<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Engineering of interlocked DNA G-quadruplexes as a robust scaffold</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Phan, Anh Tuân; Do, Ngoc Quang</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/9982">http://hdl.handle.net/10220/9982</a></td>
</tr>
</tbody>
</table>

© 2012 The Author(s). This paper was published in Nucleic Acids Research and is made available as an electronic reprint (preprint) with permission of The Author(s). The paper can be found at the following official DOI: [http://dx.doi.org/10.1093/nar/gks1304](http://dx.doi.org/10.1093/nar/gks1304). One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.
Engineering of interlocked DNA G-quadruplexes as a robust scaffold

Anh Tuấn Phan¹,* and Ngoc Quang Đô¹,²

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 \( G\text{GG}G\text{GG}G\text{G} \) 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 \textit{in vitro} selection (9,10) or rational design (11–13) and capable of forming G-quadruplexes (14–16), possess interesting biological activities, such as anticagulant (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 \textit{93del} sequence \( d(GGGGTGGGGAGGAGGGT) \), 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
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 $^1$H. Spectra were recorded at 25°C, unless otherwise specified. Experiments in H$_2$O 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 $^2$H labeling, through-bond correlations at natural abundance ([$^{13}$C,$^1$H] JRHMBC, [$^{13}$C,$^1$H] HSQC and [$^1$H,$^1$H] 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, IN$^{30-212}$ (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 μg/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 $s_2$ or $s_4$ (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).
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

Table 1. Sequences used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>93del</td>
<td>GGGGTGGGAGGAGGGT</td>
</tr>
<tr>
<td>s3</td>
<td>GGGGTGGGTGGTGAGGGT</td>
</tr>
<tr>
<td>s2</td>
<td>GGGGTGGGTGGTGAGGGT</td>
</tr>
<tr>
<td>s2-A</td>
<td>GGGGTGGTGGTGAGGGT</td>
</tr>
<tr>
<td>s4</td>
<td>GGGGTGGTGAGGGGTGGT</td>
</tr>
<tr>
<td>s4-A</td>
<td>GGGGTGGTGAGGGGTGGT</td>
</tr>
<tr>
<td>s4-Br</td>
<td>GGGGTGGTGAGGGGTGGT</td>
</tr>
<tr>
<td>s3R</td>
<td>TGGGTGGTGAGGGTGAGGGG</td>
</tr>
<tr>
<td>s2R</td>
<td>TGGGTGGTGAGGGTGAGGGG</td>
</tr>
<tr>
<td>s3R</td>
<td>TGGGTGGTGAGGGTGAGGGG</td>
</tr>
<tr>
<td>s3F</td>
<td>GGGGTGGGTGGGTGGG</td>
</tr>
</tbody>
</table>

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).
were unambiguously assigned by using site-specific low-enrichment $^{15}$N labeling, site-specific $^2$H 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-$^1$H]-HMBC and $^{13}$C-$^1$H]-HSQC) (data not shown), the classical H8/H6-H1' NOE sequential connectivity could be traced for four G-stretches (Figure 3). The intensity of intrarresidue 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$_2$O 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-tracks (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 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 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(G3G6G10G14) 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).
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 BrG, favoring a syn conformation in G-quadruplexes (31,32), without altering the general folding topology.

**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**


