4598–4608 *Nucleic Acids Research, 2008, Vol. 36, No. 14* *Published online 9 July 2008*
doi:10.1093/nar/gkn380

Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene

Kexiao Guo[1], Vijay Gokhale[2], Laurence H. Hurley[2,3,4] and Daekyu Sun[2,*]

[1]Department of Biochemistry and Molecular Biophysics, University of Arizona, [2]Department of Pharmacology and Toxicology, College of Pharmacy, University of Arizona, Tucson, AZ 85721, [3]Arizona Cancer Center, 1515 N. Campbell Ave., Tucson, AZ 85724 and [4]BIO5 Collaborative Research Institute, 1657 E. Helen St., Tucson, AZ 85721, USA

Received April 28, 2008; Revised May 28, 2008; Accepted May 29, 2008

ABSTRACT

A polyguanine/polycytosine (polyG/polyC) tract in the proximal promoter of the vascular endothelial growth factor (VEGF) gene is essential for transcriptional activation. The guanine-rich (G-rich) and cytosine-rich (C-rich) strands on this tract are shown to form specific secondary structures, characterized as G-quadruplexes and i-motifs, respectively. Mutational analysis of the G-rich strand combined with dimethyl sulfate (DMS) footprinting, a polymerase stop assay, and circular dichroism (CD) spectroscopy revealed that the G-quadruplex containing a 1:4:1 double-chain reversal loop is the most thermodynamically stable conformation that this strand readily adopts. These studies provide strong evidence that the size of loop regions plays a critical role in determining the most favored folding pattern of a G-quadruplex. The secondary structure formed on the complementary C-rich strand was also determined by mutational analysis combined with Br_2 footprinting and CD spectroscopy. Our results reveal that at a pH of 5.9 this strand is able to form an intramolecular i-motif structure that involves six C–C^+ base pairs and a 2:3:2 loop configuration. Taken together, our results demonstrate that the G-quadruplex and i-motif structures are able to form on the G- and C-rich strands, respectively, of the polyG/polyC tract in the VEGF proximal promoter under conditions that favor the transition from B-DNA to non-B-DNA conformations.

INTRODUCTION

VEGF, a key regulator of angiogenesis, plays an important role in tumor survival, growth and metastasis (1,2) and is overexpressed in many types of human cancers, including glioma, renal cell carcinoma, and ovarian and pancreatic cancer (3–6). VEGF expression is regulated by many external stimuli, such as hypoxia (7–9), growth factors (10,11), hormones (12) and cytokines (13,14). Also, the loss or inactivation of tumor suppressor genes (15,16) and the activation of oncogenes (17) are reported to cause VEGF upregulation. Studies on the molecular mechanism of VEGF gene expression have defined major *cis*-acting elements and transcription factors involved in the regulation of this expression (18–20). These studies showed that a polyG/polyC tract in the VEGF proximal promoter region (–85 to –50 bp relative to the initiation of transcription; Figure 1) is essential for basal and inducible VEGF expression. This polyG/polyC tract contains three potential Sp1 binding sites (Figure 1). It was reported that the deletion of this polyG/polyC tract dramatically decreases the basal promoter activity by about 90% and abolishes VEGF expression induced by platelet-derived growth factor (PDGF) in NIH3T3 cells (18). In another report, the deletion of this polyG/polyC tract totally abrogated VEGF promoter activity in PANC-1 cancer cells (19). Given the importance of this tract for VEGF promoter activity, we examined the DNA sequence of this region and found that the G-rich strand contains four consecutive runs of guanines separated by one or two cytosine bases, and the C-rich strand contains four consecutive runs of cytosines separated by one or two guanine bases. These sequences are consistent with the

*To whom correspondence should be addressed. Tel: +1 520 626 0323; Fax: +1 520 626 4824; Email: sun@pharmacy.arizona.edu
Correspondence may also be addressed to Laurence Hurley. Tel: +1 520 626 5622; Fax: +1 520 626 4824; Email: hurley@pharmacy.arizona.edu

general motifs that form G-quadruplex and i-motif structures.

DNA sequences containing consecutive runs of guanines are known to form intermolecular or intramolecular G-quadruplex structures. These are four-stranded structures consisting of stacked G-tetrads having a planar association of four guanines held together by Hoogsteen hydrogen bonding (21–25). DNA sequences containing consecutive runs of cytosines, at acidic pH levels, are capable of forming intermolecular or intramolecular four-stranded i-motif structures. The two parallel-stranded C–C^+ (one cytosine must be hemiprotonated at the N3 position) base-paired duplexes intercalate into each other in an antiparallel orientation. G-quadruplex and i-motif structures have been extensively studied and have been shown to form in the telomeric ends of eukaryotic chromosomes (26–30) and in the G- and C-rich sequences located in the proximal promoter regions of c-myc (21,31,32), KRAS (30,33), Rb (34) and RET (35) genes. In the present study, we demonstrate the formation of these structures in the VEGF promoter.

Within supercoiled plasmid DNA, the G-rich strand in the polyG/polyC tract of the VEGF proximal promoter is very dynamic and able to adopt non-B-DNA conformations, which are resistant to cleavage by both DNase I and S1 nuclease in the presence of K^+ or the G-quadruplex-interactive agent telomestatin (36,37). The G-rich strand of the VEGF proximal promoter contains five runs of guanines (Figure 1). Our previous work has shown that the four 5′-end runs of guanines are involved in the formation of an intramolecular G-quadruplex, whereas inclusion of the fifth run of guanines results in the formation of intermolecular G-quadruplexes (37). Therefore, we focused this study on the presumably more biologically relevant G-quadruplex formed by the first four runs of guanines at the 5′-end. Results from a polymerase stop assay and CD studies of the G-rich strand in this region show the formation of an intramolecular parallel G-quadruplex structure (36). The most thermodynamically stable G-quadruplexes formed by the G-rich strand in the VEGF proximal promoter were studied using mutational analysis, CD melting, DMS footprinting and a polymerase stop assay. This present investigation has added significance since we have recently shown that the G-quadruplex-interactive agents that are able to stabilize these G-quadruplexes repress VEGF expression in cancer cells, which offers a new approach to antiangiogenesis therapy.

Here, we demonstrate the formation of both G-quadruplex and i-motif structures in the VEGF proximal promoter and characterize their folding patterns and most likely structures using biochemical and biophysical techniques. Finally, other sequences in promoter regions that form these G-quadruplex and i-motif structures have been compared to provide additional insight into the conservation of these elements.

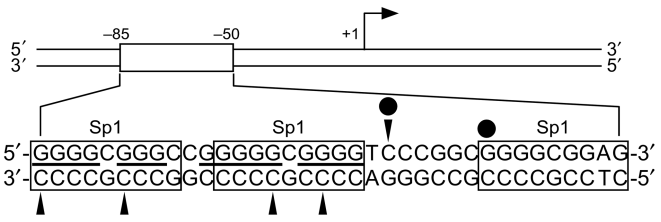

Figure 1. Schematic diagram of the proximal promoter region of the human VEGF gene, with the location and base sequence of the polyG/polyC tract in this region shown. Three Sp1 binding sites are enclosed in boxes, and the four runs of guanines are underlined. The DNase I and S1 nuclease hypersensitive sites are indicated with filled circles and arrowheads, respectively.

MATERIALS AND METHODS

Materials

The oligonucleotides used in this work were purchased from Biosearch Technologies (Novato, CA, USA), and their sequences are listed in Table 1.

CD spectroscopy

CD spectra were recorded on a Jasco-810 spectrophotometer (Jasco, Easton, MD, USA) at room temperature (~25°C), using a quartz cell of 1 mm optical path length, an instrument scanning speed of 100 nm/min, with a response time of 1 s and over a wavelength range of 200–350 nm. Each spectrum was recorded three times, smoothed and baseline-corrected for signal contributions from used buffers.

For the G-quadruplex study, the oligonucleotides were diluted to a strand concentration of 10 μM in 50 mM Tris–HCl, pH 7.4, with an appropriate concentration of KCl. In the T_m experiment with 20 mM KCl, the molar ellipticities at 262 nm at different temperatures (4–95°C) were recorded and plotted against temperatures.

Table 1. Oligonucleotides used in this study

		1 5 10 15 20
A	WT	5′-TTTTTGGGGCGGGCCGGGGGCGGGGTTTTT-3′
	C9T	5′-TTTTTGGGGCGGG**T**CGGGGGCGGGGTTTTT-3′
	C9A	5′-TTTTTGGGGCGGG**A**CGGGGGCGGGGTTTTT-3′
	C10T	5′-TTTTTGGGGCGGGC**T**GGGGGCGGGGTTTTT-3′
	C10A	5′-TTTTTGGGGCGGGC**A**GGGGGCGGGGTTTTT-3′
	G11/12T	5′-TTTTTGGGGCGGGCC**TT**GGGCGGGGTTTTT-3′
	G11T	5′-TTTTTGGGGCGGGCC**T**GGGGCGGGGTTTTT-3′
	G12T	5′-TTTTTGGGGCGGGCCG**T**GGGCGGGGTTTTT-3′
	G14T	5′-TTTTTGGGGCGGGCCGGG**T**GCGGGGTTTTT-3′
	G14A	5′-TTTTTGGGGCGGGCCGGG**A**GCGGGGTTTTT-3′
		1 5 10 15 20
B	IM1	5′-GACCCCGCCCCCGGCCCGCCCCGG-3′
	IM2	5′-GAC**TT**CGCCCCCGGCCCGCCCCGG-3′
	IM3	5′-GACCCCGC**TTT**CGGCCCGCCCCGG-3′
	IM4	5′-GACCCCGCCCCCGGC**T**CGCCCCGG-3′
	IM5	5′-GACCCCGCCCCCGGCCCGC**TT**CGG-3′
	IM6	5′-GA**T**CCCGCCCCCGGCCCGCCCCGG-3′
	IM7	5′-GACCC**T**GCCCCCGGCCCGCCCCGG-3′
	IM8	5′-GACCCCG**TT**CCCGGCCCGCCCCGG-3′
	IM9	5′-GACCCCGCCC**TT**GGCCCGCCCCGG-3′
	IM10	5′-GACCCCG**T**CCC**T**GGCCCGCCCCGG-3′
	IM11	5′-GACCCCGCCCCCGGCCCG**T**CCCGG-3′
	IM12	5′-GACCCCGCCCCCGGCCCGCCC**T**GG-3′
	T24	5′-TTTTTTTTTTTTTTTTTTTTTTTTT-3′

Bold letters indicate mutated bases.

For the i-motif study, the oligonucleotides were diluted to a strand concentration of 10 μM in 50 mM Tris–acetate at appropriate pHs. To determine the transition mid-point of the VEGF i-motif structure, the molar ellipticities at 288 nm against pHs were plotted and the mid-point was determined.

DMS footprinting

The oligonucleotides were 5′-end-labeled with ^{32}P and purified by micro-spin column 6. The ^{32}P-labeled oligonucleotides were treated with 2% DMS for 5 min in either the absence of KCl or the presence of 100 mM KCl. The DMS-treated oligonucleotides were loaded on a 16% nondenaturing polyacrylamide gel to separate the single-stranded DNA and intramolecular G-quadruplex from other intermolecular G-quadruplexes by their different electrophoretic mobilities. The DNA were recovered and subjected to piperidine cleavage (21).

Polymerase stop assay

The polymerase stop assay templates were designed by placing the VEGF proximal promoter G-rich sequence, or the various mutant G-rich sequences, in a polymerase stop assay cassette, as described previously (38). 5′-end-labeled primer p28 d(TAATACGACTCACTATAGCAA TTGCGTG) and template DNA were annealed in an annealing buffer (50 μM Tris–HCl, pH 7.5, 10 μM NaCl) by heating at 95°C for 5 min and slowly cooling to room temperature. The primer-annealed template oligonucleotides were purified by electrophoresis using a 12% nondenaturing polyacrylamide gel. The purified primer template oligonucleotides were used in a primer extension assay with *Taq* DNA polymerase, as described previously (38).

Native polyacrylamide gel electrophoresis (PAGE)

The pHs of PAGE gel solution in 1 × Tris–Acetate–EDTA and running buffer (1 × Tris–acetate–EDTA) were both adjusted to pH 5.0. Ten thousand cpm of ^{32}P-labeled IM1 or T24 in 50 mM Tris–acetate, pH 5.0, was heated at 95°C for 5 min, cooled to room temperature and incubated at room temperature for 2 h before loading on a 20% nondenaturing PAGE gel.

Br_2 footprinting experiment

The Br_2 footprinting experiment was carried out in accordance with published procedure (39) to probe the secondary structure formed by IM1. In brief, IM1 was 5′-end-labeled with ^{32}P using T4 polynucleotide kinase and [γ-^{32}P] ATP, and the labeled IM1 was gel-purified using 12% polyacrylamide gel electrophoresis under denaturing conditions (7 M urea). For the Br_2 cleavage reaction, the purified 5′-end-labeled IM1 was treated for 20 min with molecular Br_2 that was generated *in situ* by mixing an equal molar concentration (50 mM) of KBr with $KHSO_5$ in the same tube. The reactions were then terminated by adding 50 μl of stop mix containing 0.6 M Na-acetate (pH 5.2) and 10 mg/ml calf thymus DNA, and unreacted Br_2 was removed by ethanol precipitation. Following ethanol precipitation, the DNA pellet was dried and resuspended with 100 μl of freshly diluted 1 M piperidine, and the samples were heated at 90°C for 30 min to induce bromination-specific strand cleavage. Following piperidine treatments, the DNA samples were completely dried and resuspended with alkaline sequencing gel loading dye and applied to a 20% sequencing gel. The purine- and pyrimidine-specific reactions were carried out using formic acid or hydrazine to generate sequencing markers, following published procedure (39).

RESULTS

Formation of G-quadruplex structures in the G-rich strand of the VEGF proximal promoter

We have reported that the G-rich strand of the VEGF proximal promoter is able to form an unusual DNA secondary structure in the presence of KCl or G-quadruplex-interactive-agents, such as TMPyP4 and telomestatin (36). In the presence of 100 mM KCl, the CD spectrum of the wild-type G-rich sequence in the VEGF proximal promoter (WT, Table 1A) exhibits a characteristic positive peak at 264 nm and a negative peak at 240 nm, suggesting that it forms a parallel G-quadruplex structure (Figure 2A) (35,36). In the sequences studied, we added 5′ and 3′ thymidine tails and showed that these did not affect the folding patterns relative to the native sequences.

To resolve the G-quadruplex structure formed by this strand, we performed DMS footprinting to identify guanines involved in the G-tetrads, by their resistance to methylation at position N7 due to Hoogsteen hydrogen bonding (21). In this process, the G-quadruplex-forming oligonucleotides were subjected to DMS treatment in the absence or presence of KCl. The DMS-treated samples were resolved on a nondenaturing polyacrylamide gel, and the intramolecularly folded DNA samples were recovered from the gel and subjected to piperidine cleavage. In the absence of KCl, the DMS methylation pattern of the guanines along the WT sequence is consistent with an unstructured DNA form (Figure 2B, lane 3), whereas in the presence of 100 mM KCl, some guanines were well-protected against DMS methylation, while others showed enhanced methylation and subsequent cleavage (Figure 2B, lane 4). The DMS footprinting pattern of WT in the presence of 100 mM KCl demonstrated a signature for G-quadruplex formation where four tracts of three guanines (G2–G4, G6–G8, G13–G15 and G17–G19) form the four strands of a G-quadruplex with a 1:4:1 loop arrangement (two 1-base loops and one 4-base loop) consisting of C5, C9–G12 and C16, respectively (defined as G-quadruplex I; Figure 3B). Unexpectedly, G14, which is located in the run of five guanines (G11–G15), showed only modest methylation by DMS, which is consistent with our previous observation (37).

In an attempt to understand the reactivity mechanism of G14 to DMS and to extend our previous study (37), we first determined whether the two cytosines in the central loop (C9 and C10) are involved in cleavage at G14 by forming possible C–G base pairs. Sequences containing C-to-T or C-to-A mutations (C9T, C9A, C10T and C10A

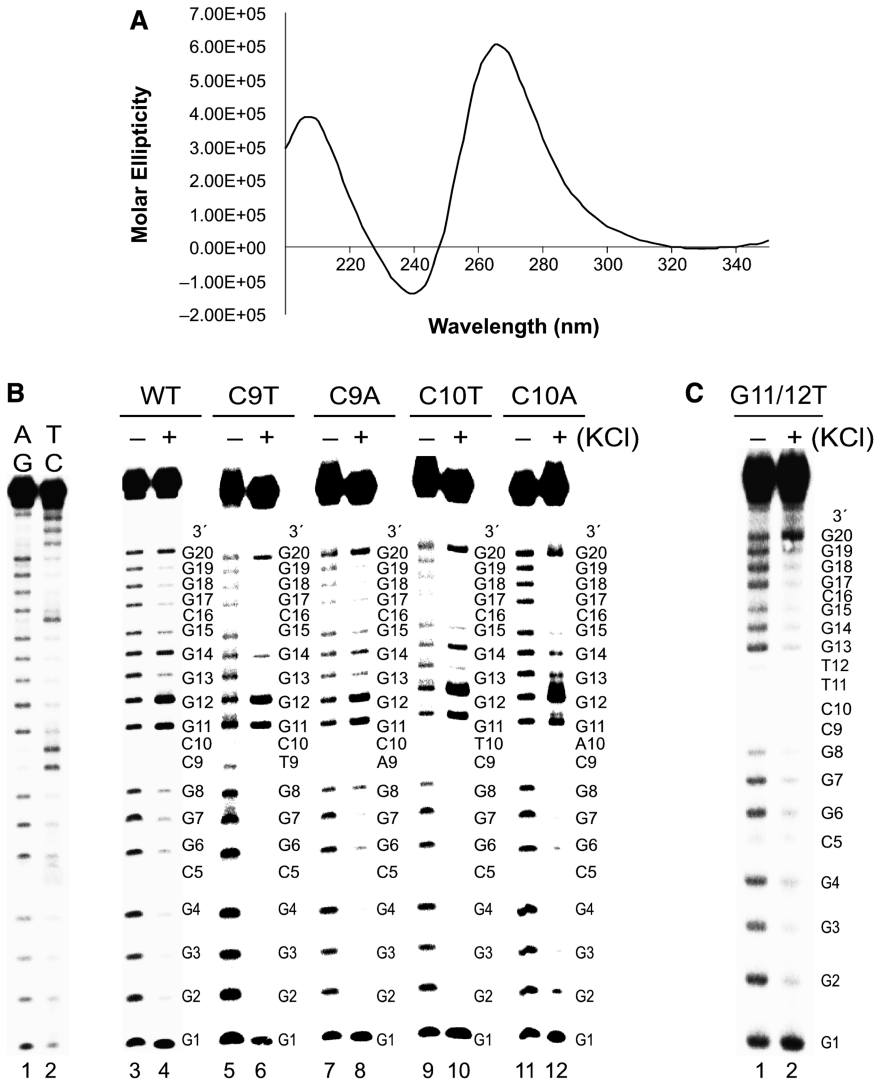

Figure 2. CD and DMS footprinting of the wild-type VEGF G-rich strand (WT) and other mutant sequences. (**A**) CD spectrum of WT in the presence of 100 mM KCl. (**B**) DMS footprinting of WT, C9T, C9A, C10T and C10A in the absence of KCl (−) or in the presence of 100 mM KCl (+). (**C**) DMS footprinting of G11/12T. The AG and TC sequencing reactions for WT are shown on the left of (B).

in Table 1A) in the central loop were examined first. In the presence of 100 mM KCl, the CD spectra of these sequences are similar to that of oligonucleotide WT (data not shown), suggesting that they all form parallel G-quadruplexes. DMS footprinting results showed that methylation at G14 still persists (Figure 2B, lanes 6, 8, 10 and 12), with some decrease in oligomers C9T and C10A compared to the WT sequence. DMS footprinting of an oligomer with G-to-T mutations at positions 11 and 12 (G11/12T in Table 1A), which forms a parallel G-quadruplex based on CD (data not shown), revealed that methylation at G14 disappeared (Figure 2C). This experiment demonstrates that the reactivity to DMS at G14 in the WT sequence is dependent on the presence of the two guanines in the loop. Thus, this reactivity of G14 to DMS comes from either the alternative use of the five consecutive guanines in forming G-quadruplexes (see below) or other possible interplay among G14, G11 and G12. However, a clear explanation for the reactivity of G14 to DMS requires detailed information on the VEGF G-quadruplex by NMR or X-ray crystallography.

DMS footprinting, polymerase stop assay and CD studies show that there are two equilibrating G-quadruplex loop isomers, of which the 1:4:1 form is the most stable

In the G-rich strand of the VEGF proximal promoter, the third run of guanines contains five guanines (G11–G15), but only three consecutive guanines are required to form a G-quadruplex. In principle, different loop isomers using these five guanines are able to form by using three different contiguous guanines, i.e. G11–G13, G12–G14 or G13–G15. To evaluate this hypothesis, we designed several mutant sequences (G11T, G12T, G14T and G14A in Table 1A). The CD spectra of these four sequences in the presence of 100 mM KCl all demonstrated parallel G-quadruplex formation (data not shown). In the presence of KCl, DMS footprinting of G11T and G12T

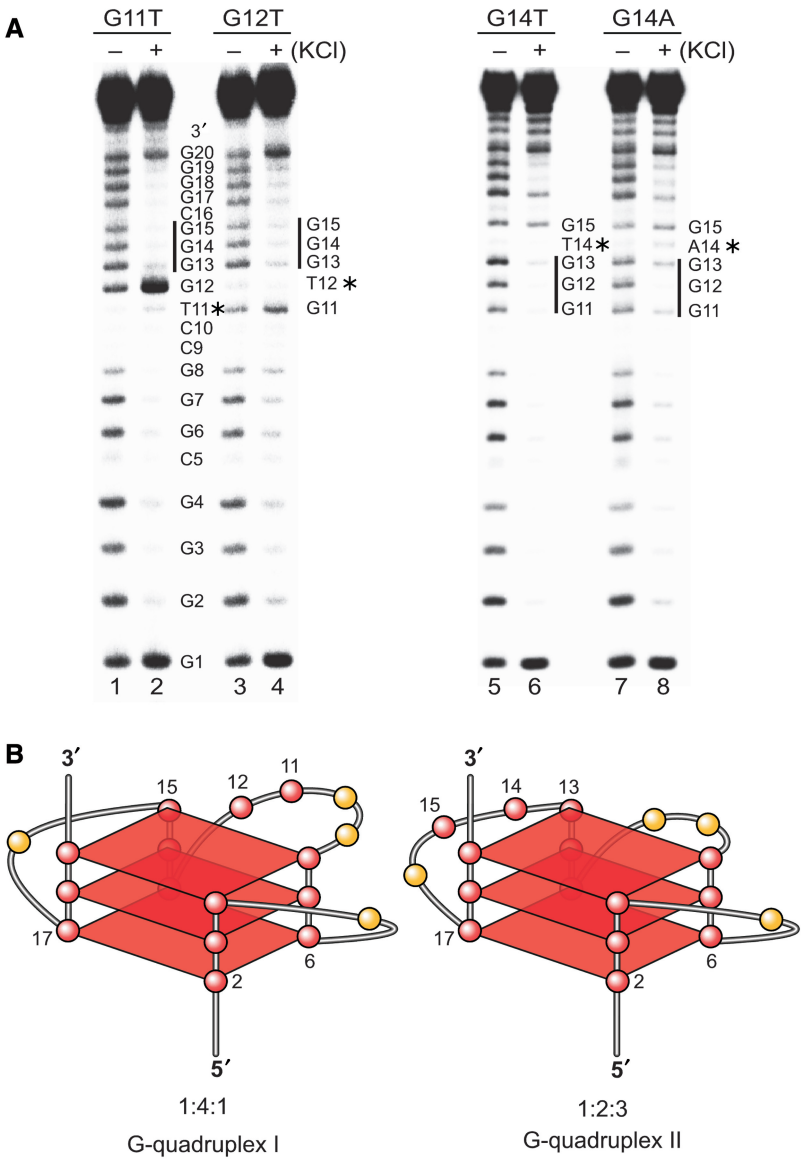

Figure 3. (**A**) DMS footprinting of oligonucleotides G11T, G12T, G14T and G14A in the absence of KCl (−) or in the presence of 100 mM KCl (+). The sequence of G11T is shown to the right of the gel. For the other sequences, only the mutated region of G11–G15 is shown. The mutation sites in each sequence are indicated with an asterisk. (**B**) The two different folding patterns (1:4:1 and 1:2:3) for the VEGF G-quadruplexes inferred from this data.

showed that four runs of three guanines (G2–G4, G6–G8, G13–G15 and G17–G19) are well protected against DMS methylation and subsequent cleavage, while the other guanines (G1, G12 and G20 in G11T, and G11 and G20 in G12T) showed enhanced DMS methylation and subsequent cleavage (Figure 3A, lanes 2 and 4). Overall, the G-quadruplex structures formed by these mutants at the 5′-end of the run of five guanines (G11T and G12T) use the same three G-tetrads, with a 1:4:1 arrangement of loops, as those formed by the WT oligomer (defined as G-quadruplex I; Figure 3B). For sequences G14T and G14A, we saw a different DMS footprinting pattern: the three contiguous guanines at the 5′-end of the third run of guanines (G11–G13) were well protected against DMS methylation (Figure 3A, lanes 6 and 8), suggesting that guanines G11–G13 are involved in G-quadruplex formation. DMS footprinting of G14T and G14A in the presence of 100 mM KCl demonstrated a pattern of G-quadruplex formation where G2–G4, G6–G8, G11–G13 and G17–G19 form the four strands that connect the three tetrads, with a 1:2:3 arrangement of loops composed of C5; C9 and C10; and G14, G15 and C16, respectively (defined as G-quadruplex II; Figure 3B).

The stabilities of G-quadruplexes I and II were compared using a polymerase stop assay (Figure 4A). Very significant stop products were seen for the sequences that form G-quadruplexes using the three guanines at the 3′-end (G13–G15) of the five guanine run with increasing concentrations of KCl (Figure 4A, Pol-WT, Pol-G11T, Pol-G12T, Pol-G11/12T), whereas only minor stop products were observed for the sequences that form G-quadruplexes using the three guanines at the 5′-end (G11–G13) (Figure 4A, Pol-G14T and Pol-G14A). We conclude that the G-quadruplex structure (G-quadruplex I)

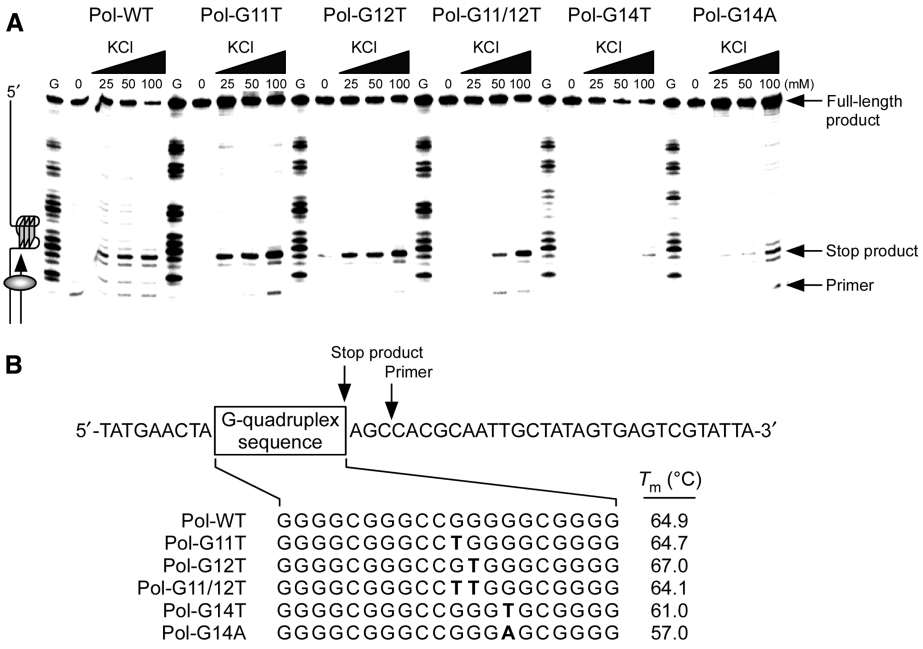

Figure 4. (**A**) Polymerase stop assay of Pol-WT, Pol-G11T, Pol-G12T, Pol-G11/12T, Pol-G14T and Pol-G14A with increasing concentrations of KCl. The primer site, stop product site and full-length product site are indicated with arrows. Sequencing reactions of guanine for each sequence are indicated at the top of each lane (G). (**B**) The polymerase stop assay cassette sequence is shown with a box indicating the different G-quadruplex sequences studied. The primer site and stop site product are shown with arrows. The G-quadruplex sequences are shown below the polymerase stop assay cassette sequence, with their T_m on the right.

formed using the three 3′-end guanines (G13–G15), which forms a 1:4:1 loop isomer species, is more stable than the one (G-quadruplex II) formed using the three 5′-end guanines (G11–G13). This is further confirmed by CD melting experiments, which showed that the G-quadruplex structures with the 1:4:1 loop arrangement (G11T, G12T, G11/12T) have higher T_m (64–67°C versus 57–61°C) than the G-quadruplexes with the 1:2:3 loop arrangement (G14T and G14A) (Figure 4B).

The C-rich strand of the VEGF proximal promoter forms an intramolecular i-motif structure

To determine whether an i-motif structure is formed by the C-rich strand of the VEGF proximal promoter, CD spectra of IM1 (Table 1B), a sequence identical to the C-rich strand of the VEGF proximal promoter, were collected at different pHs. As shown in Figure 5A, at acidic pH (≤5.9), there is a characteristic positive peak at 288 nm and a negative peak at 265 nm, with crossover at 276 nm, indicating the formation of an i-motif structure. At pH >7, the positive peak sharply decreases, showing that the i-motif stability decreases with increasing pH. This indicates that an unstructured DNA forms at neutral pH due to the deprotonation and the disruption of $C–C^+$ base-pairing. The transition mid-point pH was determined to be pH 5.8 by plotting the molar ellipticity at 288 nm versus pHs (Figure 5B).

Native PAGE was used to differentiate between the formation of intramolecular and intermolecular i-motif structures (31,40,41). A 24-mer dT oligonucleotide, T24 (Table 1B), was used as a control in this experiment. T24 and IM1 were first subjected to electrophoresis on a denaturing PAGE gel at pH 8.3, where it was found that IM1 ran as a single band with a similar mobility to T24, which is similar in size to IM1 (Figure 5C). Subsequently, we ran the same oligomers on a native PAGE gel at pH 5.0, conditions at which IM1 forms an i-motif structure as determined by CD. The results in Figure 5C show that IM1 travels as one band faster than T24, suggesting that under acidic conditions, IM1 adopts an intramolecular structure.

If an intramolecular i-motif structure is formed, all four runs of cytosines should be involved in i-motif formation. Therefore, mutations within each run of cytosines should disrupt this formation. To test this, a single C-to-T mutation, double C-to-T mutations or triple C-to-T mutations were introduced in the middle of runs of three cytosines, four cytosines or five cytosines, respectively, to disrupt the cytosine repeats and determine their involvement in i-motif formation. CD spectra of these different mutant sequences (IM2–IM5, Table 1B) and IM1 at pH 5.5 were determined. When compared to the CD spectra of IM1, the i-motif signature peaks (288 nm) of IM2–IM5 were all dramatically decreased (Figure 5D), suggesting that each run of cytosines contributes to base-pairing during i-motif formation, which further supports an intramolecular i-motif formation.

Determination of the cytosines involved in the formation of the intramolecular i-motif by CD analysis

CD analysis of wild-type and mutant sequences was previously used to characterize the i-motif structure formed by the C-rich strand on the 5′-end of the Rb gene (34).

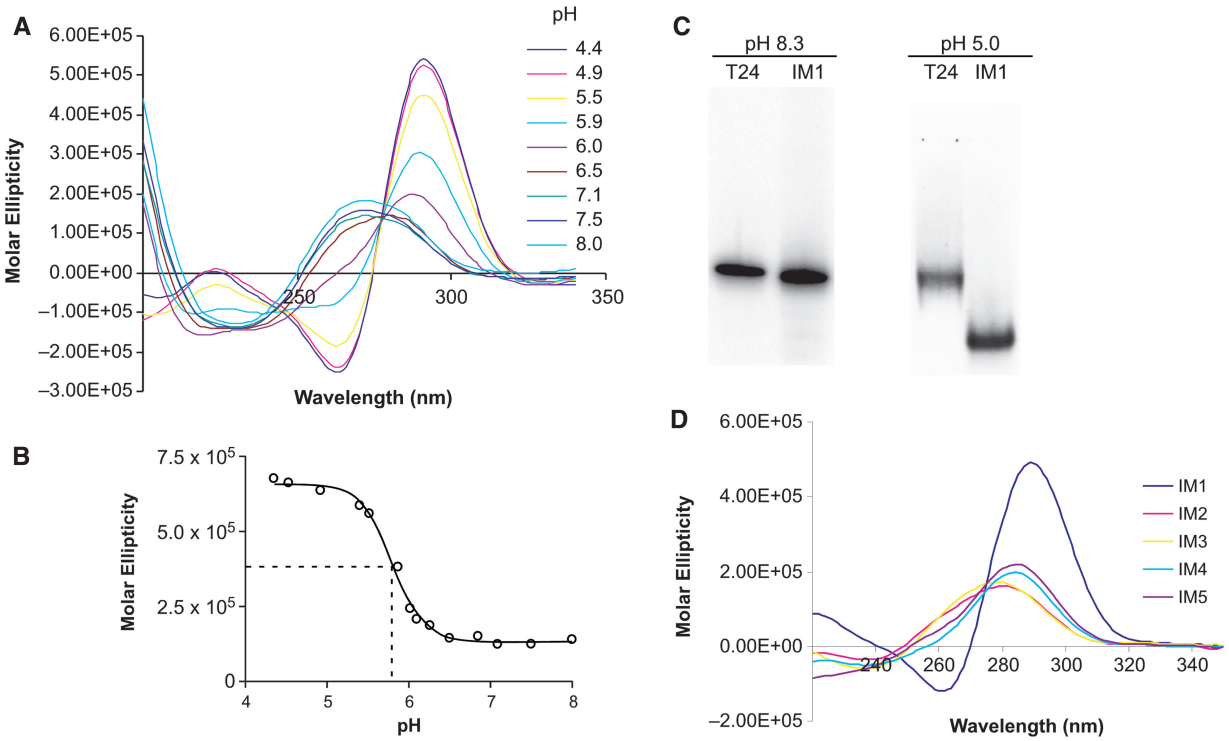

Figure 5. (**A**) CD spectra of IM1 at room temperature in 50 mM Tris–acetate buffer at different pHs (8.0–4.4). (**B**) pH dependence of the molar ellipticity at 288 nm from (A). (**C**) IM1 and T24 on denaturing PAGE gel at pH 8.3 (left) and on a nondenaturing PAGE gel at pH 5.0 (right). (**D**) CD spectra of IM1–IM5 at pH 5.5.

In the C-rich sequence of the VEGF promoter, only the third run contains three cytosines. Therefore, it was assumed that there would be only three cytosines in each run involved in the i-motif structure. To identify which three cytosines are involved in the hemiprotonated C–C^+ pairing of the i-motif, selective C-to-T mutations were introduced at specific positions on the wild-type C-rich sequence (IM6–IM12, Table 1B). The CD spectra of these sequences were collected and compared with that of IM1. IM6 and IM7 were designed so that the C at either the 5′-end (IM6) or the 3′-end (IM7) of the first run of cytosines was mutated to T. The results show that the mutation at the 3′-end (IM7) produces a greater reduction in molar ellipticity at 288 nm in comparison to the mutation at the 5′-end (IM6) (Figure 6A), suggesting that the C at the 3′-end is more important for i-motif structure formation. Using the same procedure, it was shown that in the second run of cytosines, the central three cytosines are more important for i-motif formation (Figure 6B), and in the fourth run of cytosines, the C at the 3′-end is more critical for i-motif formation (Figure 6C). On the basis of these results, the cytosines of IM1 involved in i-motif formation are proposed to be C2–C4, C7–C9, C13–C15 and C18–C20 (Figure 6D).

Determination of the predominant i-motif structure formed by the C-rich strand by Br_2 footprinting

As we previously reported, Br_2 footprinting can be used to unambiguously determine the identity of the cytosine residues involved in base pairings and intercalations to form i-motif structures (35). As shown in Figure 7 (left panel), lane 3, C2–C4, C7–C9, C13–C15 and C18–C20 are well protected, suggesting that these cytosines are involved in base pairing and intercalation, whereas cytosine residues C6, C10 and C17 showed enhanced reactivity toward Br_2, suggesting that these three residues are located in loop regions. The cytosines involved in base pairing and intercalation determined by Br_2 footprinting are the same as those determined by CD analysis. On the basis of the results of CD analysis and Br_2 footprinting, a model folding pattern for the VEGF i-motif is proposed, which involves six hemiprotonated C–C^+ base pairs formed from four consecutive antiparallel cytosine stretches with a 2:3:2 loop arrangement (Figure 7, right panel). Since there are two very narrow grooves and two very wide grooves in an i-motif (42), we speculate that the two 2-base loops run across narrow grooves and the 3-base loop runs across a wide groove. The predominant i-motif formed by the C-rich strand of the VEGF promoter adopts an antiparallel i-motif folding pattern similar to that found in the human telomeric C-strand, as determined by NMR (42).

DISCUSSION

VEGF is one of the most important angiogenic factors and is overexpressed in many types of tumors (43). VEGF and its receptors have been attractive targets for anti-angiogenesis therapy in the last decade (44). In 2004, Bevacizumab, a humanized anti-VEGF antibody, was

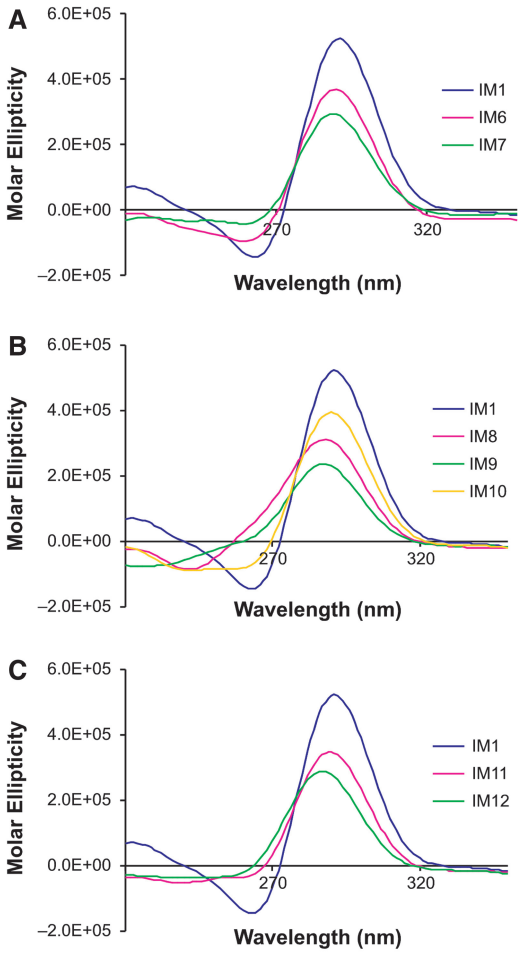

Figure 6. CD spectra of IM1, IM6 and IM7 (**A**), IM1 and IM8–IM10 (**B**) and IM1, IM11 and IM12 (**C**). The cytosines predicted to be involved in i-motif formation are underlined (**D**).

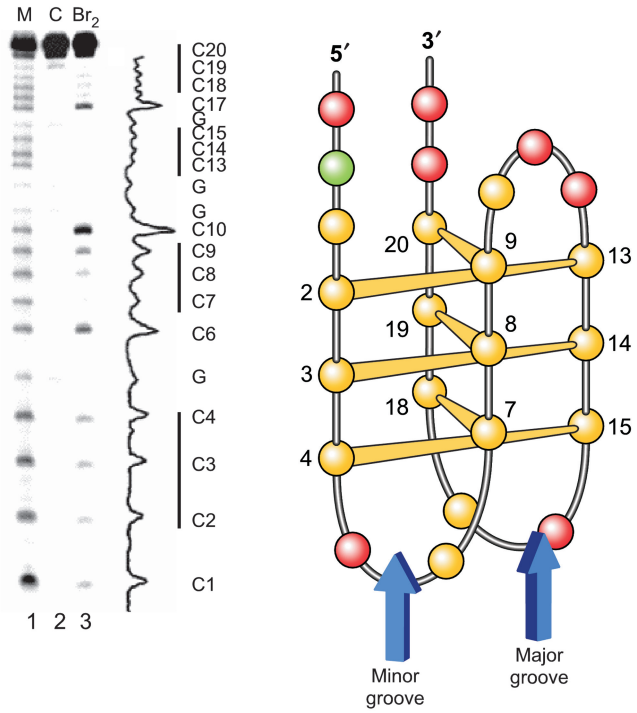

Figure 7. Autoradiogram and densitometric scanning of the autoradiogram from a Br_2 footprinting experiment to determine cytosine residues involved in base pairing and intercalation in the intramolecular i-motif folded form. Lane 1 represents the pyrimidine-specific reaction used to generate sequencing marker. Lanes 2 and 3 represent reaction without and with Br_2 (0.5 mM) in double-distilled H_2O, respectively. On the right is shown the predicted folding pattern of the predominant i-motif structure formed by IM1 based on the results of the Br_2 footprinting experiment shown on the left.

approved by the FDA as a first-line therapy for colorectal cancer treatment, and there now are small-molecule VEGF receptor inhibitors in different stages of clinical trials (45). Our previous study showed that the polyG/polyC region in the VEGF proximal promoter is very dynamic and able to unwind into single-stranded DNA and then fold into secondary DNA structures, such as G-quadruplexes (36). In this article, we have partially characterized the unique DNA secondary structures formed in the VEGF proximal promoter. The G-rich strand of the VEGF proximal promoter was shown to form a parallel G-quadruplex with three G-tetrads and three double-chain reversal loops containing 1, 4 and 1 bases. From the complementary C-rich strand, an i-motif is formed with six $C–C^+$ base pairs and three loops containing 2, 3 and 2 bases.

The four consecutive runs of guanines in the G-rich strand of the polyG/polyC tract in the VEGF proximal promoter have four, three, five and four guanines, respectively. If three consecutive guanines are randomly selected from each run of guanines to form a G-quadruplex, twelve different loop isomers could possibly be formed by the VEGF wild-type G-rich sequence. However, the results of our study revealed that a 1:4:1 loop isomer is the predominant configuration formed by this sequence. Results from the polymerase stop assay show that the 1:4:1 loop isomer is much more efficient than the 1:2:3 loop isomer in arresting DNA synthesis (Figure 4A), suggesting that the 1:4:1 loop isomer may be more biologically relevant. The CD melting experiments confirmed that the 1:4:1 loop isomer has greater stability (4–10°C) than the 1:2:3 loop isomer (Figure 4B). To enhance the production of the 1:2:3 loop isomer, a mutation is required at G14.

The major folding difference between these two loop isomers is that the 1:4:1 loop isomer possesses a $G_3N_1G_3$ motif on the 3′- and 5′-ends, in contrast to a single $G_3N_1G_3$ motif in the 1:2:3 loop isomer. Since it is known that a double-chain reversal with two edges, each made up of three guanines and connected through a 1-base loop, is very stable and is found in a variety of promoter G-quadruplexes, it is easy to rationalize why the 1:4:1 loop isomer would predominate over the 1:2:3 loop isomer (23,24,35,36,46–48). While the pathway to the 1:4:1 loop isomer is presumably through the intermediacy of two distinct 3′ and 5′ 1-base loop faces, we speculate that the 1:2:3 loop isomer occurs when one 2-base loop forms directly adjacent to a 1-base loop, thus requiring a 3-base loop to complete the G-quadruplex.

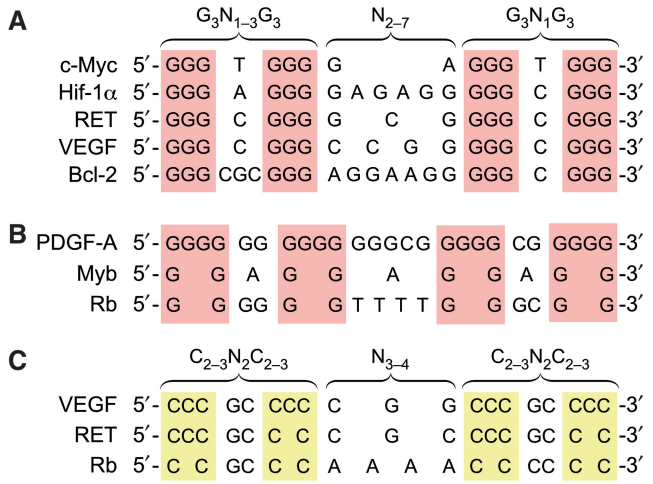

Figure 8. Comparison of truncated G-quadruplex-forming sequences (**A** and **B**) and i-motif-forming sequences (**C**) within selected gene promoters.

The DMS hypersensitivity of G14 is a unique feature of the VEGF G-quadruplex-forming element. Interpretation of the DMS footprinting pattern shows that G14 should reside in the central tetrad of the G-quadruplex. As such, it was completely unexpected that this guanine would be hypersensitive to DMS (Figure 2B). This hypersensitivity is in sharp contrast to the well-protected guanines at the adjacent positions (13 and 15) and implies that G14 must exist at some point in a non-Hoogsteen base-paired state, making this residue more accessible to attack by DMS. The DMS footprinting results of a series of oligomers containing mutant VEGF G-rich sequences suggest that G14 hypersensitivity to DMS is dependent upon the presence of G11 and G12 (Figure 2C), but not cytosines C9 and C10 (Figure 2B). However, why G14 is hyperreactive to DMS is not clear at this point, and further biophysical studies will be required to define this anomaly.

A small but growing number of promoter elements that have been shown to form G-quadruplexes, as well as a more limited number of i-motif-forming sequences, have been characterized by chemical footprinting and CD, and in a few cases NMR studies have been carried out (21,23,35,49–51). For those human promoter elements that have been shown to be critical for transcriptional activation, it is informative to compare the G-quadruplex-forming sequences and their folding patterns. The sequences for the 3-tetrad group are shown in Figure 8A. (Only the predominant G-quadruplex-forming sequence of each promoter is shown.) The similarities are striking. All the G-quadruplexes, with the exception of Bcl-2, have both a 3′- and a 5′-end $G_3N_1G_3$ motif and all form parallel G-quadruplex structures (22,24,35). The exception, Bcl-2, has a 3-base loop at the 5′-side and forms a mixed parallel/antiparallel folding pattern (23,52). However, what is strikingly different in these structures is the central loop, which differs both in size (2–7 bases) and base composition.

Outside the class of G-quadruplexes containing three tetrads are a number of structures that have either two or four tetrads (Figure 8B). In the case of PDGF-A, there are four tetrads requiring two bases to form a double-chain reversal loop on the 3′- and 5′-sides (49), so this is a simple incremental variation on the three-tetrad class that also folds to form parallel G-quadruplexes containing three double-chain reversal loops. Myb and Rb are composed of only two tetrads, and while Myb still forms a parallel-stranded G-quadruplex, it requires a heptad on one face to maintain stability (53). In contrast, the Rb forms a basket G-quadruplex with a mixed parallel/antiparallel structure (34).

In addition to that described here for VEGF, folding patterns for the i-motifs have been assigned for only RET and Rb, so the data set is more limited (Figure 8C). VEGF and RET are the same with regard to loop size (2:3:2) but differ in the number of i-motif $C–C^+$ hemiprotonated base pairs, i.e. six for VEGF and five for RET. Preliminary modeling studies suggest that the 2:3:2 loop configuration is the minimum required for bridging the antiparallel runs of cytosines (unpublished data). Rb conforms to this rule, having loop sizes of 2:4:2, but in this case only four $C–C^+$ base pairs, which may also be the minimum required for i-motif stability. So while the number of $C–C^+$ base pairs varies from four to six, the 3′- and 5′-end lateral-loop bases are two for each of these i-motifs; but in all cases the constituent bases in the central lateral loop are different.

The biological role of the i-motif on the C-rich strand in the VEGF promoter has yet to be determined. Previous studies from the Levens lab have shown that heterogeneous nuclear ribonucleoprotein K (hnRNP K) is able to bind to the C-rich strand, called the CT element, in the NHE III_1 region of the c-Myc oncogene promoter to activate c-Myc transcription (54,55). Therefore, the formation of an i-motif structure on the C-rich strand of the VEGF promoter could prevent hnRNP K binding to the single-strand to activate transcription. Thus, i-motif formation on the C-rich strand could be a silencing element for DNA transcriptional regulation.

Finally, we have also observed that G-quadruplex-interactive agents are able to stabilize the VEGF G-quadruplex in vitro and downregulate VEGF mRNA levels in cancer cells (37), suggesting that targeting the G-quadruplex formed in the VEGF promoter is a feasible strategy to modulate VEGF expression at the transcriptional level. Therefore, targeting DNA secondary structures in the VEGF promoter could be a novel approach to anti-angiogenesis drug discovery in cancer therapy.

ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health (CA109069) and the Arizona Biomedical Research Commission (0008). Dr Tracy Brooks provided valuable assistance by critiquing early versions of this article. We thank Dr Danzhou Yang for critical comments and discussion and are grateful to Dr David Bishop for preparing, proofreading and editing the final version of the article and figures. Funding to pay the Open Access publication charges for this article was provided by NIH CA109069.

Conflict of interest statement. None declared.

REFERENCES

1. Weidner,N., Semple,J.P., Welch,W.R. and Folkman,J. (1991) Tumor angiogenesis and metastasis–correlation in invasive breast carcinoma. *N. Engl. J. Med.*, **324**, 1–8.
2. Ferrara,N. (1996) Vascular endothelial growth factor. *Eur. J. Cancer*, **32A**, 2413–2422.
3. Plate,K.H., Breier,G., Weich,H.A. and Risau,W. (1992) Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo. *Nature*, **359**, 845–848.
4. Takahashi,A., Sasaki,H., Kim,S.J., Tobisu,K., Kakizoe,T., Tsukamoto,T., Kumamoto,Y., Sugimura,T. and Terada,M. (1994) Markedly increased amounts of messenger RNAs for vascular endothelial growth factor and placenta growth factor in renal cell carcinoma associated with angiogenesis. *Cancer Res.*, **54**, 4233–4237.
5. Boocock,C.A., Charnock-Jones,D.S., Sharkey,A.M., McLaren,J., Barker,P.J., Wright,K.A., Twentyman,P.R. and Smith,S.K. (1995) Expression of vascular endothelial growth factor and its receptors flt and KDR in ovarian carcinoma. *J. Natl Cancer Inst.*, **87**, 506–516.
6. Itakura,J., Ishiwata,T., Friess,H., Fujii,H., Matsumoto,Y., Buchler,M.W. and Korc,M. (1997) Enhanced expression of vascular endothelial growth factor in human pancreatic cancer correlates with local disease progression. *Clin. Cancer Res.*, **3**, 1309–1316.
7. Ladoux,A. and Frelin,C. (1993) Hypoxia is a strong inducer of vascular endothelial growth factor mRNA expression in the heart. *Biochem. Biophys. Res. Commun.*, **195**, 1005–1010.
8. Minchenko,A., Salceda,S., Bauer,T. and Caro,J. (1994) Hypoxia regulatory elements of the human vascular endothelial growth factor gene. *Cell Mol. Biol. Res.*, **40**, 35–39.
9. Forsythe,J.A., Jiang,B.H., Iyer,N.V., Agani,F., Leung,S.W., Koos,R.D. and Semenza,G.L. (1996) Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol. Cell Biol.*, **16**, 4604–4613.
10. Pertovaara,L., Kaipainen,A., Mustonen,T., Orpana,A., Ferrara,N., Saksela,O. and Alitalo,K. (1994) Vascular endothelial growth factor is induced in response to transforming growth factor-β in fibroblastic and epithelial cells. *J. Biol. Chem.*, **269**, 6271–6274.
11. Gille,J., Swerlick,R.A. and Caughman,S.W. (1997) Transforming growth factor-α-induced transcriptional activation of the vascular permeability factor (VPF/VEGF) gene requires AP-2-dependent DNA binding and transactivation. *EMBO J.*, **16**, 750–759.
12. Buteau-Lozano,H., Ancelin,M., Lardeux,B., Milanini,J. and Perrot-Applanat,M. (2002) Transcriptional regulation of vascular endothelial growth factor by estradiol and tamoxifen in breast cancer cells: a complex interplay between estrogen receptors α and β. *Cancer Res.*, **62**, 4977–4984.
13. Li,J., Perrella,M.A., Tsai,J.C., Yet,S.F., Hsieh,C.M., Yoshizumi,M., Patterson,C., Endege,W.O., Zhou,F. and Lee,M.E. (1995) Induction of vascular endothelial growth factor gene expression by interleukin-1 β in rat aortic smooth muscle cells. *J. Biol. Chem.*, **270**, 308–312.
14. Cohen,T., Nahari,D., Cerem,L.W., Neufeld,G. and Levi,B.Z. (1996) Interleukin 6 induces the expression of vascular endothelial growth factor. *J. Biol. Chem.*, **271**, 736–741.
15. Mukhopadhyay,D., Tsiokas,L. and Sukhatme,V.P. (1995) Wild-type p53 and v-Src exert opposing influences on human vascular endothelial growth factor gene expression. *Cancer Res.*, **55**, 6161–6165.
16. Pal,S., Datta,K. and Mukhopadhyay,D. (2001) Central role of p53 on regulation of vascular permeability factor/vascular endothelial growth factor (VPF/VEGF) expression in mammary carcinoma. *Cancer Res.*, **61**, 6952–6957.
17. Grugel,S., Finkenzeller,G., Weindel,K., Barleon,B. and Marme,D. (1995) Both v-Ha-Ras and v-Raf stimulate expression of the vascular endothelial growth factor in NIH 3T3 cells. *J. Biol. Chem.*, **270**, 25915–25919.
18. Finkenzeller,G., Sparacio,A., Technau,A., Marme,D. and Siemeister,G. (1997) Sp1 recognition sites in the proximal promoter of the human vascular endothelial growth factor gene are essential for platelet-derived growth factor-induced gene expression. *Oncogene*, **15**, 669–676.
19. Shi,Q., Le,X., Abbruzzese,J.L., Peng,Z., Qian,C.N., Tang,H., Xiong,Q., Wang,B., Li,X.C. and Xie,K. (2001) Constitutive Sp1 activity is essential for differential constitutive expression of vascular endothelial growth factor in human pancreatic adenocarcinoma. *Cancer Res.*, **61**, 4143–4154.
20. Abdelrahim,M., Smith,R. III, Burghardt,R. and Safe,S. (2004) Role of Sp proteins in regulation of vascular endothelial growth factor expression and proliferation of pancreatic cancer cells. *Cancer Res.*, **64**, 6740–6749.
21. 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.
22. Seenisamy,J., Rezler,E.M., Powell,T.J., Tye,D., Gokhale,V., Joshi,C.S., Siddiqui-Jain,A. and Hurley,L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126**, 8702–8709.
23. Dexheimer,T.S., Sun,D. and Hurley,L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404–5415.
24. 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 1α promoter. *Biochemistry*, **44**, 16341–16350.
25. 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.
26. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
27. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
28. Gehring,K., Leroy,J.L. and Guéron,M. (1993) A tetrameric DNA structure with protonated cytosine·cytosine base pairs. *Nature*, **363**, 561–565.
29. Ahmed,S., Kintanar,A. and Henderson,E. (1994) Human telomeric C-strand tetraplexes. *Nat. Struct. Biol.*, **1**, 83–88.
30. Manzini,G., Yathindra,N. and Xodo,L.E. (1994) Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic Acids Res.*, **22**, 4634–4640.
31. Simonsson,T., Pribylova,M. and Vorlickova,M. (2000) A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.*, **278**, 158–166.
32. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
33. 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.
34. Xu,Y. and Sugiyama,H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.*, **34**, 949–954.
35. Guo,K., Pourpak,A., Beetz-Rogers,K., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.*, **129**, 10220–10228.
36. Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
37. Sun,D., Liu,W.J., Guo,K., Rusche,J.J., Ebbinghaus,S., Gokhale,V. and Hurley,L.H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol. Cancer Ther.*, **7**, 880–889.

38. Han,H., Hurley,L.H. and Salazar,M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.*, **27**, 537–542.
39. Ross,S.A. and Burrows,C.J. (1996) Cytosine-specific chemical probing of DNA using bromide and monoperoxysulfate. *Nucleic Acids Res.*, **24**, 5062–5063.
40. Pataskar,S.S., Dash,D. and Brahmachari,S.K. (2001) Intramolecular i-motif structure at acidic pH for progressive myoclonus epilepsy (EPM1) repeat $d(CCCCGCCCCGCG)_n$. *J. Biomol. Struct. Dyn.*, **19**, 307–313.
41. Mathur,V., Verma,A., Maiti,S. and Chowdhury,S. (2004) Thermodynamics of i-tetraplex formation in the nuclease hypersensitive element of human c-myc promoter. *Biochem. Biophys. Res. Commun.*, **320**, 1220–1227.
42. Leroy,J.L., Guéron,M., Mergny,J.L. and Hélène,C. (1994) Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.*, **22**, 1600–1606.
43. Ferrara,N. and Davis-Smyth,T. (1997) The biology of vascular endothelial growth factor. *Endocr. Rev.*, **18**, 4–25.
44. Ferrara,N. and Alitalo,K. (1999) Clinical applications of angiogenic growth factors and their inhibitors. *Nat. Med.*, **5**, 1359–1364.
45. Gschwind,A., Fischer,O.M. and Ullrich,A. (2004) The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat. Rev. Cancer*, **4**, 361–370.
46. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
47. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
48. Qin,Y. and Hurley,L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, (in press).
49. Qin,Y., Rezler,E.M., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.*, **35**, 7698–7713.
50. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
51. 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.
52. 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.
53. Matsugami,A., Okuizumi,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.
54. 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.
55. Michelotti,E.F., Michelotti,G.A., Aronsohn,A.I. and Levens,D. (1996) Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell Biol.*, **16**, 2350–2360.