



**HAL**  
open science

## Bulges in left-handed G-quadruplexes

Poulomi Das, Khac Huy Ngo, Fernaldo richtia Winnerdy, Arijit Maity, Blaž Bakalar, Yves Mechulam, Emmanuelle Schmitt, Anh Tuân Phan

► **To cite this version:**

Poulomi Das, Khac Huy Ngo, Fernaldo richtia Winnerdy, Arijit Maity, Blaž Bakalar, et al..  
Bulges in left-handed G-quadruplexes. *Nucleic Acids Research*, 2021, 49 (3), pp.1724-1736.  
10.1093/nar/gkaa1259 . hal-03188426

**HAL Id: hal-03188426**

**<https://polytechnique.hal.science/hal-03188426>**

Submitted on 6 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Bulges in left-handed G-quadruplexes

Poulomi Das<sup>1</sup>, Khac Huy Ngo<sup>1</sup>, Fernaldo Richtia Winnerdy<sup>1</sup>, Arijit Maity<sup>1</sup>, Blaž Bakalar<sup>1</sup>, Yves Mechulam<sup>2</sup>, Emmanuelle Schmitt<sup>2</sup> and Anh Tuấn Phan<sup>1,3,\*</sup>

<sup>1</sup>School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore, <sup>2</sup>Laboratoire de Biologie Structurale de la Cellule (BIOC), Ecole Polytechnique, CNRS-UMR7654, Institut Polytechnique de Paris, Palaiseau 91128, France and <sup>3</sup>NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921, Singapore

Received September 30, 2020; Revised December 10, 2020; Editorial Decision December 14, 2020; Accepted December 18, 2020

## ABSTRACT

**G-quadruplex (G4) DNA structures with a left-handed backbone progression have unique and conserved structural features. Studies on sequence dependency of the structures revealed the prerequisites and some minimal motifs required for left-handed G4 formation. To extend the boundaries, we explore the adaptability of left-handed G4s towards the existence of bulges. Here we present two X-ray crystal structures and an NMR solution structure of left-handed G4s accommodating one, two and three bulges. Bulges in left-handed G4s show distinct characteristics as compared to those in right-handed G4s. The elucidation of intricate structural details will help in understanding the possible roles and limitations of these unique structures.**

## INTRODUCTION

Four guanine bases with Hoogsteen hydrogen bonds arranged symmetrically between themselves can form a planar structure commonly termed as G-tetrad (1). Such G-tetrads could stack on one another, with the help of monovalent cations, to serve as building blocks in non-canonical structures of nucleic acids known as G-quadruplexes (G4s) (2). G4s were found to exist in important genomic regions such as the telomeres, gene promoters, replication initiation sites and 5'-UTRs, which carry out essential cellular processes (3,4). Besides the biological importance of naturally occurring G4s (4), synthetic G4s are of high interest due to their potential applications in therapeutics (5) and nanotechnology (6). Polymorphism is one of the distinct features of G4 structures, allowing their structural elements—such as strand orientations, loops, and glycosidic bonds—to adopt different conformations depending on the environmental conditions (7–26). Higher-order structures can also be formed by stacking and interlocking of multiple G4 units (27–35).

Bulges are important structural elements of G4s, which form as protrusions of bases from the G-tetrad core. Bulges occur due to discontinuities (presence of non-guanine bases) in one or more G-tracts of the G4 forming sequences. Nearly a third of the experimentally detected G4s in the human genome (over 200 000 out of 700 000) could contain bulges (36). Unlike loops which connect two corners of the G-tetrad core, bulges connect two neighbouring guanines of the same G-tract in the G-tetrad core (37–40). Allowing the presence of bulges in the genomic search increases the number of putative G4 sequence drastically, therefore expanding the library of putative G4-forming sequences (39). Bulges are also important in the context of duplexes, as they can be formed due to mismatches and thus get involved in various interactions with other nucleic acids and proteins (41,42) and be implicated in cellular processes (43–45).

Similar to the B-DNA form (right-handed) and Z-DNA form (left-handed) for a duplex DNA, the backbone progression of G4 DNA is also capable of adopting both right- and left-handed configurations (46). Right-handed G4 structures are known for decades and have been studied extensively (47,48). Left-handed G4s, on the other hand, were recently discovered and remain rather unexplored. Left-handed G4s are characterized by the anticlockwise rotation of the backbone in its helical progression. They have several unique and distinct structural features, such as: (i) ability to form stable four-layered and bi-layered dimeric structures, (ii) stacking of thymine loops with the outer tetrads (T-capping) and (iii) the local sugar orientation of each base is almost perpendicular to the overall backbone progression giving rise to the 'twisted' backbone (46,49). Note that these structures are formed by natural DNA, which are to be distinguished from the mirror-image G4 structures formed by enantiomeric L-DNA (50,51). However, the circular dichroism (CD) characteristics of left-handed G4s are noticeably similar to that of L-DNA due to their resemblance in base stacking orientation (50,52). Interestingly, the CD spectrum of a new DNA conformation formed by G<sub>4</sub>C<sub>2</sub> repeats associated with neurodegenerative diseases exhibits similar features, although the structure has

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

not been resolved yet (53). Characterization of intrinsic fluorescence of DNA secondary structures demonstrated enhanced fluorescence of a left-handed G4 (54). Several families of G4 ligands were reported to exhibit enantioselectivity towards right- and left-handed G4s (55,56), while another can induce right-handed G4 conformation from a sequence favouring left-handed G4 (57).

A study on the impact of sequences to the formation of left-handed G4s revealed that the 12-nt GTGGTGGTGG TG motif (previously named *Block2Δ* (49) and termed as *LHG4motif* in this paper) was a minimal sequence capable of forming a dimeric left-handed G4 on its own and convert certain adjacent G-rich sequences, such as (GGT)<sub>4</sub>—a sequence with four G<sub>2</sub> tracts forming a right-handed parallel G4 on its own—and several other derivatives to left-handed G4s (Supplementary Figure S1) (49). *LHG4motif* was also shown to convert an adjacent antiparallel G4 to a right-handed parallel G4 (Supplementary Figure S1) (49,58). Bioinformatic search revealed high abundance of *LHG4motif* in the human genome, with over 10 000 hits, exhibiting two-order of magnitude enrichment compared to random occurrence of a 12-nt sequence (49). In this study, to understand the bulge formation in left-handed G4s, *LHG4motif* was attached to G-rich sequences with the potential of forming one or multiple bulges. We showed that *LHG4motif* could convert several adjacent sequences to left-handed G4s, including those which were unstructured on their own. We determined the crystal structure of two left-handed G4s containing one and two bulges, respectively. We have also solved the NMR solution structure of a left-handed G4 containing three bulges. Structural analysis revealed differences between bulges in right- and left-handed G4s. The high abundance of *LHG4motif* in the genome and the feasibility of bulged left-handed G4s driven by this motif can increase the chance of left-handed G4 formation in biological systems. Broadening our understanding of left-handed G4s should be useful for studying their biological functions, as well as their applications in engineering DNA nano-structures.

## MATERIALS AND METHODS

### Sample preparation

Unlabelled DNA oligonucleotides were purchased from IDT with standard desalting purification in the scales ranging from 100 nmol to 1 μmol. Sample purity was measured with ESI-MS and was >99%. The site-specific labelled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA synthesizer using reagents from Glen Research and Cambridge Isotope Laboratories. They were purified following the protocol of Glen Research (<https://www.glenresearch.com/glen-pak-dna-purification-cartridge.html>). The purified DNA samples were further dialyzed against water, 25 mM KCl and water successively. The dialyzed oligonucleotides were frozen and lyophilized. The DNA samples (concentration, 0.1–1 mM) were dissolved in a buffer containing 70 mM KCl and 20 mM KPi (pH 7), 10% D<sub>2</sub>O, 20 μM DSS. The DNA concentration was measured with UV absorption and expressed in strand molarity using the nearest neighbour approximation for the 260 nm molar extinction coefficient

of the unfolded species. DNA samples were annealed by heating at 95°C for 5 mins followed by cooling down slowly to room temperature before subjected to any measurement.

### Circular dichroism

Circular dichroism (CD) spectra were recorded on a JASCO-815 spectropolarimeter using 1-cm path length quartz cuvettes at 25°C. Scans were performed from 220 to 320 nm wavelength with a scanning speed of 100 nm/min, 1-nm data pitch, 2-nm bandwidth and 2 s digital integration time (DIT). Ten accumulations were obtained for each measurement, the spectral contribution of the buffer was subtracted using baseline correction, and data were zero-corrected at 320 nm. DNA samples with concentrations of 3–5 μM were dissolved in a buffer containing 70 mM KCl, 20 mM KPi (pH 7), 10% D<sub>2</sub>O, 20 μM DSS. Molar ellipticity of CD spectra was calculated and reported using the DNA concentration derived from the sample absorbance at 260 nm wavelength and the sample extinction coefficient calculated at 260 nm using nearest neighbour approximation for the unfolded sequence.

### UV melting

UV melting experiments were conducted on a JASCO V-650 spectrophotometer. 3–5 μM DNA samples were taken in a cuvette of a pathlength of 1-cm. UV absorption was measured at 295 and 320 nm wavelength at every 0.5°C between 20 and 90°C. Both the heating and cooling were performed at a rate of 0.1°C/min to allow slow unfolding and folding of the DNA, minimizing hysteresis. The data collected at 295 nm were subtracted from those at 320 nm for background correction and further normalized. The melting temperatures ( $T_m$ ) were determined from the normalized melting curves where the DNA folded fraction was 50%.

### NMR spectroscopy

NMR experiments were performed on a Bruker spectrometer operating at 600 MHz at 25°C. 0.1–1.5 mM DNA samples dissolved in 70 mM KCl, 20 mM KPi (pH 7), 10% D<sub>2</sub>O, 20 μM DSS were used for NMR measurements. Assignments of the imino protons of guanine residues were obtained by <sup>15</sup>N-filtered experiments using 2% site-specific labelled samples. Assignments of guanine aromatic protons was obtained via long-range through-bond correlation between imino and aromatic protons. Assignments of other protons were determined based on through-bond (TOCSY/COSY) and through-space correlation experiments. Spectra analyses were performed using the Topspin 3.5 (Bruker) and SPARKY 3.1 (59) software.

### NMR structure calculation

*NOE distance restraints.* Inter-proton distances for *3xBulge-LHG4motif* were obtained from NOESY experiments performed in H<sub>2</sub>O (mixing time, 200 ms) and D<sub>2</sub>O (mixing times, 100, 250 and 300 ms). For non-exchangeable protons, the peaks were classified as strong, medium, medium-weak and weak corresponding to the distance

restraints of  $(2.7 \pm 0.8)$ ,  $(3.8 \pm 0.9)$ ,  $(4.6 \pm 1.2)$  and  $(5.5 \pm 1.7)$  Å, respectively. Distances from exchangeable protons were classified as strong, medium and weak corresponding to the distance restraints of  $(4.0 \pm 1.2)$ ,  $(4.8 \pm 1.4)$  and  $(5.5 \pm 1.7)$  Å, respectively. Distances involving thymine methyl protons (H7#) were altered to be directed towards the methyl carbon (C7) with a 0.5 Å looser upper and lower limits as compensation.

**Dihedral restraints.** Dihedral angle restraints were imposed to the dihedral angle formed by O4'–C1'–N9–C4 of guanine residues. *Anti*-guanine residues were restricted to an angle of  $(240 \pm 70)^\circ$  or  $(240 \pm 40)^\circ$  for the outer tetrad and inner tetrad guanines, respectively.

**Hydrogen-bond restraints.** Hoogsteen hydrogen bonds between guanines were restrained using H21–N7, N2–N7, H1–O6 and N1–O6 distances, which were set to  $(2.0 \pm 0.2)$ ,  $(2.9 \pm 0.3)$ ,  $(2.0 \pm 0.2)$  and  $(2.9 \pm 0.3)$  Å, respectively.

**Planarity restraints.** Planarity restraints were used for the G1•G4•G8•G12, G2•G6•G10•G14, G16•G19•G22•G25 and G18•G21•G24•G27 tetrads.

**Distance-geometry simulated annealing.** Initial extended conformation of *3xBulge-LHG4motif* sequence was generated using the XPLOR-NIH (60) program by supplying the available standard nucleic acid topology and parameter tables. Each system was then subjected to distance-geometry simulated annealing by incorporating distance, dihedral, hydrogen bond and planarity restraints. One hundred structures were generated and subjected to further refinement.

**Distance-restrained molecular dynamics refinement.** The 100 structures obtained from the simulated annealing step were refined with a distance-restrained molecular dynamics protocol incorporating all distance restraints. The system was heated from 300 to 1000 K in 14 ps and allowed to equilibrate for 6 ps, during which force constants for the distance restraints were kept at 2 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The force constants for non-exchangeable proton and exchangeable proton restraints were then increased to 16 and 8 kcal mol<sup>-1</sup> Å<sup>-2</sup>, respectively, in 20 ps before another equilibration at 1000 K for 50 ps. Next, the system was cooled down to 300 K in 42 ps, after which an equilibration was performed for 18 ps. The coordinates were saved every 0.5 ps during the last 10.0 ps and averaged. The average structure obtained was then subjected to minimization until the gradient of energy was less than 0.1 kcal mol<sup>-1</sup>. Dihedral (50 kcal mol<sup>-1</sup> rad<sup>-2</sup>) and planarity (1 kcal mol<sup>-1</sup> Å<sup>-2</sup> for tetrads) restraints were maintained throughout the course of refinement. Ten lowest-energy structures were generated.

### Crystallization

DNA samples of *1xBulge-LHG4motif-TT* and *2xBulge-LHG4motif-TT* were prepared in 100 mM potassium cacodylate buffer (pH 7) at a concentration of ~0.8 mM. The samples were annealed by heating at 95°C for 5 min followed by slowly cooling down to room temperature. Initial screening for crystallization conditions were done at 24°C

using Matrix 1 and 2 sets of reagents (Hampton Research) in a 96-well sitting drop vapor diffusion setup at 1:1 sample to reagent proportions with the help of mosquito<sup>®</sup> LCP (ttp:labtech). Both the sequences produced crystals under multiple conditions. Crystals were flash frozen in liquid nitrogen before data collection. For *1xBulge-LHG4motif-TT* the crystals grown in 0.08 M potassium chloride, 0.04 M sodium cacodylate trihydrate pH 6.0, 55% (v/v)–2-methyl-2,4-pentanediol and 0.012 M spermine tetrahydrochloride were seen to produce the highest resolution diffraction. For the *2xBulge-LHG4motif-TT*, the crystals obtained in 0.08 M potassium chloride, 0.04 M sodium cacodylate trihydrate pH 7.0, 60% (v/v)–2-methyl-2,4-pentanediol and 0.012 M spermine tetrahydrochloride were chosen for the same reason.

### X-ray diffraction data collection and refinement

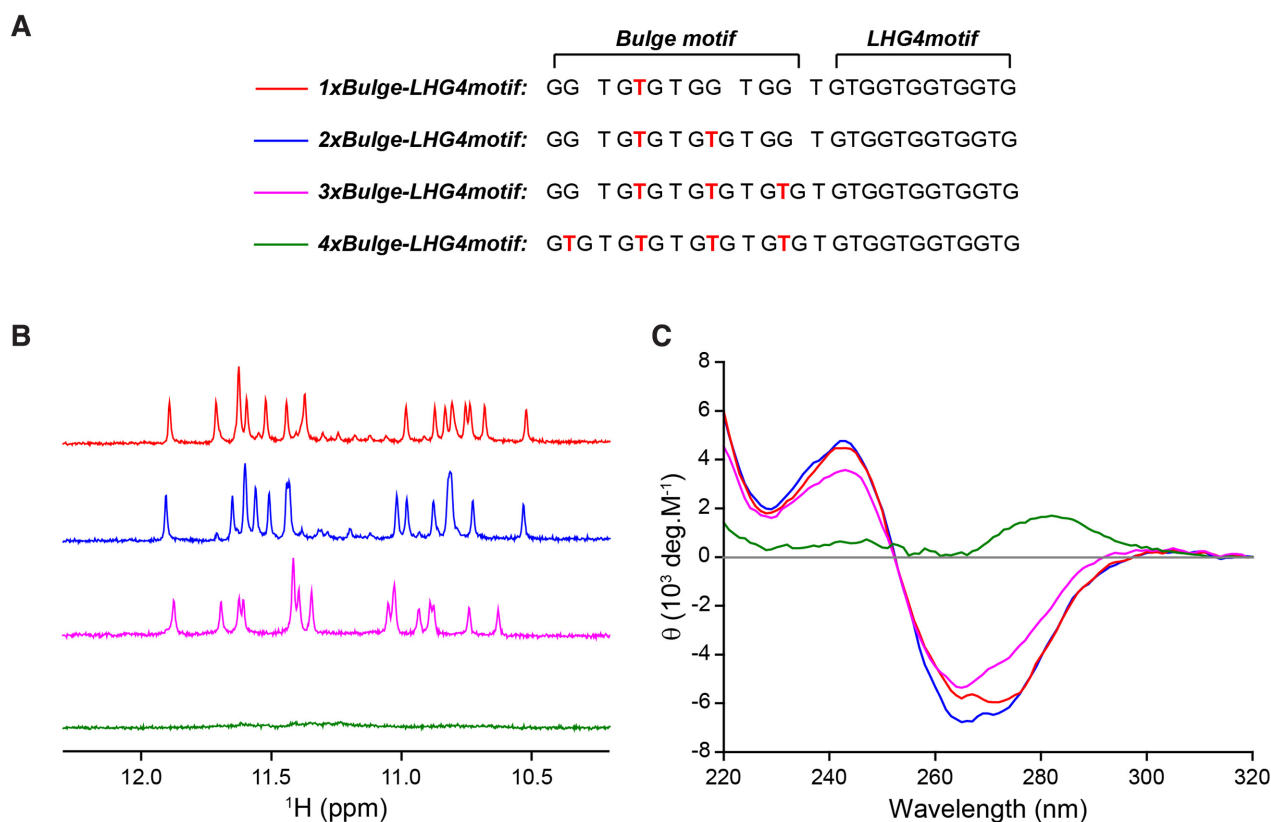
Crystal diffraction data were collected at PROXIMA 2 beamline for *1xBulge-LHG4motif-TT* and at PROXIMA 1 beamline for *2xBulge-LHG4motif-TT* of SOLEIL synchrotron, France. Native datasets were collected over 360° rotation ranges at 0.1° oscillation range. Data were processed in *P*<sub>2</sub><sub>1</sub> space group using the XDS software package (61). Molecular replacement was done using the *Z-G4* crystal structure (PDB ID: 4U5M) as a search model to obtain initial phases. One and two copies per unit cell were found respectively for *1xBulge-LHG4motif-TT* and *2xBulge-LHG4motif-TT*. The model was iteratively built through cycles of refinement using phenix (62,63) or Refmac5 (ccp4) and manual rebuilding in coot (64,65). Spermines were fitted into the structures using LigandFit function in Phenix package.

## RESULTS AND DISCUSSION

### *LHG4motif* converts various adjacent sequences to left-handed G4s

Previously, *LHG4motif* was shown to convert an adjacent sequence (TGGTGGTGGTGG) or (GGTGGTGGTGG) containing four G<sub>2</sub> tracts separated by single thymines to a left-handed G4 ((46,49) and data not shown). To explore the bulge formation in left-handed G4s, additional thymines were added in these G<sub>2</sub> tracts. Four adjacent sequences which potentially can form one or multiple bulges (Supplementary Table S1) were designed as potential bulge-forming motifs wherein thymines were inserted in the G-tracts as the following: (i) second G-tract (T5), named *1xBulge*; (ii) second and third G-tracts (T5, T9), named *2xBulge*; (iii) second, third and fourth G-tracts (T5, T9, T13), named *3xBulge* and (iv) first, second, third and fourth G-tract (T2, T6, T10, T14), named *4xBulge*. It is to be noted that this nomenclature does not indicate the corresponding bulge formation in these sequences individually. The first sequence on its own formed an unidentified structure, as indicated by six peaks observed in the imino proton region (10–12 ppm) of the one-dimensional (1D) <sup>1</sup>H NMR spectrum; all the other three sequences were essentially unstructured, as no peaks were observed in the imino proton region (Supplementary Figure S2A). The results were also supported by the CD spectroscopy data (Supplementary Figure S2B).





**Figure 1.** (A) Designed sequences attached with *LHG4motif* at their 3'-end. The potential thymine bulges are indicated in red. (B) 1D  $^1\text{H}$  NMR and (C) CD spectra of left-handed G4s potentially containing one (*1xBulge-LHG4motif*; red), two (*2xBulge-LHG4motif*; blue), three (*3xBulge-LHG4motif*; magenta) and four (*4xBulge-LHG4motif*; green) bulges. The distinct characteristics of left-handed G-quadruplexes can be observed in NMR spectra displaying sixteen peaks divided into two groups of eight peaks, and the molar ellipticity ( $\theta$ ) in the CD spectra possessing crest and trough at 240 and 270 nm, respectively. *4xBulge-LHG4motif* does not form any stable structure.

These four bulge-forming motifs were attached with *LHG4motif* at their 3'-end via a single thymine linker, resulting in sequences named as *1xBulge-LHG4motif*, *2xBulge-LHG4motif*, *3xBulge-LHG4motif* and *4xBulge-LHG4motif*, respectively (Figure 1A). Upon combining with *LHG4motif*, the first three sequences exhibited the NMR and CD spectral signatures of left-handed G4s: 1D  $^1\text{H}$  NMR spectra displayed sixteen major imino protons at 10.5–12.0 ppm regrouped into two regions with eight peaks each (Figure 1B); CD spectra showed positive and negative peaks at 240 and 265 nm, respectively (Figure 1C). The fourth sequence, *4xBulge-LHG4motif*, remained unstructured as indicated by NMR and CD spectra (Figure 1B, C). We also investigated the structural driving ability of *LHG4motif* when it is attached to the 5'-end of the designed sequences (Supplementary Table S2). The NMR and CD results suggested that two sequences, *LHG4motif-1xBulge* and *LHG4motif-3xBulge*, but not the sequence *LHG4motif-2xBulge*, were able to form left-handed G4 structures (Supplementary Figure S3). Thus, the structural driving ability of *LHG4motif* was context dependent and appeared to be stronger when attached to the 3'-end rather than the 5'-end of a G-rich sequence. Given the previous observation that a small disruption to the minimal left-handed motif (*LHG4motif*), such as a single thymine addition in a loop,

could abolish the left-handed G4 structure (49), we always kept the *LHG4motif* intact in this study.

### Crystal structure of a left-handed G4 with one bulge

X-ray crystallography was employed to investigate the structure of *1xBulge-LHG4motif*. The sequence *1xBulge-LHG4motif-TT* (Supplementary Table S3) with two additional thymines (TT) at the 3'-end was crystallized. Addition of 3'-end thymines (TT) has improved the crystal packing to provide high-quality X-ray diffraction data. The NMR and CD spectra of *1xBulge-LHG4motif* and *1xBulge-LHG4motif-TT* displayed highly similar spectral characteristics (Supplementary Figure S4), indicating the formation of the same structural conformation by these two sequences. *1xBulge-LHG4motif-TT* was found to crystallize in the  $P2_1$  space group with high-resolution diffraction to 1.18 Å (Table 1). The model fits the electron density excellently as shown in Figure 2A. The crystal structure of *1xBulge-LHG4motif-TT* features a four-layered unimolecular G4 structure separated into two blocks connected by a thymine linker. Each block has parallel left-handed backbone progression similar to the previously reported left-handed G4 structure of *Z-G4* (46). The 5'-5' stacking interface between guanine bases in successive blocks is in

**Table 1.** Data collection and refinement statistics for X-ray crystal structures of *1xBulge-LHG4motif-TT* and *2xBulge-LHG4motif-TT*

	<i>1xBulge-LHG4motif-TT</i>	<i>2xBulge-LHG4motif-TT</i>
<b>Data collection statistics</b>		
Wavelength (Å)	0.9793	0.9793
Space group	$P2_1$	$P2_1$
Cell dimensions (Å/°)	29.2, 40.0, 30.4, 90.0, 115.9, 90.0	29.1, 39.8, 55.0, 90.0, 91.3, 90.0
Resolution range (Å)	26.29–1.182 (1.224–1.182)	29.09–1.296 (1.342–1.296)
Total reflections	139 506 (12 319)	187 246 (14 484)
Unique reflections	20 747 (2060)	31 051 (2861)
Multiplicity	6.7 (6.0)	6.0 (4.9)
Completeness (%)	99.70 (99.95)	94.67 (90.68)
Mean I/σ(I)	9.00 (1.64)	6.86 (1.07)
Wilson <i>B</i> -factor (Å <sup>2</sup> )	8.52	11.81
<i>R</i> -merge	0.125 (0.938)	0.164 (1.203)
<i>R</i> -meas	0.135 (1.028)	0.180 (1.345)
<i>R</i> -pim	0.052 (0.415)	0.073 (0.589)
CC1/2	0.996 (0.5)	0.996 (0.732)
<b>Refinement statistics</b>		
Resolution range (Å)	26.29–1.182 (1.224–1.182)	29.09–1.296 (1.342–1.296)
Reflections used in refinement	20 686 (2059)	29 721 (2842)
Reflections used for <i>R</i> -free	920 (97)	1433 (135)
<i>R</i> -work	0.124 (0.178)	0.184 (0.320)
<i>R</i> -free	0.166 (0.204)	0.214 (0.354)
<b>Number of atoms</b>		
DNA	569	1178
K <sup>+</sup>	3	7
Na <sup>+</sup>	0	1
Spermine	14	28
Water	158	208
<b>Average <i>B</i>-factor of (Å<sup>2</sup>)</b>		
DNA	10.5	14.5
K <sup>+</sup>	6.1	13.9
Na <sup>+</sup>	-	12.9
Spermine	22.7	38.2
Water	33.3	26.8
<b>Rmsd</b>		
Bond angles, °	2.25	1.04
Bond length, Å	0.018	0.007

Statistics for the highest-resolution shell are shown in parentheses.

the opposite-polarity mode with partial 5/6-membered ring overlap. The thymines capping of the outer tetrads of the G4 structure (T-capping)—a distinctive feature of left-handed G4s—are also conserved in this structure (46). The upper block has three T-capping residues, while the lower block has four. The thymine residue T5 forming a bulge between the G4 and G6 stacked guanines, projecting out of the G-tetrad core, yielding the first left-handed G4 structure accommodating a bulge (Figure 2B–D).

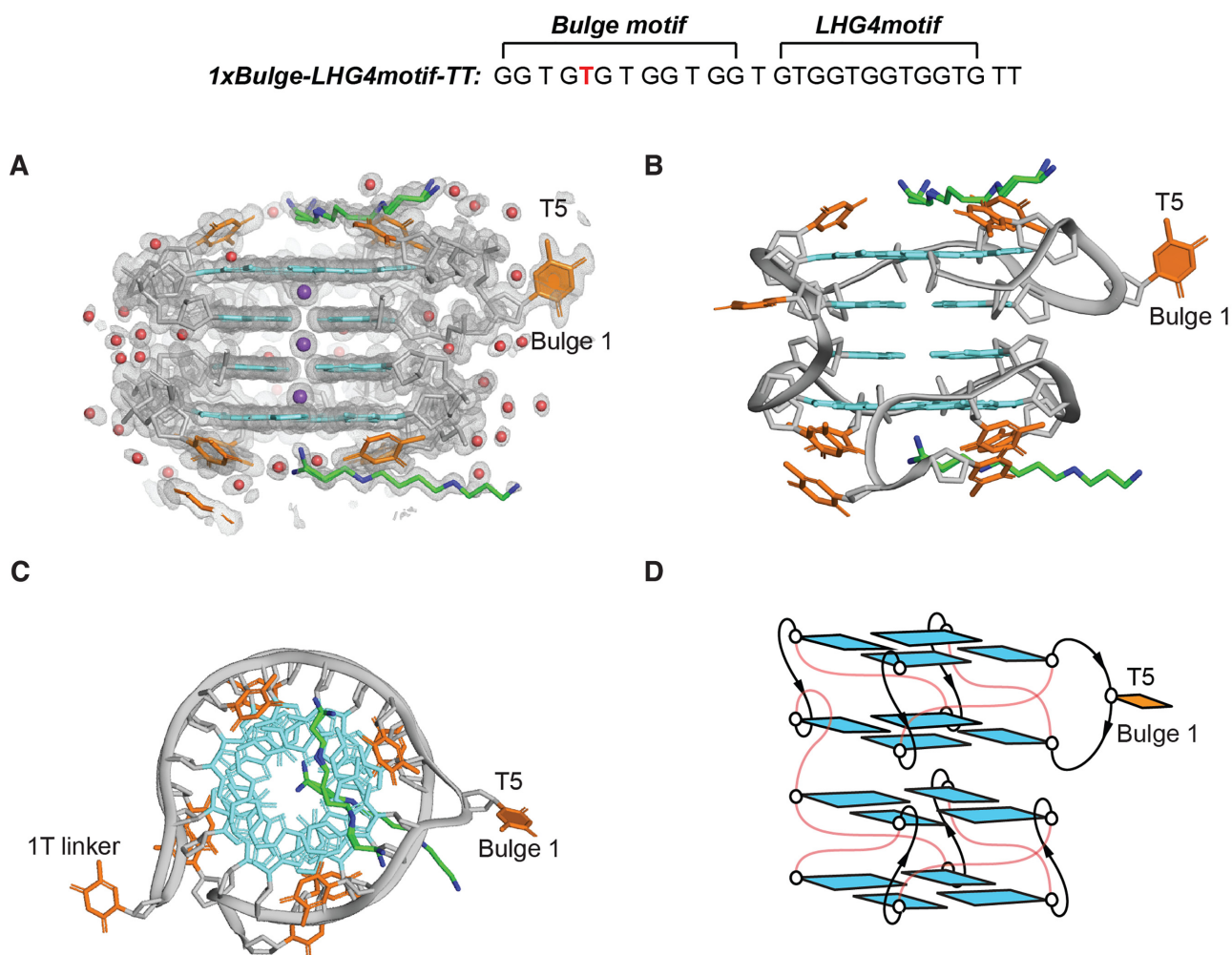
The positions of the potassium ions in between the G-tetrads are consistent with those of *Z-G4* and other reported left-handed G4 structures (46,49). The overhang additional 3'-thymines (TT) stack on the capping thymines (Supplementary Figure S5A), while a spermine molecule (from the crystallization solution) stacks over the outer tetrad in each block squeezing itself between the capping thymines (Figure 2). The accommodation of spermine stacking over the tetrads despite the presence of capping thymines establishes the possibility of ligands binding to left-handed G4s.

### Crystal structure of a left-handed G4 with two bulges

The sequence *2xBulge-LHG4motif-TT* with two thymines at the 3'-end was crystallized in the  $P2_1$  space group with two molecules in the asymmetric unit (Supplementary Fig-

ure S6). The two molecules were related by a translation. Hence, the crystals of *2xBulge-LHG4motif-TT* and *1xBulge-LHG4motif-TT* were packed in a similar fashion. The overall structures are noticeably close to each other with an rmsd of 0.112 Å for all non-H superimposed atoms excluding the bulges. Both crystal structures feature four-layered G4s separated into two blocks connected by one thymine linker. In case of *2xBulge-LHG4motif-TT*, two thymine residues T5 and T9 are projected outward between guanines G4 and G6 as well as G8 and G10, respectively, in the upper block (Figure 3). The overhang 3'-thymines (TT) stack on the capping thymines (Supplementary Figure S5B). A spermine molecule (from the crystallization solution) stacks over the outer tetrad in each block.

The crystal structure of *2xBulge-LHG4motif-TT* also displays three potassium ions coordinated in the central channel of the G4. The two molecules (A and B) in the asymmetric unit are closely similar with an rmsd of 0.132 Å for all non-H superimposed atoms. The molecule B was observed to contain one additional potassium ion in close proximity to thymine T22 and one of the spermine molecule (Supplementary Figure S6). In addition, there is a sodium ion nearby the second bulge T9 in molecule B that interacts with O2 and O4' of the bulge residue, potentially stabilizing the



**Figure 2.** Crystal structure of *1xBulge-LHG4motif-TT*. (A) Electron density  $2mF_o - DFC$  map of *1xBulge-LHG4motif-TT*. (B) Cartoon representation of the crystal structure of *1xBulge-LHG4motif-TT* with the guanine bases coloured in cyan and the thymine bases in orange. Phosphate backbone is shown in gray. Water molecules, potassium ions, spermine carbon and nitrogen atoms are shown in red, purple, green and blue, respectively. (C) Top view of the structure showing the T5 bulge in the upper block. (D) Schematics of the structure of *1xBulge-LHG4motif-TT* with thymine loops and linkers in red. The single T5 bulge is indicated and highlighted with a filled base in all the panels.

structure. The sodium ion was modelled to best fit the map, while other molecules such as water or potassium generated significant difference peaks in the  $F_o - F_c$  map. The data collection and refinement statistics of the crystal structure *2xBulge-LHG4motif-TT* are summarised in Table 1.

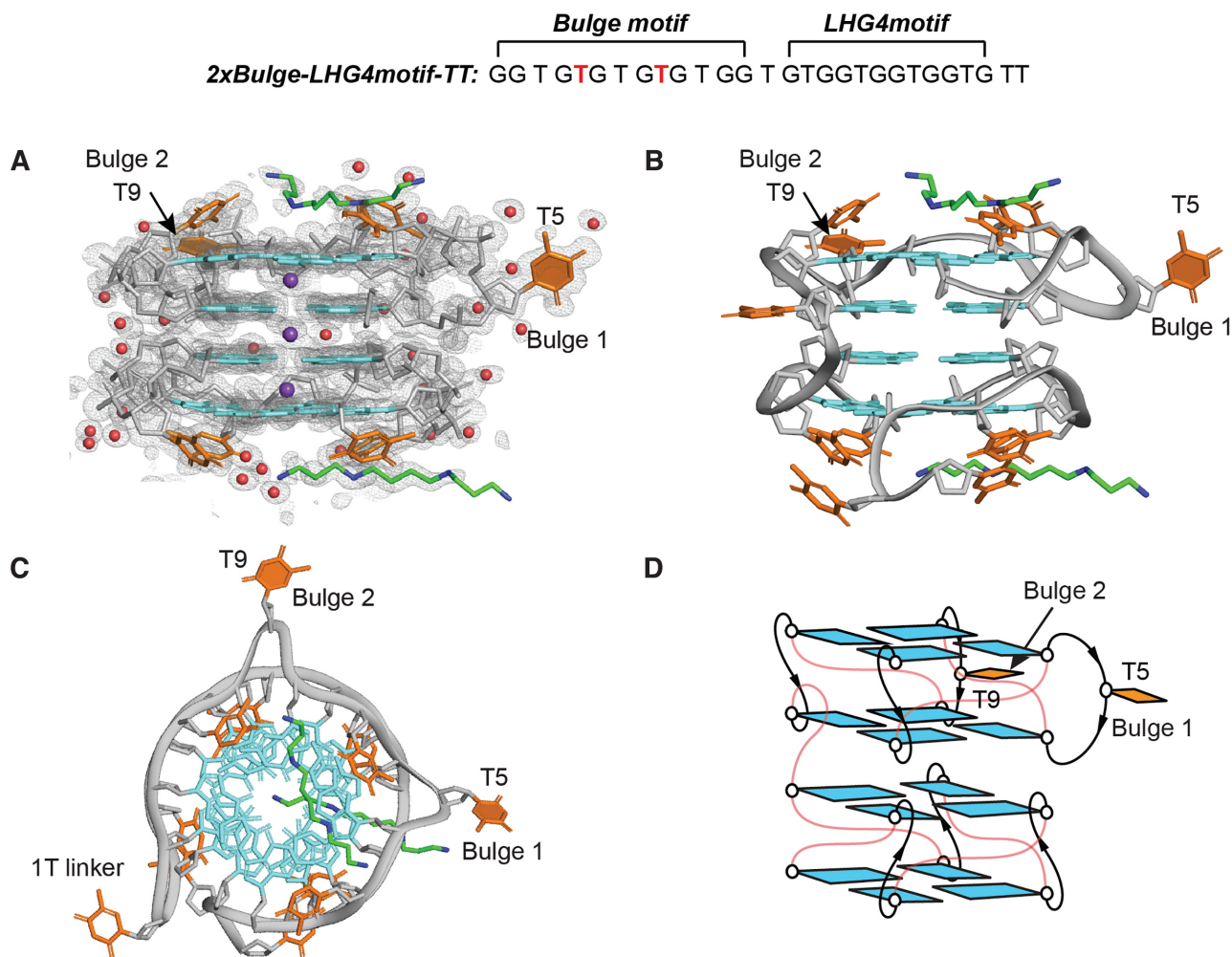
### NMR solution structure of a left-handed G4 with three bulges

As we were unable to obtain good crystals for *3xBulge-LHG4motif* and its derivatives, we proceeded to further characterize the structure of the *3xBulge-LHG4motif* sequence by NMR spectroscopy. The observation of sixteen imino proton peaks supported the formation of a four-layered left-handed G4 by this sequence. These sixteen guanine imino protons were assigned unambiguously by the site-specific 2%  $^{15}\text{N}$  labeling method (Supplementary Figure S7) (66). Guanine aromatic protons (H8) were assigned via through-bond correlations to imino (H1) protons of the same guanines in a JR-HMBC experiment (Supplementary Figure S8) (66).

To affirm the positions of guanines in the inner and outer tetrads,  $\text{D}_2\text{O}$  solvent exchange experiment was performed. The imino protons of G1, G4, G8, G12, G18, G21, G24, G27 constituting one group of eight peaks in the down-field chemical shift region disappeared immediately due to exchange with  $\text{D}_2\text{O}$  solvent, confirming their involvements in the formation of the outer tetrads. The remaining eight peaks of G2, G6, G10, G14, G16, G19, G22, G25 belonging to the other group in the upfield region remained after dissolving the sample in  $\text{D}_2\text{O}$  (Supplementary Figure S9), consistent with their positions in the inner tetrads, well protected from exchange with solvent. The  $\text{D}_2\text{O}$  exchange analysis also verified the formation of a four-layered G4 by *3xBulge-LHG4motif*.

To determine the folding topology of the *3xBulge-LHG4motif*, NOESY experiment (90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  solvent) was performed at a mixing time of 200 ms. Cyclic H1-H8 NOE connectivities between neighbouring guanines in the same tetrads gave rise to the following four-layered topology: G1•G4•G8•G12 (outer), G2•G6•G10•G14 (in-





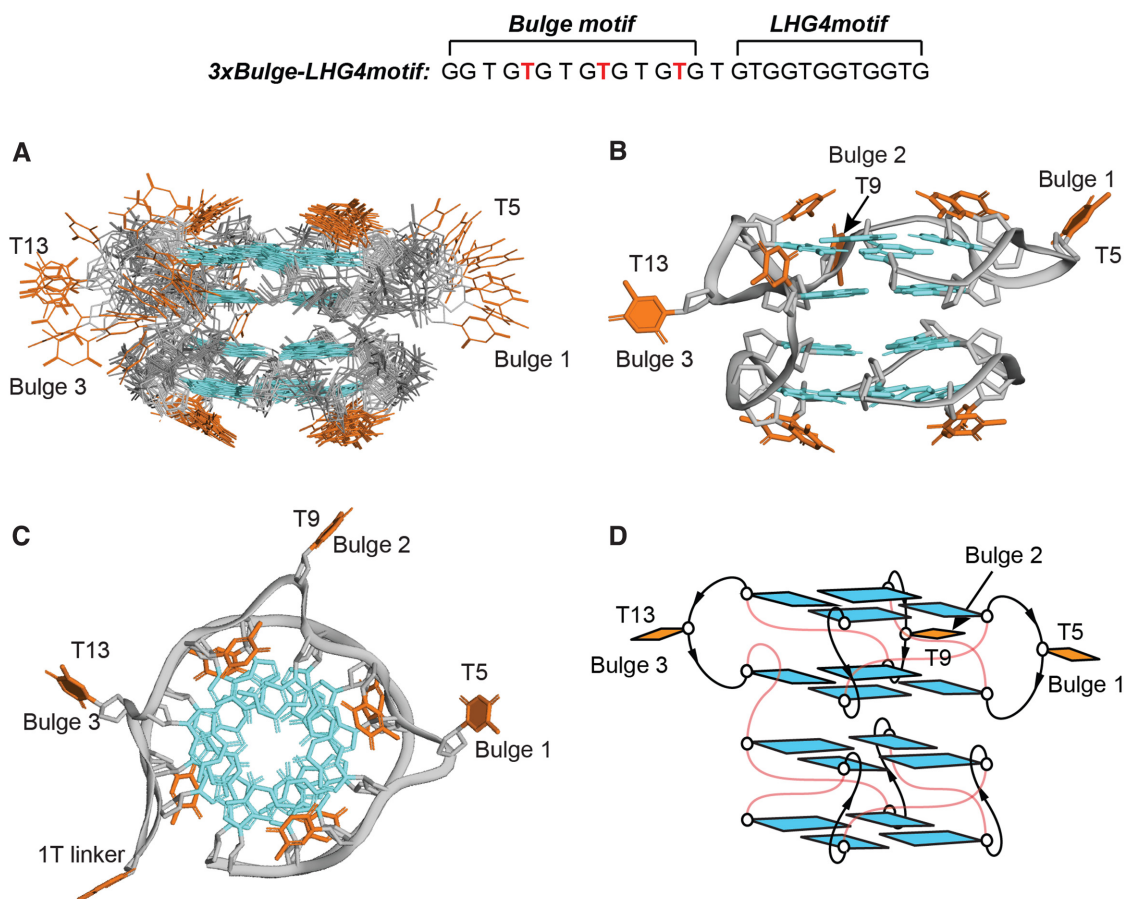
**Figure 3.** Crystal structure of *2xBulge-LHG4motif-TT*. (A) Electron density map of *2xBulge-LHG4motif-TT*. (B) Cartoon representation of the crystal structure of *2xBulge-LHG4motif-TT* with the guanine bases coloured in cyan and the thymine bases in orange. Phosphate backbone is shown in gray. Water molecules, potassium ions, sodium ions, spermine carbon and nitrogen atoms are shown in red, purple, yellow, green and blue, respectively. (C) Top view of the structure showing the bulges – T5 (Bulge 1) and T9 (Bulge 2) – in the upper block. (D) Schematics of the structure of *2xBulge-LHG4motif-TT* with thymine loops and linkers in red. The two bulges (T5, T9) are indicated and highlighted with filled bases.

ner), G16●G19●G22●G25 (inner) and G18●G21●G24●G27 (outer) (Supplementary Figure S10). The glycosidic conformations of all the participating guanine residues were found to be *anti*, deduced from the intensities of intra-residue H8-H1' cross-peaks in a NOESY spectrum performed in D<sub>2</sub>O at a mixing time of 100 ms, which showed similar moderate intensity for all residues. NOE sequential walk between H8<sub>(n)</sub>-H1'<sub>(n)</sub>-H8<sub>(n+1/n+2)</sub> observed in a NOESY spectrum performed in D<sub>2</sub>O at a mixing time of 300 ms (Supplementary Figure S11) revealed the stacking pattern of the G-tetrads, as well as the thymine loop and bulge arrangements (see caption of Figure S11). The structural schematics of *3xBulge-LHG4motif* is proposed in Figure 4D.

Finally, the NMR solution structure of *3xBulge-LHG4motif* was computed based on NMR constraints obtained from analyses of the NOESY spectra. The structure calculation provided 100 structures based on the distance, angle, hydrogen-bond and planarity constraints, out of which 10 lowest-energy structures were superimposed and presented in Figure 4 with the statistics of the computed structure shown in Table 2. The solution

structure of *3xBulge-LHG4motif* revealed a very similar structural characteristics as *1xBulge-LHG4motif* and *2xBulge-LHG4motif* with a four-layered G4 comprising of two blocks connected by a thymine linker. Both the blocks, the lower *LHG4motif* block and the upper block with three bulges, feature parallel left-handed G4 backbone progression. The three bulges T5, T9 and T13 are projecting out between the two stacked guanines in the upper block, while the thymine loops capping at the two ends distinctive to left-handed G4 structures are maintained (Supplementary Figure S12). The three bulges (T5, T9 and T13) and the thymine linker (T15) are not well defined due to lack of constraints. These residues could also be dynamic. While the glycosidic conformations of these residues were not imposed in the structure calculation, diverse conformations were observed among the ensemble of 10 lowest-energy structures, with the majority adopting an *anti*-conformation and a few in *syn*-conformation. The *LHG4motif* block adopts the same structure as reported previously including the thymine loops capping the bottom tetrad.





**Figure 4.** NMR solution structure of *3xBulge-LHG4motif*. (A) Ten superimposed lowest-energy structures. (B) Cartoon view of a representative structure. Guanine and thymine bases are shown in cyan and orange, respectively. Phosphate backbone is shown in grey. (C) Top view of the structure showing the bulges –T5 (Bulge 1), T9 (Bulge 2) and T13 (Bulge 3) – in the upper block. (D) Schematics of the structure of *3xBulge-LHG4motif* with thymine loops and linkers in red. The three bulges (T5, T9, T13) are indicated and highlighted with filled bases.

### The destabilizing effect of additional bulges: left-handed *versus* right-handed G4s

Previously, simultaneous accommodation of up to three bulges in different positions of a right-handed G4 was demonstrated (39). In that study, G4 structures with more than three bulges could not be formed, coincident with our current study where *4xBulge-LHG4motif* was unable to form any stable structure. The comparison between the costs in the thermal stability upon increasing the number of bulges is summarized (Supplementary Table S4) (39). In the right-handed G4 system, on average, the melting temperature decreased by  $\sim 17^{\circ}\text{C}$  upon addition of a second bulge in a different G-tract with respect to the first bulge (39). It further decreased by  $\sim 21^{\circ}\text{C}$  due to the addition of a third bulge in a different G-tract (39). Comparatively, in the current left-handed G4 system, the melting temperature decreased by  $\sim 16^{\circ}\text{C}$  when the second bulge (T9) was introduced in *2xBulge-LHG4motif* in addition to the first one (T5) in *1xBulge-LHG4motif*. The stability was further decreased by  $\sim 12^{\circ}\text{C}$  when the third bulge was added in *3xBulge-LHG4motif* (Supplementary Figure S13). Taken together, we found a diminished destabilizing effect of additional bulges in left-handed G4s compared to that of right-handed G4s. The comparison with no-bulge left-handed G4

structures was not attempted due to the presence of more than one conformation in the no-bulge construct.

### Effect of bulges in altering the backbone dihedral angles in the G-tetrad core: left-handed *versus* right-handed G4s

The dihedral angles of left- and right-handed G4 backbones were previously compared, showing significant differences between them (46). Here, we compare the same set of stepwise dihedral angles alterations ( $\epsilon$ ,  $\zeta$ ,  $\alpha+1$ ,  $\beta+1$  and  $\gamma+1$ ) upon addition of bulges. Specifically, we quantified the dihedral angle adjustments on the core guanine residues on the immediate 5'-side of the bulge (termed 'previous residue') as well as the core guanine residue on the immediate 3'-side of the bulge (termed 'next residue') in the systems of left- and right-handed G4s. The dihedral angle alterations are compared with the standard set of dihedral angle values of parallel G4s without bulges obtained from selected crystal structures deposited in PDB for both left- (4U5M, 6GZ6) (46,49) and right-handed systems (1KF1, 244D, 352D) (8,67,68).

The comparison results are summarized in Figure 5. The standard dihedral angle values are represented in black open circle, solid red symbols represent the dihedral angle

**Table 2.** Statistics of the computed solution structure of *3xBulge-LHG4motif*

NMR restraints		
	Exchangeable	Non-exchangeable
Distance restraints		
Intra-residue	0	365
Inter-residue	66	163
Other restraints		
Hydrogen bond		64
Dihedral angle		16
Structure statistics		
NOE violations		
Number (>2.0 Å)	0.000 ± 0.000	
Deviations from the ideal geometry		
Bond lengths (Å)	0.003 ± 0.000	
Bond angles (°)	0.684 ± 0.006	
Improper (°)	0.351 ± 0.004	
Pairwise heavy atom RMSD value (Å)		
G-tetrad core	0.636 ± 0.085	
All heavy atoms	2.023 ± 0.357	

values from the bulged systems that were significantly different from the reference values ( $>60^\circ$ ), whereas open green symbols indicate the dihedral angle values close to the standard values. The introduction of a bulge in a left-handed G4 altered the transition angles  $\alpha+1$  and  $\beta+1$  for both the ‘previous residue’ and ‘next residue’ by  $\sim 90^\circ$  and  $\sim 70^\circ$  respectively, with all other angles largely unaffected. On the other hand, the bulge in a right-handed G4 affected the angles  $\varepsilon$  and  $\gamma+1$  for the ‘previous residue’ by  $\sim 90^\circ$  and  $\sim 135^\circ$  respectively, as well as the angles  $\zeta$ ,  $\alpha+1$  and  $\gamma+1$  for the ‘next residue’ by  $\sim 180^\circ$ ,  $\sim 135^\circ$  and  $\sim 120^\circ$  respectively. We conclude that the bulges in left-handed G4s are less disruptive to the neighbouring residues in terms of dihedral angle values than their right-handed counterparts. The data of bulged left-handed G4s were taken from the crystal structures of *1xBulge-LHG4motif-TT* and *2xBulge-LHG4motif-TT* in this paper, while the bulged right-handed G4 data were taken from the deposited crystal structure 5UA3 (40).

### Effect of bulges in altering the sugar puckering in the G-tetrad core: left-handed versus right-handed G4s

The presence of a bulge also affects the sugar puckering of the core guanine residues immediately adjacent to the bulge. Here, we quantified the alteration effect of the sugar pucker of the ‘previous residue’ and ‘next residue’ by the standard pseudorotation phase angle parameter ( $P$ ) (Figure 6A) (69). Standard values of  $P$  were measured using the same set of crystal structures (PDB ID: 4U5M, 6GZ6, 1KF1, 244D and 352D) (8,46,49,67,68) for both left- and right-handed G4 systems. The  $P$  values were obtained from *1xBulge-LHG4motif-TT* and *2xBulge-LHG4motif-TT* structures for left-handed G4s and the crystal structure 5UA3 (40) for a right-handed G4. The summary of the comparison is presented in Figure 6B.

In non-bulged left-handed G4 systems, all the core guanine residues adopt  $C2'$ -endo conformations ( $135^\circ < P < 180^\circ$ ). Upon introduction of a bulge, both the ‘previous residue’ and ‘next residue’ adjacent to the bulge do not

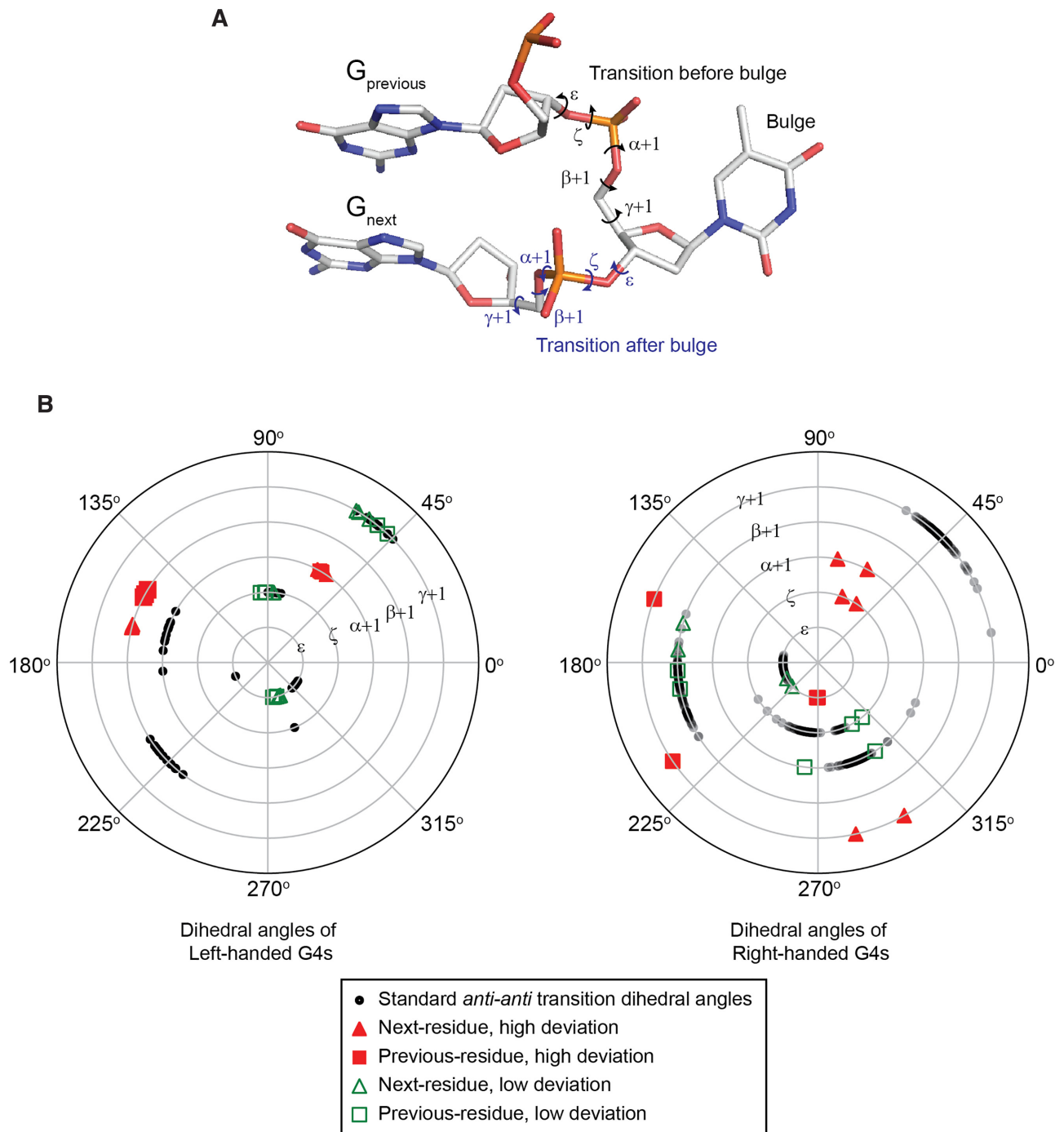
undergo changes in sugar pucker conformation, staying in  $C2'$ -endo conformations. Comparatively, in non-bulged right-handed G4 systems, the majority of the core guanine residues similarly adopt  $C2'$ -endo conformations. However, in the presence of a bulge, the ‘next residue’ sugar pucker conformation was altered to  $C3'$ -endo, with the  $P$  values around  $0^\circ$ . This analysis showed that the presence of a bulge in left-handed G4s requires less alteration in terms of sugar puckering compared to a bulge in right-handed G4s.

To complement the pseudorotation phase angle which can be difficult to visualize, we introduced the term  $z$ -deviation as an alternative parameter which is defined as the perpendicular distance between the plane containing the atoms  $C4'$ ,  $O4'$  and  $C1'$  to either  $C2'$  or  $C3'$  (Supplementary Figure S14A). Analogous comparisons between non-bulged and bulged left- and right-handed G4s are presented (Supplementary Figure S14B). Like for the  $P$  values, the only significant change was observed for the  $C2'$   $z$ -deviation values of the ‘next residue’ in right-handed bulged G4 which was negative instead of the positive values for the standard G4s. As expected, both the  $C3'$  and  $C2'$   $z$ -deviation values of bulged left-handed G4 structures are unchanged from the standard left-handed G4s.

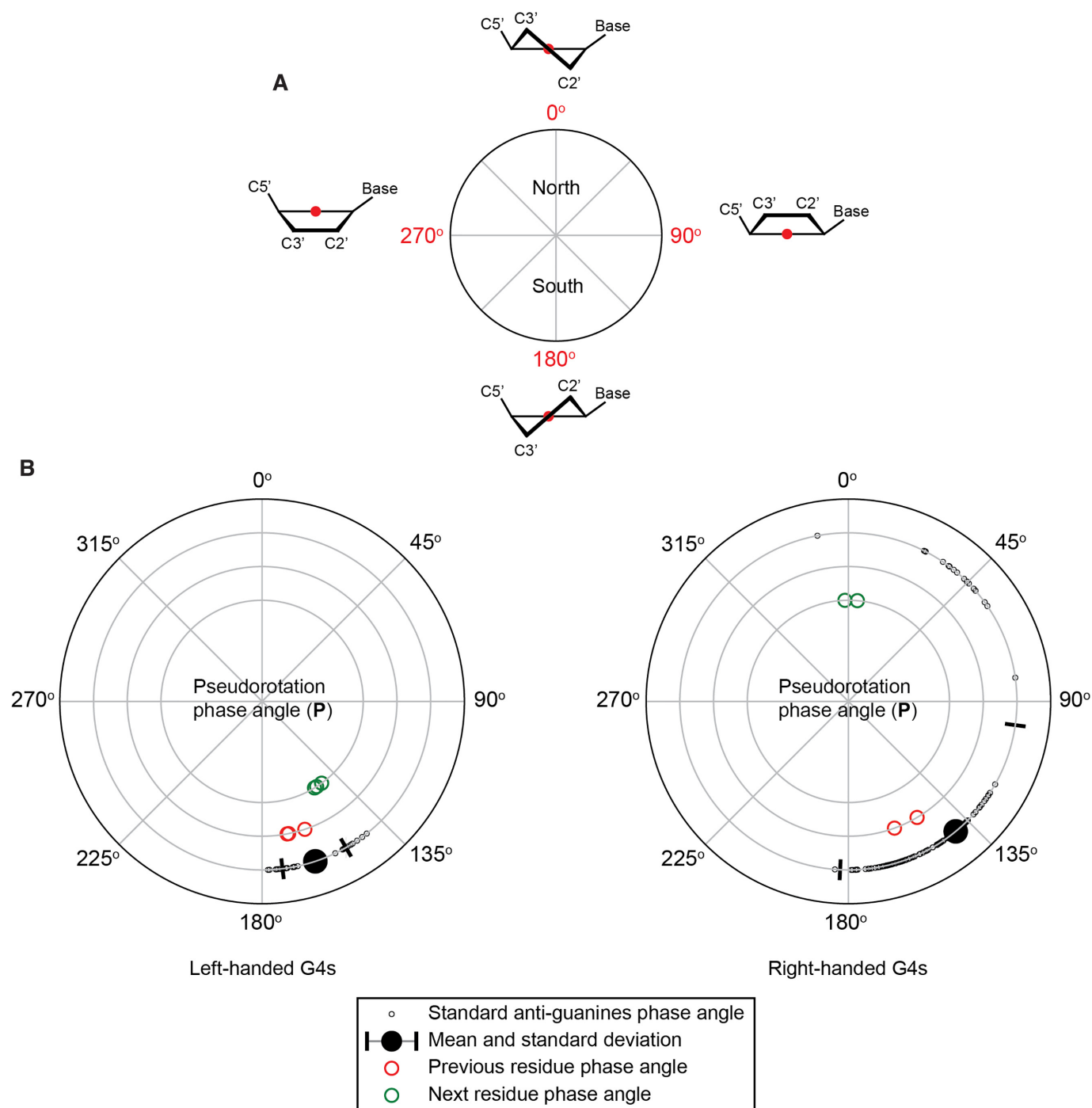
### Bulges and loops in left- and right-handed G4s

Outside of the G-tetrad core formed by four G-columns (or G-tracts), the two significant structural elements of a G4 structure are the loops and bulges. By definition, a loop is a nucleotide chain ranging from zero (only a single phosphate group) to several bases that connects two separate G-columns, while a bulge is a nucleotide chain of at least one base between two stacking guanines in the same G-column. In the context of sequences with alternating guanines and thymines, the concepts of loops and bulges might be undistinguishable based on the sequences alone, whereby the structures are required to recognize them. To do that, the stacking guanines are to be identified first. For example, in left-handed G4s the stacking guanines are defined to have a left-handed helical progression with a base rotation magnitude of  $\sim 27^\circ$  (46). In this definition, bulges and loops of a G4 structure can be readily distinguished.

This study revealed the detailed similarities and differences between the conformations of single-nucleotide bulges in left- and right-handed parallel G4s. Among the four structural elements (single-nucleotide bulges and loops in left- and right-handed parallel G4s), the loop in left-handed G4 has the most distinct feature allowing to pack a tight T-cap in an inward direction covering the ends of the G-tetrad core, whilst all the other three elements have outward projections (Supplementary Figure S15). Another distinguishing feature is the local sugar orientation (from  $5'$  to  $3'$ ) on each of the elements. Using the crystal structures of a bulged parallel left-handed G4 (*1xBulge-LHG4motif-TT*) and a bulged parallel right-handed G4 (PDB ID: 5UA3), we can see that: the local sugar orientations of the left-handed loop and the right-handed bulge are parallel to (same direction of) those of the G-tetrad guanines; while those of the left-handed bulge and the right-handed loop are anti-



**Figure 5.** Comparison of dihedral angles between bulged and non-bulged G4s in left- and right-handed systems. (A) Schematics of the transitional dihedral angles before ('previous residue') and after ('next residue') the bulge. (B) Comparison diagrams of the left- and right-handed G4 dihedral angle values. Black circles indicate standard values (PDB ID: 4U5M, 6GZ6, 1KF1, 244D and 352D) (8,46,49,67,68); solid red triangles and squares indicate high-deviation bulged values; open green triangles and squares indicate low-deviation bulged values (structures of *1xBulge-LHG4motif-TT*, *2xBulge-LHG4motif-TT* and PDB ID: SUA3) (40).



**Figure 6.** Comparison of pseudorotation phase angles between bulged and non-bulged G4s in left- and right-handed systems. (A) Sugar pucker system in terms of pseudorotation phase angle ( $P$ ) (69). (B) Comparison diagrams of the left- and right-handed G4 DNA sugar pucker in terms of  $P$  values. Black circles indicate standard values (PDB ID: 4U5M, 6GZ6, 1KF1, 244D and 352D) (8,46,49,67,68), with the mean and standard deviations denoted; red and green circles indicate the  $P$  values of the ‘previous residue’ and ‘next residue’ before and after the bulge, respectively (structures of *1xBulge-LHG4motif-TT*, *2xBulge-LHG4motif-TT* and PDB ID: 5UA3) (40).



parallel to (opposite direction of) those of the G-tetrad guanines (Supplementary Figure S15).

## CONCLUSION

Here we report on the ability of the 12-nt sequence GTG-GTGGTGGTG called *LHG4motif* to drive several DNA sequences towards left-handed G4s with bulges. We present the first crystal and NMR solution structures of left-handed G4s containing one, two and three bulges. The bulged residues were found to be pointing outwards of the G-tetrad core in all the structures. Thermal stability and geometrical analyses showed that the presence of bulges in left-handed G4s is associated with less disturbing and penalizing effect when compared to the right-handed counterparts.

## DATA AVAILABILITY

The coordinates for the X-ray crystal structures of two left-handed G4s containing one and two bulges and the NMR solution structures of a left-handed G4 containing three bulges have been deposited in the Protein Data Bank (PDB codes: 7D5D, 7D5E and 7D5F).

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank the staff of PROXIMA 1 and PROXIMA 2 beamline at SOLEIL for their assistance in data collection.

## FUNDING

Singapore National Research Foundation Investigatorship [NRF-NRFI2017-09]; Singapore Ministry of Education Academic Research Fund Tier 2 [MOE2015-T2-1092]; Nanyang Technological University (NTU Singapore) grants (to A.T.P.). Funding for open access charge: Nanyang Technological University.

*Conflict of interest statement.* None declared.

## REFERENCES

- 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.
- Maizels, N. and Gray, L.T. (2013) The G4 genome. *PLoS Genet.*, **9**, e1003468.
- Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
- Bates, P.J., Laber, D.A., Miller, D.M., Thomas, S.D. and Trent, J.O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp. Mol. Pathol.*, **86**, 151–164.
- Mergny, J.L. and Sen, D. (2019) DNA quadruple helices in nanotechnology. *Chem. Rev.*, **119**, 6290–6325.
- Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure*, **1**, 263–282.
- Haider, S., Parkinson, G.N. and Neidle, S. (2002) Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.*, **320**, 189–200.
- Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
- Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K<sup>+</sup> solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
- Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N. and Phan, A.T. (2009) Structure of the human telomere in K<sup>+</sup> solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.
- Heddi, B. and Phan, A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
- 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.
- 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.
- 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.
- 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.
- Smith, F.W. and Feigon, J. (1992) Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
- Crungej, M., Sket, P. and Plavec, J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.
- Webba da Silva, M., Trajkovski, M., Sannohe, Y., Ma'ani Hessari, N., Sugiyama, H. and Plavec, J. (2009) Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem. Int. Ed. Engl.*, **48**, 9167–9170.
- Marusic, M., Sket, P., Bauer, L., Viglasky, V. and Plavec, J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.*, **40**, 6946–6956.
- 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.
- Maity, A., Winnerdy, F.R., Chang, W.D., Chen, G. and Phan, A.T. (2020) Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.*, **48**, 3315–3327.
- 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.
- Marquevielle, J., Robert, C., Lagrabette, 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.*, **48**, 9336–9345.
- Zhang, Y., El Omari, K., Duman, R., Liu, S., Haider, S., Wagner, A., Parkinson, G.N. and Wei, D. (2020) Native de novo structural determinations of non-canonical nucleic acid motifs by X-ray crystallography at long wavelengths. *Nucleic Acids Res.*, **48**, 9886–9898.
- Haase, L. and Weisz, K. (2020) Locked nucleic acid building blocks as versatile tools for advanced G-quadruplex design. *Nucleic Acids Res.*, **48**, 10555–10566.
- Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E. and Patel, D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A x (G x G x G x G) pentad-containing dimeric DNA quadruplex involving stacked

- G(anti) x G(anti) x G(syn) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
28. Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E.S., Bowers, M.T., De Pauw, E. and Gabelica, V. (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, **130**, 10208–10216.
  29. Kuryavyi, V., Phan, A.T. and Patel, D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
  30. Sket, P. and Plavec, J. (2010) Tetramolecular DNA quadruplexes in solution: insights into structural diversity and cation movement. *J. Am. Chem. Soc.*, **132**, 12724–12732.
  31. 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.
  32. Phan, A.T., Kuryavyi, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 634–639.
  33. Marsh, T.C. and Henderson, E. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, **33**, 10718–10724.
  34. Miyoshi, D., Karimata, H., Wang, Z.M., Koumoto, K. and Sugimoto, N. (2007) Artificial G-wire switch with 2,2'-bipyridine units responsive to divalent metal ions. *J. Am. Chem. Soc.*, **129**, 5919–5925.
  35. Bose, K., Lech, C.J., Heddi, B. and Phan, A.T. (2018) High-resolution AFM structure of DNA G-wires in aqueous solution. *Nat. Commun.*, **9**, 1959.
  36. 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.
  37. Pan, B., Xiong, Y., Shi, K. and Sundaralingam, M. (2003) Crystal structure of a bulged RNA tetraplex at 1.1 Å resolution: implications for a novel binding site in RNA tetraplex. *Structure*, **11**, 1423–1430.
  38. Mukundan, V.T., Do, N.Q. and Phan, A.T. (2011) HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic Acids Res.*, **39**, 8984–8991.
  39. 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.
  40. Meier, M., Moya-Torres, A., Krahn, N.J., McDougall, M.D., Orriss, G.L., McRae, E.K.S., Booy, E.P., McEleney, K., Patel, T.R., McKenna, S.A. *et al.* (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res.*, **46**, 5319–5331.
  41. Wu, H.N. and Uhlenbeck, O.C. (1987) Role of a bulged A residue in a specific RNA-protein interaction. *Biochemistry*, **26**, 8221–8227.
  42. Hermann, T. and Patel, D.J. (2000) RNA bulges as architectural and recognition motifs. *Structure*, **8**, R47–R54.
  43. Wang, Y.H., Bortner, C.D. and Griffith, J. (1993) RecA binding to bulge- and mismatch-containing DNAs. Certain single base mismatches provide strong signals for RecA binding equal to multiple base bulges. *J. Biol. Chem.*, **268**, 17571–17577.
  44. Lilley, D.M. (1995) Kinking of DNA and RNA by base bulges. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 7140–7142.
  45. Degtyareva, N., Subramanian, D. and Griffith, J.D. (2001) Analysis of the binding of p53 to DNAs containing mismatched and bulged bases. *J. Biol. Chem.*, **276**, 8778–8784.
  46. Chung, W.J., Heddi, B., Schmitt, E., Lim, K.W., Mechulam, Y. and Phan, A.T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2729–2733.
  47. 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.
  48. 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.
  49. Bakalar, B., Heddi, B., Schmitt, E., Mechulam, Y. and Phan, A.T. (2019) A minimal sequence for left-handed G-quadruplex formation. *Angew. Chem. Int. Ed. Engl.*, **58**, 2331–2335.
  50. Tran, P.L., Moriyama, R., Maruyama, A., Rayner, B. and Mergny, J.L. (2011) A mirror-image tetramolecular DNA quadruplex. *Chem. Commun. (Camb.)*, **47**, 5437–5439.
  51. Virgilio, A., Esposito, V., Citarella, G., Mangoni, A., Mayol, L. and Galeone, A. (2011) Unprecedented right- and left-handed quadruplex structures formed by heterochiral oligodeoxyribonucleotides. *Biochimie*, **93**, 1193–1196.
  52. Kypr, J., Kejnovska, I., Rencuk, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
  53. Lat, P.K. and Sen, D. (2018) (C<sub>2</sub>G<sub>4</sub>)<sub>n</sub> repeat expansion sequences from the C9orf72 gene form an unusual DNA higher-order structure in the pH range of 5–6. *PLoS One*, **13**, e0198418.
  54. Zuffo, M., Gandolfini, A., Heddi, B. and Granzhan, A. (2020) Harnessing intrinsic fluorescence for typing of secondary structures of DNA. *Nucleic Acids Res.*, **48**, e61.
  55. Zhao, A., Zhao, C., Ren, J. and Qu, X. (2016) Enantioselective targeting left-handed Z-G-quadruplex. *Chem. Commun. (Camb.)*, **52**, 1365–1368.
  56. Zhao, C., Song, H., Scott, P., Zhao, A., Tateishi-Karimata, H., Sugimoto, N., Ren, J. and Qu, X. (2018) Mirror-image dependence: targeting enantiomeric G-quadruplex DNA using triplex metallohelices. *Angew. Chem. Int. Ed. Engl.*, **57**, 15723–15727.
  57. Fu, B., Huang, J., Chen, Y., Wang, Y., Xue, T., Xu, G., Wang, S. and Zhou, X. (2016) Right-handed and left-handed G-quadruplexes have the same DNA sequence: distinct conformations induced by an organic small molecule and potassium. *Chem. Commun. (Camb.)*, **52**, 10052–10055.
  58. Winnerdy, F.R., Bakalar, B., Maity, A., Vandana, J.J., Mechulam, Y., Schmitt, E. and Phan, A.T. (2019) NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res.*, **47**, 8272–8281.
  59. Lee, W., Tonelli, M. and Markley, J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
  60. 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.
  61. Kabsch, W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.
  62. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 213–221.
  63. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.*, **68**, 352–367.
  64. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.
  65. Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.*, **67**, 355–367.
  66. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment <sup>15</sup>N, <sup>13</sup>C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.
  67. Laughlan, G., Murchie, A.I., Norman, D.G., Moore, M.H., Moody, P.C., Lilley, D.M. and Luisi, B. (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, **265**, 520–524.
  68. Phillips, K., Dauter, Z., Murchie, A.I., Lilley, D.M. and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.*, **273**, 171–182.
  69. Altona, C. and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.*, **94**, 8205–8212.