J. Mol. Biol. (1995) **254**, 638–656

JMB

Solution Structure of a DNA Quadruplex Containing the Fragile X Syndrome Triplet Repeat

Abdelali Kettani, R. Ajay Kumar and Dinshaw J. Patel*

Cellular Biochemistry & Biophysics Program Memorial Sloan-Kettering Cancer Center, New York NY 10021, USA

**Corresponding author*

Both X-ray and NMR structural studies have defined the polymorphic nature of G-quadruplexes generated through mutual stacking of G·G·G·G tetrads by guanine rich telomeric sequences. Recently, the fragile X syndrome $d(C\text{-}G\text{-}G)_n$ triplet nucleotide repeat has been shown to form a stable quadruplex of undefined structure in monovalent cation solution. We have undertaken a structural characterization of the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ undecanucleotide to elucidate the structural alignments associated with quadruplex formation by this oligomer which contains sequence elements associated with the fragile X syndrome triplet repeat. $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ in Na^+ cation solution forms a quadruplex through dimerization of two symmetry related hairpins with the lateral connecting T_3 loops positioned at opposite ends of the quadruplex. This novel NMR-molecular dynamics based solution structure contains internal G·C·G·C tetrads sandwiched between terminal G·G·G·G tetrads. Watson-Crick G·C base-pairs within individual hairpins dimerize through their major groove edges using bifurcated hydrogen bonds to form internal G(*anti*)·C(*anti*)·G(*anti*)·C(*anti*) tetrads. Adjacent strands are anti-parallel to each other around the symmetric G-quadruplex which contains two distinct narrow and two symmetric wide grooves. By contrast, the terminal G-tetrads adopt G(*syn*)·G(*anti*)·G(*syn*)·G(*anti*) alignments. The structure of the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ quadruplex with its multi-layer arrangement of G·G·G·G and G·C·G·C tetrads greatly expands on our current knowledge of quadruplex folding topologies. Our results establish the pairing alignments that can be potentially utilized by the fragile X syndrome triplet repeat to form quadruplex structures through dimerization of hairpin stems. The formation of novel G·C·G·C tetrads through dimerization of Watson-Crick G·C base-pairs is directly relevant to the potential pairing alignments of helical stems in genetic recombination.

Keywords: DNA quadruplex; fragile X syndrome; d(C-G-G) triplet repeat; G·C·G·C tetrads

Introduction

There has been considerable interest in the formation and topology of four-stranded helices ever since it was established that G-rich sequences can form G-quadruplex structures (Gellert *et al*., 1962; reviewed by Guschlbauer *et al*., 1990). Such G-rich sequences are observed at the ends of chromosomes and these telomeric segments have the potential to form G-quadruplexes in the presence of monovalent cations (Sen & Gilbert, 1988; Sundquist & Klug, 1989; Williamson *et al*., 1989). Both X-ray crystallographic and NMR solution studies have defined the range of G-quadruplex folding topologies formed by the alignment of four separate strands (Cheong & Moore, 1992; Wang & Patel, 1993a; Gupta *et al*., 1993; Aboud-ela *et al*., 1994; Laughlan *et al*., 1994), the dimerization of a pair of hairpins (Kang *et al*., 1992; Smith & Feigon, 1992; Schultze *et al*., 1994a; Smith *et al*., 1994a) and through intramolecular folding of a single strand (Macaya *et al*., 1993; Wang *et al*., 1993; Schultze *et al*., 1994b; Padmanabhan *et al*., 1993; Wang & Patel, 1993b, 1994, 1995). Such studies have focused on the *Oxytricha* $d(T_4G_4)_n$, *Tetrahymena* $d(T_2G_4)_n$ and human $d(T_2AG_3)_n$ telomeric se-

Abbreviations used: FMR, fragile X mental retardation; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; MD, molecular dynamics.

0022–2836/95/490638–19 $12.00/0

quences. The alignments of the four strands around the G-quadruplex and the *syn*/*anti* distribution of guanine residues around individual G-tetrads and along individual strands are defined by the connectivities of the loop segments. The loop connectivities have been studied for two (n = 2) and four (n = 4) telomeric repeats and range from lateral (Kang *et al*., 1992; Macaya *et al*., 1993; Wang *et al*., 1993; Padmanabhan *et al*., 1993), diagonal (Smith & Feigon, 1992; Schultze *et al*., 1994b; Smith *et al*., 1994a; Wang & Patel, 1993b, 1995) and double chain reversal (Wang & Patel, 1994) linkages.

To date, the primary focus of four-stranded helical research has been on G-quadruplex folding topologies stabilized by G-tetrad (Figure 1(a)) formation (reviewed by Sen & Gilbert, 1991; Sundquist, 1991; Williamson, 1994; Rhodes & Giraldo, 1995; Pilch *et al*., 1995). We report below on the solution structure of a novel quadruplex that forms through the head to tail dimerization of d(G-C-G-G-T_3-G-C-G-G) hairpins in Na^+ cation solution. Our structural characterization of the d(G-C-G-G-T_3-G-C-G-G) sequence was related to the current interest in the molecular basis of the expansion of d(C-G-G)$_n$·d(C-C-G)$_n$ tracts in genomic DNA with the resulting onset of the fragile X mental retardation (FMR) syndrome (Kremer *et al*., 1991; Yu *et al*., 1991; Verkerk *et al*., 1991; Fu *et al*., 1991). The d(C-G-G)$_n$ triplet repeat is observed within the first exon of the FMR-1 gene with n ranging between 10 and 30 in normal individuals. This number increases up to ~200 in premutation carriers and expands further up to ~2000 in individuals afflicted with the fragile X syndrome. The genetic instability associated with the expansion of the d(C-G-G)$_n$ repeat to the diseased state is facilitated by hypermethylation of cytosine residues (Oberle *et al*., 1991) and results in the suppression of FMR-1 gene transcription (Pieretti *et al*., 1991) and delay in replication (Hansen *et al*., 1993) in patients with the FMR-1 syndrome. Since the phase of the d(C-G-G)$_n$ triplet repeat could be either C-G-G, G-G-C or G-C-G we chose the d(G-C-G-G-T_3-G-C-G-G) sequence which contains two (G-C-G and C-G-G) of the three possible phase alignments.

The expansion of d(C-G-G)$_n$·d(C-C-G)$_n$ repeats associated with the fragile X syndrome and d(C-A-G)$_n$·d(C-T-G)$_n$ repeats associated with neurodegenerative diseases (reviewed by Caskey *et al*., 1992; Sinden & Wells, 1992; Nelson, 1995; Sutherland & Richards, 1995) has stimulated spectroscopic and footprinting efforts to delineate the potential folding topologies adopted by such sequences. The majority of such efforts have qualitatively concluded that the triplet repeats align through hairpin formation without definitively defining the nature of the base-pairing alignment in the duplex stem segments (Mitas *et al*., 1995; Gacy *et al*., 1995). The one exception was an NMR study of the d(C-C-G-C-C-G) sequence which forms a staggered anti-parallel duplex stabilized by Watson-Crick G·C pairs and extrahelical cytosine residues located in the minor groove (Gao *et al*., 1995). This focus on duplex structures (see also, Chen *et al*., 1995) seemed to us not to directly address issues associated with the potential formation of higher order structures by human disease d(C-N-G)$_n$ repeats that could block transcription and delay replication.

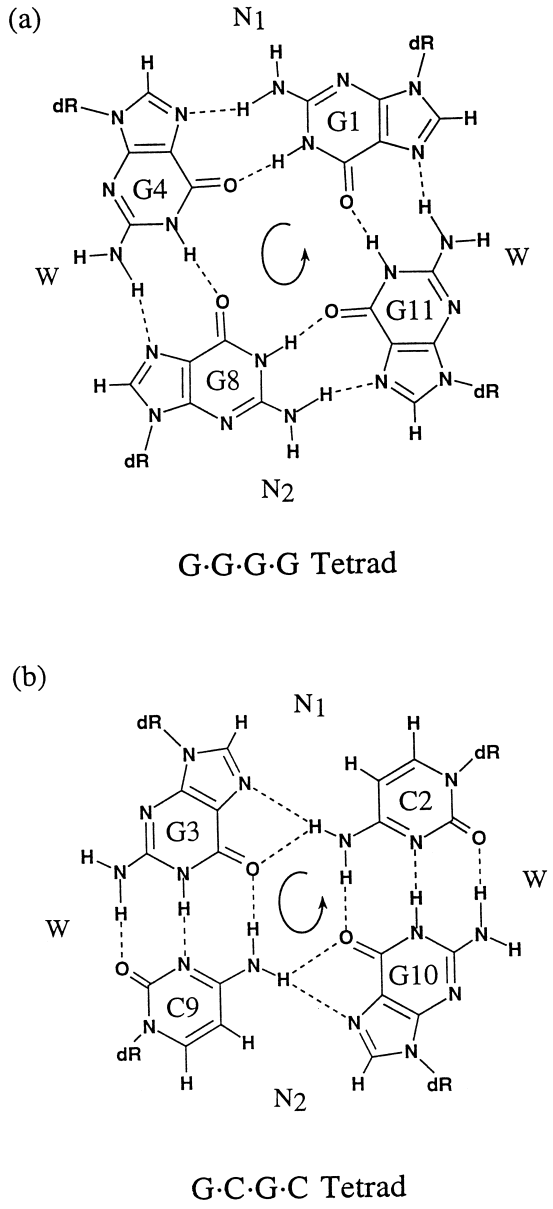

Figure 1. Schematics of the planar (a) G·G·G·G tetrad and (b) G·C·G·C tetrad alignments observed in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Na^+ containing solution.

Our attention was drawn to a recent paper by Fry & Loeb (1994) which established that the fragile X syndrome d(C-G-G)$_n$ repeat forms a stable tetrahelical structure in the presence of monovalent cations when n = 7 and, also when n = 5, for its cytosine methylated counterpart. We were therefore alert to the potential for quadruplex formation by the d(G-C-G-G-T_3-G-C-G-G) sequence which contains two G-C-G-G segments connected by a T_3 linker segment. This paper reports on a combined NMR

G1-C2-G3-G4-T5-T6-T7-G8-C9-G10-G11

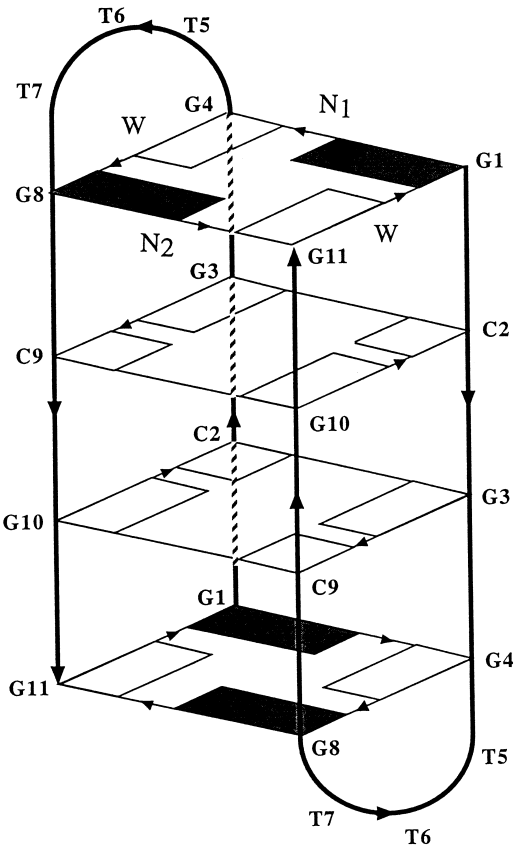

Figure 2. Schematic diagram showing the folding topology of the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Na^+ containing solution. The backbone tracing of individual hairpins is shown by thick lines and the chain directionality by thick arrows. Hydrogen bonding donor to acceptor directionalities around individual G·G·G·G tetrads are represented by arrows as are Watson-Crick G·C pairs within G·C·G·C tetrads. *Syn* guanine residues are shaded to distinguish them from *anti* guanine residues. The two distinct inter-strand narrow grooves are labelled N_1 and N_2 while the symmetric intra-strand wide grooves are labelled W.

and molecular dynamics effort which has resulted in the high resolution solution structure of the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Na^+ containing solution through the head-to-tail dimerization of two symmetry related hairpins (Figure 2) stabilized by G·C·G·C tetrads (Figure 1(b)) sandwiched between G·G·G·G tetrads (Figure 1(a)).

Results

NMR spectra

The d(G-C-G-G-T_3-G-C-G-G) sequence in 150 mM Na^+ cation exhibits the most remarkable spectra for both the exchangeable protons (6.8 to 13.5 ppm) in H_2O buffer at 0°C (Figure 3(a)) and the non-exchangeable protons (4.5 to 8.4 ppm) in ^2H_2O buffer at 20°C (Figure 3(b)) solution. We observe a single conformation with narrow resonances and a single resonance for each proton in the d(G-C-G-G-T_3-G-C-G-G) sequence.

Exchangeable protons

The exchangeable proton spectrum of the d(G-C-G-G-T_3-G-C-G-G) sequence in 150 mM Na^+ containing H_2O buffer (pH 6.5) at 0°C exhibits well resolved imino and amino proton spectra. We observe two guanine imino protons which resonate between 13.0 and 13.1 ppm in a region characteristic of Watson-Crick G·C pair formation, four guanine imino protons which resonate between 11.4 and 12.0 ppm in a region characteristic of G·G·G·G tetrad formation and four amino protons which resonate between 8.6 and 9.2 ppm in a region characteristic of hydrogen-bonded cytosine amino proton formation. There are only two cytosine residues at positions C2 and C9 in the d(G-C-G-G-T_3-G-C-G-G) sequence and hence both amino protons of each cytosine have to be hydrogen-bonded in the folded structure of this sequence in solution. We observe the three thymine imino protons which resonate between 10.4 and 11.0 ppm in a region characteristic of non-hydrogen bonded imino protons.

The guanine imino protons have been assigned following analysis of NOESY spectra of the d(G-C-G-G-T_3-G-C-G-G) and d(G-C-G-I-T_3-G-C-G-G) sequences in H_2O buffer at 0°C with the expanded contour plots of the symmetrical imino protons plotted in Figure 4 and between the imino protons and the base and amino protons plotted in Figure 5.

The imino protons of guanine residues G3 and G10 are involved in Watson-Crick G·C pair formation since they each exhibit NOEs to the cytosine amino (peaks a, a' and superpositioned peaks e, Figure 5(a) and (b)) and H5 (peaks d and h, Figure 5(a) and (b)) protons across the pair. The cytosine H5 protons can be unambiguously assigned from the non-exchangeable proton spectra (see below) permitting definitive identification of the exchangeable imino and amino protons associated with the G3·C9 and C2·G10 Watson-Crick base-pairs (Table 1). Since both cytosine amino protons of C2 (superpositioned at 9.04 ppm) and C9 (8.60, 9.17 ppm) resonate at low field and are hence hydrogen-bonded, the experimental data directly establish that the Watson-Crick G·C base-pairs must dimerize through their major groove edges to form a quadruplex while retaining strand symmetry.

The G1, G4, G8 and G11 guanine imino protons are involved in G-tetrad formation since each guanine imino proton exhibits an NOE to its Hoogsteen aligned adjacent guanine H8 proton (independently assigned from analysis of non-

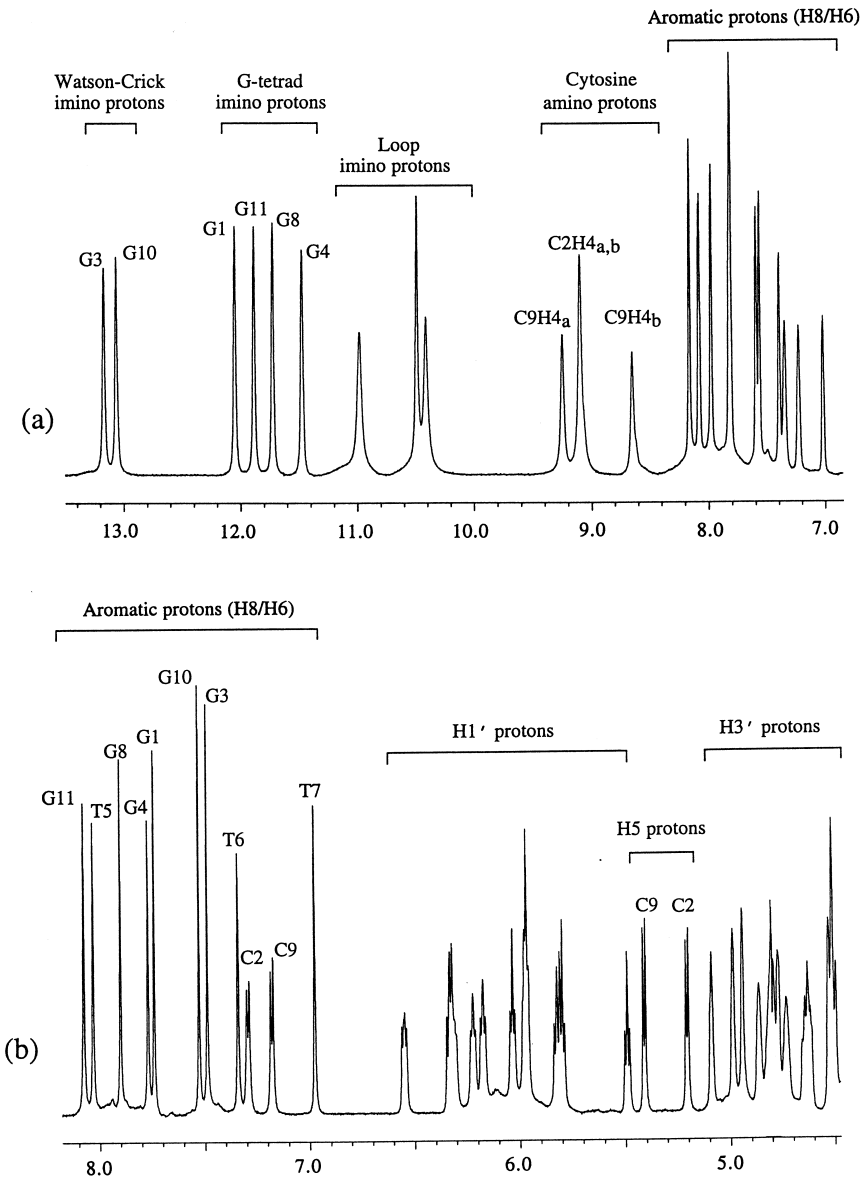

Figure 3. Proton NMR spectra of the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM NaCl, 1 mM phosphate, aqueous buffer (pH 6.5). (a) Proton spectrum (6.8 to 13.5 ppm) in H_2O buffer at 0°C. Base-paired guanine imino and cytosine amino proton assignments are listed above individual resonances. The cytosine amino proton that participates in Watson-Crick hydrogen bonding is labelled $H4_b$ and the one that participates in inter-strand hydrogen bonding is labelled $H4_a$. (b) Proton spectrum (4.5 to 8.2 ppm) in ^2H_2O buffer at 20°C. Base proton assignments are listed above individual resonances.

exchangeable proton spectra) around the G-tetrad (peaks j, x, r and o, Figure 5(a) and (b)). The directionality (clockwise or counter-clockwise) of the G1·G4·G8·G11 G-tetrad was characterized by substituting I4 for G4 and examining the resulting NOE pattern distribution. The imino proton of I4 which can be readily identified by its downfield shift to 13.88 ppm exhibits an NOE to the H8 proton of G8 (peak x, Figure 5(b)) which is consistent with a counterclockwise alignment around the G1·G4·G8·G11 G-tetrad (Figure 1(a)). The formation of a G-tetrad by the d(G-C-G-G-T_3-G-C-G-G) sequence in Na^+ solution can only occur if the undecanucleotide folds into a hairpin which in turn dimerizes to form a quadruplex. Such a quadruplex exhibits strand symmetry as required by the NMR spectral data which exhibits a single set of resonances for the d(G-C-G-G-T_3-G-C-G-G) sequence.

The imino and amino proton chemical shifts in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM Na^+ at 0°C are tabulated in Table 1. We have also recorded ^1H-^{15}N two-dimensional heteronuclear correlation spectra which have provided the corresponding ^{15}N chemical shifts in the quadruplex which are listed in Table 2.

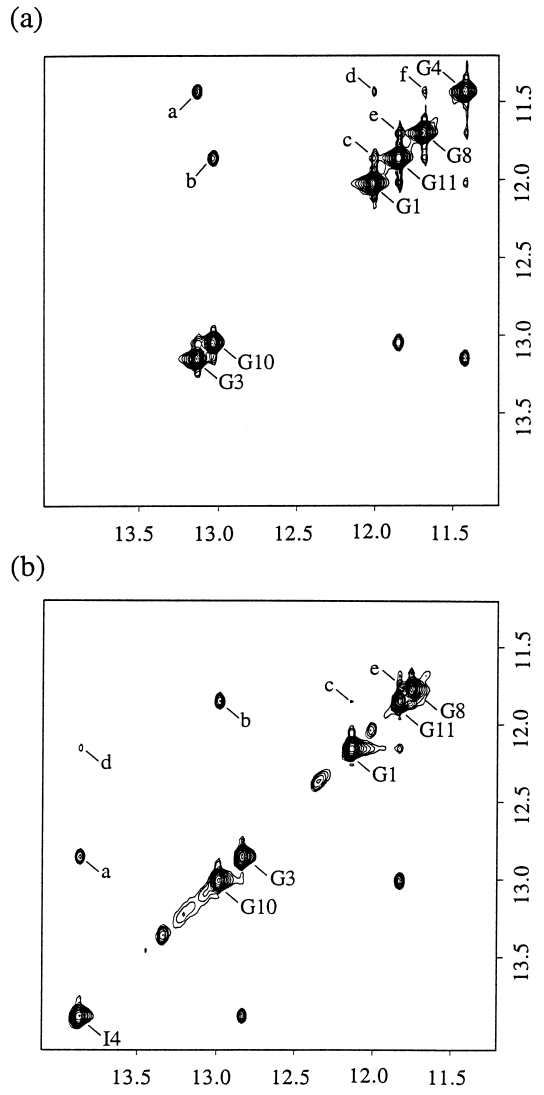

Figure 4. Expanded NOESY contour plots (250 ms mixing time) of (a) the d(G-C-G-G-T_3-G-C-G-G) and (b) the d(G-C-G-I-T_3-G-C-G-G) quadruplexes in 150 mM NaCl, H_2O buffer at 0°C. NOE cross peaks in the symmetrical imino proton (11.2 to 14.2 ppm) region. The cross peaks a to f are assigned as follows: a, G3(NH1)-G4/I4(NH1); b, G10(NH1)-G11(NH1); c, G1(NH1)-G11(NH1); d, G1(NH1)-G4/I4(NH1); e, G11(NH1)-G8(NH1); f, G8(NH1)-G4(NH1).

Non-exchangeable protons

The non-exchangeable protons in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM Na^+ containing ^2H_2O solution (pH 6.5) at 20°C have been assigned following analysis of NOESY spectra as a function of mixing time. An expanded NOESY contour plot (300 ms mixing time) correlating the base protons (6.9 to 8.2 ppm) with the base and sugar H1' protons (5.1 to 6.6 ppm) is plotted in Figure 6(a). The corresponding NOESY stacked plot (50 ms mixing time) is plotted in Figure 6(b). We have traced the NOE connectivities between the base protons and their own and 5'-flanking sugar H1' protons (reviewed by van de Ven & Hilbers, 1988) from G1 to G11 in the d(G-C-G-G-T_3-G-C-G-G) sequence as shown in Figure 6(a). There are either weak (T5-T6 step) or absent (T7-G8 step) cross peaks for the loop T_3 segment (see boxed peaks in Figure 6(a)), as well as weak cross peaks between guanine H8 and cytosine H5 protons for the G1-C2 (peak a, Figure 6(a)) and G8-C9 (peak c, Figure 6(a)) steps.

The strong NOEs between the H8 and their own H1' protons for G1 and G8 in 50 ms mixing time NOESY spectra with intensities comparable to the NOEs between the cytosine H6 and H5 protons (fixed 2.47 Å separation) of C2 and C9 (Figure 6(b)) establish the formation of *syn* glycosidic torsion angles (Patel *et al*., 1982) at G1 and G8 in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Na^+ solution. The formation of *syn* glycosidic torsion angles at G1 and G8 was also confirmed by the observation of NOEs between the H8 protons of *syn* guanine residues and the H1' protons of their 3'-linked *anti* cytosine residues for the G1-C2 step (peak g, Figure 6(a)) and the G8-C9 step (peak h, Figure 6(a)). The base and sugar H1' non-exchangeable proton assignments in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM Na^+ at 20°C are listed in Table 1. A similar analysis was extended to other regions of the NOESY spectrum and, together with the analysis of the TOCSY spectrum, yielded the sugar H2', H2", H3' and H4' assignments which are also listed in Table 1. We have also recorded ^{1}H-^{13}C two-dimensional heteronuclear correlation spectra which have provided the corresponding base and sugar ^{13}C chemical shifts in the quadruplex which are listed in Table 2.

Phosphorus spectra

The phosphorus spectrum of the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM Na^+ at 20°C exhibits well resolved resonances dispersed between –3.7 and –4.8 ppm upfield from standard trimethylphosphate. The phosphorus resonances have been assigned by identifying coupling connectivities between individual phosphorus resonances and their 5'-linked H3' protons and their 3'-linked H4' protons in a heteronuclear ^{1}H-^{31}P two-dimensional correlation spectrum. The phosphorus chemical shifts in the quadruplex are listed in Table 1 with the majority of the phosphates resonating between –3.7 and –4.5 ppm except for the phosphate at the T5-T6 step in the T_3 loop which resonates at –4.74 ppm.

NOEs between adjacent G·G·G·G and G·C·G·C tetrads

A set of same strand and cross strand NOEs are detected between the adjacently positioned G·C·G·C and G·G·G·G tetrads that define the relative alignments of these tetrads in the d(G-C-G-G/I-T_3-G-C-G-G) quadruplexes. These include NOEs between the imino protons of G3 and G4 (peak a, Figure 4(a) and (b)) and the imino protons of G10 and G11 (peak b, Figure 4(a) and (b)), as well as an NOE between the imino proton of G3 and the H2

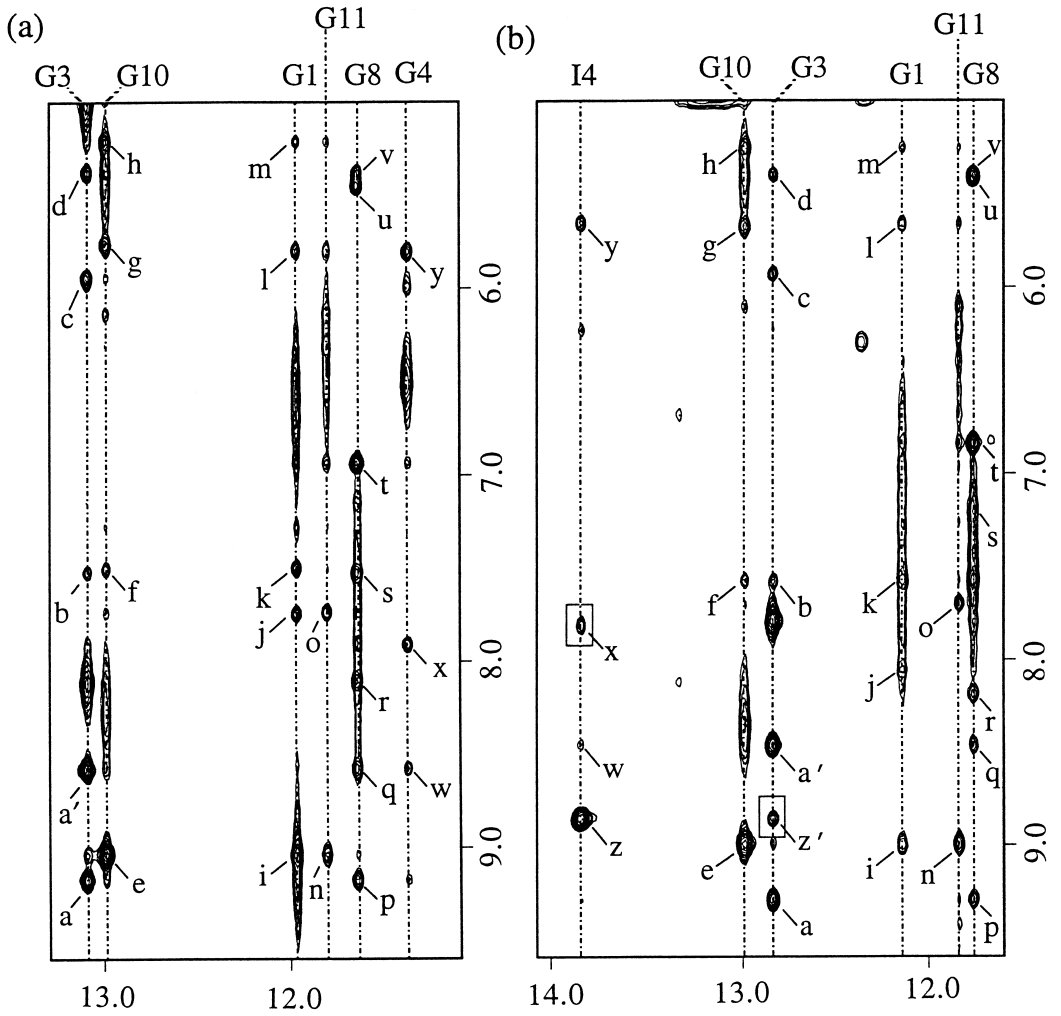

Figure 5. Expanded NOESY contour plots (250 ms mixing time) of (a) the d(G-C-G-G-T_3-G-C-G-G) and (b) the d(G-C-G-I-T_3-G-C-G-G) quadruplexes in 150 mM NaCl, H_2O buffer at 0°C. NOE cross peaks between the imino protons (11.2 to 13.2 ppm) and the base and amino protons (5.0 to 9.5 ppm). The cross peaks a to z are assigned as follows: a: G3(NH1)-C9(NH_2-a); a': G3(NH1)-C9(NH_2-b); b: G3(NH1)-G10(H8); c: G3(NH1)-G10(H1'); d: G3(NH1)-C9(H5); e: G10(NH1)-C2(NH_2-a,b); f: G10(NH1)-G3(H8); g: G10(NH1)-G3(H1'); h: G10(NH1)-C2(H5); i: G1(NH1)-C2(NH_2-a,b); j: G1(NH1)-G4/I4(H8); k: G1(NH1)-G3(H8); l: G1(NH1)-T6(H1'); m: G1(NH1)-C2(H5); n: G11(NH1)-C2(NH_2-a,b); o: G11(NH1)-G1(H8); p: G8(NH1)-C9(NH_2-a); q: G8(NH1)-C9(NH_2-b); r: G8(NH1)-G11(H8); s: G8(NH1)-G10(H8); t: G8(NH1)-T7(H6); u: G8(NH1)-T7(H1'); v: G8(NH1)-C9(H5); w: G4/I4(NH1)-C9(NH_2-b); x: G4/I4(NH1)-G8(H8); y: G4/I4(NH1)-T6(H1'); z: I4(NH1)-I4(H2); z': G3(NH1)-I4(H2).

proton of I4 (peak z', Figure 5) in the d(G-C-G-I-T_3-G-C-G-G) quadruplex. In addition, the G1 imino proton exhibits NOEs to the G3(H8) (peak k, Figure 5(a)) and C2(H5) (peak m, Figure 5(a)) protons while the G8 imino proton exhibits NOEs to the G10(H8) (peak s, Figure 5) and C9(H5) (peak v, Figure 5) protons. These NOEs should help to define the relative hydrogen bond directionalities between the adjacent G·G·G·G and G·C·G·C tetrads in the quadruplex.

NOEs within and/or between adjacent G·C·G·C tetrads

We also detect NOEs that involve exchangeable and non-exchangeable protons that could potentially originate between protons within a given G·C·G·C tetrad or between adjacent G·C·G·C tetrads in the d(G-C-G-G-T_3-G-C-G-G) quadruplex. This ambiguity is associated with strand symmetry that relates the two adjacent G·C·G·C tetrads in the interior of the quadruplex. Thus, the imino proton of G3 exhibits NOEs to the H8 and H1' protons of G10 (peaks b and c, Figure 5(a) and (b)) while the imino proton of G10 exhibits NOEs to the H8 and H1' protons of G3 (peaks f and g, Figure 5(a) and (b)). In addition, NOEs are detected between the C2(H5) and G3(H8) protons (peak b, Figure 6(a)) and between the C9(H5) and G10(H8) protons (peak d, Figure 6(a)).

Table 1. Proton chemical shifts[a] (ppm) for the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM NaCl, aqueous buffer

	NH1	NH_2	H8/H6	H5/CH_3	H1'	H2'	H2"	H3'	H4'
G1	11.96		7.74		6.22	3.13	3.02	4.99	4.30
C2		9.04, 9.04	7.29	5.21	6.18	2.30	2.87	4.86	4.39
G3	13.10	8.11, 5.00	7.49		5.80	2.24	2.52	4.82	4.29
G4	11.38		7.77		5.98	2.45	2.38	5.09	4.53
T5			8.04	2.08	6.55	2.16	2.63	4.80	4.51
T6			7.34	1.73	5.82	1.80	2.07	4.64	3.97
T7			6.98	1.59	5.49	1.92	2.52	4.62	4.04
G8	11.65		7.90		6.31	3.84	3.01	4.94	4.45
C9		9.17, 8.60	7.19	5.41	6.04	2.38	2.93	4.73	4.38
G10	12.99	8.26, 5.40	7.53		5.97	2.09	2.55	4.82	4.23
G11	11.80		8.08		6.33	2.70	2.55	4.77	4.39

[a] Exchangeable proton chemical shifts at 0°C. Non-exchangeable proton chemical shifts at 20°C.

NOEs between G·G·G·G tetrad and T-T-T loop segment

The alignment of the T5-T6-T7 loop segment relative to the G1·G4·G8·G11 tetrad is defined by a set of NOEs involving both exchangeable and non-exchangeable protons in the d(G-C-G-G/I-T_3-G-C-G-G) quadruplexes. Specifically, we detect NOEs between the G1 (imino) and T6(H1') protons (peak 1, Figure 5(a) and (b)), between the G8 (imino) and T7(H1') protons (peak u, Figure 5(a) and (b)) and the G4/I4 (imino) and T6(H1') protons (peak y, Figure 5). In addition, NOEs are detected between the G11(H8) and T7(H1') pairs (peak e, Figure 6(a)) and between the G4(H1') and T6(H6) pairs (peak f, Figure 6(a)). These NOEs should define the potential stacking of individual thymine bases over each other and over the G·G·G·G tetrad plane in the quadruplex.

Head-to-tail dimerization of hairpins

The NMR experimental data permit us to distinguish between the d(G-C-G-G-T_3-G-C-G-G) quadruplex generated through head-to-tail dimerization of hairpins (T_3 loops positioned at opposite ends of quadruplex) *versus* one generated through head-to-head dimerization of hairpins (T_3 loops positioned at the same end of quadruplex). The identification of NOEs between the G11(H8) and T7(H1') protons (peak e, Figure 6(a)) and between the G11(NH1) and T7(CH_3) protons conclusively establish that the quadruplex is formed through head-to-tail dimerization of hairpins as shown schematically in Figure 2.

Further, a head-to-tail dimerization of hairpins results in the G1·G4·G8·G11 G-tetrad planes being positioned at either end of the quadruplex (Figure 2) while a head-to-head alignment of hairpins would result in G4·G8·G4·G8 and G1·G11·G1·G11 G-tetrads being positioned at the two ends of the quadruplex. The observed NOEs between guanine imino and guanine H8 protons for the G1 and G4 pairs and the G8 and G11 pairs (Figure 5) are only consistent with the former possibility involving a head-to-tail alignment of hairpins in the d(G-C-G-G-T_3-G-C-G-G) quadruplex.

Lateral hairpin loops

The d($G_4T_4G_4$) sequence dimerizes in a head-to-tail alignment to form a G-quadruplex containing lateral loops in the crystalline state (Kang *et al*., 1992) and containing diagonal loops in solution

Table 2. Nitrogen (at 0°C), carbon and phosphorus (at 20°C) chemical shifts (ppm) for the d(G-C-G-G-T_3-G-C-G-G) quadruplex in 150 mM NaCl, aqueous buffer

Residue	N1/N4	^{31}P[c]	C8/C6	C5/CH_3	C1'	C2'	C3'	C4'
G1	148.20	–4.30	141.45		88.83	37.51	79.73	85.25
C2	110.10	–4.11	141.12	99.22	86.56	40.76	77.46	
G3	149.84	–4.25	136.90		86.23		[b]	85.58
G4	146.46	–3.87	137.23		84.93		81.38	87.86
T5	[a]	–4.74	139.50	16.69	88.83		79.73	88.50
T6	[a]	–3.81	139.82	16.69	87.86		78.46	88.50
T7	[a]	–4.42	139.82	16.37	91.75	44.01	81.68	88.18
G8	149.65	–4.25	143.07		89.48	35.24	80.06	88.83
C9	107.40	–3.75	140.47	98.57	87.86	41.73	76.81	
G10	149.70	–4.18	137.23		86.23		[b]	85.25
G11	145.84		139.50		85.26		74.86	

[a] ^{15}N chemical shifts at 157.00 and 158.82 ppm can be assigned to loop thymine N3 positions but we are unable to distinguish between T5, T6 and T7 residues.

[b] ^{13}C chemical shifts of 75.51 and 78.76 ppm can be asigned to the C3' positions of guanine but we are unable to distinguish between G3 and G10 residues.

[c] ^{31}P chemical shift corresponds to residue (*n*) in (*n*)-^{31}P-(*n* + 1) step.

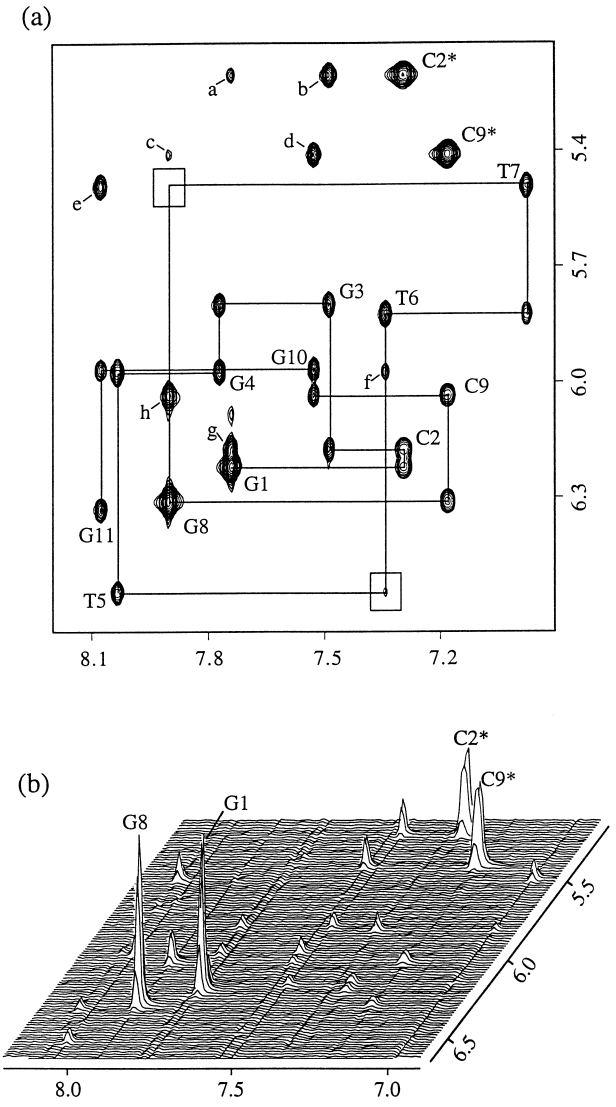

Figure 6. Expanded NOESY (a) contour plot (300 ms mixing time) and (b) stacked plot (50 ms mixing time) of the d(G-C-G-G-T_3-G-C-G-G) in 150 mM NaCl, ^2H_2O buffer at 20°C. NOE cross peaks between the base protons (6.9 to 8.2 ppm) and the sugar H1' and cytidine H5 protons (5.1 to 6.6 ppm). The NOE connectivities between the base protons and their own and 5'-flanking sugar H1' protons are traced from G1 to G11 in plot (a). The missing connectivity at the T7-G8 step is boxed, as is the weak connectivity at the T5-T6 step. Intra-residue cross peaks between H6 and H5 protons of residues C2 and C9 are designated by asterisks. Cross peaks a to h are assigned as follows: a: G1(H8)-C2(H5); b: G3(H8)-C2(H5); c: G8(H8)-C9(H5); d: G10(H8)-C9(H5); e: G11(H8)-T7(H1'); f: T6(H6)-G4(H1'); g: G1(H8)-C2(H1'); h: G8(H8)-C9(H1').

(Smith & Feigon, 1992; Schultze *et al*., 1994a). We can rule out a diagonal loop folding topology for the d(G-C-G-G-T_3-G-C-G-G) quadruplex since such an alignment would result in the formation of internal G·G·C·C tetrads while the experimental data are only consistent with formation of internal G·C·G·C tetrads. A G·G·C·C tetrad can only form through pairing of the minor groove edge of one G·C pair with the major groove edge of the other G·C pair. Only one of the two cytosine amino protons would be involved in hydrogen bonding within a G·G·C·C tetrad contrary to our demonstration that both cytosine amino protons form hydrogen bonds in the d(G-C-G-G-T_3-G-C-G-G) quadruplex.

Further, the guanine alignments around the G-tetrads would be different with a G1·G4·G8·G11 arrangement for lateral loops (Figure 2) and either G1·G4·G11·G8 or G1·G8·G11·G4 arrangements for diagonal loops. The G4 and G8 guanine residues and the G1 and G11 guanine residues would be positioned across the diagonals of the G-tetrads in the diagonal loop alignment of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. The observed NOEs between guanine imino and guanine H8 protons for the G4 and G8 pairs and for the G11 and G1 pairs (Figure 5) would not be observed for the diagonal loop folding topology but are consistent with the lateral loop folding topology where the G4 and G8 guanine residues and the G11 and G1 guanine residues are adjacent to each other in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Figure 2).

Input restraints and structure calculations

Distance restraints for the non-exchangeable protons were quantified from NOE buildup curves in NOESY experiments (at five mixing times) on the d(G-C-G-G-T_3-G-C-G-G) quadruplex in ^2H_2O solution. Distance restraints for the exchangeable protons were qualitatively deduced from NOESY data sets (two mixing times) on the d(G-C-G-G-T_3-G-C-G-G) quadruplex in H_2O solution. Hydrogen-bond distance restraints were also incorporated during the computations. Dihedral angle restraints were used for the glycosidic torsion angles (χ). (For details relating to the input restraint sets used at the various stages of the calculations see Materials and Methods.) The input restraint statistics for the structure calculations of the d(G-C-G-G-T_3-G-C-G-G) quadruplex are listed in Table 3. The distribution of distance restraints at each residue in the d(G-C-G-G-T_3-G-C-G-G) quadruplex are plotted in Figure 7(a) and (b).

The structure calculations on the d(G-C-G-G-T_3-G-C-G-G) quadruplex were undertaken in three stages. Following metric matrix distance geometry (MMDG) and simulated-annealing regularization, six structures were selected out of 100 based on their low covalent energies and relatively few input-restraint violations. These structures were refined in the second stage against distance and dihedral angle restraints using molecular dynamics (MD) calculations, and subsequently four of these were arbitrarily chosen for refinement in the third stage against NOE intensity restraints from NOESY data sets at five mixing times. (The protocols for the refinement computations are outlined in Materials and Methods.) The quality of d(G-C-G-G-T_3-G-C-G-G) structures after each stage of the computations is

Table 3. NMR restraint statistics for d(G-C-G-G-T_3-G-C-G-G) quadruplex

Stage I: Distance geometry and simulated annealing	
Non-exchangeable protons	
Number of NOE derived distance restraints	534
Intraresidue	272
Sequential ($i \rightarrow i + 1$)	240
$i \rightarrow i + 2$	22
Stage II: Distance restrained molecular dynamics[a]	
Non-exchangeable protons	
Number of NOE derived distance restraints	267
Intraresidue	136
Sequential ($i \rightarrow i + 1$)	120
$i \rightarrow i + 2$	11
Exchangeable protons	
Number of NOE derived distance restraints	44
Sequential ($i \rightarrow i + 1$)	17
$i \rightarrow i + 2$	4
Long range	23
Stage III: Relaxation matrix based NOE intensity refinement	
Number of NOE intensity restraints on non-exchangeable protons[b]	1335
Number of NOE derived distance restraints on exchangeable protons[a]	44
Additional restraints for all stages	
Hydrogen bond distance restraints	64
Torsion restraints on χ angles	16
Non-crystallographic symmetry restraint	On all heavy atoms

[a] NOE distance restraints during these stages were SUM averaged, with number of monomers set to 2.
[b] NOE intensity restraints were specified ambiguously with multiple atom selection. The number is for five NOESY data sets collected as a function of mixing time.

summarized in Table 4. The final relaxation-matrix-refined structures of the d(G-C-G-G-T_3-G-C-G-G) quadruplex exhibit very good *R*-factors and convergence characteristics.

Structure analysis

The root mean square deviation (rmsd) values at individual residues between the four relaxation-matrix-refined structures of the d(G-C-G-G-T_3-G-C-G-G) are plotted in Figure 7(c). A stereo view of the four superpositioned refined structures looking normal to the helix axis is shown in Figure 8(a). A stereo view looking down the helix axis and including only the tetrad segments is shown in Figure 8(b).

A schematic representation of the folding topology of the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Na^+-containing solution, as determined from the NMR/MD study is shown in Figure 2. This view schematically traces the chain directionality and individual strands of the d(G-C-G-G-T_3-G-C-G-G) quadruplex and identifies the guanine bases involved in G·G·G·G and G·C·G·C tetrad formation. There are two distinct narrow (N_1 and N_2) and two symmetrical wide (W) grooves in the d(G-C-G-G-T_3-G-C-G-G) quadruplex in Figure 2. One representative structure of the d(G-C-G-G-T_3-G-C-G-G) quadruplex was chosen from the four relaxation-matrix-refined structures based on its smallest rmsd value relative to the average of the four structures. This representative structure is depicted in two alternative ways in Figure 9 with the views looking into the narrow N_2 groove (Figure 9(a)) and the wide groove (Figure 9(b)) of the quadruplex.

The stacking patterns between adjacent G·G·G·G and G·C·G·C tetrads in the representative refined structure of the d(G-C-G-G-T_3-G-C-G-G) quadruplex is shown as a stereo pair in Figure 10(a). The corresponding stacking patterns between adjacent symmetry related G·C·G·C tetrads is shown as a stereo pair in Figure 10(b).

Discussion

Spectral quality and sequence dependence

The quality of proton NMR spectra of the monovalent cation stabilized G-quadruplexes containing solely G-tetrads have varied widely with good quality spectra reported for the d(TG_4T) sequence (Wang & Patel, 1992), d($G_2N_5G_2$) sequence (Wang *et al*., 1991), the *Oxytricha* d($G_4T_4G_4$) (Smith & Feigon, 1992; Schultze *et al*., 1994a) and d[$G_4(T_4G_4)_3$] (Smith & Feigon, 1992; Wang & Patel, 1995) telomere sequences, the d($G_3T_4G_3$) sequence (Scaria *et al*., 1992; Smith *et al*., 1994a) and the thrombin binding DNA aptamer sequence (Macaya *et al*., 1993; Wang *et al*., 1993; Schultze *et al*., 1994b). There was some deterioration in the proton spectral quality for the human d[$AG_3(T_2AG_3)_3$] telomere

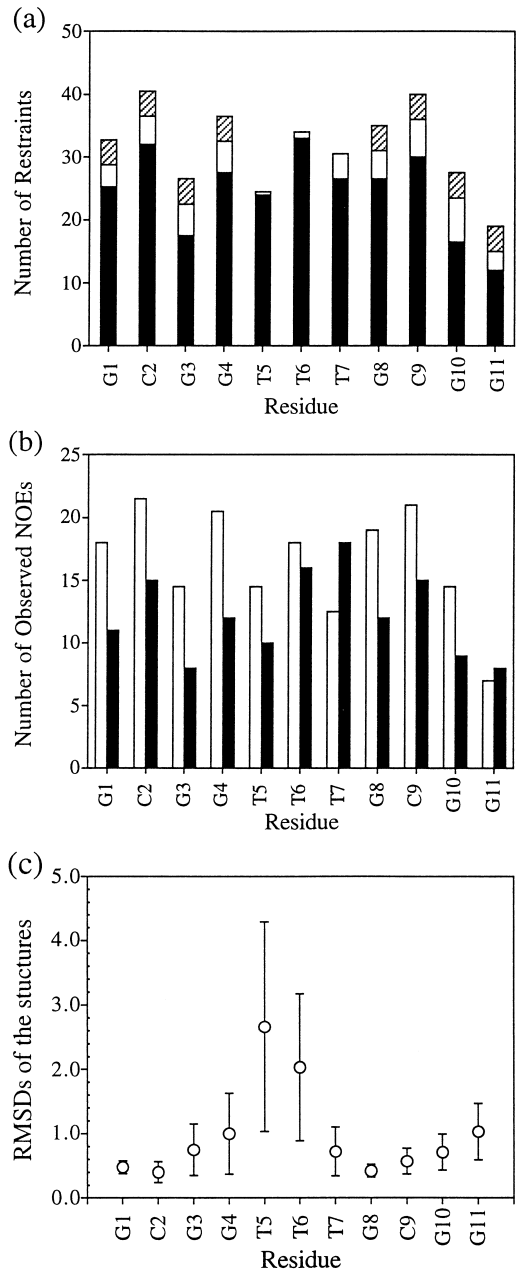

Figure 7. Distribution of restraints by residue in the structure calculations of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. (a) The restraints have been classified as follows: non-exchangeable proton restraints (filled rectangles), exchangeable proton restraints (open rectangles) and hydrogen bond restraints (hatched rectangles). Each intra-residue restraint is counted as one for that residue. Each inter-residue restraint is counted as half for that residue and half for its partner in the interaction. (b) Distribution by residue position of intra-residue (open rectangles) and inter-residue (filled rectangles) experimentally observed exchangeable proton and non-exchangeable proton restraints. (c) rmsds (Å) in the atomic coordinate positions calculated by residue for the four relaxation-matrix-refined structures and the average of these structures. A least-squared all-heavy-atom superposition of the well-defined residue segments (1–4 and 8–11) of the quadruplex was performed before the calculation. The average values (open circles) are shown along with the standard deviations (vertical bars).

sequence (Wang & Patel, 1993b) and a more dramatic deterioration for the *Tetrahymena* $d(T_2G_4)_4$ telomere sequence (Henderson *et al*., 1987; Wang & Patel, 1994). The proton spectra of the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Figure 3(a) and (b)) are quite spectacular with narrow and well resolved resonances along with the complete absence of secondary conformations and broad spectral envelopes such as observed previously for the *Tetrahymena* $d(T_2G_4)_4$ telomere sequence (Henderson *et al*., 1987; Wang & Patel, 1994). The observation of only a single set of proton resonances establishes that the two hairpins that align to form the d(G-C-G-G-T_3-G-C-G-G) quadruplex are symmetric in solution.

By contrast, the proton NMR spectra of the related d(G-G-C-G-T_3-G-G-C-G) sequence where C-G steps replaced the internal G-C steps were of poor quality in Na^+ solution due to the presence of broad resonances and multiple conformations. Thus, the precise sequence is critical for formation of quadruplexes containing G·C·G·C tetrads sandwiched between G·G·G·G tetrads as reported in this study.

Folding topology

The Na^+ stabilized d(G-C-G-G-T_3-G-C-G-G) quadruplex is formed through a head-to-tail dimerization of a pair of hairpins such that adjacent strands are anti-parallel to each other around the quadruplex (Figure 2). Such a structure exhibits strand symmetry and contains two unique narrow and two equivalent wide grooves. Such a folding topology was also reported for the X-ray structure of the K^+ stabilized *Oxytricha* $d(G_4T_4G_4)$ sequence which involved head-to-tail dimerization to form the G-quadruplex (Kang *et al*., 1992). For both structures the T_n loops are of the lateral (Kang *et al*., 1992) rather than the diagonal (Smith & Feigon, 1992; Schultze *et al*., 1994a,b; Wang & Patel, 1993b, 1995) type.

We also note that individual hairpins in the d(G-C-G-G-T_3-G-C-G-G) quadruplex involve formation of two internal Watson-Crick G(*anti*)·C(*anti*) base-pairs and two flanking Hoogsteen G(*anti*)·G(*syn*) mismatch pairs (Figure 2). The dimerization of the hairpins involve the mutual alignments of the major groove edges of the Watson-Crick G·C pairs to form the G·C·G·C tetrads (Figure 1(b)) and the major groove edges of the Hoogsteen G·G mismatch pairs to form the G·G·G·G tetrads (Figure 1(a)). It is conceivable that formation of the pair of internal G·C·G·C tetrads is facilitated by the formation of the flanking G·G·G·G tetrads and that it should be possible to generate alternate combinations of these two types of tetrads to form stable quadruplexes.

There was little change in the guanine exchangeable imino proton spectrum of the d(G-C-G-G-T_3-G-C-G-G) quadruplex on raising the temperature from 0°C (Figure 3(a)) to 40°C. Thus, the quadruplex

Table 4. Quality of the d(G-C-G-G-T_3-G-C-G-G) quadruplex structures

Parameter	DG and SA structures	Dist. restrained MD structures	Relaxation matrix MD structures
Convergence			
Pairwise r.m.s.d. values (Å) for residues 1 to 4, 8 to 11	1.82 ± 0.21	1.08 ± 0.30	0.76 ± 0.20
R-factors			
$R_{1/6}$	0.094–0.112	0.067–0.082	0.056–0.065
Unweighted R_1	0.429–0.556	0.325–0.423	0.252–0.288
Weighted R_1	1.917–3.418	1.209–2.019	1.295–1.424
Number of distance violations (>0.2 Å)	6–10	0	0[a]
r.m.s.d. values of distance restraints (Å)	0.062–0.096	0.037–0.058	0.030–0.034[a]
Number of dihedral violations (>10°)	0	0	0
r.m.s.d. values for dihedral restraints	0.003–2.054	1.841–4.915	2.828–4.049
r.m.s.d. values from ideal covalent geometry			
Bond length (Å)	0.011–0.016	0.007–0.014	0.012–0.015
Bond angles (°)	3.421–4.069	3.035–3.855	4.083–4.430
Impropers (°)	0.184–0.351	0.205–0.394	0.484–0.708

[a] The number of violations and the r.m.s.d. values for distance restraints during relaxation matrix MD refinement do not include restraints involving non-exchangeable protons for which the force constant was scaled to zero during this stage of the calculation.

structure is quite stable at physiological temperature.

We have identified *syn* guanine residues at positions G1 and G8 corresponding to G(*syn*)-C(*anti*)-G(*anti*)-G(*anti*) alignments along individual strands of the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Figure 2). We also note an alternating G(*syn*)·G(*anti*)·G(*syn*)·G(*anti*) pattern around the individual G-tetrads in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Figure 2), similar to that reported for G-quadruplexes whose adjacent strands are aligned in an anti-parallel orientation (Kang *et al*., 1992; Macaya *et al*., 1993; Wang *et al*., 1993; Schultz *et al*., 1994b; Padmanabhan *et al*., 1993).

The guanine sugar puckers within the G·G·G·G tetrads, as well as the guanine and cytosine sugar puckers within the G·C·G·C tetrads span the C1'-*exo* / C2'-*endo* / C3'-*exo* range (P = 128 to 196°) in the four refined structures of the d(G-C-G-G-T_3-G-C-G-G) quadruplex.

Helical twist and stacking between adjacent tetrads

The structure calculations on the d(G-C-G-G-T_3-G-C-G-G) quadruplex were undertaken without incorporation of tetrad planarity restraints. The bases are essentially coplanar within individual tetrads in the solution structure of the quadruplex (Figure 9(a) and (b)).

The adjacent terminal G1·G4·G8·G11 and internal C2·G3·C9·G10 tetrads are related by a right-handed helical twist of 34° with same strand stacking between bases (Figure 10(a)). We detect moderate stacking between G3 and G4 and between G10 and G11 purine pairs through overlap of their five and six-membered rings (Figure 10(a)). By contrast, there is less stacking between G1 and C2 and between G8 and C9 pairs (Figure 10(a)).

The symmetry related adjacent C2·G3·C9·G10 tetrads are related by a right-handed helical twist of 37° with cross-strand stacking between bases (Figure 10(b)). We detect moderate stacking between G3 and G10 purine pairs through overlap of their six-membered rings while no overlap is observed between equivalent C2 and equivalent C9 pyrimidine pairs (Figure 10(b)).

Previous structural studies have identified monovalent cation sites sandwiched between G-tetrad planes in G-quadruplexes (Kang *et al*., 1992; Laughlan *et al*., 1994). By contrast, it is unclear whether monovalent cation sites can be generated between adjacent tetrads in the d(G-C-G-G-T_3-G-C-G-G) quadruplex since the internal coordination sites are now composed of a mixture of favorable guanine O^6 oxygen atoms and unfavorable cytosine N^4 amino groups.

Groove dimensions

The distinct sequence composition and alternating strand directions of the 2-fold symmetric d(G-C-G-G-T_3-G-C-G-G) quadruplex subdivides the four grooves into several distinct types with unique dimensions (Figure 9) and functional edges (Figure 10). Narrow grooves are formed by the anti-parallel alignment of adjacent G1-C2-G3-G4 strands with a groove width of 6.0 Å and adjacent G8-C9-G10-G11 strands (Figure 9(a)) with a groove width of 6.3 Å (measured as the shortest phosphorus-phosphorus distance across the groove). Symmetrical wide grooves are formed by the anti-parallel alignment of adjacent G1-C2-G3-G4 and G8-C9-G10-G11 strands of each hairpin (Figure 9(b)), with a groove width of 8.5 Å. Such alternating narrow and wide grooves have also been observed in G-quadruplex structures where adjacent strands are aligned anti-parallel to each other (Kang *et al*., 1992; Macaya *et al*., 1993; Wang *et al*.,

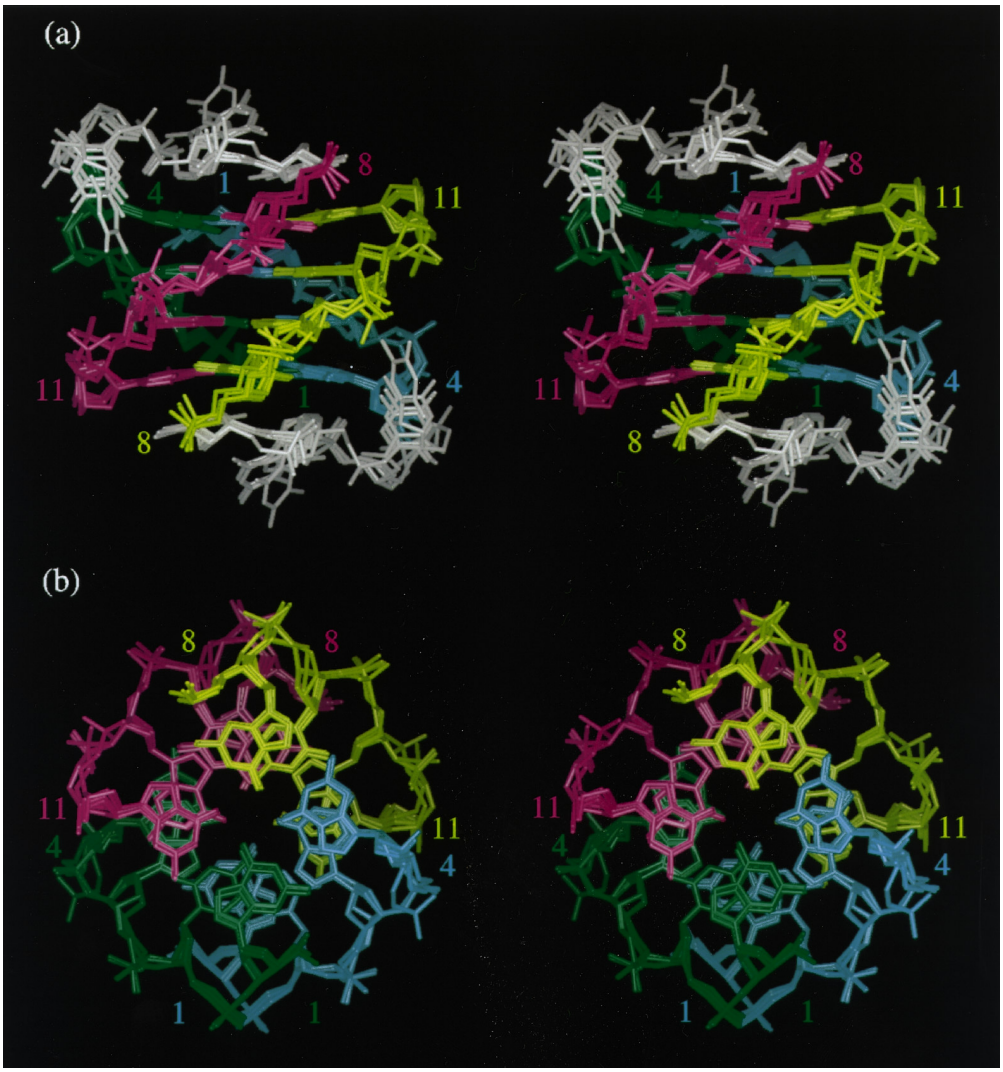

Figure 8. Stereo views of the four relaxation-matrix refined structures of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. The G1-C2-G3-G4 segments are shown in green and cyan, the T5-T6-T7 loop segments are shown in white and the G8-C9-G10-G11 segments are shown in magenta and yellow. (a) View normal to the helix axis looking into the narrow groove formed by the two G8-C9-G10-G11 segments aligned in an antiparallel orientation. (b) View looking down the helix axis showing only the tetrad segments.

1993; Schultze *et al*., 1994b; Padmanabhan *et al*., 1993).

Lateral T_3 loop segment

The T5-T6-T7 segments form lateral hairpin loops connecting G1-C2-G3-G4 and G8-C9-G10-G11 segments in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Figure 2). These loops are aligned at opposite ends of the quadruplex and are related by a 2-fold element of symmetry. The T7 base in the loop cross-strand stacks over the G11 base of the G-tetrad while the T5 base loops out into the wide groove of the quadruplex. The T6 base is also positioned over the G-tetrad but at some distance from it. The rmsd values between the four refined structures establishes that the stacked T7 base is the best defined loop residue while the looped out T5 base is the least well defined of the residues in the T_3 loop segment (Figure 7(c)).

Sequential and cross-strand NOEs between adjacent tetrads

We outlined in Results several unusual NOEs between the symmetry related C2·G3·C9·G10 tetrads that must reflect combinations of sequential and cross-strand contributions in the folded conformation of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. The folding topology of the symmetrical head-to-tail alignment of this quadruplex is shown in Figure 2. It is readily apparent that NOEs, for instance, between protons on the C9 and G10 residues could reflect the sequential interaction within the same strand in the C9-G10 step and/or reflect a cross-strand interaction within the

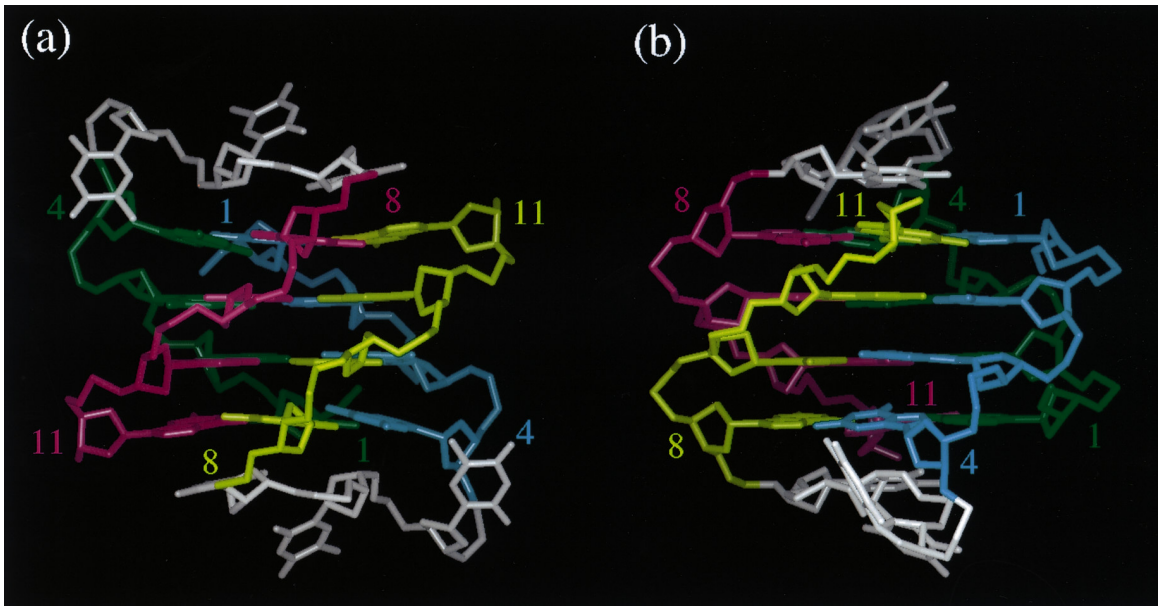

Figure 9. Views normal to the helix axis and looking (a) into the narrow N_2 groove and (b) into the wide groove in the representative relaxation matrix refined structure of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. View generated using the ''stick'' representation (INSIGHT II program). The G1-C2-G3-G4 segments are shown in green and cyan, the T5-T6-T7 segments are shown in white and the G8-C9-G10-G11 segments are shown in magenta and yellow. All hydrogen atoms and phosphate oxygen atoms in the backbone have been deleted for clarity.

C2·G3·C9·G10 tetrad. The ''SUM'' averaging option in XPLOR considers the contributions from both possibilities during the molecular dynamics refinements and it is instructive to tabulate the distances between relevant proton pairs associated with sequential and cross-strand interactions in the refined structures. These data identify the relative sequential and cross-strand contributions to the unusual NOEs observed experimentally.

We also note that the hydrogen bonds are aligned in the same direction for adjacent stacked G1·G4·G8·G11 and C2·G3·C9·G10 tetrads in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (hydrogen bond directionalities shown by small arrows within tetrad planes, Figure 2). This alignment readily explains the experimentally observed sequential NOEs between the imino protons of G3 and G4 (peak a, Figure 4) and between the imino protons of G10 and G11 (peak b, Figure 4) in the quadruplex. The d(G-C-G-G-T_3-G-C-G-G) quadruplex structure also readily explains the very weak sequential NOEs between guanine H8 and cytosine H5 protons and moderate sequential NOEs between guanine H8 and cytosine H1' protons at the G1(*syn*)-C2(*anti*) and G8(*syn*)-C9(*anti*) steps.

The NOEs between proton pairs in the symmetry related C2·G3·C9·G10 tetrads could reflect contributions from sequential and/or cross-strand contributions in the d(G-C-G-G-T_3-G-C-G-G) quadruplex. The structure was solved using the SUM averaging option in X-PLOR and the relative contributions of sequential and cross-strand NOEs to the intensity of a given cross-peak can be evaluated from the corresponding inter-proton distances in the representative refined structure of the quadruplex.

Heteronuclear chemical shift patterns

There are distinct ^{15}N and ^{13}C chemical shift differences between *syn* and *anti* guanine residues around the G1(*syn*)·G4(*anti*)·G8(*syn*)·G11(*anti*) tetrads in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Table 2). Thus, the *syn* G1 and G8 guanine residues exhibit N1 nitrogen chemical shifts that are ~3 ppm downfield relative to the *anti* G4 and G11 guanine counterparts.

Similarly, the C8 base and C1' sugar carbons for the *syn* G1 and G8 are ~4 ppm downfield relative to their *anti* G4 and G11 counterparts in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (Table 2). The latter carbon chemical shift difference patterns between *syn* and *anti* guanine residues in G-tetrads have been noted previously (Wang *et al*., 1991). Indeed, downfield guanine C8 and C1' carbon chemical shifts, given a fixed C2'-*endo* sugar conformation, are characteristic of *syn* guanine glycosidic torsion angles in nucleic acids (Ghose *et al*., 1994; Greene *et al*., 1995).

The majority of the phosphorus chemical shifts are dispersed between –3.75 and –4.45 ppm except for the phosphorus resonance corresponding to the T5-T6 step which is upfield shifted to –4.74 ppm (Table 2). This shift may reflect the contribution of the backbone to chain reversal in the T_3 loop segment, with a distinct role for the T5-T6 step due to the looping out of the T5 base into the wide groove.

G·C·G·C tetrads

The alignment of bases around the G·C·G·C tetrad in the d(G-C-G-G-T_3-G-C-G-G) quadruplex

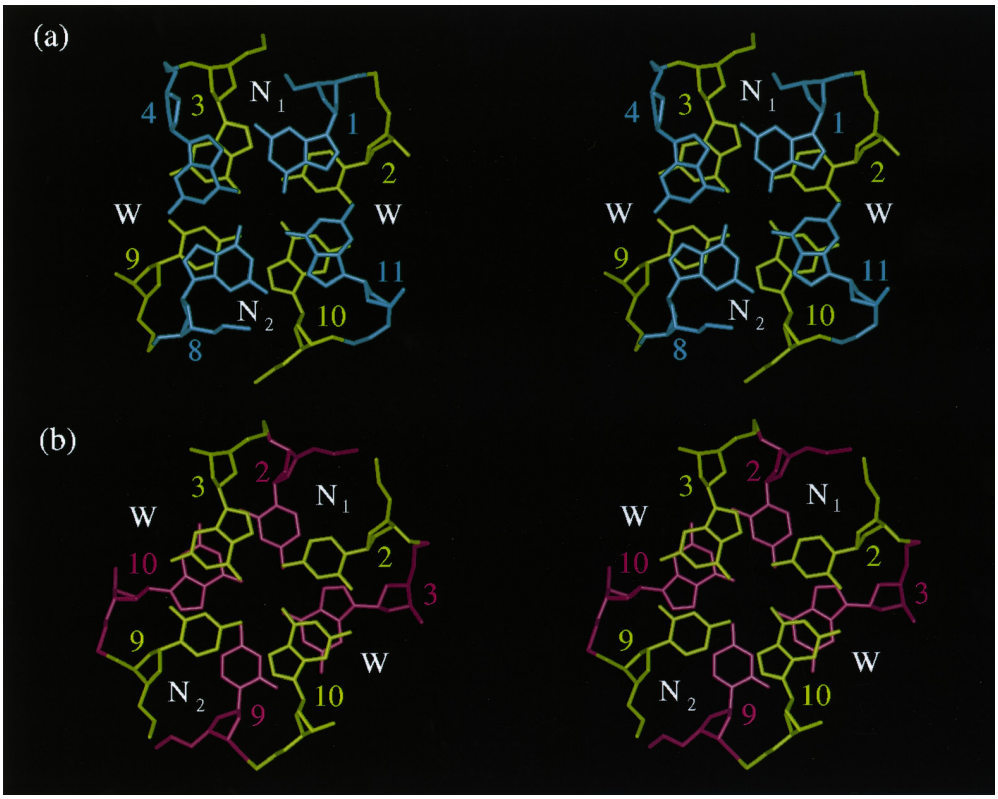

Figure 10. Two stereoviews of the overlap between adjacent tetrads in the representative refined structure of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. The two distinct narrow grooves of the quadruplex are labelled as N_1 and N_2 while the symmetric wide grooves are labelled W. (a) View down the helix axis with the G1·G4·G8·G11 tetrad (cyan) stacked on top of the C2·G3·C9·G10 tetrad (yellow). (b) View down the helix axis with adjacent C2·G3·C9·G10 tetrads (magenta and yellow) stacked on each other (yellow on top of magenta). All hydrogen atoms and phosphate oxygen atoms in the backbone have been deleted for clarity.

(Figure 2) are shown schematically in Figure 1(b). The Watson-Crick G·C base-pairs align through the major groove edges with the non-Watson-Crick paired cytosine amino protons donating bifurcated hydrogen bonds to guanine O^6 and N^7 acceptors across the tetrad. It is significant that both cytosine amino protons are involved in hydrogen bonding within the G·C·G·C tetrads consistent with both these protons resonating in the hydrogen-bonded 8.6 to 9.3 ppm region (Figure 3(a)). It should be noted that the hydrogen bonding patterns on adjacent G·C·G·C tetrads are directed in opposite directions in the d(G-C-G-G-T_3-G-C-G-G) quadruplex (hydrogen bond directionalities indicated by small arrows within tetrad planes, Figure 2).

It is noteworthy that the G·C·G·C tetrad (Figure 1(b)) involving alignment of Watson-Crick G·C base-pairs through their major groove edges was identified by O'Brien (1967) close to 30 years ago in an X-ray crystal structure of a 1:1 complex of 9-ethylguanosine and 1-methylcytosine. The key advance associated with our present contribution is that the G·C·G·C tetrads are part of a quadruplex with adjacent strands aligned anti-parallel to each other.

There are distinct differences between the G·C·G·C tetrads reported in this study for the d(G-C-G-G-T_3-G-C-G-G) quadruplex in solution (Figure 1(b)) and those reported recently for the d(G-C-A-T-G-C-T) quadruplex in the crystalline state (Leonard *et al*., 1995). In the present study on the d(G-C-G-G-T_3-G-C-G-G) quadruplex, Watson-Crick G·C base-pairing occurs within individual hairpin stems and G·C·G·C tetrad formation involved alignment of the major groove edges of the G·C base-pairs. The G·C·G·C tetrads approach planarity in this quadruplex. The T_3 containing hairpin loops are positioned at opposite ends of the quadruplex and alignment is facilitated by formation of G·G·G·G tetrads at both ends of the quadruplex. By contrast, in the previous study of the d(G-C-A-T-G-C-T) quadruplex, Watson-Crick G·C base-pairing occurs across hairpin stems and G·C·G·C tetrad formation involves alignment of the minor groove edges of the base-pairs (Leonard *et al*., 1995). The G·C·G·C tetrads are distorted significantly from planarity in this quadruplex. The A-T containing hairpin loops are positioned at one end of the quadruplex and alignment is facilitated through A·A mismatch pairing at the loop containing end of the quadruplex. It therefore appears that dimerization of Watson-Crick G·C

pairs can occur through alignment of either the major groove (this study) or minor groove (Leonard *et al*., 1995) edges to generate G·C·G·C tetrads.

Cytosine methylation and quadruplex formation

It has been established previously that expansion of the $d(C\text{-}G\text{-}G)_n \cdot d(C\text{-}C\text{-}G)_n$ repeats leading to the onset of the fragile X syndrome is facilitated by hypermethylation of cytosine residues (Oberle *et al*., 1991; Smith *et al*., 1994b). This result appears to correlate with the recent demonstration that the methylated analog $d(^5meC\text{-}G\text{-}G)_5$ forms a quadruplex in monovalent cation solution while a longer sequence is required to generate the corresponding quadruplex in the unmethylated $d(C\text{-}G\text{-}G)_7$ counterpart (Fry & Loeb, 1994). An examination of the stacking patterns between adjacent tetrads in the solution structure of the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ quadruplex establishes that incorporation of a 5-methyl group on cytosine will introduce favorable stacking interactions. Thus, the methyl group of ^{5}meC within the C2·G3·C9·G10 tetrad would stack directly over the same-strand guanine of the adjacent G1·G4·G8·G11 tetrad in one direction (Figure 10(a)) and cross-strand stack with a symmetry related cytosine of the adjacent symmetry related C2·G3·C9·G10 tetrad in the opposite direction (Figure 10(b)). Thus, the solution structure of the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ quadruplex provides a plausible molecular link between cytosine methylation and facilitation of quadruplex formation (Fry & Loeb, 1994).

Biological implications for the fragile X syndrome

We have demonstrated that the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ sequence, which contains elements of the fragile X syndrome repeat, folds into a novel quadruplex structure. This result provides a structural foundation to build on the earlier demonstration of tetraplex DNA formation by $d(C\text{-}G\text{-}G)_7$ and $d(^5meC\text{-}G\text{-}G)_5$ in monovalent cation solution (Fry & Loeb, 1994). Specifically, we have established that $d(C\text{-}G\text{-}G)_n$ sequences can potentially align through formation of one G·G·G·G and two G·C·G·C tetrads for each trinucleotide repeat unit and that this is most likely achieved through dimerization of a pair of hairpin structures. The formation of such quadruplexes, which require the head-to-tail alignment of two hairpins with anti-parallel alignment of adjacent strands, are likely to depend on the length of the $d(C\text{-}G\text{-}G)_n$ repeat. These quadruplex structures could serve as potential blockage sites for the progress of replication forks and account for the blockage at the fragile X locus observed experimentally (Hansen *et al*., 1993).

It should be pointed out that a continuous $d(C\text{-}G\text{-}G)_n$ fragile X syndrome triplet repeat can form a hairpin structure through pairing alignments in three different registers. These correspond to d(C-G-G)·d(C-G-G), d(C-G-G)·d(G-G-C) and d(C-G-G)·d(G-C-G) pairing arrangements with the first and third alignments involving two G·C pairs and one G·G mismatch while the second alignment involves two G·G and one C·C mismatches. The current study on the $d(G\text{-}C\text{-}G\text{-}G\text{-}T_3\text{-}G\text{-}C\text{-}G\text{-}G)$ quadruplex (Figure 2) represents an example of the d(C-G-G)·d(G-C-G) pairing arrangement within the hairpin stem segments and does not address the potential for the d(C-G-G)·d(C-G-G) and d(C-G-G)·d(G-G-C) pairing registries to form higher order nucleic acid structures.

It should also be noted that for a continuous $d(C\text{-}G\text{-}G)_n$ fragile X syndrome repeat to adopt the quadruplex folding topology shown in Figure 2, the G11 residue at the 3'-end of one hairpin stem would have to connect with the G1 residue at the 5'-end of the other hairpin stem. This represents the longest possible connecting distance and hence such quadruplex structures can only occur for continuous $d(C\text{-}G\text{-}G)_n$ sequences where *n* is a large number. This requirement is consistent with the demonstration that the onset of the disease is associated with increasing values of *n* as one proceeds from normal individuals, to premutagenic carriers and finally to individuals afflicted with the fragile X syndrome.

Biological implications for duplex alignments in genetic recombination

Our demonstration of adjacent G·C·G·C tetrads as part of a DNA quadruplex also has important implications for alignment of duplex segments involved in genetic recombination. Homologous DNA segments could be brought into register through G·C·G·C and A·T·A·T tetrad formation as the first step prior to the onset of strand exchange mediated through a pair of Holliday junction crossover sites. Our observation that Watson-Crick G·C pairs can dimerize through their major groove edges to form G·C·G·C tetrads (Figure 1(b)) is a key feature since A·T pairs can also potentially dimerize through their major groove edges to form A·T·A·T tetrads. The glycosidic bonds are directed towards the four corners for the G·C·G·C (Figure 1(b)) and A·T·A·T tetrads so that there are no impediments to quadruplex formation by a pair of homologous duplexes during recombination. Indeed, stereochemically feasible models have been put forward for four-stranded structures containing G·C·G·C and A·T·A·T tetrads (McGavin, 1971). The potential formation of four-stranded DNA structures during synapsis of two homologous DNAs where strand exchange is excluded has been probed through electron microscopic studies (Dombroski *et al*., 1988).

Materials and Methods

Oligonucleotide synthesis and purification

The d(G-C-G-G-T_3-G-C-G-G) oligodeoxyribonucleotide and its selectively labelled deoxyinosine analog (I4 for G4 substitution) were synthesized on a 10 μmol scale on an Applied Biosystems model 392 DNA-RNA synthesizer using solid-phase β-cyanoethylphosphoramidite chemistry. The oligodeoxyribonucleotides were deprotected and cleaved from the support following treatment with concentrated ammonium hydroxide for 16 hours at 55°C. The crude 5'-dimethoxytrityl protected products were purified by reverse-phase high performance liquid chromatography (HPLC) on a C4 column (Rainin Instrument Co.) with a linear gradient of 1%/minute of acetonitrile in 0.1 M triethylammonium acetate (pH 6.8), detritylated upon treatment with 80% (v/v) acetic acid for 30 minutes and chromatographed a second time with a linear gradient of 0.5%/minute in the same buffer like the first chromatography. The volume of the collected eluent containing the pure oligodeoxyribonucleotide was next reduced to 10 ml. The product was dialyzed against water several times and against 0.5 M NaCl, 5 mM Na_2HPO_4 at pH 6.5 for four hours and finally against 5 mM NaCl, 0.2 mM Na_2HPO_4 at pH 6.5 and lyophilized.

NMR sample preparation

Approximately 600 to 650 A_{260} units of d(G-C-G-G-T_3-G-C-G-G) were dissolved in 0.6 ml ^2H_2O (~10 mM in single strands) for non-exchangeable proton NMR experiments or in 90% H_2O/10% ^2H_2O for exchangeable proton NMR experiments. The final NaCl concentration was 0.1 to 0.15 M.

NMR experiments

All NMR experiments (NOESY, COSY-45, COSY-90, TOCSY, ^{1}H-^{31}P COSY and ^{1}H-^{13}C HMQC) in ^2H_2O were conducted at 20°C and all NMR experiments (NOESY and ^{1}H-^{15}N HMQC) in H_2O were conducted at 0°C.

The proton NOESY experiments (mixing time 80 and 250 ms) in H_2O were recorded with the H_2O signal suppressed by a jump-and-return pulse sequence as the reading pulse (Plateau & Gueron, 1982) in the hypercomplex mode (States *et al*., 1982). The carrier frequency was placed on the H_2O resonance with the maximum excitation centered at 12.0 ppm.

The proton NOESY experiments in ^2H_2O were recorded at mixing times of 50, 100, 150, 200 and 300 ms in the hypercomplex mode. The spectra were collected with 2048 complex points in the t_2 dimension and 600 real free induction decays in the t_1 dimension with a spectral width of 4.8 kHz. A repetition delay of four seconds was used and the residual H^2HO signal was suppressed by low-power continuous wave irradiation during this delay.

Homonuclear proton total correlation spectroscopy (TOCSY) experiments (Bax & Davis, 1985) were acquired with 2048 t_2 complex points and 600 t_1 increments in the hypercomplex mode with spin-lock times of 40 and 80 ms. The phase-sensitive correlation spectroscopy (COSY-45 and COSY-90) experiments (Marion & Wuthrich, 1983) were acquired with 2048 t_2 complex points and 1024 t_1 increments.

A proton-detected ^{1}H-^{13}C heteronuclear multiple quantum coherence (^{1}H-^{13}C-HMQC) experiment (Bax *et al*., 1983) at natural abundance was recorded with GARP decoupling (Shaka *et al*., 1985) during acquisition.

A proton-detected ^{1}H-^{15}N heteronuclear multiple quantum coherence (^{1}H-^{15}N-HMQC) experiment (Bax *et al*., 1983) at natural abundance was recorded with GARP decoupling (Shaka *et al*., 1985) during acquisition. The H_2O signal suppression was performed by jump-and-return pulse sequence and the maximum excitation was centered either at 12 ppm (for imino protons) or at 8.5 ppm (for amino protons).

A proton-detected ^{1}H-^{31}P heteronuclear correlation experiment (Sklenar & Bax, 1987) was recorded in the hypercomplex mode with 1024 t_2 points and 256 t_1 increments and spectral width of 6 ppm along the ^{31}P dimension.

All experiments were performed on Varian Unity Plus 600 MHz spectrometers. The data sets were processed by Varian software and the real matrices were transformed to the FELIX format. All NOESY, COSY and TOCSY data sets were apodized by a 90°-shifted sinebell function. The final matrix contained 1024×1024 or 2048×2048 real points. All proton chemical shifts were referenced relative to internal sodium 2,2-dimethyl-2-silapentane-5-sulfanate (DSS, for ^{1}H). The chemical shifts of ^{13}C, ^{15}N and ^{31}P in two-dimensional heteronuclear NMR experiments were made by indirect referencing to (3-trimethylsilyl) propionate (TSP, for ^{13}C), ammonia (for ^{15}N) and trimethylphosphate (for ^{31}P), respectively (Live *et al*., 1984). The cross peaks were picked manually and their volumes were integrated using FELIX (Biosym Technologies, Inc.).

Distance and torsion restraints

The distances between non-exchangeable protons were estimated from the buildup of NOE intensities as a function of mixing time (50, 100, 150, 200 and 300 ms) in ^2H_2O. A quadratic polynomial was fitted to the data points and the initial slope was calculated from the linear fit of the quadratic curves to the first non-zero mixing time. The cytosine H5 to H6 distance of 2.47 Å was used as the standard for calibrating initial slopes to inter-proton distances. The upper and lower bounds were uniformly assigned to ±20% of the estimated distance.

Cross peaks involving exchangeable protons in the NOESY spectra (80 and 250 ms mixing times) in H_2O were classified as strong (strong intensity at 80 ms), medium (weak intensity at 80 ms), and weak (peak observed only at 250 ms) and the corresponding proton pairs were restrained to distances of 2.7 Å (–0.7/+0.8), 3.0 Å (–1.0/+1.5), and 4.0 Å (–2.0/+2.0), respectively.

Atoms participating in hydrogen bonds in the four tetrad planes were restrained with distances corresponding to ideal hydrogen bond geometry (Saenger, 1984).

The glycosidic torsion angle (χ) was restrained to *syn* (65±25°) for residues G1 and G8 (on both strands), with a force constant of 50.0 kcal mol^{-1} rad^{-2}. Glycosidic bonds for all other residues in the tetrad planes were restrained to the *anti* conformation (220(±30)°). No torsion restraints were imposed on the loop thymine residues.

The two strands of DNA are identical in solution, as evidenced by a single set of resonances in the NMR spectra of the d(G-C-G-G-T_3-G-C-G-G) quadruplex. Hence, non-crystallographic symmetry restraints were imposed on all non-hydrogen atoms, with a force constant of 300.0 kcal mol^{-1} $Å^{-2}$, to maintain the two strands superimposable.

Distance geometry and simulated annealing regularization

Metric matrix distance geometry calculations were performed, with the program X-PLOR (Brünger, 1992) to embed and optimize initial structures. The experimentally obtained distance restraints were duplicated for the two strands of DNA, during the embedding stage of the protocol. For optimization with simulated annealing (DGSA protocol) all distance restraints were specified with the SUM averaging option in X-PLOR (Brünger, 1992; Nilges, 1995) which allows for ambiguity in the atom specification in input distance restraints. This feature is particularly useful since about 50% of the NOE cross peaks arise from a combination of intra-strand and inter-strand proton pairs, and the inherent symmetry in the system prevents strand specific assignments of the cross peaks. Out of 100 attempts at embedding and optimization, six best structures were selected based on the criterion of acceptable covalent geometry, low distance restraint violations, and favorable non-bonded energy. These six structures were further optimized with restrained molecular dynamics simulations.

Distance restrained molecular dynamics

Restrained molecular dynamics was carried out in vacuum with distance dependent dielectric constant. The dynamics was initiated at 5 K and the temperature was gradually increased to 1000 K in 5.0 ps and then equilibrated for 1.0 ps. The force constants for the distance restraints were kept at 2.0 kcal mol^{-1} $Å^{-2}$ during these stages. Subsequently, the force constants for the distance restraints were scaled up to a final value of 30 kcal mol^{-1} $Å^{-2}$ over 6.0 ps. The system was then allowed to evolve for 20.0 ps at 1000 K and then slow cooled to 300 K in 14.0 ps and equilibrated for 10.0 ps. The coordinates saved every 0.5 ps in the last 4.0 ps were averaged. The resulting structure was subject to conjugate gradient minimization till a final gradient of 0.1 kcal mol^{-1} $Å^{-1}$ was reached. All dynamics was carried out with a time step of 1.0 fs, and the force constant on the hydrogen bond related distance restraints was maintained at 60.0 kcal mol^{-1} $Å^{-2}$ throughout. All experimentally observed distance restraints were specified ambiguously with SUM averaging. Non-crystallographic symmetry restraints were maintained throughout.

Relaxation matrix intensity refinement

Four of the six distance refined structures were chosen arbitrarily for further optimization with relaxation matrix based direct NOE intensity refinement in vacuum, using X-PLOR. NOE volumes from 267 cross peaks for each of five mixing times were used as restraints, with uniform upper and lower bounds of $\pm$20%. The volume restraints were specified ambiguously with multiple atom selection. During refinement, X-PLOR restrains the sum of the corresponding calculated intensities to the observed ones. Non-crystallographic symmetry restraints were maintained throughout. The $R_{1/6}$ term (Nilges *et al*., 1991) was minimized during the refinement.

Dynamics was initiated at 5 K with distance restraints as at the end of the distance restrained molecular dynamics simulation. After heating the system to 300 K in 0.6 ps, the NOE intensity restraints were introduced by gradually scaling the force constant over 1.0 ps to a final value of 300.0. Simultaneously, the force constant on the NOE derived distance restraints on the non-exchangeable protons were scaled down to zero. The system was equilibrated at 300 K for 3.0 ps and the resulting structure was subjected to conjugate gradient minimization till a final gradient of 0.1 kcal mol^{-1} $Å^{-1}$ was reached. The distance restraints on the exchangeable protons, and on the atoms involved in hydrogen bonding in the tetrad planes were maintained throughout.

During dynamics, a cutoff of 6.0 Å was employed for computing relaxation pathways, and an isotropic correlation time of 4.0 ns was assumed.

Coordinates deposition

The coordinates for the four relaxation matrix refined structures of the d(G-C-G-G-T_3-G-C-G-G) quadruplex have been submitted to the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY, 11973, USA accession number: 228D.

Acknowledgements

The research was supported under NIH grant GM34504 to D.J.P. R.A.K. is a postdoctoral fellow of the Miriam and Benedict Wolf Cancer Research Fund.

References

Aboud-ela, F., Murchie, A. I. H., Norman, D. G. & Lilley, D. M. J. (1994). Solution structure of a parallel-stranded tetraplex formed by d(TG_4T) in the presence of sodium ions by NMR spectroscopy. *J. Mol. Biol.* **243**, 458–471.

Bax, A. & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.

Bax, A., Griffey, R. H. & Hawkins, B. L. (1983). Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J. Magn. Reson.* **55**, 301–315.

Brünger, A. T. (1992). X-PLOR User Manual, version 3.1, Yale University, New Haven.

Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G., Jr & Nelson, D. L. (1992). Triplet repeat mutations in human disease. *Science*, **256**, 784–789.

Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M. & Gupta, G. (1995). Hairpins are formed by the single strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl Acad. Sci. USA*, **92**, 5199–5203.

Cheong, C. & Moore, P. B. (1992). Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry*, **31**, 8406–8414.

Dombroski, D. F., Scraba, D. G, Bradley, R. D. & Morgan, A. R. (1988). Tetrastranded DNA as an intermediate in recombination. *Mol. Cell. Biol. (Life Sci. Adv.)* **7**, 153–158.

Fry, M. & Loeb, L. A. (1994). The fragile X-syndrome d$(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.

Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M. Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., Warren, S. T., Oostra, B. A., Nelson, D. L. & Caskey, C. T. (1991). Variation of the CGG repeat at the Fragile X site results in genetic instability:

resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.

Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*, **81**, 533–540.

Gao, X., Huang, X., Smith, G. K., Zheng, M. & Lin, H. (1995). A new anti-parallel duplex motif of DNA CCG repeats that is stabilized by extrahelical bases symmetrically located in the minor groove. *J. Am. Chem. Soc*. **117**, 8883–8884.

Gellert, M., Lipsett, M. N. & Davies, D. R. (1962). Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.

Ghose, R., Marino, J. P., Wiberg, K. B. & Prestegard, J. H. (1994). Dependence of ^{13}C chemical shifts on glycosidic torsion angles in ribonucleic acids. *J. Am. Chem. Soc*. **116**, 8827–8828.

Greene, K. L., Wang, Y. & Live, D. (1995). Influence of the glycosidic torsion angle on ^{13}C and ^{15}N shifts in guanosine nucleotides: Investigations of G-tetrad models with alternating *syn* and *anti* bases. *J. Biomol. NMR*, **5**, 333–338.

Gupta, G., Garcia, A. E., Guo, Q., Lu, M. & Kallenbach, N. R. (1993). Structure of a parallel-stranded tetramer of the *Oxytricha* telomeric DNA sequence of dT_4G_4. *Biochemistry*, **32**, 7098–7103.

Guschlbauer, W., Chantof, J. F. & Thiele, D. (1990). Four stranded nucleic acid structures 25 years later: from guanosine gels to telomere DNA. *J. Biomol. Struct. Dynym*. **8**, 491–511.

Hansen, R. S., Canfield, T. K., Lamb, M. M., Gartler, S. M. & Laird, C. D. (1993). Association of Fragile X syndrome with delayed replication of the FMR-1 gene. *Cell*, **73**, 1403–1409.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr & Blackburn, E. H. (1987). Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899–908.

Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992). Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature*, **356**, 126–131.

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richards, R. I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$. *Science*, **252**, 1711–1714.

Laughlan, G., Murchie, A. I., Norman, D. G., Moore, M. H., Moody, P. C., Lilley, D. M. & Luisi, B. (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, **265**, 520–524.

Leonard, G. A., Zhang, S., Peterson, M. R., Harrop, S. J., Helliwell, J. R., Cruse, W. B., d'Estaintot, B., Kennard, O., Brown, T. & Hunter, W. N. (1995). Self association of a DNA loop creates a quadruplex: crystal structure of d(GCATGCT) at 1.8 A resolution. *Structure*, **3**, 335–340.

Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984). Long range hydrogen mediated effects in peptides: ^{15}N NMR study of gramicidin S in water and organic solvents. *J. Am. Chem. Soc.* **106**, 1939–1941.

Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. & Feigon, J. (1993). Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl Acad. Sci. USA*, **90**, 3745–3749.

Marion, D. & Wuthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H-^1H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun*. **113**, 967–974.

McGavin, S. (1971). Models of specifically paired (homologous) nucleic acid structures. *J. Mol. Biol.* **55**, 293–298.

Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J. & Haworth, I. S. (1995). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$. *Nucl. Acids Res.* **23**, 1050–1059.

Nelson, D. L. (1995). The fragile X syndromes. *Sem. Cell Biol.* **6**, 5–12.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.

Nilges, M., Habazettl, J., Brünger, A. T. & Holak, T. A. (1991). Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* **219**, 499–510.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F. & Mandel, J. L. (1991). Instability of a 550 base pair DNA segment and abnormal methylation in Fragile X syndrome. *Science*, **252**, 1097–1102.

O'Brien, E. J. (1967). Crystal structures of two complexes containing guanine and cytosine derivatives. *Acta. Crystallog.* **23**, 92–106.

Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinski, A. (1993). The structure of α-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268**, 17651–17654.

Patel, D. J., Kozlowski, S. A., Nordheim, A. & Rich, A. (1982). Right-handed and left-handed DNA studies of B-DNA and Z-DNA by using proton nuclear Overhauser effect and phosphorus NMR. *Proc. Natl Acad. Sci. USA*, **79**, 1413–1417.

Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T. & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in Fragile X syndrome. *Cell*, **66**, 817–822.

Pilch, D. S., Plum, G. E. & Breslauer, K. J. (1995). The thermodynamics of DNA structures which contain lesions or guanine tetrads. *Curr. Opin. Struct. Biol.*, **5**, 334–342.

Plateau, P. & Gueron, M. (1982). Exchangeable proton NMR without baseline distortion using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310–7311.

Rhodes, D. & Giraldo, R. (1995). Telomere structure and function. *Curr. Opin. Struct. Biology*, in the press.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer-Verlag Inc., New York.

Scaria, P. V., Shire, S. J. & Shafer, R. H. (1992). Quadruplex structure of $d(G_3T_4G_3)$ stabilized by K^+ and Na^+ is an asymmetric hairpin dimer. *Proc. Natl Acad. Sci. USA*, **89**, 10336–10340.

Schultze, P., Smith, F. W. & Feigon, J. (1994a). Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure*, **2**, 221–233.

Schultze, P., Macaya, R. F. & Feigon, J. (1994b). Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTG-GTTGG). *J. Mol. Biol.* **235**, 1532–1547.

Sen, D. & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications in meiosis. *Nature*, **334**, 364–366.

Sen, D. & Gilbert, W. (1991). The structure of telomeric

DNA: DNA quadruplex formation. *Curr. Opin. Struct. Biol.* **1**, 435–438.

Shaka, A. J., Barker, P. B. & Freeman, R. H. (1985). Computer optimized decoupling sequence for wide-band applications and low-level operation. *J. Magn. Reson.* **64**, 547–552.

Sinden, R. R. & Wells, R. D. (1992). DNA structure, mutations and human genetic disease. *Curr. Opin. Biotechnol.* **3**, 612–622.

Sklenar, V. & Bax, A. (1987). Measurements of ^{1}H-^{31}P NMR coupling constants in double-stranded DNA fragments. *J. Am. Chem. Soc.* **109**, 7525–7526.

Smith, F. W. & Feigon, J. (1992). Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.

Smith, F. W., Lau, F. W. & Feigon, J. (1994a). d($G_3T_4G_3$) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-*syn-syn-anti* and 5'-*syn-anti-anti* N-glycosidic conformations. *Proc. Natl Acad. Sci. USA*, **91**, 10546–10550.

Smith, S. S., Laayoun, A., Lingeman, R. G., Baker, D. J. & Riley, J. (1994b). Hypermethylation of telomere-like foldbacks at codon 12 of the human c-Ha-ras gene and the trinucleotide repeat of the FMR-1 gene of Fragile X. *J. Mol. Biol.* **243**, 143–151.

States, D., Haberkorn, R. A. & Ruben, D. J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* **48**, 286–292.

Sundquist, W. I. (1991). The structures of telomeric DNA. In *Nuclic Acids and Molecular Biology* (Eckstein, F. & Lilley, D. M. J., eds), vol. 5, pp. 1–24. Springer-Verlag, Berlin.

Sundquist, W. I. & Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.

Sutherland, G. R. & Richards, R. I. (1995). Single tandem DNA repeats and human genetic disease. *Proc. Natl Acad. Sci. USA*, **92**, 3636–3641.

van de Ven, F. J. & Hilbers, C. W. (1988). Nucleic acids and nuclear magnetic resonance. *Eur. J. Biochem.* **7**, 1–38.

Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommen, G. B., Blonden, L. A., Riggins, G. J., Chastian, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A. & Warren, S. T. (1991). Identification of a gene (FMR-1) containing a C-G-G repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X syndrome. *Cell*, **65**, 905–914.

Wang, K. Y., Krawczyk, S. H., Bischofberger, N., Swaminathan, S. & Bolton, P. H. (1993). The tertiary structure of a DNA aptamer which binds to and inhibits thrombin determines activity. *Biochemistry*, **32**, 11285–11292.

Wang, Y. & Patel, D. J. (1992). Guanine residues in d(T_2AG_3) and d(T_2G_4) form parallel-stranded potassium cation stablized G-quadruplexes with *anti* glycosidic torsion angles in solution. *Biochemistry*, **31**, 8112–8119.

Wang, Y. & Patel, D. J. (1993a). Solution structure of a parallel-stranded G-quadruplex formed by d(TTGGGGT). *J. Mol. Biol.* **234**, 1171–1183.

Wang, Y. & Patel, D. J. (1993b). Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure*, **1**, 263–282.

Wang, Y. & Patel, D.J. (1994). Solution structure of the *Tetrahymena* telomeric repeat d(T_2G_4)$_4$ G-tetraplex. *Structure*, **2**, 1141–1156.

Wang, Y. & Patel, D. J. (1995). Solution structure of the *Oxytricha* telomeric repeat d[$G_4(T_4G_4)_3$] G-tetraplex. *J. Mol. Biol.* **251**, 76–94.

Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. & Patel, D. J. (1991). Multinuclear magnetic resonance studies of the Na^+ cation-stabilized complex formed by d(G-G-T-T-T-T-C-G-G) in solution: implication for G-tetrad structures. *J. Mol. Biol.* **222**, 819–832.

Williamson, J. R., Raghuraman, M. K. & Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.

Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R. & Richards, R. I. (1991). Fragile X genotype characterized by an unstable region of DNA. *Science*, **252**, 1179–1181.

Edited by P. E. Wright

(Received 9 August 1995; accepted 14 September 1995)

Supplementary material comprising one figure and one table is available from JMB Online.