HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges

Vineeth Thachappilly Mukundan¹, Ngoc Quang Do¹,² and Anh Tuan Phan¹,*

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

Received January 19, 2011; Revised June 10, 2011; Accepted June 13, 2011

ABSTRACT

T30177 is a G-rich oligonucleotide with the sequence (GTGGTGGGTGGGTGGGT) which inhibits the HIV-1 integrase activity at nanomolar concentrations. Here we show that this DNA sequence forms in K⁺ solution a dimeric G-quadruplex structure comprising a total of six G-tetrad layers through the stacking of two propeller-type parallel-stranded G-quadruplex subunits at their 5’-end. All twelve guanines in the sequence participate in the G-tetrad formation, despite the interruption in the first G-tract by a thymine, which forms a bulge between two adjacent G-tetrads. In this work, we also propose a simple analytical approach to stoichiometry determination using concentration-dependent melting curves.

INTRODUCTION

G-quadruplexes (1–5) are four-stranded structures formed by G-rich nucleic acid sequences. Formation of G-quadruplexes can affect a wide range of biological activities including telomere maintenance, gene regulation, transcription, DNA replication and repair (6). G-quadruplexes formed by synthetic oligonucleotides may possess anticancer (7–9) and anti-HIV (10–20) activity.

The basic unit of a G-quadruplex is a G-tetrad, a cyclic planar arrangement of four guanines linked by hydrogen bonds (21). The stability of a G-quadruplex depends on the stacking between different G-tetrads as well as electrostatic interaction mediated by cations, such as K⁺ or Na⁺, which are located at the center of the G-tetrad core. In most of the G-quadruplex structures reported so far, the G-tetrad core is formed by tracts of consecutive guanines (1–5). Formation of bulges has been observed in a G-quadruplex, where G-tracts are interrupted by non-guanine residues (22).

A series of G-rich DNA oligonucleotides (12–17) and their backbone-modified counterparts (11,12,23–25) have been reported to inhibit the activity of HIV-1 integrase (IN), an enzyme which is responsible for the integration of viral DNA into host genome (26). These include T30695 (or T30923) with the sequence (GGGTGGGTGGTGGGT) (14,15) and T30177 (or I100-15) with the sequence (GTGGTGGGTGGTGGGT) (11–13). These sequences were proposed to adopt a chair-type anti-parallel-stranded G-quadruplex with two G-tetrads and three edgewise (T-G) loops (12,14,27). However, this proposed folding topology is at variance with the reported CD profile showing a positive peak at 260 nm (14,28), which is characteristic of a parallel-stranded G-quadruplex (29), and the mass spectrometry detection of a dimer for these sequences (28,30). Furthermore, a related G-rich oligonucleotide 93del was determined by NMR to form an interlocked dimer comprising two parallel-stranded G-quadruplex subunits (17); sequences with GGG tracts separated by single-residue linkers were shown to form stable parallel-stranded G-quadruplexes in both experimental (31–37) and computational (33,38) studies.

In the accompanying paper (39), the T30695 oligonucleotide with the sequence (GGGTGGGTGGTGGGT) was determined to form a dimeric structure by the stacking of two propeller-type parallel-stranded G-quadruplex subunits, in which all G-tracts participate in the G-tetrad core formation. In this work, we investigate the structure of the G-rich oligonucleotide HIV-1 integrase inhibitor T30177, which differs from T30695 by the presence of a T residue that interrupts the first G-tract. Our results showed that T30177 forms a stacked dimeric G-quadruplex structure containing bulges. The knowledge of these structures might help us to uncover their inhibition mechanism against HIV-1 integrase. This work used a wide range of biophysical techniques including gel electrophoresis, CD, UV and NMR spectroscopy. We also propose a simple analytical approach to stoichiometry determination using concentration-dependent melting curves.

*To whom correspondence should be addressed. Tel: +65 6514 1915; Fax: +65 6794 1325 Email: phantuan@ntu.edu.sg

© The Author(s) 2011. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
MATERIALS AND METHODS

Sample preparation

Unlabeled and site-specific labeled DNA oligonucleotides were chemically synthesized using products from Glen Research and Cambridge Isotope Laboratories. Samples were purified following Glen Research’s protocol and then dialyzed successively against KCl solution and water. DNA oligonucleotides were dissolved in 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 absorption coefficients of the unfolded species (40).

NMR spectroscopy

NMR experiments were performed on 600 and 700 MHz NMR Bruker spectrometers. Guanine resonances were assigned by using site-specific 15N and 2H labeling (41,42) and through-bond correlations at natural abundance (43). Spectral assignments were completed by using NOESY, COSY, TOCSY and HSQC spectra (44). G-quadruplex folding topology was determined based on interproton distances obtained from NOESY experiments.

Gel electrophoresis

Electrophoresis experiment was performed at 120 V on native gels containing 20% polyacrylamide (Acrylamide: Bis-acrylamide = 37.5:1) in TBE buffer (89 mM Tris–borate, 2 mM EDTA, pH 8.3) supplemented with 3 mM KCl. Each sample contained 5 µg DNA. Gel was viewed by UV shadowing.

Circular dichroism

CD spectra were recorded on a JASCO-815 spectropolarimeter using 1 cm path-length quartz cuvette in a reaction volume of 600 µl at 20°C. Scans from 220 to 320 nm were performed with 200 nm/min, 1-nm pitch and 1-nm bandwidth. DNA concentration was 4–6 µM.

UV melting experiments

The stability of the G-quadruplex structure is measured in UV melting experiments conducted on a JASCO V-650 spectrophotometer. Absorbance at 295 nm was recorded as a function of temperature ranging from 30 to 90°C. The heating and cooling rates were 0.2°C/min. Experiments were performed with quartz cuvettes, with 1-cm path length. DNA concentration ranged from 0.5 to 200 µM. Solution contained 70 mM (or 30 mM) KCl and 20 mM potassium phosphate (pH 7.0).

RESULTS AND DISCUSSION

T30177 and T30177-I11 form G-quadruplexes in K+ solution

One-dimensional imino proton spectrum of T30177 in K+ solution (Figure 1a), similar to the one reported earlier (12), shows peaks in the range of 10.8–11.5 ppm, indicating the formation of a G-quadruplex. However, the heavy overlapping of these peaks makes further structural analysis difficult. This could be due to the structural symmetry, which arises from the repetitive nature of the sequence. We found that the DNA sequence with a single guanine-to-inosine substitution at position 11 showed greatly improved NMR spectra, and selected this sequence (henceforth designated as T30177-I11, Table 1) for further structural analysis. Similar spectral behavior was also observed (Supplementary Figure S1) for many different DNA sequences containing a single guanine-to-inosine substitution (Supplementary Table S1).

Imino proton spectrum of T30177-I11 in K+ solution (Figure 1b) shows peaks at 10.8–11.5 ppm corresponding to eleven guanine imino protons and an additional peak at...
13.8 ppm corresponding to the imino proton of the inosine. This indicates the involvement of all 12 guanine and inosine bases in the G-tetrad formation, in contrast to only eight guanines for the previously proposed structure (12). Both T30177 and T30177-III give similar CD spectra (Figure 2) with a positive peak at 260 nm, which is characteristic of parallel-stranded G-quadruplexes (29).

A simple method for stoichiometry determination: T30177-III forms a dimeric G-quadruplex

The temperature-driven folding/unfolding of T30177-III was monitored by the UV absorbance at 295 nm (45) (Figures 3a and 4a). The 295-nm absorbance decreased with an increase of temperature, as typically observed for G-quadruplexes. The heating and cooling curves were almost superimposed indicating near equilibrium processes. The melting temperature ($T_m$) was dependent on DNA concentration: in ~100 mM K⁺ solution, $T_m$ increased from 68 to 76°C when the DNA concentration increased from 5 to 150 μM (Figure 3a); in ~60 mM K⁺ solution, $T_m$ increased from 63 to 73°C when the DNA concentration increased from 0.5 to 200 μM (Figure 4a). These results indicated the formation of a multimeric G-quadruplex.

Consider the equilibrium between a monomer and a multimer ($n$-mer):

$$n[C] \rightleftharpoons [C_n]$$

where [C] and [Cₙ] are the concentrations of the monomer and the $n$-mer respectively. The equilibrium concentrations of the two species are related via the dissociation constant ($K_d$) by:

$$K_d = [C]^n/[C_n]$$

The plot of the concentration of the $n$-mer against the concentration of the monomer in log scales should be a straight line with the slope $n$ (44).

Here we propose a simple method to quantitatively analyze the concentration-dependent melting curves for stoichiometry determination. For a given temperature where the structural transition was observed (e.g. 70, 74 and 78°C, Figure 3a), concentrations of the folded and unfolded species were determined for each melting curve. The concentration of the folded species ($n$-mer) was

Table 1. DNA sequences used in this work

| Name      | Sequence (5’–3’)       |
|-----------|------------------------|
| T30177    | G T GG T GGG T GGG T  |
| T30177-III| G T GG T GGG T GGG T  |
| T30177-TT | TT G T GG T GGG T GGG |
| T30177-III-TT | TT G T GG T GGG T GGG T |

Figure 2. CD spectra of T30177 and T30177-III in K⁺ solution at 20°C.

Figure 3. UV-melting data of T30177-III. (a) Cooling and heating curves for a particular DNA concentration are shown in one color, with continuous and dashed lines respectively. Color codes are shown at the top right corner. (b) Plots of multimer concentrations against monomer concentrations at three different temperatures. Data were obtained from the melting curves in (a) with filled and open points corresponding to cooling and heating processes, respectively. Lines of slope 2 are drawn through the points. Samples contained 70 mM KCl and 20 mM potassium phosphate (pH 7.0).
plotted against the concentration of the unfolded species (monomer) in log scales (Figure 3b, Supplementary Figures S2, S3). The slope 2 would fit well the data points in ~100 mM K+ solution at each temperature (Figure 3b) [the slope of the best fits ranged from 1.74 to 2.16 (Supplementary Figure S2)], indicating the formation of a dimer by T30177-I11. At 70°C, the dissociation constant $K_d$ of this dimer is ~10 μM in this condition.

Quantitative analysis of concentration-dependent melting curves for stoichiometry determination has been previously reported (46,47). However, the proposed analysis was relatively complicated: the authors theoretically derived the analytical expression for the variation of $T_m$ as a function of DNA concentration, i.e. using the interception points of a horizontal line at 0.5 with the melting curves; furthermore, measurements of the enthalpy ($\Delta H$) associated with the transition were required for the stoichiometry determination. In contrast, our approach used interception points of vertical lines with the melting curves, and the analysis was quite simple. To our knowledge, this approach is being used here for the first time. We should note that the application of both approaches for quantitative analysis of concentration-dependent melting curves requires the assumption that these melting curves are equilibrium processes for a transition between two states: one folded and one unfolded. In practice, the equilibrium transitions can be achieved when the interconversion between these states is fast compared to the heating/cooling rate of the spectrometer.

The described analysis is based on the assumption that the observed changes in the UV absorbance reflect the multimer–monomer transition. However, in the present example (see structure below) the changes in the 295-nm UV absorbance only directly reflect the G-quadruplex unfolding, which in turn may depend on the dimer–monomer transition. In the given example, at very low DNA concentrations, monomeric intramolecular G-quadruplex can exist at a higher abundance. This can explain why the observed data points for low DNA concentrations (~1 μM) deviated from the slope 2 towards a shallower slope (Supplementary Figure S3).

Gel electrophoresis (Figure 5) showed that the migration of T30177 and T30177-I11 was similar to that of the interlocked dimeric G-quadruplex 93del comprising six G-tetrad layers (17) and slower than that of a monomeric parallel-stranded propeller-type G-quadruplex GTERT-60 with three tetrad layers (48), consistent with the formation of a dimeric G-quadruplex by T30177 and T30177-I11. This result was corroborated by NMR data (discussion below).

NMR spectral assignments of T30177-I11

The downfield-shifted peak at 13.8 ppm was assigned to the imino proton of I11 (49). Guanine imino protons were unambiguously assigned to their respective positions in the sequence (Figure 6a) using the site-specific assignments.

![Figure 4](image1.png)

**Figure 4.** UV-melting data of (a) T30177-I11 and (b) T30177-I11-TT. Cooling and heating curves for a particular DNA concentration are shown in one color, with continuous and dashed lines respectively. Color codes are shown at the top right corner. Samples contained 30 mM KCl and 20 mM potassium phosphate (pH 7.0).

![Figure 5](image2.png)

**Figure 5.** Gel electrophoresis analysis of T30177 and its derivatives. Reference markers are 93del (d[GGGTGGGAGGAGGGT]), an interlocked dimeric G-quadruplex (17) and GTERT-60 (d[AGGGIAAGGGCTGGGAGGGC]), a monomeric propeller-type G-quadruplex (48). T30177-TT has two additional thymines at the 5'-end as compared to T30177.
low-enrichment approach (41), in which one guanine at a
time was 15N-labeled at 2% (Supplementary Table S2).
These assignments further confirmed that all guanines
and inosine in the sequence participated in G-tetrad for-
mation. Guanine H8 protons were assigned independently
by site-specific 2H substitutions at the H8 position of
guanines one at a time (42,50) (Supplementary Table
S2), which led to the disappearance of a single peak
corresponding to the substituted guanine (Figure 6b).

Determination of folding topology: T30177-I11 forms a
stacked dimeric G-quadruplex
Using the complete assignments of imino (H1) and H8
protons, the G-tetrad alignments were determined from
NOESY spectra based on the specific imino-H8
connectivities within a G-tetrad (Figure 7). For example,
we observed NOE cross-peaks between G4(H1) and
G8(H8), G8(H1) and G12(H8), G12(H1) and G16(H8),
and G16(H1) and G4(H8), which established the forma-
tion of the (G4-G8-G12-G16) tetrad. In the same manner,
we determined the arrangements of (G3-G7-I11-G15)
and (G1-G6-G10-G14) tetrads (Figure 7c).

Figure 8 shows a dimeric folding topology for
T30177-I11 that satisfies the established alignments of
the three G-tetrads. This is a dimeric G-quadruplex com-
prising two identical subunits of propeller-type parallel-
stranded G-quadruplexes each containing three G-tetrad
layers, three double-chain-reversal loops (T5, T9 and T13)
and a bulge (T2). The two subunits are stacked at their
5′-end; there could be various isomers, where the two
subunits are rotated with respect to each other about
the common central helical axis. However, the broadening
of peaks at the interface (see below) and the symmet-
ic nature of the structure prevented us from definite
determination of the orientation and the detailed
structure of the stacking interface. The stacking mode
shown in Figure 8 was proposed on the basis of the
stacked dimeric structure of the homologue sequence
T30695 (39).

This folding topology is consistent with the results
of a solvent-exchange experiment showing that imino
protons belonging to the central (G3/C15G7/C15I11/C15G15) and the 5’-end (G1/C15G6/C15G10/C15G14) tetrads are the most protected (Figure 9). The glycosidic conformations of all residues are anti, as shown by the moderate intensities of H1’-H8/6 NOE cross-peaks (Figure 7a), consistent with the formation of a parallel-stranded G-quadruplex (29). NOE cross-peaks between G1 and G3 (Figure 7a) indicated continuous stacking between these bases across the bulge (T2). Note that there might be a motion at the bulge as indicated by the broadening of the H8 proton of G3. The structure of a parallel-stranded G-quadruplex of T30177 with T2 being looped out from the G-tetrad core was recently reported to be stable in a molecular dynamics simulation (38).

Dimeric interface: motion and control of stacking between the monomers

In this section, we describe the nature and stability of the dimeric interface where the stacking between two monomers occurs. Different NMR spectra provided evidence for considerable motion in this region. We observed the broadening of imino protons corresponding to guanines 1, 6, 10 and 14 (Figure 9 and data not shown). Broadening was also observed for some non-exchangeable protons of these residues (Figure 7a). Since all these guanines are located at the dimeric interface, this clearly establishes motion in this region. Possible types of motion include inter-conversion between dimer and monomer or rotation of two subunits about the central axis.

We could remove the stacking between the two G-quadruplex monomers by the addition of two extra thymine bases at the 5’-end (sequence T30177-TT, Table 1). Gel electrophoresis experiments (Figure 5) clearly showed the difference between two structures: the monomer (T30177-TT) migrated much faster than the dimer.
(T30177 or T30177-I11). The monomeric nature of T30177-TT and T30177-I11-TT were supported by the independence of their melting temperature on the DNA concentration (ranging from 0.5 to 200 μM) (Figure 4b and data not shown). Our unpublished NMR data confirmed that T30177-TT forms a monomeric propeller-type parallel-stranded G-quadruplex in this condition.

CONCLUSION
G-rich oligonucleotide T30177 (or T30177-I11) forms a dimeric structure involving two subunits of propeller-type parallel-stranded G-quadruplexes, which are stacked at their 5′-end. All guanines in the sequence participate in G-tetrad formation and there is a bulge of a T residue in each subunit. This work along with other structural studies (39), pointed to the formation of dimeric parallel-stranded G-quadruplexes comprising a total of six G-tetrad layers by various G-rich oligonucleotides that possess HIV-1 integrase inhibition activity.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

FUNDING
Singapore Ministry of Education grants (ARC30/07 and RG62/07); Singapore Biomedical Research Council (grant 07/1/22/19/542 to A.T.P.). Funding for open access charge: Waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES
1. 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.
2. Phan, A.T., Kuryavyi, V. and Patel, D.J. (2006) DNA architecture: from G to Z. Curr. Opin. Struct. Biol., 16, 288–298.
3. 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.
4. 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.
5. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. Curr. Opin. Struct. Biol., 19, 239–250.
6. Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat. Struct. Mol. Biol., 13, 1055–1059.
7. Bates, P.J., Kahlon, J.B., Thomas, S.D., Trent, J.O. and Miller, D.M. (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. J. Biol. Chem., 274, 26369–26377.
8. Jing, T., Li, Y., Xiong, W., Sha, W., Jing, L. and Tewary, 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.
9. Qi, H., Lin, C.-P., Fu, X., Wood, L.M., Liu, A.A., Tsai, Y.-C., Chen, Y., Barbieri, C.M., Pitts, D.S. and Liu, L.F. (2006) G-quadruplexes induce apoptosis in tumor cells. Cancer Res., 66, 11808–11816.
10. Wyatt, J.R., Robinson, S.J., Buckheit, R.W., Limkinkai, T., Davis, P.W., Rayner, B., Imbach, J.L. and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. Proc. Natl Acad. Sci. USA, 91, 1356–1360.
11. Ojwang, J.O., Buckheit, R.W., Pommer, Y., Mazumder, A., De Vreese, K., Estep, J.A., Reymen, D., Pallansch, L.A., Lackman-Smith, C., Wallace, T.L. et al. (1995) T30177, an oligonucleotide stabilized by an intramolecular guanosine octet, is a potent inhibitor of laboratory strains and clinical isolates of human immunodeficiency virus type 1. Antimicrob. Agents Chemother., 39, 2426–2435.
12. Rando, R.F., Ojwang, J., Elbaggar, A., Reyes, G.R., Tinde, R., McGrath, M.S. and Hogan, M.E. (1995) Suppression of human...
