J|A|C|S
COMMUNICATIONS

Published on Web 01/05/2006

An Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel G-Strands Formed in the Human BCL-2 Promoter Region in Solution

Jixun Dai,[‡] Thomas S. Dexheimer,[‡] Ding Chen,[‡] Megan Carver,[‡] Attila Ambrus,[‡] Roger A. Jones,[§] and Danzhou Yang*[,‡,⊥]

College of Pharmacy, University of Arizona, Tucson, Arizona 85721, Arizona Cancer Center, Tucson, Arizona 85724, and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854

Received August 25, 2005; E-mail: yangd@pharmacy.arizona.edu

Bcl-2 (B-cell CLL/lymphoma 2) is a potent oncoprotein that plays an essential role in cell survival and functions as an inhibitor of cell apoptosis.[1] The bcl-2 proto-oncogene was first discovered in human follicular lymphoma and has been mapped to chromosome 18q21 based on a t(14;18) translocation to the immunoglobulin heavy chain (IgH) locus at 14q32.[2] Bcl-2 has been found to be aberrantly overexpressed in a wide range of human tumors, including B-cell and T-cell lymphomas, breast, prostate, cervical, colorectal, and non-small cell lung carcinomas.[3] Elevation in bcl-2 level has also been associated with poor prognoses.[3] Thus, the bcl-2 transcriptional control has emerged as an attractive target for anticancer therapeutics.

The P1 promoter located 1386−1423 base pairs upstream of the translation start site is the major transcriptional promoter for bcl-2.[4] This is a TATA-less, GC-rich promoter that contains multiple transcriptional start sites. The 5′-end of the P1 promoter, including a highly GC-rich region, has been implicated in playing a major role in the regulation of bcl-2 transcription.[4] This GC-rich element is a 39-base-pair sequence that is located 58 to 19 base pairs upstream of the P1 promoter. Deletion or mutation of this element has been shown to increase promoter activity by 2.1-fold and 2.6-fold, respectively.[4] This 39-mer guanine-rich strand (bcl2Pu39) contains six runs of guanines, with one run of five guanines, two runs of four guanines each, and three runs of three guanines each (Figure 1A). This G-rich strand bcl2Pu39 can form a mixture of three distinct intramolecular G-quadruplexes in K^+-containing solution.[5] The G-quadruplex formed on the middle four consecutive runs of guanines (bcl2MidG4Pu23, Figure 1A) has been shown to be the most stable and is suggested to be the major G-quadruplex structure formed in the bcl-2 promoter region.[5] Moreover, the bcl2MidG4Pu23 contains a run of five guanines (Figure 1A) that gives rise to the possibility of three different loop isomers, which can each be isolated by specific dual G-to-T mutations. Of the three loop isomers, the G15T/G16T dual mutant was the most stable as shown by NMR (Figure S1). The formation of this major G-quadruplex structure was also confirmed by biochemical studies. DMS footprinting data show pronounced cleavage at the G15 and G16 in potassium solution, indicating that these two guanines are not involved in the formation of the major G-quadruplex under physiological conditions.[5] Polymerase stop assays show that the major G-quadruplex is formed with the 15,16-G-to-T mutant sequence, which is markedly more stable than the minor G-quadruplex formed with the 15,19-G-to-T mutant sequence, while the 18,19-G-to-T mutant sequence does not form a stable G-quadruplex (Figure 1A).[5] Thus, a 23-mer DNA oligomer containing the middle four G-runs with G15T/G16T dual mutations (bcl2MidG4Pu23-G15T/G16T) was prepared and used for our NMR structural analysis of this major G-quadruplex formed in the bcl-2 promoter[6] (Figure 1A).

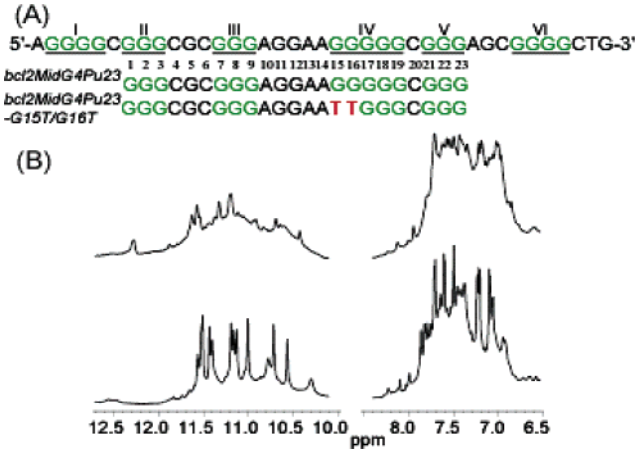

Figure 1. (A) Promoter sequence of the bcl-2 gene and its modifications. The top sequence is the wild-type bcl-2 39-mer sequence. The six G-runs are underlined and numbered using Roman numerals. Bcl2MidG4Pu23 represents the 23mer sequence containing the middle four consecutive runs of guanines, which forms the most stable G-quadruplex structure. Bcl2MidG4Pu23-G15T/G16T represents the mutant 23mer with 15,16-G-to-T mutations that forms the major G-quadruplex. (B) Imino and aromatic regions of 1D ^{1}H NMR spectra of bcl2MidG4Pu23 (upper) and bcl2MidG4Pu23-G15T/G16T (lower) at 7 °C, 20 mM K-phosphate, 40 mM KCl, pH 7.0.

Our results demonstrate a novel folding of a unique intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands. This G-quadruplex structure contains three G-tetrads connected with a single-nucleotide double-chain-reversal side loop and two lateral loops. The first three-nucleotide CGC loop in the bcl-2 promoter sequence forms a lateral loop, as opposed to a double-chain-reversal side loop observed in a similar sequence in the c-MYC promoter, and appears to largely determine the overall folding of the bcl-2 G-quadruplex. Possible rules governing various folding patterns of intramolecular G-quadruplexes are also implicated by this study.

The 1D NMR spectrum of the wild-type 23mer bcl-2 promoter sequence bcl2MidG4Pu23 in potassium solution (Figure 1B top) shows a broad envelope with some fine lines, indicating the presence of a dynamic equilibrium of multiple conformers. The imino protons at 10−12 ppm indicate the formation of G-quadruplex structures.[7] The 1D NMR spectrum of the 15,16-G-to-T mutant sequence bcl2MidG4Pu23-G15T/G16T for the major G-quadruplex structure formed in the bcl-2 promoter shows much improved line width and better resolution in potassium solution (Figure 1B, bottom). The well-resolved imino proton resonances located between 10.5 and 12 ppm clearly indicate the formation of a stable G-quadruplex structure. The melting temperature of this G-

[‡] University of Arizona.
[§] Rutgers University.
[⊥] Arizona Cancer Center.

10.1021/ja055636a CCC: $33.50

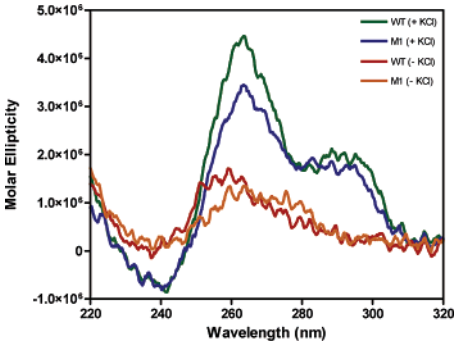

Figure 2. CD spectra of bcl2MidG4Pu23 (WT) and bcl2MidG4Pu23-G15T/G16T (M1) in the absence and presence of 100 mM KCl.

quadruplex (~75 °C) is independent of the concentration, indicating the formation of a monomeric structure (data not shown). The sharpest line widths of ~5 Hz for the imino resonances of this intramolecular G-quadruplex are observed around 20 °C. The sharp spectral line widths are also indicative of the presence of a monomeric structure. The presence of a unimolecular structure was confirmed by the EMSA experiment as well.[5] The CD spectra of the wide-type and the mutant 23mers are very similar (Figure 2).

Twelve imino peaks are present in the 10.5−12 ppm region (Figure 1B), indicating that all 12 guanines of the four consecutive guanine runs of the bcl2MidG4Pu23-G15T/G16T are involved in the intramolecular monomeric G-quadruplex formation, and that this G-quadruplex structure contains three G-tetrads. The imino and base aromatic H8 protons of the 12 guanine residues are unambiguously assigned by the site-specific low-concentration (6%) incorporation of 1, 2, 7-^{15}N, 2-^{13}C-labeled guanine nucleoside at each guanine position of the sequence.[6,8] The guanine imino H1 proton resonance has one-bond coupling to N1, and the guanine base aromatic H8 proton resonance has two-bond coupling to N7. Both the H1 and H8 protons of the site-specific labeled guanine are readily detected by 1D ^{15}N-filtered experiments. The assignment of each imino proton of the 12 guanines involved in the three G-tetrads is shown in Figure 3A. The assignment of each guanine aromatic H8 proton is shown in Figure S2.

The assignment of the imino and base H8 protons of guanines leads to the direct determination of the folding topology of this bcl-2 G-quadruplex structure. The glycosidic torsion angles of the 12 G-quadruplex guanines are defined as indicated by the intraresidue H8−H1′ NOE intensities[6] (Figure S3) and are shown in Figure 3B (bottom). In a G-tetrad plane with the Hoogsteen H-bond network, the imino proton NH1 of each guanine is in close spatial vicinity to the NH1s of the two adjacent guanines and to the base H8 of one of the adjacent guanines (Figure 3B, top). The through-space NOE connectivities of guanine H1−H1 and H1−H8 determine the arrangement and topology of a G-tetrad plane. Three G-tetrad planes, G3-G7-G19-G23, G2-G8-G18-G22, and G1-G9-G17-G21, were determined based on the NOE connectivities (Figure 3C). For example, the G3H1/G7H1, G7H1/G19H1, G19H1/G23H1, G23H1/G3H1 (GH1/GH1) and the G3H1/G7H8, G7H1/G19H8, G19H1/G23H8, G23H1/G3H8 (GH1/GH8) NOE interactions (Figure 3C) define a tetrad plane of G3-G7-G19-G23 (Figure 3B, bottom). The G-quadruplex alignment is further defined based on the inter-tetrad NOE connections from residues that position far apart in the DNA sequence. For example, the strong NOE interactions of G2H1/G9H1, G8H1/G17H1, G18H1/G21H1, and G22H1/G1H1 connect the middle and bottom G-tetrad planes, as well as reflect the reversed nucleotide glycosidic torsion angles related with the middle and bottom G-tetrads (Figures 3C and 3B). The G3H1/G8H1, G7H1/G18H1, G19H1/G22H1, and G23H1/G2H1 NOEs connect the top and middle G-tetrad planes and reflect

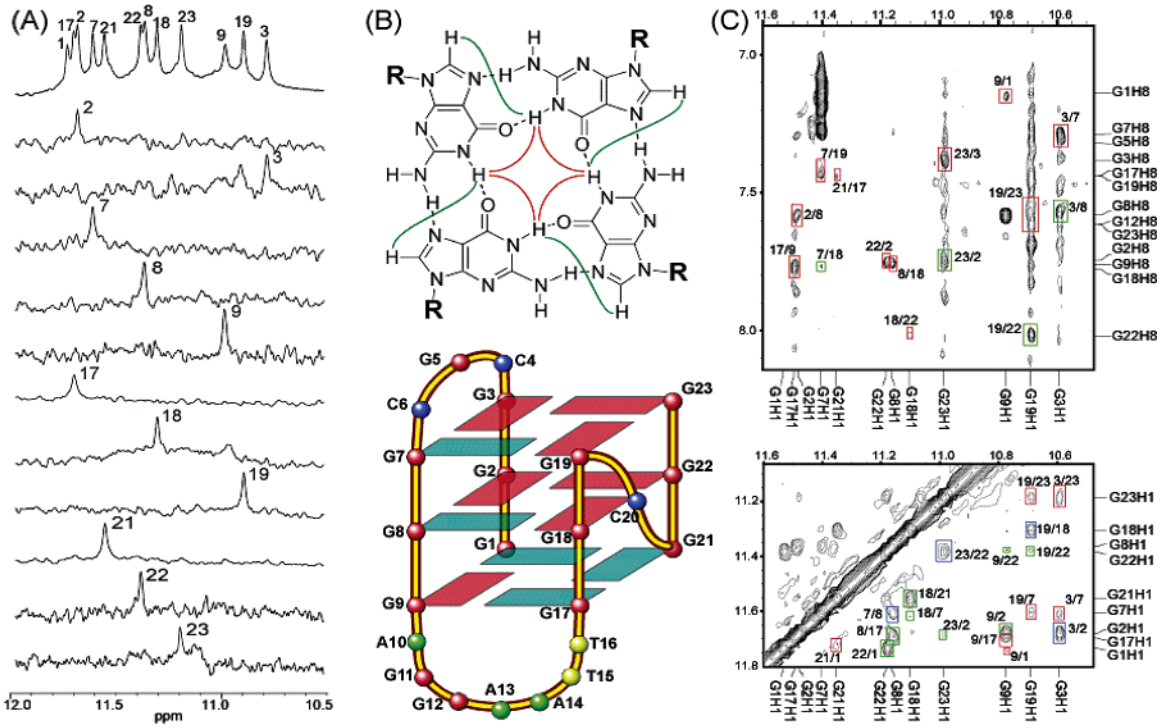

Figure 3. (A) Imino proton assignments of bcl2MidG4Pu23-G15T/G16T using 1D ^{15}N-filtered experiments on site-specific labeled oligonucleotides. The imino proton of G1 is assigned using elimination method, in combination with the exchangeable proton 2D-NOESY (see Figure 3C). (B) (Top) A G-tetrad with H1−H1 and H1−H8 connectivity pattern detectable in NOESY experiments. (Bottom) Schematic drawing of the folding topology of the bcl2MidG4Pu23-G15T/G16T G-quadruplex. Red boxes represent guanines with anti configuration, and blue boxes represent guanines with syn configuration. (C) H1−H8 region (top) and H1−H1 region (bottom) of 2D-NOESY spectrum of bcl2MidG4Pu23-G15T/G16T in H_2O at 25 °C. Red boxes represent intra-tetrad connectivity, green boxes represent inter-tetrad connectivity, and blue boxes represent sequential connectivity.

Table 1. Comparison of G-Quadruplex-Forming Sequences

Name	Sequence (5′ to 3′)						
Tet-Tel	GGG	GTTG	GGG	TTG	GGG	TT	GGG
c-MYC(a)	GGG	T	GGG	TA	GGG	T	GGG
c-MYC(b)	GGG	A	GGG	TTTTTA	GGG	T	GGG
VEGF	GGG	C	GGG	CCGG	GGG	C	GGG
HIF-1α	GGG	A	GGG	GAGAGG	GGG	C	GGG
Bcl-2	GGG	CGC	GGG	AGGAATT	GGG	C	GGG

the right-handed twist of the DNA backbone. From these data, the folding topology of the bcl2MidG4Pu23-G15T/G16T G-tetrads is determined, as shown in Figure 3B (bottom). This G-quadruplex contains mixed parallel/antiparallel G-strands, with the first, third, and fourth G-strands being parallel with each other, and the second G-strand being antiparallel with the rest of the G-strands. The first three G-strands (from 5′-end) are linked with two lateral loops (CGC and AGGAATT), while the third and fourth G-strands are linked with a single nucleotide, double-chain-reversal side loop (C). This bcl-2 G-quadruplex has a similar G-tetrad arrangement, but a different loop arrangement, as compared to that of the tetrahymena telomeric (Tet-Tel) sequence (Table 1).[9]

It is significant to note that the bcl-2 G-quadruplex represents another well-defined DNA G-quadruplex formation, in addition to c-MYC, in the promoter regions of human proto-oncogenes. In comparison with the recently determined parallel-stranded G-quadruplex structures formed in the c-MYC promoter[6,10] (Table 1, Figure S4), the third loops of both G-quadruplexes adopt single-nucleotide (nt) double-chain-reversal side loops, with residue C in bcl-2 and T in c-MYC, respectively, whereas the first and second loops adopt different conformations in the two G-quadruplex structures. In fact, one of the G-quadruplexes formed in the c-MYC promoter [c-MYC(b)] exhibits some degree of sequence similarity to the bcl-2 G-quadruplex (Table 1), both of which contain an extended second loop (7-nt in bcl-2 and 6-nt in c-MYC(b)). However, the first loop of bcl-2 is a three-nt loop (CGC) and adopts a lateral loop conformation, while the first loop of c-MYC(b) is a single-nt loop (A) and adopts a double-chain-reversal side loop conformation[10] (Figure S4). Consequently, the extended second loop of bcl-2 forms a lateral loop conformation, while that of c-MYC forms a double-chain-reversal loop. Therefore, it appears that the first three-nucleotide lateral loop determines the overall folding of the bcl-2 G-quadruplex, namely the mixed parallel/antiparallel-stranded G-quadruplex, as opposed to the parallel-stranded c-MYC G-quadruplex. Furthermore, it is interesting to note that both the bcl-2 and c-MYC promoter sequences contain the same GGGNGGG sequence motif, which forms a stable double-chain-reversal parallel-stranded structural motif. Indeed, the promoter regions of VEGF and HIF-1α also contain the same sequence motif (Table 1) and are suggested to form parallel-stranded G-quadruplexes.[10] In addition, the third double-nt loop (TT) in the Tet-Tel sequence (Table 1) forms the only double-chain-reversal side loop in the Tet-Tel G-quadruplex.[9] Thus, it appears that the single- or double-nucleotide-sized loops are more favored for the formation of the double-chain-reversal side loop. The detailed NMR solution structure determination of this predominant G-quadruplex in the bcl-2 promoter will be published in a separate full article.

In summary, our study demonstrates the formation of a unique intramolecular G-quadruplex structure in the promoter region of the human bcl-2 gene. Our NMR results define a novel folding pattern of this predominant G-quadruplex in the bcl-2 promoter, which adopts a mixed parallel/antiparallel-stranded G-quadruplex structure. This major G-quadruplex in the bcl-2 promoter represents an attractive target for the design of new anticancer drugs that specifically target this secondary structure and modulate bcl-2 gene expression. Furthermore, the bcl-2 G-quadruplex has a different folding pattern as compared to the c-MYC G-quadruplex, albeit with some degree of sequence similarity. It thus appears that different G-quadruplex structures will form based on different promoter sequences, making such regions attractive targets for pathway-specific drug design.

Acknowledgment. This research was supported by the National Institutes of Health (1K01CA83886 and 1S10 RR16659). We are grateful to Dr. Laurence H. Hurley and his lab for valuable assistance and discussion.

Supporting Information Available: Experimental methods, 1D NMR spectra of three dual G-to-T mutants of bcl2MidG4Pu23, HMQC experiments for base H8 proton assignments, and H8/H6−H1′ region of 2D-NOESY with assignments of bcl2MidG4Pu23-G15T/G16T. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Hockenbery, D.; Nunez, G.; Milliman, C.; Schreiber, R. D.; Korsmeyer, S. J. *Nature* **1990**, *348*, 334−336. (b) Vaux, D. L.; Cory, S.; Adams, J. M. *Nature* **1988**, *335*, 440−442.

(2) Yunis, J. J. *Science* **1983**, *221*, 227−236.

(3) (a) Akagi, T.; Kondo, E.; Yoshino, T. *Leuk. Lymphoma* **1994**, *13*, 81−87. (b) Joensuu, H.; Pylkkanen, L.; Toikkanen, S. *Am. J. Pathol.* **1994**, *145*, 1191−1198. (c) Tjalma, W.; De Cuyper, E.; Weyler, J.; Van Marck, E.; De Pooter, C.; Albertyn, G.; van Dam, P. *Am. J. Obstet. Gynecol.* **1998**, *178*, 113−117. (d) Pezzella, F.; Turley, H.; Kuzu, I.; Tungekar, M. F.; Dunnill, M. S.; Pierce, C. B.; Harris, A.; Gatter, K. C.; Mason, D. Y. *N. Engl. J. Med.* **1993**, *329*, 690−694. (e) McDonnell, T. J.; Troncoso, P.; Brisbay, S. M.; Logothetis, C.; Chung, L. W.; Hsieh, J. T.; Tu, S. M.; Campbell, M. L. *Cancer Res.* **1992**, *52*, 6940−6944. (f) Baretton, G. B.; Diebold, J.; Christoforis, G.; Vogt, M.; Muller, C.; Dopfer, K.; Schneiderbanger, K.; Schmidt, M.; Lohrs, U. *Cancer* **1996**, *77*, 255−264. (g) Reed, J. C.; Kitada, S.; Takayama, S.; Miyashita, T. *Ann. Oncol.* **1994**, *5* (Suppl. 1), 61−65.

(4) (a) Seto, M.; Jaeger, U.; Hockett, R. D.; Graninger, W.; Bennett, S.; Goldman, P.; Korsmeyer, S. J. *EMBO J.* **1988**, *7*, 123−131. (b) Young, R. L.; Korsmeyer, S. J. *Mol. Cell. Biol.* **1993**, *13*, 3686−3697. (c) Heckman, C.; Mochon, E.; Arcinas, M.; Boxer, L. M. *J. Biol. Chem.* **1997**, *272*, 19609−19614.

(5) Dexheimer, T. S.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2005**, submitted.

(6) (a) Dai, J. X.; Punchihewa, C.; Mistry, P.; Ooi, A. T.; Yang, D. Z. *J. Biol. Chem.* **2004**, *279*, 46096−46103. (b) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Z. *Biochemistry* **2005**, *44*, 2048−2058.

(7) (a) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164−168. (b) Matsugami, A.; Ouhashi, K.; Kanagawa, M.; Liu, H.; Kanagawa, S.; Uesugi, S.; Katahira, M. *J. Mol. Biol.* **2001**, *313*, 255−269. (c) Scaria, P. V.; Shire, S. J.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10336−10340. (d) Marathias, V. M.; Bolton, P. H. *Nucleic Acids Res.* **2000**, *28*, 1969−1977. (e) Gavathiotis, E.; Heald, R. A.; Stevens, M. F. G.; Searle, M. S. *J. Mol. Biol.* **2003**, *334*, 25−36.

(8) (a) Zhao, H.; Pagano, A. R.; Wang, W.; Shallop, A.; Gaffney, B. L.; Jones, R. A. *J. Org. Chem.* **1997**, *62*, 7832−7835. (b) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160−1161.

(9) Wang, Y.; Patel, D. J. *Structure* **1994**, *2*, 1141−1156.

(10) (a) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. *J. Am. Chem. Soc.* **2004**, *126*, 8702−8709. (b) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716. (c) Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. *Nucleic Acids Res.* **2005**, *33*, 6070−6080. (d) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry* **2005**, 44, 16341−16350.

JA055636A