BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

Loop Lengths of G-Quadruplex Structures Affect the G-Quadruplex DNA Binding Selectivity of the RGG Motif in Ewing's Sarcoma

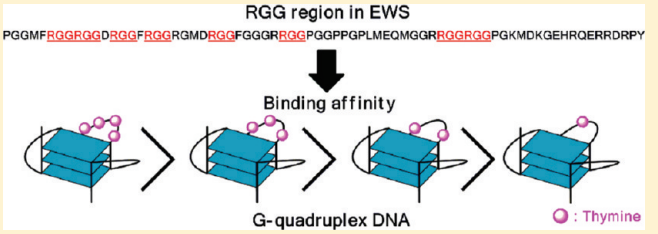
Kentaro Takahama,[†] Chieri Sugimoto,[†] Shigeki Arai,[‡] Riki Kurokawa,[‡] and Takanori Oyoshi*[,†]

[†]Faculty of Science, Department of Chemistry, Shizuoka University, 836 Ohya, Suruga, Shizuoka 422-8529, Japan

[‡]Division of Gene Structure and Function, Saitama Medical University Research Center for Genomic Medicine, 1397-1 Yamane, Hidaka, Saitama 350-1241, Japan

Supporting Information

ABSTRACT: The G-quadruplex nucleic acid structural motif is a target for designing molecules with potential anticancer properties. To achieve therapeutic selectivity by targeting the G-quadruplex, the molecules must be able to differentiate between the DNA of different G-quadruplexes. We recently reported that the Arg-Gly-Gly repeat (RGG) of the C-terminus in Ewing's sarcoma protein (EWS), which is a group of dominant oncogenes that arise due to chromosomal translocations, is capable of binding to G-quadruplex telomere DNA and RNA via arginine residues and stabilize the G-quadruplex DNA form in vitro. Here, we show that the RGG of EWS binds preferentially to G-quadruplexes with longer loops, which is not related to the topology of the G-quadruplex structure. Moreover, the G-quadruplex DNA binding of the RGG in EWS depends on the phosphate backbone of the loops in the G-quadruplex DNA. We also investigated the G-quadruplex DNA binding activity of the N- and C-terminally truncated RGG to assess the role of the regions in the RGG in G-quadruplex DNA binding. Our findings indicate that the RGG and the other arginine-rich motif of residues 617−656 of the RGG in EWS are important for the specific binding to G-quadruplex DNA. These findings will contribute to the development of molecules that selectively target different G-quadruplex DNA.



The involvement of G-quadruplex structures in cellular processes such as telomere elongation and transcriptional control has stimulated the development of drugs that are selective for different G-quadruplex structures.[1−12] The human telomere DNA and the promoter regions of some important genes, such as *c-myc*, *bcl-2*, *c-kit*, and d(GGA)$_4$ of *c-myb*, contain sequences that form different G-quadruplex structures under physiological conditions, and their G-quadruplex structures have been determined by nuclear magnetic resonance (NMR).[13−23] Previous studies reported the recognition of the human telomere DNA and subsequent inhibition of telomerase activity by analogues of the cationic porphyrin TMPyP4.[24] Moreover, small molecules, which interact with the G-quadruplex structures, modulated transcription of oncogenes *c-myc*, *bcl-2*, *c-kit*, *KRAS*, and platelet-derived growth factor receptor β.[4−12] Investigation of the transcriptional control and other biologic roles of the G-quadruplex structure selectivity by targeting the G-quadruplex in vivo requires molecules that can differentiate among the different types of G-quadruplexes. A selenium-substituted expanded porphyrin, Se2SAP, has increased selectivity for the *c-myc* G-quadruplex structure.[25] In addition, 3,8,10-trisubstituted isoalloxazines and triazole-linked acridines exhibited selective binding to *c-kit* and human telomere G-quadruplex structure, respectively.[12,26]

The engineered zinc finger protein, Gq1, binds to human telomeric G-quadruplex DNA and inhibits the action of telomerase in vitro,[27] suggesting that G-quadruplex DNA binding protein is potentially useful for investigating the biologic role of G-quadruplex DNA. Recently, we identified Ewing's sarcoma protein (EWS), derived primarily from studies of a group of dominant oncogenes that arise due to chromosomal translocations in which EWS is fused to a variety of cellular transcription factors, as a G-quadruplex DNA- and RNA-binding protein.[28−30] The specificity of G-quadruplex recognition depends on the guanidinium group of the arginine in the Arg-Gly-Gly repeat (RGG) in the C-terminus of EWS. Furthermore, inhibition of DNA polymerase on a template containing a human telomere sequence in the presence of the RGG occurs in an RGG concentration-dependent manner by the formation of a stabilized G-quadruplex DNA−RGG complex.[28] Little is known, however, about the G-quadruplex DNA recognition specificity of the RGG of EWS. To gain further insight into the interaction of the G-quadruplex DNA−RGG complex of EWS, we performed an electrophoretic mobility shift assay (EMSA) for the RGG and mutated RGG of EWS with various G-quadruplex DNAs. Here, we show that RGG3 of EWS binds preferentially to G-quadruplex DNAs with longer loops, regardless of the

Received: March 16, 2011
Revised: April 28, 2011
Published: May 11, 2011

topology of the G-quadruplex structure and the DNA sequence. We also determined that the RGG sequences between amino acids 584 and 638 and the C-terminus between amino acids 617 and 656 within RGG3 are important for the specific binding to G-quadruplex human telomere DNA.

MATERIALS AND METHODS

Plasmid Constructs. The EWS plasmid was used as a template for polymerase chain reaction (PCR) and the construction of RGG3, RGG3-N1, RGG3-N2, RGG3-N3, RGG3-C1, RGG3-C2, RGG3-N2-2, and RGG3-N2-4, which were cloned into the pGEX6P-1 vector (GE Healthcare) between the *Bam*HI and *Xho*I sites using the following sets of primers: RGG3 forward, d(CGG AAT TCG CCC CAA AGC CTG AAG GCT T), and RGG3 reverse, d(CGC TCG AGT CAC TAG TAG GGC CGA TCT CTG C), for pGEX-RGG3; RGG3-N1 forward, d(CGG AAT TCC CTG GTG GCA TGC GG), and RGG3 reverse for pGEX-RGG3-N1; RGG3-N2 forward, d(CGG AAT TCC CCG GTG GAA TGT TCA GAG G), and RGG3 reverse for pGEX-RGG3-N2; RGG3-N3 forward, d(CGG AAT TCC GGG GCA TGG ACC GAG), and RGG3 reverse for pGEX-RGG3-N3; RGG3 forward and RGG3-C1 reverse, d(GCC TCG AGT CAT CTT CCT CCC ATC TGT TCC AT), for pGEX-RGG3-C1; RGG3 forward and RGG3-C2 reverse, d(GCC TCG AGT CAG TCC ATG CCC CGG C), for pGEX-RGG3-C2; RGG3-N2 forward and RGG3-N2 reverse, d(GCC TCG AGT CAT CCT CCA CGT CCT CCT CTT C), for pGEX-RGG3-N2-2; RGG3-N2 forward and RGG3-N2-4 reverse, d(GCC ACC TCG TCT TCC TCC AC), for pGEX-RGG3-N2-4. pGEX-RGG3-N2-1, pGEX-RGG3-N2-3, and pGEX-RGG3-N2-4 were obtained by deletion in pGEX-RGG3 using a KOD -Plus- Mutagenesis Kit (TOYOBO). To construct pGEX-RGG3-N2-1, we performed PCR with pGEX-RGG3-N2 as a template and the following primers: RGG3-N2-1 forward, d(AGA AGA GGA GGA CGT GGA GG), and RGG3-N2-1 reverse, d(GCC ACC TCG TCT TCC TCC AC). pGEX-RGG3-N2-3 was generated by PCR using pGEX-RGG3-N2-2 as a template and the following primers: RGG3-N2-1 forward and RGG3-N2-1 reverse. pGEX-RGG3-N2-4 was generated by PCR using pGEX-RGG3-N2-2 as a template and the following primers: RGG3-N2-4 forward, d(TGA CTC GAG CGG CCG CAT CGT GAC), and RGG3-N2-4 reverse, d(GCC ACC TCG TCT TCC TCC AC). All constructs were verified by automated DNA sequencing. All DNA primers, d_3Htelo and d_9Htelo, which contain three and nine abasic sites, respectively instead of nucleotides in the loops within Htelo, and the other DNA oligomers were obtained form Operon Biotechnologies.

Expression and Purification of Glutathione *S*-Transferase (GST) 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 the transformants were grown at 37 °C in Luria-Bertani medium containing ampicillin (0.1 mg/mL). Protein expression was induced at an A_{600} of 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). Pellets were resuspended in W buffer [100 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), and 0.1 mM phenylmethanesulfonyl fluoride] and lysed by sonication (model UR-20P, Tomy Seiko) at 4 °C. The supernatants containing the expressed proteins were centrifuged for 15 min at 16200*g* and 4 °C, and the resulting supernatant was applied to a 1 mL GSTrap FF column (GE Healthcare). The column was washed with 20 mL of W buffer and then 10 mL of potassium buffer [50 mM Tris-HCl (pH 7.5), 100 mM KCl, 1 mM EDTA, and 1 mM DTT]. GST tags were cleaved using a buffer containing 8 units of PreScission protease (GE Healthcare) per milliliter on a column for 16 h at 4 °C, and the proteins were eluted with potassium buffer. The protein concentrations were determined using the BCA Protein Assay Kit (Thermo Scientific, Waltham, MA).

Electrophoretic Mobility Shift Assay. Electrophoretic mobility shift assays were performed as described previously.[28] ^{32}P-labeled oligonucleotide annealing and quadruplex formation were performed by heating samples to 95 °C on a thermal heating block and cooling them to 4 °C at a rate of 2 °C/min in 50 mM Tris-HCl (pH 7.5) in the presence of 100 mM KCl. Binding reactions were performed in a final volume of 20 μL using 20 fmol of the labeled oligonucleotide and 50 pmol of purified proteins in binding buffer [50 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 0.5 mM DTT, 0.1 mg/mL bovine serum albumin, 1 μg/mL calf thymus DNA, and 100 mM KCl]. The samples were incubated for 1 h at 4 °C and then loaded on a 6% polyacrylamide (19:1 acrylamide:bisacrylamide ratio) nondenaturing gel. Both the gel and the electrophoresis buffer contained 0.5× TBE buffer (45 mM Tris base, 45 mM boric acid, and 0.5 mM EDTA) with 20 mM KCl. Electrophoresis was performed at 10 V/cm for 2 h at 4 °C. The gels were exposed in a phosphorimager cassette and imaged (Personal Molecular Imager FX, Bio-Rad, Hercules, CA). Bands were quantified using ImageQuant. The data were plotted as ϕ (1 fraction of free DNA) versus the protein concentration to determine the equilibrium dissociation constants (K_d), which is equal to the protein at which half of the free DNA is bound. The DNA concentration was fixed at 1 nM, while the concentration of RGG3 added to the binding reaction was varied, as shown above each lane. K_d were extracted by nonlinear regression using Microsoft Excel 2007 and the following equation:

$$\phi = [P]/\{K_d + [P]\}$$

Circular Dichroism Spectroscopy. Circular dichroism (CD) spectroscopy was performed as described previously.[28] CD spectra were recorded on a model J-820 instrument (Jasco). The CD spectra of oligonucleotides (50 μM base concentration) with or without RGG3 (2, 1, 0) equivalent to DNA (RGG3/DNA) in 50 mM Tris-HCl (pH 7.5) and 100 mM KCl as specified were recorded using a 0.2 cm path length cell at 20 °C.

RESULTS AND DISCUSSION

The RGG3 Region of EWS Binds with High Affinity to G-Quadruplex Human Telomere DNA, *c-kit*, and *bcl-2*. RGG3 of EWS interacts with G-quadruplex DNA and RNA with structure specificity, whereas the EWS activation domain (EAD), RGG1, and RNA recognizion motif (RRM)-RGG2-zinc finger (ZnF) of EWS do not (Figure 1A).[28] In addition, the EMSA data of G-quadruplex human telomere DNA (Htelo) and RGG3 fit to a hyperbolic equation to give a K_d of 13 ± 3 nM in the presence of 100 mM KCl. To investigate the binding selectivity of RGG3 with respect to G-quadruplex structures, we performed an EMSA of RGG3 with a diverse group of G-quadruplex structures, $(GGA)_4$, *c-kit*, *c-myc*, and *bcl-2*, previously

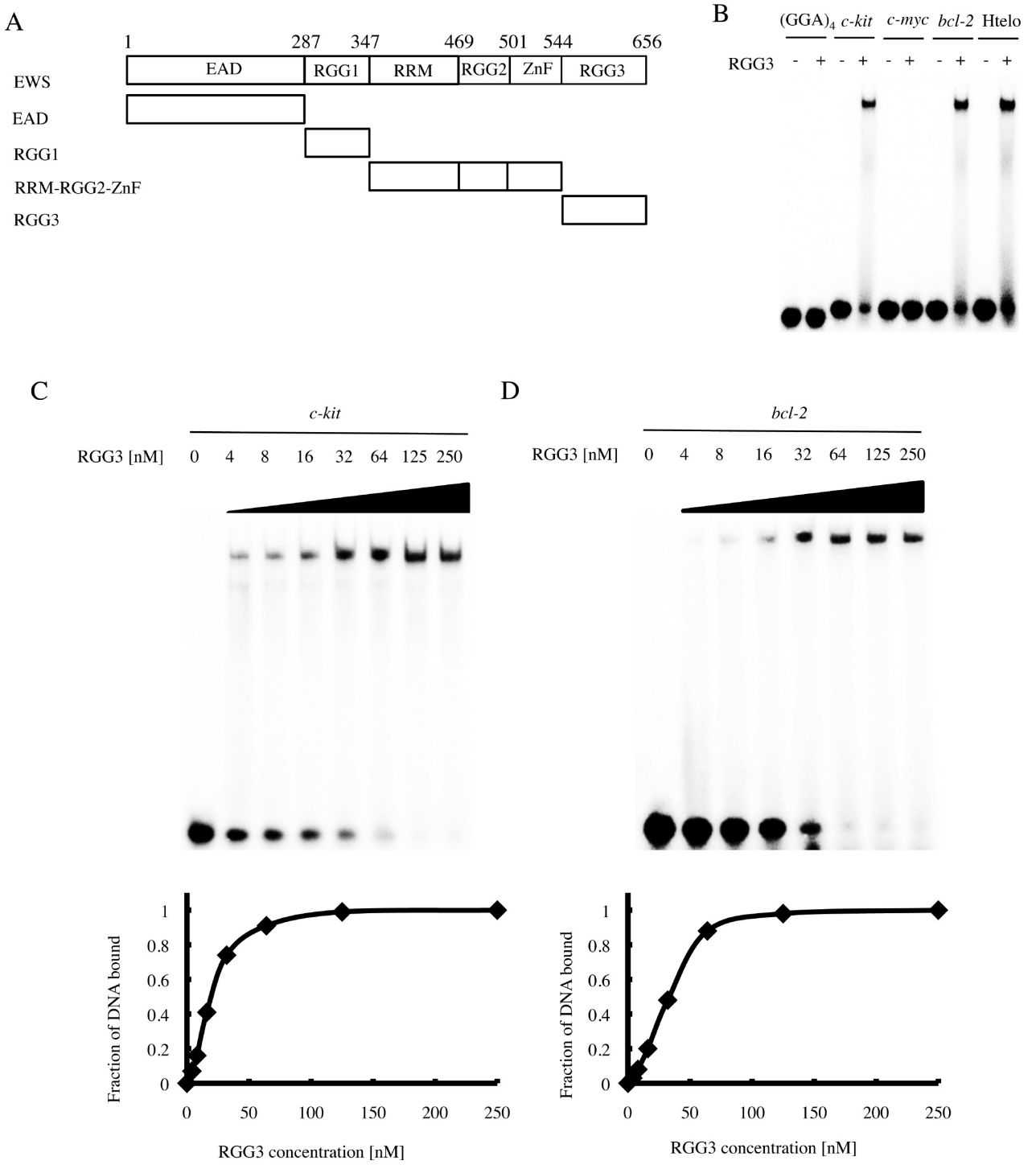


Figure 1. Structural features of EWS and G-quadruplex DNA binding affinity of RGG3. (A) Schematic representation of the deletion mutants of the EWS: EAD (residues 1−287), transcriptional activation domain; RGG1 (residues 288−347), RGG-rich motif 1; RRM (residues 348−469), RNA recognition motif; RGG2 (residues 450−501), RGG-rich motif 2; ZnF (residues 502−544), zinc finger; RGG3 (residues 545−656), RGG-rich motif 3. (B) An EMSA was performed with RGG3 (lanes 2, 4, 6, and 8) and ^{32}P-labeled $(GGA)_4$ (lanes 1 and 2), *c-kit* (lanes 3 and 4), *c-myc* (lanes 5 and 6), *bcl-2* (lanes 7 and 8), or Htelo (lanes 9 and 10). (C and D) The equilibrium binding curve was obtained by calculating the fraction of *c-kit* (C) or *bcl-2* (D) bound at varying RGG3 concentrations. The binding constant (K_d) was determined by fitting to the equation (see Materials and Methods). The DNA−protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

determined by NMR, compared with Htelo (Table 1).[13−17,19−23] In particular, $(GGA)_4$, *c-kit*, and *c-myc* form an intramolecular propeller-type parallel-strand G-quadruplex in a K^+-containing solution.[19−23] The structure of $(GGA)_4$ is composed of two intramolecular quadruplexes, each comprising a G-tetrad, a G(:A):G(:A):G(:A)G heptad, and a GGA segment, forming a trinucleotide loop.[19,20] The structure of *c-kit*, which is 87 nucleotides upstream of the transcription start site of the human

dx.doi.org/10.1021/bi2003857 |*Biochemistry* 2011, 50, 5369–5378

Table 1. Sequences of Oligonucleotides Used in EMSAs and CD Spectroscopy with the Bases of the Loops Underlined

name	sequence[a]
$(GGA)_4$	GGAGGAGGAGGA
c-kit	AGGGAGGGCGCTGGGAGGAGGG
c-myc	TGAGGGTGGGGAGGGTGGGGAA
bcl-2	GGGCGCGGGAGGAATTGGGCGGG
Htelo	AGGGTTAGGGTTAGGGTTAGGG
L444	TGGGTTTTGGGTTTTGGGTTTTGGGT
L333	TGGGTTTGGGTTTGGGTTTGGGT
L222	TGGGTTGGGTTGGGTTGGGT
L111	TGGGTGGGTGGGTGGGT
L414	TGGGTTTTGGGTGGGTTTTGGGT
L313	TGGGTTTGGGTGGGTTTGGGT
L212	TGGGTTGGGTGGGTTGGGT
L141	TGGGTGGGTTTTGGGTGGGT
L131	TGGGTGGGTTTGGGTGGGT
L121	TGGGTGGGTTGGGTGGGT
tetHtelo	TTAGGGT
d_3Htelo	AGGGTTAGGGdddGGGTTAGGG
d_9Htelo	AGGGdddGGGdddGGGdddGGG
mutHtelo	AGGGTTAGTGTTAGTGTTAGGG

[a] d indicates an abasic site.

c-kit gene, is the snapback parallel-strand G-tetrad core with four loops in which two single-residue linkers, a two-residue linker, and a five-residue segment form in the structure.[21,22] *c-myc* has three loops in its structure, in which the central loop contains two nucleotides, while the two flanking loops contain only one nucleotide.[23] On the other hand, *bcl-2* and Htelo form an intramolecular hybrid (3 + 1) G-quadruplex with three loops in K^+-containing solution.[13−17] In the case of *bcl-2*, a single-residue linker, a two-residue linker, and a six-residue linker form three loops in the structure, while all of the loops of Htelo are three-residue linkers. Recombinant RGG3 for binding to nucleic acids was expressed in *E. coli* as proteins fused to GST and purified using glutathione agarose. ^{32}P-labeled nucleic acids were first incubated for 24 h in 100 mM KCl to allow for quadruplex formation and then with GST tag-digested RGG3 for 1 h at room temperature. The protein−DNA complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography. The EMSA of these G-quadruplex DNAs and RGG3 showed that the affinities of RGG3 for *c-kit*, *bcl-2*, and Htelo G-quadruplex structures were higher than that for $(GGA)_4$ and *c-myc* (Figure 1B).

To determine the ability of RGG3 to bind the G-quadruplex, we incubated various concentrations of RGG3 with ^{32}P-labeled *c-kit*, *bcl-2*, $(GGA)_4$, and *c-myc* in a K^+ solution. As the RGG3 concentration increased, the amount of free DNA decreased, and the amount of the higher-molecular weight complex increased (Figure 1C,D). The mobility shift data were fitted to a hyperbolic equation to give K_d values of 18 ± 6 nM (*c-kit*) and 29 ± 10 nM (*bcl-2*), while the K_d of $(GGA)_4$ and *c-myc* was greater than 0.5 μM (data not shown). These results suggest that RGG3 binds preferentially to G-quadruplex DNA with long loops in a sequence- and topology-independent manner.

The RGG3 Region of EWS Binds Preferentially to G-Quadruplex DNA with a Long Loop in a Topology-Independent Manner. To investigate the effect of loop length and topology on G-quadruplex structures for RGG3 binding, we performed an EMSA of G-quadruplex DNA with various loop lengths and topologies on binding RGG3 (Table 1 and Figure 2). G-Quadruplex DNA comprising four d(GGG) repeats and three $d(T)_n$ loops (n = 1−4) forms stable G-quadruplex structures, as determined by UV melting and CD spectroscopy.[31] The G-quadruplex structure comprising three d(T) loops (L111) is purely parallel, whereas that comprising three $d(T)_2$ loops (L222) exhibited substantial polymorphism of antiparallel and parallel fold structures in a buffer consisting of 100 mM KCl and 10 mM Tris-HCl (pH 7.4) at 4 °C. On the other hand, the G-quadruplex structures comprising three $d(T)_3$ (L333) and $d(T)_4$ (L444) loops were antiparallel in buffer. An EMSA of RGG3 and these G-quadruplex structures showed that the most favorable G-quadruplex for binding contained L333 and L444, while the most unfavorable G-quadruplex contained L111 (Figure 2A). These results suggest preferential binding of RGG3 to G-quadruplex DNA with longer loops.

Figure 1B and Figure 2A show that RGG3 of EWS might bind unfavorably to parallel G-quadruplex structures, such as *c-myc*, $(GGA)_4$, L222, and L111. To determine whether RGG3 binds unfavorably to parallel G-quadruplex structures, we analyzed the effect of two $d(T)_n$ loops (n = 1−4) of the parallel G-quadruplex DNA with d(T) in the middle loop, which is a G-quadruplex structure containing two $d(T)_4$ loops with d(T) in middle loop (L414) that was reported to produce a parallel topology.[32] The CD spectra of G-quadruplex structures containing two $d(T)_n$ loops (n = 1−3) with a d(T) in the middle loop (L111, L212, and L313) shown in Figure 1S of the Supporting Information is typical of the parallel structure. An EMSA showed that the G-quadruplex with L414 had the best binding to RGG3 of EWS, while the G-quadruplex with L111 had the worst binding (Figure 2B). To further investigate the binding of RGG3 to G-quadruplex DNA with longer loops, we analyzed the effect of a $d(T)_n$ loops (n = 1−4) in the middle loop with two d(T) segments in the loops, which is a G-quadruplex structure containing $d(T)_4$ loops in the middle loop with two d(T) segments in the loops (L414) that was reported to produce a parallel topology.[32] The CD spectra of G-quadruplex structures containing $d(T)_n$ loops (n = 1−3) in the middle loop (L111, L121, and L131) with two d(T) segments in the loops shown in Figure 2S of the Supporting Information are typical of the parallel structure. An EMSA showed that the G-quadruplex with L141 had the best binding to RGG3 of EWS (Figure 2C). Taken together, these data suggest that RGG3 binds preferentially to G-quadruplexes with longer loops, regardless of differences in the topology of the G-quadruplex. Moreover, these data suggest that RGG3 associates with the longer loops of the G-quadruplex.

The RGG3 Region of EWS Binds to the Phosphate Backbone of the Loops in G-Quadruplex DNA. Recently, the crystal structure of a human telomeric G-quadruplex DNA with a d(TAGGGTTAGGG) sequence in complex with TMPyP4 showed that TMPyP4 stacks on the TTA nucleotides at part of the external loop in the parallel form.[33] On the other hand, the crystal structure of a G-quadruplex DNA with a (GGGGTTTTGGGG) sequence in complex with 3,6-disubstituted acridines showed that the acridine molecules formed a pair of hydrogen bonds with a thymine base in the diagonal loop.[34] To investigate whether RGG3 of EWS recognizes the phosphate backbone or bases of loops in the G-quadruplex, we assayed the binding of RGG3 with d(TTAGGGT) (tetHtelo) and G-quadruplex DNA containing three (d_3Htelo) or nine abasic sites

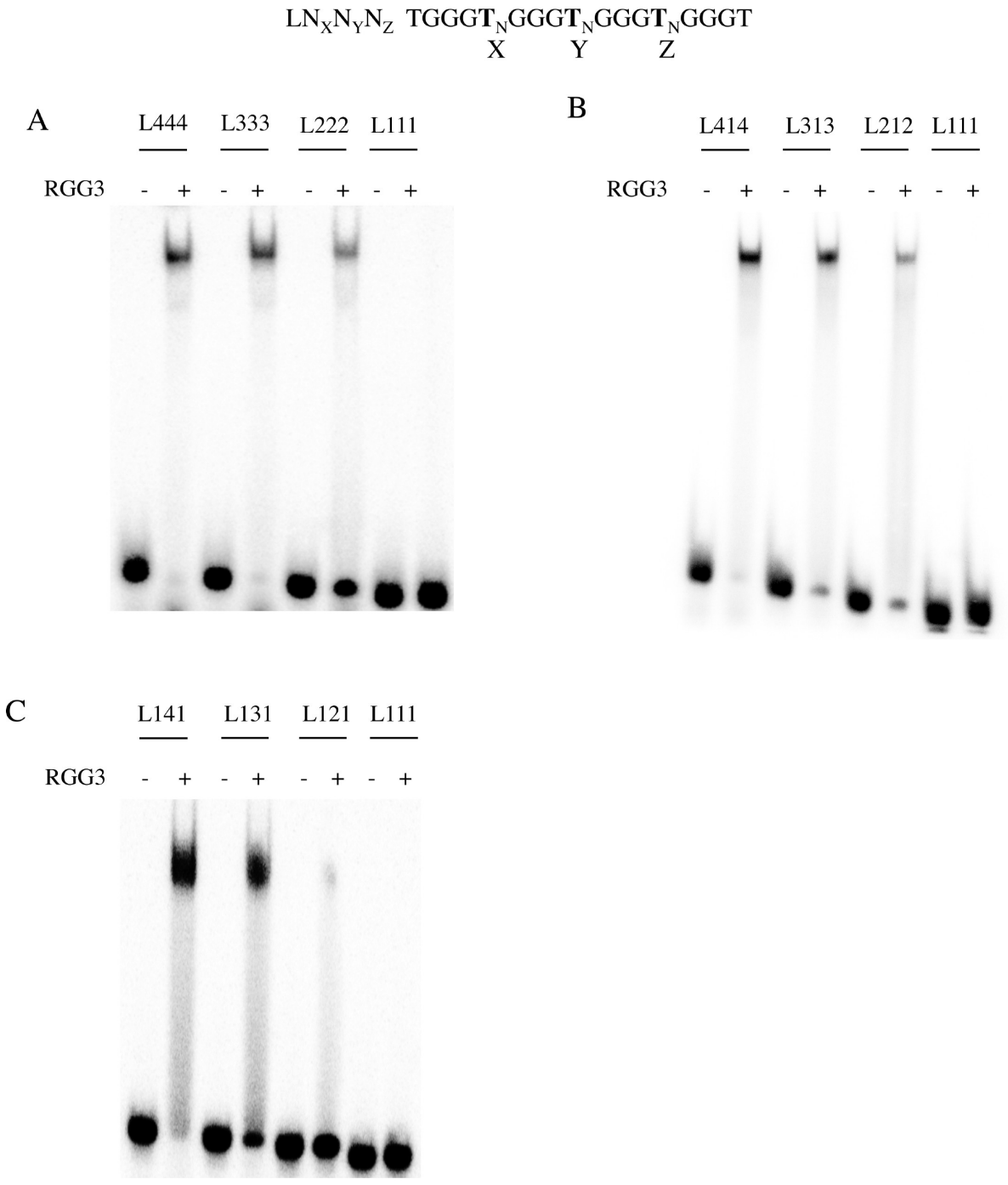


Figure 2. Effect of G-quadruplex loop length on the G-quadruplex binding affinity of RGG3. (A) An EMSA was performed using RGG3 (lanes 2, 4, 6, and 8) with ^{32}P-labeled L444 (lanes 1 and 2), L333 (lanes 3 and 4), L222 (lanes 5 and 6), or L111 (lanes 7 and 8). (B) An EMSA was performed with RGG3 (lanes 2, 4, 6, and 8) with ^{32}P-labeled L414 (lanes 1 and 2), L313 (lanes 3 and 4), L212 (lanes 5 and 6), or L111 (lanes 7 and 8). (C) An EMSA was performed with RGG3 (lanes 2, 4, 6, and 8) with ^{32}P-labeled L141 (lanes 1 and 2), L131 (lanes 3 and 4), L121 (lanes 5 and 6), or L111 (lanes 7 and 8). The DNA−protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

(d_9Htelo) instead of nucleotides in the loops within Htelo (Table 1 and Figure 3). The tetHtelo forms an intermolecular parallel-strand G-quadruplex that has no loop in its structure in a K^+-containing solution.[35] We confirmed the CD spectra of d_3Htelo, which were similar with those of Htelo as the hybrid (3 + 1) G-quadruplex form, and d_9Htelo, which results in a positive band centered at 265 nm and a negative band centered at 240 nm as the parallel form (Figure 3S of the Supporting Information).[17] EMSA studies showed that RGG3 binds to d_3Htelo or d_9Htelo despite the absence of bases in the G-quadruplex loops, but not to tetHtelo. This finding indicates that RGG3 of EWS recognizes the phosphate backbone of the loops in G-quadruplex DNA.

The RGG3 Region of EWS Induces the Formation of a G-Quadruplex Parallel Form. We have performed a CD spectroscopic analysis of Htelo in the presence of RGG3.[28]

dx.doi.org/10.1021/bi2003857 |*Biochemistry* 2011, 50, 5369–5378

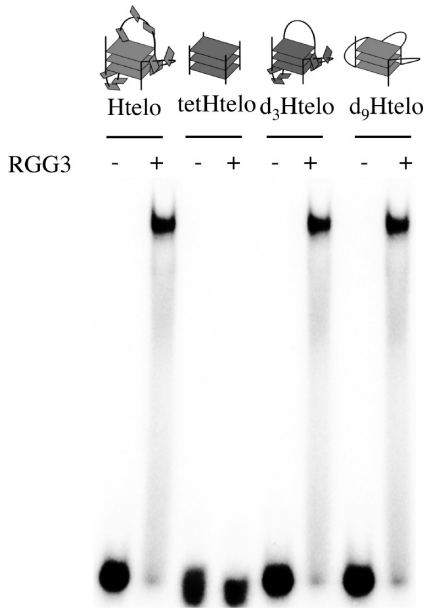


Figure 3. Effect of the number of loops and the phosphate backbone on G-quadruplex loops on the G-quadruplex binding affinity of RGG3. An EMSA was performed with RGG3 (lanes 2, 4, 6, and 8) and ^{32}P-labeled Htelo (lanes 1 and 2), tetHtelo (lanes 3 and 4), d_3Htelo (lanes 5 and 6), or d_9Htelo (lanes 7 and 8). The DNA structures used as a probe are indicated above each lane. The DNA−protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

The results indicated that the addition of 1 equiv of RGG3 increased the ellipticity and shifted the spectrum from a strong positive band to 265 nm, which is characteristic of the parallel form and consistent with the results of previous CD studies.[17] These data provided a model showing the change from the hybrid (3 + 1) G-quadruplex to the parallel form with the association of RGG3. To gain further insight into the induction of G-quadruplex formation by RGG3 of EWS, we performed a CD spectroscopic analysis with various G-quadruplex DNAs in the presence of RGG3 (Figure 4). In the absence of RGG3, the CD spectrum of L333 had a positive band at 290 nm and a shoulder in the 250−270 nm region, consistent with the results of previous CD studies of a mixed conformation in buffer containing 100 mM KCl and 50 mM Tris-HCl (pH 7.5) at 20 °C.[36] Upon addition of 1 or 2 equiv of RGG3 to L333 and *bcl-2*, we observed a positive band centered at 265 nm and a negative band centered at 240 nm (Figure 4A,B, red and yellow lines), which is characteristic of the parallel form and consistent with the results of a previous CD study.[36] These findings suggest that RGG3 binds to the L333 and *bcl-2* G-quadruplex and changes the parallel form with the association of RGG3. Incubation of L414 and *c-kit*, the structures of which were identified as the parallel form by CD and NMR, respectively, with RGG3 did not alter the positive or negative bands, however, as demonstrated by analysis of the CD spectrum (Figure 4C,D).[21,22,32] Alternatively, the CD spectrum of L444 had a positive band at 290 nm and a negative band at 260 nm in the absence of RGG3 in a buffer containing 100 mM KCl and 50 mM Tris-HCl (pH 7.5) at 20 °C (Figure 4E, black line), which is characteristic of the antiparallel form and consistent with the results of a previous CD study.[17] The addition of 1 equiv of RGG3 to L444 induced substantial polymorphism, and the addition of 2 equiv of RGG3 to L444 induced a positive band centered at 265 nm and a negative band centered at 240 nm (Figure 4E, red and yellow lines), which is characteristic of the parallel form and consistent with the results of a previous CD study.[17] These findings suggest that the excess RGG3 changes from the L444 antiparallel G-quadruplex form to the parallel form with the association of RGG3. Additionally, to determine whether RGG3 of EWS recognizes and alters the G-quadruplex structures in the presence of 100 mM Na^+, we performed an EMSA and CD spectroscopic analysis with Htelo, which is resolved as an antiparallel form by NMR, and L444 in the presence of RGG3.[37] Figure 4S of the Supporting Information shows that RGG3 binds to the Htelo and L444 G-quadruplex in the presence of Na^+ but did not alter their CD spectra.

To determine whether RGG3 significantly affects the CD spectra of the Htelo structure in the presence of a low concentration of K^+, we performed a CD analysis of recombinant RGG3 and Htelo in buffer containing 2.5 mM KCl. Incubating RGG3 with Htelo in 2.5 mM KCl resulted in a shift from a weak G-quadruplex peak at 290 nm to a strong peak at 265 nm with an increase in the RGG3 protein concentration (Figure 5S of the Supporting Information), providing evidence that RGG3 of EWS induces the formation of the Htelo G-quadruplex parallel form structure.

The Minimal G-Quadruplex Specific Binding Domain of EWS Consists of Amino Acids 584−656. We previously demonstrated that arginines between amino acids 589 and 597 within RGG3 are important for the binding of RGG3 to G-quadruplex DNA.[28] To determine the RGG3 domains responsible for specific binding to G-quadruplex DNA, we generated the RGG3 mutant proteins, i.e., RGG3, RGG3-N1 (residues 568−656), RGG3-N2 (residues 584−656), RGG3-N3 (residues 603−656), RGG3-C1 (residues 545−632), and RGG3-C2 (residues 545−606). The truncated constructs of the RGG3 N-terminus, RGG3-N1 and RGG3-N2, bound specifically to Htelo nearly as well as RGG3 in an EMSA, whereas RGG3-N3 did not bind to Htelo and nonspecifically bound to a mutated human telomere (mutHtelo) in which T was replaced with G at positions 9 and 15 to destabilize the G-quadruplex formation as previously confirmed by CD and UV spectroscopy (Figure 5).[28] Furthermore, to determine whether RGG3-N2 binding affected the stability of the G-quadruplex structure of Htelo, we performed a DNA polymerase stop assay as described previously (Figure 6S of the Supporting Information).[28] As the RGG3-N2 protein concentration increased, the level of the full-length 76-mer product decreased, and the level of the stop-site product increased. Thus, these results indicate that RGG3-N2 binds to and stabilizes the folded G-quadruplex form. On the other hand, the truncated constructs of the RGG3 C-terminus, RGG3-C1, showed slightly weaker binding to Htelo than RGG3 in an EMSA. Moreover, RGG3-C2 did not bind to Htelo and nonspecifically bound to mutHtelo. These results suggest that amino acids 584−656 and the C-terminus of RGG3 are required for specific binding to G-quadruplex DNA.

We further investigated the RGG3 region that contributes to the G-quadruplex binding specificity by comparing the behavior of RGG3-N2 mutant recombinant proteins, i.e., RGG3-N2-1 (residues 584−656, Δ617−631), RGG3-N2-2 (residues 584−638), RGG3-N2-3 (residues 584−638, Δ617−631), and RGG3-N2-4 (residues 584−617), with regard to Htelo (Figure 6). The C-terminally truncated constructs and RGG deletion mutants of RGG3-N2 did not bind to Htelo but did bind to mutHtelo (Figure 6, lanes 4−7), suggesting that the RGG and

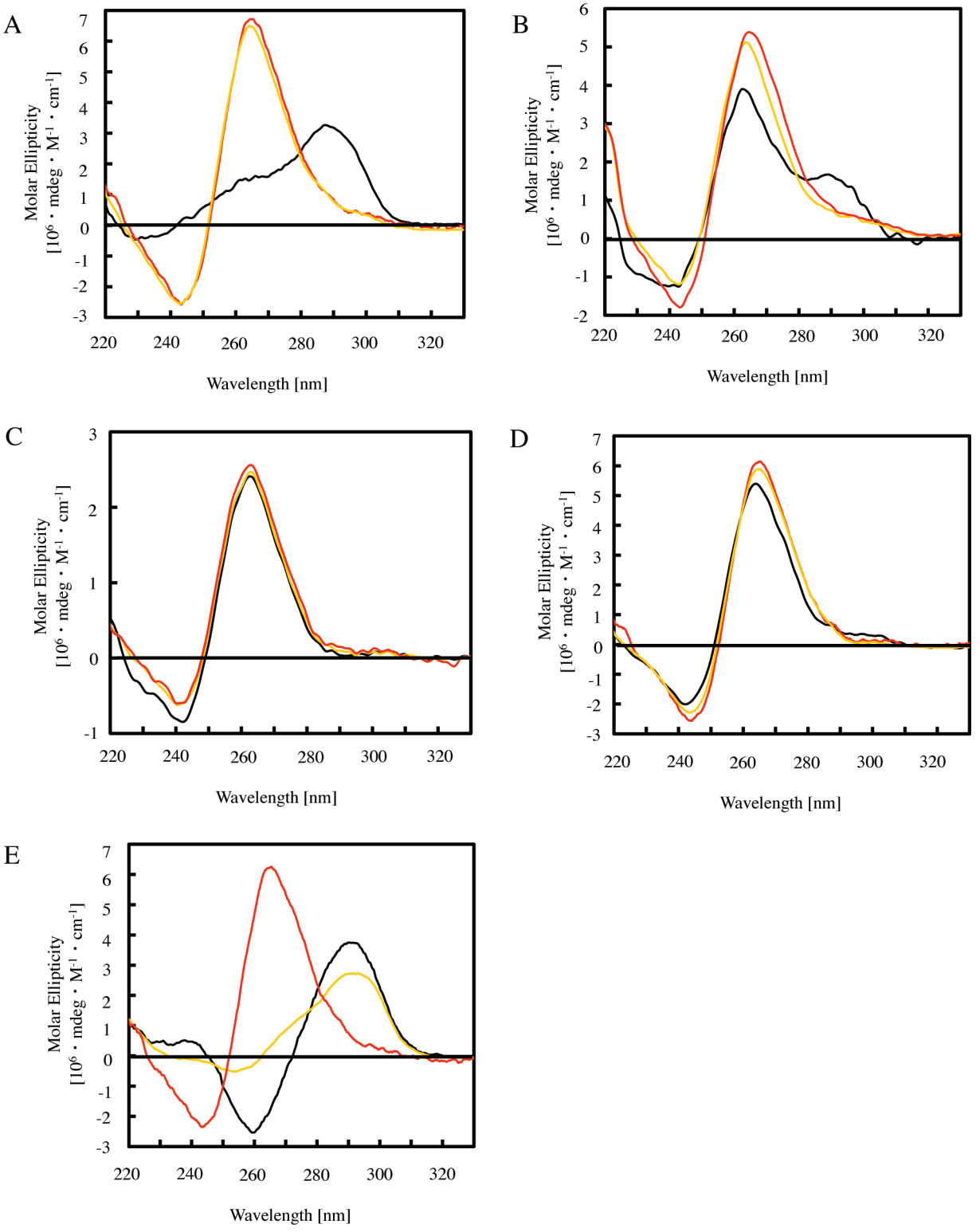


Figure 4. Circular dichroism of the G-quadruplex DNAs in the presence of RGG3. Titration of L333 (A), *bcl-2* (B), *c-kit* (C), L414 (D), and L444 (E) with RGG3 [2 (red lines), 1 (orange lines), and 0 (black lines) equiv of RGG3] in 100 mM KCl and 50 mM Tris-HCl (pH 7.5) at 20 °C. The base DNA concentration was 50 μM.

the other arginine-rich regions of amino acids 617−631 (PGGPPGPLMEQMGGR) and 638−656 (PGKMDKGEHRQERRDRPY) in RGG3 are required for specific binding of G-quadruplex DNA. In addition, simultaneous substitution of the first two phenylalanines of RGG3 with alanine reduced the level of G-quadruplex DNA binding, and substitution of the first three phenylalanines eliminated almost all G-quadruplex DNA binding (Figure 7S of the Supporting Information). The

dx.doi.org/10.1021/bi2003857 |*Biochemistry* 2011, 50, 5369–5378

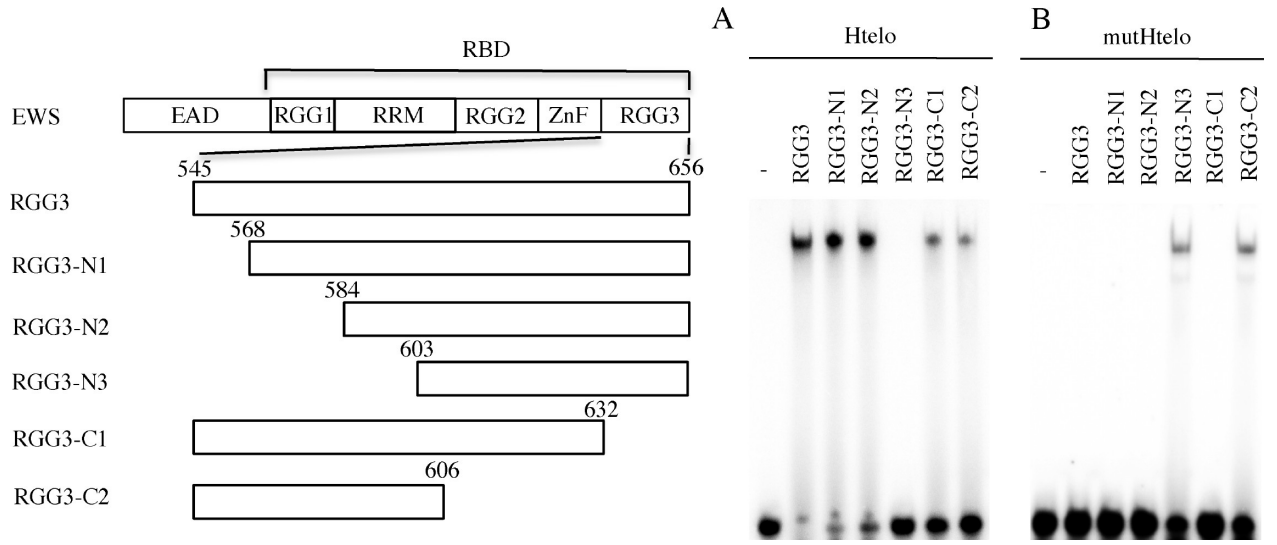


Figure 5. G-Quadruplex DNA binding specificities of RGG3 and deletion mutants of RGG3. An EMSA was performed with ^{32}P-labeled Htelo (A) or mutHtelo (B) and RGG3 (lane 2, residues 545−656), RGG3-N1 (lane 3, residues 568−656), RGG3-N2 (lane 4, residues 584−656), RGG3-N3 (lane 5, residues 603−656), RGG3-C1 (lane 6, residues 545−632), or RGG3-C2 (lane 7, residues 545−606). The DNA−protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

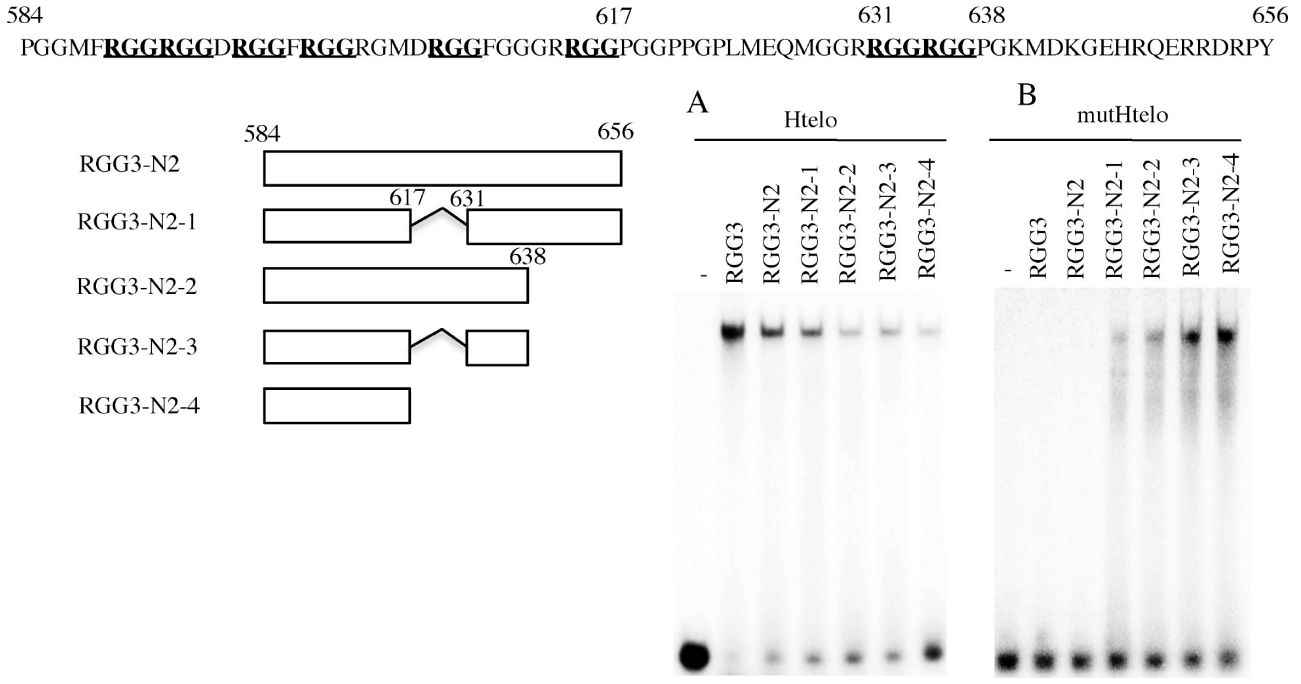


Figure 6. G-Quadruplex DNA binding specificities of RGG3 and deletion mutants of the RGG3-N2. Amino acid sequences of RGG3-N2 are described. RGGs are underlined. An EMSA was performed with ^{32}P-labeled Htelo (A) or mutHtelo (B) and RGG3 (lane 2), RGG3-N2 (lane 3), RGG3-N2-1 (lane 4, residues 584−656, Δ617−631), RGG3-N2-2 (lane 5, residues 584−638), RGG3-N2-3 (lane 6, residues 584−638, Δ617−631), and RGG3-N2-4 (lane 7, residues 584−617). The DNA−protein complexes were resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

aromatic residue of the peptide is reported to be important for specific G-quadruplex binding.[38] These findings indicate that the phenylalanines between amino acids 587 and 610 within RGG3 are important for the binding of RGG3 to G-quadruplex DNA.

In conclusion, RGG3 of EWS appears to bind preferentially to G-quadruplexes with longer loops. Moreover, RGG3 of EWS recognizes the phosphate backbone of the loop in the G-quadruplex DNA. The RGG, the other regions of amino acids 617−656, and phenylalanines in the range of amino acids 587−610 within RGG3 are important for RGG3 binding to G-quadruplex DNA. These findings might be useful in the development of molecules that selectively target different G-quadruplex DNA structures. On the basis of a combination of *in silico* and experimental approaches, Chowdhury and colleagues reported an enriched sequence with the potential to adopt the G-quadruplex motifs near transcription start sites.[39,40] These

findings suggest that G-quadruplex motif-mediated regulation is a more common mode of transcription control then previously considered. Because the protein can be easily expressed in the cells by transfection of its plasmid DNA, G-quadruplex binding protein would be useful for regulating gene expression of different G-quadruplex structures of promoters. On the other hand, EWS has been known as a very potent transcription activator depending on the EWS N-terminal domain, which binds directly to the RNA polymerase II subunit.[41] However, the function of EWS remains poorly characterized because little is known about its nucleic acid recognition specificity. The findings in this study might be useful for the identification on the role of EWS and the possible function of such G-quadruplex structure in genomic DNA.

■ ASSOCIATED CONTENT

Supporting Information. CD spectra of L414, L313, L212, and L111 (Figure 1S), CD spectra of L141, L131, L121, and L111 (Figure 2S), CD spectra of d_3Htelo and d_9Htelo (Figure 3S), DNA binding affinity of RGG3 for Htelo and L444 fold in the presence of Na^+ and CD spectra of Htelo and L444 in the presence of RGG3 (Figure 4S), effect of RGG3 on Htelo G-quadruplex structure (Figure 5S), DNA polymerase arrest assays of RGG3-N2 (Figure 6S), and effect of RGG3 phenylalanines on G-quadruplex DNA binding (Figure 7S). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Faculty of Science, Department of Chemistry, Graduate School of Science, Shizuoka University, 836 Oya, Suruga, Shizuoka 422-8529, Japan. E-mail: stohyos@ipc.shizuoka.ac.jp. Telephone: +81-54-238-4760. Fax: +81-54-237-3384.

Funding Sources

This research was supported by the Sasakawa Scientific Research Grant from The Japan Science Society and a Grant-in-Aid for Young Scientists (B) (2008, 20750130) from the Ministry of Education, Science, Sports, and Culture of Japan.

■ ABBREVIATIONS

DTT, dithiothreitol; EWSA, electrophoretic mobility shift assay; EWS, Ewing's sarcoma protein; GST, glutathione *S*-transferase; RGG, Arg-Gly-Gly repeat.

■ REFERENCES

(1) Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11593–11598.

(2) Riou, J. F., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Megnin-Chanet, F., Helene, C., and Mergny, J. L. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Natl. Acad. Sci. U.S.A. 99*, 2672–2677.

(3) Shammas, M. A., Reis, R. J. S., Akiyama, M., Koley, H., Chauhan, D., Hideshima, T., Goyal, R. K., Hurley, L. H., Anderson, K. C., and Munshi, N. C. (2003) Telomerase inhibition and cell growth arrest by G-quadruplex interactive agent in multiple myeloma. *Mol. Cancer Ther. 2*, 825–833.

(4) Waller, Z. A. E., Sewitz, S. A., Hsu, S.-T. D., and Balasubramanian, S. (2009) A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J. Am. Chem. Soc. 131*, 12628–12633.

(5) McLuckie, K. I. E., Waller, Z. A. E., Sanders, D. A., Alves, D., Rodriguez, R., Dash, J., McKenzie, G. J., Venkitaraman, A. R., and Balasubramanian, S. (2011) G-quadruplex-binding benzo[a]phenoxazines down-regulate c-kit expression in human gastric carcinoma cells. *J. Am. Chem. Soc. 133*, 2658–2663.

(6) Ou, T.-M., Lu, Y.-J., Zhang, C., Huang, Z.-S., Wang, X.-D., Tan, J.-H., Chen, Y., Ma, D.-L., Wong, K.-Y., Tang, J. C.-O., Chan, A. S.-C., and Gu, L.-Q. (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogenes c-myc by quindoline derivatives. *J. Med. Chem. 50*, 1465–1474.

(7) Wang, X.-D., Ou, T.-M., Lu, Y.-J., Li, Z., Xu, Z., Xi, C., Tan, J.-H., Huang, S.-L., An, L.-K., Li, D., Gu, L.-Q., and Huang, Z.-S. (2010) Turning off transcription of the bcl-2 gene by stabilizing the bcl-2 promoter qudruplex with quindoline derivatives. *J. Med. Chem. 53*, 4390–4398.

(8) Kang, H.-J., and Park, H.-J. (2009) Novel molecular mechanism for actinomycin D activity as an oncogenic promoter G-quadruplex binder. *Biochemistry 48*, 7392–7398.

(9) Qin, Y., Fortin, J. S., Tye, D., Gleason-Guzman, M., Brooks, T. A., and Hurley, L. H. (2010) Molecular cloning of the human platelet-derived growth factor receptor β (PDGFR-β) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-β expression. *Biochemistry 49*, 4208–4219.

(10) Bejugam, M., Gunaratnam, M., Muller, S., Sanders, D. A., Sewitz, S., Fletcher, J. A., Neidle, S., and Balasubramanian, S. (2010) Targeting the c-kit promoter G-qudruplexes with 6-substituted indenoisoquinolines. *ACS Med. Chem. Lett. 1*, 306–310.

(11) Cogoi, S., and Xodo, L. E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and ITS effect on transcription. *Nucleic Acids Res. 34*, 2536–2549.

(12) Bejugam, M., Sewitz, S., Shirude, P. S., Rodriguez, R., Shahid, R., and Balasubramanian, S. (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: Small molecule regulation of c-kit oncogenes expression. *J. Am. Chem. Soc. 129*, 12926–12927.

(13) Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006) 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. 128*, 1096–1098.

(14) Dai, J., Chen, D., Jones, R. A., Hurley, L. H., and Yang, D. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res. 34*, 5133–5144.

(15) Luu, K. N., Phan, A., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006) Structure of the human telomere in K^+ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc. 128*, 9963–9970.

(16) Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res. 34*, 2723–2735.

(17) Xu, Y., Noguchi, Y., and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)$_3$] in K^+ solution. *Bioorg. Med. Chem. 14*, 5584–5591.

(18) Palumbo, S. L., Memmott, R. M., Uribe, D. J., Krotova-Khan, Y., Hurley, L. H., and Ebbinghaus, S. W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res. 36*, 1755–1769.

(19) Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S., and Katahira, M. (2001) New quadruplex structure of GGA triplet repeat DNA: An intramolecular quadruplex composed of a G:G:G:G tetrad and G(:A):G(:A):G(:A):G heptad, and its dimerization. *Nucleic Acids Res. Suppl. 1*, 271–272.

(20) Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S., and Katahira, M. (2001) An intramolecular quadruplex of (GGA)$_4$ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol. 313*, 255–269.

(21) Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc. 129*, 4386–4392.

(22) Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc. 127*, 10584–10589.

(23) Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc. 126*, 8710–8716.

(24) Shi, D.-F., Wheelhouse, R. T., Sun, D., and Hurley, L. H. (2001) Quadruplex-interactive agents as telomerase inhibitors: Synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *J. Med. Chem. 44*, 4509–4523.

(25) Seenisamy, J., Bashyam, S., Gokhale, V., Vankayalapati, H., Sun, D., Siddiqui-Jain, A., Streiner, N., Shin-ya, K., White, E., Wilson, A. D., and Hurley, L. H. (2005) Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. *J. Am. Chem. Soc. 127*, 2944–2959.

(26) Sparapani, S., Haider, S. M., Doria, F., Gunaratnam, M., and Neidle, S. (2010) Rational design of acridine-dased ligands with selectivity for heman telomeric quadruplexes. *J. Am. Chem. Soc. 132*, 12263–12272.

(27) Patel, S. D., Isalan, M., Gavory, G., Ladame, S., Choo, Y., and Balasubramanian, S. (2004) Inhibition of human telomerease activity by an engineered zinc finger protein that G-quadruplexes. *Biochemistry 43*, 13452–13458.

(28) Takahama, K., Kino, K., Arai, S., Kurokawa, R., and Oyoshi, T. (2011) Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J. 278*, 988–998.

(29) May, W. A., and Denny, C. T. (1997) Biology of EWS/FLI and related fusion genes in Ewing's sarcoma and primitive neuroectodermal tumor. *Curr. Top. Microbiol. Immunol. 220*, 143–150.

(30) Rauscher, F. J., III (1997) Chromosome translocation-mediated conversion of a tumor suppressor gene into a dominant oncogene: Fusion of EWS1 to WT1 in desmoplastic small round cell tumors. *Curr. Top. Microbiol. Immunol. 220*, 151–162.

(31) Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc. 126*, 16405–16415.

(32) Rachwal, P. A., Findlow, S., Werner, J. M., Brown, T., and Fox, K. R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res. 35*, 4214–4222.

(33) Parkinson, G. N., Ghosh, R., and Neidle, S. (2007) Structure basis for binding of porphyrin to human telomeres. *Biochemistry 46*, 2390–2397.

(34) Campbell, N. H., Patel, M., Tofa, A. B., Ghosh, R., Parkinson, G. N., and Neidle, S. (2009) Selectivity in ligand recognition of G-quadruplex loops. *Biochemistry 48*, 1675–1680.

(35) Gavathiotis, E., Heald, R. A., Stevens, M. F. G., and Searle, M. S. (2001) Recognition and stabilization of quadruplex DNA by a potent new telomerase inhibitor: NMR studies of the 2:1 complex of a pentacyclic methylacridinium cation with $d(TTAGGGT)_4$. *Angew. Chem., Int. Ed. 24*, 4749–4754.

(36) Arora, A., and Maiti, S. (2009) Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. *J. Phys. Chem. B 113*, 8784–8792.

(37) Wang, Y., and Patel, D. J. (1993) Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure 1*, 263–282.

(38) Schouten, J. A., Ladame, S., Mason, S. J., Cooper, M. A., and Balasubramanian, S. (2003) G-quadruplex-specific peptide-hemicyanine ligands by partial combinatorial selection. *J. Am. Chem. Soc. 125*, 5594–5595.

(39) Verma, A., Yadav, V. K., Basundra, R., Kumar, A., and Chowdhury, S. (2009) Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cell. *Nucleic Acids Res. 37*, 4194–4204.

(40) Verma, A., Halder, K., Halder, R., Yadav, V. K., Rawal, P., Thakur, R. K., Mohd, F., Sharman, A., and Chowdhury, S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem. 51*, 5641–5649.

(41) Petermann, R., Mossier, B. M., Aryee, D. N., Khazak, V., Golemis, E. A., and Kover, H. (1998) Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II. *Oncogene 17*, 603–610.