

NMR structural study on the self-trimerization of d(GTTAGG) into a dynamic trimolecular G-quadruplex assembly preferentially in Na⁺ solution with a moderate K⁺ tolerance

Haitao Jing^{1,2,†}, Wenqiang Fu^{1,2,†}, Wenxuan Hu^{1,2}, Suping Xu¹, Xiaojuan Xu^{1,2}, Miao He^{1,2}, Yangzhong Liu^② and Na Zhang^{①,3,4,5,*}

¹High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230031, China, ²University of Science and Technology of China, Hefei 230026, China, ³Key Laboratory of High Magnetic Field and Ion Beam Physical Biology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China, ⁴Key Laboratory of Anhui Province for High Field Magnetic Resonance Imaging, Hefei 230031, China and ⁵High Magnetic Field Laboratory of Anhui Province, Hefei 230031, China

Received August 03, 2020; Revised January 05, 2021; Editorial Decision January 07, 2021; Accepted January 15, 2021

ABSTRACT

Vast G-quadruplexes (GQs) are primarily folded by one, two, or four G-rich oligomers, rarely with an exception. Here, we present the first NMR solution structure of a trimolecular GQ (tri-GQ) that is solely assembled by the self-trimerization of d(GTTAGG), preferentially in Na⁺ solution tolerant to an equal amount of K⁺ cation. Eight guanines from three asymmetrically folded strands of d(GTTAGG) are organized into a two-tetrad core, which features a broken G-column and two width-irregular grooves. Fast strand exchanges on a timescale of second at 17°C spontaneously occur between folded tri-GQ and unfolded single-strand of d(GTTAGG) that both species coexist in dynamic equilibrium. Thus, this tri-GQ is not just simply a static assembly but rather a dynamic assembly. Moreover, another minor tetra-GQ that has putatively tetrameric (2+2) antiparallel topology becomes noticeable only at an extremely high strand concentration above 18 mM. The major tri-GQ and minor tetra-GQ are considered to be mutually related, and their reversible interconversion pathways are proposed accordingly. The sequence d(GTTAGG) could be regarded as either a reading frame shifted single repeat of human telomeric DNA or a 1.5 repeat of *Bombyx mori* telomeric DNA. Overall, our findings provide new insight into GQs and expect more functional applications.

INTRODUCTION

As an important non-canonical DNA secondary structure *in vivo* and *in vitro*, G-quadruplexes (GQs) have been investigated in many fields, including biology (1,2), medicinal chemistry (3–5), supramolecular chemistry (6,7) and nanotechnology (8,9). A wide range of structural polymorphisms was regularly observed among GQs in terms of the number and orientation of strands, number of stacked G-tetrads, loop variations in length and spatial arrangement, combination of *syn* and *anti* glycosidic bond conformations, and dimensions of four grooves (10,11).

Vast GQs are folded primarily by one, two or four G-rich oligomers (10). Notably, only three exceptions of forming a trimolecular G-quadruplex (tri-GQ) were reported (12–14), and none of the tri-GQs had obtained an atomic resolution structure that was solved by the most structurally informative methods, neither X-ray of crystals nor nuclear magnetic resonance (NMR) in solution.

In the first report of a tri-GQ assembled by three completely different G-rich strands (12), Mergny's laboratory inventively designed short duplexes as a guide to pre-locate the G-rich tracts in spatial proximity, and then the formation of a defined tri-GQ structure was triggered by adding a GQ-compatible Na⁺ cation. However, this approach actually formed a junction between duplex segments and quadruplex segments, rather than a purely discrete tri-GQ. As Mergny pointed out, the formation of this tri-GQ assembly was still highly dependent on the guidance of covalently attached duplex segments. Currently, when using this 'duplex guidance approach', none of a purely discrete tri-GQ has been achieved whenever the guiding duplex segments are removed.

*To whom correspondence should be addressed. Tel: +86 551 6559 1582; Email: nazhang@hmfll.cas.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

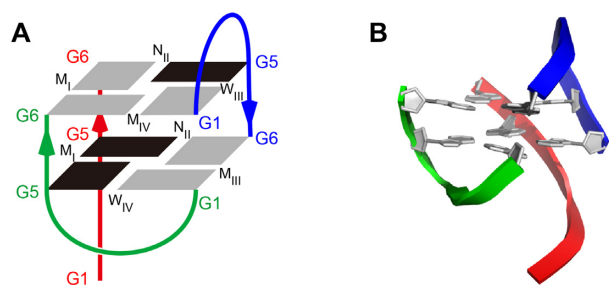


Figure 1. (A) Schematic folding topology and (B) one representative cartoon refined structure of a tri-GQ self-assembled by d(GTTAGG). Three differently folded strands are represented in blue, green and red respectively. The *syn*-guanines and *anti*-guanines are black and grey, respectively. W, M, and N represent wide groove width, medium groove width and narrow groove width, respectively. I, II, III, and IV indicate groove I, groove II, groove III and groove IV, respectively. For clarity, the illustration of a weakly formed Watson–Crick A:T base pair between red A4 and green T3, capped with the tri-GQ core, is skipped.

Telomeric DNA consists of a tandem array of the short G-rich repeat unit, d(GGGTTA) for humans (15,16) and d(GGTTA) for silkworm *Bombyx mori* (17). The sequence d(GTTAGG) that we examined in this work could be regarded as either a reading frame-shifted single repeat of human telomeric DNA or a less intact two repeat of *Bombyx mori* telomeric DNA. Here, we demonstrate, to the best of our knowledge, the first NMR solution structure of a tri-GQ assembled solely by the self-trimerization of d(GTTAGG) in Na⁺ solution (schematic Figure 1A).

This tri-GQ has two stacked G-tetrads and features a novel three-stranded folding topology, a broken G-column and two width-irregular grooves (schematic Figure 1A). In addition, NMR studies reveal that the tri-GQ of d(GTTAGG) is not just simply a static structure but rather is a dynamic assembly. Fast strand exchanges spontaneously occur on a timescale of second at 17°C, between folded tri-GQ and unfolded single-strand of d(GTTAGG), that both species coexist in dynamic equilibrium. Moreover, this tri-GQ is preferentially stabilized by Na⁺ cation, and enables a tolerance to a slightly excessive (or at least an equal) amount of K⁺ cation. Collectively, our findings make polymorphic GQ structures even more diverse, which potentially contribute to more functional applications.

MATERIALS AND METHODS

Preparation of unlabelled DNA samples

The unlabelled DNA oligonucleotides were purchased from Sangon Biotech (Shanghai) Co., Ltd. (China) and purified by high-performance liquid chromatography (HPLC) using a C18 reversed-phase separation column. Other chemicals were purchased from Thermo Fisher Scientific Inc. (China). The DNA samples were dialysed against the corresponding concentration of Na⁺ or K⁺ solution to remove the contaminant. The DNA strand concentrations were then determined by measuring the UV absorbance at 260 nm.

DNA samples were prepared mostly in 20 mM sodium phosphate buffer of final 100 mM Na⁺ at pH 6.8 if not stated. The samples were heated at 95°C for 5 min and then slowly annealed to room temperature before examination.

The sample strand concentration was 2.0 mM for the CD measurements and 0.5–24 mM for the NMR experiments, unless otherwise stated. The samples for NMR measurement in D₂O were prepared by freeze-drying and then re-dissolved in 99.96% D₂O.

Preparation of the sequence position-specifically ¹³C,¹⁵N-labelled DNA samples

We prepared three samples that were 100% ¹³C,¹⁵N-labelled specifically at G1, G5 and G6 positions respectively for d(G1T2T3A4G5G6), using our in-house enzymatic synthesis approach after appropriate optimisation and modification of the previously reported methods (18–21). The details of this in-house enzymatic synthesis approach are fully described in another methodological manuscript (Wenqiang Fu and Na Zhang, in preparation). A brief description of the preparation of sequence position-specifically labelled samples is given here.

In general, position-specific DNA labelling combines two steps of enzymatic polymerization reactions of deoxyribonucleoside triphosphates (dNTPs), which are catalysed by a Taq polymerase (Sangon, China). Taking a specific dG1-labelling for the sequence of d(G1T2T3A4G5G6) as an example, the first step utilizes a duplex that has two non-blunt ends, which are designed to function as a template for an enzymatic extension of the primer, as illustrated in the schematic Supplementary Figure S17. Two individual sequences of d(AA^uACTCGTAGCCA/rC/) (termed G1-F) and d(CGTGGCTACGAGTTTT) (termed G1-R), which are partially complementary to each other (listed in Supplementary Table S2), are slowly annealed together to form this duplex (Supplementary Figure S17). Accordingly, the dented 3'-terminus of G1-F has a ribose cytosine (rC) as the primer, while the protruding 5'-terminus of G1-R has an overhung deoxyribose cytosine as a coding template (underlined), which is ready for the first designated primer extension that is consistent with a partial filling strategy. On the other hand, instead of designing a blunt end at another terminal of this duplex template, one dented 5'-end of G1-F integrated with another protruding portion of G1-R at the 3'-end is emphasized purposely to prevent an undesired primer extension and/or a possible non-templated addition of nucleotides in the presence of a DNA blunt end.

The first elongation reaction incorporates a singly labelled guanine by using 100% uniformly ¹³C,¹⁵N-labelled dGTP (Cambridge Isotopes, USA). The enzymatic reaction was performed with 0.02 mM G1-F and G1-R, Taq polymerase (Sangon, China), 0.03 mM ¹³C,¹⁵N-labelled dGTP (Cambridge Isotopes, USA) in 1 ml of 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM MgCl₂ and 0.1% Triton X-100 (1 × reaction buffer). The mixtures were heated at 95°C for 5 min and then placed at 55°C for 30 min. The polymerization was stopped by freezing and thawing twice. After centrifugation in a 3000 K ultrafiltration tube, ultrapure water was added to the reaction system to continue centrifugation twice to remove salt ions and an excessive portion of ¹³C,¹⁵N-labelled dGTP that was unreacted in the reaction system. Next, 1 × reaction buffer, 0.04 mM unlabelled dATP, 0.08 mM unlabelled dTTP, 0.08 mM unlabelled dGTP, Taq polymerase and 0.02

mM d(CCTAACGTGGCTACGAGTTTTT) (termed G6-R) were added to the prior reaction system, which was already ultrafiltrated. The subsequent one-pot reaction was continued via a polymerase chain reaction (PCR) cycle for 2 h. The desired sample product, namely, ^{13}C , ^{15}N -labelled dG1 of d(G1T2T3A4G5G6), was detached from the template by selectively breaking the phosphodiester bond between this ribose rC and its 3'-flanking residue by alkaline cleavage. The alkaline cleavage step involved adjusting the pH of the reaction mixtures to 12 with KOH and incubating the mixtures at 90°C for 30 min. The crude products were purified using C18 reversed-phase high-performance liquid chromatography (HPLC) with an elution of various combinations of triethylammonium acetate (TEAA) buffer and acetonitrile. The overall yield of the procedure is approximately 40%, which can be readily scaled to lower or higher quantity preparations. This approach is enough to prepare 0.25 ml of a 1 mM sample, which is sufficient for an NMR microtube (Shigemi Inc.).

The preparation of the sequence-specifically labelled dG5 of d(G1T2T3A4G5G6) is similar to the strategy that was previously described using designed primers/templates of G5-F, G5-R and G6-R (Supplementary Table S2). Appealingly, the preparation of sequence-specifically labelled dG6 at the extreme 3'-end for the sequence of d(G1T2T3A4G5G6) is even more straightforward and convenient through simply accomplishing the first elongation reaction that incorporates the single labelled dG6 by using 100% uniformly ^{13}C , ^{15}N -labelled dGTP via designed primers/templates of G6-F and G6-R (Supplementary Table S2). The yield of the specially labelled dG6 sample is ~60%. All oligonucleotides that are designed as corresponding primers/templates (listed in Supplementary Table S2) were chemically synthesized by Sangon Biotech (Shanghai) Co., Ltd. In addition, similarly using this in-house enzymatic strategy, we also singly 100% isotope-labelled the sequence of d(G1-T2-T3-A4-G5-G6) specifically at dT2 or dT3 respectively, by means of 100% uniformly ^{13}C , ^{15}N -labelled dTTP (Cambridge Isotopes, USA) and the designed primers/templates accordingly.

To achieve the high sample concentration that is required for certain experimental conditions, three 100% ^{13}C , ^{15}N -labelled samples, specifically at G1, G5, and G6 respectively for d(G1T2T3A4G5G6), were mixed with an appropriate amount of conventionally unlabelled d(G1T2T3A4G5G6). The resulting samples that were 50–60% ^{13}C , ^{15}N -enriched specifically at G1, G5, and G6 respectively, were prepared with a final total strand concentration of 0.6–1.0 mM, whereas in an extreme case, the approximately 1–4% ^{13}C , ^{15}N -low enriched samples with an 18 mM strand concentration were also achieved.

Nuclear magnetic resonance (NMR) spectroscopy

NMR data were collected on 500, 600 and 850 MHz Bruker spectrometers with cryoprobes at 3–36°C. Two-dimensional total correlation spectroscopy (TOCSY) with a mixing time of 80 ms, ^1H – ^{13}C heteronuclear multiple bond correlation (HMBC), ^1H – ^{15}N heteronuclear single quantum correlation (HSQC), ^1H – ^{13}C HSQC, and nuclear Overhauser effect spectroscopy (NOESY) spectra in H_2O (containing

10% D_2O , a mixing time of 250 ms) and in 100% D_2O (mixing times of 50, 200 and 300 ms) were recorded for resonance assignment and structural identification. Two-dimensional diffusion ordered spectroscopy (DOSY) spectra were recorded with various diffusion times (60, 120, 200, 300 and 500 ms) at 32°C in H_2O (containing 10% D_2O). Water suppression either by gradient-tailored excitation (WATERGATE) or jump-return pulse sequence was employed in experiments for 10% D_2O samples, and by pre-saturation pulse sequence for 100% D_2O samples. All data sets were processed and analysed using Bruker Topspin 3.2 and CcpNmr Analysis software (version 2.4.2) (22).

Two-dimensional rotating frame Overhauser effects spectroscopy (ROESY) spectra (mixing time of 300 ms) were recorded to verify the chemical exchange phenomena. To determine the exchange rate between the unfolded portion of single stranded d(GTTAGG) and the folded trimolecular GQ, two-dimensional exchange spectroscopy (EXSY) spectra (mixing times of 50–500 ms) were obtained in D_2O at 17°C. The well-resolved exchange cross-peaks of G6 (green), G5 (blue), and G5 (red), as illustrated in the schematic Supplementary Figure S8, respectively with their corresponding unstructured d(GTTAGG) signals in each of the EXSY spectra, were extracted for integrals. Ratios between the exchange cross-peaks and corresponding diagonal peaks (I_c/I_d) were calculated and then linearly fitted as a function of mixing time (τ_m). As previously reported (23), the slope of the fitted line gives an estimate of the exchange rate k_{ex} based on a two-site exchange model.

Circular dichroism (CD) spectroscopy

CD measurements were carried out using a Chirascan (Applied Photophysics) circular dichroism spectrometer with a 1-mm path length quartz cuvette. Setting parameters: temperature, 10°C; scan range, 200–340 nm; scan rate, 1 nm/step; and time point, 0.25 s. An average of three scans was taken for each measurement, and the baseline was corrected with the same buffer. The CD data were processed by Pro-Data Viewer (Chirascan) and further plotted with OriginPro 9.0. The Chirascan (Applied Photophysics) circular dichroism spectrometer can handle much more concentrated DNA samples at sub mM without signal overflow (24).

NMR structure calculation

The distance restraints of nonexchangeable protons were derived from the ^1H – ^1H NOESY spectrum in 100% D_2O (mixing time of 300 ms) processed by CcpNmr Analysis 2.4.2 (22). The peak intensities (volume) that correspond to a known distance are used for calibration. This process is based upon the r^6 distance dependence of the NOE cross-peak intensities (25). The distance of the independent thymine H6-CH₃ cross-peak was set to 2.99 Å as the distance reference (26). The lower and upper bounds were assigned to 20–30%. The exchangeable proton restraints were inferred from NOESY spectra (mixing times of 60 and 250 ms, respectively) in H_2O . These distances were classified as strong (2.7 ± 0.9), medium (3.8 ± 1.2) or weak (5.0 ± 1.5) Å, respectively, according to the intensities of the cross-

peak volumes in the NOESY spectra. Eighteen hydrogen-bonding restraints were added to retain the hydrogen bonding within individual G-tetrads and A·T base pairs. The glycosidic χ torsion angles of *syn* and *anti* bases showed strong and medium H1'-H8/H6 cross-peaks on the 50 ms NOE spectrum in D₂O (Figure 3C), which were set to $65^\circ \pm 25^\circ$ and $220^\circ \pm 30^\circ$, respectively, as previously described (27).

The structure of this trimolecular G-quadruplex was determined by molecular dynamics- simulated annealing computations that were driven by NOE distance, dihedral angle, and hydrogen bonding restraints using the XPLOR-NIH program version 3.0.3 (28). First, iterative structure calculations were performed. A set of folded structures were generated from an extended conformation with satisfied covalent geometry by enforcing distance, torsion angle, and base-pair planarity restraints following the previously reported protocol (29,30). The force constants were scaled at 1 and 200 kcal mol⁻¹ Å⁻² for distance restraints and hydrogen bond constraints. The unambiguous restraints were incorporated, and the best structure with the lowest energy and an acceptable number of violations was employed as the structure for the next round of refinement. In this manner, an increasing number of NOE distance restraints could be included in the structure calculations. Second, structure refinement was performed (31). The best folded structure based both on their minimal energy terms and the number of NOE violations from the first step was input as a start file for refinement, and a total of 200 structures were generated. In this step, the scale of planarity was decreased to 0.5 kcal mol⁻¹ Å⁻². In total, 308 distance restraints were applied in the structural calculation. Eventually, the ten best refined structures based on the lowest energy and lowest number of NOE violations were selected as the representative solution structures for this tri-GQ of d(GTTAGG). Subsequently, these ten best representative solution structures were validated by the wwPDB validation system and deposited in the Protein Data Bank (PDB ID: 6M05). The refined structures were displayed using PyMOL (32). The chemical shifts have been deposited in the BioMagResBank under accession code 28081.

RESULTS AND DISCUSSION

A single GQ formation of d(GTTAGG) in Na⁺ solution

In the one-dimensional proton ¹H NMR spectrum of d(GTTAGG) with a 0.5 mM strand concentration in 100 mM Na⁺ solution at 3°C and pH 6.8, eight sharp guanine imino proton signals were well resolved between δ 10.8 and 12.0 ppm, the characteristic region for GQ formation (Figure 2A). The peak intensities of every guanine imino proton remained uniformly equivalent over a considerable range of temperature variations from 3 to 36°C (Supplementary Figure S1), which is consistent with the formation of a single GQ structure as the major folded species.

NMR validation of the self-trimerization of d(GTTAGG) using sequence position-specifically ¹³C,¹⁵N-labelled samples

Usually, based on a sequence-specific, low-enrichment approach (33), guanine imino protons could be unambiguously assigned using samples that are generally 2–8% ¹⁵N-

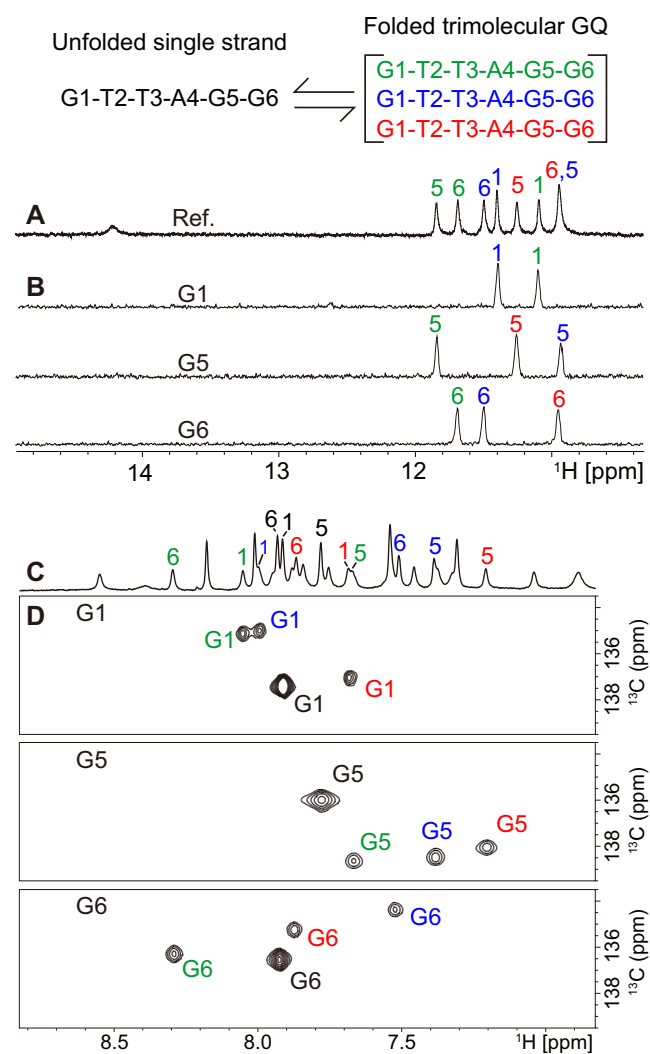


Figure 2. (A) Expanded imino proton region of the one-dimensional ¹H NMR spectrum of d(GTTAGG) as a reference. (B) Guanine imino proton assignments from ¹⁵N-filtered spectra of sequence-specifically labelled samples that are 100% ¹³C, ¹⁵N-labelled specifically at the indicated positions of d(G1T2T3A4G5G6). (C) Expanded non-exchangeable base proton region of the one-dimensional ¹H NMR spectrum of d(GTTAGG) as a reference. (D) Guanine H8 proton assignments using two-dimensional ¹H-¹³C HSQC experiments on the sequence-specifically labelled samples that are 50–60% ¹³C, ¹⁵N-enriched at the indicated positions of d(G1T2T3A4G5G6). For clarity, only the guanine resonance assignments are shown with specific residue numbers according to the top inset sequences. Three differently folded strands of the tri-GQ structure are represented by blue, green and red respectively, the same as those strands illustrated in the schematic Figure 1A, whereas the coexisting unfolded portion of single-stranded d(GTTAGG) is shown in black. Experimental conditions: strand concentration, 0.5 mM unlabelled DNA in (A); 0.4–0.6 mM 100% ¹³C, ¹⁵N-labelled DNA in (B); 3.5 mM unlabelled DNA in (C); 0.6–1.0 mM 50–60% ¹³C, ¹⁵N-enriched DNA in (D), 100 mM NaCl, 20 mM phosphate buffer (pH 6.8), 10% D₂O, 3°C.

labelled at a designated position by solid-phase synthesis. However, this chemical approach is still unable to isotopically label the last residue at the extreme 3'-end. Instead, we prepared three samples that were 100% ¹³C, ¹⁵N-labelled specifically at G1, G5, and in particular at G6 individually for sequence d(G1T2T3A4G5G6), which is based on

our in-house enzymatic approach at a considerably affordable price, as described in the labelled-sample preparation of MATERIALS AND METHOD.

All eight guanine imino protons shown in Figure 2A were assigned unambiguously to their positions in the sequence of d(G1T2T3A4G5G6) based on the relatively enhanced intensity in the ^{15}N -filtered spectra (Figure 2B). Notably, each spectrum of a singly labelled sample specifically at either G5 or G6 positions exhibited three sharp imino peaks with equivalent intensity, whereas only two sharp imino peaks were observed for that of G1 (Figure 2B). This unique pattern was consistent with the presence of three conformationally distinct strands of d(GTTAGG) that were adopted in a single tri-GQ structure, as illustrated in the schematic Figure 1A.

For this trimolecular GQ in Figure 1, except red G1 which actually acted as an overhung residue without hydrogen bonding, eight other guanine residues from three coloured strands were involved in the G-tetrad constructs. Once hydrogen-bonded, the imino proton would become sharp enough to be detectable by NMR. In this regard, only two sharp hydrogen-bonded imino proton signals for G1 from both green and blue strands were actually detected, whereas every three sharp hydrogen-bonded imino proton signals for either G5 or G6 residues respectively were observed in this asymmetrically three-stranded tri-GQ (Figure 2B).

The observed number of non-exchangeable base H8 protons at 6.8–8.8 ppm (Figure 2D), using samples 50–60% ^{13}C , ^{15}N -enriched specifically at G1, G5 and G6 individually, as prepared in MATERIALS AND METHOD, once again confirmed the presence of three differently structured strands in this tri-GQ. This three-stranded feature of tri-GQ is evidenced by an exhibition of every three structured base H8 proton signals (shown in red, green and blue), in addition to another coexisting unstructured (shown in black) base H8 proton signal for each sample, in which only a single guanine is specifically labelled (Figure 2D).

NMR validation of the self-trimerization of d(GTTAGG) using unlabelled sample

Moreover, using the conventionally unlabelled sample of d(GTTAGG) as an alternative, the NMR signals that have a characteristic region of chemical shift, such as adenine-H2 and thymine methyl CH3 protons, were monitored to validate the unique presence of three distinct strand conformations for this folded tri-GQ of d(GTTAGG). This finding was confirmed by the observations of three adenine-H2 protons with equal intensity in ^1H - ^{13}C heteronuclear single quantum correlation (HSQC) experiments and six methyl signals of thymines with equal intensity either in ^1H - ^{13}C HSQC or in total correlation spectroscopy (TOCSY) experiments (Supplementary Figure S3). Conversely, only a single set of signals that belongs to the unfolded portion of d(GTTAGG) coexisted in equilibrium (Supplementary Figure S3).

Overall, the crucial determination of three differently folded strands of d(GTTAGG) that were simultaneously adopted in this novel tri-GQ assembly was thoroughly validated by NMR.

NMR assignments of tri-GQ assembled by d(GTTAGG)

The sequence-specific assignments of guanine imino and H8 protons of d(G1T2T3A4G5G6) were first unambiguously achieved respectively using 100% and 50–60% ^{13}C , ^{15}N -labelled samples individually at G1, G5 and G6 (Figure 2). Next, all residues had to be sorted into each of three conformationally distinct strands for this folded tri-GQ of d(GTTAGG). The strand-specific assignments of non-exchangeable protons were subsequently completed by using H8/6-H1' NOE sequential connectivity, which is particularly inevitable for any asymmetrically folded structure. As a result, the non-exchangeable base H8/6 and sugar H1' proton resonances that belong to each conformationally distinct strand were eventually sorted into three kinds: blue, green and red (Figure 3B).

Subsequently, the through-bond correlation spectrum (HMBC) between imino protons and H8 protons via ^{13}C (34) further established strand-specific assignments for exchangeable guanine imino protons (Figure 3D), which is one critical step for solving any asymmetrically folded GQ structure. Accordingly, eight sharp guanine imino protons were sorted into three differently folded strands of d(G1T2T3A4G5G6) individually, including three green guanines (G1, G5, and G6), three blue guanines (G1, G5, and G6), and two red guanines (G5 and G6). For red G1, its base H8 proton assignment has been verified either by NOE sequential walking or unambiguously by using ^{13}C , ^{15}N -labelled/enriched samples specifically at G1. However, there was no observation of a through-bond correlation between this pre-assigned red G1-H8 proton and any imino proton, supporting that red G1 was not hydrogen-bonded, and thus, no imino proton signal is detected. Moreover, another validation of the overhung red G1 without hydrogen bonding was further examined upon titration with paramagnetic Cu^{2+} based on the effect of paramagnetic relaxation enhancement (PRE), as described in the Supplementary text and Figure S6.

Overall, the sorting of these guanine imino protons based on NOE sequential walking and HMBC (Figure 3B and D) was consistent with the independently observed hydrogen-bonding pattern that was unambiguously assigned using the sequence-specifically labelled samples (Figure 2). Accordingly, the same results were cross-checked by different experimental approaches, which further verified the correctness of NMR assignments. The complete chemical shift assignments of imino, amino, base and sugar protons are listed in Supplementary Table S1.

NMR structural identification of tri-GQ folded by self-trimerization of d(GTTAGG)

Analysis of characteristic NOEs between guanine imino and H8 protons (Figure 3E) revealed the formation of a trimolecular GQ, as shown in the schematic Figure 1A. Except for red G1, which actually acted as an overhanging residue without hydrogen bonding, eight other guanine residues from three conformationally distinct red, blue and green strands built up a core of two-stacked G-tetrads, namely, $\text{G5}_{\text{blue}}\cdot\text{G6}_{\text{red}}\cdot\text{G6}_{\text{green}}\cdot\text{G1}_{\text{blue}}$ and $\text{G6}_{\text{blue}}\cdot\text{G1}_{\text{green}}\cdot\text{G5}_{\text{green}}\cdot\text{G5}_{\text{red}}$ (Figure 3F). The hydrogen bond directionalities of these two G-tetrads are anti-

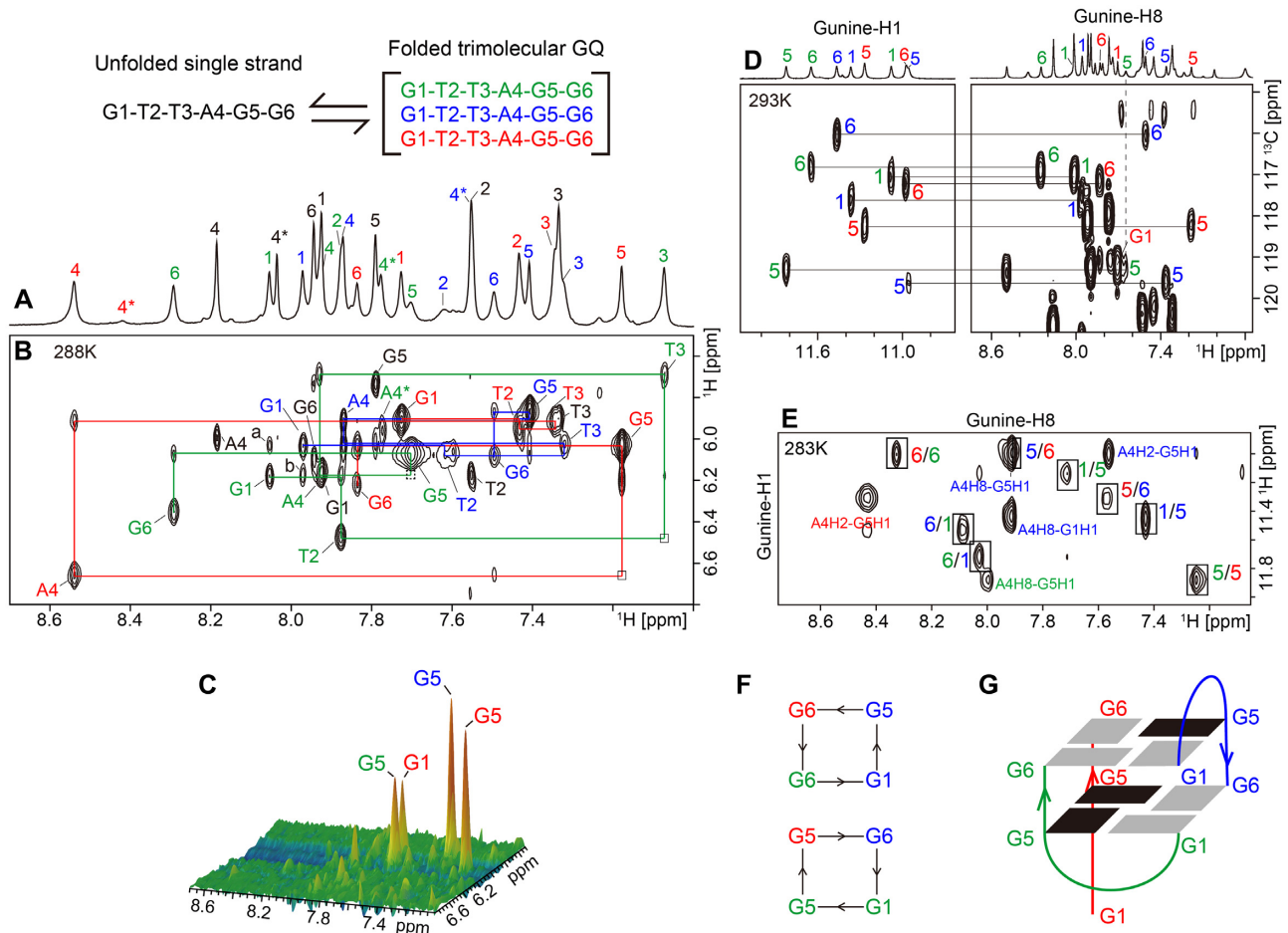


Figure 3. NMR spectral assignments to determine the tri-GQ folding topology of d(G1T2T3A4G5G6). (A) Expanded region (7.0–8.7 ppm) of the one-dimensional ^1H spectrum for the fully assigned non-exchangeable base protons of d(GTTAGG) in D_2O at 15°C . For clarity, the resonance assignments are shown with specific residue numbers according to the top inset sequences. Three conformationally asymmetric strands folded in the tri-GQ structure are represented by blue, green and red respectively, the same as those strands illustrated in the schematic Figure 1A; the coexisting unfolded single strand is shown in black. Peaks of adenine H2 protons are labelled with asterisks. (B) H8/6-H1' sequential connectivity in the NOESY spectrum (300 ms mixing time) in D_2O at 15°C . Intra-residue NOEs are labelled with residue numbers according to the top inset sequence. The missing inter-residue cross-peak marked by the open box can be observed at a lower spectral contour level. Intra-residue NOEs of unfolded species are labelled in black. The NOE cross-peaks between green G1 and blue G1 residues, which support the spatial arrangement of a broken G-column in tri-GQ, are assigned as follows: a, G1H8(green)-G1H1'(blue); b, G1H8(blue)-G1H1'(green). (C) Stacked plot of the two-dimensional NOESY spectrum (50 ms mixing time) of d(G1T2T3A4G5G6) in D_2O at 12°C , which shows strong intra-residue H8-H1' cross-peaks for the *syn* guanines. (D) HMBC spectrum that shows the through-bond correlations between imino H1 and base H8 protons from the same guanine residue via $^{13}\text{C}5$ at natural abundance. The HMBC experiment was conducted at 20°C in order to resolve the imino protons of red G6 and blue G5. (E) NOESY spectrum (250 ms mixing time) in H_2O at 10°C , which shows the inter-residue NOE cross-peaks between guanine imino H1 (vertical axis) and base H8 (horizontal axis) protons within the same layer of each G-tetrad. The characteristic H1-H8 cross-peaks are framed and labelled with a residue number of H1 and H8 protons in the first position and second position, respectively. Other strong NOE peaks attributed to adenine H2/H8 and guanine H1 are also marked. (F) Specific H1-H8 NOE connectivity pattern around a G-tetrad indicated with arrows. (G) Schematic folding topology of d(GTTAGG), the same as shown in the schematic Figure 1A. Experimental conditions: strand concentration, 13.6 mM; temperature as indicated; 100 mM NaCl; pH 6.8.

clockwise and clockwise respectively (viewed from the top, shown in Figure 3F and G).

These two stacked G-tetrads with opposite hydrogen-bond directionalities were expected to have characteristic imino-imino proton NOEs between two head-to-head guanines, each from two different G-tetrads (schematic Figure 1A). Clear observations of the corresponding imino-imino NOEs, including those of $\text{G6}_{\text{green}}-\text{G1}_{\text{green}}$, $\text{G6}_{\text{red}}-\text{G5}_{\text{green}}$, and $\text{G5}_{\text{red}}-\text{G5}_{\text{blue}}$ (Supplementary Figure S4), support the folding topology of the tri-GQ adopted by d(GTTAGG), whereas the chemical shift difference between the imino

protons of G6_{blue} and G1_{blue} is too small to detect the expected NOE.

The strong intensities of intra-residue H8-H1' NOE cross-peaks (Figure 3C) indicate the adoption of a *syn* glycosidic conformation for all three G5 residues, which belong to three differently folded strands (green, blue and red). Although it actually acted as an overhanging residue without hydrogen bonding, red G1 still adopted a *syn* glycosidic conformation. This behaviour is similar to the observations reported in the literature (35,36). Likely, this overhanging red G1 is stabilized by a hydrogen bond between the 5'-terminal

Table 1. NMR restraints and structure statistics

A. NMR Restraints	
Total number of DNA distance restraints	308
Exchangeable distance restraints	51
Nonexchangeable distance restraints	257
Inter-residue restraints	108
Intra-residue restraints	200
Hydrogen bonding restraints	18
Torsion angle restraints	18
B. Structural statistics	
NOE violations	
Number > 0.2Å	2.3±1.3
Maximum violations (Å)	0.329
rms deviation of violations	0.029±0.004
Deviation from the ideal covalent geometry	
Bond length (Å)	0.003±0.000
Bond angle (°)	0.381±0.007
Impropers (°)	0.226±0.004
Pairwise rms deviation (Å) (10 refined structures)	
G-tetrad core	0.609±0.128
All heavy atoms	1.209±0.252

hydroxyl group and N3 of guanine at the 5'-terminus in a *syn* glycosidic conformation.

Moreover, the formation of a much less stable Watson–Crick A:T base pair between A4 of red strand and T3 of green strand were detailed in the Supplementary text and Figure S18. However, this capping Watson–Crick A:T base pair is too weak to be a critical factor that impacts on the stability of entire tri-GQ assembly.

Solution structure of this tri-GQ adopted by d(GTTAGG)

The overall solution structure was calculated based on NMR restraints using X-PLOR-NIH programs. The NMR restraints and structural statistics are listed in Table 1. Ten superimposed lowest-energy structures are displayed in Figure 4A. The coordinates of these ten structures of tri-GQ have been deposited in the Protein Data Bank (accession code 6M05). The G-tetrad core of tri-GQ converged well, with a root mean squared deviation (R.M.S.D.) of 0.609 ± 0.128 Å. The edgewise loops and overhanging red G1 were more flexible than the G-tetrad core. A representative refined structure is shown as a ribbon representation in Figure 1B.

Broken G-columns composed of non-contiguous guanine backbones have been occasionally observed in GQs (37–40). Here, the tri-GQ of d(GTTAGG) with two stacked G-tetrads has three conventionally contiguous G-columns of the G5-G6 tracts from the red, blue and green strands and the fourth broken G-column, which is composed of two G1 guanines from two different strands of green and blue. The critical NOEs specifically observed between green G1 residues and blue G1 residues (Supplementary Figure S5) verified the spatial arrangement of this broken G-column in tri-GQ.

In general, the basic requirement for forming a stable GQ was the association of two-stacked tetrads with the same groove-width combinations (41,42). However, several exceptions of the width-irregular grooves, in which the width combinations were variable along a single groove, have

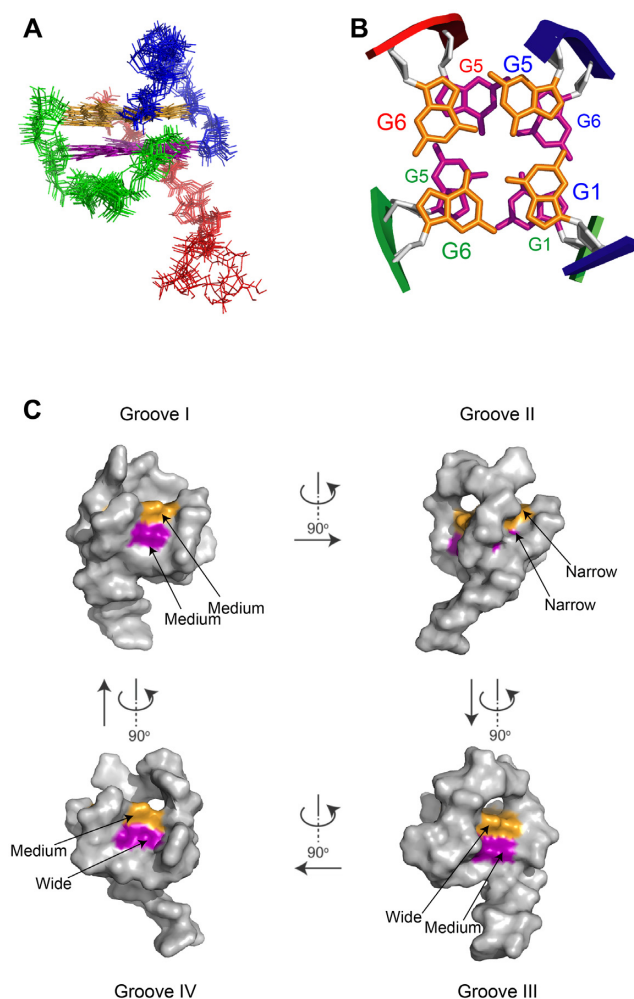


Figure 4. Solution structure of tri-GQ adopted by d(GTTAGG). (A) Superimposed stereo view of ten lowest-energy intensity-refined structures. (B) Stacking of the top tetrad G6(green)-G1(blue)-G5(blue)-G6(red) over the bottom tetrad G5(green)-G5(red)-G6(blue)-G1(green). The guanine bases of the top G-tetrad and bottom G-tetrad are shown in orange and purple, respectively. The backbones of three differently folded strands in tri-GQ are shown in green, blue and red. Guanines from the top G-tetrad are labelled using larger fonts, whereas guanines from the bottom G-tetrad are labelled using smaller fonts. (C) Surface representations of tri-GQ adopted by d(GTTAGG). The planes of the top G-tetrad and bottom G-tetrad are shown in orange and purple, respectively, and each groove width is labelled.

been observed in certain GQs that also have a broken G-column (39,40). For tri-GQ of d(GTTAGG), as shown in Figures 1 and 4C, there are two conventional width-uniform grooves: one all-medium (I) groove and another all-narrow (II) groove. Notably, the other two width-irregular grooves (III and IV) are also well adapted. Each of these width-irregular grooves contains one medium-width groove next to another wide groove (M_{III} next to W_{III} , M_{IV} next to W_{IV} , Figure 1A). This unique dimension of the width-variable grooves is expected to provide a distinctive target platform for potential GQ-specific binders.

Fast strand exchanges between the folded tri-GQ and unfolded single-strand species of d(GTTAGG)

In addition to those of three asymmetrically folded d(GTTAGG) strands, another single set of signals from the unfolded d(GTTAGG) species that coexisted in equilibrium, was also observed (Figures 2C, 2D and 3A; Supplementary Figures S2 and S3). These unstructured d(GTTAGG) signals became much more apparent at either an evaluated temperature (Supplementary Figure S2A) or a more diluted DNA concentration (Supplementary Figure S2B). A set of cross-peaks that arise from aromatic resonances was observed, notably each for a single residue of d(GTTAGG) in different folding states, between a structured strand of tri-GQ and the unstructured single strand, as shown in the expanded region of the rotating frame Overhauser effect spectroscopy (ROESY) spectrum at a mixing time of 300 ms (Figure 5B). Since they had the same phase as the diagonal peaks (Figure 5B), these cross-peaks were actually identified as exchange peaks rather than ROE peaks (43). The full spectrum of ROESY and other observed exchange cross-peaks in either expanded H1' or methyl proton regions are shown in Supplementary Figure S7.

Furthermore, EXchange spectroscopy (EXSY) spectra at various mixing times, which range from 50 to 500 ms, were also collected (Supplementary Figure S8). These observations revealed that fast strand exchanges between the structured tri-GQ and the unstructured single-strand species of d(GTTAGG), spontaneously occur on a timescale of second at 17°C as illustrated in Figure 5C. Thus, this tri-GQ is not just simply a static structure, but rather it is a dynamic assembly. This dynamic behaviour may be taken advantage of as an important feature for DNA GQ-based nanotechnology in the future. As a note, when compared with other guanines that assemble into this dynamic tri-GQ core of d(GTTAGG), blue G1 acts even more dynamic due to a relatively weaker stability, as detailed in the Supplementary text and Figures S14 and S15.

The detection of a fast local conformational switch from milliseconds to seconds by ROESY or EXSY for an intramolecular GQ has been reported (43,44). Unlike this tri-GQ of d(GTTAGG) which has two tetrads, most other reported GQs have a minimum of three tetrads; thus, they are normally too stable to permit fast millisecond/second strand exchange. Instead, a much slower strand exchange was detected by gel shifts between a tetrameric GQ and a fluoresce-labelled homologous strand, in a real-time manner in hours (45,46).

Tri-GQ folded preferentially in Na⁺ solution enables tolerance to an equal amount of K⁺ cations

Notably, the formation of a GQ was quite often determined to be much more K⁺-prone (47–51), rarely with exceptions (52,53). The detection of a K⁺-prone GQ can even be achieved in a mixture of minimal K⁺ cations and simultaneously excessive Na⁺ cations (54). Conversely, most GQs, although well prefolded in Na⁺ solution, were still unable to tolerate an equal (or even substantially lesser) amount of K⁺ cations.

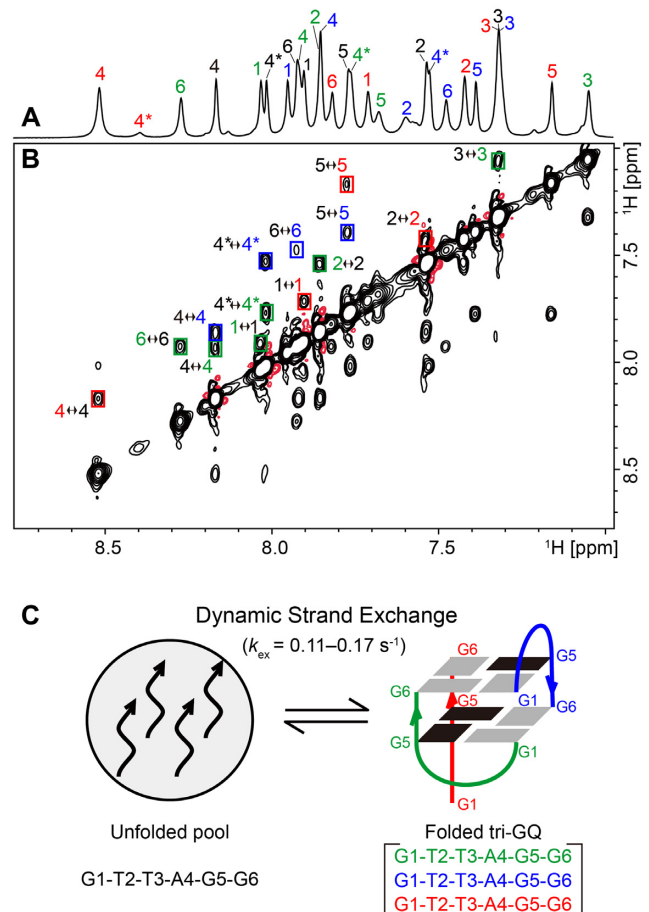


Figure 5. (A) Expanded region (7.0–8.7 ppm) of the one-dimensional ¹H spectrum for the fully assigned base H8/6 protons of d(GTTAGG) as a reference. Three different strands of the tri-GQ structure are shown in blue, green and red respectively, as shown by the schematic topology in Figure 1A, whereas the coexisting unfolded portion is shown in black. Peaks of the adenine H2 protons are labelled with stars. (B) Expanded aromatic proton region of the ROESY spectrum (mixing time: 300 ms) of d(GTTAGG) recorded at 17°C. The cross-peaks that show the same phase as that of the diagonal peaks in ROESY are due to strand exchange between the unfolded single strand species and the folded tri-GQ assembly (55). For clarity, only the observed cross-peaks that belong to a strand exchange are framed and labelled with residue numbers, whereas other expected strand exchange cross-peaks are invisible because of their poorly resolved chemical shifts. Experimental conditions: 21 mM DNA strand concentration, 100 mM NaCl, 20 mM phosphate buffer (pH 6.8), 100% D₂O. (C) Schematic of a dynamic tri-GQ assembly, in which fast strand exchanges spontaneously occur between the coexisting unstructured single strand of d(GTTAGG) and the structured tri-GQ.

Unlike only a single tri-GQ assembly adopted in Na⁺ solution alone, the folding of d(GTTAGG) in K⁺ solution alone essentially yielded a mixture of multiple structures (Figure 6B, top). To examine how much K⁺ cations could be tolerated by this tri-GQ that is well prefolded in Na⁺ solution, one-dimensional ¹H NMR titrations were carried out. Increasing amounts of K⁺ cations were titrated stepwise into a 100 mM Na⁺ solution of d(GTTAGG). Notably, the original NMR signals that belong to Na⁺-stabilized tri-GQ at 10–12 ppm were sustained at the titration point of a 100 mM equimolar mixture of K⁺ and Na⁺ solution (Figure 6A). Even when an additional 150 mM K⁺ was mixed, the

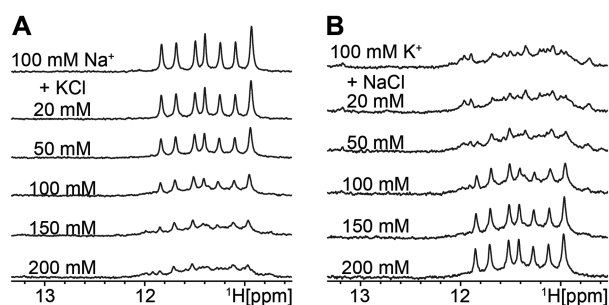


Figure 6. Expanded imino proton region of one-dimensional ^1H NMR spectra of $d(\text{GTTAGG})$ monitors the Na^+/K^+ switch titrations. (A) Starting in a 100 mM Na^+ solution and thereafter sequentially adding various amounts of K^+ cations at the indicated concentrations. (B) Reversely starting in a 100 mM K^+ solution and thereafter sequentially adding various amounts of Na^+ cations at the indicated concentrations. Experimental conditions: 2 mM DNA strand concentration, 20 mM phosphate buffer (pH 6.8), 3°C .

resulting spectrum still mostly resembled the tri-GQ structure folded in the presence of Na^+ solution exclusively (Figure 6A). The reverse titrations (Figure 6B) that were obtained by adding increasing amounts of Na^+ cations into a 100 mM K^+ solution of $d(\text{GTTAGG})$ also revealed that the formation of this tri-GQ is preferred in Na^+ solution over an equal amount of K^+ cations. Moreover, the result of Na^+/K^+ cation competitive titration by CD (Supplementary Figure S16), which is detailed in the supplementary text, showed agreement with that of the previously mentioned NMR studies above. Overall, the NMR and CD titrations demonstrated that this Na^+ -stabilized tri-GQ of $d(\text{GTTAGG})$, can be tolerant of a slightly excessive (or at least an equal) amount of K^+ cations.

Formation of the minor tetrameric GQ is favourable at an extremely high strand concentration and a high temperature

Within a GQ characteristic region of 10–12 ppm, four asterisked minor guanine imino proton peaks with an equivalent intensity for a single GQ became more evident, however, only when an extremely enhanced sample strand concentration of more than 18 mM was reached (Supplementary Figure S11A). The guanine identities of these asterisked minor imino proton peaks were confirmed by $^1\text{H}/^{15}\text{N}$ HSQC (Supplementary Figure S12) according to the characteristic ^{15}N chemical shifts at approximately 145 ppm. The minor peak intensities of every asterisked guanine imino proton remained uniformly equivalent over a considerable range of temperature variations from 28 to 36°C (Supplementary Figure S11C), which is consistent with the formation of a single GQ structure for this minor folding form.

In the DOSY spectrum at a diffusion time of 60 ms (Supplementary Figure S11B), the asterisked minor GQ diffused relatively slower than that of major tri-GQ, suggesting that this asterisked minor GQ either had a higher molecularity or a looser structure than that of major tri-GQ, which has a smaller molecularity and/or a more compact folding. In fact, the proposed tetrameric GQs, which have a two-tetrad core with four flexible overhangs, would lead to a looser structure (schematic Supplementary Figure S10).

Conversely, once the broken G-column which consists of two otherwise overhung G1 s was formed, tri-GQ was expected to have a relatively more compact structure.

In theory, an evaluated sample strand concentration favours the formation of a multiple-stranded assembly with a higher stoichiometry rather than a lower stoichiometry. Indeed, the formation of a tetrameric GQ (as asterisked in Supplementary Figure S11A) became more noticeable (but still as the minor population), however only when an extremely enhanced sample strand concentration of 18 mM or higher was reached. In addition, the temperature-variation studies showed that the structure of tetrameric GQ is relatively more stable at high temperatures, whereas tri-GQ is more favourable at low temperatures (Supplementary Figure S11C).

Proposed interconversion between the major tri-GQ and the minor tetrameric GQ with a putative (2+2) antiparallel topology

Considering the fluid and dynamic nature of strand exchange for this tri-GQ assembly, the broken column would have many chances to sense a potential competition of strand displacement by the consecutive G-tract G5–G6 of unfolded single strand $d(\text{GTTAGG})$, which coexists in a dynamic equilibrium. Once this strand displacement is achieved, a tetramolecular GQ is formed, which produces a two G-tetrad core with all four 5'-G1 residues overhung out, as proposed in schematic Supplementary Figure S10.

In an opposite process, the hydrophobic nature of DNA aromatic bases for these overhung G1 residues of a newly formed tetramolecular GQ would tend to avoid exposure to the polar water solvent. Accordingly, the spare overhung G1 residues had an inherent preference to participate, as much as possible, in a stacked G-tetrad core with the same hydrophobic nature. As a result, these two otherwise overhung residues for blue and green G1 s of this newly formed tetrameric GQ were still willing to fold back, which eventually leads to the re-construction of a broken column for tri-GQ, as proposed in the schematic Supplementary Figure S10.

These two opposite processes were reversible, and a counterbalance could be reached according to the concentration of $d(\text{GTTAGG})$. Notably, within a considerable range of sample concentrations up to 24 mM, the folded tri-GQ of $d(\text{GTTAGG})$ still rather preserved its three-stranded topology as the dominant folding form (as circled in Supplementary Figure S11A). This preservation revealed a reasonable adaption of this broken strand in tri-GQ. In routine conditions, such as at sub-millimolar strand concentrations and 17°C , the population of unfolded $d(\text{GTTAGG})$ single strands is dominant (Supplementary Figure S2), while for a comparison only between the folded species, the self-assembly of $d(\text{GTTAGG})$ intrinsically still had a preference for trimerization over tetramerization (Supplementary Figures S11A and C).

The rational determination of the asterisked minor tetrameric GQ of $d(\text{GTTAGG})$, which has a two-layer G-tetrad and a putative (2+2) antiparallel topology, is detailed in the Supplementary text and Figures S9–S13. A total of four differently folded tetrameric GQs, as proposed in Sup-

plementary Figure S9, including this putative (2+2) antiparallel GQ, are considered to have a similar stability (35,36). Thus, any of these four tetrameric GQs could be equally formed if only directly via the tetramerization of the unfolded single strand of d(GTTAGG). However, the ultimate observation of only one single minor tetrameric GQ that specifically has a putative (2+2) antiparallel topology suggested that the pre-existing major tri-GQ enables a guidance impact on the formation of this minor (2+2) antiparallel tetrameric GQ, as proposed in the schematic Supplementary Figure S10. As an experimentally supportive observation, this folded tri-GQ of d(GTTAGG) can further associate with another short G-rich fragment of a completely different sequence and readily reassemble into a heteromolecular tetra-GQ (unpublished data, as a separate manuscript in preparation by Haitao Jing and Na Zhang).

CONCLUSION

In this work, we described the first NMR solution structure of tri-GQ that was adopted by the preferential self-trimerization of d(GTTAGG) in Na⁺ solution with moderate K⁺ tolerance. This tri-GQ has a novel three-stranded folding topology with a broken G-column and width-irregular grooves. Notably, NMR studies reveal that the tri-GQ of d(GTTAGG) is not just a static structure but rather a dynamic assembly. Fast strand exchanges spontaneously occur on a timescale of second at 17°C between structured tri-GQ and unstructured single-strand of d(GTTAGG), that both species coexist in dynamic equilibrium. Moreover, another minor GQ that has a putatively tetrameric (2+2) antiparallel topology becomes noticeable only at an extremely high strand concentration above 18 mM. The major tri-GQ and the minor tetra-GQ are considered to be mutually related, and their reversible interconversion pathways are proposed accordingly. The polymorphism has been regarded as an inherent feature for GQs; our findings will make GQ structures even more diverse with potentially more functional applications.

DATA AVAILABILITY

The coordinates of 10 structures of trimolecular GQ have been deposited in the Protein Data Bank (accession code 6M05). The chemical shifts have been deposited in the BioMagResBank under accession code 28081.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

All the NMR experiments were performed at Steady High Magnetic Field Laboratory, Chinese Academy of Sciences. A portion of this work was supported by the High Magnetic Field Laboratory of Anhui Province.

FUNDING

National Key Research and Development Program of China [2016YFA0400900]; National Natural Science

Foundation of China [U1932157, 21372223, U1232145]; Major/Innovative Program of Development Foundation of Hefei Center for Physical Science and Technology [2018ZYFX004]. Funding for open access charge: National Natural Science Foundation of China [U1932157].

Conflict of interest statement. None declared.

REFERENCES

- Huppert, J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, **37**, 1375–1384.
- Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
- Han, H. and Hurley, L.H. (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.*, **21**, 136–142.
- Sissi, C. and Palumbo, M. (2014) Telomeric G-quadruplex architecture and interactions with potential drugs. *Curr. Pharm. Des.*, **20**, 6489–6509.
- 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.
- Davis, J.T. and Spada, G.P. (2007) Supramolecular architectures generated by self-assembly of guanosine derivatives. *Chem. Soc. Rev.*, **36**, 296–313.
- 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.
- Krishnan, Y. and Simmel, F.C. (2011) Nucleic acid based molecular devices. *Angew. Chem. Int. Ed. Engl.*, **50**, 3124–3156.
- 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.
- Karsisiotis, A.I., O'Kane, C. and Webba da Silva, M. (2013) DNA quadruplex folding formalism—a tutorial on quadruplex topologies. *Methods*, **64**, 28–35.
- Zhou, J., Bourdoncle, A., Rosu, F., Gabelica, V. and Mergny, J.L. (2012) Tri-G-quadruplex: controlled assembly of a G-quadruplex structure from three G-rich strands. *Angew. Chem. Int. Ed. Engl.*, **51**, 11002–11005.
- Singh, A. and Kukreti, S. (2018) A triple stranded G-quadruplex formation in the promoter region of human myosin beta (Myh7) gene. *J. Biomol. Struct. Dyn.*, **36**, 2773–2786.
- Kumar, M., Kaushik, M. and Kukreti, S. (2018) A topological transition from bimolecular quadruplex to G-triplex/tri-G-quadruplex exhibited by truncated double repeats of human telomere. *Eur. Biophys. J.*, **47**, 903–915.
- Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6622–6626.
- Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989) Conservation of the Human Telomere Sequence (Ttagg)_n among Vertebrates. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 7049–7053.
- Okazaki, S., Tsuchida, K., Maekawa, H., Ishikawa, H. and Fujiwara, H. (1993) Identification of a pentanucleotide telomeric sequence, (TTAGG)_n, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.*, **13**, 1424–1432.
- Zimmer, D.P. and Crothers, D.M. (1995) NMR of enzymatically synthesized uniformly ¹³C/¹⁵N-labeled DNA oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 3091–3095.
- Smith, D.E., Su, J.Y. and Jucker, F.M. (1997) Efficient enzymatic synthesis of ¹³C,¹⁵N-labeled DNA for NMR studies. *J. Biomol. NMR*, **10**, 245–253.
- Masse, J.E., Bortmann, P., Dieckmann, T. and Feigon, J. (1998) Simple, efficient protocol for enzymatic synthesis of uniformly ¹³C,¹⁵N-labeled DNA for heteronuclear NMR studies. *Nucleic Acids Res.*, **26**, 2618–2624.

21. Mer, G. and Chazin, W.J. (1998) Enzymatic synthesis of region-specific isotope-labeled DNA oligomers for NMR analysis. *J. Am. Chem. Soc.*, **120**, 607–608.
22. Skinner, S.P., Goult, B.T., Fogh, R.H., Boucher, W., Stevens, T.J., Laue, E.D. and Vuister, G.W. (2015) Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallogr. D-Struct. Biol.*, **71**, 154–161.
23. Levitt, M.H. (2012) In: *2nd ed., repr. with corr.* Wiley, Hoboken, NJ.
24. Kocman, V. and Plavec, J. (2014) A tetrahelical DNA fold adopted by tandem repeats of alternating GGG and GCG tracts. *Nat. Commun.*, **5**, 5831.
25. Abragam, A. and Abragam, A. (1961) In: *The Principles of Nuclear Magnetism*. Oxford University Press.
26. Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.
27. Wan, C., Fu, W., Jing, H. and Zhang, N. (2019) NMR solution structure of an asymmetric intermolecular leaped V-shape G-quadruplex: selective recognition of the d(G2NG3NG4) sequence motif by a short linear G-rich DNA probe. *Nucleic Acids Res.*, **47**, 1544–1556.
28. Schwieters, C.D., Bermejo, G.A. and Clore, G.M. (2018) Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.*, **27**, 26–40.
29. Bermejo, G.A., Clore, G.M. and Schwieters, C.D. (2016) Improving NMR Structures of RNA. *Structure*, **24**, 806–815.
30. Nielsen, J.T., Arar, K. and Petersen, M. (2009) Solution structure of a locked nucleic acid modified quadruplex: introducing the V4 folding topology. *Angew. Chem. Int. Ed. Engl.*, **48**, 3099–3103.
31. Clore, G.M. and Kuszewski, J. (2003) Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation. *J. Am. Chem. Soc.*, **125**, 1518–1525.
32. DeLano, W.L. and Bromberg, S. (2004) In: *PyMOL User's Guide*. DeLano Scientific LLC, San Carlos, California.
33. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment (15)N,(13)C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.
34. 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.
35. Cang, X., Sponer, J. and Cheatham, T.E. 3rd (2011) Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.*, **39**, 4499–4512.
36. Sponer, J., Mladek, A., Spackova, N., Cang, X., Cheatham, T.E. 3rd and Grimme, S. (2013) Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *J. Am. Chem. Soc.*, **135**, 9785–9796.
37. Chen, Y., Agrawal, P., Brown, R.V., Hatzakis, E., Hurley, L. and Yang, D. (2012) The major G-quadruplex formed in the human platelet-derived growth factor receptor beta promoter adopts a novel broken-strand structure in K⁺ solution. *J. Am. Chem. Soc.*, **134**, 13220–13223.
38. Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
39. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.*, **1**, 167–173.
40. Pavc, D., Wang, B.F., Spindler, L., Drevensek-Olenik, I., Plavec, J. and Sket, P. (2020) GC ends control topology of DNA G-quadruplexes and their cation-dependent assembly. *Nucleic Acids Res.*, **48**, 2749–2761.
41. Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738–9745.
42. Webba da Silva, M., Trajkovski, M., Sannohe, Y., Ma'ani Hessari, N., Sugiyama, H. and Plavec, J. (2009) Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem. Int. Ed. Engl.*, **48**, 9167–9170.
43. Karg, B., Haase, L., Funke, A., Dickerhoff, J. and Weisz, K. (2016) Observation of a dynamic G-Tetrad flip in intramolecular G-Quadruplexes. *Biochemistry*, **55**, 6949–6955.
44. Adrian, M., Ang, D.J., Lech, C.J., Heddi, B., Nicolas, A. and Phan, A.T. (2014) Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. *J. Am. Chem. Soc.*, **136**, 6297–6305.
45. Tran, P.L., Moriyama, R., Maruyama, A., Rayner, B. and Mergny, J.L. (2011) A mirror-image tetramolecular DNA quadruplex. *Chem. Commun. (Camb.)*, **47**, 5437–5439.
46. Moriyama, R., Shimada, N., Kano, A. and Maruyama, A. (2011) DNA assembly and re-assembly activated by cationic comb-type copolymer. *Biomaterials*, **32**, 2351–2358.
47. Hardin, C.C., Watson, T., Corregan, M. and Bailey, C. (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry*, **31**, 833–841.
48. Mergny, J.-L., Phan, A.-T. and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
49. Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
50. Sacca, B., Lacroix, L. and Mergny, J.L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.*, **33**, 1182–1192.
51. Marusic, M. and Plavec, J. (2015) The effect of DNA sequence directionality on G-quadruplex Folding. *Angew. Chem. Int. Ed. Engl.*, **54**, 11716–11719.
52. Tran, P.L., Mergny, J.L. and Alberti, P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res.*, **39**, 3282–3294.
53. Lat, P.K. and Sen, D. (2018) (C2G4)_n repeat expansion sequences from the C9orf72 gene form an unusual DNA higher-order structure in the pH range of 5–6. *PLoS One*, **13**, e0198418.
54. Sun, H., Xiang, J., Gai, W., Liu, Y., Guan, A., Yang, Q., Li, Q., Shang, Q., Su, H., Tang, Y. *et al.* (2013) Quantification of the Na⁺/K⁺ ratio based on the different response of a newly identified G-quadruplex to Na⁺ and K⁺. *Chem. Commun. (Camb.)*, **49**, 4510–4512.
55. Mao, X.A. and Gmeiner, W.H. (2005) NMR study of the folding-unfolding mechanism for the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *Biophys. Chem.*, **113**, 155–160.