12534–12555 *Nucleic Acids Research, 2020, Vol. 48, No. 22* *Published online 2 December 2020*
doi: 10.1093/nar/gkaa1126

SURVEY AND SUMMARY

Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back

Prakash Kharel [1], Gertraud Becker[1], Vladimir Tsvetkov[2,3,4] and Pavel Ivanov [1,5,*]

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA, [2]Computational Oncology Group, I. M. Sechenov First Moscow State Medical University, Moscow 119146, Russia, [3]Federal Research and Clinical Center for Physical-Chemical Medicine, Federal Medical Biological Agency, Moscow 119435, Russia, [4]A. V. Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences, Moscow 117912, Russia and [5]Harvard Initiative for RNA Medicine, Boston, MA 02115, USA

Received September 16, 2020; Revised October 23, 2020; Editorial Decision October 28, 2020; Accepted November 06, 2020

ABSTRACT

Guanine-quadruplexes (G4s) are non-canonical four-stranded structures that can be formed in guanine (G) rich nucleic acid sequences. A great number of G-rich sequences capable of forming G4 structures have been described based on *in vitro* analysis, and evidence supporting their formation in live cells continues to accumulate. While formation of DNA G4s (dG4s) within chromatin *in vivo* has been supported by different chemical, imaging and genomic approaches, formation of RNA G4s (rG4s) *in vivo* remains a matter of discussion. Recent data support the dynamic nature of G4 formation in the transcriptome. Such dynamic fluctuation of rG4 folding-unfolding underpins the biological significance of these structures in the regulation of RNA metabolism. Moreover, rG4-mediated functions may ultimately be connected to mechanisms underlying disease pathologies and, potentially, provide novel options for therapeutics. In this framework, we will review the landscape of rG4s within the transcriptome, focus on their potential impact on biological processes, and consider an emerging connection of these functions in human health and disease.

INTRODUCTION

G-quadruplexes (G4s) are non-canonical secondary structures formed by G-rich nucleic acid sequences. A German physician and chemist Ivar Bang in 1910 reported that guanosine monophosphate and its analogues, but no other nucleosides, readily form gel in minerally impure solutions (1). This unique property of guanine derivatives was largely overlooked until the 1960s when Gellert *et al*. used fiber crystallography to arrive at a model of four guanine bases arranged in a planar tetrameric square arrangement via Hoogsteen base pairing to form a G-quartet (or G-tetrad (Figure 1A)), which explained the gooey nature of the guanine solution (2). Interest in the structural arrangements of G4s was only ignited in the late 1980s when Henderson *et al.* observed a higher order structure formed by G:G interaction in telomeric DNA sequence (3) and, separately, Sen and Gilbert discovered that short G-rich DNA oligonucleotides can self-assemble to form G4 structures under physiological salt concentrations indicating a potential regulatory role of such structures during meiosis (4). It is clear from multiple studies that the presence of cations is key to G4 formation and their stability, at least *in vitro* (5,6). Stacks of G-quartets (Figure 1B and D) are stabilized by cations centrally coordinated to O6 of the guanines with stabilizing preference in the following order: $Sr^{2+} > Ba^{2+} > K^+ > Na^+$, NH_4^+, $Rb^+ >> Li^+$. The larger cations (i.e. K^+) are located between the plane of two quartets and coordinate eight carbonyl oxygens whereas the smaller cations (i.e. Na^+) are deep-seated in the middle of a quartet and coordinate only four carbonyl oxygens (Figure 1A). Furthermore, smaller cations, like Li^+, do not favor G4 formation (Figure 1C) while cations like Ca^{2+}, Cs^+, Mg^{2+} were found to destabilize the G4s, probably through the interactions with electron acceptor groups of the guanine Hoogsteen edge. In contrast to the key role of the cations in the core of the G-quartets, G4 stability is only slightly affected by ions masking the negatively charged sugar-phosphate backbone (7).

The number of stacks of G-quartets in a G4 is specified by the sequence and length of each individual G-tract and their connecting loops (Figure 1B and D) (8). The feasible intramolecular G4s (Figure 1B) are designated as

*To whom correspondence should be addressed. Tel: +1 617 525 1233; Email: pivanov@rics.bwh.harvard.edu

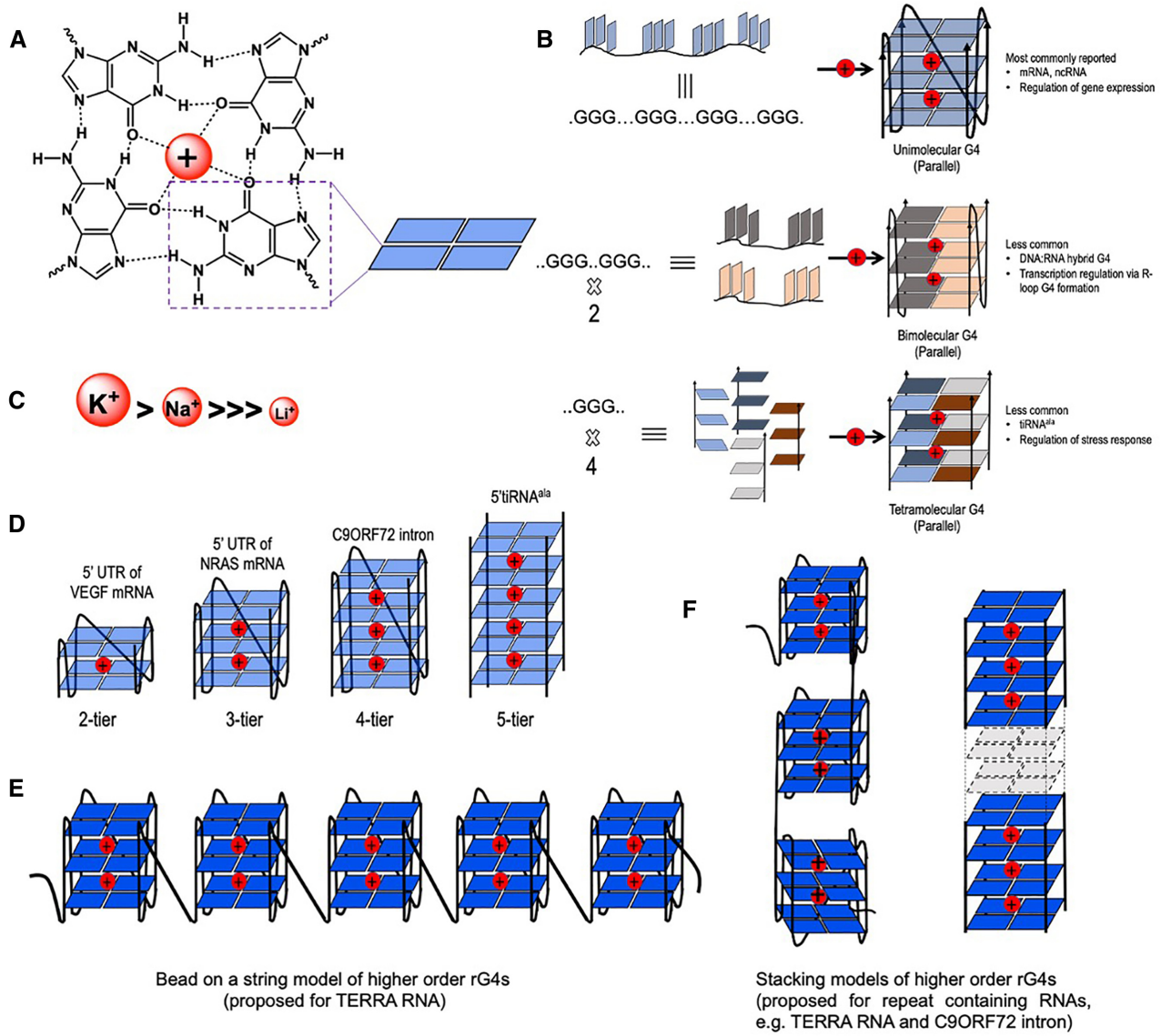

Figure 1. RNA G-quadruplex architecture, complexity and formation. (**A**) G-quartet, a square planar arrangement of 4 Gs stabilized by a cation, is a structural unit of G-quadruplexes. (**B**) The presence of structurally diverse RNA G-quadruplexes in various RNA species reflects their roles in various aspects of RNA metabolism. (**C**) Most commonly used cations in G4 studies, and their G4 stabilizing order. (**D**) Various types of commonly reported rG4s based on different number of G-quartet stacking. (**E**, **F**) Proposed complex rG4 structures in pG4 repeat containing sequences like *TERRA* and *C9ORF72* intronic RNAs.

'GxNyGxNyGxNyGx'. Here, 'Gx ($x \geq 2$)' indicates the number of Gs in each G-tract, whereas 'Ny ($y \geq 1$)' specifies nucleotide combinations of different lengths found in the linking loops. Most bimolecular G4s consist of two identical strands comprised of the same 'NyGxNyGxNy' combination whereas the tetramolecular ones consist of 'NyGxNy' sequence (Figure 1B). Based on biophysical studies of different G4s, algorithms using sequence motifs such as GxNyGxNyGxNyGx were developed and deployed to predict putative G4 structures in genomic DNA (8,9). Early models assumed loop lengths no longer than seven and a requirement for four continuous stretches of Gs (8,10). Subsequently, non-conventional G4s with longer loop lengths and discontinuities in G-stretches causing bulges were observed (11,12) leading to alternative predictive models (13,14). In addition, there are more complex models where G4s are either arranged in a 'beads on a string model' or in a 'stacking model' (Figure 1D). It is possible to apply machine learning to predict genome-wide G4 forming propensity with the availability of big datasets (15). Computational prediction algorithms indicate up to 375 000 G4s can be formed in the human genome simultaneously (10,16). Small molecule assisted high-throughput G4-seq analysis (using G4-stabilizing ligand pyridostatin (PDS)) revealed ∼700 000 G4s in the human genome, with ∼450 000 G4s that were not predicted by bioinformatics, including non-canonical G4s (17). Mapping of G4s in chromatin by G4 ChIP-sequencing with an anti-G4 antibody retrieved ∼10 000 G4s in highly transcribed regulatory nucleosome-depleted chromatin regions, indicating that G4s are mostly

suppressed in chromatin and may influence the occupancy and positioning of nucleosomes (18). In general, putative G4s (pG4s) are clustered in certain genomic regions, namely telomeres, gene promoters and DNA replication origins (19–21). Various *in vitro* and *in cella* studies suggest the crucial role of DNA G4s (dG4s) in the maintenance of chromosomal integrity, and the regulation of replication, transcription and recombination (13,22,23).

Given the prevalence of dG4s in the genome, it was no surprise that similar sequences in RNA can also form RNA G4 structures (rG4s) (24). Pioneer work from the Balasubramanian lab employing rG4-seq in vitro to profile RNA G4s in the polyadenylated (poly(A)) fraction of RNA (∼17 000 transcripts) determined that about 2500 genes contain ∼3700 putative rG4s. Even more putative rG4s sites distributed in ∼6000 genes were found under treatment with PDS (25). Consistent with whole genome sequencing of human dG4s, rG4 enrichment was found in the 5′ and 3′ untranslated regions (UTRs) of mRNA, as well as within the 5′ end of the first intron in pre-mRNAs (reviewed in (26–28)). rG4-motifs were also found in protein-coding regions, albeit less frequently. Consequently, rG4s are implicated in transcriptional regulation, mRNA processing, the regulation of translation and RNA translocation. rG4s were also indicated in non-coding RNAs (ncRNAs)–especially long ncRNAs (lncRNA), telomeric repeat associated RNA (TERRA), the RNA component of the telomerase (TERC) and microRNAs–with yet mainly obscure biological significance (29–31). Most studied rG4s act as translational repressors, however some were found to modulate mRNA polyadenylation and splicing, dictate neuritic mRNA targeting, and more (26–28). Certain rG4s could contribute to genomic stability by building RNA:DNA hybrid G4s or rG4:RNA-binding protein (RBP) complexes, which regulate genome processes, such as recombination, telomere homeostasis or gene expression (reviewed in (28,32)). Additionally, previous studies excluded ribosomal RNA (rRNA)—the most abundant cellular RNA—however it was the first type of RNA shown to form highly stable G4s *in vitro* (33).

Despite their apparent similarity, the assumption that rG4s are identical to the dG4s may be misleading because of the chemical, environmental, and structural differences between the two nucleic acids. While DNA mostly exists in a double-stranded form, RNA is mostly single-stranded or adopts different types of secondary structures (hairpins, loops, bulges, pseudoknots, etc.) and even has the potential to form tertiary structures. In addition, at least in eukaryotes, the subcellular localization of RNA and DNA further distinguish them: while DNA is located in the nucleus and mitochondria bound to histones and other auxiliary factors, RNA is found in both the nucleus and cytoplasm as well as in cytoplasmic organelles (i.e. the mitochondrion and endoplasmic reticulum) yielding a great diversity in protein-binding partners. Even at the molecular level, rG4s are more compact, less hydrated, and often more thermodynamically stable than their DNA counterparts (24,34). Over the last decade, G4s and their biology—long regarded as a mere *in vitro* phenomenon—received deeper attention: their structure, function, and *in cella* behavior not only in G-rich DNA and RNA, but also in nucleic acid analogues, were examined. The presence of G4s in viral, prokaryotic and eukaryotic genomes, as well as transcriptomes, suggest their evolutionary conservation, and thus regulatory relevance for numerous biological processes. Attention is now being paid to rG4s, which show enhanced stability and propensity for G4 formation. Multiple independent *in silico*, *in vitro*, *in vivo* and *in cella* studies signify the biological importance of rG4s. The aim of this review is to summarize the state-of-the-art knowledge of rG4 structure-function relationships and their broadened biological implications.

RNA G-QUADRUPLEX-TOPOLOGICAL DIVERSITY AND DYNAMICS

G4s can be intramolecular (unimolecular) or intermolecular (bimolecular or tetramolecular) (Figure 1B) and can adopt a wide diversity of topologies arising from different combinations of strand direction, as well as length and loop composition (35). Although the G-quartet organization is the signature of each G4 core, it is accompanied by a huge number of particular steric conformations depending on the primary sequence and physiochemical conditions. The right-handed helical forms of G4 were the only known orientation of both dG4s and rG4s until Chung *et al.* reported the existence of a left-handed dG4 structure *in vitro* for a well-known G4 aptamer AS1411 (36). All possible G4 topologies feature a four-stranded arrangement, which exhibits a regular rise and twist between the G-quartet planes and spans four grooves of variable width (8). This layout is accomplished either intramolecularly, provided that four or more G-tracts are present in one nucleic acid strand, or intermolecular, where G4s almost entirely arise from two (bimolecular) or four (tetramolecular) different strands (Figure 1B). Although commonly believed to be thermodynamically unstable, trimolecular G4 structures have been reported in the context of *in vitro* studies of the assembly of higher-order dG4 species (37,38). Using various *in vitro* biophysical techniques, Li *et al*. reported a unique type of intramolecular dG4 that forms with one G2 tract bearing a guanine vacancy along with three regular G3 tracts (39). The authors demonstrated that such a structure can recruit a guanine derivative from the environment to fulfil the vacancy. Given that there are almost as many G-vacancy containing pG4s in the genome as the regular pG4s, the impact of such potential vacancy filling G4s in the cellular G4 dynamics could be enormous (39). It is therefore plausible to assume the formation of rG4s from analogous RNA sequences which could have a wider impact on gene regulation. The possibility of formation of left-handed G4s, trimolecular G4s, and space-filling G4s open up a different dimension in the topological complexity of the transcriptome. G4 polymorphism is determined by the combination of strand orientation (parallel, antiparallel or mix), loop length, base composition and topology between successive G-tracts (propeller, lateral or diagonal), as well as conformation of the glycosidic bonds in each guanine. The G4 structure is parallel if each G-tract is oriented in the same direction, antiparallel if the alternate G-tracts run in the opposite direction, or it can be mixed. A correlation exists between the mutual orientation and length of G-tracts and the N-glyosidic *syn–anti*

conformation of guanosines inside each G-tetrad and along the G4 axis. Given the chemical properties of RNAs, rG4s are generally more stable than dG4s of the same sequence (24). The additional 2′-OH group in the ribose ring populates a C3′-endo sugar pucker conformation, which significantly organizes the hydration shell and the hydrogen bond network and promotes a parallel G4 topology in rG4s (29). Parallel G4s tend to dimerize or multimerize, as observed among quadruplexes in TERRA RNA (Figure 1E, F) (40–42). Generally, parallel G4s energetically favor G-residues in anti-conformation, are structurally more or less homogeneous and show four almost identical grooves (reviewed in (43)). In contrast, anti-parallel G4s carry guanines in both anti- and syn-conformation, which are arranged following to the particular G4 topology and orientation of participating G-tracts. Reports from the Xu lab suggested the possibility of formation of antiparallel rG4s in both native and modified RNA oligos *in vitro* (44,45). In addition to strand orientation and glyosidic bond orientation pattern, loop variation also affects the topology and dynamics of G4s. Although not directly involved in the G-quartet formation, the nature of the loops is a limiting factor for G4 folding and their thermodynamic stability. Formation of one or another loop type is dependent on the nucleotide sequence, the number of G-quartets and strand orientation within the G4. Two adjacent parallel strands require a strand-reversal loop, termed propeller or double-chain reversal loop, which consequently links the upper G-quartet with the lower ones without inverting the nucleic acid chain direction (46). On the other hand, adjacent antiparallel G-tracts are connected edgewise by lateral loops, which form the transition between opposed-sense G-tracts (47).

Structural studies using X-ray crystallography (XRC) (48,49) and nuclear magnetic resonance (NMR) spectroscopy (50) have provided detailed insights into the structure of G4s, primarily based on the human telomeric repeat or sequences derived from the promoter regions of certain human genes. Both XRC and NMR can provide insight into the folding dynamics of rG4s under different ionic conditions and their interaction with small molecule or protein ligands. Despite of the fundamental role of structural analysis tools in determining rG4 folding and their interaction with specific ligands, their use is often limited which makes modeling studies based on molecular dynamics (MD) simulations a valuable tool to predict the most probable topology of rG4s (Figure 2) and their interaction with ligands (51–53). While using the structure prediction tools, published homologous structures (based on XRC and NMR) are searched in the existing databases (sequences with same number of G-repeats separated by the same number of nucleotides to set a basis for the model, mismatched nucleotides in the model loops are then replaced by those from the desired sequence using one or more suitable modeling programs such as Biovia Discovery Studio (https://www.3ds.com/), SYBYL-X (Certara, USA), and NAB (reviewed in (54)). The built-in programming languages within these software packages can be modified to create scripts that allow one to carry out various structural and molecular rearrangements, docking, geometric transformations, and molecular mechanics optimization in force fields suitable for describing relevant interatomic interactions (55). To visualize models and MD calculations, VMD (http://www.ks.uiuc.edu/Research/vmd/) and USCF Chimera (https://www.cgl.ucsf.edu/chimera/) packages can also be used (51). In the absence of a homologue, assumptions based on the sequence analysis can be made about the possible topology of the G4 core, the number of G-quartets involved, the number of nucleotides in the loops, and the nature of loops. While simulating the G4 topologies using MD tools, the nature of the loop is considered a key parameter; for example, in a general sequence GxNyGxNyGxNyGx, $y = 1$ results in propeller loops, $y = 2$ causes deformation of the G-quartets if they have a diagonal and right lateral orientation, $y = 3$ generally results in lateral orientation (56,57). Since there is a lot more structural data available for dG4s, those can be cautiously used as base models for rG4 structure prediction using MD simulations since replacing thymine in loops with uracil can change the core topology (58,59). It should be noted that often alternative models can be built for a given sequence. In this case, the model with the lowest free energy would be the model of choice.To score the deformation of the structure of the quadruplex core, one can use the parameter system described by Tsvetkov *et al.* (60). Indeed, such approach has already been used to build models for nucleic acid secondary structures including G4s (61) and i-motifs (62,63). Figure 2 summarizes different models for diverse topologies of rG4s predicted based on their comparison with similar dG4 structures from the Protein Data Bank (PDB) and MD simulations. Despite of the flexibility in structure prediction, MD approaches also have their own limitations such as, force fields used for structure optimization and MD simulation only approximately describe interatomic interactions and examination of obtained model stabilities does not allow to sample the entire conformational space within reasonable simulation time (55,64). With the ongoing rapid advancement in *in silico* approaches, it is possible that modeling approaches will provide accurate and efficient structural information making such approaches key components of rG4 structural analysis and advancing rG4 targeting therapeutics.

G4 folding/unfolding likely follow multi-exponential kinetics involving various intermediates of which only a few finishes as G4s. Moreover, identical G-rich sequences can form different energetically analogous G4s, which exist in dynamic equilibrium with each other. Apart from that, G4-flanking nucleotide fragments may significantly affect these structural polymorphisms. We are far from a complete understanding of rG4 folding dynamics, nevertheless, a few studies shed light on peculiar folding kinetics (65) and folding-unfolding dynamics of rG4s. While dG4 folding is a complex multifaceted process that results in various topologies (66), the folding of rG4s has been mostly looked at as a relatively straightforward process that results in almost only all-parallel monomorphic topology (67). Nevertheless, the molecular understanding of the formation of anti-parallel hairpin like structures which result in an all parallel rG4 topology is challenging and largely overlooked. Using MD simulations in parallel G-hairpins with propeller loops, Havrila *et al.* demonstrated that a series of rearrangements of the H-bond interactions, starting from compacted anti-parallel hairpin-like structures to propeller loops could happen before reaching to the rG4 structure (68). Since the

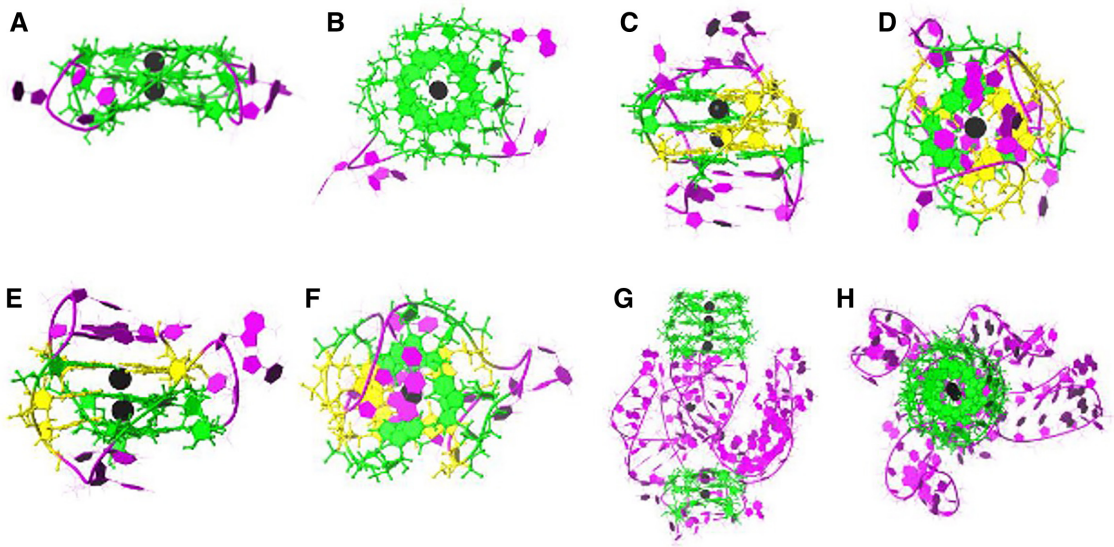

Figure 2. Models of rG4 types based on structural predictions. (**A**, **B**) A three-tier parallel rG4 modeled for 5′-GGGAGGGGCGGGUCUGGG, lateral and top view, respectively. (**C**, **D**) An anti-parallel rG4, modeled for 5′-UUAGGGUUAGGGUUAGGGUUAGGGUUA based on analogous dG4 PDB: 2mbj (208). (**E**, **F**) A proposed hybrid rG4, modeled for 5′-UAGGGUUAGGGUUAGGGUUAGGG based on analogous dG4 PDB: 2jsm (209). (**G**, **H**) A tetramolecular rG4, modeled for 5′-GGGGGUGUAGCUCAGUGGUAGAGCGCGUGC based on the sequence of 5′-tiRNA^{Ala} (details in Figure 6E). The nucleotides are colored as follows: Gs in syn conformation in quartets are yellow, Gs in anti-conformation in quartets are green, the remaining nucleotides in loops are magenta. K^+ ions are shown in black. Gs in quartets are depicted by 'ball and stick' model. The models were created using the method described in 'RNA G-quadruplex-topological diversity and dynamics'. The rG4 models shown in C, D and E, F have the same number of tetrads and loops, but their topology is different.

functions of many regulatory RNAs depend on how their 3D structure changes in response to a diverse array of cellular conditions (69), it is really important to understand the dynamic nature of rG4s to understand their molecular and cellular functions. In this respect, rG4 interactions with proteins, which may help a quadruplex to fold or destabilize it (such as RNA helicases), are of high biological significance.

EVOLUTIONARY DISTRIBUTION OF RNA G-QUADRUPLEXES

The evolutionary context of G4 distribution is mostly studied in the genome. The Balasubramanian lab recently analyzed genomes of 12 species ranging from bacteria to humans in the presence of a physiological K^+ concentration or a G4 stabilizing ligand to map pG4s (70). The authors found a striking observation in human, mouse and *Trypanosoma brucei*, where they detected a strong enrichment of G4s at gene promoters and in 5′UTRs (Untranslated Regions). Other eukaryotes like *Caenorhabditis elegans*, *Danio rerio* and *Drosophila melanogaster*, however, showed depletion of experimentally observed G4s at these and other (e.g. exons, 3′UTR) intragenic regions but showed an enrichment in ncRNA regions. The authors also reported that other species they studied (*Saccharomyces cerevisiae*, *Leishmania major* and *Plasmodium falciparum*) showed depletion of observed G4s at intragenic as well as ncRNA regions. The last group, *Rhodobacter sphaeroides*, *Escherichiae coli* and *Arabidopsis thaliana*, did not show enrichment or depletion of experimentally observed G4s at any genomic regions. Ding *et al.* studied the distribution of pG4s densities in the genomes of many organisms (71). Their analysis suggests that although the overall G4 distribution amongst species with a large genome is pretty consistent throughout the genome, there is a higher density of pG4s within the promoters and UTRs. The authors found that the organisms with large genomes ($>10^9$ bp; log(Mb) > 3) have similar GC% (36–44%) and possess pG4s with densities ranging from 0.1–0.5 per kb, among which 30–50% of the pG4s have four G tracts implying the possibility of formation of very stable G4s. The mouse and chicken genomes contain slightly more pG4s with >4 G tracts (50%). The human genome has a pG4 density of 0.23 per kb. In the large genome cohort, ∼45% of the pG4s were found to have more than four G tracts implying the favorable selectivity of the additional G tracts during evolution may maintain G4 folding in the event of damage to the principal G4 tracts. pG4s within the human genome and transcriptome are enriched at the transcription start site (TSS), the UTRs, and the 5′ end of the first intron, and depleted in coding regions. The coding regions of most genes are depleted of G4 motifs (22). Lee *et al.* recently integrated multiple sources of human genomic data and demonstrated that pG4s in 5′ and 3′ mRNA UTRs are selectively constrained, and enriched for *cis*-expression quantitative trait loci and RBP interactions, implicating rG4s in the mRNA UTRs as important cis-regulatory elements (72).

However, in organisms with a smaller genome there is greater variability in pG4 density, ranging from nearly ∼0–2.5 pG4s per kb, and greater variability in the percentage of pG4s with >4 G tracts, ranging between 20% and 50%. The low pG4s frequency in *E. coli*, yeast, and *A. thaliana* were previously reported (70). On the other hand, the genomes of the stress resistant tardigrade, as well as rice, a com-

mon plant species, have pG4s densities similar to that of large genome species (3 pG4s/kb). Some microorganisms that had >0.5 pG4s/kb include *Arabidopsis johnsonii* found in the human microbiome, the radiation resistant bacteria *Deinococcus radiodurans*, the thermophile *Thermus aquaticus*, and the single cell green algae *Chlamydomonas reinhardtii*. Among the analyzed species, the organisms that had the highest pG4s densities also possessed the highest GC% in their genomes (71). Among the 1627 bacterial genomes analyzed using the G4-search engine G4Hunter, Bartas *et al.* reported that the highest distribution of pG4s lies around the tRNA locus, inside the transfer mRNA, and rRNA locus (73). Although the pG4s were present in all the sequences analyzed, the density differed significantly across evolutionary groups (a range of 0.013 in *Lacinutrix venerupis* to 14.213 *Thermus oshimai JL-2* pG4s per kb). The highest frequency of pG4s was detected in the subgroup of Deinococcus-Thermus, and the lowest frequency in Thermotogae. The analysis revealed that the pG4s are non-randomly distributed, are favored in various evolutionary groups and are enriched mostly in ncRNA segments followed by mRNAs suggesting a unique and non-random localization of pG4s in bacterial genomes.

Lavezzo *et al.* analyzed the known genome data of all human infecting viruses and found that the occurrence and localization of pG4s are features characteristic of each virus class and family (74). The authors demonstrated that such sequences in both DNA and RNA viral genomes are orderly arranged such that two-thirds of viral classes can be assigned by just looking at the distribution of pG4s. pG4s were observed to be denser in single stranded genome viruses, and overall viral pG4 distribution was significantly higher than predicted based on the random distribution expected relative to GC percentage in the genome. In a separate study Biswas *et al*. showed that alpha-herpesviruses such as herpes simplex virus-1 (HSV-1) and varicella-zoster virus genomes had a higher G4-density than human and mouse genomes, indicating that G4 formation in these viruses could have a greater impact on gene regulation (75). Overall, rG4s are increasingly being recognized as important *cis*-regulatory elements across evolutionary tree.

MAPPING OF RNA G-QUADRUPLEXES *IN VIVO*

The majority of G4 prediction has been done computationally (reviewed in (9)), while validation tests have mostly been done *in vitro* using various biophysical (NMR (76), CD spectroscopy (77), UV spectroscopy (78), etc.) and biochemical (electrophoretic gel shift assay, enzymatic & chemical footprinting, polymerase stop assays, etc.) techniques (79), and *in cella* using G4 specific fluorescent ligands (80,81) or antibodies (82). Recent advances in high-throughput sequencing technology has allowed for transcriptome-wide identification of both dG4s (83) and rG4s *in cella* (84,85) under various conditions.

RNA G4s, like other stable secondary structures, can make reverse transcriptase stall during reverse transcription (RT) making it possible to track the presence of rG4s (Figure 3A) using the RT stop approach. Comparison of rG4-permissive conditions (e.g. K^+ or G4 stabilizing ligands such as PDS, carboxy-PDS, BRACO-19) with non-rG4-permissive conditions (like Li^+) allows us to map the position of rG4s detected as RT stops. In combination with next generation sequencing techniques, this approach has been recently used to map transcriptome-wide rG4 distribution (85,86).

RNA immunoprecipitation techniques such as individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation (iCLIP) (87) and PhotoActivatable Ribonucleoside-enhanced Cross-Linking and ImmunoPrecipitation (parCLIP) (88,89) have been applied to study transcriptome-wide binding sites of rG4 interacting proteins (Figure 3C). Although CLIP methods lack definitive proof of rG4 folding, they are indeed valuable to find the site that is potentially folded into a G4 *in vivo*. Another approach to study rG4s in cells is to use G4-specific antibodies in combination with fluorophore tagged secondary antibodies (Figure 3E). BG4 antibody was employed to visualize rG4s in the cytoplasm of fixed human cells. Elevated BG4 signal has been observed upon depletion of rG4 helicases, and during the S-phase of the cell-cycle. In addition, G4 specific small molecules (fluorescently tagged or those with intrinsic fluorescence) have been used to probe cellular rG4 structures (Figure 3F) (80,81). Derivatives of pyridostatin and PhenDC incubated with live cells have subsequently been conjugated to a fluorescence probe using ′click chemistry′ after fixation to visualize rG4s. Laguerre *et al.* used N-TASQ probe for the detection of rG4s in live cells. The authors showed N-TASQ has apparent high affinity for rG4s in cells since it cannot penetrate the nuclear membrane making it useful to visualize cytoplasmic rG4s (90). Ligands with intrinsic fluorescence, such as QUMA-1, were recently developed tools with a potential for live cell visualization of rG4s (91).

Application of reactive small molecules to discriminate among different RNA structures provides an alternative to mapping rG4s *in vitro* and *in cella*. Selective 2′-Hydroxylacylation Analyzed by Primer Extension (SHAPE) is a powerful tool in RNA structure mapping that has recently been applied to map rG4 structure formation based on their differential reactivity towards RNA nucleosides with or without the rG4 formation (Figure 3B) (92). *In vivo* DMS treatment in combination with *in vitro* rG4 folding can be used to map the position of rG4s (Figure 3D) in the cells. N7 of each guanine in a G-quartet is Hoogsteen hydrogen-bonded and protected from methylation by DMS which allows the formation of rG4s under permissive conditions *in vitro*. This, in combination with the RT stop assay, has been used as powerful tool to map rG4 folding *in vivo* (92).

Peroxidase proximity labeling with ascorbate peroxidase (APEX) is a relatively new, yet very powerful tool, to map the local proteome or transcriptome within a cellular compartment, based on the engineered peroxidase-mediated biotinylation of local biomolecules (93). While it remains to be seen whether an APEX-like system can be tuned to map cellular rG4s, the Sen lab showed that the intrinsic peroxidase like activity of a G4-hemin complex can be exploited to biotinylate G4s (Figure 3G) (94). Apparently, rG4s also can actively biotinylate themselves under these conditions. Under optimal condition, this method, in combination with next generation sequencing, has the potential to map *in vivo* rG4 dynamics.

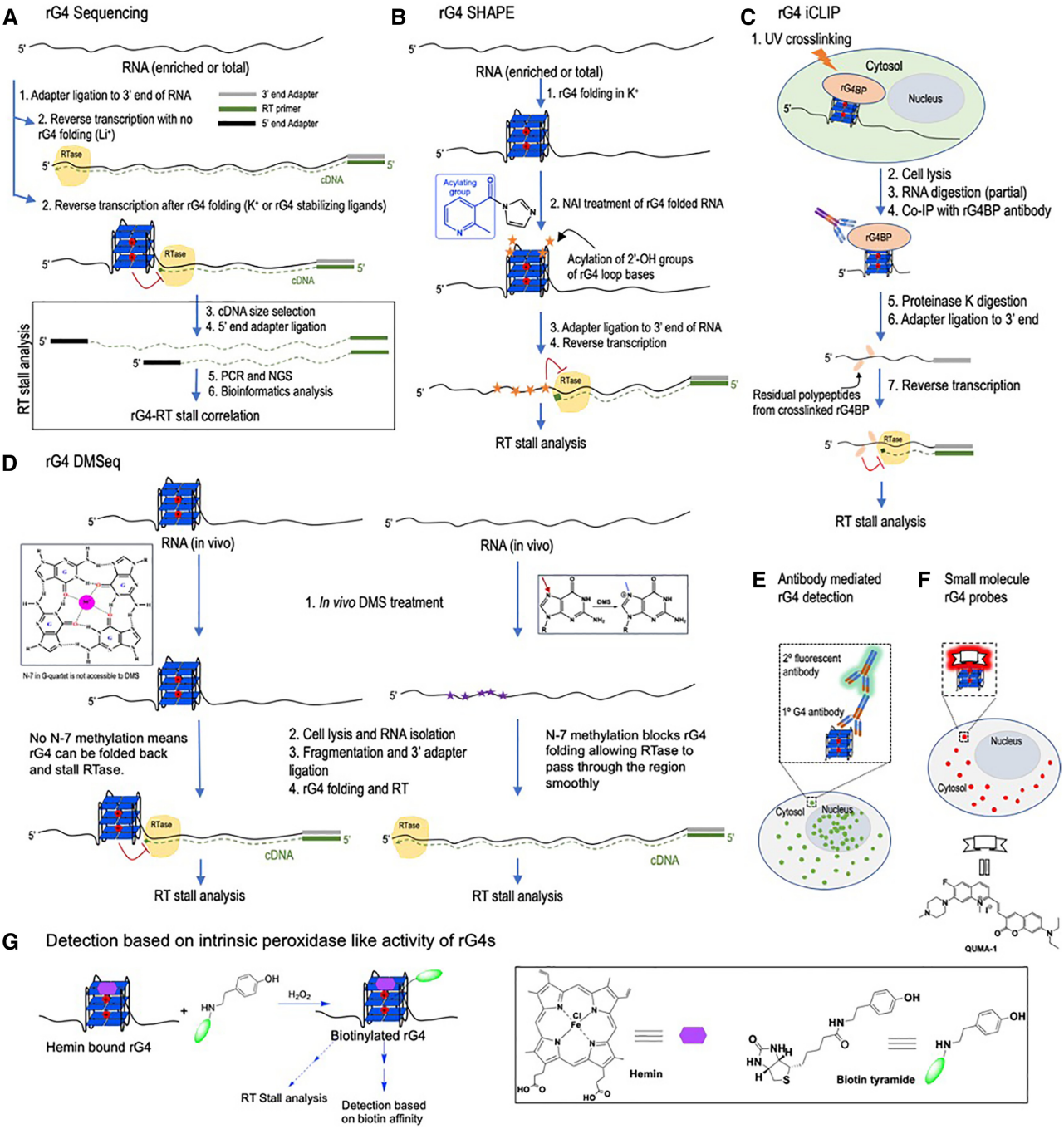

Figure 3. Recent developments in cellular rG4 mapping tools that have been used in tandem with high throughput sequencing technologies. (**A**) rG4 sequencing. (**B**) SHAPE probing in combination with other validation methods can be used to map the presence of rG4s. (**C**) iCLIP-RNA-seq can be used to map potential rG4BP binding regions within the transcriptome. (**D**) Differential DMS reactivity of N7 of Gs can be a valuable tool in mapping *in cella* rG4 formation. (**E**) G4 antibodies can detect rG4s in fixed cells. (**F**) An rG4 specific small molecule with intrinsic fluorescence properties. (**G**) Hemin-bound rG4 can show peroxidase like activity which can be exploited to biotinylate rG4s for their potential *in cella* detection.

RNA G-QUADRUPLEX FOLDING IS DYNAMIC IN THE CELLS

The physiological intracellular concentrations of free metal cations are in principle sufficient to maintain rG4 folding when compared to conditions typically used for *in vitro* studies. While rG4s are readily assembled into extremely stable structures in test tubes, even under temperatures exceeding the physiological tolerance, evidence for their existence *in vivo* remained uncertain and was only supported by few studies based on antibody based imaging (82), G4 ligands and fluorescent probes (90,95–97). Although these approaches are practical, they have several limitations including the possibility for rG4 folding to occur during preparation for microscopy analysis, cross-reactivity with other

molecules, or ligands shifting the equilibrium of unfolded G-rich sequences/partially folded structures towards fully folded rG4s or vice versa (stabilization versus destabilization). Moreover, these approaches are not capable to determine quantitatively the folding state of a specific G-rich region of interest.

Even though the details of *in vivo* rG4 formation are not understood in depth, based on the *in vitro* studies it is not difficult to speculate on a crucial role for rG4 folding-unfolding dynamics in myriad of cellular biochemical processes. rG4 structures may undergo protein-, small molecule- or ion-assisted folding, unfolding and refolding in a continuous fashion inside the cells. Under favorable ionic environment, some rG4 structures are thermodynamically more favorable than kinetically favorable secondary structures, for example hairpins (98). The dynamic switching between hairpin structures and rG4s has been proposed to regulate translation of rG4 containing mRNAs (see below). Using a prokaryotic system, Endoh and Sugimoto recently demonstrated that the functional role of thermodynamically favorable rG4s can be negatively affected by metastable kinetically favored structures demonstrating the importance of RNA folding dynamics in the regulation of gene expression (99).

Unexpectedly, one study brought a skepticism of *in vivo* occurrence of rG4s (92). Their DMS treatment- and SHAPE-based analysis of rG4s in poly(A)-containing transcriptome from mouse embryonic cells and yeasts, suggested that rG4s are globally unfolded in eukaryotic cells under steady-state conditions while they are readily detected in folded state *in vitro*. This would suggest two possible scenarios in which rG4s do not exist *in vivo* or they are actively unfolded/kept in an unfolded state by some cellular machinery. However, it is possible that the experimental conditions used in this study were not sensitive enough to capture the transient nature of rG4s *in vivo* (reviewed in (85)).

Recent data support the latter scenario and uphold the idea that rG4s exist and are dynamic in live cells. The dynamic behavior of rG4s is supported by data based on the selective and continuous tracking of rG4 folding and unfolding using the fluorescent probe QUMA-1 (Figure 3F) to visualize their dynamic nature in live cells (91).This was further endorsed by results based on capturing global snapshots of transiently folded rG4s at the transcriptome-wide level by a G4-specific ligand followed by target sequencing (86). The Monchaud lab used biotinylated Template-Assembled Synthetic G-quartets (BioTASQ) rG4 pull down followed by RT-stop mapping using RNA-seq approach to provide evidence in favor of transient existence of rG4s in mammalian cells (86,100). Moreover, in plants, *in vivo* RNA chemical structure profiling determined that hundreds of rG4s are strongly folded in rice and *Arabidopsis*, and their formation contributes to gene regulation and plant growth (101). Finally, loss-of-function studies on the 3′-5′ DEAH-box helicase DHX36, which possesses robust dG4/rG4 unfolding activity, results in the folding of rG4s in 5′ and 3′ UTRs of DHX36 mRNA targets. Consequently, rG4 formation on these mRNAs increase their stability with concurrent decrease in their translation efficiency (89).

Together with proteome-wide approaches aiming to identify bona fide RBPs (102) and RNA affinity approaches using rG4 sequences aiming to identify specific rG4-binding proteins (rG4BPs) (87,103,104,105), these studies corroborate the view that rG4s are dynamic *in vivo*, and their dynamics are largely dependent on interaction with specific protein factors.

RNA G-QUADRUPLEX INTERACTING PROTEINS

A plethora of *in vitro* studies and some *in cella* studies demonstrate the association of G4 sequences with DNA or RNA binding proteins (103,106,107). Although *in vitro* experiments indicate the interplay between G4 and non-G4 structures is largely controlled by their ionic environment (5), within cells the dynamics could be totally different, where proteins could solely dictate or contribute to the metal assisted G4 folding. The dynamics of rG4s *in vivo* is presumably largely controlled by rG4BPs directly or indirectly. G-quadruplex recognition by a protein is a multistep process involving main binding domains recognizing the G4 structure with the assistance of interactions from neighboring disordered regions (108). In some cases, previously unstructured regions of RBPs become ordered upon canonical RNA binding to stabilize G4-interacting conformations (109). Additionally, RNA G4s may directly or indirectly (via disruption of other secondary structures) impede protein binding and movement along RNA. An rG4BP can either stabilize, destabilize (unwind), or prevent the formation of rG4s (Figure 4A). Many of the reported rG4BPs were either suggested based on dG4 based studies or *in vitro* pull-down experiments. Recently, pull-down experiments using rG4 oligos with or without the G4 stabilizing ligands in the cellular environment have identified new protein players in rG4 biology (86,87). The analysis of reported rG4 interacting proteins reveal the presence of certain domains, motifs, or unstructured regions in the established or predicted binding regions of the rG4BPs (Supplementary file and Figure 4B-C). By virtue of their chemical nature, RRM (RNA-recognition motif) and RGG (Arginine-Glycine-Glycine) motifs within the proteins are mostly involved in the interaction with rG4s, and hence are the most commonly reported.

RGG domain containing rG4BPs

RGG domain (also called RGG/RG motif or GAR domain) collectively describes regions rich in RGG or RGX repeats. More than 1000 human nucleic acid binding proteins contain RGG/RG domains, which are crucial in the regulation of various physiological processes such as transcription, pre-mRNA splicing, DNA damage signaling, mRNA translation and the regulation of apoptosis (110). The arginine residue within the RGG domain not only mediates hydrogen bonding to nucleic acids, but also provides amino-aromatic interactions to build tertiary interactions necessary to position other domains (109,111). Interestingly, other positively charged amino acids (lysine and histidine) are not found within RGG/RG motifs, suggesting that the arginine residue confers distinct properties to the motif besides its positive charge. Glycine within the RGG domain enhances structural flexibility providing target speci-

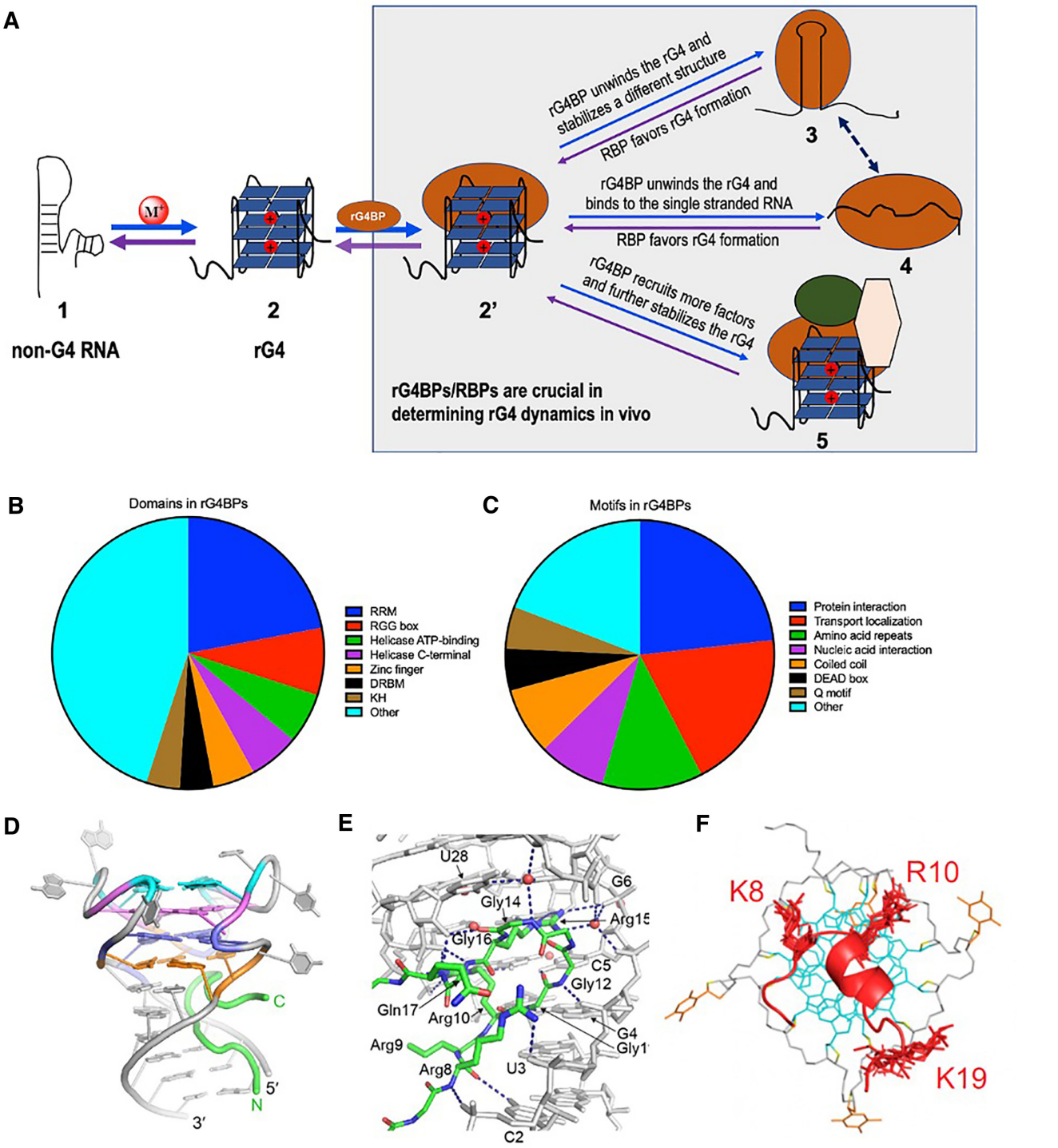

Figure 4. RNA G-quadruplex binding proteins are key in regulating rG4 dynamics *in vivo*. (**A**) Different possibilities of rG4–rG4BP interactions—unfolded RNA folds into an rG4 with the help of rG4BP (1,2) or the later can interact with pre-folded rG4, potentially resulting in one of the following consequences—stabilization of the rG4 or melting of rG4 to transit into an alternate structure (3) or a non-structured form (4), or recruitment of other binding factors to further stabilize the rG4 (5). (**B**, **C**) Classification of reported proteins composition based on their loosely defined domain and motif compositions (supplementary file available for the analyzed proteins). (**D**) Crystal structure of the complex between the FMRP RGG box (RGGGGR peptide) and sc1 RNA quadruplex-duplex junction, peptide is in green color and RNA is in gray color with G-quartets and the mixed tetrad in orange. (**E**) Hydrogen bonding pattern between peptide and nucleic acids (complex from D). (**F**) NMR solution structure of G4 binding segment of DHX36 with human telomeric parallel dG4, with detailed intermolecular interactions between peptide positively charged side-chains of K8, R10 and K19 and the DNA phosphate backbone. DHX36, red; guanines, cyan; thymines, orange; DNA backbone, gray; O4′ atoms, yellow. D–F are reproduced with permission from the National Academy of Sciences of the United States of America (114,118).

ficity (109). HNRNPs (Heterogenous Nuclear Ribonucleoproteins), FXR (Fragile X mental retardation syndrome-regulated) proteins, and DEAD box helicases are the most widely recognized rG4BPs which contain RGG or RGG like domain(s) (105,112–114). Interestingly, HNRNPH1 recognizes telomeric rG4 via its loop base, emphasizing the importance of loop sequence in the identity of rG4s (112). The C-terminal RGG-domain of HNRNPU was shown to bind telomeric DNA allowing for the formation of G4s (115).

It was initially reported that the RGG domain present in human fragile X mental retardation protein (FMRP) potently binds rG4s with high affinity in contrast to its paralogs, FXR1P and FXR2P, which also contain RGG motives (116). This makes FMRP unique in its ability to recognize G4s, also suggesting the FMRP RGG domain may play a unique, non-redundant, role in the pathophysiology of Fragile X mental retardation syndrome. Additionally, Menon *et al.* demonstrated using *in vitro* assays that an FMRP RGG-derived peptide at lower concentration helps stabilize rG4 of *MAP1 (*microtubule associated protein 1B) mRNA *in vitro*, yet at a higher peptide:RNA ratio unwinds rG4 structure (117). In separate studies, the FMRP RGG box has been shown to bind RNA duplex–quadruplex junctions, mixed tetrads, and duplex regions rather than G-quartets directly, by forming base-specific RNA contacts using arginines and glycines (113,118). A crystallographic study revealed specific recognition of a rG4 by a β-turn in the RGG motif of FMRP (Figure 4D-E) (105). Later, it was shown that in addition to FMRP, FXR1P binds rG4s with high affinity and specificity (119).

Many helicases of the DEAH (Asp-Glu-Ala-His) box family interact with G4s in an RGG dependent manner (23,114,120). Using a pull down/mass spectrometry approach, Herdy *et al.* demonstrated that 55% of the proteins associated with *NRAS* rG4 contain an RGG domain (87). The authors further validated that rG4 recognition by some of these proteins (DDX3X and DDX17) indeed requires RGG as evidenced by abrogation of their binding in protein mutants. Reversibly, it was shown that DDX3X interacting mRNA targets were significantly enriched in rG4s (87). Small-angle X-ray scattering and nuclear magnetic resonance spectroscopy revealed general principles for G4 specificity by an RNA helicase associated with AU-rich element (RHAU, DHX36), which preferentially binds parallel G4s both in DNA and RNA (121). Using a parallel dG4 binding to DHX36, it was shown that not only the Gs from the G-quartets but other bases outside the quartets as well involve in recognition and binding of the G4 to a particular peptide sequence (114). For example, the positively charged side-chains of Lysine 8 (K8), Arginine 10 (R10), and Lysine 19 (K19) are each located above a groove in G4-DHX36 complex, and electrostatic interactions are favored by the proximity between these residues and the phosphate groups of G16, G4 and G8/G12, respectively (Figure 4F). While the residues directly contacting the G-quartet core are among the most conserved residues across various species, these positively charged residues are conserved in higher organisms. Interestingly, Tippana *et al.* used DHX36-rG4 binding to demonstrate highly asymmetric folding-unfolding mechanism of rG4s (122). The authors showed that, unlike in dG4s, DHX36 displays an ATP-independent unfolding of rG4 followed by an ATP-dependent refolding.

Another DEAD box helicase, translation initiation factor eIF4A is suggested to unwind rG4 structures in mRNA 5′-UTRs encoding many transcription factors, epigenetic regulators and key oncogenes (123). Waldron *et al.*, however, showed that eIF4A also alleviates the translational repression mediated by classical secondary structures, even more so than by G4s in the 5′ UTRs of eIF4A-sensitive mRNAs bearing (G2C)4 stretches (124). Gao *et al.* used a dG4 pull down to identify G4-interacting proteins in yeast and demonstrated that some of the DEAD-box RNA helicases (Dbp2, Ded1 and Mss11) can bind to both dG4s and rG4s but can only unwind rG4s (125).

In some instances tyrosine can replace arginine to form a YGG box which could use a similar mechanism of RNA binding as RGG boxes (102). Recently a bioinformatic analysis identified a common denominator for several dG4-binding proteins, called Novel Interesting Quadruplex Interaction Motif (NIQI) (126). Similar to the RG-rich domain of FRM1 (RRGDGRRRGGGGRGQGGRGRGG GFKG), which specifically binds and stabilizes G4 structure, the NIQI motif (RGRGRGRGGGSGGSGGRGRG) mainly harbors disorder-promoting residues. Thus, presence of NIQI motifs suggests the importance of intrinsically disordered regions among G4BPs, which enable hydrogen bonding interactions with a wide portfolio of G4s, suggesting a NIQI motif search could be useful in searching for novel G4BPs in general, including rG4BPs. Altogether, this indicates a wide variation in the composition and selectivity of RGG and RGG-like domains in rG4 recognition. Adaptability of the RGG motif could be critical for high-affinity recognition of rG4s in mRNAs, for instance, as arginine methylation is frequently observed in RGG boxes (110).

RRM domain containing rG4BPs

The RNA recognition motif (RRM) domain (also called RNP-type RNA-binding domains or ribonucleoprotein motif) mostly recognizes ssRNA during RNA processing and transport. The RRM is the most common RNA-binding domain of 75–85 amino acids containing two highly conserved short sequence motifs known as RNP1 and RNP2. It has a typical βαββαβ topology that forms a four-stranded β-sheet packed against two α-helices with RNP1/RNP2 located in the two middle β-strands (127). RRMs of the HNRNP family of proteins show high affinity for single stranded guanine repeats tracts. Given their plasticity, HNRNPs interact with structured RNAs as well, promoting tremendous functional diversity (127). Different recognition modes enable HNRNPF and HNRNPH to bind both G-rich oligos and rG4s. HNRNPF contains two glycine-rich regions and three quasi-RNA recognition motifs (qRRMs), which, unlike canonical RRMs, lack aromatic and basic residues in RNP1 and RNP2 motifs. Loop residues of qRRMs recognize and maintain G-tract binding in ssRNA conformation, and in addition, HNRNPF shows some specificity towards rG4 recognition (128). Liu and Xu reported that HNRNPA1 binds to the loop of the telomeric rG4 by recognizing the base of the loop's nucleotides pro-

viding evidence that the nucleotide composition could be a key factor for HNRNPA1 specifically recognizing the rG4s (112).

Zinc finger (ZnFs) domain containing rG4BPs

ZnFs are typically ∼30 amino acid long ββα topology domains, in which certain cysteine and histidine residues coordinate metal ions, usually zinc, to stabilize a rigid fold, which rarely undergoes conformational change upon target binding (129). Certain ZnF proteins, which are usually abundant in glycine and arginine, are G4-binding proteins. Cellular nucleic acid binding protein (CNBP/ZNF9), which harbors seven CCHC-type ZnF domains in addition to one RGG motif, preferentially binds to the G-rich regions of target mRNAs and enhances their translation (130). Both ZnF-domains and the RGG domain are required for high-affinity binding of CNBP to G-rich RNA sequences. CNBP binding to pG4s prevents rG4 formation *in vitro* suggesting that CNBP supports translation by resolving stable rG4 structures in mRNAs. Lin-28 homolog A (Lin28) contains two ZnF motifs as well as a cold-shock domain (CSD) which together form critical contacts with guanine and adenine bases in distinct G-rich RNA regions (131). Similarly, ZnF and RRM domains in splicing factor LARK and its homologue RBM4 specifically bind G4s in the promoter regions and may regulate transcription (132).

Recently, a significant effort has been put into identifying the rG4 interactome. For example, rG4 proteome pull down by different labs (87,103,104) have identified novel rG4BPs, implying a wider significance (post-transcriptional regulation, translation regulation, splicing regulation, etc.) of rG4-proteome dynamics. Independently, using a SILAC (stable isotope labeling by amino acids in cell culture)-mass spectrometry approach, Serikawa *et al.* identified a broad range of rG4 interacting proteins which showed a diverse pattern depending upon the rG4s they used for the pull-down (133). The majority of the candidates they found belong to the HNRNP family, discussed above. While many of them recognize rG4s directly, the interplay of HNRNPH/F with other RNA-binding proteins, such as RNA helicases DDX5, DDX17 and DHX36, enhances their affinity for rG4s (105,134). Similarly, DDX5 and DDX17 associate with RNA binding protein (RBM) 4 and 14 which are known to interact with rG4s. Other rG4BPs with RRM domains are splicing proteins, including serine/arginine-rich splicing factors (SRSF) 1 and 9 (104).

As the number of studies aiming to identify the rG4 proteome increases, we should focus on identifying the specific rG4 interactome since in the studies using wild type rG4 versus non-rG4 pull-down approach, results could still be questionable. While the field is growing, we so far understand that there is no common motif or structural element in the identified proteins that broadly defines them as rG4BPs, rather it is the presence of polar residues with inherent flexible randomness in the peptide composition that makes a protein suitable to regulate dynamic rG4 structures. Common approaches to study rG4–protein interactions use pull-down experiments using mutant oligos or G4-non-permissive salt conditions as negative controls, which can unexpectedly bias the end results because in one case it is possible to get a different RNA structure (in RNA mutants) while the other case can result in a different chemical environment where proteins behave differently (under non-physiological salt conditions). An alternative approach to reduce these biases is to use control oligos with chemical variations that only abolish the possibility of G4 formation while keeping rest of the chemical environment same (e.g. using 7-deazaguanine containing oligos). This approach helps distinguishing *bona fide* G4 binding proteins from the G-rich sequence binding proteins (135,136).

PROPOSED FUNCTIONS OF RNA G-QUADRUPLEXES IN REGULATION OF GENE EXPRESSION

As pG4s are commonly found in regulatory regions of mRNAs such as 3′- and 5′-UTRs or in introns of pre-mRNAs, it immediately hints at the potential impact of rG4s on RNA metabolism. Consequently, the number of proposed roles played by rG4s in different aspects of gene regulation is constantly increasing. Although they are implicated in various processes ranging from genome stability to RNA biogenesis, transport, and regulation of transcription and translation, it should also be noted that many of these implications are bases on *in vitro* studies awaiting further examination *in vivo*.

RNA G-quadruplexes in RNA transcription and co-transcriptional processing

The nascent RNA strand can base pair with its template DNA during transcription to form a structure called R-loop (137). G-rich nascent RNA with two or more G-tracts can base pair with the non-template DNA strand to form RNA:DNA hybrid G4s in the R-loop structure (Figure 5A) (138,139). Regions downstream of transcription start sites are enriched with pG4s, with an average of 73 putative hybrid G4s per gene (139,140). Formation of hybrid G4s in R-loops was confirmed using T7 RNA polymerase in *in vitro* transcription (141), where R-loop formation was shown to inhibit transcription. *In vitro* crosslinking, site-specific mutagenesis, and luciferase reporter assays were also used to show transcriptional blockage by R-loop G4 formation (142). A recent study showed that R-loops and the hybrid G4s can be formed post-transcriptionally in the mouse immunoglobulin heavy chain (IHC) locus (143). The authors demonstrated a unique manner of R-loop formation where the rG4 in the spliced introns are resolved by DHX1 to allow them to form hybrid G4s in the switch region of the DNA locus, thereby promoting the class switch recombination in the mouse IHC locus (Figure 5B). This interesting approach of gene regulation could be paused to stabilize the spliced intronic rG4s to modulate the DDX1 and AMD mediated IgH class switch recombination. The nascent RNA and non-template DNA strand of mitochondrial *CSB II* can co-transcriptionally form a stable DNA–RNA hybrid G4, which was suggested to promote transcription termination (144). DNA:RNA hybrid G4s could play a role in transcription termination as pG4s are proposed terminator sequences that cause RNA Polymerase II transcription to pause (145,146). Particularly, R-loops formed behind elongating polymerase II are prevalent over G-rich pause sites

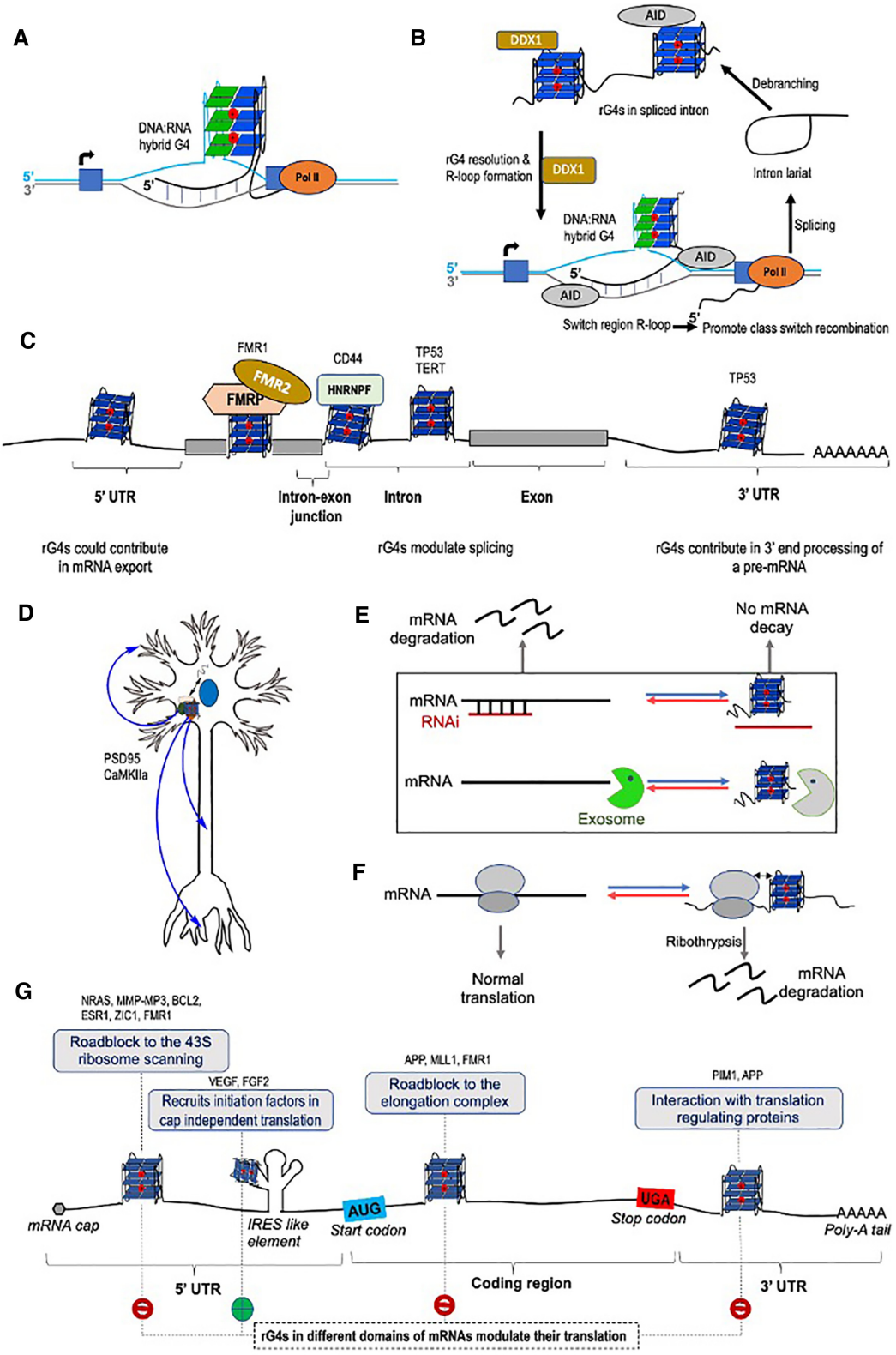

Figure 5. rG4s are implicated in different layers of gene regulation. (**A**) Proposed transcriptional regulation model via a DNA:RNA hybrid G4 formation in the R-loops. (**B**) The RNA Helicase DDX1 Converts rG4 into R-Loops to Promote IgH Class Switch Recombination. (**C**) rG4s regulate mRNA maturation. (**D**) rG4s regulate mRNA transport. (**E**, **F**) rG4s regulate mRNA stability. (**G**) rG4s modulate mRNA translation.

positioned downstream of poly(A) signals and are capable of G4 formation. Senataxin (SETX) helicase plays a key role in this mechanism by resolving R-loops allowing 5′→3′ exonuclease Xrn2 access to the 3′ cleavage poly(A) sites causing nascent RNA release, 3′ cleavage product degradation and finally polymerase termination. The depletion of SETX causes R-loops to be stabilized downstream of poly(A) signals, preventing efficient pause-mediated termination.

RNA G-quadruplexes regulate mRNA maturation, transport, and translation

Although RNA pG4s are more abundant in UTRs, they are also present in the intronic (especially in the first intron) and exonic regions of pre-mRNAs and the coding region of mature mRNAs (reviewed in (26–28)). rG4s present in pre-mRNA introns or exons can enhance, inhibit or alternate their splicing via recruitment of specific rG4BPs or steric interference with regulatory elements in the vicinity of the rG4 (Figure 5C) (reviewed in (26,27). For example, two exon-located rG4s within *FMRP* pre-mRNA act a potent splicing enhancer (147). Interestingly, while an intronic rG4 in the human telomerase (*hTERT*) pre-mRNA works as a splicing silencer (148), an intronic rG4 in *TP53* pre-mRNA acts as a splicing enhancer (149). Lastly, rG4s regulate alternative splicing, partially mediated via recruitment of HNRNPH and HNRNPF. A recent study showed that an rG4-binding ligand results in thousands of alternative splicing events in cells (150). rG4s in the 3′ UTR of pre-mRNA are speculated to affect mRNA maturation by modulating their 3′ end processing (151,152). Zarudnaya *et al.* proposed that such pG4s are abundant in the transcriptome and can act as auxiliary downstream elements (DSE) that can help modulate alternate adenylation (151). It was later shown that a unique rG4 is formed within the DSE region of the UTR in p53 pre-mRNA and is recognized by HNRNP F/H to allow 3′ end formation of p53 mRNAs after DNA damage (153).

rG4s can be crucial in mRNA localization (154). rG4s in the 3′ UTR of many neuronal mRNAs (for example, *APP*, *BDNF*, *GLRA1*, *PSD95*, and *αCaMKII*) facilitate their dendritic localization (Figure 5D). The deletion of rG4 sequence from *PSD95* and *αCaMKII* mRNAs resulted in the loss of dendritic translocation of mRNAs and also the loss of neurite signals (154). Similarly, TDP-43 has been shown to bind rG4 mRNAs to transport them for local translation within the neurons (155). In addition, rG4s present in mRNA 3′ UTR could play a role in the regulation of RNA interference by regulating miRNA binding (Figure 5E) (156). Ibrahim *et al*. recently proposed a novel mechanism for canonical mRNA decay mediated by deadenylation and decapping followed by exonucleolysis from the 3′ and 5′ ends. The authors showed that human mRNAs are subject to repeated, cotranslational, ribosome-phased, endonucleolytic cuts at the exit site of the mRNA ribosome channel, in a process called ribothrypsis, and proposed rG4s as potential ribothrypsis triggers (Figure 5F) (157).

rG4s within mature mRNAs mostly inhibit their translation, although some reports suggest stimulatory roles of rG4s on translation (reviewed in (26,27) and Figure 5G). A number of cell-free translation and cell-based reporter assays showed that rG4s in the 5′ UTR of mRNAs cause reduction in the efficiency of their translation (reviewed in (26,27)). rG4 density, thermodynamic stability and position relative to the 5′ cap have all been shown to differentially influence translation (158). rG4s are often located near the beginning of 5′ UTRs, suggesting a role in translation initiation. Depletion or inhibition of eukaryotic initiation factor 4A (eIF4A), a helicase that unwinds RNA secondary structures and facilitates the recruitment of the 43S preinitiation complex, reduced the translation efficiency of mRNAs, indicating that rG4s directly influence recruitment of, or scanning by, the ribosome (reviewed in (159)). On the other hand, unresolved rG4s in 5′ UTRs can promote the formation of 80S ribosomes on alternative, upstream start codons, thus inhibiting the translation of the main open reading frame (160). rG4s in *FGF2*, *α-Syn* and *VEGF* mRNAs are proposed to stimulate translation as a part of an internal ribosome entry site (IRES) or IRES-like elements, potentially by helping recruit the 40S ribosome (Reviewed in (159)). However, the nature of cellular IRESes and mechanisms of their activity are currently debated (159,161), and thus the role of rG4s in IRES-dependent translation is far from clear.

rG4s in the coding regions have much lower abundance than in UTRs (162), and when present act as translational repressors. For example, rG4 within the ORF of *APP* mRNA inhibits its translation via association with FMRP, a known translational silencer (163). It is postulated that ORF-located rG4s, due to their extreme stability, may obstruct elongating ribosomes and, consequently, cause their stalling and/or disassociation as in the case of the virally encoded EBNA1 transcript (164), or stimulate frameshifting as shown in studies using mRNA reporters (165). Nonetheless, some ORF-situated rG4s were also described to stimulate translation. One such rG4 is located in *MLL1/4* mRNA, which is recognized by the RGG-containing factor AVEN in complex with helicase DHX36 (166). Although the molecular mechanism is unknown, it is possible that DHX36 stimulates *MLL1/4* mRNA translation via its rG4-resolving activity, thus removing roadblocks for elongating ribosomes (Figure 5G).

Although less studied, rG4s in the 3′ UTR of mRNA are shown to inhibit translation (e.g. *PIM1, APP*) without apparent effect on mRNA stability (167). Although mechanisms of such inhibition are not known, rG4s may attract RBPs that act as translational repressors (Figure 5G). Beaudoin and Perreault showed that 3′-UTR mRNA rG4s increase the efficiency of alternative polyadenylation, leading to the expression of shorter transcripts (167). The authors also showed that alternatively polyadenylated transcripts possess the ability to interfere with the miRNA regulatory network of a specific mRNA (167).

RNA G-quadruplexes regulate non-coding RNA biogenesis and functions

Noncoding RNAs not only regulate gene expression at the levels of transcription, RNA processing, and translation, they protect genomes from foreign nucleic acids and guide DNA synthesis or genome rearrangement (168). Many ncRNAs exploit their base pairing to achieve their function

while others use their structural features or act as RNA–protein complexes. As in case of mRNA, selected ncRNA areas are also enriched with pG4 sequences (27). Except relatively well studied *TERRA* rG4 (29,40) (Figure 6A), the presence and functions of rG4s in ncRNAs have only recently gained attention. An increasing number of reports suggest the existence of putative rG4s in diverse classes of ncRNAs, including lncRNAs (169), rRNAs (33) (Figure 6B), miRNAs (31) (Figure 6C), piRNAs (170,171) (Figure 6D) and tRNA-derived RNAs (135) (Figure 6E), although functional consequences for their presence in these RNAs are largely unknown.

TERRA RNA is the main lncRNA component of the telomere that has been shown to contribute in telomere homeostasis via its rG4 dependent interaction with telomeric protein TRF2 (172). The interactions between TRF2, TERRA rG4 and telomeric dG4 could occur simultaneously (Figure 6A) making this interaction a key in regulating telomere homeostasis. Another telomeric lncRNA TERC can also form rG4 structure and it was reported that G4 helicase RHAU is associated with TERC *in vivo* suggesting the dynamic involvement of TERC rG4 in an active telomerase holoenzyme (173). Although the pG4s are predicted widely in other lncRNAs (174), relatively smaller progress has been made to identify their presence and function. Nevertheless, GSEC lncRNA was shown to bind DHX36 helicase via its G4 forming sequence and modulates colon cancer cell migration (169). With the amount of diversity both in distribution and function of lncRNAs, we speculate that much more effort will be put to identify such targets and their physiological and pathophysiological roles.

In miRNAs and piRNAs, rG4s have been shown to regulate biogenesis of these ncRNAs in addition to impeding their target recognition (Figure 6C-D) (31,170,171). The formation of an rG4 instead of the canonical stem-loop in certain precursor microRNAs (both pri- and pre-) can inhibit their maturation by Drosha or Dicer (30,31,175). Consequently, rG4-mediated regulation allows for manipulation of the production of mature microRNAs. rG4s' formation within mature microRNA or their target mRNA sequences can subsequently alter the regulation of the target mRNA. rG4s in precursor piRNA clusters have been shown to regulate piRNA biogenesis (170). The interaction of RNA helicase MOV10L1 with rG4s is necessary for the subsequent endolytic cleavage of precursors to produce mature piRNA (170). Multiple retrotransposons in the human genome harbor pG4s, which may promote their transposition as observed in long interspersed element 1 (176). The involvement of rG4s in histone modification was recently implicated in a CLIP approach experiment that showed Polycomb repressive complex 2 (PRC2) binding to rG4s and catalyzing the gene-repressive histone modification H3K27me3 (177).

In response to stress, transfer RNAs are cleaved in their anticodon loops by the ribonuclease angiogenin to produce tRNA halves called tRNA-derived stress-induced RNAs (tiRNAs), a ncRNA class that participates in translational control under stress (reviewed in (178,179)). Selected tiRNAs derived from the 5′halves of $tRNA^{Ala}$ and $tRNA^{Cys}$ (5-$tiRNA^{Ala/Cys}$) assemble intermolecular rG4s (135). These tetramolecular rG4–tiRNAs target protein synthesis by inhibition of cap-dependent translation via direct interaction with the translation initiation factor eIF4G (Figure 6E) (180). As a consequence of translation inhibition, cells promote the formation of Stress Granules (SGs), membraneless cytoplasmic foci with pro-survival functions (181,182).

RNA G-QUADRUPLEXES IN BIOMOLECULAR CONDENSATES AND THEIR CONNECTION TO DISEASES

The earliest description of the ability of guanosine monophosphate and its derivatives to self-assemble into gels upon reaching certain concentrations is one of the most prominent features of G-rich nucleic acids (1). Such gelation represents connectivity transition that involves a gradual increase in interactions between monomeric units and culminates in the formation of crosslinked interaction networks. Gelation is related to liquid–liquid phase separation (LLPS), a fundamental physical phenomenon describing a density transition of an initially homogenous solution into a dense and a dilute liquid phase that can stably coexist. This phenomenon underlies many biological processes including the formation of membraneless compartments (reviewed in (183)), notably RNA-containing subcellular entities (called RNA granules) such as the nucleolus in the nucleus (184) and SGs in the cytoplasm (181).

Structurally distinct RNAs are capable of undergoing phase separation themselves or together with interacting proteins as a part of ribonucleoprotein (RNP) complexes (185). During phase separation RNPs interact with other complexes, RNAs and proteins, via multiple weak interactions to promote high local RNA:protein concentrations that trigger the formation of RNA granules. These RNA granules are dynamic and constantly exchange their content (proteins, RNAs) with the surrounding environment. Interestingly, many rG4-binding proteins (e.g. RNA helicases, FXR proteins, HNRNPs, etc.) are also associated with SGs, specific cytoplasmic foci that assemble in cells in response to stress. Different RNA features contribute to their ability to assist in or promote biomolecular condensation via LLPS, e.g. their GC-richness (186). With their intrinsic ability to undergo gelation at high local concentrations, dynamic nature *in vivo* and promiscuity to interact with multiple RBPs, these RNAs are excellent candidates to contribute to biomolecular condensation. We have recently shown that functional tetramolecular RG4s assembled from 5′tiRNAs (135) actively promote the formation of SGs by (1) inhibition of mRNA translation, thus converting translationally-arrested mRNAs into substrates for LLPS (180,187), and by (2) building molecular scaffolds to attract rG4BPs into their proximity to nucleate SG formation (187).

Interestingly, specific pathological conditions (especially neurological disorders) are associated with patient mutations in which repeats of nucleotides (tri-, tetra-, hexa- nucleotides) increase in copy number in comparison to healthy individuals (188). Such expansions can occur in both coding and non-coding parts of genes, which when transcribed, produce repeat-containing RNAs that are commonly associated with cytotoxicity and the assembly of RNA foci. Mechanistically, these RNA foci are built by RNA repeats,

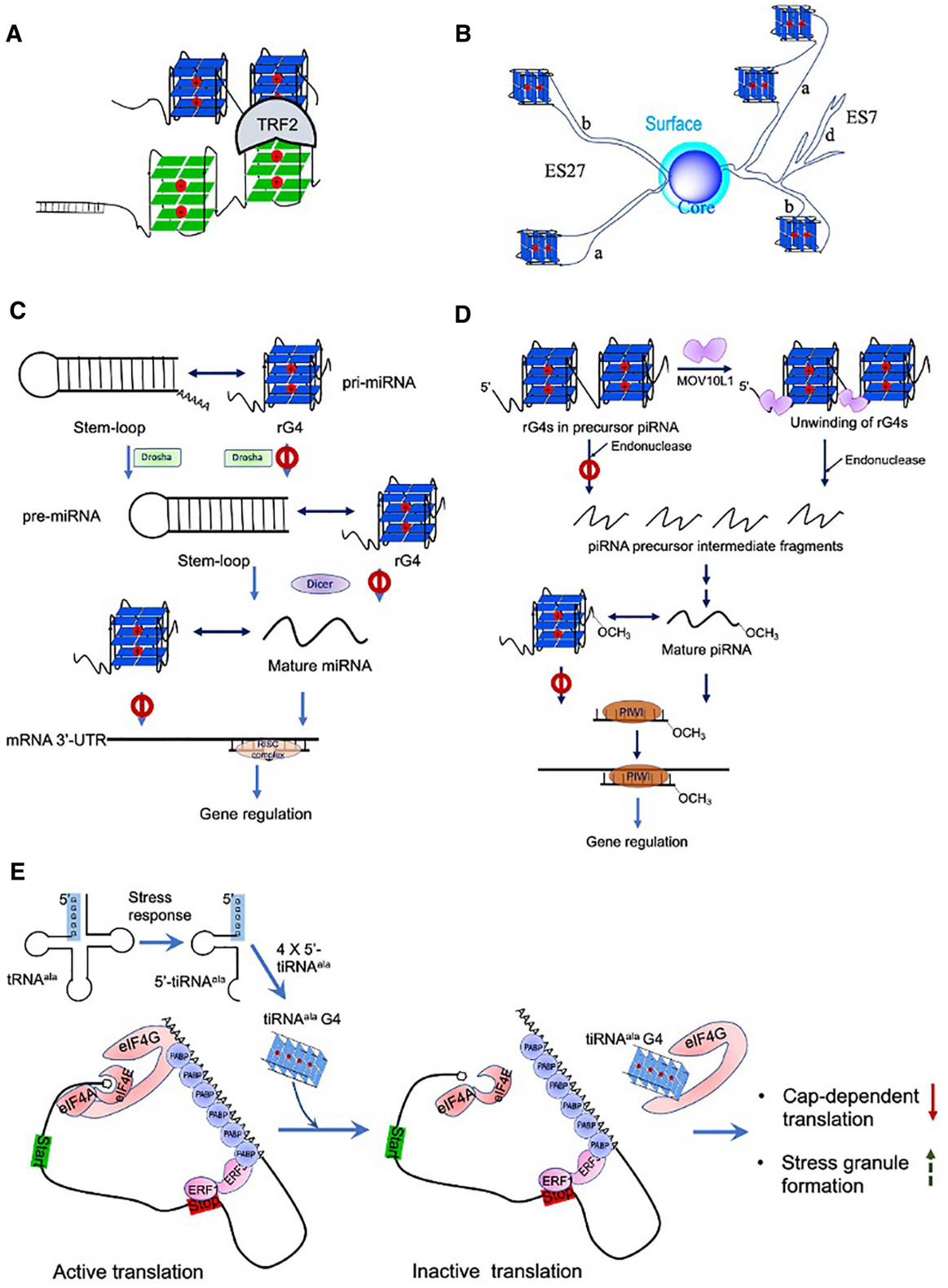

Figure 6. rG4s influence ncRNA biology. (**A**) TRF2 interacts with both TERRA rG4s and telomere end-situated dG4s to influence telomere biology. (**B**) rG4s are present in different expansion segments of human rRNA (ES7 and ES27 are indicated). (**C**) rG4s modulate miRNA biogenesis and their target recognition. (**D**) rG4s affect piRNA biogenesis and their target recognition. (**E**) rG4s formed by 5′tiRNA^{ala} in response to cellular stress inhibit cap-dependent mRNA translation.

which act as a template to attract and sequester specific RBPs. One of such characteristic repeats is the intronic GGGGCC (G4C2) hexanucleotide repeat in the C9ORF72 gene that is the most common cause of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia (ALS/FTD), neurodegenerative diseases with no cure (189,190). r(G4C2) repeats readily assemble into hairpin and G4 structures *in vitro* (191,192). Our work revealed that rG4C2 repeats promote RNA condensation and assembly of RNPs in lysates (193). In cells, r(G4C2) assemble cytoplasmic foci, which resemble SGs both compositionally and dynamically, and nuclear foci, which are a hallmark of ALS/FTD patient cells with C9ORF72 mutations (193). Importantly, the formation of r(G4C2)-mediated foci is mainly dependent on G4 structures of the repeat, where repetitive pattern of rG4s drives r(G4C2) condensation, interaction with RBPs and phase separation. The r(G4C2) repeat length is directly proportional to the degree of r(G4C2)-mediated condensation and foci formation. This is important because r(G4C2) repeats exceeding a threshold length also correlate with ALS/FTD pathogenesis (189).These studies suggest that rG4 targeting of r(G4C2) repeats by rG4-ligands may benefit ALS/FTD patients. This hypothesis awaits further examination.

RNA G-QUADRUPLEXES AS THERAPEUTIC TARGETS

As we understand the wider role of RNA structural elements in different molecular, cellular and pathophysiological levels, it is possible that they are potential targets to modulate gene expression in the context of disease (194). Indeed, a significant effort has already been made to develop therapeutics that target rG4s using different mechanisms. The common approaches involve small molecule therapeutics that recognize rG4s based upon their planar interaction with the G-quartets or groove binding or loop recognition ability, and oligonucleotides (or their analogs) that complementarily bind directly to the rG4s or adjacent to them to modulate their formation (Reviewed in (27)).

First report of using rG4 stabilizing ligands to target mRNA 5′UTR rG4s *in cella* came from the Hartig lab almost a decade ago, where the authors used luciferase reporter assays to demonstrate the dose dependent translation inhibition of TRF2 translation upon treatment with G4 stabilizing bisquinolinium compounds (360A, PhenDC3, and PhenDC6) (195). Thereafter, many labs have used numerous small molecules to show their impact in stabilizing or destabilizing rG4s in mostly *in vitro* and using reporter systems *in cella*. Miglietta *et al.* used 4,11-bis(2-aminoethylamino)anthra[2,3-*b*]furan-5,10-dione and its analog to target 5′ UTR KRAS rG4 (196). The authors showed KRAS down-regulation induces apoptosis together with a dramatic reduction of cell growth and colony formation in pancreatic cancer cells. Simone *et al.* demonstrated that small molecules that target r(G4C2) repeats can ameliorate C9orf72 ALS/FTD pathology suggesting a therapeutic potential of rG4 targeting in ALS/FTD cure (197). Recently, using adult neural stem cell and mouse derived neural progenitor cells, it was shown that while both PDS and carboxy-PDS (cPDS) both reduce cell proliferation, the mechanism of their action is different (198). As reported earlier, PDS mediated blockage of cell proliferation is achieved by increased DNA damage while rG4 interacting cPDS could diminish cell proliferation through a mechanism that promotes cell cycle exit and the production of oligodendrocyte progenitors. Small molecules like TMPyP4, NMM, BRACO-19, MMQ1, NiSal-3im (199), RHPS4, etc. bind to both dG4s and rG4s, while others like QUMA-1 (91) and carboxyPDS (82) have a higher affinity towards rG4s, and small molecules such as NMM selectively bind parallel G4s (reviewed in (200), making them good candidates to recognize rG4s.

In addition, rG4 targeting complementary oligonucleotide treatment is getting attention to modulate rG4 dynamics. The Basu lab used a locked nucleic acid modified oligo that targets pre-miRNA-92b to lock it in a rG4 containing non-canonical form, as a mean to stop Dicer mediated miRNA-92b maturation and upregulate PTEN expression in non-small cell lung carcinoma cells (201). With the advancement in delivery mechanisms in recent years, such strategy could find stronger practical ground as more selective rG4 targeting approach.

Quite recently an rG4 targeting aptamer was developed against human TERC RNA, binding of which prevents native rG4-protein interactions between TERC and nucleolin (202). Besides targeting the rG4s in cellular transcriptome, engineered induced rG4s (203) and those present in synthetic aptamers, such as AIR3 aptamer that targets interleukin-6 receptor (204) add newer dimension in the therapeutic uses of rG4s. In addition, several RNA aptamers that are evolved against various small molecule dyes such as Mango and Spinach aptamers have one or more G-quartets or G4s embedded in them as a key element that recognizes planar small molecules (205,206).

CONCLUSIONS, PERSPECTIVES AND FUTURE DIRECTIONS

For many decades, formation of rG4s has been merely considered an *in vitro* phenomenon with limited impact on biology. However, since rG4s are evolutionarily conserved (at least throughout eukaryotes), mechanisms that regulate their formation in live cells have also evolved, otherwise there would be a selection pressure to remove them if they pose any potential problem. Recent advances in the development of highly sensitive genome-, transcriptome- and proteome-wide approaches allowing for the identification of new RNA regulatory motifs and RBP-binding sites *in vivo* have boosted research in many fields of RNA biology, including the rG4 field. Consequently, we are currently witnessing an incredible explosion of studies that shed light on molecular details of rG4 dynamics and association with proteins in live cells. Which specific roles (if any) do these newly identified rG4s play in cellular physiology? Are there specific cellular conditions that change rG4 dynamics and signaling pathways that regulate rG4-mediated functions? Although rG4s are still difficult to study *in vivo*, we are positioned to answer these and other questions in near future.

Another important aspect of future rG4 research is structural. The number of solved rG4 structures is lagging behind their DNA counterparts and is significantly lower than

the number of *in silico* predicted pG4s in the transcriptome. As evidence continues growing that rG4s contribute to multiple biological processes, understanding of the structure-function relationship between specific rG4 and their ligands is important. While high-resolution structures remain a golden standard for studying RNA molecules, the development and application of computer-based approaches will certainly aid in predicting potential rG4 structures, allowing for rapid analysis of rG4:ligand interactions.

An intriguing question is how the intrinsic property of rG4s to self-assemble *in vitro* contribute to the ability of rG4s to undergo phase transitions and phase separations in the context of complex macromolecular complexes such as RNA granules. Which identities of rG4s (sequence itself, number of tetrads, length and type of loops) determine the biophysical properties that grant their participation in supramolecular biological systems through non-covalent association with other RNAs, proteins and RNA-protein complexes? What are the consequences of the altered rG4 dynamics in biomolecular condensation? How feasible is their application in chemical biology and nanotechnology (207)?

Finally, an emerging theme in the rG4 field is its potential relevance to human health and disease. The presence of rG4s in the important regulatory regions of mRNA, amplification of sequences able to assemble rG4s in transcripts of disease-associated genes and their appearance within the non-coding transcriptome is well documented. Careful assessment of rG4 contribution to specific pathophysiological conditions is required, preferentially in *in vivo* settings such as in animal models. This may open new therapeutic avenues to counteract human pathologies by targeting rG4s and their protein partners.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Paul Anderson and Ivanov lab members for helpful critiques.

FUNDING

National Institutes of Health [R01 GM126150 to P.I.]. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Lagnado,J. (2013) The story of quadruplex DNA – it started with a Bang! *The Biochemist*, **35**, 44–46.
2. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
3. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. *Cell*, **51**, 899–908.
4. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
5. Bhattacharyya,D., Mirihana Arachchilage,G. and Basu,S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4**, 38–38.
6. Largy,E., Mergny,J.L. and Gabelica,V. (2016) Role of alkali metal ions in G-quadruplex nucleic acid structure and stability. *Met. Ions Life Sci.*, **16**, 203–258.
7. Kim,B.G., Shek,Y.L. and Chalikian,T.V. (2013) Polyelectrolyte effects in G-quadruplexes. *Biophys. Chem.*, **184**, 95–100.
8. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
9. Lombardi,Puig and Londoño-Vallejo,A. (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.*, **48**, 1–15.
10. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
11. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.
12. Sengar,A., Vandana,J.J., Chambers,V.S., Di Antonio,M., Winnerdy,F.R., Balasubramanian,S. and Phan,A.T. (2019) Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. *Nucleic Acids Res.*, **47**, 1564–1572.
13. Spiegel,J., Adhikari,S. and Balasubramanian,S. (2020) The structure and function of DNA G-quadruplexes. *Trends Chem.*, **2**, 123–136.
14. Bedrat,A., Lacroix,L. and Mergny,J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.
15. Sahakyan,A.B., Chambers,V.S., Marsico,G., Santner,T., Di Antonio,M. and Balasubramanian,S. (2017) Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.*, **7**, 14535.
16. Wong,H.M., Stegle,O., Rodgers,S. and Huppert,J.L. (2010) A toolbox for predicting g-quadruplex formation and stability. *J. Nucleic Acids*, **2010**, 564946.
17. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
18. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
19. Smith,F.W. and Feigon,J. (1992) Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
20. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, **59**, 871–880.
21. Huppert,J.L. and Balasubramanian,S. (2006) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
22. Maizels,N. and Gray,L.T. (2013) The G4 Genome. *PLoS Genet.*, **9**, e1003468.
23. Lerner,L.K. and Sale,J.E. (2019) Replication of G Quadruplex DNA. *Genes (Basel)*, **10**, 95.
24. Saccà,B., Lacroix,L. and Mergny,J.-L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.*, **33**, 1182–1192.
25. Kwok,C.K., Marsico,G., Sahakyan,A.B., Chambers,V.S. and Balasubramanian,S. (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods*, **13**, 841–844.
26. Fay,M.M., Lyons,S.M. and Ivanov,P. (2017) RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.*, **429**, 2127–2147.
27. Kharel,P., Balaratnam,S., Beals,N. and Basu,S. (2020) The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *WIREs RNA*, **11**, e1568.
28. Cammas,A. and Millevoi,S. (2016) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.*, **45**, 1584–1595.

29. Collie,G.W., Haider,S.M., Neidle,S. and Parkinson,G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.*, **38**, 5569–5580.
30. Rouleau,S.G., Garant,J.M., Bolduc,F., Bisaillon,M. and Perreault,J.P. (2018) G-Quadruplexes influence pri-microRNA processing. *RNA Biol*, **15**, 198–206.
31. Mirihana Arachchilage,G., Dassanayake,A.C. and Basu,S. (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.*, **22**, 262–272.
32. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
33. Mestre-Fos,S., Penev,P.I., Suttapitugsakul,S., Hu,M., Ito,C., Petrov,A.S., Wartell,R.M., Wu,R. and Williams,L.D. (2019) G-quadruplexes in human ribosomal RNA. *J. Mol. Biol.*, **431**, 1940–1955.
34. Zaccaria,F. and Fonseca Guerra,C. (2018) RNA versus DNA G-quadruplex: the origin of increased stability. *Chemistry*, **24**, 16315–16322.
35. Webba da Silva,M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738–9745.
36. Chung,W.J., Heddi,B., Schmitt,E., Lim,K.W., Mechulam,Y. and Phan,A.T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2729–2733.
37. Zhou,J., Bourdoncle,A., Rosu,F., Gabelica,V. and Mergny,J.-L. (2012) Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed.*, **51**, 11002–11005.
38. Singh,A. and Kukreti,S. (2018) A triple stranded G-quadruplex formation in the promoter region of human myosin β(Myh7) gene. *J. Biomol. Struct. Dyn.*, **36**, 2773–2786.
39. Li,X.-M., Zheng,K.-w., Zhang,J.-y., Liu,H.-h., He,Y.-d., Yuan,B.-f., Hao,Y.-h and Tan,Z. (2015) Guanine-vacancy–bearing G-quadruplexes responsive to guanine derivatives. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 14581–14586.
40. Martadinata,H. and Phan,A.T. (2013) Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry*, **52**, 2176–2183.
41. Kolesnikova,S. and Curtis,E.A. (2019) Structure and function of multimeric G-quadruplexes. *Molecules*, **24**, 3074.
42. Xiao,C.-D., Ishizuka,T., Zhu,X.-Q., Li,Y., Sugiyama,H. and Xu,Y. (2017) Unusual topological RNA architecture with an eight-stranded helical fragment containing A-, G-, and U-tetrads. *J. Am. Chem. Soc.*, **139**, 2565–2568.
43. Dolinnaya,N.G., Ogloblina,A.M. and Yakubovskaya,M.G. (2016) Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry (Moscow)*, **81**, 1602–1649.
44. Xiao,C.-D., Ishizuka,T. and Xu,Y. (2017) Antiparallel RNA G-quadruplex formed by human telomere RNA containing 8-bromoguanosine. *Sci. Rep.*, **7**, 6695.
45. Xiao,C.-D., Shibata,T., Yamamoto,Y. and Xu,Y. (2018) An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem. Commun.*, **54**, 3944–3946.
46. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
47. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic. Acids. Res.*, **26**, 1167–1172.
48. Collie,G.W., Sparapani,S., Parkinson,G.N. and Neidle,S. (2011) Structural basis of telomeric RNA quadruplex−acridine ligand recognition. *J. Am. Chem. Soc.*, **133**, 2721–2728.
49. Chen,M.C., Murat,P., Abecassis,K., Ferré-D'Amaré,A.R. and Balasubramanian,S. (2015) Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.*, **43**, 2223–2231.
50. Martadinata,H. and Phan,A.T. (2013) Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K+ solution. *Biochemistry*, **52**, 2176–2183.
51. Monsen,R.C. and Trent,J.O. (2018) G-quadruplex virtual drug screening: a review. *Biochimie*, **152**, 134–148.
52. Haider,S. and Neidle,S. (2010) In: Baumann,P. (ed). *G-Quadruplex DNA: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 17–37.
53. Cheatham,T.E. 3rd, Brooks,B.R. and Kollman,P.A. (2001) Molecular modeling of nucleic acid structure. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0710s06.
54. Macke,T.J. and Case,D.A. (1997) In: *Molecular Modeling of Nucleic Acids*. American Chemical Society, Vol. **682**, pp. 379–393.
55. Sponer,J., Cang,X. and Cheatham,T.E. 3rd. (2012) Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods*, **57**, 25–39.
56. Cang,X., Šponer,J. and Cheatham,I.I.I.T.E. (2011) Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy Analysis. *J. Am. Chem. Soc.*, **133**, 14270–14279.
57. Mori,T., Oguro,A., Ohtsu,T. and Nakamura,Y. (2004) RNA aptamers selected against the receptor activator of NF-kappaB acquire general affinity to proteins of the tumor necrosis factor receptor family. *Nucleic Acids Res.*, **32**, 6120–6128.
58. Arora,A. and Maiti,S. (2009) Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J. Phys. Chem. B*, **113**, 10515–10520.
59. Joachimi,A., Benz,A. and Hartig,J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.*, **17**, 6811–6815.
60. Tsvetkov,V., Pozmogova,G. and Varizhuk,A. (2016) The systematic approach to describing conformational rearrangements in G-quadruplexes. *J. Biomol. Struct. Dyn.*, **34**, 705–715.
61. Varizhuk,A.M., Protopopova,A.D., Tsvetkov,V.B., Barinov,N.A., Podgorsky,V.V., Tankevich,M.V., Vlasenok,M.A., Severov,V.V., Smirnov,I.P., Dubrovin,E.V. *et al.* (2018) Polymorphism of G4 associates: from stacks to wires via interlocks. *Nucleic Acids Res.*, **46**, 8978–8992.
62. Tsvetkov,V.B., Zatsepin,T.S., Belyaev,E.S., Kostyukevich,Y.I., Shpakovski,G.V., Podgorsky,V.V., Pozmogova,G.E., Varizhuk,A.M. and Aralov,A.V. (2018) i-Clamp phenoxazine for the fine tuning of DNA i-motif stability. *Nucleic Acids Res.*, **46**, 2751–2764.
63. Protopopova,A.D., Tsvetkov,V.B., Varizhuk,A.M., Barinov,N.A., Podgorsky,V.V., Klinov,D.V. and Pozmogova,G.E. (2018) The structural diversity of C-rich DNA aggregates: unusual self-assembly of beetle-like nanostructures. *Phys. Chem. Chem. Phys.*, **20**, 3543–3553.
64. Haider,S. (2018) Computational methods to study G-quadruplex–ligand complexes. *J. Indian Inst. Sci.*, **98**, 325–339.
65. Zhang,A.Y.Q. and Balasubramanian,S. (2012) The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J. Am. Chem. Soc.*, **134**, 19297–19308.
66. Chu,J.F., Chang,T.C. and Li,H.W. (2010) Single-molecule TPM studies on the conversion of human telomeric DNA. *Biophys. J.*, **98**, 1608–1616.
67. Zhang,D.H., Fujimoto,T., Saxena,S., Yu,H.Q., Miyoshi,D. and Sugimoto,N. (2010) Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*, **49**, 4554–4563.
68. Havrila,M., Stadlbauer,P., Kührová,P., Banáš,P., Mergny,J.-L., Otyepka,M. and Šponer,J. (2018) Structural dynamics of propeller loop: towards folding of RNA G-quadruplex. *Nucleic Acids Res.*, **46**, 8754–8771.
69. Ganser,L.R., Kelly,M.L., Herschlag,D. and Al-Hashimi,H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.*, **20**, 474–489.
70. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., Antonio,M.D. and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.
71. Ding,Y., Fleming,A.M. and Burrows,C.J. (2018) Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes. *Sci. Rep.*, **8**, 15679–15679.
72. Lee,D.S.M., Ghanem,L.R. and Barash,Y. (2020) Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat. Commun.*, **11**, 527.
73. Bartas,M., Čutová,M., Brázda,V., Kaura,P., Šťastný,J., Kolomazník,J., Coufal,J., Goswami,P., Červeň,J. and Pečinka,P.

(2019) The presence and localization of G-quadruplex forming sequences in the domain of bacteria. *Molecules*,**24**, 1711.
74. Lavezzo,E., Berselli,M., Frasson,I., Perrone,R., Palù,G., Brazzale,A.R., Richter,S.N. and Toppo,S. (2018) G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLoS Comput. Biol.*, **14**, e1006675.
75. Biswas,B., Kandpal,M., Jauhari,U.K. and Vivekanandan,P. (2016) Genome-wide analysis of G-quadruplexes in herpesvirus genomes. *BMC Genomics*, **17**, 949.
76. Mathad,R.I. and Yang,D. (2011) G-quadruplex structures and G-quadruplex-interactive compounds. *Methods Mol. Biol.*, **735**, 77–96.
77. Del Villar-Guerra,R., Trent,J.O. and Chaires,J.B. (2018) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **57**, 7171–7175.
78. Mergny,J.-L., Phan,A.-T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
79. Sun,D. and Hurley,L.H. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.*, **608**, 65–79.
80. Di Antonio,M., Rodriguez,R. and Balasubramanian,S. (2012) Experimental approaches to identify cellular G-quadruplex structures and functions. *Methods*, **57**, 84–92.
81. Manna,S. and Srivatsan,S.G. (2018) Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv*, **8**, 25673–25694.
82. Biffi,G., Di Antonio,M., Tannahill,D. and Balasubramanian,S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem*, **6**, 75–80.
83. Hänsel-Hertsch,R., Spiegel,J., Marsico,G., Tannahill,D. and Balasubramanian,S. (2018) Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.*, **13**, 551–564.
84. Raguseo,F., Chowdhury,S., Minard,A. and Antonio,Di, M. (2020) Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.*, **56**, 1317–1324.
85. Kwok,C.K., Marsico,G. and Balasubramanian,S. (2018) Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harb. Perspect. Biol.*, **10**, a032284.
86. Yang,S.Y., Lejault,P., Chevrier,S., Boidot,R., Robertson,A.G., Wong,J.M.Y. and Monchaud,D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.*, **9**, 4730.
87. Herdy,B., Mayer,C., Varshney,D., Marsico,G., Murat,P., Taylor,C., D'Santos,C., Tannahill,D. and Balasubramanian,S. (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.*, **46**, 11592–11604.
88. Benhalevy,D., Gupta,S.K., Danan,C.H., Ghosal,S., Sun,H.W., Kazemier,H.G., Paeschke,K., Hafner,M. and Juranek,S.A. (2017) The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.*, **18**, 2979–2990.
89. Sauer,M., Juranek,S.A., Marks,J., De Magis,A., Kazemier,H.G., Hilbig,D., Benhalevy,D., Wang,X., Hafner,M. and Paeschke,K. (2019) DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.*, **10**, 2421.
90. Laguerre,A., Hukezalie,K., Winckler,P., Katranji,F., Chanteloup,G., Pirrotta,M., Perrier-Cornet,J.-M., Wong,J.M.Y. and Monchaud,D. (2015) Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.*, **137**, 8521–8525.
91. Chen,X.C., Chen,S.B., Dai,J., Yuan,J.H., Ou,T.M., Huang,Z.S. and Tan,J.H. (2018) Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angew. Chem. Int. Ed. Engl.*, **57**, 4702–4706.
92. Guo,J.U. and Bartel,D.P. (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, **353**, aaf5371.
93. Fazal,F.M., Han,S., Parker,K.R., Kaewsapsak,P., Xu,J., Boettiger,A.N., Chang,H.Y. and Ting,A.Y. (2019) Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell*, **178**, 473–490.
94. Einarson,O.J. and Sen,D. (2017) Self-biotinylation of DNA G-quadruplexes via intrinsic peroxidase activity. *Nucleic Acids Res.*, **45**, 9813–9822.
95. Chen,S.-B., Hu,M.-H., Liu,G.-C., Wang,J., Ou,T.-M., Gu,L.-Q., Huang,Z.-S. and Tan,J.-H. (2016) Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe. *J. Am. Chem. Soc.*, **138**, 10382–10385.
96. Lyu,K., Chen,S.-B., Chan,C.-Y., Tan,J.-H. and Kwok,C.K. (2019) Structural analysis and cellular visualization of APP RNA G-quadruplex. *Chem. Sci.*, **10**, 11095–11102.
97. Xu,S., Li,Q., Xiang,J., Yang,Q., Sun,H., Guan,A., Wang,L., Liu,Y., Yu,L., Shi,Y. *et al.* (2015) Directly lighting up RNA G-quadruplexes from test tubes to living human cells. *Nucleic. Acids. Res.*, **43**, 9575–9586.
98. Bugaut,A., Murat,P. and Balasubramanian,S. (2012) An RNA hairpin to G-quadruplex conformational transition. *J. Am. Chem. Soc.*, **134**, 19953–19956.
99. Endoh,T. and Sugimoto,N. (2019) Conformational dynamics of the RNA G-quadruplex and its effect on translation efficiency. *Molecules*, **24**, 1613.
100. Renard,I., Grandmougin,M., Roux,A., Yang,S.Y., Lejault,P., Pirrotta,M., Wong,J.M.Y. and Monchaud,D. (2019) Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. *Nucleic Acids Res.*, **47**, 5502–5510.
101. Yang,X., Cheema,J., Zhang,Y., Deng,H., Duncan,S., Umar,M.I., Zhao,J., Liu,Q., Cao,X., Kwok,C.K. *et al.* (2020) RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biol.*, **21**, 226.
102. Castello,A., Fischer,B., Eichelbaum,K., Horos,R., Beckmann,B.M., Strein,C., Davey,N.E., Humphreys,D.T., Preiss,T., Steinmetz,L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
103. Serikawa,T., Spanos,C., von Hacht,A., Budisa,N., Rappsilber,J. and Kurreck,J. (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5′ UTR of tumor-associated mRNAs. *Biochimie*, **144**, 169–184.
104. von Hacht,A., Seifert,O., Menger,M., Schütze,T., Arora,A., Konthur,Z., Neubauer,P., Wagner,A., Weise,C. and Kurreck,J. (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.*, **42**, 6630–6644.
105. Herviou,P., Le Bras,M., Dumas,L., Hieblot,C., Gilhodes,J., Cioci,G., Hugnot,J.-P., Ameadan,A., Guillonneau,F., Dassi,E. *et al.* (2020) hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nat. Commun.*, **11**, 2661.
106. Huang,Z.-L., Dai,J., Luo,W.-H., Wang,X.-G., Tan,J.-H., Chen,S.-B. and Huang,Z.-S. (2018) Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J. Am. Chem. Soc.*, **140**, 17945–17955.
107. Brázda,V., Hároníková,L., Liao,J.C.C. and Fojta,M. (2014) DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.*, **15**, 17493–17517.
108. McRae,E.K.S., Booy,E.P., Padilla-Meier,G.P. and McKenna,S.A. (2017) On characterizing the interactions between proteins and guanine quadruplex structures of nucleic acids. *J. Nucleic Acids*, **2017**, 9675348.
109. Ozdilek,B.A., Thompson,V.F., Ahmed,N.S., White,C.I., Batey,R.T. and Schwartz,J.C. (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.*, **45**, 7984–7996.
110. Rajyaguru,P. and Parker,R. (2012) RGG motif proteins: modulators of mRNA functional states. *Cell Cycle*, **11**, 2594–2599.
111. Zagrovic,B., Bartonek,L. and Polyansky,A.A. (2018) RNA-protein interactions in an unstructured context. *FEBS Lett.*, **592**, 2901–2916.
112. Liu,X. and Xu,Y. (2018) HnRNPA1 specifically recognizes the base of nucleotide at the loop of RNA G-quadruplex. *Molecules*, **23**, 237.
113. Phan,A.T., Kuryavyi,V., Darnell,J.C., Serganov,A., Majumdar,A., Ilin,S., Raslin,T., Polonskaia,A., Chen,C., Clain,D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an

RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.*, **18**, 796–804.
114. Heddi,B., Cheong,V.V., Martadinata,H. and Phan,A.T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9608–9613.
115. Izumi,H. and Funa,K. (2019) Telomere function and the G-quadruplex formation are regulated by hnRNP U. *Cells*, **8**, 390.
116. Darnell,J.C., Fraser,C.E., Mostovetsky,O. and Darnell,R.B. (2009) Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs. *Hum. Mol. Genet.*, **18**, 3164–3177.
117. Menon,L., Mader,S.A. and Mihailescu,M.-R. (2008) Fragile X mental retardation protein interactions with the microtubule associated protein 1B RNA. *RNA*, **14**, 1644–1655.
118. Vasilyev,N., Polonskaia,A., Darnell,J.C., Darnell,R.B., Patel,D.J. and Serganov,A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5391–E5400.
119. Bensaid,M., Melko,M., Bechara,E.G., Davidovic,L., Berretta,A., Catania,M.V., Gecz,J., Lalli,E. and Bardoni,B. (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic. Acids. Res.*, **37**, 1269–1279.
120. Chen,M.C., Tippana,R., Demeshkina,N.A., Murat,P., Balasubramanian,S., Myong,S. and Ferré-D'Amaré,A.R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, **558**, 465–469.
121. Chen,W.-F., Rety,S., Guo,H.-L., Dai,Y.-X., Wu,W.-Q., Liu,N.-N., Auguin,D., Liu,Q.-W., Hou,X.-M., Dou,S.-X. *et al.* (2018) Molecular mechanistic insights into Drosophila DHX36-mediated G-quadruplex unfolding: a structure-based model. *Structure*, **26**, 403–415.
122. Tippana,R., Chen,M.C., Demeshkina,N.A., Ferré-D'Amaré,A.R. and Myong,S. (2019) RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.*, **10**, 1855.
123. Wolfe,A.L., Singh,K., Zhong,Y., Drewe,P., Rajasekhar,V.K., Sanghvi,V.R., Mavrakis,K.J., Jiang,M., Roderick,J.E., Van der Meulen,J. *et al.* (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, **513**, 65–70.
124. Waldron,J.A., Raza,F. and Le Quesne,J. (2018) eIF4A alleviates the translational repression mediated by classical secondary structures more than by G-quadruplexes. *Nucleic Acids Res.*, **46**, 3075–3087.
125. Gao,J., Byrd,A.K., Zybailov,B.L., Marecki,J.C., Guderyon,M.J., Edwards,A.D., Chib,S., West,K.L., Waldrip,Z.J., Mackintosh,S.G. *et al.* (2019) DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA. *Chem. Commun.*, **55**, 4467–4470.
126. Brazda,V., Cerven,J., Bartas,M., Mikyskova,N., Coufal,J. and Pecinka,P. (2018) The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules*, **23**, 2341.
127. Stefl,R., Skrisovska,L. and Allain,F.H.T. (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.*, **6**, 33–38.
128. Dominguez,C. and Allain,F.H.T. (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res.*, **34**, 3634–3645.
129. Kluska,K., Adamczyk,J. and Krężel,A. (2018) Metal binding properties, stability and reactivity of zinc fingers. *Coord. Chem. Rev.*, **367**, 18–64.
130. Benhalevy,D., Gupta,S.K., Danan,C.H., Ghosal,S., Sun,H.-W., Kazemier,H.G., Paeschke,K., Hafner,M. and Juranek,S.A. (2017) The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.*, **18**, 2979–2990.
131. O'Day,E., Le,M.T., Imai,S., Tan,S.M., Kirchner,R., Arthanari,H., Hofmann,O., Wagner,G. and Lieberman,J. (2015) An RNA-binding protein, Lin28, recognizes and remodels G-quartets in the microRNAs (miRNAs) and mRNAs it regulates. *J. Biol. Chem.*, **290**, 17909–17922.
132. Niu,K., Xiang,L., Jin,Y., Peng,Y., Wu,F., Tang,W., Zhang,X., Deng,H., Xiang,H., Li,S. *et al.* (2019) Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res.*, **47**, 7306–7320.
133. Serikawa,T., Spanos,C., von Hacht,A., Budisa,N., Rappsilber,J. and Kurreck,J. (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5′ UTR of tumor-associated mRNAs. *Biochimie*, **144**, 169–184.
134. Dardenne,E., Polay Espinoza,M., Fattet,L., Germann,S., Lambert,M.-P., Neil,H., Zonta,E., Mortada,H., Gratadou,L., Deygas,M. *et al.* (2014) RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.*, **7**, 1900–1913.
135. Lyons,S.M., Gudanis,D., Coyne,S.M., Gdaniec,Z. and Ivanov,P. (2017) Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.*, **8**, 1127.
136. Weldon,C., Behm-Ansmant,I., Hurley,L.H., Burley,G.A., Branlant,C., Eperon,I.C. and Dominguez,C. (2017) Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat. Chem. Biol.*, **13**, 18–20.
137. Crossley,M.P., Bocek,M. and Cimprich,K.A. (2019) R-loops as cellular regulators and genomic threats. *Mol. Cell*, **73**, 398–411.
138. Skourti-Stathaki,K. and Proudfoot,N.J. (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.*, **28**, 1384–1396.
139. Kuznetsov,V.A., Bondarenko,V., Wongsurawat,T., Yenamandra,S.P. and Jenjaroenpun,P. (2018) Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acids Res.*, **46**, 7566–7585.
140. Xiao,S., Zhang,J.-Y., Zheng,K.-W., Hao,Y.-H. and Tan,Z. (2013) 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.*, **41**, 10379–10390.
141. Belotserkovskii,B.P., Soo Shin,J.H. and Hanawalt,P.C. (2017) Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res.*, **45**, 6589–6599.
142. Zheng,K.W., Xiao,S., Liu,J.Q., Zhang,J.Y., Hao,Y.H. and Tan,Z. (2013) Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.*, **41**, 5533–5541.
143. Ribeiro de Almeida,C., Dhir,S., Dhir,A., Moghaddam,A.E., Sattentau,Q., Meinhart,A. and Proudfoot,N.J. (2018) RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol. Cell*, **70**, 650–662.
144. Zheng,K.-w., Wu,R.-y., He,Y.-d., Xiao,S., Zhang,J.-y., Liu,J.-q., Hao,Y.-h. and Tan,Z. (2014) A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.*, **42**, 10832–10844.
145. Gromak,N., West,S. and Proudfoot,N.J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell. Biol.*, **26**, 3986–3996.
146. Skourti-Stathaki,K., Proudfoot,N.J. and Gromak,N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell*, **42**, 794–805.
147. Didiot,M.-C., Tian,Z., Schaeffer,C., Subramanian,M., Mandel,J.-L. and Moine,H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.*, **36**, 4902–4912.
148. Wong,M.S., Wright,W.E. and Shay,J.W. (2014) Alternative splicing regulation of telomerase: a new paradigm? *Trends Genet.*, **30**, 430–438.
149. Marcel,V., Tran,P.L., Sagne,C., Martel-Planche,G., Vaslin,L., Teulade-Fichou,M.P., Hall,J., Mergny,J.L., Hainaut,P. and Van Dyck,E. (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, **32**, 271–278.

150. Fujita,K.-I., Ishizuka,T., Mitsukawa,M., Kurata,M. and Masuda,S. (2020) Regulating divergent transcriptomes through mRNA splicing and its modulation using various small compounds. *Int. J. Mol. Sci.*, **21**, 2026.
151. Zarudnaya,M.I., Kolomiets,I.M., Potyahaylo,A.L. and Hovorun,D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.*, **31**, 1375–1386.
152. Shi,Y. and Manley,J.L. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.*, **29**, 889–897.
153. Decorsière,A., Cayrel,A., Vagner,S. and Millevoi,S. (2011) 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.*, **25**, 220–225.
154. Subramanian,M., Rage,F., Tabet,R., Flatter,E., Mandel,J.-L. and Moine,H. (2011) G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.*, **12**, 697–704.
155. Ishiguro,A., Kimura,N., Watanabe,Y., Watanabe,S. and Ishihama,A. (2016) TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells*, **21**, 466–481.
156. Rouleau,S., Glouzon,J.-P.S., Brumwell,A., Bisaillon,M. and Perreault,J.-P. (2017) 3′ UTR G-quadruplexes regulate miRNA binding. *RNA*, **23**, 1172–1179.
157. Ibrahim,F., Maragkakis,M., Alexiou,P. and Mourelatos,Z. (2018) Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat. Struct. Mol. Biol.*, **25**, 302–310.
158. Kumari,S., Bugaut,A. and Balasubramanian,S. (2008) Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5′ UTR of the NRAS proto-oncogene. *Biochemistry*, **47**, 12664–12669.
159. Leppek,K., Das,R. and Barna,M. (2018) Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.*, **19**, 158–174.
160. Murat,P., Marsico,G., Herdy,B., Ghanbarian,A., Portella,G. and Balasubramanian,S. (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.*, **19**, 229.
161. Jackson,R.J. (2013) The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol.*, **5**, a011569.
162. Mirihana Arachchilage,G., Hetti Arachchilage,M., Venkataraman,A., Piontkivska,H. and Basu,S. (2019) Stable G-quadruplex enabling sequences are selected against by the context-dependent codon bias. *Gene*, **696**, 149–161.
163. Crenshaw,E., Leung,B.P., Kwok,C.K., Sharoni,M., Olson,K., Sebastian,N.P., Ansaloni,S., Schweitzer-Stenner,R., Akins,M.R., Bevilacqua,P.C. *et al.* (2015) Amyloid precursor protein translation is regulated by a 3′UTR guanine quadruplex. *PLoS One*, **10**, e0143160.
164. Murat,P., Zhong,J., Lekieffre,L., Cowieson,N.P., Clancy,J.L., Preiss,T., Balasubramanian,S., Khanna,R. and Tellam,J. (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.*, **10**, 358–364.
165. Yu,C.-H., Teulade-Fichou,M.-P. and Olsthoorn,R.C.L. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.*, **42**, 1887–1892.
166. Thandapani,P., Song,J., Gandin,V., Cai,Y., Rouleau,S.G., Garant,J.-M., Boisvert,F.-M., Yu,Z., Perreault,J.-P., Topisirovic,I. *et al.* (2015) Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *eLife*, **4**, e06234.
167. Beaudoin,J.-D. and Perreault,J.-P. (2013) Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.*, **41**, 5898–5911.
168. Cech,T.R. and Steitz,J.A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, **157**, 77–94.
169. Matsumura,K., Kawasaki,Y., Miyamoto,M., Kamoshida,Y., Nakamura,J., Negishi,L., Suda,S. and Akiyama,T. (2017) The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*, **36**, 1191–1199.
170. Zhang,X., Yu,L., Ye,S., Xie,J., Huang,X., Zheng,K. and Sun,B. (2019) MOV10L1 binds RNA G-quadruplex in a structure-specific manner and resolves it more efficiently than MOV10. *iScience*, **17**, 36–48.
171. Balaratnam,S., Hettiarachchilage,M., West,N., Piontkivska,H. and Basu,S. (2019) A secondary structure within a human piRNA modulates its functionality. *Biochimie*, **157**, 72–80.
172. Biffi,G., Tannahill,D. and Balasubramanian,S. (2012) An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.*, **134**, 11974–11976.
173. Lattmann,S., Stadler,M.B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2011) The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.*, **39**, 9390–9404.
174. Jayaraj,G.G., Pandey,S., Scaria,V. and Maiti,S. (2012) Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol*, **9**, 81–86.
175. Pandey,S., Agarwala,P., Jayaraj,G.G., Gargallo,R. and Maiti,S. (2015) The RNA stem–loop to G-quadruplex equilibrium controls mature microRNA production inside the cell. *Biochemistry*, **54**, 7067–7078.
176. Lexa,M., Steflova,P., Martinek,T., Vorlickova,M., Vyskot,B. and Kejnovsky,E. (2014) Guanine quadruplexes are formed by specific regions of human transposable elements. *BMC Genomics*, **15**, 1032.
177. Wang,X., Goodrich,K.J., Gooding,A.R., Naeem,H., Archer,S., Paucek,R.D., Youmans,D.T., Cech,T.R. and Davidovich,C. (2017) Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell*, **65**, 1056–1067.
178. Lyons,S.M., Fay,M.M., Akiyama,Y., Anderson,P.J. and Ivanov,P. (2017) RNA biology of angiogenin: current state and perspectives. *RNA Biol*, **14**, 171–178.
179. Advani,V.M. and Ivanov,P. (2019) Translational control under stress: reshaping the translatome. *Bioessays*, **41**, e1900009.
180. Lyons,S.M., Kharel,P., Akiyama,Y., Ojha,S., Dave,D., Tsvetkov,V., Merrick,W., Ivanov,P. and Anderson,P. (2020) eIF4G has intrinsic G-quadruplex binding activity that is required for tiRNA function. *Nucleic Acids Res.*, **48**, 6223–6233.
181. Riggs,C.L., Kedersha,N., Ivanov,P. and Anderson,P. (2020) Mammalian stress granules and P bodies at a glance. *J. Cell Sci.*, **133**, jcs242487.
182. Advani,V.M. and Ivanov,P. (2020) Stress granule subtypes: an emerging link to neurodegeneration. *Cell. Mol. Life Sci.*, **77**, 4827–4845.
183. Feng,Z., Chen,X., Wu,X. and Zhang,M. (2019) Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *J. Biol. Chem.*, **294**, 14823–14835.
184. Lafontaine,D.L.J., Riback,J.A., Bascetin,R. and Brangwynne,C.P. (2020) The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.*, doi:10.1038/s41580-020-0272-6.
185. Fay,M.M. and Anderson,P.J. (2018) The role of RNA in biological phase separations. *J. Mol. Biol.*, **430**, 4685–4701.
186. Jain,A. and Vale,R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243–247.
187. Lyons,S.M., Achorn,C., Kedersha,N.L., Anderson,P.J. and Ivanov,P. (2016) YB-1 regulates tiRNA-induced stress granule formation but not translational repression. *Nucleic Acids Res.*, **44**, 6949–6960.
188. Khristich,A.N. and Mirkin,S.M. (2020) On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.*, **295**, 4134–4170.
189. Renton,A.E., Majounie,E., Waite,A., Simón-Sánchez,J., Rollinson,S., Gibbs,J.R., Schymick,J.C., Laaksovirta,H., van Swieten,J.C., Myllykangas,L. *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.
190. Balendra,R. and Isaacs,A.M. (2018) C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.*, **14**, 544–558.
191. Fratta,P., Mizielinska,S., Nicoll,A.J., Zloh,M., Fisher,E.M., Parkinson,G. and Isaacs,A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, **2**, 1016.
192. Reddy,K., Zamiri,B., Stanley,S.Y., Macgregor,R.B. Jr. and Pearson,C.E. (2013) 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.*, **288**, 9860–9866.
193. Fay,M.M., Anderson,P.J. and Ivanov,P. (2017) ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Rep.*, **21**, 3573–3584.
194. Disney,M.D. (2019) Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J. Am. Chem. Soc.*, **141**, 6776–6790.
195. Halder,K., Largy,E., Benzler,M., Teulade-Fichou,M.-P. and Hartig,J.S. (2011) Efficient suppression of gene expression by targeting 5′-UTR-based RNA quadruplexes with bisquinolinium compounds. *ChemBioChem*, **12**, 1663–1668.
196. Miglietta,G., Cogoi,S., Marinello,J., Capranico,G., Tikhomirov,A.S., Shchekotikhin,A. and Xodo,L.E. (2017) RNA G-quadruplexes in Kirsten Ras (KRAS) oncogene as targets for small molecules inhibiting translation. *J. Med. Chem.*, **60**, 9448–9461.
197. Simone,R., Balendra,R., Moens,T.G., Preza,E., Wilson,K.M., Heslegrave,A., Woodling,N.S., Niccoli,T., Gilbert-Jaramillo,J., Abdelkarim,S. *et al.* (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol. Med.*, **10**, 22–31.
198. Goldberg,D.C., Fones,L., Vivinetto,A.L., Caufield,J.T., Ratan,R.R. and Cave,J.W. (2020) Manipulating adult neural stem and progenitor cells with G-quadruplex ligands. *ACS Chem. Neurosci.*, **11**, 1504–1518.
199. Bonnat,L., Dejeu,J., Bonnet,H., Génnaro,B., Jarjayes,O., Thomas,F., Lavergne,T. and Defrancq,E. (2016) Templated formation of discrete RNA and DNA:RNA hybrid G-quadruplexes and their interactions with targeting ligands. *Chemistry*, **22**, 3139–3147.
200. Umar,M.I., Ji,D., Chan,C.-Y. and Kwok,C.K. (2019) G-quadruplex-based fluorescent turn-on ligands and aptamers: from development to applications. *Molecules*, **24**, 2416.
201. Mirihana Arachchilage,G., Kharel,P., Reid,J. and Basu,S. (2018) Targeting of G-quadruplex harboring pre-miRNA 92b by LNA rescues PTEN expression in NSCL cancer cells. *ACS Chem. Biol.*, **13**, 909–914.
202. Umar,M.I. and Kwok,C.K. (2020) Specific suppression of D-RNA G-quadruplex–protein interaction with an L-RNA aptamer. *Nucleic Acids Res.*, **48**, 10125–10141.
203. Bhattacharyya,D., Nguyen,K. and Basu,S. (2014) Rationally induced RNA:DNA G-quadruplex structures elicit an anticancer effect by inhibiting endogenous eIF-4E expression. *Biochemistry*, **53**, 5461–5470.
204. Szameit,K., Berg,K., Kruspe,S., Valentini,E., Magbanua,E., Kwiatkowski,M., Chauvot de Beauchêne,I., Krichel,B., Schamoni,K., Uetrecht,C. *et al.* (2016) Structure and target interaction of a G-quadruplex RNA-aptamer. *RNA Biol*, **13**, 973–987.
205. Trachman,R.J. and Ferré-D'Amaré,A.R. (2019) Tracking RNA with light: selection, structure, and design of fluorescence turn-on RNA aptamers. *Q. Rev. Biophys.*, **52**, e8.
206. Jeng,S.C.Y., Chan,H.H.Y., Booy,E.P., McKenna,S.A. and Unrau,P.J. (2016) Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers. *RNA*, **22**, 1884–1892.
207. Stefan,L. and Monchaud,D. (2019) Applications of guanine quartets in nanotechnology and chemical biology. *Nat. Rev. Chem.*, **3**, 650–668.
208. Lim,K.W., Ng,V.C.M., Martín-Pintado,N., Heddi,B. and Phan,A.T. (2013) Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.*, **41**, 10556–10562.
209. Phan,A.T., Kuryavyi,V., Luu,K.N. and Patel,D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K + solution †. *Nucleic Acids Res.*, **35**, 6517–6525.