ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization

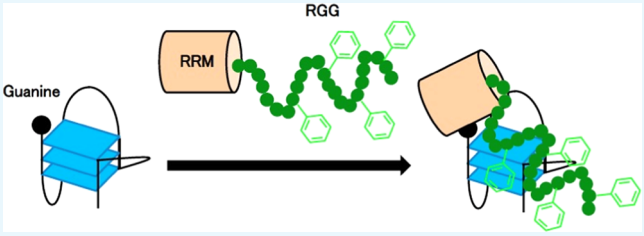
Tatsuki Masuzawa and Takanori Oyoshi*

Cite This: *ACS Omega* 2020, 5, 5202−5208

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: G-quadruplexes have important biologic functions that are regulated by G-quadruplex-binding proteins. In particular, G-quadruplex structures are folded or unfolded by their binding proteins and affect transcription and other biologic functions. Here, we investigated the effect of the RNA recognition motif (RRM) and arginine−glycine−glycine repeat (RGG) domain of nucleolin on G-quadruplex formation. Our findings indicate that Phe in the RGG domain of nucleolin is responsible for G-quadruplex binding and folding. Moreover, the RRM of nucleolin potentially binds to a guanine-rich single strand and folds the G-quadruplex with a 5′-terminal and 3′-terminal single strand containing guanine. Our findings contribute to our understanding of how the RRM and RGG domains contribute to G-quadruplex folding and unfolding.



INTRODUCTION

G-quadruplexes in DNA and RNA have important biologic roles, such as regulation of transcription, translation, DNA replication, telomere elongation, and histone modification.[1−3] Each function of G-quadruplexes is thought to be regulated by G-quadruplex-binding proteins.[4] Some of these binding proteins, including heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), nucleolin, cold-inducible mRNA-binding protein (CIRBP), translocated in liposarcoma (also known as fused in sarcoma, TLS/FUS), and Ewing's sarcoma (EWS), have common nucleic-acid-binding domains, such as the RNA recognition motif (RRM) and arginine−glycine−glycine repeat (RGG) domain.[5−11] Previous bioinformatic analysis revealed that the RGG domain is an evolutionarily conserved sequence and at least 31 different protein isoforms contain the domain.[12,13] The RRM is one of the most highly conserved nucleic-acid-binding domains, presenting in approximately 0.5−1% of human genes and comprising one four-stranded antiparallel β-sheet and two α-helixes packed against the β-sheet.[14,15] The RGG domain with the RRM of hnRNP A1 and nucleolin is a G-quadruplex-binding domain.[5−7] hnRNP A1 and its derivative UP1 are involved in telomere maintenance and transcription, and nucleolin regulates transcription, translation of G-quadruplex-containing regions, and suppression of virus replication.[16−24] Furthermore, nucleolin interacts with the G-quadruplex-formed C9orf72 hexanucleotide repeat expansion, which causes the neurodegenerative disease amyotrophic lateral sclerosis.[25] On the other hand, only the RGG domains of CIRBP, TLS/FUS, and EWS have been shown to bind the G-quadruplex.[8−10] In particular, the formation of the ternary complex between TLS/FUS and the G-quadruplex of telomeric DNA and telomeric repeat-containing RNA (TERRA) leads to telomere shortening and trimethylation of histone H3 at lysine 9 and histone H4 at lysine 20 at the heterochromatin of telomeres.[9]

The conformational changes of G-quadruplex structures regulated by G-quadruplex-binding proteins are important for gene expression and replication. DNA and RNA helicases unwind G-quadruplex structures and enhance transcription, translation, or replication.[1,2,26,27] NM23-H2, which unfolds G-quadruplex DNA in the promoter regions, is necessary for its transcriptionally active form.[28] The RRM and RGG domains in the C-terminal region of nucleolin are necessary for inducing G-quadruplex formation of the *c-myc* promoter sequence.[6] The RGG domain of nucleolin is especially important for inducing a stable G-quadruplex. Moreover, the RGG domain of the C-terminal domain in TLS/FUS and EWS is important for folding G-quadruplex telomere DNA.[9,10] On the other hand, the RRM of hnRNP A1 unfolds G-quadruplex telomere DNA, and the RGG domain of hnRNP A1 enhances the G-quadruplex unfolding activity of RRM.[5] Therefore, G-quadruplex-binding proteins containing the RRM and RGG domains that are involved in folding or unfolding G-quadruplex structures have been identified and their functions investigated. The reasons for the differences in the effects of these domains on G-quadruplex formation, however, are unknown. Here, we investigated the effect of the RRM and RGG domains in nucleolin on G-quadruplex formation. Our

Received: December 10, 2019
Accepted: February 20, 2020
Published: March 5, 2020

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202−5208

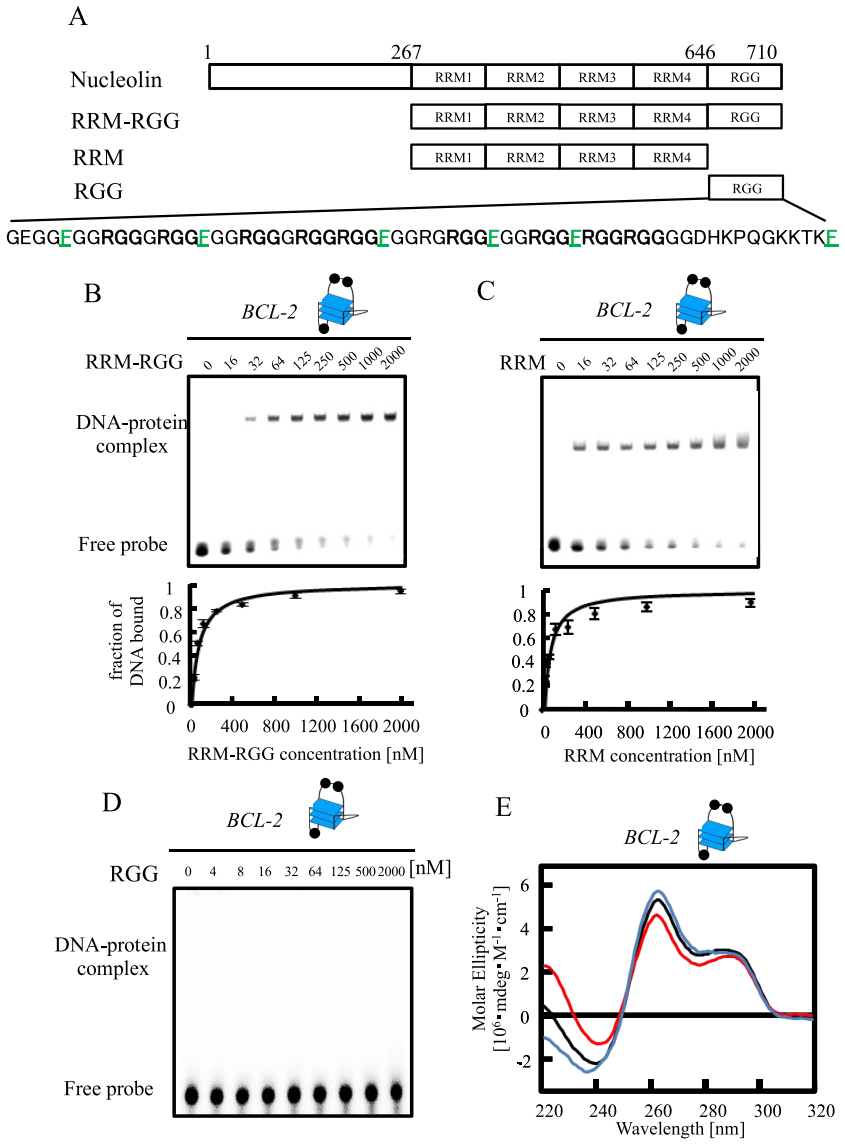


Figure 1. Effects of the RRM and RGG domains in nucleolin on *BCL*-2 G-quadruplex folding. (A) Schematic illustration of nucleolin and its truncated mutants. (B–D) The equilibrium binding curve was obtained by calculating the fraction of *BCL*-2 at varying RRM–RGG domain (B), RRM (C), or RGG domain (D) concentrations. The dissociation constant (K_d) was determined by fitting the data to the appropriate equation. The DNA–protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography. (E) Circular dichroism spectra of *BCL*-2 with RRM or the RRM–RGG domain. Line colors: black, *BCL*-2; red, *BCL*-2 and RRM; and blue, *BCL*-2 and RRM–RGG domain. The concentrations of DNA and protein were both 2.5 μM. The G-quadruplex structure is indicated in (B–E), respectively. Black circles in the cartoons of the G-quadruplex represent a guanine residue.

findings indicate that the Phe of the RGG domain in nucleolin is responsible for G-quadruplex binding and folding. Moreover, the RRM of nucleolin folds G-quadruplex structures with guanine residues in the G-quadruplex terminal single strands and loops. Our findings suggest potential mechanisms underlying the effects of the RRM and RGG domains of nucleolin for G-quadruplex folding and unfolding.

MATERIALS AND METHODS

Plasmid Constructs. The nucleolin plasmid was used as a template for polymerase chain reaction. The RRM-RGG (267–709) and RRM (267–645) cDNAs of nucleolin were cloned into the pGEX6P-1 vector (GE Healthcare, Chicago, IL) between the EcoRI and XhoI sites using the following sets of primers to express an *N*-terminal glutathione S-transferase (GST) fusion protein: for RRM-RGG, forward 1 d(CGG AAT TCT TCA ATC TCT TTG TTG GAA ACC) and reverse 1 d(CGC TCG AGC TAT TCA AAC TTC GTC TTC); for RRM, forward 1 and reverse 2 d(CGC TCG AGT TAT CTT TGA GAA TCT TCT CTG GAG AC). RRM-RGGF/A1, RRM-RGGF/A2, RRM-RGGF/A3, and RRM-RGGF/A4 were obtained by replacing Phe with Ala in RRM-RGG using a KOD-Plus- Mutagenesis Kit (Toyobo, Japan) with the RRM-RGG in the pGEX6P-1 vector used as the template and the following primers: for RRM-RGGF/A1, F/A forward 1 d(CGG CGC TGG AGG ACG AGG TGG TGG T) and F/A reverse 1 d(CCT CTG CCT CCA CCA CGA CCC CCG A); for RRM-RGGF/A2, F/A forward 1, F/A reverse 1, F/A forward 2 d(AGG AGC TGG TGG CAG AGG CCG GGG A), and F/A reverse 2 d(CCT CGG CCT CCT CTA CCA CCA CCT CGT C); for RRM-RGGF/A3, F/A forward 1, F/A reverse 1, F/A forward 2, F/A reverse 2, and F/A forward 3

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202–5208

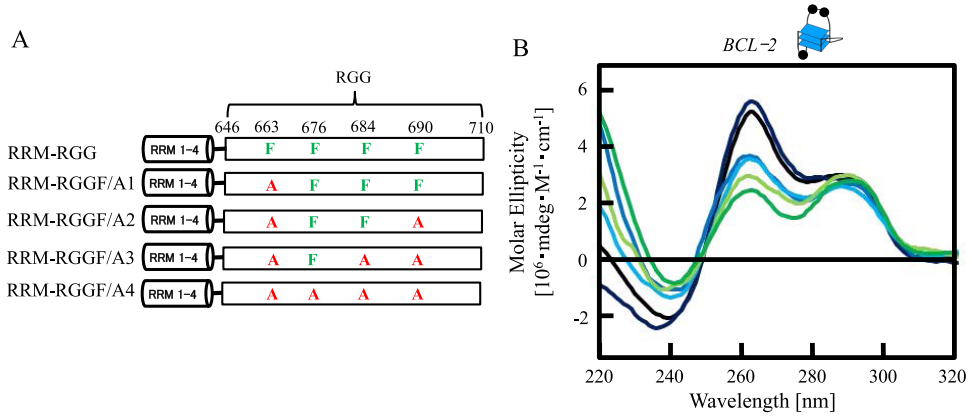


Figure 2. Effects of Phe in the RGG domain of nucleolin on *BCL*-2 G-quadruplex folding. (A) Schematic illustration of RRM−RGG and mutated RRM−RGG domains (RRM-RGGF/A1, RRM-RGGF/A2, RRM-RGGF/A3, and RRM-RGGF/A4). (B) Circular dichroism spectra of *BCL*-2 with RRM-RGG or mutated RRM-RGG domains. Line colors: black, *BCL*-2; deep blue, *BCL*-2 and RRM-RGG; blue, *BCL*-2 and RRM-RGGF/A1; light blue, *BCL*-2 and RRM-RGGF/A2; yellow green, *BCL*-2 and RRM-RGGF/A3; and green, *BCL*-2 and RRM-RGGF/A4. The concentrations of DNA and protein were both 2.5 μM. The G-quadruplex structure is indicated in (B). Black circles in the cartoon of the G-quadruplex represent a guanine residue.

d(AGG CGC TGG AGG GCG AGG AGG CGC C) and F/A reverse 3 d(CCC CGG CCT CTG CCA CCA AAT CCT CCT C); for RRM-RGGF/A4, F/A forward 1, F/A reverse 1, F/A forward 2, F/A reverse 2, and F/A forward 3, F/A reverse 3 F/A, forward 4 d(AGG CGC CCG AGG AGG CAG AGG AGG A), and F/A reverse 4 d(CCT CGC CCT CCA AAG CCT CCC CGG C). All constructs were verified by automated DNA sequencing. All DNA oligomers were obtained from Operon Biotechnologies (Japan).

Expression and Purification of Glutathione S-Transferase Fusion Proteins. All recombinant proteins were fused at the N-terminus to GST and overexpressed in *Escherichia coli*. The *E. coli* strain BL21 (DE3) pLysS-competent cells were transformed with the vectors, and transformants were grown at 37 °C in Luria Bertani medium containing ampicillin (0.1 mg/mL). Protein expression was induced at A_{600} = 0.6 with 0.1 mM isopropyl β-D-1-thiogalactopyranoside. The cells were then grown for an additional 16 h at 25 °C and harvested by centrifugation (6400*g* for 20 min). The *E. coli* pellets were resuspended in the following buffers: WK buffer (20 mM potassium phosphate [pH 7.0], 150 mM KCl, 1 mM dithiothreitol [DTT], 1 mM ethylenediaminetetraacetic acid (EDTA), and 0.1 mM phenylmethanesulfonyl fluoride) or WLi buffer (20 mM Tris-HCl [pH 7.5], LiCl 20 mM, 1 mM DTT, 1 mM EDTA, and 0.1 mM phenylmethanesulfonyl fluoride). The supernatants containing the expressed proteins were lysed by sonication (model UR-20P, Tomy Seiko, Japan) and centrifuged at 16 200*g* for 15 min at 4 °C. The supernatant and glutathione agarose (MilliporeSigma, Burlington, MA) were incubated with gentle mixing for 1 h at 4 °C; the resin was washed with WKT buffer (20 mM potassium phosphate [pH 7.0], 150 mM KCl, 1 mM DTT, 1 mM EDTA acid, and 1 v/v % Triton X-100) or WLiT buffer (20 mM Tris-HCl [pH 7.5], 20 mM LiCl, 1 mM DTT, 1 mM EDTA, and 1 v/v % Triton X-100) at 4 °C. GST-tags were cleaved using buffer containing 8 units/mL PreScission protease (GE Healthcare) on a resin for 16 h at 4 °C, and the protein was eluted with K buffer (20 mM potassium phosphate [pH 7.0]) or Li buffer (20 mM Tris-HCl [pH 7.5], 20 mM LiCl). The protein concentrations were determined using a BCA Protein Assay Kit (Thermo Scientific, Altham, MA). All proteins were stored at 4 °C and used within 12 h of purification.

Electrophoretic Mobility Shift Assay. ^{32}P-Labeled oligonucleotide annealing and G-quadruplex formation were induced by heating samples to 95 °C on a thermal heating block and cooling to 4 °C at a rate of 2 °C/min in K buffer or Li buffer. Binding reactions were performed in a final volume of 20 μL using 1 nM labeled oligonucleotide with various concentrations of each purified protein with 0.1 mg/mL bovine serum albumin in K buffer or Li buffer. After incubating the samples for 30 min at 4 °C, they were loaded on a 6% polyacrylamide (acrylamide/bisacrylamide = 19:1) nondenaturing gel. Both the gel and the electrophoresis buffer contained 0.5x TBE buffer (45 mM Tris base, 45 mM boric acid, and 0.5 mM EDTA) with or without 20 mM KCl. Electrophoresis was performed at 10 V/cm for 100 min at 4 °C. The gels were exposed in a phosphorimager cassette and imaged by a Personal Molecular Imager FX (Bio-Rad Laboratories, Hercules, CA). To determine the equilibrium dissociation constants (K_d), the data from four replicate experiments were plotted as ϕ (1 fraction of free DNA) versus the protein concentration, which is equal to the protein at which half of the free DNA is bound. The K_d was extracted by nonlinear regression using Microsoft Excel 2011 and the following equation: $\phi = [P]/\{K_d + [P]\}$.

Circular Dichroism Spectroscopy. Circular dichroism (CD) spectra were recorded on a model J-820 instrument (Jasco). The CD spectra of DNA (2.5 μM strand concentration) with each protein (2.5 μM) in 20 mM potassium phosphate (pH 7.0) were recorded using a 0.2 cm pathlength cell at 20 °C. For each spectrum, the spectrum of the corresponding buffer was subtracted, and these data were not further processed (e.g., by smoothing).

RESULTS AND DISCUSSION

RRM−RGG Domain in Nucleolin Slightly Folded the *BCL*-2 G-Quadruplex Structure and RRM Did Not Fold It. Nucleolin has four RRM−RGG domains as a nucleic-acid-binding region and binds to G-quadruplex DNA (Figure 1A).[6,7] To investigate the G-quadruplex-binding abilities of the RRM and RGG domains in nucleolin, we compared the

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202−5208

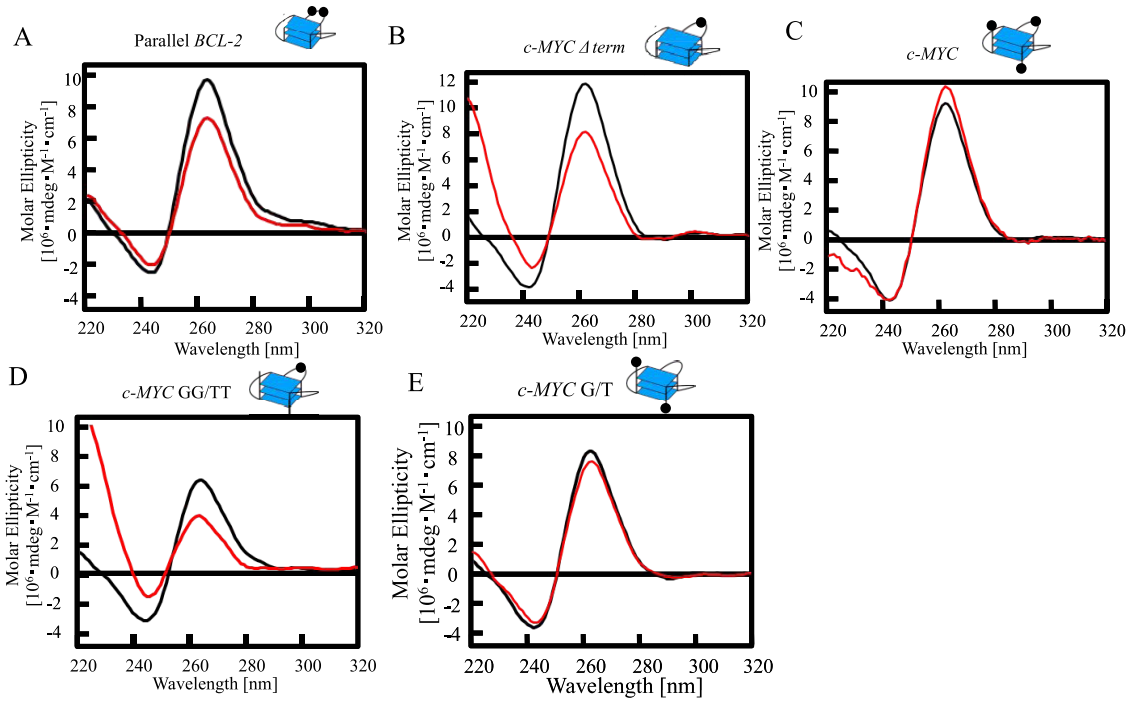


Figure 3. Properties of RRM in nucleolin for binding to single-stranded DNA and folding several parallel G-quadruplex DNAs. Circular dichroism spectra of Parallel *BCL*-2 (black) or *c*- Parallel *BCL*-2 with RRM (red) (A), *c-MYC* Δterm (black) or *c-MYC* Δterm with RRM (red) (B), *c-MYC* (black) or *c-MYC* with RRM (red) (C), *c-MYC* GG/TT (black) or *c-MYC* GG/TT with RRM (red) (D), and *c-MYC* G/TT (black) or *c-MYC* G/T with RRM (E). The concentrations of DNA and protein were both 2.5 μM. The G-quadruplex structure is indicated. Black circles in the cartoons of each G-quadruplex represent a guanine residue.

binding of the RRM−RGG domain and RRM alone to the modified promoter sequence of the *bcl*-2 gene (*BCL*-2) (Figure 1B−D). As *BCL*-2 forms an intramolecular hybrid (3 + 1) G-quadruplex and nucleolin activates *bcl*-2 expression in living cells, we used *BCL*-2 as G-quadruplex DNA in Figures 1 and 2 to investigate G-quadruplex binding and folding.[6,21,29] All purified proteins discussed in this article were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Figure S1). The electrophoretic mobility shift assay (EMSA) data of the RRM−RGG domain or RRM with various concentrations of *BCL*-2 were fitted to a hyperbolic equation to give a K_d of 79.6 ± 4.0 and 73.6 ± 7.1 nM, respectively (Figure 1B,C). EMSA of the RGG domain alone with *BCL*-2 did not show obvious binding (K_d > 2000 nM, Figure 1D). These results suggest that the binding activity of nucleolin to *BCL*-2 depends mainly on the RRM.

To investigate the effect of nucleolin on the G-quadruplex structure, we performed CD spectroscopy studies of *BCL*-2 with 1 equivalent of the RRM−RGG domain or RRM in nucleolin (Figure 1E). The CD of *BCL*-2 with the RRM−RGG domain or RRM showed a positive peak at 262 nm, consistent with the result of a previous CD spectroscopy study of *BCL*-2,[8] which was slightly increased in the presence of the RRM−RGG domain and slightly decreased in the presence of the RRM. These findings indicate that while the RGG domain has an important role in the induction of the *BCL*-2 G-quadruplex structure, the RGG domain does not show strong binding to the G-quadruplex structure and RRM did not induce G-quadruplex formation of *BCL*-2. These data are not consistent with previously reported findings that both the RRM−RGG domain and RRM in nucleolin induce the G-quadruplex structure of the promoter sequence of the *c-myc* gene (*c-MYC*).[4] Taken together, these findings suggest that the effects of the RRM in nucleolin for G-quadruplex folding might differ depending on the G-quadruplex structure and its DNA sequence.

Phe in the RGG Domain of Nucleolin Contributes to G-Quadruplex Binding and Folding. The RGG domain is a G-quadruplex-binding domain in several proteins.[5−11] NMR-based-binding assays revealed that each Phe and Tyr of the C-terminal of the RGG domain in TLS/FUS plays a central role in binding to G-quadruplex telomeric DNA and TERRA.[30] The RGG domain in nucleolin contains a Phe adjacent to four Arg-Gly-Gly sequences (Figure 1A). To evaluate the role of the adjacent Phe in the RGG domain for G-quadruplex binding, RRM−RGG domains with the Phe substituted with an Ala were designed and expressed. The substitutions of Phe with Ala at different positions and the resulting mutated proteins (RGGF/A1, RRM-RGGF/A2, RRM-RGGF/A3, and RRM-RGGF/A4) are shown in Figure 2A. The EMSA of RRM-RGGF/A1, RRM-RGGF/A2, RRM-RGGF/A3, and RRM-RGGF/A4 with various concentrations of *BCL*-2 fitted to a hyperbolic equation gave K_ds of 75.5 ± 9.5, 71.8 ± 6.9, 171.8 ± 9.1, and >500 nM, respectively (Figure S2A−D). These data indicate that the G-quadruplex-binding affinities of RRM-RGGF/A1 and RRM-RGGF/A2 were essentially the same as that of RRM−RGG, but the binding affinities of RRM-RGGF/A3 and RRM-RGGF/A4, in which three or four of the Phe were replaced with Ala, were decreased. Thus, the increase of the number of Phe-to-Ala substitutions in the RGG domain decreased G-quadruplex-binding affinities except for the Phe to Ala point mutation of residue 690.

To estimate the role of Phe in the RGG domain for G-quadruplex folding, we performed CD spectroscopy studies of *BCL*-2 to examine the effect of substituting Phe with Ala within the RGG domain of the RRM−RGG domain (Figure 2A,B).

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202−5208

The CD spectrum of *BCL*-2 with RRM-RGGF/A1 shows a decreased positive peak at 262 nm compared with the RRM−RGG domain. As the number of Phe-to-Ala substitutions in the RGG domain increased, the positive peak at 262 nm of *BCL*-2 decreased, except for the Phe to Ala point mutation of residue 690. These findings indicate that replacing Phe with Ala at amino acids 663, 676, and 684 affected G-quadruplex folding but the Phe to Ala substitution at amino acid 690 did not. Comparing the CD spectrum of *BCL*-2 with RRM and RRM-RGGF/A3, RRM-RGGF/A4 strongly decreased the positive peak at 262 nm (Figures 1E and 2). This finding suggests that the RGG domain in which Phe was substituted with Ala enhances the G-quadruplex unfolding activity of RRM.

RRM-Mediated Unfolding Does Not Depend on the Differences between Parallel and Hybrid G-Quadruplexes. The negative effect of RRM in nucleolin for G-quadruplex folding of *BCL*-2 is not consistent with a previous report that the RRM in nucleolin induces the G-quadruplex structure of the promoter sequence of *c-MYC*.[4] These findings suggest that the effects of RRM in nucleolin on G-quadruplex folding might differ depending on the G-quadruplex structure or its DNA sequence. Previous NMR analyses revealed that *BCL*-2 forms a (3 + 1) hybrid type G-quadruplex and *c-MYC* forms a parallel-type G-quadruplex.[29,31] To investigate the effects of the RRM in nucleolin for folding different G-quadruplex structures, we performed CD spectroscopy studies of parallel-type G-quadruplexes based on *BCL*-2 with RRM (Figure 3A). The number of bases in the loops within each G-quadruplex is an important factor in determining the topology of G-quadruplex structures.[11,32] The (3 + 1) hybrid G-quadruplex of *BCL*-2 contains one, three, and seven bases in each loop (Figure 1D). CD spectroscopy of mutated *BCL*-2, in which three bases in the loop were changed to one base (parallel *BCL*-2), showed a shift in the spectrum with a decrease in the strong positive band at 265 nm (Figure 3A). This finding is characteristic of the parallel form and consistent with the results of a previous CD study.[8] These results suggest that the G-quadruplex DNA-folding activities of RRM in nucleolin do not depend on the G-quadruplex topology.

RRM Folded G-Quadruplex with 5′-Terminal and 3′-Terminal Single Strands Containing Guanine. We then investigated the effects of different DNA sequences, not topology, for RRM-mediated folding or unfolding of G-quadruplex structures. A comparison of the DNA sequences of each previous NMR structure of *BCL*-2 and *c-MYC* revealed that *c-MYC* has single strands containing three bases at both the 5′- and 3′-terminals of the G-quadruplex whereas *BCL*-2 does not (Table 1).[29,31] To investigate the effects of the 5′- and 3′-terminal single strands of G-quadruplex on RRM-mediated folding, we performed CD spectroscopy studies of mutated G-quadruplexes based on *c-MYC* in the presence of RRM (Figure 3B−E). The CD spectrum of mutated *c-MYC* without 5′- and 3′-terminal single strands (*c-MYC* Δterm) with RRM showed a decrease in the positive peak at 265 nm, even though the *c-MYC* with RRM showed a slight increase in the positive peak (Figure 3B,C). These results suggest that RRM binds to 5′- and 3′-terminal single strands and folds the G-quadruplex structure.

To characterize the DNA-binding base of RRM in nucleolin, we analyzed the binding ability of RRM to homo-oligo-DNAs such as dA_{10}, dT_{10}, dG_{10}, and dC_{10} by EMSA (Figure S3). All homo-oligo-DNAs were folded in buffer containing Li^+ instead of K^+ to inhibit the formation of the G-quadruplex of dG_{10}. The EMSA showed that RRM binds well to dG_{10}. The DNA sequence of each 5′- and 3′-terminal single strand in *c-MYC* contains one guanine among the three bases. We hypothesize that guanine in the 5′- and 3′-terminals of G-quadruplex needs the RRM to fold the G-quadruplex. To investigate the effect of the guanine in the 5′- and 3′-terminals of the G-quadruplex for RRM-mediated folding and unfolding of the G-quadruplex, we performed CD spectroscopy studies of *c-MYC* mutated 5′- and 3′-terminals (*c-MYC* GG/TT) with RRM (Figure 3D). *c-MYC* GG/TT comprises *c-MYC* with thymine substituted for a guanine in the 5′- and 3′-terminals. The CD spectrum of *c-MYC* GG/TT with RRM showed a decrease in the positive peak at 265 nm. The CD spectrum of *c-MYC* with thymine substituted for a guanine in the loop (*c-MYC* G/T) with RRM was not changed compared to that of the *c-MYC* without RRM (Figure 3E). These results suggest that RRM mainly binds to guanine in the 5′- and 3′- terminal single strands and folds the G-quadruplex structure.

Table 1. Sequence and T_m of Oligonucleotides Used in EMSA and CD[a]

oligo-DNA	5′ sequence 3′
BCL-2	GGGCGCGGGAGGAATTGGGCGGG
parallel *BCL*-2	GGGCGGGAGGAATTGGGCGGG
c-MYC	TGA GGGTGGGGAGGGTGGGGAA
c-MYC Δterm	GGGTGGGGAGGGTGGG
c-MYC GG/TT	TTAGGGTGGGGAGGGTGGGTAA
c-MYC G/T	TGAGGGTGGGTAGGGTGGGGAA

[a]Underlined guanines form a G-tetrad.

CONCLUSIONS

This article reports the amino acids in the RGG domain that are important for G-quadruplex folding and the role of RRM in nucleolin for G-quadruplex folding. Our findings suggest that the RRM domain of nucleolin preferentially binds to the 5′-and 3′-terminal single strands containing guanine in the G-quadruplex and Phe in the RGG domain contribute to the G-quadruplex folding. The RRM of nucleolin folds the G-quadruplex with guanine-containing single strands, but the RRM unfolds G-quadruplexes without guanines in the single strands of the 5′- and 3′-terminals. The RRMs of hnRNP A1 and hnRNP D are reported to bind and unfold G-quadruplexes.[5,33] The X-ray structure of two RRMs in hnRNP A1 with single-stranded telomere DNA revealed direct interactions with d(TAGG) and d(TTAGG) in RRM1 and RRM2, respectively.[34] The recognition of d(TAG) in d(TTAGGG) by the RRM of hnRNP D was determined by NMR.[33] These data suggest that RRMs of hnRNP A1 and hnRNP D bind to guanine forming a G-tetrad and induces unfolding of the G-quadruplex. This article shows that the RRM in nucleolin bound preferentially to guanine and unfolded the G-quadruplex when it did not contain single-stranded DNA (*BCL*-2) or if the single-stranded DNA contained thymine in place of guanine (*c-MYC* GG/TT). The RRM of nucleolin might unfold G-quadruplexes with a mechanism similar to that of the RRMs in hnRNP A1 and hnRNP D.

Figure S2 shows that the Phe of the RGG domain in nucleolin is important for G-quadruplex binding, even if the RGG domain alone does not show obvious binding to the G-quadruplex. Mutated RGG domains in which some Phe were

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202−5208

substituted with Ala (RRM-RGGF/A3 and RRM-RGGF/A4) decreased the ability of RRM to bind to the G-quadruplex. The RGG domains in the C-terminals of EWS and TLS/FUS bind G-quadruplexes and the substitutions of Tyr or Phe by Ala in the RGG domains decrease G-quadruplex binding.[10,35] These findings suggest that the aromatic amino acids in the RGG domain of TLS/FUS and EWS are important for G-quadruplex binding. The RGG domains of TLS/FUS and EWS are thought to bind loops in G-quadruplexes. Furthermore, NMR studies of the RGG domain in TLS/FUS with G-quadruplex telomere DNA indicate that Phe in this domain interacts with the G-tetrad.[30] The RGG domain in CIRBP binds to the G-quartet plane of a G-quadruplex and the loss of Phe in this domain results in decreased binding.[8] The Phe in the RGG domain of nucleolin might bind to G-tetrad or loops with a similar mechanism as in TLS/FUS and CIRBP.

The RGG domain is known as both a G-quadruplex folding and unfolding domain.[5,6,9−11,36] The RGG domains in the C-terminals of EWS and TLS/FUS fold G-quadruplexes, whereas the RGG domain in hnRNP A1 promotes RRM-mediated unfolding of G-quadruplexes.[5] Figure 2 shows that Phe in the RGG domain of nucleolin contributes to G-quadruplex stabilization except for the Phe of residue 690. The RGG domain in hnRNP A1 contains Phe but it might not be able to achieve G-quadruplex folding. The findings of the present study will contribute to reveal the roles of the RRM and RGG domains conserved in many nucleic-acid-binding proteins and to elucidate their biologic functions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.9b04221.

sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) of truncated or mutated proteins based on nucleolin (Figure S1); binding affinity of the RRM-RGG domain and its mutants to *BCL*-2 (Figure S2); binding ability of RRM in nucleolin to dA_{10}, dT_{10}, dG_{10}, or dC_{10} (Figure S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Takanori Oyoshi − *Department of Chemistry, Graduate School of Science, Shizuoka University, Shizuoka 422-8529, Japan;* orcid.org/0000-0003-4920-6790; Phone: +81-54-238-4760; Email: oyoshi.takanori@shizuoka.ac.jp; Fax: +81-54-237-3384

Author

Tatsuki Masuzawa − *Department of Chemistry, Graduate School of Science, Shizuoka University, Shizuoka 422-8529, Japan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.9b04221

Author Contributions

T.O. conceived the study, T.M. prepared the protein and performed the EMSA assay, and C.D. T.O. and T.M. analyzed the data and wrote the manuscript.

Funding

This work was supported by the JGC-S Scholarship Foundation, a Grant-in-Aid for Scientific Research (C) (No. 17K05930 to T.O.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. S.C. Lee at the National Cheng University in Taiwan for the nucleolin cDNA.

■ ABBREVIATIONS

hnRNP A1, heterogeneous nuclear ribonucleoprotein A1; CIRBP, cold-inducible mRNA-binding protein; EWS, Ewing's sarcoma; TLS/FUS, translocated in liposarcoma/fused in sarcoma; RRM, RNA recognition motif; RGG, arginine−glycine−glycine repeat

■ REFERENCES

(1) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The structure and function of DNA G-quadruplexes. *Trends Chem.* **2020**, *2*, 123−136.

(2) Mukherjee, A. K.; Sharma, S.; Chowdhury, S. Non-duplex G-quadruplex structures emerge as mediators of epigenetic modifications. *Trends Genet.* **2019**, *35*, 129−144.

(3) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-quadruplexes in biology: Principles and molecular mechanisms. *J. Mol. Biol.* **2017**, *429*, 2127−2147.

(4) Brázda, V.; Haronikova, L.; Liao, J. C. C.; Fojta, M. DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* **2014**, *15*, 17493−17517.

(5) Ghosh, M.; Singh, M. RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res.* **2018**, *46*, 10246−10261.

(6) Gonzalez, V.; Hurley, L. H. The C-terminus of nucleolin promotes the formation of the c-myc G-quadruplex and inhibition c-myc promoter activity. *Biochemistry* **2010**, *49*, 9706−9714.

(7) González, V.; Guo, K.; Hurley, L. H.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **2009**, *284*, 23622−23635.

(8) 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.

(9) 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.

(10) 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.

(11) Takahama, K.; Sugimoto, C.; Arai, S.; Kurokawa, R.; Oyoshi, T. Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's Sarcoma. *Biochemistry* **2011**, *50*, 5369−5378.

(12) Thandapani, P.; O'Connor, T. R.; Bailey, T. L.; Richard, S. Defining the RGG/RG motif. *Mol. Cell* **2013**, *50*, 613−623.

(13) Sun, X.; Ali, M. S. S. H.; Moran, M. The role of interactions of long non-coding RNAs and heterogeneous nuclear ribonucleoproteins in regulating cellular functions. *Biochem. J.* **2017**, *474*, 2925−2935.

(14) Cléry, A.; Blatter, M.; Allain, F. H.-T. RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290−298.

(15) Maris, C.; Dominguez, C.; Allain, F. H.-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118−2131.

(16) Flynn, R. L.; Centore, R. C.; O'Sullivan, R. J.; Rai, R.; Tse, A.; Songyang, Z.; Chang, S.; Karlseder, J.; Zou, L. TERRA and hnRNPA1 orchestrate an RPA-to-POT 1 switch on telomeric single-stranded DNA. *Nature* **2011**, *471*, 532−536.

(17) Redon, S.; Zemp, I.; Lingner, J. A three-state model for the regulation of telomerase by TERRA and hnRNPA1. *Nucleic Acids Res.* **2013**, *41*, 9117−9128.
(18) Paramasivam, M. P.; Membrino, A.; Cogoi, S.; Fukuda, H.; Nakagama, H.; Xodo, L. E. Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: implications for transcription. *Nucleic Acids Res.* **2009**, *37*, 2841−2853.
(19) Fukuda, H.; Katahira, M.; Tsuchiya, N.; Enokizono, Y.; Sugimura, T.; Nagano, M.; Nakagama, H. Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12685−12690.
(20) Grinstein, E.; Du, Y.; Santourlidis, S.; Christ, J.; Uhrberg, M.; Wernet, P. Nucleolin regulates gene expression in CD34-positive hematopoietic cells. *J. Biol. Chem.* **2007**, *282*, 12439−12449.
(21) Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H. A mechanosensor mechanism controls the G-quadruplex/i-Motif molecular switch in the MYC promoter NHE III_1. *J. Am. Chem. Soc.* **2016**, *138*, 14138−14151.
(22) Abdelmohsen, K.; Tominaga, K.; Lee, E. K.; Srikantan, S.; Kang, M.-J.; Kim, M. M.; Selimyan, R.; Martindale, J. L.; Yang, X.; Carrier, F.; Zhan, M.; Becker, K. G.; Gorospe, M. Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. *Nucleic Acids Res.* **2011**, *39*, 8513−8530.
(23) Bian, W.-X.; Xie, Y.; Wang, X.-N.; Xu, G.-H.; Fu, B.-S.; Li, S.; Long, G.; Zhou, X.; Zhang, X.-L. Binding of Cellular nucleolin with the viral core RNA G-quadruplex structure suppresses GCV replication. *Nucleic Acids Res.* **2019**, *47*, 56−68.
(24) Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palu, G.; Fabris, D.; Richter, S. N. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* **2015**, *43*, 8884−8897.
(25) Haeusler, A. R.; Donnelly, C. J.; Periz, G.; Simko, E. A. J.; Shaw, P. G.; Kim, M.-S.; Maragakis, N. J.; Troncoso, J. C.; Pandey, A.; Sattler, R.; Rothstein, J. D.; Wang, J. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **2014**, *507*, 195−200.
(26) Mendoza, O.; Bourdoncle, A.; Boule, J.-B.; Brosh, R. M., Jr.; Mergny, J.-L. G-quadruplex and helicases. *Nucleic Acids Res.* **2016**, *44*, 1989−2006.
(27) Tippana, R.; Chen, M. C.; Demeshkina, N. A.; Ferre-D Amare, A. R.; Myong, S. RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.* **2019**, *10*, No. 1855.
(28) Dexheimer, T. S.; Carey, S. S.; Zuohe, S.; Gokhale, V. M.; Hu, X.; Murata, L. B.; Maes, E. M.; Weichsel, A.; Sun, D.; Meuillet, E. J.; Montfort, W. R.; Hurley, L. H. NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III_1. *Mol. Cancer Ther.* **2009**, *8*, 1363−1377.
(29) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **2006**, *128*, 1096−1098.
(30) Kondo, K.; Mashima, T.; Oyoshi, T.; Yagi, R.; Kurokawa, R.; Kobayashi, N.; Nagata, T.; Katahira, M. Plastic roles of phenylalanine and tyrosine residues of TLS/FUS in complex formation with the G-quadruplexes of telomeric DNA and TERRA. *Sci. Rep.* **2018**, *8*, No. 2864.
(31) Phan, A. T.; Modi, Y. S.; Patel, D. J. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.
(32) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-length-dependent folding of G-quadruplex. *J. Am. Chem. Soc.* **2004**, *126*, 16405−16415.
(33) 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.
(34) 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.
(35) Takahama, K.; Miyawaki, A.; Shitara, T.; Mitsuya, K.; Morikawa, M.; Hagihara, M.; Kino, K.; Yamamoto, A.; Oyoshi, T. G-quadruplex DNA- and RNA-specific-binding proteins engineered from the RGG domain of TLS/FUS. *ACS Chem. Biol.* **2015**, *10*, 2564−2569.
(36) Liu, X.; Ishizuka, T.; Bao, H.-L.; Wada, K.; Takeda, Y.; Iida, K.; Nagasawa, K.; Yang, D.; Xu, Y. Structure-dependent binding of hnRNPA1 to telomere RNA. *J. Am. Chem. Soc.* **2017**, *139*, 7533−7539.

https://dx.doi.org/10.1021/acsomega.9b04221
ACS Omega 2020, 5, 5202−5208