JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

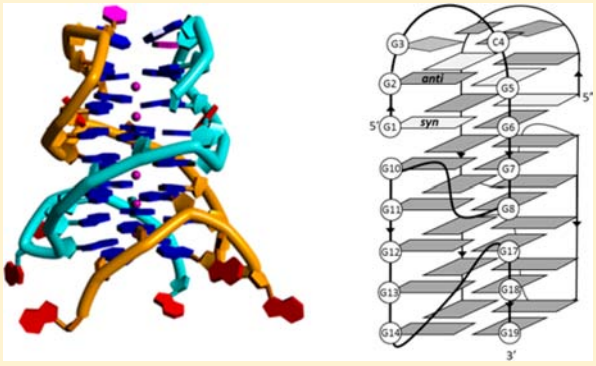
Crystal Structure of a Promoter Sequence in the *B-raf* Gene Reveals an Intertwined Dimer Quadruplex

Dengguo Wei, Alan K. Todd, Mire Zloh, Mekala Gunaratnam, Gary N. Parkinson, and Stephen Neidle*

UCL School of Pharmacy, University College London, London WC1N 1AX, United Kingdom

Supporting Information

ABSTRACT: The sequence d(GGGCGGGGAGGGGGAAGGGA) occurs in the promoter region of the *B-raf* gene. An X-ray crystallographic study has found that this forms an unprecedented dimeric quadruplex arrangement, with a core of seven consecutive G-quartets and an uninterrupted run of six potassium ions in the central channel of the quadruplex. Analogy with previously reported promoter quadruplexes had initially suggested that in common with these a monomeric quadruplex was to be expected. The structure has a distorted G·C·G·C base quartet at one end and four flipped-out adenosine nucleosides at the other. The only loops in the structure are formed by the cytosine and by the three adenosines within the sequence, with all of the guanosines participating in G-quartet formation. Solution UV and circular dichroism data are in accord with a stable quadruple arrangement being formed. 1D NMR data, together with gel electrophoresis measurements, are consistent with a dimer being the dominant species in potassium solution. A single-chain intramolecular quadruplex has been straightforwardly constructed using molecular modeling, by means of a six-nucleotide sequence joining 3′ and 5′ ends of each strand in the dimer. A human genomic database search has revealed a number of sequences containing eight or more consecutive short G-tracts, suggesting that such intramolecular quadruplexes could be formed within the human genome.



INTRODUCTION

The concept that guanine-rich polymeric nucleic acid sequences can form four-stranded helical structures originates with the fiber-diffraction studies of Gellert, Lipsett, and Davies.[1] The underlying motif in such structures is the G-quartet, comprising four in-plane guanine bases held together by eight hydrogen bonds.[2] Subsequent crystallographic and NMR studies have shown that discrete oligonucleotide sequences containing runs of (typically) four short G-tract repeats can form analogous four-stranded, highly stable structures, termed quadruplexes.[3,4] Telomeric DNA at the ends of eukaryotic chromosomes has been of particular interest since it consists of quadruplex-forming sequences, typified by the human telomeric tandem repeat d(TTAGGG).[5] G-tract repeating sequences can be found naturally within human and other genomes, with bioinformatics studies showing that putative quadruplex-forming sequences (typically represented as $G_mX_nG_mX_oG_mX_pG_m$, where X_{n-p} are intervening sequences of varying length and composition) are widely present in the human and other genomes.[6] There is currently considerable interest in understanding their (diverse) biological roles and in exploiting them for therapeutic and diagnostic ends.[7] More recently the existence of quadruplexes within human cells has been validated using labeled antibody and cross-linking approaches.[8] They are over-represented in promoter sequences, especially of genes involved in proliferation, notably in many oncogenes and proto-oncogenes implicated in human cancers.[9] Although it is common for such a sequence to comprise more than four G-tracts, which may contain several potential overlapping sites separated by length of duplex DNA, particular attention has focused on simple four G-repeat sequences that have been shown, at least in vitro, to form stable quadruplexes. Examples include promoter quadruplexes within the *c-myc*,[10] *c-kit*,[11] *k-ras*,[12] *VEGF*,[13] *hTERT*,[14] *HIF*,[15] and androgen receptor[16] genes, all of which contain more than one quadruplex-forming site. The *hTERT* promoter has been found to be mutated in a number of human cancers, including melanoma, where mutations that occur in putative quadruplex-forming regions of the promoter result in altered expression of hTERT,[17] providing further evidence of the significant role played by quadruplexes in transcriptional regulation. All of these promoter quadruplexes have been reported as existing in a monomeric form.

It is now apparent that quadruplex three-dimensional structures can vary widely in terms of how the G-quartet motifs are held together. The intervening sequences can themselves introduce added structural features and complexity, as observed in the *c-kit* quadruplex structures that have been characterized.[18−20] In general, though, a particular promoter quadruplex sequence comprising four G-tracts is normally assumed to form a discrete intramolecular quadruplex. This assumption is consistent with all the molecular structures reported to date, notably NMR structures for quadruplexes

Received: October 7, 2013
Published: December 2, 2013

from the *c-myc*,[21] *c-kit*,[18,19] and *bcl2*[22] oncogenes, and a crystal structure of one of the two *c-kit* quadruplexes.[20]

The concept of targeting promoter quadruplexes with small molecules has been proposed as a novel approach to the down-regulation of gene expression.[9c,10b] For cancer targets, this approach could, for example, be used to inhibit otherwise up-regulated signal transduction pathways, leading to potential decreased tumor growth. A wide variety of small molecules has been examined to develop such a therapeutic strategy. Most of these have been obtained from screens of small compound libraries and/or devised from general quadruplex structural considerations rather than from detailed structural knowledge of an individual targeted quadruplex.[23] NMR structures are currently available[21b,24] for two small-molecule promoter quadruplexes complexes, both involving sequences in the *c-myc* gene promoter.

As part of an effort to establish possible quadruplex targets in a number of genes of therapeutic relevance yet unreported to contain potential quadruplex-forming sequences, we report here on structural studies of a sequence from the *B-raf* gene. This gene encodes a kinase whose aberrant expression when mutated is implicated in a number of human cancers, notably malignant melanoma.[25] A number of small-molecule drugs (notably Sorafenib and Vemurafenib) have recently been developed that target the B-raf kinase,[26] and these have shown highly promising results in clinical trial, although issues of resistance remains an obstacle to long-term clinical efficacy. A search of the ENSEMBL database using our in-house software[6a,27] has located a putative G-quadruplex sequence on the opposite strand of the sequence d($C_3TTC_5TC_4GC_3$), five bases from the transcription start site within the 5′-untranslated region of the protein coding transcript ID ENST00000288602. The sequence is also close to two other *B-raf* transcripts in the ENSEMBL database, being 31 bases upstream of the transcription start site of transcript ID ENST00000469930 and 82 bases upstream of the transcription start site of transcript ID ENST00000497784. The similar sequence d($C_5TTC_5GCTC_7GCAC_5$) occurs 180 bases upstream of this region. Several other loci in this gene, such as in a number of introns, also contain putative quadruplex sequences, but this particular promoter sequence is a plausible target for transcriptional regulation in view of its proximity to the transcription start site. We report here on a crystal-structure analysis of the complementary strand of this sequence together with associated biophysical studies. These demonstrate that this sequence forms both in the crystalline state and in solution a highly stable bimolecular quadruplex arrangement.

MATERIALS AND METHODS

Crystallization and Data Collection. The *B-raf* DNA sequence used for crystallization trials was purchased from MWG (Germany) (HPLC purified). A 2 mM solution of the sequence d(GGGCGGGG-AGGGGGAAGGGA), containing 20 mM potassium cacodylate buffer at pH 6.5 and 50 mM KCl, was heated to 368 K before annealing by slow cooling to room temperature. Crystals were grown by the hanging-drop vapor-diffusion method. A volume of 1 μL of a premixed drop solution containing 10% MPD, 10 mM $MgCl_2$, 100 mM NaCl, and 50 mM sodium cacodylate at pH 6.5 was added to 1 μL of 1 mM B-raf oligonucleotide solution. The drop solution was equilibrated against a well solution at 20 °C containing 50% MPD, 20 mM $MgCl_2$, 200 mM NaCl, and 50 mM sodium cacodylate. Crystals appeared after 3 weeks. These had large thin flat plate morphology. Two data sets were collected at 105 K on two single flash-frozen crystals at the ESRF and the Diamond synchrotron facility respectively. Data were processed and scaled using Xia2 and Scala software in the CCP4 package.[28]

Structure Solution and Refinement. A starting-point G-quartet geometry was obtained from the human telomeric quadruplex crystal structure[29] (PDB code 1K8P) by editing-out the loops, and two revised quadruplex models, each containing three contiguous G-quartets, were fitted into one asymmetric unit by molecular replacement methods using the PHASER program.[30] With guidance from the $2F_o - F_c$ and $F_o - F_c$ difference electron density maps, two close strands from the top and the bottom quadruplexes were connected, and one strand in each quadruplex was split. Other parts of the quadruplex structures were progressively built on the basis of the difference electron density maps derived from the 2.37 Å diffraction data collected at the ESRF synchrotron. The 1.99 Å data set collected at the Diamond facility was used for high-resolution refinement, and the orientation of the guanine and cytosine residues on the loop was finally confirmed during this process. After further refinement, a total of 31.5 water molecules (ie 31 were assigned full occupancy and one was assigned 0.5 occupancy) were included in the final structure. Model building and restrained refinement were performed using the Coot[31] and Refmac5[32] programs respectively. Final R and R_{free} values are 0.213 and 0.242, respectively. Coordinates and structure factor data have been deposited in the Protein Data Bank with the PDB id 4H29. Programs Chimera[33] (http://www.cgl.ucsf.edu/chimera/) and Discovery Studio Visualizer (www.accelyrs.com) were used for final structure visualization and structural analysis. The crystallographic data is detailed in Table 1.

Table 1. Crystallographic Data Collection and Refinement Statistics for the *B-raf* Quadruplex[a]

sequence	d(GGGCGGGGAGGGGGAAGGGA)
data collection	
space group	$C222_1$
cell dimensions	
a, *b*, *c* (Å)	33.460, 47.680, 137.320
wavelength (Å)	0.9198
resolution (Å)	25.44−1.99
R_{merge}	0.055 (0.641)
I/σ	10.4 (1.8)
completeness (%)	98.3 (97.4)
redundancy	4.0 (4.1)
total no. of reflections	30 813 (2285)
no. of unique reflections	7754 (557)
refinement	
resolution (Å)	25.44 --1.99
no. of reflections	7374
R_{work}/R_{free}	0.213/0.242
no. of ions	6
no. of water molecules	31.5
overall B factor (Å^2)	50.43
rms deviations in:	
bond length (Å)	0.008
bond angle (deg)	0.86
PDB ID	4H29

[a]Values in parentheses refer to the highest resolution shell, 2.04−1.99 Å

UV Melting. A significant hyperchromic shift at 295 nm upon melting is characteristic for G-quadruplex structures. Ultraviolet melting point experiments with the B-raf sequence were conducted on a Varian Bio-300 UV/vis spectrophotometer with a Varian temperature controller. Samples were prepared to final oligonucleotide concentrations of 2−20 μM in the pH 7.0 buffer solution containing 10, 65, 120 mM, and 200 mM K^+. A volume of 1.5 mL of DNA

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319−19329

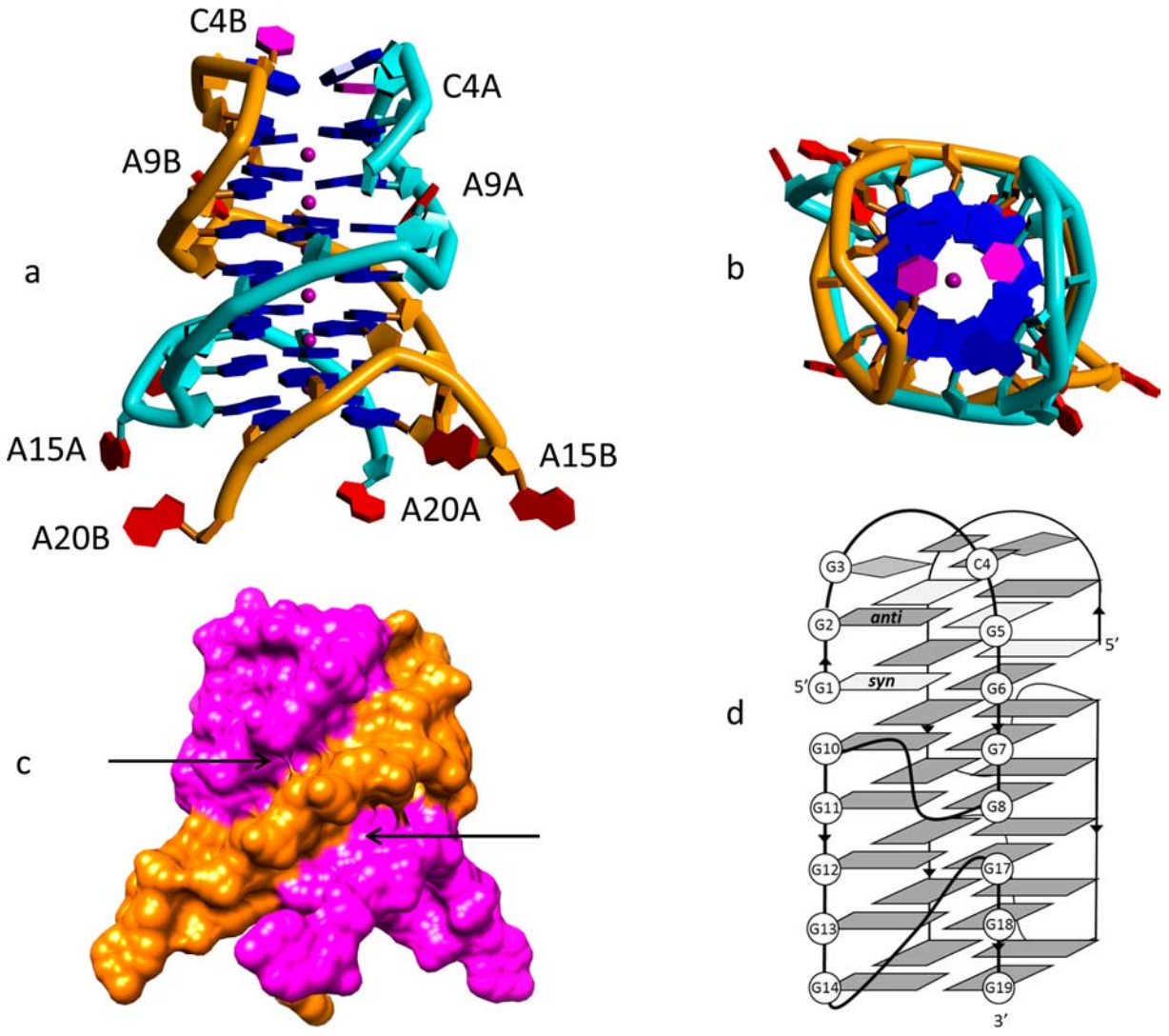


Figure 1. (a) Cartoon representation of the *B-raf* quadruplex, highlighting the two strands in the crystallographic asymmetric unit. The potassium ions are drawn as purple spheres, guanine bases are colored dark blue, adenines are red and cytosines are mauve. (b) View down the line of six potassium ions, emphasizing their colinearity. (c) Solvent-accessible surface view of the *B-raf* quadruplex, with the two strands colored orange and mauve. The two narrow grooves are indicated by arrows. (d) Schematic representation of the topology and base stacking in the structure, showing the seven G-quartets and the G·C·G·C quartet. The quadruplex has approximate 2-fold symmetry, so nucleotide G1 at the start of the second strand is at the 5′ position on the right-hand side of the figure.

solution was transferred to a quartz cell with a path length of 1 cm, which was then sealed. UV thermal difference spectra were obtained by carrying out a wavelength scan from 340 to 220 nm above (85 °C) and below (20 °C) the thermal transition temperature, and plotting the difference in absorbance at each wavelength. The thermal transition temperatures at different conditions were derived by plotting the melting curves and annealing curve. Samples were first heated to 90 °C with a temperature gradient of 10 °C/min and held at 90 °C for 5 min without data collection before being cooled to 5 °C with a temperature gradient of 0.5 °C/min, during which absorbance data were recorded. Samples were then returned to 90 °C with an identical temperature gradient, again with data collection. The data were then normalized and smoothed, with melting temperatures being calculated using the first derivative method.

Circular Dichroism. Experiments were performed on an Applied Photophysics Ltd. (Leatherhead, UK) Chirascan Plus spectrometer at King's College London. DNA samples were prepared at 4 μM in 20 mM potassium phosphate buffer at pH 7.0 or 120 mm K^+ solution (20 mM potassium phosphate and 100 mm KCl). Spectra were measured between 360 and 200 nm at different temperatures and recorded with a 1.0 nm step size, a 1.0 s measurement time-per-point and a spectral bandwidth of 1.0 nm. All spectra were buffer baseline corrected.

Nuclear Magnetic Resonance. The sample for the NMR analysis was prepared by dissolving 200 μM DNA in 0.6 mL of the solution containing 90% H_20, 10% D_2O, 10 mM potassium phosphate buffer or 120 mM K^+ (100 mM KCl and 20 mM potassium phosphate buffer (pH = 7.0)). NMR spectra were acquired using a Bruker Avance 500 MHz spectrometer equipped with a cryoprobe. Water suppression was achieved using an excitation sculpting pulse sequence. The sample was heated in the magnet from 298 to 358 K at 5 °C/min, and NMR spectra were recorded at every 10 °C. The sample was cooled down to room temperature at 2 °C per min, and NMR spectra were acquired at 30 °C intervals. The sample was equilibrated for a period of 600 s at each temperature. 128 scans were acquired for each spectrum with a relaxation delay of 2 s between scans.

Gel Electrophoresis. Electrophoresis experiments were performed with a 10 × 7 cm^2 native gel containing 20% acrylamide in TBE buffer, pH 8.3 supplemented with 50 mM KCl. Each sample contains 5 μL DNA at concentrations of 0.3−0.8 mM. Gels were viewed by UV shadowing after staining with 0.1% SYBR-green II.

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319−19329

Table 2. Nucleotide Backbone Torsion Angles and Sugar Puckers in the Two Strands of the *B-raf* Quadruplex, Calculated by the 3DNA Program[41a]

Nu		α	β	γ	δ	ε	ζ	χ	glycosidic angle	sugar pucker
G1	strand A	–	–	52	136	−166	−89	**51**	*syn*	C2′-*endo*
	strand B	–	–	52	145	−157	−92	**53**	*syn*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	–	–	0	9	9	3	**2**		
G2	strand A	−71	164	62	134	−175	−82	−113	*anti*	C1′-*exo*
	strand B	−64	162	55	137	161	−81	−104	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	7	2	7	3	24	1	9		
G3	strand A	−87	168	91	146	**−101**	106	−95	*anti*	C2′-*endo*
	strand B	−103	−161	84	152	**−40**	144	−85	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	16	31	7	6	61	38	10		
C4	strand A	−178	**−132**	94	154	−174	**−111**	−162	*anti*	C2′-*endo*
	strand B	−118	**146**	57	132	−156	**73**	−118	*anti*	C1′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	60	**82**	36	22	18	**176**	44		
G5	strand A	**-86**	**47**	176	138	172	−93	70	*syn*	C2′-*endo*
	strand B	**84**	**−156**	−169	156	154	−82	78	*syn*	C3′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	**170**	**157**	15	18	18	11	8		
G6	strand A	**−46**	178	**42**	145	−175	−134	−94	*anti*	C2′-*endo*
	strand B	**166**	−170	**176**	130	−169	−98	−129	*anti*	C1-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	**148**	12	**134**	14	6	36	35		
G7	strand A	−55	170	56	132	174	−92	−121	*anti*	C2′-*endo*
	strand B	−87	−179	69	136	−173	−112	−118	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	32	19	13	4	13	20	3		
G8	strand A	−70	−157	47	148	−84	116	−104	*anti*	C2′-*endo*
	strand B	−57	−175	46	145	−83	118	−108	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	13	18	1	3	1	2	4		
A9	strand A	126	−172	72	106	−127	−84	95	*syn*	O4′-*endo*
	strand B	128	−173	68	101	−114	−92	91	*syn*	C4′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	2	1	4	5	13	8	4		
G10	strand A	−84	−170	60	87	−171	−75	−178	*anti*	C4′-*exo*
	strand B	−58	−171	38	117	−165	−106	−159	*anti*	C1′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	26	1	22	30	6	31	19		
G11	strand A	−62	−154	45	146	−173	−98	−113	*anti*	C2′-*endo*
	strand B	−30	−169	24	154	−171	−99	−106	*anti*	C3′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	32	15	21	8	2	1	7		
G12	strand A	−63	−173	40	146	−162	−137	−98	*anti*	C2′-*endo*
	strand B	−66	−179	43	139	−164	−132	−101	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	3	6	3	7	3	5	3		
G13	strand A	−44	170	31	129	180	−95	−114	*anti*	C2′-*endo*
	strand B	−51	170	40	134	−165	−109	−114	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	7	0	9	5	15	14	0		
G14	strand A	−79	−165	51	135	−136	79	−107	*anti*	C1′-*exo*
	strand B	−65	−175	34	152	−130	75	−99	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	14	10	17	17	6	4	8	*syn*	
A15	strand A	57	170	44	141	−114	57	−94	*anti*	C2′-*endo*
	strand B	89	176	34	149	−142	79	−110	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	32	6	10	8	28	22	16		
A16	strand A	**96**	178	**−175**	160	−85	−149	77	*syn*	C3′-*exo*
	strand B	**−133**	152	**76**	124	−77	−153	90	*syn*	C1′-*exo*
	$\lvert\Delta_{A\text{-}B}\rvert$	**131**	26	**109**	36	8	4	13		
G17	strand A	63	−147	53	144	−170	−105	−134	*anti*	C2′-*endo*
	strand B	60	−135	65	141	−168	−99	−136	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	3	12	12	3	2	6	2		
G18	strand A	−57	−164	37	147	−174	−100	−106	*anti*	C2′-*endo*
	strand B	−67	−166	41	151	−165	−120	−105	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	10	2	4	4	9	20	1		
G19	strand A	−61	−167	27	147	−133	**-146**	−94	*anti*	C2′-*endo*
	strand B	−60	−176	32	152	−94	**-82**	−97	*anti*	C2′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	1	9	5	5	39	64	3		
A20	strand A	77	109	−80	151	–	–	−54	*syn*	C2′-*endo*
	strand B	**−124**	128	−65	114	–	–	−120	*anti*	C1′-*endo*
	$\lvert\Delta_{A\text{-}B}\rvert$	**159**	19	15	37			66		
B-DNA		**−30**	136	31	143	−141	−161	−98	*anti*	C2′-*endo*

[a]Major differences between strands A and B are shown in bold. ⟨Esd⟩ ± 1−2°.

dx.doi.org/10.1021/ja4101358 | *J. Am. Chem. Soc.* 2013, 135, 19319−19329

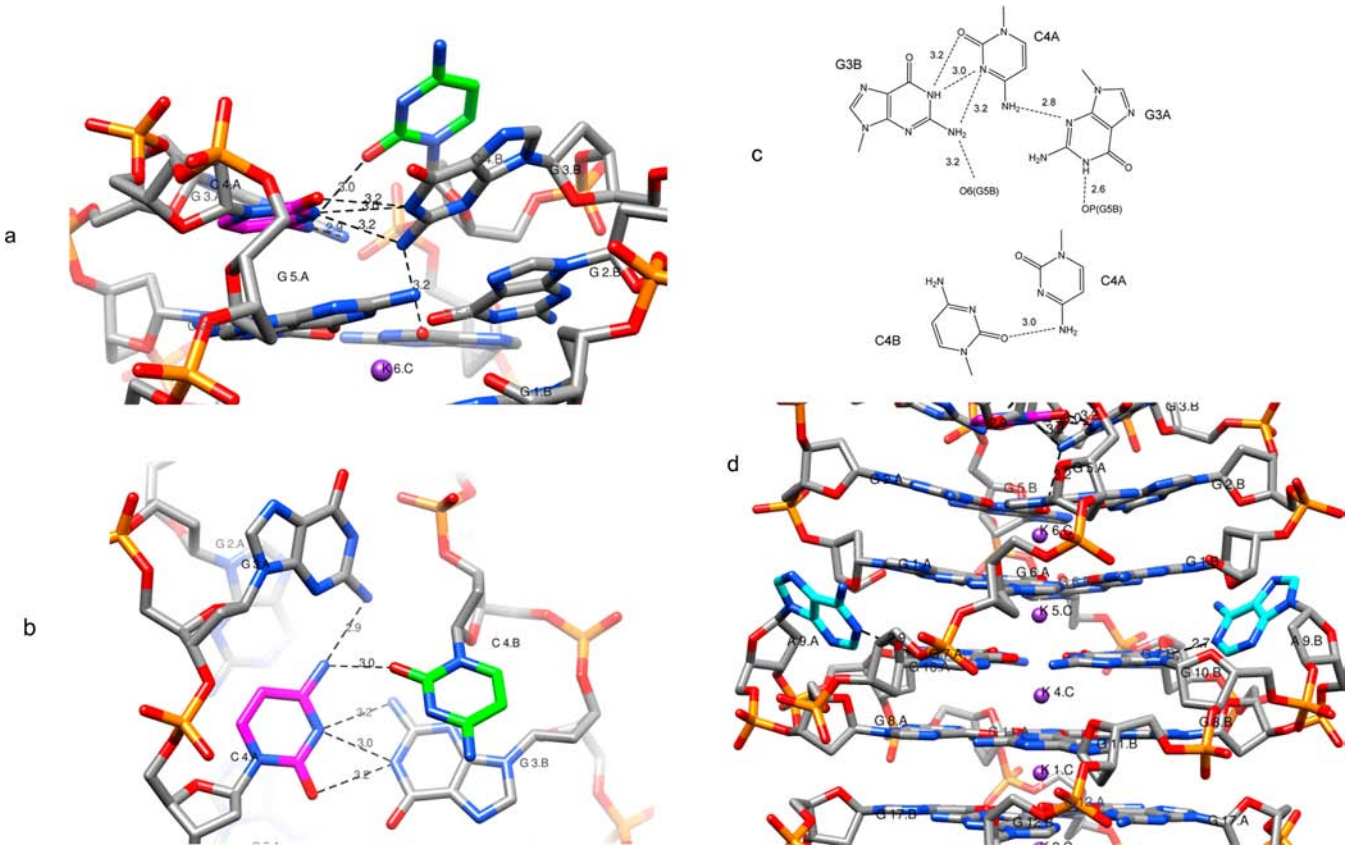


Figure 2. (a, b) Two views of the arrangement of the two guanines and two cytosines forming the distorted G·C·G·C quartet at the top end of the *B-raf* quadruplex. Carbon atoms of the cytosine base C4A are colored mauve, and those of C4B are colored green. Note in particular the distinct orientations of the two cytosines, even though both C4A and C4B are stacked onto their adjacent guanine bases G3A and G3B respectively, as seen in (a). The hydrogen bond from G3B to the preceding G-quartet is visible in this figure. (c) Details (top) of the hydrogen bonding interactions between the cytosine bases of C4A and C4B, and the adjacent guanine bases, and (bottom) interactions between the two cytosine bases alone. (d) View showing the A9A and A9B loops with the carbon atoms of the adenine bases colored cyan. The intramolecular hydrogen bonds are highlighted, involving N1 of A9A to N1 of G7A and from N1 of A9B to N2 of G7B.

RESULTS

Overall Features of the Quadruplex Structure. The *B-raf* DNA sequence crystallizes in space group $C222_1$, with the asymmetric unit containing two strands of the 20-mer sequence d(GGGCGGGGAGGGGGAAGGGA). This unit of 40 nucleotides comprises the two crystallographically independent chains intertwined together to form a bimolecular quadruplex. The central core of this unprecedented cylinder-like dimeric quadruplex structure comprises seven consecutive G-quartets. There are six potassium ions running down the central axis of the quadruplex, each situated between pairs of adjacent G-quartets (Figure 1a, b), in an almost completely linear arrangement. All the G-quartets have the expected pattern of hydrogen-bonding, with donor−acceptor distances corresponding to eight hydrogen bonds in each quartet. The G-quartets in the interior of the central core are coplanar, whereas those at the termini have individual guanines propeller-twisted out of plane: for example, G5A and G5B, in the G-quartet formed by G2A/G5A/G2B/G5B are twisted by 13° (as measured by the twist between the N1-N7 vectors in each base).

The overall structure has approximate 2-fold symmetry, which is broken, in particular, by the markedly nonequivalent conformations of two nucleotides in each strand. Thus the two quadruplex chains have overall a high degree of structural conservation with each other (the nucleotide torsion angles are listed in Table 2), for the majority of the 20 nucleotides in each strand, though there are some notable differences, as discussed below. The overall shape of the structure is of a quasi-4-fold helical stem, with four adenines (A15 and A20 from each strand), protruding from one end of the stem as four approximately equidistant legs (Figure 1a, c). Seventeen out of 20 nucleotides in each strand are involved in hydrogen-bonding and stacking interactions (Figure 1d: see the detailed description below), and are thus essential for the integrity of the structure. Almost every base involved in hydrogen-bonding interactions contacts bases in both its own strand and the second strand. In particular, 14 out of the 15 guanines in each strand participate in forming G-quartets (Figure 1d).

The Structure Contains a Distorted G·C·G·C Quartet. At the opposite end of the structure to the adenine feet, significant differences between chains A and B are apparent in the conformations of the four nucleotides G3A, G3B and C4A, C4B. These are all involved in two unusual lateral loops, involving C4A and C4B respectively, and can be categorized as single-nucleotide loops, although these cytosines are involved in tertiary interactions (see below). The two cytosines are each held in between two guanosines that form a complex nucleobase stacked and hydrogen-bonded interface between the two strands (Figures 2a, b). One guanine in each strand does not participate in G-quartet hydrogen bonding, G3A in strand A and G3B in strand B. Instead, both are hydrogen-bonded

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319−19329

to cytosine C4A, which in turn is hydrogen-bonded to C4B, to form a highly deformed G·C·G·C base quartet (Figure 2a–c) at the top end of the molecule. Atom N2 of G3B is involved in two hydrogen bonds to C4A (Figures 2a–c); although the distance to the O2 atom of C4A is long (3.2 Å), the arrangement is typical of a bifurcated hydrogen bond arrangement. The two guanines in this quartet are thus not structurally identical, with one (G3A) being oriented out of the plane of the G3B-C4A pairing. There is also a hydrogen bond between the G3A N1 atom and an adjacent backbone phosphate oxygen atom. The two cytosine nucleotides C4A and C4B have distinct backbone conformations with the two torsion angles β and ζ (Table 2), each differing by ~90° and 180°, respectively. These differences result in one of the cytosine bases, C4B, being stacked, at a 3.4 Å separation, with the adjacent G3B (Figure 2a), whereas C4A stacks onto the nonequivalent guanine base of G5A and forms the central element of a G·C·G triplet. The bases of C4A and C4B are still able to form a C·C base pair with a single hydrogen bond holding them together, albeit in a highly nonplanar arrangement (with a propeller twist angle of 62°). The N2 substituent of G3B is close to the O6 substituent of G5B on the preceding G-quartet. Although the distance involved (3.2 Å) is long for a hydrogen bond, the orientation of G3B and the twisted hydrogen-bond angles around N2 suggest that this is a stereochemically plausible hydrogen bond. There are no potassium ions or water molecules between this G-quartet and the G·C·G·C base quartet.

The Adenosine Loops. The topology of the overall structure cannot readily be described in the straightforward manner that single-chain intramolecular quadruplexes have been discussed to date. The individual strands form three distinct folded-back domains, each with the strands being parallel to each other: G5-G8, G10-G14, and G17-G19. The short stretch G1-G2 is antiparallel to these. The strands form two equivalent pairs of narrow grooves that extend along most of the structure (Figure 1c). The groove widths, as measured by P···P distances, vary from 8.0 to >9 Å. Each strand forms two loops, both comprising solely adenosine nucleotides: (1) A two-nucleotide propeller (strand-reversal) loop involving A15 and A16, which is hydrogen-bonded to the edge of a guanine base in an adjacent G-quartet (Figure S1, Supporting Information). The adenines in both loops are oriented away from the quadruplex core. Both of the A15 nucleotides are in an *anti* glycosidic angle conformation whereas the A16 pair are both in a *syn* conformation. This change ensures that A16A and A16B are each stacked over the sugar ring of the A15 nucleotide, rather than being close to the phosphate group. The backbones of A16A and A16B have nonidentical conformations (Table 2), with the two affected angles α and γ showing *gauche*$^+$–*trans* correlated differences, analogous to observations of correlated backbone angles in B-DNA oligonucleotide structures[34,35] and most recently in simulation studies on quadruplexes.[36,37] (2) A single-adenosine loop comprising A9. Each adenine base is tucked into the G-quartet core by means of the A9 *syn* glycosidic conformation: for both loops the adenosines A9A and A9B are also each hydrogen-bonded to the edge of a guanine base (G7A, G7B) in an adjacent G-quartet (Figure 2d). There is a subtle difference between the two hydrogen bonds: one is between N1 of A9A and N1 of G7A (2.9Å), and the other is between N1 of A9B and N2 of G7B (2.7 Å).

The two single-adenosine loops A9A and A9B are between successive G-quartet layers and contrast with propeller loops observed in, for example, the human intramolecular telomeric quadruplex, in which the trinucleotide TTA loops are able to bridge three successive G-quartets along the sides of the quadruplex core. By contrast the two-adenosine loops A15A, A16A and A15B, A16B do span three quartets and resemble the TTA propeller loops. All other potential loop nucleotides such as C4A and C4B or some of the guanosines that are at the ends of the longer individual G-tracts, are involved in base–base hydrogen bonding and are thus fully incorporated into the core of the structure. The sharp kinks in the backbone of each chain are around nucleotides G3, G5, G8, G14, A15 and G17. These are necessary to achieve the fold-back of succeeding nucleotides such that maximum G-quartet formation can occur. These kinks are achieved by a variety of backbone torsion angle changes from the approximately B-DNA-like backbone conformation of the majority of nucleotides in this structure (Table 2), as well as changes in sugar pucker from the standard C2′-*endo* form in the conformationally regular part of the quadruplex. For example, the kinks at G3 and G5 are achieved by changes in torsion angles ε in G3 itself, α and ζ in C4, α, β and γ in G5, in α and γ in G6, and in the *syn* sugar pucker for G5A and G5B. The differences between backbone conformations for the two strands are most pronounced in this region and again are often correlated changes. The backbone kink at A15 is mostly achieved by changes in several angles in G14 and A16 as well as in A15 itself, where α and ζ change by ca. 90° and 140°, respectively, compared to the more quadruplex nucleotides in this structure. The kink at G17 is mostly achieved by changes in angles α and ε, and by a glycosidic angle change to *syn* in the preceding nucleotide A16.

Solution Studies. A decrease in UV absorbance at 295 nm was observed upon heating. Analogous behavior occurs for other quadruplex-forming sequences, and is normally considered to be diagnostic for G-quadruplex formation. The decrease in absorbance for the solution containing 120 mM K^+ is less than that for that with 10 mM K^+, indicating that the *B-raf* quadruplex is more stable in 120 mM K^+ (Figure 3a). The melting curves exhibit a single melting transition at 295 nm, with little difference between the melting and annealing curves (Figure S2). The 71 °C value of T_m derived from the UV melting curves further indicates the high stability of the *B-raf* DNA quadruplex in 120 mM K^+. No significant change in melting temperature was apparent over a 2–22 μM range of strand concentrations (data not shown). The T_m value from the UV/vis studies is 72.5 ± 0.15 °C, based on curve fitting, at a *B-raf* DNA quadruplex concentration of 15 μM, and 71.3 °C at a 4 μM concentration, both in 120 mM K^+.

G-quadruplexes exhibit characteristic circular dichroism (CD) spectra, whose features depend on the topology of the structure (parallel or antiparallel). Parallel folds exhibit a maximum positive signal at around 260 nm with a corresponding negative signal around 240 nm. Antiparallel quadruplex species show a characteristic positive signal at around 295 nm with a negative signal at around 260 nm. The *B-raf* sequence shows a positive signal at around 260 nm and a negative signal around 240 nm (Figure 3b), with a small positive peak at 295 nm. These CD spectral features are apparent both when the temperature is lowered, and when the K^+ concentration is increased, suggesting that the structure is mostly parallel under all these conditions (Figures 3b, S3). The small "antiparallel" signal is more prominent in 10 mM K^+ solution than that in 120 or 200 mM K^+ (Figure 3b), and is consistent with the short antiparallel stretch (G1-G2) observed in the crystal structure.

1D NMR studies have been performed on the *B-raf* sequence. The imino proton spectrum showed sharp peaks

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319–19329

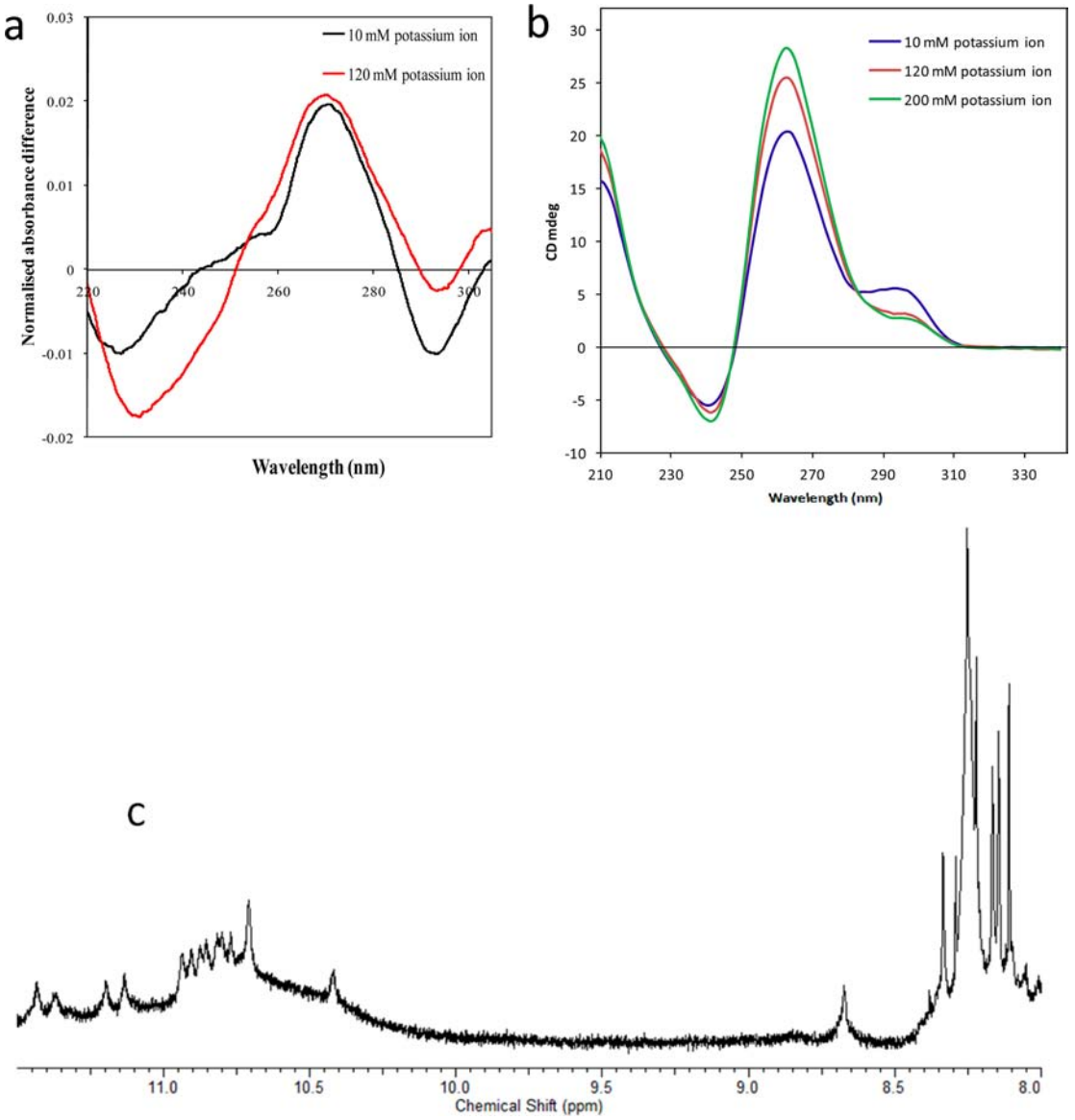


Figure 3. (a) UV thermal difference spectrum of the *B-raf* sequence in solution with two different concentration of potassium ions. (b) CD spectra of the annealed *B-raf* sequence in 10, 120, and 200 mM potassium ion buffer at 25 °C. (c) Imino and aromatic region of 1D ^{1}H NMR spectrum of the *B-raf* quadruplex sequence at 200 μM oligonucleotide concentration, 10 mM potassium phosphate and pH 7. The spectrum was acquired using a Bruker Avance 500 MHz NMR instrument at 298 K.

between 10 and 12 ppm, which are characteristic of G-quadruplex formation in solution (Figure 3c). At 25 °C, there are thirteen sharp imino proton peeks between 10 and 12 ppm, and the integration of the peak at 10.72 ppm corresponds to the two imino protons. This suggests there are fourteen distinct guanines from a single strand involved in G-quartets in the overall G-quadruplex structure and that the oligonucleotide chains in the dimer are symmetric in solution, since only one set of peaks was observed for each guanine. A large diffuse peak between 10 and 11 ppm may reflect other conformations. With increasing temperatures, the number of imino proton peaks decreased to twelve at 45 °C (Figure S4), in qualitative agreement with UV experiments, giving a melting point for the *B-raf* sequence of 50 °C in low K^+ conditions (Figure S2). The NMR spectrum at 85 °C showed nine peaks, which indicated the high stability of the core of this quadruplex structure. The hydrogen bond network of the quadruplex is not completely disrupted at this temperature, although the conformation of the G-quadruplex may be slightly changed, as suggested by changes in the peak positions. The closely similar NMR spectra acquired during the cooling process indicates that the process is reversible and thus that the sequence folds into a very similar conformation. On cooling the sample down to 25 °C, thirteen main peaks appeared again between 10 and 12 ppm, although the multiple-conformation peak decreased and the presence of minor peaks at 11.35 ppm suggest formation of a second, minor conformation. The 1D NMR spectrum of the *B-raf* quadruplex in 120 mM K^+ solution exhibited a very similar set of peaks to that in 10 mM K^+ solution (Figure S4), although the peaks are not straightforward to assign. This suggests that the oligonucleotide adopts similar, but not exactly identical conformations in the two salt conditions.

The gel mobility behavior of the *B-raf* sequence is shown in Figure S5. The weak band assigned as a single-stranded species corresponds to only a small fraction of the total DNA, whereas the prominent band corresponds to a dimer species, with 40 nucleotides. The presence of the dimer is in accord with both

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319−19329

Table 3. Examples of Human Genome Sequences Containing Eight or More Short G-Tracts, Obtained from Searches of the Ensembl Human Genome Resource (www.ensembl.org)

Gene: RIMS3
ENSG00000117016 regulating synaptic membrane exocytosis 3 [Source:HGNC Symbol;Acc:21292]
Transcript: ENST00000372684 (CCDS30687); Location: within intron 1−2
Sequence: GGGGGAGGGGCGGGGAGGGGGAGGGGCGGGAAGGGGAGGGGAGGAGATGGGAGGGCAGGCGGGGGAGGGACTGGG
Gene: KIF21B
ENSG00000116852 kinesin family member 21B [Source:HGNC Symbol;Acc:29442]
Transcript: ENST00000332129 (CCDS30965); Location: within intron 26−27
Sequence: GGGAAGGGAAGGGAAAGGAAGGGAGGGGAGGGGAGGGGAGGGGAGGGGAAGGGGAGGGGAGTGGGGGAGGG
Gene: RPL37A
ENSG00000197756 ribosomal protein L37a [Source:HGNC Symbol;Acc:10348]
Transcript: ENST00000446558; Location: within intron 3−4
Sequence: CCCCTTCCCCTCCCCTCTCTTCCCCATCCCCTTCCCCTCCCCTCTCTTCCCCATCCCCTTCCCCTCCCCTCTCTTCCCC
Gene: SENP2
ENSG00000163904 SUMO1/sentrin/SMT3 specific peptidase 2 [Source:HGNC Symbol;Acc:23116]
Transcript: ENST00000296257 (CCDS33902); Location: within the 3′-untranslated region
Sequence: CCCCCCCCCCTCCCACTCCCCCTCCCTCCCTCTCCCCTCCCCCTCCCTCCCTCCCCCCTCCCTCCCTCCCC-CCCCTCCCCCTCCCTCCCCCTCCCTCTCTCCCTAACACCCACCCTCCCCCTTTCCCCCACCCTCCCCCCTTCTCCCCTTT-CCCCACCCTCCCTCCCCCCTTCCCCCTCCCCCCCCCCCCCCC
Gene: TRIM13
ENSG00000204977 tripartite motif containing 13 [Source:HGNC Symbol;Acc:9976]
Transcript: ENST00000378182 (CCDS9423); Location: 9 bases upstream of the transcription start site
Sequence: GGGCGACGGGGAAGGGAGGGGGAGGGGGAGGGGAAAGGGAGGGGGAGGGGGAGGGGAAAGGGAGGGGGAGGGGGGAAG-GGACGGGAGGGGAGGGGGAAGGGACGGGAGGGGAGGGGGAAGGGACGGGAGGGGAAGGGGAAGCGGGAGAGGGAAGGGGG
Gene: SDK2
ENSG00000069188 sidekick cell adhesion molecule 2 [Source:HGNC Symbol;Acc:19308]
Transcript: ENST00000392650 (CCDS45769); Location: within intron 29−30
Sequence: GGGGAAGGGGAAGGGGAAGGGAAGGGAGGAGAGGGGAGGGGAGGGGGAAGGGAAAGGGAGGGGAGGGAAGG-GGGGAAAAAGGGAGGGAGGGAGGGAGGGAGGG
Gene: TMEM238
ENSG00000233493
transmembrane protein 238 [Source:HGNC Symbol;Acc:40042]; Transcript: ENST00000444469 (CCDS54323); Location: within intron 1−2
Sequence: GGGGAGGGGAGGGGAGGAGGGGGGAAGGGGGAGGGGAGGGGGAAGCAGGGAGGGAAGGGGGGAAGGGAGGGG-AGGGGGAAGGGGGAAGGGGGAGGGGACGGGCAAGGGGGGAGGGGAGGGGGGAAGGGGGGAGGGGAGGGGAGGGG
Gene:PCBP3
ENSG00000183570 poly(rC) binding protein 3 [Source:HGNC Symbol;Acc:8651]
Transcript: ENST00000400314 (CCDS42974) Location: Within intron 2−3
Sequence: CCCCTTCCCCTTCCCCCTCCCCGTCCCCTGCCCCTCCCCTCCCTCTCCCTCCTCCCCCTCCTC
Gene: BCL2
ENSG00000183570 Apoptosis Regulator [Source:HGNC Symbol;Acc:990]
Transcript: ENST00000333681 (CCDS11981); Location: within intron 2−3
Sequence GGGAGGGAGGGAGGGAGGGAGACAGGGAGGGAGAGAGAGGG

the crystallographic and NMR data. This suggests that the intermolecular dimer is the major species in K^+ solution, and not an intramolecular quadruplex comprising a single strand.

DISCUSSION

The biophysical data and the crystal structure concur in finding that the *B-raf* sequence d(GGGCGGGGAGGGGGAAGGGA) forms a stable dimer, which is the dominant species found to be present in solution and in the crystalline state. The number of resolved imino protons observed in the 1D NMR data between 10 and 12 ppm is consistent with the seven G-quartets found in the crystal structure. Although only a full NMR structure analysis will provide definitive proof that the structures in solution and the crystal have the same topology, it is difficult to envisage an alternative dimer arrangement that also contains seven G-quartets.

Putative Quadruplex Dimers in the Human Genome. The *B-raf* promoter sequence in situ cannot form this structure as the physiological dimer since there is only one copy of this sequence present at this genome locus. It is possible that a dimer could be involved in recombination at the promoter sequence, as has been discussed for the *c-kit*2 quadruplex;[19] however, we have not investigated this possibility for the *B-raf* quadruplex.

This structure does suggest that where two discrete quadruplex sequences, each containing four G-tracts, occurs in a promoter or elsewhere within a genome, then a dimer fold, such as is seen here, is possible. The likelihood of this occurring may depend on the percentage of guanines since, at least in the present structure, the preponderance of guanines results in seven G-quartets being formed, as the dominating feature of the structure. The formation of a monomeric intramolecular quadruplex from a promoter sequence cannot therefore be assumed in all cases and suggests that future studies on such sequences, for example as small-molecule targets, need to verify the nature of the quadruplex at the outset.

We have used qualitative molecular modeling (HYPERCHEM: www.hyper.com) to show that the two strands in the

dx.doi.org/10.1021/ja4101358 | J. Am. Chem. Soc. 2013, 135, 19319−19329

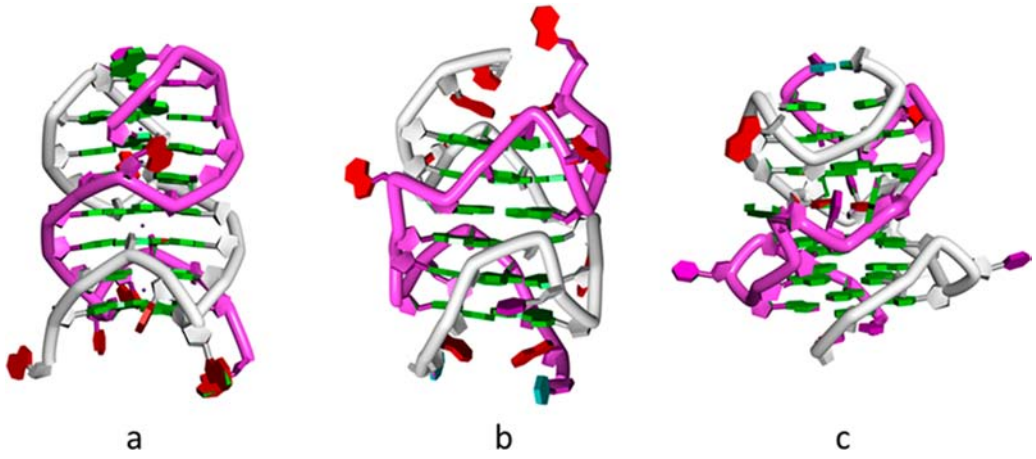


Figure 4. Comparison of three dimeric quadruplex structures, shown in cartoon representation. The two strands in each structure are colored gray and mauve. (a) the *B-raf* structure reported here, (b) the *N-myc* quadruplex[40] (PDB id 2LED), with six consecutive G-quartets, and (c) the *c-kit*2 quadruplex[19] (PDB id 2KYO), with two sets of three G-quartets, separated by an A·A base pair.

present structure can be readily linked 3′ to 5′ with a short six-nucleotide stretch of arbitrary DNA sequence (Figure S6) to form a stereochemically plausible single-nucleotide chain arrangement, with the nucleotides in the linker shown having been assigned an (arbitrary) B-DNA conformation. The arrangement shown here preserves the observed fold of the quadruplex, although a rearrangement to an alternative topology driven by this long loop, cannot be discounted. A shorter five-nucleotide stretch could also be accommodated in a similar manner, but at the cost of some perturbation in the geometry of the adjacent A-loop. The results of a search for sequences with eight or more G-tracts in the human genome and some hits are shown in Table 3. A number of the sequences are notable in having (at least) two sets of four three-guanine tracts separated by A/G-rich sequences, which themselves could form a stable stem-loop type of secondary structure, analogous to that observed in the *c-kit*1 crystal and NMR structures.[18,20]

Other Quadruplex Dimers. It is well-established that monomeric intramolecular quadruplexes such as that formed by human telomeric DNA and RNAs, can dimerize in solution and in the crystalline state by stacking end-to-end.[29,37−39] The *B-raf* dimer observed here is quite distinct from these and more resembles those determined by NMR methods, for the quadruplexes formed by the *c-kit*2 sequence[19] (PDB id 2KYO) and one from the intron of the *N-myc* gene,[40] with PDB id 2LED). Each strand spans the entire length of all three structures (Figure 4). The *c-kit*2 and *N-myc* quadruplexes both have six stacked quartets, with an intervening central A·A base pair in the former structure, whereas the *B-raf* quadruplex has seven consecutive G-quartets. This reflects the higher percentage of guanines in this structure. Remarkably the entire backbone in the *B-raf* structure has a significant overlap with the *N-myc* structure with a rmsd of 1.9 Å. The *c-kit*2 quadruplex dimer has two distinct regions of overlap, residues A/B 14,15,16,18,19,20 with a rmsd of 0.9 Å and A/B 2,3,4,6,7,8 with a significantly larger rmsd of 2.9 Å due to *syn*/*anti* mismatches for residues A/B 2,6, giving an overall rmsd of 3.9 Å for the top three and bottom three stacked quartets.

■ CONCLUSIONS

We have shown here that a promoter sequence in the *B-raf* gene can form a highly stable G-quadruplex arrangement. Contrary to expectations, in view of the presence of four discrete G-tracts, the quadruplex is a dimer with the backbones intertwined, rather than being the anticipated monomeric quadruplex species, as observed for a number of other promoter quadruplexes. Evidence is given that the dimer is present in solution, where it appears to be the dominant species, by contrast with two other recently characterized quadruplex dimers. Both the *c-kit*2 and *N-myc* dimers are in equilibrium in solution with monomeric quadruplexes. The difference is likely due to the exceptional high percentage of guanosine nucleotides in the *B-raf* structure. This has the effect of maximizing the possible number of G-quartets; the *B-raf* structure comprises an exceptionally large quadruplex core of seven consecutive G-quartets, whose formation would favor dimer over monomer species. Future studies on other putative promoter quadruplexes (and those located in other genomic loci) may need to challenge the normal assumptions of monomer character before undertaking, for example, searches for small-molecule modulators of quadruplex biological function. The modeling study undertaken here, in conjunction with a sequence search, has indicated that rather larger quadruplex structures than the widely studied conventional 3/4-quartet intramolecular species, may have a stable existence and are worthy of further study.

The existence of a stable *B-raf* quadruplex dimer suggests that it might have potential utility as a quadruplex aptamer agent.[42,43] A preliminary evaluation of the effects of the *B-raf* sequence on the proliferation of a small panel of cancer cell lines[44] supports this suggestion. It was found that the quadruplex selectively inhibits the growth of A549 human lung adenocarcinoma cells, with an IC_{50} value for 96 h exposure of 6.4 μM, whereas MCF7 breast carcinoma cells and WI38 normal fibroblast cells were unaffected up to significantly higher concentrations of the sequence (see Supporting Information for experimental details).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Further experimental information and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

s.neidle@ucl.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Cancer Research UK for support (Programme Grant No. C129/A4489 to S.N. and a China Fellowship to D.W.), and to the Diamond Light Source and the ESRF, Grenoble for access to synchrotron facilities. Tony Reszka and Drs. Gavin Collie and Stephan Ohnmacht (School of Pharmacy, UCL) are thanked for much guidance on instrument use, Dr. Ambrose Cole (Birkbeck College) for assistance with crystallographic data collection, and Dr. Tam T. T. Bui (King's College London) for running the CD spectra.

■ REFERENCES

(1) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013−2018.

(2) Davies, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668−698. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402−5415. Phan, A. T. *FEBS J.* **2010**, *277*, 1107−1117. Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat. Rev. Genetics* **2012**, *13*, 770−780.

(3) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876−880. Neidle, S. *Curr. Opin. Struct. Biol.* **2009**, *19*, 239−250. Campbell, N.; Collie, G. W.; Neidle, S. *Curr. Protoc. Nucleic Acid Chem.* **2012**, Ch. 17, Unit 17.6.

(4) Moyzis, R. K.; Buckingham, J. M.; Cram, I. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6622−6626. Meyne, J.; Ratliff, R. L.; Moyzis, R. K. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7049−7053. Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364−366. Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825−829. Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. *Biochemistry* **1987**, *51*, 899−908.

(5) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263−282. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723−2735. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927−4740. Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 4301−4309. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963−9970. Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715−5719. Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11−24.

(6) (a) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901−2907. (b) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908−2916.

(7) Oganesian, L.; Bryan, T. M. *BioEssays* **2007**, *29*, 155−165. Monchaud, D.; Teulade-Fichou, M.-P. *Org. Biomol. Chem.* **2008**, *6*, 627−636. Franceschin, M. *Eur. J. Org. Chem.* **2009**, 2225−2238. Ou, T.-M.; Lu, Y-J.; Tan, J.-H.; Huang, Z.-S.; Wong, K.-Y.; Gu, L.-Q. *ChemMedChem* **2008**, *3*, 690−713. Neidle, S. *Therapeutic Applications of Quadruplex Nucleic Acids*; Academic Press: San Diego, 2011.

(8) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* **2012**, *8*, 301−310. Müller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. *Nat. Chem.* **2010**, *2*, 1095−1098. Biffi, G.; Tannahill, D; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182−186.

(9) (a) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406−413. (b) Yuan, L.; Tian, T.; Chen, Y.; Yan, S.; Xing, X.; Zhang, Z.; Zhai, Q.; Xu, L.; Wang, S.; Weng, X.; Yuan, B.; Feng, Y.; Zhou, X. *Sci. Rep.* **2013**, *3*, 1811. (c) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.

(10) (a) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167−1172. (b) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593−11598.

(11) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584−10589. Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry* **2006**, *45*, 7854−7860.

(12) Cogoi, S.; Xodo, L. E. *Nucleic Acids Res.* **2006**, *34*, 2536−2549. Cogoi, S.; Paramasivam, M.; Filichev, V.; Géci, I.; Pedersen, E. B.; Xodo, L. E. *J. Med. Chem.* **2009**, *52*, 564−568. Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K. K.; Xodo, L. E. *J. Biol. Chem.* **2010**, *285*, 22003−22016.

(13) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878−10891. Lim, K. W.; Lacroix, L.; Yue, D. J.; Lim, J. K.; Lim, J. M.; Phan, A. T. *J. Am. Chem. Soc.* **2010**, *132*, 12331−12342.

(14) Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. *Nucleic Acids Res.* **2005**, *33*, 6070−6080. Sun, D.; Guo, K.; Shin, Y. J. *Nucleic Acids Res.* **2011**, *39*, 1256−1265.

(15) De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry* **2005**, *44*, 16341−16350. Lombardo, C. M.; Welsh, S. J.; Strauss, S. J.; Dale, A. G.; Todd, A. K.; Nanjunda, R.; Wilson, W. D.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5984−5988.

(16) Mitchell, T.; Ramos-Montoya, A.; Di Antonio, M.; Murat, P.; Ohnmacht, S.; Micco, M.; Jurmeister, S.; Fryer, L.; Balasubramanian, S.; Neidle, S.; Neal, D. E. *Biochemistry* **2013**, *52*, 1429−1436.

(17) Huang, F. W.; Hodis, E.; Xu, M. J.; Kryukov, G. V.; Chin, L.; Garraway, L. A. *Science* **2013**, *339*, 957−959. Horn, S.; Figl, A.; Rachakonda, P. S.; Fischer, C.; Sucker, A.; Gast, A.; Kadel, S.; Moll, I.; Nagore, E.; Hemminki, K.; Schadendorf, D.; Kumar, R. *Science* **2013**, *339*, 959−961.

(18) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 4386−4392.

(19) Kuryavyi, V.; Phan, A. T.; Patel, D. J. *Nucleic Acids Res.* **2010**, *38*, 6757−6773.

(20) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. *Nucleic Acids Res.* **2012**, *40*, 4691−4700.

(21) (a) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167−173. (b) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. *Nucleic Acids Res.* **2011**, *39*, 9023−9033.

(22) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. *J. Am. Chem. Soc.* **2006**, *128*, 1096−1098.

(23) See, for example: Liu, J. N.; Deng, R.; Guo, J. F.; Zhou, J. M.; Feng, G. K.; Huang, Z. S.; Gu, L. Q.; Zeng, Y. X.; Zhu, X. F. *Leukemia* **2007**, *21*, 1300−1302. Gunaratnam, M.; Swank, S.; Haider, S. M.; Galesa, K.; Reszka, A. P.; Beltran, M.; Cuenca, F.; Fletcher, J. A.; Neidle, S. *J. Med. Chem.* **2009**, *52*, 3774−3783. McLuckie, K. I. E.; Waller, Z. A. E.; Sanders, D. A.; Alves, D.; Rodriguez, R.; Dash, J.; McKenzie, G. J.; Venkitaraman, A. R.; Balasubramanian, S. *J. Am. Chem. Soc.* **2011**, *133*, 2658−2663. Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. *J. Biol. Chem.* **2011**, *286*, 41018−41027.

(24) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. *J. Am. Chem. Soc.* **2011**, *133*, 17673−17680.

(25) Davies, H.; Bignell, G. R.; Cox, C.; Stephens, P.; Edkins, S.; Clegg, S.; Teague, J.; Woffendin, H.; Garnett, M. J.; Bottomley, W.; Davis, N.; Dicks, E.; Ewing, R.; Floyd, Y.; Gray, K.; Hall, S.; Hawes, R.; Hughes, J.; Kosmidou, V.; Menzies, A.; Mould, C.; Parker, A.; Stevens, C.; Watt, S.; Hooper, S.; Wilson, R.; Jayatilake, H.; Gusterson, B. A.; Cooper, C.; Shipley, J.; Hargrave, D.; Pritchard-Jones, K.; Maitland, N.; Chenevix-Trench, G.; Riggins, G. J.; Bigner, D. D.; Palmieri, G.; Cossu, A.; Flanagan, A.; Nicholson, A.; Ho, J. W.; Leung, S. Y.; Yuen, S. T.; Weber, B. L.; Seigler, H. F.; Darrow, T. L.; Paterson, H.; Marais, R.; Marshall, C. J.; Wooster, R.; Stratton, M. R.; Futreal, P. A. *Nature* **2002**, *417*, 949−954. Wan, P. T.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R. *Cell* **2004**, *116*, 855−867.

(26) See, for example: Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.;

Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A. *Cancer Res.* **2004**, *64*, 7099−7109. Flaherty, K. T.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; O'Dwyer, P. J.; Lee, R. J.; Grippo, J. F.; Nolop, K.; Chapman, P. B. *N. Engl. J. Med.* **2010**, *363*, 809−819. Sosman, J. A.; Kim, K. B.; Schuchter, L.; Gonzalez, R.; Pavlick, A. C.; Weber, J. S.; McArthur, G. A.; Hutson, T. E.; Moschos, S. J.; Flaherty, K. T.; Hersey, P.; Kefford, R.; Lawrence, D.; Puzanov, I.; Lewis, K. D.; Amaravadi, R. K.; Chmielowski, B.; Lawrence, H. J.; Shyr, Y.; Ye, F.; Li, J.; Nolop, K. B.; Lee, R. J.; Joe, A. K.; Ribas, A. *N. Engl. J. Med.* **2012**, *366*, 707−714. Flaherty, K. T.; Hodi, F. S.; Fisher, D. E. *Nat. Rev. Cancer* **2012**, *12*, 349−361.

(27) Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2011**, *39*, 4917−4927.

(28) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr.* **2011**, *D67*, 235−242.

(29) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876−880.

(30) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.

(31) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr.* **2010**, *D66*, 486−501.

(32) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr.* **2011**, *D67*, 355−367.

(33) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

(34) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113−124.

(35) Djuranovic, D.; Hartmann, B. *Biopolymers* **2004**, *73*, 356−368.

(36) Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Sponer, J.; Neidle, S.; Haider, S. *Nucleic Acids Res.* **2013**, *41*, 2723−2135.

(37) Zhu, H.; Xiao, S.; Liang, H. *PLoS One* **2013**, *8*, e71380.

(38) Collie, G. W.; Parkinson, G. N.; Neidle, S.; Rosu, F.; De Pauw, E.; Gabelica, V. *J. Am. Chem. Soc.* **2010**, *132*, 9328−9334.

(39) Do, N. Q.; Phan, A. T. *Chemistry* **2012**, *18*, 14752−14759.

(40) Trajkovski, M.; da Silva, M. W.; Plavec, J. *J. Am. Chem. Soc.* **2012**, *134*, 4132−4141.

(41) Lu, X.-J.; Olson, W. K. *Nat. Protoc.* **2008**, *3*, 1215−1227.

(42) Gatto, B.; Palumbo, M.; Sissi, C. *Curr. Med. Chem.* **2009**, *16*, 1248−1265.

(43) Tseng, T. Y.; Wang, Z. F.; Chien, C. H.; Chang, T. C. *Nucleic Acids Res.* **2013**, *41*, 10605−10618.

(44) Gunaratnam et al., to be published.