

# Duplex formation in a G-quadruplex bulge

Thi Quynh Ngoc Nguyen<sup>1</sup>, Kah Wai Lim<sup>1</sup> and Anh Tuan Phan<sup>1,2,\*</sup><sup>1</sup>School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore and <sup>2</sup>NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921, Singapore

Received August 03, 2020; Revised August 21, 2020; Editorial Decision August 21, 2020; Accepted August 31, 2020

## ABSTRACT

**Beyond the consensus definition of G-quadruplex-forming motifs with tracts of continuous guanines, G-quadruplexes harboring bulges in the G-tetrad core are prevalent in the human genome. Here, we study the incorporation of a duplex hairpin within a bulge of a G-quadruplex. The NMR solution structure of a G-quadruplex containing a duplex bulge was resolved, revealing the structural details of the junction between the duplex bulge and the G-quadruplex. Unexpectedly, instead of an orthogonal connection the duplex stem was observed to stack below the G-quadruplex forming a unique quadruplex–duplex junction. Breaking up of the immediate base pair step at the junction, coupled with a narrowing of the duplex groove within the context of the bulge, led to a progressive transition between the quadruplex and duplex segments. This study revealed that a duplex bulge can be formed at various positions of a G-quadruplex scaffold. In contrast to a non-structured bulge, the stability of a G-quadruplex slightly increases with an increase in the duplex bulge size. A G-quadruplex structure containing a duplex bulge of up to 33 nt in size was shown to form, which was much larger than the previously reported 7-nt bulge. With G-quadruplexes containing duplex bulges representing new structural motifs with potential biological significance, our findings would broaden the definition of potential G-quadruplex-forming sequences.**

## INTRODUCTION

Guanine-rich oligonucleotides have a propensity to adopt non-canonical nucleic acid structures known as G-quadruplexes in the presence of monovalent cations, such as potassium and sodium (1–3). The G-quadruplex is a four-stranded helical structure containing multiple stacked G•G•G•G tetrads (G-tetrads), each formed by association of four guanines through Hoogsteen hydrogen bonding (4,5). G-quadruplexes can assume a variety of topologies, as defined by different combinations of strand

orientations and loop configurations (6–8). There are four distinct basic core topologies, namely all-parallel-stranded, (3+1) hybrid, antiparallel-stranded up–up–down–down, and antiparallel-stranded up–down–up–down; the core strands are connected by loops, of which there are three general types: edgewise (or lateral), diagonal and double-chain-reversal (or propeller).

Over the past decades, numerous G-quadruplex motifs were found in the functional regions of the human genome (9), such as the telomeres (10–15), RNA transcripts (16–19), and promoter regions of a wide range of critical genes, e.g. *MYC* (20,21), *KIT* (22), *KRAS* (23–27), *BCL2* (28), *hTERT* (29–31), *VEGF* (32), *HIF-1 $\alpha$*  (33), *RET* (34,35) and *PARP1* (36). There is growing evidence showing the formation of G-quadruplexes in human cells (37–39) and their regulatory roles in biological processes, including telomere length maintenance, replication, transcription, splicing and translation (40).

Most of reported G-quadruplex structures are adopted by sequences following the consensus motif  $G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+$  (41,42). However, recent studies have shown that some sequences, which do not strictly conform to this algorithm, are capable of forming stable G-quadruplexes (43–46). For example, G-quadruplex motifs from the human *MYC* promoter (47) and *KIT* promoter (48) were shown to form unusual G-quadruplex structures, which contain a G-column made of discontinuous guanine residues. Various motifs of long-loop G-quadruplex structures containing duplex stem-loops were demonstrated to assemble into quadruplex–duplex hybrids, which displayed high stability (43,45,49,50). These duplex stem-loop-containing G-quadruplex motifs were found at regulatory important regions of the human genome (51), such as the promoter region of cancer-associate genes *RAB3D* and *RAB12*, the 5'-UTR region of the *PIMI* oncogene and the intron region of myelin transcription factor. Duplex formation was shown to help drive the G-quadruplex folding process (43,45,49,52–57).

On the other hand, the formation of G-quadruplexes with bulges was reported (44). It was shown that a G-quadruplex can tolerate the presence of bulges (between adjacent guanine residues) within a G-column of the G-tetrad core, wherein increasing in the bulge size resulted in a significant decrease in the G-quadruplex stability. The

\*To whom correspondence should be addressed. Tel: +65 6514 1915; Fax: +65 6795 7981; Email: phantuan@ntu.edu.sg

formation of bulges in a G-quadruplex extended significantly the number of possible G-quadruplex-forming sequences. Recently, over 200 000 distinct bulge-containing G-quadruplex motifs were experimentally detected in the human genome by a sequencing-based method, suggesting the biological relevance of such motifs (58). Several reported bulge-containing G-quadruplex structures located at regulatory genomic regions, such as the *PARP1* promoter (36), upstream of *BCL2* promoter (59), 5'-terminal region of human telomerase RNA (hTERC) (60).

Here, we present the solution NMR structure of a G-quadruplex containing a duplex bulge, which provides the structural details of the junction between the duplex bulge and the G-quadruplex. Unexpectedly, instead of an orthogonal connection, the duplex stem was observed to stack below the G-quadruplex, forming a unique quadruplex–duplex junction. The influence of the size and position of the duplex bulge on the structure and stability of the G-quadruplex scaffold was investigated. This study revealed that a duplex bulge can be formed at various positions of a G-quadruplex scaffold. In contrast to a non-structured bulge (44), the stability of a G-quadruplex slightly increases with increasing duplex bulge size. A G-quadruplex structure containing a duplex bulge of up to 33 nt in size was shown to form. Our finding would further broaden the definition of G-quadruplex-forming sequences. These quadruplexes containing duplex bulges represent new structural duplex–quadruplex hybrid motifs with potential biological significance.

## MATERIALS AND METHODS

### DNA sample preparation

Unlabeled and site-specific  $^{15}\text{N}$ -labeled DNA oligonucleotide were synthesized on an ABI 394 DNA/RNA synthesizer using products from Glen Research and Cambridge Isotope Laboratories. The oligonucleotides were then purified following the Glen Research protocol and dialyzed successively against water, 20 mM KCl, and water again. Samples were then dried in a lyophilizer. Unless otherwise stated, the oligonucleotides were dissolved in a buffer containing 30 mM KCl and 20 mM KPi (pH 7.0).

### NMR spectroscopy

NMR experiments were performed on 600 and 700 MHz Bruker spectrometers at 25°C, unless otherwise specified. DNA samples were dissolved in a buffer containing 30 mM KCl and 20 mM KPi (pH 7). The DNA concentration of NMR samples was typically 0.2–1.0 mM. Resonances for guanine residues were assigned unambiguously using 2–4% site-specific  $^{15}\text{N}$  labeling. Spectral assignments were completed by NOESY, COSY, TOCSY and  $^{13}\text{C}$ – $^1\text{H}$ -HSQC experiments. Spectra were processed and analyzed by using the program FELIX (Felix NMR, Inc.).

### Structure calculation

Structures of the duplex–quadruplex construct *B4-dx2* were calculated using the program XPLOR-NIH (61). Hydrogen-bond, inter-proton distance, dihedral, planarity

and repulsive restraints were imposed during structure calculations. Inter-proton distance restraints, deduced from NOESY experiments at various mixing times (100, 200 and 300 ms), were classified as very strong, strong, medium, weak and very weak. Initially, an extended conformation of the sequence was generated randomly. Structure embedding was performed to produce a family of 100 structures. Next, the 100 structures were subjected to simulated-annealing regularization and refinement. The distance constraints of exchangeable and non-exchangeable protons, G-tetrad hydrogen bonding, Watson-Crick hydrogen bonding and negative constraints were incorporated in this step of calculation. 100 structures obtained from distance-geometry simulated-annealing calculation were subjected to NOE-restrained simulated-annealing refinement in XPLOR. The system was heated up from 300 to 1000 K in 14 ps, and allowed to equilibrate for 6 ps, wherein the force constant was maintained at 2 kcal mol $^{-1}$  Å $^{-2}$  for exchangeable and non-exchangeable proton distance restraints, and 8 kcal mol $^{-1}$  Å $^{-2}$  for repulsive distance restraints. Subsequently, the force constants were scaled to 32, 16, 8 and 0 kcal mol $^{-1}$  Å $^{-2}$  for hydrogen-bond, non-exchangeable proton, exchangeable proton and repulsive distance restraints, respectively. Next, the system was allowed to equilibrate at 1000 K for 200 ps, then gradually cooled down to 300 K in 42 ps. After that, the system was subjected to NOE-restrained molecular dynamics calculation for 18 ps. During the last 10 ps of NOE-restrained molecular dynamics calculation, the coordinates of the systems were saved every 0.5 ps and averaged. The resulting averaged structures were submitted to Powel minimization until the energy gradient was smaller than 0.1 Kcal/mol. Dihedral (50 kcal mol $^{-1}$  Å $^{-2}$ ) and planarity (1 kcal mol $^{-1}$  Å $^{-2}$  for G-tetrads; 0.5 kcal mol $^{-1}$  Å $^{-2}$  for Watson-Crick base pairs) restraints were maintained throughout the course of refinement. Out of 100 conformers, 10 best structures with the lowest overall energy were selected. The structures were displayed using the program PyMOL (DeLano Scientific LLC).

### Circular dichroism

Circular dichroism (CD) experiments were performed on a JASCO-815 spectropolarimeter. Unless otherwise stated, CD spectra were recorded at 25°C over the range of 220–320 nm using a 1-cm path-length quartz cuvette with a reaction volume of 500  $\mu\text{l}$ . The bandwidth, scan speed and data pitch were set at 2 nm, 200 nm/min and 1 nm, respectively. The oligonucleotide concentration was typically 3–5  $\mu\text{M}$ . For each sample, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm. For CD melting experiments, cooling and heating were successively performed across the temperature range of 15–95°C. The CD signals were recorded at intervals of 0.5°C and at the wavelength of 263 nm. Two baselines corresponding to the completely folded (low temperature) and completely unfolded (high temperature) states were manually drawn in order to determine the fractions of folded and unfolded species during the melting/folding process. The melting temperature ( $T_m$ ) is defined as the temperature for which there are equal fractions of folded and unfolded species. For each sequence,

**Table 1.** DNA sequences used in this study

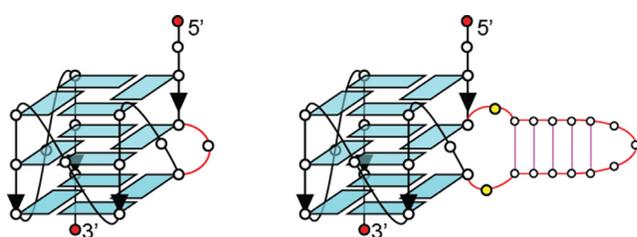
Name	Sequence <sup>a,b</sup>	$T_m$ (°C) <sup>c</sup>	Bulge size (nt)
<i>dx2</i>	<u>ATCTGAGAATCAGAT</u>	–	–
<i>B4-T</i>	TT <b>GG</b> (T) <b>G T GGG T GGG T GGG T</b>	77 <sup>d</sup>	1
<i>B4-15T</i>	TT <b>GG</b> (TTTTTTTTTTTTTTTT) <b>G T GGG T GGG T GGG T</b>	33.5	15
<i>B4-dx2</i>	TT <b>GG</b> ( <u>ATCTGAGAATCAGAT</u> ) <b>G T GGG T GGG T GGG T</b>	46.2	15
<i>B4-dx3</i>	TT <b>GG</b> ( <u>ATCTGATATTAGAATAATATCAGAT</u> ) <b>G T GGG T GGG T GGG T</b>	49.2	25
<i>B4-dx4</i>	TT <b>GG</b> ( <u>ATCTGATATTAGCAGGAAGCTGCTAATATCAGAT</u> ) <b>G T GGG T GGG T GGG T</b>	50.5	33
<i>B3-dx2</i>	TT <b>G</b> ( <u>ATCTGAGAATCAGAT</u> ) <b>GG T GGG T GGG T GGG T</b>	54.5	15
<i>B7-dx2</i>	TT <b>GGG T G</b> ( <u>ATCTGAGAATCAGAT</u> ) <b>GG T GGG T GGG T</b>	40.9	15
<i>B8-dx2</i>	TT <b>GGG T GG</b> ( <u>ATCTGAGAATCAGAT</u> ) <b>G T GGG T GGG T</b>	47.2	15
<i>B12-dx2</i>	TT <b>GGG T GGG T GG</b> ( <u>ATCTGAGAATCAGAT</u> ) <b>G T GGG T</b>	45.6	15

<sup>a</sup>G-tracts are shown in bold-face.

<sup>b</sup>Individual bulges are enclosed by parentheses with complementary tracts underlined.

<sup>c</sup>The melting temperatures ( $T_m$ ) of samples in a buffer containing 30 mM KCl and 20 mM KPi (pH 7) were monitored by the CD signals at 263 nm.

<sup>d</sup> $T_m$  was obtained from reference (44).



**Figure 1.** Schematics of a G-quadruplex containing a single-nucleotide bulge (left) and a duplex bulge (right).

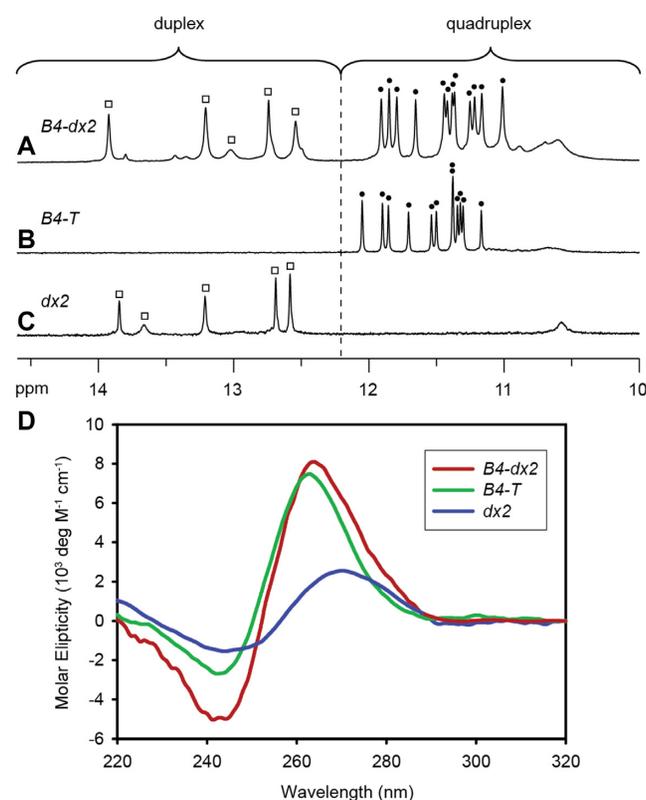
the average  $T_m$  value from the folding and unfolding experiments is presented. For clarity, only the folding curves are presented.

## RESULTS AND DISCUSSION

### Inclusion of a duplex hairpin within a bulge of a G-quadruplex

Previously, it was shown that the sequence TTGGTGTGGGTGGGTGGGT (henceforth denoted as *B4-T*, bulge underlined; Table 1) could form a stable all-parallel-stranded G-quadruplex structure with a bulge T5 residue positioned between G4 and G6 of the first G-column (Figure 1, left) (44). To investigate the potential expansion of the bulge for an inclusion of a duplex stem, T5 was replaced by a duplex hairpin-forming motif *dx2*, giving rise to the sequence TTGG(ATCTGAGAATCAGAT)GTGGGTGGGTGGGT (henceforth denoted as *B4-dx2*; duplex hairpin *dx2* enclosed by parentheses with complementary sequences underlined; Table 1).

One-dimensional (1D) imino proton NMR spectrum of *B4-dx2* showed twelve major peaks at ~10.5–12 ppm (marked by filled circles, Figure 2A), characteristic of guanine imino protons involved in the formation of G-tetrads. Five additional imino proton peaks were observed at ~12.5–14 ppm (marked by open squares, Figure 2A), indicative of the duplex hairpin formation. Distribution patterns of these imino proton peaks closely resembled those of the individual components, G-quadruplex (*B4-T*; Figure 2B) and duplex hairpin (*dx2*; Figure 2C), indicative of the formation of a quadruplex–duplex hybrid with a duplex hairpin located within the bulge of the all-

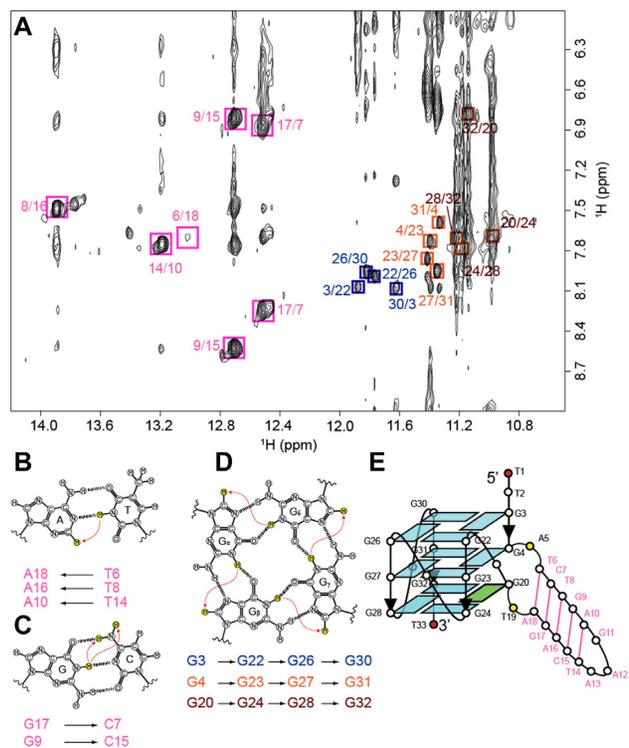


**Figure 2.** 1D imino proton NMR spectra of (A) *B4-dx2*, (B) *B4-T* and (C) *dx2*. Imino protons of the G-tetrad core are indicated by filled circles, whereas imino protons of the duplex are indicated by open squares. (D) CD spectra of *B4-T* (green), *dx2* (blue) and *B4-dx2* (red).

parallel-stranded G-quadruplex scaffold (Figure 1, right). CD spectrum of *B4-dx2* showed an intense positive peak at ~260 nm and a negative trough at ~240 nm (Figure 2D), supporting the formation of an all-parallel-stranded G-quadruplex (62–64). Overall, these observations corroborated the quadruplex–duplex hybrid nature of *B4-dx2*.

### NMR spectral assignments of *B4-dx2*

We proceeded to elucidate the solution structure of *B4-dx2*. Guanine imino and aromatic protons were unambiguously



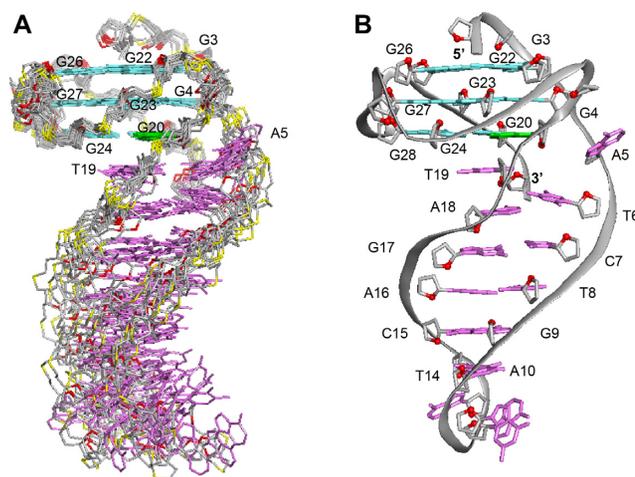
**Figure 3.** NMR structural characterization of *B4-dx2*. (A) NOESY spectrum (mixing time, 200 ms) showing cross-peaks for the identification of three G-tetrads and five Watson-Crick base pairs. Signature NOE cross-peaks for the Watson-Crick base pairs and G-tetrads are framed and labeled with the residue numbers of the imino and H8 protons at the first and second positions, respectively. (B) NOE between the thymine imino proton and the adenine H2 proton in an A•T base pair. (C) NOEs between the guanine imino proton and the cytosine amino protons in a G•C base pair. (D) NOEs between H8 and imino protons within a G-tetrad. (E) Schematic diagram of *B4-dx2*.

assigned (Supplementary Figure S1) through site-specific  $^{15}\text{N}$ -labeled (4%) samples (65). Spectral assignment of proton resonances of *B4-dx2* was completed by using through-bond (COSY and TOCSY) and through-space (NOESY) NMR experiments. The imino proton of the isolated guanine G20 was clearly detected at the region of 10.5–12 ppm (Supplementary Figure S1), reflecting its participation in the G-tetrad formation. Other 11 guanines involved in G-tetrads are G3–G4, G22–G24, G26–G28 and G30–G32.

### Folding topology of *B4-dx2*

Within a G-tetrad layer, the imino proton of a guanine is in proximity of the H8 proton of the adjacent guanine and therefore would exhibit a NOE cross-peak. By observing such cross-peaks, the folding topology of a G-quadruplex can be defined (65). The NOEs of G3(H1)–G22(H8), G22(H1)–G26(H8), G26(H1)–G30(H8) and G30(H1)–G3(H8) (Figure 3A) pointed to the formation of the tetrad plane of G3•G22•G26•G30. Similarly, the other two tetrads, G4•G23•G27•G31 and G20•G24•G28•G32, were defined (Figure 3D).

Signature NOE cross-peaks can be observed in Watson–Crick base pairs due to the proximity between respective



**Figure 4.** NMR solution structure of *B4-dx2*. (A) Ten superimposed lowest-energy structures. (B) Ribbon view of a representative structure. *Anti* guanines in the G-tetrads are colored in cyan; *syn* guanine in the G-tetrads, green; bases in the duplex stem, magenta; backbone and sugar, gray; O4' atoms, red; phosphorus atoms, yellow.

protons: NOE between the thymine imino proton and the adenine H2 proton in an A•T base pair (Figure 3A,B), as well as NOEs between the guanine imino proton and the cytosine amino protons in a G•C base pair (Figure 3A, C). These signature NOE cross-peaks were observed for five base pairs (T6•A18, G17•C7, T8•A16, G9•C15 and T14•A10), pointing to the formation of the duplex stem (Figure 3E).

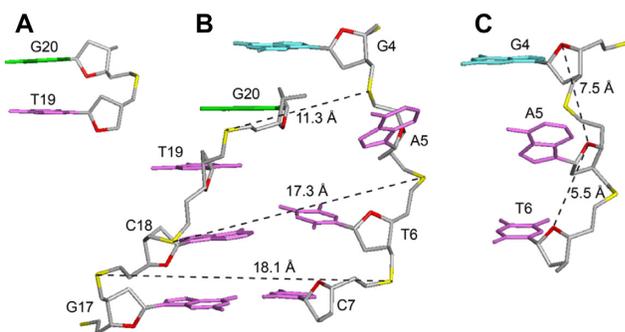
The formation and arrangement of three G-tetrads and five Watson–Crick base pairs validated the topology of *B4-dx2* as an all-parallel-stranded G-quadruplex involving three continuous tracts of guanines: G22–G24, G26–G28, G30–G31 and one discontinuous tract: G3, G4 and G20 (Figure 3E). This discontinuous tract results in a bulge between G4 and G20, forming a 5-bp duplex stem. Four G-tracts are connected by three 1-nt double-chain-reversal loops: T21, T25, and T29. NOEs observed between the protons of G20 and those of G4, e.g. G20(H8)–G4(H8), G20(H8)–G4(H1'), G20(H8)–G4(H2'/H2''), G20(H8)–G4(H3'), G20(H1')–G4(H1') and G20(H1)–G4(H1), support the position of G20 in the G-tetrad core below G4. The typical NOE sequential connectivities ( $n$ )H8 and ( $n - 1$ )H1'/H2'/H2''/H3'/H4' are clearly observed along the G-columns and two strands of the duplex hairpin (Supplementary Figure S2). These NOE sequential connectivities are broken at the transition point from the first G-column to the bulge and at the three double-chain-reversal loops (T21, T25 and T29). Moderate intensities of the intra-residue H8–H1' NOE cross-peaks reflect *anti* glycosidic conformations for all guanine residues except G20, which adopts a *syn* conformation as seen by its strong intra-residue H8–H1' NOE cross-peak.

### Solution structure of *B4-dx2*

The structure of *B4-dx2* (Figure 4) was calculated on the basis of NMR restraints (Table 2) using the XPLOR-NIH program (61). The G-tetrad core of *B4-dx2* adopts a three-

**Table 2.** Statistics of the computed structures of *B4-dx2*

A. NMR Restraint			
Distance restraints	D <sub>2</sub> O	H <sub>2</sub> O	
Intraresidue	508	9	
Sequential ( <i>i, i+1</i> )	194	17	
Long-range ( <i>i, ≥i+2</i> )	25	46	
Other restraints			
Hydrogen-bond		72	
Dihedral		22	
Repulsive		15	
B. Structure statistics			
NOE violation			
Number (>0.2 Å)	0.500 ± 0.972		
Maximum violation (Å)	0.189 ± 0.063		
Rmsd of violation (Å)	0.015 ± 0.002		
Deviations from the ideal covalent geometry			
Bond length (Å)	0.003 ± 0.000		
Bond angles (°)	0.686 ± 0.009		
Impropers (°)	0.340 ± 0.006		
Pairwise all heavy atom rmsd (Å)			
All heavy atoms in G-tetrad core	0.486 ± 0.120		
All heavy atoms in duplex hairpin	0.578 ± 0.154		
All heavy atoms	1.086 ± 0.269		

**Figure 5.** Structural features of *B4-dx2* at the quadruplex–duplex junction. (A) Stacking between T19 (magenta) and G20 (green). (B) Backbone strand separation at the junction for the duplex stem and the G-tetrad core (C) Stretching of backbone between the G4 and A5 residues.

layered all-parallel-stranded scaffold, with three single-nucleotide double-chain-reversal loops (T21, T25 and T29). The bases of these loops project outwards from the G-tetrad core.

The first two thymine residues (T1 and T2) are positioned on top of the G-tetrad core; the last thymine T33 is stacked below the bottom G-tetrad. The duplex at the bulge formed between G4 and G20 includes five Watson–Crick base pairs, while the immediate possible base pair step at the quadruplex–duplex junction, A5–T19, is broken up (Figure 5B). The duplex hairpin shows continuous stacking under the 3'-end G-tetrad, with T19 directly sandwiched between the tetrad and duplex while the base of A5 projects outward from the groove. Stacking between the duplex and the quadruplex is reflected by the generally upfield-shifted imino proton peaks of the G-tetrad core of *B4-dx2* (Figure 2A) as compared to those of *B4-T* (Figure 2B) in the region of 11–12 ppm. The placement of T19 right below the 6-membered ring of G20 of the G-tetrad (Figure 5A) is consistent with the up-field shift of resonances belonging to T19 and G20. Stacking between T19 and G20 is supported by

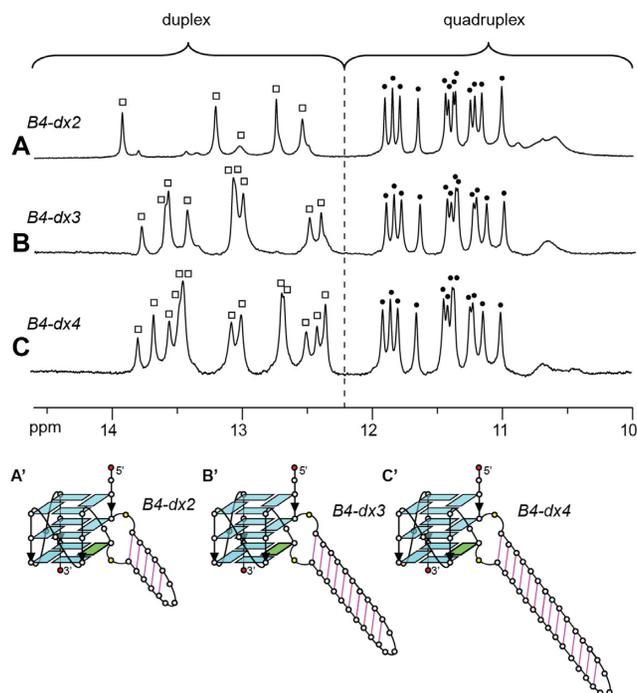
the following observed NOEs: T19(H2'/H2'')–G20(H1), T19(H1')–G20(H1), T19(CH<sub>3</sub>)–G20(H1), T19(CH<sub>3</sub>)–G24(H1), T19(CH<sub>3</sub>)–G24(H8) and T19(H6)–G24(H8). To accommodate a duplex hairpin insertion within the context of a quadruplex bulge, the strand separation (defined as the distance between matching pairs of phosphate groups) at the A5–T19 step is narrower (11.3 Å), as compared to the next T6–A18 step (17.3 Å) and a normal Watson–Crick base pair step (~18 Å) within the duplex stem (Figure 5B). Moreover, the stepwise O4'-to-O4' distance between G4 and A5 at the junction was found to be 7.5 Å, which is considerably more stretched as compared to that observed within a duplex (5–6 Å) (Figure 5C). This configuration of A5 permits a progressive helical transition between the quadruplex and duplex segments, thus maximizing the stability of the overall structure.

Previously, various structural configurations have been observed at the quadruplex–duplex junction to bridge the transition between the two distinct helical segments. A duplex stem, when placed in different loops, can be connected to a quadruplex in either orthogonal or co-axial orientation (43). In the former case, the base pair step at the junction was found to flip open, while in the latter case continuous stacking between G-tetrads and base pairs in the duplex was observed (Supplementary Figure S3). On one hand, the duplex connection to a bulge resembles orthogonal connection, leading to the broken first base pair step. On the other hand, in *B4-dx2* only A5 was projected outward from the overall helix and continuous base stacking between the quadruplex and duplex was observed. Across other quadruplex–duplex hybrid structures, various adaptor base pairs or structural elements have been observed to bridge the two helical domains (Supplementary Figure S4) (43,50,66,67).

### Effects of the bulge size on the structure and stability of a G-quadruplex containing a duplex bulge

To explore the effects of larger duplex bulges on the structure and stability of a G-quadruplex scaffold, we investigated similar sequences containing duplex bulges of 25 and 33 nt (denoted as *B4-dx3* and *B4-dx4*, respectively; Table 1). In 1D imino proton NMR spectra of *B4-dx2*, *B4-dx3* and *B4-dx4*, twelve major peaks were observed at the 10.5–12 ppm region for all sequences, while the number of peaks at the 12–14 ppm region increased with increasing duplex stem length (Figure 6). The similarity of spectral patterns for *B4-dx2*, *B4-dx3* and *B4-dx4* in the region of 10.5–12 ppm suggested the formation of the same G-quadruplex scaffold. CD spectra of these three structures also showed the signature of an all-parallel-stranded G-quadruplex structure with a positive peak at ~260 nm and a negative trough at ~240 nm (Supplementary Figure S5).

These structures exhibited a high thermal stability, with the melting temperature ( $T_m$ ) ranging from 46 to 50.5°C in the presence of ~60 mM K<sup>+</sup> (Table 1). Previously, it has been shown that increasing the bulge size led to decreasing stability of the G-quadruplex (44). In this study, as observed in the case of *B4-dx2*, *B4-dx3* and *B4-dx4*, increasing the size of a duplex bulge did not lead to a decrease in the stability of the G-quadruplex. On the contrary, the melt-

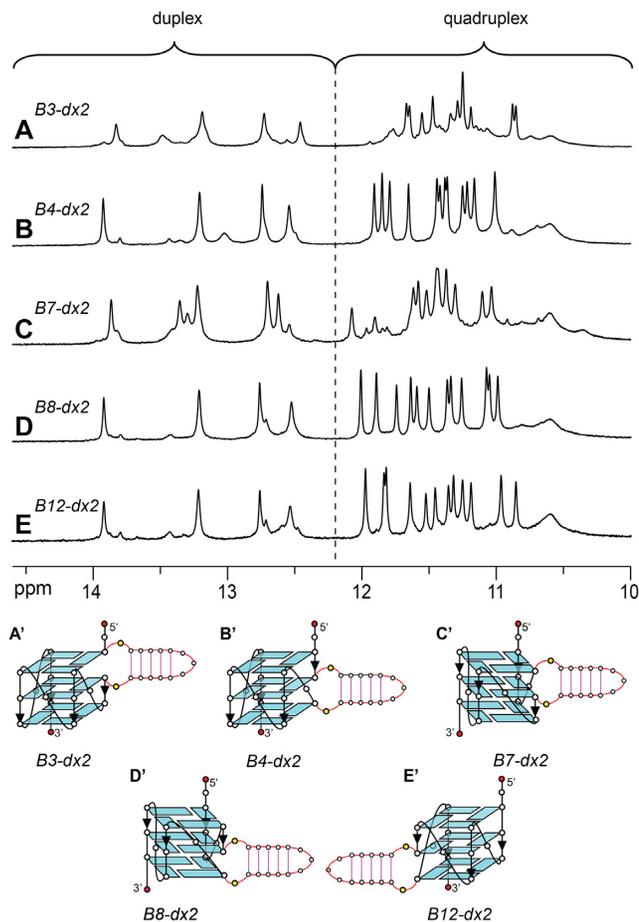


**Figure 6.** 1D imino proton NMR spectra and schematics of G-quadruplexes containing a duplex bulge of different sizes: (A, A') *B4-dx2* (15 nt), (B, B') *B4-dx3* (25 nt) and (C, C') *B4-dx4* (33 nt).

ing temperature slightly increased from 46.2°C to 50.5°C, as the duplex bulge size was increased from 15 to 33 nt (Supplementary Figure S6). A similar trend was also observed when a duplex of increasing size was incorporated within a G-quadruplex loop (49). The melting temperature of a G-quadruplex was observed to increase by ~13°C (Supplementary Figure S6) when a non-structured bulge (*B4-15T*; Table 1) is replaced by a duplex bulge with the same length (*B4-dx2*), indicating that a G-quadruplex with a duplex bulge exhibited significantly higher thermal stability than a G-quadruplex containing a non-structured bulge of the same length.

### Insertion of a duplex bulge at different positions of a G-quadruplex

To investigate the effect of different placements of a duplex bulge on the structure and stability of a G-quadruplex scaffold, the duplex hairpin *dx2* was inserted at various bulge positions within the G-tetrad core (sequences *B3-dx2*, *B4-dx2*, *B7-dx2*, *B8-dx2* and *B12-dx2*, Table 1). Similar CD spectra, showing a positive peak at ~260 nm and a negative peak at ~240 nm, were observed for all five sequences (Supplementary Figure S7), suggesting that these sequences adopt an all-parallel-stranded G-quadruplex fold. In all cases, imino proton peaks were observed in both G-quadruplex (10.5–12 ppm) and duplex (12–14 ppm) characteristic regions (Figure 7), supporting the formation of G-quadruplex–duplex hybrids. In the cases of *B4-dx2*, *B8-dx2* and *B12-dx2*, a major conformation with similar spectral patterns was observed, while in the cases of *B3-dx2* and *B7-dx2*, significant minor conformations were also observed.



**Figure 7.** 1D imino proton NMR spectra and schematics of quadruplex–duplex hybrids with a duplex hairpin inserted at different bulge positions: (A, A') *B3-dx2* with duplex at the bulge position 3, (B, B') *B4-dx2* with duplex at the bulge position 4, (C, C') *B7-dx2* with duplex at the bulge position 7, (D, D') *B8-dx2* with duplex at the bulge position 8, and (E, E') *B12-dx2* with duplex at the bulge position 12.

The stability of these structures was found to depend on the position of the duplex bulge, with  $T_m$  ranging from 40.9°C to 54.5°C in the presence of ~60 mM  $K^+$  (Table 1).

### Broadening the definition of G-quadruplex forming sequences

It was previously shown that sequences with disrupted G-tracts can form G-quadruplexes with bulges of a size ranging from 1 to 7 nt (44). Among over 700 000 G-quadruplexes experimentally detected in the human genome, over 200 000 of which were G-quadruplexes with bulges (58). In this study, we have shown that the size of a bulge can be extended to 33 nt or even longer if base pairs are formed within the bulge. Such a duplex bulge can exist at different positions within a G-quadruplex scaffold, significantly increasing the space of G-quadruplex-forming sequences beyond the standard  $G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+$  motif previously defined for potential G-quadruplex-forming sequences (41,42). This has broadened the definition of G-quadruplex-forming sequences and can greatly increase the number of potential G-quadruplex-forming sites in the genome.

### A unique structure for specific G-quadruplex targeting

Targeting G-quadruplexes was shown to be a promising strategy for anti-cancer therapy (68). Numerous G-quadruplex-binding ligands have been developed with the current generation of ligands exhibiting high affinity toward G-quadruplexes and high selectivity between G-quadruplex and other DNA structures (69–79). However, these ligands still show poor selectivity between different G-quadruplex structures. As most G-quadruplex-binding ligands stack on the top and bottom G-tetrads, it is challenging to achieve specific targeting of a single G-quadruplex amongst over 700 000 potential G-quadruplex structures across the human genome (58). There has been increasing interest in designing specific G-quadruplex-binding ligands, where additional structural features besides the G-tetrad core, such as loops and bulges, could provide insights for the design of highly specific ligands. The G-quadruplex structure with a duplex bulge in this study presents a new type of quadruplex–duplex hybrid which can serve as a unique structure for specific G-quadruplex targeting. Such a motif was identified in the *BCL2* proximal promoter (59), suggesting its biological relevance and therapeutic potential. The *BCL2* protein plays an important role in the regulation of programmed cell death and its overexpression is correlated with a large number of cancers. There are two GC-rich regions identified in the upstream of the *P1* promoter of *BCL2* (denoted as *Pu39* and *PIG4*) (59). The major G-quadruplex form adopted by *Pu39*, which was shown to be a *BCL2* transcription activator, was a parallel G-quadruplex with one 13-nt and two 1-nt unstructured loops (80). On the other hand, *PIG4* can form two interchangeable parallel G-quadruplex conformers: a duplex stem-loop G-quadruplex and a duplex stem-bulge G-quadruplex, which were shown to be transcription repressors (59). *BCL2* contains multiple transcription start sites and the formation of *Pu39* or *PIG4* G-quadruplexes could be correlated with the transcription start sites being used (59). The existence of a duplex stem in the *PIG4* G-quadruplexes could provide a unique feature for these G-quadruplexes as compared to *Pu39* or other G-quadruplexes for the binding of protein or small molecule to precisely regulate the transcription of *BCL2*. Unique structural features of a G-quadruplex with a duplex bulge, such as a broken G-column induced by a bulge, can allow the intercalation of a ligand, while the minor groove of a duplex bulge can serve as a substrate for specific targeting by ligands with duplex sequence selectivity (81–84). Pyrrole-imidazole polyamides, for example, can be programmed to target a duplex sequence with high selectivity (85,86). Highly specific targeting of a G-quadruplex structure containing a duplex bulge can be achieved by simultaneous targeting the duplex part, the G-quadruplex part and the junction between the duplex and the G-quadruplex.

### CONCLUSION

In this study, we showed that a duplex stem can be incorporated into a bulge of a G-quadruplex. The NMR solution structure of a G-quadruplex containing a duplex bulge was presented, showing a unique quadruplex–duplex junction with the duplex bulge stacking below the 3'-end G-tetrad. Breaking up of the immediate base pair step at the

quadruplex–duplex junction, coupled with a narrowing of the duplex groove within the context of the bulge, led to a progressive transition between the quadruplex and duplex segments. Duplex bulges can occur at various positions of a G-quadruplex scaffold and increasing the length of a duplex bulge does not lead to a decrease in the stability of the overall G-quadruplex core. The formation of a long duplex bulge within G-quadruplexes would expand the current knowledge on predicting G-quadruplex-forming sequences. Potential existence of such structures in the human genome may serve as unique targets for designing ligands for specific G-quadruplex targeting.

### DATA AVAILABILITY

The coordinates for the NMR solution structure of *B4-dx2* have been deposited in the Protein Data Bank (PDB ID: 7CLS).

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### FUNDING

Singapore National Research Foundation Investigatorship [NRF-NRFI2017-09]; Nanyang Technological University grants (to A.T.P.). Funding for open access charge: Singapore National Research Foundation.

Conflict of interest statement. None declared.

### REFERENCES

- Davis, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
- Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
- Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738–9745.
- Lightfoot, H.L., Hagen, T., Tatum, N.J. and Hall, J. (2019) The diverse structural landscape of quadruplexes. *FEBS Lett.*, **593**, 2083–2102.
- Winnerdy, F.R. and Phan, A.T. (2020) Quadruplex structure and diversity. *Annu. Rep. Med. Chem.*, **54**, 45–73.
- Maizels, N. and Gray, L.T. (2013) The G4 genome. *PLoS Genet.*, **9**, e1003468.
- Dai, J., Carver, M. and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie.*, **90**, 1172–1183.
- Phan, A.T. (2010) Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.
- Lim, K.W., Ng, V.C., Martin-Pintado, N., Heddi, B. and Phan, A.T. (2013) Structure of the human telomere in Na<sup>+</sup> solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.*, **41**, 10556–10562.
- Galer, P., Wang, B., Sket, P. and Plavec, J. (2016) Reversible pH switch of two-quartet G-quadruplexes formed by human telomere. *Angew Chem. Int. Ed. Engl.*, **55**, 1993–1997.

14. Russo Krauss, I., Ramaswamy, S., Neidle, S., Haider, S. and Parkinson, G.N. (2016) Structural insights into the quadruplex–duplex 3' interface formed from a telomeric repeat: a potential molecular target. *J. Am. Chem. Soc.*, **138**, 1226–1233.
15. Liu, C., Zhou, B., Geng, Y., Yan Tam, D., Feng, R., Miao, H., Xu, N., Shi, X., You, Y., Hong, Y. *et al.* (2019) A chair-type G-quadruplex structure formed by a human telomeric variant DNA in K<sup>+</sup> solution. *Chem. Sci.*, **10**, 218–226.
16. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the *NRAS* proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
17. Gomez, D., Guedin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P. and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.
18. Endoh, T., Kawasaki, Y. and Sugimoto, N. (2013) Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chem. Int. Ed. Engl.*, **52**, 5522–5526.
19. Rouleau, S., Jodoin, R., Garant, J.-M. and Perreault, J.-P. (2020) In: Seitz, H., Stahl, F. and Walter, J.-G. (eds). *Catalytically Active Nucleic Acids*, Springer International Publishing, Cham, pp. 1–20.
20. Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
21. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.
22. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
23. Cogoi, S. and Xodo, L.E. (2006) G-quadruplex formation within the promoter of the *KRAS* proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
24. Podbešek, P. and Plavec, J. (2016) *KRAS* promoter oligonucleotide with decoy activity dimerizes into a unique topology consisting of two G-quadruplex units. *Nucleic Acids Res.*, **44**, 917–925.
25. Kerkour, A., Marquevielle, J., Ivashchenko, S., Yatsunyk, L.A., Mergny, J.L. and Salgado, G.F. (2017) High-resolution three-dimensional NMR structure of the *KRAS* proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem.*, **292**, 8082–8091.
26. Marquevielle, J., Robert, C., Lagrabet, O., Wahid, M., Bourdoncle, A., Xodo, L.E., Mergny, J.-L. and Salgado, G.F. (2020) Structure of two G-quadruplexes in equilibrium in the *KRAS* promoter. *Nucleic Acids Res.*, doi:10.1093/nar/gkaa387.
27. Ou, A., Schmidberger, J.W., Wilson, K.A., Evans, C.W., Hargreaves, J.A., Grigg, M., O'Mara, M.L., Iyer, K.S., Bond, C.S. and Smith, N.M. (2020) High resolution crystal structure of a *KRAS* promoter G-quadruplex reveals a dimer with extensive poly-A  $\pi$ -stacking interactions for small-molecule recognition. *Nucleic Acids Res.*, **48**, 5766–5776.
28. Dai, J.X., Chen, D., Jones, R.A., Hurley, L.H. and Yang, D.Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human *BCL2* promoter region. *Nucleic Acids Res.*, **34**, 5133–5144.
29. Palumbo, S.L., Ebbinghaus, S.W. and Hurley, L.H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the *hTERT* core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.*, **131**, 10878–10891.
30. Lim, K.W., Lacroix, L., Yue, D.J.E., Lim, J.K.C., Lim, J.M.W. and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the *hTERT* promoter. *J. Am. Chem. Soc.*, **132**, 12331–12342.
31. Micheli, E., Martufi, M., Cacchione, S., De Santis, P. and Savino, M. (2010) Self-organization of G-quadruplex structures in the *hTERT* core promoter stabilized by polyaminic side chain perylene derivatives. *Biophys. Chem.*, **153**, 43–53.
32. Sun, D.Y., Guo, K.X., Rusche, J.J. and Hurley, L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human *VEGF* gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
33. De Armond, R., Wood, S., Sun, D.Y., Hurley, L.H. and Ebbinghaus, S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 $\alpha$  promoter. *Biochemistry*, **44**, 16341–16350.
34. Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the *RET* oncogene. *J. Am. Chem. Soc.*, **129**, 10220–10228.
35. Tong, X.T., Lan, W.X., Zhang, X., Wu, H.M., Liu, M.L. and Cao, C.Y. (2011) Solution structure of all parallel G-quadruplex formed by the oncogene *RET* promoter sequence. *Nucleic Acids Res.*, **39**, 6753–6763.
36. Sengar, A., Vandana, J.J., Chambers, V.S., Di Antonio, M., Winnerdy, F.R., Balasubramanian, S. and Phan, A.T. (2019) Structure of a (3+1) hybrid G-quadruplex in the *PARP1* promoter. *Nucleic Acids Res.*, **47**, 1564–1572.
37. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
38. Yamaoki, Y., Kiyoshi, A., Miyake, M., Kano, F., Murata, M., Nagata, T. and Katahira, M. (2018) The first successful observation of in-cell NMR signals of DNA and RNA in living human cells. *Phys. Chem. Chem. Phys.*, **20**, 2982–2985.
39. Krafcikova, M., Dzatko, S., Caron, C., Granzhan, A., Fiala, R., Loja, T., Teulade-Fichou, M.-P., Fessl, T., Hänsel-Hertsch, R., Mergny, J.-L. *et al.* (2019) Monitoring DNA–ligand interactions in living human cells using NMR spectroscopy. *J. Am. Chem. Soc.*, **141**, 13281–13285.
40. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
41. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
42. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
43. Lim, K.W. and Phan, A.T. (2013) Structural basis of DNA quadruplex–duplex junction formation. *Angew. Chem. Int. Ed. Engl.*, **52**, 8566–8569.
44. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.
45. Lim, K.W., Nguyen, T.Q.N. and Phan, A.T. (2014) Joining of multiple duplex stems at a single quadruplex loop. *J. Am. Chem. Soc.*, **136**, 17969–17973.
46. Bedrat, A., Lacroix, L. and Mergny, J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746–1759.
47. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human *MYC* promoter. *Nat. Chem. Biol.*, **1**, 167–173.
48. Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
49. Lim, K.W., Khong, Z.J. and Phan, A.T. (2014) Thermal stability of DNA quadruplex–duplex hybrids. *Biochemistry*, **53**, 247–257.
50. Butovskaya, E., Heddi, B., Bakalar, B., Richter, S.N. and Phan, A.T. (2018) Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc.*, **140**, 13654–13662.
51. Lim, K.W., Jenjaroenpun, P., Low, Z.J., Khong, Z.J., Ng, Y.S., Kuznetsov, V.A. and Phan, A.T. (2015) Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **43**, 5630–5646.
52. Huang, Y.C., Cheng, A.K.H., Yu, H.-Z. and Sen, D. (2009) Charge conduction properties of a parallel-stranded DNA G-quadruplex: implications for chromosomal oxidative damage. *Biochemistry-US*, **48**, 6794–6804.
53. Zhou, J., Bourdoncle, A., Rosu, F., Gabelica, V. and Mergny, J.L. (2012) Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed. Engl.*, **51**, 11002–11005.

54. Yu, Z., Gaerig, V., Cui, Y., Kang, H., Gokhale, V., Zhao, Y., Hurley, L.H. and Mao, H. (2012) Tertiary DNA structure in the single-stranded *hTERT* promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J. Am. Chem. Soc.*, **134**, 5157–5164.
55. Mendoza, O., Porrini, M., Salgado, G.F., Gabelica, V. and Mergny, J.L. (2015) Orienting tetramolecular G-quadruplex formation: the quest for the elusive RNA antiparallel quadruplex. *Chemistry*, **21**, 6732–6739.
56. Karg, B., Mohr, S. and Weisz, K. (2019) Duplex-guided refolding into novel G-quadruplex (3+1) hybrid conformations. *Angew. Chem. Int. Ed. Engl.*, **58**, 11068–11071.
57. Nguyen, T.Q.N., Lim, K.W. and Phan, A.T. (2020) Folding kinetics of G-quadruplexes: Duplex stem loops drive and accelerate G-quadruplex folding. *J. Phys. Chem. B*, **124**, 5122–5130.
58. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P. and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
59. Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M. and Yang, D. (2016) A new G-quadruplex with hairpin loop immediately upstream of the human *BCL2* P1 promoter modulates transcription. *J. Am. Chem. Soc.*, **138**, 2563–2570.
60. Martadinata, H. and Phan, A.T. (2014) Formation of a stacked dimeric G-quadruplex containing bulges by the 5'-terminal region of human telomerase RNA (*hTERC*). *Biochemistry*, **53**, 1595–1600.
61. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
62. Kypr, J., Kejnovska, I., Rencuk, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
63. Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A. and Webba da Silva, M. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed. Engl.*, **50**, 10645–10648.
64. Del Villar-Guerra, R., Trent, J.O. and Chaires, J.B. (2018) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **57**, 7171–7175.
65. Adrian, M., Heddi, B. and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11–24.
66. Huang, H., Suslov, N.B., Li, N.S., Shelke, S.A., Evans, M.E., Koldobskaya, Y., Rice, P.A. and Piccirilli, J.A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.*, **10**, 686–691.
67. Russo Krauss, I., Spiridonova, V., Pica, A., Napolitano, V. and Sica, F. (2016) Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding. *Nucleic Acids Res.*, **44**, 983–991.
68. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: A novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
69. Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627–636.
70. Dai, J., Carver, M., Hurley, L.H. and Yang, D. (2011) Solution structure of a 2:1 quindoline-*c-MYC* G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.*, **133**, 17673–17680.
71. Collie, G.W., Promontorio, R., Hampel, S.M., Micco, M., Neidle, S. and Parkinson, G.N. (2012) Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *J. Am. Chem. Soc.*, **134**, 2723–2731.
72. Nicoludis, J.M., Miller, S.T., Jeffrey, P.D., Barrett, S.P., Rablen, P.R., Lawton, T.J. and Yatsunyk, L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*, **134**, 20446–20456.
73. Chung, W.J., Heddi, B., Tera, M., Iida, K., Nagasawa, K. and Phan, A.T. (2013) Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative. *J. Am. Chem. Soc.*, **135**, 13495–13501.
74. Bazzicalupi, C., Ferraroni, M., Bilia, A.R., Scheggi, F. and Gratteri, P. (2013) The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.*, **41**, 632–638.
75. Chung, W.J., Heddi, B., Hamon, F., Teulade-Fichou, M.P. and Phan, A.T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC<sub>3</sub>. *Angew. Chem. Int. Ed. Engl.*, **53**, 999–1002.
76. Kotar, A., Wang, B., Shivalingam, A., Gonzalez-Garcia, J., Vilar, R. and Plavec, J. (2016) NMR structure of a Triangulenium-based long-lived fluorescence probe bound to a G-quadruplex. *Angew. Chem. Int. Ed. Engl.*, **55**, 12508–12511.
77. Wirmer-Bartoschek, J., Bendel, L.E., Jonker, H.R.A., Grun, J.T., Papi, F., Bazzicalupi, C., Messori, L., Gratteri, P. and Schwalbe, H. (2017) Solution NMR structure of a ligand/hybrid-2-G-quadruplex complex reveals rearrangements that affect ligand binding. *Angew. Chem. Int. Ed. Engl.*, **56**, 7102–7106.
78. Liu, W., Zhong, Y.-F., Liu, L.-Y., Shen, C.-T., Zeng, W., Wang, F., Yang, D. and Mao, Z.-W. (2018) Solution structures of multiple G-quadruplex complexes induced by a platinum(II)-based tripod reveal dynamic binding. *Nat. Commun.*, **9**, 3496.
79. Calabrese, D.R., Chen, X., Leon, E.C., Gaikwad, S.M., Phyto, Z., Hewitt, W.M., Alden, S., Hilimire, T.A., He, F., Michalowski, A.M. et al. (2018) Chemical and structural studies provide a mechanistic basis for recognition of the *MYC* G-quadruplex. *Nat. Commun.*, **9**, 4229.
80. Agrawal, P., Lin, C., Mathad, R.I., Carver, M. and Yang, D. (2014) The major G-quadruplex formed in the human *BCL-2* proximal promoter adopts a parallel structure with a 13-nt loop in K<sup>+</sup> solution. *J. Am. Chem. Soc.*, **136**, 1750–1753.
81. Nguyen, T.Q.N., Lim, K.W. and Phan, A.T. (2017) A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs. *Sci. Rep.*, **7**, 11969.
82. Asamitsu, S., Bando, T. and Sugiyama, H. (2019) Ligand design to acquire specificity to intended G-quadruplex structures. *Chemistry*, **25**, 417–430.
83. Asamitsu, S., Obata, S., Phan, A.T., Hashiya, K., Bando, T. and Sugiyama, H. (2018) Simultaneous binding of hybrid molecules constructed with dual DNA-binding components to a G-quadruplex and its proximal duplex. *Chemistry*, **24**, 4428–4435.
84. Mandal, S., Kawamoto, Y., Yue, Z., Hashiya, K., Cui, Y., Bando, T., Pandey, S., Hoque, M.E., Hossain, M.A., Sugiyama, H. et al. (2019) Submolecular dissection reveals strong and specific binding of polyamide-pyridostatin conjugates to human telomere interface. *Nucleic Acids Res.*, **47**, 3295–3305.
85. Dervan, P.B. and Edelson, B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Chem. Biol.*, **13**, 284–299.
86. Kawamoto, Y., Bando, T. and Sugiyama, H. (2018) Sequence-specific DNA binding Pyrrole-imidazole polyamides and their applications. *Bioorg. Med. Chem.*, **26**, 1393–1411.