ARTICLE

OBC
www.rsc.org/obc

Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)$_4$ containing the human telomeric repeat: evidence for A-tetrad formation from NMR and molecular dynamics simulations

Evripidis Gavathiotis and Mark S. Searle*
School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD

Received 21st January 2003, Accepted 27th March 2003
First published as an Advance Article on the web 17th April 2003

The structure of the intermolecular DNA quadruplex d(TTAGGGT)$_4$, based on the human telomeric DNA sequence d(TTAGGG), has been determined in solution by NMR and restrained molecular dynamics simulations. The core GGG region forms a highly stable quadruplex with G-tetrads likely stabilised by K^+ ions bound between tetrad plains. However, we have focused on the conformation of the adenines which differ considerably in base alignment, stability and dynamics from those in previously reported structures of d(AGGGT)$_4$ and d(TAGGGT)$_4$. We show unambiguously that the adenines of d(TTAGGGT)$_4$ are involved in the formation of a relatively stable A-tetrad with well-defined glycosidic torsion angles (*anti*), hydrogen bonding network (adenine 6-NH_2–adenine N1) defined by interbase NOEs, and base stacking interactions with the neighbouring G-tetrad. All of these structural features are apparent from NOE data involving both exchangeable and non-exchangeable protons. Thus, context-dependent effects appear to play some role in dictating preferred conformation, stability and dynamics. The structure of d(TTAGGGT)$_4$ provides us with a model system for exploiting in the design of novel telomerase inhibitors that bind to and stabilise G-quadruplex structures.

Introduction

Nucleic acids can adopt a wide range of different conformations other than the well established A and B duplex forms.[1] Telomeric DNA,[2–5] gene promoter regions[6–7] and immunoglobulin switch regions[8] have been the focus of attention because they contain continuous repeats of G-rich sequences that have the ability to form G-quadruplex structures.[9–11] In addition, a number of proteins have been identified that exhibit specific recognition of G-quadruplexes[12–13] or promote G-quadruplex formation.[14] Of particular interest are the telomeric DNA sequences found at the ends of chromosomes. These contain tandem repeats of guanine-rich DNA sequences of several kilobases that protect the ends from recombination, nuclease degradation and end-to-end fusions.[2–3] Examples of such sequence repeats are TTAGGG, TTGGGG, TTTTGGGG, TTTAGGG and TTTTAGGG, which are found in telomeres of human, *Tetrahymena*, *Oxytricha*, *Arabidopsis*, *Chrorella* and *Chlamydomona*, respectively.[3] Telomeric sequences have the potential to form structures held together by guanine tetrads by either the intramolecular folding of a repetitive sequence, formation of a hairpin dimer or association of single strands to form a tetramer. Thus, G-quadruplex formation may be involved in capping the chromosome end with a structure resistant to nucleases or to association of chromosomes. Recent studies suggest that the G–quadruplex structure plays a role in interfering with telomerase action suggesting it as a potential therapeutic target for telomerase inhibition in cancer therapy.[15–16] The basic unit of the G-quadruplex is the G-tetrad (Fig. 1a), which consists of a square planar arrangement of guanines hydrogen-bonded through their Watson–Crick and Hoogsteen edges. The O6 carbonyl groups are directed towards the interior of the G-tetrad and require the presence of a monovalent cation such as Na^+ or K^+ to stabilise the structure.[17] The precise localisation of Na^+ is well defined in the high resolution crystal structure of d(TG$_4$T) were the bound metal ions are octahedrally co-ordinated between G-tetrad planes.[18] An NMR approach, involving direct observation of $^{15}NH_4^+$ bound to the quadruplex d(G$_4$T$_4$G$_4$)$_2$, located the cation binding site equidistant from each G tetrad.[19]

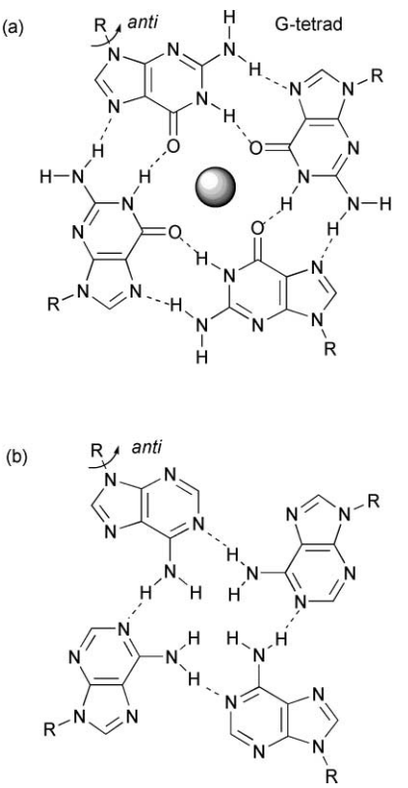

Fig. 1 Schematic drawing of the G-tetrad (a) with a bound K^+ in the central channel, and (b) the A-tetrad with the hydrogen bonding pattern dictated by NOE data.

The human telomeric sequence d[AGGG(TTAGGG)$_3$] forms an intramolecular quadruplex structure in Na^+ solution, with each strand having one antiparallel and one parallel neighbouring strand and *syn–syn–anti–anti* glycosidic torsion angles in each G-tetrad.[20] A more recent X-ray structure has shown that the same sequence is able to fold in the presence of K^+ to form a quite different parallel-stranded conformation with the

DOI: 10.1039/b300845m

Table 1 Chemical shift assignments for d(TTAGGGT)$_4$ at 298 K, pH 7.0

	H1′	H2′/H2″	H3′	H4′	H5′/H5″	H6/H8	H2/H5/CH$_3$	NH$_2$	NH
T1	6.00	2.10/2.34	4.64	4.00	3.65/3.65	7.41	1.67	—	nd
T2	6.25	2.06/2.34	4.74	4.06	3.93/3.93	7.33	1.78	—	nd
A3	6.28	2.86/2.92	5.10	4.44	4.16/4.10	8.43	8.09	7.10	—
G4	6.01	2.67/2.91	5.05	4.49	4.27/4.27	7.95	—	9.15/5.70	11.67
G5	6.03	2.66/2.74	5.04	4.51	4.30/4.30	7.79	—	9.10/5.90	11.27
G6	6.27	2.57/2.70	4.91	4.52	4.27/4.27	7.70	—	9.00/5.75	11.04
T7	6.07	2.17/2.19	4.49	4.23	4.07/4.07	7.36	1.63	—	nd

A–T-rich loops on the outside of the core G-tetrad region.[21] A number of parallel and anti-parallel stranded intermolecular G-quadruplex structures have now been studied in solution by a combined NMR and molecular dynamics approach.[22–26] These structures consist of right-handed helices with all the residues adopting *anti* glycosidic torsion angles and predominantly C2′-*endo* sugar pucker conformations. Each G-tetrad is well-defined, adopting a coplanar alignment with strong stacking between adjacent G-tetrads. However, the terminal thymine residues are more dynamic, sampling multiple conformations in solution with little evidence to suggest formation of a stable T-tetrad. Recently, a parallel-stranded solution structure of d(TGGTGGC)$_4$ in K^+solution[27] revealed stable T-tetrad formation when sandwiched between G-tetrads at the center of the quadruplex despite underwinding of the right-handed helix and poor stacking across the T4–G5 step.

The increasing appreciation that DNA quadruplexes may play an important and versatile role in biological processes has been illustrated by the formation of tetrads from other than simple G_n repeats, including mixed tetrads of the form G–C–G–C and A–T–A–T.[28,29] Stacking interactions with guanine tetrads by the neighbouring nucleotides in parallel-stranded quadruplexes has been explored with the NMR quadruplex structures of d(AGGGT)$_4$ and d(TAGGGT)$_4$ in the presence of K^+ ions.[30] This comparison study revealed that although the G_3 segments in the quadruplexes have largely similar structures, only the adenine residues in the d(AGGGT)$_4$ structure appear to form an A-tetrad and have good stacking with the adjacent G-tetrad. Surprisingly, thymine and adenine residues in the d(TAGGGT)$_4$ quadruplex structure appear to sample a range of conformations resulting in poor stacking.[30] Thus, formation of the tetrad alignment by other residues except guanines is highly context-dependent. We have carried out a full NMR-molecular dynamics analysis to determine the structure of the parallel-stranded quadruplex structure d(TTAGGGT)$_4$, which contains the full human telomeric repeat TTAGGG. The extra 3′-terminal thymine residue prevents the possible formation of aggregates in solution through end-to-end stacking between G-tetrads of different molecules. In contrast to previous studies, we show that the conformation of the adenines is well-defined forming significant stacking interactions with the adjacent G-tetrads consistent with the formation of a hydrogen bonded A-tetrad.

Results and discussion

Exchangeable protons assignments

Three distinct exchangeable guanine N1 imino protons are observed in H_2O spectra in the region 11–12 ppm; these persist at >320 K, and exchange slowly after dilution of the sample into D_2O solution, consistent with the high kinetic stability and low solvent accessibility previously identified in a number of NMR studies.[20,23,26,31–34] Resonance assignments for the exchangeable and non-exchangeable protons were based on 2D NOE data collected at various mixing times 70–400 ms, and from through-bond scalar coupling interactions observed in TOCSY and DQF-COSY spectra. Chemical shift assignments (Table 1) for all protons of d(TTAGGGT)$_4$ were obtained following standard assignment procedures; nucleotides are numbered sequentially from the 5′- to 3′-end, as d(T1–T2–A3–G4–G5–G6–T7).

Expanded regions (A) and (B) of the NOESY spectrum of d(TTAGGGT)$_4$ in H_2O solution, recorded at 300 ms mixing time and 293 K, are plotted in Fig. 2. NOEs between adjacent guanine imino protons in the sequence, G4NH–G5NH and G5NH–G6NH, are evident in Fig. 2A. In addition, guanine imino protons show NOEs with their own base protons and also to their 5′ flanking base protons in the A3–G4–G5–G6 part of the d(TTAGGGT)$_4$ quadruplex structure (Fig. 2B). The G6NH at 11.04 ppm exhibits NOEs to the base protons of G6H8 and G5H8, and the 11.27 ppm imino proton of the G5 exhibits NOEs with the base protons of G5H8 and G4H8. These NOEs are evidence of the interstrand interactions between the guanine residues that are involved in G-tetrad formation, as they cannot be accounted for by NOEs within the same strand. Importantly, we also observe NOEs from the base protons A3H8 and A3H2 to the imino proton of G4 at 11.67

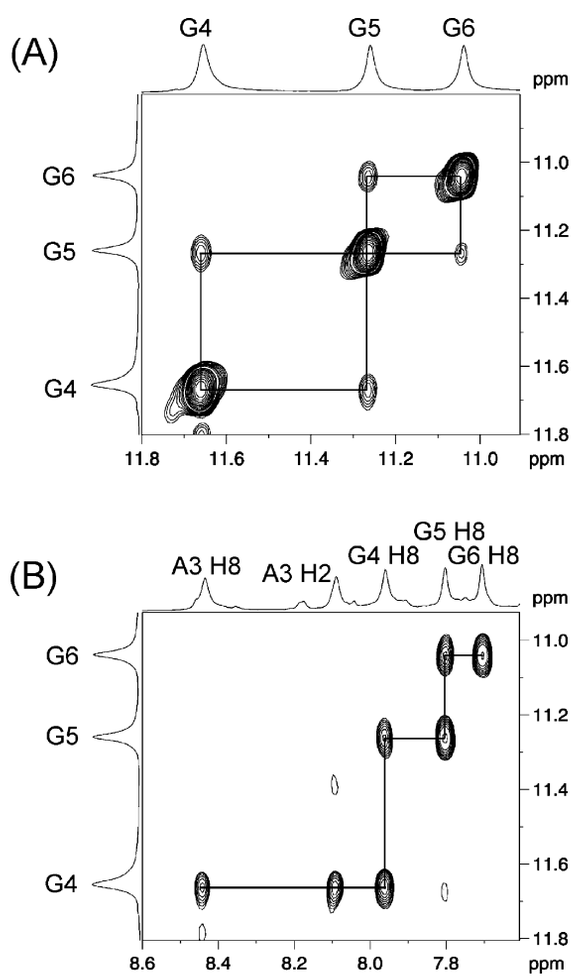

Fig. 2 Portions of the NOESY spectrum of d(TTAGGGT)$_4$ at 293 K in H_2O solution illustrating (A) NH–NH NOEs between adjacent G-tetrads involving hydrogen bonded G N1–H resonances, and (B) G N1–H NOEs to base H6/H8 illustrating base stacking between G and A-tetrads.

ppm, providing evidence for good stacking interactions across the A3pG4 step.

The observation of two distinct amino protons for each guanine is consistent with slow rotation about the C–NH_2 bond. One hydrogen of the amino group resonates near 6.0 ppm and is exposed to the solvent while the other hydrogen near 9.0 ppm (see Table 1) exhibits an NOE with the base H8 proton of a guanine within the same tetrad, confirming that G residues form G-tetrads with the hydrogen-bonding alignment shown in Fig. 1A. All of the assigned NOEs for $d(TTAGGGT)_4$ confirm that the nucleotides in the sequence form a symmetric right-handed twisted parallel-stranded quadruplex. A resonance for the 6-NH_2 group of A3 is also clearly identified in the 1D NMR spectrum at 7.1 ppm and exhibits an NOE to G4NH but also shows unambiguously an NOE with the base A3H2 (Fig. 3a). The latter NOE from A3NH_2 to A3H2 can only be attributed to an interaction between A residues within an A-tetrad and is consistent with a pattern of hydrogen bonding involving preferentially the N1 and 6-NH_2 of adjacent adenine bases (Fig. 1b). An averaged chemical shift is observed for the A3 6-NH_2 group at 7.1 ppm mid-way between the values observed for the non-equivalent guanine 2-NH_2 groups (see above). The data suggest that rotation about the C–NH_2 bond of adenine is fast on the chemical shift time scale. However, the A3 6-NH_2 resonance persists at high temperature indicative of some degree of protection from solvent through interactions within the A-tetrad. The thymine imino protons are in fast exchange with the solvent and are not observed even at low temperature. There is no direct evidence of hydrogen bond pairing of the thymine residues that would suggest the formation of a hydrogen-bonded T-tetrad by T1, T2 or T7. Thus, the NMR data suggests that only the purine residues of the sequence form tetrads in the quadruplex structure.

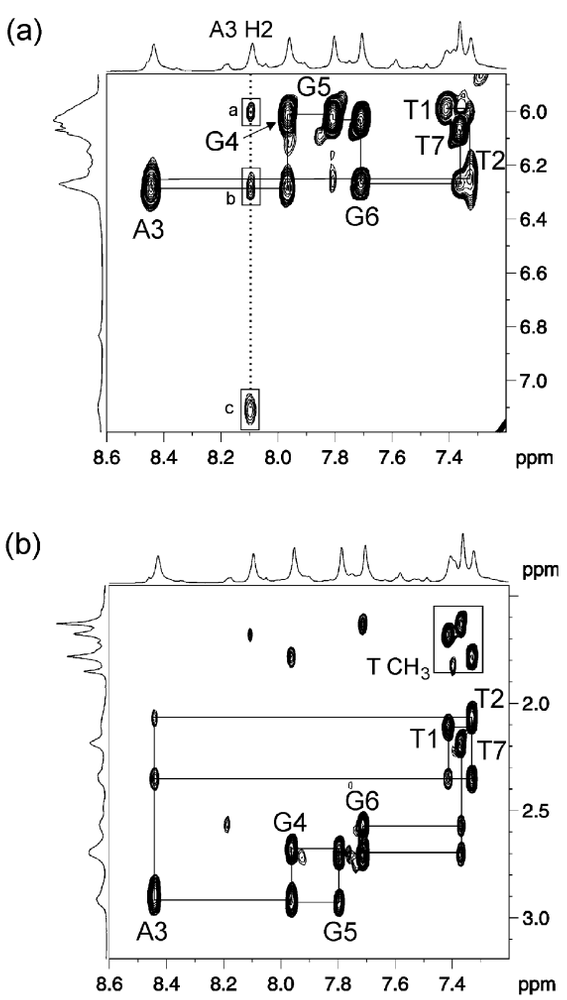

Fig. 3 (a) Portion of the NOESY spectrum illustrating sequential NOE connectivities between base H6 or H8 and deoxyribose H1′ resonances. Intranucleotide NOEs are labelled according to sequence; other NOEs are assigned as follows: a, A3H2–G4H1′; b, A3H2–A3H1′, and c, A3–NH_2–A3 H2. (b) Portion of the NOESY spectrum illustrating sequential NOE connectivities between base H6 or H8 and deoxyribose H2′/2″ resonances; intranucleotide NOEs are labelled according to the assignment of base H6 or H8.

Non exchangeable protons assignments

Expansion of the NOESY spectrum in Fig. 3a shows the pathway followed for the sequential assignment of the sugar H1′ protons and the base H8/H6 protons of each nucleotide in the $d(T1–T2–A3–G4–G5–G6–T7)_4$ quadruplex structure. All of the sequential connectivities from H8/H6 base proton to its 5′-flanking H1′ sugar proton are evident. Expansion of the NOESY spectrum in Fig. 3b shows NOE connectivities between the base H8/H6 protons and their own and their 5′-flanking H2′/H2″ sugar protons. The relative intensities of the NOE cross peaks between base H8/H6 protons to sugar protons H1′, H2′, H2″, H3′ are similar not only for the core G4, G5, G6, but also for the T1, T2, A3 and T7 residues in the $d(TTAGGGT)_4$ quadruplex structure. This suggests that the nucleotides along the four strands in the symmetrical quadruplex structure are in close proximity to their adjacent nucleotides on the same strand and adopt a B-DNA-like twisted helical geometry. The non-exchangeable NOE assignments show clearly that the T1, T2 and T7 nucleotides are not randomly oriented, even though it is unlikely that they hydrogen bond with their interstrand neighbouring nucleotides, but they are involved in stacking with neighbouring nucleotides within the same strand. Whether these interactions contribute to the overall stability of the quadruplex structure is unclear. In addition, all the sequential assignment pathways observed in NOESY spectra follow only one direction from the base to the 5′ flanking sugar H1′, H2′/H2″ and H3′ protons, which indicates that each individual strand has the right-handed helical alignment with coupling constants consistent with S-type sugar conformations.[35]

The intensity of intranucleotide NOEs between base H6 or H8 and sugar H1′ that define glycosidic torsion angles were quantified in a short mixing time (70 ms) NOESY spectrum and were found to be weaker than reference NOEs between base H6 and CH_3 (3.0 Å) in any of the thymine residues. All of the NOE restraints determined for the base H6/H8–H1′ sugar protons have a distance >3.0 Å; *anti* and *syn* glycosidic conformations are differentiated by distances of 3.7 Å and 2.5 Å, respectively, between the H8 and H1′ protons, indicating that all residues in the $d(TTAGGGT)_4$ quadruplex adopt *anti* glycosidic torsion angles. This is in agreement with earlier conclusions from studies of intermolecular G-quadruplex structures, except for $d(AGGGT)_4$ where the terminal A has been proposed to adopt a *syn* conformation.[30] The data presented here for $d(TTAGGGT)_4$ are not consistent with the adenines adopting the *syn* conformation.

Thermodynamic stability of $d(TTAGGGT)_4$

Intrastrand NOEs between adjacent nucleotides outside the AGGG core of the quadruplex show some evidence for stacking of Ts with neighbouring nucleotides within the same strand. Whether these interactions, and even A-tetrad stacking interactions, contribute to the overall stability of the quadruplex structure can be assessed from thermodynamic analysis of the quadruplex melting transition. The melting temperature T_m of $d(TTAGGGT)_4$ (single stranded and quadruplex equally populated) is ~60 °C at a strand concentration of 6.4 mM, comparable to that for $d(AGGGT)_4$ at a similar concentration.[30] The presence of resonances from both the single-stranded and quadruplex forms of the d(TTAGGGT) in the same spectrum enables the change in the population of the folded and unfolded

states of the quadruplex d(TTAGGGT)$_4$ to be determined as a function of temperature. The aromatic peaks of the A3 and T7 residues were used for calculating the populations giving very similar estimates. The free energy of association $\Delta G°$ was calculated at each temperature based on relative populations and the temperature-dependence of $\Delta G°$ used to derive values for $\Delta H°$ and $\Delta S°$. Data points in the range 278–338 K gave a linear correlation (R=0.998) indicating that $\Delta H°$ is independent of temperature (ΔC_p= 0). At 298 K we determined the following thermodynamic parameters for the formation of d(TTAGGGT)$_4$: $\Delta H° = -256$ kJ mol^{-1}, $\Delta S° = -763$ J K^{-1}mol^{-1}, $\Delta G° = -28.9$ kJ mol^{-1}. These values agree very well with previous calorimetric experiments with the parallel four-stranded quadruplex d(TGGGT)$_4$ in K^+ solution at a similar ionic strength, which also has a core structure of three G-tetrads ($\Delta H° = -263$ kJ mol^{-1}, $\Delta S° = -777$ J K^{-1}mol^{-1}, $\Delta G° = -28.9$ kJ mol^{-1}).[36] The data suggest that formation of the core G-quadruplex, stabilised by bound K^+ ions, probably provides the majority of the enthalpic driving force and that additional interactions from the flanking non-G nucleotides make a very modest energetic contribution.

Solution structure of d(TTAGGGT)$_4$

The average minimised structure of d(TTAGGGT)$_4$ calculated from the last 500 ps of a 1 ns restrained MD simulation, guided by 178 NOE restraints per strand, is shown in Fig. 4. The right-handed parallel quadruplex structure shows the guanine residues to have an almost planar conformation and all form a well-defined G-tetrad conformation. Although the NMR refinement did not use any hydrogen bond restraints between the adenine residues on adjacent strands, the structure shows clearly the formation of a planar hydrogen bonded tetrad with the base alignment dictated by specific NOE interactions. In contrast, the thymine bases do not form any specific hydrogen bonded alignment. These general features of the structure agree with our NMR evidence and emphasize that the purine residues in the structure are able to form tetrads, stack on top of each other and stabilize the quadruplex structure, while the thymine residues have a more flexible conformation.

The sequence-dependent variation of helical parameters for d(TTAGGGT)$_4$ have been calculated over the last 500 structures (500 ps) of the restrained MD simulation using the CURVES program (Table 2).[37] The overall four-fold symmetry is clearly apparent. The glycosidic torsion angles χ are in the *anti* conformation for all the nucleotides including A3 (centered on −119°), in agreement with previous studies of parallel-stranded structures.[23,24] However, the data for the A3 tetrad contrasts with that of Patel *et al.*[30] for d(AGGGT)$_4$, where the adenine was shown to be in the *syn* conformation on the basis of quantitative NOE calculations. These differences suggest an additional role for the 5′-terminal thymines in modulating nucleotide conformation and dynamics. Hoogsteen N1–O6 and N2–N7 hydrogen bonds between the guanine bases of the three G-tetrads are stable in all of the structures. The hydrogen bonds formed by the N1–N6 atoms of the adenine bases have a little more flexibility, but they still are well defined in all calculated structures, though no specific H-bond restraints were included. The larger RMSD in hydrogen bond distances within the A-tetrads compared with the G-tetrads (3.12 ± 0.16 *versus* 2.96 ± 0.06 Å) are consistent with larger amplitude breathing motions that may account for the reduced barrier to rotation about the C–NH_2 bond of adenine such as to result in the averaging of the amino proton chemical shifts. The thymines do not form any recognisable hydrogen bonding alignment during the MD simulation. The AGGG core of the quadruplex is very stable maintaining the channel through the centre of the structure which is occupied by K^+ ions binding between neighbouring G-tetrads (Fig. 5). Metal ions were placed in these positions during the initial structural modelling and remain tightly bound throughout the MD simulations. Although the exact number of potassium ions in the quadruplex structure has not been addressed experimentally in this study, others have established the K^+binding stochiometry, consistent with this model.[32]

The pseudorotation phase angle P that defines the sugar conformation shows that the S-type geometry is prefered in all cases with a mean value of 132°, although some variations amongst residues are observed. This value is rather smaller than mean values observed for the canonical B-DNA conformation. The purine bases of the AGGG segment have axial rise values of 3.0 Å, 3.5 Å and 3.7 Å. The larger values for the latter G–G steps reflect the increased separation resulting from metal binding between G-tetrads. The helical twist of 25°, 29°, 26° for the A3–G4, G4–G5, G5–G6 base steps, show that the helical structure is more unwound than a canonical B-DNA duplex structure (~36°), but is in the range of helical twist observed for guanine residues in other G-tetrad structures (25–30°).[23–26] Fig. 6 shows the stacking of the tetrads at the A3–G4 (a), G4–G5 (b) and G5–G6 (c) steps and the conformation of the A3-tetrad (d) in the average minimised structure (Fig. 5). Generally, base

Table 2 Helical parameters from CURVES analysis of 500 structures taken at 1 ps intervals over the final 500 ps of the NOE-restrained MD simulation of d(TTAGGGT)$_4$. Average value and standard deviation (parentheses) over all four strands for helical twist, axial rise for the d(AGGG)$_4$ segment, glycosidic torsion angle χ, and pseudorotation phase angle P for all nucleotides of d(TTAGGGT)$_4$

	Twist/°	Rise/Å
A3–G4	25 (3)	3.0 (0.3)
G4–G5	29 (2)	3.5 (0.2)
G5–G6	26 (2)	3.7 (0.2)
	χ angle/°	Phase P/°
T1	−146 (17)	140 (19)
T2	−136 (14)	100 (14)
A3	−119 (10)	133 (21)
G4	−117 (9)	134 (10)
G5	−111 (9)	143 (12)
G6	−115 (10)	145 (9)
T7	−109 (15)	127 (24)

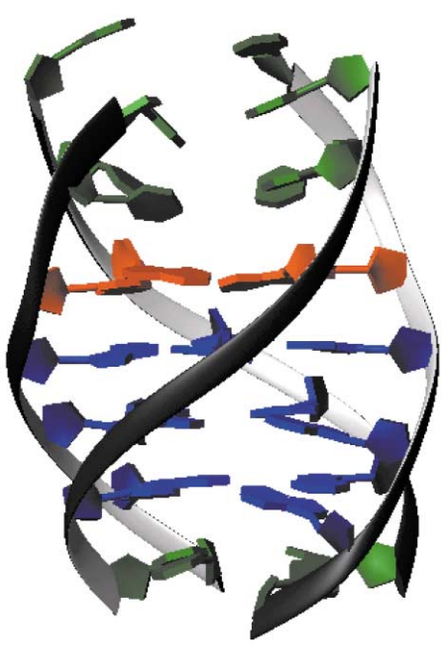

Fig. 4 Average NOE-restrained energy minimised structure of d(TTAGGGT)$_4$. The core AGGG sequence forms a well-ordered structure but the 5′- and 3′-terminal Ts show no evidence for stable T-tetrad formation.

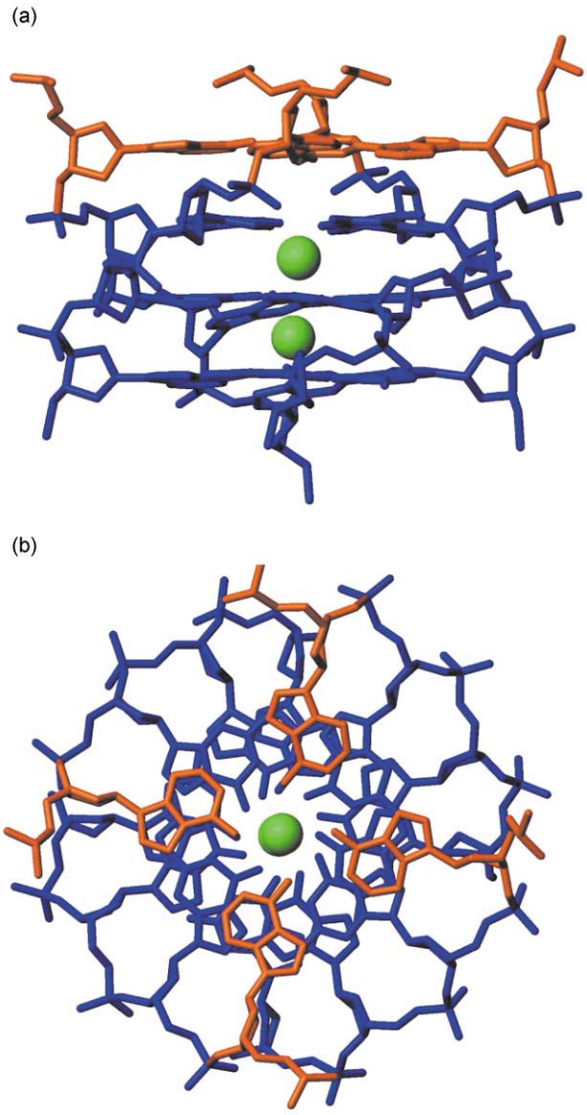

Fig. 5 (a) Structure of the AGGG core showing the location of bound K^+ ions equidistant between the plains of the G-tetrads; view from above (b) showing the K^+ channel and the helical twist.

stacking is observed between purine residues on the same strand, as in the B-DNA structure. The guanine nucleotides of the G5-tetrad show the best overlap with the G4- and G6-tetrads and this may explain why the G5NH resonance exchanges more slowly than the imino protons of the G4 and G6 nucleotides when dissolved in D_2O solution.

The conformation of the A-tetrad, in terms of the hydrogen bonding arrangement and stacking interactions, is well-defined by the NOE data. In particular, the A3NH_2–A3H2 NOE can only be rationalised as an interaction between adenine bases within an A-tetrad conformation in which the A3 6-NH_2 is hydrogen bonded to the N1 of the adjacent adenine base. The formation of an A-tetrad has been reported previously only for d(AGGGT)$_4$.[30] Patel *et al.*[30] suggested that A-tetrads can have two distinct patterns of base alignment distinguished by the observation of NOEs between ANH_2–AH2 or ANH_2–AH8. Observation of both NOEs suggested a dynamic behaviour of the A-tetrad exchanging between the two possible hydrogen-bonding arrangements. Exchange between the two alignments was also evident in their structure calculations, resulting in two different base stacking patterns with the adjacent G-tetrad perhaps reflecting the lack of conformational preference in the absence of other 5′-capping nucleotides. Our studies of d(TTAGGGT)$_4$ support the preferential formation of the A-tetrad alignment shown in Fig. 1b. The 6–NH_2 groups of the adenines point into the central cavity of the quadruplex structure in a similar way to that observed for the O6 atoms of the guanines. Base stacking across the A3–G4 step shows partial stacking of the six-membered ring of the adenine with the five-membered ring of the guanine nucleotide of the same strand. This differs from that reported for d(AGGGT)$_4$ as a consequence of the difference in glycosidic torsion angles of the adenine residues in these two structures.[30] Thus, context-dependent effects appear to partially dictate whether adenine nucleotides have a *syn* conformation, as observed for d(AGGGT)$_4$,[30] or the *anti* conformation observed for d(TAGGGT)$_4$[30] and in this study of d(TTAGGGT)$_4$. On-going studies strongly support context-dependent effects on A-tetrad stability and conformation, and a potential role for metal ion binding in A-tetrad stabilisation. The structure of d(TTAGGGT)$_4$ also provides us with a model system for exploiting in the design of novel telomerase inhibitors. Such work has recently been reported and is on-going.[37]

Materials and methods

DNA samples

The oligonucleotide d(TTAGGGT) was synthesized and purified as previously described,[38] and shown to be >95% pure by ^{1}H NMR spectroscopy. The NMR sample of d(TTAGGGT) was prepared at a concentration of single strand of 6.4 mM (or 1.6 mM in quadruplex) with a final salt concentration of 100 mM KCl and 10 mM K_2HPO_4.

NMR experiments and NOE restraints

NMR data were collected on a Bruker DRX500 spectrometer equipped with an broad-band inverse-detection probe with z-field gradients. Data were processed and integrated using Bruker XWINNMR software. 1D NMR spectra were collected as 16384 data points with a recycle delay of 1.5 s. Standard phase-sensitive 2D NMR pulse sequences were used to record NOESY, TOCSY and DQF-COSY spectra with spectral widths of 20 ppm at temperatures in the range 5–65 °C and at NOESY mixing times of 300 ms, 200 ms, 150 ms, 100 ms and 70 ms. Typically 2048 data points were collected in t_2 and 512 increments in t_1 each of 64 transients with a recycle delay of 1.5 s. In all cases solvent suppression was achieved using the WATERGATE sequence.[39] 2D data sets were zero-filled to 4K × 2K prior to Fourier transformation and apodised with a shifted sinebell squared window function. Interproton distances were derived from integration of NOE cross peak volumes in 150 ms, 100 ms and 75 ms NOESY data sets in D_2O and from 200 ms and 100 ms data in H_2O solutions. Distances were determined by extrapolation to zero mixing time to account for the possible effects of spin-diffusion using the method of Baleja *et al.*,[40] using the thymine H6–CH_3 reference distance (3.0 Å) for NOEs involving base protons, and the sugar H2′–H2″ fixed distance (1.85 Å) for NOEs involving only sugar protons. For well resolved non-exchangeable cross peaks the distances were given upper and lower error bounds of 15% of the calculated distance, while for the exchangeable cross peaks 25% was used. Hydrogen-bond restraints were included for atoms involved in the ideal hydrogen-bonding geometry of the G-tetrad. The heavy atom–heavy atom distance restraints for O6–N6 and N7–N7 distances were set to 2.85 ± 0.10 Å. Distance restraints were checked for large geometrical inconsistencies by comparing visually with the distances of the energy-minimised quadruplex structure derived from unrestrained molecular dynamics simulation using MolMol software.[41] In order to distinguish ambiguities in NOE restraints between interstrand and intrastrand NOE interactions resulting from the symmetry of the quadruplex structure the energy-minimised structure was considered. Interstrand restraints were manually determined by matching their calculated distance with the distances of the

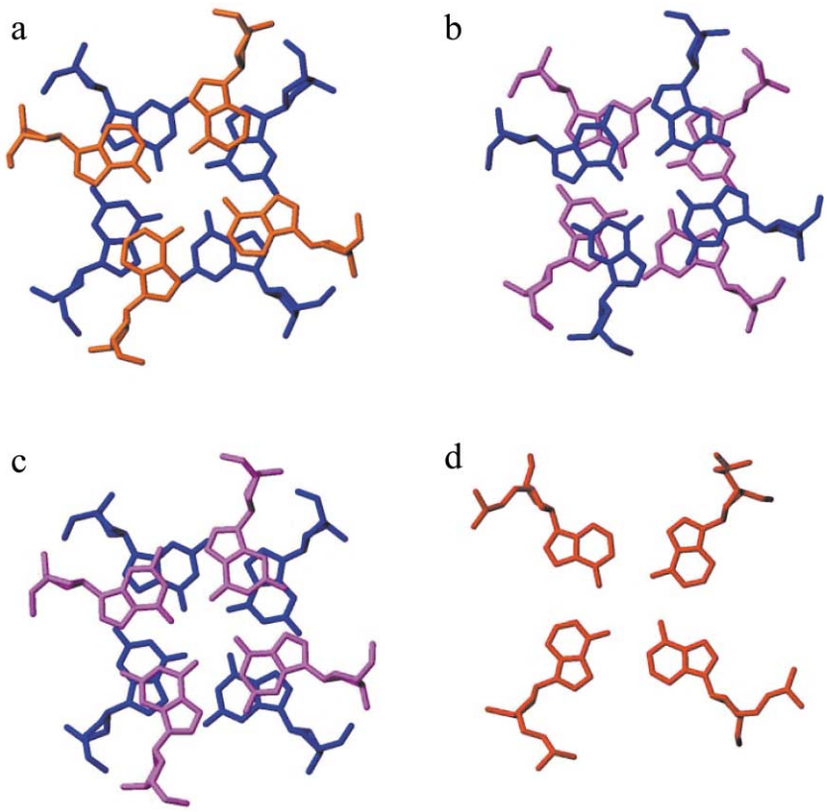

Fig. 6 Base-tetrad stacking interactions with neighbouring tetrads taken from the averaged energy-minimised structure: (a) A3–G4, (b) G4–G5, (c) G5–G6; base arrangement in the isolated A-tetrad (d).

initial energy-minimised structure. Those distance restraints that included violations over 0.5 Å were adjusted after restrained energy minimisation was carried out to derive a set of restraints that were consistent with the geometrical limits of the structure. A total set of 728 restraints was used for energy minimizations and restrained molecular dynamics simulations.

Thermodynamic stability

The concentrations of the single stranded and four stranded structures of d(TTAGGGT) were estimated from integrals of the slowly exchanging resonances in the 1D ^{1}H NMR spectra at various temperatures (278–338 K). The uncertainties in the integrals where estimated to be <10%. The equilibrium constant K_{eq}, is expressed as:

$$K_{eq} = [\text{quadruplex}]/[\text{single strand}]^4 = (f_Q c_t/4)/(f_S c_t)^4$$

where f_Q and f_S are the fractions of quadruplex and single strand, respectively, and c_t is the total concentration of oligonucleotide in terms of single strands. The temperature-dependence of the free energy of association $\Delta G^\circ = -RT \ln(K_{eq})$ gives a linear plot that was fitted to $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ to determine the thermodynamic parameters for melting.

Structure calculations

Energy minimisations and restrained molecular dynamics calculations were performed on an Origin 200 Silicon Graphics Server using the AMBER 6 suite of programs[42] employing the AMBER 94 force field with modifications[43] and the Particle Mesh Ewald (PME) method[44] for the treatment of long-range electrostatics. The initial co-ordinates for the starting model of the quadruplex were taken from the NMR structure of the d(TTGGGGT)$_4$ quadruplex[24] (Protein Data Bank accession entry number 139d), choosing randomly one of the deposited structures. The starting model of the d(TTAGGGT) quadruplex was generated by replacing the guanine nucleotides in the third position of the sequence from the 5′ end with adenine nucleotides using the LEAP module of AMBER 6. The DNA molecule was solvated in a periodic TIP3 water box of approximate dimensions 60 Å × 60 Å × 60 Å, which extended to a distance of 10 Å from any solute atom and contained 5030 water molecules. Two internal potassium ions were manually positioned in the central channel between adjacent G-quartets using standard parameters for the AMBER force field. The potassium ions were positioned equidistant from two adjacent G-quartets to allow octahedral coordination with guanine carbonyl oxygen O6 atoms. The quadruplex system was neutralised externally requiring 22 potassium ions placed at the most negative locations using Coloumbic potential terms with the LEAP module. All the potassium ions including those placed in the central channel were treated as part of the solvent. Energy minimisations and restrained molecular dynamics were carried out using the SANDER module of AMBER 6. Calculations with SANDER were performed with a 2 fs time step, with the SHAKE algorithm (tolerance 0.00005 Å) applied to all bonds to remove bond stretching, and a 9 Å cut off to the Lennard Jones interactions. The restrained molecular dynamics were performed at 300 K and a constant pressure of 1.0 atm with isotropic position scaling utilising the Berendsen algorithm for temperature coupling. Translational and rotational motions were removed every 100 fs. All calculations were carried out with the PME method using a 9 Å cut-off for direct space non-bonded calculations and a 0.00001 Ewald convergence tolerance for the inclusion of long-range electrostatics in our calculations.

The quadruplex system was allowed to equilibrate fully before beginning the molecular dynamics calculations. Minimisation was performed with 50 steps of steepest descent and 5000 steps of conjugate gradient to first the water and counterions, with the DNA coordinates frozen, followed by a further 5000 steps on all the components of the system. Next, 10 ps unrestrained molecular dynamics were run at 100K on the water alone with the DNA and potassium ions constrained,

followed for another 10 ps to allow the potassium ions to move. In the following 5 ps of dynamics the temperature of the system was increased from 100 K to 300 K. In the next runs, each of them of 10 ps dynamics, the DNA force constant is gradually reduced from 100 to 50, 25, 10, 5 and 2.5 kcal mol^{-1} $Å^{-2}$. The equilibration step ends with 100 ps of dynamics on the whole fully unrestrained system. The system now is fully equilibrated and NOE restraints can be applied to the quadruplex system. Distance restraints were introduced gradually on the system over the first 10 ps of 100 ps MD run with the temperature stable at 300K and PME on. All NOE restraints were introduced in the form of square well potentials with a force constant of 50 kcal mol^{-1} $Å^{-1}$ for the hydrogen-bond restraints and 30 kcal mol^{-1} $Å^{-1}$ for all the other NOE distance restraints. A total of 1000 ps simulation was performed under the same conditions, after which the system was energy minimized also with NOE restraints. Calculated structures satisfied the majority of the NOE restraints from the set of 728 restraints. The average minimised structure had no restraint violation >0.3 Å, and only 4 violations >0.2 Å with a total restraint violation energy penalty of 33 kcal mol^{-1}. The RMS restraint violation over all restraints was 0.03 ± 0.07 Å. Time-averaged structures were calculated with the CARNAL module of AMBER. Helical structural properties, sugar pucker and backbone torsion angles have been analysed using CURVES.[36] Helical parameters were measured individually for each strand of the G-quadruplex and the four strands were found to have very similar values. The co-ordinates of the averaged energy minimised structure and an ensemble of 10 NMR structures have been deposited in the PDB (code: 1NP9; rcsb code: rcsb018075). The 10 structures show an RMSD to the mean structure of 1.62(±0.22) Å, with an RMS deviation from ideal covalent bond lengths of 0.011 Å, and from ideal bond angles of 3.1°.

Acknowledgments

We thank the EPSRC of the UK and AstraZeneca for financial support. NMR facilities in the School of Chemistry were funded by the EPSRC. We are grateful to John Keyte in the School of Biomedical Sciences for oligonucleotide synthesis.

References

1 S. Neidle, *Oxford Handbook of Nucleic Acid Structure*, 1998, Oxford University Press, New York.
2 E. H. Blackburn, *Nature*, 1991, **350**, 569–573.
3 V. A. Zakian, *Science*, 1995, **270**, 1601–1607.
4 C. W. Greider and E. H. Blackburn, *Sci. Am.*, 1996, **274**, 92–97.
5 D. E. Shippen, *Curr. Opin. Genet. Dev.*, 1993, **3**, 759–763.
6 T. Simonsson, P. Pecinka and M. Kubista, *Nucleic Acids Res.*, 1998, **26**, 1167–1172.
7 K. J. Woodford, R. M. Howell and K. Usdin, *J. Biol. Chem.*, 1994, **269**, 27029–27035.
8 A. Shimizu and T. Honjo, *Cell*, 1984, **36**, 801–803.
9 J. Feigon, K. M. Koshlap and F. W. Smith, *Nuclear Magn. Reson. Nucleic Acids*, 1995, **261**, 225–255.
10 R. H. Shafer, *Prog. Nucleic Acid Res. Mol. Biol.*, 1998, **59**, 55–94.
11 D. J. Patel, S. Bouaziz, A. Kettani, Y. Wang, 1998 *Oxford Handbook of Nucleic Acid Structure*, p. 389–453, Oxford University Press, New York.
12 K. Walsh and A. Gualberto, *J. Biol. Chem.*, 1992, **267**, 13714–13718.
13 R. Erlitzki and M. Fry, *J. Biol. Chem.*, 1997, **272**, 15881–15890.
14 W. I. Sundquist and S. Heaphy, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 3393–3397.
15 E. Raymond, J. C. Soria, E. Izbicka, F. Boussin, L. Hurley and D. D. Von Hoff, *Invest. New Drugs*, 2000, **18**, 123–137.
16 H. Han, R. J. Bennett and L. H. Hurley, *Clin. Cancer Res.*, 1999, **5**, 607.
17 J. R. Williamson, *Curr. Opin. Struct. Biol.*, 1993, **3**, 357–362.
18 K. Phillips, Z. Dauter, A. I. H. Murchie, D. M. J. Lilley and B. Luisi, *J. Mol. Biol.*, 1997, **273**, 171–182.
19 N. V. Hud, V. Sklenar and J. Feigon, *J. Mol. Biol.*, 1999, **286**, 651–660.
20 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
21 G. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
22 Y. Wang and D. J. Patel, *Structure*, 1994, **2**, 1141–1156.
23 Y. Wang and D. J. Patel, *J. Mol. Biol.*, 1993, **234**, 1171–1183.
24 V. M. Marathias and P. H. Bolton, *Nucleic Acids. Res.*, 2000, **28**, 1967–1977.
25 G. Gupta, A. E. Garcia, Q. Guo, M. Lu and N. R. Kallenbach, *Biochemistry*, 1993, **32**, 7098–7103.
26 F. Aboulela, A. I. H. Murchie, D. G. Norman and D. M. J. Lilley, *J. Mol. Biol.*, 1994, **243**, 458–471.
27 P. K. Patel and R. V. Hosur, *Nucleic Acids Res.*, 1999, **27**, 2457–2464.
28 N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin and D. J. Patel, *J. Mol. Biol.*, 2001, **312**, 1073–1088.
29 A. Kettani, A. Kumar and D. J. Patel, *J. Mol. Biol.*, 1995, **254**, 638–656.
30 P. K. Patel, A. S. R. Koti and R. V. Hosur, *Nucleic Acids Res.*, 1999, **27**, 3836–3843.
31 Y. Wang and D. J. Patel, *Biochemistry*, 1992, **31**, 8112–8119.
32 Y. Wang, C. Delossantos, X. L. Gao, K. Greene, D. Live and D. J. Patel, *J. Mol. Biol.*, 1991, **222**, 819–832.
33 R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc. Natl. Acad. Sci, U. S. A,*, 1993, **90**, 3745–3749.
34 P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Structure*, 1994, **2**, 221–233.
35 V. J. Wijk, B. D. Huckriede, J. H. Ippel and C. Altona, *Methods Enzymol.*, 1992, **211**, 286–306.
36 R. Z. Jin, B. L. Gaffney, C. Wang, R. A. Jones and K. J. Breslauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 8832–8836.
37 R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.*, 1989, **6**, 655–667.
38 E. Gavathiotis, R. Heald, M. F. G. Stevens and M. S. Searle, *Angew. Chem.*, 2001, **40**, 4749–4751.
39 M. Piotto, V. Saudek and V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661–665.
40 J. D. Baleja, J. Moult and B. D. Sykes, *J. Magn. Res.*, 1990, **87**, 375–384.
41 R. Koradi, M. Billeter and K. Wuthrich, *J. Mol. Graphics*, 1996, **14**, 51–55.
42 D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. L. Darden, K. M. Marz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Dingh, P. K. Weinerand P. A. Kollman, 1999, AMBER 6, University of California, San Francisco.
43 T. E. Cheatham, P. Cieplak and P. A. Kollman, *J. Biomol. Struct. Dyn.*, 1999, **16**, 845–862.
44 T. E. Cheatham, J. L. Miller, T. Fox, T. A. Darden and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 4193–4194.