THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 34, Issue of August 26, pp. 21858–21869, 1994
Printed in U.S.A.

Structure and Stability of Human Telomeric Sequence*

(Received for publication, March 4, 1994, and in revised form, May 25, 1994)

Pichumani Balagurumoorthy‡ and Samir K. Brahmachari‡§¶

From the ‡Molecular Biophysics Unit and the §Centre for Genetic Engineering, Indian Institute of Science, Bangalore 560 012, India

Telomeric DNA of a variety of vertebrates including humans contains the tandem repeat $d(TTAGGG)_n$. We have investigated the structural properties of the human telomeric repeat oligonucleotide models $d(T_2AG_3)_4$, $d(G_3T_2A)_3G_3$, and $d(G_3T_2AG_3)$ using CD, gel electrophoresis, and chemical probing techniques. The sequences $d(G_3T_2A)_3G_3$ and $d(T_2AG_3)_4$ assume an antiparallel G quartet structure by intramolecular folding, while the sequence $d(G_3T_2AG_3)$ also adopts an antiparallel G quartet structure but by dimerization of hairpins. In all the above cases, adenines are in the loop. The TTA loops are oriented at the same end of the G tetrad stem in the case of hairpin dimer. Further, the oligonucleotide $d(G_3T_2AG_3)$ forms a higher order structure by the association of two hairpin dimers via stacking of G tetrad planes. Here we show that N-7 of adenine in the hairpin dimer is Hoogsteen hydrogen-bonded. The partial reactivity of loop adenines with DEPC in $d(T_2AG_3)_4$ suggests that the intramolecular G quartet structure is highly polymorphic and structures with different loop orientations and topologies are formed in solution. Intra- and interloop hydrogen bonding schemes for the TTA loops are proposed to account for the observed diethyl pyrocarbonate reactivities of adenines. Sodium-induced G quartet structures differ from their potassium-induced counterparts not only in stability but also in loop conformation and interactions. Thus, the overall structure and stability of telomeric sequences are modulated by the cation present, loop sequence, and the number of G tracts, which might be important for the telomere function.

Telomeres are the physical ends of eukaryotic chromosomes comprising specialized nucleoprotein structures essential for stability and complete replication of chromosomes (1, 2). Telomeric DNA consists of tandemly repeated sequences with one strand being rich in guanines (3). The G-rich strand running in a 5′ to 3′ direction toward the molecular terminus of the chromosome falls under the subset consensus $d(T_{1-3}(T/A)G_{3-4})$. The repeats $(T_4G_4)_n$ and $(T_2G_4)_n$ occur in the telomeres of protozoans *Oxytricha nova* and *Tetrahymena thermophila*, respectively (4–6), and $(T_2AG_3)_n$ is present in telomeres of more than 100 vertebrates examined so far including human (7). A unique feature of telomeres is that two repeats of G-rich strand extend beyond the duplex DNA producing a 3′ single-stranded overhang 12–16 nucleotides in length (8–10).

A conserved feature of telomeric DNA in diverse eukaryotes is the occurrence of stretches containing three to four contiguous guanines. Guanosine mononucleotides, poly(dG) and poly(rG), have been known to exhibit unusual aggregation properties and form multistranded structures containing G quartets stabilized by a cyclic Hoogsteen type hydrogen bonding (11–16). Such higher order structures were thought to be biologically irrelevant until the recent realization that G-rich sequences found at the chromosomal telomeres, immunoglobulin switch regions, and recombinational hotspots can adopt G quartet structures *in vitro* (17–21).

Structures of synthetic oligonucleotides corresponding to telomeric G-rich strands have been investigated extensively in the recent years because of their unusual cohesive properties (22). Non-denaturing gel mobility and NMR studies of Henderson *et al.* (17) indicated a compact folded structure for $d(T_2G_4)_4$ with some of the guanines in unusual *syn* conformation. Subsequently, gel mobility and dimethyl sulfate reactivity experiments on G-rich single-stranded oligonucleotides suggested a parallel four-stranded structure for the oligonucleotides derived from immunoglobulin switch regions (18), an antiparallel hairpin dimer for $d(T_2G_4)_2$ (23), an intramolecular G quartet structure for $d(T_4G_4)_4$ (24), and a fold-back G quartet structure for the oligo$(dG)_{27}$ and $(dG)_{32}$ fragments (25). In most of these studies, the inaccessibility of N-7 positions of guanines to methylation by dimethyl sulfate and anomalous electrophoretic mobility on non-denaturing gels were used as diagnostics for the G quadruplex formation. We demonstrated that telomeric sequences $d(G_4T_nG_4)$ where n = 1–4, containing 2 isolated tracts of guanine residues adopt either a parallel four-stranded or an antiparallel G quartet structure depending on the number of intervening thymine residues by spectroscopic, gel electrophoretic, and chemical foot-printing methods (26, 27). Further, the formation and stability of these G quartet structures were found to depend on the presence of either Na^+ or K^+ (27, 28). In the high resolution crystal structure, $d(G_4T_4G_4)$ forms a G quartet structure by dimerization of hairpins with alternating *syn-anti-syn-anti* conformation of guanines both along the polynucleotide chain as well as around the G tetrad (29) as proposed by us from model building studies (27). On the other hand, from NMR studies (30) the same sequence in solution has been found to adopt a hairpin dimer with loops oriented diagonally across the G quartet stem with guanines in alternating *syn-anti-syn-anti* conformation along the polynucleotide chain.

Thus, the oligomers corresponding to the G-rich strand of telomeres adopt quadruplex structures of three distinct types depending on the number of G tracts. That is, sequences with one run of guanines form a tetramolecular structure with all the four strands running parallel; two stretches of guanines adopt a bimolecular structure by dimerization of hairpins while four copy sequences form an intramolecular G quartet structure. The polarity of adjacent strings of guanines in the quadruplex structures of two and four copy sequences is decided by the loop orientation. In the crystal structure of $d(G_4T_4G_4)$, the hairpin dimer has thymine loops oriented at opposite ends of the G tetrad stem with planes of loops being parallel to each other (29) resulting in the antiparallel arrangement of all ad-

* This work was supported by grants from the Department of Biotechnology, India and DST (ILTP) India. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed.

jacent strands. On the other hand, NMR studies of Smith and Feigon (30) indicate that loops in the hairpin dimer of $d(G_4T_4G_4)$ are oriented diagonally across the G quartet, apparently at opposite ends of the tetrad stem leading to alternating parallel and antiparallel adjacent strands. Our earlier studies on a series of oligonucleotides $d(G_4T_nG_4)$ where n = 1–4 showed that sequences with $n \geq 2$ form antiparallel hairpin dimers whose stability increases with an increase in the number of thymines (27). Model building and molecular mechanics studies revealed the possibility of interloop T:T base pairing if the two thymine loops occur at the same end of the tetrad stem as in the intramolecular G quartet structure of $d(G_4T_4)_3G_4$ (27). Similar interloop T:T base pairing was observed in the intramolecular G quartet structure of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG) (31). NMR studies on r(UGGGGU) indicated a parallel tetraplex for this sequence in the presence of K^+ ions with an additional unusual U quartet stabilized by N-3···O-4 hydrogen bonding (32). Scaria *et al.* (33) proposed G quartet structures with asymmetric dimerization of hairpins for the sequence $d(G_3T_4G_3)$ in KCl solution. Recently, a single copy human telomeric sequence d(TTAGGG) has been shown to adopt a parallel four-stranded structure in K^+ solution with all guanines in *anti* conformation (34). Thus x-ray (29), spectroscopic (22, 27, 30), calorimetric (35), and theoretical (36) studies conducted so far on several oligonucleotide models have indicated that the quadruplex structure is overwhelmingly polymorphic, although the G tetrad is conserved as the basic structural motif. Thus, it is of interest to study how the sequence, conformation, and orientation of loops determine the overall structure and stability of the telomeric sequences.

Here we have investigated the structures formed by the human telomeric repeat sequences $d(T_2AG_3)_4$, $d(G_3T_2AG_3)$, and $d(G_3T_2A)_3G_3$ by spectroscopic, gel electrophoretic, and chemical probing methods. Using the above oligonucleotide models, we show that a human telomeric sequence adopts an antiparallel G quartet structure with adenines in the loop. $KMnO_4$ probing indicates that loops are at the same end of the G tetrad stem in the antiparallel hairpin dimer of $d(G_3T_2AG_3)$. Sodium-induced G quartet structures differ from their potassium-induced counterparts not only in the stability, but also in loop conformation and interactions. Results of diethylpyrocarbonate reactivity experiments provide evidence for the intra- and interloop hydrogen bonding interactions between T and A residues in the TTA loops of the quadruplex structures formed by the above sequences. Finally, we show that at a higher concentration the antiparallel hairpin dimer of $d(G_3T_2AG_3)$ further dimerizes to form a higher order structure via stacking of G tetrad planes.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification—Oligonucleotides were synthesized on Pharmacia DNA synthesizer Gene Assembler Plus employing β-cyanoethyl phosphoramidite chemistry and purified on a 20% polyacrylamide gel containing 8 M urea. The desired bands were excised under UV, eluted with sterile water, and desalted by passing twice through Pharmacia NAP columns equilibrated with sterile water. Some of the oligonucleotides were also purified on Sep-Pak C18 (Waters) cartridge using an acetonitrile/water mixture as the eluent and were found to give results identical to that obtained from oligomers purified on NAP columns.

CD Measurements—CD data were collected on a Jasco-J 500 A spectropolarimeter interfaced with a data processor DP 501. Samples were prepared by heating the oligonucleotide at a concentration of 2 A_{260} units (corresponding to strand concentrations of 7.9, 8.9, and 20.5 μM for $d(T_2AG_3)_4$, $d(G_3T_2A)_3G_3$, and $d(G_3T_2AG_3)$, respectively), in the presence of the respective cation at 95 °C for 3 min and then gradually cooling to room temperature. In CD melting studies, annealed samples were equilibrated at room temperature for several hours before the experiment. The temperature of the sample was maintained using a circulating water bath (Pharmacia). Samples were equilibrated at each temperature for 20 min prior to scan. The monophasic melting transitions were assumed to reflect helix-coil equilibrium, and thermodynamic parameters were obtained from van't Hoff plots using a two-state model (37). Thermodynamic parameters were derived from the melting curves generated by monitoring a 295-nm CD band which is characteristic of an antiparallel G quartet structure as a function of temperature, and at this wavelength the contribution from parallel G quartet structure is zero. Equilibrium constants at various temperatures were calculated using the following expression (38):

$$K = \frac{\alpha}{nC_T^{n-1}(1-\alpha)^n} \quad \text{(Eq. 1)}$$

where α is the fraction of helix, and n is the molecularity of the complex (1 for 3.5 and four copy sequences and 2 for 1.5 copy sequence) and C_T is the total strand concentration. In concentration-dependent CD, experiment samples were annealed at the appropriate strand concentration and equilibrated at room temperature for several hours before the CD measurement.

Non-denaturing Gel Electrophoresis—Oligonucleotides 5′ end-labeled with [γ-^{32}P]dATP using polynucleotide kinase (Bangalore Genei) were mixed with unlabeled oligonucleotide to give appropriate final concentrations, annealed as mentioned above. The samples were analyzed on a 12% non-denaturing polyacrylamide gel containing either 70 mM NaCl or 70 mM KCl pre-equilibrated at the indicated temperatures for 1 h. Electrophoresis was done in 0.5 × TBE (45 mM Tris-borate, 1 mM EDTA) buffer containing 70 mM NaCl or 70 mM KCl at 8 V/cm for a period of 12 h.

$KMnO_4$ Modification—$KMnO_4$ modification experiments were done as described earlier (27). Essentially, oligonucleotide samples in the appropriate salt-containing buffer were treated with 1 μl of 5 mM $KMnO_4$ solution, diluted from freshly prepared 100 mM solution. After an incubation for 5 min on ice, the reaction was arrested by the addition of 2 μl of allyl alcohol. Samples were ethanol precipitated along with *Escherichia coli* tRNA, and pellets were dried in a Speed Vac.

Dimethyl Sulfate (DMS)[1] Modification—Labeled oligonucleotides (2 μg) annealed as mentioned above were treated with 1 μl of 1:25 DMS in 20 μl of reaction volume for 10 min on ice, followed by the addition of 6 μl of DMS stop solution (1 M β-mercaptoethanol and 1.5 M sodium acetate, pH 7.0) containing 6 μg of tRNA and ethanol precipitation. In control experiments labeled oligonucleotides were heat denatured in a no salt buffer just before adding DMS to ensure complete absence of any secondary structure.

Diethyl Pyrocarbonate (DEPC) Modification—Oligonucleotides ^{32}P labeled at the 5′ end were annealed with 2 μg of unlabeled oligonucleotide in the presence of respective cation (70 mM Na^+ or 70 mM K^+ in 2 mM sodium cacodylate buffer and 0.1 mM EDTA, pH 7.2). 1 μl of DEPC was added, vortexed vigorously, and incubated at the desired temperatures (4 and 22 °C) for 25 min with occasional shaking to keep the solution saturated with DEPC (39). Reactions were quenched by adding 7 μg of *E. coli* tRNA, followed by ethanol precipitation.

Piperidine Cleavage—After chemical modifications by $KMnO_4$, DMS, and DEPC, DNA pellets were dried and treated with 40 μl of 1 M piperidine at 90 °C for 25 min. Subsequently, samples were dried in a Speed Vac, dissolved in 90% formamide containing bromphenol blue and xylene cyanol, and analyzed on a 20% polyacrylamide gel containing 8 M urea using 1 × TBE as the running buffer. Chemical probing experiments were repeated at least three times, and the reactivity data were found to be highly reproducible.

RESULTS AND DISCUSSION

Human Telomeric Oligonucleotides Adopt Antiparallel G Quartet Structure with Adenines in the Loop in Na^+—CD spectroscopy has been used to distinguish between parallel and antiparallel G quartet structures formed by oligonucleotides corresponding to the G-rich strand of telomeric DNA (26, 27, 40). We have previously shown that the antiparallel G quartet structure has a characteristic CD signature with 295 nm positive and a 265 nm negative bands (26, 27). Several studies have shown that the parallel four-stranded structure of G cluster oligonucleotides exhibits a CD spectrum with a positive band around 260 nm and a negative band around 240 nm (16, 26, 27, 40–44). The temperature-dependent CD spectra of the four copy human telomeric repeat $d(T_2AG_3)_4$ in 70 mM NaCl are

[1] The abbreviations used are: DMS, dimethyl/sulfate; DEPC, diethyl pyrocarbonate; bp, base pair(s).

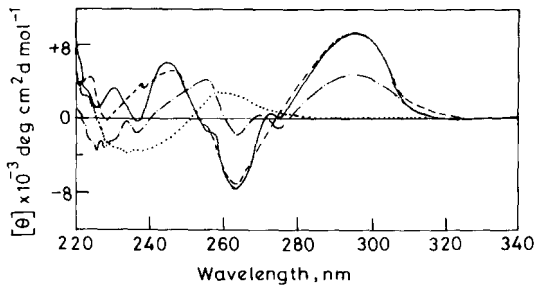


FIG. 1. **CD spectra of $d(T_2AG_3)_4$ in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl.** ——, 5 °C; - - - -, 20 °C; —·—·—, 50 °C; ····, 70 °C.

shown in Fig. 1. In 70 mM NaCl the sequence $d(T_2AG_3)_4$ exhibits a CD spectrum with a positive band at 295 nm and a negative band around 265 nm characteristic of an antiparallel G quartet structure. The antiparallel G quartet structure of $d(T_2AG_3)_4$ exhibits considerable stability, and the intensity of 295 nm positive band decreases with an increase in temperature and reaches zero at 70 °C. There are two possible intramolecular antiparallel G quartet structures for the four copy human telomeric sequence $d(T_2AG_3)_4$ as shown in Fig. 2. The first possibility is that adenine residues can be in the loop region resulting in a structure having three G tetrads in the stem region. The second possibility is that adenine residues also participate in the cyclic Hoogsteen type hydrogen bonding with the G residues resulting in two tetrads containing only guanines (inner ones) and two tetrads (outer ones) with both A and G residues involved in cyclic Hoogsteen type hydrogen bonding. In order to determine whether adenine residues are in the stem or loop region, we synthesized two more truncated 3.5 and 1.5 copy human telomeric sequences $d(G_3T_2A)_3G_3$ and $d(G_3T_2AG_3)$, respectively. These two truncated human telomeric sequences do not have as many adenine residues as the number of guanine tracts and hence cannot accommodate A residues in the tetrad stem. The temperature-dependent CD spectra of $d(G_3T_2A)_3G_3$ and $d(G_3T_2AG_3)$ in 70 mM NaCl are shown in Fig. 3 (*A* and *B*). Both of these truncated human telomeric oligonucleotides exhibit a positive band at 295 nm and a negative band around 265 nm for an antiparallel G quartet structure. The 1.5 copy human telomeric repeat $d(G_3T_2AG_3)$ adopts an antiparallel hairpin dimer similar to that of $d(G_4T_4G_4)$ as inferred from the CD spectrum with 295 nm positive and 265 nm negative bands (27). The similarity in the CD spectra of $d(T_2AG_3)_4$, $d(G_3T_2A)_3G_3$, and $d(G_3T_2AG_3)$ in terms of band positions and the higher stability of $d(G_3T_2A)_3G_3$ over $d(T_2AG_3)_4$ led us to conclude that adenine residues are in the loop region of the antiparallel G quartet structure formed by the above human telomere oligonucleotide models. The four copy *Tetrahymena* telomeric sequence $d(T_2G_4)_4$ showed a CD spectrum similar to $d(G_4T_2G_4)$ with positive bands at 295 and 260 nm indicating that these sequences form a mixture of both parallel and antiparallel G quartet structures in NaCl solution, and the antiparallel G quartet structure contains 2 thymine residues in the loop region (27). Therefore, if adenine residues were in the quadruplex stem in the intramolecular G quartet structure of $d(T_2AG_3)_4$ with 2 thymine residues in the loop region, then this sequence would have also exhibited a CD spectrum similar to that of $d(G_4T_2G_4)$ and $d(T_2G_4)_4$ with two positive bands, one at 295 nm and another at 260 nm (27). Thus, we conclude that the human telomeric sequences adopt antiparallel G quartet structures with TTA loops. This conclusion is further supported by chemical probing.

Effect of K^+ on the Structure and the Stability of Human Telomeric Oligonucleotides—Guanosine mononucleotides have been known to form ordered structures by aggregation of G tetrad alignments in a monovalent cation-dependent fashion (12). The tetraplex structures formed by the G-rich sequences

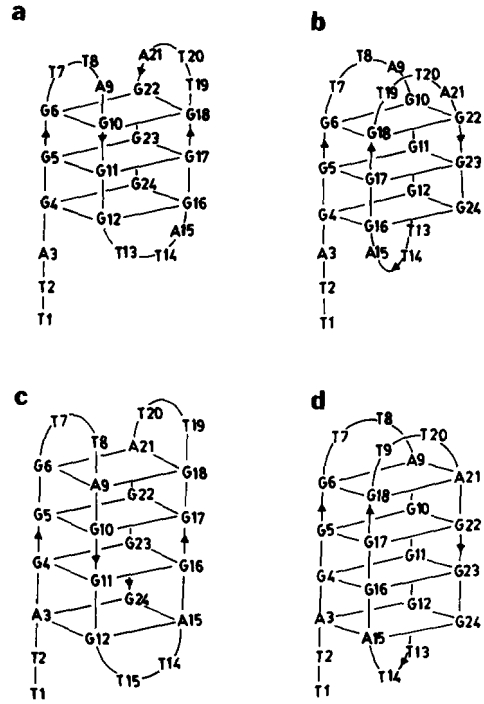


FIG. 2. **Schematic diagram showing the various possible folded quadruplex structures of $d(T_2AG_3)_4$.** *a*, edge-looped structure with A in the loop; *b*, diagonal-looped structure with A in the loop; *c*, edge-looped structure with A in the tetrad stem; *d*, diagonal-looped structure with A in the tetrad stem.

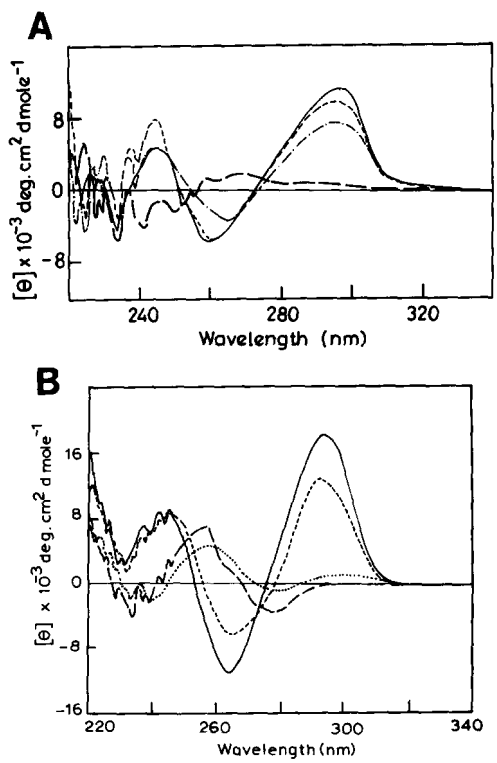


FIG. 3. *A*, CD spectra of $d(G_3T_2A)_3G_3$ in 2 mM sodium cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl at 5 °C (——), 20 °C (- - - -), 50 °C (—·—·—), and 70 °C (– –). *B*, CD spectra of $d(G_3T_2AG_3)$ in 2 mM sodium cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl at 5 °C (——), 20 °C (- - - -), 50 °C (····), and 70 °C (– –).

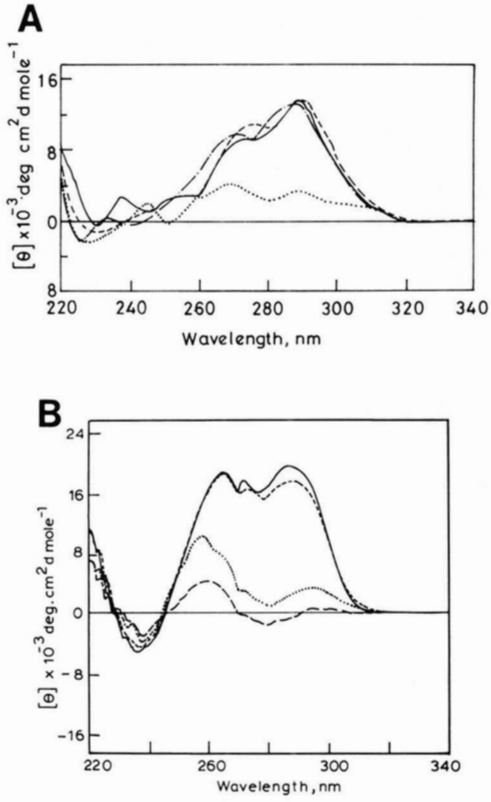


FIG. 4. *A*, CD spectra of $d(T_2AG_3)_4$ in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM KCl at 5 °C (——), 20 °C (- - - -), 50 °C (–·–·–), and 70 °C (····). *B*, CD spectra of $d(G_3T_2AG_3)$ in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM KCl at 5 °C (——), 20 °C (- - - -), 50 °C (····), and 70 °C (– –).

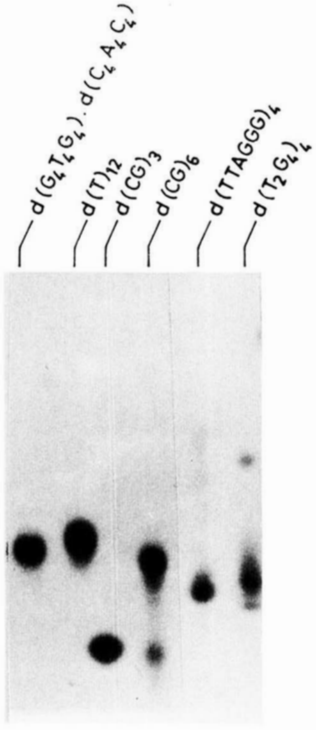


FIG. 5. **Analysis of four copy human telomeric repeat $d(T_2AG_3)_4$ on a 12% non-denaturing polyacrylamide gel containing 70 mM KCl at 22 °C.** $d(T_2G_4)_4$ which is known to migrate anomalously faster was also loaded as a positive control.

corresponding to the immunoglobulin switch regions depend on the balance of Na^+/K^+ concentration (28). Earlier we have shown that K^+ ions shift the equilibrium toward the parallel

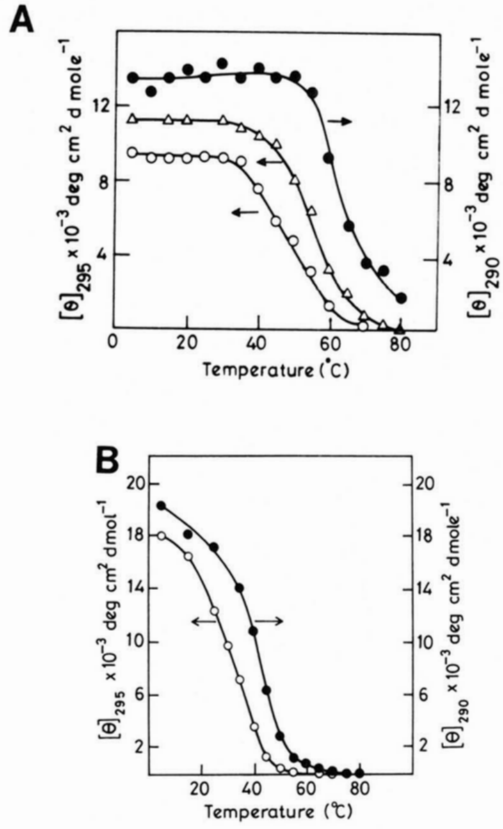


FIG. 6. *A*, CD melting profiles of $d(T_2AG_3)_4$ in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl (○), 70 mM KCl (●), and $d(G_3T_2A)_3G_3$ in 70 mM NaCl (△). *B*, CD melting profiles of $d(G_3T_2AG_3)$ in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing 70 mM NaCl (○) and 70 mM KCl (●).

four-stranded structure for the sequences $d(G_4T_nG_4)$, n = 1–4 (27). In order to examine if this observation is valid for the human telomeric sequences where A residues are in the loop, we have carried out temperature-dependent CD experiments for $d(T_2AG_3)_4$ and $d(G_3T_2AG_3)$ in 70 mM KCl (Fig. 4, *A* and *B*). The oligonucleotide $d(T_2AG_3)_4$ shows a positive band around 295 nm with a shoulder around 270 nm indicating that even in K^+ solution a major fraction of molecules adopts an antiparallel G quartet structure. In contrast, the 1.5 copy sequence gives both 295 nm as well as 260 nm positive bands indicating the presence of some amount of parallel tetraplex structure.

In addition, the analysis of $d(T_2AG_3)_4$ in non-denaturing polyacrylamide gel containing 70 mM KCl (Fig. 5) confirmed that a major portion of four copy human telomeric repeat migrates much faster than the single strand $d(T)_{12}$ and 12-bp duplexes in accordance with the mobility of an intramolecular folded G quartet structure. Thus, we conclude that even in the presence of K^+ ion a major portion of $d(T_2AG_3)_4$ molecules in solution adopt antiparallel intramolecular G quartet structures.

Stability of the G Quartet Structure Depends on the Copy Number and the Cation—It was of interest to dissect the various structural contributions toward the overall stability of the G quartet structures formed by the telomeric oligonucleotides. The thermal stability of human telomeric sequences $d(G_3T_2AG_3)$, $d(G_3T_2A)_3G_3$, and $d(T_2AG_3)_4$ have been investigated in 70 mM Na^+ or K^+ containing buffer by monitoring the ellipticity at 295 nm as a function of temperature. The melting curves are shown in Fig. 6 (*A* and *B*). The oligonucleotide $d(T_2AG_3)_4$ which forms an antiparallel intramolecular G quartet structure melts at 49 °C in 70 mM NaCl, whereas its K^+-induced structure melts at 63 °C (Fig. 6*A*). Thermodynamic parameters obtained by van't Hoff analysis of melting profiles

TABLE I
Thermodynamic parameters for the formation of Na^+ and K^+-induced G-quartet structures of $d(TTAGGG)_4$

Total strand concentration 7.9×10^{-6} M.

Cation M^+	Ionic radius	M–O distance	T_m	ΔH	ΔS	ΔG^a
	A		°C	*kcals/mol*	*cal/deg/mol*	*kcals/mol*
Na^+	0.96	2.3	49	−38	−119	−2.5
K^+	1.33	2.8	63	−49	−147	−5.2

[a] Calculated at 25 °C.

TABLE II
Thermodynamic parameters for the formation folded back hairpin and intramolecular G quartet structures by truncated human telomeric sequences

Total strand concentration for $d(G_3T_2AG_3)$ is 20.5×10^{-6} M and for $d(G_3T_2A)_3G_3$ is 8.9×10^{-6} M.

Sequence	Cation	T_m	ΔH	ΔS	ΔG^a
		°C	*kcal/mol*	*cal/deg/mol*	*kcal/mol*
$d(G_3T_2AG_3)$	Na^+	31	−44	−124	−7.1
$d(G_3T_2AG_3)$	K^+	42	−38	−100	−8.2
$d(G_3T_2A)_3G_3$	Na^+	55	−35	−107	−3.1

[a] Calculated at 25 °C.

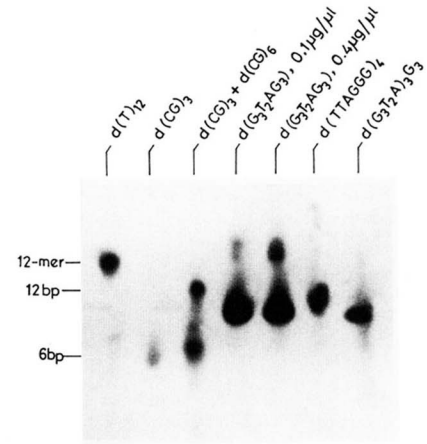


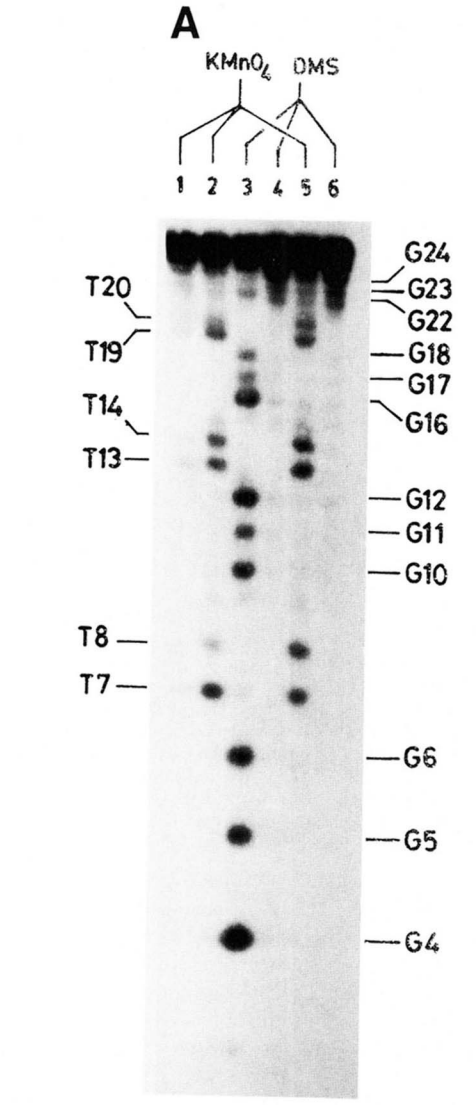
FIG. 7. **Analysis of four copy and truncated human telomeric sequences on 12% non-denaturing gel containing 70 mM NaCl.** The retarded band of $d(G_3T_2AG_3)$ increases in intensity with increase in strand concentration.

for the intramolecular and hairpin dimer structures are given in Tables I and II. K^+-induced intramolecular G quartet structure of the four copy $d(T_2AG_3)_4$ sequence is more stable than its sodium counterpart by a ΔG of 2.7 kcal/mol. This additional stability of the K^+-induced structure can be attributed to the size selective stabilization of the G tetrads by monovalent cations as proposed for the 5′ and 3′ GMP gels (12). Under identical conditions, the antiparallel hairpin dimer of 1.5 copy sequence $d(G_3T_2AG_3)$ melts at 31 and 42 °C in 70 mM NaCl and KCl, respectively (Fig. 6*B*). However, K^+-induced structure of $d(G_3T_2AG_3)$ is only marginally more stable (by a ΔG of 1.1 kcals/mol) than its Na^+-induced structure. Thus, the difference in stability between the Na^+- and K^+-induced structures is more pronounced in the case of the four copy sequence. However, truncation of the four copy sequence increases the T_m by 6 °C, and as evident from Table II the contribution for the higher stability of this sequence is entropic in origin due to the removal of dangling nucleotides at the 5′ end. These observations suggest that the 3.5 copy sequence forms an intramolecular G quartet structure which is more compact and stable than the four copy sequence. This is confirmed by the analysis of a 3.5 copy sequence on non-denaturing polyacrylamide gel containing Na^+. As seen in Fig. 7, $d(G_3T_2A)_3G_3$ migrates faster than $d(T_2AG_3)_4$ indicating that the intramolecular G quartet structure of $d(G_3T_2A)_3G_3$ is indeed more compact than that of $d(T_2AG_3)_4$. Thus, the copy number and the cation were found to influence the structure and the stability of G quartet structures formed by these human telomeric oligonucleotides. In order to get insight into the structural origins of this difference in stabilities, we carried out chemical probing experiments which revealed several interesting features.

$KMnO_4$ Reactivity of $d(T_2AG_3)_4$ Reveals Interloop Interactions—The accessibility of 5,6-double bond in thymines to oxidation by $KMnO_4$ is a direct method of determining whether or not the thymines are exposed (45–47). MnO_4^- anion attacks the C=C bond in thymine from one side of the plane of the heterocyclic ring and forms a cyclic intermediate, hydrolysis of which results in a *cis*-5,6-diol derivative. Thus, the ability of MnO_4^- to form a cyclic manganate ester intermediate necessitates thymines to be in a destacked or exposed state (27, 47). We have exploited this property to probe the loop structure in the sodium and potassium ion-induced intramolecular folded G quartet structures of $d(T_2AG_3)_4$. Using $KMnO_4$ probing it was possible to distinguish between parallel four-stranded and antiparallel hairpin dimer structures for $d(G_4T_nG_4)$ where n = 1–4 sequences (27). Recently, using $KMnO_4$ probing an oligonucleotide duplex $d(C_6A_5G)\cdot d(CT_5G_6)$ was found to have a kink at the CA/TG doublet junction which was responsible for the inherent curvature and its unusual electrophoretic mobility (48).

Results of $KMnO_4$ and DMS reactivity of $d(T_2AG_3)_4$ are shown in Fig. 8 (*A* and *B*). DMS reactivity of guanines in no salt buffer is used as a control (*lane 3*). Protection of guanines from methylation at N-7 by DMS in both 70 mM NaCl or KCl (*lanes 4* and *6*) confirms the formation of a structure containing guanine tetrads. The reactivity of thymines in $d(T_2AG_3)_4$ toward $KMnO_4$ in both 70 mM NaCl and KCl solutions (*lanes 2* and *5*) suggests that they are exposed to the solvent thus forming a loop in the intramolecular antiparallel G quartet structure. However, in the Na^+-induced G quartet structure of $d(T_2AG_3)_4$, T8 and T20 are significantly less reactive as compared to their adjacent thymines T7 and T19 (Fig. 8*B*), indicating that T8 and T20 are stacked in the loop region. On the other hand, the pair of thymines T13 and T14 located at the "one-loop" end of the intramolecular G quartet structure are identical in reactivity. The differential reactivity of thymines is observed only at the end of the quadruplex stem where two loops occur side by side indicating that in addition to stacking interactions in the loop region there are interloop interactions as well in the Na^+-induced G quartet structure of four copy human telomeric sequence $d(T_2AG_3)_4$.

In contrast, in the K^+-induced structure this difference in reactivity between adjacent thymines disappears (*lane 5*, Fig. 8*A*), and the $KMnO_4$ reactivities for all the three pairs of thymines are equal regardless of the end where they are present (Fig. 8*B*). Thus, our $KMnO_4$ probing experiments indicate stacking and interloop interactions in the Na^+-induced intramolecular G quartet structure of $d(T_2AG_3)_4$ but not in the K^+-induced structure. These stacking and hence interloop interactions get disrupted exposing the thymines to the solvent when the sample is annealed in the presence of KCl. The higher reactivity of thymines toward $KMnO_4$ in the presence of K^+ is expected only if K^+ binds in the loop region. At the apex of the quadruplex, in the loop region, the phosphodiester backbone makes a sharp turn as a result of which phosphates are oriented very close to each other (27, 29, 30). In order to overcome the intrastrand phosphate-phosphate repulsion, binding of the ions in the loop may be mandatory. In fact, in the crystal structure of $d(G_4T_4G_4)$, a K^+ ion has been located between the second and third thymines where the backbone makes a sharp turn (29). These results indicate that not only the stability but also



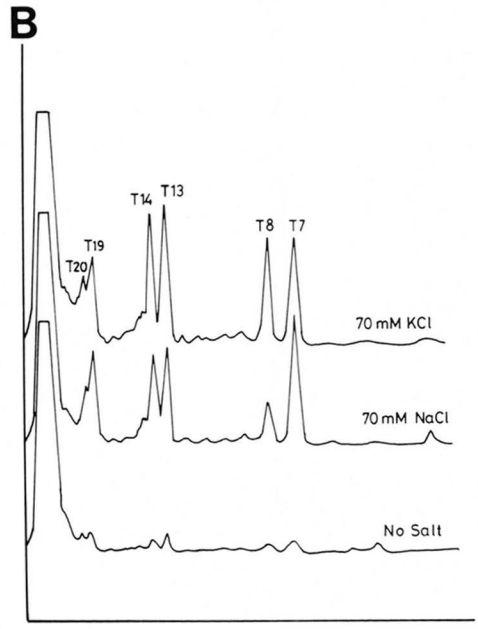


FIG. 8. ***A*, reactivity of four copy human telomeric oligonucleotide $d(T_2AG_3)_4$ with $KMnO_4$ and DMS.** $KMnO_4$ reactivity (*lanes 1, 2,* and *5*) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing no added salt (*lane 1*), 70 mM NaCl (*lane 2*), and 70 mM KCl (*lane 5*). DMS reactivity (*lanes 3, 4,* and *6*) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing no added salt (*lane 3*), 70 mM NaCl (*lane 4*), and 70 mM KCl (*lane 6*). *B*, densitometric scans showing the $KMnO_4$ reactivity difference among the thymines under different conditions.

the overall structure of the quadruplex including the orientation and structure of the loops depend on the counter ion. These observations provided us the impetus to know whether the difference in reactivity of adjacent thymines toward $KMnO_4$ arising due to interloop interactions could be used to predict the orientation of loops in the hairpin dimer.

Loops Are at the Same End in the Hairpin Dimer of $d(G_3T_2AG_3)$—The thymines in $d(G_3T_2AG_3)$ which forms an antiparallel hairpin dimer are probed with $KMnO_4$ to get insight into the orientation of the loops and also to find out if there is any interloop interaction as found in the case of $d(T_2AG_3)_4$. $KMnO_4$ and DMS reactivity of $d(G_3T_2AG_3)$ in 70 mM NaCl and 70 mM KCl are shown in Fig. 9 (*A* and *B*). Complete protection of all guanines from DMS reaction both in the presence of Na^+ as well as K^+ suggests the formation of G tetrads through hydrogen bonding involving N-7 of guanines (*lanes 5* and *7*, Fig. 9*A*). The thymines in the above sequence show reactivity toward $KMnO_4$ under both Na^+ and K^+ conditions (*lanes 3* and *6*, Fig. 9*A*) indicating that they are in the loop region. Interestingly, also in this case we observe a significant difference in reactivity between adjacent thymines T4 and T5 in the sodium-induced hairpin dimer of $d(G_3T_2AG_3)$. As can be seen in Fig. 9*B*, in 70 mM NaCl solution T5 is much less reactive than T4. However, the magnitude of the difference in reactivity is less as compared with that observed in the folded G quartet structure of $d(T_2AG_3)_4$ as evident from the densitometric scans. A similar reactivity difference between adjacent thymines toward $KMnO_4$ is observed in the two TTA loops occurring at the same end of the intramolecular G quartet structure of $d(T_2AG_3)_4$. From these results we conclude that in Na^+ solution the majority of the $d(G_3T_2AG_3)$ molecules adopt a hairpin dimer with loops at the same end as shown in Fig. 10*c*, with a minor proportion being present which does not have loops at the same end (Fig. 10 (*a* and *b*)). This interpretation is further supported by DEPC probing and concentration-dependent gel mobility and CD studies. Model building and molecular mechanics studies carried out[2] to explore the various possibilities of loop orientation and interloop interactions also buttressed our notion of the possible hydrogen bonding potentialities in the loop region of $d(G_3T_2AG_3)$, as both complementary nucleotides A and T are present in the loop.

N-7 Position of Adenines in the Hairpin Dimer of $d(G_3T_2AG_3)$ Is Hydrogen-bonded—Diethylpyrocarbonate carbethoxylates purines (A > > G) preferentially at N-7, when it is exposed to the solvent either in a single-stranded state or in *Z* conformation (49), whereas purines in B-DNA are resistant to DEPC, as N-7 positions are not exposed to the solvent (49). The unpaired adenines occurring in the loop region of supercoil-stabilized cruciforms have also been shown to be hypersensitive to DEPC (50). Thus, DEPC can be used as a probe to determine whether N-7 positions of purines are exposed to the solvent or not. Guanines involved in the tetrad formation are resistant to DEPC modification indicating that N-7 positions of guanines are Hoogsteen hydrogen bonded in the G-tetrad (23). Thus, DEPC has proven to be a reliable and sensitive reagent to determine if N-7 positions of purines are involved in hydrogen bonding or not.

The results of DEPC reactivity experiments carried out at 4 and 22 °C on $d(G_3T_2AG_3)$ under various conditions are shown in Fig. 11 (*A-D*). As expected, the adenine residue in $d(G_3T_2AG_3)$ is unreactive toward DEPC (*lane 2*) when it is duplexed with the

[2] D. Mohanty and M. Bansal, manuscript in preparation.

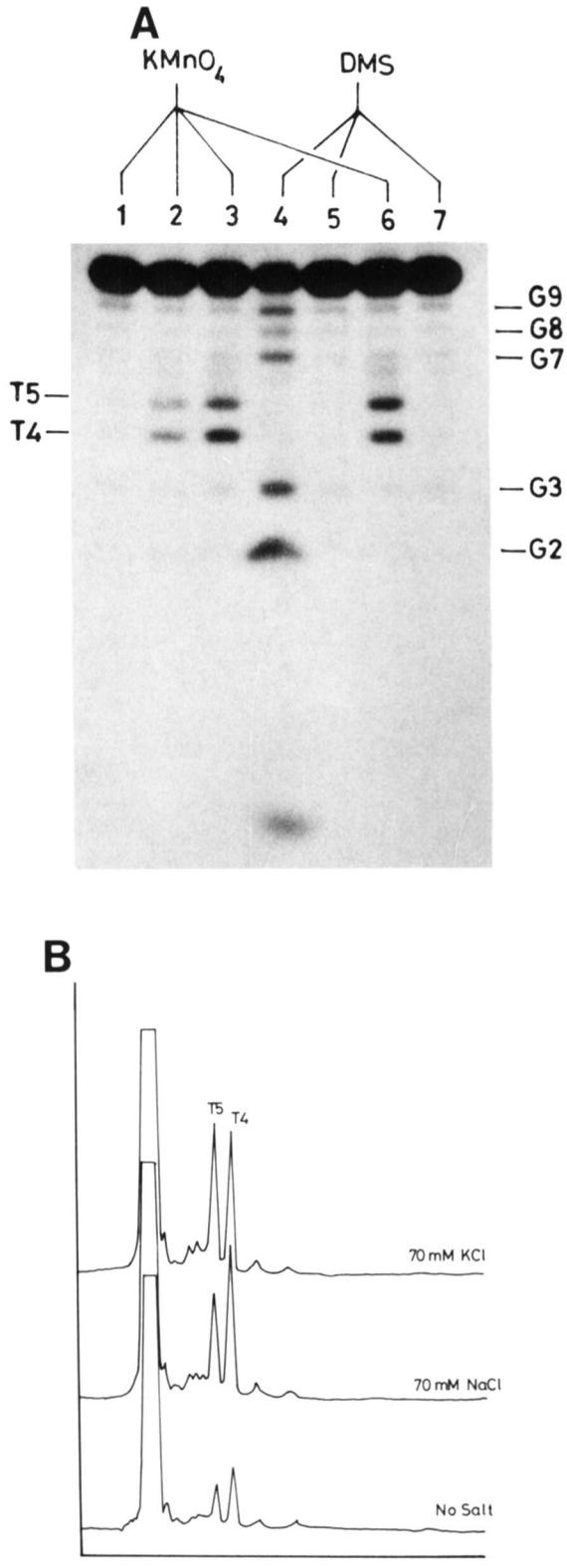


FIG. 9. *A*, reactivity of truncated human telomeric oligonucleotide $d(G_3T_2AG_3)$ toward $KMnO_4$ and DMS. $KMnO_4$ reactivity (*lanes 1–3* and 6) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing 70 mM NaCl in the presence of complementary oligonucleotide $d(C_3AT_2)_3$ (*lane 1*), no salt (*lane 2*), 70 mM NaCl (*lane 3*), and 70 mM KCl (*lane 6*). DMS reactivity (*lanes 4*, *5*, and *7*) in cacodylate buffer containing no added salt (*lane 4*), 70 mM NaCl (*lane 5*), and 70 mM KCl (*lane 7*). *B*, densitometric scans showing the $KMnO_4$ reactivities of thymines in $d(G_3T_2AG_3)$ under different conditions.

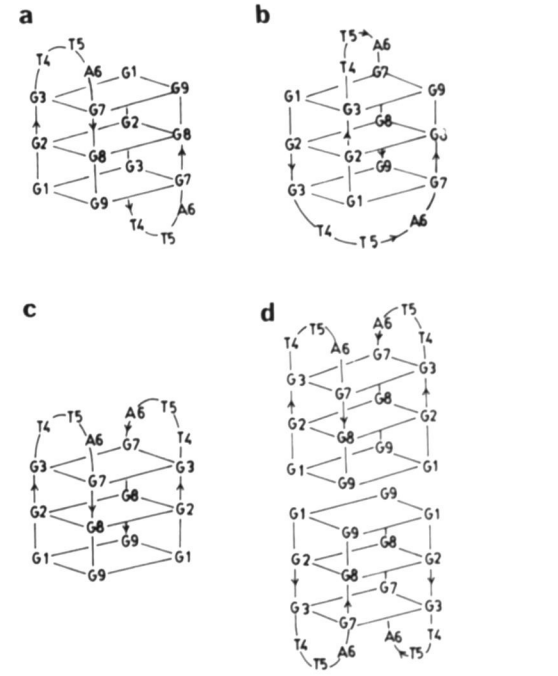


FIG. 10. **Schematics showing the various possible hairpin dimers with different loop orientations and topologies.** *a*, hairpin dimer with loops at opposite ends of the tetrad stem oriented along the edges; *b*, loops at opposite ends oriented diagonally across the quadruplex stem; *c*, loops at the same end of the quadruplex stem; *d*, super structure containing two units of hairpin dimer formed via stacking of G tetrads end-to-end.

complementary C-rich oligonucleotide $d(C_3A_2T)_3$. At 4 °C, all the guanine residues in $d(G_3T_2AG_3)$ are protected from DEPC modification both in the presence of Na^+ and K^+ indicating that guanines have their N-7 positions hydrogen bonded, a prerequisite for G tetrad formation. Surprisingly, the single adenine residue A_6 in $d(G_3T_2AG_3)$ is also completely protected from DEPC reaction at 4 °C in 70 mM NaCl (*lane 4*) in spite of being in the loop and exposed to the solvent. This strongly suggests that N-7 of A is hydrogen-bonded. At 22 °C, A of $d(G_3T_2AG_3)$ in 70 mM NaCl strongly reacts with DEPC (Fig. 11*B*), suggesting that hydrogen bonds are disrupted making the N-7 accessible. (It is surprising to note that adenine residues in $d(G_3T_2AG_3)$ under no salt conditions also do not exhibit DEPC reactivity at 4 °C which might be due to the aggregation of the oligonucleotide under these conditions.) Since the hairpin dimer of $d(G_3T_2AG_3)$ has a T_m of 31 °C, it appears that at 22 °C loop hydrogen bonds are disrupted prior to the global melting of G tetrad stem. Thus, from DEPC probing experiments we conclude that in an Na^+-induced hairpin dimer of $d(G_3T_2AG_3)$ the N-7 position of A is hydrogen-bonded. On the other hand, even at 22 °C, the DEPC reactivity of A in the K^+-induced structure is lower than that in the Na^+-induced structure. This can be attributed to the higher stability of the K^+-induced structure.

G Quartet Structure of $d(G_3T_2AG_3)$ Forms a Stacked Dimer in Solution—Having established that the TTA loops are at the same end in the antiparallel hairpin dimer of $d(G_3T_2AG_3)$ with intra- and interloop interactions between the A and T residues, we realized the possibility that the two hairpin dimers can also stack end-to-end via G tetrad planes to form a super structure-containing dimer of a dimer as shown in Fig. 10*d*. The oligonucleotide $d(G_3T_2AG_3)$ has been examined on a non-denaturing polyacrylamide gel containing 70 mM NaCl at two different concentrations. As can be seen in Fig. 7, a major portion of the 9-mer $d(G_3T_2AG_3)$ migrates ahead of 12-mer single strand $d(T)_{12}$ and 12-bp duplex $d(CG)_6$. However, a significant fraction

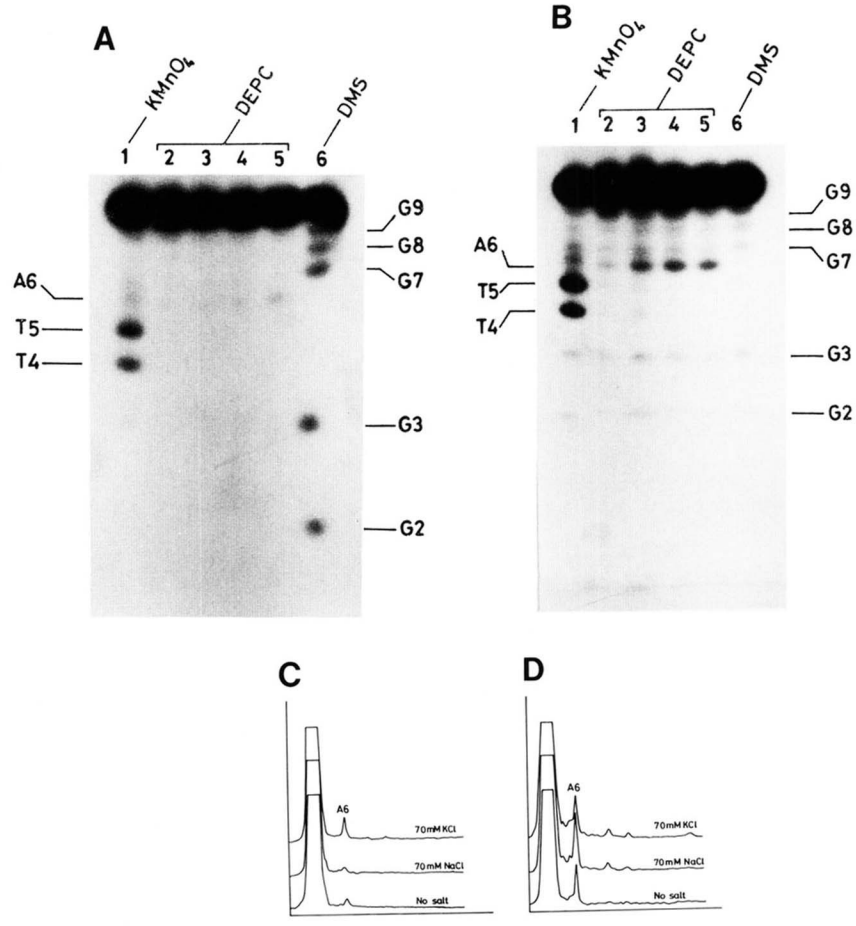


FIG. 11. **Reactivity of truncated human telomeric oligonucleotide d($G_3T_2AG_3$) with DEPC at 4 °C (*A*) and 22 °C (*B*).** DEPC reactivities (*lanes 2–5*) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing 70 mM NaCl in the presence of complementary oligonucleotide d($C_3A_2T)_3$ (*lane 2*), no added salt (*lane 3*), 70 mM NaCl (*lane 4*), and 70 mM KCl (*lane 5*). $KMnO_4$ reactivity in 70 mM KCl (*lane 1*). *Lane 6* is the Maxam-Gilbert DMS reaction done at room temperature to identify G residues. Densitometric scans used for comparison of the extent of reactivities of the adenine residue in d($G_3T_2AG_3$). *C*, 4 °C; *D*, 22 °C.

of the molecules move slower than the d(T_{12}) single strand and d$(CG)_6$-bp duplex. The intensity of the slow moving band increases as the concentration of the oligonucleotide d($G_3T_2AG_3$) is increased from 0.1 to 0.4 μg/μl during annealing. This can be interpreted only as the aggregation of G quartet structures containing two hairpin dimers into a super structure consisting of two such units, probably through stacking along the ends through G planes as shown in Fig. 10*d*.

CD spectra of d($G_3T_2AG_3$) were measured at increasing concentrations of oligonucleotide in order to see any dimerization which will be reflected in the molar ellipticity at 295 nm. The results of the concentration-dependent CD experiments are shown in Fig. 12. There is an increase in 295 nm ellipticity with an increase in strand concentration. The *inset* shows the change in the 295 nm ellipticity as a function of DNA concentration. These observations support our conclusion drawn from non-denaturing gel that the 1.5 copy sequence indeed forms a dimer of a dimer in solution. The concentration of oligonucleotide used in all the previous experiments in this paper corresponds to a state where the predominant species is just a hairpin dimer. But our concentration-dependent experiments suggest that at a higher oligonucleotide concentration d($G_3T_2AG_3$) adopts a dimer of a dimer in the presence of Na^+ ions with TTA loops oriented at the same end of the G tetrad stem.

N-7 of Adenines in the Folded G Quartet Structure of d$(T_2AG_3)_4$ Are Partially Protected—The results of DEPC modification of d$(T_2AG_3)_4$ done at 4 and 22 °C are shown in Fig. 13 (*A-D*). When the DEPC modification is performed in the presence of complementary C-rich strand d($C_3A_2T)_3$, the central adenines (A9 and A15) do not exhibit any reactivity, whereas adenines (A3 and A21) show equal reactivity (*lane 2*). Since the oligonucleotide d($C_3A_2T)_3$ has only three repeats of complementary sequence, after duplexation there is an equal probability

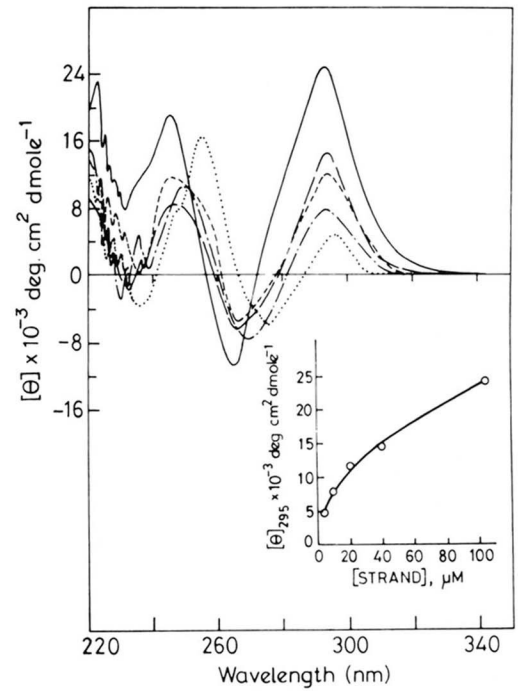


FIG. 12. **Concentration-dependent CD spectra of d($G_3T_2AG_3$) in 2 mM sodium cacodylate, 0.1 mM EDTA, pH 7.2, 70 mM NaCl at 22 °C.** ····, 4 μM; –·–·–, 10 μM; - - - -, 20 μM; – –, 40 μM; and ——, 100 μM. *Inset* shows the variation in molar ellipticity at 295 nm as a function of strand concentration.

that any of the two adenines, either A3 or A21, would be single-stranded. We have used the DEPC reactivity of these adenines (A3 and A21) as external controls, and the sum of the reactivi-

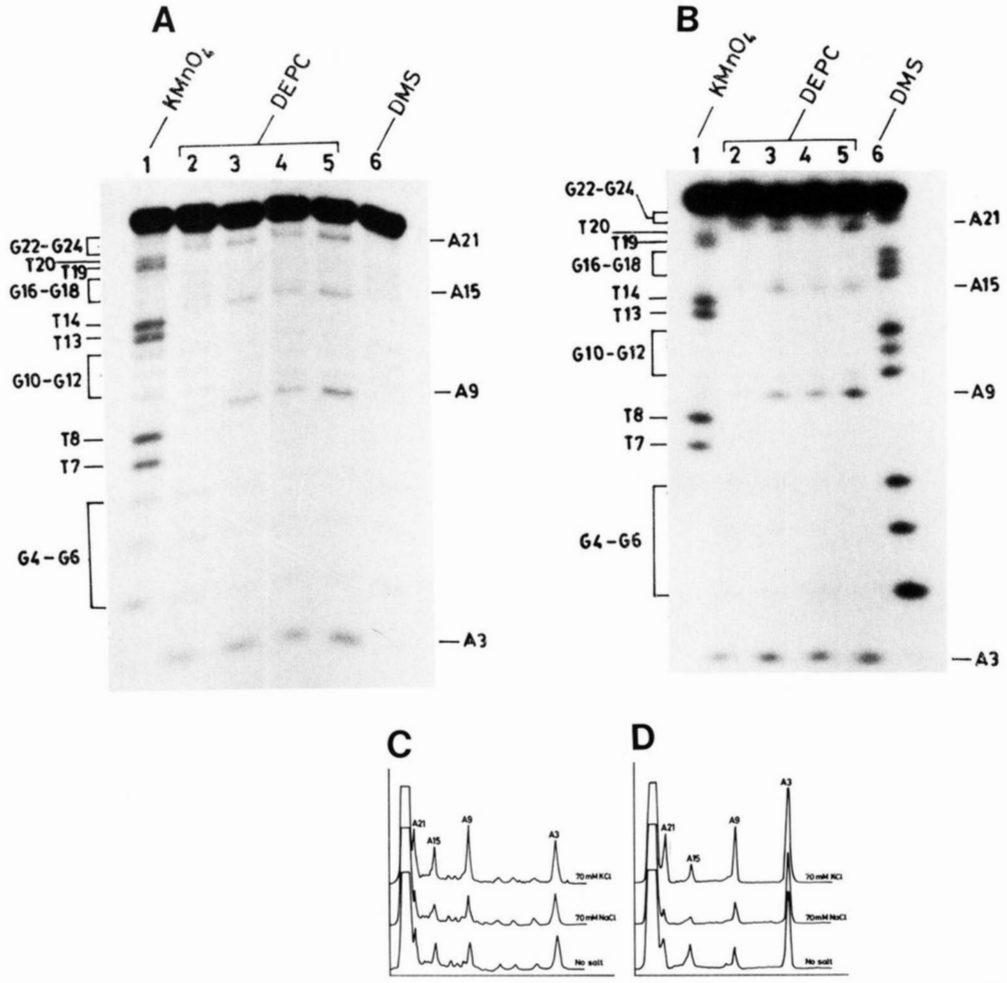


FIG. 13. **DEPC reactivity of four copy human telomeric oligonucleotide $d(T_2AG_3)_4$ at 4 °C (*A*) and 22 °C (*B*).** DEPC reactivities (*lanes 2–5*) in 2 mM cacodylate, 0.1 mM EDTA, pH 7.2, containing 70 mM NaCl in the presence of the complementary strand $d(C_3A_2T)_3$ (*lane 2*), no added salt (*lane 3*), 70 mM NaCl (*lane 4*), and 70 mM KCl (*lane 5*). *Lane 1* is the $KMnO_4$ reaction in 70 mM KCl, and *lane 6* is the Maxam-Gilbert DMS reaction done at room temperature to identify G residues. Since in 70 mM KCl, at room temperature, all the thymines in the human telomeric oligonucleotide exhibited reactivity, this has been used as a control. Densitometric scans showing the relative reactivities of adenines with DEPC at 4 °C (*C*) and 22 °C (*D*) in $d(T_2AG_3)_4$.

ties of A3 and A21 in the presence of complementary sequence matches with the DEPC reactivity of A3 for the G-rich strand in Na^+ or K^+ indicating that A3 in the intramolecular G quartet structure is indeed single-stranded. Thus, A3 serves as an internal reference for comparison of the reactivities of other loop adenines. As seen in Fig. 13 (*A* and *B*) in 70 mM NaCl, adenine (A15) at the "one-loop" end is significantly less reactive as compared with A3 (*lane 4* in Fig. 13, *A* and *B*) and the other adenines (A9 and A21) at the "two-loop" end indicating that N-7 of A15 is hydrogen-bonded. The adenines (A9 and A21) exhibit slightly higher DEPC reactivity as compared with the adenine (A15) at the one-loop end (*lane 4*). DEPC reactivities of these adenines are still much less with respect to the internal control A3. Thus, the loop adenines are protected to a different extent depending on whether they are at the one- or two-loop end. These observations indicate the existence of different kinds of stacking and hydrogen bonding interactions between A and T residues at each end of the intramolecular G quartet structure. A similar trend is observed for the K^+-induced G quartet structure except that adenines are relatively more reactive. This is consistent with the results of our $KMnO_4$ probing that K^+ ions bind in the loop and disrupt stacking and hydrogen bonding interactions rendering adenines more reactive. At 22 °C, a similar trend in DEPC reactivity is observed (Fig. 13*B*) except for the slight increase as compared with that at 4 °C.

Intra- and Interloop Hydrogen Bonding in the TTA Loops—Our results on chemical probing experiments thus suggest intra- and interloop hydrogen bonding interactions in the TTA loops of the Na^+-induced intramolecular and hairpin dimeric G quartet structures of $d(T_2AG_3)_4$ and $d(G_3T_2AG_3)$, respectively. At the two-loop end of the intramolecular G quartet structure of $d(T_2AG_3)_4$, A and T residues in the TTA loop can form a ATAT tetrad if the loop goes along the edge of the G tetrad. On the other hand, if the loop goes diagonally across the tetrad stem, then the tetrad would be AATT as a result of loops running parallel. Thus, different loop orientations result in two altogether different tetrad arrangements in the two-loop end of the intramolecular G quartet structure of $d(T_2AG_3)_4$. Although the adenines in the intramolecular G quartet structure are exposed to the solvent by being in the loop, we observe reduced reactivity indicating that their N-7 positions are involved in hydrogen bonding interactions. Various possible intra- and interloop hydrogen bonding arrangements that are possible in the tetrads of TTA loops are shown in Fig. 14 (*a-d*). At the one-loop end of the intramolecular G quadruplex, the Hoogsteen hydrogen bonding between A15 and T13 could explain the protection of A15 from DEPC (Fig. 14*c*). However, interaction of thymine residues of the TTA tail of the intramolecular G quartet structure with the N-7 of A15 at the one-loop end cannot be ruled out. Nevertheless, as shown in Fig. 7 the slower electrophoretic

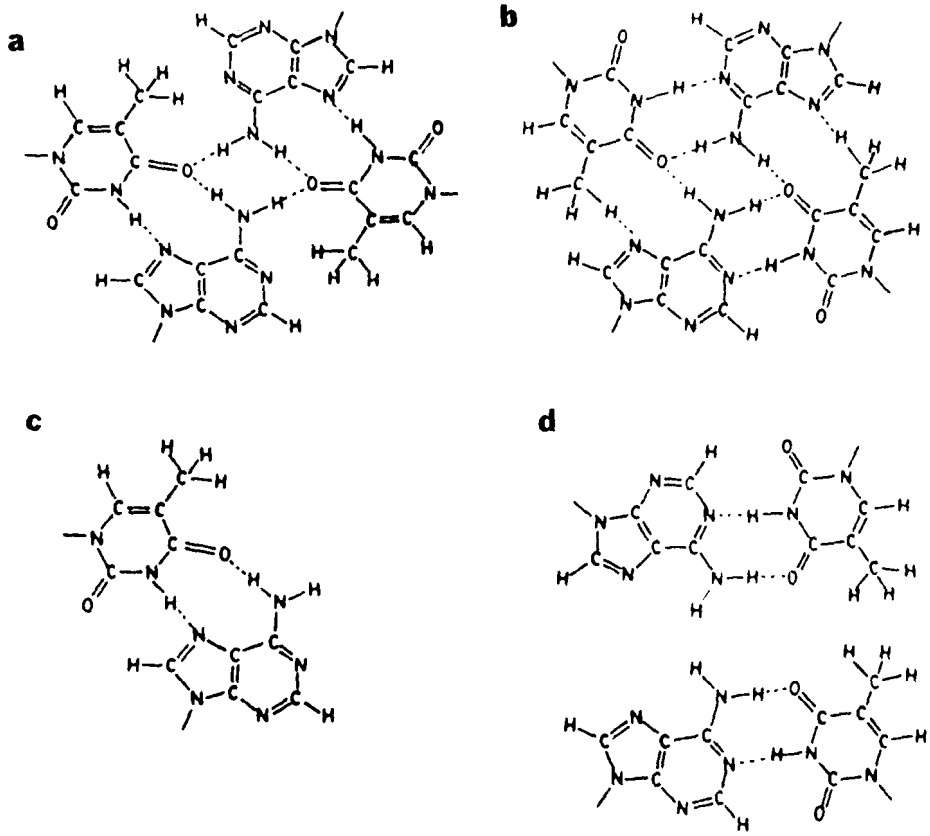


FIG. 14. **Intra- and interloop hydrogen bonding schemes for the TTA loops.** *a*, ATAT tetrad with intraloop Hoogsteen type hydrogen bonding when all adjacent strands are antiparallel as in the hairpin dimer containing loops at the same end. *b*, interloop Watson-Crick hydrogen bonding with CH_3 group of thymine protecting the N-7 of adenine which accounts for the partial protection of adenines (A9 and A21) at the two-loop end in an intramolecular G quartet structure with all adjacent strands arranged antiparallel. *c*, Hoogsteen hydrogen bonding between (T13 and A15) at the one-loop end of the intramolecular G quartet structure of $d(T_2AG_3)_4$. *d*, intraloop Watson-Crick hydrogen bonding for the TTA loops when the loop goes diagonally across the G tetrad stem where N-7 cannot participate in hydrogen bonding.

mobility of $d(T_2AG_3)_4$ compared with that of $d(G_3T_2A)_3G_3$ indicates that the 5′ TTA tail in the intramolecular G quartet structure is not folded, and hence either T1 or T2 cannot be involved in hydrogen bonding with A15. This is further supported by the higher stability of $d(G_3T_2A)_3G_3$ than $d(T_2AG_3)_4$ which is entropic in origin. Therefore, the protection of A15 at the one-loop end from DEPC is less likely to arise from hydrogen bonding interactions between thymines in 5′ TTA tail and A15 in the TTA loop. Intraloop Watson-Crick hydrogen bonding between A and T residues with hydrogen of the CH_3 group interacting with N-7 of adenine in the other loop could explain the partial protection of adenines A9 and A21 at the two-loop end (Fig. 14*b*) observed for the intramolecular folded G quartet structure of $d(T_2AG_3)_4$. This type of C–H...N interaction would be weaker than the conventional N–H...N and O–H...N hydrogen bonds. On the other hand, the complete protection of adenines in the hairpin dimer of $d(G_3T_2AG_3)$ could be explained by Hoogsteen N–H.....N Hoogsteen hydrogen bonding across the TTA loops as this is stronger than the former (Fig. 14*a*). Another possibility which can explain the partial protection of adenines at the two-loop end is that the four copy sequence adopts a mixture of edge- and diagonal-looped structures in solution. In the diagonal-looped structure, as a result of alternation of parallel and antiparallel adjacent strands, the loop tetrad will be AATT. In such a tetrad, N-7 of adenines cannot be involved in hydrogen bonding as shown in Fig. 14*d*. But if the loop goes diagonally, then the distance between A15 and T13 would be too large to accommodate a hydrogen bond,[2] thus A15 would have exhibited pronounced DEPC reactivity. A15 exhibits reduced DEPC reactivity *vis-a-vis* other loop adenines and it seems possible that in solution $d(T_2AG_3)_4$ forms a mixture of both diagonal- and edge-looped structures. The observed lesser reactivity of A15 could also be due to stacking of A15 at the one-loop end. Thus, our present results strongly suggest that the intramolecular G quartet structure of human telomeric repeat is highly polymorphic, and the polymorphism arises mainly from the variability in the TTA loop orientations and interactions leading to a mixture of both diagonal- and edge-looped structures in solution. After the completion of the present article, the solution structure for $d(AG_3(T_2AG_3)_3$ was reported by Wang and Patel (51). From detailed NMR and molecular dynamics studies, they have arrived at a diagonal-looped intramolecular G quartet with loop adenines stacked over the guanines for $d(AG_3(T_2AG_3)_3$ in Na^+ solution (51). However, from NMR data no intra- and interloop hydrogen bonding interactions in the TTA loops could be proposed which was made possible by our chemical probing data. Kallenbach and co-workers (52) showed that the *Arabidopsis* telomeric sequences $d(T_3AG_3)_4$ and $d(T_3AG_3)_2$ form more stable structures than the corresponding oligonucleotides $d(T_4G_3)_4$ and $d(T_4G_3)_2$ where adenines are replaced by thymines. Similar intra- and interloop hydrogen bonding interactions reported here for the human telomeric sequences could be responsible for such higher stability of adenine containing *Arabidopsis* telomeric oligonucleotides. Thus, from our present chemical probing results we conclude that the four copy human telomeric sequence $d(T_2AG_3)_4$ forms a mixture of both edge- and diagonal-looped intramolecular G quartet structures in solution with intra- and interloop hydrogen bonding interactions in the TTA loops. We have established that the loops are at the same end of the G tetrad stem in the hairpin dimer of 1.5 copy sequence. As the N-7 position of A6 is completely protected from DEPC modification, there has to be a ATAT tetrad leading us to the conclusion that all adjacent strands are antiparallel in the hairpin dimer. Therefore, we conclude that $d(G_3T_2AG_3)$ adopts an antiparallel hairpin dimer with TTA loops forming intra- and interloop Hoogsteen type hydrogen bonding which are at the same end of the tetrad stem, and $d(T_2AG_3)_4$ forms a mixture of both edge- and diagonal-looped intramolecular G quartet structures in Na^+ solution.

Loop Sequence, Number of G Tracts, and the Monovalent Cation Decide the Overall Structure of the Telomeric Repeat—Results of our temperature-dependent CD experiments suggest that the human telomeric oligonucleotides reported here adopt an antiparallel G quartet structure, either a hairpin dimer or an intramolecular G quartet structure with TTA loops depending upon the number of G tracts in the presence of Na^+. $KMnO_4$ reactivity experiments have indicated that the TTA loops are at the same end of the tetrad stem in the hairpin dimer of 1.5 copy sequence and also the existence of interloop interactions. In addition, we have demonstrated the versatility of $KMnO_4$ footprinting in deducing the topology and orientation of TTA loops in the quadruplex structures of human telomeric repeats. As a consequence of TTA loops being oriented at the same end of the tetrad stem, intra- and interloop hydrogen bonding interactions further contribute toward the overall stability of these quadruplex structures. Thus, results reported in this paper allow us to understand the factors which control the formation of various G quartet structures with different loop orientations and topologies depending on the pathway of folding. There are similarities as well as striking differences between the high resolution x-ray crystal structure (29) and the solution NMR structure of $d(G_4T_4G_4)$ (30). Although in both x-ray and NMR structures the thymine loops are oriented at the opposite ends of the tetrad stem, the former has planes of loops parallel to each other resulting from loops spanning the edges of the tetrad planes. On the other hand, in the NMR structure loops are oriented diagonally across the tetrad plane, and the planes of the loops are perpendicular to each other indicating that diagonal-looped structure of $d(G_4T_4G_4)$ cannot be formed by simple dimerization of hairpins. Here we report a G quartet structure with a radically different loop structure for $d(G_3T_2AG_3)$ with loops oriented at the same end of the tetrad stem, apparently along the edges of the tetrad planes. A folding scheme is proposed as shown in Fig. 15 (*A* and *B*) to rationalize the formation of various quadruplex structures with different loop orientations and topologies depending on the sequence, number of guanine tracts, environmental factors, and concentration of oligonucleotide used in the experiment. From the results reported here and elsewhere (22, 29, 30), it appears that the loop sequence determines the orientation of loops and whether they are going to be at the same or opposite ends. In the case of the 1.5 copy human telomeric sequence, the presence of complementary nucleotides A and T in the intervening region results in the formation of a hairpin dimer with loops at the same end, due to interloop hydrogen bonding interactions as indicated by the DEPC reactivity experiments. This arrangement facilitates further dimerization of hairpin dimers through stacking of G tetrad planes via end-to-end as shown in Fig. 15*A*. In solution the formation of either edge- or diagonal-looped intramolecular G quartet structure for 3.5 and 4 copy sequences depends on where the initial G:G-base pairing nucleates. Thus, as shown in Fig. 15*B* nucleation between central G tracts in the four copy sequence would lead only to edge-looped structure. On the other hand, if the initiation of base pairing takes place in the adjacent pairs of G tracts, then the end product could be either an edge- or diagonal-looped structure. For two and 1.5 copy sequences at low oligonucleotide concentration, hairpin formation is favored leading to only an edge-looped structure, whereas at a higher oligonucleotide concentration intermolecular interactions between strands could result in a diagonal-looped structure. The fact that a diagonal-looped structure is not a simple hairpin dimer necessitates intermolecular interactions between the strands for its formation. The folding pathway shown here (Fig. 15*A*) explains the differences seen in the crystal and the NMR solution structure for $d(G_4T_4G_4)$. Thus,

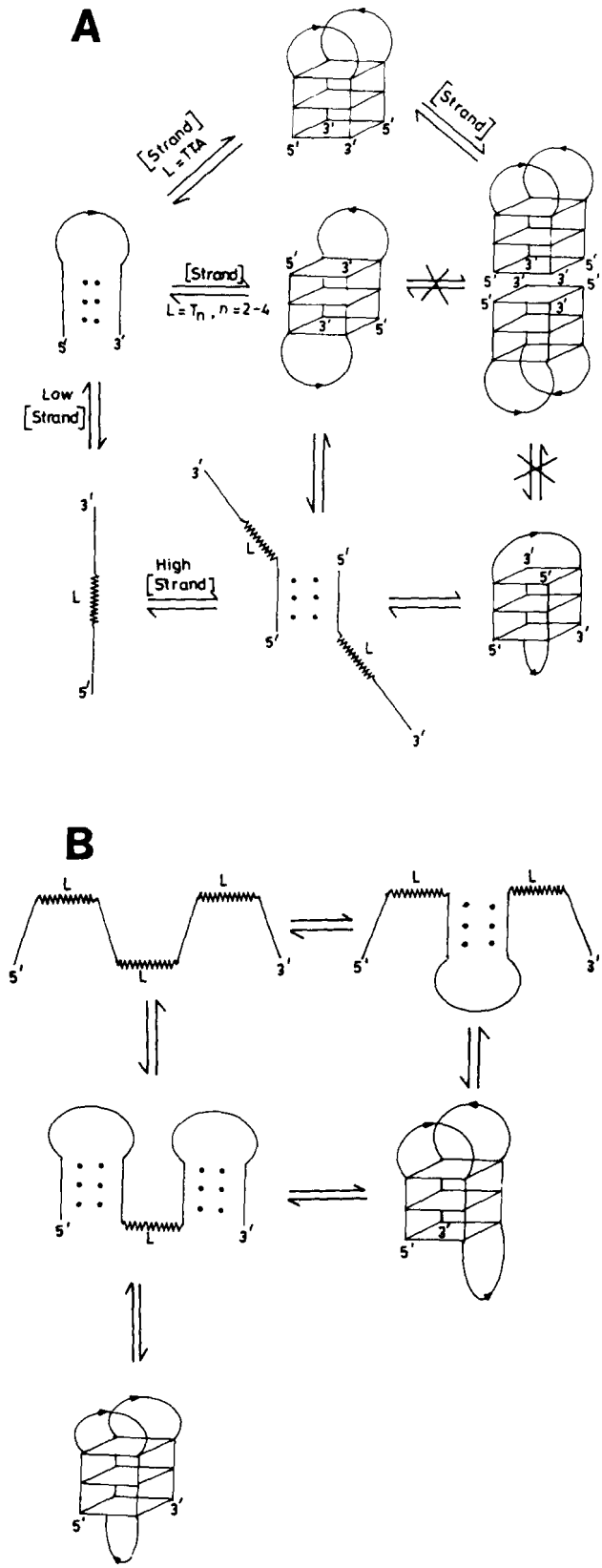


FIG. 15. **Folding pathway proposed to rationalize the formation of G quadruplex structures for sequences with two tracts of guanine residues (*A*) and with four tracts of guanine residues (*B*) leading to G quartet structures with different loop orientations and topologies.**

loop sequence, cation, and the concentration of oligonucleotide decide the overall structure and stability of the telomeric sequences.

Conclusions—Results reported in this paper support the formation of intra- and intermolecular antiparallel G quartet structures induced by sodium and potassium for the human telomeric repeats $d(T_2AG_3)_4$ and $d(G_3T_2AG_3)$ which differ in their loop structure and stability. Here we have shown that the overall structure, loop conformation, and thermodynamic stabilities of G quartet structures formed by the telomeric sequences are governed by many factors like cation type, number of repeats, and the number of intervening hydrophilic residues which form spacers or loops between the G tracts. These results are interesting in light of the observation that the ability of telomerase to act on G-rich telomeric sequences very much depends on the number of repeats and the cation type which promotes folding and alters stability.

What is the biological relevance of these structures *in vivo*? The intramolecular G quartet structure formed by the *Oxytricha* telomeric repeat has been shown to be a very poor substrate for telomerase (53). Nonetheless, such structures may be important in the regulation of telomere elongation *in vivo*, as suggested by Cech and co-workers (53). The presence of two repeats of the G-rich strand at the telomeric ends would allow the pairing of chromosomes through dimerization of hairpins, whereas the occurrence of four such repeats at the chromosomal end would prevent such pairing by the formation of intramolecular folded G quadruplex structure. The C-rich RNA strand in telomerase assembly can destabilize the G quartet structure by forming Watson-Crick duplex. This event in turn might depend on the stability of G quartet structure which can be modulated by the fluctuation in the Na^+/K^+ concentration in the cell and the number of such repeats present at the chromosomal terminus. Recently, a β-subunit of *Oxytricha* telomere binding protein has been shown to promote and stabilize the formation of G quartet structure by inducing dimerization of the duplexes bearing two repeats of T_4G_4 as 3′ overhangs (54). Such proteins in the nucleus, may act as mediators in regulating the end replication by binding to a G quartet structure transiently formed and protecting it from the invasion of RNA strand of telomerase. Earlier it was shown that short stretches of $(TG)_n$ sequences, abundant in eukaryotic genome (55), with the potential to form paranemic DNA joints such as alternating right-handed B and left-handed *Z* structures (56) detected in the vicinity of translocation breakpoints are associated with homologous recombination (57, 58). In addition to this, $d(G_mN_nG_m)$ sequence motifs have also been found to occur in the immunoglobulin switch regions and recombination hotspots (19–21) suggesting the plausible involvement of G quartet structures in genetic recombination. It seems plausible that hairpins stabilized by G:G-base pairing could loop out from two duplexes involved in recombination and form a G quartet structure which might initiate the strand exchange between the duplexes.

Acknowledgments—We thank the Department of Biotechnology (India) for the oligonucleotide synthesis facility at Indian Institute of Science, Dr. M. Bansal and D. Mohanty for useful discussions, and Sowmya Raghavan for editing the manuscript.

REFERENCES

1. Blackburn, E. H., and Szostak, J. W. (1984) *Annu. Rev. Biochem.* **53,** 163–194
2. Blackburn, E. H. (1990) *J. Biol. Chem.* **265,** 5919–5921
3. Blackburn, E. H. (1991) *Nature* **350,** 569–573
4. Blackburn, E. H., and Gall, J. G. (1978) *J. Mol. Biol.* **120,** 33–53
5. Klobutcher, L. A., Swanton, M. T., Donnini, P., and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 3015–3019
6. Oka, Y., Shiota, S., Nakai, S., Nishida, Y., and Okubo, S. (1980) *Gene (Amst.)* **10,** 301–306
7. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 6622–6626
8. Greider, C. W., and Blackburn, E. H. (1985) *Cell* **43,** 405–413
9. Greider, C. W., and Blackburn, E. H. (1989) *Nature* **337,** 331–337
10. Shippen-Lentz, D., and Blackburn, E. H. (1989) *Mol. Cell. Biol.* **9,** 2761–2764
11. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) *Proc. Natl. Acad. Sci. U. S. A.* **48,** 2013–2018
12. Pinnavaia, T. J., Miles, H. T., and Becker, E. D. (1975) *J. Am. Chem. Soc.* **97,** 7198- 7200
13. Sasisekharan, V., Zimmermann, S. B., and Davies, D. R. (1975) *J. Mol. Biol.* **92,** 171- 179
14. Arnott, S., Chandrasekharan, R., and Martilla, C. M. (1974) *Biochem. J.* **141,** 537- 543
15. Zimmermann, S. B., Cohen, G. H., and Davies, D. R. (1975) *J. Mol. Biol.* **92,** 181–192
16. Guschlbauer, W., Chantof, J. F., and Thiele, D. (1990) *J. Biomol. Struct. Dyn.* **8,** 491–511
17. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., and Blackburn, E. H. (1987) *Cell* **51,** 899–908
18. Sen, D., and Gilbert, W. (1988) *Nature* **334,** 364–366
19. Pluta, A. F., and Zakian, V. A. (1989) *Nature* **337,** 429–433
20. Cherry, J. M., and Blackburn, E. H. (1985) *Cell* **43,** 747–758
21. Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A., and Williams, K. (1985) *Cell* **43,** 759–768
22. Williamson, J. R., (1993) *Curr. Opin. Struct. Biol.* **3,** 357–362
23. Sundquist, W. I., and Klug, A. (1989) *Nature* **342,** 825–829
24. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) *Cell* **59,** 871–880
25. Panyutin, I. G., Kovalsky, O. I., Budowsky, E. I., Dickerson, R. E., Rikhirev, M. E., and Lipanov, A. A. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 867–870
26. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1991) *J. Biomol. Struc. Dyn.* **8,** 15 (abstr.)
27. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) *Nucleic Acids Res.* **20,** 4061–4067
28. Sen, D., and Gilbert, W. (1990) *Nature* **344,** 410–414
29. Kang, C., Zhang, X., Ratliff, R., Moyzis, R., and Rich, A. (1992) *Nature* **356,** 126–131
30. Smith, F. W., and Feigon, J. (1992) *Nature* **356,** 164–168
31. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3745–3749
32. Cheong, C., and Moore, P. B. (1992) *Biochemistry* **31,** 8406–8414
33. Scaria, P. V., Shire, S. J., and Shafer, R. D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 10336- 10340
34. Wang, Y., and Patel, D. J. (1992) *Biochemistry* **31,** 8112–8119
35. Jin, R., Breslauer, K. J., Jones, R. A., and Gaffney, B. L. (1990) *Science* **250,** 543–546
36. Mohanty, D., and Bansal, M. (1993) *Nucl. Acids Res* **21,** 1767- 1774
37. Albergo, D. D., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) *Biochemistry* **20,** 1409–1413
38. Marky, L. A. and Breslauer, K. J. (1987) *Biopolymers* **26,** 1601- 1620
39. Johnston, B. H. (1992) *Methods Enzymol.* **211,** 127
40. Lu, M., Guo, Q., and Kallenbach, N. R. (1992) *Biochemistry* **31,** 2455–2459
41. Gray, D. M., and Bollum, F. J. (1974) *Biopolymers* **13,** 2087–2102
42. Marck, C., and Thiele, D. (1978) *Nucleic Acids Res.* **45,** 1017–1028
43. Jin, G., Gaffney, B. L., Wang, C., Jones, R. A., and Breslauer, K. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 8832–8836
44. Guo, Q., Lu, M., and Kallenbach, N. R. (1993) *Biochemistry* **32,** 3596–3603
45. Friedman, T., and Brown, D. M. (1978) *Nucleic Acids Res.* **5,** 615- 622
46. Burton, K., and Reilly, W. T. (1966) *Biochem. J.* **98,** 70–77
47. Hayatsu, H., and Ukita, T. (1967) *Biochem. Biophys. Res. Commun.* **29,** 556–561
48. Nagaich, A. K., Bhattacharya, D., Brahmachari, S. K., and Bansal, M. (1994) *J. Biol. Chem.* **269,** 7824–7833
49. Johnston, B. H., and Rich, A. (1985) *Cell* **42,** 713–724
50. Furlong, J. C., and Lilley, D. M. J., (1986) *Nucleic Acids Res.* **14,** 3995–4007
51. Wang, Y., and Patel, D. J., (1993) *Structure* **1,** 263–282
52. Guo, Q., Lu, M., and Kallenbach, N. R. (1992) *J. Biol. Chem.* **267,** 15293–15300
53. Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M. (1991) *Nature* **350,** 718–720
54. Fang, G., and Cech, T. R. (1993) *Biochemistry* **32,** 11646–11657
55. Tripathi, J., and Brahmachari, S. K. (1991) *J. Biomol. Struc. Dyn.* **9,** 387–397
56. Brahmachari, S. K., Mishra, R. K., Bagga, R., and Ramesh, N. (1989) *Nucleic Acids Res.* **17,** 7273–7281
57. Boehm, T., Graw, L. M., Kees, U. R., Spurr, N., Lavenir, I., Foster, A., and Rabbitts, T. H. (1989) *EMBO. J* **8,** 2621–2631
58. Holliday, R. (1989) *Trends Genet.* **5,** 355–356