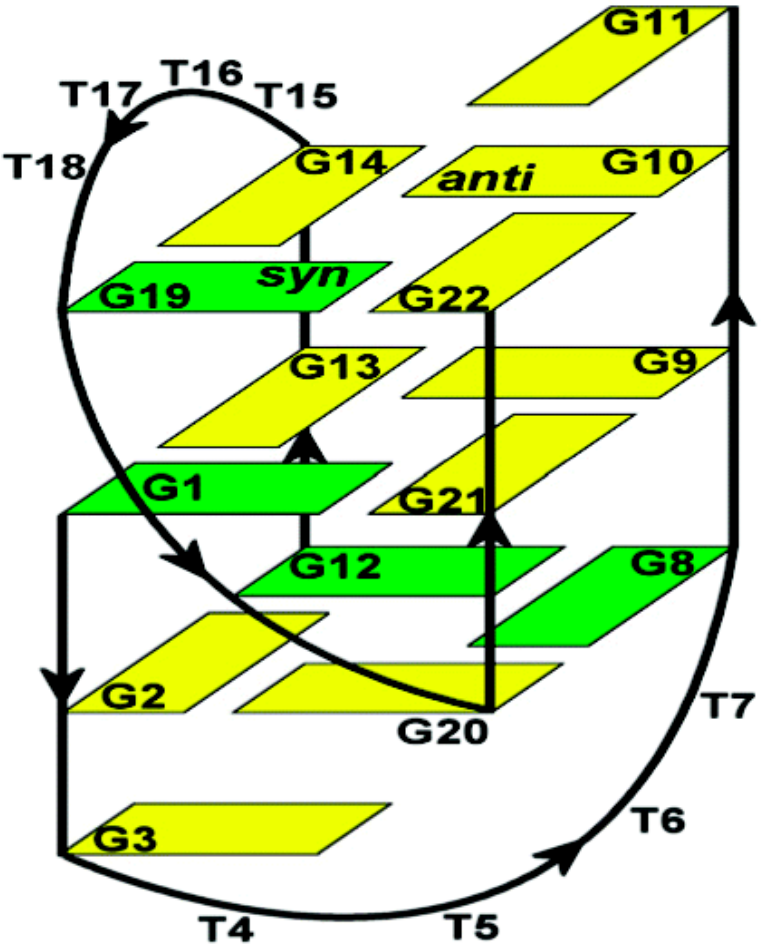
Article

Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations

Martin rnugelj, Primo ket, and Janez Plavec

J. Am. Chem. Soc., **2003**, 125 (26), 7866-7871• DOI: 10.1021/ja0348694 • Publication Date (Web): 07 June 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

J|A|C|S
ARTICLES
Published on Web 06/07/2003

Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations

Martin Črnugelj, Primož Šket, and Janez Plavec*

Contribution from the NMR center, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

Received February 26, 2003; E-mail: janez.plavec@ki.si

Abstract: NMR study has shown that DNA oligonucleotide d($G_3T_4G_4$) adopts an asymmetric bimolecular G-quadruplex structure in solution. The structure of d($G_3T_4G_4$)$_2$ is composed of three G-quartets, overhanging G11 residue and G3, which is part of the loop. Unique structural feature of d($G_3T_4G_4$)$_2$ fold is the orientation of the two loops. Thymidine residues T4-T7 form a diagonal loop, whereas T15-T18 form an edge type loop. The G-quadruplex core of d($G_3T_4G_4$)$_2$ consists of two stacked G-quartets with *syn-anti-anti-anti* alternation of dG residues and one G-quartet with *syn-syn-anti-anti* alternation. Another unusual structural feature of d($G_3T_4G_4$)$_2$ is a leap between G19 and G20 over the middle G-quartet and chain reversal between G19 and G20 residues. The presence of one antiparallel and three parallel strands reveals the hitherto unknown G-quadruplex folding motif consisting of antiparallel/parallel strands and diagonal as well as edge type loops. Further examination of the influence of different monovalent cations on the folding of d($G_3T_4G_4$) showed that it forms a bimolecular G-quadruplex in the presence of K^+, Na^+, and NH_4^+ ions with the same general fold.

Introduction

DNA can adopt several secondary structures besides the well-known canonical B-type duplex. Multistranded DNA structures have attracted considerable attention during the last years. Four-stranded DNA structures have been suggested to play a role in a number of biological processes such as telomere function, genetic recombination, transcription, and replication, and they have been also established in vitro in DNA sequences associated with human diseases.[1−4] The potential role of G-quadruplex in vivo has stimulated development of therapeutic strategies targeted into the design of small molecules that specifically stabilize their formation.[5−9] Recent study has provided evidence of G-quadruplex formation in vivo.[10] Telomeres, protein-DNA complexes located at the ends of chromosomes, contain DNA with G-rich sequences that can form four-stranded G-quadruplex structures in vitro.[11−13] It has been suggested that formation of G-quadruplex DNA structures may play a role in regulating the maintenance of telomere length by the ribonucleoprotein telomerase.[4,14−17] Telomeres have been implicated in fundamental health problems such as cancer, organ failure, premature aging, and early death, which make them an attractive therapeutic target.[2] Recently, DNA duplex-quadruplex exchange has been exploited in the design of nanodevices that could perform linear and rotary movements.[18,19] Understanding factors that govern DNA self-assembly processes and details of their 3D structures at the molecular level could add considerably to the route toward construction of artificial nanomechanical devices.

The most studied four-stranded motif in DNA is the G-quadruplex, where four guanines are hydrogen-bonded through their Hoogsteen and Watson−Crick sides thus forming a planar array. Numerous biophysical and structural studies have shown that two or more hydrogen-bonded arrays of four guanines are perfectly predisposed to stack on each other thus forming a G-quadruplex.[20] Intermolecular G-quadruplexes are formed through association of two or four separate strands. The first

(1) Arthanari, H.; Bolton, P. H. *Chem. Biol.* **2001**, *8*, 221−230.
(2) Neidle, S.; Read, M. A. *Biopolymers* **2001**, *56*, 195−208.
(3) Shafer, R. H.; Smirnov, I. *Biopolymers* **2001**, *56*, 209−227.
(4) Rezler, E. M.; Bearss, D. J.; Hurley, L. H. *Curr. Opin. Pharmacol.* **2002**, *2*, 415−423.
(5) Han, H. Y.; Hurley, L. H. *Trends Pharmacol. Sci.* **2000**, *21*, 136−142.
(6) Mergny, J. L.; Lacroix, L.; Teulade-Fichou, M. P.; Hounsou, C.; Guittat, L.; Hoarau, M.; Arimondo, P. B.; Vigneron, J. P.; Lehn, J. M.; Riou, J. F.; Garestier, T.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3062−3067.
(7) Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4844−4849.
(8) Kim, M. Y.; Vankayalapati, H.; Kazuo, S.; Wierzba, K.; Hurley, L. H. *J. Am. Chem. Soc.* **2002**, *124*, 2098−2099.
(9) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593−11598.
(10) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8572−8577.
(11) Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703−730.
(12) Bryan, T. M.; Cech, T. R. *Curr. Opinion Cell Biol.* **1999**, *11*, 318−324.
(13) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876−880.
(14) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621−628.
(15) Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.* **2002**, *61*, 1154−1162.
(16) Kerwin, S. M.; Chen, G.; Kern, J. T.; Thomas, P. W. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 447−450.
(17) Mergny, J. L.; Riou, J. F.; Mailliet, P.; Teulade-Fichou, M. P.; Gilson, E. *Nucleic Acids Res.* **2002**, *30*, 839−865.
(18) Alberti, P.; Mergny, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1569−1573.
(19) Li, J. W. J.; Tan, W. H. *Nano Lett.* **2002**, *2*, 315−318.
(20) Hardin, C. C.; Perry, A. G.; White, K. *Biopolymers* **2000**, *56*, 147−194.

10.1021/ja0348694 CCC: $25.00

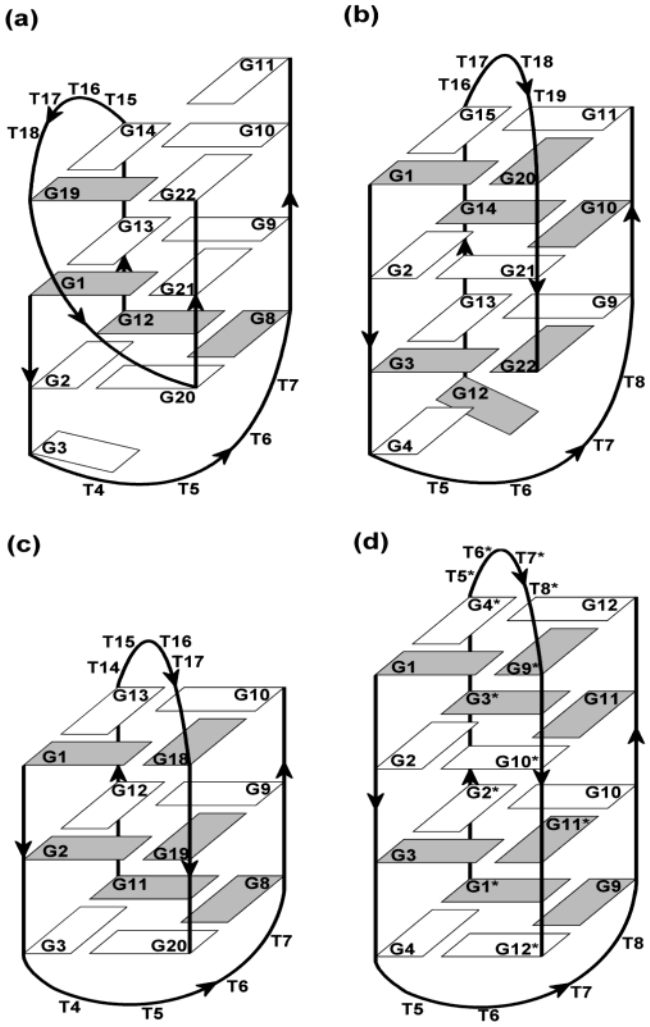


Figure 1. Folding topologies of G-quadruplex structures formed by (a) $d(G_3T_4G_4)_2$, (b) $d(G_4T_4G_3)_2$,[28] (c) $d(G_3T_4G_3)_2$,[26,27] and (d) $d(G_4T_4G_4)_2$.[21,23−25] Strand directions are indicated by arrows. The guanine bases are shown as rectangles, where shaded rectangles represent *syn* nucleobases. For clarity, thymine bases are not shown.

high-resolution structures of G-quadruplexes were reported over a decade ago for the sequence $d(G_4T_4G_4)$ by both X-ray crystallography and NMR spectroscopy.[21−23] Subsequent NMR and crystallographic studies have shown that topology and strand orientations within G-quadruplex structures depend on the sequence, strand concentration and number of G-rich repeats (e.g., unimolecular vs multimolecular). DNA oligonucleotides containing two G-tracts separated by a short run of thymine residues have been shown to often adopt bimolecular G-quadruplexes, where the thymine residues act as hairpin loops such that the two G tracts of each strand fold-back to form the G-quartets at the core of the quadruplex. Sequence details and the nature of metal ions play a major role in the formation of G-quadruplex structures. For example, $d(G_4T_4G_4)$ forms a symmetric G-quadruplex containing four stacked G-quartets with two T_4 segments looped diagonally across the two faces of the stacked G-quartets (Figure 1d).[21,23−25] The removal of both 5′ and 3′ terminal dG residues from $d(G_4T_4G_4)$ results in the removal of the central G-quartet to create an asymmetric fold of $d(G_3T_4G_3)_2$ (compare Figure 1, parts c and d).[26,27] We have recently shown that $d(G_4T_4G_3)$, a sequence with only 3′ terminal dG residue missing from $d(G_4T_4G_4)$, forms G-quadruplex that consists of three G-quartets while two dG residues reside on the same side of the end G-quartet (Figure 1b).[28]

In the present work, we describe a completely unprecedented folding topology and loop orientations of the sequence with double G-rich repeats. The 3D folding topology of $d(G_3T_4G_4)_2$ is significantly different from those of the closely related molecules $d(G_4T_4G_4)_2$, $d(G_3T_4G_3)_2$, and $d(G_4T_4G_3)_2$ under similar conditions. The present study provides both insight into the role of 5′ terminal dG residue in the formation of G-quadruplex by sequences containing two G-tracts, and clearly shows that there are new rules to be discovered that govern the folding of DNA G-rich sequences. Another important aspect of this study is examination of the influence of different monovalent cations on the folding of sequences with two G-tracts like $d(G_3T_4G_4)$. Monovalent cations stabilize G-quadruplex structures typically through the coordination with eight carbonyl groups.[23,29] Understanding metal−ion interactions with nucleic acids is of utmost general interest since some G-rich sequences adopt similar structures, whereas others form different structures or assemble in some other more complex way in the presence of different counterions.[28,30−36] The present study shows that $d(G_3T_4G_4)$ adopts the same G-quadruplex structure in the presence of K^+, Na^+, and NH_4^+ ions.

Methods

Experimental Details. Oligonucleotides were synthesized on an Expedite 8909 synthesizer using phosphoramidite chemistry following the manufacturer's protocol and deblocked with concentrated aqueous ammonia. DNA was purified on 1.0 m Sephadex G25 column. Fractions containing only full-length oligonucleotide were pooled, lyophilized, redissolved in 1 mL H_2O and dialyzed extensively against LiCl solution. Sample concentrations were between 1 and 5 mM in strand (0.5−2.5 mM in G-quadruplex). NMR spectra were collected on a Varian Unity Inova 600 MHz NMR spectrometer. We recorded NOESY spectra in H_2O and ^2H_2O at several mixing times, double-quantum filtered COSY (DQF−COSY) and other standard 1D and 2D NMR experiments in ^2H_2O at 278 and 298 K. Spectra were processed and analyzed using the FELIX program (Accelrys Inc.) and VNMR 6.1B software (Varian).

NMR Restraints. Distance restraints for nonexchangeable protons were derived from NOESY spectra acquired at mixing times of 80, 150, and 250 ms. The upper and lower bounds were assigned to ±30%. NOE cross-peaks involving exchangeable protons were derived from NOESY spectra in 10% ^2H_2O at mixing times of 80 and 300 ms. These

(21) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164−168.
(22) Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126−131.
(23) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189−200.
(24) Schultze, P.; Smith, F. W.; Feigon, J. *Structure* **1994**, *2*, 221−233.
(25) Smith, F. W.; Feigon, J. *Biochemistry* **1993**, *32*, 8682−8692.
(26) Strahan, G. D.; Shafer, R. H.; Keniry, M. A. *Nucleic Acids Res.* **1994**, *22*, 5447−5455.
(27) Smith, F. W.; Lau, F. W.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10 546−10 550.
(28) Crnugelj, M.; Hud, N. V.; Plavec, J. *J. Mol. Biol.* **2002**, *320*, 911−924.
(29) Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171−182.
(30) Hud, N. V.; Smith, F. W.; Anet, F. A. L.; Feigon, J. *Biochemistry* **1996**, *35*, 15 383−15 390.
(31) Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018−3028.
(32) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 619−636.
(33) Bouaziz, S.; Kettani, A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 637−652.
(34) Miyoshi, D.; Nakao, A.; Toda, T.; Sugimoto, N. *FEBS Lett.* **2001**, *496*, 128−133.
(35) Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998**, *75*, 968−981.
(36) Keniry, M. A.; Strahan, G. D.; Owen, E. A.; Shafer, R. H. *Eur. J. Biochem.* **1995**, *233*, 631−643.

were classified as strong (strong intensity at 80 ms), medium (weak intensity at 80 ms) and weak (peak observed only at 300 ms) and were restrained to distances of 2.7 (±0.9), 3.8 (±1.2), and 5.0 (±1.5) Å, respectively. A total of 414 distance restraints, of which 79 involved exchangeable protons, were derived from NOESY spectra corresponding to an average of 19 restraints per residue. In each of the three G-quartets, an additional eight hydrogen-bond restraints were applied to retain hydrogen-bonding within individual G-quartet. In addition, 66 torsion angle restraints for sugar moieties were derived from $^3J_{HH}$ coupling constants, which were obtained from P. E. COSY spectra and interpreted using PSEUROT procedure.[37]

Structure Determination. All calculations were performed using the AMBER 7[38] program with a Cornell et al. (1999) force field.[39,40] Five different initial structures were created manually using molecular modeling tools within the InsightII package (Accelrys Inc.) and the XLEAP module within AMBER 7 package.[38] For each structure, 10 simulated annealing calculations with NMR restraints using a generalized Born (GB) implicit solvation model[41−43] were performed for 60 ps with a time step of 1 fs. Initial velocities were assigned from a Maxwellian distribution at 300 K. In the first 5 ps, the molecule was heated to 700 K with NOE distance force constants concomitantly increased from 25 to 50 kcal mol^{-1} Å^{-2}. These conditions were maintained for 30 ps, after which the temperature was reduced over the course of 25 ps to 0 K. Force constants for torsion angle restraints were increased from 100 kcal mol^{-1} rad^{-2} to 200 kcal mol^{-1} rad^{-2} in the first 5 ps and then maintained at this value for the rest of the simulation. The temperature of the system was maintained by the Berendsen coupling algorithm with varying time constant (0.4 ps to heat up the system, 1.0 ps during the sampling of the conformational space, and 0.2 ps for the last part of simulation). All atoms within a 20 Å radius were included in nonbonded interactions. SHAKE algorithm was applied on the hydrogen atoms with tolerance of 0.0005 Å. The resulting fifty structures were subjected to 1000 steps of conjugate gradient minimization. All 50 of the final structures were very similar (rmsd below 1.3 Å) and satisfied experimental NMR restraints equally well (all violations of NOE distance restraints were below 0.4 Å). Eight structures of d($G_3T_4G_4$)$_2$, which best satisfied the NMR restraints and had the lowest energies were analyzed using the Carnal module of AMBER 7 program.

Results

The formation of a G-quadruplex structure by d($G_3T_4G_4$) was initially followed by monitoring proton resonances as a function of increasing potassium ion concentration. Twelve narrow and well-resolved guanine imino resonances were observed in the 11.18 to 11.95 ppm range of the ^{1}H NMR spectrum (Figure 2), which is characteristic of G-quadruplex formation with guanine residues involved in Hoogsteen base paring. The number of imino protons indicated that the structure possibly contains three G-quartets or two full and two incomplete G-quartets.

The perusal of aromatic-anomeric region of 2D NOESY spectra clearly ascertained four nucleobases in *syn* orientation,

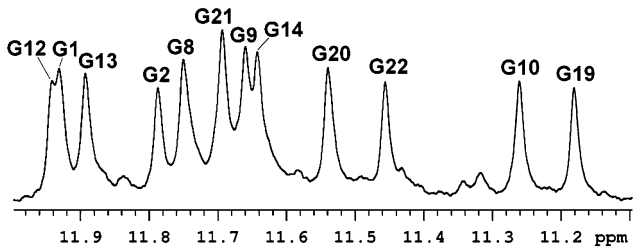


Figure 2. Imino region of the ^{1}H NMR spectrum of d($G_3T_4G_4$)$_2$ at 25 °C in 10% ^2H_2O (10 mM KCl, pH = 5.5). The sample was 4 mM in strand (formation of G-quadruplex was not concentration dependent).

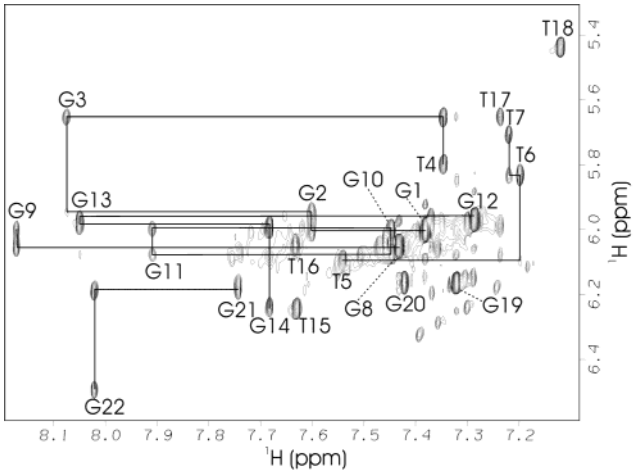


Figure 3. Portion of a NOESY spectrum of d($G_3T_4G_4$)$_2$ at 25 °C and τ_m = 400 ms showing the aromatic-H1′ sequential connectivities.

which was consistent with a fold-back structure.[11,26,28,44,45] Four observed *syn* residues clearly implied *syn-anti* alternation inside G-quartets and along the strands in d($G_3T_4G_4$)$_2$ that is different from closely related structures of d($G_4T_4G_4$)$_2$, d($G_3T_4G_3$)$_2$, and d($G_4T_4G_3$)$_2$. The dimeric nature of d($G_3T_4G_4$)$_2$ fold was confirmed by NMR diffusion measurements. The plots of natural logarithm of intensities of aromatic, sugar and methyl resonances as a function of square of gradient strength in PFG experiments afforded straight lines with slopes giving translational diffusion constants (D_t). The comparison of translational diffusion constants for d($G_3T_4G_4$)$_2$ (D_t = 1.8 (±0.1) × 10^{-6} cm^2 s^{-1}) with those of the closely related known G-quadruplex structures of d($G_4T_4G_4$)$_2$[21,23−25] (D_t = 1.5 × 10^{-6} cm^2 s^{-1}) and d($G_4T_4G_3$)$_2$[28] (D_t = 1.7 × 10^{-6} cm^2 s^{-1}) clearly showed that d($G_3T_4G_4$) sequence folded into a dimeric structure.

The NMR resonance assignment of d($G_3T_4G_4$)$_2$ was challenging not only because the 5′-3′ sequential H8/H6−H1′ NOE connectivities were interrupted at guanine residues with *syn* conformation (Figure 3), but also due to the unexpected structural features. The examination of NOESY spectra at several mixing times, however, enabled us to initially identify the following 5′-3′ sequential NOE connectivities: G*syn*-G*anti*-G*anti*-T*anti*, G*syn*-G*anti*-G*anti*, T*anti*-T*anti*-T*anti*, G*syn*-G*anti* and two G*anti*-G*anti* steps (Figure 3).

Methyl resonances of T7 and T18 showed NOEs to their neighboring G-quartets (see Figure S1). Examination of imino-aromatic region of NOESY spectrum clearly established the hydrogen-bonding network within the three G-quartets (see

(37) Haasnoot, C.; de Leeuw, F. A. A. M.; Huckriede, D.; van Wijk, J.; Altona, C., PSEUROT program, version 6.0, Leiden University: Leiden, Netherlands, 1993.
(38) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Chetham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Mertz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A., AMBER program, version 7, University of California: San Francisco, 2002.
(39) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(40) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(41) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.
(42) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275−291.
(43) Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317−331.
(44) Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1995**, *251*, 76−94.
(45) Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225−255.

Figure S2). The use of sequential NOE information and assuming possible topologies based on the known bimolecular G-quadruplex structures did not enable us to propose the dimeric fold-back structure of d($G_3T_4G_4$)$_2$ that would agree with all available NMR data. We have furthermore synthesized 10 dI and dU analogues of d($G_3T_4G_4$) sequence where single guanine or thymine residue was replaced in a systematic fashion.[45,46] Inosine for guanine substitution resulted in a large downfield shift of the imino proton and a smaller shift of H8 proton which assisted the identification of these resonances for the guanine base that has undergone substitution (see Figure S3). At the same time, substitution of a single guanine by inosine introduced a new aromatic proton (i.e., H2) that clarified its spatial proximities with respect to the neighboring nucleobases. Individual uridine for thymine substitution introduced a new proton (i.e., H5) that gave strong cross-peak in the aromatic-anomeric region of DQF-COSY spectrum which made it readily distinguishable from the intraresidual and sequential NOEs in that region. In addition, uridine for thymine substitution resulted in disappearance of its methyl resonance. Although the above systematic substitutions enabled us to assign all thymine residues, there were still some ambiguities in the assignment of guanines due to the chemical shift changes in dI analogues. To complete the assignment, we resorted to the residue specific ^{15}N labeled (15%) samples[47] of d($G_3T_4G_4$)$_2$. Partial ^{15}N isotopic labeling enabled us to unequivocally assign specific imino resonance to a particular guanine in a sequence with the use of one-bond ^{15}N$-^{1}$H correlation established by HSQC experiment (see Figure S4). Furthermore, partially ^{15}N labeled guanine helped us to establish correlations via two bonds in ^{15}N$-^{1}$H HMBC experiments, which gave the residue specific assignment of H8 resonances.

The search for topology model of d($G_3T_4G_4$)$_2$ prompted us to search for new rules of parallel and antiparallel strand orientations.[26,45,48] If d($G_3T_4G_4$)$_2$ was to assume dimeric G-quadruplex with antiparallel strands, then G22 and T15-T18 loop residues would be on the opposite sides of the G-quadruplex structure. The observed NOE contacts between G22 and T18 residues and medium NOE between G22(H8) and G19(NH) clearly indicated their spatial proximity (see Figures S1 and S2). These structural restraints could only be fulfilled by uncommon reversal in the strand direction.

335 distance restraints associated with the nonexchangeable protons were quantified from NOESY spectra in $^{2}H_2O$ at three mixing times. Additional 87 distance restraints involving exchangeable protons were calculated from NOESY spectra in 10% $^{2}H_2O$ at two mixing times. 24 hydrogen-bond restraints were applied within the three G quartets to retain hydrogen-bonds between Hoogsteen paired guanine bases (i.e., 8 restraints per quartet). In addition, 66 torsion angle restraints for sugar moieties were derived from $^{3}J_{HH}$ coupling constants. The sugar moieties of all residues showed a strong bias of over 70% toward South-type conformers (i.e., around C2′-*endo* puckered form) and were uniformly constrained during calculations to assume puckering in the South region of conformational space. The above experimental restraints were used to calculate the structure of d($G_3T_4G_4$)$_2$ G-quadruplex by using simulated annealing protocol. The resulting fifty structures were subjected to energy minimization. All fifty final structures were very similar (rmsd below 1.3 Å) and satisfied experimental NMR restraints equally well (all violations of NOE distance restraints were below 0.4 Å). Eight representative structures of d($G_3T_4G_4$)$_2$, which best satisfied NMR restraints and had the lowest energies are shown in Figure 4a.

Perusal of the lowest energy structures of d($G_3T_4G_4$)$_2$ shows that three G-quartets adopt nearly identical structures, whereas T_4 loops exhibit some conformational flexibility. In particular, G3-T7 loop that spans the diagonal of the end G-quartet is less well defined (shown in red in Figure 4a). Residues T15-T18 constituting an edge type loop are better defined due to the larger amount of NOE contacts (shown in green in Figure 4a). Figure 4b shows that G3-T7 loop adopts the structure in which T6 and T7 stack on the neighboring G12 and G20 residues, respectively. T4 and T5 are stacked, whereas G3 is not involved in any stacking interactions with its purine ring (Figure 4b). Residues T15 and T18 of the edge type loop are stacked on G14 and G19 residues, respectively, that constitute the neighboring G-quartet, whereas T16 and T17 are stacked on each other (Figure 4c).

Discussion

The solution structure of d($G_3T_4G_4$)$_2$ that is consistent with all NMR data is composed of three G-quartets, overhanging G11 residue and G3, which is part of the loop (Figure 4a). Unique structural feature of d($G_3T_4G_4$)$_2$ fold is the orientation of the two loops. Thymidine residues T4-T7 form a diagonal loop, whereas T15-T18 form an edge type loop (Figure 4, parts b and c). To the best of our knowledge, this is the first bimolecular G-quadruplex comprised of diagonal as well as edge type loop. The G-quadruplex core consists of two stacked G-quartets with *syn-anti-anti-anti* dGs and one G-quartet with *syn-syn-anti-anti* alternation. The G-quadruplexes of closely related d($G_4T_4G_3$), d($G_4T_4G_4$), and d($G_3T_4G_3$) sequences exhibit *syn-syn-anti-anti* alternation inside G-quartets.[21,24−28] Another unusual structural feature of d($G_3T_4G_4$)$_2$ is a leap and chain reversal between G19 and G20 residues. Although this sharp turn resembles the so-called V-shaped scaffold,[46] it is the leap between G19 and G20 over the middle G-quartet that makes the structure of d($G_3T_4G_4$)$_2$ so unique. NMR restrained simulated annealing calculations show that these unique structural features can be easily accommodated and do not disturb the planarities of the G-quartets consisting of G10-G14-G19-G22 and G2-G20-G8-G12 (Figure 4a). The presence of one antiparallel and three parallel strands suggests that the folding into a G-quadruplex is directed toward maximizing the number of G-quartets. In comparison, d(TAG_3TTAG_3T) sequence has been shown to form G-quadruplex consisting of three G-quartets with parallel strands and TTA loops that project outward adopting a shape of a molecular propeller.[13] The loops in this structure that has been determined by X-ray crystallography connect the top of one strand with the bottom of the other which ensures that all strands are parallel. Similarly, intramolecular G-quadruplex formed by the four-repeat human telomere DNA sequence d[AG_3($TTAG_3$)$_3$] has all four strands in parallel arrangement.[13] Earlier analysis by NMR of Na^+ form of the same 22-mer sequence in solution has shown a completely different G-quadruplex topology with

(46) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. *J. Mol. Biol.* **2001**, *311*, 1063−1079.
(47) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160−1161.
(48) Marathias, V. M.; Bolton, P. H. *Biochemistry* **1999**, *38*, 4355−4364.

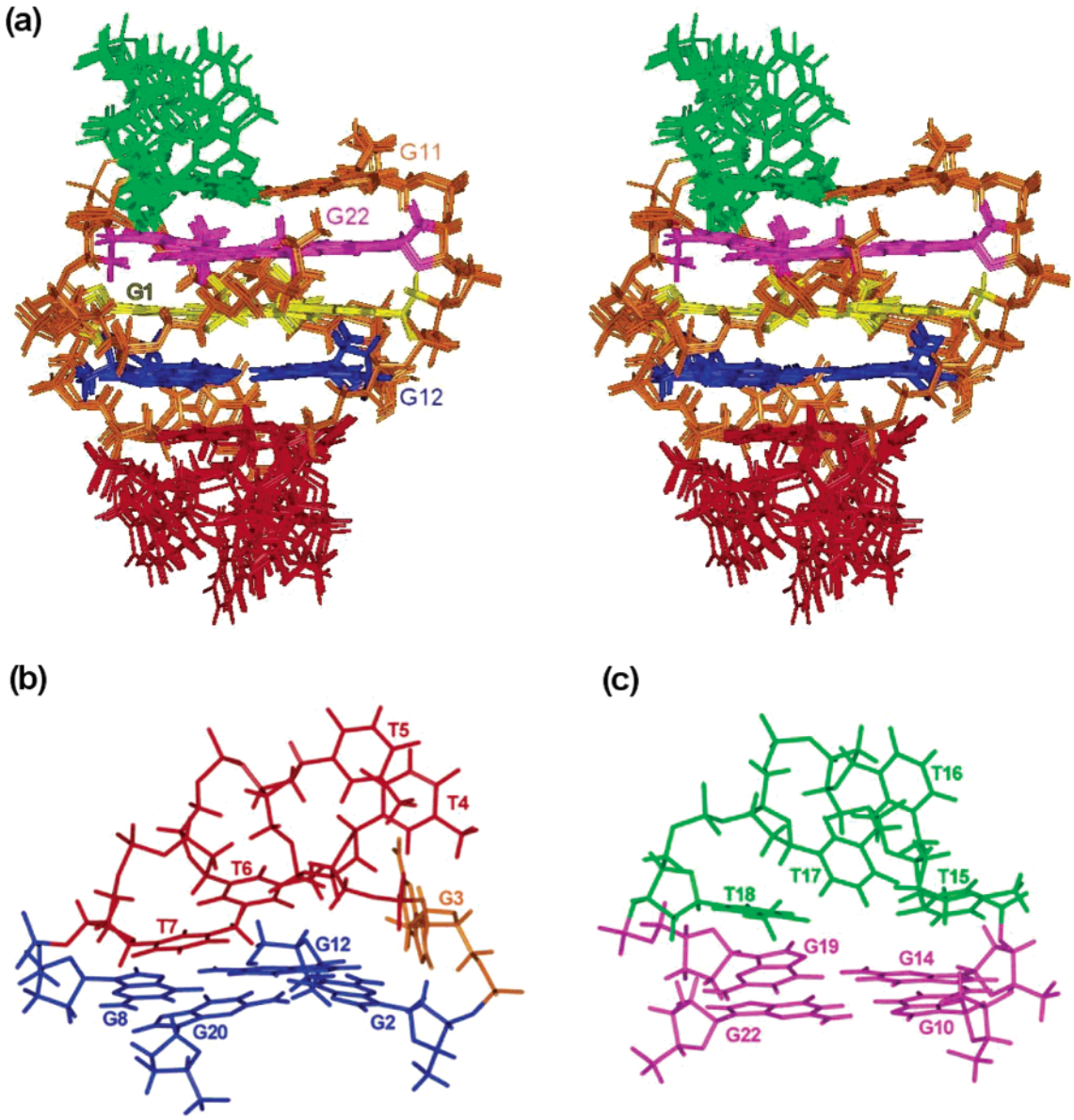


Figure 4. (a) Stereoview of superposition of eight refined structures of $d(G_3T_4G_4)_2$ G-quadruplex. End residues of each strand are labeled. Pairwise all atom rmsd between eight structures was 0.76 ± 0.42 Å. Without G3-T7 loop residues rmsd reduces to 0.45 ± 0.11 Å. Views of the (b) G3-T7 diagonal loop and (c) T15-T18 edge type loop with the neighboring G-quartets in the representative structure of the $d(G_3T_4G_4)_2$ G-quadruplex.

antiparallel strands.[49] In K^+-stabilized structure, all loops connect the top of one strand with the bottom of the other, whereas Na^+ stabilized structure has one diagonal and two edge type loops that connect opposing G-rich tracts. The distinct topologies that are adopted in the presence of K^+ and Na^+ ions with their different loop orientations could have important implications in further intermolecular assembly and higher order packing. Loops on the outside of the G-quadruplex core probably facilitate stacking of individual G-quadruplexes though large planar surfaces. On the other hand, diagonal and edge type loops are expected to hinder G-quadruplexes in stacking up on each other.

A chain reversal in dimeric G-quadruplex structure of $d(TAG_3TTAG_3T)_2$ that consists of two human telomere repeat is achieved with particular conformation of TTA loops.[13] In the structure of $d(G_3T_4G_4)_2$, there is a leap between the two dGs over the central G-quartet without any extra intervening residues. The comparison of structural information to date on bimolecular G-quadruplexes shows that structures formed from telomeric DNA sequence can be divided in four categories. First, sequences such as $d(G_3T_4G_3)$ and $d(G_4T_4G_4)$ adopt topologies with two antiparallel and two parallel strands and two diagonal loops.[21,23−27] Second, category exhibits antiparallel/parallel pattern and two edge type loops adopted by sequence of $d(G_3CT_4G_3C)$.[32,33] and third, having all-parallel strands and external loops adopted by sequences with human telomere repeats.[13] In the present work, we reveal the hitherto unknown fourth category having one antiparallel and three parallel strands with diagonal as well as edge type loops.

We have also examined the influence of different monovalent cations on the folding of $d(G_3T_4G_4)$. It turned out that $d(G_3T_4G_4)$ forms a bimolecular G-quadruplex in the presence of K^+, Na^+, and NH_4^+ ions with the same general fold. A related $d(G_4T_4G_3)$

(49) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263−282.

sequence forms G-quadruplex in the presence of Na^+ ions whereas it does not fold into a single structure in the presence of K^+ or NH_4^+ ions.[28] Related $d(G_3T_4G_3)$ and $d(G_4T_4G_4)$ sequences both adopt the same folds in the presence of K^+, Na^+, and NH_4^+ ions.[30,31] Further studies on specific indications of the metal ion binding into the cavities of $d(G_3T_4G_4)_2$ are in progress.

Conclusions

The present study describes a novel folding pattern of a DNA oligonucleotide $d(G_3T_4G_4)$, which adopts an asymmetric bimolecular G-quadruplex structure in solution. The solution structure of $d(G_3T_4G_4)_2$ that is consistent with all NMR data is composed of three G-quartets, overhanging G11 residue and G3, which is part of the loop. Unique structural feature of $d(G_3T_4G_4)_2$ fold is the orientation of the two loops. Thymidine residues T4-T7 form a diagonal loop, while T15-T18 form an edge type loop. To the best of our knowledge this is the first bimolecular G-quadruplex where the end G-quartets are spanned by diagonal as well as edge type loop in a single structure.

$d(G_3T_4G_4)$ forms a bimolecular G-quadruplex in the presence of K^+, Na^+, and NH_4^+ ions with the same general fold. Our preliminary investigations of the cation binding properties of $d(G_3T_4G_4)_2$ also indicate that the number of cation binding sites within $d(G_3T_4G_4)_2$ and the binding dynamics of cations differs from those determined for other related G-rich DNA sequences.

Four-stranded DNA structures have been implicated in many biological processes. Recently their self-assembly was utilized in the design of nanomotors where conformational change leads to motion. Understanding the factors that influence details of particular conformation and its 3D structure can add significantly to our insight into molecular recognition and intermolecular assembly. As more structural data becomes available we will get insights into the pathways for folding and unfolding of G-quadruplex structures. We feel that the results of our study will be important in the design of new selective ligands that will stabilize the formation of G-quadruplex structures and potentially inhibit telomerase function through novel binding modes exploiting unique loop conformations and chemical functionalities that are accessible in the grooves.

Acknowledgment. We thank the Ministry of Education, Science and Sport of the Republic of Slovenia (Grant No. J1-3309-0104) and European Commission (Contract No. ICA1-CT-2000-70034) for their financial support.

Supporting Information Available: Regions of NOESY spectra, spectra of oligonucleotides with single dI substitutions and spectra of ^{15}N labeled samples. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0348694