

# A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity

SunMi L. Palumbo<sup>1</sup>, Regan M. Memmott<sup>1</sup>, Diana J. Uribe<sup>1</sup>, Yulia Krotova-Khan<sup>1</sup>, Laurence H. Hurley<sup>1,2,3</sup> and Scot W. Ebbinghaus<sup>1,\*</sup>

<sup>1</sup>Arizona Cancer Center, University of Arizona, 1515 N. Campbell Ave., Tucson, AZ 85724-5024, <sup>2</sup>College of Pharmacy, University of Arizona, 1295 N. Martin Ave., Tucson, AZ 85721 and <sup>3</sup>BIO5 Institute, University of Arizona, 1657 E. Helen St., Tucson, AZ 85721, USA

Received May 31, 2007; Revised and Accepted November 13, 2007

## ABSTRACT

The c-myb promoter contains multiple GGA repeats beginning 17 bp downstream of the transcription initiation site. GGA repeats have been previously shown to form unusual DNA structures in solution. Results from chemical footprinting, circular dichroism and RNA and DNA polymerase arrest assays on oligonucleotides representing the GGA repeat region of the c-myb promoter demonstrate that the element is able to form tetrad:heptad:heptad:tetrad (T:H:H:T) G-quadruplex structures by stacking two tetrad:heptad G-quadruplexes formed by two of the three (GGA)<sub>4</sub> repeats. Deletion of one or two (GGA)<sub>4</sub> motifs destabilizes this secondary structure and increases c-myb promoter activity, indicating that the G-quadruplexes formed in the c-myb GGA repeat region may act as a negative regulator of the c-myb promoter. Complete deletion of the c-myb GGA repeat region abolishes c-myb promoter activity, indicating dual roles of the c-myb GGA repeat element as both a transcriptional repressor and an activator. Furthermore, we demonstrated that Myc-associated zinc finger protein (MAZ) represses c-myb promoter activity and binds to the c-myb T:H:H:T G-quadruplexes. Our findings show that the T:H:H:T G-quadruplex-forming region in the c-myb promoter is a critical *cis*-acting element and may repress c-myb promoter activity through MAZ interaction with G-quadruplexes in the c-myb promoter.

## INTRODUCTION

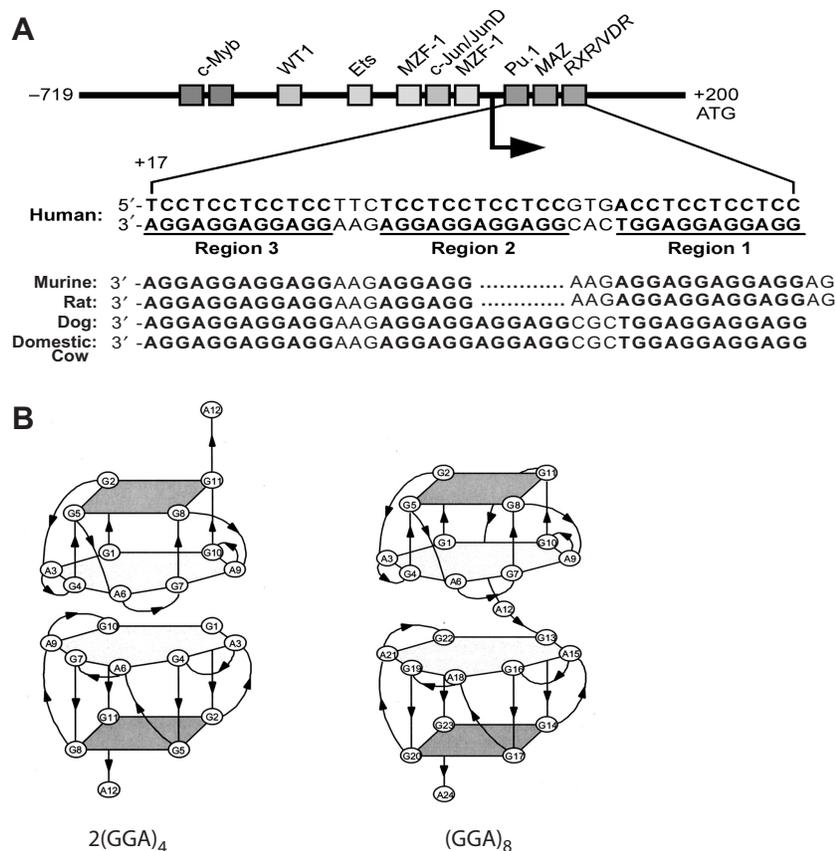
The c-myb proto-oncogene, the cellular homologue of the transforming v-myb oncogene of avian leukemia viruses,

encodes a critical transcription factor for proliferation, differentiation and survival of haematopoietic progenitor cells (1). High levels of the gene product c-Myb prevent haematopoietic stem cells from both differentiation and apoptosis (2,3). Because of the critical role of c-Myb in determining cell fate, c-Myb expression levels are tightly controlled in normal cells, showing high levels in immature, proliferating haematopoietic cells and undetectable levels in differentiated cells (4). c-Myb is over expressed in many leukemias and some solid tumors, and plays a critical role in leukemogenesis by maintaining cells in a proliferative state and by preventing terminal differentiation (5–7). Luger and coworkers showed that an antisense oligonucleotide (ODN) against c-myb mRNA could eliminate leukemia cells as a bone marrow purging agent, resulting in cytogenetic remissions in CML and showing that c-myb is a potential therapeutic target for leukemia treatment (8).

Cellular levels of c-myb are regulated at the transcriptional level by several mechanisms. One important mechanism is blocking mRNA elongation in intron I, which attenuates c-myb mRNA elongation in a cell type-dependent manner (9–12). c-Myb transcription is also controlled by a number of transcription factors. For example, c-Myb acts as a negative regulator of its own expression in a lineage-dependent manner (13). WT1 (Wilms Tumor 1), MZF1 (myeloid zinc finger 1) and PU.1 also down regulate c-myb promoter activity, and the Ets and c-Jun/JunD transcription factors activate the c-myb promoter in a lineage-specific manner (Figure 1A) (14–18).

The c-myb promoter contains a purine-rich region with three copies of four GGA repeats, [3(GGA)<sub>4</sub>], located 17 bps downstream of the transcription initiation site on the bottom strand, and the GGA repeat region is highly conserved in mammals. Nuclear magnetic resonance (NMR) and molecular modeling studies showed that (GGA)<sub>4</sub> DNA sequences similar to those found in the human c-myb promoter can form an unusual secondary

\*To whom correspondence should be addressed. Tel: +1-267-305-1279; Fax: +267-305-6428; Email: scot\_ebbinghaus@merck.com



**Figure 1.** The *c-myc* promoter and key transcription factors that regulate *c-myc* expression and T:H:H:T G-quadruplexes formed by 2(GGA)<sub>4</sub> or (GGA)<sub>8</sub> (19–22). **(A)** Transcription initiation site (+1) is indicated with an arrow. Inhibitors: *c-Myb* (–596~–618), WT1 (–390~–428), myeloid zinc finger 1 (MZF-1, –50~–64 and –162~–178) and PU.1 binding site (+15~+20). Activators: Ets (–251~–270) and *c-Jun/JunD* (–138~–162). Putative transcription factors that bind to the GGA repeat region include MAZ, RXR and VDR, which have multiple binding sites in the GGA repeat region. The myb GGA repeat sequences in different species are compared, and R1 and R3 are most conserved in the five different species. **(B)** *Left:* (GGA)<sub>4</sub> forms a T:H and two T:Hs intermolecularly dimerize to form a T:H:H:T G-quadruplex (taken from reference 19). *Right:* (GGA)<sub>8</sub> forms two T:Hs that dimerize intramolecularly, resulting in a T:H:H:T G-quadruplex (modified from reference 21).

DNA structure related to guanine quadruplexes and composed of a guanine tetrad (T) stacked onto a guanine-adenine heptad (H) (Figure 1B) (19,20). This unique G-quadruplex was the first identified G-quadruplex structure with pseudo-double chain reversal loops in which the loop base (A) is part of the heptad. DNA sequences with two adjacent (GGA)<sub>4</sub> units, (GGA)<sub>8</sub>, can form a very stable ( $T_m = 86^\circ\text{C}$ ) higher ordered structure by intramolecularly stacking two T:H G-quadruplexes on the heptad plane, resulting in a tetrad:heptad:heptad:tetrad (T:H:H:T) DNA structure (Figure 1B) (21,22). These structures form in the presence of physiologic concentrations of potassium ions and at neutral pH.

GGA repeats found in some promoters are shown to be a critical *cis* element for protein:DNA interaction. The eight GGA repeats (GGA)<sub>8</sub> found in the avian SPARC promoter were shown to be a critical positive regulatory region with multiple Sp1/Sp3 binding sites (23). In addition, the human homologue of SPARC, the BM-40 gene, also contains GGA repeats and the region is required to exert promoter activity of the BM-40 gene (24). These studies show that GGA repeats in gene promoters can play critical roles in regulating gene

expression, and suggest that the formation of secondary DNA structures may be important to the function of *cis*-acting GGA repeats.

In the present study we investigated if the *c-myc* GGA repeat region can form T:H:H:T G-quadruplexes and if the region is a critical regulator of promoter activity. We report that the *c-myc* GGA repeat region forms T:H:H:T G-quadruplexes involving two (GGA)<sub>4</sub> regions and acts as a repressor element of the *c-myc* promoter. However, the GGA repeat region is also essential for basal promoter activity of the *c-myc* gene. We identified a transcription factor, Myc-associated zinc finger protein (MAZ), that binds to the double-stranded and G-quadruplex conformations of the GGA repeat region and represses *c-myc* promoter activity. Our findings suggest that the *c-myc* G-quadruplexes act as a negative regulator of the promoter and that MAZ may repress *c-myc* promoter activity by binding to the *c-myc* T:H:H:T G-quadruplexes.

## MATERIALS AND METHODS

### Oligonucleotides

The oligonucleotide (ODN) containing the pyrazolopyrimidine analog of guanine (PPG) was synthesized by

**Table 1.** Summary of ODNs used in this study

		DMS protection	CD spectra	EMSA	DNA pol arrest	RNA pol arrest
c-Myb GGA repeat region	Wildtype	I			XI	XXI
	R1 mut	II			XII	
	R2 mut	III			XIII	
	R3 mut	IV			XIV	
	(GGA) <sub>8</sub> control		V	V	XVII	
	R1 and 2		VI	VI		
	PPG R1 and 2		VII	VII		
	R1 duplex			VIII		
	R2 duplex			IX		
	R3 duplex			X		
	R2 and 3				XV	XXV
	R3				XVI	XXVI

For sequence details, see Supplementary Data.

Bio-Synthesis, Inc. (Lewisville, TX, USA). All other ODNs were synthesized by Operon Biotechnologies, Inc. (Huntsville, AL, USA). ODNs were gel purified and the concentrations were determined by using a spectrophotometer. A summary and sequences of the ODNs are shown in Table 1 and Supplementary Data, respectively.

#### Circular dichroism spectroscopy

ODNs for circular dichroism (CD) were prepared at 5 or 1.25  $\mu$ M in 50 mM Tris-HCl (pH 7.5) in the presence or absence of 140 mM KCl. Samples were denatured at 95°C and slowly cooled to room temperature. CD spectra were measured by a Jasco-810 spectropolarimeter (Jasco, Easton, MD, USA) using a quartz cell of 1-mm optical path length, an instrument scanning speed of 100 nm/min, with a response of 1 s, over a range of 200–350 nm. A set of three scans was averaged for each sample at 25°C.

#### Dimethylsulfate protection assay

ODNs were 5'-end labeled, diluted to 100 nM in 40  $\mu$ l of 10 mM Tris-HCl (pH 8.0), denatured at 95°C, and slowly cooled down to 37°C in the presence or absence of 140 mM KCl. An amount of 1  $\mu$ g of dIdC was added to each sample, and the sample was treated with 0.5% dimethylsulfate (DMS) as a final concentration for 5 min at RT. The reaction was quenched by adding 50  $\mu$ l of a stop solution (1.5 M sodium acetate (pH 7.0), 1 M  $\beta$ -mercaptoethanol, 250  $\mu$ g/ml yeast tRNA), and the DMS-treated ODNs were cleaved by 10% (v/v) piperidine, and separated on a 10% sequencing gel.

#### Polymerase arrest assays

DNA-polymerase arrest assays were performed as previously described (25). Briefly, a primer was annealed to the template bearing the sequence of interest and gel purified. The asymmetric primer-template duplex was incubated in the absence or presence of 140 mM KCl at 37°C overnight, and primer extension was conducted with Taq polymerase (Fermentas, Hanover, MD, USA) at 47°C for 20 min.

RNA polymerase arrest assays were conducted by using the *in vitro* transcription kit T7 MEGashortscript from

Ambion (Austin, TX, USA) according to the manufacturer's protocol. The top sequence of T7 promoter was annealed to template ODNs bearing the bottom sequence of T7 promoter followed by the c-myb GGA repeat region. *In vitro* transcription from the T7 promoter site was conducted in the presence of rNTPs and <sup>32</sup>P alpha UTP for 2 h at 37°C. The reaction was stopped by digesting DNA templates with DNase I, and the transcription products were separated on a 12% denaturing gel.

#### Cell lines

The T-cell acute lymphoblastic leukemia cell line CCRF-CEM was purchased from ATCC (Manassas, VA, USA); the chronic myelogenous leukemia (CML) cell line K562 was purchased from Coriell Cell Repositories (Camden, NJ, USA); and the genetically modified human embryonic kidney cell line Flp-In<sup>TM</sup> 293 was purchased from Invitrogen (Carlsbad, CA, USA). Cell lines were maintained as recommended by the suppliers.

#### Site-directed mutagenesis

The wild-type c-myb reporter plasmid pMybWT was constructed by subcloning the 900 bp c-myb promoter in the plasmid LB178 into a pGL3 basic vector (Promega, Madison, WI, USA). All the mutant c-myb reporter plasmids, except for pMybDelR2, were generated by using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's protocol. For pMybDelR2, mutagenic PCR amplified the reporter plasmid pMybWT excluding Region 2 GGA repeats in the promoter, and then the PCR product was ligated in order to circularize it. The c-myb GGA repeat region and adjacent flanking region in the reporter plasmids were sequenced in order to confirm that there were no unintended mutations introduced during mutagenesis.

#### Generation of Flp-In 293 MybWT luciferase reporter stable cells

The c-myb reporter plasmid pFRT/MybWT/Luc for stable transfection was generated by cloning the BgIII/XbaI fragment containing the wild-type myb promoter

and luciferase coding sequence from the pMybWT construct at the BglII and SpeI sites of a pcDNA5/FRT vector, in which the CMV promoter of the vector has been deleted (Invitrogen, Carlsbad, CA, USA). The sequence of the insert was confirmed by sequencing.

The Flp-In 293 MybWT reporter stable cell line was generated by following the manufacturer's protocol. Briefly, the pFRT/MybWT/Luc construct and pOG44 at a ratio of 1:9 were transfected into the Flp-in 293 cells (Invitrogen, Carlsbad, CA, USA) by using ExGen500 (Fermentas, Glen Burnie, MD, USA) according to the company's protocol. Transfected cells were selected and maintained with hygromycin at 150 µg/ml in DMEM (Cellgro, Herndon, VA, USA). After the selection, stable transfectants were pooled and expanded for use in experiments. Successful integration of the reporter construct at the flp recombinase target was inferred from the phenotype of the cells (gain of hygromycin resistance, loss of Zeocin<sup>®</sup> resistance and gain of luciferase expression).

#### Transient transfection and luciferase assay

CCRF-CEM cells in exponential growth phase were plated at a density of  $6 \times 10^5$  cells/well of a 24-well plate in 500 µl OPTI MEM (Invitrogen, Carlsbad, CA, USA). Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) was used as a transfection reagent, and transient transfection was conducted according to the manufacturer's protocol, using 10 ng of the renilla luciferase reporter plasmid pRL SV40 and 1 µg of the c-myb reporter plasmids. Firefly and renilla luciferase activities were measured by using a dual luciferase reporter assay system (Promega, Madison, WI, USA) 24 h after transfection. Relative luciferase activities were acquired by normalizing the ratio of the firefly luciferase activity to the renilla luciferase activity of the mutant construct with that of the wild-type construct. Each transfection of the c-myb report plasmids was done in duplicate, repeated at least three times. Flp-In 293 stable cells were plated at 50% confluency in a complete medium, and 0–1.2 µg of the FLAG-tagged-MAZ expression plasmid BRB112 were transfected by using Lipofectamine 2000. Twenty-four hours after transfection, cells were lysed and luciferase activities were measured. Since luciferases in Flp-In 293 stable cells are stably expressed, total protein amount in the lysate was used to normalize luciferase activity of the samples. Each transfection of the c-myb reporter plasmids was done in triplicate, repeated at least three times. The data are presented as the average relative luciferase activity compared to the wild-type c-myb promoter from six luciferase assays per plasmid.

#### Nuclear extract preparation

K562 cells were harvested, resuspended in RSB buffer (10 mM NaCl, 3 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 7.5, 0.05% NP-40 and protease inhibitors), vortexed, incubated on ice and spun down. The pellet was washed once with RSB buffer and resuspended in RIPA buffer (50 mM Tris-HCl (pH 7.5), 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA and protease inhibitors). Then, samples were vortexed, incubated on ice and

spun down. The supernatant containing nuclear proteins was stored at  $-80^{\circ}\text{C}$ .

#### Expression of FLAG-MAZ

FLAG-tagged-MAZ was expressed from the plasmid BRB112 in BL21 (DE3) pLysS cells (Invitrogen, Cedar Creek, TX, USA) according to the manufacturer's protocol. Briefly, BRB112 was transformed into BL21 (DE3) pLysS cells. Overnight culture was prepared from a single colony and diluted to 1:100 in fresh Luria-Bertani broth with 0.4% glucose. Pre- and post-induction incubations were 2 h each at 37°C, and 1 mM of IPTG was used for induction of FLAG-MAZ expression. After induction, bacterial cells were harvested and lysed, and FLAG-MAZ was purified by a FLAG affinity column according to the manufacturer's protocol (Sigma-Aldrich, St Louis, MO, USA). MAZ was also synthesized by using the TnT<sup>®</sup> coupled Reticulocyte lysate systems according to the manufacturer's protocol (Promega, Madison, WI, USA). Briefly, 2 µg of the plasmid BRB112 was used as a DNA template in a final volume of 50 µl, and coupled transcription/translation reaction was conducted at 30°C for 90 min.

#### Electrophoretic mobility shift assay (EMSA) and competition EMSA

For EMSA with purified FLAG-MAZ, 1 µg of purified FLAG-MAZ was incubated in a binding buffer (10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM dithiothreitol (DDT), 1 mM EDTA (pH 8.0) and 5% glycerol) in the presence of 0.2 pmol of labeled, double-stranded probes and incubated for 20 min at room temperature. For EMSA with MAZ synthesized by the TnT<sup>®</sup> coupled Reticulocyte lysate system, the previously described methods in (26) were followed. A total of 10 µl of reticulocyte lysate containing newly synthesized MAZ was used per sample in the presence of the labeled probes at 12.5 nM. For competition EMSAs, 5 µg of K562 nuclear extract was incubated with 0, 5 and 20 pmol of double-stranded cold competitors in the binding buffer with 1 µg of dIdC for 20 min at room temperature, and then 0.2 pmol of the labeled probes were added to the reaction and incubated for 20 min at room temperature. Probe-protein complexes were separated on a 4% non-denaturing gel with 2.4% glycerol in 0.25X TBE at 4°C.

#### DNase I protection assay

A 260 bp GGA repeat region probe was prepared by digesting pMybWT with BanII and SfcI and labeling by Klenow filling. Approximately 18 000 cpm of the probe was incubated in a binding buffer (25 mM Tris-HCl (pH 8.0), 50 mM KCl, 6.25 mM MgCl<sub>2</sub>, 0.5 mM EDTA (pH 8.0), 10% glycerol, 0.5 mM DTT) for 10 min on ice in the presence of FLAG peptides, or FLAG-MAZ. After adding CaCl<sub>2</sub> and MgCl<sub>2</sub> to 2.5 mM and 5 mM as final concentrations, respectively, 0.01u of DNase I (Invitrogen, Carlsbad, CA, USA) was added to each sample and incubated 2 min at room temperature. DNase I reaction was quenched by adding a stop solution (100 mM NaCl, 15 mM EDTA, 0.5% SDS, 50 µg/ml

yeast tRNA as final concentrations). The labeled probes were recovered via phenol/chloroform extraction followed by ethanol precipitation and separated on a 6% denaturing gel.

### Statistical analysis

Microsoft Excel was used to perform statistical analysis. Unpaired, two-tailed *t*-tests were used to determine statistical differences in relative luciferase activities of the wild-type and mutant c-myb reporter constructs, and a value of  $P < 0.01$  was considered significant.

## RESULTS

### The c-myb GGA repeat region forms T:H:H:T G-quadruplexes

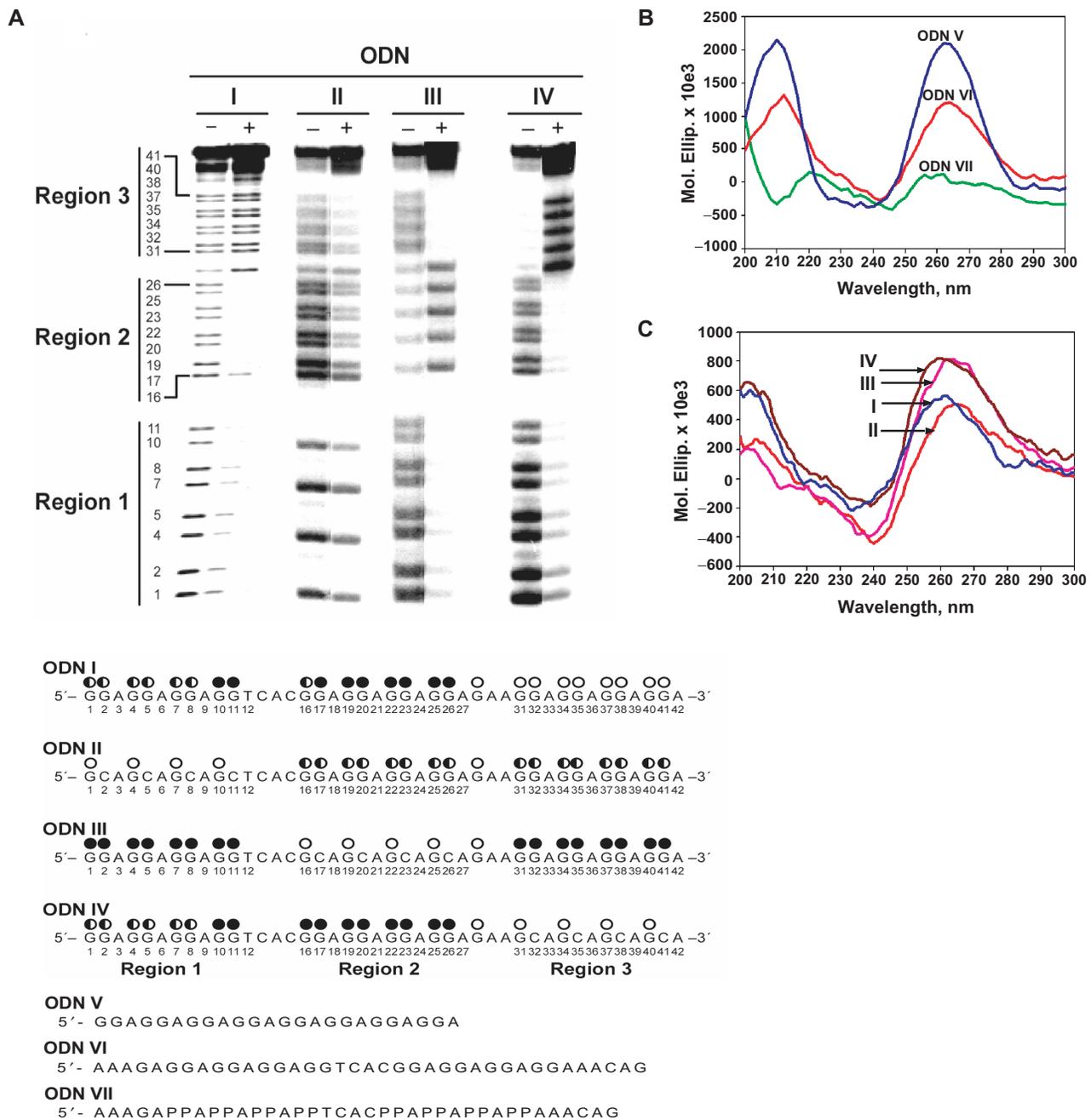
ODNs with the sequence (GGA)<sub>4</sub> have been shown to form a secondary structure composed of a guanine tetrad stacked on a guanine-adenine heptad. Two tetrad:heptad (T:H) G-quadruplexes can dimerize by stacking interactions on their heptad planes to form a T:H:H:T structure (Figure 1B, *left*, taken from reference 19) (19,20). This T:H:H:T G-quadruplex can be derived intramolecularly from a (GGA)<sub>8</sub> ODN and was shown to be very stable, producing a melting temperature of 86°C (22). In this structure, the two T:H G-quadruplex-forming core sequences, GGAGGAGGAGG, are linked by a single adenine nucleotide that does not form any hydrogen bonds to stabilize the T:H:H:T higher order structure (Figure 1B, *right*, modified from reference 21) (21,22).

Since the c-myb promoter contains three nearly perfect tandem (GGA)<sub>4</sub> repeats (Figure 1A), we speculated that each could be considered a T:H-forming unit and would form T:H:H:T structures with slightly longer linker sequences than in the previously characterized (GGA)<sub>8</sub> ODN (Figure 1B, *right*, modified from reference 21). We performed DMS footprinting studies with an ODN containing all three GGA repeat regions (ODN I) and ODNs in which each region was mutated to (GCA)<sub>4</sub> to prevent secondary structure formation (ODNs II, III and IV) (Figure 2A). These studies show potassium-dependent partial and complete DMS footprints in R1 and R2 in ODN I, respectively, and in the two unaltered GGA repeat regions of ODNs II, III and IV (Figure 2A). This pattern of DMS protection suggests that T:H:H:T structures are formed within the c-myb GGA repeats, and further suggests that any two of the three (GGA)<sub>4</sub> repeats, such as R2/R3 of ODN II, R1/R3 of ODN III and R1/R2 of ODN IV, can form the dimer structure, implying that there is a possible competition for dimerization among the three GGA repeat regions. Significantly, R1/R3 in ODN III and R1/R2 in ODNs I and IV form a strong T:H:H:T dimer and R2/R3 in ODN II forms a less stable T:H:H:T. Stronger protection of R1 in ODN I compared to ODN III is probably explained by competition for dimerization partners in ODN I, where three T:H:H:T structures (R1/R2, R1/R3 and R2/R3) can potentially form and probably exist in equilibrium. We found that the R1 and R3 GGA repeat sequences are

perfectly conserved across multiple mammalian species, including mice, rats, dogs, cows and humans (Figure 1A), suggesting that R1 and R3 are evolutionarily conserved to preserve a DNA secondary structure that controls myb expression.

CD has been used to deduce the relative orientation (parallel vs antiparallel) of DNA strands in G-quadruplexes (27), and most forms of G-quadruplexes produce very characteristic CD spectra. The CD spectrum of a T:H:H:T from a (GGA)<sub>8</sub> oligonucleotide (ODN V) is shown in Figure 2B, and essentially reproduces the CD signature for the T:H:H:T structure that was reported in correlation with the NMR studies from which the T:H:H:T was identified (22). ODN VI is an ODN representing the c-myb promoter and containing GGA repeat regions 1 and 2 with the flanking sequence and intervening sequence between the GGA repeats derived from the c-myb promoter. ODN VI produces the characteristic CD signature of a T:H:H:T with a strong peak at 264 nm in the presence of potassium, indicating T:H:H:T G-quadruplex formation by the ODN. In order to verify that the CD signature of ODN VI is due to G-quadruplex formation by the ODN, we tested a negative control oligonucleotide (ODN VII) with 7-deazaguanine analog substitutions to prevent G-quadruplex formation. PPG is a pyrazolopyrimidine analog of guanine in which the N7 and C8 atoms of guanine are interposed. N7-dependent secondary DNA structures, such as G-quadruplexes, are usually not formed, but the PPG retains the same electron density in the ring system as guanine, and PPG substitutions do not alter the optical activity (absorption of UV light) compared to the unfolded form of the guanine oligo. As shown in Figure 2B, the PPG-substituted oligonucleotide ODN VII did not produce the CD signature of a T:H:H:T, confirming that the CD signature of ODN VI is due to T:H:H:T formation. ODNs I through IV also produce CD spectra characteristic of a T:H:H:T G-quadruplex in a potassium-dependent manner (Figure 2C). The T:H:H:T stability inferred from the signature peaks agrees with the DMS footprinting results, showing the lowest signature peak on ODN II and high peaks on ODNs III and IV. When taken together with the other biochemical studies presented here, the CD spectra provide strong supporting evidence for the formation of a T:H:H:T structure by the GGA repeat regions of the c-myb promoter.

DNA and RNA polymerase stop assays have been used as a means to show the relative stability of DNA secondary structure formation in a DNA template strand and to deduce the location and stability of G-quadruplexes in single-stranded templates and plasmid DNA (28,29). Because the GGA repeat region of c-myb occurs on the transcribed strand downstream of the transcription start site, RNA polymerase arrest could occur in the c-myb promoter if a stable G-quadruplex structure was formed. Thus, both DNA and RNA polymerase arrest sites at these sequences are potentially relevant to the disruption of enzymatic function that could be encountered in cells by G-quadruplex formation in the c-myb GGA repeat region. The results show that DNA polymerase (Figure 3A and B) and RNA polymerase

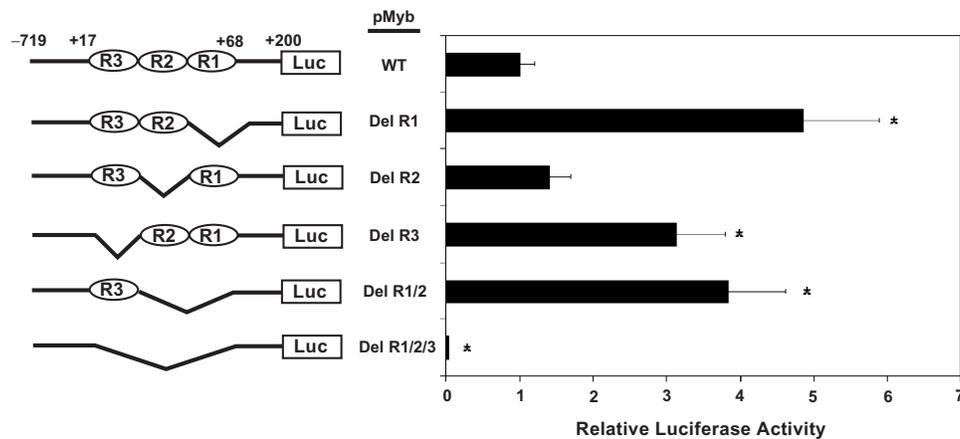


**Figure 2.** DMS protection assay and CD spectra. (A) DMS methylation patterns are indicated by open circles (methylated), half-shaded circles (partially methylated) and full-shaded circles (unmethylated). The wild-type GGA repeat sequence (ODN I) shows protection from DMS methylation in R1 and R2 in the presence of potassium. The (GGA)<sub>4</sub> to (GCA)<sub>4</sub> mutants (ODNs II, III and IV) show protection in the two intact (GGA)<sub>4</sub> repeats in the presence of potassium. Note that Ps in ODN VII means PPGs. (B) Circular dichroism spectra of ODNs V, VI and VII in the presence of KCl. ODNs V and VI show CD signature of G-quadruplex but not ODN VII. (C) CD spectra of ODNs I, II, III and IV. ODN II shows the lowest G-quadruplex signature peak and ODNs III and IV show a high signature peak.

(Figure 3C) arrest sites (designated S1 and S2) are observed precisely at the beginning of the third and second GGA repeat regions in a potassium-dependent manner. These results are explained by the polymerase arrest at S1 from T:H:H:T by the R2/R3 and/or R1/R3 regions and the arrest at S2 by the R1/R2 GGA repeat regions. The presence of two arrest sites indicates that the c-myc promoter can form a mixture of stable secondary structures due to the competition between GGA repeat

regions. In addition, our DMS footprinting data showing DMS cleavage in both R1 and R3 regions of ODN I (Figure 2A) support our conclusion that there is an equilibrium between R1/R2 and R2/R3 that results in polymerase arrest at S2 and S1, respectively. Furthermore, our findings imply that polymerase was not arrested by a T:H, because otherwise there would be an arrest site before R1. Significantly, in the full-length (FL) c-myc GGA repeat templates (ODNs XI and XXI), polymerase





**Figure 4.** Luciferase activity driven by the wild-type and GGA-deleted c-myc promoters in CCRF-CEM cells. Various c-myc promoters were transiently transfected into CCRF-CEM cells. Deletion of R1, R3, or both R1 and R2 from the c-myc promoter increases luciferase activity in CCRF-CEM cells. The R1, R2 and R3 deletion mutant pMybDelR1/2/3 markedly reduces luciferase activity. \* $P < 0.01$

arrest at the first arrest site (S1) is only partial, and the second arrest site (S2) is more pronounced, consistent with the data from the DMS protection assays that the T:H:H:T formed between GGA repeat regions 1 and 2 is the most preferred, when all three regions are intact.

When one of the GGA repeat regions is mutated (ODNs XII, XIII and XIV in Figure 3A), T:H:H:T formation occurs with the two intact GGA repeats and results in a single arrest site at either S1 or S2. This observation is comparable to the single arrest site (S) that is observed before the GGA repeats in a template containing the (GGA)<sub>8</sub> sequence (ODN XVII in Figure 3B) that was used as a positive control for the T:H:H:T structure (22), and indicates that any two (GGA)<sub>4</sub> repeats can form a T:H:H:T regardless of their positions in the c-myc GGA repeat region. Of note, although any two (GGA)<sub>4</sub> motifs are able to form a T:H:H:T, the relative amount of arrest product compared to FL primer extension products varies in ODNs XI, XII, XIII and XIV, which provides an estimate of the stability of the T:H:H:T structures formed by the GGA repeat sequences. For example, in Figure 3A, the arrest products at S1 and S2 are substantial compared to the amount of the FL primer extension product in ODN XI (representing the WT myb sequence), whereas the arrest product at S1 in ODN XII is only partial compared to the FL extension product. Our DMS protection assay data (Figure 2A) and the percentage of arrest products (Figure 3A) demonstrate that mutation of GGA repeat region 1 leads to a weak T:H:H:T structure formed between R2 and R3.

Deletion studies in DNA and RNA polymerase arrest assays demonstrated that dimerization of two T:H G-quadruplexes is required for stable quadruplex formation. When one of the GGA repeat regions is deleted (ODNs XV and XXV in Figure 3B and C), T:H:H:T formation takes place by dimerization of two T:H quadruplexes and causes a single arrest site in both DNA and RNA polymerase assays. However, when two GGA repeat regions are deleted (ODNs XVI and XXVI), a single T:H structure could form, and this structure partially arrests RNA polymerase (Figure 3C) but not

DNA polymerase (Figure 3B), suggesting that the T:H structure is not as stable as the T:H:H:T structure, in support of the concept that stacking interactions between the heptad planes of two T:H subunits lead to a more stable higher order T:H:H:T structure.

In summary, these biochemical assays demonstrate that two (GGA)<sub>4</sub> motifs from the c-myc promoter can form a stable potassium-dependent secondary DNA structure capable of arresting the progress of DNA and RNA polymerases and provide evidence for the formation of a T:H:H:T structure by the c-myc promoter.

#### The c-myc GGA repeat region plays a critical role in c-myc promoter activity

In order to investigate the role of the GGA repeat region in c-myc promoter activity, we performed luciferase assays using wild-type and mutant c-myc promoters. The reporter plasmid pMybWT bears the wild-type c-myc promoter from -719 to +200 with the three intact GGA repeat regions. The mutant reporter plasmids pMybDelR1, pMybDelR2 and pMybDelR3 are deletion mutants in which a (GGA)<sub>4</sub> motif has been individually deleted as shown in Figure 4. We also generated serial deletion mutant plasmids, pMybDelR1/2 and pMybDelR1/2/3, in which two or all three GGA repeat regions have been deleted. Figure 4 shows relative firefly luciferase activities in CCRF-CEM cells normalized by renilla luciferase activities in a dual luciferase assay system. Deletion of GGA repeat region 1 or both regions 1 and 2 in the c-myc promoter increased relative luciferase activities by 3- to 5-fold. In ODNs representing the wild-type c-myc sequence, R1 and R2 were the two GGA repeat regions involved in T:H:H:T formation, and deletion of R1 or R1/2 was shown to result in an unstable secondary structure with the remaining region(s). In contrast, deletion of R2 has no significant effect on c-myc expression. In ODNs, mutation of R2 resulted in a stable T:H:H:T between R1 and R3. Taken together, these results support the conclusion that stable T:H:H:T formation involving two GGA repeat regions represses c-myc expression. The deletion of region 3 also

significantly increased c-myb promoter activity, even though corresponding mutations in R3 in ODNs representing the c-myb promoter resulted in a stable T:H:H:T between R1/R2. The effect of deleting R3 may be explained by the newly created transcription factor binding site that is introduced as a result of deletion of R3 and then subsequent rejoining. We found such a transcription factor binding site in pMybDelR3, ILF1 (winged-helix transcription factor IL-2 enhancer binding factor), by using the computer program MatInspector from GenoMatix ([www.genomatix.de](http://www.genomatix.de)). ILF1 is constitutively expressed and serves as a transcriptional activator of the IL-2 promoter in the T-cell leukemia cell line Jurkat (30). Since the cell line CCRF-CEM used in this study is a T-cell leukemia cell line, it is possible that the ILF1 site created in the GGA repeat region of pMybDelR3 may act as a transcriptional activator, which would result in an increase of promoter activity of pMybDelR3. We also examined the sequence of pMybDelR1 and pMybDelR1/2 to see if there are any transcription sites created as a result of deletion of R1 or both R1 and R2. However, MatInspector did not find any newly introduced transcription sites due to the deletion. There are transcription factor binding sites that have been lost in the deletion mutant constructs pMybDelR1, pMybDelR2 and pMybDelR1/2. Nevertheless, considering luciferase activity of pMybDelR1/2/3, the loss of transcription factor binding sites does not seem to cause an increased promoter activity.

Although deletion of R1/2 and R3 individually increased c-myb promoter activity, the additional 15 bp deletion of region 3 from pMybDelR1/2, which results in complete removal of all the GGA repeats from the c-myb promoter, showed almost complete loss of luciferase activity in CCRF-CEM cells that ranges only 1–3% of the pMybWT construct. It is important to note that luciferase activity was still measurable (~50-fold over background) from the pMybDelR1/2/3 plasmids. A very low level of c-myb expression in response to deletion of the entire GGA repeat region was also observed in two other leukemia cell lines, Jurkat and K562 (data not shown). Because of the striking loss of promoter activity by deletion of the GGA repeats in three cell lines tested, another pMybDelR1/2/3 construct was tested to control for spurious effects from site-directed mutagenesis, and these results were confirmed with a second construct plasmid. At least one (GGA)<sub>4</sub> repeat region is required for c-myb promoter activity, and we speculate that duplication of the (GGA)<sub>4</sub> motif is used by the myb promoter to finely tune c-myb expression by equilibrium between secondary structure formation and transcription factor binding to double-stranded DNA. Collectively, these data show that the c-myb GGA repeats are critical for regulating c-myb expression and suggest that T:H:H:T secondary DNA structures formed by adjacent (GGA)<sub>4</sub> motifs are likely negative regulators of the c-myb promoter.

#### The transcription factor MAZ binds to the c-myb GGA repeat region

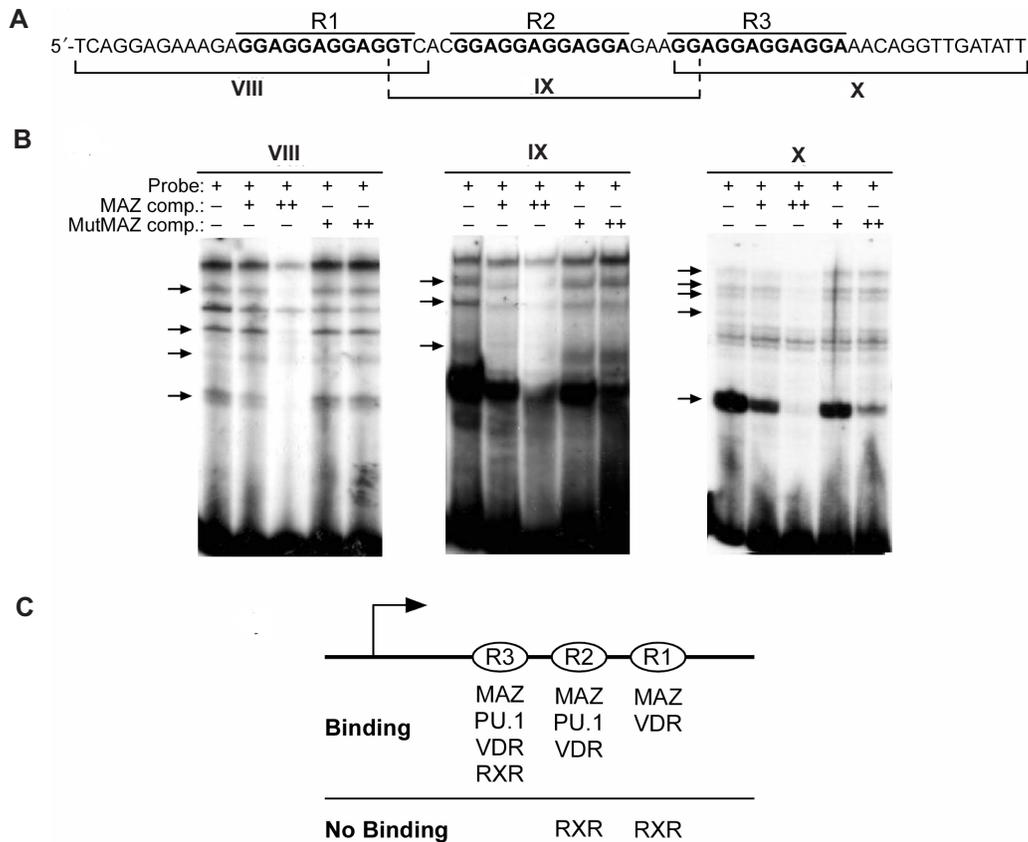
Since our data demonstrated that the c-myb GGA repeat region is a critical *cis* element for controlling c-myb

promoter activity, we investigated the possible protein:DNA interactions in the region. First, we searched for putative transcription factors for the c-myb GGA repeat region from +2 to +73 with MatInspector (Supplementary Data). MAZ, RXR, PU.1 and VDR were identified as transcription factors that have multiple potential binding sites in the GGA repeat region. We performed EMSAs to evaluate the binding of nuclear proteins from the leukemia cell line K562 to the myb GGA repeats in the presence of competitor oligos with the consensus binding sites for these transcription factors. Since the sequence of the GGA repeat region is repetitive, we divided the GGA repeat region into three regions (VIII, IX and X), each of which includes one of the (GGA)<sub>4</sub> repeat regions (R1, R2 and R3, respectively) for competition EMSAs (Figure 5B) and forms a stable duplex with a  $T_m > 64^\circ\text{C}$ . In the competition EMSA for investigating MAZ binding, we first incubated K562 cell nuclear extract with 0, 25X or 100X molar excess of a non-labeled double-stranded competitor bearing a consensus MAZ binding site or mutant MAZ binding site, and then the labeled double-stranded probes ODN VIII, IX and X, representing the three (GGA)<sub>4</sub> regions of the c-myb promoter, were added to the reaction. The competition EMSAs in Figure 5B demonstrate that the c-myb GGA duplex repeat region interacts with multiple nuclear proteins, but several of the protein:DNA complexes were eliminated by the MAZ consensus competitor but not the MAZ mutant competitor. We also used consensus binding sequences for PU.1, RXR and VDR in competition EMSAs. These data are summarized in Figure 5C, and the other competition EMSAs are shown as Supplementary Data. These data provide preliminary evidence that MAZ, PU.1, RXR and VDR have one or more binding sites in the c-myb GGA repeats.

To confirm MAZ binding to the c-myb promoter, we purified MAZ protein for EMSAs and DNase I footprinting studies. Figure 6A shows that MAZ binds to all three duplex probes representing the three GGA repeat regions of the c-myb promoter and the two MAZ-probe complexes indicates that MAZ may bind to the duplex probes as a monomer or a dimer. DNase I digestion was performed on a 260 bp fragment of the c-myb promoter in the presence of either a negative control FLAG peptide (Figure 6B, lane 2) or purified FLAG-MAZ (Figure 6B, lane 3). Only FLAG-MAZ showed protection from DNase I digestion spanning the GGA repeat region but not FLAG peptide, confirming our competition EMSA data that MAZ has multiple binding sites in the c-myb GGA repeats.

#### MAZ binds to the G-quadruplexes formed in the c-myb GGA repeat region and represses promoter activity

MAZ was reported to bind to the G-quadruplex formed by the G-rich region of the diabetes susceptibility locus *IDDM2* (26). To test if MAZ can also bind to the G-quadruplexes formed in the c-myb GGA repeat region, we performed EMSAs with ODNs folded into T:H:H:T structures in the presence of MAZ in rabbit reticulocyte lysate as described in (26). First, we tested ODN VI, which



**Figure 5.** Competition EMSA with K562 nuclear extract. (A) The DNA probes used for EMSAs are illustrated under the corresponding GGA repeat sequence. The probes are double stranded, but only the G-rich sequence is shown. (B) The results from competition EMSAs using a MAZ competitor are shown. The protein:DNA complexes completely or almost completely eliminated by the MAZ competitor are indicated with arrows. (C) Summary of competition EMSA results. The putative transcription factors shown to interact or not to interact with the c-myb GGA repeat region in competition EMSAs are listed.

contains c-myb GGA repeat regions 1 and 2 and should form a more stable T:H:H:T than ODN I due to the lack of competition for dimerization in ODN VI. ODN V (perfect (GGA)<sub>8</sub> sequence) and ODN VII (PPG-substituted analog of ODN VI) were studied as positive and negative controls, respectively. Figure 7A shows a MAZ protein:DNA interaction that is specific for the T:H:H:T structure. The band denoted as the T:H:H:T-MAZ complex is observed with ODNs V and VI, but not with ODN VII, the PPG-substituted control, and only in the presence of MAZ. Non-specific protein:DNA interactions, presumably from proteins in the reticulocyte lysate, are seen with all three ODNs in the presence or absence of MAZ. We also confirmed that MAZ binds to the G-quadruplex(es) formed by the native c-myb promoter with all three (GGA)<sub>4</sub> repeats (Figure 7B). This result confirms that MAZ binds to the equilibrium structure(s) formed when all three GGA repeat regions are present, but the free probe migrates more slowly than ODNs V, VI and VII because the ODN is longer (Supplementary Data).

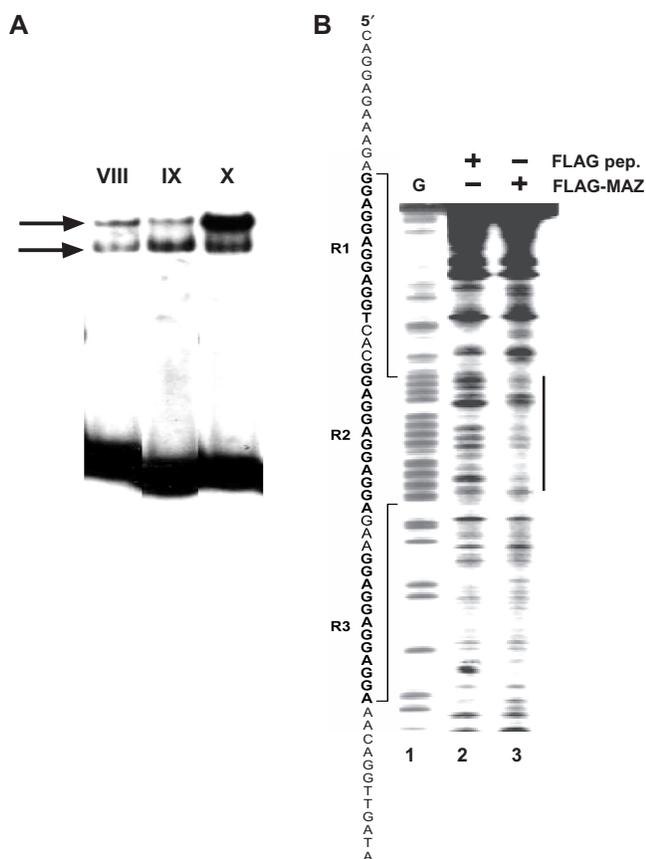
We investigated the role of MAZ in regulation of c-myb expression via transfection of the MAZ expression plasmid BRB112 in Flp-In 293 Myb WT reporter stable cells (Figure 8). These Flp-in 293 cells contain the stably transfected wild-type c-myb promoter and luciferase

coding sequence in the genome, which are originally from pMybWT (Figure 8). Forced overexpression of MAZ decreased the activity of the stably transfected myb promoter, demonstrating that MAZ is a repressor of the c-myb promoter. We also treated the same myb stable cells with a pool of three MAZ siRNAs to evaluate the effect of MAZ knockdown on c-myb promoter activity. However, siRNA treatment did not change c-myb promoter activity (data not shown). It is probably because many other transcription factors interact with and compete each other for binding to the c-myb GGA repeat region (Figure 5). Knocking down MAZ would open more opportunity for other competitor proteins to bind to the region, which may result in no net changes in c-myb promoter activity.

Collectively, this study presents the first evidence that the c-myb GGA repeat region forms T:H:H:T G-quadruplexes capable of binding to a transcriptional repressor protein. The GGA repeat region acts as a repressor of c-myb expression by folding into a G-quadruplex structure and recruiting transcriptional repressors such as MAZ, which are needed for basal transcription.

## DISCUSSION

In support of previous NMR studies (19–22), our findings from DMS protection, CD and polymerase arrest assays



**Figure 6.** Electrophoretic mobility shift and DNase I protection assays with FLAG-MAZ. (A) EMSA. The complexes of FLAG-MAZ and the c-myb probes are indicated with arrows. (B) DNase I protection assay. Lane 1: sequencing for guanines, lane 2: FLAG peptide and lane 3: FLAG-MAZ. The area protected from DNase I digestion is marked with a solid line.

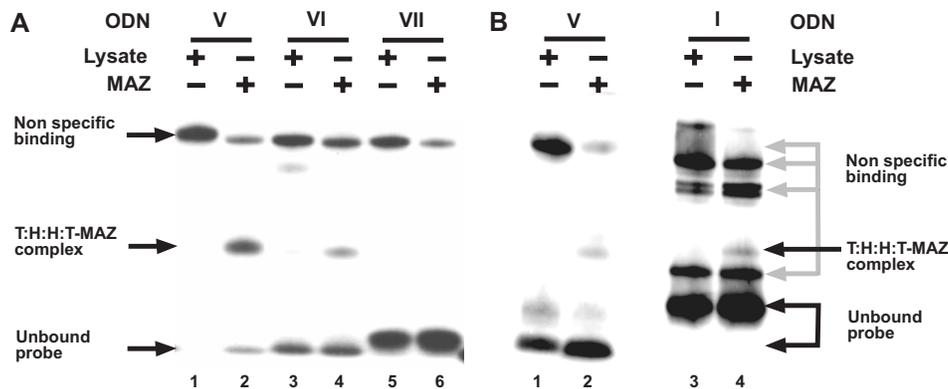
show that the GGA repeat region in the c-myb promoter is the first example of a genomic sequence that forms T:H:H:T G-quadruplex-like structures, involving two adjacent  $(GGA)_4$  repeats. Several G-quadruplex regions have been identified in promoters and shown to be critical for regulation of promoter activity. The G-quadruplexes formed in the c-myc and Kras promoters inhibit promoter activity (28,29), whereas the G-quadruplex formed in the human insulin gene activates the promoter (31), which suggests that these G-quadruplex structures can alter gene expression. Our promoter-reporter transfection studies showed that the c-myb G-quadruplex-forming region is critical for regulation of c-myb expression and that the deletion of  $(GGA)_4$  motifs that destabilize the G-quadruplex increase promoter activity. The GGA repeat region of the c-myb gene is highly conserved across different mammalian species (Figure 1A), and the R1 and R3 GGA repeat regions are absolutely conserved in all five mammalian genomes queried. This finding implies that R1 and R3 are critical for regulation of myb expression, and our finding that deletion of either R1 or R3 significantly alters c-myb transcription supports this implication. Conversely, GGA repeat region R2 is not

evolutionarily conserved, and deletion of this region has no effect on c-myb transcription in our studies. The c-myb GGA repeat region also contains numerous transcription factor binding sites in these three GGA repeat regions, and the c-myb promoter may regulate its activity by altering the stability of T:H:H:Ts via changing dimerization partners.

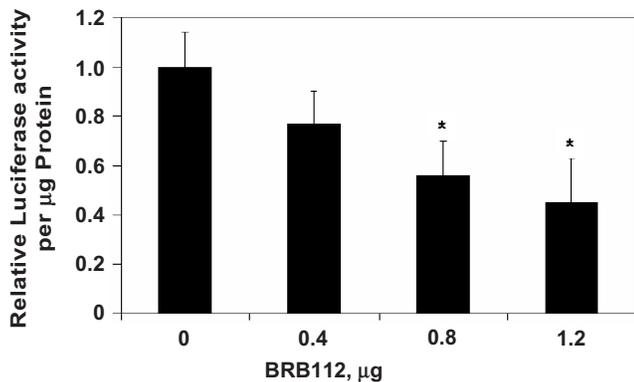
The mechanism by which G-quadruplexes can be transcriptional activators or inhibitors is not clear. There are many studies showing that promoter regions that support G-quadruplex formation are usually also binding sites for one or more important transcription factors (32–36). For instance, the c-myc NHE G-quadruplex-forming region involves various DNA-binding proteins, such as the double-strand binding protein Sp1 (32), the single-strand binding proteins CNBP (33) and hnRNP K (34), and possibly yet-to-be-discovered proteins that can recognize the G-quadruplex formed by this sequence. These protein interactions are probably in fine balance and competing for potentially complementary or exclusive binding sites. Interaction with transcription factors specific for a certain DNA conformation would shift the equilibrium to the DNA conformation at the region, which would be critical for determining promoter activity. Supporting this model of transcriptional regulation by recognition of different DNA conformations, the ErbB2 Hr-DNA formation in the GGA repeats that is homologous to those in the c-myb promoter displaced the binding of the critical transcription factor Ets in the ErbB2 GGA repeat region, conversely, Ets binding inhibited Hr-DNA formation (37). This observation suggests that secondary DNA structures and protein:DNA interactions can influence each other to either activate or suppress gene expression. In agreement with this concept, our EMSA results demonstrate that the c-myb GGA repeat region actively interacts with many nuclear proteins (Figure 5) when it is double stranded, which might mean that T:H:H:T formation downregulates promoter activity by simply displacing double-strand binding nuclear proteins.

The c-myb GGA repeat region is a hot spot for nuclear protein interaction as shown in Figure 5. Several previous studies showed that GGA repeats in a gene promoter were critical transcriptional activator binding sites (23,24). Similarly, our luciferase assay results demonstrated that the GGA repeat region is essential for c-myb promoter activity, which is abolished when the GGA repeats are deleted (Figure 4), probably due to the loss of transcription factor binding sites. These results indicate that the GGA repeat region has to be available for transcriptional activators to bind to the promoter and may imply that G-quadruplex formation, if stabilized by a small molecule, could also markedly inhibit c-myb expression, making the region unavailable for transcription factor binding. The discovery and development of small molecule ligands for G-quadruplexes is the subject of investigation by several research groups (38,39).

We identified MAZ as one of the transcription factors that bind to the GGA repeat region of c-myb, and showed that MAZ can bind to both double-stranded DNA and T:H:H:T conformations of the c-myb promoter. MAZ was first identified as a GA box binding transcription



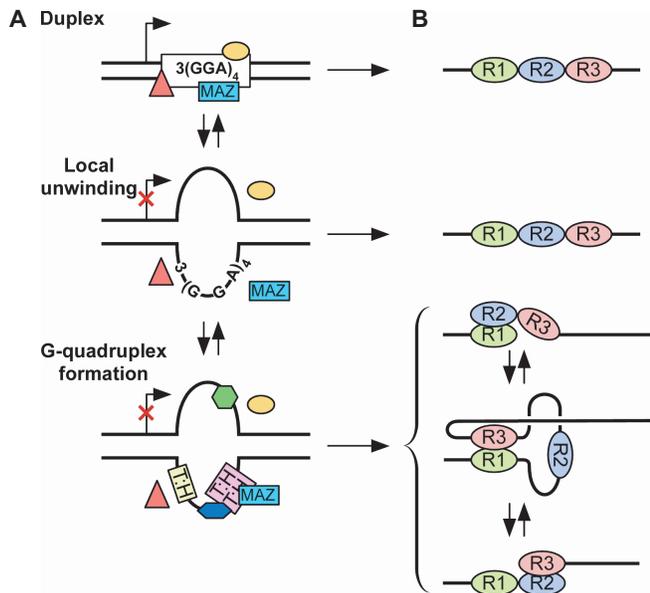
**Figure 7.** MAZ binding to the G-quadruplexes formed by GGA repeat DNA. (A) EMSA with ODNs V, VI and VII in the presence or absence of MAZ in rabbit reticulocyte lysate. MAZ in rabbit reticulocyte lysate binds to the T:H:H:T G-quadruplexes formed in ODNs V and VI but does not bind to the ODN VII that cannot form a G-quadruplex. DNA-protein complexes formed with MAZ or nonspecific lysate protein and the probes are indicated with arrows. (B) EMSA with ODNs V and I. ODN V, a positive control for MAZ binding to a T:H:H:T, confirms that MAZ also binds to the wild-type c-myc GGA repeat sequence ODN I. Due to the different size of ODNs I and V, the free probes show different mobility.



**Figure 8.** Transfection of the MAZ expression plasmid BRB112 into the Flp-In 293 Myb WT reporter stable cells. 0–1.2  $\mu\text{g}$  of BRB112 were transfected into the Flp-In 293 Myb WT reporter stable cells and luciferase activities were measured 24h after transfection. MAZ overexpression resulted in inhibition of c-myc promoter activity. \* $P < 0.01$ .

factor in the c-myc promoter and was shown to control P1 and P2 promoter activities of the c-myc gene (40–43). MAZ is constitutively expressed in various tissues (44), but abnormal MAZ expression is found in the terminal phase of CML, suggesting that overexpression of MAZ plays a role in the progression of CML (45). The functional role of MAZ is dependent on the target gene, and some promoters are activated by MAZ (46–48), while others are repressed (49,50). One possible explanation for the dual roles of MAZ may be that different cofactors interact with MAZ depending on the target promoter. Song *et al.* (50) showed that MAZ acts as a negative regulator via physical interaction with HDAC (histone deacetylase complex), and Jordan-Sciutto *et al.* demonstrated that MAZ interacts with FAC1 (Fetal Alzheimer's clone 1), a truncation of the chromatin remodeler BPTF (bromodomain and PHD domain transcription factor) (51,52). Interestingly, MAZ and the MAZ-like protein THZif-1 (triple-helix-binding zinc-finger protein-1) have

now been shown to bind to secondary DNA structures (26,53). The G-quadruplex structure formed in the IDDM2 locus, a critical regulator of IDDM2 promoter activity, is recognized by MAZ; however, MAZ does not bind to the unfolded form of the G-quadruplex-forming region (26). Moreover, the MAZ-like protein THZif-1 was shown to interact with the triplex formed in the c-myc NHE, which is both crucial for c-myc promoter activity, and more recently investigated by several groups as a G-quadruplex forming element (28,53–55). These observations suggest that regulation of promoter activity by MAZ and related proteins may involve recognition of secondary DNA structures. Some proteins, such as RAP1 (Repressor-activator protein 1), Thrombin and DNA topoisomerase I, have been shown to stabilize G-quadruplexes (56–58). Other proteins, including RecQ helicases, RPA (Replication Protein A) and POT1 (Protection of Telomere 1), may inhibit secondary structure formation or serve to resolve secondary DNA structures (59–61). In the case of G-quadruplexes formed in gene promoters, we propose that these genes are regulated by mechanisms involving single-stranded, double-stranded and folded DNA-binding proteins that may inhibit or promote G-quadruplex formation. Furthermore, it is possible that some proteins can bind to more than one DNA conformations, as we have shown with MAZ in this study (Figures 6 and 7). Since MAZ has dual binding activities to both double-stranded and G-quadruplex DNA, regulation of promoter activity by MAZ might be determined by the cofactors that interact with MAZ. It is not known how MAZ recruits different cofactors, but it may depend on levels of MAZ in the nucleus or on posttranslational modifications, such as phosphorylation of MAZ. It is also possible that MAZ binding to double-stranded or G-quadruplex DNA could result in changes in available transactivation domains of MAZ, which may allow different groups of transcription factors to interact with MAZ to determine its role. Moreover, the c-myc GGA repeat region is the first identified G-quadruplex-forming element in a transcribed



**Figure 9.** Working model for regulation of c-myc promoter activity via G-quadruplex formation in the c-myc GGA repeat region. (A) The equilibrium between the duplex form, opened form and G-quadruplex structure at the c-myc GGA repeat region. In the duplex form, transcriptional repressors and activators bind to the c-myc GGA repeat region in order to regulate c-myc expression levels, and MAZ possibly acts as a repressor of c-myc promoter by binding to this form. In the G-quadruplex structure, duplex-binding proteins are dissociated from the GGA repeat region and G-quadruplex binding proteins regulate G-quadruplex stability. MAZ may downregulate c-myc promoter activity via G-quadruplex binding at the GGA repeat region. (B) Schematic representation of G-rich strand of the GGA repeat region. When locally unwound the c-myc GGA repeat region could form T:H:H:T G-quadruplexes with any two of three (GGA)<sub>4</sub> repeats in the region.

region that indicates that T:H:H:T formation in the c-myc promoter may play a role as a transcriptional repressor not only by changing transcription factor interaction but also by acting as a roadblock to transcriptional machinery. Forced overexpression of MAZ dramatically downregulates c-myc promoter activity (Figure 8), indicating that MAZ is a transcriptional repressor of the c-myc promoter. However, these data do not show whether repression of c-myc expression is due to MAZ binding to duplex DNA or to the T:H:H:T structure.

The c-myc GGA repeat region is very complex, involving DNA conformational changes from double-stranded to G-quadruplex forms that subsequently alter transcription factor binding to the region. In spite of this complexity, we have shown here the importance of the T:H:H:T G-quadruplex-forming region of the c-myc promoter in controlling promoter activity, identified MAZ as a T:H:H:T binding transcription factor, and presented preliminary evidence that MAZ is a repressor of the c-myc promoter. Further study is needed to discriminate DNA conformation-dependent roles for MAZ in regulation of the c-myc promoter, to identify the cofactors that bind to MAZ depending on DNA structure, and to elucidate what determines MAZ binding to either G-quadruplex or double-stranded DNA.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Dr Linda Boxer at Stanford University School of Medicine and Dr Kelly Jordan-Sciutto at the University of Pennsylvania for providing the plasmid LB178 from which we obtained the c-myc promoter and the plasmid BRB112 that is used for expression of FLAG-MAZ, respectively. We are also grateful to David Bishop for proofreading and helping with some of the figures. Funding for this work was provided by grants from the U. S. Army Medical Research and Materiel Command awards (W81XWH-04-1-0560 and W81XWH-06-1-0233) and the Arizona Biomedical Research Commission (0021). Funding to pay the Open Access publication charges for this article was provided by NIH CA94166.

*Conflict of interest statement.* None declared.

## REFERENCES

- Oh, I.H. and Reddy, E.P. (1999) The myb gene family in cell growth, differentiation and apoptosis. *Oncogene*, **18**, 3017–3033.
- Todokoro, K., Watson, R.J., Higo, H., Amanuma, H., Kuramochi, S., Yanagisawa, H. and Ikawa, Y. (1988) Down-regulation of c-myc gene expression is a prerequisite for erythropoietin-induced erythroid differentiation. *Proc. Natl Acad. Sci. USA*, **85**, 8900–8904.
- Yi, H.K., Nam, S.Y., Kim, J.C., Kim, J.S., Lee, D.Y. and Hwang, P.H. (2002) Induction of apoptosis in K562 cells by dominant negative c-myc. *Exp. Hematol.*, **30**, 1139–1146.
- Gonda, T.J. and Metcalf, D. (1984) Expression of myb, myc and fos proto-oncogenes during the differentiation of a murine myeloid leukaemia. *Nature*, **310**, 249–251.
- Ramsay, R.G., Thompson, M.A., Hayman, J.A., Reid, G., Gonda, T.J. and Whitehead, R.H. (1992) Myb expression is higher in malignant human colonic carcinoma and premalignant adenomatous polyps than in normal mucosa. *Cell Growth Differ.*, **3**, 723–730.
- Griffin, C.A. and Baylin, S.B. (1985) Expression of the c-myc oncogene in human small cell lung carcinoma. *Cancer Res.*, **45**, 272–275.
- Slamon, D.J., Boone, T.C., Murdock, D.C., Keith, D.E., Press, M.F., Larson, R.A. and Souza, L.M. (1986) Studies of the human c-myc gene and its product in human acute leukemias. *Science*, **233**, 347–351.
- Luger, S.M., O'Brien, S.G., Ratajczak, J., Ratajczak, M.Z., Mick, R., Stadtmauer, E.A., Nowell, P.C., Goldman, J.M. and Gewirtz, A.M. (2002) Oligodeoxynucleotide-mediated inhibition of c-myc gene expression in autografted bone marrow: a pilot study. *Blood*, **99**, 1150–1158.
- Bender, T.P., Thompson, C.B. and Kuehl, W.M. (1987) Differential expression of c-myc mRNA in murine B lymphomas by a block to transcription elongation. *Science*, **237**, 1473–1476.
- Watson, R.J. (1988) A transcriptional arrest mechanism involved in controlling constitutive levels of mouse c-myc mRNA. *Oncogene*, **2**, 267–272.
- Watson, R.J. (1988) Expression of the c-myc and c-myc genes is regulated independently in differentiating mouse erythroleukemia cells by common processes of premature transcription arrest and increased mRNA turnover. *Mol. Cell Biol.*, **8**, 3938–3942.
- Yuan, W. (2000) Intron 1 rather than 5' flanking sequence mediates cell type-specific expression of c-myc at level of transcription elongation. *Biochim. Biophys. Acta*, **1490**, 74–86.
- Guerra, J., Withers, D.A. and Boxer, L.M. (1995) Myb binding sites mediate negative regulation of c-myc expression in T-cell lines. *Blood*, **86**, 1873–1880.

14. McCann,S., Sullivan,J., Guerra,J., Arcinas,M. and Boxer,L.M. (1995) Repression of the c-myc gene by WT1 protein in T and B cell lines. *J. Biol. Chem.*, **270**, 23785–23789.
15. Perrotti,D., Melotti,P., Skorski,T., Casella,I., Peschle,C. and Calabretta,B. (1995) Overexpression of the zinc finger protein MZF1 inhibits hematopoietic development from embryonic stem cells: correlation with negative regulation of CD34 and c-myc promoter activity. *Mol. Cell Biol.*, **15**, 6075–6087.
16. Bellon,T., Perrotti,D. and Calabretta,B. (1997) Granulocytic differentiation of normal hematopoietic precursor cells induced by transcription factor PU.1 correlates with negative regulation of the c-myc promoter. *Blood*, **90**, 1828–1839.
17. Sullivan,J., Feeley,B., Guerra,J. and Boxer,L.M. (1997) Identification of the major positive regulators of c-myc expression in hematopoietic cells of different lineages. *J. Biol. Chem.*, **272**, 1943–1949.
18. Nicolaidis,N.C., Correa,I., Casadevall,C., Travali,S., Soprano,K.J. and Calabretta,B. (1992) The Jun family members, c-Jun and JunD, transactivate the human c-myc promoter via an Ap1-like element. *J. Biol. Chem.*, **267**, 19665–19672.
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 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 tetrad and a G(A):G(A):G(A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
21. Matsugami,A., Ouhashi,K., Ikeda,T., Okuzumi,T., Sotoya,H., Uesugi,S. and Katahira,M. (2002) Unique quadruplex structures of d(GGA)4 (12-mer) and d(GGA)8 (24-mer)—direct evidence of the formation of non-canonical base pairs and structural comparison. *Nucleic Acids Res Suppl.*, **2**, 49–50.
22. Matsugami,A., Okuzumi,T., Uesugi,S. and Katahira,M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(A):G(A):G(A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
23. Chamboredon,S., Briggs,J., Vial,E., Hurault,J., Galvagni,F., Oliviero,S., Bos,T. and Castellazzi,M. (2003) v-Jun downregulates the SPARC target gene by binding to the proximal promoter indirectly through Sp1/3. *Oncogene*, **22**, 4047–4061.
24. Hafner,M., Zimmermann,K., Pottgiesser,J., Krieg,T. and Nischt,R. (1995) A purine-rich sequence in the human BM-40 gene promoter region is a prerequisite for maximum transcription. *Matrix Biol.*, **14**, 733–741.
25. De Armond,R., Wood,S., Sun,D., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 $\alpha$  promoter. *Biochemistry*, **44**, 16341–16350.
26. Lew,A., Rutter,W.J. and Kennedy,G.C. (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc. Natl Acad. Sci. USA*, **97**, 12508–12512.
27. Keniry,M.A. (2000) Quadruplex structures in nucleic acids. *Biopolymers*, **56**, 123–146.
28. 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. USA*, **99**, 11593–11598.
29. 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.
30. Nirula,A., Moore,D.J. and Gaynor,R.B. (1997) Constitutive binding of the transcription factor interleukin-2 (IL-2) enhancer binding factor to the IL-2 promoter. *J. Biol. Chem.*, **272**, 7736–7745.
31. Hammond-Kosack,M.C., Dobrinski,B., Lurz,R., Docherty,K. and Kilpatrick,M.W. (1992) The human insulin gene linked polymorphic region exhibits an altered DNA structure. *Nucleic Acids Res.*, **20**, 231–236.
32. DesJardins,E. and Hay,N. (1993) Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human c-myc promoters. *Mol. Cell Biol.*, **13**, 5710–5724.
33. Michelotti,E.F., Tomonaga,T., Krutzsch,H. and Levens,D. (1995) Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene. *J. Biol. Chem.*, **270**, 9494–9499.
34. Takimoto,M., Tomonaga,T., Matunis,M., Avigan,M., Krutzsch,H., Dreyfuss,G. and Levens,D. (1993) Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. *J. Biol. Chem.*, **268**, 18249–18258.
35. Ji,L., Mochon,E., Arcinas,M. and Boxer,L.M. (1996) CREB proteins function as positive regulators of the translocated bcl-2 allele in t(14;18) lymphomas. *J. Biol. Chem.*, **271**, 22687–22691.
36. Heckman,C., Mochon,E., Arcinas,M. and Boxer,L.M. (1997) The WT1 protein is a negative regulator of the normal bcl-2 allele in t(14;18) lymphomas. *J. Biol. Chem.*, **272**, 19609–19614.
37. Scott,G.K., Chang,C.H., Erny,K.M., Xu,F., Fredericks,W.J., Rauscher,F.J.,III, Thor,A.D. and Benz,C.C. (2000) Ets regulation of the erbB2 promoter. *Oncogene*, **19**, 6490–6502.
38. Neidle,S. and Thurston,D.E. (2005) Chemical approaches to the discovery and development of cancer therapies. *Nat. Rev. Cancer*, **5**, 285–296.
39. Dexheimer,T.S., Sun,D., Fry,M. and Hurley,L.H. (2006) DNA quadruplexes and gene regulation. In Neidle,S. (ed.), *Quadruplex Nucleic Acids*, Royal Society of Chemistry, RSC Publishing, Cambridge; pp. 180–207.
40. Pyrc,J.J., Moberg,K.H. and Hall,D.J. (1992) Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites within the c-myc promoter. *Biochemistry*, **31**, 4102–4110.
41. Komatsu,M., Li,H.O., Tsutsui,H., Itakura,K., Matsumura,M. and Yokoyama,K.K. (1997) MAZ, a Myc-associated zinc finger protein, is essential for the ME1a1-mediated expression of the c-myc gene during neuroectodermal differentiation of P19 cells. *Oncogene*, **15**, 1123–1131.
42. Bossone,S.A., Asselin,C., Patel,A.J. and Marcu,K.B. (1992) MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc. Natl Acad. Sci. USA*, **89**, 7452–7456.
43. Izzo,M.W., Strachan,G.D., Stubbs,M.C. and Hall,D.J. (1999) Transcriptional repression from the c-myc P2 promoter by the zinc finger protein ZF87/MAZ. *J. Biol. Chem.*, **274**, 19498–19506.
44. Song,J., Murakami,H., Tsutsui,H., Tang,X., Matsumura,M., Itakura,K., Kanazawa,I., Sun,K. and Yokoyama,K.K. (1998) Genomic organization and expression of a human gene for Myc-associated zinc finger protein (MAZ). *J. Biol. Chem.*, **273**, 20603–20614.
45. Daheron,L., Salmeron,S., Patri,S., Brizard,A., Guilhot,F., Chomel,J.C. and Kitzis,A. (1998) Identification of several genes differentially expressed during progression of chronic myelogenous leukemia. *Leukemia*, **12**, 326–332.
46. Parks,C.L. and Shenk,T. (1996) The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1. *J. Biol. Chem.*, **271**, 4417–4430.
47. Parks,C.L. and Shenk,T. (1997) Activation of the adenovirus major late promoter by transcription factors MAZ and Sp1. *J. Virol.*, **71**, 9600–9607.
48. Williams,L.J. and Abou-Samra,A.B. (2000) The transcription factors SP1 and MAZ regulate expression of the parathyroid hormone/parathyroid hormone-related peptide receptor gene. *J. Mol. Endocrinol.*, **25**, 309–319.
49. Karantzoulis-Fegaras,F., Antoniou,H., Lai,S.L., Kulkarni,G., D'Abreo,C., Wong,G.K., Miller,T.L., Chan,Y., Atkins,J. et al. (1999) Characterization of the human endothelial nitric-oxide synthase promoter. *J. Biol. Chem.*, **274**, 3076–3093.
50. Song,J., Ugai,H., Kanazawa,I., Sun,K. and Yokoyama,K.K. (2001) Independent repression of a GC-rich housekeeping gene by Sp1 and MAZ involves the same cis-elements. *J. Biol. Chem.*, **276**, 19897–19904.
51. Jordan-Sciutto,K.L., Dragich,J.M., Caltagarone,J., Hall,D.J. and Bowser,R. (2000) Fetal Alz-50 clone 1 (FAC1) protein interacts with the Myc-associated zinc finger protein (ZF87/MAZ) and alters its transcriptional activity. *Biochemistry*, **39**, 3206–3215.

52. Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J. and Shiekhattar, R. (2003) Isolation of human NURF: a regulator of *Engrailed* gene expression. *EMBO J.*, **22**, 6089–6100.
53. Sakatsume, O., Tsutsui, H., Wang, Y., Gao, H., Tang, X., Yamauchi, T., Murata, T., Itakura, K. and Yokoyama, K.K. (1996) Binding of THZif-1, a MAZ-like zinc finger protein to the nuclease-hypersensitive element in the promoter region of the *c-MYC* protooncogene. *J. Biol. Chem.*, **271**, 31322–31333.
54. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical studies of the *c-MYC* NHE III<sub>1</sub> promoter: model quadruplex interactions with a cationic porphyrin. *Biophys. J.*, **92**, 2007–2015.
55. Yang, D. and Hurley, L.H. (2006) Structure of the biologically relevant G-quadruplex in the *c-MYC* promoter. *Nucleosides Nucleotides Nucleic Acids*, **25**, 951–968.
56. Giraldo, R., Suzuki, M., Chapman, L. and Rhodes, D. (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl Acad. Sci. USA*, **91**, 7658–7662.
57. Baldrich, E. and O'Sullivan, C.K. (2005) Ability of thrombin to act as molecular chaperone, inducing formation of quadruplex structure of thrombin-binding aptamer. *Anal. Biochem.*, **341**, 194–197.
58. Arimondo, P.B., Riou, J.F., Mergny, J.L., Tazi, J., Sun, J.S., Garestier, T. and Helene, C. (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.*, **28**, 4832–4838.
59. Mohaghegh, P., Karow, J.K., Brosh, J.R., Jr, Bohr, V.A. and Hickson, I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.*, **29**, 2843–2849.
60. Salas, T.R., Petrusseva, I., Lavrik, O., Bourdoncle, A., Mergny, J.L., Favre, A. and Saintome, C. (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res.*, **34**, 4857–4865.
61. Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.