<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>G-quadruplex structure of an anti-proliferative DNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Do, Ngoc Quang; Chung, Wan Jun; Truong, Thi Hong Anh; Heddi, Brahim; Phan, Anh Tuân</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/42925">http://hdl.handle.net/10220/42925</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2017 The Author(s). Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact <a href="mailto:journals.permissions@oup.com">journals.permissions@oup.com</a></td>
</tr>
</tbody>
</table>
G-quadruplex structure of an anti-proliferative DNA sequence

Ngoc Quang Do1,2,†, Wan Jun Chung1,†, Thi Hong Anh Truong1, Brahim Heddi1 and Anh Tuân Phan1,*

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

Received February 23, 2017; Revised March 28, 2017; Editorial Decision March 29, 2017; Accepted May 24, 2017

ABSTRACT

AGRO100 (also known as AS1411) is a G-rich oligonucleotide that has long been established as a potent anti-cancer aptamer. However, the structure of AGRO100 remained unresolved, due to the co-existence of multiple different G-quadruplex conformations. We identified a DNA sequence named AT11, derived from AGRO100, which formed a single major G-quadruplex conformation and exhibited a similar anti-proliferative activity as AGRO100. The solution structure of AT11 revealed a four-layer G-quadruplex comprising of two propeller-type parallel-stranded subunits connected through a central linker. The stacking between the two subunits occurs at the 3′-end of the first block and the 5′-end of the second block. The structure of the anti-proliferative DNA sequence AT11 will allow greater understanding on the G-quadruplex folding principles and aid in structural optimization of anti-proliferative oligonucleotides.

INTRODUCTION

Guanine-rich oligonucleotides are capable of folding into four-stranded structures called G-quadruplexes in the presence of cations (1). A folded intramolecular G-quadruplex consists of two main elements: core and loops (2–4). The core comprises of stacked G-G-G-G tetrad (or G-tetrad) layers, while the loops are linker sequences connecting the strands of the G-tetrad core. G-quadruplex structures are highly polymorphic, depending on the relative orientations of strands and types of loops (1–4). G-quadruplex polymorphism was also recently extended to motifs with bulges, duplex stems and missing guanines (5–7). In many cases, a single oligonucleotide could adopt multiple G-quadruplex conformations (4).

Over the past decades, numerous G-quadruplex-forming oligonucleotides have been reported as aptamers for a wide range of targets, such as thrombin (8–10), HIV (11–19) and cancer cells (20–25). The crystal structure of the complex between thrombin and a DNA aptamer (10) showed that the G-quadruplex formation was required for the recognition and biological activity of the oligonucleotide.

Most notably, AGRO100 (also known as AS1411), a G-rich oligonucleotide (sequence, GGTGGTGGTGGTG TGTTGGTGGTGGG), first discovered by Bates et al. (20), has garnered considerable research interest in the field of anticancer therapeutics (26,27). AGRO100 was reported to have inhibition activity against a large number of malignant cell lines with GI50 in the low micromolar range (26) and underwent Phase II clinical trials on renal cell carcinoma (28). The binding of AGRO100 to nucleolin, a protein participating ribosomal biogenesis (29), was proposed to be crucial for the anticancer activity of this oligonucleotide (30). Recently, AGRO100 has been reported to possess inhibitory activity against HIV (18,19).

Previous reports showed that AGRO100 adopted a mixture of multiple G-quadruplexes in solution, hindering structural studies (31). Using size-exclusion chromatography, Dailey et al. suggested the presence of at least eight species in this mixture (31). We previously determined the structure of a DNA sequence modified from AGRO100, revealing the formation of a left-handed G-quadruplex conformation (32).

In this study, we have identified a DNA sequence named AT11, derived from AGRO100, which formed a single major G-quadruplex conformation and exhibited a similar anti-proliferative activity as AGRO100. The solution structure AT11 revealed a novel G-quadruplex fold consisting of two propeller-type subunits. The structure of AT11 provides the basis for rational structural design and optimization of anti-proliferative oligonucleotides.

†To whom correspondence should be addressed. Tel: +65 6514 1915; Fax: +65 6795 7981; Email: phantuan@ntu.edu.sg
†These authors contributed equally to the paper as first authors.

Present addresses:
Ngoc Quang Do, Hanoi University of Pharmacy, Vietnam.
Brahim Heddi, Laboratoire de Biologie et de Pharmacologie Appliquée, UMR 8113 CNRS, ENS Paris-Saclay, Université Paris Saclay, France

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
MATERIALS AND METHODS

DNA sample preparation

Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized on an ABI 394 DNA/RNA synthesizer using products from Glen Research and Cambridge Isotope. DNA was purified by a reverse-phase PolyPak cartridge (Glen Research). Samples were then dialyzed successively against 25 mM KCl solution and against water. DNA oligonucleotides were dissolved in solution containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0). DNA concentrations were determined by measuring the UV absorbance at 260 nm.

Cell proliferation assay

The following sequences, AT11, d(TGG TGG TGG TTG TGG TGG TGG TGG T), AGRO100, d(GGT GGT GGT GGT TGT GGT GGT GGT G) and a control sequence CRO (20), d(CCT CCT CCT TCT CCT CCT CCT CCT), were used in the cell proliferation assay. Oligonucleotides were purchased from Integrated DNA Technologies (Singapore), then dialyzed against water, followed by freeze-drying. The dried samples were dissolved in a buffer containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0) to a final DNA stock concentration of 0.5 mM and stored at −20°C for further usage.

Human cancer cell line A549 and human fibroblast cell line WI-38 were obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). Cells were cultured in complete medium containing Dulbecco’s modified Eagle’s medium (high glucose, GlutaMAX™ supplement, pyruvate) (Gibco), 10% fetal bovine serum (Gibco), 1% Penicillin-Streptomycin (Sigma). Cell proliferation assay kit was purchased from ATCC (Rockville, MD, USA) including MTT reagent (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide) and stop detergent reagent solution.

In cell proliferation assay, cells were cultured in a 96-well plate at a low density of 1000 and 2000 cells per well for A549 and WI-38, respectively. Each well contained 100 μl complete medium and incubated at 37°C, 5% CO2 for 18–20 h to allow cell adherence and growth. The medium was then replaced with a new complete medium containing an oligonucleotide at different concentrations (1, 5, 10 and 20 μM) followed by 5 days incubation at 37°C, 5% CO2. Complete medium without additional oligonucleotides was used as control. In cell proliferation assay, 10 μl MTT was added to each well and incubated in 37°C for 4 h. The reaction was stopped by adding 100 μl stop detergent reagent solution and incubated at 37°C overnight to solubilize formazan crystal. Infinite 200 PRO Microplate Reader (Tecan, Switzerland) was used to record the absorbance at 570 nm, while the reference wavelength was 650 nm. Each experiment was repeated three times to have an average value of the measurements.

Circular dichroism

Circular dichroism (CD) spectra were recorded at 25°C using a JASCO-815 spectropolarimeter with a 1-cm path length quartz cuvette containing a solution volume of 500 μl. Spectra (220–320 nm) were taken with a scan speed of 200 nm/min. DNA (~5 μM) was dissolved in a buffer containing 70 mM KCl and 20 mM potassium phosphate (pH 7.0). For each measurement, an average of three scans was taken and the spectral contribution of the buffer was subtracted.

UV melting experiments

Melting experiments were performed on a JASCO V-650 UV-visible spectrophotometer (Tokyo, Japan). DNA concentration was ~5 μM. Melting curves were monitored by the UV absorbance at 295 nm (33). Melting temperatures (Tm) reported represent an average over heating and cooling cycles.

NMR spectroscopy

Nuclear magnetic resonance (NMR) experiments were performed on 600 and 700 MHz Bruker spectrometers at 25°C, unless otherwise stated. NMR spectra were recorded either directly or after following an annealing procedure (heating the sample at high temperature (~95°C) and slowly cooling down to room temperature). The concentration of DNA samples was typically 0.1–2.0 mM. Solution contained 70 mM KCl and 20 mM potassium phosphate (pH 7.0). Spectra analyses were performed using SpinWorks software (http://home.cc.umanitoba.ca/~wolowiec/spinworks/), FE-LIX (Felix NMR, Inc.) and SPARKY (34) programs.

NMR-restrained structure calculation

The structure of AT11 was calculated using the XPLOR-NIH program (35). NMR-restrained computations were performed as described below. Structures were displayed using the PyMOL program (36).

NOE distance restraints. Distances between the non-exchangeable protons of AT11 were obtained from NOESY experiments at various mixing times (100 and 300 ms) recorded in D2O, those involving exchangeable protons were obtained from NOESY spectrum (mixing time, 200 ms) recorded in H2O solution. The NOE cross-peaks were classified as strong, medium and weak and given corresponding distance restraints as previously described (32).

Dihedral restraints. The glycosidic bonds of all 16 guanines were restrained according to their anti conformation. The glycosidic χ torsion angle for experimentally determined anti guanine residues were fixed at 240° ± 40°.

Hydrogen-bond and planarity restraints. Hoogsteen hydrogen bonds between guanines were restrained using H21–N7, N2–N7, H1–O6 and N1–O6 distances, which were set to 1.99 ± 0.1, 2.9 ± 0.1, 1.95 ± 0.1, 2.95 ± 0.1Å respectively. Planarity restrained was used for the G2-G5-G8-G11, G3-G6-G9-G12, G15-G18-G21-G24 and G17-G20-G23-G26 tetrads.
Repulsive restraints. Repulsive restraints (4.0–20.0 Å) were applied on pairs of protons that do not exhibit cross-peaks in a well-resolved region of the NOESY spectrum. A total of 17 repulsive restraints involving thymine residues were used: 9 between sugar protons and methyl protons, 1 between an aromatic proton and a methyl proton, 4 between sugar protons and T(H6) protons and 3 between sugar protons.

Distance geometry simulated annealing. An initial extended conformation of AT11 was generated using the XPLOR program. The system was then subjected to distance geometry simulated annealing by incorporating the hydrogen-bond, distance, dihedral, planarity and repulsive restraints. A total of 100 structures were generated and subjected to further refinements.

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 5 ps and allowed to equilibrate for 1 ps, during which force constants for the distance restraints were kept at 2 kcal.mol⁻¹.A⁻². The force constants for non-exchangeable proton, exchangeable proton and repulsive distance restraints were then increased to 64 kcal.mol⁻¹.A⁻² in 26 ps. Next, the system was cooled down to 300 K in 14 ps, after which an equilibration was performed for 10 ps. The coordinates were saved every 0.5 ps during the last 40 ps and averaged. The average structure obtained was then subjected to minimization until the gradient of energy was <0.1 kcal.mol⁻¹. Dihedral (50 kcal.mol⁻¹.rad⁻²) and planarity (1 kcal.mol⁻¹.A⁻² for tetrad) restraints were maintained throughout the course of refinement. Ten lowest-energy structures were chosen.

Data deposition

The coordinates for the AT11 G-quadruplex have been deposited in the Protein Data Bank (accession code 2N3M).

RESULTS

Favoring a single G-quadruplex conformation from AGRO100 by sequence modifications

We recorded NMR and CD spectra of AGRO100 in K⁺ containing solution (Figure 1). The imino proton NMR spectrum of AGRO100 displayed multiple humps and peaks at 10.5–12.0 ppm (Figure 1A), indicating the presence of multiple G-quadruplex species, consistent with a previous report showing that AGRO100 could adopt at least eight different G-quadruplex conformations (31). CD spectrum of AGRO100 showed a minimum at 240 nm and a positive peak at 260 nm (Figure 1B), characteristic of the formation of parallel-stranded G-quadruplexes.

The complexity of NMR spectra of AGRO100 prevented its further structural elucidation. Sequence modifications were performed to favor a single G-quadruplex conformation and simplify the NMR spectra. The sequence AT11, derived from AGRO100 by an addition of thymine nucleotides to both 5' and 3' ends of the DNA sequence and a single G-to-T modification at position 11, exhibited a well-resolved NMR spectrum consistent with the formation of a single major G-quadruplex conformation (Figure 1). Addition of a thymine nucleotide to the 5' and 3' ends of the sequence was aimed to prevent stacking of G-quadruplexes at the 5'/3'-end surfaces which might lead to aggregation (15,37–40), while a G-to-T substitution reduced the number of guanines in the sequence from 17 to 16 with the aim to favor a single G-quadruplex fold (Supplementary Figure S1). Indeed, the imino proton spectrum of AT11 showed 16 peaks at 11.0–11.7 ppm, consistent with the formation of a G-quadruplex containing four G-tetrad layers. The same major conformation was formed by AT11 in the presence of varying K⁺ concentration (Supplementary Figure S2). The CD profile of AT11 was similar to that of AGRO100, showing a minimum at 240 nm and a positive peak at 260 nm (Figure 1C), suggesting the formation of a right-handed parallel-stranded G-quadruplex.

AT11 exhibits anti-proliferative activity

We tested the effect of different oligonucleotides on the proliferation of the human lung cancer cells A549. Both G-quadruplex-forming sequences AGRO100 and AT11 inhibited the proliferation of the A549 cells over a range of oligonucleotide concentrations from 5 to 20 μM, display-
Figure 3. NMR spectral assignments of AT11. (A) Examples of the assignment of guanine imino protons through site-specific 15N-labeling. (B) Examples of the assignment of guanine aromatic protons through site-specific deuterium labeling. Labeled guanine residues were indicated with their corresponding residue number. (C) Through-bond correlations between guanine imino and aromatic protons at natural abundance, using long-range J-couplings. Weak/missing signals are framed in boxes. DNA concentration was 0.2–0.5 mM for (A and B) and ~1 mM for (C).

NMR structure of AT11 revealed a G-quadruplex with two parallel-stranded subunits

Spectral assignment. Unambiguous assignment of the imino and aromatic protons of AT11 was performed using site-specific low-enrichment (4 or 2%) 15N-labeling of guanines and site-specific deuterium labeling of guanines, and through-bond correlation experiment (1H-13C-HMBC) (Figure 3 and Supplementary Figure S4). The remaining protons of AT11 were assigned accordingly using data obtained from through-bond (TOCSY, COSY, 1H-13C-HSQC and through-space (NOESY) correlation experiments (Supplementary Figure S5).

Folding topology of AT11. The folding topology of AT11 was determined by analyzing NOE cross-peaks between the imino (H1) and aromatic (H8) protons of the guanine bases (Figure 4A). From H1-H8 NOEs, the alignments of four G-tetrads, G2-G5-G8-G12, G3-G6-G9-G15, G17-G20-G23-G26 and G18-G21-G24-G27, were established (Figure 4). Solvent exchange data showed that imino proton peaks of the G3-G6-G9-G15 and G17-G20-G23-G26 tetrads were more protected than the counterparts from the other G-tetrads (Supplementary Figures S5 and S7). All together the data indicated the formation of a four-layer G-quadruplex structure as shown in Figure 4. The structure comprises of two 2-layered propeller-type parallel-stranded subunits connected by a single-residue linker (T16). H8/H6-H1’ NOE sequential connectivity could be traced from T1 to T28, including the T1-G2-G3, G5-G6, G8-G9, T11-G12, G17-G18, G20-G21, G23-G24 and G26-G27-T28 segments, with several broken connections due to the presence of propeller loops, bulge and linker (Supplementary Figure S5). All guanines show weak intra-residue H8-H1’ NOEs comparing to strong intra-residue thymine H6-CH3 NOEs, reflecting their anti glycosidic conformation. A two-residue bulge forms between G12 and G15 of the first subunit. The two subunits exhibit the 3’-to-5’ stacking mode (Figure 4B). Such an arrangement is supported by solvent exchange data and several NOEs observed at the interface between the two subunits, including: G20(H8)-G3(H1’), G23(H8)-G6(H1’), G26(H8)-G9(H1’) and G17(H8)-G15(H1’) (Supplementary Figures S5 and S7). The parallel-stranded folding topology of the AT11 G-quadruplex is consistent with its CD data (see above).

Solution structure of AT11. Following the unambiguous assignment of AT11 resonances, the solution structure of AT11 complex was computed (Table 1). The lowest-energy refined structures are presented in Figure 5. AT11 consists of two parallel-stranded propeller-type G-quadruplex...
Figure 5. (A and C) Ten superimposed lowest-energy solution structures and (B and D) a representative structure of AT11. (A and B) Side view. (C and D) Top view.

Figure 6. (A) AT11 consists of two propeller-type G-quadruplex subunits. (B) Stacking interface between the two subunits of AT11.

blocks linked via a thymine residue (T16). The top G-quadruplex block contains three propeller loops (T4, T7 and T10–T11) and a bulge (T13–T14), while the bottom G-quadruplex block contains three single-nucleotide propeller loops (T19, T22 and T25). All guanine residues adopt the anti conformation. Partial overlap between the five- and six-membered rings of the guanine bases across the interface of the two G-quadruplex blocks (G3/G20, G6/G23, G9/G26 and G15/G17) (Figure 6) was observed to be similar to that between the two continuous G-tetrad layers.

Effects of bulge and linker size on structure and thermal stability

Linker/bulge/loop elements in the AT11 G-quadruplex include five single-residue (T4, T7, T19, T22 and T25) and one two-residue (T10–T11) propeller loops, a two-residue bulge (T13–T14) and a one-residue linker between the two subunits (T16). To analyze the relationship between bulge/linker size and the thermal stability of AT11, modifications were performed on AT11 by varying the bulge/linker size from zero to two thymines (Table 2). It was observed that by decreasing either the bulge or linker size of AT11, an increase in thermal stability was achieved and vice versa, consistent with previous studies (5). For example, the sequence AT11-B0, a derivative of AT11 with no bulge (or

Table 1. NMR restraints and structure statistics

<table>
<thead>
<tr>
<th>Distance restraints</th>
<th>Exchangeable</th>
<th>Non-exchangeable</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA restraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-residue</td>
<td>0</td>
<td>141</td>
</tr>
<tr>
<td>Inter-residue</td>
<td>47</td>
<td>171</td>
</tr>
<tr>
<td>Other Restraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen-bond restraints</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Torsion angle restraints</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Repulsive restraints</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

| Table 2. NOE violations, Deviations from standard geometry, Improper angles, Pairwise all heavy atom r.m.s.d. values, G-tetrad core

<table>
<thead>
<tr>
<th>NOE violations</th>
<th>Deviations from standard geometry</th>
<th>Improper angles</th>
<th>Pairwise all heavy atom r.m.s.d. values</th>
<th>G-tetrad core</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers (&gt;0.2 Å)</td>
<td>0.000 ± 0.000</td>
<td>0.004 ± 0.000</td>
<td>0.709 ± 0.016</td>
<td>0.378 ± 0.007</td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>1.611 ± 0.256</td>
<td></td>
<td></td>
<td>1.010 ± 0.210</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td></td>
<td>0.000 ± 0.000</td>
<td>0.000 ± 0.000</td>
<td>0.000 ± 0.000</td>
</tr>
<tr>
<td>Improper (°)</td>
<td></td>
<td></td>
<td></td>
<td>0.378 ± 0.007</td>
</tr>
<tr>
<td>Pairwise all heavy atom r.m.s.d. values (Å)</td>
<td>1.611 ± 0.256</td>
<td></td>
<td></td>
<td>1.010 ± 0.210</td>
</tr>
<tr>
<td>G-tetrad core</td>
<td></td>
<td></td>
<td></td>
<td>0.378 ± 0.007</td>
</tr>
</tbody>
</table>
bulge size zero), exhibits an increase in the melting temperature by 16 °C compared to that of AT11.

**DISCUSSION**

5′-3′ versus 5′-5′ stacking between propeller-type G-quadruplex subunits

In a previous study by Matsugami et al. (41), the NMR structure of a G-quadruplex adopted by the d(GGA)$_8$ sequence in K$^+$ solution was reported. Both the d(GGA)$_8$ (41) and AT11 sequence (this work) fold into similar G-quadruplexes, comprising of two parallel-stranded blocks connected by a single-residue linker. However, in the d(GGA)$_8$ structure, the stacking of two blocks occurs between two 5′-end surfaces, while in AT11, the stacking occurs between the 3′-end of the first block and the 5′-end of the second block. Formation of a G-A-G-A-G-A-G heptad at the stacking interface of d(GGA)$_8$ may be the reason for this difference in the stacking mode.

Right-handed versus left-handed G-quadruplexes

Recently, we reported on a left-handed G-quadruplex called Z-G4 (32) formed by another sequence derived from AGRO100. Like AT11, Z-G4 also forms a structure consisting of two parallel-stranded two-layer G-quadruplex subunits. However, unlike AT11, Z-G4 involves the stacking between the 5′-ends of the two subunits and, particularly, the strand progression in the left-handed orientation.

Structure-activity relationship

It has been suggested that the cancer-selective anti-proliferative activity of G-rich oligonucleotides is associated with their interaction with proteins such as nucleolin, a protein abundant on the surface of many cancer cells (30). On the other hand, a recent study suggested that the anti-proliferative activity of G-rich oligonucleotides is mainly contributed by the guanine-containing degradation products (42). Further studies will be required to answer the question of whether and how the G-quadruplex structures of G-rich oligonucleotides contribute to their cellular uptake and anti-proliferative activity.

**CONCLUSION**

The AT11 sequence, derived from AGRO100 by a simple base substitution, was shown to favor a single major G-quadruplex conformation and exhibit a similar anti-proliferative activity as AGRO100. We have determined the structure of AT11, which adopts a four-layer G-quadruplex, which comprises of two propeller-type parallel-stranded subunits. The stacking between the two subunits occurs at the 5′-end of the first block and the 5′-end of the second block.

**ACCESSION NUMBER**

PDB ID: 2N3M.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

Singapore Ministry of Education Academic Research Fund Tier 2 [MOE2015-T2-1-092]; National Research Foundation Investigatorship [NRF-NRFI2017-09]; Nanyang Technological University (to A.T.P.). Funding for open access charge: Singapore Ministry of Education Academic Research Fund Tier 2 [MOE2015-T2-1-092].

Conflict of interest statement. None declared.

**REFERENCES**


34. Goddard, T.D. and Kneller, D.G. *SPARKY 3*, University of California, San Francisco.


