doi:10.1006/jmbi.2000.3524 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 297, 627–644

JMB

A Dimeric DNA Interface Stabilized by Stacked A·(G·G·G·G)·A Hexads and Coordinated Monovalent Cations

Abdelali Kettani[1], Andrey Gorin[1], Ananya Majumdar[1], Thomas Hermann[1], Eugene Skripkin[1], Hong Zhao[2], Roger Jones[2] and Dinshaw J. Patel[1]*

[1]*Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center New York, NY, 10021, USA*

[2]*Department of Chemistry Rutgers, the State University of New Jersey, Piscataway, NJ 08855, USA*

**Corresponding author*

We report on the identification of an A·(G·G·G·G)·A hexad pairing alignment which involves recognition of the exposed minor groove of opposing guanines within a G·G·G·G tetrad through sheared G·A mismatch formation. This unexpected hexad pairing alignment was identified for the d(G-G-A-G-G-A-G) sequence in 150 mM Na^+ (or K^+) cation solution where four symmetry-related strands align into a novel dimeric motif. Each symmetric half of the dimeric ''hexad'' motif is composed of two strands and contains a stacked array of an A·(G·G·G·G)·A hexad, a G·G·G·G tetrad, and an A·A mismatch. Each strand in the hexad motif contains two successive turns, that together define an S-shaped double chain reversal fold, which connects the two G-G steps aligned parallel to each other along adjacent edges of the quadruplex. Our studies also establish a novel structural transition for the d(G-G-A-G-G-A-N) sequence, N = T and G, from an ''arrowhead'' motif stabilized through cross-strand stacking and mismatch formation in 10 mM Na^+ solution (reported previously), to a dimeric hexad motif stabilized by extensive inter-subunit stacking of symmetry-related A·(G·G·G·G)·A hexads in 150 mM Na^+ solution. Potential monovalent cation binding sites within the arrowhead and hexad motifs have been probed by a combination of Brownian dynamics and unconstrained molecular dynamics calculations. We could not identify stable monovalent cation-binding sites in the low salt arrowhead motif. By contrast, five electronegative pockets were identified in the moderate salt dimeric hexad motif. Three of these are involved in cation binding sites sandwiched between G·G·G·G tetrad planes and two others, are involved in water-mediated cation binding sites spanning the unoccupied grooves associated with the adjacent stacked A·(G·G·G·G)·A hexads. Our demonstration of A·(G·G·G·G)·A hexad formation opens opportunities for the design of adenine-rich G-quadruplex-interacting oligomers that could potentially target base edges of stacked G·G·G·G tetrads. Such an approach could complement current efforts to design groove-binding and intercalating ligands that target G-quadruplexes in attempts designed to block the activity of the enzyme telomerase.

Keywords: A·(G·G·G·G)·A hexad formation; G·(A-G) triads; G-quadruplex recognition; sheared G·A mismatch formation; uniform ^{13}C,^{15}N-labeled DNA

Introduction

The polymorphic nature of DNA has been known for quite some time (reviewed by Neidle, 1994), as is its ability to form multi-stranded architectures (reviewed by Patel *et al.*, 1999). The role of

Abbreviations used: HPLC, high pressure liquid chromatography; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; COSY, correlated spectroscopy; BD, Brownian dynamics; MD, molecular dynamics; MMLV, murine mammary leukemia virus; TOCSY, total correlated spectroscopy.

E-mail address of the corresponding author: pateld@mskcc.org

0022-2836/00/030627–18 $35.00/0

cations in inducing structural transitions remains an area of active interest ever since the demonstration of a cation dependent transition from right-handed *B*-DNA to left-handed *Z*-DNA (Wang *et al*., 1979). More recently, a cation-dependent conformational transition in slow exchange on the NMR time-scale has been reported for G·G·G·G and G·C·G·C tetrad-containing quadruplexes on proceeding from Na^+ (Kettani *et al*., 1998) to K^+ (Bouaziz *et al*., 1998) containing solution. We now report on a novel structural transition for the d(G-G-A-G-G-A-N) sequence, N = G and T, on proceeding from 10 mM Na^+ solution to 150 mM Na^+ solution. This sequence contains a pair of tandem all purine G-G-A steps, a triplet repeat (Mishima *et al*., 1997) identified in micro satellite DNA belonging to the rat polymorphic immunoglobulin receptor gene (Koch *et al*., 1995; Aoki *et al*., 1997), and in portions of human and mouse cellular DNA which cross-hybridizes with the internal direct repeat IR3 repetitive region of Epstein-Barr virus (Heller *et al*., 1985).

Our previous report on the solution structure of d(G-G-A-G-G-A-T) in low salt (10 mM Na^+) solution established formation of a two-stranded arrowhead motif structure aligned solely through mismatch formation (Kettani *et al*., 1999). We now report on the solution structure of d(G-G-A-G-G-A-G) in 150 mM Na^+ solution and establish formation of a novel dimeric four-stranded architecture (shown schematically in Figure 1(a)) stabilized by a pair of stacked A·(G·G·G·G)·A hexads (Figure 1(b)) at the dimer interface. The recognition of G·G·G·G tetrads using sheared G·A mismatches to form A·(G·G·G·G)·A hexads could represent a novel approach for targeting G-quadruplexes. The biological relevance of this approach is based on the premise that telomerase activity can be blocked by ligand-induced stabilization of G-quadruplex structures in telomeric DNA.

Results

NMR spectra

We have recorded nuclear magnetic resonance (NMR) spectra of d(G-G-A-G-G-A-N), N = T or G, as a function of Na^+ concentration. We observe a single conformation for the N = T analog in 10 mM NaCl, H_2O solution as reflected in the exchangeable proton NMR spectrum recorded at 0 °C in Figure 2(a). The same conformation also predominates for the N = G analog, but it is in equilibrium with a minor conformer which predominates in moderate salt solution. The low salt conformer of d(G-G-A-G-G-A-N) has been shown by our group to be a two-stranded ''arrowhead'' duplex motif aligned solely through mismatch formation (Kettani *et al*., 1999). We observe a totally different NMR spectrum associated with a single conformation for the N = G analog in 150 mM NaCl, H_2O solution as reflected in the exchange-

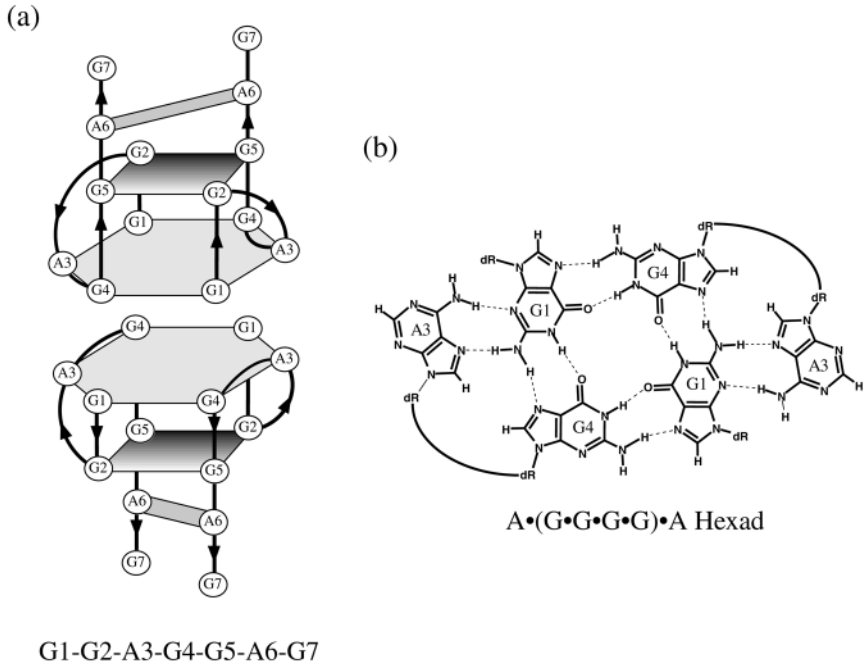

Figure 1. (a) Schematic diagram showing the dimeric hexad motif for the multi-stranded d(G-G-A-G-G-A-G) architecture in 150 mM Na^+ containing solution. The backbone tracing of individual strands is shown in bold lines with the chain directionality represented by bold arrows. The broken bar represents the A6·A6 mismatch, the shaded rectangle represents the G2·G5·G2·G5 tetrad and the dashed hexagon represents the A3·(G1·G4·G1·G4)·A3 hexad. (b) Schematic of the planar A3·(G1·G4·G1·G4)·A3 hexad observed in the d(G-G-A-G-G-A-G) quadruplex in 150 mM Na^+ containing solution.

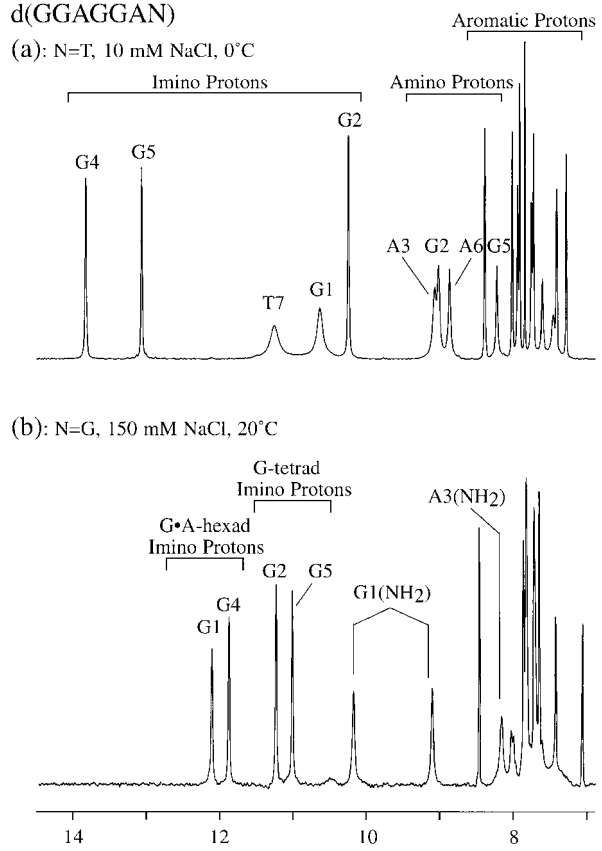

Figure 2. Proton NMR spectra (6.9 to 14.5 ppm) of the d(G-G-A-G-G-A-N) sequence as a function of NaCl. (a) N = T in 10 mM NaCl, 2 mM phosphate (pH 6.6) at 0 °C. This structure corresponds to a two-stranded interlocked mismatch-aligned ''arrowhead'' DNA motif (Kettani *et al.*, 1999). (b) N = G in 150 mM NaCl, 2 mM phosphate, pH 6.6 at 20 °C. This structure corresponds to a dimeric four-stranded hexad motif, stabilized through formation of stacked A·(G·G·G·G)·A hexads. Imino and amino proton assignments are listed over the spectra.

able proton NMR spectrum recorded at 20 °C (Figure 2(b)). The same conformation, under structural characterization in this study, also predominates for the N = T analog, but it is in equilibrium with a minor conformer that predominates in low salt solution. These low (10 mM) and moderate (150 mM) Na^+ cation conformers of d(G-G-A-G-G-A-N), N = T or G, are in slow exchange on the NMR time-scale at intermediate salt concentrations. The same monovalent cation concentration-dependent structural transition for d(G-G-A-G-G-A-N), N = T or G is also observed in K^+ solution.

The NMR spectrum (5.5 to 12.5 ppm) of d(G-G-A-G-G-A-G) sequence in 150 mM NaCl, H_2O solution at 20 °C contains four imino protons between 10.9 and 12.2 ppm and amino proton resonances at 10.2 and 9.15 ppm (Figure 2(b)). These narrow and unusually downfield shifted amino proton chemical shifts are suggestive of a unique pairing alignment associated with a novel architecture and hence we have undertaken systematic NMR studies on unlabeled, site-specific ^{13}C,^{15}N-labeled and uniform ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) sequence.

Strand stoichiometry

We have approached the stoichiometry of the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl using a gel filtration method. This approach is based on the size dependent retention time of DNA oligomers on a high pressure liquid chromatography column (HPLC) which has been calibrated by recording the log of the known number of residues of several nucleic acid oligomers under non-denaturing conditions as a function of eluted volume (Leroy *et al.*, 1993). We used an analytical Synchropak GPC peptide column with the elution buffer being 150 mM NaCl, 10 mM phosphate at ambient temperature. The calibration plot based on a range of oligomer samples of varying length (filled circles) is shown in Figure 3(a). The elution volume for the native d(G-G-A-G-G-A-G) structure was determined by rapidly applying to the column an aliquot of the NMR sample that was diluted in chromatographic buffer. The observed elution volume corresponds approximately to a 28-mer (Figure 3(a)) consistent with the d(G-G-A-G-G-A-G) 7-mer sequence forming a four-stranded structure in 150 mM NaCl solution. The control complementary d(G-G-A-G-G-A-G)·d(C-T-C-C-T-C-C) duplex, migrated, as expected, as an approximately 14-mer (Figure 3(a)) while the d(G-G-A-G-G-A-G) sequence prepared under single-strand conditions (see the legend to Figure 3(a)) migrated as an approximately 7-mer.

We have also attempted to measure strand stoichiometry of the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl by undertaking a systematic study of the concentration dependence of this folded form relative to the unstructured single-stranded form detectable in slow equilibrium at lower concentrations. A log-log plot of the multimer versus single strand concentrations for d(G-G-A-G-G-A-G) samples equilibrated for 12 weeks at ambient temperature is shown in Figure 3(b) and yields a strand stoichiometry for the folded form of 3.80(±0.3). These data independently favor a four-stranded folded form of the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl solution.

Exchangeable proton resonance assignments

The unambiguous assignment of a subset of exchangeable imino and amino proton resonances and non-exchangeable purine H-8 and H-2 proton resonances was achieved through specific incorporation of ^{15}N-labeled and ^{15}N, ^{13}C-labeled nucleoside (*G, N^1, N^2, N^7-^{15}N-labeled guanine; #G, N^1, N^2, N^7-^{15}N, C^2-^{13}C-labeled guanine; and *A, N^1, N^6, N^7-^{15}N-labeled adenine) phosphonates within the d(G1-G2-A3-G4-G5-A6-G7) sequence. Labels

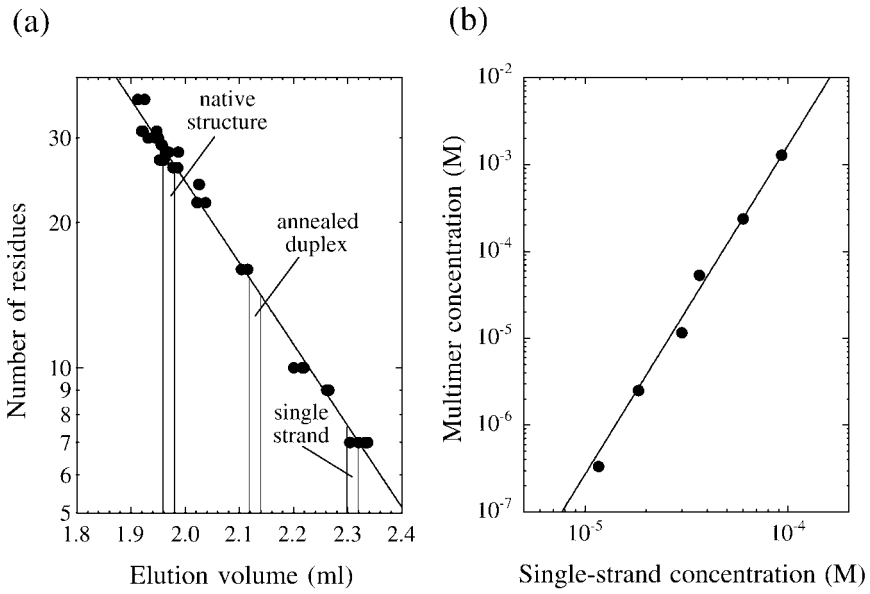

Figure 3. (a) A plot of the log of the known number of residues of several nucleic acid species under non-denaturing conditions as a function of eluted volume on an analytical Synchropak GPC peptide column with the elution buffer being 150 mM NaCl, 10 mM phosphate at ambient temperature. The calibration samples (filled circles) included DNA oligomer strands of varying length, as well as several guanine rich sequences that form G-quadruplexes either through tetramerization of four strands or through dimerization of hairpins. The straight line represents an exponential fit to this data. The elution volume for the single-stranded form corresponds to an aliquot of the NMR sample that was diluted in chromatographic buffer, heated to 100 °C for five minutes and then cooled on ice, prior to application to the column. The elution volume for the annealed duplex corresponds to a mixture of the d(G-G-A-G-G-A-G) sequence and its Watson-Crick complementary partner in chromatographic buffer. The elution volume for the native structure corresponds to an aliquot of the NMR sample that was diluted in chromatographic buffer and rapidly loaded onto the column. (b) A plot of the concentration of the d(G-G-A-G-G-A-G) multimer *versus* the concentration of the d(G-G-A-G-G-A-G) single strand as monitored by the average of the three most resolved peaks (A3(H-2), A6(H-2) and G2(H-8) protons) in the multimeric state and the two most resolved peaks in the single-stranded state. The samples were equilibrated for 12 weeks at ambient temperature prior to recording the spectra at 25 °C. The slope of the curve is 3.80(±0.3) based on the best fit to the data.

$^{\#}$G4 and *A6 were incorporated into the sequence in one set, while labels *G2 and $^{\#}$G5 were incorporated in a second set. The proton spectrum (5.5 to 12.5 ppm) of the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl, H_2O solution at 0 °C is plotted in Figure 4(a). Additional broad amino proton resonances between 9.0 to 11.0 ppm are observed in this spectrum recorded at 0 °C (Figure 4(a)) compared to its counterpart recorded at 20 °C (Figure 2(b)). An example of an one dimensional difference spectrum with site specifically $^{\#}$G4 and *A6-labeled d(G-G-A-G-G-A-G) recorded at 0 °C is presented in Figure 4(b). These labeled analogs result in the splitting of the imino proton resonance (one bond coupling to N^1), the amino proton resonances (one bond coupling to N^2 for guanine bases and N^6 for adenine bases), as well as the base H-8 (two bond coupling to N^7) and H-2 (two bond coupling to N^1 for adenine bases) resonances. The difference spectrum in Figure 4(b) permits the assignment of the imino, amino and H-8 protons of G4 and the H-8 and H-2 protons of A6.

The assignment of the imino and amino protons in the d(G-G-A-G-G-A-G) sequence was completed following analysis of the expanded nuclear Overhauser enhancement spectroscopy (NOESY) spectrum recorded in 150 mM NaCl, H_2O at 0 °C (Figure 4(c)). The key nuclear Overhauser enhancement (NOE) cross-peaks are labeled and their assignments listed in the caption to Figure 4(c). The chemical shift differences ($\Delta\delta$) between the two amino protons of G2 ($\Delta\delta = 2.6$ ppm), A3 ($\Delta\delta = 1.45$ ppm), G4 ($\Delta\delta = 5.0$ ppm) and G5 ($\Delta\delta = 3.5$ ppm) are large (see Table S1, Supplementary Material), indicative of one hydrogen-bonded and one exposed amino proton for each of these residues in the d(G-G-A-G-G-A-G) architecture. By contrast, both amino protons of G1 (10.2 and 9.15 ppm) resonate to low field, indicative of both participating in hydrogen bond formation.

Non-exchangeable proton resonance assignments

The assigned imino protons could be next correlated to non-exchangeable H8 protons within individual guanine rings following recording of an HCCNH-TOCSY experiment (Fiala *et al.*, 1996; Simorre *et al.*, 1996; Sklenar *et al.*, 1996) on the uniform ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) sequence recorded in 150 mM NaCl, H_2O at 10 °C. Such

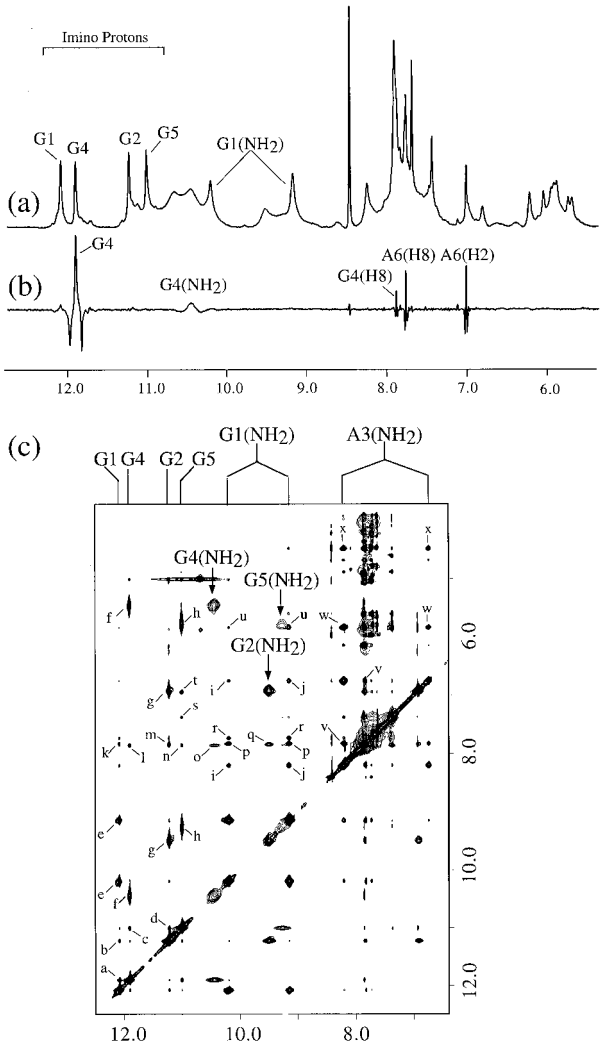

Figure 4. (a) Proton NMR spectrum (5.5 to 12.5 ppm) of the d(G-G-A-G-G-A-G) sequence in H_2O buffer (150 mM NaCl, 2 mM phosphate (pH 6.6)) at 0 °C. The imino and amino protons are assigned over the spectrum. We observe both narrow (designated by asterisks) and broad amino protons between 9.0 and 10.8 ppm in this spectrum recorded at 0 °C. (b) Proton difference spectrum recorded at 0 °C of ^{15}N-decoupled and ^{15}N-coupled spectra of the d(G-G-A-G-G-A-G) sequence ^{15}N-labeled at the N^1, N^2 and N^7 positions and C^{13}-labeled at the C^2 position of G4 and ^{15}N-labeled at the N^1, N^6 and N^7 positions of A6. The ^{15}N-labeling approach readily identifies the imino proton of G4 (one bond coupling to N^1), the amino protons of G4 (one bond coupling to N^2), the H8 protons of G4 and A6 (two bond coupling to N^7) and the H2 proton of A6 (two bond coupling to N^1). (c) An expanded NOESY (200 ms mixing time) contour plot correlating NOEs between imino, amino and non-exchangeable protons in the d(G-G-A-G-G-A-G) sequence (5 mM in single strand) in 150 mM NaCl, 2 mM phosphate, H_2O, pH 6.6 at 20 °C. The cross peaks a to x are assigned as follows: a, G1(NH1)-G4(NH1); b, G1(NH1)-G2(NH1); c, G4(NH1)-G5(NH1); d, G2(NH1)-G5(NH1); e and e′, G1(NH1)-G1(NH_2); f and f′, G4(NH1)-G4(NH_2); g and g′, G2(NH1)-G2(NH_2); h and h′, G5(NH1)-G5(NH_2); i, i′, j and j′, G1(NH_2)-A3(NH_2); k, G1(NH1)-G4(H8); l, G4(NH1)-G1(H8); m, G2(NH1)-G5(H8); n, G5(NH1)-G1(H8); o, G4(NH_2)-G1(H8); p and p′, G1(NH_2)-G4(H8); q, G2(NH_2)-G5(H8); r and r′, G1(NH_2)-A3(H8); s, G5(NH1)-G2(H8); t, G5(NH1)-A6(H2); u and u′, G1(NH1)-G1(H1′); v and v′, A3(NH_2)-G4(H8); w and w′, A3(NH_2)-G1(H1′); x and x′, A3(NH_2)-G1(H4′).

imino proton to H8 proton correlations for residues G1, G2, G4 and G5 are shown in Figure 5(a).

Next, the sequential NOE connectivities between the non-exchangeable base protons and their own and 5′-linked sugar H1′ protons can be traced without interruption in the expanded nuclear Overhauser enhancement spectroscopy (NOESY) contour plot of the d(G-G-A-G-G-A-G) sequence recorded in 150 mM NaCl, ^2H_2O at 20 °C (Figure 5(b)). We do not observe any strong purine H8 to its own sugar H1′ NOE values at short (50 ms) mixing times, ruling out *syn* glycosidic torsion angles (Patel *et al.*, 1982) at any of the adenine or guanine bases within the d(G-G-A-G-G-A-G) structural fold. All the base and sugar protons have been assigned (Table S1, Supplementary Material), with the most shifted resonance associated with the sugar H4′ proton of G2, whose chemical shift of 2.83 ppm is dramatically upfield of the 4.3(±0.4) ppm chemical shift for unperturbed sugar H4′ protons.

G·G·G·G tetrad formation

The imino protons of G1 (12.1 ppm), G2 (11.2 ppm), G4 (11.9 ppm) and G5 (11.0 ppm) resonate between 11.0 and 12.2 ppm in a region characteristic of G·G·G·G tetrad formation (Wang *et al.*, 1991, Smith & Feigon, 1992). The observed NOE values between the imino proton of G2 and the H8 proton of G5 (peak m, Figure 4(c)) and between the imino proton of G5 and the H8 of G2 (peak s, Figure 4(c)) are consistent with formation of a G2·G5·G2·G5 tetrad. The observed NOE values between the imino proton of G1 and the H8 proton of G4 (peak k, Figure 4(c)) and between the imino proton of G4 and the H8 of G1 (peak l, Figure 4(c)) are consistent with formation of a G1·G4·G1·G4 tetrad. These results support formation of an architecture stabilized by G2·G5·G2·G5 and G1·G4·G1·G4 tetrads for d(G1-G2-A3-G4-G5-A6-G7) in 150 mM NaCl solution.

A·(G·G·G·G)·A hexad formation

In addition, the observed NOE values between the amino proton of G1 and the H8 proton of A3 (peaks r and r′, Figure 4(c)) and between the amino protons of A3 and the sugar H1′ (peaks w and w′, Figure 4(c)) and H4′ (peaks x and x′, Figure 4(c)) protons of G1 are consistent with sheared G1·A3 mismatch formation (Li *et al.*, 1991; Heus & Pardi, 1991; Chou *et al.*, 1997). Since we have already established formation of a G1·G4·G1·G4 tetrad,

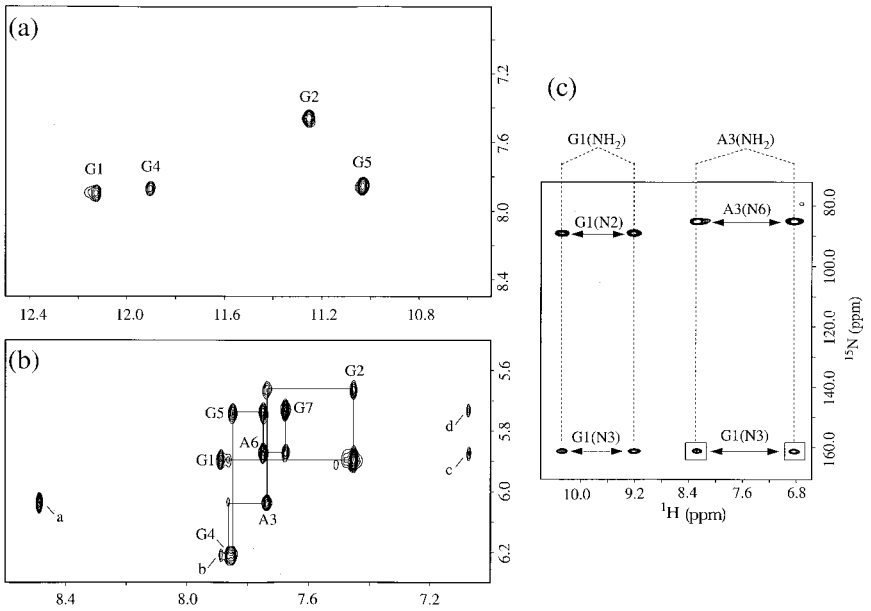

Figure 5. (a) Contour plot showing intranucleotide NH1-H8 cross-peaks of G1, G2, G4 and G5 recorded using a modified version of the HCCNH-TOCSY pulse sequence (Fiala *et al.*, 1996) on uniformly ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) sequence (2 mM in single strand) in 150 mM NaCl, 2 mM phosphate, H_2O (pH 6.6) at 10 °C. (b) An expanded NOESY (300 ms mixing time) contour plot correlating the base (7.0 to 8.6) and sugar (5.5 to 6.3) protons in the d(G-G-A-G-G-A-G) sequence (5 mM in single strand) in 150 mM NaCl, 2 mM phosphate, ^2H_2O, at 20 °C. The lines trace the NOE connectivities between the base protons and their own and 5′-flanking sugar H1′ protons from G1 to G7 in the sequence. The cross peak a to d are assigned as follows: a, A3(H2)-A3(H1′); b, G1(H8)-G4(H1′); c, A6(H2)-A6(H1′); d, A6(H2)-G7(H1′). (c) Expanded $^2J_{NN}$ HNN-COSY contour plot correlating two bond coupling connectivities between donor and acceptor nitrogen atoms within N-H···N hydrogen bond across the A3(N^6 donor)·G1(N^3 acceptor) mismatch pair (boxed cross-peaks) in uniformly ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) sequence (2 mM in single strand) in 150 mM NaCl, 2 mM phosphate, H_2O (pH 6.6) at 10 °C.

these observed NOE patterns imply formation of an A3·(G1·G4·G1·G4)·A3 hexad (Figure 1(b)).

A·A mismatch formation

An analysis of the NOE data on the d(G-G-A-G-G-A-G) fold in 150 mM NaCl solution were suggestive of pairings between the symmetry related adenine bases across a potential A6·A6 mismatch alignment. The available markers on adenine are the H2, H8 and N^6 amino protons and alternate A·A mispairing alignments should be distinguishable based on NOE values between the amino protons and non-exchangeable H2 or H8 protons. Unfortunately, we observe a broadened average amino proton resonance for A6, presumably due to intermediate rotation about the exocyclic C-N amino bond. To overcome this limitation, we have replaced A6 by N^6-methyl-A6 in the sequence, which can be incorporated without perturbation of the NMR markers characteristic of structure formation. The N-methylation of the exocyclic amino group freezes the rotation about the exocyclic C-N bond resulting in distinct proton and CH_3 resonance markers for monitoring potential pairing alignments. We observe NOE values between the H2 proton and the N^6CH_3 protons of the A6 residues across the A6·A6 mismatch pair in an expanded NOESY contour plot on this analog in 150 mM NaCl, ^2H_2O buffer (Figure S2a, Supplementary Material). We have also observed related NOE cross-peaks between the H2 and N^6H protons of N^6CH_3-A6 in NOESY spectra recorded in H_2O solution. These and other NOE values involving the resolved N^6CH_3 and N^6H protons of A6 are only consistent with a reversed A6·A6 mismatch orientation of adenines aligned through their Watson-Crick edges (Figure S2b, Supplementary Material).

Hydrogen-bonding alignments from $^2J_{NN}$ scalar couplings

Two bond $^2J_{NN}$ scalar couplings have been shown to correlate imino donor ^{15}N nucleus and acceptor ^{15}N nucleus across hydrogen bonds in Watson-Crick pairs (Dingley & Grzesiek, 1998; Pervushin *et al.*, 1998) and amino donor ^{15}N nucleus and acceptor ^{15}N nucleus across hydrogen bonds in mismatch pairs (Majumdar *et al.*, 1999a; Kettani *et al.*, 1999) in nucleic acids. We have applied this approach, based on HNN-COSY (Majumdar *et al.*, 1999a) and newly developed H(CN)N(H)-COSY (Majumdar *et al.*, 1999b) experiments, to define pairing alignments in the uniform ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) sequence in

150 mM NaCl solution. Some of the experimental data has been published recently in a methodological study reporting on the application of an H(CN)N(H)-COSY pulse sequence approach for observation of internucleotide N-H···N hydrogen bonds in the absence of directly detectable exchangeable protons (Majumdar *et al*., 1999b).

The directionality of the hydrogen bonding alignment around the G2·G5·G2·G5 tetrad was defined by the observation of a $^2J_{NN}$ scalar coupling between the N^2 donor amino protons of G2 and the N^7 acceptor of G5 (see Figure 3(a) of Majumdar *et al*., 1999b). The N^2 amino protons of G5 were broad and this prevented identification of the potential $^2J_{NN}$ scalar coupling between the N^2 donor amino protons of G5 and the N^7 acceptor of G2, even at low temperature (0 °C). This problem was overcome through use of the H(CN)N(H)-COSY pulse sequence approach for a hydrogen bonding system of the type $N_dH\cdots N_a$-CH. The donor N_d nitrogen (N^2 of G5 in this case) is correlated with the corresponding CH proton (H8 of G2 in this case) associated with the acceptor N_a nitrogen (N^7 of G2 in this case) (see Figure 3(b) of Majumdar *et al*., 1999b).

The directionality of the hydrogen bonding alignment around the G1·G4·G1·G4 tetrad was defined by the observation of a $^2J_{NN}$ scalar coupling between the N^2 donor amino protons of G1 and the N^7 acceptor of G4 (see Figure 3(a) of Majumdar *et al*., 1999b). The N2 amino protons of G4 were broad; however, we were able to correlate the N^2 nitrogen of G4 with the H8 proton of G1 using the H(CN)N(H)-COSY pulse sequence approach (see Figure 3(b) of Majumdar *et al*., 1999b).

The formation of the A3·(G1·G4·G1·G4)·A3 hexad (Figure 1(b)) was defined through identification of sheared G1·A3 mispairs. Thus, $^2J_{NN}$ scalar couplings were observed between the N^2 amino donor protons of G1 and the N^7 acceptor of A3 (see Figure 3(a) of Majumdar *et al*., 1999b) and between the N^6 amino donor of A3 and the N^3 acceptor of G1 (labeled boxed peaks, Figure 5(c)).

Alternate models

Model building suggests two alternate potential folding topologies for the architecture of d(G-G-A-G-G-A-G) in 150 mM NaCl solution. The stoichiometry measurements and NMR experimental data require formation of a four-stranded architecture containing symmetrically related A3·(G1·G4·G1·G4)·A3 hexads, G2·G5·G2·G5 tetrads and A6·A6 mismatches. These conditions could be satisfied by the alternate folding topologies shown in Figures 1(a) and S1 (Supplementary Material), which differ primarily in the nature of the linkage at the A3-G4 steps. A consequence of these alternate linkages is that the topology shown in Figure 1(a) involves a dimer of symmetry-related two-stranded motifs where A3·(G1·G4·G1·G4)·A3 hexad and G2·G5·G2·G5 tetrad alignments form within individual two-stranded motifs, while all four strands are required to form the same hexad and tetrad alignments for the folding topology shown in Figure S1.

We experimentally observe NOE values from the A3(H2) proton to the A3(H1′, H2′) protons (moderate intensity) and to the other A3 sugar protons (weak intensity) (Table S2, Supplementary Material). These base-sugar NOE values cannot be within a given A3 residue due to proton separations of > 5 Å for an anti glycosidic bond alignment and must reflect spatial interactions between A3 residues on different strands. This implies that A3 residues on adjacent A3·(G1·G4·G1·G4)·A3 hexads must be positioned over each other and one could rule out models that do not satisfy this criterion. Further, we observe moderate to weak NOE values between the sugar protons of G1 and G4 (Table S2, Supplementary Material) which, based on their separation in sequence, are likely to be between strands rather than within the same strand. This implies that the sugar rings of the G1 and G4 residues on adjacent A3·(G1·G4·G1·G4)·A3 hexads must also be positioned over each other and one could similarly rule out models that do not satisfy this additional criterion.

Distance restraints and molecular dynamics calculations

Distance restraints associated with exchangeable protons (total of 66) were qualitatively deduced from NOESY experiments in H_2O at two mixing times, while those associated with their non-exchangeable proton counterparts (total of 213) were quantified from NOE buildup curves in ^2H_2O at five mixing times as described in Materials and Methods. The observation of a single set of narrow resonances for the d(G-G-A-G-G-A-G) sequence at temperatures down to 0 °C was consistent with formation of a folded architecture containing four strands defined by a pair of 2-fold symmetry axis. Therefore, non-crystallographic symmetry restraints were used during the computations. All distance restraints were classified as ambiguous during the distance-restrained molecular dynamics computations since we have not distinguished between intra-strand and inter-strand restraints between proton pairs within the folded architecture. Experimentally defined hydrogen bonding alignments from $^2J_{NN}$ scalar couplings discussed above were used to restrain the mispairs, tetrads and hexads, with the folding models incorporating these hydrogen bonding alignments outlined in Figures 1(a) and S1 (Supplementary Material) considered independently during the computations.

The solution structure of the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl was solved by molecular dynamics computations guided by hydrogen bonding and NOE distance restraints. 60 starting structures were generated for the d(G-G-A-G-G-A) 6-mer segment as sets of four randomized chains

separated by space intervals of 50 Å. We omitted G7 from the computations since there were too few restraints to define this terminal residue. The protocol outlined in the Methods section involved initial torsion space dynamics at 20,000 K followed by cartesian space dynamics at 300 K. The structural statistics following distance refinement of structures incorporating hydrogen bonding alignments outlined in Figures 1(a) and S1 are listed in Table S3 (Supplementary Material). It is clear that the structural statistics for distance refined structures support the folding topology shown in Figure 1(a) and are incompatible with the folding topology in Figure S1(a), since the latter exhibit poorer convergence rates, higher van der Waals energy values, greater number of violations and increased deviations from ideal covalent geometry. A subset of 12 distance refined structures of the dimeric d(G-G-A-G-G-A-G) hexad motif excluding G7 (folding topology in Figure 1(a)) were identified based on a combination of low NOE energies and fewest NOE violations.

Intensity restraints and NOE back calculations

The subset of 12 converged distance-refined structures corresponding to the folding topology in Figure 1(a) were next refined against the non-exhangeable proton NOE intensities associated with NOESY spectra recorded at five mixing times. These computations utilized an energy minimization with back calculation protocol described in Materials and Methods. The NOE violations, deviations from covalent geometry and pairwise r.m.s.d. values for the 10 lowest energy intensity-refined structures of the dimeric d(G-G-A-G-G-A-G) hexad motif are listed in Table 1.

Table 1. NMR restraints and structural statistics for the intensity refined structures of the dimeric d(G-G-A-G-G-A-G) hexad motif, excluding G7

A. *NMR restraints*	
Distance restraints	
Intra-residue distance restraints[a]	124
Sequential (i, $i+1$) distance restraints[a]	97
Long range $\geqslant(i, i+2)$ distance restraints[a]	58
Other restraints	
Hydrogen bonding restraints[b]	88
Glycosidic torsion angle restraints[c]	24
Intensity restraints	
Non-exchangeable protons (total for five mixing times)	1065
B. *Structural statisitics in complex following intensity refinement*	
NOE violations	
Number >0.2 Å	2.7(±1.0)
Maximum violations Å	0.32(±0.02)
r.m.s.d. of violations Å	0.042(±0.02)
NMR R-factor ($R_{1/6}$)	0.051(±0.002)
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.014(±0.0001)
Bond angles (deg.)	2.15(±0.05)
Impropers (deg.)	0.18(±0.02)
Pairwise all heavy atom r.m.s.d. values (10 refined strucutres)	
Dimeric d(G-G-A-G-G-A-G) hexad motif, excluding G7 (Å)	0.97(±0.30)

[a] All distance restraints were set as ambiguous between intra and inter-strand contributions.

[b] These hydrogen bonding restraints are based on experimental NOE and $^2J_{NN}$ coupling data.

[c] Residues G1 to A6 were restrained to χ values in the 210(±40)° range, characteristic of anti glycosidic torsion angles identified experimentally.

Dimeric ''hexad'' motif

A view of the ten lowest energy intensity refined structures of the dimeric d(G-G-A-G-G-A-G) hexad motif (excluding G7) are shown in stereo in Figure 6(a). Individual strands are colored in yellow, cyan, green and orange with backbone phosphorus atoms in red. The individual symmetry related components within the dimer contain an A3·(G1·G4·G1·G4)·A3 hexad stacked on a G2·G5·G2·G5 tetrad which is in turn stacked on a A6·A6 mismatch (see the schematic in Figure 1(a)). The dimer interface contains a pair of stacked A3·(G1·G4·G1·G4)·A3 hexads.

The directionalities of the individual strands in the dimeric hexad motif are shown in Figure 6(b) (see also the schematic in Figure 1(a)). A space filling view of this dimeric hexad motif with one of the strands in yellow and the others in white is shown in Figure 6(c).

A stick representation of one symmetric half of the dimeric d(G-G-A-G-G-A-G) hexad motif (excluding G7) is shown in Figure 7(a). Note that the backbone spanning the G1-G2-A3-G4-G5 segment undergoes two successive turns resulting in an S-shaped double chain reversal motif. The pairing alignments of the A3·(G1·G4·G1·G4)·A3 hexad, reversed A6·A6 mismatch and G2·G5·G2·G5 tetrad are shown in Figure 7(b), (c) and (d), respectively.

The stacking geometries between the reversed A6·A6 mismatch and G2·G5·G2·G5 tetrad, between the G2·G5·G2·G5 tetrad and A3·(G1·G4·G1·G4)·A3 hexad and between the A3·(G1·G4·G1·G4)·A3 hexads across the dimer interface are shown in Figure 8(a), (b) and (c), respectively.

Identification of potential cation binding sites

A combination of Brownian dynamics (BD) and unconstrained molecular dynamics (MD) simulations using protocols described in Materials and Methods, have been used to probe potential metal cation binding sites within the dimeric d(G-G-A-G-G-A-G) hexad structure. Five electronegative pockets suitable for the binding of monovalent cations (Figure 9(a)) were predicted by BD simulations of cation diffusion (Hermann & Westhof, 1998) in the electrostatic field of the dimeric hexad motif. Three of these sites are sandwiched between the planes of the hexad and tetrad base alignments similar to the positions of monovalent ions in other G tetrad motifs (Kang *et al.*, 1992; Laughlan *et al.*, 1994; Hud

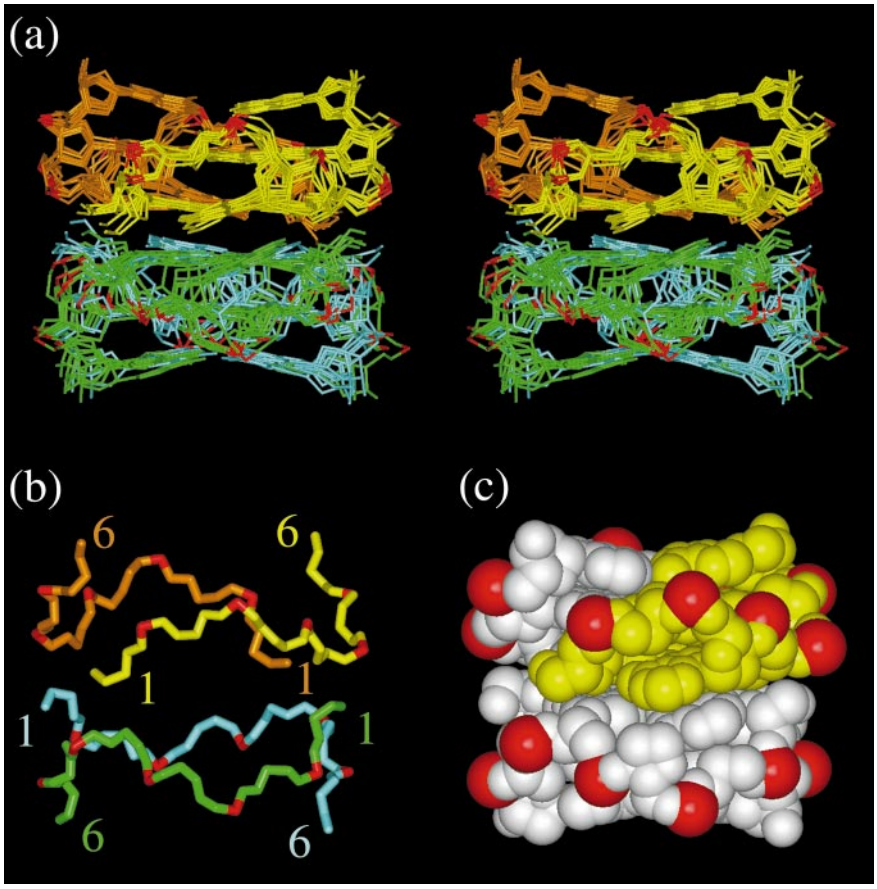

Figure 6. (a) Superposed stereo views of 10 intensity refined structures of the dimeric d(G-G-A-G-G-A-G) hexad motif, excluding G7. The individual strands (yellow, cyan, green, orange), the backbone phosphorus atoms (red). The exocyclic phosphate oxygens are deleted in the interests of clarity. The best fit superposition was achieved for the top half (strands, yellow and orange) of the dimeric hexad motif. (b) A stick depiction of the backbone of a representative distance refined structure of the dimeric hexad motif, excluding G7. (c) A space-filling view of a representative distance refined structure of the dimeric hexad motif, excluding G7. One of the strands (yellow) and the other three (white), the backbone phosphorus atoms (red).

et al., 1998). Two additional electronegative pockets were predicted for the two symmetry-related, unoccupied grooves of the stacked core structure (Figure 9(a)). The groove sites are centered between the hexad planes and adjacent phosphate groups of the DNA backbone, approximately at positions where adenines dock into the central tetrads in the other two grooves. The phosphate groups and O4′ atoms of symmetry-related pairs of G5 and N3 atoms of pairs of G4 contribute to the electronegative potential in the tetrad grooves (Figure 9(b)).

Metal ion binding in the calculated electronegative pockets of the dimeric hexad motif was further investigated by unconstrained MD simulations. A complex of the dimeric hexad structure with five Na^+ placed at the predicted sites was submitted to MD simulations in explicit solvent water. The stability of the Na^+ at their potential binding sites and their hydration were followed over two 1 ns MD trajectories calculated at 300 K. The DNA structure remained close to the experimental NMR structure in both trajectories attested by a r.m.s.d. of around 2.0 Å after 1 ns simulation. The three Na^+ sandwiched between the hexad and tetrad planes were stably retained at their binding positions directly coordinated to O6 carbonyl oxygen atoms of the guanine bases in the stacked tetrads. The two Na^+ in the tetrad grooves also remained bound to the DNA stabilized by a network of hydrogen bonds between the hydration shell of the cations and the nucleic acid (Figure 9(b)). Two water molecules were found hydrogen-bonded simultaneously with the N3 atom of G4 and the O4′ of the adjacent G5, both of which contribute significantly to the electronegative pockets in the tetrad grooves. Two other water molecules formed stable hydrogen bonds with the phosphate groups of G5.

Discussion

Our NMR studies on the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl has identified a new DNA architecture termed the dimeric hexad motif. The formation of A3·(G1·G4·G1·G4)·A3 hexads and the S-shaped double chain reversal of the backbone spanning the G1-G2-A3-G4-G5 segment represent the most unexpected features of this

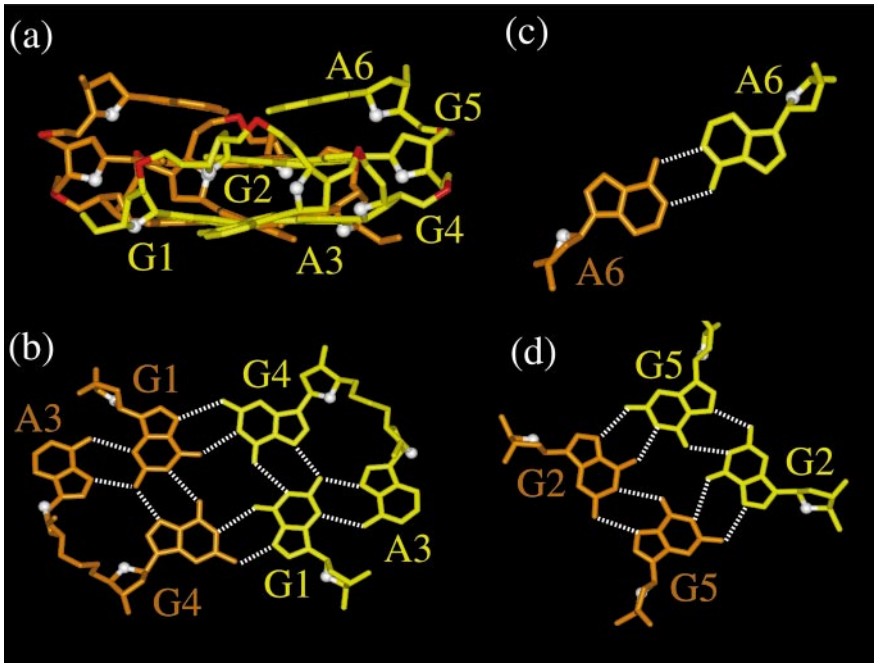

Figure 7. (a) A stick representation of one symmetric half of a representative intensity refined structure of the dimeric d(G-G-A-G-G-A-G) hexad motif, excluding G7. (b) Pairing alignments in the A3·(G1·G4·G1·G4)·A3 hexad. Note formation of sheared G1·A3 mismatches and A3-G4 platforms between residues of the same strand. (c) Pairing alignment along the Watson-Crick edges of the A6·A6 mismatch. (d) Pairing alignments in the G2·G5·G2·G5 tetrad.

four-stranded architecture. Individual hexads in turn form through alignment of pairs of G1·(A3-G4) triads. The structural features associated with the dimeric d(G-G-A-G-G-A-G) motif are discussed below, along with an analysis of the origin of the NaCl concentration dependence, which results in a switch from a two-stranded arrowhead motif in 10 mM Na^+ reported previously (Kettani *et al.*, 1999) to a four-stranded dimeric hexad motif in 150 mM NaCl.

Structure determination

The main challenge during structure determination was to distinguish between two alternate folding topologies represented schematically in Figures 1(a) and S1 (Supplementary Material). The structures based on the folding topology in Figure 1(a) were able to satisfy the experimentally observed NOE values between the H-2 and sugar protons of A3 and between the sugar protons of G1 and G4 on adjacent A3·(G1·G4·G1·G4)·A3 hexads across the dimer interface (Table S2). By contrast, severe violations were observed for structures associated with the folding topology in Figure S1 (Supplementary Material), due to the large separation in the computed structures between H-2 and sugar protons of A3 on adjacent hexads across the dimer interface.

There was an improvement in the number of >0.2 Å violations on proceeding from distance refined (9.7(±1.5), Table S3) to intensity refined (2.7(±1.0), Table 1) structures of the dimeric hexad motif. The structures are well defined within each symmetric half of the dimeric hexad motif, but show some variability in the relative alignments across the dimeric interface (the dimer hexad motif refined structures in Figure 6(a) represent superpositioning for the yellow and orange strands only).

Guanine glycosidic torsion angles

The sugar ring oxygen atoms are represented by white balls for purine bases in the A6·A6 mismatch (top step, Figure 7(a)), the G2·G5·G2·G5 tetrad (middle step, Figure 7(a)) and the A3·(G1·G4·G1·G4)·A3 hexad (lower step, Figure 7(a)). It is striking that all purines, including G1, G2, G4 and G5, experimentally adopt anti conformations, consistent with the parallel orientation (Cheong & Moore, 1992; Wang & Patel, 1993; Gupta *et al.*, 1993; Aboul-ela *et al.*, 1994; Laughlan *et al.*, 1994) of the pair of G1-G2 and the pair of G4-G5 steps connecting the G2·G5·G2·G5 tetrad and A3·(G1·G4·G1·G4)·A3 hexad within each symmetry-related half of the folded dimeric hexad motif architecture (see the schematic in Figure 1(a)).

G·(A-G) triad

Residues G1, A3 and G4 from the same strand are involved in G1·(A3-G4) triad formation (Figure 7(b)) in the dimeric d(G-G-A-G-G-A-G) hexad architecture. Kuryavyi & Jovin (1995, 1996) put forward the concept of base triads where two adjacent bases are coplanar (A3 and

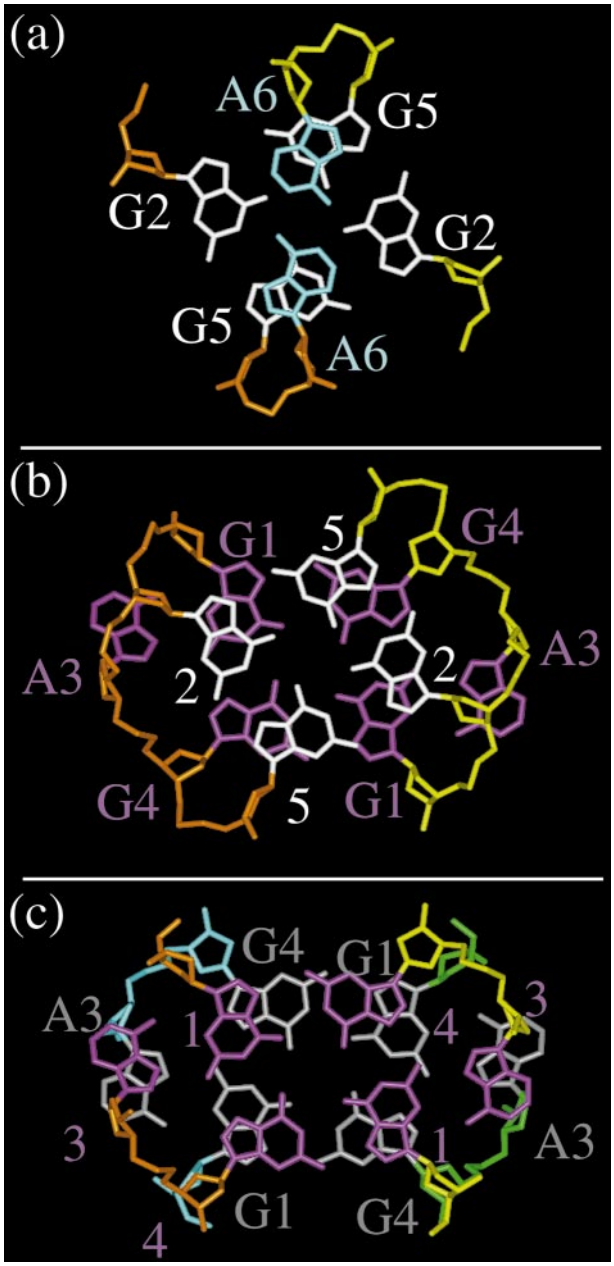

Figure 8. Stacking alignments in a representative intensity refined structure of the dimeric d(G-G-A-G-G-A-G) hexad motif, excluding G7. (a) Overlap geometry between the A6·A6 mismatch (in cyan) and the G2·G5·G2·G5 tetrad (white). (b) Overlap geometry between the G2·G5·G2·G5 tetrad (white) and the A3·(G1·G4·G1·G4)·A3 hexad (magenta). (c) Overlap geometry between adjacent A3·(G1·G4·G1·G4)·A3 hexads (magenta and grey).

G4 in our case) and in turn pair in the same plane with a non-adjacent third base (G1 in our case). Experimental support for the coplanar alignment of adjacent bases emerged with the identification of three separate A(anti)-A(anti) platforms in the crystal structure of the P4-P5-P6 domain of the *Tetrahymena* group I intron (Cate *et al*., 1996). Subsequently, an A·(T-A) triad containing a T(anti)-A(anti) platform (Kettani *et al*., 1997) and a T·(A-A) triad containing an A(*syn*)-A(anti) platform (Kuryavyi *et al*., 2000) have been reported in multi-stranded DNA architectures. In addition, U·(A-A) (Kalurachchi *et al*., 1997), G·(A-C) (Zimmerman *et al*., 1997) and U·(G-U) (Wimberly *et al*., 1999; Conn *et al*., 1999) triads have been identified to date in RNA folds.

The G1·(A3-G4) triad (Figure 7(b)) in the dimeric d(G-G-A-G-G-A-G) hexad motif has several features that have not been identified in triads reported previously. This is the first example of an all purine triad, in contrast to all previous triads which contain atleast one pyrimidine (Kettani *et al*., 1997; Kuryavyi *et al*., 2000; Kalurachchi *et al*., 1997; Zimmerman *et al*., 1997; Wimberly *et al*., 1999; Conn *et al*., 1999). Further, both bases in the platform (A3-G4 in our case) form hydrogen bonds with the interacting base (G1 in our case) (Figure 7(b)). This could reflect the availability of Watson-Crick and Hoogsteen edges within purine bases for hydrogen bonding alignment. Indeed, G1 pairs with A3 through sheared G·A mismatch formation while G1 pairs with G4 through Hoogsteen G·G mismatch formation (Figure 7(b)). Thirdly, this is also the first example of two triads aligning to form a hexad, thus establishing the versatility of triads as a module for formation of multi-stranded architectures.

A·(G·G·G·G)·A hexad formation

All residues within the d(G-G-A-G-G-A) segment of the dimeric hexad motif are involved in pairing alignments either as non-canonical pairs, tetrads or hexads. The Watson-Crick and Hoogsteen edges of guanine are involved in pairing around the G·G·G·G tetrad. This leaves the minor group edge of guanine exposed for specific recognition. Sheared G·A mismatch formation involves alignment of the minor groove edge of guanine with the major groove edge of adenine (Li *et al*., 1991; Heus & Pardi, 1991). Thus, adenine bases are ideally suited for targeting G·G·G·G tetrads through sheared G·A mismatch formation. The pairing alignment around the A·(G·G·G·G)·A hexad in a representative refined structure of the dimeric d(G-G-A-G-G-A-G) hexad motif is shown in Figure 7(b). Note that the Watson-Crick, major groove and minor groove edges of G1 are all involved in hydrogen bonding, so that this residue is anchored in place through six hydrogen bonds (Figure 7(b)).

Successive reverse turns resulting in double chain reversal

Individual strands of the dimeric d(G-G-A-G-G-A-G) hexad motif undergo a sharp turn within the G1-G2-A3 segment and a gradual turn within

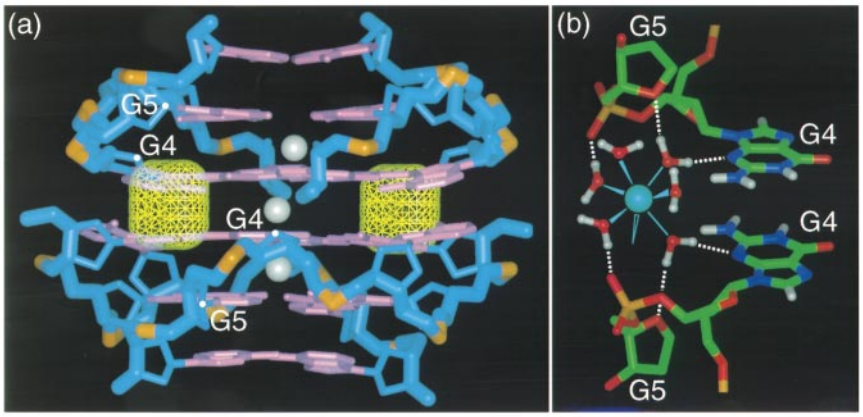

Figure 9. Metal ion binding sites of the dimeric d(G-G-A-G-G-A-G) hexad motif. (a) Five electronegative pockets suitable for cation binding were predicted by Brownian dynamics simulations. Three of these sites (grey spheres) are located between the planes of the hexad and tetrad alignments (pink). Two additional symmetry-related electronegative pockets (yellow density grid) are found in the grooves centered between the base edges of the hexad planes. (b) Hydrated Na^+ were stably bound at the predicted sites in the grooves during molecular dynamics simulations of the fully solvated dimeric hexad structure. Binding of Na^+ (blue) in the grooves is mediated by a symmetric hydrogen bond network between water molecules of the hydration shell and bases and sugars of the dimeric hexad motif. Six of the water molecules coordinated to Na^+ show persisting residence over the 1 ns MD simulations whereas occupation of a seventh coordination site (open wedge) is fluctuating.

the G2-A3-G4-G5 segment (see schematic in Figures 1(a)). In the case of reversal in the G1-G2-A3 segment, a single G2 residue is involved in turn formation, with the chain reversal closed by formation of a sheared G1·A3 mismatch pair. Such single residue turns have been reported previously for G-N-A segments closed by a sheared G·A mismatch pair (Hirao *et al.*, 1994; Zhu *et al.*, 1995; Chou *et al.*, 1996; Yoshizawa *et al.*, 1997). They are characterized by an unusual upfield shift for the H4′ proton of the turn residue (Zhu *et al.*, 1995), as has been observed for the H4′ proton of G2 resonating at 2.83 ppm (Table S1) in the d(G-G-A-G-G-A-G) sequence in 150 mM NaCl. It should be noted that there are differences in the orientation of the sugar ring and backbone of A3 in our G1-G2-A3 turn (Figure 7(a)) and the corresponding A residue in more classical G-N-A turns (Hirao *et al.*, 1994; Zhu *et al.*, 1995; Chou *et al.*, 1996; Yoshizawa *et al.*, 1997). This difference reflects the participation of A3 also in a second turn associated with the G2-A3-G4-G5 segment in our structure.

In the case of reversal in the G2-A3-G4-G5 segment, the turn occurs at the A3-G4 platform step and is closed by a Hoogsteen G2·G5 mismatch pair. In essence, the two successive turns within each individual strand of the dimeric head motif defines an S-shaped double chain reversal architecture, which connects the G1-G2 and G4-G5 steps aligned parallel to each other along adjacent edges of the quadruplex.

Stacking

The dimeric d(G-G-A-G-G-A-G) hexad has mismatches stacking on tetrads, which in turn stack on hexads, within each symmetric half of the architecture. Thus, both base-base and base-sugar stacking could contribute to stabilization of the global fold. There is extensive stacking between the A6 residues of the A6·A6 mismatch and the G5 residues of the adjacent G2·G5·G2·G5 tetrad (Figure 8(a)) in the refined structure of the dimeric hexad motif. The H2 proton of A6 at 7.07 ppm is positioned over the G2·G5 mismatch pair accounting for its observed upfield shift. The stacking between the adjacent G2·G5·G2·G5 tetrad and A3·(G1·G4·G1·G4)·A3 hexad is primarily between guanine bases within individual strands, while the purine rings of A3 residues are partially stacked over the sugar rings of G2 residues (Figure 8(b)). This latter stacking could account for the observed upfield shift for the H4′ proton of G2 which resonates at 2.83 ppm. There is considerable overlap between the five-membered purine rings of G1 and G4 and between the five-membered purine rings of A3 residues resulting in extensive stacking between symmetry related A3·(G1·G4·G1·G4)·A3 hexads across the dimer interface (Figure 8(c)).

The experimental data establish that the d(G-G-A-G-G-A-N) sequence can switch from a two-stranded ''arrowhead'' motif in 10 mM NaCl (Kettani *et al.*, 1999) to a four-stranded dimeric hexad motif in 150 mM NaCl solution (this study). Note that a sharp chain reversal is observed for both conformers, with reversal at A3 for the arrowhead motif (Kettani *et al.*, 1999) and at G2 for the dimeric hexad motif. Intra- and inter-strand stacking appears to be quite extensive for both motifs, but the complete stacking between adjacent A·(G·G·G·G)·A hexads in the dimeric ''hexad'' motif (Figure 8(c)), explains in part why the moderate salt hexad motif is more stable than its low salt arrowhead counterpart.

Na^+ dependence of arrowhead to hexad transition

The switch of the d(G-G-A-G-G-A-N) sequence from an arrowhead motif in 10 mM NaCl (Kettani *et al.*, 1999) to a dimeric hexad motif in 150 mM NaCl (this study) may reflect the role of alkali cations in stabilizing G·G·G·G tetrad-contianing architectures. We have also observed the same transition with K^+ as cation as a function of KCl concentration. Both X-ray (with Na^+ and K^+) (Laughlan *et al.*, 1994; Kang *et al.*, 1992) and NMR (with NH_4^+) (Hud *et al.*, 1998) studies have identified monovalent cations that are sandwiched between adjacent G·G·G·G tetrads in DNA quadruplexes. Nevertheless, we were interested in elucidating whether additional metal ion binding sites, beyond the classical ones coordinating the inwardly pointing carbonyls on adjacent G·G·G·G tetrads, could be found within the novel fold of the dimeric hexad motif.

The identification and positioning of monovalent cations in the tetrad grooves (Figure 9(a)) determined from the combined Brownian and molecular dynamics simulations was not anticipated and suggests that such metal ion sites might be involved in stabilization of the dimeric hexad motif by bridging between the two symmetry-related DNA subunits. The two cation binding sites in the grooves, along with the metal ion binding sites between the hexad and tetrad planes, could account for the experimentally observed dependence of the dimeric hexad structure on moderate salt concentrations. By contrast, no comparable stable cation binding sites were found by prediction methods applied on the arrowhead motif which is the folded form of the d(G-G-A-G-G-A-N) sequence at low salt concentrations (Kettani *et al.*, 1999).

Earlier investigations have reported on cation concentration dependent transitions between Watson-Crick hairpin and G·G·G·G tetrad-mediated quadruplex in G + C-rich sequences (Hardin *et al.*, 1992). Our research on cation concentration dependent structural transitions between the arrowhead (Kettani *et al.*, 1999) and dimeric hexad motifs in guanine plus adenine rich sequences has resulted in the identification of both novel DNA architectures and unanticipated cation binding sites.

Biological implications

Here, we report on a novel multi-stranded DNA architecture which we have designated the dimeric hexad motif. A key discovery is our demonstration for the first time of a noncovalent pair of stacked A·(G·G·G·G)·A hexads that form the base of a diamond-shaped architecture involving hexads, stacked on tetrads which are in turn stacked on mismatches. A second discovery defines in molecular terms the conformational transition associated with the Na^+ (or K^+) concentration dependent switch from the two-stranded arrowhead motif (Kettani *et al.*, 1999) to the four-stranded dimeric hexad motif. Both the arrowhead (Kettani *et al.*, 1999) and dimeric hexad motifs contain novel pairing alignments along with sharp turns and point to the structural and functional diversity of multi-stranded DNA architectures.

The ends of chromosomes contain guanine-rich repeats called telomeres (reviewed by Rhodes & Giraldo, 1995). For humans the repeat is $d(T_2AG_3)_n$, with all telomeres containing a protruding single strand overhang at the 3′-end. Telomere maintenance depends on the enzyme telomerase, which is a reverse transcriptase containing an RNA template. The guanine-rich telomere overhangs can adopt a range of folds ranging from a linear extended conformation required for binding to telomerase and telomere-binding proteins to higher order structures involving mismatch and quadruplex formation. Thus, a shift in equilibrium between G-quadruplex and extended folds for guanine-rich telomeric sequences towards the former architecture, could represent an approach for potential inhibition of telomerase activity. This implies that G-quadruplex-binding ligands may have potential as antitumor drugs and considerable effort has focused on the identification and characterization of small ligands (Chen *et al.*, 1996), intercalation-capable porphyrins (Wheelhouse *et al.*, 1998; Anantha *et al.*, 1998; Arthanari *et al.*, 1998; Haq *et al.*, 1999) and chromophores (Fedoroff *et al.*, 1998; Perry *et al.*, 1998) that can target the grooves or insert between tetrad planes of the quadruplex, respectively.

Here, we demonstrate for the first time that there is a stereochemical possibility for targeting G-quadruplexes site-specifically along their G·G·G·G tetrad edges through A·(G·G·G·G)·A hexad formation. This could be achieved through sheared G·A mismatch formation, a robust mismatch alignment, that has been increasingly observed to stabilize novel folds and recognition elements in RNA (Wimberly *et al.*, 1993; Szewczak *et al.*, 1993; Fan *et al.*, 1996; Cai *et al.*, 1998; Legault *et al.*, 1998) and DNA (Chou *et al.*, 1997; Lin & Patel, 1997; Lin *et al.*, 1998; Sheppard *et al.*, 1999).

A future challenge concerns the design of adenine rich G-quadruplex interacting sequences for potential targeting of the G·G·G·G tetrad edges through A·(G·G·G·G)·A hexad formation. Thus, could more than two adjacent stacked G·G·G·G tetrad planes associated with telomeric G-quadruplex folds be targeted through sheared G·A mismatch formation? To what extent would formation of such a postulated stacked array be modulated by the directionalities of individual strands around the G-quadruplex and the anti/*syn* distribution of guanine glycosidic torsion angles? Model building indicates that both the Watson-Crick and Hoogsteen edges of adenines can target the minor groove edges of anti guanine bases around a G·G·G·G tetrad through G·A mismatch formation but are occluded due to steric effects from targeting syn

guanine bases through similar mismatch alignments. Would adenine rich G-quadruplex interacting sequences be restricted to targeting the minor groove edges of anti guanine bases along a single strand or could they through appropriate positioning of anti and syn adenine bases target the minor groove edges of anti guanine bases on adjacent strands within individual grooves of the quadruplex? In the longer term, can adenine rich G-quadruplex interacting sequences be designed to target human telomeric DNA quadruplex folds formed by the $d(T_2AG_3)_n$ repeat? These and related questions are currently under active investigation.

Materials and Methods

Preparation of unlabeled and selectively labeled DNA

The d(G-G-A-G-G-A-G) sequence was synthesized on a 10 μmol scale on an Applied Biosystems 392 DNA synthesizer using solid phase β-cyanoethylphosphoamidite chemistry and was subsequently purified by HPLC. The details of the synthesis and purification of the d(G1-G2-A3-G4-G5-A6-G7) sequence labeled with ^{15}N at N^1, N^2 and N^7 and/or ^{13}C at C^2 of guanine and with ^{15}N at N^1, N^6 and N^7 of adenine as labeled G4 and A6 pairs and as labeled G2 and G5 pairs will be reported elsewhere. The DNA oligomers in 5 ml volume were dialyzed against five changes of H_2O, followed by dialysis against 150 mM NaCl, 2 mM Na-phosphate buffer (pH 6.6) and final dialysis against 15 mM NaCl, 0.2 mM phosphate buffer (pH 6.6) and lyophilized.

Preparation of the uniformly ^{13}C,^{15}N-labeled nucleoside 5′-triphosphates

Escherichia coli XL-Blue cells bearing transformed high copy pVK plasmid (pUC derivative) were grown in 50% (v/v) Celtone-CN ($^{13}C > 98\%$, $^{15}N > 98\%$) and 50% Martek 9-CN ($^{13}C > 98\%$, $^{15}N > 98\%$) media purchased from Martek Biosciences Corp. in the presence of 200 mg/l ampicillin (Sigma). The cells were harvested at late log phase. The pH of the media was adjusted to 7.5 with Na-phosphate buffer and fresh *E. coli* cells were inoculated. Growth cycles were repeated five times till the yield dropped below 0.5 A_{600}/ml. The yield was about 20 g wet cells per liter of media. Cell pellets were disrupted using a French press and total polymeric nucleic acids in the disrupted cells were next separated from proteins by phenol extraction. The polymeric DNA and RNA were hydrolyzed to the monophosphates and the ribo-NMPs were separated from dexoy-NMPs on a boronate column according to the general procedure of Batey *et al*. (1995) for isolation of labeled RNA. The mixture of uniformly labeled ^{13}C,^{15}N dNMPs were separated by reverse-phase HPLC on a semi-preparative C-18 column and next converted to dNTPs using published enzymatic procedures (Smith *et al*., 1997). Specifically, we isolated glutathione-S-transferase/yeast deoxythymidylate kinase fusion protein (GST-dTMP kinase) for conversion of TMP to TTP from *E. coli* strain DH10β transformed corresponding recombinant plasmid GEX-2 derivative (kindly provided by Dr F. Jucker of Nexstar Pharmaceuticals).

Preparation of uniformly ^{13}C, ^{15}N-labeled d(G-G-A-G-G-A-G)

A modified version of the Zimmer & Crothers (1995) procedure as described previously (Kettani *et al*., 1999) was used for the enzymatic synthesis of uniformly ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G). The template consisted of a 23-mer sequence of deoxy residues terminating in a ribo residue, of which a central 16-mer segment formed a self-complementary duplex linked to d(C-T-C-C-T-C-C) 5′-overhangs. The in-house prepared uniformly ^{13}C,^{15}N-labeled dNTPs were used as building blocks in the *in vitro* polymerization reaction catalyzed by murine mammary leukemia virus (MMLV) reverse transcriptase (Gibco-BRL). The uniformly ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) 7-mer was cleaved at the rbio-deoxy step by alkaline hydrolysis following completion of the *in vitro* polymerization and separated from the unlabeled 23-mer template using 22% denaturing polyacrylamide electrophoresis. The DNA 7-mer bands were eluted from the gel by ''crush-and-soak'' procedure (Chen & Ruffner, 1996) and purified as described above for the non-labeled samples.

NMR data collection and processing

NMR data on the d(G-G-A-G-G-A-G) 7-mer in H_2O and ^2H_2O buffer (150 mM NaCl, 2 mM phosphate (pH 6.6)) were collected on a Varian 600 MHz Unity *Inova* NMR spectrometer. Proton assignments are based on homonuclear NOESY, correlation spectroscopy (COSY), total correlated spectroscopy (TOCSY) and HCCNH-TOCSY (Fiala *et al*., 1996; Simorre *et al*., 1996; Sklenar *et al*., 1996) experiments. Data sets were processed and analyzed using the FELIX program (Molecular Simulations). Two bond $^2J_{NN}$ scalar couplings between imino and amino donors and nitrogen acceptors in uniformly ^{13}C,^{15}N-labeled d(G-G-A-G-G-A-G) in 150 mM NaCl were monitored in HNN-COSY and H(CN)N(H)-COSY contour plots using a pulse sequence described in the literature (Dingley & Grzesiek, 1998; Pervushin *et al*., 1998; Majumdar *et al*., 1999a,b).

Distance restraints

The distances between non-exchangeable protons were estimated from the buildup curves of cross-peak intensities in NOESY spectra at five different mixing times (50, 100, 150, 200 and 300 ms) in ^2H_2O and given bounds of ±30% with distances referenced relative to the sugar H1′-H2″ distance of 2.20 Å. Exchangeable proton restraints are based on NOESY data sets at two mixing times (60 and 200 ms) in H_2O. Cross-peaks involving exchangeable protons were classified as strong (strong intensity at 60 ms), medium (weak intensity at 60 ms) and weak (observed only at a mixing time of 200 ms) and proton pairs were then restrained respectively to distances of 3.0(±0.9) Å, 4.0(±1.2) Å and 6.0(±1.8) Å. Since the experimental NMR data are consistent with a four-stranded motif containing a pair of 2-fold symmetry axis, non-crystallographic symmetry restraints were imposed on all heavy atoms.

Structure calculations

The structure of the d(G-G-A-G-G-A-G) sequence (excluding G7) in 150 mM NaCl was determined by MD-simulated annealing computations driven by NOE

distance and hydrogen bonding restraints using X-PLOR package, version 3.8 (Brunger, 1992). At the initial stage of the refinement, torsional molecular dynamics was undertaken at high temperature. The molecules were equilibrated at 20,000 K (30,000 steps over 3 ps) and then cooled very slowly to 1000 K (40,000 steps over 20 ps). The potential energy function included a repulsive force field, NOE and hydrogen bond distance restraints, glycosidic bond (χ) dihedral angle restraints and a non-crystallographic symmetry potential. The force constant for NOE distance restraints was maintained at a value of 30 kcal mol^{-1} $Å^{-2}$, while for hydrogen bonds restraints the value was 50 kcal mol^{-1} $Å^{-2}$. All NOE distance restraints were considered as ambiguous and treated with the ''sum'' averaging option (Nilges *et al*., 1991; Nilges, 1995). Dihedral angle restraints (210(±40)°, with force constant of 50 kcal mol^{-1} rad^{-2}) were imposed on glycosidic torsion angles for the residues G1, G2, A3, G4, G5 and A6 shown experimentally to adopt anti conformations. The force constant for non-crystallographic symmetry was maintained at 30 kcal mol^{-1} $Å^{-2}$.

These computations were followed by lower temperature cartesian space molecular dynamics guided by the hydrogen bonding and NOE distance restraints with changes in the potential energy function: the repulsive force field was replaced with Lennard-Jones potentials and planarity restraints were included for tetrad and hexad planes with low weights of 5 kcal mol^{-1} $Å^{-2}$ and 10 kcal mol^{-1} $Å^{-2}$, respectively. During this stage of the dynamics, the structures were further cooled from 1000 K to 300 K (20,000 steps over 10 ps) and minimized until the gradient of energy was less than 0.1 kcal mol^{-1}. It should be noted that computations repeated without planarity restraints resulted in the same low energy structures, but exhibited a lower convergence rate.

The refinement protocol started from sixty different initial structures. The initial structures were generated as sets of four chains, each six nucleotides long (excluding underdefined G7), randomized for all dihedral angles, and separated by space intervals of 50 Å. The convergence rate following dynamics was good (given that the system contained four independent chains) for the case where the computations were guided by hydrogen bonding restraints associated with the topoplogy shown in Figure 1(a): 12 structures out of 60 emerged with the same fold and pairwise r.m.s.d. values less than 1 Å between members of the group. Non-converged structures were separated from that group by large gaps (in total more than 100 kcal) in all components of the potential energy (van der Waals, NOE violations, covalent geometry). By contrast, the convergence rates, van der Waals energy, number of NOE violations and deviations from ideal covalent geometry were poor (Table S3, Supplementary Material) for the case where the computations were guided by hydrogen bonding restraints associated with the topoplogy shown in Figure S1 (Supplementary Material).

The 12 converged distance refined structures corresponding to the folding topology shown schematically in Figure 1(a) were used as the starting point for subsequent X-PLOR based energy minimization with back-calculation of the NOESY spectra. The relaxation matrix was set up for the nonexchangeable protons, with the exchangeable imino and amino protons exchanged for deuterons. A total of 1065 non-exchangeable intensity values from NOESY data sets at five mixing times in ^2H_2O buffer (213 nonexchangeable intensities per mixing time) were included with force constant of 500 kcal mol^{-1}. Planarity restraints were lifted at this stage while distance restraints were retained with 30 % bounds and the same weights as before. During minimization, the NMR R-factor ($R_{1/6}$) improved from the initial value of 10 % to 5.1 % while retaining structure convergence and stereochemistry.

Prediction and simulation of cation binding

Brownian dynamics simulations of cation diffusion (Hermann & Westhof, 1998) were used to determine electronegative pockets as potential binding sites for cations in refined NMR structures of the d(G-G-A-G-G-A-G) sequence in 10 mM (Kettani *et al*., 1999) and 150 mM NaCl solution. The electrostatic field of the DNA was calculated with the nonlinear Poisson-Boltzmann equation using partial atomic charges from the AMBER 94 force field (Cornell *et al*., 1995). Sites meeting electrostatic and sterical criteria for the binding of metal cations were identified by simulations of the diffusion of +1 charged spherical probes in the electrostatic field of the DNA. Simulations were performed with the UHBD program (Madura *et al*., 1994) and analyzed as has been described previously (Hermann & Westhof, 1998).

The MD simulations under realistic solvent conditions were done with the AMBER software and force field (Cornell *et al*., 1995). The five Na^+ were placed at the predicted sites followed by immersion of the DNA complex in a solvent box containing ~3000 molecules of SPC/E water molecules and 15 Na^+ conterions randomly added at distances of larger than 5 Å from any atom of the solute. The system was prepared for the simulations at 300 K by applying a multi-step procedure adapted from previously published protocols (Hermann *et al*., 1998) redistributing first the solvent molecules and bulk ions at 300 K while both DNA and bound Na^+ were fixed, followed by an intermediate phase in which the constraints on the bound Na^+ were slowly reduced to zero, and a final phase, initiated with a restart at 10 K and slow heating to 300 K without any constraints. Two independent MD trajectories of 1 ns length, obtained by slight modifications in the equilibration protocol, were run with a time step of 2 fs and subsequently analyzed with the Carnal module of the AMBER software. Periodic boundary conditions were applied throughout along with the Particle Mesh Ewald method for the evaluation of electrostatic interactions.

Coordinates deposition

Coordinates (accession number: 1eeg) of the dimeric d(G-G-A-G-G-A-G) hexad motif have been deposited in the RCSB Protein Data Bank.

Acknowledgments

This research was supported by NIH grants GM-34504 to D.J.P. and GM-48802 and GM-31486 to R.A.J. We thank Timothy Ellis for preparation of dNTPs and Dr Weidong Hu for devoloping an improved version of the pulse sequence for the HCCNH-TOCSY experiment.

References

Aboul-ela, F., Murchie, A. I., Norman, D. G. & Lilley, D. M. (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.

Anantha, N. V., Azam, M. & Sheardy, R. D. (1998). Porphyrin-binding to quadruplex T_4G_4. *Biochemistry,* **37**, 2709-2714.

Aoki, T., Koch, K. S. & Leffert, H. L. (1997). Attenuation of gene expression by a trinucleotide repeat-rich tract from the terminal exon of the rat hepatic polymeric immunoglobulin receptor gene. *J. Mol. Biol.* **267**, 229-236.

Arthanari, H., Basu, S., Kawano, T. L. & Bolton, P. H. (1998). Fluorescent dyes specific for quadruplex DNA. *Nucl. Acids Res.* **26**, 3724-3728.

Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995). Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol.* **261**, 300-322.

Bouaziz, S., Kettani, A., Zhao, H., Jones, R. & Patel, D. J. (1998). A K^+ cation dependent conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.* **282**, 637-652.

Brunger, A. T. (1992). *X-PLOR: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. & Patel, D. J. (1998). Solution structure of P22 transcriptional antitermination N-peptide-*boxB* RNA complex. *Nature Struct. Biol.* **5**, 203-212.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. E. & Doudna, J. A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science,* **273**, 1696-1699.

Chen, Q., Kuntz, I. D. & Shafer, R. H. (1996). Spectroscopic recognition of guanine dimeric hairpin quadruplexes by a carbocyanine dye. *Proc. Natl Acad. Sci. USA,* **93**, 2635-2639.

Chen, Z. & Ruffner, D. E. (1996). Modified crush and soak method for recovering oligodeoxynucleotides from polyacrylamide gels. *Biotechniques,* **21**, 820-822.

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

Chou, S.-H., Zhu, L., Gao, Z., Cheng, J-W. & Reid, B. R. (1996). Hairpin loops consisting of single adenine residues closed by a sheared A·A and G·G pairs formed by the DNA triplets AAA and GAG: Solution structure of the d(GTAC*AAA*GTAC). *J. Mol. Biol.* **264**, 981-1001.

Chou, S.-H., Zhu, L. & Reid, B. R. (1997). Sheared purine·purine pairing in biology. *J. Mol. Biol.* **267**, 1055-1067.

Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein-RNA complex. *Science,* **284**, 1171-1174.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.

Dingley, A. J. & Grzesiek, S. (1998). Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{NN}$ couplings. *J. Am. Chem. Soc.* **120**, 8293-8297.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996). Molecular recognition in the FMN-internal loop RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.

Fedoroff, O. Y., Salazar, M., Han, H., Chemeris, V. V., Kerwin, S. M. & Hurley, L. H. (1998). NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry,* **37**, 12367-12374.

Fiala, R., Jiang, F. & Patel, D. J. (1996). Direct correlation of exchangeable and nonexchangeable protons on purine bases in ^{13}C, ^{15}N-labeled RNA using HCCNH-TOCSY experiment. *J. Am. Chem. Soc.* **118**, 689-690.

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 dT_4G_4. *Biochemistry,* **32**, 7098-7103.

Haq, I., Trent, J. O., Chowdhry, B. Z. & Jenkins, T. C. (1999). Intercalative G-tetraplex stabilization of telomeric DNA by a cation porphyrin. *J. Am. Chem. Soc.* **121**, 1768-1779.

Hardin, C. C., Watson, T., Corregan, M. & Bailey, C. (1992). Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d($CGCG_3GCG$). *Biochemistry,* **31**, 833-841.

Heller, M., Flemington, E., Kieff, E. & Deininger, P. (1985). Repeat arrays in cellular DNA related to the Epstein-Barr virus IR3 repeat. *Mol. Cell. Biol.* **5**, 457-465.

Hermann, T. & Westhof, E. (1998). Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure,* **6**, 1303-1314.

Hermann, T., Auffinger, A. & Westhof, E. (1998). Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* **27**, 153-165.

Heus, H. A. & Pardi, A. (1991). Structural features that give rise to unusual stability of RNA hairpins containing GNRA loops. *Science,* **253**, 191-194.

Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. & Miura, K. (1994). Most compact hairpin turn structure exerted by a short DNA fragment d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat. *Nucl. Acids Res.* **22**, 576-582.

Hud, N. V., Schultze, P. & Feigon, J. (1998). Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes. *J. Am. Chem. Soc.* **120**, 6403-6404.

Kalurachchi, K., Uma, K., Zimmermann, R. A. & Nikonowicz, E. P. (1997). Structural features of the binding site for ribosomal protein S8 in *E. coli* 16 S rRNA defined using NMR spectroscopy. *Proc. Natl Acad. Sci. USA,* **94**, 2139-2144.

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

Kettani, A., Bouaziz, S., Wang, W., Jones, R. A. & Patel, D. J. (1997). *Bombyx mori* single repeat telomeric DNA sequence forms a G-quadruplex capped by base triads. *Nature Struct. Biol.* **4**, 382-389.

Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R. & Patel, D. J. (1998). Solution structure of a Na^+ cation stabilized DNA quadruplex containing G·G·G·G and G·C·G·C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* **282**, 619-636.

Kettani, A., Bouaziz, S., Skripkin, E., Majumdar, A., Wang, W., Jones, R. A. & Patel, D. J. (1999). Inter-

locked mismatch-aligned arrowhead DNA motifs. *Structure,* **7**, 803-815.

Koch, K. S., Gleiberman, A. S., Aoki, T., Leffert, H. L., Feren, A., Jones, A. L. & Fodor, E. J. (1995). Discordant expression and variable numbers of neigboring GGA- and GAA-rich triplet repeats in the 3′-untranslated regions of two groups of messenger RNAs encoded by the rat polymorphic immunoglobulin receptor gene. *Nucl. Acids. Res.* **23**, 1098-1112.

Kuryavyi, V. V. & Jovin, T. M. (1995). Triad DNA: a model for trinucleotide repeats. *Nature Genet.* **9**, 339-341.

Kuryavyi, V. V. & Jovin, T. M. (1996). Triangular complementarity of the triad-DNA duplex. In *Proceedings to the Ninth Conversation in Biomology Structure and Dynamics* (Sarma, R. H. & Sarma, M. H., eds), pp. 91-103, Adenine Press, Guilderland, New York.

Kuryavyi, V., Kettani, A., Wang, W., Jones, R. & Patel, D. J. (2000). A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *J. Mol. Biol.* **295**, 455-469.

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

Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. (1998). NMR structure of the bacteriophage λ N peptide/*boxB* RNA complex: recognition of a GNRA fold by an arginine rich motif. *Cell,* **92**, 289-299.

Leroy, J. L., Gehring, K., Kettani, A. & Gueron, M. (1993). Acid multimers of oligocytidine strands: stoichiometry, base-pair characterization and proton exchange properties. *Biochemistry,* **32**, 6019-6031.

Li, Y., Zon, G. & Wilson, W. D. (1991). NMR and molecular evidence for a G·A mismatch base-pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA,* **88**, 26-30.

Lin, C. H. & Patel, D. J. (1997). Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. *Chem. Biol.* **4**, 817-832.

Lin, C. H., Wang, W., Jones, R. A. & Patel, D. J. (1998). Formation of an amino acid-binding pocket through adaptive zippering-up of a large DNA hairpin loop. *Chem. Biol.* **5**, 555-572.

Madura, J. D., Davis, M. E., Gilson, M. K., Wade, R. C., Luty, B. A. & McCammon, J. A. (1994). Biological applications of electrostatic calculations and Brownian dynamics simulations. *Rev. Comput. Chem.* **5**, 229-267.

Majumdar, A., Kettani, A. & Skripkin, E. (1999a). Observation and measurement of $^2J_{NN}$ coupling constants between ^{15}N nuclei with widely separaterd chemical shifts. *J. Biomol. NMR,* **14**, 67-70.

Majumdar, A., Kettani, A., Skripkin, E. & Patel, D. J. (1999b). Observation of internucleotide N-H··N hydrogen bonds in the absence of directly detectable protons. *J. Biomol. NMR,* **15**, 207-211.

Mishima, Y., Kaizu, H. & Kominami, R. (1997). Pairing of DNA fragments containing $(GGA:TTC)_n$ repeats and promotion by high mobility group protein 1 and histone H1. *J. Biol. Chem.* **272**, 26578-26584.

Neidle, S., (ed). (1994). *DNA Structure and Recognition*, Oxford University Press, Oxford.

Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE cross peaks and disulfide connectivities. *J. Mol. Biol.* **245**, 645-660.

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

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.

Patel, D. J., Bouaziz, S., Kettani, A. & Wang, Y. (1999). Structures of guanine-rich and cytosine-rich quadruplexes formed *in vitro* by telomeric, centromeric and triplet repeat disease sequences. In *Oxford Handbook of Nucleic Acid Structures* (Neidle, S., ed.), pp. 389-453, Oxford University Press, Oxford.

Perry, P. J., Reszka, A. P., Wood, A. A., Read, M. A., Gowan, S. M., Dosanjh, H. S., Trent, J. O., Jenkins, T. C., Kelland, L. R. & Neidle, S. (1998). Human telomerase inhibitors by regioisomeric disubstituted amidoanthracene-9,10-diones. *J. Med. Chem.* **41**, 4873-4884.

Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. & Wuthrich, K. (1998). NMR scalar couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. Natl Acad. Sci. USA,* **95**, 14147-14151.

Rhodes, D. & Giraldo, R. (1995). Telomere structure and function. *Curr. Opin. Struct. Biol.* **5**, 311-312.

Sheppard, W., Cruse, W. B., Fourme, R., de la Fortelle, E. & Prange, T. (1999). A zipper-like duplex in DNA: the crystal structure of d(GCGAAAGCT) at 2.1 Å resolution. *Structure,* **6**, 849-861.

Simorre, J. P., Zimmermann, G. R., Mueller, L. & Pardi, A. (1996). Correlation of guanosine exchangeable and nonexchangeable base protons in $^{13}C,^{15}N$-labeled RNA with an HNC-TOCSY-CH experiment. *J. Biomol. NMR,* **7**, 153-156.

Sklenar, V., Dieckmann, T., Butcher, S. E. & Feigon, J. (1996). Through bond correlation of imino and aromatic resonances in ^{13}C, ^{15}N-labeled RNA *via* heteronuclear TOCSY. *J. Biomol. NMR,* **7**, 83-87.

Smith, D. E., Su, J.-Y. & Jucker, F. M. (1997). Efficient enzymatic synthesis of $^{13}C,^{15}N$-labeled DNA for NMR studies. *J. Biomol. NMR,* **19**, 245-253.

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

Szewczak, A. A., Moore, P. B., Chan, Y. L. & Wool, I. G. (1993). The conformation of the sarcin/ricin loop from 28 S ribosomal RNA. *Proc. Natl Acad. Sci. USA,* **90**, 9581-9585.

Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979). Molecular structure of a left-handed DNA fragment at atomic resolution. *Nature,* **282**, 680-686.

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

Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. & Patel, D. J. (1991). Multinuclear NMR studies of Na^+ cation stabilized complex formed by d(G-G-T-T-T-T-C-G-G) in solution: implications for G-tetrad structures. *J. Mol. Biol.* **222**, 259-279.

Wheelhouse, R. T., Sun, D., Han, H., Han, F. X. & Hurley, L. H. (1998). Cationic porphyrins as telomerase inhibitors: the interaction of tetra-(N-

methyl-4-pyridyl)-porphine with quadruplex DNA. *J. Am. Chem. Soc.* **120**, 3261-3262.
Wimberly, B., Varani, G. & Tinoco, I., Jr (1993). The conformation of loop E of eukaryotic 5 S ribosomal RNA. *Biochemistry*, **32**, 1078-1087.
Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell*, **97**, 491-502.
Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K. & Hirao, I. (1997). GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins. *Biochemistry*, **36**, 4761-4767.
Zhu, L., Chou, S-H., Xu, J. & Reid, B. R. (1995). Structure of a single-cytidine hairpin loop formed by the DNA triplet GCA. *Nature Struct. Biol.* **2**, 1012-1017.
Zimmer, D. P. & Crothers, D. M. (1995). NMR of enzymatically synthesized uniformly ^{13}C, ^{15}N-labeled DNA oligonucleotides. *Proc. Natl Acad. Sci. USA*, **92**, 3091-3095.
Zimmerman, G. R., Jenison, R. D., Wick, C. L., Simmore, J. P. & Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA aptamer. *Nature Struct. Biol.* **4**, 644-649.

Edited by P. E. Wright

(Received 17 December 1999; accepted 12 January 2000)

http://www.academicpress.com/jmb

Supplementary material comprising of three Tables and two Figures is available from JMB Online