© 1992 Oxford University Press

Nucleic Acids Research, Vol. 20, No. 15 4061–4067

Hairpin and parallel quartet structures for telomeric sequences

P.Balagurumoorthy, Samir K.Brahmachari*, Debasisa Mohanty, Manju Bansal and V.Sasisekharan
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Received March 19, 1992; Revised and Accepted June 10, 1992

ABSTRACT

The role of thymine residues in the formation of G-quartet structures for telomeric sequences has been investigated using model oligonucleotides of the type $d(G_4T_nG_4)$, with n = 1–4. Sequences $d(G_4T_3G_4)$ and $d(G_4T_4G_4)$ adopt a G-quartet structure formed by hairpin dimerization in 70 mM NaCl as judged by a characteristic circular dichroism signature with a 295 nm positive and 265 nm negative bands while $d(G_4TG_4)$ adopts a parallel G-quartet structure like $d(G_{12})$ which exhibits a strong positive band at 260 nm and a negative band at 240 nm. The sequence $d(G_4T_2G_4)$ exhibits a mixture of both conformations. The stability of hairpin G-quartet structures decreases with decrease in the number of intervening thymine residues. Potassium permanganate, a single strand specific probe has been used to establish the presence of loops composed of T residues in the hairpin G quartet structures formed by the oligonucleotides $d(G_4T_nG_4)$ with n = 2 – 4 in 70 mM NaCl. The formation of hairpin G quartet structure for the above sequences is further supported by the enhanced electrophoretic mobility observed on non-denaturing polyacrylamide gels. Human telomeric sequence $d(TTAGGG)_4$ which showed enhanced electrophoretic mobility like *Tetrahymena* telomeric sequence $d(T_2G_4)_4$ also exhibited a characteristic CD spectrum for a folded-back G-quartet structure. A detailed model for G-quartet structure involving hairpin dimer with alternating *syn-anti-syn-anti* conformation for the guanine residues both along the chain as well as around the G tetrad with at least two thymine residues in the loop is proposed. Intermolecular association of short telomeric sequences reported here provides a possible model for chromosomal pairing.

INTRODUCTION

Telomeres, the linear ends of eukaryotic chromosomes have repeat sequences consisting of runs of G residues, interspersed with short stretches of T's or A's which extend at the 3′ end leading to a single stranded overhang. In *Tetrahymena*, *Oxytricha* and *Stylonychia*, $d(T_nG_4)$ repeats (with n = 2 or 4) are observed (1–3), whereas in mouse, human and *Arabidopsis* $d(T_nAG_3)$ (n = 2 or 3) repeats are present (4–6). The presence of guanine tracts containing 3–4 guanine residues is a highly conserved feature of telomeres and suggests the existence of a common structural element which may be involved in telomere function (7–10). Models involving guanine tetrads in a square planar array have been proposed recently for the telomeric sequences, either with a four stranded parallel arrangement of the phosphodiester chains or folded-back hairpin quartet structures with a four stranded guanine stem and thymine loop (9,10,11–13). Although several hydrogen bonding schemes are possible for a G:G base pair, the tetrad structure is restricted to Hoogsteen type pairing (14,15) as has been proposed for poly(G) on the basis of fibre diffraction data (16,17). Recent studies have also indicated that G-rich telomeric sequences self-associate to form a parallel four stranded structure, termed G4-DNA (9,10), while a tetraplex structure has been suggested for nonameric synthetic oligonucleotides of the type d(GGTTXTTGG), (where X = A,C,G or T) from spectroscopic and calorimetric studies (18). Gel electrophoretic mobility, and methylation protection studies on $d(T_4G_4)_4$ and $d(T_2G_4)_4$ sequences and dimethyl sulfate methylation experiments on oligo d(G) suggested folded-back G-quartet structures (11,12,19). Since no detailed models have been reported for the parallel or folded-back hairpin structures, we have carried out spectroscopic, gel electrophoretic, chemical foot-printing and model building studies for a series of synthetic oligonucleotides of the type $d(G_4T_nG_4)$ (where n=1 to 4). These studies enabled us to characterize the role of thymine residues in stabilizing parallel or folded hairpin structures for the telomeric sequences. We observe that the sequences $d(G_{12})$ and $d(G_4TG_4)$ adopt parallel four stranded structures, whereas the sequences $d(G_4T_nG_4)$ with $n \geq 2$ form folded-back quartet structures similar to that of $d(TTAGGG)_4$ with a characteristic CD spectrum. Our model building and molecular mechanics studies indicate that folded-back G-quartet structures are energetically more favourable than parallel tetraplex structures and can be formed not only by $d(G_4T_3G_4)$ and $d(G_4T_4G_4)$, but also by $d(G_4T_2G_4)$.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Preparation of Oligonucleotides

Oligonucleotides were synthesized on Pharmacia automated DNA synthesizer Gene Assembler Plus employing β-cyanoethyl phosphoramidite chemistry. The oligonucleotides were purified on 20% polyacrylamide gel containing 8 M urea. The gel was uv shadowed and the desired oligonucleotide bands were eluted with sterile water. Oligonucleotides were completely desalted by passing twice through Pharmacia NAP columns equilibrated with sterile water. The oligonucleotide $d(G_4T_4G_4)$ was also purified on Sep-Pak C18 (Waters) cartridge using 50% aetonitrile in water as eluant and found to exhibit CD spectrum which is identical with that of oligonucleotide desalted on NAP column.

Circular Dichroism Measurements and Analysis of CD Melting Profiles

CD spectra were recorded on JASCO-J 500 A spectropolarimeter equipped with a data processor DP 501. Samples were prepared by heating at 95°C for three minutes in 20 OD_{260} strand concentration, followed by gradual cooling to room temperature. Prior to CD scan samples were diluted to give strand concentrations of the order of 2 OD_{260} and equilibrated in the CD cell holder for 30 minutes. In CD melting studies diluted samples were equilibrated at room temperature for several hours to obtain equilibrium spectra. Then the samples were cooled to low temperature (5°C) and the CD spectra were recorded after equilibrating the sample for 20 minutes at each temperature.

The monophasic cooperative melting transitions are assumed to reflect helix-coil equilibrium. Assuming molecularity of 2 (since a folded-back hairpin quartet involves two strands) equilibrium constants at different temperatures were calculated from the melting profiles using a two state model(20,21). Thermodynamic parameters were obtained from van't Hoff plots.

Non-denaturing Gel Electrophoresis Experiments

Gel purified oligonucleotides were 5′ end labelled with γ-^{32}P ATP and polynucleotide kinase. Labelled oligonucleotides of appropriate counts were mixed with the corresponding unlabelled oligonucleotide to give final concentration of 2 OD_{260} units in 20 μl volume. The samples were dried in a speedvac and redissolved in 20 μl of 2 mM cacodylate, 0.1 mM EDTA, pH7.2 and 70 mM NaCl. The oligonucleotide solutions were heat denatured at 95°C and slowly cooled to room temperature. Samples were immediately loaded onto a 12% polyacrylamide gel preequilibrated at the indicated temperatures for an hour. The gel as well as the running buffer (0.5×TBE) contained 70 mM NaCl. The gels were run for 10−14 hours at a constant voltage of 8 volts per cm until the bromophenol blue reached 19 cm from the well. The gel was maintained at the indicated temperatures during the course of the experiment using a Lauda circulating waterbath.

$KMnO_4$ Probing

Labelled oligonucleotides were annealed as mentioned above in 10 mM Tris, 0.1 mM EDTA, pH 7.4, containing 70 mM NaCl in 20 μl volume at a DNA concentration of 2 OD_{260}. To this 1 μl 5 mM $KMnO_4$ solution, diluted from freshly prepared 100 mM solution was added. After incubating the reaction mixture for 5 minutes at room temperature the reaction was stopped by adding 2 μl allyl alcohol. Samples were ethanol precipitated along with 6 μg of *E.coli* tRNA and the pellets were dried in a speedvac. The sample was treated with 10% piperidine at 90°C for 30 minutes. Samples were repeatedly dissolved in water and dried to remove piperidine completely. The dried solids were dissolved in 90% formamide dye containing bromophenol blue and xylene cyanol markers, heat denatured at 95°C for 3 minutes, quick-chilled on ice and electrophoresed on 20% denaturing gel containing 8 M urea in 1× TBE buffer.

Model Building Studies

The preliminary models were built using DTMM software package (22) on an IBM-PC compatible system, starting from planar guanine tetrads and fixing sugar and phosphate groups following stereochemical guidelines (23). These models were subsequently energy minimised using all atom AMBER force field (24).

RESULTS AND DISCUSSION

Hairpin G-Quartet Structure Has a Characteristic CD Spectrum

The CD spectra of $d(G_4T_nG_4)$ (n=1 to 4), $d(TTAGGG)_4$ and $d(G)_{12}$ in 70 mM NaCl are shown in figure 1. The sequences $d(G_{12})$ and $d(G_4TG_4)$ exhibited a strong positive band at 260 nm and negative band at 240 nm, similar to poly d(G) which is known to adopt a parallel four-stranded structure (25,26). On the other hand, $d(G_4T_3G_4)$ and $d(G_4T_4G_4)$, under similar conditions of experiment, show CD spectra with 295 nm positive and 265 nm

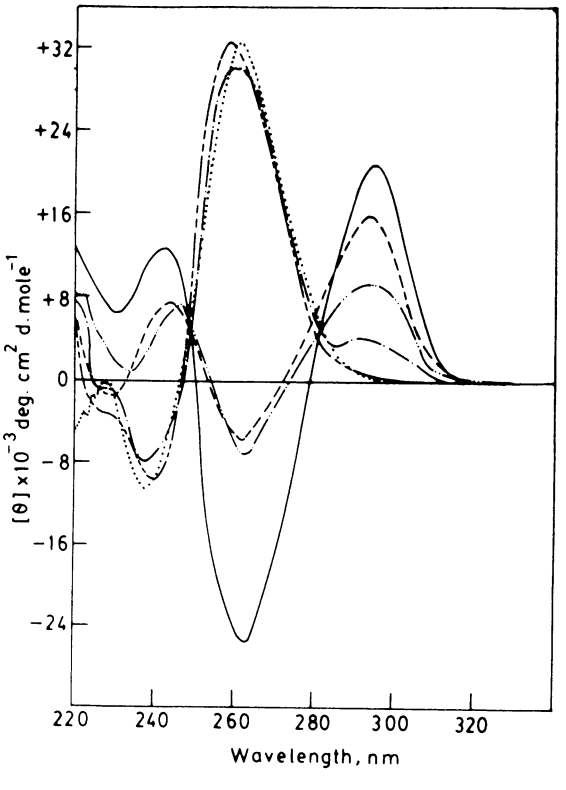

Figure 1. CD spectra of $d(G_4T_4G_4)$ (——), $d(G_4T_3G_4)$ (– – –), $d(G_4T_2G_4)$ (—·—), $d(G_4TG_4)$ (· · · ·), $d(G_{12})$ (— – – —) and $d(TTAGGG)_4$ (—· ·—) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl at 22°C. Samples were annealed at 20 OD_{260} concentration, diluted to 2 OD_{260} and equilibrated for 30 minutes before the scan.

negative bands, while $d(G_4T_2G_4)$ gives a CD spectrum with the shoulder at 295 nm along with the positive band at 260 nm and the negative band at 240 nm. The intensity of the 295 nm positive band increases, with a concomitant decrease in the 265 nm negative band, as the number of thymine residues in the sequences increases from 2 to 4, which suggests that the 295 nm positive band arises due to an ordered structure which is different from that of the $d(G_4TG_4)$ and $d(G_{12})$, which adopt a parallel four stranded structure. As shown in figure 2, $d(G_4T_2G_4)$ exhibits an increase in the intensity of 295 nm CD band on heating to 80°C and rapid cooling to room temperature. $d(G_4T_2G_4)$ apparently exhibits a mixture of both conformations at room temperature. Cech and coworkers (12) demonstrated the formation of folded-back G-quartet structures for the sequences $d(T_2G_4)_4$ and $d(T_4G_4)_4$ from electrophoretic and UV cross linking experiments. The human telomeric sequence $d(TTAGGG)_4$ that shows enhanced electrophoretic mobility in non-denaturing gel as does $d(T_2G_4)_4$, also exhibits a CD spectrum with a 295 nm positive band and a negative band at 263 nm (figure 1). This suggests that the 295 nm positive band arises due to a folded-back hairpin G-quartet structure. Hence the characteristic CD spectra with 295 nm positive band reported here for $d(G_4T_4G_4)$ and $d(G_4T_3G_4)$ sequences can be interpreted as being a spectroscopic signature for the folded back G-quartet structure.

In the case of $d(G_4T_2G_4)$ which shows both 260 nm and 295 nm bands, there is probably a mixture of two structures. The 295 nm positive band for the sequence $d(G_4T_2G_4)$ emerges only upon diluting the sample from 20 OD_{260} to 2 OD_{260} concentration. It is interesting to note that after dilution the 295 nm positive band gradually increases as a function of time accompanied by a decrease in 260 nm ellipticity and reaches a plateau after 5 hours as shown in figure 2 (inset). The CD spectrum of $d(G_4T_2G_4)$ at low concentration obtained after the attainment of equilibrium with a strong 295 nm CD band is also shown in figure 2.This can be interpreted as a transition from four-stranded structure to a hairpin quartet structure although detailed analysis is needed to ascertain the exact pathway of the conversion. The four-stranded structure for $d(G_4T_2G_4)$ at higher concentration is apparently kinetically preferred over the hairpin quartet structure. It is known that the kinetic barrier for the formation of a hairpin loop can be overcome by heating the oligonucleotide above its melting temperature at a low concentration in a low ionic strength buffer and cooling to room temperature (27). Here we have shown that the heating and cooling of the oligonucleotide $d(G_4T_2G_4)$ in 2 OD_{260} concentration shifts the equilibrium towards the hairpin G-quartet structure as manifested by the increase in the 295 nm positive band, whereas the CD spectra of $d(G_4T_4G_4)$ and $d(G_4T_3G_4)$ do not undergo any significant change on heating and cooling. The equilibrium CD spectrum of $d(G_4T_2G_4)$ (figure 2) does not change with time up to 12 hours indicating that the hairpin G-quartet structure once formed is thermodynamically more stable than the four-stranded parallel G-quartet structure. The oligonucleotide $d(T_2G_4)_4$ corresponding to the telomere of *Tetrahymena* also exhibited a CD spectrum similar to that of $d(G_4T_2G_4)$, with 295nm and 260 nm positive bands after heating and cooling at 2 OD_{260} concentration in 70 mM NaCl (figure 2). Recently, similar structural transition from intramolecular folded structure to intermolecular multistranded structure was reported for $d(T_2G_4)_4$ under varying ionic conditions (28). However Hardin *et al* (28) did not observe the characteristic CD spectrum of an intramolecular folded quadruplex structure for the human telomeric sequence $d(TTAGGG)_4$ that we report. On the other hand, $d(G_4TG_4)$ and $d(G_{12})$ which do not show the 295 nm band under our experimental conditions , do not favour a folded-back quartet structure even after heating and cooling, although they could take up such a structure by incorporating

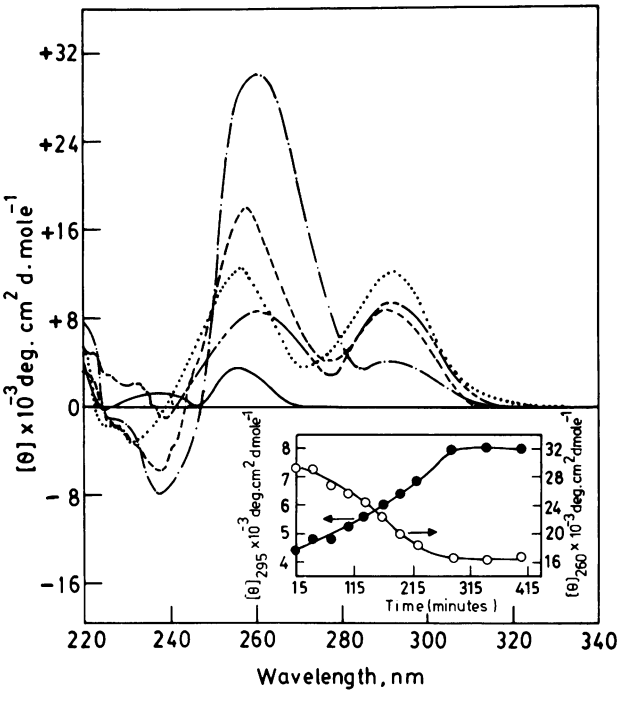

Figure 2. CD spectra of $d(G_4T_2G_4)$ at 2 OD_{260} concentration in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl. (—·—) 22°C ; (——) 80°C. (····) is the spectrum obtained after heating to 80°C and cooling down to 22°C. (– – – –) after attainment of equilibrium (several hours). (— - - —) $d(T_2G_4)_4$, after heating and cooling at 2 OD_{260} concentration. Inset shows the variation of CD at 295 nm (—●—) and 260 nm (—○—) of $d(G_4T_2G_4)$ as a function of time at 22°C after diluting the sample from 20 OD_{260} to 2 OD_{260} in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl.

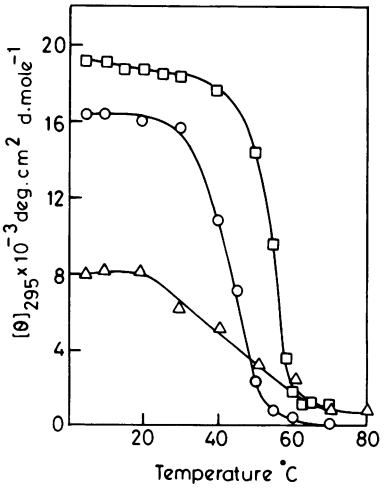

Figure 3. CD melting curves of $d(G_4T_4G_4)$ (□), $d(G_4T_3G_4)$ (○), $d(G_4T_2G_4)$ (△) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl showing the variation in the molar ellipticity at 295 nm with temperature.

Table 1. Thermodynamic parameters for the formation of folded-back hairpin quartet structures.

Sequence	T_m (°C)	ΔH (kcal/mol)	ΔS (cal/mol.deg)	ΔG^+ (kcal/mol)
$d(G_4T_3G_4)$	44	−71	−201	−11
$d(G_4T_4G_4)$	55	−97	−275	−15

* Total concentration of strands = 1.7×10^{-5} M.
† Calculated at 25°C

G residues in the loops, as proposed earlier for oligo d(G) fragments (19).

CD Melting Studies

The stability of the folded-back structures for the sequences $d(G_4T_4G_4)$, $d(G_4T_3G_4)$ and $(G_4T_2G_4)$ was studied by monitoring the 295 nm positive band with change in temperature (figure 3). $d(G_4T_4G_4)$ melts cooperatively with a T_m of 55°C, and $d(G_4T_3G_4)$ with a T_m of 44°C. Thermodynamic parameters obtained by analysis of the melting profiles for the sequences $d(G_4T_3G_4)$ and $d(G_4T_4G_4)$ are given in table 1. It is evident from the ΔG and ΔS values that the larger the number of T residues in the loop the higher is the stability of the hairpin dimer. Hilbers *et al* (29) have reported that the melting temperature of palindromic oligonucleotides $d(ATCCTAT_nTAGGAT)$ which form hairpins in solution remains more or less constant for $n \geq 4$ and gradually decreases for higher values of n up to 7. The difference of 11°C between the melting temperatures of $d(G_4T_4G_4)$ and $d(G_4T_3G_4)$ inspite of the stem being identical in both structures arises from the loop region. Hare and Reid (30) have shown from NMR studies that in a hairpin loop formed by the sequence d(CGCGTTTTTCGCG) the stacking interactions from the stem could extend into the T residues of the loop thus increasing the overall stability of the hairpin structure. We could envisage that additional stability of 4 kcal/mol of $d(G_4T_4G_4)$ over $d(G_4T_3G_4)$ also originates from specific interactions in the loop. The interactions of the T residues with adjacent residues in the loop may be hydrophobic and/or by H-bonding as predicted by earlier workers (30). As expected, the small population of folded back G-quartet structure observed for $d(G_4T_2G_4)$ shows a broad melting profile as monitored at 295 nm with a T_m of 41°C.

The nonameric DNA sequences d(GGTTXTTGG), with X being A, C, G or T, have been reported to form tetraplex structures, rather than the hairpin quartet structure, even though these sequences exhibit a 295 nm positive CD band (18). However, in the light of the results presented in this paper the higher stability observed for d(GGTTTTTGG) over d(GGTTGTTGG) as reported by Jin *et al* (18) cannot be explained on the basis of a tetraplex structure since replacement of G by T in the position X in such a structure would lead to a loss of at least four hydrogen bonds. On the other hand the stability of the quartet structure formed by dimerization of hairpins would have very little effect, if G is replaced by T in the middle of the nonameric sequence forming the loop.

Effect of K^+ Ion

It has been shown by earlier workers, that the addition of K^+ ions stabilises the G-quartet structure for both the parallel and folded-back hairpin arrangements (10−12). All the sequences studied by us denatured completely above 70°C in 70 mM NaCl, as evident from the collapse of the CD bands. However, with

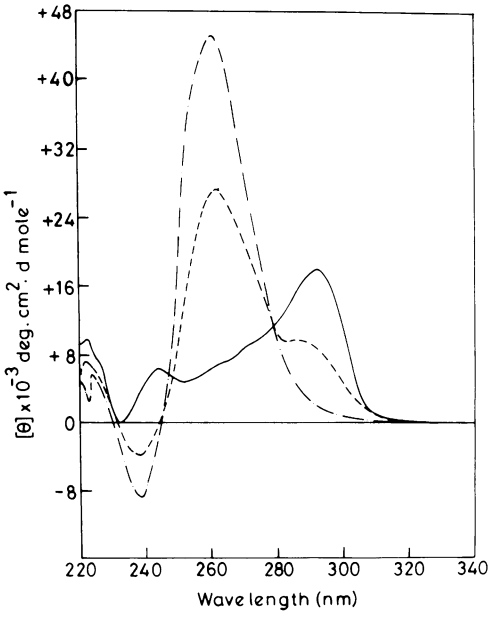

Figure 4. CD spectra in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM KCl at 22°C. $d(G_4T_4G_4)$ (——); $d(G_4T_3G_4)$ (− − − −); $d(G_4T_2G_4)$ (—·—).

the addition of 35 mM KCl to a solution of $d(G_4T_4G_4)$ in 35 mM NaCl, the structure was stable even up to 80°C, showing no appreciable change in $[\theta]_{295}$ molar ellipticity value. If the oligonucleotides were annealed in 70 mM KCl in absence of Na^+ ion there was a progressive increase in the 260 nm positive CD band with concomitant decrease in the 295 nm positive band for $d(G_4T_3G_4)$ and $d(G_4T_4G_4)$ and complete diappearance of the 295 nm shoulder for $d(G_4T_2G_4)$ (figure 4). This suggests that in presence of K^+ ion alone the formation of hairpin structure is not favoured and the equilibrium is shifted towards a parallel four stranded structure. This indicates that while Na^+ ion is essential for nucleation of the folded-back G-quartet structures, addition of K^+ ion could impart further stability to the hairpin G-quartet structure.

Anomalous Gel Electrophoretic Mobility of Telomeric Sequences

The electrophoretic mobilities of telomeric oligonucleotides in a non-denaturing gel containing 70 mM NaCl are shown in figure 5. The oligonucleotides corresponding to human and *Tetrahymena* telomeric repeats $d(TTAGGG)_4$ and $(T_2G_4)_4$ exhibited enhanced electrophoretic mobility at 22°C (figure 5a). At 22°C, these two 24-mer oligonucleotides migrated faster than single stranded $d(T)_{12}$ on a non-denaturing gel whereas at 60°C (figure 5b) the above two telomeric sequences exhibited normal mobility and

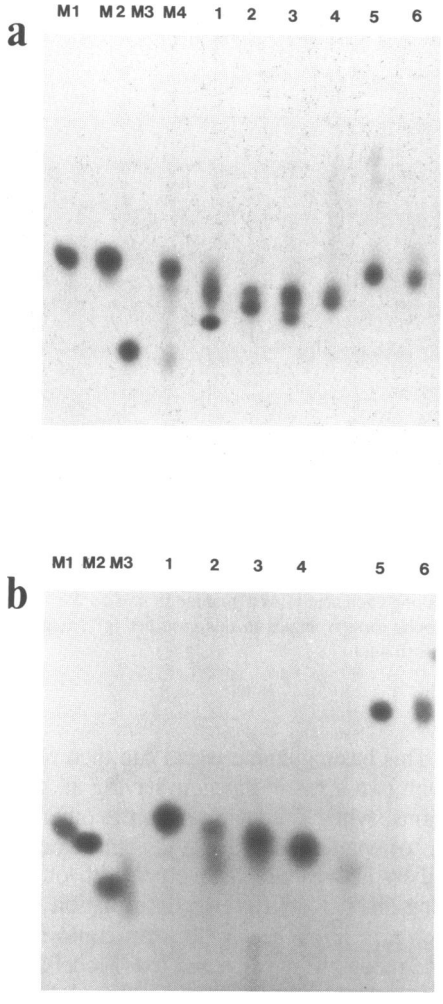

Figure 5. Electrophoresis of telomeric oligonucleotides in 12% non-denaturing gel containing 70 mM NaCl. (a) 22°C and (b) 60°C Lanes M1, WC duplex $d(G_4T_4G_4).d(C_4A_4C_4)$; M2, $d(T)_{12}$; M3, 6-mer duplex $d(CG)_3$; M4, 12-mer duplex $d(CG)_6$; 1, $d(G_4T_4G_4)$; 2, $d(G_4T_3G_4)$; 3, $d(G_4T_2G_4)$; 4, $d(G_4TG_4)$; 5, $d(TTAGGG)_4$; 6, $d(T_2G_4)_4$.

migrated as would be expected for a 24-mer. The oligonucleotides $d(G_4T_4G_4)$, $d(G_4T_3G_4)$ and $d(G_4T_2G_4)$ migrated as two distinct species at 22°C (figure 5a) whereas at 60°C, they exhibited mobilities as expected from their respective lengths with the disappearance of the fast moving component (figure 5b). These results are in good agreement with those of Henderson *et al* (31) who for the first time observed enhanced gel mobility in telomeric oligonucleotides which has been attributed to the formation of stable intramolecular folded G-quartet structures (12,31). These observations and our present data suggest that human telomeric repeat $d(TTAGGG)_4$ like *Tetrahymena* telomeric sequence $d(T_2G_4)_4$ adopts intramolecular folded G-quartet structure with enhanced electrophoretic mobility. The fact that the sequences $d(G_4T_4G_4)$, $d(G_4T_3G_4)$ and $d(G_4T_2G_4)$ migrated slower than the 6bp duplex $d(CG)_3$ and much faster than 12bp duplexes $d(CG)_6$ and $d(G_4T_4G_4).d(C_4A_4C_4)$ as shown in figure 5a suggests that the fast moving component must be a hairpin dimer. The fastmoving component of the sequence $d(G_4T_4G_4)$ migrates ahead of those of $d(G_4T_3G_4)$ and $d(G_4T_2G_4)$ indicating that the hairpin dimer structure adopted by $d(G_4T_4G_4)$ is more compact and stable under the conditions of electrophoresis than the other

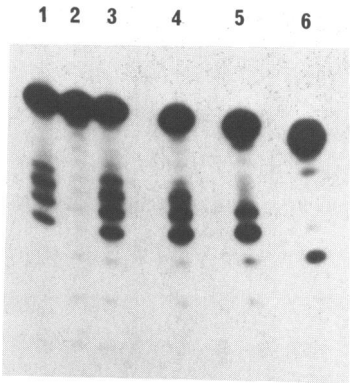

Figure 6. $KMnO_4$ reactivity of thymines in the oligonucleotides $d(G_4T_nG_4)$, n=1−4. Lane 1, $KMnO_4$ modification performed for heat denatured $d(G_4T_4G_4)$ in no salt buffer; lane 2, WC duplex $d(G_4T_4G_4).d(C_4A_4C_4)$ in 70 mM NaCl; lanes 3−6 are $d(G_4T_4G_4)$,$d(G_4T_3G_4)$,$d(G_4T_2G_4)$ and $d(G_4TG_4)$ annealed in 70 mM NaCl respectively; after modification and piperidine treatment cleavage products were resolved on 20% polyacrylamide gel containing 8 M urea and autoradiographed.

two sequences. $d(G_4TG_4)$ did not exhibit a fast moving component under the conditions of electrophoresis suggesting that it does not form a hairpin structure.

$KMnO_4$ Reacts More with Thymines in $d(G_4T_nG_4)$, n=2−4

This interpretation is further supported by the chemical probing of T residues in $d(G_4T_nG_4)$, n=1−4 with $KMnO_4$. OsO_4 and $KMnO_4$ react preferentially with thymines in the single stranded regions of DNA (32−34). We have used $KMnO_4$ to probe the status of thymines in the structures formed by the sequences $d(G_4T_nG_4)$. As expected the thymine residues in the WC duplex $d(G_4T_4G_4).d(C_4A_4C_4)$ showed no reactivity with $KMnO_4$ as shown in figure 6. On the contrary, the T residues in the oligonucleotides $d(G_4T_nG_4)$ with n=2−4 exhibited high reactivity with $KMnO_4$ in 70 mM NaCl (figure 6). This shows that thymines in the structures formed by these sequences (n=2−4) are more accessible to chemical attack by $KMnO_4$, while the single T residue in $d(G_4TG_4)$ is much less reactive with $KMnO_4$ indicating that the structure adopted by this sequence is different from that of $d(G_4T_nG_4)$, n=2−4. Inspite of the fact that $KMnO_4$ at very low concentrations is selective only to thymines (34), the fourth G residue from the 5′end of the above sequences becomes increasingly sensitive to $KMnO_4$ as the number of T residues decreases from 4−1 for reasons which are not clear. From these results and gel mobility studies we infer that the sequences $d(G_4T_nG_4)$, n=2−4 adopt G quartet structures formed by dimerization of hairpin loops with thymines present in the loop being more accessible to $KMnO_4$ oxidation. The thymine residues present in the hairpin dimeric structure formed by the *Tetrahymena* telomeric oligonucleotide were also found to be accessible to OsO_4 oxidation by Sundquist and Klug (11). The sequence $d(G_4TG_4)$ where a single thymine residue is not sufficient to form a loop showed decreased reactivity as expected. $KMnO_4$ reactivity and gel mobility studies are consistent with the circular dichroism data.

Model Building

Detailed model building studies were carried out to investigate the various possible hairpin quartet structures. A folded back hairpin structure necessarily implies an antiparallel arrangement

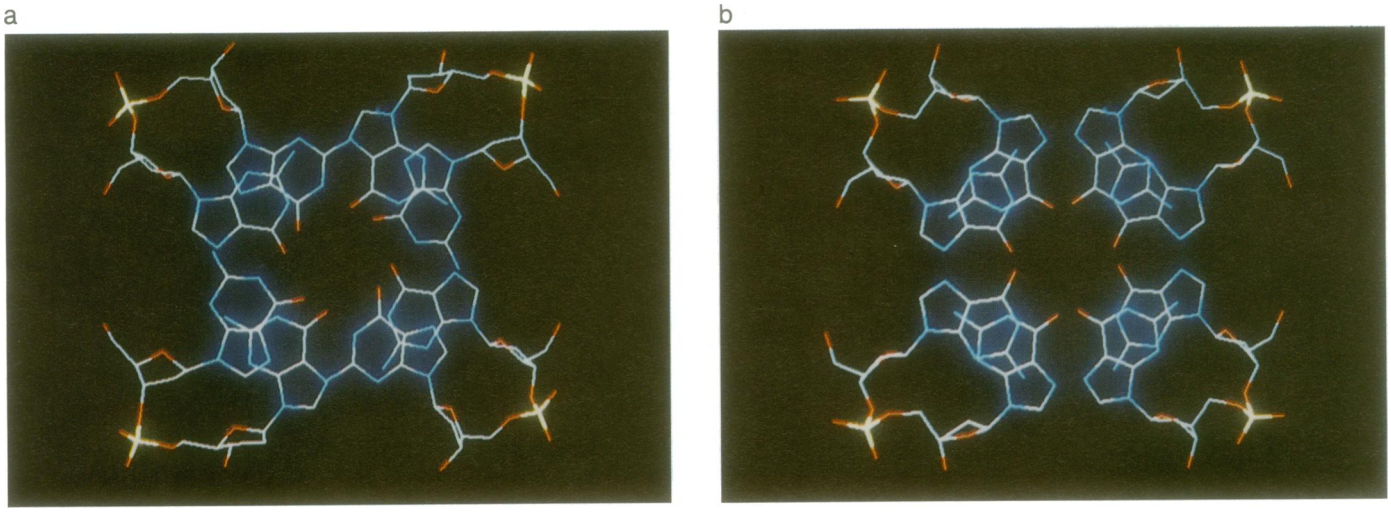

Figure 7: Line diagrams showing stacking arrangement of two guanine tetrads. (a) Normal stacking corresponding to Model I. The backbone torsion angles (α, β, γ, δ, ϵ, ζ) in the arm with *anti* conformation about the glycosidic bond are in the region (**t, t, t, $_1$E, t, g$^-$**), while the arm with *syn* orientations have the conformation (**g$^-$, t, g$^+$, $_1$E, t, g$^-$**). (b) Inverted stacking corresponding to Model II. The 5′→3′ chain direction along each arm joins a guanine in *anti* conformation to a guanine in *syn* conformation and has better base overlap than in the neighbouring *syn-anti* step. The backbone torsion angles in this case are (**g$^-$, t, g$^+$, ^{2}E, t, g$^-$**) for guanine in *anti* conformation, while they are (**g$^-$, g$^-$, g$^-$, ^{2}E, t, g$^-$**) for guanine in the *syn* conformation.

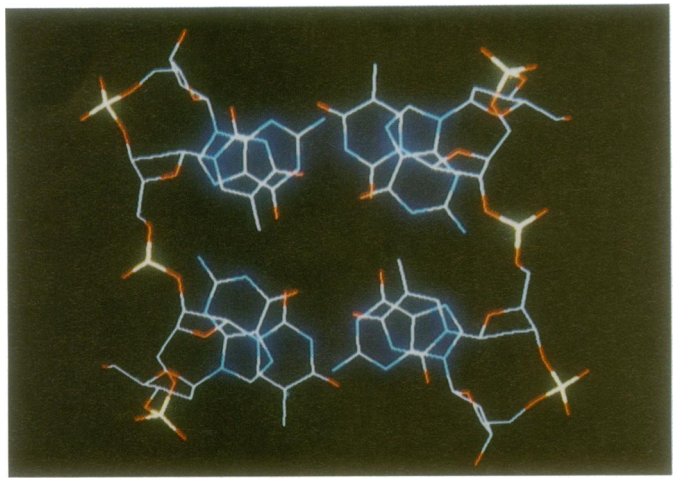

Figure 8: Line diagram showing the stacking of the thymine tetrad on the guanine tetrad for d($G_4T_2G_4$) in the hairpin dimer structure when the T-loops are on the same end of the G-tetrad stem. The thymine residues in this model form two pairs of N3 → O4 hydrogen bonds between the two hairpins. The 5′→3′ chain direction in the loop, joins a guanine in *syn* conformation to a guanine in *anti* conformation, but it is also possible to connect a guanine in *anti* conformation to a guanine in *syn* conformation in each of the hairpins.

of adjacent chains in the stem region and in principle, many different combinations of arrangements are possible for the parallel and antiparallel arms in the stem region of the two hairpins (11,13,31). We have considered two types of stereochemically viable hairpin structures, with the adjacent phosphodiester chains across any pair of hydrogen bonded guanines running in antiparallel directions and with a minimum of two thymine residues in the loop. The antiparallel sense of adjacent arms, combined with the four fold rotational symmetry of the guanine tetrad leads to the neighbouring hydrogen bonded bases being oriented in nearly opposite conformations about the glycosidic bond (i.e. *syn-anti-syn-anti*), in contrast to the parallel stranded structure, in which all guanine bases are in the *anti* orientation. This basic guanine tetrad can then repeat along the molecule, with two arms in the stem having all guanines in the *syn* orientations, while the guanines in the other two arms are in the *anti* orientation (11,13,19). Thus the molecule is characterised by a two-fold symmetry, with one arm of each hairpin having distinctly different conformation from the other arm, but each arm has a nearly uniform conformation (Model I). The second model we have considered (Model II) has the same basic tetrad structure with adjacent guanines being in *syn-anti-syn- anti* orientations, but along the polynucleotide chain also there is an alternation of guanine orientation with respect to the sugar phosphate backbone (as in Z-DNA) (35). However, the alternation occurs by an inversion of the complete guanine tetrad, while retaining the orientation of the sugar phosphate backbone (unlike Z-DNA where adjacent sugars are oriented in opposite directions) (35). The alternating *anti-syn* conformation in successive guanine tetrads leads to an inverted stacking arrangement of the bases, and there is a dinucleotide repeat along the polynucleotide chain in the stem region. However, all four arms have identical conformations in this structure. Thus both these models are different from the schematic model proposed earlier (13), with alternating *syn-anti-syn-anti* glycosidic orientation along a strand, but with the bases in either all *syn* or all *anti* orientations within a guanine tetrad.

Preliminary models were built for the quartet stem region, using a planar guanine tetrad with Hoogsteen type hydrogen bonds. In the model I, the successive guanine tetrads were positioned as in the fibre models of poly(G) (i.e. rise = 3.4Å and twist = 30°) and joined so that adjacent arms in the stem were antiparallel. The overall arrangement of bases in the quartet remains identical in model II, but the neighbouring tetrads were rotated through 180° about an axis in the plane of the basepairs. Two slightly different backbone conformations were required to form the *syn* to *anti* and *anti* to *syn* linkages across the phosphodiester bonds, leading to small differences in helical twist for the two steps. However, all four arms in the stem have identical conformations in this model and all the sugars have C2′-endo pucker.

Normal stacking of G tetrads as in Model I, as well as inverted stacking, as in model II, for the energy minimised structures lead to good base overlap as seen in figure 7 (a and b). A two residue thymine link could be generated for both models, with the residue at the 5′ end of the loop being stacked over the preceding G while the thymine at 3′ end was slightly destacked in the starting models. A comparison of energy values of hairpin quartet and parallel tetraplex structures indicates that while the parallel four-stranded structure is marginally more favourable than the quartet structure with antiparallel chains, for an all G sequence, the presence of intervening thymine residues makes the folded-back hairpin structure more favourable. It is interesting to note that the various NMR studies also indicate alternating *syn-anti* conformation of guanines along a strand, as in Model II (36,37). The nmr experiments are however unable to distinguish between a hairpin dimer model and a four stranded quartet model with a bulge in the T region (36,37).

It is gratifying to note that subsequent to the submission of this manuscript the X-ray crystal structure (38) for the sequence $d(G_4T_4G_4)$ has been reported and shows many features similar to our Model II *viz* hairpin dimer formation with alternating *anti-syn* glycosidic orientation along the strand as well as within the G-tetrad. The two T-loops are found to be on opposite ends of the G-tetrad in the crystal structure. However, if the loops are on the same end of the G-tetrad stem, then we find that the thymine residues in the loops of the two hairpins can also form a tetrad structure (fig. 8). In the intramolecular folded quadruplex structures, as in $d(T_2G_4)_4$ and $d(T_4G_4)_4$ at least two hairpin loops have to be on the same end of the tetraplex stem. In fact the four thymine residues in the two loops of energy minimised $d(G_4T_2G_4)$ dimer stack over the last guanine tetrad in the stem region with the thymines in the two hairpins being linked through a pair of hydrogen bonds (fig. 8). The differences seen in the arrangement of the T-loops in the crystal (38) and the recently reported NMR (39) structures for $d(G_4T_4G_4)$ and $d(G_4T_4)_3G_4$ suggest that in solution different structures with varying loop orientations are possible.

CONCLUSIONS

Thus, studies reported here show that even short stretches of $d(G_4T_nG_4)$ ($n \geq 2$) sequences have an intrinsic propensity to form stable hairpin dimers and the stability of these hairpin quartet structures increases with increase in the number of intervening thymine residues. It is believed that such structures may be formed by association of telomeric ends of two chromosomes as suggested by Sundquist and Klug (11). Presence of two repeats of $d(T_nG_4)$ at the 3′ protruding end of the chromosomes could lead to pairing of the chromosomes while occurrence of four repeats could reduce such pairing potential due to the possibility of intramolecular folding. Stabilization of a hairpin dimer or parallel four stranded structure would then depend not only on Na^+/K^+ switch (9,10), but also on the nature of the sequence repeat. Since $d(G_mN_nG_m)$ sequence motifs have been observed in DNA sequences involved in recombination (40–42) it also seems plausible that these G-rich sequences within the DNA double helix could loop out, leading to the formation of G-quartet structure involving two hairpins during recombination.

ACKNOWLEDGEMENTS

Authors thank DBT (India) for the oligonucleotide synthesis facility at IISc. Financial assistance to SKB from DST (ILTP) India, and to MB from DBT (India) is gratefully acknowledged. DM is a recipient of CSIR (India) fellowship. Authors thank the anonymous referees for valuable suggestions.

REFERENCES

1. Blackburn, E.H. & Gall, J.G. (1978) J.Mol.Biol. **120**, 33–53.
2. Klobutcher, L.A., Swanton, M.T., Donnini, P. & Prescott, D.M. (1981) Proc.Natl.Acad.Sci. U.S.A. **78**, 3015–3019.
3. Oka, Y., Shiota, S., Nakai, S., Nishida, Y. & Okubo, S. (1980) Gene, **10**, 301–306.
4. Kipling, D. & Cooke, H.J. (1990) Nature **347**, 400–402.
5. Moyzis, R.K., Buckingham,J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. & Wu, J.R. (1988) Proc.Natl.Acad.Sci.USA. **85**, 6622–6626.
6. Richards, E.J. & Ausubel, F.M. (1988) Cell **53**, 127–136.
7. Henderson, E.R. & Blackburn, E.H. (1989) Mol.Cell.Biol. **9**, 345–348.
8. Blackburn, E.H. (1990) J.Biol.Chem. 265, 5919–5921.
9. Sen, D & Gilbert, W. (1988) Nature **334**, 364–366.
10. Sen, D. & Gilbert, W. (1990) Nature **344**, 410–414.
11. Sundquist, W.I. & Klug, A. (1989) Nature **342**, 825–829.
12. Williamson, J.R., Raghuraman, M.K. & Cech, T.R. (1989) Cell **59**, 871–880.
13. Guschlbauer, W., Chantot, J.F. and Thiele, D. (1990) J. Biomol. Struct. Dynam., **8**, 491–511.
14. Sasisekharan, V., Zimmermann, S.B. & Davies, D.R. (1975) J.Mol.Biol. **92**, 171–179.
15. Pinnavaia, T.J., Miles, H.T. & Becker, E.D. (1975) J.Amer.Chem.Soc. **97**, 7198–7200.
16. Arnott, S., Chandrasekharan, R. & Martilla, C.M. (1974) Biochem.J. **141**, 537–543.
17. Zimmermann, S.B., Cohen, G.H. & Davies, D.R. (1975) J.Mol.Biol. **92**, 181–192.
18. Jin, R., Breslauer, K.J., Jones, R.A. & Gaffney, B.L. (1990) Science **250**, 543–546.
19. Panyutin, I.G., Kovalsky, O.I., Budowsky, E.I., Dickerson, R.E., Rikhirev, M.E. & Lipanov, A.A. (1990) Proc.Natl.Acad.Sci.USA. **87**, 867–870.
20. Albergo, D.D., Marky, L.A., Breslauer, K.J. and Turner, D.H. (1981) Biochemistry **20**, 1409–1413.
21. Mishra, R.K., Latha, P.K. and Brahmachari, S.K. (1988) Nucleic Acids Res. **16**, 4651–4665.
22. Desktop Molecular Modeller, Version 1.2 (Oxford University Press, 1989).
23. Bansal, M. and Sasisekharan, V. (1986) in Szabo, G.N.(ed) Theoretical Chemistry of Biological Systems, Elsevier, Amsterdam, New York, pp 127–288.
24. AMBER Ver. 3.0 (1987) Molecular Mechanics Program from University of California, San Francisco.
25. Gray, D.M. & Bollum, F.J. (1974) Biopolymers **13**, 2087–2102.
26. Marck, C. & Thiele, D. (1978) Nucl. Acids Res. **45**, 1017–1028.
27. Xodo,L.E., Manzini, G., Quadrifoglio, F., van der Marel, G., and van Boom, J.H. (1988) Biochemistry **27**, 6321–6326.
28. Hardin, C.C., Henderson, E., Watson, T. & Prosser, J.K. (1991) Biochemistry **30**, 4460–4472.
29. Hilbers, C.W., Haasnoot, C.A.G., de Bruin, S.H., Joordens, J.J.M., van der Marel, G.A. & van Boom, J.H. (1985) Biochimie **67**, 685–695.
30. Hare, D.R., & Reid, B.R. (1986) Biochemistry **25**, 5341–5350.
31. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I.,Jr. and Blackburn, E.H. (1987) Cell, **51**, 899–908.
32. Friedman, T. & Brown, D.M. (1978) Nucl.Acids.Res. **5**, 615–622
33. Burton, K. & Reilly, W.T. (1966) Biochem.J. **98**, 70–77.
34. Hayatsu, H. & Ukita, T. (1967) Biochem.Biophy.Res.Com. **29**, 556–561.
35. Rich, A., Nordheim, A. and Wang, A. H. -J. (1984) Ann. Rev.Biochem. **53**, 791–846.
36. Wang, Y., Jin, R., Gaffney, B., Jones, R.A. and Breslauer, K.J. (1991) Nucl. Acids Res. **19** , 4619–4622.
37. Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. and Patel, D. J (1991) J. Mol. Biol. **222** , 819–832.
38. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. and Rich, A. (1992) Nature **356**, 126–131.
39. Smith, F.W. & Feigon, J. (1992) Nature **356**, 164–168.
40. Pluta, A.F. & Zakian, V.A. (1989) Nature **337**, 429–433.
41. Cherry, J.M. & Blackburn, E.H. (1985) Cell **43**, 747–758.
42. Herrick, G., Cartinhour,S., Dawson, D., Ang, D., Sheets, R., Lee, A. & Williams, K. (1985) Cell **43**, 759–946.