J|A|C|S
ARTICLES
Published on Web 04/04/2006

Topology Variation and Loop Structural Homology in Crystal and Simulated Structures of a Bimolecular DNA Quadruplex

Pascale Hazel, Gary N. Parkinson, and Stephen Neidle*

Contribution from the Cancer Research U.K. Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, U.K.

Received December 19, 2005; E-mail: stephen.neidle@pharmacy.ac.uk

Abstract: The topology of DNA quadruplexes depends on the nature and number of the nucleotides linking G-quartet motifs. To assess the effects of a three-nucleotide TTT linker, the crystal structure of the DNA sequence d($G_4T_3G_4$) has been determined at 1.5 Å resolution, together with that of the brominated analogue d($G_4^{Br}UTTG_4$) at 2.4 Å resolution. Both sequences form bimolecular intermolecular G-quadruplexes with lateral loops. d($G_4^{Br}UTTG_4$) crystallized in the monoclinic space group $P2_1$ with three quadruplex molecules in the asymmetric unit, two associating together as a head-to-head stacked dimer, and the third as a single head-to-tail dimer. The head-to-head dimers have two lateral loops on the same G-quadruplex face and form an eight-G-quartet stack, with a linear array of seven K^+ ions between the quartets. d($G_4T_3G_4$) crystallized in the orthorhombic space group $C222$ and has a structure very similar to the head-to-tail dimer in the $P2_1$ unit cell. The sequence studied here is able to form several different folds; however, all four quadruplexes in the two structures have lateral loops, in contrast to the diagonal loops reported for the analogous quadruplex with T_4 loops. A total of seven independent T_3 loops were observed in the two structures. These can be classified into two discrete conformational classes, suggesting that these represent preferred loop conformations that are independent of crystal-packing forces.

G-quadruplex structures are formed from guanine-rich nucleic acid sequences.[1−5] They comprise Hoogsteen hydrogen-bonded guanine quartet motifs formed from the guanine tracts in these sequences, together with intervening sequences which form extrahelical loops and help to hold together the G-quartets. Quadruplexes can be formed by the intramolecular folding of a single strand or by the intermolecular association of two or four separate strands.

Potential quadruplex-forming G-rich sequences occur throughout the genomes of many species,[6,7] as well as at the telomeric ends of eukaryotic chromosomes. There is considerable diversity within these sequences, especially in the loops of nontelomeric ones, suggestive of diversity in quadruplex structures themselves. In addition, quadruplexes from any one sequence are often capable of conformational plurality; for example, the human telomeric repeat T_2AG_3 forms parallel G-quadruplex structures in the crystal[8] and mixed parallel/antiparallel structures in solution.[9−12] Other G-rich sequences which have been shown to form G-quadruplex structures include the nuclease hypersensitivity promoter element III1 (NHE III1) upstream of transcription start site in the *c-myc* gene,[13−17] a promoter sequence in the human insulin gene,[18] sequences from immunoglobulin switch regions,[19] and in promoter sequences of several muscle-specific genes.[20]

NMR,[9,10,16,17] melting studies,[21,22] and CD spectroscopy[11,23] are commonly used to obtain information about the forming and particular topology of G-quadruplexes in solution. In several cases, multiple conformations have been found for a single sequence, notably ones containing the human telomeric repeat T_2AG_3; these are suggestive of multiple quadruplex substates or folding pathways.[9−12] Similarly, the *c-myc* G-quadruplex has also been suggested to convert between parallel and mixed

(1) Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441−478.
(2) Simonsson, T. *Biol. Chem.* **2001**, *382*, 621−628.
(3) Phan, A. T.; Mergny, J.-L. *Nucleic Acids Res.* **2002**, *30*, 4618−4625.
(4) Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275−283.
(5) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668−698.
(6) Huppert, J.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2901−2907.
(7) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901−2907.
(8) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876−880.
(9) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263−282.
(10) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021−15027.
(11) Rujan, I. N.; Meleney, J. C.; Bolton, P. H. *Nucleic Acids Res.* **2005**, *33*, 2022−2031.
(12) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14629−14634.
(13) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167−1172.
(14) Rangan, A.; Fedoroff, O. Y.; Hurley, L. H. *J. Biol. Chem.* **2001**, *276*, 4640−4646.
(15) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593−11598.
(16) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.
(17) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048−2058.
(18) Catasti, P.; Chen, X.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *J. Mol. Biol.* **1996**, *264*, 534−545.
(19) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364−366.
(20) Yafe, A.; Etzioni, S.; Weisman-Shomer, P.; Fry, M. *Nucleic Acids Res.* **2005**, *33*, 2887−2900.
(21) Mergny, J.-L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74−78.
(22) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584−10589.
(23) Dapic, V.; Abdomerovic, V.; Marrington, R.; Peberdy, J.; Rodger, A.; Trent, J. O.; Bates, P. J. *Nucleic Acids Res.* **2003**, *31*, 2097−2107.

10.1021/ja058577+ CCC: $33.50

parallel/antiparallel quadruplex forms, this time upon addition of a ligand.[24] There is also evidence that ligands are capable of inducing a particular fold in human telomeric DNA quadruplexes that optimizes quadruplex binding to the ligand.[25,26]

The structure of a number of G-quadruplexes has been solved using X-ray crystallography and NMR spectroscopy, revealing their diverse topologies and conformational flexibility.[3,4,26] Differences in strand polarities can affect the location of the linkers, or loops, between G-rich segments. Thus, parallel G-strands are necessarily linked with double-strand reversal (propeller) loops, as seen in the d[$AG_3(T_2AG_3)_3$] and d(TAGGGTTAGGGT) crystal and NMR structures[8,10] and the *c-myc* NMR structures.[16,17,27] Antiparallel G-strands can be linked either by diagonal or lateral loops, depending on whether the strands are adjacent or diagonally opposed. In general the nature of the loops (sequence and size) plays a key role in determining the topology of quadruplexes.[28−30] Loop residues can themselves form stacking and hydrogen-bonding interactions, further stabilizing or destabilizing particular G-quadruplex folds.[29−31]

The crystal structure of the bimolecular quadruplex formed by d($G_4T_4G_4$) has been previously reported by us[32] and compared with other high-resolution structure analyses of this quadruplex.[33−35] NMR and X-ray structures concur in showing that d($G_4T_4G_4$) forms in both Na^+ and K^+ environments an identically folded bimolecular G-quadruplex with two diagonal loops each containing four thymine nucleotides. The quadruplex from the closely related sequence d($G_3T_4G_3$) has the same fold,[36] although other small changes to this sequence can result in major structural changes, as found for the bimolecular quadruplexes formed from the sequences d($G_4T_4G_3$), d($G_3T_4G_4$), and d($T_2G_4T_2G_4T$), respectively.[37−39]

We have previously studied the effects of systematic changes in loop length on intramolecular quadruplex folding and topology,[29] using a combined experimental biophysics and molecular dynamics (MD) simulation analysis approach. We now report crystallographic and molecular simulation studies on loop changes in a bimolecular quadruplex and have examined the effects of a conservative shortening of the T_4 loop between the two G-tracts in the *Oxytricha* sequence by one nucleotide to T_3, i.e., to form the sequence d($G_4T_3G_4$). An understanding of the influence of loop size and sequence is important for the establishment of general guidelines for the folding of G-quadruplexes,[6,7,29] with the ultimate aim of being able to predict the topology of individual G-quadruplex structures solely from their sequence. This is relevant to the successful design of ligands which can be targeted to specific potential G-quadruplex-forming sequences throughout the genome.

Materials and Methods

Materials and Crystal Preparation. Native and 5-^{Br}dU derivative DNA oligonucleotides were purchased from Eurogentec Ltd. (U.K.) and used without further purification. Solutions of 2 mM DNA were annealed at 85 °C for 5 min in 50 mM K^+ cacodylate buffer, pH 7.0, and 80 mM KCl. The solutions were then left to cool to room temperature. Crystals of d($G_4T_3G_4$) appeared after 1 month at 11 °C in a hanging drop from an in-house crystallization screen. The drop contained 2 μL of 1 mM DNA solution and 2 μL of the reservoir solution (0.3 M NaI, 0.05 M Na^+ cacodylate buffer pH 6.5, and 15% PEG 400). These native crystals in space group *C*222 diffracted to 1.5 Å resolution, and intensity data were collected at the ESRF, Grenoble, beamline ID14-4. We were unable to solve the structure by molecular replacement using a variety of models.

d($G_4{}^{Br}UTTG_4$) was crystallized under oil from drops containing 2 μL of 1 mM DNA and 2 μL of the above reservoir solution. Crystalline plates appeared within a few days from drops under 20 μL of a 1:2 mixture of paraffin/silicon oil. These crystals diffracted poorly, and obtaining single crystals proved difficult; however, they were indexed in space group *C*2, and intensity data could be collected on an in-house *R*-axis IV detector. The data are not of high quality, and the structure is not reported here. Smaller crystals in the space group $P2_1$ appeared after 1 month in drops under a 2:1 mixture of paraffin/silicon oil. These crystals diffracted to 2.8 Å, and MAD diffraction data were collected at four wavelengths at the ESRF, Grenoble, beamline ID14-4.

Crystal Structure Determinations. The SHELXD/E program[40] was used for the initial heavy atom search of the $P2_1$ data set, using only the peak wavelength (SAD), due to poor merging statistics with the data collected at other wavelengths. This could be due to loss of the bromine atoms during the collection. The initial electron density map calculated from heavy atom positions was of poor quality. However, two rows of three and seven discrete spherical peaks were apparent and were interpretable as the positions of K^+ ions within the quadruplex central channel. This enabled the positioning of the three quadruplexes in the asymmetric unit. Model building was carried out using the Coot program,[41] and refinement used the Refmac5 program[42] within the CCP4 package. The final *R* factor and R_{free} were 16.8% and 24.2%, respectively. Crystal structure coordinates and structure factor data have been deposited in the Protein Data Bank with code 2AVH.

The $P2_1$ solution provided models for molecular replacement solution of the other structure. The d($G_4T_3G_4$) structure was solved in space group *C*222 using a single strand (i.e., half a complete quadruplex) from the $P2_1$ solution as a model for molecular replacement using the EPMR program.[43] The tenth best solution had a correlation coefficient of 28% and an *R* factor of 58%, and packed correctly. The solution was refined using the Refmac5 program, with only slight modifications to the loop region necessary, to a final *R* factor and R_{free} of 19.8% and 20.8% respectively. Crystal structure coordinates and structure factor

(24) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. *J. Am. Chem. Soc.* **2004**, *126*, 8702−8709.
(25) Rossetti, L.; Franceschin, M.; Bianco, A.; Ortaggi, G.; Savino, M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2527−2533.
(26) Kim, M.-Y.; Gleason-Guzman, M.; Izbicka, E.; Mishioka, D.; Hurley, L. H. *Cancer Res.* **2003**, *63*, 3247−3256.
(27) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167−173.
(28) Smirnov, I.; Shafer, R. H. *Biochemistry* **2000**, *39*, 1462−1468.
(29) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405−16415.
(30) (a) Risitano, A.; Fox, K. R. *Nucleic Acids Res.* **2004**, *32*, 2598−2606. (b) Cevec, M.; Plavec, J. *Biochemistry* **2005**, *44*, 15238−15246. (c) Balagurumoorthy, P.; Brahmacahari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids. Res.* **1992**, *20*, 4061−4067. (d) Petraccone, L.; Erra, E.; Esposito, V.; Randazzo, A.; Mayol, L.; Nasti, L.; Barone, G.; Giancola, C. *Biochemistry* **2004**, *43*, 4877−4884.
(31) Keniry, M. A.; Owen, E. A.; Shafer, R. H. *Nucleic Acids Res.* **1997**, *25*, 4389−4392.
(32) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189−200.
(33) Horvath, M. P.; Schultz, S. C. *J. Mol. Biol.* **2001**, *310*, 367−377.
(34) Schultze, P.; Smith, F. W.; Feigon, J. *Structure* **1994**, *2*, 221−233.
(35) Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018−3028.
(36) Smith, F. W.; Lau, F. W.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10546−10550.
(37) Cmugelj, M.; Hud, N. V.; Plavec, J. *J. Mol. Biol.* **2002**, *320*, 911−924.
(38) Cmugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2003**, *125*, 7866−7871.
(39) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Mol. Biol.* **2004**, *338*, 93−102.
(40) Schneider, T. R.; Sheldrick, G. M. *Acta Crystallogr.* **2002**, *D58*, 1772−1779.
(41) Emsley, P.; Cowtan, K. *Acta Crystallogr.* **2004**, *D60*, 2126−2132.
(42) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr.* **1997**, *D53*, 240−255.
(43) Kissinger, C. R.; Gehlhaar, D. K.; Smith, B. A.; Bouzida, D. *Acta Crystallogr.* **2001**, *D57*, 1474−1479.

Table 1. Crystallographic Data

sequence	$d(G_4T_3G_4)$	$d(G_4{}^{Br}UTTG_4)$
space group	*C*222	$P2_1$
cell (*a*, *b*, *c* (Å)	27.9, 37.1, 43.7	31.7, 33.3, 79.2
α, β, γ (deg))	90, 90, 90	90, 91.5, 90
wavelength (Å)	0.9792	0.9199 (peak)
resolution range (Å)	43.6−1.5	79.1−2.4
completeness (%)	98.9	94.6
R_{merge} on *I*	0.05	0.04
$I/\sigma(I)$	23.2	12.1
no. unique reflections	4066	6534
R (%)	19.8	16.8
R_{free} (%)	20.8	24.2
rmsd bond distances (Å)	0.004	0.008
rmsd angles (Å)	1.15	1.73
DNA strands/ asymmetric unit	1	6
water molecules	54	155
K^+ ions	4	14

data have been deposited in the Protein Data Bank with code 2AVJ. Crystallographic data for the two structures is given in Table 1.

Modeling and Simulation Studies. The X-ray structures were subjected to 4 ns MD simulations. Head-to-tail and head-to-head dimers were both simulated after replacing the ^{Br}U with T residues. Five separate, fully solvated, 4 ns simulations were carried out. Each of the three dimers in the $d(G_4{}^{Br}UTTG_4)_2$ $P2_1$ asymmetric unit was separately simulated, and two head-to-head dimers were also simulated as stacked in the crystal structure. The structure of the $d(G_4T_3G_4)_2$ quadruplex from the *C*222 crystal structure was also simulated. Further details of energetic calculations are available as Supporting Information.

Results

Crystal Structures. In the $P2_1$ crystals of $d(G_4{}^{Br}UTTG_4)$, the asymmetric unit is composed of six $d(G_4{}^{Br}UTTG_4)$ strands that associate into three crystallographically and structurally distinct bimolecular G-quadruplex molecules. A head-to-tail antiparallel dimer and two head-to-head parallel hairpin dimers are formed (Figure 1, parts a and b). The head-to-tail dimer has a lateral loop on each quadruplex face. The two head-to-head dimers have two lateral loops on the same quadruplex face and stack with each other via an interface of stacked G-quartets.

There are crystal-packing interactions between the loop residues and the central guanines in the *C*222 $d(G_4T_3G_4)$ structure so that the stacking of a thymine with guanine causes the G-quartets to be slightly buckled out of coplanarity (Figure 1c). This, however, does not affect the nature of the hydrogen-bonding interactions between guanine residues in the G-quartets. The $d(G_4T_3G_4)$ structure is closely similar to that of the $P2_1$ $d(G_4{}^{Br}UTTG_4)$ head-to-tail dimer, suggesting that bromination of the uracil loop residue does not affect either the crystallization or quadruplex folding of the thymine analogue.

Structure of the G-Quartets. Both head-to-tail dimeric quadruplexes in the $P2_1$ and *C*222 space groups have a pattern of alternating syn−anti glycosidic conformations down the chains and around the G-quartets. This creates two alternating groove widths, wide and narrow, with the loops joining strands over the wide grooves. The two stacked head-to-head quadruplexes in the $P2_1$ $d(G_4{}^{Br}UTTG_4)$ structure also have alternating syn−anti bases down the strands; however, now the pattern for each G-quartet is syn−syn−anti−anti, due to the different chain polarity required for the two loops to be adjacent. This in turn leads to three different groove widths, a narrow, two medium, and a wide groove. These grooves are identical in both head-to-head dimers, with an rms deviation of only 0.5 Å between the G-quartets. The loops are located over the narrow and wide grooves.

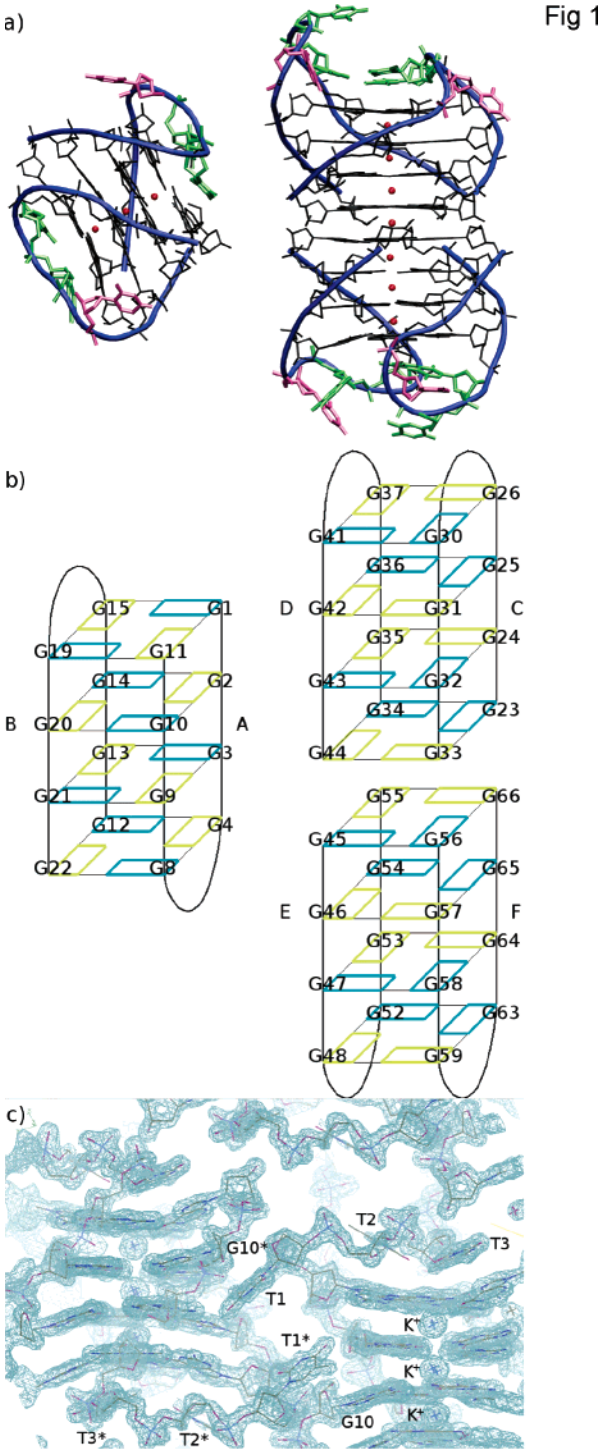

Figure 1. (a) Diagram of the three quadruplexes in the $P2_1$ $d(G_4{}^{Br}UTTG_4)$ structure, shown in their orientation in the crystallographic asymmetric unit. Strands A and B form the head-to-tail dimer, and strands C, D, E, and F form the two stacked head-to-head dimers. The *C*222 structure loop A has the same topology and atom numbering as the $P2_1$ loop A shown here. Loop residues have been omitted, guanines in the syn conformation are colored yellow, and those in the anti conformation are colored green. The 5′ residue of each G-strand is shown. (b) The two stacked head-to-head dimeric quadruplexes in the $P2_1$ asymmetric unit shown alongside the single head-to-tail dimer. (c) Crystal packing within the *C*222 crystal unit cell. The $2F_o - F_c$ electron density map has been contoured at the 1.8σ level.

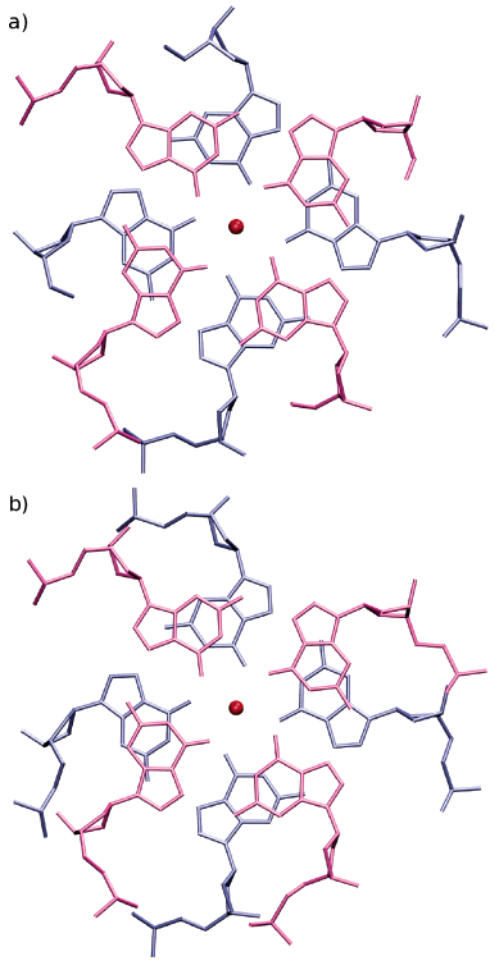

Figure 2. (a) Inter- and (b) intraquadruplex G-quartet stacking in the $P2_1$ d($G_4{}^{Br}UTTG_4$) dimer of head-to-head dimers. K^+ ions between the two G-quartets are shown as red spheres.

The two head-to-head dimers in the $P2_1$ crystal structure stack 5′ to 3′ so that the G-quartet stacking is identical between the two quadruplexes and within a single quadruplex (Figure 2, parts a and b). Such stacking of bimolecular G-quadruplexes has not been previously observed in crystal structures, as loop residues generally prohibit direct stacking of two quadruplex faces.

The Hoogsteen hydrogen bonds between guanine residues are conserved in all these G-quadruplex structures. K^+ ions are found between each of the G-quartets, including between the terminal G-quartets of the two stacked dimer quadruplexes. In every case the ions are equidistant from the O6 guanine atoms, as previously reported.[32] However, unlike the d($G_4T_4G_4$) structure, no ions were observed within the loop regions of any of the quadruplexes here.

The electron density in both the *C*222 and $P2_1$ structures showed clearly defined water molecules located in most quadruplex grooves. The hydration patterns are very similar to those previously described for the d($G_4T_4G_4$) quadruplex, which contains similar groove widths (narrow, medium, and wide),[32,33] although in the present structures the hydrogen bonds with groove water molecules involve both N2 and N3 atoms along the edges of guanine bases, whereas only N2 involvement was observed in the d($G_4T_4G_4$) structure. Potassium and arsenic ions (as cacodylate from the buffer) were also located in several positions within the quadruplex grooves. The cacodylate ions were unequivocally identified due to their strong anomalous peaks, although the spherical shape of their density did not allow the individual oxygen and carbon atoms from the cacodylate ion to be identified and placed in electron density. The channel of seven potassium ions in the head-to-head $P2_1$ dimer structure is continuous and almost linear, with an ion midway between the two G-quartets at the interface providing the linkage between the two discrete quadruplexes. It is remarkable that their backbones are almost exactly in register so that the addition of linking phosphate groups could be accomplished without requiring any relative rotation of the two halves of the dimer. This results in the dimer having the appearance of an eight-quartet step, almost continuous, right-handed four-stranded helix.

Table 2. Torsion Angles (deg) for the Thymine Residues Comprising the Loops in the *C*222 d($G_4T_3G_4$) and $P2_1$ d($G_4{}^{Br}UTTG_4$) Crystal Structures

angle	α	*β*	*γ*	*δ*	*ϵ*	*ζ*	X
$P2_1$ loop A							
U5	57	−142	74	135	−107	−65	−129
T6	−57	−126	64	137	−160	164	−119
T7	−66	−168	48	115	55	−82	−172
> $P2_1$ loop B							
U16	1	106	−62	179	−111	−73	−99
T17	75	99	47	104	−159	164	−135
T18	−54	172	55	153	−101	76	−167
> $P2_1$ loop C							
U27	−164	112	56	137	−177	64	−86
T28	−19	−137	71	132	−141	−72	−137
T29	−74	−172	65	144	−172	−82	−108
> $P2_1$ loop D							
U38	48	−129	66	145	−105	−62	−127
T39	54	147	67	143	−140	164	−111
T40	−105	179	66	141	−96	76	−165
> $P2_1$ loop E							
U49	−45	100	173	100	88	−48	155
T50	−7	−51	150	86	−131	−176	−140
T51	−101	180	57	147	−90	66	−156
> $P2_1$ loop F							
U60	−65	116	165	145	−145	−145	−112
T61	69	−139	68	141	−139	−85	−130
T62	−41	−180	35	138	173	−73	−106
> *C*222 loop A							
T5	68	−128	78	148	−99	−62	−129
T6	64	120	55	126	−157	170	−119
T7	−73	−174	59	158	−107	75	−172

Loop Conformations. A total of five occurrences of loops bridging a wide quadruplex groove are observed in the *C*222 and $P2_1$ structures (the single *C*222 loop in the asymmetric unit, two loops from the head-to-tail $P2_1$ d($G_4{}^{Br}UTTG_4$) dimer, and one loop from each of the head-to-head $P2_1$ dimers). There are a further two occurrences of loops bridging narrow grooves in the $P2_1$ d($G_4{}^{Br}UTTG_4$) head-to-head dimers. This yields seven independent loop conformations, each participating in different crystal-packing interactions. However, despite the different packing environments, there is marked conservation of loop conformation in the various quadruplexes (Table 2). One recurring conformation (type 1) is shown in Figure 3a, which is found in the *C*222 d($G_4T_3G_4$) loop and both the $P2_1$ d($G_4{}^{Br}UTTG_4$) head-to-head (loops A and B) and head-to-tail dimers (loop D). The first loop thymine (^{Br}uracil in $P2_1$) is pointing away from the G-quartet surface and folds down into a quadruplex groove, its depth within the groove dependent on the packing interactions with other residues, leading to some variation in the backbone conformation for this thymine between

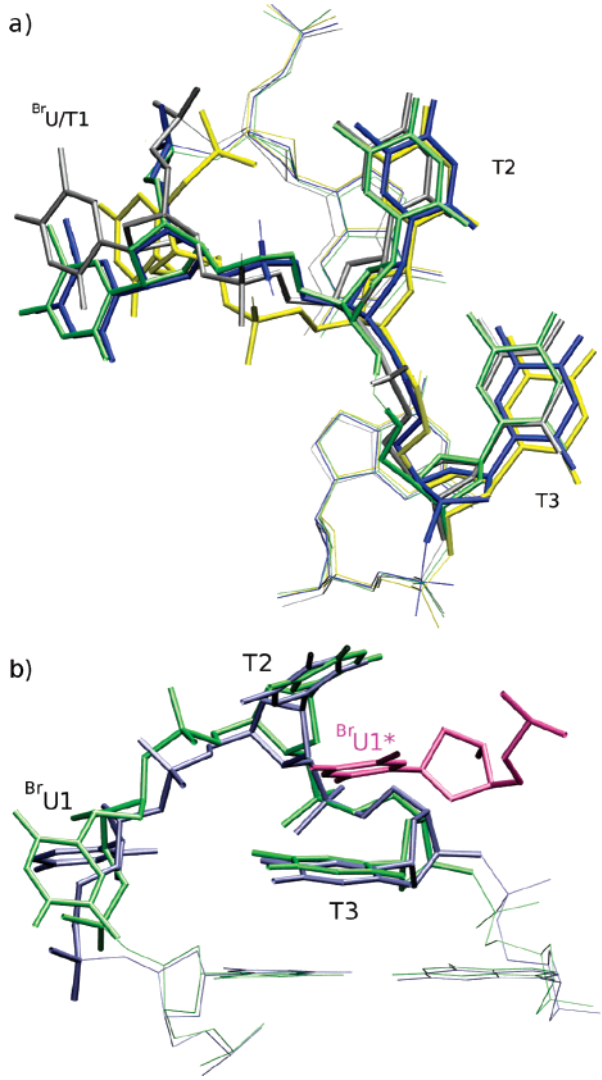

Figure 3. (a) Overlay of the *C*222 T_3 loop (grey) with the $P2_1$ ^{Br}UTT loops A (blue), B (yellow), and D (green). The structures were overlaid using the two upper guanine residues only. The loop geometry shown here is referred to as type 1 in the text. (b) Overlap of loops C (green) and F (blue) from the $P2_1$ structure. The structures were overlaid using the guanine residues only. This loop geometry is referred to as type 2 in the text.

individual loops. The second and third thymines in each loop stack on top of the G-quartet plane. There are no hydrogen-bonding interactions between these residues. However, the stabilization gained by stacking with the guanines presumably contributes to all four loops having very similar conformations, especially for thymines 2 and 3 in each case. This loop conformation is always formed over the wide quadruplex groove. The loop over the fourth wide groove in the *P*21 d(G_4^{Br}UTTG_4) structure (loop E) is less well defined, and although the third thymine stacks on the G-quartet in a similar position as above, the first loop residue (^{Br}uracil) does not lie in the groove but in a more mobile position above the quartets.

The two loops bridging narrow grooves (loops C and F in the $P2_1$ d(G_4^{Br}UTTG_4) structure) adopt a different set of conformations (type 2), probably due to the fact that these grooves are too narrow for two thymine residues to lie side-by-side on top of the G-quartet plane. This is reflected in large differences between P(thymine1)···O3′(thymine3) distances for each pair of loops at the end of the dimer, 12.6 Å versus 15.2 Å for loops A and C. Both loops C and F have a single thymine

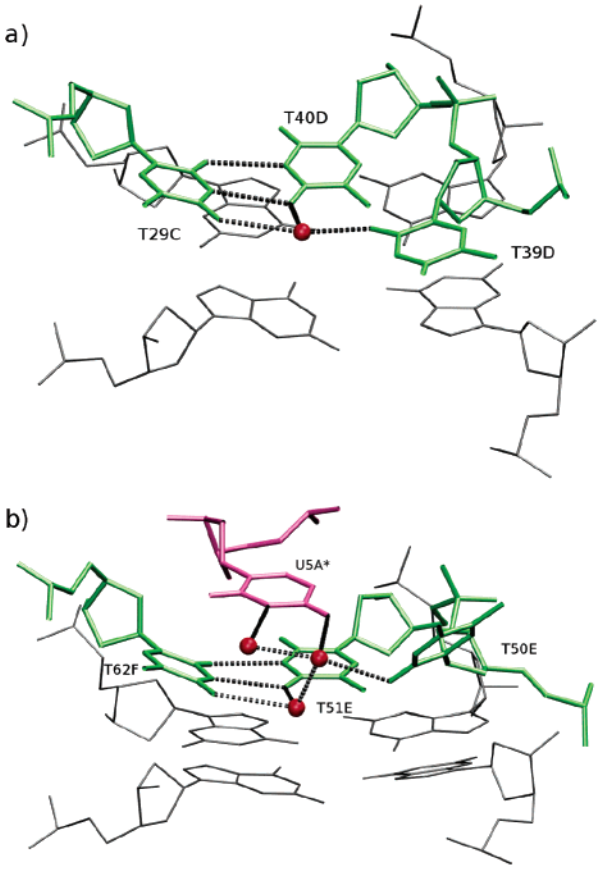

Figure 4. Interactions between (a) adjacent loops C and D in the $P2_1$ d(G_4^{Br}UTTG_4) structure and (b) loops E and F. Residue numbering corresponds to that of Figure 1b, and hydrogen bonds are indicated by dotted lines. An asterisk indicates a residue from a symmetry-related molecule.

(thymine1) stacking on the terminal G-quartet. In both cases the second thymine of the loop forms a stacking interaction with another loop residue, and the third thymine is pointing toward the solvent and is less well defined. Figure 3b shows the overlap of loops C and F in the $P2_1$ d(G_4^{Br}UTTG_4) structure.

The proximity of the two loops at each end of the head-to-head dimers enables loop···loop stabilizing interactions to occur. Thymine 3 from the type 2 loops C and F forms a hydrogen-bonded N3···O4 thymine···thymine base pair with thymine 3 from the loops D and E, respectively. The C−D base pair is also involved in a water-mediated triplet interaction with thymine 2 from loop D, which extends over the G-quartet surface (Figure 4a). Loops E and F also have similar interactions, although in this case hydrogen bonding occurs over two planes (Figure 4b). These loop−loop interactions result in complex structures at the head-to-head ends (Figure 1a).

Modeling. Figure 5 shows selected average loop conformations over the final 2−4 ns of simulation of the head-to-tail and head-to-head quadruplexes. The head-to-tail dimer in the d(G_4^{Br}UTTG_4)$_2$ $P2_1$ crystal structure, formed by strands A and B, did not change its structure significantly during the simulations, and the average structures of the loops shown in Figure 5, parts a and b, are very close to the X-ray conformation. Stability of the structures was estimated by calculating rms deviations with respect to the starting (crystal) structures. These rmsd values have been added to the table of energy values in the Supporting Information. This suggests that the type 1 loop, in which two residues stack with the G-quartets and the first

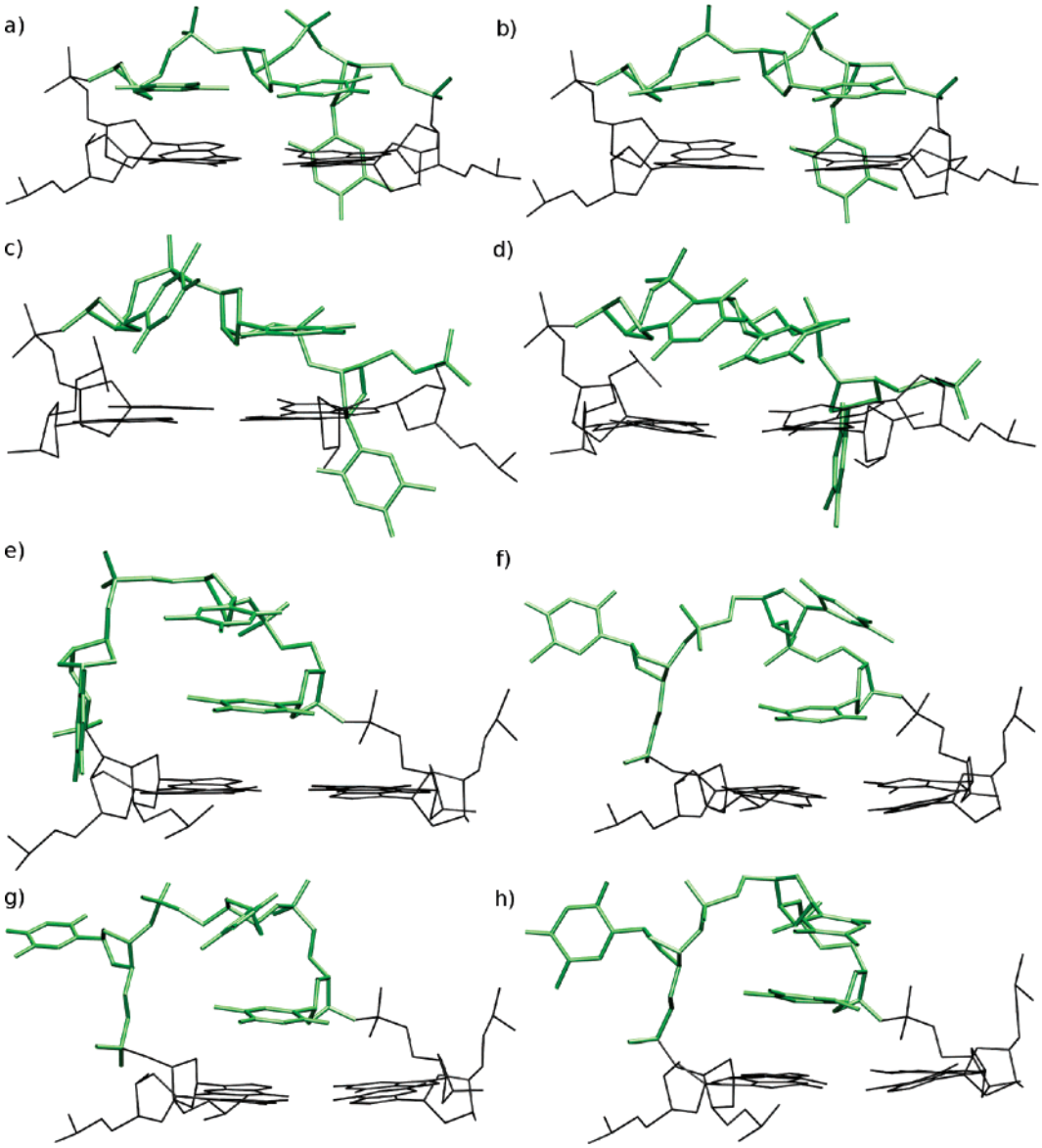

Figure 5. Average loop structures over the last 2−4 ns of dynamics for the $P2_1$ dimer simulations (^{Br}U residues were replaced with T). (a) Loop A, (b) loop B, (c) loop D, (d) loop D, stacked dimer simulation, (e) loop C, (f) loop C, stacked dimer simulation, (g) loop F, and (h) loop F, stacked dimer simulation.

residue is located in the groove, is not a crystallographic artifact and so may be a stable form in solution. The head-to-tail dimers, in both $d(G_4{}^{Br}UTTG_4)_2$ and $d(G_4T_3G_4)_2$ crystal structures, had slightly buckled G-quartets (Figure 6a). During the simulations, in the absence of crystal-packing interactions, the G-quartets became planar, as shown by the $d(G_4T_3G_4)_2$ structure averaged over the last 2−4 ns of simulation in Figure 6b. These simulations suggest that the slight distortions of the G-quartet planes in the crystal are caused by crystal-packing interactions, rather than any loop effects. Moreover, in simulated solution, the first T residues of each loop moved closer to the floor of the wide quadruplex groove, again probably due to the absence of T···T intermolecular crystal-packing interactions compared to those of the crystal.

Not all the type 1 loops simulated remained in the starting conformation, as shown in Figure 5, parts c and d. Loop D of the head-to-head dimer formed by strands C and D was more flexible during the simulations, both alone in solution and as part of the two-quadruplex stack. The general loop conformation remained similar, although the third T residue stacked with a loop residue rather than with the G-quartets. This second conformation was stable over the course of the dynamics. Loops C and F formed type 2 loops, shown in Figure 3b, which were involved in crystal-packing interactions. During the simulations, the first and second T residues formed a stable stack over the G-quartets (Figure 5e−h). The third T residue was more flexible and generally more exposed to the solvent. Despite the fact that the loops over the narrow groove in the head-to-head dimers are involved in crystal-packing interactions, stable conformations were adopted in the simulations, showing that such loops can be stable even outside the environment of the crystal. The small changes in loop conformations during the dynamics affected the loop−loop interactions between adjacent loops of the head-to-head dimers. Hydrogen bonding between loops E and F (Figure 4b) was not maintained during either of the strand E and F head-to-head dimer simulations. This did not affect the stability of the loops, as stacking interactions with the G-quartets were conserved. Loops C and D also formed hydrogen-bonding interactions (Figure 4a), and the direct hydrogen bonds were conserved during the simulations.

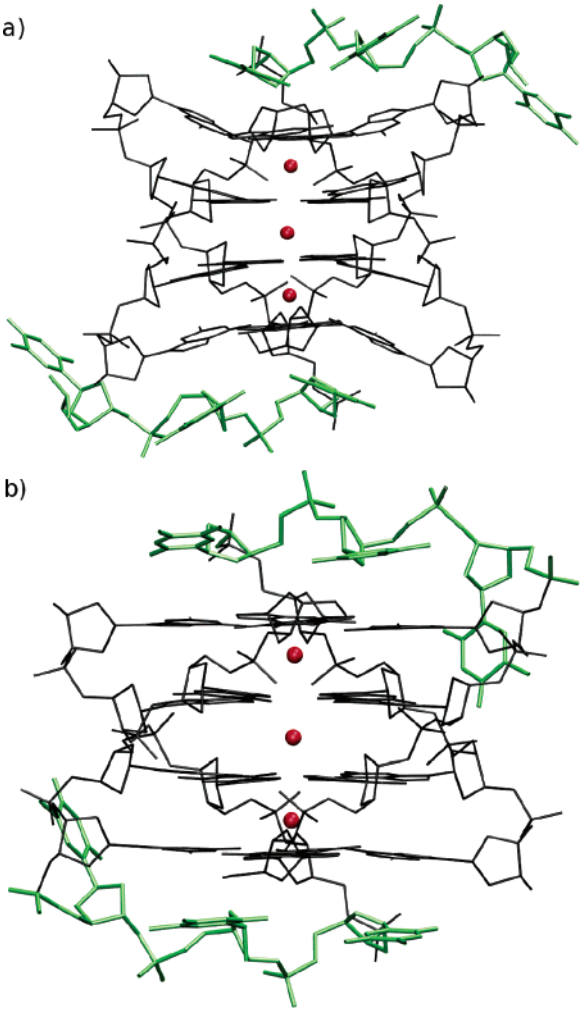

Figure 6. (a) $d(G_4T_3G_4)_2$ *C*222 X-ray structure and (b) $d(G_4T_3G_4)_2$ simulated structure, averaged over the final 2 ns of simulation.

Energetics of the experimental structures were calculated from the MD simulations, using the MM-PBSA method. These are discussed in detail in the Supporting Information. Overall, the calculations show that the head-to-head dimers tend to be less favorable structures than the head-to-tail dimers. The free-energy calculations do show a slight preference for the type 1 loop conformation, which was the most frequently observed experimentally. The simulations suggest that both type 1 and type 2 loop conformations are stable in solution and, hence, not only in the context of crystal-packing interactions.

Discussion

The structures reported here expand the repertoire of quadruplex crystal structures and show the large structural variability of G-quadruplexes, even within the constraints of a single sequence. The same sequence forms different structures, even within the same crystal, suggesting strongly that there are not large energy differences between the head-to-head and head-to-tail bimolecular quadruplexes, at least with the T_3 loop. It is not clear whether the formation of adjacent loops on the same G-quadruplex face would be favorable without the stabilization gained by two quadruplexes forming an eight-G-quartet stack with seven K^+ ions, as seen in the $P2_1$ $d(G_4{}^{Br}UTTG_4)$ crystal. The formation of dimers has been reported previously for some quadruplex structures, notably for an aptamer forming an interlocked dimeric parallel-stranded quadruplex,[44] as well as for the parallel human 22-mer intramolecular quadruplex.[8] In both instances the loops in the structures are distant from the G-quartets that form the interface between individual quadruplexes. The sequence $d(T_2G_4T_2G_4T)$ forms interconverting parallel (head-to-head) and antiparallel (head-to-tail) bimolecular quadruplexes in solution[39] that are analogous to the two types of bimolecular structure reported here. These two species were identified by this NMR study to be of comparable stability, and by analogy the finding here of both in the one crystal, the $P2_1$ form, similarly suggests that there are not large energy differences between the head-to-head and head-to-tail bimolecular quadruplexes containing the T_3 loop and that they may also be present in solution.

The presence of three quadruplexes in the $P2_1$ $d(G_4{}^{Br}UTTG_4)$ crystal asymmetric unit provides six independent loop conformations, each in a different crystal-packing environment. The repeated loop conformations are therefore likely to be due to real structural preferences and not to crystal-packing effects. This is especially the case for the loop structure of type 1 that has a residue lying in the quadruplex groove and two residues stacking on the G-quartet plane. The two different folding topologies, head-to-head and head-to-tail, also showed that the T_3 loop can bridge both wide and narrow groove widths, although this leads to different loop conformations.

This addition to the library of quadruplex structures determined to date highlights some of the challenges in targeting ligands to specific G-quadruplex-forming sequences. The crystallographic data shows that this one sequence $d(G_4T_3G_4)$ can fold with apparently equal facility into two different topologies, and the simulation studies suggest that these are preserved in solution. A number of other G-quadruplexes have been shown to adopt more than one conformation in solution.[10,12,39] This flexibility could make designing ligands to target a potential G-quadruplex with a T_3 loop[6,7] challenging to achieve. However, there are several key features that are conserved between the head-to-head and head-to-tail quadruplex structures. First, all loops are connecting strands in a lateral manner, and join adjacent quadruplex strands. This is in contrast to the NMR and X-ray structures of the $d(G_4T_4G_4)$ bimolecular quadruplex,[32−35] in which the loops have been found to always adopt diagonal conformations. Second, we observe here just two categories of preferred loop conformations, again probably a consequence of the constraints imposed by the lateral connection of the T_3 stretch of sequence. The type 1 loop conformation is also closely similar (Figure 7) to that reported for the T_3 loop from a solution NMR study[45] of the fragile X syndrome quadruplex sequence $d(GCGGT_3GCGG)$.

This conservation of structure could be used in the design of specific ligands to quadruplexes with shorter loop lengths and offers some scope for targeting specific G-quadruplex sequences. The complex nature of the loop structures in the parallel hairpin quadruplex (Figure 1a) suggests that this could be a site for highly selective ligand recognition. Searching the human genome[7] with the sequence $G_4T_3G_4$ finds only 135 occurrences and, more rarely, a total of just eight hits occur close to transcription start sites. The existence of analogous hairpin

(44) Phan, A. T.; Kuryavyi, V.; Ma, J.-B.; Faure, A.; Andreola, M.-L.; Patel, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 634−639.
(45) Kettani, A.; Kumar, R. A.; Patel, D. J. *J. Mol. Biol.* **1995**, *254*, 638−656.

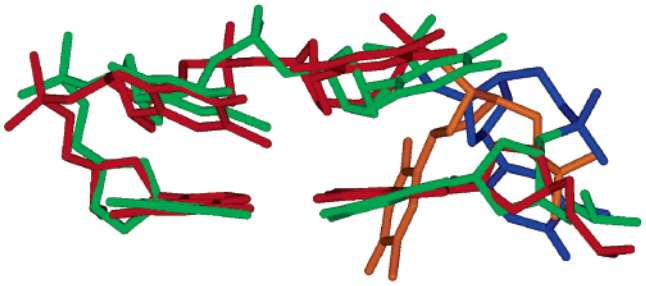

Figure 7. Overlay of the T_3 loop and attached guanine nucleotides of the *C*222 crystal structure (shown in green with the 5′ thymine in dark blue) with those from the d(GCGGT$_3$GCGG) NMR structure (ref 45), shown in red with the 5′ thymine in orange.

bimolecular quadruplexes has been recently demonstrated for promoter sequences in the human sarcomeric mitochondrial creatine kinase (sMtCK) gene, and the myogenic determination protein MyoD gene,[20] and have been suggested to play a role in the regulation of their expression. Bimolecular quadruplexes cannot be formed from a single DNA strand, unless neighboring hairpins can associate together, as has been observed during transcription[20] for sMtSK. Ligand binding generally stabilizes quadruplex formation, and the artificial regulation of gene expression by targeting rarely occurring bimolecular quadruplexes that have distinctive structures with small molecules[14,15] is becoming a realistic concept.

Acknowledgment. This work has been supported by a Program Grant from Cancer Research U.K. and by a Project Grant from the Association for International Cancer Research. We are very grateful to Dr. Alan Todd for the genome search and Dr. Mark Roe (The Institute of Cancer Research) for assistance with data collection at ESRF.

Supporting Information Available: Summary of molecular modeling. This material is available free of charge via the Internet at http://pubs.acs.org.

JA058577+