Published online 20 February 2012 Nucleic Acids Research, 2012, Vol. 40, No. 11 4727–4741
doi:10.1093/nar/gks068

SURVEY AND SUMMARY

5′-UTR RNA G-quadruplexes: translation regulation and targeting

Anthony Bugaut[1] and Shankar Balasubramanian[1,2,3,*]

[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, [2]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Center, Cambridge CB2 0RE and [3]School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK

Received November 7, 2011; Revised January 17, 2012; Accepted January 19, 2012

ABSTRACT

RNA structures in the untranslated regions (UTRs) of mRNAs influence post-transcriptional regulation of gene expression. Much of the knowledge in this area depends on canonical double-stranded RNA elements. There has been considerable recent advancement of our understanding of guanine(G)-rich nucleic acids sequences that form four-stranded structures, called G-quadruplexes. While much of the research has been focused on DNA G-quadruplexes, there has recently been a rapid emergence of interest in RNA G-quadruplexes, particularly in the 5′-UTRs of mRNAs. Collectively, these studies suggest that RNA G-quadruplexes exist in the 5′-UTRs of many genes, including genes of clinical interest, and that such structural elements can influence translation. This review features the progresses in the study of 5′-UTR RNA G-quadruplex-mediated translational control. It covers computational analysis, cell-free, cell-based and chemical biology studies that have sought to elucidate the roles of RNA G-quadruplexes in both cap-dependent and -independent regulation of mRNA translation. We also discuss protein *trans*-acting factors that have been implicated and the evidence that such RNA motifs have potential as small molecule target. Finally, we close the review with a perspective on the future challenges in the field of 5′-UTR RNA G-quadruplex-mediated translation regulation.

INTRODUCTION

G-quadruplex structures

While the discovery of canonical DNA double helix structure comprising Watson–Crick base pairing has provided the basis for our understanding of the genetic code (1), it is becoming increasingly evident that non-Watson–Crick interactions between bases and non-canonical nucleic acids structures have importance in biology. More than 40 years before the elucidation of the DNA double helix, Bang (2) reported the observation that guanylic acid (GMP) forms gels at high concentration in aqueous solution. Fifty years later, Gellert and co-workers (3) collected X-ray fiber diffraction data revealing that the structural basis for this phenomenon was the formation of regular hydrogen-bonded helices based on the assembly of tetrameric units, now know as guanine (G)-quartets. A G-quartet is formed by four G bases arranged in a square planar cyclic hydrogen-bonding pattern, where each guanine is both the donor and acceptor of two hydrogen bonds, providing a central site where the oxygen lone pair of the carbonyl groups can coordinate with metal cations (Figure 1A). Beyond GMP monomers, G-quartets can also arise intermolecularly between G-rich strands or intramolecularly within some G-rich nucleic acid sequences. When several G-quartets can form proximally within a single strand of nucleic acids they can stack upon each other, by mean of π–π interactions, to form a 3D structure, called a G-quadruplex (Figure 1B) (4). Such structures are further stabilized by physiologically abundant monovalent cations, particularly K^+ or Na^+, which fit in or coordinate between the G-quartets.

From DNA to RNA G-quadruplexes

G-quadruplex formation by biologically relevant nucleic acid sequences remained largely unexplored until it was discovered that the ends of human chromosomes, the telomeres, are composed of tandem repeats of a G-rich DNA sequence d(TTAGGG) (5). It was subsequently demonstrated by NMR spectroscopy that the single-stranded G-rich 3′ overhang of human telomeres is prone to form intramolecular G-quadruplex *in vitro* (6). In addition, a key paper by Zahler *et al.* (7) demonstrated an interplay between *Oxytricha nova* telomeric sequence

*To whom correspondence should be addressed. Tel: +44 1223 336347; Fax: +44 1223 336913; Email: sb10031@cam.ac.uk
Correspondence may also be addressed to Anthony Bugaut. Tel: +44 1223 746686; Fax: +44 1223 336913; Email: ab605@cam.ac.uk

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

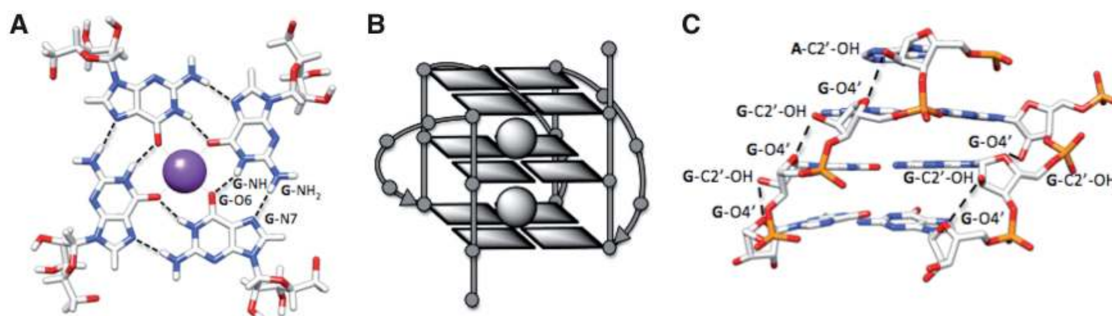

Figure 1. Schematic representations of (**A**) a G-quartet arrangement, (**B**) a G-quadruplex nucleic acids structure and (**C**) the intermolecular hydrogen bonding network (dash lines) between the ribose C2′ hydroxyl groups and the O4′ sugar oxygens of RNA G-quartet-forming residues [adapted from (53), PDB: 3IBK].

G-quadruplex formation and telomerase (i.e. the enzyme responsible for the elongation of chromosome ends during DNA replication) function *in vitro*, which stimulated further consideration of the potential function of DNA G-quadruplex in biology. Owing to the strong associations between telomerase over-expression and human cancers (8), extensive research effort has been devoted to investigate the relevance of G-quadruplex formation in telomere maintenance (9,10). This has led to the proposal of a new anticancer strategy based on small molecule agents that could target and stabilize G-quadruplex structures at telomeres with a view to inhibit telomerase function and thus impart cell death (11,12).

In addition to studies on G-quadruplex formation at telomeres, computational searches have been performed to identify G-rich sequences with potential to form G-quadruplex structures (putative quadruplex sequence, PQS) in the human genome and other genomes (13–26). Such bioinformatics studies revealed a prevalence of PQSs in genomes with a notable enrichment in gene promoter regions, indicating possible roles for DNA G-quadruplex-forming sequences as regulatory elements of transcription (17–20,22,24). This association of PQSs with gene promoters was consistent with an original hypothesis of Simonsson *et al*. (27) linking G-quadruplex formation with transcriptional control of the *C-MYC* gene. Subsequently, Hurley and co-workers (28) showed that mutational destabilization of the *C-MYC* promoter quadruplex led to a 3-fold increase in the transcription activity of a plasmid-based reporter assay. Data have continued to accumulate in the literature in support of transcription–associated roles for DNA PQSs, particularly for oncogenes, and the potential for such motifs to act as drug targets (29).

While much attention in the G-quadruplex field has been focused on DNA, in the 1990s, there were early reports of G-quadruplex formation from biologically relevant RNA molecules. For example, a 19-nt oligonucleotide derived from *Escherichia coli* 5S RNA was found to form an extremely stable K^+-stabilized tetrameric aggregate *in vitro* that was dependent on a UG_4U sequence motif at its 3′ end (30). The authors noted 'This complex is so stable that it would be surprising if similar structures do not occur in nature'. Based on 2D NMR spectra and molecular dynamics data, the UG_4U sequence motif was further proposed to form a parallel-stranded tetraplex structure containing four stacked G-quartets and at least one U-quartet (31,32). Other early studies linked formation of RNA G-quadruplex structures to biological events in viruses, such HIV or herpes simplex virus (33,34), and four-stranded RNA structures formed from G-rich sequences were proposed to be involved in the processing and translational control in human transcripts (35–43). A noteworthy recent example of the potential importance of RNA G-quadruplex emerged from the discovery of ~100- to 9000-nt G-rich telomeric repeat-containing RNAs (TERRA), which arise from RNA polymerase II transcription of the C-rich strand human DNA telomeres (44,45). TERRA transcripts are thought to play a key role in chromatin remodeling and in the regulation of telomerase activity (46). Unsurprisingly, G-rich oligonucleotides derived from TERRA sequence have been shown to adopt G-quadruplex structures *in vitro* (47–54). While synthetic oligonucleotide probes having human TERRA sequence and light-switching pyrene moieties at their ends have been used to show G-quadruplex formation in living cells (55), proof of G-quadruplex structures in native TERRA transcripts *in vivo* and details of their functional significance are still in need of robust experimental support. In addition to TERRA, G-quadruplex formation has also been proposed at the 5′ extremity of human telomerase RNA (hTR) (56). Interestingly, two recent studies demonstrated that hTR is bound *in vivo* by the RHAU/DHX36 protein, which is an RNA helicase that exhibits G-quadruplex substrate preference (57,58).

Numerous biophysical studies have shown that RNA sequences form G-quadruplex structures that are thermodynamically much more stable than their DNA counterparts (59–64). Some structural explanations of the higher stability of RNA versus DNA G-quadruplexes have recently been provided by X-ray crystallography studies on intermolecular RNA and DNA G-quadruplex structures from human telomeric repeats (53,54). In RNA G-quadruplexes, the ribose C2′ hydroxyl groups participate in an extended network of hydrogen bonds, including interactions with phosphate and oxygen backbone atoms, O4′ sugar oxygen and H-bond acceptors N2 groups of quartet-forming guanines (Figure 1C). As compared to DNA, these additional interactions

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

contribute to the increased thermodynamic stability of RNA G-quadruplexes by both decreasing the entropic cost associated with the number of ordered water molecules in the RNA G-quadruplex grooves and enhancing the favorable enthalpic contribution to the free energy of RNA G-quadruplex formation. Another noteworthy difference is that whereas intramolecular DNA G-quadruplex structures are generally highly polymorphic (65,66), currently available biophysical data indicate that intramolecular RNA G-quadruplexes have a preference to adopt a parallel-stranded conformation regardless of the sequences and the experimental conditions (60–64). However, it should be noted that such data have been predominantly obtained from circular dichroism (CD) spectroscopic measurements on rather short synthetic RNA sequences with closely related tandem repeat G-rich patterns. Recently, a high-resolution NMR structure of 1:1 complex between a 36-nt G-rich *in vitro* selected RNA aptamer and a peptide from the RGG domain of the human fragile X mental retardation protein (FMRP) has revealed the formation of a three-G-quartet RNA G-quadruplex structure adopting an unprecedented fold, which contains a mixed-junction quartet that connects the quadruplex to a flanking duplex stem (67). Due to the relative paucity of RNA G-quadruplex studies, it is conceivable that further structural variations in RNA G-quadruplex structures will be discovered in due course.

RNA G-quadruplexes in the 5′-UTRs of mRNAs

When compared to genomic double-helical DNA, transcribed RNA molecules are relatively unconstrained and can readily fold intramolecularly to form a wide variety of nucleic acid structures, which ultimately dictate the fate and function of the RNA. A particular example is the formation of secondary and tertiary structures within the 5′ untranslated regions (UTRs) of mRNAs, which has been shown to play important roles in the post-transcriptional regulation of gene expression (68–72). Whereas much of the scientific history of this field has been based on RNA structures that depend on canonical Pu/Py base pairs, numerous other non-covalent interactions can exist between the nucleotides that constitute a molecule of RNA, including the formation of G-quadruplexes. In 2007, we reported an RNA G-quadruplex-forming sequence within the 5′-UTR of the *NRAS* human proto-oncogene and demonstrated that formation of the RNA G-quadruplex inhibits protein expression *in vitro* (73). At the same time, a computational search of all annotated 5′-UTRs of the human transcriptome, in the latest available version of the Ensembl database at that time (version 40, NCBI build 36), identified ~3000 5′-UTRs that contained one or more RNA PQSs, including several other proto-oncogenes (73). This allowed us to propose that 5′-UTR RNA G-quadruplex formation could be a more general mechanism to regulate mRNA translation (73). In support of this proposition, a more detailed follow-up bioinformatics study revealed a significant enrichment in RNA G-quadruplex-forming motifs in the 5′-UTRs of human genes as compared to the whole transcriptome (74). Moreover, a recent bioinformatics analysis revealed a similar prevalence and distribution for RNA PQSs in the genome of the plant *Arabidopsis thaliana* (75).

In the past few years, there has been a rapid emergence of interest on RNA G-quadruplexes in the 5′-UTRs of mRNAs. Collectively, these studies suggest that RNA G-quadruplexes exist in many mRNAs, including genes of clinical interest for human diseases, and that such structures play a role in regulating the level of translation of their host gene into the corresponding protein product. In addition, recent publications have provided proof-of-concept for the inhibition of translation by small molecules that target G-quadruplexes in the 5′-UTR of RNA transcripts. While recent reviews have broadly covered some structural and/or biological progresses in RNA G-quadruplex (66,76,77), herein we discuss in detail the collective advancements in the emerging area of 5′-UTR RNA G-quadruplex-mediated translational control. Specifically, we cover the experimental evidence that support roles for 5′-UTR G-quadruplexes in cap-dependent and -independent regulation of mRNA translation, plus the putative protein *trans*-acting factors that might be involved in 5′-UTR G-quadruplex-mediated translation regulation, and also the evidence that such RNA motifs may serve as molecular targets for synthetic small molecules to affect gene expression. Finally, we close the review with a perspective on the future challenges for this field.

5′-UTR RNA G-QUADRUPLEXES IN TRANSLATION REGULATION

The translation of an mRNA into its protein product involves three major steps: initiation (assembly of a ribosome on the mRNA platform), elongation (protein synthesis) and termination (disassembly of the ribosome) (78). Initiation is believed to be the rate limiting and most regulated step of the whole process. The canonical model for eukaryotic translation initiation (also called a 'scanning' mechanism) can be briefly described by: (i) the assembly of a pre-initiation ribosomal complex (43S) at a modified nucleotide cap analog that forms the mRNA 5′ end, (ii) scanning of the complex along the 5′-UTR of the mRNA until it reaches the initiator AUG codon and (iii) release of the initiation factors (eIFs) and recruitment of the 60S ribosomal subunit to form a competent ribosome (80S) that proceeds to protein synthesis. Alternatively, certain mRNAs have the capacity to use another form of translation initiation that does not involve the cap nucleotide analog, but rather relies on internal ribosome entry sites (IRESs) (79). Translation initiation by IRES involves a mechanism in which translation of specific mRNAs is carried out using non-coding RNA sequences/structures that substitute the 5′ cap and some eIFs. This alternative mechanism for translation initiation is particularly important under circumstances where cap-dependent translation is compromised, such as cell stress, growth, differentiation, mitosis, apoptosis or viral infection. Both translation initiation mechanisms are

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

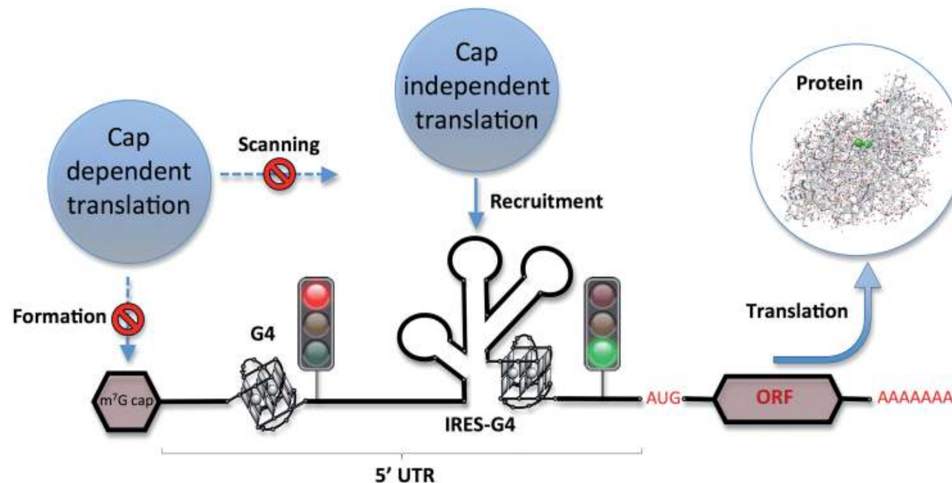

Figure 2. Schematic illustration of the possible roles of 5′-UTR RNA G-quadruplex formation in cap-dependent and cap-independent regulation of translation initiation. Red light indicates translation inhibition. Green light indicates translation enhancement.

influenced by 5′-UTR RNA structural elements, including G-quadruplex structures (Figure 2).

5′-UTR RNA G-quadruplexes in cap-dependent translation

Several studies have demonstrated that thermodynamically stable RNA hairpin structures in the 5′-UTRs of mRNAs can impair eukaryotic cap-dependent translation by compromising the assembly of the translation initiation machinery at the 5′ cap of the mRNA and/or by perturbing its scanning process toward an AUG translation start codon (80–83). However, the variety of RNA structures may be substantially broader than stem–loop formation, and could involve a wide range of non-Watson–Crick hydrogen bonding interactions, including Hoogsteen pairing between guanine bases that lead to the formation of G-quadruplex structures. Given that folded RNA G-quadruplexes are thermodynamically very stable under near physiological ionic conditions *in vitro* suggests that they are likely to form *in vivo*.

In 2001, Moine and co-workers provided a first glimpse of the potential of an intramolecular 5′-UTR RNA G-quadruplexes to influence translation efficiency. They identified a 35-nt RNA G-quadruplex structure, as determined by footprinting experiments, within the RGG-coding region of the *FMRP* as a specific binding site for the FMRP protein to its own mRNA (39). When this sequence was inserted in the 5′-UTR of a luciferase reporter mRNA, this resulted in a 1.5-fold decrease in *in vitro* translation efficiency as compared to control mRNAs, either lacking or containing an incomplete G-quadruplex sequence.

Subsequently, in 2007, we reported the presence of an RNA G-quadruplex-forming sequence naturally occurring in the 254-nt 5′-UTR of the p21 GTPase encoding human *NRAS* mRNA (73). This motif is conserved, in both its sequence and its position relative to the translation start site, across the 5′-UTRs of human, chimpanzee, macaque, mouse, rat and dog *NRAS* genes. Using CD spectroscopy and UV-melting experiments, we demonstrated that the 18-nt G-rich sequence folded into a RNA G-quadruplex *in vitro*. We then inserted the *NRAS* 5′-UTR in front of the *Firefly luciferase* reporter gene and directly downstream of a minimal T7 promoter in a plasmid vector, and generated control constructs that either lacked the PQS or contained mutations that disrupt G-quadruplex formation. Translation of the *in vitro* transcribed mRNAs in rabbit reticulocyte lysates revealed that the RNA G-quadruplex structure in its natural context within the *NRAS* 5′-UTR inhibits translation by ~80% (73). This result suggested that native RNA G-quadruplex structures in 5′-UTRs could act as regulatory elements of translation.

In a following publication, the same year, Wieland and Hartig (84) provided evidence consistent with RNA G-quadruplex formation in bacteria. Using artificial rationally designed G-rich elements introduced adjacent to the Shine–Dalgarno sequence of a GFP reporter gene, they demonstrated that sequences with potential to fold into four-stranded structures inhibited gene expression in *Escherichia coli*, presumably by abrogating access to the ribosome binding site. Importantly, the levels of GFP expression correlated well with the thermal stabilities of the inserted G-quadruplex structures, and G-quadruplex structures of moderate stabilities were shown to behave as thermo-regulators of GFP expression in bacteria. Although further work is needed to address the existence and importance of naturally occurring RNA G-quadruplex elements in prokaryotic translation initiation, these findings suggested that RNA G-quadruplex-forming sequences could be useful as tools in synthetic biology.

The first demonstration of translation inhibition by a 5′-UTR RNA G-quadruplex in living eukaryotic cells was published in 2008 (85). The authors investigated a 27-nt G-quadruplex-forming sequence identified within the 719-nt 5′-UTR of the mRNA of the human *Zic-1*. UV melting and CD spectroscopy experiments indicated that this sequence is likely to adopt a highly stable RNA G-quadruplex structure under near physiological ionic conditions. Using an established

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

dual-luciferase plasmid-based assay, the authors showed that the 27-nt G-rich fragment of the UTR repressed protein synthesis in HeLa cells by ~80% while a mutated version of the sequence, which did not form a stable quadruplex structure, did not influence translation. As identical levels of mRNA were detected using quantitative RT-PCR, it was concluded that the decrease in protein synthesis was due to repression of translation rather than a consequence of reduced transcription. In addition, western-blot analysis revealed that expression of Zic-1 from a plasmid vector was strongly reduced by a 73-nt fragment of the UTR containing the G-quadruplex motif.

Subsequent to these reports, additional studies have shown the presence of G-quadruplex-forming sequences within the mRNA 5′-UTRs of several human genes (Table 1). A comparable strategy, based on biophysical analysis of the G-rich sequence, mutagenesis and reporter gene-based expression assays, has been employed to confirm the capability of the sequences to fold into RNA G-quadruplex structures and to modulate translation (Table 1). The genes studied included the matrix metalloproteinase *MT3-MMP* (86), the estrogen receptor *ESR1* (87), the apoptotic regulator *BCL2* (88), the telomere shelterin protein *TRF2* (89) and the α-secretase *ADAM10* (90). Down-regulation of translation by rationally designed RNA G-quadruplex-forming sequences containing four to six G_3 tracts and oligo(U) loops of various length has also been reported (91). This study indicated that both loop-length and the number of G_3 repeats could influence *in cellulo* protein expression. Both shorter loop-lengths and increased numbers of G_3 tracts correlated with greater reductions in expression of a luciferase reporter gene in mammalian cells, which also correlated with the thermal stabilities of the RNA G-quadruplexes. In our own studies, we have explored the effect of RNA G-quadruplex position and stability in the *NRAS* 5′-UTR on protein synthesis (92). We observed a marked difference in translation inhibition by the G-quadruplex-forming sequence depending on its position relative to 5′ cap of the *NRAS* 5′-UTR. We found that the *NRAS* G-quadruplex motif is inhibitory of *in vitro* translation only when it is located sufficiently close to the 5′ cap (within ~50–100 nt), illustrating the importance of studying 5′-UTR RNA G-quadruplex-forming sequences in their natural context before drawing conclusions on their biological function.

Recently, Beaudoin and Perreault (93) have provided an in-depth analysis on naturally occurring 5′-UTR RNA G-quadruplex structures, which included *in silico*, *in vitro* and *in cellulo* experiments. Extending on a bioinformatics study previously reported (74), a database of all 5′-UTR PQS in the protein-coding genes of 18 different organisms was constructed. Next, PQSs within the 5′-UTRs of nine human genes encoding proteins involved in various biological processes were further investigated (Table 1). First, their abilities to fold into RNA G-quadruplex structures *in vitro* were evaluated using CD and also by the application of in-line probing, a technique originally developed in the to study riboswitches (94). Based on these experiments, six sequences were observed to form G-quadruplex. Each of them was shown to significantly inhibit translation in their natural context within the full-length 5′-UTR, using plasmid-based dual-luciferase assays in HEK 293 cells. Comparative analysis of the nine PQSs revealed that the three sequences that failed to form a G-quadruplex structure contained significantly more cytosines than those that folded into G-quadruplex. The authors proposed that the large number of cytosines most likely increases the propensity of those sequences to adopt stem-loop structures based on canonical GC base pairing, thus competing out G-quadruplex formation. This hypothesis was then substantiated through an elegant mutagenesis study in which several C-to-A substitutions were introduced to lower the G/C ratios and to destabilize the stem–loop structures. In all three cases, mutations were identified that rescued G-quadruplex formation *in vitro*, and in one case the C/A change was shown to induce translation repression. These results raise questions on whether any single nucleotide polymorphisms (SNPs) within 5′-UTRs map to RNA PQSs and therefore have potential to influence function. Through a bioinformatic search a total of 143 SNPs within 116 5′-UTR RNA PQSs were found in the human transcriptome (93). One SNP, consisting of a G-to-C substitution, was identified within one of the initially studied candidates, namely the *AASDHPPTT* 5′-UTR PQS (Table 1). This SNP was then shown to be detrimental to RNA G-quadruplex formation by in-line probing, and was demonstrated to increase the *in cellulo* translation efficiency of a reporter gene as compared to the wild-type (G-quadruplex-forming) sequence (93). This result suggests that SNPs within 5′-UTR RNA G-quadruplex-forming sequences can in principle cause differential gene expression between individuals.

5′-UTR RNA G-quadruplexes in cap-independent translation

Most examples in literature have described 5′-UTR RNA PQSs that inhibit translation, leading to the proposal that 5′-UTR RNA G-quadruplex are ‘predictable’ inhibitory elements of gene expression (91). However, there are cases where RNA G-quadruplex formation has been shown to actually promote translation. In 2003, Bonnal *et al.* (40) reported an analysis of the *cis*-acting elements defining a IRES present within the 5′-UTR of the human fibroblast growth factor 2 (*FGF2*) mRNA that revealed a G-quadruplex motif as a structural determinant of IRES activity. Here, chemical and enzymatic footprinting experiments were employed to probe the structure of the 484-nt *FGF2* mRNA 5′-UTR and an RNA G-quadruplex structure composed of five G-quartets was identified. Deletion analysis using bicistronic plasmid constructs in human liver adenocarcinoma SK-Hep1 cells demonstrated that the G-quadruplex-forming sequence (Table 1) is part of a 176-nt RNA module that is necessary and sufficient to confer IRES activity. While the RNA sequences and structural features of cellular IRESs remain largely unknown (79), this study provided the first example of an RNA G-quadruplex structure involved in an IRES element.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

Table 1. RNA G-quadruplex-forming sequences identified within human mRNA 5′-UTR that were experimentally shown to modulate translation efficiency

Gene	Protein product	RNA G-quadruplex-forming sequence (5′ to 3′)	Evidence for RNA G-quadruplex formation	*In vitro*/*in cellulo*	Change in protein expression[a]	Ref.
Cap-dependent translation						
NRAS	GTPase Nras	GGGAGGGGCGGGUCUGGG	UV-melting/CD	*In vitro* (RRL)	~70% (↓)	(73)
ZIC1	Zinc finger protein Zic-1	GGGUGGGGGGGGCGGGGGA-GGCCGGGG	UV-melting/CD	*In cellulo* (HeLa)	~80% (↓)	(85)
MT3-MMP	Matrix metallopeptidase 16	GAGGGAGGGAGGGAGAGGGA	CD/CD-melting/RNase T1 footprinting	*In cellulo* (HeLa)	~55% (↓)	(86)
ERS1	Estrogen receptor α	GGGUAGGGGCAAAGGGGCTGGGG	CD/CD melting	*In vitro* (RRL)	~85% (↓)	(87)
EBAG9	Estrogen receptor binding site associated antigen 9	GGAGCCUCCGCCGGGCGGGCGGG-GAGGG GGAGGGGGCAGGUUUUGA	CD/CD meting/in-line probing	*In cellulo* (HEK293)	~45% (↓)	(93)
FZD2	Frizzled family receptor	GGGGAAGAAGCGCAGUCUCCGG-UGGGGG CGGGGGCGGGGGGGC-GCCAAGGAGCCGGG	CD/CD-meting/in-line probing	*In cellulo* (HEK293)	~60% (↓)	(93)
BARHL1	BarH-like 1 homeobox protein	GGGGCAGCGGCGGCUGGGGUUG-GGGGUG GGUGGGGAGCUUUUGG-GG	CD/CD-meting/in-line probing	*In cellulo* (HEK293)	~50% (↓)	(93)
NCAM2	Neural cell adhesion molecule 2	GGAGGAGCGGCGGGGCUGCGGG-CGGCUGG GGCACCGCGGGAGCG-GCGGCGGCGG	CD/CD-meting/in-line probing	*In cellulo* (HEK293)	~35% (↓)	(93)
THRA	Thyroid hormone receptor α	GGGUGCUGUGCCCUAGGGCCUG-GGUGGCAG GGGGUGGGUGGCCU-GUGGG	CD/CD-meting/in-line probing	*In cellulo* (HEK293)	~35% (↓)	(93)
AASDHPPT	4′-phosphopan–etheinyl transferase	GGGGGCGGGCUGGGAGGGCUGU-CGGUGGGC CAGUCUGC	CD/CD-meting/in-line probing	*In cellulo* (HEK293)	~55% (↓)	(93)
BCl-2	Apoptosis regulator Bcl-2	GGGGGCCGUGGGGUGGGAGCUG-GGG	UV/UV-melting/CD	*In cellulo* (MCF10A, MCF7, HGC27)	~50% (↓)	(88)
TRF2	Telomeric repeat binding factor 2	CGGGAGGGCGGGGAGGGC	UV-melting/CD	*In cellulo* (293T)	~65% (↓)	(89)
ADAM10	α-secretase	GGGGGACGGGUAGGGGCGGAGGUAGGGG	CD/CD-melting	*In vitro* (RRL)	~70% (↓)	(90)
Cap-independent translation						
FGF-2	fibroblast growth factor 2	GGGGGAGGAGAACUGGGGGCGC-GGGGGCUGG UGGGUGUCGGGGG	Enzymatic and chemical footprinting[b]	*In cellulo* (SK-Hep1)	~60% (↑)	(40)
VEGF	vascular endothelial growth factor	GGAGGAGGGGAGGAGGA	CD/Enzymatic and chemical footprinting[b]	*In cellulo* (HeLa)	~80% (↑)	(95)

[a]% change in protein expression a reporter gene (luciferase or GFP) induced by the G-quadruplex-forming-sequence when compared to constructs where the sequence has been deleted or mutated. ↓ indicates a reduction and ↑ indicates an increase in protein expression.
[b]Footprinting experiments performed in the native context of the full-length 5′-UTR.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

In a more recent study, Morris *et al.* (95) have provided structural and functional evidence that a G-rich sequence within the 5′-UTR of human vascular endothelial growth factor (*hVEGF*) mRNA adopts an RNA G-quadruplex structure that is essential for IRES-mediated translation initiation. The *hVEGF* mRNA possesses a long (~1000-nt), GC-rich 5′-UTR that harbors two separate IRESs. Using RNase T1 and dimethylsulfate to probe to the 293-nt IRES-A structure, a 17-nt G-rich sequence (Table 1) was identified that adopts a G-quadruplex structure. Mutational analysis of the *hVEGF* IRES-A in the context of a bicistronic dual-luciferase reporter vector demonstrated that RNA G-quadruplex formation is essential for cap-independent translation initiation in HeLa cells. However, it was notable that not every mutation tested led to a complete loss of IRES activity. Only a quadruple mutant, lacking sufficient Gs to adopt an intramolecular G-quadruplex structure, was close to being completely inactive. Analysis of the primary sequence (GGAGGAGGGGGAGGAGGA) suggested that more than 20 two-quartet intramolecular G-quadruplexes are theoretically possible by using different combinations of the five G-stretches (underlined). In principle, such redundancy can still allow G-quadruplex formation despite the introduced mutations. The authors proposed that the *hVEGF* 5′-UTR PQS could act as a 'switchable' element using different G-stretches to promote G-quadruplex structure formation and to tune IRES-mediated translation initiation.

5′-UTR RNA G-QUADRUPLEXES AND PROTEINS

The number of reported RNA G-quadruplex-binding proteins is still relatively scarce compared to DNA G-quadruplex-binding proteins (96,97), and most of the biological data that link protein binding to 5′-UTR RNA G-quadruplexes with biological functions are mostly correlative rather than definitive.

The FMRP is possibly the best-studied RNA G-quadruplex-binding protein. FMRP is an RNA-binding protein whose aberrant expression is linked to a common form of intellectual disability, the fragile X syndrome (FXS) (98). An *in vitro* selection for RNA molecules that preferentially bind to the FMRP identified a large number of sequences containing PQSs (37), and indeed when FMRP-containing ribonucleoprotein complexes were immunoprecipitated from mouse brain ~70% of the associated mRNAs contained a PQS (38). Of particular importance in the regulation of FMRP expression is the demonstration that FMRP binds to its own mRNA via a G-quadruplex structure in the coding region (39). In addition, a large number of putative FMRP mRNA targets have been identified, which contain computationally predicted G-quadruplex motifs. Experimental evidence for the involvement of FMRP and related proteins in G-quadruplex-mediated mRNA metabolism has recently been reviewed (99). At least two mRNA targets of FMRP harbor one or more PQS in their 5′-UTR, namely the *MAP1B* and the *PP2A* genes (37,100,101). In both cases, absence of FMRP was associated with elevated levels of protein expression, suggesting that FMRP binds to the 5′-UTR G-quadruplex structures to repress translation, possibly by stabilizing the G-quadruplex structure and thus perturbing translation initiation (100,101).

In a recent study, we proposed that another class of translational regulatory protein, Pat1, selectively binds to RNA G-quadruplex structures (102). The Pat proteins are conserved RNA-binding proteins that have been shown to participate in processing (P)-bodies formation, which are ribonucleoprotein cytoplasmic foci involved in translation repression and mRNA decay (103). As in the case of FMRP, the *Xenopus* Pat proteins, xPat1a and xPat1b, as well the human Pat1b, were shown to preferentially bind poly r(G) *in vitro* (102). In competition assays with xPat1a and xPat1b, we found that an oligonucleotide of the *NRAS* RNA G-quadruplex sequence (Table 1), used as model RNA G-quadruplex-forming sequence, but not a mutated sequence that is unable to form a G-quadruplex, could efficiently compete for poly r(G) binding. Furthermore, a biotinylated *NRAS* RNA G-quadruplex, but not the mutated sequence, was able to isolate xPat1a protein from a *Xenopus* oocytes lysate using streptavidin-coated beads. Altogether, these results suggest that Pat1 proteins preferentially bind G-quartet containing RNA structures. Interestingly, other identified protein components of P-bodies, as for example, the yeast protein Stm1 and the exoribonuclease XRN1p, have been shown to preferentially interact with RNA G-quadruplex structures *in vitro* (104,105). However, more work is still needed to elucidate whether such interactions between proteins and 5′-UTR RNA G-quadruplexes is actually associated with P-body formation.

Because of the high thermodynanic stability of RNA G-quadruplexes *in vitro*, it is likely that resolving such structures *in vivo* would require specialized helicases. RHAU (also known as DHX36 or G4R1) is a member of the human DEAH-box family of RNA helicases that has been shown to bind with high affinity to RNA G-quadruplex *in vitro* and to unwind G-quadruplex structures more efficiently than double-stranded nucleic acids (106,107). RHAU has been shown to associate with mRNA *in cellulo* (108), and was identified as the main source of tetramolecular RNA G-quadruplex-resolving activity in HeLa cell lysates (107). In a recent study, Lattmann *et al.* (58) over-expressed RHAU in HeLa cells and employed RNA immunoprecipitation coupled with genome-wide microarray analysis (RIP-chip) to identify and quantify RNAs associated with the RHAU helicase. RHAU was found to be associated with 106 RNAs, of which >50% contained at least one PQS. Furthermore, the PQS density per transcript showed a small but significant correlation with the level of RNA enrichment by RHAU. These findings provide clear indications that RNA G-quadruplexes are physiologically relevant targets of RHAU in cells. Of particular note was the observation that nearly three-quarters of the identified PQSs were found within the 5′- and/or 3′-UTRs of mRNAs. Even though no role for RHAU in translation regulation has yet been demonstrated so far, this raises the interesting possibility that RHAU could be

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

involved in G-quadruplex-mediated translational control, or other aspects of RNA metabolism such as splicing or turnover. Recently, the human DHX9 helicase, which displays a high similarity with RHAU, has also been shown to bind and unwind RNA G-quadruplexes *in vitro* (109).

FXS is genetically characterized by expansion of a CGG repeat in the 5′-UTR of the *FMR1* gene, which is thought to cause loss of FMRP. Fry and co-workers (110) have provided some evidence that the CGG repeats can form G-quadruplex structures with translational modulatory function. They showed that a 99-nt r(CGG) sequence positioned upstream of a *luciferase* reporter gene in a *FMR1* promoter-driven plasmid repressed translation in HEK293 cells. However, normal translation efficiency could be restored upon co-expression of CBF-A and hnRNP A2 proteins, which were previously shown to destabilize bimolecular DNA tetraplex *in vitro* (111,112). These results suggest that a balance between G-quadruplex formation through the $r(CGG)_n$ repeats in the 5′-UTR *FMR1* mRNA and endogenous G-quadruplex-disrupting proteins controls FMRP expression.

5′-UTR RNA G-QUADRUPLEXES AS SMALL MOLECULE TARGETS

Achieving control of gene expression by using agents that target nucleic acids represents a major goal in the life sciences. One such approach is to target and inhibit mRNA function(s). Oligonucleotide-based agents such as antisenses, aptamers and, most notably, small interfering RNAs represent promising molecules for the silencing of mRNAs; however their poor pharmacological properties still represent an issue. Thus, interventions based on ‘drug-like’ synthetic small molecules could provide an alternative strategy to overcome the disadvantages associated with nucleic acids-based approaches, particularly with a view to ultimately translating chemical biology studies towards therapeutics (113).

Small molecules that bind to structural elements within the 5′-UTRs of mRNAs have been explored with a view to interfering with gene expression at the translation level (114). In eukaryotic translation studies, examples have mainly been limited to *in vitro* selected small molecule-binding RNA aptamers. Artificial systems were produced by inserting the aptamer sequences upstream of reporter genes. These systems were shown to be responsive to the small molecules, and provided proof-of-principle for small molecule-mediated translation inhibition by targeting structural elements in the 5′-UTRs of mRNA (115,116). These studies have provided proof-of-principle that small molecules known to bind to defined structural elements within the 5′-UTRs of eukaryotic mRNAs can be used to control the translation rate of transcripts. However, an important consideration in targeting cellular mRNAs with small molecules relates to identifying RNA structural elements that are amenable to selective molecular recognition. The special structural features of G-quadruplexes and the discovery of RNA G-quadruplex-forming-sequence in the 5′-UTRs of numerous genes, including several proto-oncogenes, led us to propose that such RNA motifs could be suitable targets for small molecules (73).

In principle small molecules targeting of 5′-UTR RNA G-quadruplex structures could modulate the mRNA translation according three mechanisms: (i) by stabilizing the RNA G-quadruplex structure and thus impairing the assembly and/or the scanning process of the 43S ribosomal complex; (ii) by destabilizing the RNA G-quadruplex structure and thus stimulating translation or (iii) by interfering with a biologically essential RNA G-quadruplex/protein binding event.

Using our previously developed *NRAS* 5′-UTR luciferase system, we evaluated the translational effect of several small molecules from our ligand collection that had previously been established to exhibit selective binding to DNA G-quadruplexes as compared to double-stranded DNA. In this study, we compared the *in vitro* translation efficiencies of both the wild-type *NRAS* 5′-UTR and a control 5′-UTR in the presence of increasing concentration of the G-quadruplex ligands (117). The control mRNA was derived from the wild-type *NRAS* 5′-UTR by deleting the 18-nt G-quadruplex sequence from the wild-type *NRAS* 5′-UTR. In an initial screening study, we found several molecules that had no effect on translation and others that inhibited translation in a G-quadruplex-independent manner. For example, the well-studied G-quadruplex ligand TMPyP4 showed translational inhibition without any RNA G-quadruplex specificity in this assay, probably because of its poor selectivity for G-quadruplex versus other nucleic acids structures (118,119). However, we also identified molecules that displayed selective translational inhibition depending on the presence of the RNA G-quadruplex. Among these, a pyridine-2,6-bis-quinolino-dicarboxamide derivative (RR82, Figure 3) reduced the translational efficiency of the *NRAS* 5′-UTR by ~50% at 1.25 μM concentration; whereas under the same conditions the control still exhibited ~80% translation efficiency. We then tested structural variants of this molecule and found that a *para*-fluorophenyl substituent at pyridine C4 (RR110, Figure 3) considerably improved G-quadruplex selectivity. At 10 μM concentration, RR110 had no effect on the translational efficiency of the control mRNA, but inhibited that *NRAS* 5′-UTR translation by ~40%. This inhibitory effect was retained in the presence of large excess of double-stranded DNA or hairpin RNA competitors, and it also compared favorably with studies based on *in vitro* selected RNA aptamers (115,116). Using hydrogen-deuterium exchange followed by ^{1}H-NMR, we demonstrated that RR110 stabilizes the *NRAS* RNA G-quadruplex and we performed an mRNA stability assay to show that the small molecule did not exert its effect by altering the mRNA degradation rate (117). In addition, we also showed that an RNA G-quadruplex motif that did not display any intrinsic translational inhibitory activity could become a translational inhibitory element upon binding to a G-quadruplex ligand, suggesting that a G-quadruplex-binding small molecule can trigger a translational effect (117). Collectively, these

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

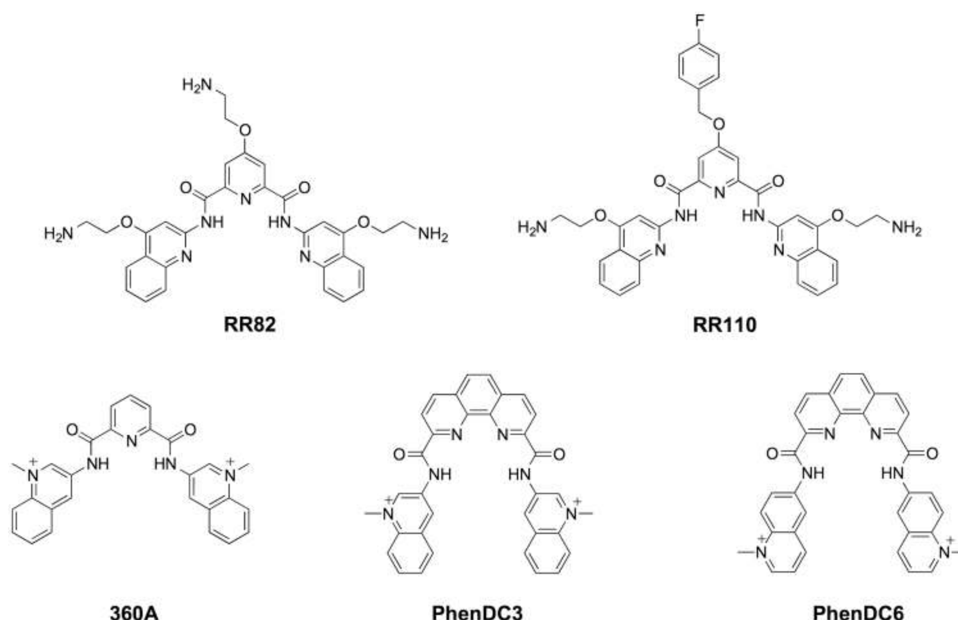

Figure 3. Chemical structures of synthetic molecules that have been demonstrated to exert selective RNA G-quadruplex mediated translation inhibition.

observations support a mechanism whereby stabilization of the *NRAS* RNA G-quadruplex affects initiation of the translation process and provided proof-of-concept that a small molecule G-quadruplex ligand is able to modulate translation via selectively binding to a 5′-UTR RNA G-quadruplex.

In a subsequent study, Gomez *et al.* (89) evaluated the potency of three bisquinolinium compounds (360A, PhenDC3 and PhenDC6, Figure 3) to bind the *TRF2* RNA G-quadruplex and to alter translation *in vitro*. These molecules were selected on the basis of previous studies showing their high efficacy to bind and stabilize DNA G-quadruplexes (120,121). The potential of these molecules to bind and stabilize the *TRF2* RNA G-quadruplex was first assessed in a FRET melting competition assay using a doubly labeled oligonucleotide that mimics the human telomeric DNA G-quadruplex. For all compounds, ligand-induced stabilization of the DNA G-quadruplex structure was strongly decreased in the presence of an excess of the *TRF2* RNA G-quadruplex-forming sequence, but was not affected by a 26-bp DNA duplex or a *TRF2* mutated RNA sequence that is unable to form a G-quadruplex. These results indicate that the three compounds are able to selectively bind the *TRF2* RNA G-quadruplex. Using an *in vitro* coupled transcription-translation assay, the authors went on to demonstrate that the binding of the ligands to the *TRF2* RNA G-quadruplex correlates with inhibition of mRNA translation. For example, at 3 μM concentration, the most potent compound in this assay, PhenDC3 (Figure 3), inhibited GFP protein expression from a 5′-UTR *TRF2* reporter plasmid by about 4-fold; whereas it did not affect the GFP expression from a control plasmid containing G-quadruplex-disrupting mutations. Furthermore, mRNA level analyses showed that protein synthesis inhibition was not due to a transcriptional effect, but rather arose from ligands interacting with RNA G-quadruplex to interfere with the translation process.

In a following study, Hartig and co-workers (122) used the same three bisquinolinium compounds to investigate whether these molecules could be used to target a 5′-UTR RNA G-quadruplex in a cellular context. In this study, fluorescence intercalator displacement (FID) and CD melting experiments were used to demonstrate the binding to and stabilization of rationally designed RNA G-quadruplex-forming sequences of the form $(G_3U_{1-3})_4$. FID experiments revealed high association constants to RNA G-quadruplexes, in the 10^7–$10^8\,M^{-1}$ range, for the three compounds. HEK293 cells were transfected with plasmid reporter constructs expressing the *Renilla* luciferase gene under control of 5′-UTRs containing different G-quadruplex sequences of the form $(G_3U_{1-3})_{4-6}$, before treating cells with increasing concentrations of the compounds. In all cases, a dose-dependent reduction in luciferase expression was observed as compared to untreated cells. The most notable effect was a 86% decrease in luciferase expression for a $(G_3U)_6$ containing plasmid when treated with 10 μM of PhenDC3 compound (Figure 3). In contrast, similar treatments showed little effect on the expression of plasmids containing G-rich control sequences unable to fold into G-quadruplex structures. G-quadruplex ligands did not affect mRNA abundance of the investigated reporter genes. Taken together, these results indicate that the bisquinolinium molecules can inhibit protein expression in a RNA G-quadruplex-dependent fashion within a cellular context.

In addition to the above-described studies using 'drug-like' small molecules, Ito *et al.* (123) have recently demonstrated the use of a short (9-nt long) G-rich RNA to

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

inhibit translation of a EGFP reporter mRNA in living cells through the formation of an intermolecular G-quadruplex in its 5′-UTR.

Whereas all these studies have exploited plasmid-based reporter assays, there is one report were an observed decrease in *Aurora A* protein level that accompanied a M-phase cell cycle arrest in cells treated with an hexaoxazole macrocycle G-quadruplex binder was hypothesized to arise through stabilization of an endogenous 5′-UTR G-quadruplex in the mRNA sequence of the *Aurora A* gene (124,125). However, no direct evidence for that has yet been provided, and further studies are required to confirm this possibility.

Although all the examples discussed so far suggest a mechanism by which small molecule stabilization of a 5′-UTR RNA G-quadruplex structure inhibits translation, there is also an example where a G-quadruplex ligand has been shown to increase the translation efficiency of a 5′-UTR RNA G-quadruplex-containing mRNA. In this study, Ofer *et al.* (126) showed that the cationic porphyrin TMPyP4 affected the electrophoretic mobility in non-denaturing agarose gel of an RNA construct containing a 5′-UTR $r(CGG)_{33}$ derived from the *FMR1* mRNA, which they previously proposed to fold into intramolecular G-quadruplex (110). In contrast, gel migration of a control RNA construct that did not contain the $r(CGG)_{33}$ tract was unaffected. Furthermore, the presence of potassium ions was required to observe an electophoretic mobility shift. Based on these results and on the previous observation that TMPyP4 destabilized bimolecular $(CGG)_n$ tetraplex structures (127), the authors proposed that the small molecule slowed electrophoretic migration through G-quadruplex unfolding (126). When assessed in functional assays, both by itself and synergistically with the CBF-A and hnRNP A2 proteins, TMPyP4 increased translation efficiency of a 5′-UTR $r(CGG)_{99}$ luciferase reporter mRNA *in vitro* in rabbit reticulocyte lysates and *in cellulo* in HEK293 tissue cultured cells. Since TMPyP4, but not its positional isomers TMPyP2 or TMPyP3, which were both previously shown to be unable to unfold $(CGG)_n$ G-quadruplexes (127), enhanced translation efficiency, these results suggested that the elevated translation efficiency of the 5′-UTR $r(CGG)_{99}$ reporter mRNA was a consequence of the ability of TMPyP4 to destabilize the translational blocking 5′-UTR G-quadruplex structure. Even though G-quadruplex formation by the $r(CGG)_n$ repeat element is still a subject of debate (128), and the selectivity of TMPyP4 for G-quadruplex is limited (118,119), these findings offer an alternative perspective on RNA G-quadruplex targeting by small molecule that warrant further investigation.

Finally, very recently, while this manuscript was under review, Xodo and co-workers (129) have reported on an alkyl derivative of TMPyP4 (TMpyP4-C14), which efficiently enter cells and preferentially localize into the cytoplasm. In this study, the authors demonstrated that TMPyP4-C14 binds to G-quadruplex structures in the 5′-UTR of *KRAS* mRNA and, upon photoactivation, selectively induces mRNA degradation, resulting in a about 90% downexpression of *KRAS* protein in pancreatic cancer cells.

CONCLUSION, CHALLENGES AND PERSPECTIVES

A great deal of scientific research has been carried out during the past 15–20 years to study the roles of RNA structures on gene expression regulation at the translational level (68–72,80–83). These studies have mainly concentrated on structural arrangements based on stem–loop formation, while a relatively early report demonstrated that an RNA G-quartet-based structure in an IRES motif within the 5′-UTR of a human *FGF2* gene transcript had a role in cap-independent translation initiation. It is <5 years ago since we reported that a naturally occurring intramolecular RNA G-quadruplex structure in the 5′-UTR of the human *NRAS* transcript inhibits translation (73). Furthermore a computational search of 5′-UTRs human transcripts for RNA PQSs led us to postulate that translational regulatory G-quadruplex structures in the mRNA 5′-UTRs may be common (73). Subsequently there have been many publications on additional genes that have reinforced our initial observation on the *NRAS* RNA G-quadruplex (Table 1). In addition to mRNA 5′-UTRs, RNA PQSs have also been computationally identified in 3′-UTRs, and in pre-mRNAs near splicing, transcription termination and polyadenylation sites, indicating that RNA G-quadruplexes might also regulate other stages of RNA metabolism (42,43,74). Indeed, several studies published within the past 4 years have provided experimental evidence that RNA G-quadruplex structures in introns affect the splicing and expression patterns of *Bcl-xL*, *FMR1* and *TP53* human genes (130–132). Furthermore, RNA G-quadruplex formation has recently been shown to control human mitochondrial transcription termination and mRNA localization in cortical neurites (133,134).

Whereas several computational algorithms exist that can reasonably predict G-quadruplex formation within DNA, such a tool for predicting RNA G-quadruplex more accurately would be beneficial. The detailed rules governing RNA G-quadruplex formation are just starting to be discovered. Given the high propensity of single-stranded RNA to fold into secondary structures and engage in long-range interactions, rules for G-quadruplex formation in RNA may be complex. The sequence and the local structural context could of particular importance for RNA G-quadruplex formation. Beaudoin and Perreault (93) have provided evidence that the presence of cytosine tracks within a putative G-quadruplex-forming sequence could be detrimental to G-quadruplex formation, presumably because they favor the formation of competing stem structures. In addition, a recent report by Patel and co-workers (67) has exemplified the importance of sequence context by showing the formation of a RNA G-quadruplex structure with an unprecedented fold, which contains a mixed-junction quartet that connects between the G-quadruplex and a flanking duplex stem. It is noteworthy that while several software packages exist to computationally predict RNA

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

secondary structures, none of them takes account of RNA G-quadruplex formation. Advancements in the prediction of RNA G-quadruplex formation would enhance future research in this field.

While evidence is emerging to suggest critical roles for RNA G-quadruplexes in biological processes, it is striking that none of the high-resolution crystallographic or NMR structures of naturally occurring complex RNA molecules, such as tRNAs, rRNAs or ribozymes, reported so far have revealed incident G-quartet structures. In a recent bioinformatics search of the model plant species *A. thaliana*, RNA PQSs were shown to be strongly underrepresented in non-coding RNA molecules, such as tRNAs, rRNAs and sno RNAs (75). These absences may suggest that RNA G-quadruplexes are unsuitable as long-lived architectural elements. Perhaps their roles are confined to be transient and regulatory? In light of the recent findings presented here, RNA G-quadruplex structures in the 5′-UTRs of mRNAs seem to represent key elements in translational regulation. A simple view would be that the presence of a RNA G-rich sequence with the required thermodynamic parameters is sufficient to form a stable G-quadruplex structure that affects translation. However, RNA G-quadruplex formation within the context of a 5′-UTR is likely to be in equilibrium with alternative secondary/tertiary structures, possibly involving cellular co-factors, and more complex mechanisms for translational regulation by 5′-UTR RNA G-quadruplexes could be involved (93). If this view is correct, an important and experimentally challenging goal will be to demonstrate where and when 5′-UTR RNA G-quadruplexes naturally exist within cellular mRNAs. For example, they may only be present at very defined points of the cell cycle or in particular cell states. Further evidence is needed to help us better understand whether such structures are merely incidental or tightly regulated.

The discovery of naturally occurring proteins with preferential RNA G-quadruplex recognition properties goes some way towards helping address the existence and function of naturally occurring RNA G-quadruplexes. Immunoprecipitation followed by genome-wide microarray analysis studies of the RNA-binding FMPR protein and the RHAU helicase have provided some evidence for RNA G-quadruplex in cellular mRNAs (38,58). The discovery of more RNA G-quadruplex-binding proteins will be a major step to further elucidate the occurrence and biological functions of 5′-UTR RNA G-quadruplexes. It should also be noted that more recent approaches to map RNA structures on a genome-wide scale based on enzymatic degradation of RNA populations have not yet taken account of RNA G-quadruplexes nor have they considered the cellular context of RNA structure bound to proteins *in vivo* (135–137). The application and refinement of such approaches and other methods, including SHAPE and CLIP-SEQ (138,139), could shed light on the existence and nature of RNA G-quadruplexes in cells.

There is now a growing body of evidence that has established a link between deregulation of translational control and disruption of normal cell behavior in human diseases, especially cancers (140–142). Due to their intricate roles in translational regulation, RNA *cis*-regulatory elements located in the 5′-UTRs are considered to be active players in translational regulation breakdowns (71,72). Despite the recent elucidation of translational control by 5′-UTR RNA G-quadruplexes and the identification of RNA G-quadruplex in a large number of genes, including numerous proto-oncogenes, no explicit link between RNA G-quadruplex formation/disruption and cancer development has yet been established; future studies may provide further insights.

The recent demonstrations that small molecule G-quadruplex binding ligands can selectively target RNA G-quadruplexes and derail translation initiation opens up a new and attractive avenue in RNA-directed drug design. Clearly, more research will be needed to rigorously validate RNA G-quadruplexes as drug targets for therapeutics applications and explore how selective ligands can be for a given RNA G-quadruplex. Part of the upcoming challenge will be to better understand the mechanistic effects and selectivity factors on endogenous mRNAs in the all complexity of a cellular and, ultimately, an *in vivo* environment. However, it is clear that the RNA G-quadruplex motif represents a structurally attractive scaffold for small molecule targeting and given the promising early insights into their functional effects, this represents an attractive and fertile area for future research.

ACKNOWLEDGEMENTS

We thank Dr David Tannahill for critical reading of this manuscript and Dr Pierre Murat for assistance with graphical material.

FUNDING

Research in Balasubramanian laboratory on RNA G-quadruplex is funded by the Biotechnology and Biological Sciences Research Council of the UK and Cancer Research UK. Funding for open access charge: Cancer Research UK.

Conflict of interest statement. None declared.

REFERENCES

1. Watson,J.D. and Crick,F.H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*, **171**, 964–967.
2. Bang,I. (1910) Untersuchungen über die Guanylsaüre. *Biochemische Zeitschrift*, **26**, 293–311.
3. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
4. Neidle,S. and Balasubramanian,S. (2006) *Quadruplex Nucleic Acids*. RSC Biomolecular Sciences, Cambridge, UK.
5. Moyzis,R.K., Buckingham,J.M., Cram,L.S., Dani,M., Deaven,L.L., Jones,M.D., Meyne,J., Ratliff,R.L. and Wu,J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl Acad. Sci. USA*, **85**, 6622–6626.
6. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

7. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
8. Kim,N.W., Piatyszek,M.A., Prowse,K.R., Harley,C.B., West,M.D., Ho,P.L., Coviello,G.M., Wright,W.E., Weinrich,S.L. and Shay,J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
9. Mergny,J.L., Riou,J.F., Mailliet,P., Teulade-Fichou,M.P. and Gilson,E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res.*, **30**, 839–865.
10. Xu,Y. (2011) Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.*, **40**, 2719–2740.
11. Salazar,M., Thompson,B.D., Kerwin,S.M. and Hurley,L.H. (1996) Thermally induced DNA.RNA hybrid to G-quadruplex transitions: possible implications for telomere synthesis by telomerase. *Biochemistry*, **35**, 16110–16115.
12. Sun,D., Thompson,B., Cathers,B.E., Salazar,M., Kerwin,S.M., Trent,J.O., Jenkins,T.C., Neidle,S. and Hurley,L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.*, **40**, 2113–2116.
13. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
14. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
15. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
16. Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
17. Du,Z., Kong,P., Gao,Y. and Li,N. (2007) Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun.*, **354**, 1067–1070.
18. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic. Acids Res.*, **35**, 406–413.
19. Zhao,Y., Du,Z. and Li,N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett.*, **581**, 1951–1956.
20. Hershman,S.G., Chen,Q., Lee,J.Y., Kozak,M.L., Yue,P., Wang,L.S. and Johnson,F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **36**, 144–156.
21. Eddy,J. and Maizels,N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.*, **36**, 1321–1333.
22. Verma,A., Halder,K., Halder,R., Yadav,V.K., Rawal,P., Thakur,R.K., Mohd,F., Sharma,A. and Chowdhury,S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem.*, **51**, 5641–5649.
23. Todd,A.K. and Neidle,S. (2008) The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.*, **36**, 2700–2704.
24. Du,Z., Zhao,Y. and Li,N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **18**, 233–241.
25. Yadav,V.K., Abraham,J.K., Mani,P., Kulshrestha,R. and Chowdhury,S. (2008) QuadBase: genome-wide database of G4 DNA–occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.*, **36**, D381–D385.
26. Du,Z., Zhao,Y. and Li,N. (2009) Genome-wide colonization of gene regulatory elements by G4 DNA motifs. *Nucleic Acids Res.*, **37**, 6784–6798.
27. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
28. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
29. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
30. Kim,J., Cheong,C. and Moore,P.B. (1991) Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature*, **351**, 331–332.
31. Cheong,C. and Moore,P.B. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry*, **31**, 8406–8414.
32. Pan,B., Shi,K. and Sundaralingam,M. (2006) Crystal structure of an RNA quadruplex containing inosine tetrad: implications for the roles of NH2 group in purine tetrads. *J. Mol. Biol.*, **363**, 451–459.
33. Sundquist,W.I. and Heaphy,S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl Acad. Sci. USA*, **90**, 3393–3397.
34. Horsburgh,B.C., Kollmus,H., Hauser,H. and Coen,D.M. (1996) Translational recoding induced by G-rich mRNA sequences that form unusual structures. *Cell*, **86**, 949–959.
35. Christiansen,J., Kofod,M. and Nielsen,F.C. (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Res.*, **22**, 5709–5716.
36. Oliver,A.W., Bogdarina,I., Schroeder,E., Taylor,I.A. and Kneale,G.G. (2000) Preferential binding of fd gene 5 protein to tetraplex nucleic acid structures. *J. Mol. Biol.*, **301**, 575–584.
37. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
38. Brown,V., Jin,P., Ceman,S., Darnell,J.C., O'Donnell,W.T., Tenenbaum,S.A., Jin,X., Feng,Y., Wilkinson,K.D., Keene,J.D. *et al*. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*, **107**, 477–487.
39. Schaeffer,C., Bardoni,B., Mandel,J.L., Ehresmann,B., Ehresmann,C. and Moine,H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.*, **20**, 4803–4813.
40. Bonnal,S., Schaeffer,C., Créancier,L., Clamens,S., Moine,H., Prats,A.C. and Vagner,S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.*, **278**, 39330–39336.
41. Gomez,D., Lemarteleur,T., Lacroix,L., Mailliet,P., Mergny,J.L. and Riou,J.F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.*, **32**, 371–379.
42. Kostadinov,R., Malhotra,N., Viotti,M., Shine,R., D'Antonio,L. and Bagga,P. (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D124.
43. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
44. Azzalin,C.M., Reichenbach,P., Khoriauli,L., Giulotto,E. and Lingner,J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798–801.
45. Schoeftner,S. and Blasco,M.A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell. Biol.*, **10**, 228–236.
46. Schoeftner,S. and Blasco,M.A. (2010) Chromatin regulation and non-coding RNAs at mammalian telomeres. *Semin. Cell. Dev. Biol.*, **21**, 186–193.
47. Xu,Y., Kaminaga,K. and Komiyama,M. (2008) Human telomeric RNA in G-quadruplex structure. *Nucleic Acids Symp. Ser.*, 175–176.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

48. Xu,Y., Kimura,T. and Komiyama,M. (2008) Human telomere RNA and DNA form an intermolecular G-quadruplex. *Nucleic Acids Symp. Ser.*, 169–170.
49. Xu,Y., Kaminaga,K. and Komiyama,M. (2008) G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution. *J. Am. Chem. Soc.*, **130**, 11179–11184.
50. Martadinata,H. and Phan,A.T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *J. Am. Chem. Soc.*, **131**, 2570–2578.
51. Xu,Y., Ishizuka,T., Kimura,T. and Komiyama,M. (2010) A U-tetrad stabilizes human telomeric RNA G-quadruplex structure. *J. Am. Chem. Soc.*, **132**, 7231–7233.
52. Collie,G.W., Parkinson,G.N., Neidle,S., Rosu,F., De Pauw,E. and Gabelica,V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.*, **132**, 9328–9334.
53. 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.
54. 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.
55. Xu,Y., Suzuki,Y., Ito,K. and Komiyama,M. (2010) Telomeric repeat-containing RNA structure in living cells. *Proc. Natl Acad. Sci. USA*, **107**, 14579–14584.
56. Gros,J., Guédin,A., Mergny,J.L. and Lacroix,L. (2008) G-Quadruplex formation interferes with P1 helix formation in the RNA component of telomerase hTERC. *Chembiochem*, **9**, 2075–2079.
57. Sexton,A.N. and Collins,K. (2011) The 5′ guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol. Cell. Biol.*, **31**, 736–743.
58. 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.
59. Saccá,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.
60. Arora,A. and Maiti,S. (2009) Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J. Phys. Chem. B*, **113**, 10515–10520.
61. 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.
62. Zhang,D.H. and Zhi,G.Y. (2010) Structure monomorphism of RNA G-quadruplex that is independent of surrounding condition. *J. Biotechnol.*, **150**, 6–10.
63. 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.
64. Zhang,A.Y., Bugaut,A. and Balasubramanian,S. (2011) A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, **50**, 7251–7258.
65. 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.
66. Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
67. 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.
68. Gray,N.K. and Hentze,M.W. (1994) Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.*, **19**, 195–200.
69. van der Velden,A.W. and Thomas,A.A. (1999) The role of the 5′ untranslated region of an mRNA in translation regulation during development. *Int. J. Biochem. Cell. Biol.*, **31**, 87–106.
70. Mignone,F., Gissi,C., Liuni,S. and Pesole,G. (2002) Untranslated regions of mRNAs. *Genome Biol.*, **3**, REVIEWS0004.
71. Pickering,B.M. and Willis,A.E. (2005) The implications of structured 5′ untranslated regions on translation and disease. *Semin. Cell. Dev. Biol.*, **16**, 39–47.
72. Chatterjee,S. and Pal,J.K. (2009) Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biol. Cell*, **101**, 251–262.
73. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
74. Huppert,J.L., Bugaut,A., Kumari,S. and Balasubramanian,S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
75. Mullen,M.A., Olson,K.J., Dallaire,P., Major,F., Assmann,S.M. and Bevilacqua,P.C. (2010) RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. *Nucleic Acids Res.*, **38**, 8149–8163.
76. Ji,X., Sun,H., Zhou,H., Xiang,J., Tang,Y. and Zhao,C. (2011) Research progress of RNA quadruplex. *Nucleic Acid Ther.*, **21**, 185–200.
77. Collie,G.W. and Parkinson,G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
78. Gebauer,F. and Hentze,M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.*, **5**, 827–835.
79. Baird,S.D., Turcotte,M., Korneluk,R.G. and Holcik,M. (2006) Searching for IRES. *RNA*, **12**, 1755–1785.
80. Kozak,M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA*, **83**, 2850–2854.
81. Kozak,M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.*, **9**, 5134–5142.
82. Koromilas,A.E., Lazaris-Karatzas,A. and Sonenberg,N. (1992) mRNAs containing extensive secondary structure in their 5′ non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.*, **11**, 4153–4158.
83. Babendure,J.R., Babendure,J.L., Ding,J.H. and Tsien,R.Y. (2006) Control of mammalian translation by mRNA structure near caps. *RNA*, **12**, 851–861.
84. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
85. Arora,A., Dutkiewicz,M., Scaria,V., Hariharan,M., Maiti,S. and Kurreck,J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.
86. Morris,M.J. and Basu,S. (2009) An unusually stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313–5319.
87. Balkwill,G.D., Derecka,K., Garner,T.P., Hodgman,C., Flint,A.P. and Searle,M.S. (2009) Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry*, **48**, 11487–11495.
88. Shahid,R., Bugaut,A. and Balasubramanian,S. (2010) The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, **49**, 8300–8306.
89. Gomez,D., Guédin,A., Mergny,J.L., Salles,B., Riou,J.F., Teulade-Fichou,M.P. and Calsou,P. (2010) A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.
90. Lammich,S., Kamp,F., Wagner,J., Nuscher,B., Zilow,S., Ludwig,A.K., Willem,M. and Haass,C. (2011) Translational Repression of the Disintegrin and Metalloprotease ADAM10 by a Stable G-quadruplex Secondary Structure in Its 5′-Untranslated Region. *J. Biol. Chem.*, **286**, 45063–45072.
91. Halder,K., Wieland,M. and Hartig,J.S. (2009) Predictable suppression of gene expression by 5′-UTR-based RNA quadruplexes. *Nucleic Acids Res.*, **37**, 6811–6817.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

92. 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.
93. Beaudoin,J.D. and Perreault,J.P. (2010) 5′-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.*, **38**, 7022–7036.
94. Regulski,E.E. and Breaker,R.R. (2008) In-line probing analysis of riboswitches. *Methods Mol. Biol.*, **419**, 53–67.
95. Morris,M.J., Negishi,Y., Pazsint,C., Schonhoft,J.D. and Basu,S. (2010) An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.*, **132**, 17831–17839.
96. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.*, **12**, 4336–4351.
97. Sissi,C., Gatto,B. and Palumbo,M. (2011) The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie*, **93**, 1219–1230.
98. Gécz,J., Shoubridge,C. and Corbett,M. (2009) The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet.*, **25**, 308–316.
99. Melko,M. and Bardoni,B. (2010) The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. *Biochimie*, **92**, 919–926.
100. Lu,R., Wang,H., Liang,Z., Ku,L., O'donnell,W.T., Li,W., Warren,S.T. and Feng,Y. (2004) The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proc. Natl Acad. Sci. USA*, **101**, 15201–15206.
101. Castets,M., Schaeffer,C., Bechara,E., Schenck,A., Khandjian,E.W., Luche,S., Moine,H., Rabilloud,T., Mandel,J.L. and Bardoni,B. (2005) FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum. Mol. Genet.*, **14**, 835–844.
102. Marnef,A., Maldonado,M., Bugaut,A., Balasubramanian,S., Kress,M., Weil,D. and Standart,N. (2010) Distinct functions of maternal and somatic Pat1 protein paralogs. *RNA*, **16**, 2094–2107.
103. Marnef,A. and Standart,N. (2010) Pat1 proteins: a life in translation, translation repression and mRNA decay. *Biochem. Soc. Trans.*, **38**, 1602–1607.
104. Bashkirov,V.I., Scherthan,H., Solinger,J.A., Buerstedde,J.M. and Heyer,W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell. Biol.*, **136**, 761–773.
105. Van Dyke,M.W., Nelson,L.D., Weilbaecher,R.G. and Mehta,D.V. (2004) Stm1p, a G4 quadruplex and purine motif triplex nucleic acid-binding protein, interacts with ribosomes and subtelomeric Y′ DNA in Saccharomyces cerevisiae. *J. Biol. Chem.*, **279**, 24323–24333.
106. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.
107. Creacy,S.D., Routh,E.D., Iwamoto,F., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2008) G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **283**, 34626–34634.
108. Chalupníková,K., Lattmann,S., Selak,N., Iwamoto,F., Fujiki,Y. and Nagamine,Y. (2008) Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.*, **283**, 35186–35198.
109. Chakraborty,P. and Grosse,F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, **10**, 654–665.
110. Khateb,S., Weisman-Shomer,P., Hershco-Shani,I., Ludwig,A.L. and Fry,M. (2007) The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.
111. Weisman-Shomer,P., Cohen,E. and Fry,M. (2002) Distinct domains in the CArG-box binding factor A destabilize tetraplex forms of the fragile X expanded sequence d(CGG)n. *Nucleic Acids Res.*, **30**, 3672–3681.
112. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
113. Aboul-ela,F. (2010) Strategies for the design of RNA-binding small molecules. *Future Med. Chem.*, **2**, 93–119.
114. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with small molecules. *Chem. Rev.*, **108**, 1171–1224.
115. Werstuck,G. and Green,M.R. (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science*, **282**, 296–298.
116. Harvey,I., Garneau,P. and Pelletier,J. (2002) Inhibition of translation by RNA-small molecule interactions. *RNA*, **8**, 452–463.
117. Bugaut,A., Rodriguez,R., Kumari,S., Hsu,S.T. and Balasubramanian,S. (2010) Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org. Biomol. Chem.*, **8**, 2771–2776.
118. Ren,J. and Chaires,J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.
119. Murat,P., Bonnet,R., Van der Heyden,A., Spinelli,N., Labbé,P., Monchaud,D., Teulade-Fichou,M.P., Dumy,P. and Defrancq,E. (2010) Template-assembled synthetic G-quadruplex (TASQ): a useful system for investigating the interactions of ligands with constrained quadruplex topologies. *Chemistry*, **16**, 6106–6114.
120. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.F., Mergny,J.L. *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.
121. De Cian,A., Delemos,E., Mergny,J.L., Teulade-Fichou,M.P. and Monchaud,D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.*, **129**, 1856–1857.
122. 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.
123. Ito,K., Go,S., Komiyama,M. and Xu,Y. (2011) Inhibition of translation by small RNA-stabilized mRNA structures in human cells. *J. Am. Chem. Soc.*, **133**, 19153–19159.
124. Tsai,Y.C., Qi,H., Lin,C.P., Lin,R.K., Kerrigan,J.E., Rzuczek,S.G., LaVoie,E.J., Rice,J.E., Pilch,D.S., Lyu,Y.L. *et al.* (2009) A G-quadruplex stabilizer induces M-phase cell cycle arrest. *J. Biol. Chem.*, **284**, 22535–22543.
125. Satyanarayana,M., Kim,Y.A., Rzuczek,S.G., Pilch,D.S., Liu,A.A., Liu,L.F., Rice,J.E. and LaVoie,E.J. (2010) Macrocyclic hexaoxazoles: influence of aminoalkyl substituents on RNA and DNA G-quadruplex stabilization and cytotoxicity. *Bioorg. Med. Chem. Lett.*, **20**, 3150–3154.
126. Ofer,N., Weisman-Shomer,P., Shklover,J. and Fry,M. (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
127. Weisman-Shomer,P., Cohen,E., Hershco,I., Khateb,S., Wolfovitz-Barchad,O., Hurley,L.H. and Fry,M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res.*, **31**, 3963–3970.
128. Sobczak,K., Michlewski,G., de Mezer,M., Kierzek,E., Krol,J., Olejniczak,M., Kierzek,R. and Krzyzosiak,W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
129. Faudale,M., Cogoi,S. and Xodo,L.E. (2012) Photoactivated cationic alkyl-substituted porphyrin binding to g4-RNA in the 5′-UTR of KRAS oncogene represses translation. *Chem. Commun.*, 874–876.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022

130. Hai,Y., Cao,W., Liu,G., Hong,S.P., Elela,S.A., Klinck,R., Chu,J. and Xie,J. (2008) A G-tract element in apoptotic agents-induced alternative splicing. *Nucleic Acids Res.*, **36**, 3320–3331.
131. 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.
132. 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.
133. Wanrooij,P.H., Uhler,J.P., Simonsson,T., Falkenberg,M. and Gustafsson,C.M. (2010) G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc. Natl Acad. Sci. USA*, **107**, 16072–16077.
134. 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.
135. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
136. Zheng,Q., Ryvkin,P., Li,F., Dragomir,I., Valladares,O., Yang,J., Cao,K., Wang,L.S. and Gregory,B.D. (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. *PLoS Genet.*, **6**, pii: e1001141.
137. Underwood,J.G., Uzilov,A.V., Katzman,S., Onodera,C.S., Mainzer,J.E., Mathews,D.H., Lowe,T.M., Salama,S.R. and Haussler,D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.
138. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA*, **106**, 97–102.
139. Wang,Z., Tollervey,J., Briese,M., Turner,D. and Ule,J. (2009) CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods*, **48**, 287–293.
140. Clemens,M.J. and Bommer,U.A. (1999) Translational control: the cancer connection. *Int. J. Biochem. Cell. Biol.*, **31**, 1–23.
141. Meric,F. and Hunt,K.K. (2002) Translation initiation in cancer: a novel target for therapy. *Mol. Cancer. Ther.*, **1**, 971–979.
142. Malina,A., Cencic,R. and Pelletier,J. (2011) Targeting translation dependence in cancer. *Oncotarget*, **2**, 76–88.

Downloaded from https://academic.oup.com/nar/article/40/11/4727/2409040 by U.S. Department of Justice user on 17 August 2022