditions such as during specific stages of the the cell cycle, under specific stress stimuli (TP53 and DNA damage), disease conditions (C9ORF72 in ALS/FTD) or in a cell type-specific manner (neurite-specific G4-containing transcripts in neurons). Moreover, the intriguing finding that RNA G4s are disassembled in living cells suggest important and active roles of G4-binding proteins in RNA G4 biology. It re-emphasizes the importance of identifying and structure-function analysis of *bona fide* G4-interacting proteins. One intriguing aspect of RNA G4/G4-binding protein interaction is their potential functional convergence in many pathological conditions such as cancer and neurodegenerative diseases. For example, amplification of G4 RNA repeats in ALS/FTD (*C9ORF72* gene) and Fragile × syndrome (*FMR1* gene) is evident. In these same diseases, changes in expression levels or mutations can be found in the G4-binding proteins such as FUS/TLS, hnRNP A1/A2 and FMRP. This may indicate that G4-binding RBPs are uniquely important for normal neuron physiology; when mutated or in cells with RNA G4 expansions these proteins cannot exert their normal functions.

As evidence based on *in vitro* studies supports the involvement of G4s in versatile biological functions, understanding the molecular mechanisms of RNA G4-mediated regulation *in vivo* requires special attention. In contrast to DNA G4s that affect nuclear events and act as roadblocks for efficient transcription or inhibit telomere extension, RNA G4s may act in both the nucleus and the cytoplasm, exerting a wide range of affects on RNA metabolism. Thus, *in vivo* structural and functional studies of RNA G4s need to be developed. Recently, an approach called “in cell-NMR” was developed to study proteins and nucleic acids in their native physiological environment [171; 172]. This approach was used to study G4s of radiolabeled oligonucleotides injected into *Xenopus laevis* oocytes [173]. Another promising approach that we are currently developing in our lab is CRISPR/CAS9 genome editing to directly manipulate G4s in their natural genetic context, e.g. to introduce site-specific

mutations that disrupt G4 formation and/or addition of epitopes for detection of G4-regulated products. Such an approach has an advantage over the widely used reporter assays because it does not introduce additional perturbations in living cells such as high expression of candidate RNAs or expression in unnatural genomic context (such as plasmids).

The unique chemical properties of RNA G4s suggest it is possible to develop small molecules that select RNA rather than DNA G4s[174]. This is exemplified by the recent discovery of a small molecule, carboxypyridostatin, that exhibits high molecular specificity for RNA over DNA G4s[175]. Taking into consideration a growing number of diseases that implicate RNA G4s in their pathogenesis, this area is of great therapeutic importance and potential. Selective intervention targeting only RNA G4s may be a strategically important therapeutic approach for specific diseases. As thousands of DNA G4 ligands are reported, the same molecules can be screened for their interactions with RNA G4s to allow rational design of selective RNA G4 ligands. Similarly, therapeutic approaches developed to manipulate activities of *bona fide* G4-binding proteins will be promising.

From our perspective, RNA G4s are particularly interesting structures for future biophysical, chemical and functional investigations. RNA G4 biology will continue to develop as an exciting branch of experimental biology with unique perspectives in molecular medicine, and represents an attractive area for future research.

Acknowledgments

We thank members of the Ivanov and Anderson labs for helpful discussion and feedback on this manuscript. This work is supported by the National Institutes of Health [NS094918 to PI and F32GM119283 to SML].

References

1. Bang I. Examination or the guanyle acid. Biochemische Zeitschrift. 1910; 26:293–311.
2. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. Proc Natl Acad Sci U S A. 1962; 48:2013–8. [PubMed: 13947099]
3. Ralph RK, Connors WJ, Khorana HG. Secondary structure and aggregation in deoxyguanosine oligonucleotides. J. Amer. Chem. Soc. 1962; 84:2265–2266.
4. Forman SL, Fettinger JC, Pieraccini S, Gottarelli G, Davis JT. Toward Artificial Ion Channels: A Lipophilic G-Quadruplex. J. Am. Chem. Soc. 2000; 122:4060–4067.
5. Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CL, Miles HT, Becker ED. Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. J. Am. Chem. Soc. 1978; 100:3625–3627.
6. Pinnavaia TJ, Miles HT, Becker ED. Self-assembled 5'-guanosine monophosphate, nuclear magnetic resonance evidence for a regular, ordered structure and slow chemical exchange. J. Am. Chem. Soc. 1975; 97:7198–7200. [PubMed: 1184896]
7. Guiset Miserachs H, Donghi D, Borner R, Johannsen S, Sigel RK. Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. J Biol Inorg Chem. 2016; 21:975–986. [PubMed: 27704222]
8. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 2008; 36:5482–515. [PubMed: 18718931]
9. Halder K, Hartig JS. RNA quadruplexes. Met Ions Life Sci. 2011; 9:125–39. [PubMed: 22010270]
10. Hardin CC, Watson T, Corregan M, Bailey C. Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). Biochemistry. 1992; 31:833–41. [PubMed: 1731941]

11. Joachimi A, Benz A, Hartig JS. A comparison of DNA and RNA quadruplex structures and stabilities. Bioorg Med Chem. 2009; 17:6811–5. [PubMed: 19736017]
12. Malgowska M, Czajczynska K, Gudanis D, Tworak A, Gdaniec Z. Overview of the RNA G-quadruplex structures. Acta Biochim Pol. 2016
13. Miyoshi D, Nakao A, Toda T, Sugimoto N. Effect of divalent cations on antiparallel G-quartet structure of d(G4T4G4). FEBS Lett. 2001; 496:128–33. [PubMed: 11356196]
14. Venczel EA, Sen D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. Biochemistry. 1993; 32:6220–8. [PubMed: 8512932]
15. Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A. Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. J Phys Chem B. 2005; 109:3594–605. [PubMed: 16851398]
16. Wong A, Wu G. Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5'-monophosphate: a solid-state NMR study. J Am Chem Soc. 2003; 125:13895–905. [PubMed: 14599230]
17. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006; 34:5402–15. [PubMed: 17012276]
18. Ji X, Sun H, Zhou H, Xiang J, Tang Y, Zhao C. Research Progress of RNA Quadruplex. Oligonucleotides. 2011
19. Krishnan-Ghosh Y, Stephens E, Balasubramanian S. A PNA4 quadruplex. J Am Chem Soc. 2004; 126:5944–5. [PubMed: 15137741]
20. Randazzo A, Esposito V, Ohlenschlager O, Ramachandran R, Mayola L. NMR solution structure of a parallel LNA quadruplex. Nucleic Acids Res. 2004; 32:3083–92. [PubMed: 15181173]
21. Russo Krauss I, Parkinson GN, Merlino A, Mattia CA, Randazzo A, Novellino E, Mazzarella L, Sica F. A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an all-LNA G-quadruplex. Acta Crystallogr D Biol Crystallogr. 2014; 70:362–70. [PubMed: 24531470]
22. Zhang DH, Fujimoto T, Saxena S, Yu HQ, Miyoshi D, Sugimoto N. Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. Biochemistry. 2010; 49:4554–63. [PubMed: 20420470]
23. Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. Nat Chem Biol. 2014; 10:686–91. [PubMed: 24952597]
24. Olsen CM, Lee HT, Marky LA. Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. J Phys Chem B. 2009; 113:2587–95. [PubMed: 19014184]
25. Olsen CM, Marky LA. Energetic and hydration contributions of the removal of methyl groups from thymine to form uracil in G-quadruplexes. J Phys Chem B. 2009; 113:9–11. [PubMed: 19198041]
26. Pandey S, Agarwala P, Maiti S. Effect of loops and G-quartets on the stability of RNA G-quadruplexes. J Phys Chem B. 2013; 117:6896–905. [PubMed: 23683360]
27. Agarwala P, Kumar S, Pandey S, Maiti S. Human telomeric RNA G-quadruplex response to point mutation in the G-quartets. J Phys Chem B. 2015; 119:4617–27. [PubMed: 25763809]
28. Tomasko M, Vorlickova M, Sagi J. Substitution of adenine for guanine in the quadruplex-forming human telomere DNA sequence G(3)(T(2)AG(3))(3). Biochimie. 2009; 91:171–9. [PubMed: 18852018]
29. Lee JY, Kim DS. Dramatic effect of single-base mutation on the conformational dynamics of human telomeric G-quadruplex. Nucleic Acids Res. 2009; 37:3625–34. [PubMed: 19359361]
30. Liu W, Fu Y, Zheng B, Cheng S, Li W, Lau TC, Liang H. Kinetics and mechanism of conformational changes in a G-quadruplex of thrombin-binding aptamer induced by Pb2+ J Phys Chem B. 2011; 115:13051–6. [PubMed: 21950308]
31. Moriwaki H. Complexes of cadmium ion with guanine bases detected by electrospray ionization mass spectrometry. J Mass Spectrom. 2003; 38:321–7. [PubMed: 12644994]
32. Arora A, Maiti S. Differential biophysical behavior of human telomeric RNA and DNA quadruplex. J Phys Chem B. 2009; 113:10515–20. [PubMed: 19572668]

33. Blackburn EH. Structure and function of telomeres. Nature. 1991; 350:569–73. [PubMed: 1708110]
34. Cech TR. Beginning to understand the end of the chromosome. Cell. 2004; 116:273–9. [PubMed: 14744437]
35. Henderson E, Hardin CC, Walk SK, Tinoco I Jr, Blackburn EH. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. Cell. 1987; 51:899–908. [PubMed: 3690664]
36. Sundquist WI, Klug A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature. 1989; 342:825–9. [PubMed: 2601741]
37. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature. 1988; 334:364–6. [PubMed: 3393228]
38. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 2005; 33:2908–16. [PubMed: 15914667]
39. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res. 2008; 36:6260–8. [PubMed: 18832370]
40. Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, Ladame S, Balasubramanian S, Neidle S. Putative DNA quadruplex formation within the human c-kit oncogene. J Am Chem Soc. 2005; 127:10584–9. [PubMed: 16045346]
41. Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. 1998; 26:1167–72. [PubMed: 9469822]
42. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 2005; 33:2901–7. [PubMed: 15914666]
43. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. 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. 2002; 99:11593–8. [PubMed: 12195017]
44. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 2006; 34:2536–49. [PubMed: 16687659]
45. Hsu ST, Varnai P, Bugaut A, Reszka AP, Neidle S, Balasubramanian S. A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. J Am Chem Soc. 2009; 131:13399–409. [PubMed: 19705869]
46. Patel DJ, Phan AT, Kuryavyi V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res. 2007; 35:7429–55. [PubMed: 17913750]
47. Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. J Am Chem Soc. 2007; 129:4386–92. [PubMed: 17362008]
48. Rodriguez R, Muller S, Yeoman JA, Trentesaux C, Riou JF, Balasubramanian S. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J Am Chem Soc. 2008; 130:15758–9. [PubMed: 18975896]
49. Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem. 2014; 6:75–80. [PubMed: 24345950]
50. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013; 5:182–6. [PubMed: 23422559]
51. Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S. G-quadruplex structures mark human regulatory chromatin. Nat Genet. 2016; 48:1267–72. [PubMed: 27618450]
52. Isalan M, Patel SD, Balasubramanian S, Choo Y. Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. Biochemistry. 2001; 40:830–6. [PubMed: 11170401]
53. Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun. 2013; 4:1796. [PubMed: 23653208]
54. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015; 43:8627–37. [PubMed: 26350216]

55. Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett. 2015; 589:1653–68. [PubMed: 25979174]
56. Beaudoin JD, Perreault JP. 5'-UTR G-quadruplex structures acting as translational repressors. Nucleic Acids Res. 2010; 38:7022–36. [PubMed: 20571090]
57. Weldon C, Eperon IC, Dominguez C. Do we know whether potential G-quadruplexes actually form in long functional RNA molecules? Biochem Soc Trans. 2016; 44:1761–1768. [PubMed: 27913687]
58. Vorlickova M, Kejnovska I, Sagi J, Renciuk D, Bednarova K, Motlova J, Kypr J. Circular dichroism and guanine quadruplexes. Methods. 2012; 57:64–75. [PubMed: 22450044]
59. Mergny JL, Lacroix L. UV Melting of G-Quadruplexes. Curr Protoc Nucleic Acid Chem. 2009 Chapter 17, Unit 17 1.
60. Adrian M, Heddi B, Phan AT. NMR spectroscopy of G-quadruplexes. Methods. 2012; 57:11–24. [PubMed: 22633887]
61. Kypr J, Kejnovska I, Renciuk D, Vorlickova M. Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res. 2009; 37:1713–25. [PubMed: 19190094]
62. Malgowska M, Gudanis D, Kierzek R, Wyszko E, Gabelica V, Gdaniec Z. Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. Nucleic Acids Res. 2014; 42:10196–207. [PubMed: 25081212]
63. Yang D, Okamoto K. Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med Chem. 2010; 2:619–46. [PubMed: 20563318]
64. Li Q, Xiang JF, Yang QF, Sun HX, Guan AJ, Tang YL. G4LDB: a database for discovering and studying G-quadruplex ligands. Nucleic Acids Res. 2013; 41:D1115–23. [PubMed: 23161677]
65. Kralovicova J, Lages A, Patel A, Dhir A, Buratti E, Searle M, Vorechovsky I. Optimal antisense target reducing INS intron 1 retention is adjacent to a parallel G quadruplex. Nucleic Acids Res. 2014; 42:8161–73. [PubMed: 24944197]
66. Marcel V, Tran PL, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou MP, Hall J, Mergny JL, Hainaut P, Van Dyck E. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. Carcinogenesis. 2011; 32:271–8. [PubMed: 21112961]
67. Ribeiro MM, Teixeira GS, Martins L, Marques MR, de Souza AP, Line SR. G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. Hum Genet. 2015; 134:37–44. [PubMed: 25204874]
68. Weldon C, Behm-Ansmant I, Hurley LH, Burley GA, Branlant C, Eperon IC, Dominguez C. Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. Nat Chem Biol. 2016
69. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol. 2007; 3:218–21. [PubMed: 17322877]
70. Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, Balasubramanian S, Khanna R, Tellam J. G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. Nat Chem Biol. 2014; 10:358–64. [PubMed: 24633353]
71. Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, Schatz JH, Rodrigo CM, Zhao C, Rondou P, de Stanchina E, Teruya-Feldstein J, Kelliher MA, Speleman F, Porco JA Jr, Pelletier J, Ratsch G, Wendel HG. RNA G-quadruplexes cause eIF4A–dependent oncogene translation in cancer. Nature. 2014; 513:65–70. [PubMed: 25079319]
72. Ivanov PV, Zvereva MI, Shpanchenko OV, Dontsova OA, Bogdanov AA, Aglyamova GV, Lim VI, Teraoka Y, Nierhaus KH. How does tmRNA move through the ribosome? FEBS Lett. 2002; 514:55–9. [PubMed: 11904181]
73. Kwok CK, Sahakyan AB, Balasubramanian S. Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. Angew Chem Int Ed Engl. 2016; 55:8958–61. [PubMed: 27355429]

74. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 2005; 127:4223–31. [PubMed: 15783204]
75. Wassarman KM, Steitz JA. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. Mol Cell Biol. 1992; 12:1276–85. [PubMed: 1372090]
76. Wells SE, Hughes JM, Igel AH, Ares M Jr. Use of dimethyl sulfate to probe RNA structure in vivo. Methods Enzymol. 2000; 318:479–93. [PubMed: 10890007]
77. Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc. 2006; 1:1610–6. [PubMed: 17406453]
78. Zvereva MI, Ivanov PV, Teraoka Y, Topilina NI, Dontsova OA, Bogdanov AA, Kalkum M, Nierhaus KH, Shpanchenko OV. Complex of transfer-messenger RNA and elongation factor Tu. Unexpected modes of interaction. J Biol Chem. 2001; 276:47702–8. [PubMed: 11595738]
79. Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science. 2016:353.
80. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods. 2016; 13:841–4. [PubMed: 27571552]
81. Morris MJ, Basu S. An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. Biochemistry. 2009; 48:5313–9. [PubMed: 19397366]
82. Kutyavin IV, Lokhov SG, Afonina IA, Dempcy R, Gall AA, Gorn VV, Lukhtanov E, Metcalf M, Mills A, Reed MW, Sanders S, Shishkina I, Vermeulen NM. Reduced aggregation and improved specificity of G-rich oligodeoxyribonucleotides containing pyrazolo[3,4-d]pyrimidine guanine bases. Nucleic Acids Res. 2002; 30:4952–9. [PubMed: 12433999]
83. Murchie AI, Lilley DM. Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. Nucleic Acids Res. 1992; 20:49–53. [PubMed: 1738603]
84. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Molecular Biology of the Cell. 4. New York: Garland Science; 2002.
85. Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, Brosh RB Jr, Sen D, Lansdorp PM. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 2017
86. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature. 2014; 505:696–700. [PubMed: 24270811]
87. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature. 2014; 505:701–5. [PubMed: 24336214]
88. Loughrey D, Watters KE, Settle AH, Lucks JB. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. 2014:42.
89. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Chang HY. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015; 519:486–90. [PubMed: 25799993]
90. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
91. Beaudoin JD, Jodoin R, Perreault JP. New scoring system to identify RNA G-quadruplex folding. Nucleic Acids Res. 2014; 42:1209–23. [PubMed: 24121682]
92. Benhalevy D, Gupta SK, Danan CH, Ghosal S, Sun HW, Kazemier HG, Paeschke K, Hafner M, Juranek SA. The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. Cell Rep. 2017; 18:2979–2990. [PubMed: 28329689]

93. von Hacht A, Seifert O, Menger M, Schutze T, Arora A, Konthur Z, Neubauer P, Wagner A, Weise C, Kurreck J. Identification and characterization of RNA guanine-quadruplex binding proteins. Nucleic Acids Res. 2014; 42:6630–44. [PubMed: 24771345]
94. Zamiri B, Reddy K, Macgregor RB Jr, Pearson CE. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J Biol Chem. 2014; 289:4653–9. [PubMed: 24371143]
95. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature. 2014; 507:195–200. [PubMed: 24598541]
96. Takahama K, Miyawaki A, Shitara T, Mitsuya K, Morikawa M, Hagihara M, Kino K, Yamamoto A, Oyoshi T. G-Quadruplex DNA- and RNA-Specific-Binding Proteins Engineered from the RGG Domain of TLS/FUS. ACS Chem Biol. 2015; 10:2564–9. [PubMed: 26360301]
97. Takahama K, Oyoshi T. Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. J Am Chem Soc. 2013; 135:18016–9. [PubMed: 24251952]
98. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile × mental retardation protein targets G quartet mRNAs important for neuronal function. Cell. 2001; 107:489–99. [PubMed: 11719189]
99. Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C, Moine H. The fragile × mental retardation protein binds specifically to its mRNA via a purine quartet motif. EMBO J. 2001; 20:4803–13. [PubMed: 11532944]
100. Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D, Darnell RB, Patel DJ. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. Nat Struct Mol Biol. 2011; 18:796–804. [PubMed: 21642970]
101. Bensaid M, Melko M, Bechara EG, Davidovic L, Berretta A, Catania MV, Gecz J, Lalli E, Bardoni B. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. Nucleic Acids Res. 2009; 37:1269–79. [PubMed: 19136466]
102. Melko M, Douguet D, Bensaid M, Zongaro S, Verheggen C, Gecz J, Bardoni B. Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability. Hum Mol Genet. 2011; 20:1873–85. [PubMed: 21330300]
103. Biffi G, Tannahill D, Balasubramanian S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. J Am Chem Soc. 2012; 134:11974–6. [PubMed: 22780456]
104. O'Day E, Le MT, Imai S, Tan SM, Kirchner R, Arthanari H, Hofmann O, Wagner G, Lieberman J. An RNA-binding Protein, Lin28, Recognizes and Remodels G-quartets in the MicroRNAs (miRNAs) and mRNAs It Regulates. J Biol Chem. 2015; 290:17909–22. [PubMed: 26045559]
105. Emara MM, Ivanov P, Hickman T, Dawra N, Tisdale S, Kedersha N, Hu GF, Anderson P. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. J Biol Chem. 2010; 285:10959–68. [PubMed: 20129916]
106. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. Angiogenin-induced tRNA fragments inhibit translation initiation. Mol Cell. 2011; 43:613–23. [PubMed: 21855800]
107. Ivanov P, O'Day E, Emara MM, Wagner G, Lieberman J, Anderson P. G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. Proc Natl Acad Sci U S A. 2014
108. Ivanov PPA. Stress-induced Ribonucleases. Ribonucleases. Nucleic Acids and Molecular Biology. 2011; 26:115–134.
109. Lyons SM, Achorn C, Kedersha NL, Anderson PJ, Ivanov P. YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression. Nucleic Acids Res. 2016; 44:6949–60. [PubMed: 27174937]

110. Creacy SD, Routh ED, Iwamoto F, Nagamine Y, Akman SA, Vaughn JP. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. J Biol Chem. 2008; 283:34626–34. [PubMed: 18842585]
111. Lattmann S, Stadler MB, Vaughn JP, Akman SA, Nagamine Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. Nucleic Acids Res. 2011; 39:9390–404. [PubMed: 21846770]
112. Lattmann S, Giri B, Vaughn JP, Akman SA, Nagamine Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. Nucleic Acids Res. 2010; 38:6219–33. [PubMed: 20472641]
113. Sexton AN, Collins K. The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. Mol Cell Biol. 2011; 31:736–43. [PubMed: 21149580]
114. Reddy K, Zamiri B, Stanley SY, Macgregor RB Jr, Pearson CE. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J Biol Chem. 2013; 288:9860–6. [PubMed: 23423380]
115. Khateb S, Weisman-Shomer P, Hershco I, Loeb LA, Fry M. Destabilization of tetraplex structures of the fragile × repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. Nucleic Acids Res. 2004; 32:4145–54. [PubMed: 15302914]
116. Khateb S, Weisman-Shomer P, Hershco-Shani I, Ludwig AL, Fry M. The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile × premutation mRNA. Nucleic Acids Res. 2007; 35:5775–88. [PubMed: 17716999]
117. Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Manley JL. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife. 2016:5.
118. Mori K, Lammich S, Mackenzie IR, Forne I, Zilow S, Kretzschmar H, Edbauer D, Janssens J, Kleinberger G, Cruts M, Herms J, Neumann M, Van Broeckhoven C, Arzberger T, Haass C. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. 2013; 125:413–23. [PubMed: 23381195]
119. Grand CL, Han H, Munoz RM, Weitman S, Von Hoff DD, Hurley LH, Bearss DJ. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. Mol Cancer Ther. 2002; 1:565–73. [PubMed: 12479216]
120. Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. 2008; 36:144–56. [PubMed: 17999996]
121. Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, Mohd F, Sharma A, Chowdhury S. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. J Med Chem. 2008; 51:5641–9. [PubMed: 18767830]
122. Xiao S, Zhang JY, Zheng KW, Hao YH, 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; 41:10379–90. [PubMed: 23999096]
123. Zheng KW, Xiao S, Liu JQ, Zhang JY, Hao YH, 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; 41:5533–41. [PubMed: 23585281]
124. Wanrooij PH, Uhler JP, Shi Y, Westerlund F, Falkenberg M, Gustafsson CM. A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. Nucleic Acids Res. 2012; 40:10334–44. [PubMed: 22965135]
125. Zheng KW, Chen Z, Hao YH, Tan Z. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. Nucleic Acids Res. 2010; 38:327–38. [PubMed: 19858105]

126. Gromak N, West S, Proudfoot NJ. Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol Cell Biol. 2006; 26:3986–96. [PubMed: 16648491]

127. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011; 42:794–805. [PubMed: 21700224]

128. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic Acids Res. 2003; 31:1375–86. [PubMed: 12595544]

129. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008; 65:1099–122. [PubMed: 18158581]

130. Bagga PS, Ford LP, Chen F, Wilusz J. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. Nucleic Acids Res. 1995; 23:1625–31. [PubMed: 7784220]

131. Dalziel M, Nunes NM, Furger A. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. Mol Cell Biol. 2007; 27:1568–80. [PubMed: 17189425]

132. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev. 2015; 29:889–97. [PubMed: 25934501]

133. Decorsiere A, Cayrel A, Vagner S, Millevoi S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. Genes Dev. 2011; 25:220–5. [PubMed: 21289067]

134. Kikin O, Zappala Z, D'Antonio L, Bagga PS. GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. Nucleic Acids Res. 2008; 36:D141–8. [PubMed: 18045785]

135. Kostadinov R, Malhotra N, Viotti M, Shine R, D'Antonio L, Bagga P. GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. Nucleic Acids Res. 2006; 34:D119–24. [PubMed: 16381828]

136. Didiot MC, Tian Z, Schaeffer C, Subramanian M, Mandel JL, Moine H. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. Nucleic Acids Res. 2008; 36:4902–12. [PubMed: 18653529]

137. Gomez D, Lemarteleur T, Lacroix L, Mailliet P, Mergny JL, Riou JF. Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. Nucleic Acids Res. 2004; 32:371–9. [PubMed: 14729921]

138. Subramanian M, Rage F, Tabet R, Flatter E, Mandel JL, Moine H. G-quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep. 2011; 12:697–704. [PubMed: 21566646]

139. Kanai Y, Dohmae N, Hirokawa N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron. 2004; 43:513–25. [PubMed: 15312650]

140. Babendure JR, Babendure JL, Ding JH, Tsien RY. Control of mammalian translation by mRNA structure near caps. RNA. 2006; 12:851–61. [PubMed: 16540693]

141. Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 2012; 40:4727–41. [PubMed: 22351747]

142. Shahid R, Bugaut A, Balasubramanian S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. Biochemistry. 2010; 49:8300–6. [PubMed: 20726580]

143. Halder K, Wieland M, Hartig JS. Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. Nucleic Acids Res. 2009; 37:6811–7. [PubMed: 19740765]

144. Bonnal S, Schaeffer C, Creancier L, Clamens S, Moine H, Prats AC, Vagner S. A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. J Biol Chem. 2003; 278:39330–6. [PubMed: 12857733]

145. Morris MJ, Negishi Y, Pazsint C, Schonhoft JD, Basu S. An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. J Am Chem Soc. 2010; 132:17831–9. [PubMed: 21105704]

146. Terenin IM, Smirnova VV, Andreev DE, Dmitriev SE, Shatsky IN. A researcher's guide to the galaxy of IRESs. Cell Mol Life Sci. 2016

147. Lee SC, Zhang J, Strom J, Yang D, Dinh TN, Kappeler K, Chen QM. G-Quadruplex in the NRF2 mRNA 5' Untranslated Region Regulates De Novo NRF2 Protein Translation under Oxidative Stress. Mol Cell Biol. 2017:37.

148. Koukouraki P, Doxakis E. Constitutive translation of human alpha-synuclein is mediated by the 5'-untranslated region. Open Biol. 2016; 6:160022. [PubMed: 27248657]

149. Bolduc F, Garant JM, Allard F, Perreault JP. Irregular G-quadruplexes Found in the Untranslated Regions of Human mRNAs Influence Translation. J Biol Chem. 2016; 291:21751–21760. [PubMed: 27557661]

150. Arora A, Suess B. An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. RNA Biol. 2011; 8:802–5. [PubMed: 21734463]

151. Crenshaw E, Leung BP, Kwok CK, Sharoni M, Olson K, Sebastian NP, Ansaloni S, Schweitzer-Stenner R, Akins MR, Bevilacqua PC, Saunders AJ. Amyloid Precursor Protein Translation Is Regulated by a 3'UTR Guanine Quadruplex. PLoS One. 2015; 10:e0143160. [PubMed: 26618502]

152. Westmark CJ, Malter JS. FMRP mediates mGluR5-dependent translation of amyloid precursor protein. PLoS Biol. 2007; 5:e52. [PubMed: 17298186]

153. Endoh T, Kawasaki Y, Sugimoto N. Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. Angew Chem Int Ed Engl. 2013; 52:5522–6. [PubMed: 23589400]

154. Endoh T, Sugimoto N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. Sci Rep. 2016; 6:22719. [PubMed: 26948955]

155. Yu CH, Teulade-Fichou MP, Olsthoorn RC. Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. Nucleic Acids Res. 2014; 42:1887–92. [PubMed: 24178029]

156. Baranov PV, Atkins JF, Yordanova MM. Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. Nat Rev Genet. 2015; 16:517–29. [PubMed: 26260261]

157. Thandapani P, Song J, Gandin V, Cai Y, Rouleau SG, Garant JM, Boisvert FM, Yu Z, Perreault JP, Topisirovic I, Richard S. Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. Elife. 2015:4.

158. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LP. Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A. 2011; 108:260–5. [PubMed: 21173221]

159. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW 3rd, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron. 2013; 77:639–46. [PubMed: 23415312]

160. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C, Haass C, Edbauer D. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. 2013; 339:1335–8. [PubMed: 23393093]

161. Fratta P, Mizielinska S, Nicoll AJ, Zloh M, Fisher EM, Parkinson G, Isaacs AM. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep. 2012; 2:1016. [PubMed: 23264878]

162. Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, Fostvedt E, Jansen-West K, Belzil VV, Desaro P, Johnston A, Overstreet K, Oh SY, Todd PK, Berry JD, Cudkowicz ME, Boeve BF, Dickson D, Floeter MK, Traynor BJ, Morelli C, Ratti A, Silani V, Rademakers R, Brown RH, Rothstein JD, Boylan KB, Petrucelli L, Disney MD. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. Neuron. 2014; 83:1043–50. [PubMed: 25132468]

163. Bhattacharyya D, Diamond P, Basu S. An Independently folding RNA G-quadruplex domain directly recruits the 40S ribosomal subunit. Biochemistry. 2015; 54:1879–85. [PubMed: 25730526]

164. Jayaraj GG, Pandey S, Scaria V, Maiti S. Potential G-quadruplexes in the human long non-coding transcriptome. RNA Biol. 2012; 9:81–6. [PubMed: 22258148]

165. Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. Cancer Cell. 2016; 29:452–63. [PubMed: 27070700]

166. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013; 154:26–46. [PubMed: 23827673]

167. Matsumura K, Kawasaki Y, Miyamoto M, Kamoshida Y, Nakamura J, Negishi L, Suda S, Akiyama T. The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. Oncogene. 2016

168. Zhang R, Lin Y, Zhang CT. Greglist: a database listing potential G-quadruplex regulated genes. Nucleic Acids Res. 2008; 36:D372–6. [PubMed: 17916572]

169. Mirihana Arachchilage G, Dassanayake AC, Basu S. A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. Chem Biol. 2015; 22:262–72. [PubMed: 25641166]

170. Pandey S, Agarwala P, Jayaraj GG, Gargallo R, Maiti S. The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. Biochemistry. 2015; 54:7067–78. [PubMed: 26554903]

171. Hansel R, Foldynova-Trantirkova S, Lohr F, Buck J, Bongartz E, Bamberg E, Schwalbe H, Dotsch V, Trantirek L. Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. J Am Chem Soc. 2009; 131:15761–8. [PubMed: 19824671]

172. Serber Z, Corsini L, Durst F, Dotsch V. In-cell NMR spectroscopy. Methods Enzymol. 2005; 394:17–41. [PubMed: 15808216]

173. Hansel R, Foldynova-Trantirkova S, Dotsch V, Trantirek L. Investigation of quadruplex structure under physiological conditions using in-cell NMR. Top Curr Chem. 2013; 330:47–65. [PubMed: 22760824]

174. Collie G, Reszka AP, Haider SM, Gabelica V, Parkinson GN, Neidle S. Selectivity in small molecule binding to human telomeric RNA and DNA quadruplexes. Chem Commun (Camb). 2009:7482–4. [PubMed: 20024253]

175. Di Antonio M, Biffi G, Mariani A, Raiber EA, Rodriguez R, Balasubramanian S. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. Angew Chem Int Ed Engl. 2012; 51:11073–8. [PubMed: 23038154]

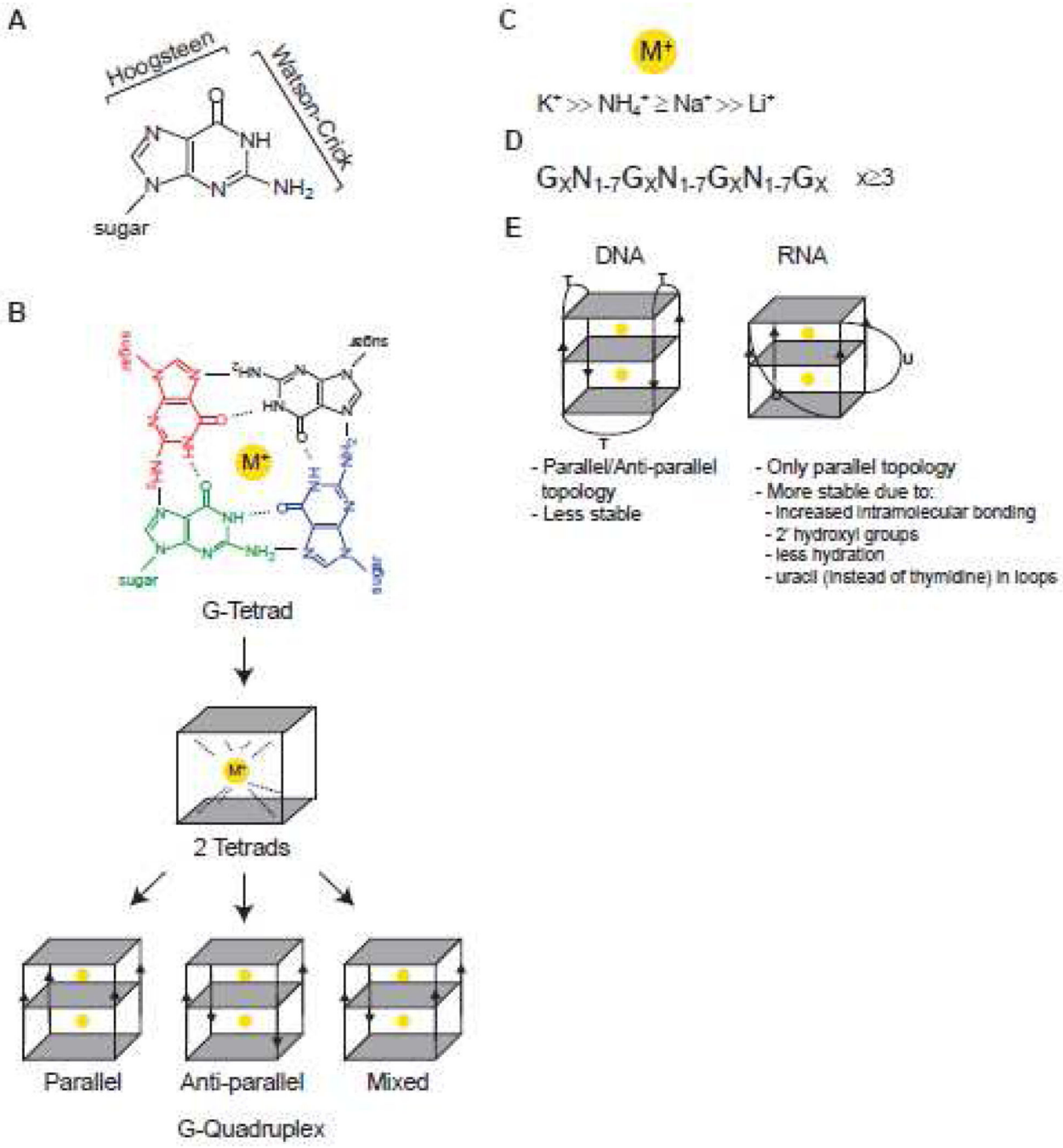


Figure 1. G-quartets and G-quadruplexes

A. Guanine atoms participating in Watson-Crick and Hoogsteen base pairing interactions. **B**. *Upper*: The structure of G-quartet (G-tetrad) showing four-fold symmetrical arrangement of coplanar guanines. Hydrogen bonds between each pair of guanines involves four donor/acceptor atoms (the N1, N7, N2 and O6 atoms). Overall arrangement has a total of eight hydrogen bonds (four N2-H…N7 and four N1-H…O6 bonds). Four carbonyl oxygen (O6) atoms form a negatively charged core in the center of G-quartet that favors binding of monovalent cations (M^+). The sugar moieties are removed for clarity. *Middle*: 2 tetrads

coordinated by a monovalent cation are shown. *Lower*: G-quadruplex structures with different topologies are shown. Note that G4 can be inter- and intra-molecular. **C**. Preferential binding of monovalent cations to G4s. **D**. Simplified prediction algorithm used to identify potential G4 motif (**G_X-$N_{1–7}$-G_X-$N_{1–7}$-$G_X$$N_{1–7}$-$G_X$**, where *x* is 3–6 nucleotides and *N* corresponds to any nucleotide (A, G, T, C or U)). **E**. Summary of the main differences between DNA and RNA G4s.

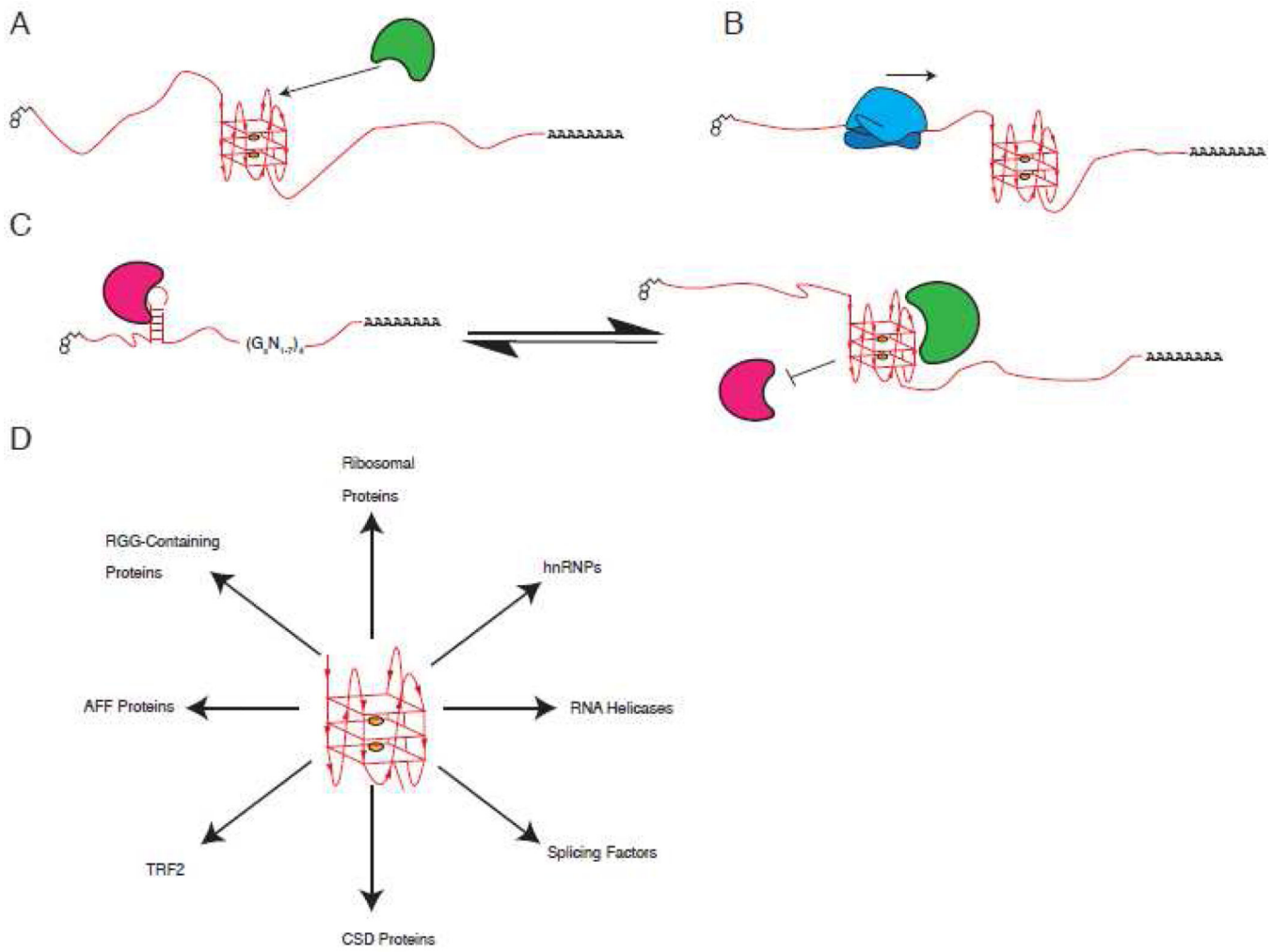


Figure 2. Possible mechanisms of action underlying RNA G4 functions

A. G4s act as specific binding sites for regulatory or structural proteins. **B**. G4s act as barriers or kinetic traps for movement of proteins or protein machines along RNA. **C**. G4s regulate the formation of alternative secondary structures on RNA, recognized by different proteins. Arrows indicate possible bidirectional shift of equilibrium between G4 and hairpin conformation. **D**. Examples of proposed G4-specific binding proteins.

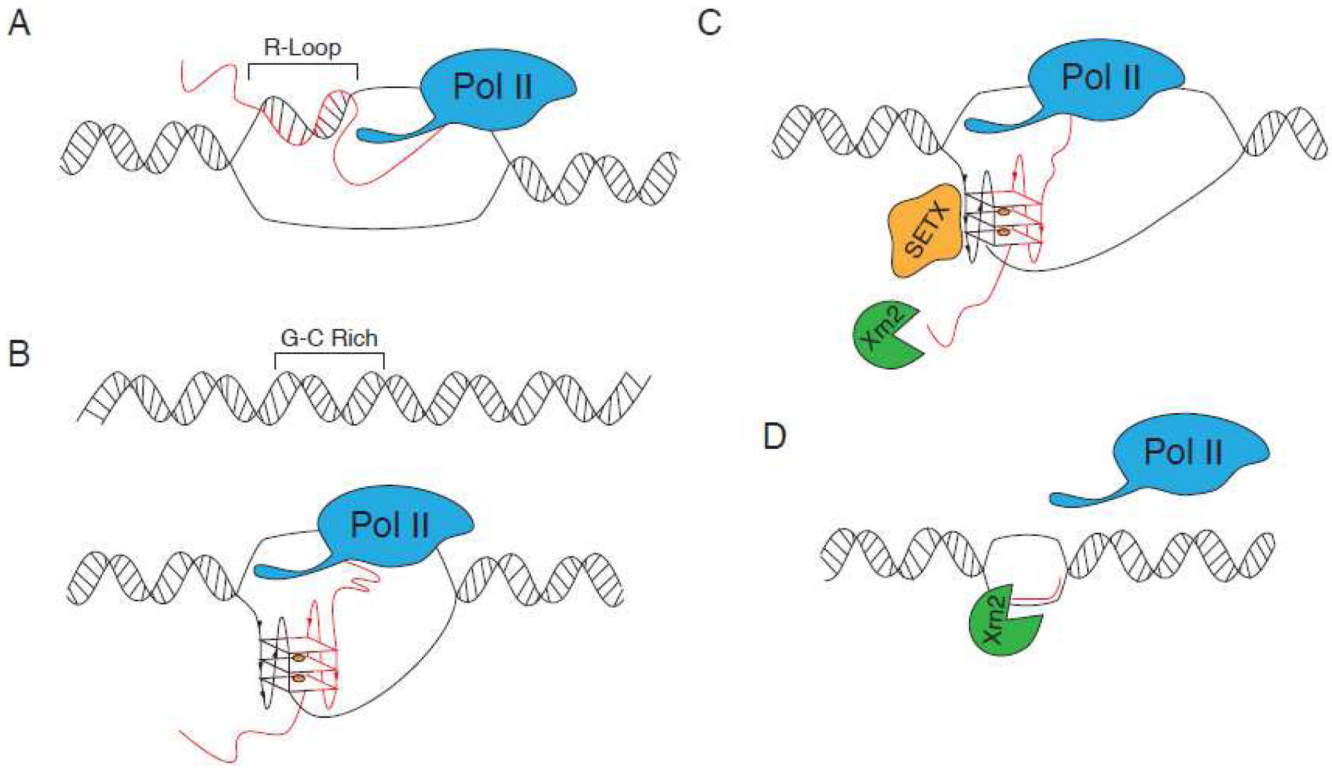


Figure 3. Proposed roles of RNA G4s in transcriptional regulation

A. R-loops, structures that contain an RNA-DNA hybrid and displaced single-stranded DNA, can form during transcription when RNA emerging from the transcription machinery hybridizes with the DNA template (RNA transcript shown in red). **B**. As few as two tandem G-tracks (guanine-rich sequences) on a non-template DNA strand are capable of forming hybrid G4 with guanine-rich transcript. **C–D**. G4s can act as terminator sequences to cause Pol II transcription to pause. Hybrid DNA/RNA G4s are very stable and require assistance of specialized enzymes to unwind G4s and assist with transcription termination. In mammals, helicase senataxin (SETX) cooperates with exoribonuclease Xrn2 to resolve hybrid G4s, promote degradation of 3'-end RNA cleavage product (red) and release of Pol II from DNA.

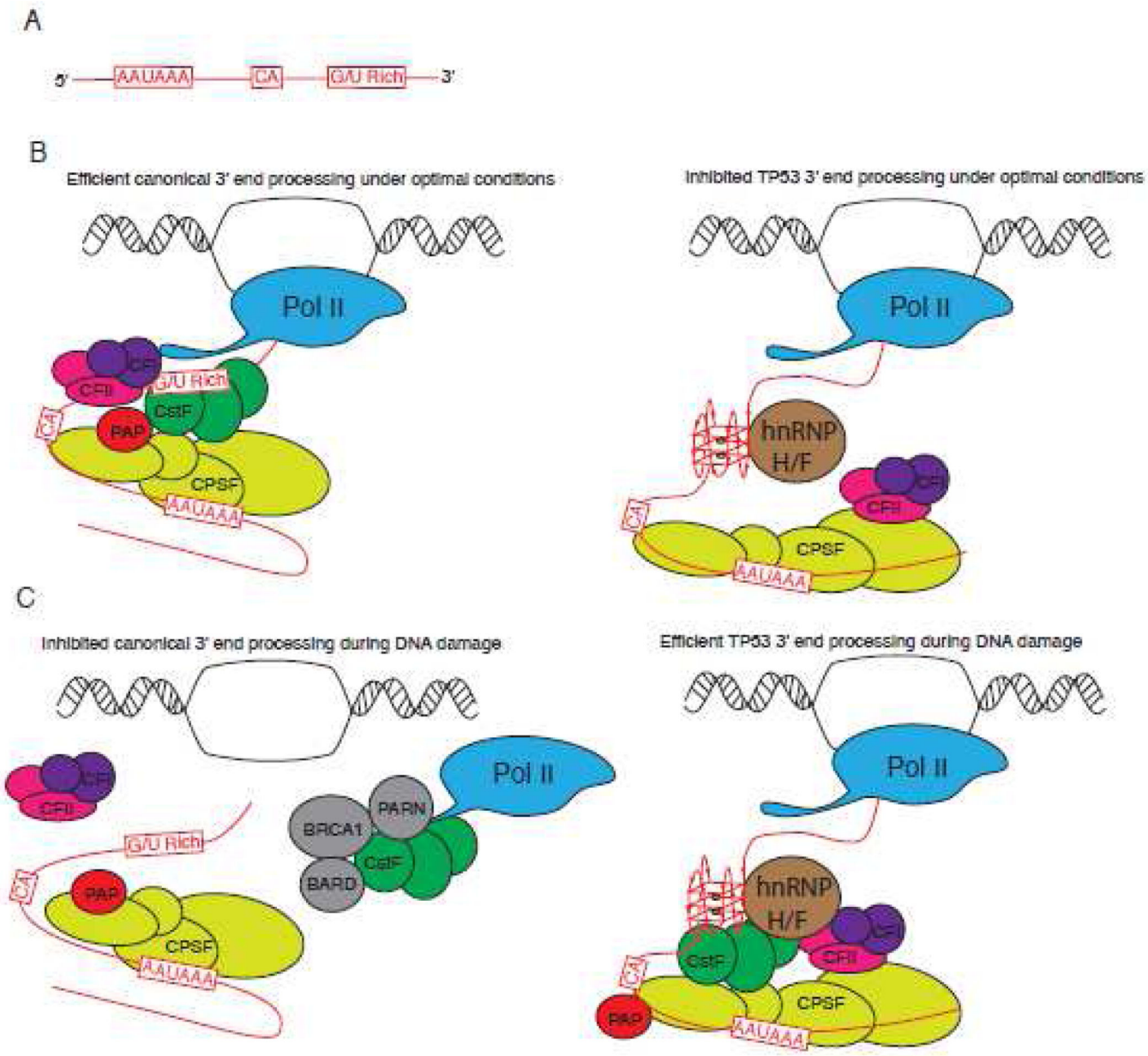


Figure 4. Proposed roles of RNA G4s in 3'-end mRNA processing

A. Three primary sequence elements that define the polyadenylation site. These pre-mRNA *cis*-elements include the hexamer AAUAAA polyadenylation signal, CA cleavage site and G/U-rich downstream element. **B**. Under optimal conditions, multi-subunit cleavage/polyadenylation machinery (CFI/CFII/CstF/CPSF) assembles on these *cis*-elements to promote efficient 3'-end processing and global transcription (left panel). 3'-end processing of specific stress-responsive mRNAs, such as encoding DNA damage factor p53, is however inhibited. The downstream G/U-rich element of *TP53* pre-mRNA assembles a G4 that is recognized by hnRNP H/F. Binding of hnRNP H/F interferes with efficient recruitment of cleavage/polyadenylation machinery and inhibits 3'-end processing of *TP53* pre-mRNA (right panel). **C**. Under DNA damage, specific factors sequester the essential

polyadenylation factor CSTF into an inactive complex thus inhibiting 3'-end processing (left panel). In contrast, hnRNP H/F bound to G4 of *TP53* pre-mRNA associates with CstF thus protecting it from sequestration and promoting TP53 3'-end processing and expression of p53 (right panel).

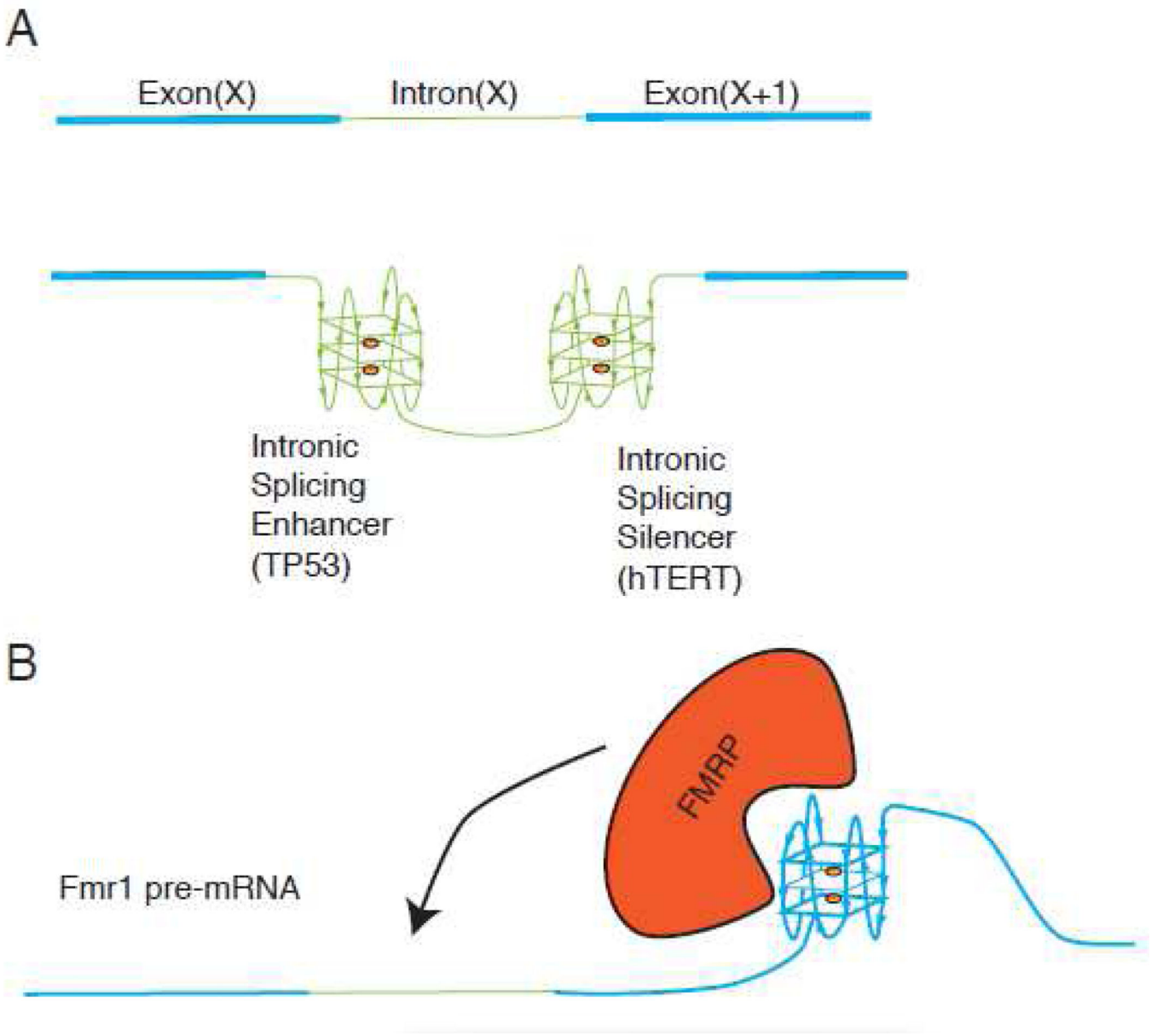


Figure 5. Proposed roles of RNA G4s in splicing regulation

A. Exonic and intronic regions of a putative pre-mRNA are shown as blue (Exon (X) and Exon (X+1)) or green (Intron X), respectively. Intronic G4s can either enhance or silence intron splicing thus acting as intronic splicing enhancer (such as in intron of *TP53* pre-mRNA) or intronic splicing silencer (such as in intron of *hTERT* pre-mRNA). **B**. Exon-located G4s can act as exonic splicing enhancers. FMRP recognizes G4s in its own pre-mRNA (*FMR1) to* regulate its splicing pattern and production of protein isoforms.

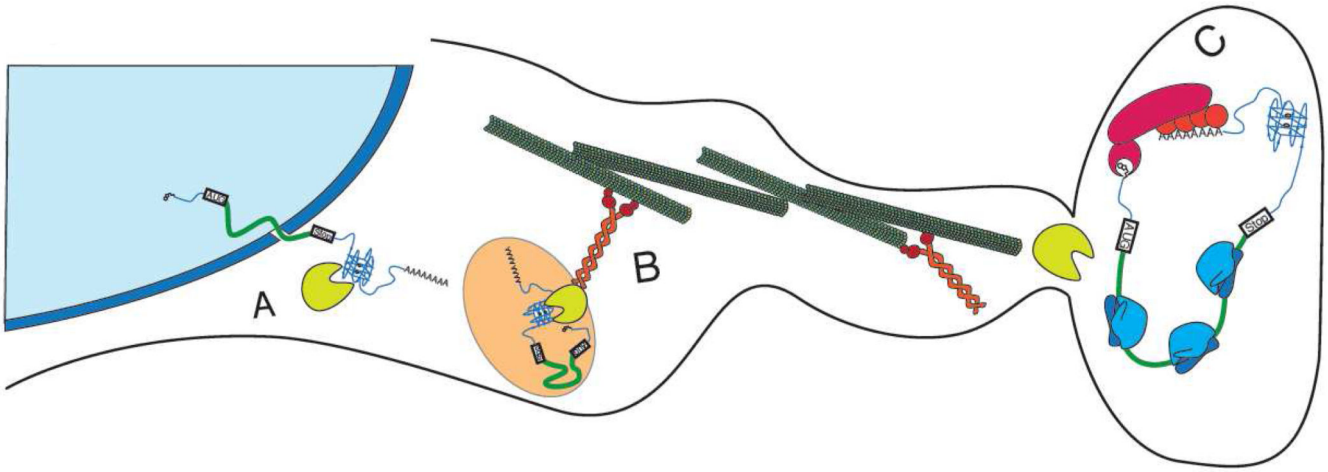


Figure 6. Proposed roles of RNA G4s in mRNA localization

A. In specialized cells such as neurons, processed G4-containing mRNAs are exported from the nucleus to the cytoplasm where specific RBPs (e.g. hnRNP U, FUS/TLS and FMRP) recognize and bind to their G4 structures. **B**. G4-bound RBPs assemble into large mRNPs (such as neuronal granules) that specialize in mRNA transport using molecular motors such as microtubule-associated kinesins. **C**. Upon arrival to the destination site (e.g. dendritic synapses), G4-containing mRNPs remodel allowing local translation.

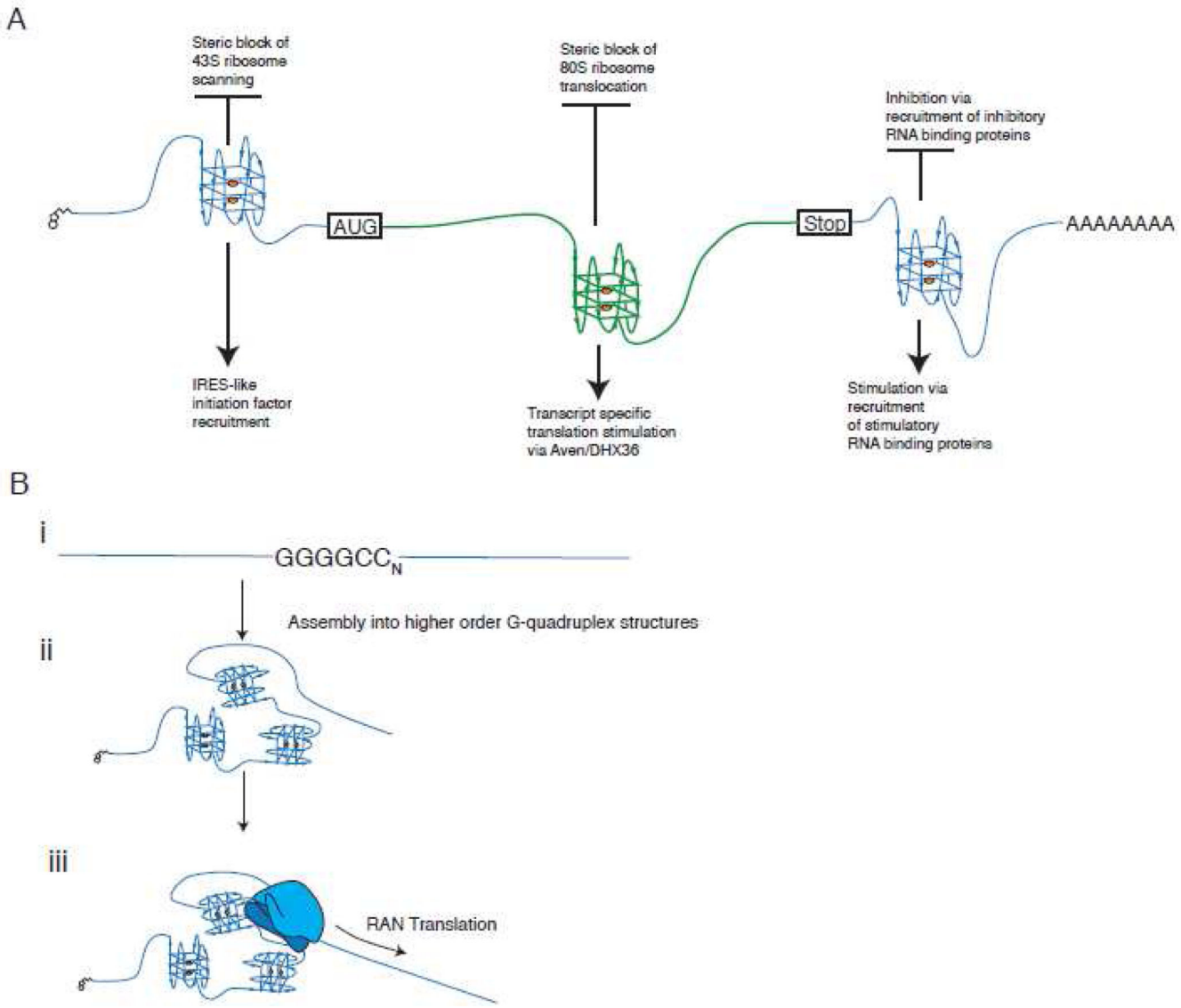


Figure 7. Proposed roles of RNA G4s in protein synthesis

A. Proposed roles of G4s in mRNA translation. G4s can be found in both non-coding (5'- and 3'-UTRs) and coding (ORF) regions of mRNA where they can inhibit or stimulate translation. In 5'-UTRs, G4s are commonly found near mRNA 5'-cap structures, where they inhibit translation initiation by mechanisms that may involve interference with cap binding or inhibition of 43S pre-initiation complex scanning (e.g. in *NRas* mRNA). G4-mediated stimulation of translation is proposed in an IRES-like manner (e.g. in 5'-UTRs of *VEGF* or *FGF2* mRNAs), although strong experimental evidences are still lacking. In 3'-UTRs, G4s can both inhibit and stimulate translation by recruitment of translational silencers or stimulators, the identity of which is still unknown. As in non-coding parts, G4s located in the ORF can both block (most likely acting as roadblocks for elongating ribosomes) or promote translation by recruitment of specific protein complexes (such as RBP Aven and RNA helicase DHX36). **B**. Hypothetical role of G4s in repeat-associated non-AUG (RAN) translation, an unconventional translation mechanism. In ALS patients, hexameric

r(GGGGCC)$_n$ repeats located in intron of *C9ORF72* gene are amplified (i). These repeats are capable of forming higher order G4 structures that make them extremely stable (ii). Some r(GGGGCC)n transcripts undergo RAN translation via direct recruitment of translationally competent ribosomal complexes to produce di-peptide proteins (iii).

Table 1

Protein	Supporting Data	Functional implications
U2AF65	Pulldowns confirmed by SPR spectroscopy show that U2AF65 associates with *ARPC2* and *MMP16* RNA G4s, yet not to the mutated control [93].	Expected to affect splicing, yet no functional data to date.
SRSF1 (ASF/SF2)	Pulldowns confirmed by SPR spectroscopy show SRSF2 associates with *ARPC2* RNA G4, yet not to the mutated control and did not associate with *MMP16* RNA G4, suggesting specificity in G4 affinity[93]; binds GGGGCC repeats in gel shift assay and this is sensitive to disruption by TMPyP4 [94].	Expected to affect splicing, yet no functional data to date.
Nucleolin	Pulldown confirmed by SPR spectroscopy shows Nucleolin binds *APRC2* RNA G4 but not the mutated control [93], Nucleolin preferentially binds to GGGGCC RNA in G4 over the hairpin conformation [95].	Expected to affect mRNA stability. Indirect evidences.
FUS/TLS	Electrophoretic mobility shift assays (EMSAs) and Isothermal Titration Calorimetry show FUS/TLS binds telomeric DNA sequences and TERRA RNA via the RGG domain [96]. Tyrosines in the FUS/TLS RGG domain recognize the 2'OH groups in the RNA backbone [97].	Regulate histone modifications at telomeres [96].
FMRP	FMRP binds RNA forming G-quartets in the presence of K^+ but not Li^+, consensus sequence $DWGGN_{0-2}DWGGN_{0-1}DWGGN_{0-1}DWGG$ [98; 99]. NMR of FMRP with a G-quartet forming RNA shows FMRP binds duplex - G-quadruplex junction with the RGG domain [100].	Expected to regulate mRNA stability and translation.
FMR2 (AFF2)	Filter binding assay using exonic splicing enhancer shows FMR2 binds G-quartets in the presence of K^+ and moderately Na^+ but not Li^+, this interaction occurs through the C-terminus of FMR2 [101; 102].	Alternative splicing by recognition of the exonic splicing enhancer [101].
AFF3, AFF4	Filter binding assays were used to show that the C-terminal domains of AFF3/AFF4 bind the G-quartet forming exonic splicing enhancer in the presence K^+ but not Li^+ [102].	Alternative splicing by recognition of the exonic splicing enhancer [102].
TRF2	ELISAs with immobilized TRF2 bind G-quadruplex containing oligos including *TERRA*, *Bcl-2*, *NRas* [103].	Potential regulation of telomeric ends, although no strong evidence to date [103].
Lin28	Analysis of the top 50 RNAs bound by Lin28 indicates an enrichment in potentially forming G-quadruplex RNAs. EMSAs coupled with NMM fluorescence showed that Lin28 binds to and remodels target G-quadruplexes [104].	Remodels G-quartets, affects mRNA stability, affects miR metabolism [104].
YB1	Using pulldowns, YB1 was shown to use its cold shock domain (CSD) to bind G-quadruplex forming tRNA-derived, stress induced RNAs (tiRNAs) [118]. Pulldowns with the *ARPC2* G4 RNA and mutated versions, YB1 was shown to bind G-rich sequences [93].	Stress granule formation and translation inhibition initiated by tiRNAs[105; 106; 107; 108; 109]
RHAU (DHX36 or G4 resolvase)	RHAU unwinds various RNA G4s; EMSAs with a synthetic G4 and purified RHAU; RHAU-depleted cell lysates lack G4 unwinding activities [110], associates and unwinds the G4 containing *TERC* RNA [111], requires both the helicase domains and the N-term regions for tight bindings to G4s [112].	RHAU interacts with telomerase holoenzyme, potential role in telomere biology [111; 113].
DHX9 (RNA Helicase A, RHA)	EMSAs showing recombinant DHX9 unwinds G4 forming oligos [119].	Expected to affect transcript stability. No available data.
eIF4A	eIF4A effects the translation of a reporter construct with G4 containing sequences and *hippuristanol*, a drug that inhibits eIF4A activity, increases translation of RNAs containing G4s. Ribosomal profiling revealed $(CGG)_4$, capable of forming G4, as the eIF4A consensus motif [71].	Alters translation efficiency of mRNAs with G4 in the 5' UTR[71].
hnRNP A1	hnRNP A1 binds GGGGCC repeats in EMSA assays and this interaction is disrupted by TMPyP4, which distorts G4 structures [94; 114].	Expected to regulate mRNA splicing. No functional data available
hnRNP A2	EMSAs show that hnRNP A2 destabilizes CGG from folding into a G4 structure [115]. hnRNP A2 disrupts G4 forming sequences (FMR1 5' UTR) to promote the translation of a reporter[116].	Promotes the translation of FMR1 by preventing G4 from forming [116].
hnRNP A3	hnRNP A3 identified in pulldown experiments to bind GGGGCC although not under any conditions that test G4 specificity. hnRNPA3 localizes to GGGGCC foci that have since been identified as G4 containing [117; 118].	Expected to regulate mRNA splicing. No functional data available

Protein	Supporting Data	Functional implications
hnRNP H	Pulldowns from nuclear extract shown hnRNP H interacts with GGGGCC RNA and this is dependent on G4 structure as indicated by 7-deaza derivative variants, and G4 stabilization by porphyrins and K^+ ions. Also colocalizes with BG4 (G4-specific antibody) in patient samples with many GGGGCC repeats [117].	Sequestration of hnRNP H to GGGGCC-G4 foci causes alterations in splicing in patients with ALS where GGGGCC repeats are increased [117].
CNBP/ZNF9	PAR-CLIP indicated the CNBP consensus binding sites can also fold into G4s and filter binding assays show that CNBP interacts with G-rich sequences *in vitro*. CNBP promotes the translation of mRNAs with G-rich sequences by preventing G4 formation as assessed by circular dichroism, reporter assays, and ribosome profiling [92].	Promotes the translation of G4 forming by prevent G4 formation [92].