JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K^+ Solution

Prashansa Agrawal,[‡] Clement Lin,[‡] Raveendra I. Mathad,[‡] Megan Carver,[‡] and Danzhou Yang*[,‡,§,∥,†]

[‡]Department of Pharmacology and Toxicology, College of Pharmacy, [†]Department of Chemistry, [§]BIO5 Institute, [∥]The Arizona Cancer Center, University of Arizona, 1703 East Mabel Street, Tucson, Arizona 85721, United States

Supporting Information

ABSTRACT: The human BCL-2 gene contains a 39-bp GC-rich region upstream of the P1 promoter that has been shown to be critically involved in the regulation of BCL-2 gene expression. Inhibition of BCL-2 expression can decrease cellular proliferation and enhance the efficacy of chemotherapy. Here we report the major G-quadruplex formed in the Pu39 G-rich strand in this BCL-2 promoter region. The 1245G4 quadruplex adopts a parallel structure with one 13-nt and two 1-nt chain-reversal loops. The 1245G4 quadruplex involves four nonsuccessive G-runs, I, II, IV, V, unlike the previously reported bcl2 MidG4 quadruplex formed on the central four G-runs. The parallel 1245G4 quadruplex with the 13-nt loop, unexpectedly, appears to be more stable than the mixed parallel/antiparallel MidG4. Parallel-stranded structures with two 1-nt loops and one variable-length middle loop are found to be prevalent in the promoter G-quadruplexes; the variable middle loop is suggested to determine the specific overall structure and potential ligand recognition site. A limit of 7 nt in loop length is used in all quadruplex-predicting software. Thus, the formation and high stability of the 1245G4 quadruplex with a 13-nt loop is significant. The presence of two distinct interchangeable G-quadruplexes in the overlapping region of the BCL-2 promoter is intriguing, suggesting a novel mechanism for gene transcriptional regulation and ligand modulation.

The BCL-2 (B-cell CLL/lymphoma 2) gene product is a mitochondrial membrane protein that plays an essential role in cell survival; it functions as an inhibitor of programmed cell death, or apoptosis.[1−3] BCL-2 has been found to be aberrantly overexpressed in a wide range of human tumors.[4−9] Elevation of BCL-2 levels has also been associated with poor prognosis and has been found to interfere with the traditional cancer therapeutics.[10,11] Inhibition of BCL-2 expression by small molecules,[12,13] peptidomimetics,[14] or antisense oligonucleotides[15,16] has been shown to reduce cellular proliferation and to enhance chemotherapy efficacy. It has been shown that gene amplification or translocation can be equally common mechanisms causing BCL-2 overexpression in human cancer cells;[17] thus, effective modulation of BCL-2 expression offers promise for cancer therapeutics.

The human BCL-2 gene has two promoters, P1 and P2. The major promoter, P1, is located 1386−1423 base pairs upstream of the translation start site.[18] This is a TATA-less, GC-rich promoter that contains multiple transcriptional start sites and is positioned within a nuclease hypersensitive site. The 5′-end of the P1 promoter has been implicated in playing a major role in the regulation of BCL-2 transcription,[19] including a 39-bp GC-rich element that is located 57−19 base pairs upstream of the P1 promoter. Multiple transcription factors have been reported to bind to or regulate BCL-2 gene expression through this region, such as Sp1,[18] WT1,[20] E2F,[21] and NGF.[22] We have shown that the guanine-rich strand of the DNA in this region can form G-quadruplex structures, which can be stabilized by G-quadruplex-interactive agents.[23,24] A second G-quadruplex-forming sequence has been found in the BCL-2 P1 promoter region; however, its stability appears to be much lower.[25] The 39-mer G-rich strand of this GC-rich element in the BCL-2 promoter region contains six runs of 3−5 consecutive guanines (Pu 39, Figure 1A) and has the potential to form 15 intramolecular G-quadruplex structures, using different combinations of four G-tracts. Our previous studies have shown that, on the three segments comprising the four successive runs of guanines of Pu39, i.e., I−IV (1234), II−V (2345), and III−VI (3456), the G-quadruplex formed on the central four G-runs (MidG4) appears to be the most stable in K^+.[24] We have determined the folding pattern and molecular structure of MidG4, which forms a three-tetrad mixed parallel/antiparallel G-quadruplex with three loops of 1, 7, and 3 nt.[23,26]

However, when we carried out dimethyl sulfate (DMS) footprinting on the full-length Pu39 in the presence of K^+, we found that the G-run III, which is required for the MidG4 structure, is cleaved by DMS (Figure 1B left). In contrast, the G-runs I, II, IV, V are clearly protected from DMS cleavage (Figure 1B left). DMS footprinting can probe the formation of G-quadruplexes as the guanine residues involved in the G-tetrad formation are protected from the DMS cleavage.[27] Thus, it is indicated that the major G-quadruplex formed in K^+ in the full-length Pu39 involves the four nonsuccessive G-runs, i.e., I, II, IV, and V, which we named 1245G4. We then carried out DMS footprinting of the wild-type Pu30 sequence comprising the 5′ five G-runs. The results showed that, consistent with the full-length Pu39 data, the major G-quadruplex formed in Pu30 did not involve the G-run III for the tetrad formation, or the 5′ two guanines of the G-run IV (Figure 1B right). Therefore, we

Received: November 21, 2013
Published: January 22, 2014

   dx.doi.org/10.1021/ja4118945 | J. Am. Chem. Soc. 2014, 136, 1750−1753

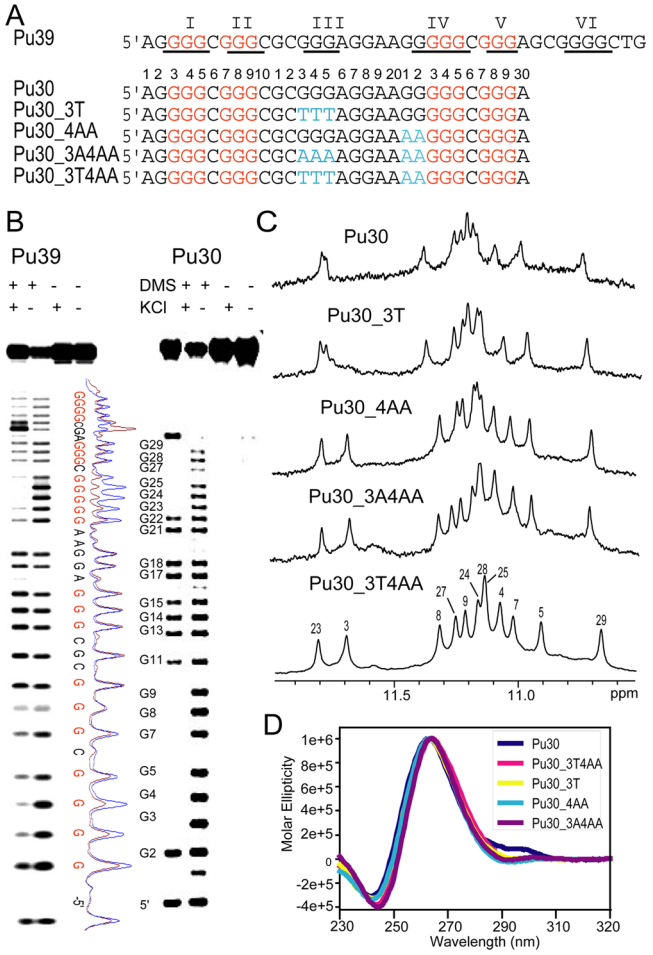

Figure 1. (A) The promoter sequence of the BCL-2 gene and its modifications. The top sequence is the 39-mer wild-type G-rich sequence (Pu39). The six G-runs with three or more guanines are underlined and numbered. Pu30 is the wild-type 30-mer G-rich sequence containing the I–V G-runs; the numbering used in this study is shown for Pu30. The guanine residues that are involved in the tetrad formation of the major BCL-2 G-quadruplex 1245G4 are shown in red. The mutations are shown in cyan. (B) DMS footprinting of the wild-type Pu39 with densitometric scans (left) and Pu30 (right). (C) Imino regions of 1D ^{1}H NMR spectra of BCL-2 promoter sequences at 25 °C in 45 mM K^{+}, pH 7.0. (D) CD spectra of Pu30 sequences in 95 mM K^{+}.

prepared Pu30 sequences with modifications at G-runs III and IV (Figure 1A), including Pu30_3T with G-to-T mutations at G-run III, Pu30_4AA with G-to-A mutations at the first two guanines (G21 and G22) of G-run IV, and Pu30_3T4AA with mutations at both G-runs III and IV. The 1D ^{1}H NMR spectra of the Pu30 sequences in K^{+} solution are shown in Figure 1C. Importantly, the imino proton regions at 10–12 ppm, characteristic of the formation of G-quadruplex structures,[26,28] were very similar for all Pu30 sequences (Figure 1C), indicating that the same G-quadruplex is formed, and that G-run III and G21 and G22 of G-run IV were not involved in the tetrad formation. We have also carried out CD study of the wild-type and mutant Pu30 sequences, which were almost identical (Figure 1D). Therefore, the NMR and CD results were in good agreement with the DMS footprinting data (Figure 1B), that the G-run III and first two guanines of G-run IV are not involved in the G-tetrad formation of the major G-quadruplex (1245G4) formed in the BCL-2 promoter. We have also recorded 1D ^{1}H NMR spectra of Pu39 and Pu30 (Figure S1), which showed that the imino protons of Pu30 could be detected in the major well-defined conformation in Pu39, although the NMR spectrum of Pu39 is not as well-resolved and appears to have higher-order structures co-present.

Pu30_3T4AA showed the best NMR spectral quality (Figure 1C), and was chosen for NMR structural analysis. For other sequences, however, the well-resolved spectrum can only be observed for the fresh sample at low DNA concentration; at higher concentration or with longer time, higher-order structures appeared to form as indicated by clearly elevated baseline (Figure S2). We prepared Pu30_3T4AA oligonucleotides with 6% incorporation of ^{15}N-labeled guanine at each guanine position. The H1 proton of guanine is one-bond connected to N1, the guanine H8 proton is two-bond connected to N7 (Figure 2A); they are both readily detected for the site-specific labeled guanine by 1D ^{15}N-edited NMR experiments. With such experiments, the H1 protons (Figure 2B left) and H8 protons (Figure 2B right) of each guanine were unambiguously assigned. Pu30_3T4AA contains 16 guanines (Figure 1A); only 12 imino peaks are observed in the 10.5–12 ppm region (Figure 2B left), indicating the formation of a three-tetrad G-quadruplex. G2 of G-run I, G11, G17, and G18 are not involved in the tetrad formation, consistent with the DMS footprinting data (Figure 1B right).

The assignment of the imino H1 and base H8 protons of guanines led to the direct determination of the folding topology of 1245G4 in the BCL-2 promoter (Figure 2C). A G-tetrad plane is connected with Hoogsteen H-bonds; the imino H1 protons of adjacent guanines, as well as the imino H1 and one adjacent guanine H8 protons, are in close spatial vicinity and detectable by NOE connections (Figure 2A). The arrangement and topology of a G-tetrad plane can thus be determined by guanine H1–H1 and H1–H8 NOEs (Figure 2C). For example, the NOE interactions between G3H1/G7H1, G7H1/G23H1, G23H1/G27H1, and G27H1/G3H1 (Figure 2D), and G7H8/G3H1, G23H8/G7H1, G27H8/G23H1, and G3H8/G27H1 (Figure 2E) defined a G-tetrad plane of G3-G7-G23-G27 (Figure 2C). The other two G-tetrad planes, i.e., G4-G8-G24-G28 and G5-G9-G25-G29 (Figure 2C), can be similarly defined. The G-quadruplex folding is further supported by intertetrad NOEs. These NOEs, such as G3H8/G28H1, G7H8/G4H1, G23H8/G8H1, and G27H8/G24H1 (Figure 2E), connect the top and middle G-tetrad planes (Figure 2C), while G4H8/G29H1, G8H8/G5H1, G24H8/G9H1, and G28H8/G25H1 (Figure 2E) connect the middle and bottom planes (Figure 2C), and reflect the right-handed twist of the DNA backbone. Our NMR results thus showed that the 1245G4 of the BCL-2 promoter is a 3-tetrad, parallel-stranded, intramolecular G-quadruplex structure, with three chain-reversal loops of 1, 13, and 1 nt, respectively (Figure 2C). All the tetrad-guanines are in *anti* glycosidic configuration, as shown by the medium intensities of intraresidue H8–H1′ NOEs (Figure S3). The CD spectra of Pu30 sequences showed a positive maximum at 264 nm and a negative minimum at 240 nm (Figure 1D), which are characteristic of parallel G-quadruplex structures,[29] supporting the parallel-stranded folding of the 1245G4. The melting temperature of Pu30_3T4AA was determined to be 71 °C as measured by the CD melting experiments at 264 nm (Figure S4) and NMR variable temperature study (Figure S5). The G-quadruplex formed by Pu30_3T4AA appeared to be of unimolecular nature as indicated by the concentration-independent melting temperature measured by CD and NMR.

The presence of a 13-nt loop in the major G-quadruplex 1245G4 formed in the BCL-2 promoter is unexpected and

dx.doi.org/10.1021/ja4118945 | *J. Am. Chem. Soc.* 2014, 136, 1750–1753

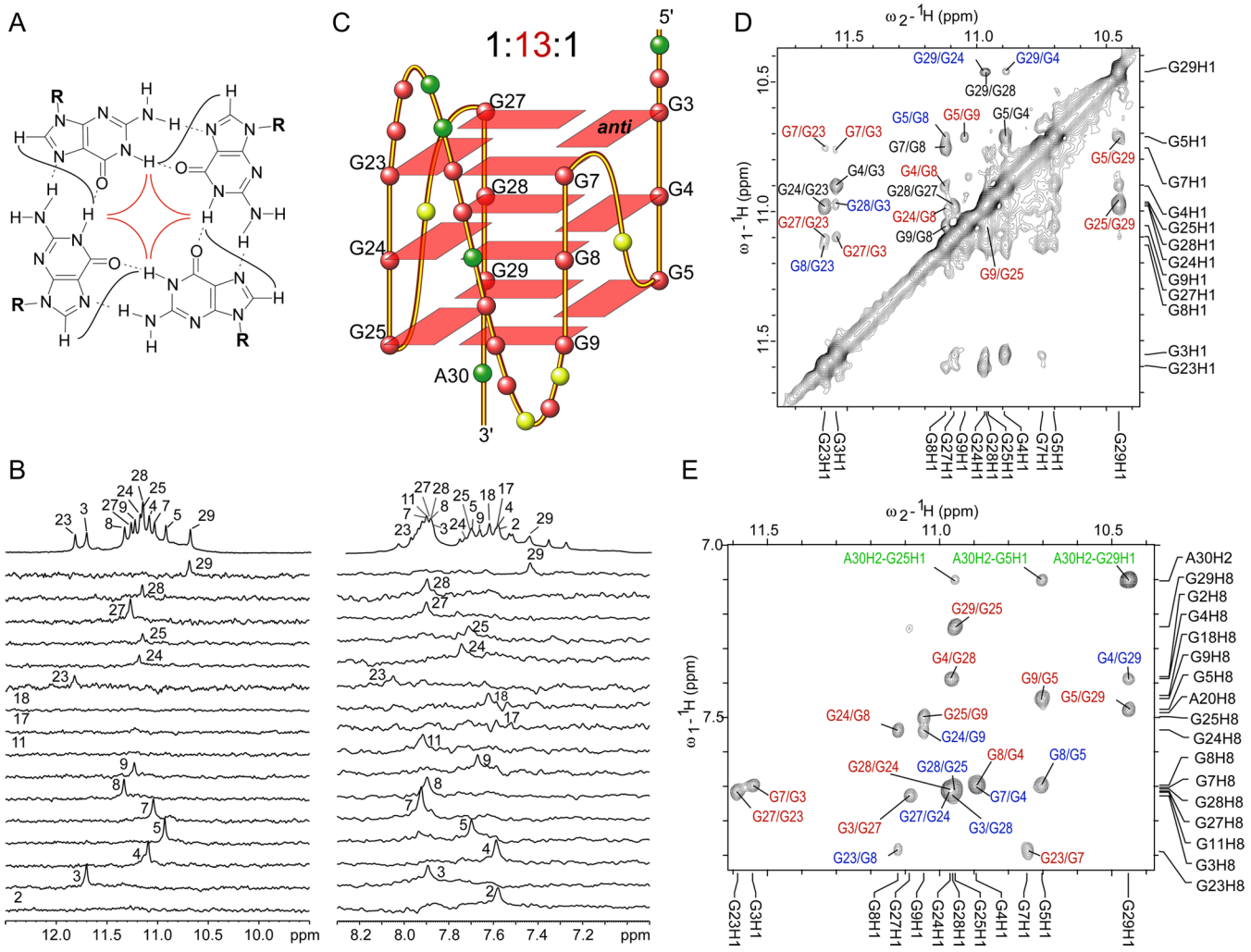

Figure 2. (A) A G-tetrad with detectable H1−H1 and H1−H8 NOE connectivity. (B) Imino H1 and aromatic H8 proton assignments of Pu30_3T4AA by 1D ^{15}N-filtered experiments using site-specific labeled oligonucleotides at 25 °C. (C) Schematic drawing of the folding topology of the major G-quadruplex 1245G4 formed in the BCL-2 promoter sequence Pu30 (G = red, A = green, C = yellow). (D) H1−H1 region and (E) H1−H8 region of the 2D-NOESY spectrum of Pu30_3T4AA in H_2O at 5 °C with a mixing time of 200 ms. The proton assignments are shown on the sides. Intratetrad NOEs are in red, intertetrad NOEs in blue, sequential intertetrad NOEs in black, and NOEs with flanking bases in green. Conditions: 25 mM K-phosphate, 70 mM KCl, pH 7.0.

remarkable. DNA G-quadruplex secondary structures have been found to be overpresented in the human gene promoter regions as transcriptional regulators and are considered as a novel class of molecular targets for cancer therapeutics.[30] Intriguingly, parallel-stranded structures are found to be prevalent in the promoter G-quadruplexes, including those of c-MYC,[31−33] VEGF,[28] c-KIT21,[34] HIF-1α,[35] RET,[36] and hTERT[37,38] (Figure 3). It has been shown that the chain-reversal loops in parallel G-quadruplexes greatly favor the short loop lengths, such as the

		G3NG3				G3NG3		
Myc2345	5'-	GGG	T	GGG	GA	GGG	T	GGG -3'
Myc1234	5'-	GGG	A	GGG	TG	GGG	A	GGG -3'
Myc1245	5'-	GGG	A	GGG	TGGGGA	GGG	T	GGG -3'
VEGF	5'-	GGG	C	GGG	CCGG	GGG	C	GGG -3'
HIF-1α	5'-	GGG	A	GGG	AGAGG	GGG	C	GGG -3'
c-KIT21	5'-	GGG	C	GGG	CGCGA	GGG	A	GGG -3'
RET	5'-	GGG	C	GGG	GCG	GGG	C	GGG -3'
hTERT	5'-	GGG	A	GGG	GCT	GGG	A	GGG -3'
Bcl2-1245	5'-	GGG	C	GGG	CGCGGGAGGAAGG	GGG	C	GGG -3'

Figure 3. G-quadruplex-forming promoter sequences.

G_3NG_3 motif with 1-nt loop.[39−42] The G_3NG_3 forms a robust parallel-stranded structural motif with 1-nt loop, which was first shown in the major c-MYC promoter quadruplex structure,[31] and is present in all other loop isomers formed in the c-MYC promoter.[29,43] Indeed, most parallel-stranded promoter G-quadruplexes contain three tetrads and three chain-reversal loops, including two 1-nt loops and a variable-length middle loop (Figure 3). In the major c-MYC promoter G-quadruplex, a parallel structure with two 1-nt loops and one 2-nt middle loop, the 2-nt middle loop stays entirely in the groove.[31] In our recent study of the major human VEGF promoter G-quadruplex, a parallel structure with two 1-nt loops and a 4-nt middle loop (Figure 3), the 4-nt middle loop was found to stretch over the 5′ tetrad forming a unique capping structure with the 5′ flanking segment.[28] It is thus suggested that each parallel G-quadruplex adopts unique capping and loop structures by its specific middle loop and flanking segments, which likely determine the specific recognition sites of proteins or small molecules. By having two 1-nt loops, it appears that a stable parallel G-quadruplex can contain a more extended middle loop. However, the longest loop that has been reported to-date in a parallel G-quadruplex

dx.doi.org/10.1021/ja4118945 | J. Am. Chem. Soc. 2014, 136, 1750−1753

structure formed on naturally occurring sequences is 6 nt long.[33] In fact, in all the available G-quadruplex-predicting software, a limit of 7 nt is used for loop lengths.[44,45] The formation and high stability of the 1245G4 G-quadruplex with a 13-nt chain-reversal loop in the BCL-2 promoter sequence is thus significant and would expand the current knowledge of DNA G-quadruplexes.

It is highly intriguing that the two stable intramolecular structures formed in the BCL-2 promoter, i.e., 1245G4 and MidG4, adopt completely different folding structures. The presence of two distinct interchangeable G-quadruplexes in the overlapping region of the BCL-2 promoter could be important for the precise regulation of gene transcription, as each G-quadruplex is likely to be recognized by different proteins leading to different gene modulation. While thermodynamically the 1245G4 quadruplex is slightly more stable than MidG4 (T_m of 66 °C), the MidG4 quadruplex could be kinetically more favored due to its shorter loop-lengths. In addition, the two interchangeable G-quadruplexes may be recognized by different small molecules and proteins, which could differentially modulate BCL-2 gene transcription.

■ ASSOCIATED CONTENT

Supporting Information

Materials and methods, 1D ^{1}H NMR spectra of Pu39 and Pu30 sequences, and H8/H6−H1′ region of the nonexchangeable 2D-NOESY, CD melting and NMR VT studies, of Pu30_3T4AA. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

yang@pharmacy.arizona.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research is supported by the National Institutes of Health (CA122952 and GM083117 to D.Y.).

■ REFERENCES

(1) Hockenbery, D.; Nunez, G.; Milliman, C.; Schreiber, R. D.; Korsmeyer, S. J. *Nature* **1990**, *348*, 334−336.
(2) Vaux, D. L.; Cory, S.; Adams, J. M. *Nature* **1988**, *335*, 440−442.
(3) Yunis, J. J. *Science* **1983**, *221*, 227−236.
(4) Akagi, T.; Kondo, E.; Yoshino, T. *Leuk. Lymphoma* **1994**, *13*, 81−87.
(5) Joensuu, H.; Pylkkanen, L.; Toikkanen, S. *Am. J. Pathol.* **1994**, *145*, 1191−1198.
(6) 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.
(7) Pezzella, F.; Turley, H.; Kuzu, I.; Tungekar, M. F.; Dunnill, M. S.; et al. *N. Engl. J. Med.* **1993**, *329*, 690−694.
(8) McDonnell, T. J.; Troncoso, P.; Brisbay, S. M.; Logothetis, C.; Chung, L. W.; et al. *Cancer Res.* **1992**, *52*, 6940−6944.
(9) Baretton, G. B.; Diebold, J.; Christoforis, G.; Vogt, M.; Muller, C.; et al. *Cancer* **1996**, *77*, 255−264.
(10) Desoize, B. *Anticancer Res.* **1994**, *14*, 2291−2294.
(11) Reed, J. C.; Kitada, S.; Takayama, S.; Miyashita, T. *Ann. Oncol.* **1994**, *5* (Suppl 1), 61−65.
(12) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; et al. *Nature* **2005**, *435*, 677−681.
(13) Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X.; et al. *J. Med. Chem.* **2001**, *44*, 4313−4324.
(14) Tzung, S. P.; Kim, K. M.; Basanez, G.; Giedt, C. D.; Simon, J.; et al. *Nat. Cell Biol.* **2001**, *3*, 183−191.
(15) Marshall, J.; Chen, H.; Yang, D.; Figueira, M.; Bouker, K. B.; et al. *Ann. Oncol.* **2004**, *15*, 1274−1283.
(16) Klasa, R. J.; Gillum, A. M.; Klem, R. E.; Frankel, S. R. *Antisense Nucleic Acid Drug Dev.* **2002**, *12*, 193−213.
(17) Rantanen, S.; Monni, O.; Joensuu, H.; Franssila, K.; Knuutila, S. *Leuk. Lymphoma* **2001**, *42*, 1089−1098.
(18) Seto, M.; Jaeger, U.; Hockett, R. D.; Graninger, W.; Bennett, S.; Goldman, P.; Korsmeyer, S. J. *EMBO J.* **1988**, *7*, 123−131.
(19) Young, R. L.; Korsmeyer, S. J. *Mol. Cell. Biol.* **1993**, *13*, 3686−3697.
(20) Heckman, C.; Mochon, E.; Arcinas, M.; Boxer, L. M. *J. Biol. Chem.* **1997**, *272*, 19609−19614.
(21) Gomez-Manzano, C.; Mitlianga, P.; Fueyo, J.; Lee, H. Y.; Hu, M.; et al. *Cancer Res.* **2001**, *61*, 6693−6697.
(22) Liu, Y. Z.; Boxer, L. M.; Latchman, D. S. *Nucleic Acids Res.* **1999**, *27*, 2086−2090.
(23) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. Z. *J. Am. Chem. Soc.* **2006**, *128*, 1096−1098.
(24) Dexheimer, T. S.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2006**, *128*, 5404−5415.
(25) Onyshchenko, M. I.; Gaynutdinov, T. I.; Englund, E. A.; Appella, D. H.; Neumann, R. D.; Panyutin, I. G. *Nucleic Acids Res.* **2009**, *37*, 7570−7580.
(26) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. Z. *Nucleic Acids Res.* **2006**, *34*, 5133−5144.
(27) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593−11598.
(28) Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. *Nucleic Acids Res.* **2013**, *41*, 10584−10592.
(29) Hatzakis, E.; Okamoto, K.; Yang, D. Z. *Biochemistry* **2010**, *49*, 9152−9160.
(30) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.
(31) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Z. *Biochemistry* **2005**, *44*, 2048−2058.
(32) 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.
(33) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.
(34) Hsu, S. T.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2009**, *131*, 13399−13409.
(35) De Armond, R.; Wood, S.; Sun, D. Y.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry* **2005**, *44*, 16341−16350.
(36) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2007**, *129*, 10220−10228.
(37) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878−10891.
(38) Lim, K. W.; Lacroix, L.; Yue, D. J. E.; Lim, J. K. C.; Lim, J. M. W.; Phan, A. T. *J. Am. Chem. Soc.* **2010**, *132*, 12331−12342.
(39) Chen, Y.; Yang, D. Z. *Curr. Protoc. Nucleic Acid Chem.* **2012**, *50*, 17.15.11−17.15.17.
(40) Rachwal, P. A.; Brown, T.; Fox, K. R. *FEBS Lett.* **2007**, *581*, 1657−1660.
(41) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901−2907.
(42) Bugaut, A.; Balasubramanian, S. *Biochemistry* **2008**, *47*, 689−697.
(43) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. Z. *Nucleic Acids Res.* **2011**, *39*, 9023−9033.
(44) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908−2916.
(45) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. *Nucleic Acids Res.* **2010**, *38*, 7858−7868.