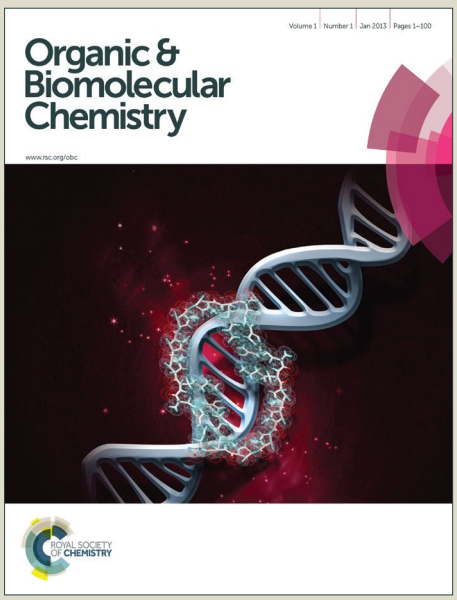
Organic & Biomolecular Chemistry

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Journal Name Organic & Biomolecular Chemistry Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Tale of RNA G-quadruplex

Prachi Agarwala [a,b], Satyaprakash Pandey [a,b] and Souvik Maiti [a,b,c]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

G-quadruplexes are non-canonical secondary structures found in guanine rich regions of DNA and RNA. Reports have indicated the wide occurrence of RNA G-quadruplexes across the transcriptome in various regions of mRNAs and non-coding RNAs. RNA G-quadruplexes have been implicated to play important role in translational regulation, mRNA processing events and maintenance of chromosomal end integrity. In this review, we summarize the structural and functional aspects of RNA G-quadruplexes with emphasis on the recent progress to understand the protein/trans factors binding these motifs. With the revelation of importance of these secondary structures as regulatory modules in biology, we also have evaluated various advancements towards targeting these structures and challenges associated with it. Apart from this, numerous potential applications of this secondary motif have also been discussed.

Introduction

G-quadruplexes are non canonical four stranded secondary structures found in nucleic acid sequences rich in guanine residues. These structures are formed of stacks of G-quartets-cyclic planar arrangement of four Hoogsteen hydrogen bonded guanine residues (Fig. 1).[1] Bang *et al* demonstrated that high concentrations of guanylic acid (GMP) can form gels in aqueous solution as early as 1910.[2] Fifty years later, Gellert *et al* revealed G-quadruplex structure based on X-ray fiber diffraction studies,[3] where each guanine residue in G-quartet acts as both acceptor and donor of two hydrogen bonds.

G-quadruplexes exhibit diverse topologies depending on monovalent cation (K^+ or Na^+), glycosidic conformation (*syn* or *anti*), number of molecules of nucleic acid involved in their formation (intramolecular, bimolecular or tetramolecular), relative orientation of the strands (parallel or antiparallel), number of stacking G-quartets and nucleotide sequence.[4, 5] Monovalent cations stabilise these structures by interacting with the negatively charged carbonyl oxygen atoms located at the centre of G-quartets stacks.[6] These monovalent cations stabilise the structure in order of $K^+ > Na^+ > Li^+$.[7]

There are about 3,76,000 DNA G-quadruplex forming sequences in the human genome. Whilst initial studies dealt with understanding the role of G-quadruplexes at telomere ends, DNA G-quadruplexes are found enriched in the promoters of proto-oncogenes, immunoglobulin heavy chain switch regions and mutational hotspots; and have been implicated in the maintenance of chromosomal integrity and regulation of replication, transcription and recombination processes.[8] Owing to the functional association between telomeres, oncogenes and cancer, extensive research on targeting DNA G-quadruplexes for therapeutic intervention is an ongoing active area of research.[9, 10]

Albeit considerable amount of work has been conducted on DNA G-quadruplex, recent years encountered a gradual shift towards understanding the G-quadruplexes at the RNA level. In 1991, Kim *et al* showed that 19 nt sequence situated at 3′ terminus of 5S rRNA of *Escherichia coli* forms a G-quadruplex.[11] Later, RNA G-quadruplex was reported to be present in the 3′ untranslated region (UTR) of insulin-like growth factor II (IGF II) mRNA located just downstream of endonucleolytic cleavage site.[12] Subsequently, various studies have reported the presence of RNA G-quadruplex in the mRNA, long non coding RNAs and telomeric ends.[5, 13, 14] It is sited in untranslated region (UTR), intronic and coding regions of mRNA further cementing its regulatory potential.[15]

The existence of RNA G-quadruplex is more inevitable than DNA counterpart. DNA exists in double standed condition, being base paired with the complementary strand. However, RNA being single stranded can adopt complex higher order architecture and is majorly localised in the cytoplasm.[16] Various RNA secondary structures such as hairpin, loop, bulge with functional implications have been reported in literature and discussed in detail elsewhere.[17-20] RNA G-quadruplex representing one of such structures have been associated with translational regulation, 3′ end processing, transcription termination, alternative splicing, mRNA localization, protein

binding and telomeric RNA biology.[15, 16, 21] Recently, Balasubramanian and coworkers for the first time demonstrated visual localisation of RNA G-quadruplex within the cytoplasm of human cells using G-quadruplex specific antibody.[22] With information on biological importance of RNA G-quadruplex pouring in recent years, here we have reviewed the structural and functional aspects of RNA G-quadruplexes and highlighted various applications of this structure.

Structure

RNA G-quadruplex forms a more thermodynamically stable, compact and less hydrated structure than DNA G-quadruplex (Fig. 2).[23, 24] Conceptually, the fundamental difference between RNA and DNA quadruplexes is the presence of ribose sugar instead of deoxyribose sugar and the presence of uracil instead of thymine residue. The presence of 2′ hydroxyl group in its ribose sugars causes more intramolecular interactions within the structure leading to enhanced stability of RNA G-quadruplex. In matter of fact, these 2′ hydroxyl groups exert conformational constraints on G-quadruplex topology preventing it from attaining *syn*-conformation, a prerequisite for antiparallel topology. Instead it restricts orientation of base about glycosidic bond to anti- conformation, imparting C3′ endo puckering and limiting the topology of RNA G-quadruplex to parallel conformation (Fig. 1B).[25] This implies less dependency of RNA G-quadruplex topology on the environmental conditions as compared to DNA G-quadruplex.[26, 27] DNA G-quadruplex can form both parallel and antiparallel conformations. In parallel conformation, the strands in G-quadruplex are oriented in the same direction and propeller loops connects the top of one strand with the bottom of another strand. On the other hand, strands are oriented in opposite directions in antiparallel conformation and lateral loops connects adjacent strands or/and diagonal loops connects diagonally opposite aiding antiparallel arrangement of strands. Overall, this monomorphic parallel topology reduces the diversity of RNA G-quadruplex structures as compared to DNA G-quadruplex.[28]

Using naturally occurring sequences of bcl-2 gene, Sugimoto and coworkers compared the stability of RNA and DNA G-quadruplex and revealed their significantly different hydration pattern in spite of their exposition of parallel conformation.[28] On the similar line, studies on naturally occurring RNA sequences and their DNA counterparts affirm more stable nature of RNA quadruplex.[23, 24, 28] The inherent chemistry of 2′ hydroxyl group of ribose sugar confers increased stability to RNA G-quadruplex, whose significance is well highlighted in reports where it is replaced with different OH analogues. Substitution of 2´-OH group by other chemically modified analogues culminated into destabilisation of G-quadruplex.[29] RNA quadruplex was found to be more robust than its DNA counterpart even in the presence of polyethylene glycol.[26, 27] Thus, higher stability of RNA G-quadruplex is attributed to the 2´-OH group of the ribose sugar which acts as scaffold for an ordered network of water molecules and bonding patterns.[25]

Mergny and group performed an elegant experiment by substituting thymine with uracil in the loop to G-quadruplex and revealed that it stabilises G-quadruplexes and makes it less hydrated. The removal of methyl group stabilizes the stacking interactions and releases the structured water molecules. By thymine to uracil substitution, the authors also suggested the importance of loop residues in the G-quadruplex stability.[30]

In addition, various independent studies have demonstrated the effect of varying loop lengths and composition on the stability of G-quadruplexes at both DNA and RNA level.[31-39] All of them demonstrated- shorter the loop, greater is the stability of RNA/DNA G-quadruplex.[39] In contrast to more importance of middle quartet in maintaining DNA G-quadruplex stability,[40, 41] all three G-quartets were found to contribute equally to the stability of RNA G-quadruplex.[42] Type of monovalent cation also affects G-quadruplex stability. Potassium ions are preferred over sodium ions because dehydration of the cations prior to their entry in the central channel of stacks formed by G-tetrads causes more energy loss for Na^+ than for K^+.[5, 43]

Recently, studies on folding dynamics by Mullen *et al* suggested that RNA G-quadruplex formation with two G-quartets show positive cooperativity, high dependence on K^+ concentration and very few populating intermediate states.[44] On the other hand, G-quadruplex formation with three G-quartets demonstrate less dependence on K^+ ions concentration and no or even negative cooperativity. Intentional increase in intermediate state population was also reported to broaden the range of K^+ ions concentration detected by RNA G-quadruplex.[45]

Although several NMR and crystallography studies have been performed on DNA G-quadruplexes,[46, 47] only few studies have been conducted on RNA G-quadruplex. NMR spectrum of a RNA quadruplex was first reported by Uesugi group for RNA G-quadruplex sequence R14 (GGAGGUUUUGGAGG) present in mRNAs of immunoglobin regions.[48] The reported NMR structure elucidates the dimeric parallel quadruplex structure of R14 comprising two stacks of G-quartets. The same group further conducted comparative studies between R14 and D14 sequences and revealed that D14 forms an antiparallel DNA quadruplex with diagonal loops in contrast to R14's parallel RNA quadruplex conformation with propeller loops.[49] Solution based studies of NMR were further extended to the crystallography studies by Collie *et al* where the 2′ hydroxyl group and rigid C3′-*endo* sugar puckering in RNA G-quadruplex were shown to redefine the hydration structure in grooves and hydrogen bonding patterns (Fig. 1C).[50] Thereafter, the crystal structure of TERRA G-quadruplex- acridine small molecule complex was also accounted by the same group where 2′ hydroxyl group of ribose sugar was found to be involved in binding with acridine.[51] All of the above mentioned NMR and crystallography studies have been performed on

intermolecular RNA G-quadruplexes. However, no such studies have been conducted on intramolecular RNA G-quadruplex yet.

In addition, telomeric repeat containing RNA (TERRA) has been studied by electrospray ionisation mass spectroscopy (ESI-MS). These studies demonstrate formation of higher order assemblies of RNA quadruplexes of variable lengths (12 mer, 22mer, 45mer) of human telomeric RNA sequences.[52] This stacking nature of RNA quadruplexes is comparable with "beads on a string" model for the G-quadruplex assembly.[53, 54] T1 RNase digestion of TERRA generating a ladder of bands with size in integral multiple of 24 nucleotides further supported this model.[55] Lately, NMR structure of TERRA also revealed 5′-5′ end stacking of two RNA G-quadruplexes.[56]

Function

While the main focus of research has been on DNA G-quadruplexes and their potential role in biology, in recent years the role of RNA G-quadruplexes as regulatory elements of gene expression is beginning to emerge (Table 1 and Fig. 3). In this regard, the initial emphasis was on G-quadruplexes located in the 5′ UTR of mRNA, which is known to be involved in translational regulation.

1. Translational regulation

The presence of G-quadruplex structures in the 5′ UTR of cellular mRNAs is generally associated with its role in translational repression. The first example of translational repression by an RNA G-quadruplex was reported by Balasubramanian group in the 5′ UTR of gene transcript of the human NRAS (Neuroblastoma RAS) proto-oncogene.[13] Using a cell-free translation reporter assay, they demonstrated that thermodynamically stable G-quadruplex in NRAS mRNA inhibits gene expression. However, their studies were performed under *in vitro* conditions. Subsequently our group for the first time examined the role of RNA G-quadruplex present in the 5′ UTR of the human Zic-1 (Zinc finger of the cerebellum 1) zinc finger protein in eukaryotic cells using dual luciferase reporter assay and reported its inhibitory role at the translational level.[57]

Genome-wide computational analysis revealed the presence of 4141 5′ UTR G-quadruplex motifs in 5′ UTR of mRNAs [58]. Following NRAS and Zic-1, the functional relevance of this secondary structure has been validated experimentally in genes – FGF-2 (Fibroblast growth factor-2),[59] NRAS,[13] Zic-1,[57] Exon C of human and bovine ESR (estrogen receptor α),[60, 61] MT3-MMP (Membrane-type 3 matrix metalloproteinase),[62] Bcl2 (B-cell lymphoma 2),[63] TRF-2 (Telomeric repeat-binding factor 2),[64] VEGF (Vascular endothelial growth factor),[65] ADAM-10 (Disintegrin and metalloproteinase domain-containing protein 10),[66] cyclin D3[67] and TGFβ2 (Transforming growth factor β2).[68] RNA G-quadruplex in the 5′ UTR of most of the aforesaid genes was shown to repress gene expression at the translational level.[69] In most of the above studies and in a study by Halder *et al* showed that G-quadruplex-forming sequences alone when present in 5′ UTRs exhibit translational repressive activity.[70] Recently, transcriptome scale ribosomal profiling of helicase eIF4A-dependent transcripts revealed the enrichment of 5′ UTR sequences with propensity to form RNA G-quadruplexes, indicating the probable role of this factor in unwinding of this secondary structure during translation process.[71]

Albeit most studies indicate translational inhibitory role of 5′ UTR G-quadruplex, few examples indicate its contrasting role. The first translational up-regulating role of RNA G-quadruplex was revealed in the 5′-UTR of mRNA associated with cap-independent translation. Bonnal *et al* was pioneer to show gene –up-regulating role of G-quadruplex motif in internal entry of ribosomes (IRES) sequence.[59] They reported that a G-quadruplex structure in conjunction with two stem-loop structures located within the FGF-2 IRES was required for translation initiation. Subsequently, Basu and coworkers showcased a unique example wherein a quadruplex is exclusively required for cap-independent translation initiation in human VEGF IRES.[65] The authors here reported that presence of multiple G tracts in IRES-A might direct initiation process to depend on a conformationally flexible G-quadruplex switch toggling in response to a biological stimulus. They gave it a grade of a tunable translation initiator equivalent to a riboswitch, which generally changes conformation in response to metabolite binding. However, these studies highlighted RNA G-quadruplex role in IRES and not in cap dependent translation. In this direction, our group for first time established the activating role of RNA G-quaduplex in 5′ UTR of TGFβ2 where it was demonstrated to augment gene expression.[68] Subsequently, we also indicated the translational up-regulating role of this structure in 5′ UTR of FOXE3 (Forkhead box E3).[72]

Furthermore, in order to show gene regulatory role of G-quadruplexes, Wieland and Hartig designed and inserted a series of quadruplex forming sequences surrounding the Shine-Dalgarno sequence of a GFP reporter gene.[73, 74] They showed that these motifs readily fold *in vivo* and modulate gene expression if positioned within the ribosome binding site (RBS), thereby providing a novel way to interfere with translational initiation in bacteria. Moreover, even less stable G-quadruplexes composed of only two stacks of G tetrads was reported to exhibit pronounced gene regulating activity when placed at critical positions in ribosome binding site.

Given, the hypothesis that position of G-quadruplex in 5′ UTR might affect its function, two independent groups shifted the position of G-quadruplex in 5′ UTR to understand position specific effect of G-quadruplexes. Kumari et al reported that shifting this motif in NRAS 5′ UTR at the position far away from cap region causes loss of its function whereas Huppert et al demonstrated that this shift in mRNA does not change the function of RNA G-quadruplex.[70, 75] The difference in outcome of shift may be attributed to the different mRNAs or different experimental setups they employed for their studies. To further

delineate the position specific role of G-quadruplex, our group assessed its role when located at 5′ end of naturally occurring mRNAs and found that G-quadruplex exhibited varied functions in different genes.[72] It inhibited gene expression in AKTIP (AKT interacting protein) and CTSB (Cathepsin B), but augmented protein production in FOXE3 suggesting that neighbouring nucleotide sequence, *trans* factors and intracellular conditions might affect the function of G-quadruplex. From above studies, we hypothesise that inhibitory G-quadruplex may itself act as a roadblock impeding the movement of per-initiation complex or may recruit some inhibitory factors repressing mRNA translation (Fig. 4A and Fig. 4B). On the other hand, activating G-quadruplex may engage activating factors to augment gene expression (Fig. 4C). More studies are required to decipher the mechanistic understanding of G-quadruplexes in this direction.

RNA G-quadruplexes influence the gene expression not only via their location in UTRs but also in the coding sequence of mRNA. Upon evaluating the role of G-quadruplex in open reading frame (ORF), it was revealed that it suppresses translation and the extent of gene suppression depends on its stability.[76, 77] G-quadruplex present in ORF of human estrogen receptor α was also shown to halt translational elongation and influence protein folding and proteolysis.[78, 79] Subsequently, ORF G-quadruplex was demonstrated to cause frameshift mutation of one nucleotide leading to the recoding process.[80, 81] In essence, these studies show that G-quadruplex can impede the process of translation at elongation level despite of helicase activity of ribosome.

In contrast to 5′ UTR G-quadruplexes, very few studies have been performed to unravel the translational regulatory role of G-quadruplex in 3′ UTR of mRNAs. G-quadruplex in the proto-oncogene PIM1 3′ UTR was reported to cause a translational suppression, though the mechanism of suppression is unclear.[82]

While above studies discussed upon the implication of intramolecular RNA G-quadruplexes on gene translation, Ito *et al* studies indicated intermolecular RNA G-quadruplex influence upon translational process. Using reporter systems, they demonstrated that G-rich RNA can bind G-rich sequences in UTRs or coding region of mRNAs through RNA-RNA interaction forming intermolecular G-quadruplex that can cause gene inhibition *in cellulo* conditions.[83] In the similar line, a recent report by Basu and coworkers indicated that DNA oligonucleotide induced hybrid DNA: RNA G-quadruplex formation in 5′ UTR and protein coding region of eIF4E mRNA causes reduced eIF4E protein production in human cancer cells.[84] Essentially, these reports extended the concept of small nucleic acid mediated gene regulation that can occur through intermolecular structure formation in living cells.

2. 3′ end processing and alternative polyadenylation

G-quadruplex function in physiology was initiated with the discovery of a G quadruplex structure downstream of an endonucleolytic cleavage site in the 3′ UTR of IGF2 (insulin-like growth factor 2) mRNA.[12, 85] Since this transcript undergoes a specific cleavage reaction at a site just 5′ to this motif, it suggested the role of G-quadruplex in targeting mRNA- processing events.[85] Subsequently, G-quadruplex formation was found to assist mRNA 3′ end processing in p53 in the event of DNA damage (Fig. 5). Decorsiere *et al* demonstrated that the presence of stable G-quadruplex structure downstream from the p53 pre-mRNA cleavage site is critical for maintaining p53 3′ end processing efficiency following UV irradiation mediated DNA damage. This *cis* element via interaction with the hnRNP H/F protein maintains p53 expression and leads to apoptosis following DNA damage.[86]

Lately, G-quadruplexes located in the 3′ UTR of LRP5 (Low-density lipoprotein receptor-related protein 5) and FXR1 (fragile X mental retardation 1) mRNAs were shown to increase the efficiency of alternative polyadenylation leading to generation of several short transcripts. These short transcripts production was further implicated to impair miRNA regulation of FXR1 mRNA.[87]

3. Mitochondrial transcription termination

The role of 3′ UTR G-quadruplexes has been further explored in the diverse areas of gene regulation. It has been implicated to aid transcriptional termination of mitochondrial light strand in manner analogues to Rho-independent transcription termination in prokaryotes, where a G-quadruplex motif substitutes the hairpin loop formed in bacterial mRNA.[88] It was also exhibited to generate primer for heavy strand DNA replication.

4. mRNA localisation

3′ UTR RNA G-quadruplexes have been shown to act as structural "zipcodes" assisting the process of mRNA localization in neurons.[89] G quadruplexes in the 3′ UTR of two mRNAs- PSD-95 (Postsynaptic density protein 95) and CaMKIIa (Calcium/calmodulin-dependent protein kinase IIa) are required for their localisation from soma to dendrites of cortical neurites. Since these mRNAs were previously described as targets of FMRP, the authors demonstrated that FMRP depletion did not abolish the transport thus negating the possibility of FMRP role in ferrying the RNA to the neurites. Bioinformatic analyses further showed the presence of putative G-quadruplex sequence in the 3′ UTR of 30% highly conserved dendritic mRNAs suggesting that these neurite-targeting elements are important for the establishment and maintenance of cell polarity.[89] More studies on these dendritic mRNAs as well as other mRNAs would reveal the mechanistic understanding of zipcoding function of RNA G-quadruplex.

5. Alternative Splicing

G-quadruplexes located in intronic regions of mRNA also affect alternative splicing- a mechanism of producing multiple protein products from a limited number of genes. T*rans* elements namely proteins of the hnRNP family play crucial

roles in splicing (Fig. 5). Genome wide studies indicated that G-rich tracts in introns of many genes exhibit potentiality to form G-quadruplexes and thus affect their splicing and expression patterns.[90] The first report on *cis*-regulatory element G-quadruplex role in alternative splicing demonstrated that hnRNP F binds to a G-rich tract in the pre-mRNA which could form a G quadruplex structure and is essential in splicing of the scr N1 exon.[91] Subsequently, G-quadruplex mediated regulation of alternate splicing was demonstrated to occur in a number of physiologically important genes such as B-tropomyosin,[92] hTERT (human telomerase reverse transcriptase),[93] FMR1 (Fragile X mental retardation 1),[94] p53,[95] BACE-1 (beta-site APP-cleaving enzyme-1) [96] and PAX9 (Paired box gene 9).[97] Interestingly, alternative splicing was found to be different in p53 compared to rest of the examples cited above, as the authors showed that G-quadruplex in p53 is located in another intron (intron 3) than the one affected by alternative splicing (intron 2). Additionally, they showed that G-quadruplex presence in p53 did not induce a complete shift from one splicing pattern to the other, but rather modified the equilibrium between the two spliced products in a subtle manner, leading to differential expression of transcripts encoding distinct p53 isoforms;[95] further indicating the ability of G-quadruplex motifs to fine tune the levels of different isoforms. There is a need to explore the prevalence of this kind of splicing regulation in other genes as well.

Proteins interacting with the RNA G-quadruplex

The number of RNA G-quadruplex binding proteins reported is relatively very less in contrast to proteins known binding to DNA G-quadruplex.[98, 99] Among them, FMRP is most extensively studied RNA G-quadruplex binding protein.

1. FMRP

In recent years, the role of RNA G-quadruplex upon interacting with two RNA-binding proteins: FMRP (Fragile X Mental Retardation Protein) and FMR2P (Fragile X Mental Retardation 2 protein) has been extensively studied.[100-102] Impaired production of these proteins is associated with two forms of mental diseases: the Fragile X Mental Retardation syndrome (FXS) and the FRAXE-associated mental retardation (FRAXE).[103]

FMRP protein has been shown to bind RNA targets harboring a G quadruplex motif with high affinity and specificity via its arginine–glycine–glycine (RGG) box.[101, 104] RNA binding studies have revealed that an array of FMRP's RNA targets have G-quadruplexes in their 5′ UTR,[105, 106] coding regions [107, 108] and 3′ UTR.[109, 110] RNA targets of FMRP, particularly Microtubules Associated Protein 1b (MAP1B) and catalytic subunit of Protein Phosphatase 2A (PP2Ac) evinced the presence of one or more G-quadruplex structures in their 5′ UTR.[101, 106] FMRP binds to these motifs in MAP1B mRNA and inhibits its translation during active synaptogenesis in neonatal brain development. In the absence of FMRP, increased MAP1B protein expression is shown to cause abnormally enhanced microtubule stability, which in turn interferes with the normal development of dendritic spines.[105] FMRP also negatively regulates PP2Ac mRNA translation by binding with its 5′ UTR G-quadruplex. In the absence of FMRP, the expression of PP2Ac increases and thereby alters actin remodelling in fibroblast cell lines.[106] Additionally, a report by Khateb *et al* showed that the longer (CGG)n sequences in 5′ UTR of FMR1 gene in FXS form G-quadruplex secondary structures, blocking the migration of the 40S subunit and thereby leading to impeded translation of FMR1 mRNA. Indeed, expression of the quadruplex disrupting proteins hnRNP A2 or CBFA were shown to alleviate this translational block.[111]

RNA targets of FMRP harbouring G-quadruplex in their 3′ UTR - PSD95 [110] and Sem3F (Semaphorin 3F),[109, 112] mRNAs have been reported.[103] Melko *et al* demonstrated that the association of Sem3F mRNAs to polyribosomes is decreased in absence of FMRP, thereby leading to their reduced expression. Hence, they indirectly suggested that FMRP in this instance may act as translational activator by sequestration or may influence mRNA stability.[103]

Interestingly, FMRP was found to regulate its own activity by binding to G-quadruplex present in coding region of its mRNA and inhibiting its translation by a negative feedback loop.[107] Likewise, FMRP binds to G-quadruplex in coding region of APP (Amyloid Precursor Protein) and regulate its translation.[108] Besides this modulation in gene expression, a G-quadruplex structure present in exon 15 of FMR1 gene is depicted essential for alternative splicing of exon 14,[94] where it has been shown to be a potent exonic splicing enhancer.[102]

2. Helicase proteins

RHAU is DEAH box containing RNA helicase that unwind both RNA and DNA G-quadruplex.[113-117] Although N-terminal region of RHAU protein known as RHAU specific motif (RSM) harbors G-quadruplex recognition ability, its helicase activity is found to require the full length protein to unfold G-quadruplex structure.[118, 119] This helicase remodel the G-quadruplex located in 5′ end of telomerase RNA and resolve G-quadruplex increasing the expression of Yin Yang 1 gene.[117, 120] In contrary, RHAU (DHX36) RNA helicase was recently reported to bind to the G-quadruplex present in the 3′ UTR of PITX1 mRNA and suppress its expression.[121] The authors further emphasised that these changes are brought about in association with argonaute-2 protein, a part of RISC silencing complex involved in microRNA mediated gene regulation, indicating a novel role of RHAU in microRNA mediated gene regulation of G-quadruplex containing 3′ UTR in some mRNAs. Another helicase DHX9 also has been reported to unwind RNA and DNA G-quadruplexes.[122] This increasing repertoire of G-quadruplex resolving helicases would likely increase our overall understanding of mechanism of unfolding

and maintenance of equilibrium between folded and unfolded states of this structure.

3. Other proteins

Hexanucleotide repeat expansion (HRE), (GGGGCC)n in the non-coding region of C9orf72 is associated with neurodegenerative diseases- amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). This HRE can form DNA G-quadruplex and promote transcriptionally induced RNA.DNA hydrid called R-loop formation.[123] Increase in number of repeats in HRE under pathological conditions display increased structural polymorphism causing increased transcriptional pauses, thereby leading to accumulation of abortive transcripts containing HRE. These transcripts forming RNA G-quadruplex were found to entrap several ribonucleoproteins such as nucleolin leading to nucleolar stress and molecular cascades culminating into ALS and FTD pathologies.[123] Recently, these transcripts were shown to undergo repeat associated non-ATG translation generating C9RAN proteins which are specifically detected in cerebrospinal fluid of ALS patients, providing it a status of a biomarker of this disease.[124]

In addition, TLS/FUS- telomere binding protein has been shown to bind telomeric DNA G-quadruplex and TERRA RNA G-quadruplex via its RGG domain to form a ternary complex. In matter of fact, these G-quadruplexes has been illustrated to serve as a scaffold for recruitment of these proteins to telomere, thereby regulating telomere length by histone modifications of telomere.[125]

Furthermore, Frantz and Gilbert group purified and characterized two activities from yeast, G4p1 and G4p2 and found that both these activities bind G-quadruplex structures.[126, 127] Subsequently, Xenopus Pat 1 proteins were also demonstrated to bind RNA G-quadruplexes.[128]

Recently, pull down studies followed by matrix assisted laser desorption ionisation time of flight mass spectroscopy (MALDI-TOF MS) were utilised to identify the interacting partners of G-quadruplex located in 5′ UTR of MMP6 and ARPC2 (Actin-related protein 2/3 complex subunit 2) mRNAs.[129] Binding partners were found to predominantly comprise ribosomal proteins, heterogeneous nuclear ribonucleoproteins, nucleolin and splicing factors. Few novel proteins previously unknown to bind nucleic acids were also identified.[129] However, the mode of action of these proteins was not elucidated in these studies.

Telomeric RNA G-quadruplexes

In biological systems, functions of G-quadruplexes were first elucidated in the context of telomeric regions of the chromosomes, where highest genomic prevalence of this secondary structure occurs.[130] The telomeric regions (2–10 kb) of mammalian chromosomes consists of tandem repeats of the sequence d(TTAGGG) and a G-rich strand that forms 100–200 nucleotide long 3′ overhang called G-tail. Telomeres and their associated enzyme telomerase are vital elements for the maintenance of chromosomal end integrity. A protein complex shelterin further upholds the integrity of telomeric ends.[131] Dynamic resolution and formation of G-quadruplex structures in tandem repeats of the sequence d(TTAGGG) were proposed to aid chromosomal end maintanence.[132] Since these telomeric G-quadruplexes are implicated in cancer biology, a considerable research effort has been focused on therapeutic implications and targeting of such structures.[9]

Initially telomeric region was thought to be transcriptionally silent, but a landmark study in 2007 provided an unequivocal evidence that the C-rich strand of the telomeric DNA is in fact actively transcribed by DNA-dependent RNA polymerase II (Pol II), giving rise to transcripts of variable lengths (0.1- 9.0 kb) known as telomeric repeat containing RNA (TERRA or telRNA).[133] TERRA was observed to exclusively localize to the nucleus near the telomeric regions.[134, 135] It has been identified in humans, mice, zebrafish and yeast, indicating its evolutionary conservation and hinting towards its probable important role within cells.[133, 136, 137] These long non coding RNAs are involved in a number of key cellular processes, including regulation of telomere length and inhibition of telomerase.[138 139]

Since TERRA consists of consecutive repeats of UUAGGG, extensive studies were carried out to assess G-quadruplex formation in these telomeric RNA.[52] These studies evinced that TERRA RNA could form stable G-quadruplexes and further evidences of their *in vivo* existence provided clues about their unknown roles in telomere biology.[135] Employing pull down studies and MALDI-TOF MS, a wide array of proteins has been identified to bind TERRA RNA.[140, 141] These proteins were deduced to regulate the localization and local concentration of the TERRA RNA. The shelterin proteins TRF1 and TRF2 were shown capable of binding directly to TERRA, as well as to elements of the origin recognition complex and H3TK9. In particular, TRF2 binds simultaneously to both TERRA RNA G-quadruplex and telomeric DNA G-quadruplex.[142] Moreover, proteins belonging to the hnRNP class were also found enriched in the pool of binding partners. Pull down studies with a poly UUAGGG tract indicated binding of hnRNP A1 to G-quadruplex structure.[141] Thus, these reports make protein binding and thus function mediated via the formation of G-quadruplex structure in TERRA plausible.

TERRA provides an excellent example of a non coding RNA that is involved in maintenance of chromosomal end integrity by most probably acting as a scaffold for recruiting proteins. With growing literature on long non coding RNA with unknown functions and bioinformatic prediction of G-quadruplex in non coding RNAs,[14] we speculate that non coding RNA harbouring G-quadruplexes might be a more universal phenomenon with probable important functions.

Tools for detecting RNA G-quadruplex

G-quadruplexes are detected by biophysical techniques such as UV-vis spectroscopy, Circular Dichroism (CD), Fluorescence spectroscopy, single molecule FRET spectroscopy and electrospray ionisation mass spectroscopy.[52] Biophysical characterisation of G-rich sequences is performed in buffer containing potassium/ sodium salts. A hypochromic shift with increasing temperature at 295 nm is a characteristic signature of G-quadruplex during UV melting.[143] CD scans of RNA G-quadruplex shows a positive peak at 262 nm and a negative peak at 240 nm of wavelength, indicating its parallel topology. Thermal difference spectrum of G-quadruplex where differential absorbance of the structure in unfolded and folded condition is plotted against wavelength indicates two positive peaks and a negative peak (at 295 nm).[144] Fluorescence resonance energy transfer study of G-quadruplex forming sequence with a pair of acceptor/donor fluorophores covalently attached to the 5′ and 3′ ends is also used.[41] Besides biophysical characterisation, polyacrylamide gel electrophoresis, reverse transcription assay and footprinting analysis of the sequences are employed.[44, 88] As RNase T1 enzyme cleaves after guanine residue in single stranded condition and G-quadruplex secondary structure confers protection to involved guanine residues against RNase T1 digestion, RNase T1 footprinting technique is widely used.[62] DMS footprinting and in line probing are also performed to indicate the guanine residues involved in G-quadruplex formation.[59, 69] Furthermore, to detect the folding topologies of G-quadruplex at the atomic level, techniques such as NMR spectroscopy and crystallography are employed.[48, 50] These techniques have been discussed in detail elsewhere.[145]

In order to show the existence and function of G-quadruplex in cells, various controls are included in the experiments. For example, G-quadruplex sequence with point mutations that disrupt G-quadruplex formation are used as control to elucidate its function.[13, 57, 89] Also proteins and small molecule known to bind these structures are used to unravel its role which are discussed in detail in next section. Recently, Balasubramanian and co-workers developed a single chain fragment antibody for visualisation of RNA G-quadruplex in human cells.[22] However, more probe molecules to identify and study this structure in cells are highly required.

Targeting RNA G-quadruplex

Manipulation of cellular events by targeting nucleic acid secondary structure is a promising strategy to control gene expression.[99, 145] Oligonucleotide based targeting agents interfere with gene function but exhibit poor pharmacological properties.[146] Small molecule based intervention methodology overcomes the disadvantages associated with nucleic acid based approaches, thereby translating chemical biology towards developing therapeutics.[147, 148]

As G-quadruplex at the telomeric ends and promoter regions are well established targets for cancer therapeutics, extensive research on the ligands targeting DNA G-quadruplex has been performed in last few decades.[149] Small molecules such as porphyrin (TmPyP4), acridine, pentacridium, quinacridine, telomestatin, naphthalene diiamide, bisquinolium and their derivatives are found to selectively bind and stabilise DNA G-quadruplex.[150, 151] Various crystal structures of DNA G-quadruplex-small molecule are reported.[150] Even few of above ligands have been tested to bind and modulate DNA G-quadruplex function in cellular conditions.[151]

Small molecules synthesised to selectively bind to G-quadruplex typically exhibit following features: (1) A large planar aromatic core in order to maximise π-stacking interactions with G-quartets of G-quadruplex. (2) A positive charge to neutralise the negatively charged phosphate groups of nucleic acids backbone. (3) Positive charged side chains with functional groups to enhance contacts with grooves/loops/inner G-quartet core of G-quadruplex.[145] However till date, only few studies on ligand targeting RNA G-quadruplex have been conducted.[152] Since various clinically important genes are known to harbour RNA G-quadruplex, designing small molecules that stabilise or destabilise this structure acts as an attractive strategy towards drug discovery against various diseases including cancer.[145] Especially G-quadruplexes in telomeric RNA containing repeat (TERRA) transcribed from the subtelomeric loci present at chromosomal ends and untranslated region of mRNA of oncogenes are plausible best suitable targets for drug discovery. Moreover, the parallel topology of RNA G-quadruplex limits its conformational diversity making selective targeting less challenging.[152]

Initially, Collie *et al* studied the selectivity of few small molecules against DNA and RNA G-quadruplexes and found that BRACO-19 was more selective for telomeric DNA G-quadruplex and a naphthalene diamide derivative was more selective towards RNA G-quadruplex.[153] Thereafter, the same group reported the crystal structure of TERRA G-quadruplex-acridine small molecule complex where increased involvement of 2′ hydroxyl group of ribose sugar in binding with acridine was revealed (Fig. 6).[51] In contrast to DNA G-quadruplex-acridine complex, the loops of RNA G-quadruplex were shown to interact with acridine. Till date, this is the only crystal structure of ligand-RNA G-quadruplex complex available. Nevertheless, ESI mass spectroscopy has also been employed to understand the binding behaviour of three natural alkaloids (nitidine, palmatine, and jatrorrizine) to RNA G-quadruplex structure present in 5′ UTR of Bcl2. They are found to bind Bcl2 RNA quadruplex with high affinity and binding stoichiometry ranging from 1:1 to 3:1.[154] ESI-TOF-MS experiments also demonstrated that chelerythrine can bind both DNA and RNA G-quadruplex.[155] However, no NMR or crystallography studies have been performed on mRNA G-quadruplex.

Pyridostatin is a known molecule to stack over the terminal G-quartets of DNA G-quadruplex stabilising the structure and to

cause growth arrest in cancer cells by inducing extensive DNA damage.[156] Through a template-directed in situ 'click chemistry' approach, a derivative of this molecule – carboxypyridostatin was discovered to exhibit more binding preference and selectivity towards RNA G-quadruplex over DNA G-quadruplex.[157] In fact, this selectivity of carboxypyridostatin was later exploited in conjunction with G-quadruplex structure specific antibody probe for the direct visualisation of RNA G-quadruplexes in the cytoplasm of the cell.[22] Recently, Basu and co-workers screened a library of RNA G-quadruplex forming sequences with varying loop length for their binding to aminoglycoside kanamycin A.[158] They found that only few RNA G-quadruplex with two nucleotide loop length with specific combination were able to bind to kanamycin A.

Ligands binding to G-quadruplex can behave in varying manners: (a) It can stabilise G-quadruplex, thereby intensifying its function. (b) It can destabilise the structure, thus minimising its role. (c) It can interfere with the binding of trans factor to G-quadruplex, consequently hindering its activity (Fig. 7).[152] In this regard, Balasubramanian and coworkers for the first time demonstrated the proof of concept for small molecule mediated gene regulation. They showed that small molecule-bisquinolium derivatives RR82 and RR110 are potent binders and stabilisers of RNA G-quadruplex present in 5′ UTR of NRAS mRNA enhancing inhibition of NRAS gene expression.[148] In addition, three bisquinolium derivatives-PhenDC3, PhenDC6 and 360A with binding affinity similar towards RNA G-quadruplex and DNA G-quadruplex were found to perform small molecule- RNA G-quadruplex mediated gene regulation.[63, 64, 81, 95, 159] Balasubramanian group also showed that ligand TmPyP4 does not exhibit RNA G-quadruplex selectivity and thereby, do not inhibit NRAS translation.[148] On the contrary, Basu and coworkers demonstrated that TmPyP4 destabilises RNA G-quadruplex present in 5′ UTR of MT3-MMP gene relieving the suppressive role of RNA G-quadruplex. Recently, this molecule was shown to distort RNA G-quadruplex structures formed by r(GGGGCC)n repeats of C9orf72 gene interrupting their interaction with various proteins.[160] Moreover, a photoactivable alkyl substituted TmPyP4 was developed to bind KRAS G-quadruplex and cause mRNA degradation and translation suppression upon photoactivation.[161] TmPyP4 tetrachloride was further demonstrated to bind TERRA RNA G-quadruplex with binding affinity of magnitude one order higher than that of telomeric DNA G-quadruplex using ESI-TOF-MS.[162] These studies, thus, indicated that a molecule against RNA G-quadruplex such as TmPyP4 can behave differently in different gene context.

Employing robust approaches to identify ligands for RNA G-quadruplex is of great interest from drug discovery point of view.[163] Various high throughput screening platforms such as virtual docking[164-166] and fluorescence resonance energy transfer (FRET) based methods[167-169] are used for identifying the binders. However, these screening were majorly utilised to identify molecules against DNA G-quadruplex. Lacroix et al used fluorescence based approach and quadruplex-duplex competition to screen various compounds and identified selective RNA G-quadruplex binders.[170] Recently, Garavis *et al* used ^{19}F- NMR fragment based screening of various compounds to identify ligands that can bind with TERRA RNA G-quadruplex with high selectivity.[164] Some of these RNA G-quadruplex ligands identified were also observed to bind DNA G-quadruplexes with parallel topology.

Template assisted synthetic G-quartets (TASQs) are prototype G-quadruplex ligands that do not fall into the category of "small molecules having planar aromatic rings". These TASQs undergo conformational rearrangement upon binding to G-quadruplex, ensuring a high level of selectivity.[171] A deviation of this prototype, pyrene template-assisted synthetic G-quartets (PyroTASQ) ligand was recently reported to emit fluorescence upon binding to DNA/ RNA G-quadruplex.[172]

Current research on ligands is mainly restricted to small molecules having planar aromatic rings that primarily undergo extensive π-stacking interactions with the end G-quartets of G-quadruplex.[5] There is a requirement to explore ligands that can bind to grooves, loops and other regions of G-quadruplex. As proteins are known to bind G-quadruplex, another line of research has led to advancements towards developing peptides and antibodies towards this secondary structure. Peptides provide a good alternative to generally used heterocyclic molecules that stack with terminal G-quartets of DNA G-quadruplex. Peptides with amino groups exhibit potential to interact with loops, grooves and phosphate backbone of the G-quadruplex structure. The prime peptide developed against this structure was a single chain antibody fragment that recognises DNA G-quadruplex structures formed by telomeric repeat and was employed to show *in vivo* evidence of G-quadruplex structures within the macronucleus of *Stylonychia lemnae* for the first time.[173] Recently, Balasubramanian and coworkers have identified a single chain fragment selective for G-quadruplex using peptide display library and have used them as valuable visual tool to demonstrate the presence of both DNA and RNA G-quadruplex in the human cells (Fig. 8).[22, 174] In another study, chemically modified peptides with zinc finger motifs were identified as suitable ligands for G-quadruplex on screening phage display library.[175] These peptides were shown to inhibit telomerase activity when administered at the nanomolar range.[176]

Besides this, few other peptides were identified to bind G-quadruplex structure. Human antimicrobial peptide-cathelicidin peptide LL37 was reported as a potent binder of DNA G-quadruplex structure.[177] In addition, RGG (Arginine-glycine-glycine) domain of TLS protein has been demonstrated to bind to both telomeric RNA and DNA G-quadruplex.[125] Interestingly, on substituting tyrosine for phenylalanine in RGG domain of TLS protein, binding specificity of engineered RGG domain was found to greatly increase for RNA G-quadruplex and drastically decrease for DNA G-quadruplex.[178] Indeed, this report is a paradigm where modifications incorporated in the G-quadruplex specific peptide is shown to alter its affinity to

either RNA or DNA G-quadruplex.Thus lately, although there has been emerging interest to use RNA G-quadruplex as a potential molecular target for small molecule therapeutic agents and few molecules have been developed in this direction, most of the compounds identified against RNA G-quadruplex resulted from rational modification of molecules already tested against DNA G-quadruplexes. Therefore, the major challenge remains to identify/design novel ligands against RNA G-quadruplex. Novel ligands may be developed with a careful understanding of the inherent differences between DNA and RNA G-quadruplex structures. Exploiting effects of ribose hydroxyl group on RNA G-quadruplex structure can form the basis for ligand design for discerning differences between DNA and RNA G-quadruplex.[179]

Applications

Various G-quadruplex forming oligonucleotides functioning as aptamers are being investigated for targeting molecules of therapeutic importance. Compared to monoclonal antibodies and other targeting tools, G-quadruplex aptamers offers several advantages of non-immunogenicity, heat stability, biostability, cost effectiveness, ease of chemical synthesis, enhanced cellular uptake efficiency and flexibility for introducing chemical modifications. In this regard, the most extensive work has been performed on RNA G-quadruplex based aptamers against prion protein, the agent causing transmissible neurodegenerative encephalopathy known as Creutzfeldt-Jakob disease.[180] This aptamer targeting N-terminal region of prion protein (PrPc) comprises of G-quadruplex structure; and substitution of guanine bases with uracil residues was revealed to result in complete loss of PrPc binding, strongly suggesting that G-quadruplex is indispensible for PrPc recognition.[181, 182]

In addition, numerous G-quadruplex based RNA aptamer have been identified against other pathological diseases. In the process of developing an aptamer against rheumatoid arthritis using SELEX, a G-quadruplex based RNA aptamer highly selective against human receptor activator of NF-kB (RANK) was identified.[183] This aptamer was also found to target other TNF receptor family proteins, such as TRAIL-R2, CD30, NGFR and osteoprotegerin, a decoy receptor for RANK. Furthermore, an aptamer had been successfully developed as trypanocidal agents to treat sleeping sickness.[184] This 2′-NH2 modified aptamer possessing a stable G-quadruplex directly binds live trypanosomes at the parasite surface, thereby targeting it for immunoglobulin attack. In addition, several G-rich sequences were identified during *in vitro* screening of RNA sequences targeting HIV integrase enzyme.[185] Most of these sequences harboured the motif capable of forming stable G-quadruplex structure.

Furthermore, G-quadruplex based aptamer are being developed to bind with important biomolecules and cofactors of metabolic pathways such as riboflavin [186] and thyroxine.[187] Recently, RNA aptamers have also been used as diagnostic tools for the detection of SARS coronavirus.[188] G-quadruplex in spinach RNA aptamer also has been shown to bind GFP like ligand and activates its fluorescence, suggesting its usage in imaging techniques.[189, 190] RNA G-quadruplex has been already used as a thermoregulator to detect changes in temperature in *E.coli*. Like DNA G-quadruplex, it exhibit traits required for its usage as biosensor, molecular beacon, nanoswitch, nanowire, nanodevice [191-194] and become an important module in the field of synthetic biology, which has been discussed in detail elsewhere.[195]

Conclusion

With a large number of RNA G-quadruplexes present across the transcriptome, it becomes imperative to understand biological importance of these structural motifs. Although the research progress executed in this arena is highly commendable, much remains to be explored. The mechanism of G-quadruplex mediated gene regulation requires special attention. Recent progress in unravelling the function of RNA G-quadruplex has projected it as suitable candidate for therapeutic intervention with considerable efforts being invested to manipulate cellular events by strategically targeting this secondary structure.[145, 152] In this regard, optimization of suitable drug candidates that can interact selectively with RNA quadruplexes is of utmost importance.[99] An array of ligands targeting DNA quadruplexes may be screened for suitable interactions with RNA counterparts as well as novel ligands may be designed to selectively interact with RNA G-quadruplexes. Though few ligands have been identified, a better understanding of the thermodynamic basis of these studied interactions would further open a new avenue for developing better therapeutic molecules targeting these secondary structures. Moreover, understanding the structural contribution of versatile G-quadruplex on physiological functioning remains uncovered *in vivo* scenario. More animal model studies are required in this direction. Along with fundamentals and their therapeutic application in the form of aptamers, RNA G-quadruplex research possesses the potential to contribute immensely to the field of synthetic biology. We envision that increased availability of high-resolution structures of RNA G-quadruplexes in concert with an increased understanding of *in vivo* G-quadruplex topologies will expand the scope of identification of G-quadruplex ligands and their utility in the future.

Acknowledgement

This work was financially supported by the BSC0123 project (Project: Genome Dynamics in Cellular Organization, Differentiation and Enantiostasis) from the Council of Scientific and Industrial Research (CSIR), Government of India. The authors also like to thank Dr. Mary Krishna Ekka for modelling crystal structures of RNA G-quadruplex.

Notes

[a]Proteomics and Structural Biology Unit, CSIR-Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India, Tel: +91 11 27666156; Fax: +91 11 27667437; E-mail: souvik@igib.res.in
[b]Academy of Scientific and Innovative Research (AcSIR), Anusandhan Bhawan, 2 Rafi Marg, New Delhi 110001, India
[c]CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, India.

References

1. F. W. Smith and J. Feigon, *Nature*, 1992, **356**, 164.
2. J. T. Davis, *Angew Chem Int Ed Engl*, 2004, **43**, 668.
3. M. Gellert, M. N. Lipsett and D. R. Davies, *Proc Natl Acad Sci U S A*, 1962, **48**, 2013.
4. M. A. Keniry, *Biopolymers*, 2000, **56**, 123.
5. D. J. Patel, A. T. Phan and V. Kuryavyi, *Nucleic Acids Res*, 2007, **35**, 7429.
6. C. C. Hardin, T. Watson, M. Corregan and C. Bailey, *Biochemistry*, 1992, **31**, 833.
7. S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res*, 2006, **34**, 5402.
8. T. Simonsson, *Biol Chem*, 2001, **382**, 621.
9. D. Monchaud and M. P. Teulade-Fichou, *Org Biomol Chem*, 2008, **6**, 627.
10. M. C. Nielsen and T. Ulven, *Curr Med Chem*, 2010, **17**, 3438.
11. J. Kim, C. Cheong and P. B. Moore, *Nature*, 1991, **351**, 331.
12. J. Christiansen, M. Kofod and F. C. Nielsen, *Nucleic Acids Res*, 1994, **22**, 5709.
13. S. Kumari, A. Bugaut, J. L. Huppert and S. Balasubramanian, *Nat Chem Biol*, 2007, **3**, 218.
14. G. G. Jayaraj, S. Pandey, V. Scaria and S. Maiti, *RNA Biol*, 2012, **9**, 81.
15. S. Millevoi, H. Moine and S. Vagner, *Wiley Interdiscip Rev RNA*, 2012, **3**, 495.
16. X. Ji, H. Sun, H. Zhou, J. Xiang, Y. Tang and C. Zhao, *Nucleic Acid Ther*, 2011, **21**, 185.
17. P. Svoboda and A. Di Cara, *Cell Mol Life Sci*, 2006, **63**, 901.
18. A. Jambhekar and J. L. Derisi, *RNA*, 2007, **13**, 625.
19. P. C. Bevilacqua and J. M. Blose, *Annu Rev Phys Chem*, 2008, **59**, 79.
20. C. J. McManus and B. R. Graveley, *Curr Opin Genet Dev*, 2011, **21**, 373.
21. O. Kikin, Z. Zappala, L. D'Antonio and P. S. Bagga, *Nucleic Acids Res*, 2008, **36**, D141.
22. G. Biffi, M. Di Antonio, D. Tannahill and S. Balasubramanian, *Nat Chem*, 2014, **6**, 75.
23. A. Joachimi, A. Benz and J. S. Hartig, *Bioorg Med Chem*, 2009, **17**, 6811.
24. A. Arora and S. Maiti, *J Phys Chem B*, 2009, **113**, 10515.
25. C. F. Tang and R. H. Shafer, *J Am Chem Soc*, 2006, **128**, 5966.
26. D. Miyoshi, A. Nakao and N. Sugimoto, *Biochemistry*, 2002, **41**, 15017.
27. D. Miyoshi, H. Karimata and N. Sugimoto, *J Am Chem Soc*, 2006, **128**, 7957.
28. D. H. Zhang, T. Fujimoto, S. Saxena, H. Q. Yu, D. Miyoshi and N. Sugimoto, *Biochemistry*, 2010, **49**, 4554.
29. B. Sacca, L. Lacroix and J. L. Mergny, *Nucleic Acids Res*, 2005, **33**, 1182.
30. C. M. Olsen and L. A. Marky, *J Phys Chem B*, 2009, **113**, 9.
31. P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J Am Chem Soc*, 2004, **126**, 16405.
32. A. Risitano and K. R. Fox, *Nucleic Acids Res*, 2004, **32**, 2598.
33. M. Cevec and J. Plavec, *Biochemistry*, 2005, **44**, 15238.
34. N. Kumar, B. Sahoo, K. A. Varun and S. Maiti, *Nucleic Acids Res*, 2008, **36**, 4433.
35. A. Guedin, A. De Cian, J. Gros, L. Lacroix and J. L. Mergny, *Biochimie*, 2008, **90**, 686.
36. A. Arora and S. Maiti, *J Phys Chem B*, 2009, **113**, 8784.
37. A. Guedin, J. Gros, P. Alberti and J. L. Mergny, *Nucleic Acids Res*, 2010, **38**, 7858.
38. A. Y. Zhang, A. Bugaut and S. Balasubramanian, *Biochemistry*, 2011, **50**, 7251.
39. S. Pandey, P. Agarwala and S. Maiti, *J Phys Chem B*, 2013, **117**, 6896.
40. M. Tomasko, M. Vorlickova and J. Sagi, *Biochimie*, 2009, **91**, 171.
41. J. Y. Lee and D. S. Kim, *Nucleic Acids Res*, 2009, **37**, 3625.
42. P. Agarwala, S. Kumar, S. Pandey and S. Maiti, *J Phys Chem B*, 2015.
43. N. V. Hud, F. W. Smith, F. A. Anet and J. Feigon, *Biochemistry*, 1996, **35**, 15383.
44. M. A. Mullen, S. M. Assmann and P. C. Bevilacqua, *J Am Chem Soc*, 2012, **134**, 812.
45. C. K. Kwok, M. E. Sherlock and P. C. Bevilacqua, *Angew Chem Int Ed Engl*, 2013, **52**, 683.
46. C. Kang, X. Zhang, R. Ratliff, R. Moyzis and A. Rich, *Nature*, 1992, **356**, 126.
47. G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876.
48. H. Liu, M. Kanagawa, A. Matsugami, Y. Tanaka, M. Katahira and S. Uesugi, *Nucleic Acids Symp Ser*, 2000, 65.
49. H. Liu, A. Kugimiya, T. Sakurai, M. Katahira and S. Uesugi, *Nucleosides Nucleotides Nucleic Acids*, 2002, **21**, 785.
50. G. W. Collie, S. M. Haider, S. Neidle and G. N. Parkinson, *Nucleic Acids Res*, 2010, **38**, 5569.
51. G. W. Collie, S. Sparapani, G. N. Parkinson and S. Neidle, *J Am Chem Soc*, 2011, **133**, 2721.
52. G. W. Collie, G. N. Parkinson, S. Neidle, F. Rosu, E. De Pauw and V. Gabelica, *J Am Chem Soc*, 2010, **132**, 9328.
53. H. Martadinata and A. T. Phan, *J Am Chem Soc*, 2009, **131**, 2570.
54. L. Petraccone, *Top Curr Chem*, 2013, **330**, 23.
55. A. Randall and J. D. Griffith, *J Biol Chem*, 2009, **284**, 13980.
56. H. Martadinata and A. T. Phan, *Biochemistry*, 2013, **52**, 2176.
57. A. Arora, M. Dutkiewicz, V. Scaria, M. Hariharan, S. Maiti and J. Kurreck, *RNA*, 2008, **14**, 1290.
58. J. L. Huppert, A. Bugaut, S. Kumari and S. Balasubramanian, *Nucleic Acids Res*, 2008, **36**, 6260.
59. S. Bonnal, C. Schaeffer, L. Creancier, S. Clamens, H. Moine, A. C. Prats and S. Vagner, *J Biol Chem*, 2003, **278**, 39330.
60. K. Derecka, G. D. Balkwill, T. P. Garner, C. Hodgman, A. P. Flint and M. S. Searle, *Biochemistry*, 2010, **49**, 7625.

61. G. D. Balkwill, K. Derecka, T. P. Garner, C. Hodgman, A. P. Flint and M. S. Searle, *Biochemistry*, 2009, **48**, 11487.
62. M. J. Morris and S. Basu, *Biochemistry*, 2009, **48**, 5313.
63. R. Shahid, A. Bugaut and S. Balasubramanian, *Biochemistry*, 2010, **49**, 8300.
64. D. Gomez, A. Guedin, J. L. Mergny, B. Salles, J. F. Riou, M. P. Teulade-Fichou and P. Calsou, *Nucleic Acids Res*, 2010, **38**, 7187.
65. M. J. Morris, Y. Negishi, C. Pazsint, J. D. Schonhoft and S. Basu, *J Am Chem Soc*, 2010, **132**, 17831.
66. S. Lammich, F. Kamp, J. Wagner, B. Nuscher, S. Zilow, A. K. Ludwig, M. Willem and C. Haass, *J Biol Chem*, 2011, **286**, 45063.
67. H. Y. Weng, H. L. Huang, P. P. Zhao, H. Zhou and L. H. Qu, *RNA Biol*, 2012, **9**, 1099.
68. P. Agarwala, S. Pandey, K. Mapa and S. Maiti, *Biochemistry*, 2013, **52**, 1528.
69. J. D. Beaudoin and J. P. Perreault, *Nucleic Acids Res*, 2010, **38**, 7022.
70. K. Halder, M. Wieland and J. S. Hartig, *Nucleic Acids Res*, 2009, **37**, 6811.
71. A. L. Wolfe, K. Singh, Y. Zhong, P. Drewe, V. K. Rajasekhar, V. R. Sanghvi, K. J. Mavrakis, M. Jiang, J. E. Roderick, J. Van der Meulen, J. H. Schatz, C. M. Rodrigo, C. Zhao, P. Rondou, E. de Stanchina, J. Teruya-Feldstein, M. A. Kelliher, F. Speleman, J. A. Porco, J. Pelletier, G. Ratsch and H. G. Wendel, *Nature*, 2014.
72. P. Agarwala, S. Pandey and S. Maiti, *Biochim Biophys Acta*, 2014, **1840**, 3503.
73. M. Wieland and J. S. Hartig, *Chem Biol*, 2007, **14**, 757.
74. M. Wieland and J. S. Hartig, *Nat Protoc*, 2009, **4**, 1632.
75. S. Kumari, A. Bugaut and S. Balasubramanian, *Biochemistry*, 2008, **47**, 12664.
76. T. Endoh, Y. Kawasaki and N. Sugimoto, *Angew Chem Int Ed Engl*, 2013, **52**, 5522.
77. P. Murat, J. Zhong, L. Lekieffre, N. P. Cowieson, J. L. Clancy, T. Preiss, S. Balasubramanian, R. Khanna and J. Tellam, *Nat Chem Biol*, 2014, **10**, 358.
78. T. Endoh, Y. Kawasaki and N. Sugimoto, *Nucleic Acids Res*, 2013, **41**, 6222.
79. T. Endoh, Y. Kawasaki and N. Sugimoto, *Methods*, 2013, **64**, 73.
80. T. Endoh and N. Sugimoto, *Anal Chem*, 2013, **85**, 11435.
81. C. H. Yu, M. P. Teulade-Fichou and R. C. Olsthoorn, *Nucleic Acids Res*, 2014, **42**, 1887.
82. A. Arora and B. Suess, *RNA Biol*, 2011, **8**, 802.
83. K. Ito, S. Go, M. Komiyama and Y. Xu, *J Am Chem Soc*, 2011, **133**, 19153.
84. D. Bhattacharyya, K. Nguyen and S. Basu, *Biochemistry*, 2014, **53**, 5461.
85. F. C. Nielsen and J. Christiansen, *Scand J Clin Lab Invest Suppl*, 1995, **220**, 37.
86. A. Decorsiere, A. Cayrel, S. Vagner and S. Millevoi, *Genes Dev*, 2011, **25**, 220.
87. J. D. Beaudoin and J. P. Perreault, *Nucleic Acids Res*, 2013, **41**, 5898.
88. P. H. Wanrooij, J. P. Uhler, T. Simonsson, M. Falkenberg and C. M. Gustafsson, *Proc Natl Acad Sci U S A*, 2010, **107**, 16072.
89. M. Subramanian, F. Rage, R. Tabet, E. Flatter, J. L. Mandel and H. Moine, *EMBO Rep*, 2011, **12**, 697.
90. R. Kostadinov, N. Malhotra, M. Viotti, R. Shine, L. D'Antonio and P. Bagga, *Nucleic Acids Res*, 2006, **34**, D119.
91. H. Min, R. C. Chan and D. L. Black, *Genes Dev*, 1995, **9**, 2659.
92. P. Sirand-Pugnet, P. Durosay, E. Brody and J. Marie, *Nucleic Acids Res*, 1995, **23**, 3501.
93. D. Gomez, T. Lemarteleur, L. Lacroix, P. Mailliet, J. L. Mergny and J. F. Riou, *Nucleic Acids Res*, 2004, **32**, 371.
94. M. C. Didiot, Z. Tian, C. Schaeffer, M. Subramanian, J. L. Mandel and H. Moine, *Nucleic Acids Res*, 2008, **36**, 4902.
95. V. Marcel, P. L. Tran, C. Sagne, G. Martel-Planche, L. Vaslin, M. P. Teulade-Fichou, J. Hall, J. L. Mergny, P. Hainaut and E. Van Dyck, *Carcinogenesis*, 2011, **32**, 271.
96. J. F. Fisette, D. R. Montagna, M. R. Mihailescu and M. S. Wolfe, *J Neurochem*, 2012, **121**, 763.
97. M. M. Ribeiro, G. S. Teixeira, L. Martins, M. R. Marques, A. P. de Souza and S. R. Line, *Hum Genet*, 2014.
98. M. Fry, *Front Biosci*, 2007, **12**, 4336.
99. C. Sissi, B. Gatto and M. Palumbo, *Biochimie*, 2011, **93**, 1219.
100. V. Brown, P. Jin, S. Ceman, J. C. Darnell, W. T. O'Donnell, S. A. Tenenbaum, X. Jin, Y. Feng, K. D. Wilkinson, J. D. Keene, R. B. Darnell and S. T. Warren, *Cell*, 2001, **107**, 477.
101. J. C. Darnell, K. B. Jensen, P. Jin, V. Brown, S. T. Warren and R. B. Darnell, *Cell*, 2001, **107**, 489.
102. M. Bensaid, M. Melko, E. G. Bechara, L. Davidovic, A. Berretta, M. V. Catania, J. Gecz, E. Lalli and B. Bardoni, *Nucleic Acids Res*, 2009, **37**, 1269.
103. M. Melko and B. Bardoni, *Biochimie*, 2010, **92**, 919.
104. E. Blackwell, X. Zhang and S. Ceman, *Hum Mol Genet*, 2010, **19**, 1314.
105. R. Lu, H. Wang, Z. Liang, L. Ku, T. O'Donnell W, W. Li, S. T. Warren and Y. Feng, *Proc Natl Acad Sci U S A*, 2004, **101**, 15201.
106. M. Castets, C. Schaeffer, E. Bechara, A. Schenck, E. W. Khandjian, S. Luche, H. Moine, T. Rabilloud, J. L. Mandel and B. Bardoni, *Hum Mol Genet*, 2005, **14**, 835.
107. C. Schaeffer, B. Bardoni, J. L. Mandel, B. Ehresmann, C. Ehresmann and H. Moine, *EMBO J*, 2001, **20**, 4803.
108. C. J. Westmark and J. S. Malter, *PLoS Biol*, 2007, **5**, e52.
109. L. Menon and M. R. Mihailescu, *Nucleic Acids Res*, 2007, **35**, 5379.
110. P. K. Todd, K. J. Mack and J. S. Malter, *Proc Natl Acad Sci U S A*, 2003, **100**, 14374.
111. S. Khateb, P. Weisman-Shomer, I. Hershco-Shani, A. L. Ludwig and M. Fry, *Nucleic Acids Res*, 2007, **35**, 5775.
112. T. L. Evans, A. C. Blice-Baum and M. R. Mihailescu, *Mol Biosyst*, 2012, **8**, 642.
113. J. P. Vaughn, S. D. Creacy, E. D. Routh, C. Joyner-Butt, G. S. Jenkins, S. Pauli, Y. Nagamine and S. A. Akman, *J Biol Chem*, 2005, **280**, 38117.
114. S. D. Creacy, E. D. Routh, F. Iwamoto, Y. Nagamine, S. A. Akman and J. P. Vaughn, *J Biol Chem*, 2008, **283**, 34626.
115. S. Lattmann, M. B. Stadler, J. P. Vaughn, S. A. Akman and Y. Nagamine, *Nucleic Acids Res*, 2011, **39**, 9390.
116. A. N. Sexton and K. Collins, *Mol Cell Biol*, 2011, **31**, 736.
117. E. P. Booy, M. Meier, N. Okun, S. K. Novakowski, S. Xiong, J. Stetefeld and S. A. McKenna, *Nucleic Acids Res*, 2012, **40**, 4110.

118. S. Lattmann, B. Giri, J. P. Vaughn, S. A. Akman and Y. Nagamine, *Nucleic Acids Res*, 2010, **38**, 6219.
119. M. Meier, T. R. Patel, E. P. Booy, O. Marushchak, N. Okun, S. Deo, R. Howard, K. McEleney, S. E. Harding, J. Stetefeld and S. A. McKenna, *J Biol Chem*, 2013, **288**, 35014.
120. W. Huang, P. J. Smaldino, Q. Zhang, L. D. Miller, P. Cao, K. Stadelman, M. Wan, B. Giri, M. Lei, Y. Nagamine, J. P. Vaughn, S. A. Akman and G. Sui, *Nucleic Acids Res*, 2012, **40**, 1033.
121. E. P. Booy, R. Howard, O. Marushchak, E. O. Ariyo, M. Meier, S. K. Novakowski, S. R. Deo, E. Dzananovic, J. Stetefeld and S. A. McKenna, *Nucleic Acids Res*, 2014, **42**, 3346.
122. P. Chakraborty and F. Grosse, *DNA Repair (Amst)*, 2011, **10**, 654.
123. A. R. Haeusler, C. J. Donnelly, G. Periz, E. A. Simko, P. G. Shaw, M. S. Kim, N. J. Maragakis, J. C. Troncoso, A. Pandey, R. Sattler, J. D. Rothstein and J. Wang, *Nature*, 2014, **507**, 195.
124. Z. Su, Y. Zhang, T. F. Gendron, P. O. Bauer, J. Chew, W. Y. Yang, E. Fostvedt, K. Jansen-West, V. V. Belzil, P. Desaro, A. Johnston, K. Overstreet, S. Y. Oh, P. K. Todd, J. D. Berry, M. E. Cudkowicz, B. F. Boeve, D. Dickson, M. K. Floeter, B. J. Traynor, C. Morelli, A. Ratti, V. Silani, R. Rademakers, R. H. Brown, J. D. Rothstein, K. B. Boylan, L. Petrucelli and M. D. Disney, *Neuron*, 2014, **83**, 1043.
125. K. Takahama, A. Takada, S. Tada, M. Shimizu, K. Sayama, R. Kurokawa and T. Oyoshi, *Chem Biol*, 2013, **20**, 341.
126. J. D. Frantz and W. Gilbert, *J Biol Chem*, 1995, **270**, 9413.
127. J. D. Frantz and W. Gilbert, *J Biol Chem*, 1995, **270**, 20692.
128. A. Marnef, M. Maldonado, A. Bugaut, S. Balasubramanian, M. Kress, D. Weil and N. Standart, *RNA*, 2010, **16**, 2094.
129. A. von Hacht, O. Seifert, M. Menger, T. Schutze, A. Arora, Z. Konthur, P. Neubauer, A. Wagner, C. Weise and J. Kurreck, *Nucleic Acids Res*, 2014, **42**, 6630.
130. R. H. Shafer and I. Smirnov, *Biopolymers*, 2000, **56**, 209.
131. A. R. Pinto, H. Li, C. Nicholls and J. P. Liu, *Front Biosci (Landmark Ed)*, 2011, **16**, 187.
132. M. Salazar, B. D. Thompson, S. M. Kerwin and L. H. Hurley, *Biochemistry*, 1996, **35**, 16110.
133. C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto and J. Lingner, *Science*, 2007, **318**, 798.
134. B. Luke and J. Lingner, *EMBO J*, 2009, **28**, 2503.
135. Y. Xu, Y. Suzuki, K. Ito and M. Komiyama, *Proc Natl Acad Sci U S A*, 2010, **107**, 14579.
136. S. Schoeftner and M. A. Blasco, *Nat Cell Biol*, 2008, **10**, 228.
137. B. Luke, A. Panza, S. Redon, N. Iglesias, Z. Li and J. Lingner, *Mol Cell*, 2008, **32**, 465.
138. B. Pan, Y. Xiong, K. Shi, J. Deng and M. Sundaralingam, *Structure*, 2003, **11**, 815.
139. B. Pan, Y. Xiong, K. Shi and M. Sundaralingam, *Structure*, 2003, **11**, 825.
140. I. Lopez de Silanes, M. Stagno d'Alcontres and M. A. Blasco, *Nat Commun*, 2010, **1**, 33.
141. Z. Deng, J. Norseen, A. Wiedmer, H. Riethman and P. M. Lieberman, *Mol Cell*, 2009, **35**, 403.
142. G. Biffi, D. Tannahill and S. Balasubramanian, *J Am Chem Soc*, 2012, **134**, 11974.
143. J. L. Mergny, A. T. Phan and L. Lacroix, *FEBS Lett*, 1998, **435**, 74.
144. J. L. Mergny, J. Li, L. Lacroix, S. Amrane and J. B. Chaires, *Nucleic Acids Res*, 2005, **33**, e138.
145. G. W. Collie and G. N. Parkinson, *Chem Soc Rev*, 2011, **40**, 5867.
146. S. T. Crooke, *Annu Rev Med*, 2004, **55**, 61.
147. J. R. Thomas and P. J. Hergenrother, *Chem Rev*, 2008, **108**, 1171.
148. A. Bugaut, R. Rodriguez, S. Kumari, S. T. Hsu and S. Balasubramanian, *Org Biomol Chem*, 2010, **8**, 2771.
149. S. A. Ohnmacht and S. Neidle, *Bioorg Med Chem Lett*, 2014, **24**, 2602.
150. A. T. Phan, V. Kuryavyi and D. J. Patel, *Curr Opin Struct Biol*, 2006, **16**, 288.
151. Y. Xu, *Chem Soc Rev*, 2011, **40**, 2719.
152. A. Bugaut and S. Balasubramanian, *Nucleic Acids Res*, 2012, **40**, 4727.
153. G. Collie, A. P. Reszka, S. M. Haider, V. Gabelica, G. N. Parkinson and S. Neidle, *Chem Commun (Camb)*, 2009, 7482.
154. W. Tan and G. Yuan, *Rapid Commun Mass Spectrom*, 2013, **27**, 560.
155. L. P. Bai, M. Hagihara, K. Nakatani and Z. H. Jiang, *Sci Rep*, 2014, **4**, 6767.
156. R. Rodriguez, K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian and S. P. Jackson, *Nat Chem Biol*, 2012, **8**, 301.
157. M. Di Antonio, G. Biffi, A. Mariani, E. A. Raiber, R. Rodriguez and S. Balasubramanian, *Angew Chem Int Ed Engl*, 2012, **51**, 11073.
158. G. Mirihana Arachchilage, M. J. Morris and S. Basu, *Chem Commun (Camb)*, 2014, **50**, 1250.
159. K. Halder, E. Largy, M. Benzler, M. P. Teulade-Fichou and J. S. Hartig, *Chembiochem*, 2011, **12**, 1663.
160. B. Zamiri, K. Reddy, R. B. Macgregor, Jr. and C. E. Pearson, *J Biol Chem*, 2014, **289**, 4653.
161. M. Faudale, S. Cogoi and L. E. Xodo, *Chem Commun (Camb)*, 2012, **48**, 874.
162. L. P. Bai, J. Liu, L. Han, H. M. Ho, R. Wang and Z. H. Jiang, *Anal Bioanal Chem*, 2014, **406**, 5455.
163. S. Zhang, Y. Wu and W. Zhang, *ChemMedChem*, 2014, **9**, 899.
164. M. Garavis, B. Lopez-Mendez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas and C. Gonzalez, *ACS Chem Biol*, 2014, **9**, 1559.
165. E. Largy, N. Saettel, F. Hamon, S. Dubruille and M. P. Teulade-Fichou, *Curr Pharm Des*, 2012, **18**, 1992.
166. B. Pagano, S. Cosconati, V. Gabelica, L. Petraccone, S. De Tito, L. Marinelli, V. La Pietra, F. S. di Leva, I. Lauri, R. Trotta, E. Novellino, C. Giancola and A. Randazzo, *Curr Pharm Des*, 2012, **18**, 1880.
167. D. Renciuk, J. Zhou, L. Beaurepaire, A. Guedin, A. Bourdoncle and J. L. Mergny, *Methods*, 2012, **57**, 122.
168. A. Benz, V. Singh, T. U. Mayer and J. S. Hartig, *Chembiochem*, 2011, **12**, 1422.
169. K. M. Rahman, K. Tizkova, A. P. Reszka, S. Neidle and D. E. Thurston, *Bioorg Med Chem Lett*, 2012, **22**, 3006.
170. L. Lacroix, A. Seosse and J. L. Mergny, *Nucleic Acids Res*, 2011, **39**, e21.
171. R. Haudecoeur, L. Stefan, F. Denat and D. Monchaud, *J Am Chem Soc*, 2013, **135**, 550.

172. A. Laguerre, L. Stefan, M. Larrouy, D. Genest, J. Novotna, M. Pirrotta and D. Monchaud, *J Am Chem Soc*, 2014, **136**, 12406.
173. C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps and A. Pluckthun, *Proc Natl Acad Sci U S A*, 2001, **98**, 8572.
174. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat Chem*, 2013, **5**, 182.
175. M. Isalan, S. D. Patel, S. Balasubramanian and Y. Choo, *Biochemistry*, 2001, **40**, 830.
176. S. D. Patel, M. Isalan, G. Gavory, S. Ladame, Y. Choo and S. Balasubramanian, *Biochemistry*, 2004, **43**, 13452.
177. J. Jana, R. K. Kar, A. Ghosh, A. Biswas, S. Ghosh, A. Bhunia and S. Chatterjee, *Mol Biosyst*, 2013, **9**, 1833.
178. K. Takahama and T. Oyoshi, *J Am Chem Soc*, 2013, **135**, 18016.
179. Y. Xu and M. Komiyama, *Methods*, 2012, **57**, 100.
180. R. C. Olsthoorn, *Nucleic Acids Res*, 2014, **42**, 9327.
181. S. Weiss, D. Proske, M. Neumann, M. H. Groschup, H. A. Kretzschmar, M. Famulok and E. L. Winnacker, *J Virol*, 1997, **71**, 8790.
182. S. Sekiya, F. Nishikawa, K. Noda, P. K. Kumar, T. Yokoyama and S. Nishikawa, *Nucleic Acids Symp Ser (Oxf)*, 2005, 361.
183. T. Mori, A. Oguro, T. Ohtsu and Y. Nakamura, *Nucleic Acids Res*, 2004, **32**, 6120.
184. M. Homann, M. Lorger, M. Engstler, M. Zacharias and H. U. Goringer, *Comb Chem High Throughput Screen*, 2006, **9**, 491.
185. M. Metifiot, O. Leon, L. Tarrago-Litvak, S. Litvak and M. L. Andreola, *Biochimie*, 2005, **87**, 911.
186. C. T. Lauhon and J. W. Szostak, *J Am Chem Soc*, 1995, **117**, 1246.
187. D. Levesque, J. D. Beaudoin, S. Roy and J. P. Perreault, *Biochem J*, 2007, **403**, 129.
188. J. Tan, C. Vonrhein, O. S. Smart, G. Bricogne, M. Bollati, Y. Kusov, G. Hansen, J. R. Mesters, C. L. Schmidt and R. Hilgenfeld, *PLoS Pathog*, 2009, **5**, e1000428.
189. H. Huang, N. B. Suslov, N. S. Li, S. A. Shelke, M. E. Evans, Y. Koldobskaya, P. A. Rice and J. A. Piccirilli, *Nat Chem Biol*, 2014, **10**, 686.
190. K. D. Warner, M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey and A. R. Ferre-D'Amare, *Nat Struct Mol Biol*, 2014, **21**, 658.
191. T. Li, E. Wang and S. Dong, *J Am Chem Soc*, 2009, **131**, 15082.
192. C. Zhao, Y. Song, J. Ren and X. Qu, *Biomaterials*, 2009, **30**, 1739.
193. D. Miyoshi, Z. M. Wang, H. Karimata and N. Sugimoto, *Nucleic Acids Symp Ser (Oxf)*, 2005, 43.
194. Y. Krishnan and F. C. Simmel, *Angew Chem Int Ed Engl*, 2011, **50**, 3124.
195. P. Agarwala, S. Pandey and S. Maiti, *Chembiochem*, 2013, **14**, 2077.

Table 1. Biological functions of RNA G-quadruplexes

ROLE	mRNA	PROBABLE FUNCTIONS	REFERENCES
1. Translational regulation (a) Repression	NRAS	G-quadruplex in the 5′ UTR of these cellular mRNAs represses translation.	13
	Human and bovine ESR		60, 61
	Zic-1		57
	BCL2		63
	TRF-2		64
	MT3-MMP		62
	ADAM-1		66
	Cyclin D3		67
	AKTIP		72
	CTSB		72
	Estrogen receptor	G-quadruplex present in ORF of this mRNA halt translational elongation.	78, 79
	PIM1	G-quadruplex in the 3′ UTR of this cellular mRNA is associated with its role in translational repression.	82
(b) Augmentation	FGF-2	G-quadruplex is essential for IRES-A mediated cap-independent translation initiation.	59
	VEGF		65
	TGFβ2	G-quadruplex in the 5′ UTR of these cellular mRNAs increases translation.	68
	FOXE3		72
2. 3′ end processing	IGF-I	This transcript undergoes a specific cleavage reaction in vivo at a site just 5′ to G-quadruplex motif.	12, 85
	p53	G-quadruplex downstream from the p53 pre-mRNA cleavage site is crucial for maintaining p53 3′ end processing efficiency following UV irradiation.	86
Alternative polyadenylation	LRP5 and FXR1	G-quadruplexes located in 3′ UTR of these genes increases alternative polyadenylation efficiency generating several short transcripts.	87
3. Transcription termination	Control region-conserved sequence block II (CSB II) of mitochondria	G-quadruplex helps in transcription termination and the mechanism is similar to Rho-independent transcription termination in prokaryotes.	88
4. mRNA localisation	PSD-95 and CaMKIIa	G-quadruplex in the 3′-UTR of these mRNA are neurite-targeting elements that aid transport of these mRNAs from soma to dendrites.	89
5. Alternative Splicing	src N1	G-quadruplexes affect the splicing and gene expression by providing binding sites for proteins required for splicing such as hnRNP protein family.	91
	Tropomyosin		92
	hTERT		93
	FMR1		94
	BACE1		96
	PAX9		97
	p53	G-quadruplex in intron 3 modulates the splicing of intron 2, leading to differential expression of transcripts encoding distinct p53 isoforms.	95

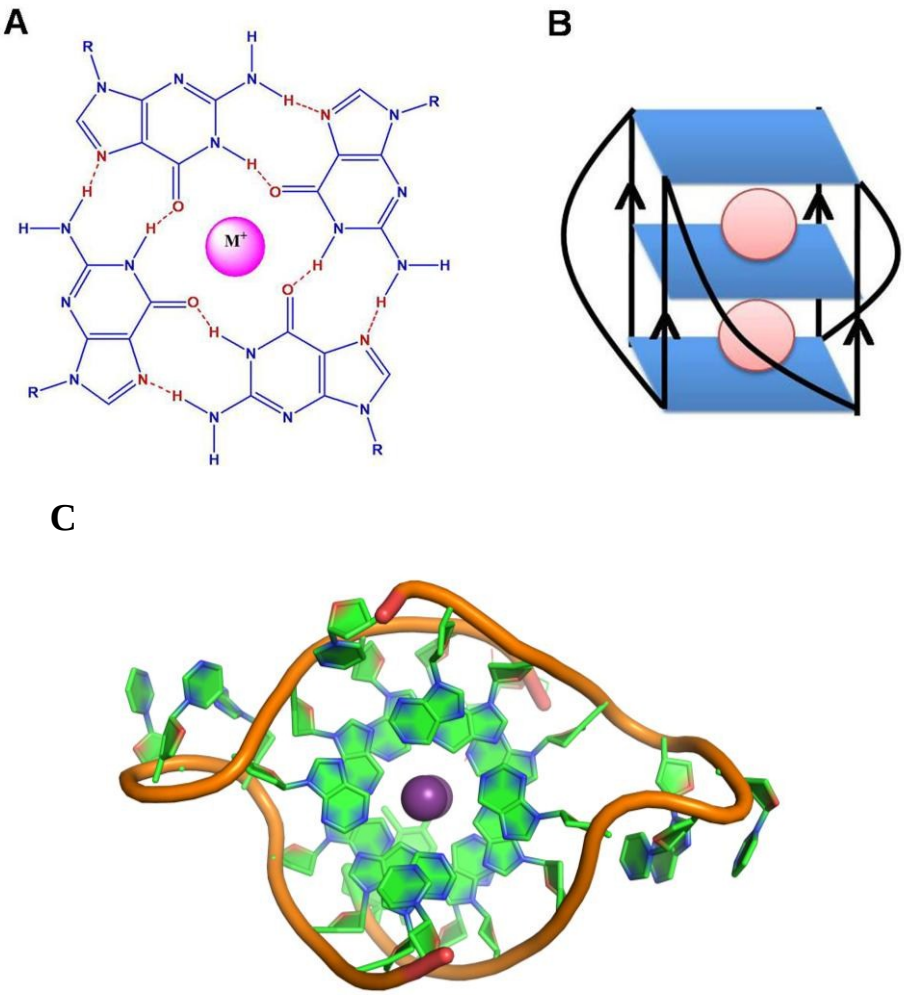


Figure 1. (A) Schematic representation of a G-quartet (B) Schematic representation of RNA G-quadruplex with parallel topology (C) Crystal structure of intermolecular RNA G-quadruplex (PDB ID – 3IBK).[50] Ribose sugar and nitrogenous bases are represented in green colour and phosphate backbone is represented in orange colour. Potassium ions (K^+) are shown in violet colour.

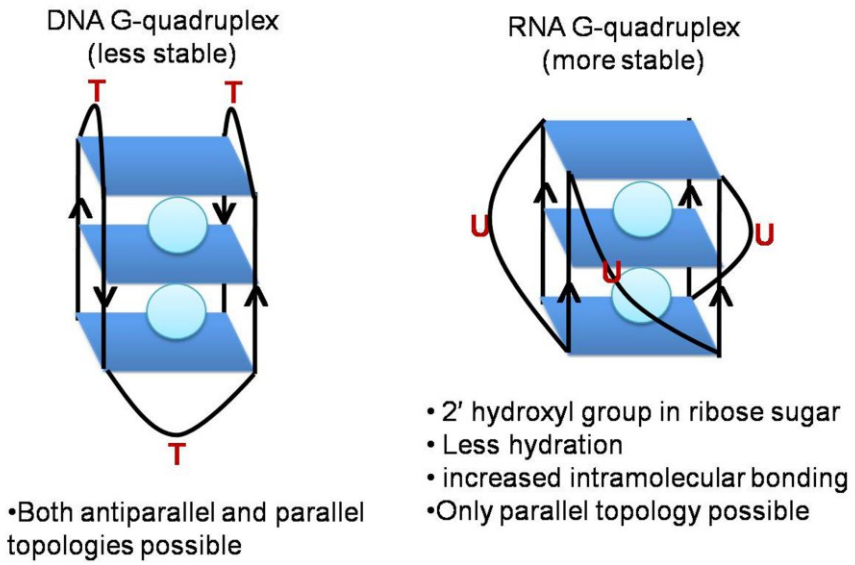


Figure 2. Differences between DNA G-quadruplex and RNA G-quadruplex.

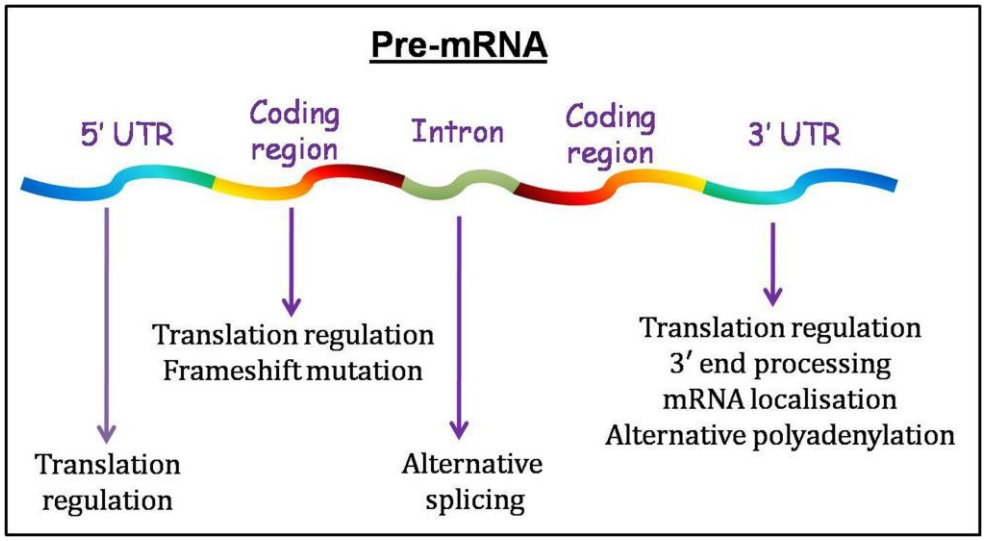


Figure 3. Functions of G-quadruplex located in different regions of pre-mRNA.

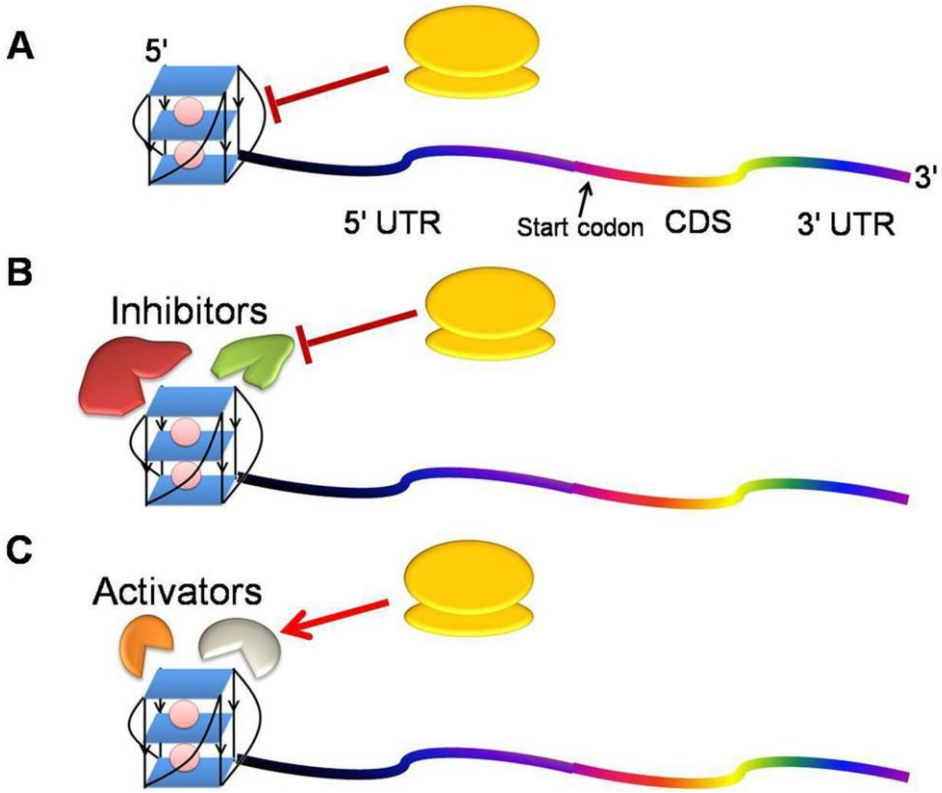


Figure 4. Probable mechanisms for regulation of gene expression by RNA G-quadruplex in 5′ UTR of mRNA. (A) G-quadruplex itself may interfere with the recruitment of preinitiation complex and subsequently delays the translation. (B) This structure may decoy inbibitors which in turn hinder the translation process. (C) G-quadruplex may employ factors that actively engage the translation machinery to translate mRNA into the protein.

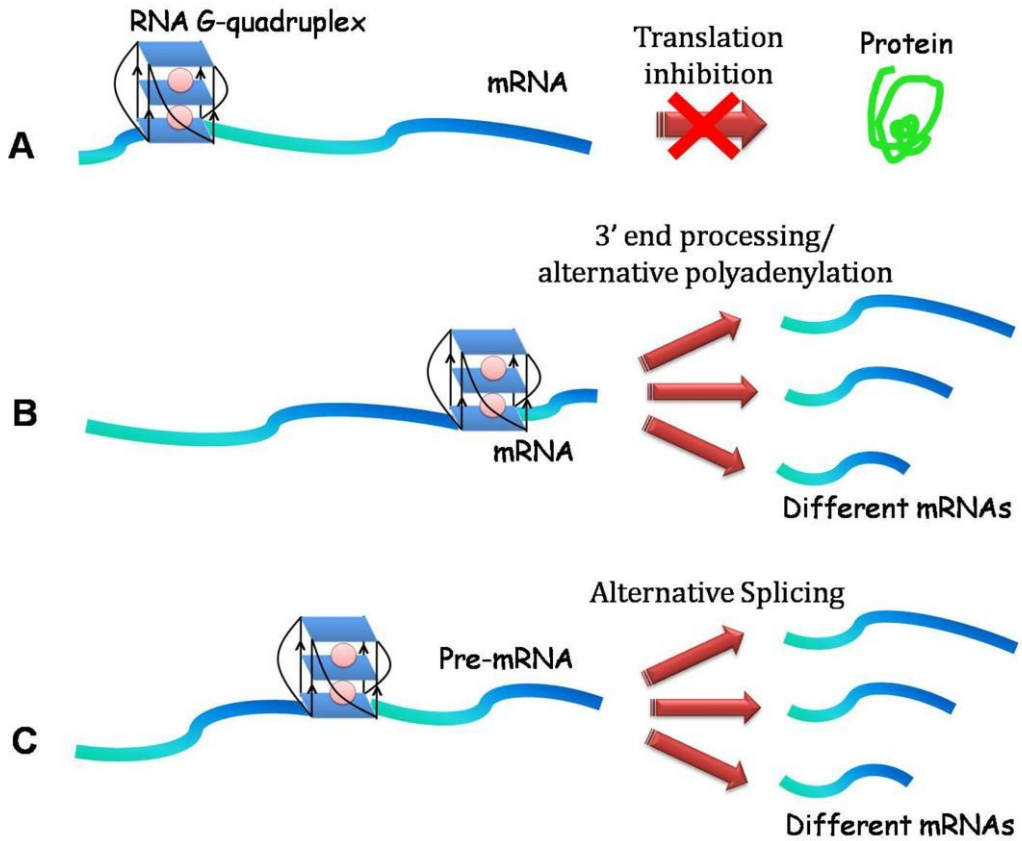


Figure 5. Functions of RNA G-quadruplex. (A) Translation repression. (B) 3′ end processing and alternative polyadenylation when located in 3′ UTR of mRNA. (C) Alternative splicing when present in introns.

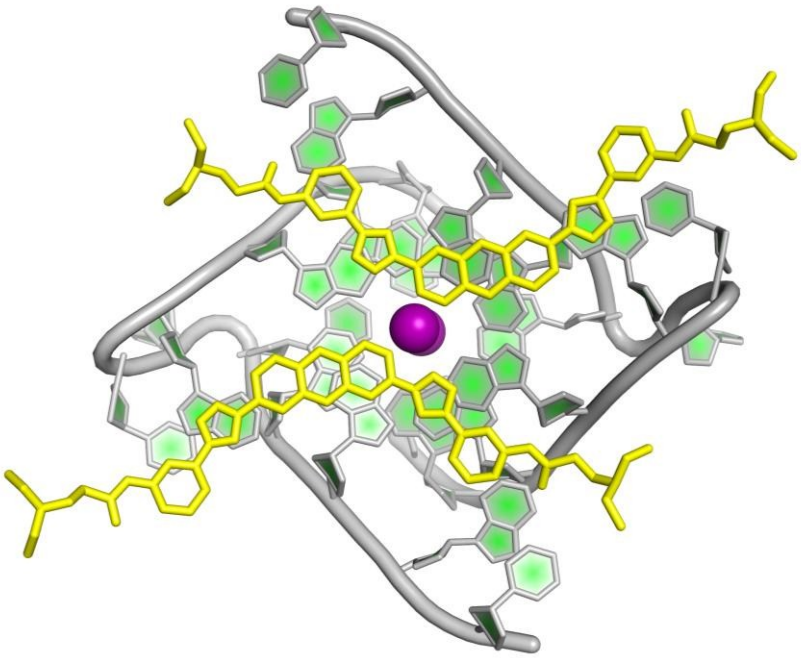


Figure 6. Crystal structure of RNA G-quadruplex-acridine complex (PDB ID - 3MIJ).[51] The grey, dark green and light green represents the phosphate backbone, ribose sugar and nitrogenous bases of RNA oligonucleotide, respectively. The acridine molecules N,N′-((1,10-(acridine-3,6-diyl)bis(1*H*-1,2,3-triazole-4,1-diyl))bis(3,1-phenylene))bis-(2-(diethylamino)acetamide) are shown in yellow colour. Potassium ions (K^+) are depicted in violet colour.

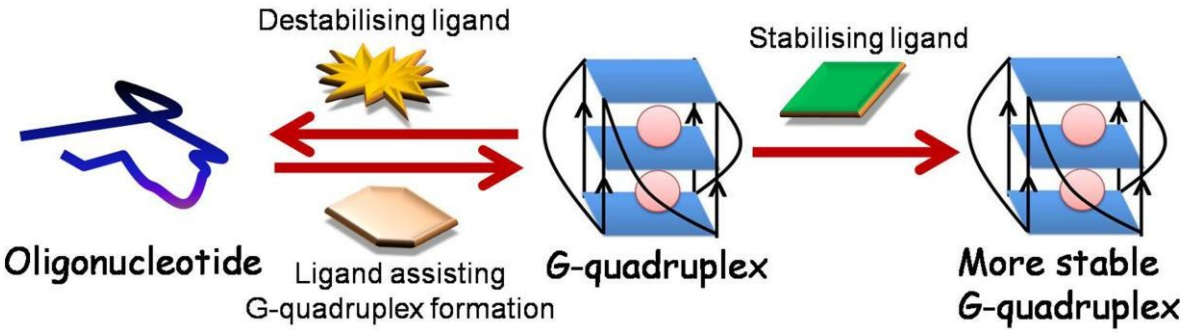


Figure 7. A ligand can stabilise, destabilise RNA G-quadruplex or can aid RNA G-quadruplex formation from oligonucleotide.

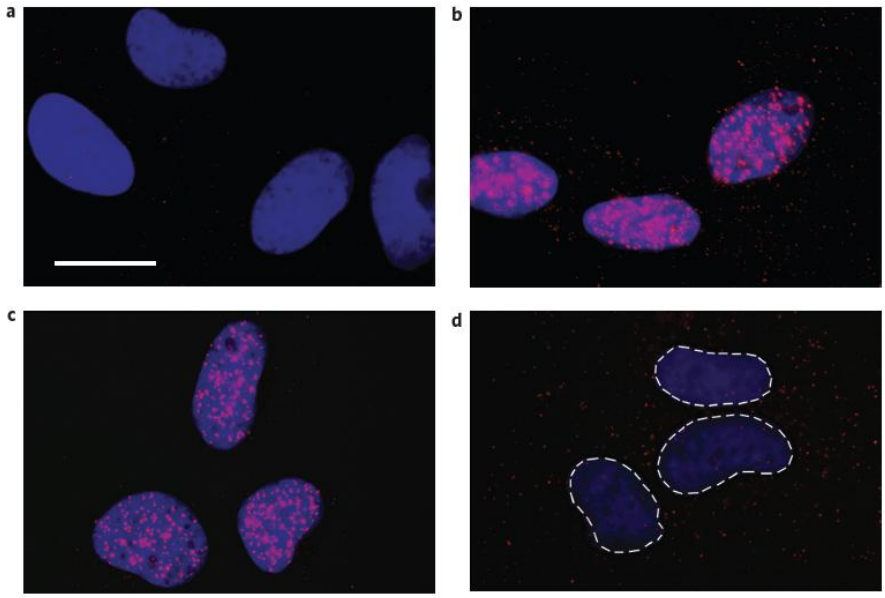


Figure 8. Visual identification of RNA G-quadruplex in the human cells. G-quadruplex specific antibody BG4 (shown in red) was used to detect RNA G-quadruplex in the cytoplasm of SV40- transformed MRC-5 fibroblast using indirect immunoflorescence microscopy. Nucleus is stained with DAPI and cytoplasm is seen as dark region around the nuclei. (a) No primary antibody (b) BG4 staining shows presence of G-quadruplex in both nucleus and cytoplasm (c) Loss of BG4 staining in cytoplasm after RNase A treatment (d) Loss of BG4 staining in nucleus after DNase treatment. Reprinted by permission from Macmillan Publishers Ltd: [Nature Chemistry],[22] copyright (2014).

Biographic sketch and photograph

Prachi Agarwala did her B.Sc. in Zoology (Honours) from the Presidency College, Kolkata and M.Sc. in Zoology from University of Calcutta, Kolkata, India. She is currently pursuing her Ph.D. degree under the supervision of Dr. Souvik Maiti at the CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. Her current research focuses on understanding the function G-quadruplex in Untranslated Region of mRNA.

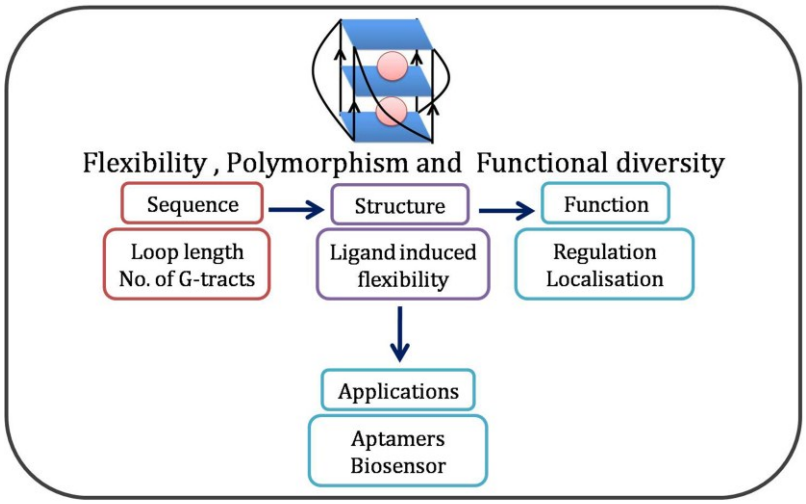
Satyaprakash Pandey obtained his B.Sc. degree in Biotechnology and M.Sc. degree in Life Sciences from University of Mumbai, India. He is currently pursuing a Ph.D. degree under the supervision of Dr. Souvik Maiti at the CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. His present research focuses on studying the role of G-quaduplex in non-coding RNA.

Souvik Maiti completed his M. Sc. in Chemistry from Jadavpur University, Kolkata, India, and Ph.D. in polymer chemistry from CSIR-Institute of Chemical Technology, Hyderabad, India. Subsequently, he performed postdoctoral research on biophysical aspects of DNA secondary structures and development of DNA-based nanomaterials in Professor Luis A. Marky's laboratory at University of Nebraska Medical Center, Omaha, Nebraska and Professor Francis Rondelez's laboratory, Curie University, Paris, France, respectively. He is currently a Principal Investigator in the Proteomics and Structural Biology Unit of CSIR-Institute of Genomics and Integrative Biology, New Delhi, India. His present research interests involve understanding the structural and functional aspects of G-quadruplexes in non-coding RNA and targeting these structures as well as miRNAs by chemically engineered oligonucleotides or small molecules.

Table of Content

Tale of RNA G-quadruplex

Prachi Agarwala, Satyaprakash Pandey, Souvik Maiti



RNA secondary structure G-quadruplex with its malleable nature can execute diverse biological functions, can be manipulated and be used for various applications.