

Engineering of interlocked DNA G-quadruplexes as a robust scaffold

Anh Tuân Phan^{1,*} and Ngoc Quang Do^{1,2}

¹School of Physical and Mathematical Sciences and ²School of Biological Sciences, Nanyang Technological University, Singapore

Received May 18, 2012; Revised November 12, 2012; Accepted November 16, 2012

ABSTRACT

Interlock is a structural element in DNA G-quadruplexes that can be compared with the commonly used complementary binding of ‘sticky ends’ in DNA duplexes. G-quadruplex interlocking can be a basis for the assembly of higher-order structures. In this study, we formulated a rule to engineer (3+1) interlocked dimeric G-quadruplexes and established the folding topology of the designed DNA sequences by nuclear magnetic resonance spectroscopy. These interlocked G-quadruplexes are very stable and can serve as compact robust scaffolds for various applications. Different structural elements can be engineered in these robust scaffolds. We demonstrated the anti-HIV inhibition activity of the newly designed DNA sequences.

INTRODUCTION

Guanine-rich (G-rich) nucleic acid sequences can adopt special structures called G-quadruplexes based on stacking of G•G•G•G tetrads (1–6). In nature, G-rich DNA and RNA sequences are found in many important genomic regions, such as the telomeres, oncogenic promoters and RNA 5'-untranslated regions (1–6). G-quadruplexes formed by these natural sequences have been established as attractive anticancer targets (7,8). On the other hand, some engineered G-rich oligonucleotides, obtained by *in vitro* selection (9,10) or rational design (11–13) and capable of forming G-quadruplexes (14–16), possess interesting biological activities, such as anticoagulant (9), anticancer (11–13) and anti-HIV activities (10). G-quadruplexes can also have potential applications in chemistry, material sciences and nanotechnology (1). For example, these scaffolds can be used as a drug delivery system or a basis for the design of new materials and nanodevices; they can also be used to guide charge transfer or to support catalysts (17–23). Studies have shown that G-quadruplex structures are highly polymorphic (1–6). Therefore, it is important to control their

different structural elements and folding topologies for various applications.

Interlock (16,24,25) is a structural element in G-quadruplexes that can be compared with the commonly used complementary binding of ‘sticky ends’ in DNA duplexes. G-quadruplex interlocking can be a basis for higher-order structures, such as G-wires (26). In the structure of interlocked G-quadruplexes (16,24,25), G-tetrads at the interface between the two subunits are completed by (i) two guanines from each subunit (in ‘2+2’ type) (24,25) or (ii) three guanines from one subunit and one guanine from another subunit (in ‘3+1’ type) (16).

The *93del* sequence d(GGGGTGGGAGGGT), a HIV-1 integrase inhibitor (10), forms in K⁺ solution an extremely stable ‘3+1’-type interlocked dimeric G-quadruplex with totally six G-tetrad layers (16) (Figure 1a). Dimer formation is achieved through mutual pairing of G1 from one subunit with G2, G6 and G13 of the other subunit. In each subunit, there are (i) four parallel-stranded G-tracts (or legs) and (ii) three single-residue double-chain-reversal loops bridging two or three G-tetrad layers. The first leg is long with four Gs; the second and fourth legs are medium with three Gs; the third one is short with only two Gs. The 5'-end guanine from the long leg (first) of one subunit compensates the short leg (third) of the other subunit in completing the G-tetrads at the interlocking interface (Figure 1a). Here, we aim to determine a rule to control this G-quadruplex interlocking motif.

METHODS

Sample preparation

Unlabeled and site-specific labeled DNA oligonucleotides were chemically prepared using products from Glen Research and Cambridge Isotope Laboratories. Samples were purified following the protocol from Glen Research and then dialyzed successively against KCl and water solution. DNA oligonucleotides were dissolved in a solution containing 70 mM potassium chloride and 20 mM potassium phosphate, pH 7.0. DNA concentration was expressed in strand molarity using a nearest-neighbor

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

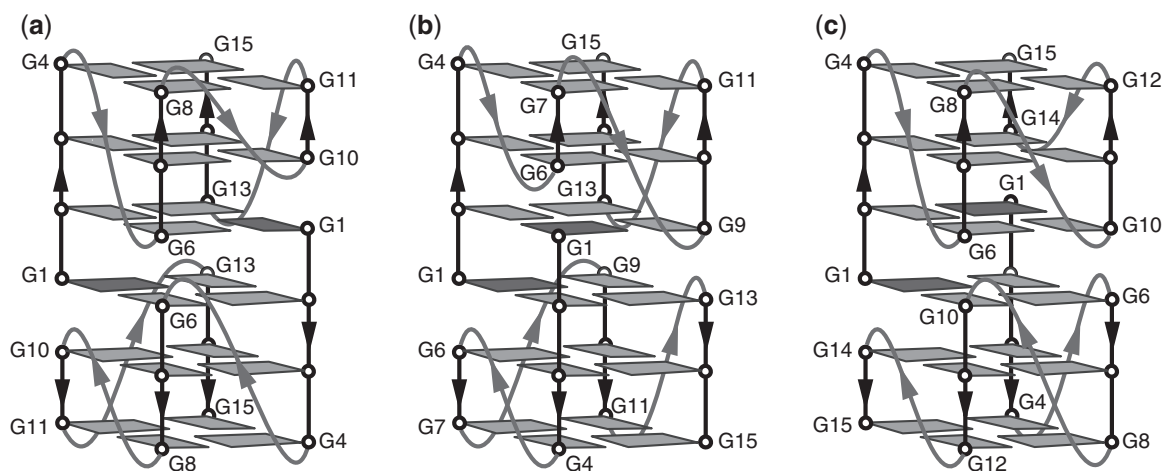


Figure 1. Schematic structures of dimeric (3+1) interlocked G-quadruplexes: (a) *93del*, (b) *s2* and (c) *s4*. *Anti* and *syn* guanines are shown in light and dark grey, respectively.

approximation for the extinction coefficients of the unfolded species.

Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) experiments were performed on Bruker Avance spectrometers operating at 600 or 700 MHz for ^1H . Spectra were recorded at 25°C, unless otherwise specified. Experiments in H_2O used the jump-and-return (JR)-type water suppression for detection. Guanine resonances were assigned unambiguously by using site-specific low-enrichment ^{15}N labeling, site-specific ^2H labeling, through-bond correlations at natural abundance (^{13}C , ^1H) JRHMB, ^{13}C , ^1H HSQC and ^1H , ^1H TOCSY) and NOESY. Thymine resonances were identified by the through-bond H6- CH_3 correlations from TOCSY spectra. The folding topologies and other structural information of G-quadruplexes were obtained from NOESY spectra.

Protein preparation

The pET28b plasmid containing the mutated core domain of HIV-1 integrase, IN⁵⁰⁻²¹²(F185H), was expressed in *Escherichia coli* strain BL21 (DE3-Gold) as described previously (16). The cells were grown at 37°C in BL medium containing 50 $\mu\text{g}/\text{ml}$ of kanamycin until the 600-nm optical density reached 0.5–0.8, then induced by adding 0.5 mM IPTG and let grow in shaking culture for 3 h. The cells were harvested by centrifugation at 8000 rpm for 10 min at 4°C. The cell pellet from 1 L culture was suspended in 20 ml of lysis buffer (25 mM HEPES, pH 7.5, 0.5 M NaCl, 2 mM 2-mercaptoethanol, 5 mM imidazole, 0.5% Triton X-100, 1 mM PMSF). The cells were sonicated and the debris was discarded by centrifugation at 14 000 rpm for 20 min at 4°C. The collected supernatant was filtered through a 0.2- μm filter and applied to a HisTrapTM HP column. The column was then washed successively by elution buffer (25 mM HEPES, pH 7.5, 0.5 M NaCl, 2 mM 2-mercaptoethanol, 10% glycerol), which contained 20, 150, 400 and 800 mM imidazole, respectively, with five column-volumes for each imidazole

concentration. The protein was collected at 400 mM imidazole fraction and dialyzed overnight at 4°C in a buffer containing 25 mM HEPES, pH 7.5, 0.5 M NaCl, 1 mM DTT, 1 mM EDTA, 10% (wt/vol) glycerol. The sample was frozen in liquid nitrogen and stored at -80°C .

HIV-1 integrase activity test

The reaction mixture contained 20 mM HEPES (pH 7.5), 10 mM MnCl_2 , 30 mM NaCl, 10 mM DTT, 0.05% Nonidet-P40, 600 nM HIV-1 integrase and 200 nM DB-Y1. The DB-Y1 substrate (5'-TGCTAGTTCTAGCAGGCCCTTGGGCCGGCGCTTGCGCC-3') used in the reaction (purchased from 1st BASE Pte. Ltd., Singapore) was labeled with 6-FAM at the 5'-end. After incubating at 37°C for 90 min, the reaction was stopped by adding an equal volume of 98% deionized formamide, 10 mM EDTA, pH 8.0 and heated at 90°C for 3 min. For inhibition tests, the inhibitor was added into the mixture and incubated for 30 min before adding DB-Y1. The product was analyzed by electrophoresis on a 15% polyacrylamide gel supplemented with 7 M urea and viewed under 495-nm light.

RESULTS AND DISCUSSION

Rational design of G-rich sequences that form (3+1) interlocked G-quadruplexes

We propose that in general a (3+1) interlocked G-quadruplex might be formed by sequences containing one long G-tract, one short G-tract and two medium G-tracts. In the next section, we demonstrate by NMR that (3+1) interlocked dimeric G-quadruplexes (Figure 1) can be formed by sequences *s2* or *s4* (Table 1), where the first G-tract is long (with four Gs) and, respectively, the second or fourth G-tract is short (with two Gs), whereas the remaining G-tracts are medium (with three Gs). Single-residue linkers were chosen to connect G-tracts because they could form stable double-chain-reversal loops bridging two or three G-tetrad layers (16).

Table 1. Sequences used in this study

Name	Sequence (5'-3')
<i>93del</i>	GGGGTGGGAGGAGGGT
<i>s3</i>	GGGGTGGGTGGTGGGT
<i>s2</i>	GGGGTGGTGGGTGGGT
<i>s2-A</i>	GGGGAGGTGGGTGGGT
<i>s4</i>	GGGGTGGGTGGGTGGT
<i>s4-A</i>	GGGGTGGGTGGGAGGT
<i>s4-Br</i>	^{Br} GGGGTGGGTGGGTGGT
<i>s1R</i>	TGGTGGGTGGGTGGGG
<i>s2R</i>	TGGGTGGTGGGTGGGG
<i>s3R</i>	TGGGTGGGTGGTGGGG
<i>s3F</i>	GGGGTGGGTGGGTGGGT

We also ask whether (3 + 1) interlocked G-quadruplexes could be formed by sequences containing one long G-tract (four Gs), one short G-tract (two Gs) and two medium G-tracts (three Gs) but with the 3'-end fourth G-tract being long (instead of the 5'-end first G-tract) (sequences *s1R*, *s2R* and *s3R*; Table 1). Another question is whether a short G-tract is required to control the (3 + 1) interlocking motif; in other words, could a sequence with the first G-tract long (four Gs) and three medium G-tracts (three Gs) (sequence *s3F*, Table 1) form an interlocked G-quadruplex?

Verification of (3 + 1) interlocked G-quadruplex formation by NMR

We used NMR to establish the interlocked G-quadruplex topologies of *s2* and *s4*. Guanine imino and H8 protons

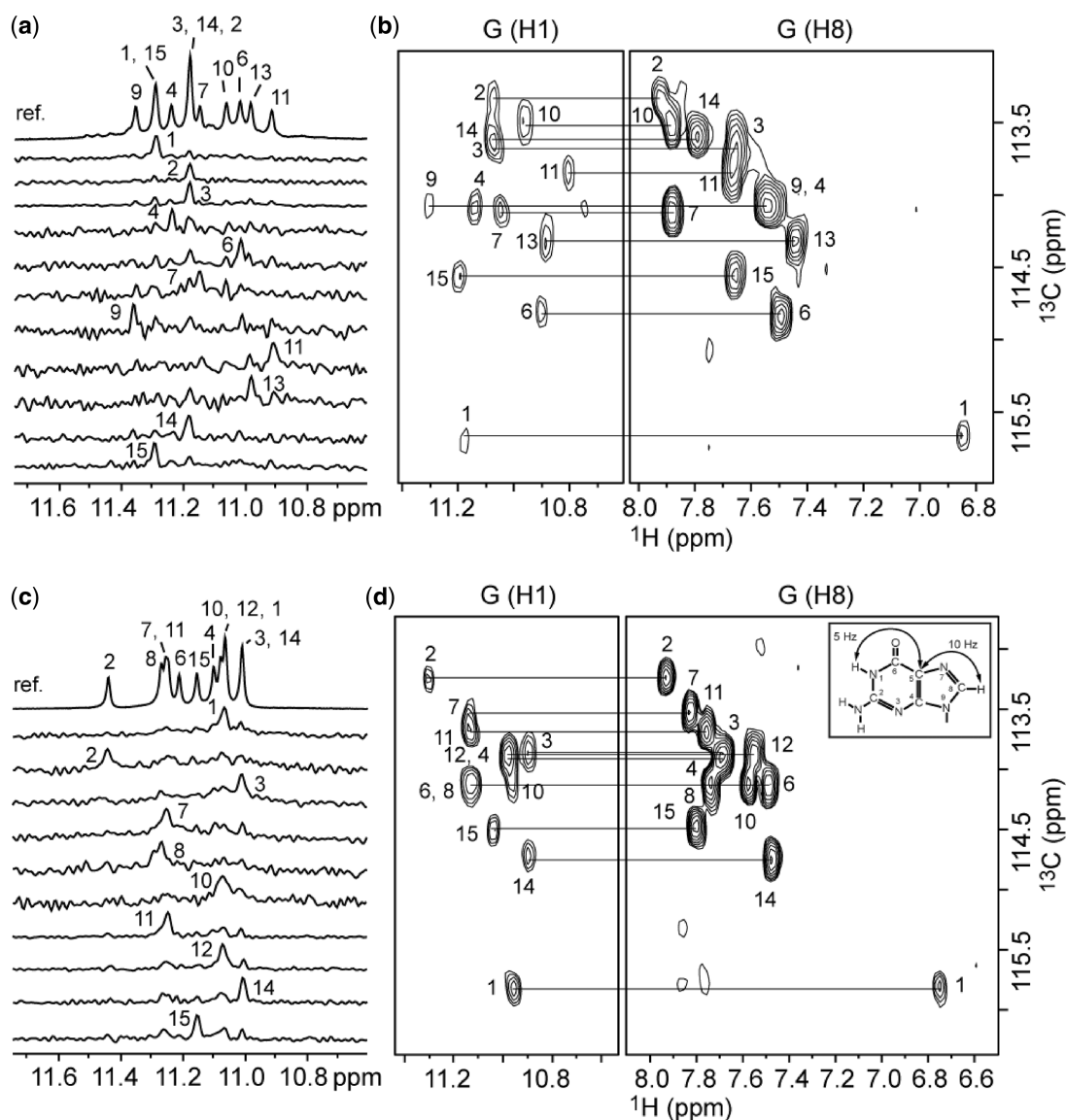


Figure 2. Resonance assignments of *s2* and *s4* in K^+ solution at 25°C. (a and c) Imino protons of *s2* and *s4*, respectively, were assigned by site-specific 2% ¹⁵N-labeling. (b and d) H8 protons of *s2* and *s4*, respectively, were assigned by through-bond correlations between imino and H8 protons via ¹³C5 at natural abundance, using long-range J couplings shown as insert in (d).

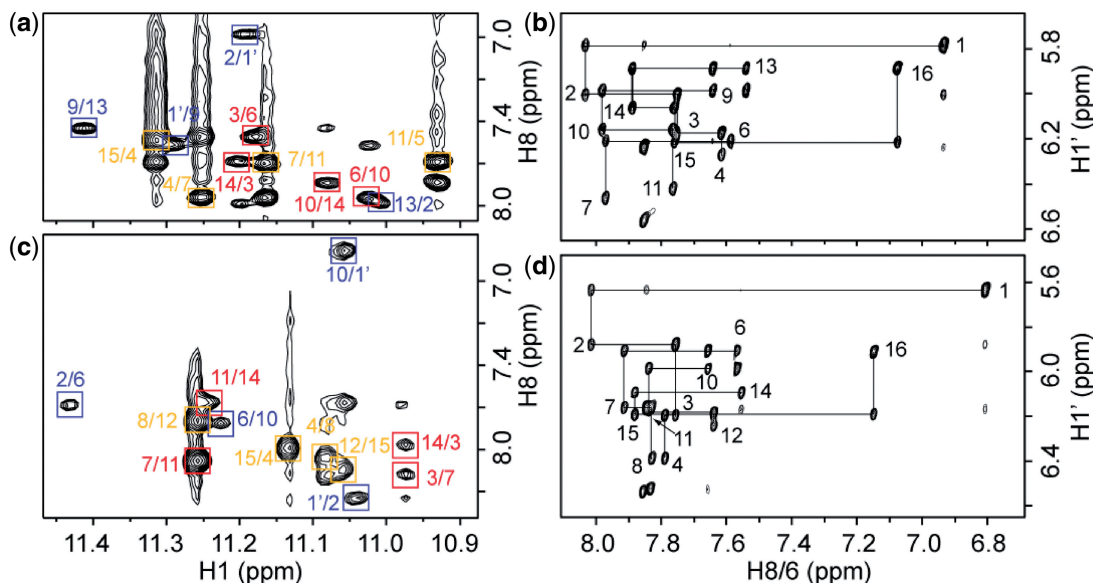


Figure 3. Determination of the G-quadruplex folding topology for the *s2* and *s4* sequences in K^+ solution. (a and c) The imino-H8 proton region of NOESY spectra (mixing time, 200 ms) of *s2* and *s4*, respectively at 25°C. The characteristic guanine imino-H8 cross-peaks for G-tetrads (colored red, yellow and blue, respectively) are framed and labeled with the imino proton assignment in the first position and that of the H8 proton in the second position. (b and d) The H8/6-H1' proton region of NOESY spectra (mixing time, 300 ms) of *s2* and *s4*, respectively at 25°C. The assignments and H8/6-H1' NOE sequential connectivities are shown.

were unambiguously assigned by using site-specific low-enrichment ^{15}N labeling, site-specific 2H labeling and through-bond correlations at natural abundance (27–30) (Figure 2 and Supplementary Figure S1). With the help of these unambiguous assignments and other through-bond correlation experiments (COSY, TOCSY, $[^{13}C-^1H]$ -HMBC and $[^{13}C-^1H]$ -HSQC) (data not shown), the classical H8/H6-H1' NOE sequential connectivity could be traced for four G-stretches (Figure 3). The intensity of intrasidue H8-H1' Nuclear Overhauser Effect (NOE) cross-peaks indicated a *syn* glycosidic conformation for G1 and *anti* conformations for the other guanines (Supplementary Figure S2). Specific NOEs between imino and H8 protons established G-tetrad alignments (Figure 3) and confirmed the designed folding topology of *s2* and *s4* (Figure 1). Each structure comprises six G-tetrad layers formed by interlocking of two parallel-stranded G-quadruplex subunits. The folding topologies of *93del*, *s2* and *s4* differ from each other by the relative positions of the two interlocking points: in *93del* these two points are across the G-tetrad core diagonal, whereas in *s2* and *s4* these two points are of different adjacent configurations (Figure 1). These proposed folding topologies are consistent with deuterium exchange data, which showed that imino protons of the middle G-tetrads are well protected from the exchange with solvent (Supplementary Figure S3; data not shown). These interlocked G-quadruplexes are all extremely stable, as some imino protons persist in D_2O solution even after 5 min at 100°C (Supplementary Figure S3).

In contrast to the situation of *s2* and *s4*, NMR imino proton spectra of sequences with the fourth G-tract being long (instead of the first one) (sequences *s1R*, *s2R* and *s3R*,

Table 1) did not indicate the formation of a well-defined G-quadruplex (Supplementary Figure S4). For a sequence with one long (first G-tract) and three medium G-tracts (sequence *s3F*, Table 1), the imino proton spectrum (Supplementary Figure S4) indicated the formation of more than one G-quadruplex conformations, which could be different possible interlocked or non-interlocked G-quadruplexes. This result shows that the presence of a short G-tract is important to control the formation of a well-defined (3 + 1) interlocked G-quadruplex.

Engineering of structural elements in (3 + 1) interlocked G-quadruplexes

Different structural elements can be engineered in these robust interlocked scaffolds. In *s2* and *s4*, all the non-G linkers are a thymine (T) residue, whereas in *93del* two linkers are an adenine (A) residue (Table 1). In the latter structure, the adenine before the short leg (namely, A9) participates in the formation of a $A \bullet (G \bullet G \bullet G \bullet G)$ pentad (Supplementary Figure S5). We attempted to design a $A \bullet (G \bullet G \bullet G \bullet G)$ pentad in the *s2* and *s4* interlocked G-quadruplexes. The 1D NMR spectra of the corresponding sequences, *s2-A* and *s4-A*, where the T residue before the short leg was substituted by an A, indicated the formation of $A \bullet (G \bullet G \bullet G \bullet G)$ pentads in these structures (Supplementary Figure S6). The formation of these pentads was further supported by NOESY experiments, e.g. the formation of the $A5 \bullet (G3 \bullet G6 \bullet G10 \bullet G14)$ pentad in *s2-A* was supported by the detection of a NOE cross-peak between the A5(H8) and G6(H8) protons (Supplementary Figure S7). Reversely, A-to-T substitutions of *93del* (i.e. sequence *s3*, Table 1) abolished the $A \bullet (G \bullet G \bullet G \bullet G)$ pentad but retained the general folding topology (Supplementary Figure S6).

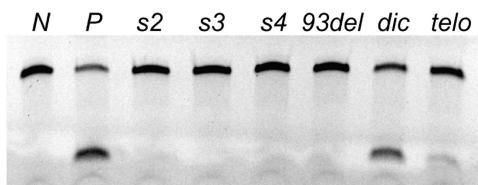


Figure 4. HIV-1 integrase inhibition disintegration assays. Lanes 'N' and 'P' are controls without and with HIV-1 integrase alone, respectively. Other lanes are inhibition assays with added inhibitors as indicated, where *dic* is the Dickerson d(CGCGAATTCGCG) dodecamer and *telo* is the human telomeric sequence d[TT(GGGTTA)₃GGGA]. The concentrations of the enzyme, substrate and inhibitors were 0.6, 0.2 and 3.0 μ M, respectively. The DNA substrate (dumbbell DB-Y1) and disintegration product (see 'Methods' section) are monitored as bands of 38 nt (top band) and 14 nt (bottom band), respectively (33,34).

In these (3+1) interlocked G-quadruplex topologies, G1 adopts a *syn* conformation. NMR data (Supplementary Figure S8) showed that this residue in *s4* could be substituted by a ^{Br}G, favoring a *syn* conformation in G-quadruplexes (31,32), without altering the general folding topology.

HIV-1 integrase inhibition activity of (3+1) interlocked G-quadruplexes

We demonstrated the inhibition activity of the three newly engineered sequences *s2*, *s3* and *s4* against HIV-1 integrase in the reverse 'disintegration' assays (33,34). Our results (Figure 4) indicated that these molecules have an HIV-1 integrase inhibition activity comparable with that of *93del* and higher than that of negative controls (a duplex DNA or a G-quadruplex DNA formed by a human telomeric sequence). Note, however, that this conclusion should only be taken as qualitative, because the sample concentrations used in the assays were in the micromolar range, much higher than the HIV-1 integrase inhibition activity of *93del* (10).

CONCLUSION

We have engineered three different (3+1) interlocked dimeric G-quadruplexes. These structures are very stable and can serve as compact robust scaffolds for various applications. Different structural elements can be added in these robust interlocked scaffolds. The principle for interlocked G-quadruplex formation can be used for the assembly of higher-order structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–8.

FUNDING

Singapore Biomedical Research Council, Singapore Ministry of Education and Nanyang Technological University (to A.T.P.). Funding for open access charge: Nanyang Technological University.

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.
- Phan, A.T., Kuryavyi, V. and Patel, D.J. (2006) DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.*, **16**, 288–298.
- 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.
- Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19**, 239–250.
- Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.
- Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. and Hurley, L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.*, **40**, 2113–2116.
- Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
- Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. and Toole, J.J. (1992) Selection of single-stranded-DNA molecules that bind and inhibit human thrombin. *Nature*, **355**, 564–566.
- de Soultrait, V.R., Lozach, P.Y., Altmeyer, R., Tarrago-Litvak, L., Litvak, S. and Andréola, M.L. (2002) DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *J. Mol. Biol.*, **324**, 195–203.
- Jing, N., Li, Y., Xiong, W., Sha, W., Jing, L. and Tweardy, D.J. (2004) G-quartet oligonucleotides: a new class of signal transducer and activator of transcription 3 inhibitors that suppresses growth of prostate and breast tumors through induction of apoptosis. *Cancer Res.*, **64**, 6603–6609.
- Qi, H., Lin, C.-P., Fu, X., Wood, L.M., Liu, A.A., Tsai, Y.-C., Chen, Y., Barbieri, C.M., Pilch, D.S. and Liu, L.F. (2006) G-quadruplexes induce apoptosis in tumor cells. *Cancer Res.*, **66**, 11808–11816.
- 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.
- Schultze, P., Macaya, R.F. and Feigon, J. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGT TGGTGTGGTTGG). *J. Mol. Biol.*, **235**, 1532–1547.
- Martino, L., Virno, A., Randazzo, A., Virgilio, A., Esposito, V., Giancola, C., Bucci, M., Cirino, G. and Mayol, L. (2006) A new modified thrombin binding aptamer containing a 5'-5' inversion of polarity site. *Nucleic Acids Res.*, **34**, 6653–6662.
- Phan, A.T., Kuryavyi, V., Ma, J.-B., Faure, A., Andréola, 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. USA*, **102**, 634–639.
- Huang, Y.C. and Sen, D. (2010) A contractile electronic switch made of DNA. *J. Am. Chem. Soc.*, **132**, 2663–2671.
- Chinnapen, D.J.F. and Sen, D. (2004) A deoxyribozyme that harnesses light to repair thymine dimers in DNA. *Proc. Natl Acad. Sci. USA*, **101**, 65–69.
- Wieland, M. and Hartig, J.S. (2006) Turning inhibitors into activators: A hammerhead ribozyme controlled by a guanine quadruplex. *Angew. Chem. Int. Ed. Engl.*, **45**, 5875–5878.
- Tang, Z., Goncalves, D.P.N., Wieland, M., Marx, A. and Hartig, J.S. (2008) Novel DNA catalysts based on G-quadruplex recognition. *ChemBiochem*, **9**, 1061–1064.
- Cai, J., Shapiro, E.M. and Hamilton, A.D. (2009) Self-assembling DNA quadruplex conjugated to MRI contrast agents. *Bioconjug. Chem.*, **20**, 205–208.

22. Li, T., Wang, E.K. and Dong, S.J. (2009) Potassium-lead-switched G-quadruplexes: A new class of DNA logic gates. *J. Am. Chem. Soc.*, **131**, 15082–15083.
23. Alberti, P., Bourdoncle, A., Sacca, B., Lacroix, L. and Mergny, J.L. (2006) DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem.*, **4**, 3383–3391.
24. Majumdar, A., Kettani, A., Skripkin, E. and Patel, D.J. (2001) Pulse sequences for detection of NH₂...N hydrogen bonds in sheared G. A mismatches via remote, non-exchangeable protons. *J. Biomol. NMR*, **19**, 103–113.
25. Krishnan-Ghosh, Y., Liu, D.S. and Balasubramanian, S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.*, **126**, 11009–11016.
26. 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.
27. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment ¹⁵N, ¹³C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.
28. Phan, A.T. (2000) Long-range imino proton-¹³C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175–178.
29. Huang, X.N., Yu, P.L., LeProust, E. and Gao, X.L. (1997) An efficient and economic site-specific deuteration strategy for NMR studies of homologous oligonucleotide repeat sequences. *Nucleic Acids Res.*, **25**, 4758–4763.
30. Lim, K.W., Alberti, P., Guédin, A., Lacroix, L., Riou, J.-F., Royle, N.J., Mergny, J.-L. and Phan, A.T. (2009) Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.
31. Xu, Y., Noguchi, Y. and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)₃] in K⁺ solution. *Bioorgan. Med. Chem.*, **14**, 5584–5591.
32. Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.*, **35**, 6517–6525.
33. Chow, S.A., Vincent, K.A., Ellison, V. and Brown, P.O. (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723–726.
34. Do, N.Q., Lim, K.W., Teo, M.H., Heddi, B. and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.