

A minimal sequence for left-handed G-quadruplex formation

Blaž Bakalar,^{[a],[b]} Brahim Heddi,^{[a],[c]} Emmanuelle Schmitt,^[d] Yves Mechulam,^[d] Anh Tuân Phan^{*[a]}

Abstract: Recently, we observed the first example of a left-handed G-quadruplex structure formed by natural DNA, named Z-G4. We analysed the Z-G4 structure and inspected its primary 28-nt sequence in order to identify motifs that convey the unique left-handed twist. Using CD, NMR spectroscopy and X-ray crystallography, we revealed a minimal sequence motif of 12 nt GTGGTGGTGGTG for the left-handed DNA G-quadruplex formation, which is found to be highly abundant in the human genome. A systematic analysis of thymine loop mutations revealed a moderate sequence tolerance, which would further broaden the space of sequences prone to left-handed G-quadruplex formation.

G-quadruplexes are four-stranded structures of nucleic acids. They are formed by piling up multiple G-tetrads, each being a coplanar arrangement of four guanines.^[1] G-quadruplexes are involved in essential cellular processes.^[2] Studies have shown that G-quadruplex-forming sequences possess anti-cancer^[3] and anti-HIV^[4] activities. Such peculiar structures have also been used in nanotechnology.^[5]

Several G-quadruplex folding topologies have been reported that differ by their relative strand orientations and loop types.^[6] The structural variability observed for G-quadruplexes depends on the primary sequence,^[7] ions type,^[6b, 6c] salt concentration^[8] and environment.^[9] Diverse new G-quadruplex folds have been found recently including G-quadruplexes with bulges^[10] and duplex hairpins.^[11] Related alternative conformations of G-rich DNA include a hairpin,^[12] triplex^[13] and alternative tetraplex.^[14] Bioinformatic and sequencing analyses show that G-quadruplexes are present in important regulatory region of the human genome.^[15]

All these diverse structures were found with a right-handed helical twist, except for a G-quadruplex formed by a 28-nt DNA sequence, named Z-G4, derived from the aptamer AGRO100 (or AS1411) that was found with an unprecedented left-handed helicity^[16] further enlarging the variability of G-quadruplex structural space. However, little is understood about the

sequence requirement for left-handed G-quadruplex formation. Unlike the left-handed duplex Z-DNA, the formation of the left-handed G-quadruplex does not require high salt concentration or base modifications, hence sequence specific requirement would explain the formation of a left-handed G-quadruplex.

We explored how the DNA sequence affects the handedness of G-quadruplex structure and identified a minimal sequence sufficient for the formation of a left-handed G-quadruplex.

Sequence analysis of Z-G4. The Z-G4 structure is composed of two continuously connected blocks of parallel left-handed two-layered G-quadruplexes (Figure 1). The sequences of the two blocks differ slightly, whereas the 5'-end block (*Block 1*, T(GGT)₄T) is composed of four continuous G-doublets, the 3'-end block (*Block 2*, GT(GGT)₃GTT) has three G-doublets and two separated single guanines (underlined). In the Z-G4 structure, the two split Gs in *Block 2* (G15 and G26) stack upon each other in the same way as G-doublets (Figure 1) which is facilitated by the backbone circling all the way around the G-tetrad core. We examine which of these two sequences (*Block 1* or *Block 2*) infers the left-handed helicity.

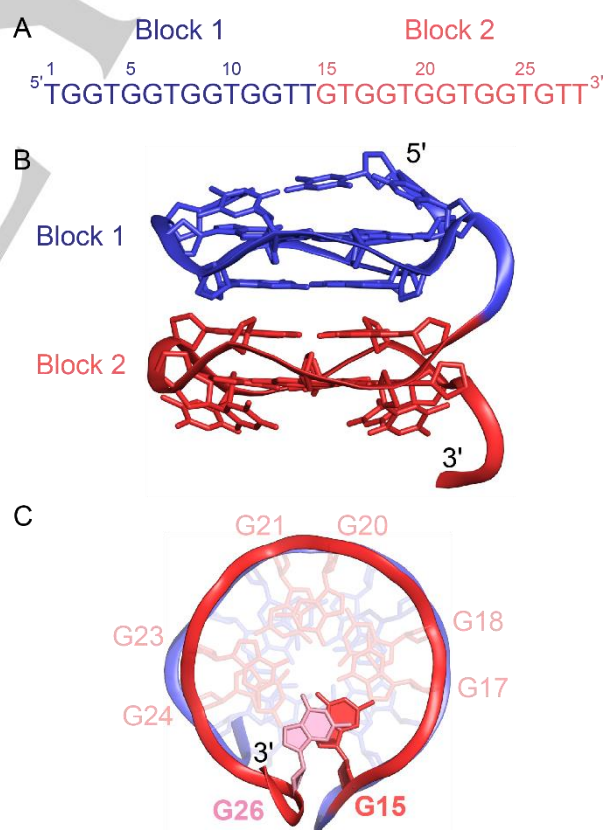


Figure 1. (A) Sequence and (B) structure of the left-handed Z-G4 structure.^[16] (C) Bottom view of the Z-G4 structure, with split Gs in *Block 2* being highlighted. The two constitutive blocks are coloured in blue (*Block 1*) and red (*Block 2*) respectively. Some thymine residues are not shown for clarity.

[a] B. Bakalar, Dr. B. Heddi, Prof. Dr. A. T. Phan
School of Physical and Mathematical Sciences,
Nanyang Technological University,
Singapore 637371 (Singapore)
E-mail: phantuan@ntu.edu.sg

[b] B. Bakalar
School of Biological Sciences,
Nanyang Technological University,
Singapore 637551 (Singapore)

[c] Dr. B. Heddi
Laboratoire de Biologie et Pharmacologie Appliquée,
CNRS, Ecole Normale Supérieure Paris-Saclay, Cachan 94235
(France)

[d] Dr. E. Schmitt, Dr. Y. Mechulam
Laboratoire de Biochimie, UMR 7654
CNRS, Ecole Polytechnique, Palaiseau 91128 (France)

Block 1 and Block 2 display right- and left-handed G-quadruplex features, respectively. The NMR imino proton spectra of *Block 1* (Figure S1) and *Block 2* (Figure S2) show peaks from 10.5 to 11.8 ppm, indicating the formation of G-quadruplexes. The distribution of imino protons of *Block 2*, with peaks being separated into two groups, is highly similar to that observed for *Z-G4*. A particular feature of the *Z-G4* structure is the formation of hydrogen bonds between O4' of the capping thymines and the guanine amino protons of the adjacent tetrad, resulting in sharp amino protons at 9–10 ppm;^[17] such protons are also observed for *Block 2* (Figure S2).

The CD spectrum of *Block 1* (Figure 2) has a positive peak at 260 nm and a negative peak at 240 nm, characteristic of parallel right-handed G-quadruplexes.^[18] The CD spectrum of *Block 2* has a negative peak at 270 nm and a positive peak at 250 nm (Figure 2), similar to that of the left-handed G-quadruplex, *Z-G4*. Altogether, these results indicate that *Block 2* forms a left-handed G-quadruplex.

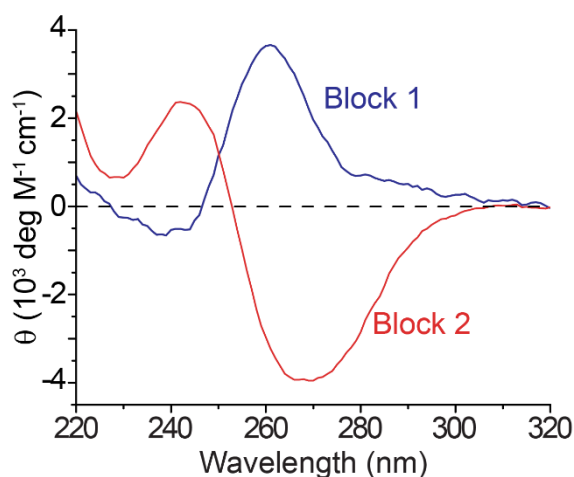


Figure 2. CD spectra of *Block 1* (blue) and *Block 2* (red). θ refers to the molar ellipticity.

Block 2 forms a dimeric left-handed G-quadruplex. The *Block 2* sequence and a variant without two Ts at the 3'-end (Table S1) display very similar CD and NMR spectra (Figure S2 and Figure S3), indicating the formation of the same structure. The latter sequence was successfully crystallized and the crystal structure was determined to 2.3 Å resolution (Table 1, PDB accession number 6FQ2). In our structure, the asymmetric unit is composed of two G-quadruplexes (subunits A and B) stacked on the 5'-5' interface. Each subunit is a two-layered parallel-stranded G-quadruplex structure (Figure 3) with a left-handed helicity. The two subunits can be superimposed with r.m.s.d.=0.77 Å for 253 atom pairs, reflecting adjustment of the thymine residues in the two blocks (Figure S4), probably due to packing interactions in the crystal involving different thymine residues for subunits A (T2 and T8) and B (T5 and T11) (Figure S5). The two G-tetrads in each subunit are formed by G1•G10•G7•G4 and G3•G12•G9•G6 with identical hydrogen-bond directionalities. The stacking interface of the subunits is 5/6 ring^[19] with G1 and G10 from one subunit stacking above G10 and G1 from the other subunit, respectively.

Thymines cap the outer G-tetrads of both subunits, allowing for hydrogen bond formation between the O4' sugar of thymines and amino proton of outer tetrad guanines. Stacking between two consecutive guanines in the tetrad core is partial 5/6-member ring with a twist mean angle value of $(-25.2 \pm 0.6)^\circ$. Dihedral backbone angles are summarized in Table 2, consistent with the values found for *Z-G4*. Three K⁺ ions are found between the G-tetrads, 12 molecules of water are located around the grooves and on top of G-tetrads. In both subunits, an intramolecular hydrogen bond could be formed between backbone oxygens of the first and last residues (Figure S6), likely further stabilizing the folded structure. Adding a thymine to the 5'-end of *Block 2* abolished the formation of the left-handed G-quadruplex (Figure S2), indicating that the dimerization of *Block 2* through 5'-5' stacking is important for the stability of this structure.

Overall the structural features are very similar to those found in *Z-G4*. Comparison of two *Block 2* subunits with the 3' half of *Z-G4* (Figure S7) shows a high similarity with an all-atom r.m.s.d. < 1.0 Å.

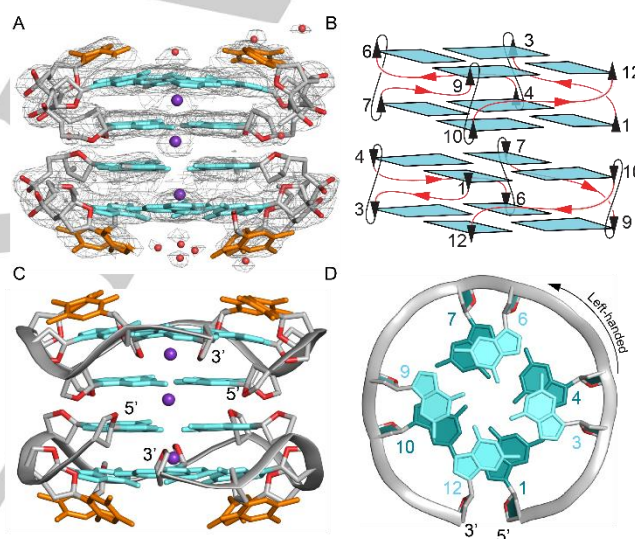


Figure 3. (A) Crystal structure of the left-handed G-quadruplex formed by *Block 2*. The electron density corresponds to the final 2mFo-DFc map contoured at 1.0 σ . (B) Schematic representation of the crystal structure with numbered guanine residues and marked backbone progression direction. (C) Ribbon representation of the crystal structure. (D) Top view of the crystal structure.

Sequence containing two *Block 2* motifs forms a monomeric left-handed G-quadruplex. Since *Block 2* requires the 5'-5' stacking dimerization in order to form a stable structure, we further connected two *Block 2* sequences with a TT linker. Remarkably, the new *2xBlock 2* sequence (Table S1) preserved both the CD and NMR spectral patterns of *Block 2* described above. The observation of a negative CD peak at 270 nm and 16 imino proton NMR peaks from 10.5 to 11.8 ppm (Figure S8) indicated the formation of a single, left-handed G-quadruplex structure.

We have successfully crystallized *2xBlock 2* and solved the crystal structure to 2.0 Å resolution (Table 1, PDB accession number 6GZ6). The structure of *2xBlock 2* exhibits very similar features as compared to the *Block 2* dimer (Figure 4). Compared

to *Z-G4*, where the linker between *Block 1* and *Block 2* connects two neighbouring inner tetrads^[16] (Figure 1), the connecting linker of *2xBlock 2* spans three tetrads, from the outer tetrad of one block to the inner tetrad of the other block (Figure 4). The hydrogen bond observed between dG1-O5' and dG12-O3' of *Block 2* was not seen in *Z-G4* and *2xBlock 2*, indicating that it is likely an additional feature in *Block 2* and not critical for left-handed G-quadruplex formation. The crystal packing of *2xBlock 2* is similar to that observed for *Z-G4* and *Block 2*; where the interactions are mediated by thymine residues in the T-cap and linker (Figure S9). Eight out of eleven thymine residues are engaged in crystal packing interactions.

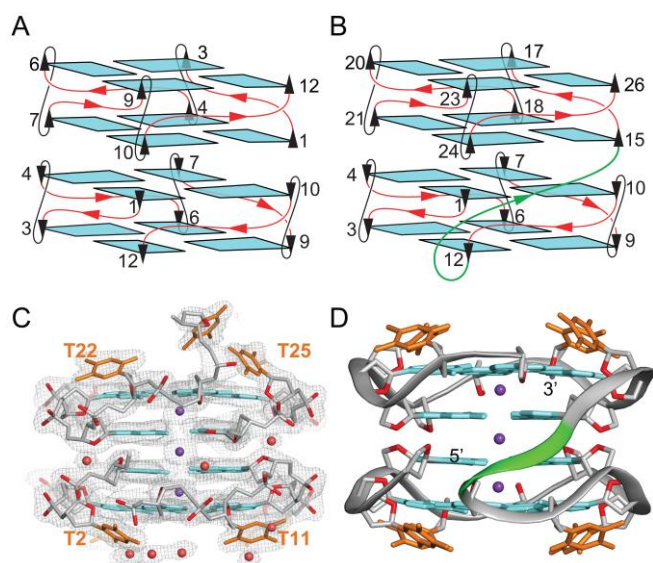


Figure 4. (A) Schematic representation of *Block 2* for comparison. (B) Schematic representation of the *2xBlock 2*. (C) Crystal structure of the left-handed G-quadruplex formed by *2xBlock 2*. The electron density corresponds to the final 2mFo-DFc map contoured at 1.0 σ (D) Ribbon representation of the crystal structure.

Table 2. Average backbone dihedral angles of *Block 2*, *2xBlock 2* and *Z-G4*

Backbone dihedral	<i>Block 2</i> ^[a]	<i>2xBlock 2</i> ^[a]	<i>Z-G4</i> ^[b]
ϵ	(321 \pm 2) $^\circ$	(323 \pm 3) $^\circ$	(316 \pm 2) $^\circ$
ζ	(82 \pm 1) $^\circ$	(84 \pm 4) $^\circ$	(86 \pm 3) $^\circ$
$\alpha+1$	(159 \pm 4) $^\circ$	(164 \pm 6) $^\circ$	(161 \pm 6) $^\circ$
$\beta+1$	(232 \pm 6) $^\circ$	(222 \pm 4) $^\circ$	(226 \pm 5) $^\circ$
$\gamma+1$	(53 \pm 2) $^\circ$	(51 \pm 6) $^\circ$	(54 \pm 3) $^\circ$

[a] Average values and standard deviations obtained from the guanine dinucleotide steps. [b] Values reported in^[16].

Table 1. Data collection and refinement statistics for *Block 2* and *2xBlock 2*

Data collection	<i>Block 2</i>	<i>2xBlock 2</i>
Space group	P3 ₂	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	30.56, 30.56, 53.66	29.57, 37.63, 54.28
α , β , γ ($^\circ$)	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution (Å)	26.46 2.31	30.93 2.01
R _{sym}	0.031 (0.230) ^[a]	0.089 (0.668)
I / σ	16.4 (3.4)	18.3 (3.95)
Completeness (%)	97.7 (96.9)	98.9 (93.6)
Redundancy	2.8 (2.3)	13.0 (13.4)
CC1/2 (%)	100 (93)	100 (96)
Refinement		
Resolution (Å)	26.46 – 2.31	30.93 (2.01)
Number of reflections	2405	4285
Twin fraction	0.51 (k, h, -l)	
R _{work} / R _{free}	0.196 / 0.227	0.185 / 0.231
Number of atoms		
DNA	506	569
K ⁺	3	3
Waters	12	10
B-factors (Å ²)		
DNA	44.5	31.0
K ⁺	27.4	22.1
Water	26.6	31.0
r.m.s.d.		
Bond lengths (Å)	0.004	0.008
Bond angles ($^\circ$)	0.697	0.997

[a] values in parentheses are for the highest resolution shell

Loop and linker substitutions show moderate sequence tolerance of the left-handed G-quadruplex motif. To determine the sequence flexibility for left-handed G-quadruplex formation, we carried out a systematic analysis of thymine loop substitutions and monitored the characteristic CD and NMR spectral patterns for the structure formation. We chose *2xBlock 2Δ* (a variant of *2xBlock 2* without a T at the 3'-end) as our reference scaffold, since it does not require dimerization and has a cleaner NMR spectrum as compared to *Block 2* (Figure S8). The results are summarized in Table 3, while the list of sequences studied is shown in Table S2. Based on the CD data, the left-handed G-quadruplex formation was maintained when the linker connecting the two blocks of *2xBlock 2Δ* was varied in length (2-4 nt) or composition (Figure S10). In order to better translate the results to the minimal left-handed motif, the loop substitutions were carried out in a single block or at equivalent thymine positions in both blocks simultaneously (Table S2). CD and NMR data show that the left-handed G-quadruplex fold was maintained, when a T-to-C substitution was done at each of the four loop positions (2, 5, 8 and 11) of a single block (Figure S11). Thymine loop substitutions to either C, A, or TT preserved the left-handed G-quadruplex fold even when all four substitutions are carried out in a single block (Figure S12). Furthermore, one T-to-C or T-to-A substitution in both blocks did not alter the fold (Figure S13). In the cases of two T-to-C or T-to-A substitutions in each block, NMR spectra suggested the presence of left-handed G-quadruplexes, however, the intensity of the characteristic negative CD peak at 270 nm was low (Figure S14). Other cases of multiple loop substitutions in both blocks (Table 3) abolished the left-handed G-quadruplex formation (Figure S15).

COMMUNICATION

Table 3. Results of *2xBlock 2A* thymine loop substitutions

	Connecting loop modifications					
	TT	TA	AT	AA	TTT	TTTT
	✓	✓	✓	✓	✓	✓
	Single block position modifications					
	T2 to C	T5 to C	T8 to C	T11 to C		
	✓	✓	✓	✓		
	Substitution in single block					
	T to C	T to A	T to C	T to TT		
1 substitution	✓	✓	✓	✓		
2 substitutions	✓	✓	✓	✓		
3 substitutions	✓	✓	✓	✓		
4 substitutions	✓	✓	✓	✓		
	Substitution in both blocks					
	T to C	T to A	T to C	T to TT		
2 substitution	✓	✓	✓	✗		
4 substitutions	—	—	—	✗		
6 substitutions	✗	✗	✗	✗		
8 substitutions	✗	✗	✗	✗		

✓ - Left-handed G4 observed by NMR and CD
 — - Left-handed G4 suggested by NMR (CD patterns unclear)
 ✗ - Left-handed G4 not observed

The sequence analysis of *Z-G4* together with new experimental evidence shows that a shorter motif of 12-nt GT(GGT)₃G is essential and sufficient for the formation of a G-quadruplex with a left-handed helicity. It is also this short motif which is able to change the helicity of the otherwise right-handed G-quadruplex forming *Block 1*, within the structure of *Z-G4*.

The split-guanines feature of the *Block 2* and *2xBlock 2* sequences represents the only major difference when compared to the sequence of *Block 1* and we can reasonably assume that the unusual split stacking guanines convey the left-handed twist. More computational and experimental data are needed to fully understand the effect of the split stacked guanines on DNA helicity.

The motif deviates significantly from the previously proposed model sequences, such as G₃₊N₁₋₇G₃₊N₁₋₇G₃₊N₁₋₇G₃₊, used for finding putative G4-forming sequences.^[20] Interestingly, a bioinformatic search for the GT(GGT)₃G motif in the human genome found over 10,000 hits, representing a two-order of magnitude enrichment compared to the occurrence of a 12-nt motif by chance, which may be partly due to the repetitive nature of the motif. Sequence tolerance would further broaden the space of sequences prone to left-handed G-quadruplex formation.

Acknowledgements

This research was funded by Singapore Ministry of Education Academic Research Fund Tier 2 (MOE2015-T2-1-092) and Tier 3 (MOE2012-T3-1-001), National Research Foundation Investigatorship (NRF-NRF12017-09) and grants from Nanyang Technological University to A.T.P.

Keywords: DNA • circular dichroism • G-quadruplex • left-handed • X-ray structure

- [1] a) M. Gellert, M. N. Lipsett, D. R. Davies, *Proc Natl Acad Sci U S A* **1962**, *48*, 2013-2018; b) D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364-366.
 [2] D. Rhodes, H. J. Lipps, *Nucleic Acids Res* **2015**, *43*, 8627-8637.

- [3] a) P. J. Bates, J. B. Kahlon, S. D. Thomas, J. O. Trent, D. M. Miller, *J Biol Chem* **1999**, *274*, 26369-26377; b) T. Simonsson, M. Henriksson, *Biochem Biophys Res Commun* **2002**, *290*, 11-15; c) H. Qi, C. P. Lin, X. Fu, L. M. Wood, A. A. Liu, Y. C. Tsai, Y. Chen, C. M. Barbieri, D. S. Pilch, L. F. Liu, *Cancer Res* **2006**, *66*, 11808-11816; d) N. Q. Do, W. J. Chung, T. H. A. Truong, B. Heddi, A. T. Phan, *Nucleic Acids Res* **2017**, *45*, 7487-7493.
 [4] a) A. T. Phan, V. Kuryavyi, J. B. Ma, A. Faure, M. L. Andreola, D. J. Patel, *Proc Natl Acad Sci U S A* **2005**, *102*, 634-639; b) N. Q. Do, K. W. Lim, M. H. Teo, B. Heddi, A. T. Phan, *Nucleic Acids Res* **2011**, *39*, 9448-9457; c) D. Musumeci, C. Riccardi, D. Montesarchio, *Molecules* **2015**, *20*, 17511-17532; d) M. Metifiot, S. Amrane, J. L. Mergny, M. L. Andreola, *Biochimie* **2015**, *118*, 173-175; e) R. Perrone, E. Butovskaya, S. Lago, A. Garzino-Demo, C. Pannecouque, G. Palu, S. N. Richter, *Int J Antimicrob Agents* **2016**, *47*, 311-316.
 [5] a) D. Miyoshi, H. Karimata, Z. M. Wang, K. Koumoto, N. Sugimoto, *J Am Chem Soc* **2007**, *129*, 5919-5925; b) Y. C. Huang, D. Sen, *J Am Chem Soc* **2010**, *132*, 2663-2671; c) S. P. Liu, S. H. Weisbrod, Z. Tang, A. Marx, E. Scheer, A. Erbe, *Angew Chem Int Ed Engl* **2010**, *49*, 3313-3316; d) G. I. Livshits, A. Stern, D. Rotem, N. Borovok, G. Eidelstein, A. Migliore, E. Penzo, S. J. Wind, R. Di Felice, S. S. Skourtis, J. C. Cuevas, L. Gurevich, A. B. Kotlyar, D. Porath, *Nat Nanotechnol* **2014**, *9*, 1040-1046; e) J. Thevarpadam, I. Bessi, O. Binas, D. P. Goncalves, C. Slavov, H. R. Jonker, C. Richter, J. Wachtveitl, H. Schwalbe, A. Heckel, *Angew Chem Int Ed Engl* **2016**, *55*, 2738-2742.
 [6] a) J. T. Davis, *Angew Chem Int Ed Engl* **2004**, *43*, 668-698; b) S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res* **2006**, *34*, 5402-5415; c) D. J. Patel, A. T. Phan, V. Kuryavyi, *Nucleic Acids Res* **2007**, *35*, 7429-7455.
 [7] a) M. Webba da Silva, M. Trajkovski, Y. Sannohe, N. Ma'ani Hessari, H. Sugiyama, J. Plavec, *Angewandte Chemie International Edition* **2009**, *48*, 9167-9170; b) L. Hu, K. W. Lim, S. Bouaziz, A. T. Phan, *J Am Chem Soc* **2009**, *131*, 16824-16831; c) P. Agrawal, C. Lin, R. I. Mathad, M. Carver, D. Yang, *J Am Chem Soc* **2014**, *136*, 1750-1753; d) M. Marusic, J. Plavec, *Angew Chem Int Ed Engl* **2015**, *54*, 11716-11719.
 [8] a) N. Q. Do, A. T. Phan, *Chemistry* **2012**, *18*, 14752-14759; b) M. Trajkovski, M. W. da Silva, J. Plavec, *J Am Chem Soc* **2012**, *134*, 4132-4141.
 [9] a) D. Miyoshi, H. Karimata, N. Sugimoto, *J Am Chem Soc* **2006**, *128*, 7957-7963; b) M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane, J. O. Trent, *Journal of the American Chemical Society* **2010**, *132*, 17105-17107; c) B. Heddi, A. T. Phan, *J Am Chem Soc* **2011**, *133*, 9824-9833.
 [10] a) V. T. Mukundan, A. T. Phan, *J Am Chem Soc* **2013**, *135*, 5017-5028; b) B. De Nicola, C. J. Lech, B. Heddi, S. Regmi, I. Frasson, R. Perrone, S. N. Richter, A. T. Phan, *Nucleic Acids Res* **2016**, *44*, 6442-6451.
 [11] a) Z. Yu, V. Gaerig, Y. Cui, H. Kang, V. Gokhale, Y. Zhao, L. H. Hurley, H. Mao, *J Am Chem Soc* **2012**, *134*, 5157-5164; b) K. W. Lim, A. T. Phan, *Angew Chem Int Ed Engl* **2013**, *52*, 8566-8569.
 [12] M. Gajarsky, M. L. Zivkovic, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomaska, J. Sponer, J. Plavec, L. Trantirek, *J Am Chem Soc* **2017**, *139*, 3591-3594.
 [13] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, M. Parrinello, *Angew Chem Int Ed Engl* **2013**, *52*, 2269-2273.
 [14] V. Kocman, J. Plavec, *Nat Commun* **2017**, *8*, 15355.
 [15] a) N. Maizels, L. T. Gray, *PLoS Genet* **2013**, *9*, e1003468; b) A. Bedrat, L. Lacroix, J. L. Mergny, *Nucleic Acids Res* **2016**, *44*, 1746-1759; c) V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat Biotechnol* **2015**, *33*, 877-881.
 [16] W. J. Chung, B. Heddi, E. Schmitt, K. W. Lim, Y. Mechulam, A. T. Phan, *Proc Natl Acad Sci U S A* **2015**, *112*, 2729-2733.
 [17] M. Adrian, F. R. Winnerdy, B. Heddi, A. T. Phan, *Biophys J* **2017**, *113*, 775-784.
 [18] J. Kyrp, I. Kejnovska, D. Renciu, M. Vorlickova, *Nucleic Acids Res* **2009**, *37*, 1713-1725.
 [19] C. J. Lech, B. Heddi, A. T. Phan, *Nucleic Acids Res* **2013**, *41*, 2034-2046.

- [20] a) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res* **2005**, *33*, 2908-2916; b) A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res* **2005**, *33*, 2901-2907.

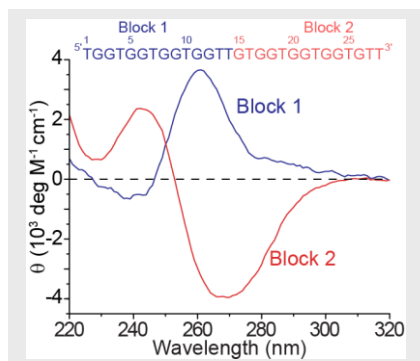
WILEY-VCH

COMMUNICATION

Layout 1:

COMMUNICATION

12-nt sequence GTGGTGGTGGTG was identified as a minimal sequence motif for the left-handed DNA G-quadruplex formation. Such a motif is abundant in the human genome.



*Blaž Bakalar, Brahim Heddi,
Emmanuelle Schmitt, Yves Mechulam,
Anh Tuân Phan**

Page No. – Page No.

**A minimal sequence for left-handed
G-quadruplex formation**