G-Quadruplex-Binding Proteins

Subjects: Biology Contributor: Xiao Sun, Rongxin Zhang

G-quadruplexes (G4s) are non-canonical secondary nucleic acid structures. Sequences with the potential to form G4s are abundant in regulatory regions of the genome including telomeres, promoters and 5' non-coding regions, indicating they fulfill important genome regulatory functions. In recent years, an increasing number of G-quadruplex-binding proteins have been identified with biochemical experiments. G4-binding proteins are involved in vital cellular processes such as telomere maintenance, DNA replication, gene transcription, mRNA processing. Therefore, G4-binding proteins are also associated with various human diseases.

Keywords: G-quadruplex ; G-quadruplex-binding protein ; drug target ; biological functions ; structural properties

1. G-Quadruplex-Binding Proteins

1.1. Detection of G-Quadruplex-Binding Proteins

The G4 structure is highly dynamic in vivo and depends on the cell type and chromatin state ^[1]. Meanwhile, the formation and unwinding of G4 structures across the whole genome and transcriptome are directly or indirectly regulated by G4BPs, thereby affecting various biological processes ^{[1][2]}. Thus, the identification and an in-depth study of G4BPs can provide a full explanation of G4–protein interactions and their biological roles in vivo. These studies will further inspire the development of medical applications for these proteins.

G4BPs are most often identified by biochemical experiments. The commonly used methods include affinity chromatography, quantitative methods based on mass spectrometry, and fluorescence energy resonance transfer (FRET) technology. Affinity chromatography is often used in combination with mass spectrometry to separate proteins that bind to specific G4 motifs ^[3]. For instance, this method was applied to the identification of proteins binding to G4s in the 5' UTR of tumor-associated mRNA ^[4]. FRET is a spectroscopic technique that provides information about the conformation and dynamics of biomolecules. It has been widely used because this technology can detect whether there is a direct interaction between G4 structures and proteins in vivo ^{[5][6]}. In addition to performing biochemical experiments, computational analyses could be conducted to identify G4BPs. For example, putative G4 motifs can be predicted at known binding sites of nucleic acid-binding proteins, or the computational modeling of structural features may be exploited to discover new G4BPs ^{[1][2][8][9]}.

Recently, great advances have been made in the identification of G4BPs. The affinity purification experiments do not take into consideration the native chromatin state, so Shankar Balasubramanian et al. pioneered a co-binding-mediated protein profiling (CMPP) approach for the exploration of DNA G4BPs in living cells ^[10]. Researchers designed small-molecule ligands that specifically target DNA G4 in cells so that the probes could approach G4BPs with minimal interference with G4-protein interactions and enable labeling by subsequent photoproximity crosslinking ^[10]. The strategy was employed to identify hundreds of potential G4BPs, and finally in vitro experiments confirmed the binding specificity of several candidate proteins. Overall, this approach laid the foundations for the subsequent investigation of new G4BPs.

In conclusion, the detection methods are made up of in vivo, in vitro and in silico approaches. Generally, in vivo and in silico approaches are employed to identify potential G4BPs while in vitro approaches are utilized to confirm G4-protein interactions.

1.2. DNA G-Quadruplex-Binding Proteins

Recent discoveries related to the involvement of DNA G4BPs in the regulation of cellular fundamental functions will be discussed in the following section.

1.2.1. Telomeric G-Quadruplex-Binding Proteins

The human telomeric sequence was one of the first sequences discovered to form G4 structures. Telomeres are nucleoprotein complexes that constitute the ends of eukaryotic chromosomes, which play crucial roles in maintaining the integrity and stability of the genome ^{[11][12]}. When specific proteins bind to telomeric DNA, these proteins can prevent not only the degradation of the chromosome ends by nucleases, but also the recognition of them as broken fragments by the DNA repair mechanism ^{[12][13][14]}.

Telomeric DNA is highly conserved between vertebrates and consists of the identical TTAGGG short repeat sequences with a guanine-rich single-stranded 3' overhang ^[12]. These repeat sequences have the potential to form a G4 structure. The experiments using G4 specific antibodies and G4 ligands also confirmed the existence of G4 structures at telomeres in vivo ^[11].

Mammalian telomeric DNA is bound by a protein complex called shelterin, which protects the DNA termini from being considered as damaged and prevents the triggering of the repair mechanism (**Figure 1**) ^{[15][16][17]}. The proteins TRF1 and TRF2 (Telomere Repeat Binding Factor 1 and 2) in shelterin bind to double-stranded telomeric DNA; POT1 (Protection of Telomeres protein 1) binds to 3' overhang of telomeric repeats and regulates the folding and unwinding of the G4 structures with its heterodimeric partner TPP1 (TIN2 Interacting Protein) ^{[11][18][19][20]}. TPP1 connects POT1 to TRF1 and TRF2 via TIN2 (TRF1-interacting Nuclear protein 2) ^[21]. In addition, the study found that the helicases WRN (Werner syndrome ATP-dependent helicase) and BLM (Bloom syndrome protein) of the RecQ family are recruited to the telomeres and unfold the G4 structures to maintain the integrity of telomeres and ensure telomere replication ^{[22][23]}. WRN colocalizes with TRF2 and POT1, and both WRN and BLM can bind to POT1 with high affinity, which indicates that the telomeric DNA-binding proteins are essential for the recruitment of helicases ^[24].



Figure 1. Schematic diagram of the telomere-associated protein complexes shelterin and CST. Shelterin and CST play crucial roles in telomere maintenance. TPP1-POT1 subunit of shelterin regulates the folding and unwinding of G4 structures. CST could resolve and prevent the formation of G4 structures.

Mammalian cells also contain another telomere-associated protein complex called CST (CTC1-STN1-TEN1), which plays a crucial role in efficient telomere replication and in the maintenance of telomere length (**Figure 1**) ^{[21][25][26][27]}. Human CST is a single-stranded DNA-binding protein complex that helps to solve the genome-wide replication problems ^{[21][26]}. For example, GC-rich regions of the genome may induce obstacles to DNA replication, because DNA polymerase may stall at a G4. Experiments confirmed that CST could bind to the G4s and unfold them ^[21]. G4 structures possibly form on the lagging strand template at telomeres where WRN, BLM and POT1 all participate in G4 removal ^{[21][24][28][29]}. However, the presence of CST could make the replication of double-stranded telomeric DNA more effective as this complex unwinds G4 structures more rapidly than POT1 ^[21].

Other telomere-binding proteins also have similar functions as is the case for these two protein complexes.

1.2.2. G-Quadruplex-Binding Proteins Involved in Replication

The G4 structure has a dual effect on the process of DNA replication. On one hand, the G4 structure has been demonstrated to support the initiation of DNA replication at the replication origin ^[30]. Furthermore, the G4 structure may prevent the uncoupling of the leading- and lagging-strand polymerases, thereby protecting proper replication ^[31]. On the other hand, the G4 could hinder the progression of the replication fork and influence DNA synthesis, which may lead to

mutations and deletions in the genome. Consequently, helicases usually unfold the G4 structures before replication to maintain genome stability ^[31].

FANCJ (Fanconi anemia complementation group J) is a 5'–3' DNA helicase, which is involved in various biological processes such as DNA damage repair, G4 resolution, homologous recombination and genome stability maintenance ^[32]. FANCJ can unfold and remove G4 structures for efficient DNA replication while its absence will stop replication at G4s and eventually lead to DNA damage ^[33]. Studies have shown that FANCJ might promote replication at G4s by two independent mechanisms ^[34]. One mechanism is that FANCJ may cooperate with polymerase REV1 to aid replication at the replication fork ^[35]. REV1 destabilizes the G4 structures so that FANCJ can unwind them from the other side of the G4 structures. Second, WRN or BLM may assist FANCJ to bind and unfold the G4s from the opposite direction in order to promote replication synergistically ^{[34][36][37]}.

The helicase Pif1 from yeasts is able to bind and unfold G4 structures to support DNA replication. It is not clear whether Pif1 can play a role unwinding G4 structures on both chains or if it has a binding preference for the G4 structure at a certain chain ^[31]. However, recent studies have suggested that the ubiquitin ligase complex protein Mms1 is not only a DNA G4-binding protein, but also assists Pif1 to bind to a specific G4 structure located on the lagging strand. It could be observed that the absence of Mms1 leads to a reduction in Pif1 binding and slow replication at G4 motifs, and finally causes G4-dependent genome instability ^[38].

1.2.3. G-Quadruplex-Binding Proteins Involved in Transcription

It has been found that about 50% of human genes contain G4 motifs near their promoter region, which indicates that G4s play an essential role in the regulation of gene expression ^[24]. When DNA G4 is located at the first intron downstream of the transcription start site (TSS), it blocks the RNA polymerase and suppresses transcription ^[39]. However, recent studies have shown that endogenous G4s in promoters are prominent binding sites for multiple transcription factors and are thus invariably linked to high transcription levels ^{[39][40]}. Notably, G4s and their associated transcription factors cooperate to shape the cell-specific transcription factors among the 56 DNA G4-binding proteins in the G4IPDB (G4 Interacting Proteins DataBase) ^[42]. For example, SP1 (Specificity protein 1) is a zinc finger transcription factor, which can bind to the G4 structures on the *c-KIT* promoter and regulate the expression of a variety of housekeeping genes ^[39]. MAZ (Myc-associated zinc finger) and PARP-1 (Poly [ADP-ribose] polymerase 1) interact with the G4 structures upstream of the transcription start site of *KRAS*, and both of them are activators of *KRAS* ^{[1][24][43]}.

The G4 motif occurs more frequently in proto-oncogenes and regulatory genes than in housekeeping genes and tumor suppressor genes $^{[24][44][45]}$. The first reported G4 on the promoter is formed in the nuclease hypersensitivity element III1 (NHE III1) which locates upstream of the P1 promoter of the proto-oncogene *c-MYC* $^{[32][46]}$. This guanine-rich region controls 85–90% of the transcriptional activation of the gene, and can fold into an intramolecular parallel G4 as a transcriptional repressor element $^{[47]}$. In addition to *c-MYC*, many genes have been demonstrated to form G4 structures in the promoter regions, such as proto-oncogenes *VEGF* $^{[48]}$, *KRAS* $^{[49]}$, *BCL-2* $^{[50]}$ and *c-KIT* $^{[51]}$; human platelet-derived growth factor receptor *PDGFR-* β $^{[52]}$; human telomerase reverse transcriptase *hTERT* $^{[53]}$ and other genes $^{[32]}$. In particular, the G4s in the promoter regions of the proto-oncogenes have been most intensively studied so far $^{[1]}$.

Nucleolin (NCL) is a multifunctional phosphoprotein that is most abundant in the nucleolus. Nucleolin is mainly associated with ribosome biosynthesis and also involved in chromatin remodeling, transcriptional regulation, G4 binding and apoptosis ^[47]. Nucleolin can bind to the *c-MYC* G4 with high affinity and promote the formation and stabilization of G4 structures. The luciferase assay results also proved that the overexpression of nucleolin could contribute remarkably to a reduction in *c-MYC*-driven transcription ^[47]. Another protein NM23-H2 which belongs to the NM23 family of nucleoside diphosphate kinase (NDPK) has a completely different structure effect on G4s from nucleolin. It has a variety of functions, including kinase activity, promoter binding, transcriptional regulation and DNA repair ^[54]. Experiments have confirmed that NM23-H2 could bind to the *c-MYC* G4 to promote the unfolding of the G4 structure, thereby activating the transcription of *c-MYC* ^[54].

The tumor suppressor protein p53 functions in apoptosis, DNA repair, cell cycle regulation and aging. As a transcriptional regulator, p53 can inhibit the expression of cell cycle regulatory and growth promoting genes via multiple mechanisms and plays a key role in tumor suppression ^[55]. Previous studies have found that wild-type p53 (wtp53) and several types of mutant p53 (mutp53) have the ability to selectively bind *c-MYC* and *hTERT* promoter G4s ^[56], and the C-terminal region of p53 is essential for the recognition of the G4. Accordingly, the interaction between p53 and G4 structures in promoter regions of p53 target genes may play an important role in p53-mediated transcriptional regulation ^[55].

1.2.4. Other DNA G-Quadruplex-Binding Proteins

Direct evidence has demonstrated that the endogenous human G4 DNA landscape is dynamically shaped by chromatin relaxation or cell status $\frac{[57][58]}{100}$. Indeed, several G4BPs also function in chromatin structure regulation and histone modification $\frac{[59][60]}{100}$. For example, various epigenetic and chromatin remodeling enzymes bind selectively to DNA G4 $^{[1]}$. Genomic binding sites of the chromatin remodeling protein ATR-X colocalize with GC-rich tandem repeats and CpG islands (CGI) that have the potential to form G4 structures $\frac{[61][62]}{100}$.

Guanine-rich sequences are very common around CpG islands, with a high distribution rate of up to 80% ^{[32][63][64]}. The presence of G4 structures is closely related to the hypomethylation of CpG islands in the human genome. Studies have revealed that DNMT1 (DNA methyltransferase 1) interacts with these G4 sites, which is consistent with the results observed in biophysical experiments. Specifically, DNMT1 shows a higher binding affinity to G4 compared with double-stranded, single-stranded or hemimethylated DNA ^[65]. Biochemical analyses demonstrated that G4 structures inhibit the enzymatic activity of DNMT1, and the formation of G4 also hinders DNMT1 to protect specific CpG islands from methylation and inhibit local methylation ^[65].

In addition, it has been found that G4s colocalize with CTCF (CCCTC-binding factor) binding sites in CpG islands and interact with CTCF in vitro. G4 is also crucial to the localization of CTCF ^[66]. CTCF is frequently recruited to CpG islands that are usually hypomethylated. Furthermore, the enrichment of G4s at CpG islands maintains CGI hypomethylation, which may explain the correlation between CpG islands and CTCF ^[66]. CTCF also functions as a chromatin remodeling factor with the capability of nucleosome repositioning; therefore, G4 can facilitate the binding of CTCF to genomic DNA by recruiting chromatin proteins ^[60].

1.3. RNA G-Quadruplex-Binding Proteins

It is easier for single-stranded RNA to form G4s in guanine-rich regions, and G4 is also an important structural characteristic of mRNA [11][67]. Recently, in vitro experiments combining high-throughput sequencing with reverse transcriptase stalling at RNA G4s (rG4) have found more than 13,000 loci with the potential to form rG4 structures in the human transcriptome; and immunofluorescence using G4 specific antibodies demonstrated rG4 formation in cells [68][69]. Notably, the highest abundance of rG4 is in functional regions including 5' and 3'-UTR [67]. All these observations of the enrichment of rG4s in functionally important regions suggest that they play crucial roles in transcription termination, alternative splicing, translational regulation, and chromosome integrity maintenance [4][70].

A substantial number of proteins interacting with rG4s have been identified by biochemical experiments, for example hnRNPs, ribosomal proteins and splicing factors ^{[4][11]}. Although there are DNA and RNA G4 specific proteins, their binding proteins have a significant overlap due to structural similarities between DNA and RNA G4s ^[11]. Basically, the discrimination between DNA and RNA G4BPs may depend on their different biological functions. It was found that the fragile X mental retardation protein (FMRP) could bind to the G4s in its own mRNA coding region, thereby regulating its own translation through a negative feedback pathway ^[71]. Additionally, FMRP is likely to interact with G4s in other mRNAs for translation repression by the recruitment of translation inhibitors, miRNA pathway activation, and direct interaction with ribosomes ^[4]. FRAXE-associated mental retardation protein FMR2 could also bind to G4s in mRNAs and function in alternative splicing ^[72].

The rG4 in the region where proto-oncogene *NRAS* 5'-UTR folds into a stable intramolecular parallel G4 structure and it has been demonstrated that it represses translation in vitro ^[4]. The study revealed that DEAD box helicase DDX3X involved in several pathways of RNA biology could bind to *NRAS* rG4s and the mutations of DDX3X are associated with tumorigenesis, especially medulloblastoma ^[4]. In addition, some helicases such as DHX36 (DEAH-Box Helicase 36) and DDX21 are able to bind and unfold rG4 structures. Another multifunctional helicase DHX9 shows a binding affinity for several secondary nucleic acid structures including G4s, but it is more inclined to bind RNA substrates. Therefore, helicases with the function of recognition and resolution of rG4s may play essential roles in post-transcriptional biological processes such as mRNA translation, transportation and stability ^[67].

Although the vast majority of rG4s are present in mRNAs, others are also detected in long non-coding RNAs (IncRNAs) including nuclear paraspeckle assembly transcript 1 (*NEAT1*). *NEAT1* is involved in gene regulation as a scaffold for the assembly of paraspeckles ^[73]. An upregulation of *NEAT1* could be observed in the majority of solid tumors such as lung cancer, esophageal cancer and hepatocellular carcinoma, and *NEAT1* also plays a critical role in neurodegenerative diseases and viral infection ^{[74][75]}. Evidence has shown that nascent *NEAT1* transcripts interact directly with the non-POU domain-containing octamer-binding protein (NONO) through its conserved rG4 motifs. The primary paraspeckle formation

is required for the recruitment of NONO to *NEAT1* transcripts which stabilizes *NEAT1* and lays the foundation for the recruitment of additional protein components to facilitate subsequent steps of assembly and maturation ^[75].

2. Structural Properties of G-Quadruplex-Binding Proteins

2.1. RGG Domain

The RGG (Arginine-Glycine-Glycine) domain, also termed the RGG/RG motif or GAR (glycine-arginine-rich) domain is composed of repeat sequences rich in RGG or RG and is highly conserved in evolution (**Figure 2**) ^{[76][77]}. Researchers have discovered RGG/RG motifs in more than 1000 human proteins which influence transcription, precursor mRNA splicing, DNA damage signaling pathways, mRNA translation, and apoptosis ^[76]. A study analyzed the amino acid composition of 77 human G4-binding proteins ^[8]. Compared with a random subset of the human proteome and a well-defined group of nucleic acid binding proteins, the study demonstrated a significant enrichment of glycine and arginine and also high abundance in RR, GR and RG in G4BPs. Research was conducted to investigate the presence of a conserved RG-rich motif, which is a typical characteristic of G4BPs ^[8].



Figure 2. Structural properties of G-quadruplex-binding proteins. RGG/RG motifs are from NCL, hnRNP U and CIRBP. The RRM domain structure is derived from Protein Data Bank with structure code 2KRR (NCL). RRM domain is an $\alpha\beta$ sandwich structure composed of one four-stranded antiparallel β -sheet and two α -helices packed against the β -sheet. The OB-fold domain structure is derived from Protein Data Bank with structure code 5W2L (CTC1). OB-fold domain is a β -barrel formed by five antiparallel β -sheets.

The RGG domain is usually found in G4BPs and it has been shown to mediate G4-protein interactions. For example, hnRNP U contains the RGG domain $^{[12]}$. The C-terminal region of nucleolin composed of RNA-binding domain (RBD) 3 and 4 and the RGG domain is essential for the recognition of the *c-MYC* NHE III1 sequence and the promotion of G4 formation $^{[9]}$. In addition, more than half of the newly identified *NRAS* rG4BPs contain the GAR domain which has been proved to be critical for *NRAS* rG4-DDX3X interaction $^{[67]}$.

The short residue gap between RGG repeats in the RGG domain frequently contains aromatic amino acids. The research on the binding mechanisms of the RGG domain revealed that the small segment RGG motif in this domain greatly contributes to the G4 binding affinity. Huang et al. found that the internal arrangement of RGG repeats and gap amino acids are more fundamental to G4-protein interactions than the length of RGG peptides and numbers of RGG repeats ^[9].

Experiments demonstrated that the peptide 12 with seven RGG repeats could efficiently bind to DNA G4s. Based on the above results, they discovered that the cold-inducible RNA-binding protein (CIRBP) containing peptide 12 could bind G4s both in vitro and in vivo, and this RGG peptide is essential for the G4 recognition of CIRBP ^[9]. The team provided a great deal of insight into the interaction between the RGG peptide and G4s, and identified a new G4-binding protein based on the exploration of G4-binding RGG motifs. In summary, this approach also adds a new dimension to the discovery of other G4BPs.

2.2. RRM Domain

Several G4BPs, such as hnRNPs, nucleolin, CIRBP, TLS/FUS (translocated in liposarcoma, also known as fused in sarcoma), and EWS (Ewing's sarcoma), have shared structural features, such as RNA recognition motifs (RRM) and RGG domains ^[78]. RRM, also known as the RNA-binding domain (RBD) or ribonucleoprotein domain (RNP), is one of the most highly conserved nucleic acid binding domains that occurs in approximately 0.5–1% of human genes and folds into an $\alpha\beta$ sandwich structure composed of one four-stranded antiparallel β -sheet and two α -helices packed against the β -sheet (**Figure 2**) ^{[79][80][81]}. Proteins with RRM are implicated in the regulation of transcription, translation, RNA processing, RNA export and stability ^[82], and they are also common in G4BPs.

The RRM and RGG domains at the C-terminal of nucleolin are necessary to inhibit and induce the formation of the G4 on the *c-MYC* promoter. The RRM in nucleolin can form G4s with guanine-containing single strands, but it unfolds G4s without guanines in the single strands of the 5' and 3' terminals ^[B2]. The RRMs of hnRNP A1 and hnRNP D are able to bind and unfold G4s. The crystal structure of the two RRMs of hnRNP A1 with single-stranded telomeric DNA showed that RRM1 and RRM2 interact directly with d(TAGG) and d(TTAGG), respectively ^[B3]. The RRM of hnRNP D could recognize d(TAG) in d(TTAGGG) determined by NMR ^[B4]. A recent study indicated that a novel G4-binding protein SLIRP (stem-loop interacting RNA binding protein) also contains the RRM domain, which is required for efficient interaction between DNA G4s and SLIRP ^[B5]. Furthermore, the sequence alignment for the RRMs derived from SLIRP and other G4BPs such as hnRNP A1 and nucleolin showed similar amino acid composition of these domains ^[B5]. The findings of these studies shed light on the roles of the RRM domain conserved in many nucleic acid binding proteins and contribute greatly to the exploration of its biological functions.

2.3. OB-Fold Domain

Oligonucleotide/oligosaccharide binding (OB)-fold is a β -barrel structure comprising a five-stranded antiparallel β -sheet, and this barrel is capped by an α -helix located between the third and fourth strands (**Figure 2**) ^[86]. The OB-fold structure is highly dynamic, and the dynamic properties enable OB-fold containing proteins to participate in multiple cellular pathways, including the re-initiation of DNA synthesis and the maintenance of genome stability ^[87].

Replication protein A (RPA) is a single-stranded DNA-binding complex with three subunits which unfolds the G4s and is involved in various biological processes such as DNA replication, repair and recombination. Although both RPA and POT1-TPP1 can bind to telomeric overhangs, RPA is more abundant in cells ^[11]. The CST complex resembles RPA in that they harbor comparable arrays of OB-folds and possess small subunits with similar structures ^[21]. Since CST contains multiple OB-folds (one each in STN1 and TEN1, and seven in CTC1), it was estimated that CST could play distinct roles in replication using a dynamic binding mechanism similar to that observed in RPA ^{[21][88][89]}. The dynamic properties of RPA binding due to the microscopic dissociation and re-association of individual OB-folds allow RPA to diffuse along the single-stranded DNA and to melt unwanted DNA secondary structures ^[21]. In addition, POT1 also contains the OB-fold domain, and FRET has shown that it is critical for gradual G4 unfolding ^[28].

DHX36 can bind DNA and RNA G4 structures with high affinity. It is a multifunctional helicase involved in G4-dependent transcriptional and post-transcriptional regulation, and plays a critical role in heart development, hematopoiesis and embryogenesis in mice ^[90]. The DHX36-specific motif at the N-terminal of the protein forms a DNA-binding-induced α -helix that together with the OB-fold-like subdomain selectively binds to parallel G4s ^[90].

References

- 1. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. Trends Chem. 2020, 2, 123–136.
- 2. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. 2020, 21, 459–474.

- Matsumoto, K.; Okamoto, K.; Okabe, S.; Fujii, R.; Ueda, K.; Ohashi, K.; Seimiya, H. G-quadruplex-forming nucleic acids interact with splicing factor 3B subunit 2 and suppress innate immune gene expression. Genes Cells 2021, 26, 65–82.
- 4. Serikawa, T.; Spanos, C.; von Hacht, A.; Budisa, N.; Rappsilber, J.; Kurreck, J. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. Biochimie 2018, 144, 169–184.
- Lee, C.Y.; McNerney, C.; Myong, S. G-Quadruplex and Protein Binding by Single-Molecule FRET Microscopy. Methods Mol. Biol. 2019, 2035, 309–322.
- Scalabrin, M.; Frasson, I.; Ruggiero, E.; Perrone, R.; Tosoni, E.; Lago, S.; Tassinari, M.; Palu, G.; Richter, S.N. The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. Sci. Rep. 2017, 7, 45244.
- McRae, E.K.S.; Booy, E.P.; Padilla-Meier, G.P.; McKenna, S.A. On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. J. Nucleic Acids 2017, 2017, 9675348.
- Brazda, V.; Cerven, J.; Bartas, M.; Mikyskova, N.; Coufal, J.; Pecinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. Molecules 2018, 23, 2341.
- 9. Huang, Z.L.; Dai, J.; Luo, W.H.; Wang, X.G.; Tan, J.H.; Chen, S.B.; Huang, Z.S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. J. Am. Chem Soc. 2018, 140, 17945–17955.
- 10. Zhang, X.; Spiegel, J.; Martinez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical profiling of DNA G-quadruplexinteracting proteins in live cells. Nat. Chem. 2021, 13, 626–633.
- 11. Brazda, V.; Haronikova, L.; Liao, J.C.; Fojta, M. DNA and RNA quadruplex-binding proteins. Int. J. Mol. Sci. 2014, 15, 17493–17517.
- 12. Izumi, H.; Funa, K. Telomere Function and the G-Quadruplex Formation are Regulated by hnRNP U. Cells 2019, 8, 390.
- 13. Takahama, K.; Kino, K.; Arai, S.; Kurokawa, R.; Oyoshi, T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. FEBS J. 2011, 278, 988–998.
- 14. Takahama, K.; Takada, A.; Tada, S.; Shimizu, M.; Sayama, K.; Kurokawa, R.; Oyoshi, T. Regulation of Telomere Length by G-Quadruplex Telomere DNA- and TERRA-Binding Protein TLS/FUS. Chem. Biol. 2013, 20, 341–350.
- 15. Chen, Y. The structural biology of the shelterin complex. Biol. Chem. 2019, 400, 457–466.
- 16. Stewart, J.A.; Chaiken, M.F.; Wang, F.; Price, C.M. Maintaining the end: Roles of telomere proteins in end-protection, telomere replication and length regulation. Mutat. Res.-Fund. Mol. M 2012, 730, 12–19.
- 17. Arnoult, N.; Karlseder, J. Complex interactions between the DNA-damage response and mammalian telomeres. Nat. Struct. Mol. Biol. 2015, 22, 859–866.
- 18. Baumann, P.; Cech, T.R. Pot1, the putative telomere end-binding protein in fission yeast and humans. Science 2001, 292, 1171–1175.
- 19. Wang, F.; Podell, E.R.; Zaug, A.J.; Yang, Y.; Baciu, P.; Cech, T.R.; Lei, M. The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature 2007, 445, 506–510.
- Chaires, J.B.; Gray, R.D.; Dean, W.L.; Monsen, R.; DeLeeuw, L.W.; Stribinskis, V.; Trent, J.O. Human POT1 unfolds Gquadruplexes by conformational selection. Nucleic Acids Res. 2020, 48, 4976–4991.
- 21. Bhattacharjee, A.; Wang, Y.; Diao, J.; Price, C.M. Dynamic DNA binding, junction recognition and G4 melting activity underlie the telomeric and genome-wide roles of human CST. Nucleic Acids Res. 2017, 45, 12311–12324.
- 22. Wu, W.; Rokutanda, N.; Takeuchi, J.; Lai, Y.; Maruyama, R.; Togashi, Y.; Nishikawa, H.; Arai, N.; Miyoshi, Y.; Suzuki, N.; et al. HERC2 Facilitates BLM and WRN Helicase Complex Interaction with RPA to Suppress G-Quadruplex DNA. Cancer Res. 2018, 78, 6371–6385.
- Budhathoki, J.B.; Ray, S.; Urban, V.; Janscak, P.; Yodh, J.G.; Balci, H. RecQ-core of BLM unfolds telomeric Gquadruplex in the absence of ATP. Nucleic Acids Res. 2014, 42, 11528–11545.
- 24. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015, 43, 8627–8637.
- Miyake, Y.; Nakamura, M.; Nabetani, A.; Shimamura, S.; Tamura, M.; Yonehara, S.; Saito, M.; Ishikawa, F. RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. Mol. Cell 2009, 36, 193–206.
- 26. Zhang, M.M.; Wang, B.; Li, T.F.; Liu, R.; Xiao, Y.N.; Geng, X.; Li, G.; Liu, Q.; Price, C.M.; Liu, Y.; et al. Mammalian CST averts replication failure by preventing G-quadruplex accumulation. Nucleic Acids Res. 2019, 47, 5243–5259.

- Surovtseva, Y.V.; Churikov, D.; Boltz, K.A.; Song, X.Y.; Lamb, J.C.; Warrington, R.; Leehy, K.; Heacock, M.; Price, C.M.; Shippen, D.E. Conserved Telomere Maintenance Component 1 Interacts with STN1 and Maintains Chromosome Ends in Higher Eukaryotes. Mol. Cell 2009, 36, 207–218.
- 28. Hwang, H.; Buncher, N.; Opresko, P.L.; Myong, S. POT1-TPP1 regulates telomeric overhang structural dynamics. Structure 2012, 20, 1872–1880.
- 29. Leon-Ortiz, A.M.; Svendsen, J.; Boulton, S.J. Metabolism of DNA secondary structures at the eukaryotic replication fork. DNA Repair 2014, 19, 152–162.
- 30. Valton, A.L.; Prioleau, M.N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? Trends Genet. 2016, 32, 697–706.
- 31. Sauer, M.; Paeschke, K. G-quadruplex unwinding helicases and their function in vivo. Biochem. Soc. Trans. 2017, 45, 1173–1182.
- 32. Sun, Z.Y.; Wang, X.N.; Cheng, S.Q.; Su, X.X.; Ou, T.M. Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid(-)Protein Interaction. Molecules 2019, 24, 396.
- Castillo Bosch, P.; Segura-Bayona, S.; Koole, W.; van Heteren, J.T.; Dewar, J.M.; Tijsterman, M.; Knipscheer, P. FANCJ promotes DNA synthesis through G-quadruplex structures. EMBO J. 2014, 33, 2521–2533.
- 34. Sarkies, P.; Murat, P.; Phillips, L.G.; Patel, K.J.; Balasubramanian, S.; Sale, J.E. FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res. 2012, 40, 1485–1498.
- 35. Eddy, S.; Ketkar, A.; Zafar, M.K.; Maddukuri, L.; Choi, J.Y.; Eoff, R.L. Human Rev1 polymerase disrupts G-quadruplex DNA. Nucleic Acids Res. 2014, 42, 3272–3285.
- 36. Mendoza, O.; Bourdoncle, A.; Boule, J.B.; Brosh, R.M., Jr.; Mergny, J.L. G-quadruplexes and helicases. Nucleic Acids Res. 2016, 44, 1989–2006.
- 37. Suhasini, A.N.; Brosh, R.M. Fanconi anemia and Bloom's syndrome crosstalk through FANCJ-BLM helicase interaction. Trends Genet. 2012, 28, 7–13.
- Schwindt, E.; Paeschke, K. Mms1 is an assistant for regulating G-quadruplex DNA structures. Curr. Genet. 2018, 64, 535–540.
- 39. Kim, N. The Interplay between G-quadruplex and Transcription. Curr. Med. Chem. 2019, 26, 2898–2917.
- 40. Spiegel, J.; Cuesta, S.M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol. 2021, 22, 117.
- Lago, S.; Nadai, M.; Cernilogar, F.M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S.N. Promoter Gquadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. Nat. Commun. 2021, 12, 3885.
- 42. Mishra, S.K.; Tawani, A.; Mishra, A.; Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci. Rep. 2016, 6, 38144.
- Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K.K.; Xodo, L.E. The KRAS promoter responds to Mycassociated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GAelement. J. Biol. Chem. 2010, 285, 22003–22016.
- 44. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007, 35, 406–413.
- 45. Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. 2006, 34, 3887–3896.
- 46. Simonsson, T.; Pecinka, P.; Kubista, M. DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. 1998, 26, 1167–1172.
- Gonzalez, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplexbinding protein. J. Biol. Chem. 2009, 284, 23622–23635.
- 48. Sun, D.; Guo, K.; Rusche, J.J.; Hurley, L.H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplexinteractive agents. Nucleic Acids Res. 2005, 33, 6070–6080.
- 49. Cogoi, S.; Xodo, L.E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 2006, 34, 2536–2549.
- 50. Dexheimer, T.S.; Sun, D.; Hurley, L.H. Deconvoluting the structural and drug-recognition complexity of the Gquadruplex-forming region upstream of the bcl-2 P1 promoter. J. Am. Chem. Soc. 2006, 128, 5404–5415.

- Rankin, S.; Reszka, A.P.; Huppert, J.; Zloh, M.; Parkinson, G.N.; Todd, A.K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. J. Am. Chem. Soc. 2005, 127, 10584–10589.
- 52. Qin, Y.; Fortin, J.S.; Tye, D.; Gleason-Guzman, M.; Brooks, T.A.; Hurley, L.H. Molecular cloning of the human plateletderived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. Biochemistry 2010, 49, 4208–4219.
- 53. Palumbo, S.L.; Ebbinghaus, S.W.; Hurley, L.H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. J. Am. Chem Soc. 2009, 131, 10878–10891.
- 54. Thakur, R.K.; Kumar, P.; Halder, K.; Verma, A.; Kar, A.; Parent, J.L.; Basundra, R.; Kumar, A.; Chowdhury, S. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. Nucleic Acids Res. 2009, 37, 172–183.
- 55. Petr, M.; Helma, R.; Polaskova, A.; Krejci, A.; Dvorakova, Z.; Kejnovska, I.; Navratilova, L.; Adamik, M.; Vorlickova, M.; Brazdova, M. Wild-type p53 binds to MYC promoter G-quadruplex. Biosci. Rep. 2016, 36, e00397.
- Quante, T.; Otto, B.; Brazdova, M.; Kejnovska, I.; Deppert, W.; Tolstonog, G.V. Mutant p53 is a transcriptional co-factor that binds to G-rich regulatory regions of active genes and generates transcriptional plasticity. Cell Cycle 2012, 11, 3290–3303.
- 57. Hansel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. Nat. Genet. 2016, 48, 1267–1272.
- 58. Guilbaud, G.; Murat, P.; Recolin, B.; Campbell, B.C.; Maiter, A.; Sale, J.E.; Balasubramanian, S. Local epigenetic reprogramming induced by G-quadruplex ligands. Nat. Chem. 2017, 9, 1110–1117.
- Zyner, K.G.; Simeone, A.; Flynn, S.M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S. G-quadruplex DNA structures in human stem cells and differentiation. Nat. Commun. 2022, 13, 142.
- 60. Hou, Y.; Li, F.Y.; Zhang, R.X.; Li, S.; Liu, H.D.; Qin, Z.H.S.; Sun, X. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. Epigenetics-Us 2019, 14, 894–911.
- Law, M.J.; Lower, K.M.; Voon, H.P.; Hughes, J.R.; Garrick, D.; Viprakasit, V.; Mitson, M.; De Gobbi, M.; Marra, M.; Morris, A.; et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a sizedependent manner. Cell 2010, 143, 367–378.
- 62. Wang, Y.; Yang, J.; Wild, A.T.; Wu, W.H.; Shah, R.; Danussi, C.; Riggins, G.J.; Kannan, K.; Sulman, E.P.; Chan, T.A.; et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nat. Commun. 2019, 10, 943.
- 63. Cayrou, C.; Coulombe, P.; Vigneron, A.; Stanojcic, S.; Ganier, O.; Peiffer, I.; Rivals, E.; Puy, A.; Laurent-Chabalier, S.; Desprat, R.; et al. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. 2011, 21, 1438–1449.
- 64. Besnard, E.; Babied, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.M.; Lemaitre, J.M. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. 2012, 19, 837–844.
- Mao, S.Q.; Ghanbarian, A.T.; Spiegel, J.; Martinez Cuesta, S.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. 2018, 25, 951–957.
- Tikhonova, P.; Pavlova, I.; Isaakova, E.; Tsvetkov, V.; Bogomazova, A.; Vedekhina, T.; Luzhin, A.V.; Sultanov, R.; Severov, V.; Klimina, K.; et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. Int. J. Mol. Sci. 2021, 22, 7090.
- 67. Herdy, B.; Mayer, C.; Varshney, D.; Marsico, G.; Murat, P.; Taylor, C.; D'Santos, C.; Tannahill, D.; Balasubramanian, S. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. Nucleic Acids Res. 2018, 46, 11592–11604.
- 68. Kwok, C.K.; Marsico, G.; Sahakyan, A.B.; Chambers, V.S.; Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Methods 2016, 13, 841–844.
- 69. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and selective chemical targeting of RNA Gquadruplex structures in the cytoplasm of human cells. Nat. Chem. 2014, 6, 75–80.
- 70. Agarwala, P.; Pandey, S.; Maiti, S. The tale of RNA G-quadruplex. Org. Biomol. Chem. 2015, 13, 5570–5585.
- 71. Schaeffer, C.; Bardoni, B.; Mandel, J.L.; Ehresmann, B.; Ehresmann, C.; Moine, H. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. EMBO J. 2001, 20, 4803–4813.

- 72. Bensaid, M.; Melko, M.; Bechara, E.G.; Davidovic, L.; Berretta, A.; Catania, M.V.; Gecz, J.; Lalli, E.; Bardoni, B. FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. Nucleic Acids Res. 2009, 37, 1269–1279.
- Clemson, C.M.; Hutchinson, J.N.; Sara, S.A.; Ensminger, A.W.; Fox, A.H.; Chess, A.; Lawrence, J.B. An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. Mol. Cell 2009, 33, 717–726.
- 74. Simko, E.A.J.; Liu, H.; Zhang, T.; Velasquez, A.; Teli, S.; Haeusler, A.R.; Wang, J. G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural IncRNA. Nucleic Acids Res. 2020, 48, 7421–7438.
- 75. Tassinari, M.; Richter, S.N.; Gandellini, P. Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. Nucleic Acids Res. 2021, 49, 3617–3633.
- 76. Thandapani, P.; O'Connor, T.R.; Bailey, T.L.; Richard, S. Defining the RGG/RG motif. Mol. Cell 2013, 50, 613–623.
- 77. Kharel, P.; Becker, G.; Tsvetkov, V.; Ivanov, P. Properties and biological impact of RNA G-quadruplexes: From order to turmoil and back. Nucleic Acids Res. 2020, 48, 12534–12555.
- 78. Masuzawa, T.; Oyoshi, T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. ACS Omega 2020, 5, 5202–5208.
- 79. Arumugam, S.; Miller, M.C.; Maliekal, J.; Bates, P.J.; Trent, J.O.; Lane, A.N. Solution structure of the RBD1,2 domains from human nucleolin. J. Biomol. Nmr 2010, 47, 79–83.
- 80. Clery, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: Boring? Not quite. Curr. Opin. Struct. Biol 2008, 18, 290–298.
- Maris, C.; Dominguez, C.; Allain, F.H. The RNA recognition motif, a plastic RNA-binding platform to regulate posttranscriptional gene expression. FEBS J. 2005, 272, 2118–2131.
- Oyoshi, T.; Masuzawa, T. Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins. Biochem. Biophys. Res. Commun. 2020, 531, 39–44.
- Ding, J.; Hayashi, M.K.; Zhang, Y.; Manche, L.; Krainer, A.R.; Xu, R.M. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. 1999, 13, 1102–1115.
- Enokizono, Y.; Konishi, Y.; Nagata, K.; Ouhashi, K.; Uesugi, S.; Ishikawa, F.; Katahira, M. Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. J. Biol. Chem. 2005, 280, 18862–18870.
- Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. J. Am. Chem. Soc. 2017, 139, 12426–12429.
- Murzin, A.G. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for nonhomologous sequences. EMBO J. 1993, 12, 861–867.
- 87. Nguyen, D.D.; Kim, E.Y.; Sang, P.B.; Chai, W. Roles of OB-Fold Proteins in Replication Stress. Front. Cell Dev. Biol. 2020, 8, 574466.
- Lim, C.J.; Barbour, A.T.; Zaug, A.J.; Goodrich, K.J.; Mckay, A.E.; Wuttke, D.S.; Cech, T.R. The structure of human CST reveals a decameric assembly bound to telomeric DNA. Science 2020, 368, 1081–1085.
- Shastrula, P.K.; Rice, C.T.; Wang, Z.; Lieberman, P.M.; Skordalakes, E. Structural and functional analysis of an OB-fold in human Ctc1 implicated in telomere maintenance and bone marrow syndromes. Nucleic Acids Res. 2018, 46, 972– 984.
- 90. Chen, M.C.; Tippana, R.; Demeshkina, N.A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D'Amare, A.R. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. Nature 2018, 558, 465–469.

Retrieved from https://encyclopedia.pub/entry/history/show/55452