

Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins

著者	Oyoshi Takanori, Masuzawa Tatsuki
journal or publication title	Biochemical and Biophysical Research Communications
volume	531
number	1
page range	39-44
year	2020-10-08
出版者	Elsevier
権利	(C) 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
URL	http://hdl.handle.net/10297/00027717

doi: 10.1016/j.bbrc.2020.02.178

Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins

Takanori Oyoshi*, Tatsuki Masuzawa

Department of Chemistry, Graduate School of Science, Shizuoka University, 836 Ohya Suruga, Shizuoka, 422-8529, Japan

*To whom correspondence should be addressed. Tel: +81-54-238-4760; Fax: +81-54-237-3384; Email: oyoshi.takanori@shizuoka.ac.jp

Abstract

The functions of local conformations of non-B form secondary structures in DNA and RNA, such as the G-quadruplex, are thought to be regulated by DNA and RNA binding proteins. Recent studies reported that G-quadruplexes regulate epigenetics through G-quadruplex binding proteins. G-quadruplex binding proteins control the role of G-quadruplexes in cells and regulate G-quadruplex formation. We discuss histone modifications through G-quadruplex RNA with G-quadruplex binding proteins and G-quadruplex binding protein modulation of G-quadruplex conformations. G-quadruplex RNA is involved in telomere maintenance and transcription via histone modification. Furthermore, the biologic function and formation of G-quadruplexes are regulated through folding or unfolding by

G-quadruplex binding proteins. In particular, in this review, we focus on the G-quadruplex RRM and RGG binding domains.

Keywords: G-quadruplex binding protein, epigenetics, RGG domain, RRM domain

1. Introduction

G-quadruplex structures are formed at specific G-rich regions in the genome, mRNA, and non-coding RNA [1,2]. These structures comprise two or more stacked guanine (G)-tetrad planes and a monovalent cation such as K^+ or Na^+ . G-quadruplex DNA regulates transcription, telomere elongation, epigenetics, and replication, which has been extensively reviewed [3-5]. Recent reviews have focused on the functions of G-quadruplexes for regulating epigenetics [6-8]. G-quadruplex DNA binds to G-quadruplex binding proteins, which interact with histones or DNA modification enzymes in the promoter and telomere, and regulate epigenetics through both histone and DNA modifications. Furthermore, G-quadruplexes interact directly with chromatin remodeling factors. The formation of histone-modified proteins is disrupted by G-quadruplex helicases. These functions and related proteins were reviewed previously [9].

G-quadruplex RNA is thought to have biologic functions, such as transcriptional regulation, RNA processing, mRNA localization and translation regulation [10,11]. While these functions have been reviewed, little attention has been focused on how G-quadruplex RNA affects histone modification and how G-quadruplex structures are regulated by

G-quadruplex binding proteins. Recent results revealed that G-quadruplex RNA also regulates histone modification through its binding proteins. Moreover, G-quadruplex structures are folded or unfolded by their binding proteins and these systems might affect transcription and other biologic functions. These studies have clarified the mechanisms and led to the development of medicine targeting these structures. Here, we discuss recent studies of the epigenetics controlled by G-quadruplex RNA with its binding proteins and G-quadruplex folding or unfolding proteins. In this review, we focus on the RNA recognition motif (RRM) and Arg-Gly-Gly (RGG) domains.

2. G-quadruplex RNA binding proteins regulating epigenetics

Translocated in liposarcoma/Fused in sarcoma (TLS/FUS) binds to telomere DNA and telomeric repeat-containing RNA (TERRA) with G-quadruplex structure specificity and together they form a ternary complex (Figure 1A) [12]. Moreover, TLS/FUS directly interacts with Suv4-20h2 and TERRA to promote histone modifications with the histone H3 modifying enzyme. These interactions facilitate the trimethylation of histone H4 at lysine 20 (H4K20me3) and histone H3 at lysine 9 (H3K9me3), respectively, and act to maintain heterochromatin and control telomere length [12,13]. A recent individual-nucleotide-resolution UV cross-linking and immunoprecipitation assay revealed TLS/FUS binding to G-rich RNA with the potential to form a G-quadruplex in nascent RNAs [14]. This is a promising method for revealing the novel biologic functions of TLS/FUS that depend on G-quadruplex RNA.

TRF2 is a component of the shelterin complex in the telomere region and contributes to telomere maintenance. TRF2 has G-quadruplex TERRA and double-strand telomere DNA binding abilities, and is important for H3K9me3 depending on TERRA (Figure 1B) [15,16]. These findings suggest that G-quadruplex RNA binding proteins, such as TLS/FUS and TRF2, promote methylation at a specific Lys in telomeric histones through G-quadruplex TERRA.

G-quadruplex TERRA might regulate both methylation and demethylation of histones in the telomere. Lysine-specific histone demethylase (LSD1) is an essential epigenetic regulator that catalyzes the removal of mono- and dimethyl functional groups from histone H3 at the position 4 and 9 Lys [17]. LSD1 is a TERRA-binding epigenetic regulator with SUV39H1 [18]. Binding and kinetics assays of LSD1 with TERRA suggest that G-quadruplex TERRA inhibits LSD1-catalyzed histone demethylation as a noncompetitive inhibitor (Figure 1C) [19].

Polycomb repressive complex 2 (PRC2) binds to TERRA and is responsible for catalyzing the trimethylation of histone H3 at Lys 27 (H3K27me3) (Figure 1D) [20]. This core complex contains enhancer of zeste homolog 2 (EZH2), suppressor of zeste, embryonic ectoderm development, and retinoblastoma binding protein 4. EZH2 contains the SET domain, which catalyzes the transfer of methyl groups to histone H3 at Lys 27 [21]. The complex binds to unstructured G-rich RNA and G-quadruplex RNA. EZH2 mainly contributes to G-quadruplex RNA binding with other subunits, including embryonic ectoderm development [22,23]. The localization of PRC2 at the telomere depends on TERRA, and PRC2-dependent H3K27me3 at the telomere is required for other heterochromatin

modifications, such as H3K9me3 and H4K20me3. These findings indicate that PRC2 might be an essential histone modification enzyme for telomeric heterochromatin. On the other hand, G-quadruplex RNA that specifically removes PRC2 from genes in human and mouse cells inhibits methylation at H3K27 (Figure 1E) [14]. H3K27me modified by PRC2 in the promoter suppresses transcription, but PRC2 transfers from the chromatin to G-quadruplex nascent precursor mRNA upon gene activation, which removes PRC2 from the promoter by G-quadruplex RNA and decreases methylation at H3K27. Thus, PRC2 regulates methylation at H3K27 in specific genes via G-quadruplex RNA.

3. G-quadruplex folding proteins with RRM and RGG domains

Stabilization and unfolding of G-quadruplex structures may be involved in the regulation of transcription, telomere extension, replication, and epigenetics [24–26]. Here we discuss protein motifs with a focus on the RRM and RGG domains regulating G-quadruplex structures and their molecular mechanisms.

The RGG domain is frequently found in G-quadruplex binding proteins [27]. The RGG domain lies in the C-terminal region of TLS/FUS and (Ewing's sarcoma) EWS binds and stabilizes G-quadruplex telomeres [12,28]. These proteins contain an RRM domain and three RGG domains as DNA and RNA binding domains (Figure 2). An electrophoretic mobility shift assay showed that only the C-terminal RGG domain interacts with and stabilizes G-quadruplex telomere DNA and TERRA. Moreover, the RGG domain in G-quadruplex telomere DNA and TERRA recognizes the G-quadruplex loop [29]. In the

case of EWS, the RGG domain binds preferentially to a G-quadruplex with longer loops, which is not related to the topology (parallel, antiparallel, and hybrid) of the G-quadruplex structure. The RGG domain of EWS binds to intramolecular G-quadruplexes despite the absence of bases in the loops, but it does not bind to intermolecular G-quadruplexes without loops. These findings indicate that the RGG domain of EWS recognizes the sugar and the phosphate backbone of the G-quadruplex loops. Similarly, the RGG domain in the C-terminal region of TLS/FUS binds to G-quadruplex telomere DNA, TERRA, and intramolecular G-quadruplexes without bases in the loops, but not to intermolecular G-quadruplexes without loops. Interestingly, the RGG domain of TLS/FUS forms a ternary complex with G-quadruplex telomere DNA and TERRA (Figure 3A) [12]. Binding assays with mutated RGG domains of TLS/FUS revealed that Phe and Tyr in this domain contribute to G-quadruplex DNA and RNA binding, respectively [30,31]. Especially, substituting Tyr for Phe in the RGG domain of TLS/FUS, which recognizes the 2'-OH of the riboses of the G-quadruplex loops, converts its binding specificity to G-quadruplex RNA [30]. Furthermore, nuclear magnetic resonance (NMR) studies of RGG3 with the G-quadruplex indicate that Phe and Tyr in this domain interact with the G-tetrad. RGG3 uses both Phe and Tyr to bind G-quadruplex DNA or RNA in the binary complex of RGG3 with G-quadruplex DNA or RNA, but Tyr binds exclusively to G-quadruplex RNA while Phe binds exclusively to G-quadruplex DNA in ternary complexes [32].

Arg in the RGG domain is important for the G-quadruplex binding of the RGG domain. Simultaneous substitutions of Arg with Lys within the RGG3 of EWS reduces G-quadruplex

binding despite the basic nature of the Lys side chain [28]. Arg can interact with nucleotide bases via hydrogen binding and π -stacking [33]. Arg in the RGG domain might contribute to G-quadruplex binding interactions, thereby affecting G-quadruplex folding and unfolding.

A recent study suggested that the G-quadruplex-specific binding abilities of TLS/FUS require RGG3 with a β -spiral structure stabilized by adjacent proline- and arginine-rich regions, whereas the RGG repeat region alone is not sufficient for G-quadruplex binding (Figure 3B) [34]. The β -spiral formation of the RGG domain is consistent with findings of previous spectroscopic analysis and molecular modeling demonstrating that the RGG repeat forms a helical β -spiral comprising several β -turn structures [35]. Similarly, the G-quadruplex DNA binding activity of RGG3 in EWS is decreased by the truncation of proline- and arginine-rich regions in the RGG domain [29]. Not only Arg, Phe, and Tyr, but also secondary structures of the RGG domain might be important for G-quadruplex-specific binding of the RGG domain. The RGG domain of human fragile X mental retardation protein (FMRP), which has a central role in fragile X syndrome, binds to *in vitro*-selected G-rich RNA [36,37]. Solution and crystal structures of the RGG domain in FMRP with G-quadruplex-containing duplex RNA reveal that the β -turn in the RGG domain recognizes the groove of the duplex-quadruplex junction. These findings indicate that the β -turn of the RGG domain might be a key structure for recognizing the G-quadruplex.

The RGG domain in TLS/FUS and EWS stabilizes the folded G-quadruplex formation [12,28,31]. Each RGG domain changes the hybrid (3 + 1) form of telomere DNA to a parallel form. The hybrid (3 + 1) form of human telomere G-quadruplex DNA in K^+ is converted to a

stable parallel-stranded form in K⁺-containing solutions with 40% (w/v) PEG200 due to water depletion [38–40]. The RGG domain might promote the stability of the G-quadruplex DNA conformation by such a molecular crowding effect.

Recent studies revealed that some peptides containing an RGG domain in cold-inducible RNA-binding protein (CIRBP) is able to form G-quadruplexes [41]. Huang and coworkers identified the RGG domain within CIRBP as the G-quadruplex binding peptide from 12 types of RGG domains included in transcription factors, helicases, and RNA processing factors. This peptide binds on the G-quartet plane of G-quadruplexes. CIRBP containing this RGG domain and RRM might bind both G-quartets and G-quadruplex loops and thus not disturb the DNA conformation (Figures 2, 3C) [41].

RRM is one of the most conserved nucleic acid binding domains; it is present in approximately 0.5%–1% of human genes and folds into an $\alpha\beta$ sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology [42,43]. The folded RRM comprises one four-stranded antiparallel β -sheet $\beta_4\beta_1\beta_3\beta_2$ and two α -helices packed against the β -sheet. RRM-containing proteins are involved in RNA processing, RNA export and stability, and regulation of transcription and translation, and are frequently found in G-quadruplex binding proteins. Nucleolin contains four RRM and RGG domains as nucleic acid binding domains and represses *c-myc* and HIV-1 LTR promoter activity (Figure 2) [44–46]. The C-terminal region of nucleolin, which contains two RRM and RGG domains, is necessary for the repression and induction of G-quadruplex formation of the *c-myc* promoter sequence. The RRM in nucleolin seems to

be essential for binding to nucleic acids, whereas the RGG domain in nucleolin is most important for inducing a stable G-quadruplex of the *c-myc* promoter.

Recently, steroid receptor RNA activator stem-loop interacting RNA binding protein (SLIRP), which contains an RRM domain as a nucleic acid binding domain, was identified as a novel G-quadruplex-interacting protein (Figure 2) [47]. Sequence alignment of the RRM of SLIRP and other G-quadruplex binding proteins, hnRNP A1 and nucleolin, revealed that they have similar amino acid sequences. In the case of SLIRP, SLIRP does not alter the folding of the G-quadruplex structure, which is not consistent with the effect of the RRM in hnRNP A1. The G-quadruplex unfolding ability of the RRM in hnRNP A1 is explained in section 4.

On the other hand, the RGG domain without RRM in nucleolin is not sufficient for stable binding to the G-quadruplex, whereas the RGG domain of TLS/FUS, EWS, FMRP, and CIRBP is sufficient [44]. These findings suggest that the G-quadruplex binding and folding activities of RGG domains in these proteins are different, even though these RGG domains have an RGG repeat and aromatic amino acids.

4. G-quadruplex unfolding proteins with the RRM and RGG domains

G-quadruplex folding proteins such as nucleolin contain RRM and RGG domains, but G-quadruplex unfolding protein, hnRNP A1, also contains these domains. hnRNP A1 contains two RRM domains and one RGG domain, and regulates telomere length (Figure 2) [48,49]. The crystal structure of unwinding protein 1 (UP1), which is an amino-terminal

proteolytic fragment of hnRNP A1 that contains two RRM domains, revealed sequence-specific recognition of the single-stranded DNA of human telomeres d(TTAGGGTTAGGG) [50]. In this complex, direct interaction with d(TAGG) is observed in the N-terminal RRM and with d(TTAGGG) in the C-terminal RRM. UP1 unfolds not only G-quadruplex telomere DNA, but also G-quadruplexes of the minisatellite repeat and human *KRAS* promoter [51,52]. Moreover, the RGG domain in hnRNP A1 enhances the G-quadruplex unfolding activity of UP1 [48]. NMR analysis of hnRNA A1 and G-quadruplex telomere DNA showed that the RGG domain in hnRNA A1 weakly binds to the G-quadruplex depending on the loop nucleotide. Arg residues in the RGG domain in hnRNA A1 are important for the G-quadruplex DNA unfolding effect and promote the unfolding of UP1 (Figure 4).

5. Conclusion

G-quadruplex RNA regulates histone modifications with G-quadruplex RNA binding proteins through various mechanisms: histone modifying enzyme is recruited by TERRA with G-quadruplex RNA binding proteins (TLS/FUS, TRF2), TERRA inhibits G-quadruplex RNA binding proteins and catalyzes histone demethylation (LSD1), histone modifying enzyme with G-quadruplex RNA binding ability is recruited by G-quadruplex RNA (PRC2), and histone modifying enzyme with G-quadruplex RNA binding ability is removed by G-quadruplex RNA (PRC2) (Figure 1). Interestingly, G-quadruplex binding proteins form a complex, but the functions and regulated mechanisms of the complex are unknown [11]. To

clarify the functions of G-quadruplexes with binding proteins *in vivo*, it will be necessary to further investigate the role of G-quadruplex protein complexes.

G-quadruplex binding helicases that disrupt the G-quadruplex are important for DNA replication and epigenetics [7,8]. NM23-H2, which unfolds G-quadruplex DNA in the promoter regions, is necessary for its transcriptionally active form [24]. hnRNP A1 unfolds G-quadruplex telomere DNA and enhances the actions of telomerase [49]. Several G-quadruplex binding proteins involved in the folding or unfolding G-quadruplex have been identified and their functions investigated, but the essential roles of their folding or unfolding G-quadruplexes are essentially unknown. The G-quadruplex folding or unfolding proteins are not confined to proteins containing RRM and RGG domains, but also to those with hnRNP K with KH domains, hnRNP F with qRRM, TRF2 with TRFH, and HMGB1 with HMG boxes as G-quadruplex folding proteins and CNBP with zinc finger and POT1 with OB fold as G-quadruplex unfolding proteins [25,26,53-57]. The G-quadruplex binding mechanism(s) and the biologic functions of other motifs require further investigation to elucidate their essential roles.

The RRM and RGG domains are conserved in both G-quadruplex folding and unfolding proteins. The RRM of hnRNP A1 promotes G-quadruplex unfolding, whereas the RRM of CIRBP, nucleolin, and SLIRP exhibits no G-quadruplex unfolding activities. RRM is a single-strand DNA/RNA binding motif and the RRM of hnRNP A1 recognizes d(TAGG) and d(TTAGGG) in telomere DNA d(TTAGGG)_n. G-quadruplex DNA might be unfolded by these proteins because their RRM domain binds to guanine to form G-tetrads. The

combination of the RRM recognition sequence and G-quadruplex, including loops and 5' and 3' single strands, might be one mechanism involved in G-quadruplex formation.

The RGG domain of TLS/FUS, EWS, and nucleolin promotes G-quadruplex folding, whereas the RGG domain of hnRNP A1 promotes RRM-mediated G-quadruplex unfolding. Each RGG domain of TLS/FUS, EWS, CIRBP, and FMRP has G-quadruplex binding activities, whereas none of the RGG domains of nucleolin and hnRNP A1 are able to interact strongly with G-quadruplexes. Moreover, the RGG domain of TLS/FUS, EWS, and hnRNP A1 mainly interact with the G-quadruplex loops, whereas the RGG domain of CIRBP mainly interacts with the G-tetrad. NMR analysis of the RGG domain in TLS/FUS with the G-quadruplex revealed an interaction with the G-tetrad. These findings indicate that the RGG domain might bind to both the G-tetrad and G-quadruplex loops. G-quadruplex binding activities of the RGG domain require Arg and aromatic amino acids in the region, yet these are not sufficient for binding. More detailed thermodynamic and structural data of the RGG domain and the G-quadruplex are required to elucidate the molecular mechanisms of RGG domain activities.

Acknowledgements

This work was supported by a JGC-S Scholarship Foundation, a Grant-in-Aid for Scientific Research (C) (26410176 to T.O.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

- [1] R. Hansel-Hertsch, M.D. Antonio, S. Balasubramanian, DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* 18 (2017) 279–284.
- [2] S.Y. Yang, P. Lejault, S. Chevrier, R. Boidot, A.G. Robertson, J.M.Y. Wong, D. Monchaud, Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* 9 (2018) 4730.
- [3] H.J. Lipps, D. Rhodes, G-quadruplex structure: *in vivo* evidence and function. *Trends Cell Biol.* 19 (2009) 414–422.
- [4] P. Murat, S. Balasubramanian, Existence and consequences of G-quadruplex structure in DNA. *Curr. Opin. Genet. Dev.* 25 (2014) 22–29.
- [5] D. Rhodes, H.J. Lipps, G-quadruplex and their regulatory roles in biology. *Nucleic Acids Res.* 43 (2015) 8627–8637.
- [6] A. Varizhuk, E. Isaakova, G. Pozmogova, DNA G-quadruplex (G4s) modulate epigenetic (Re) programming and chromatin remodeling. *Bioessays* 41 (2019) e1900091.
- [7] A.K. Mukherjee, S. Sharma, S. Chowdhury, Non-duplex G-quadruplex structures emerge as mediators of epigenetic modifications. *Trends Genet.* 35 (2019) 129–144.
- [8] J. Spiegel, S. Adhikari, S. Balasubramanian, The structure and function of DNA G-quadruplexes. *Trends Genet.* 2 (2020) 123–136.

- [9] O. Mendoza, A. Bourdoncle, J.B. Boule, R.M. Brosh Jr, J.L. Mergny, G-quadruplexes and helicases. *Nucleic Acids Res.* 44 (2016) 1989–2006.
- [10] A. Cammas, S. Millevoi, RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* 45 (2017) 1584–1595.
- [11] M.M. Fay, S.M. Lyons, P. Ivanov, RNA G-quadruplexes in biology: Principles and molecular mechanisms. *J. Mol. Biol.* 429 (2017) 2127–2147.
- [12] K. Takahama, A. Takada, S. Tada, M. Shimizu, K. Sayama, R. Kurokawa, Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.* 20 (2013) 341–350.
- [13] R. Benetti, S. Gonzalo, I. Jaco, G. Schotta, P. Klatt, T. Jenuwein, M.A. Blasco, Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J. Cell Biol.* 178 (2007) 925–936.
- [14] M. Beltran, M. Tavares, N. Justin, G. Khandelwal, J. Ambrose, B.M. Foster, K.B. Worlock, A. Tvardovskiy, S. Kunzelmann, J. Herrero, T. Bartke, S.J. Gamblin, J.R. Wilson, R.G. Jenner, G-tract RNA removes polycomb repressive complex 2 from genes. *Nat. Struct. Mol. Biol.* 26 (2019) 899–909.
- [15] Z. Deng, J. Norseen, A. Wiedmer, H. Riethman, P.M. Lieberman, TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* 35 (2009) 430–413.

- [16] G. Biffi, D. Tannahill, S. Balasubramanian, An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.* 134 (2012) 11974–11976.
- [17] Y. Shi, F. Lan, C. Matson, P. Mulligan, J.R. Whetstine, P.A. Cole, R.A. Casero, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119 (2004) 941–953.
- [18] A. Porro, S. Feuerhahn, J. Lingner, TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres. *Cell Rep.* 6 (2014) 765–776.
- [19] A. Hirschi, W.J. Martin, Z. Luka, L.V. Loukachevitch, N.J. Reiter, G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. *RNA* 22 (2016) 1250–1260.
- [20] J.J. Montero, I. Lopez-Silanes, D. Megias, M.F. Fraga, A. Castells-Garcia, M.A. Blasco, TERRA recruitment of polycomb to telomeres is essential for histone trimethylation marks at telomeric heterochromatin. *Nat. Commun.* 9 (2018) 1548.
- [21] P. Joshi, E.A. Carrington, L. Wang, C.S. Ketel, E.L. Miller, R.S. Jones, J.A. Simon, Dominant alleles identify SET domain residues required for histone methyltransferase of polycomb repressive complex 2. *J. Biol. Chem.* 283 (2008) 27757–27766.
- [22] X. Wang, K.J. Goodrich, A.R. Gooding, H. Naeem, S. Archer, R.D. Paucek, D.T. Youmans, T.R. Cech, C. Davidovich, Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* 65 (2017) 1056–1067.

- [23] Y. Long, B. Bolanos, L. Gong, W. Liu, K.J. Goodrich, X. Yang, S. Chen, A.R. Gooding, K.A. Maegley, K.S. Gajiwala, A. Brooun, T.R. Cech, X. Liu, Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *eLIFE* 6 (2017) e31558.
- [24] T.S. Dexheimer, S.S. Carey, S. Zouhe, V.M. Gokhale, X. Hu, L.B. Murata, E.M. Maes, A. Weichsel, D. Sun, E.J. Meuillet, W.R. Montfort, L.H. Hurley, NM23-H2 may play an indirect role in transcriptional activation of *c-myc* gene expression but does not cleave the nuclease hypersensitive element III₁. *Mol. Cancer Ther.* 8 (2009) 1363–1377.
- [25] A.J. Zaug, E.R. Podell, T.R. Cech, Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl. Acad. U.S.A.* 102 (2005) 10864–10896.
- [26] S. Ray, J.N. Bandaria, M.H. Qureshi, A. Yildiz, H. Balci, G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 2990–2995.
- [27] P. Thandapani, T.R. O'Connor, T.L. Bailey, S. Richard, Defining the RGG/RG motif. *Mol. Cell* 50 (2013) 613–623.
- [28] K. Takahama, K. Kino, S. Arai, R. Kurokawa, T. Oyoshi, Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.* 278 (2011) 988–998.

- [29] K. Takahama, C. Sugimoto, S. Arai, R. Kurokawa, T. Oyoshi, Loop lengths of G-quadruplex structure affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry* 50 (2011) 5369–5378.
- [30] K. Takahama, T. Oyoshi, Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: Tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. *J. Am. Chem. Soc.* 135 (2013) 18016–18019.
- [31] K. Takahama, A. Miyawaki, T. Shitara, K. Mitsuya, M. Morikawa, M. Hariwara, K. Kino, A. Yamamoto, T. Oyoshi, G-quadruplex DNA- and RNA-specific-binding proteins engineered from the RGG domain of TLS/FUS. *ACS Chem. Biol.* 10 (2015) 2564–2569.
- [32] K. Kondo, T. Mashima, T. Oyoshi, R. Yagi, R. Kurokawa, N. Kobayashi, T. Nagata, M. Katahira, Plastic roles of phenylalanine and tyrosine residues of TLS/FUS in complex formation with the G-quadruplexes of telomeric DNA and TERRA. *Sci. Rep.* 8 (2018) 2864.
- [33] P.A. Chong, R.M. Vernon, J.D. Forman-Kay, RGG/RG motif regions in RNA binding and phase separation. *J. Mol. Biol.* 430 (2018) 4650–4665.
- [34] R. Yagi, T. Miyazaki, T. Oyoshi, G-quadruplex binding ability of TLS/FUS depends on the β -spiral structure of the RGG domain. *Nucleic Acids Res.* 46 (2018) 5894–5901.
- [35] C.C. Burd, G. Dreyfuss, Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265 (1994) 615–621.

- [36] A.T., Phan, V. Kuryavyi, J.C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, T., A. Plonskai, C. Chen, D. Clain, R.B. Darnell, D.J. Patel, Structure–function studies of FMRP RGG peptide recognition of an RNA duplex–quadruplex junction. *Nat. Struct. Mol. Biol.* 18 (2011) 796–804.
- [37] N. Vasilyev, A. Polonskaia, J.C. Darnell, R.B. Darnell, D.J. Patel, A. Serganov, Crystal structure reveals specific recognition of a G–quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) E5391–E5400.
- [38] B. Heddi, A.T. Phan, Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.* 133 (2011) 9824–9833.
- [39] D. Miyoshi, H. Karimata, N. Sugimoto, Hydration regulates thermodynamic of G–quadruplex formation under molecular crowding conditions. *J. Am. Chem. Soc.* 133 (2006) 7957–7963.
- [40] Y. Xue, Z.Y. Kan, Q. Wang, Y. Yao, J. Liu, Y.H. Hao, Z. Tan, Human telomeric DNA forms parallel–stranded intramolecular G–quadruplex in K^+ solution under molecular crowding condition. *J. Am. Chem. Soc.* 129 (2007) 11185–11191.
- [41] Z.–L. Huang, J. Dai, W.–H. Luo, X.–G. Wang, J.–H. Tan, S.–B. Chen, Z.–S. Huang, Identification of G–quadruplex–binding protein from the exploration of RGG Motif/G–quadruplex interactions. *J. Am. Chem. Soc.* 140 (2018) 17945–17955.
- [42] A. Clery, M. Blatter, F.H. Allain, RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 18 (2008) 290–298.

- [43] C. Maris, C. Dominguez, F. H. Allain, The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 272 (2005) 2118–2131.
- [44] V. Gonzalez, L.H. Hurley, The C-terminus of nucleolin promotes the formation of the *c-MYC* G-quadruplex and inhibits *c-MYC* promoter activity. *Biochemistry* 49 (2010) 9706–9714.
- [45] V. Gonzalez, K. Guo, L. Hurley, D. Sun, Identification and characterization of nucleolin as a *c-myc* G-quadruplex-binding protein. *J. Biol. Chem.* 284 (2009) 23622–23635.
- [46] E. Tosoni, I. Frasson, M. Scalabrin, R. Perrone, E. Butovskaya, M. Naodai, G. Palu, D. Fabris, S.N. Richter, Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* 43 (2015) 8884–8897.
- [47] P. Williams, L. Li, X. Dong, Y. Wang, Identification of SLIRP as a G-quadruplex-binding protein. *J. Am. Chem. Soc.* 139 (2017) 12426–12429.
- [48] M. Ghosh, M. Singh, RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res.* 46 (2018) 10246–10261.
- [49] H. LaBranche, S. Dupuis, Y. Ben-David, M.R. Bani, R.J. Wellinger, B. Chabot, Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet.* 19 (1998) 199–202.

- [50] J. Ding, M.K. Hayashi, Y. Zhang, L. Manche, A.R. Krainer, R.-M. Xu, Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* 13 (1999) 1102–1115.
- [51] M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama, L.E. Xodo, Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human *KRAS* promoter: implications for transcription. *Nucleic Acids Res.* 37 (2009) 2841–2853.
- [52] H. Fukuda, M. Katahira, N. Tsuchiya, Y. Enokizono, T. Sugimura, M. Nagao, H. Nakagama, Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc. Natl. Acad. U.S.A.* 99 (2002) 12685–12690.
- [53] J. Qiu, J. Liu, S. Chen, T.-M. Ou, J.-H. Tan, L.-Q. Gu, Z.-S. Huang, D. Liu, Role of hairpin-quadruplex DNA secondary structural conversion in the promoter of hnRNP K in gene transcription. *Org. Lett.* 17 (2015) 4584–4587.
- [54] B. Samatanga, C. Dominguez, I. Jelesarov, F.H.-T. Allain, The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Res.* 41 (2013) 2505–2516.
- [55] I.M. Pedroso, W. Hayward, T.M. Fletcher, The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures. *Nucleic Acids Res.* 37 (2009) 1541–1554.
- [56] J. Amato, T.W. Madanayake, N. Iaccarino, E. Novellino, A. Randazzo, L.H. Hurley, B. Pagano, HMGB1 binds to the *KRAS* promoter G-quadruplex: a new player in oncogene transcriptional regulation? *Chem. Commun.* 54 (2018) 9442–9445.

[57] A.P. David, A. Pipier, F. Pascutti, A. Binolfi, A.M.J. Weiner, E. Challier, S. Heckel, P. Calsou, D. Gomez, N.B. Calcaterra, P. Armas, CNBP controls transcription by unfolding DNA G-quadruplex structures. *Nucleic Acids Res.* 47 (2019) 7901–7913.

Figure Legends

Figure 1

Models of the roles of TLS/FUS, TRF2, LSD1, and PRC2 in G-quadruplex RNAs for histone modifications. (A) Recruitment of Suv4-20h2 and TERRA with Histone H3-modifying enzyme to telomeres by TLS/FUS results in increased trimethylation of histone H4 at Lys 20 and histone H3 at Lys 9, respectively. (B) Recruitment of TERRA with histone H3-modifying enzyme to telomeres by TRF2 results in increased H3K9-trimethylation. (C) TERRA inhibits LSD1-catalyzed histone demethylation of histone H3 at Lys 4/9 as a noncompetitive inhibitor. (D) PRC2 binds to TERRA and catalyzes trimethylation of histone H3 at Lys 27 in telomeres. (E) G-quadruplex RNA within nascent precursor mRNA removes PRC2 from genes, thus depleting trimethylation of histone H3 at Lys 27.

Figure 2

Architecture of TLS/FUS, EWS, CIRBP, nucleolin, SLIRP, hnRNP A1, and UP1 as G-quadruplex folding or unfolding proteins, including the RRM domain with or without an RGG domain.

Figure 3

Models of G-quadruplex binding proteins with G-quadruplex. (A) The ternary complex of RGG3 in TLS/FUS with telomere DNA and TERRA. (B) β -turn-dependent G-quadruplex binding of RGG3 in TLS/FUS. Orange: Pro-rich region, Green: RGG repeat, Blue: Arg-rich region. (C) RRM and RGG in CIRBP binding to telomere DNA.

Figure 4

Model of RGG domain-assisted recognition and unfolding of telomere G-quadruplex DNA unfolding protein hnRNP A1.

Figure 2

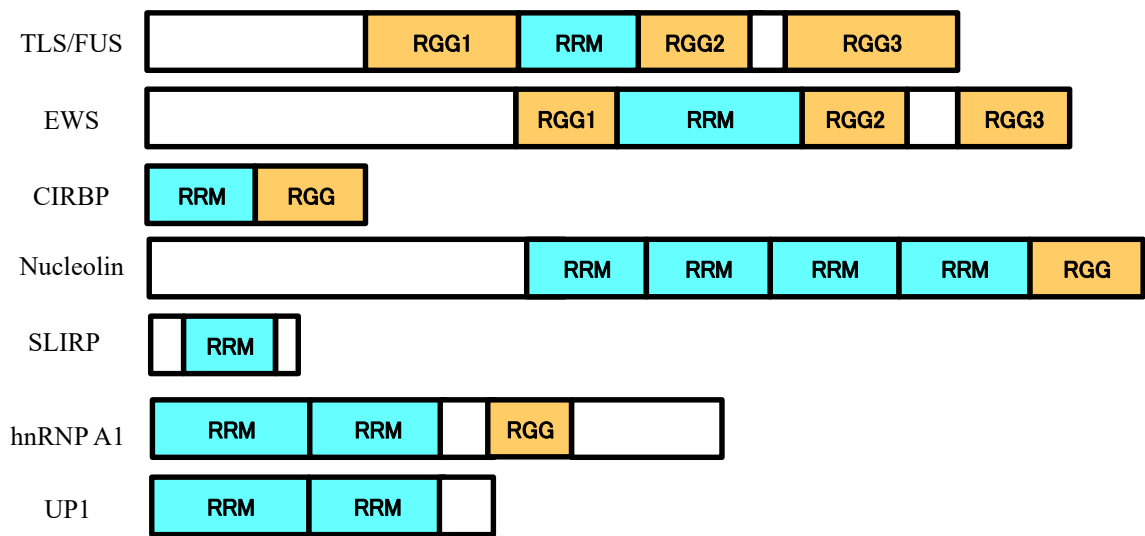


Figure 3

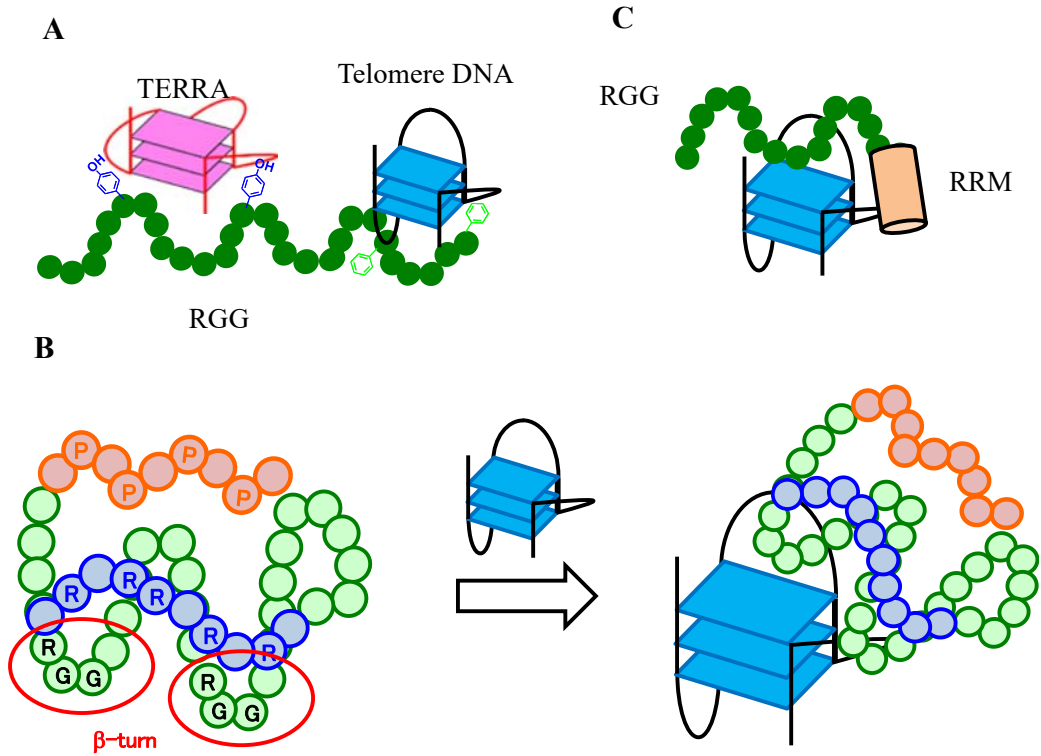


Figure 4

