The EMBO Journal vol.13 no.4 pp.993-1001, 1994

Tetraplex folding of telomere sequences and the inclusion of adenine bases

Alastair I.H.Murchie and David M.J.Lilley[1]

CRC Nucleic Acid Structure Research Group, Department of Biochemistry, The University, Dundee DD1 4HN, UK

[1]Corresponding author

Communicated by D.M.J.Lilley

Telomeres are required for eukaryotic chromosome stability. They consist of regularly repeating guanine-rich sequences, with a single-stranded 3′ terminus. Such sequences have been demonstrated to have the propensity to adopt four-stranded structures based on a tetrad of guanine bases. The formation of an intramolecular foldback tetraplex is associated with markedly increased mobility in polyacrylamide. Most telomeric sequences are based either on a repeat of d(T_nGGGG) or d(T_nAGGG) sequences. We have used a combination 7-deazaguanine or 7-deaza-adenine substitution, chemical modification and gel electrophoresis to address the following aspects of intramolecular tetraplex formation. (i) Intramolecular tetraplex formation by d(TTTTGGGG)$_4$ sequences is prevented by very low levels of 7-deazaguanine substitution. This confirms the important role of guanine N^7 in the formation of the tetraplex. (ii) The sequences d(TTAGGG)$_4$ and d(TTTTAGGG)$_4$ fold into tetraplexes. By contrast, the electrophoretic behaviour of d(TTTTGGGA)$_4$, d(TTTTAGAG)$_4$ and d(TTTTGAGA)$_4$ does not indicate formation of stable intramolecular tetraplexes under available conditions. (iii) Selective 7-deazaguanine and 7-deaza-adenine substitutions in d(TTTTAGGG)$_4$ give results consistent with tetraplex folding by the formation of three G_4 tetrads, with the adenine bases formally part of the single-stranded loops, where they probably interact with thymine bases. These results demonstrate that eukaryotic cells appear to have selected just those sequences that can adopt the tetraplex conformation for their telomeres, while those that cannot have been avoided. This suggests that the conformation may be significant in the function of the telomere, such as attachment to nuclear structures.

Key words: DNA structure/guanine quartet/telomere/tetrad

Introduction

Telomeres are the ends of eukaryotic chromosomes (reviewed in Blackburn, 1991). In general they comprise many copies of a simple repeating sequence, such as the d(TTTTGGGG)$_n$ sequence of *Oxytricha*, and some examples of known sequences are presented in Table I (Zakian, 1989). The chromosome ends are neither covalently closed as a hairpin, nor covalently attached to protein, but instead the G-rich strand appears to project single-stranded for ~12–20 nt in the 3′ direction (Klobutcher *et al.*, 1981). The telomeric ends are elongated in the cell by the action of telomere terminal transferase (telomerase) (Greider and Blackburn, 1985), and the telomeres are essential in maintenance of chromosome stability. Shortening of telomeres is associated with ageing and tumorigenesis (Harley *et al.*, 1990; Hastie *et al.*, 1990). Despite the single-stranded character of the telomeric projections, they do not behave as conventional single-stranded DNA (Blackburn, 1986), suggesting that they are protected by folding into a structure that resists attack.

It has been known for some time that telomeric sequences can adopt novel conformations based on association between guanine bases (Henderson *et al.*, 1987). Two different types of structure have been observed in the presence of monovalent cations, based on the inter- or intramolecular assembly of four strands, held together by the formation of tetrads of guanine bases that are hydrogen bonded between N^1H and O^6, and N^2H and N^7, as first proposed by Gellert *et al.* (1962) (Figure 1). Tetraplex formation may occur either as a parallel intermolecular association of four-strands (Sen and Gilbert, 1988, 1990) or as an antiparallel association of either two hairpin-forming strands (Sundquist and Klug, 1989) or a single strand that is folded back to comprise three loops (Williamson *et al.*, 1989). NMR studies show that the parallel structure is based on a backbone conformation in which all glycosyl torsion angles are *anti* (Aboul-ela *et al.*, 1992; Jin *et al.*, 1992; Wang and Patel, 1992), while in the antiparallel structure there is an alternation of *syn* and *anti* glycosyl angles (Wang *et al.*, 1991; Smith and Feigon, 1992). A variety of potential isomeric forms of the antiparallel structures exists, depending on how the loop regions connect the different oligoguanine blocks of the tetraplex, and structures have been presented in which the loops connect either neighbouring (Kang *et al.*, 1992) or diametrically separated strands (Smith and Feigon, 1992) (see Figure 1). The parallel tetraplex structure has also been observed in RNA (Kim *et al.*, 1991), where its conformation is closely similar to that of parallel-stranded guanine tetraplex DNA (Cheong and Moore, 1992), and RNA tetraplex formation has been suggested to be important in retroviral genome dimerization (Sundquist and Heaphy, 1993).

Thus there is a wealth of data revealing the propensity

Table I. Telomere repeat sequences found in a variety of organisms

Organism	Sequence repeat
Tetrahymena	... T_2G_4 ...
Oxytricha	... T_4G_4 ...
Paramecium	... T_2G_3(G or T) ...
Homo sapiens	... T_2AG_3 ...
Arabidopsis	... T_3AG_3 ...
Chlamydomonas	... T_4AG_3 ...

Information from Blackburn (1991) and Zakian (1989).

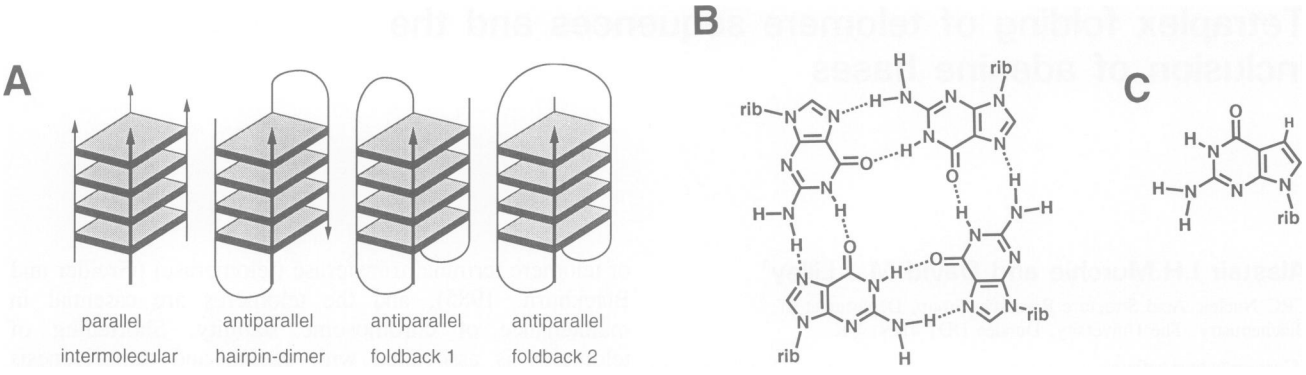

Fig. 1. Tetraplex formation by oligoguanine sequences, based on the formation of hydrogen bonded guanine tetrads. (**A**) Oligoguanine sequences can assemble in a variety of different conformations that are based on the G_4 tetrad (indicated by the shaded blocks). The polarity of the strands is shown by the arrows. When four separate strands assemble in the presence of monovalent cations, they adopt a parallel conformation (Sen and Gilbert, 1988; Aboul-ela *et al.*, 1992; Jin *et al.*, 1992; Wang and Patel, 1992). By contrast, intramolecular tetraplex association requires the formation of antiparallel structures. Sequences containing two oligoguanine blocks can form hairpin structures that then dimerize by tetraplex formation (Sundquist and Klug, 1989); these can exist in a number of isomers (Kang *et al.*, 1992; Smith and Feigon, 1992), only one of which is shown. The loops comprise a number (generally between two and four) of consecutive formally single-stranded thymine bases. Single strands containing four oligoguanine blocks may fold intramolecularly by virtue of guanine tetrad formation (Williamson *et al.*, 1989); two isomers of this structure are possible that differ in their connectivity. In the foldback 1 structure each oligoguanine strand is antiparallel to both nearest neighbours, while in the foldback 2 structure the nearest neighbours are inequivalent, one parallel and one antiparallel. The latter has been observed by NMR spectroscopy (Smith and Feigon, 1992). (**B**) The structure of a guanine tetrad. (**C**) The structure of 7-deazaguanine, used in these and earlier (Murchie and Lilley, 1992) studies to examine the role of guanine N^7 in the formation of tetraplex structures. Note that the N^7 position of guanine has been replaced by CH, and is therefore unavailable as a hydrogen bond acceptor.

of oligoguanine sequences to undergo either inter- or intramolecular tetraplex formation in the presence of monovalent cations. However, examination of Table I shows that about half of the known sequences of eukaryotic telomeres are based on a repeated oligoguanine block that is preceded at the 5′ end by an adenine base. Thus the human telomere has a sequence based on $(T_2AG_3)_n$, while that of *Chlamydomonas* is based on $(T_4AG_3)_n$. This raises the question of how the adenine base might affect any folding of the guanine-rich strand and whether the adenine might participate directly in tetraplex formation, by the formation of an AGAG tetrad for example. If the tetraplex conformation of the telomeric sequences fulfils a functional role in the cell, the widespread occurrence of telomeres based on the AGGG repeat would indicate that these sequences should adopt a structure that is essentially the same as that adopted by GGGG repeats. In this work we have sought answers to the following questions concerning intramolecular tetraplex formation. (i) Can we obtain confirmation of the importance of the N^7 of guanine in the formation of the foldback structure? (ii) Do sequences based on AGGG repeats fold in the presence of potassium to generate a foldback structure? (iii) When adenine bases are substituted into other positions within the basic GGGG repeat, can these species also adopt the foldback structure? It is striking that all natural telomeric sequences appear to be based upon either GGGG or AGGG repeats, suggesting that there may be a fundamental impediment to the inclusion of adenine bases in positions other than the 5′ end of the block. (iv) Do the adenine bases of the AGGG repeats play a direct role in base tetrad formation, or do they exist as part of the loops?

We have employed a combination of gel electrophoresis, chemical probing and base substitution to obtain answers to these questions. We conclude that adenine substitution at the 5′ guanine of GGGG sequences allows the formation of the foldback structure, while adenine substitution elsewhere in the block is more destabilizing. The 5′ adenine bases do not appear to participate directly in tetraplex formation.

Results

Formation of intramolecular foldback structures by GGGG and AGGG sequences

Gel electrophoresis provides a simple method for the study of intramolecular foldback tetraplex formation by sequences such as $(T_4G_4)_4$. In this approach, a series of multimers of the repeating block are made, $(T_4G_4)_n$, where n = 1, 2, 3 or 4, and the electrophoretic mobilities of the four species are compared under conditions where intramolecular tetraplex formation is favoured, i.e. in the presence of potassium ions. Since the foldback structure requires the participation of four G_4 blocks, the species with $n < 4$ migrate normally, while $(T_4G_4)_4$ folds to generate a species that has anomalously fast mobility. In the presence of 25 mM potassium ions, the $(T_4G_4)_4$ species acquires a mobility in 15% polyacrylamide, which is intermediate between those of $(T_4G_4)_2$ and $(T_4G_4)_3$ (Figure 2A), in agreement with the observations of Cech and coworkers (Williamson *et al.*, 1989; Zahler *et al.*, 1991). The compact shape of the foldback structure reduces the frictional drag on the $(T_4G_4)_4$ sequence. In the absence of added potassium ions, the mobility of the $(T_4G_4)_4$ species is significantly slower than that of $(T_4G_4)_3$, as expected for the unfolded conformation (data not shown).

The human telomere is based on a different repeating sequence, namely $(T_2AG_3)_n$. We found that this also exhibited anomalously fast mobility in polyacrylamide (Figure 2B), closely similar to that of the $(T_4G_4)_4$ sequence. This is in agreement with the observations of others on the gel mobility of repeating sequences based on an AGGG block (Balagurumoorthy *et al.*, 1992; Guo *et al.*, 1992; Petracek and Berman, 1992). Clearly the human sequence has the potential to fold in a manner closely similar to the $(T_4G_4)_n$ sequence.

Role of guanine N^7 in the $(T_4G_4)_4$ structure

The fast mobility of the $(T_4G_4)_4$ species is thought to be due a foldback structure based on guanine tetrads (Figure 1). The

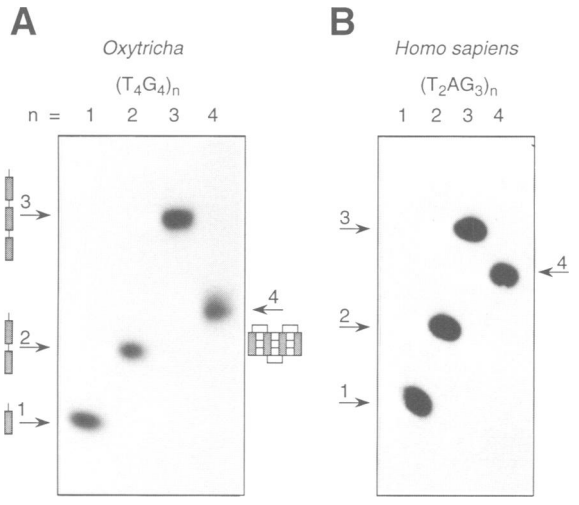

Fig. 2. Intramolecular folding of *Oxytricha* $(T_4G_4)_4$ and human $(T_2AG_3)_4$ telomere sequences in the presence of potassium ions. (**A**) Electrophoretic analysis of intramolecular tetraplex formation by $(T_4G_4)_4$. The series of oligonucleotides $(T_4G_4)_n$ was synthesized, where n = 1, 2, 3 and 4. These were 5′-^{32}P radioactively labelled, and the four species were separately electrophoresed in 15% polyacrylamide gels containing 25 mM potassium ions at 20°C. The autoradiograph of the gel is presented. Note that the species $(T_4G_4)_n$, where n = 1, 2 and 3, migrate with progressively lower mobilities consistent with their increasing length, while $(T_4G_4)_4$ exhibits a marked increase in mobility that is indicative of intramolecular folding into a tetraplex. This is not observed in the absence of added potassium ions. (**B**) Electrophoretic analysis of intramolecular tetraplex formation by $(T_2AG_3)_4$. The electrophoretic mobilities of $(T_2AG_3)_n$, where n = 1, 2, 3 and 4 were compared in 20% polyacrylamide gels containing 25 mM potassium ions at 4°C. $(T_2AG_3)_4$ exhibits a marked increase in mobility relative to the shorter species, indicating potassium-dependent intramolecular folding into a tetraplex conformation.

integrity of the tetrads is maintained by hydrogen bonds donated from the imino H^1 and amino H^2 protons on each guanine base to the O^6 and N^7 respectively of the next base around the tetrad, making a total of eight hydrogen bonds per tetrad. We have previously shown that a powerful way to analyse the importance of the latter hydrogen bond is by means of chemical substitution of 7-deazaguanine (see Figure 1C) for guanine (Murchie and Lilley, 1992), thus removing the acceptor; we showed that this substitution completely prevented parallel, interstrand tetraplex formation. We therefore decided to apply this method to the foldback structure in order to analyse the role of individual guanine N^7 positions on folding of the tetraplex.

We carried out a comparative electrophoretic mobility analysis of $(T_4G_4)_4$ sequences with selective 7-deazaguanine substitution (Figure 3A). It is clear that all the modified sequences exhibit changes in mobility compared with the unmodified DNA in the presence of 10 mM potassium ions. This is not due to the trivial effect of base substitution; all the substituted forms had the same mobility in the absence of potassium ions and the mobilities of the three-block species $(T_4G_4)_3T$ with and without complete deaza substitution of the final block were virtually identical (Figure 3B). In the presence of potassium ions, substitution of all the guanine bases in the 3′ T_4G_4 sequence block (compound 4-gggg, see Table II) reduced the mobility of the four-block species to a level significantly slower than that of $(T_4G_4)_3$, strongly suggesting that this reflects that of a completely unfolded $(T_4G_4)_4$ species. The difference in mobility of the $(T_4G_4)_4$ species with and without modification corresponds to a retardation ratio (RR, defined as the migration difference as a fraction of the migration of the faster species) of 0.26. Folding the substituted species 4-gggg into a tetrad would require that each of the four tetrads contain one 7-deazaguanine, and lose one hydrogen bond out of eight. Interestingly, lower levels of substitution also resulted in destabilization of the folded tetraplex. Thus substitution of two guanine bases in the final block (compound 4-gGGg) lead to an equivalent, slower mobility, as did substitution of a single guanine in the 3′ block (compound 4-GGGg), or the next block in the 5′ direction (compound 3-GgGG). The latter two molecules would only lose one guanine N^7 position out of a total of 16, yet the destabilizing effect is clear. One substituted molecule did exhibit a significant mobility shift; this species (4-GGgG) had a single substitution in the 3′ block, resulting in a smeared mobility intermediate between those of the high mobility $(T_4G_4)_4$ species and the other substituted forms. The smear suggests a dynamic interconversion between forms, on a time-scale similar to the rates of equilibration in the gel pores. At higher potassium concentrations, the singly-substituted species exhibited faster mobilities relative to the $(T_4G_4)_n$ ladder and to the multiply substituted species (data not shown). This indicates that the degree of destabilization of the tetraplex depends on the extent of substitution.

Effects of inclusion of adenine bases at different positions in the purine block

The fast electrophoretic mobility of $(T_2AG_3)_4$ indicates that this species may adopt a compact geometry similar to that of $(T_4G_4)_4$, suggesting that the substitution of the 5′ guanine of the repeating block by adenine does not prevent some type of tetraplex formation. We were therefore curious to see whether adenine substitution in other positions of the purine block might still permit a folded conformation to be adopted. We prepared a series of variants of the basic sequence $(T_4XXXX)_4$, where XXXX was GGGG, AGGG, GGGA, AGAG or GAGA, and analysed their relative electrophoretic mobilities in polyacrylamide in the presence of 25 mM potassium ions (Figure 4). A number of differences in the electrophoretic mobilities of the $(T_4XXXX)_4$ species are apparent; these were not due to different mobilities of the $(T_4XXXX)_n$ comparative species, because the variation in the $(T_4XXXX)_3$ species corresponded to RR $\leq$ 0.04, except for the $(T_4GAGA)_3$ species (see below).

While the substitution of guanine by adenine at the 5′ end of the block (AGGG) permitted normal folding in the presence of potassium ions, the equivalent substitution at the 3′ end, to generate the isomeric species GGGA, had a more destabilizing effect. A smear is seen between the positions corresponding to folded and unfolded conformations, suggesting a medium–slow interconversion between forms. This would also be consistent with the enhanced chemical reactivity of the guanine bases of this sequence (see below). However, we have not found conditions under which the $(T_4GGGA)_4$ species migrated as a sharp band with a mobility equal to that of $(T_4G_4)_4$.

The AGAG substitution appeared to prevent folding totally (RR = 0.22), while the GAGA substitution led to a species of well defined, but intermediate, mobility. This is explored further in Figure 5, where the mobility of the $(T_4GAGA)_4$ species is compared against its own corresponding $(T_4GAGA)_n$ (n = 1, 2, 3) species. The $(T_4GAGA)_4$ species

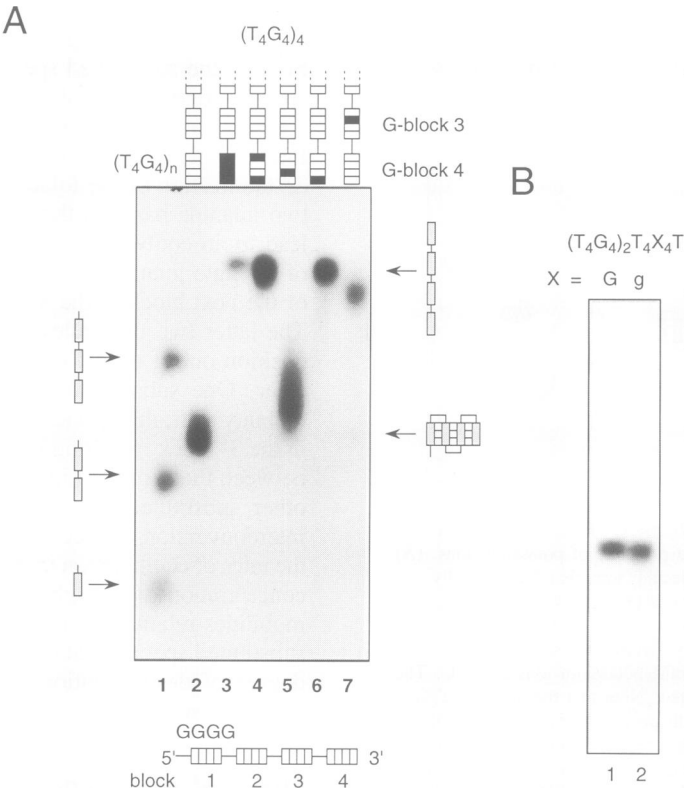

Fig. 3. Gel electrophoretic analysis of intramolecular folding of $(T_4G_4)_4$ sequences with selective 7-deazaguanine substitution; the importance of guanine N^7. (**A**) $(T_4G_4)_4$ was synthesized with the selective replacement of guanine bases by 7-deazaguanine (Table II). Each was 5′-^{32}P radioactively labelled and analysed by electrophoresis in 15% polyacrylamide in the presence of 10 mM potassium ions at 20°C. The mobilities of the modified species can be compared with the unmodified $(T_4G_4)_4$, and the ladder of $(T_4G_4)_n$ where n = 1, 2 and 3. The positions of the base substitutions are indicated above the gel lanes by the black boxes. All but the final substitution are located in the G_4 block at the 3′ end on the molecule. Track 1, mixture of $(T_4G_4)_n$ where n = 1, 2 and 3; track 2, $(T_4G_4)_4$; track 3, 4-gggg; track 4, 4-gGGg; track 5, 4-GGgG; track 6, 4-GGGg; track 7, 3-GgGG. (**B**) Effect of 7-deazaguanine substitution on the electrophoretic mobility of $(T_4G_4)_3$. 5′-^{32}P radioactively labelled DNA was electrophoresed under conditions identical to those used for panel A. The effect on the base substitution on mobility is negligible, even for this extreme case where a complete block of four guanines was substituted. Track 1, $(T_4G_4)_2T_4GGGGT$; track 2, $(T_4G_4)_2T_4ggggT$ (g = 7-deazaguanine).

exhibited fast mobility, but much slower compared with the relative mobilities of $(T_4G_4)_4$ (see track 7) species; the latter molecules migrated with a speed intermediate between those of the two- and three-block species, while the $(T_4GAGA)_4$ species migrated rather more slowly than its corresponding three-block species. This pattern of mobility for the $(T_4GAGA)_4$ species was largely independent of conditions; we have observed equivalent behaviour in the presence of 25 mM sodium, lithium, rubidium ions or caesium ions, or 10 mM magnesium, calcium or strontium ions. This indicates a general ion-screening effect, rather than the selective binding characteristic of tetraplex structures. The modest enhancement of mobility of $(T_4GAGA)_4$ was reversed by substitution of adenine by 7-deaza-adenine or inosine in the 3′ GAGA block, but not by substitution of guanine by 7-deazaguanine.

The electrophoretic mobility of the GAGA species suggests that some type of folding process can occur in this sequence; however, this is unlikely to be an intramolecular tetraplex for several reasons. First, all the $(T_4GAGA)_n$ species exhibited faster mobility than their $(T_4G_4)_n$ equivalents [e.g. $(T_4GAGA)_3$ migrated more quickly than $(T_4G_4)_3$ by RR = 0.09]; thus the folding does not require the presence of four blocks. Second, the relative retardation (compared with unfolded forms) of $(T_4GAGA)_4$ (RR = 0.1) is much less than that of $(T_4G_4)_4$ (RR = 0.26), suggesting a different geometry. Third, the mobility enhancement is observed in the presence of a wide variety of monovalent and divalent metal ions, By contrast, in the presence of magnesium ions $(T_4G_4)_4$ exhibits a mobility characteristic of the unfolded conformation.

Table II. Summary of 7-deazaguanine, 7-deaza-adenine and inosine substitutions employed in these studies

Substituted sequences based on $d(T_4G_4)_4$	
4-gggg	$d(T_4G_4)_3T_4ggggT$
4-gGGg	$d(T_4G_4)_3T_4gGGgT$
4-GGgG	$d(T_4G_4)_3T_4GGgGT$
4-GGGg	$d(T_4G_4)_3T_4GGGgT$
3-GgGG	$d(T_4G_4)_2T_4GgGGT_4G_4T$
Substituted sequences based on $d(T_4AG_3)_4$	
4-aGGG	$d(T_4AG_3)_3T_4aGGGT$
4-AGGg	$d(T_4AG_3)_3T_4AGGgT$
4-AGgG	$d(T_4AG_3)_3T_4AGgGT$
Substituted sequences based on $d(T_4GAGA)_4$	
4-gaga	$d(T_4GAGA)_3T_4gaga$
4-GaGa	$d(T_4GAGA)_3T_4GaGa$
4-gAgA	$d(T_4GAGA)_3T_4gAgA$
4-IAIA	$d(T_4GAGA)_3T_4IAIA$

7-deaza substitutions are indicated by the use of lower case letters and inosine is indicated by I.

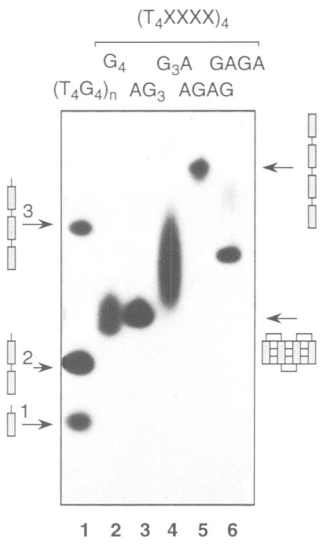

Fig. 4. Gel electrophoretic analysis of intramolecular folding of $(T_4XXXX)_4$ sequences, where the XXXX oligopurine block has been replaced by adenine-containing sequences. $(T_4XXXX)_4$ was synthesized, where XXXX was GGGG, AGGG, GGGA, AGAG or GAGA. Each was 5′-^{32}P radioactively labelled and analysed by electrophoresis in 15% polyacrylamide in the presence of 25 mM potassium ions at 4°C. Track 1, mixture of $(T_4G_4)_n$ where n = 1, 2 and 3; track 2, $(T_4G_4)_4$; track 3, $(T_4AG_3)_4$; track 4, $(T_4G_3A)_4$; track 5, $(T_4AGAG)_4$; track 6, $(T_4GAGA)_4$.

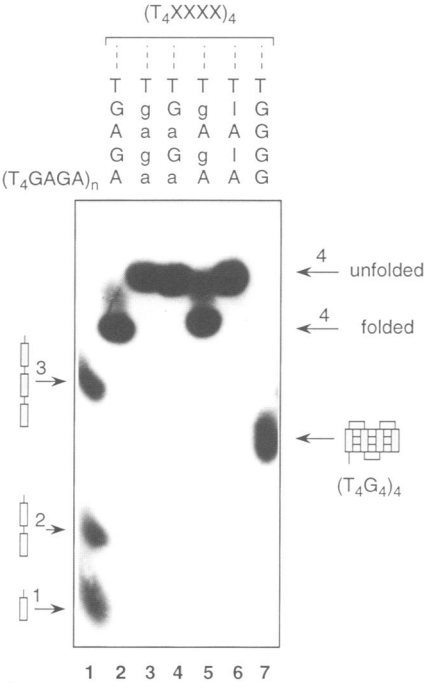

Fig. 5. Gel electrophoretic analysis of intramolecular folding of $(T_4GAGA)_4$ sequences. $(T_4GAGA)_4$ was synthesized, with a variety of 7-deaza-adenine, 7-deazaguanine and inosine substitutions in the 3′ GAGA block. Each was 5′-^{32}P radioactively labelled and analysed by electrophoresis in 15% polyacrylamide in the presence of 25 mM potassium ions at 4°C. Note the great difference in electrophoretic mobility between $(T_4GAGA)_4$ and $(T_4G_4)_4$. The enhanced mobility of $(T_4GAGA)_4$ is abolished by base substitution with 7-deaza-adenine and inosine, but not by 7-deazaguanine. Track 1, mixture of $(T_4GAGA)_n$ where n = 1, 2 and 3; track 2, $(T_4GAGA)_4$; track 3, 4-gaga; track 4, 4-GaGa; track 5, 4-gAgA; track 6, 4-IAIA; track 7, $(T_4G_4)_4$.

Thus the only substitution that appears to permit folding into the compact geometry without evidence for exchange processes is the AGGG sequence, i.e. that found in a number of telomere sequences.

Chemical modification of adenine-substituted sequences

Dimethyl sulfate (DMS) methylates guanine bases at N^7 if this position is accessible, and reactivity is severely diminished in guanine tetraplex structures (Sen and Gilbert, 1988; Sundquist and Klug, 1989; Williamson *et al.*, 1989). We have therefore used DMS to probe guanine N^7 accessibility in the $(T_4XXXX)_4$ sequences under conditions identical to those used for the electrophoretic experiments. 5′-^{32}P-labelled oligonucleotides were reacted with DMS in the presence of 25 mM potassium ions at 15°C. The DNA was then reacted with piperidine at 90°C to cleave N^7-methylguanine adducts, electrophoresed in a polyacrylamide sequencing gel and subjected to autoradiography (Figure 6A). The $(T_4G_4)_4$ sequence exhibited a pattern of very strong protection of the inner two guanine bases of each G_4 tract and a weaker protection of the outer guanines. The three guanine bases of each block of the $(T_4AG_3)_4$ sequence were all equally protected, to about the same extent as the outer bases of the $(T_4G_4)_4$ sequence. By contrast, the three guanine bases of the $(T_4G_3A)_4$ sequence were more strongly reactive, indicative of a lower stability of this structure. Similarly, the two guanine bases of the $(T_4AGAG)_4$ and $(T_4GAGA)_4$ sequences were fully reactive to DMS under the same conditions (data not shown).

In order to examine the accessibility of the adenine bases in the $(T_4AG_3)_4$ and $(T_4G_3A)_4$ sequences, we employed the chemical probe diethyl pyrocarbonate (Leonard *et al.*, 1971; Herr, 1985; Johnston and Rich, 1985; Furlong and Lilley, 1986), which reacts with purine bases in the order of reactivity A > G. 5′-^{32}P-labelled oligonucleotides were modified in 25 mM potassium ions at 15°C under identical conditions, and subjected to piperidine cleavage, sequencing gel electrophoresis and autoradiography (Figure 6B). The adenine bases in each block of both the $(T_4AG_3)_4$ and $(T_4G_3A)_4$ sequences were reactive, suggesting that they were accessible to the reagent.

Participation of adenine in the $(T_4AG_3)_4$ structure

The ability of the $(T_4AG_3)_4$ and $(T_2AG_3)_4$ species to fold into compact conformations that are tetraplex structures of some type poses the question of whether the adenine participates in the tetraplex or not. One could envision two broad types of tetraplex structures based on the side loop foldback 1 structure, depending on whether or not the adenine bases are included (Figure 7). In one form the adenine simply becomes part of the loop connecting tetraplex segments, and the tetraplex comprises three normal G_4 tetrads. In the other possibility, the adenine participates in tetrad formation, requiring the formation of two AGAG tetrads separated by two G_4 tetrads, making a total of four tetrads in this structure.

The structures of A–G mismatches suggests some possible ways in which AGAG tetrads might be constructed. Two main types of A–G mismatch have been observed by X-ray crystallography or NMR of DNA duplexes, based on A*syn*–G*anti* pairing (Brown *et al.*, 1986; Webster *et al.*, 1990) and A*anti*–G*anti* (Patel *et al.*, 1984; Privé *et al.*, 1987) pairing respectively. Individual A–G pairs can form the basis of AGAG tetrads by formation of hydrogen bonds

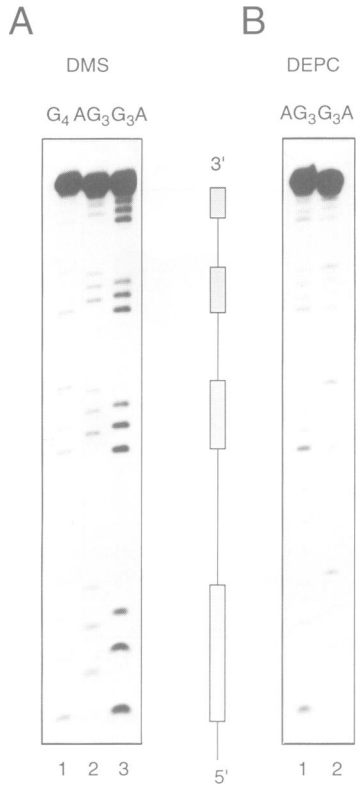

Fig. 6. Chemical modification of adenine-substituted species. (**A**) DMS modification: $(T_4G_4)_4$, $(T_4AGGG)_4$ and $(T_4GGGA)_4$ were $5'$-^{32}P radioactively labelled, and reacted with DMS in the presence of 25 mM potassium ions at 15°C. The DNA was then treated with piperidine at 90°C, electrophoresed in a polyacrylamide sequencing gel and subjected to autoradiography. The oligopurine blocks are indicated (stippled). Track 1, $(T_4G_4)_4$; track 2, $(T_4AGGG)_4$; track 3, $(T_4GGGA)_4$. (**B**) Diethyl pyrocarbonate modification: $(T_4AGGG)_4$ and $(T_4GGGA)_4$ were $5'$-^{32}P radioactively labelled, and reacted with diethyl pyrocarbonate in the presence of 25 mM potassium ions at 15°C. The DNA was then treated with piperidine at 90°C, electrophoresed in a polyacrylamide sequencing gel, and subjected to autoradiography. The oligopurine blocks are indicated (stippled). Track 1, $(T_4AGGG)_4$; track 2, $(T_4GGGA)_4$.

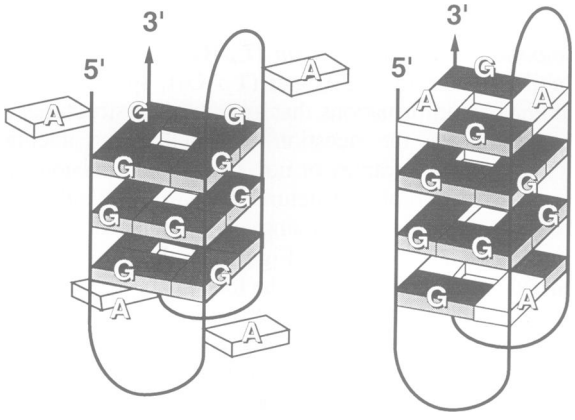

Fig. 7. Schematic to show the potential participation of adenine in the foldback intramolecular tetraplex conformation of $(T_4AG_3)_4$. Two types of structure can be drawn, based on the side-looped foldback 1 structure. In one structure (left) the adenines form part of the formally single-stranded loops, leaving the guanine bases to form three conventional G_4 tetrads. In the other structure (right), the adenines participate directly in tetraplex formation by forming two AGAG tetrads, one on either side of the two inner G_4 tetrads. The tetrads are indicated by the shaded blocks, with the adenine bases indicated by lighter shading.

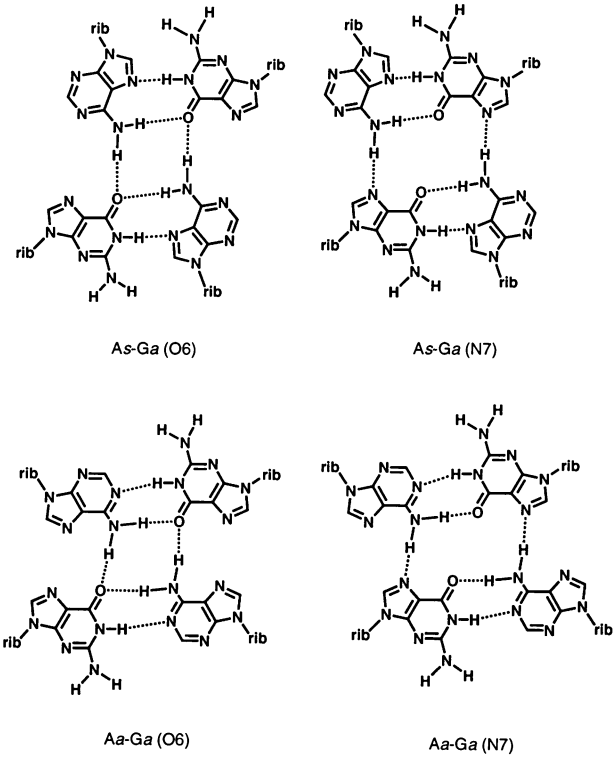

Fig. 8. Possible tetrad structures of two adenine and two guanine bases. Two types of tetrad have been considered, based upon A*syn*−G*anti* or A*anti*−G*anti* interactions. Two alternative tetrads are shown for each, depending on whether the N^6H proton of adenine is accepted by the O^6 or N^7 of guanine.

between adenine N^6H and guanine O^6. Alternative forms of these tetrads may be constructed by using the N^7 of G as the acceptor, which has the advantage of greater colinearity, at the expense of a poorer acceptor. These four tetrads are illustrated in Figure 8. The two tetrads based on the A*syn*−G*anti* pairing (termed A*s*−G*a* tetrads) might be accommodated within the antiparallel structure, in which there is alternation of *syn* and *anti* glycosyl bond torsion angles (Kang *et al.*, 1992; Smith and Feigon, 1992). By contrast, it is more difficult to envisage how the tetrads based on A*anti*−G*anti* pairing (termed A*a*−G*a* tetrads) might be accommodated.

We have employed 7-deaza-adenine and 7-deazaguanine substitutions to examine the possible role of the N^7 position in the $(T_4AG_3)_4$ structure. We can predict the response of tetraplex folding to 7-deaza substitution in different positions, according to the various models for the structure.

(i) Three G_4 tetrads. In this model all the guanine N^7 positions are used for accepting hydrogen bonds, while the adenine N^7 is not required as these bases are formally single-stranded (even if they base-pair with loop thymine bases, as seems likely, the N^7 position is not required). Thus any guanine modification should be destabilizing, while no adenine substitution should affect the stability of the tetraplex.

(ii) Two G_4 tetrads and two AGAG tetrads. This structure has two conventional G_4 tetrads formed from guanine bases in the second and third positions, and two AGAG tetrads which could be of the type discussed above. Depending on which model is correct, different adenine and guanine substitutions should have different effects. In the

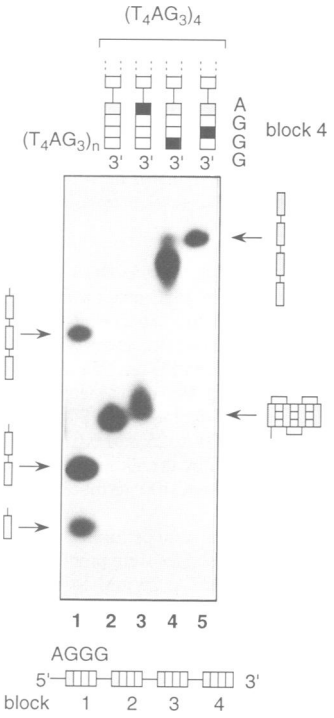

Fig. 9. Probing the participation of adenine in the foldback intramolecular tetraplex conformation of $(T_4AG_3)_4$. $(T_4AG_3)_4$ was synthesized with the selective replacement of guanine or adenine bases by 7-deazaguanine or 7-deaza-adenine respectively (Table II). Each was 5′-^{32}P radioactively labelled and analysed by electrophoresis in 15% polyacrylamide in the presence of 25 mM potassium ions at 4°C. The mobilities of the modified species can be compared with the unmodified $(T_4AG_3)_4$ and the ladder of $(T_4AG_3)_n$ where n = 1, 2 and 3. The positions of the base substitutions are indicated above the gel lanes by the black boxes; all base modifications were introduced into the 3′ AGGG block. Track 1, mixture of $(T_4AG_3)_n$ where n = 1, 2 and 3; track 2, $(T_4AG_3)_4$; track 3, 4-aGGG; track 4, 4-AGGg; track 5, 4-AGgG.

A*s*−G*a* tetrads, N^7 of adenine accepts a hydrogen bond from N^1H of guanine; the importance of N^7 of guanine depends on which acceptor of the adenine N^6H proton is used. In the A*a*−G*a* tetrad, the N^7 of adenine is not used as an acceptor, and the N^7 of guanine is only used in the alternative structure, when it accepts the adenine N^6H proton.

The experimental results are shown in Figure 9. 7-Deaza-adenine substitution (4-aGGG) had no effect on electrophoretic mobility. This suggests that this substitution does not affect folding, although we cannot formally exclude the formation of an alternative tetraplex conformation with identical mobility, however unlikely. By contrast, 7-deazaguanine substitution at either position 4 or 3 in the 3′ block (4-AGgG or 4-AGGg) led to prevention of folding. These results are completely consistent with the three G_4 tetrad model, and exclude three of the four AGAG tetrad models. The substitution of 7-deaza-adenine with retention of folding makes the formation of tetrads employing either of the A*s*−G*a* tetrads improbable. The structure based on A*a*−G*a* (O^6) tetrads is also unlikely, because the folding is prevented by substitution of the 3′ 7-deazaguanine base. The only structure that remains is the one based on the A*a*−G*a* (N^7) tetrad. However, as we discuss below, we feel that structures based on A*a*−G*a* tetrads are improbable for an antiparallel tetraplex.

Discussion

Base substitution is a powerful way of distinguishing models of nucleic acid folding. We have confirmed the importance of guanine N^7 in the formation of the folded structure of $(T_4G_4)_4$. This demonstrates the important role of the N^2H-N^7 hydrogen bond in the stability of the structure, consistent with the known structure of guanine tetrads (Wang *et al.*, 1991; Aboul-ela *et al.*, 1992; Jin *et al.*, 1992; Kang *et al.*, 1992; Smith and Feigon, 1992; Wang and Patel, 1992). The structure appears to be extremely sensitive to this modification; inclusion of a single 7-deazaguanine base (e.g. 4-GGgG) can result in a reversion to normal electrophoretic mobility. Since this potentially removes just one hydrogen bond out of a total of 32, it seems that the substitution may interfere with a cooperative folding of the entire structure. This is supported by the results with the AGGG sequences, showing that these molecules appear to fold by the formation of three guanine tetrads (see below); thus three perfect tetrads (total of 24 hydrogen bonds) are more stable than four tetrads with a single substitution (total 31 hydrogen bonds), although there may be additional stabilization conferred by the adenine bases in the loops. This all points to a destabilization caused by the inclusion of 7-deazaguanine that corresponds to more than the loss of a single hydrogen bond. This could be a consequence of steric repulsion caused by the C^7H proton of the 7-deazaguanine base.

The possible role of adenine bases in sequences based on TTTTAGGG repeats have been investigated by purine N^7 substitutions. The results are consistent with the formation of a structure based on three G_4 tetrads. Formation of a tetraplex that includes AGAG tetrads based on A*syn*−G*anti* interaction is unlikely, because folding is prevented by 7-deazaguanine substitution, but not by 7-deaza-adenine substitution. Inclusion of AGAG tetrads based on A*anti*−G*anti* interaction (in which guanine N^7 acts as a hydrogen bond acceptor) remains consistent with our data, but seems improbable on stereochemical grounds, because of the requirement for alternation of *syn* and *anti* glycosyl bonds in an antiparallel tetraplex. Moreover, the reactivity of the adenine bases to diethyl pyrocarbonate suggests that they are accessible in the structure. This agrees with the observations that adenine substitution at the 5′ guanine of G_4 blocks reduces the stability of antiparallel tetraplex structures (Hardin *et al.*, 1991; Petracek and Berman, 1992), while the stability of the foldback tetraplex is unaffected by the change of $(T_4G_3)_4$ to $(T_3AG_3)_4$ (Guo *et al.*, 1992). Our conclusions are fully consistent with a recent crystallographic study of the sequence AGGGTTAGGG, which associates as a hairpin dimer based on a tetraplex comprising three G_4 tetrads (A.Rich, personal communication).

Two isomeric forms of foldback tetraplexes are possible, that are distinguished by having loops that connect chains that are located at adjacent corners, or traverse the diagonal (foldback structures 1 and 2, see Figure 1). If the adenine bases were to participate in tetraplex formation with guanine bases, these would be required to be AGAG or AAGG tetrads for side and diagonal looped structures respectively. It is difficult to propose a stereochemically reasonable tetrad structure for the latter. An energetic preference for the three-G_4 structure could therefore reflect the formation of a diagonal-looped structure (foldback structure 2 of Figure 1), as has been observed by NMR studies of a intramolecular

tetraplex based on the *Oxytricha* repeat TTTTGGGG (Smith and Feigon, 1992).

We have shown that two types of DNA sequence exhibit the anomalously fast gel electrophoretic mobility that reflects the formation of the intramolecular tetraplex. Sequences based on GGGA repeats may be in a folded ↔ unfolded equilibrium, but we have not found conditions where this is completely folded into an intramolecular tetraplex; this sequence clearly is of significantly lower stability. The AGAG- and GAGA-based sequences cannot adopt the intramolecular tetraplex geometry. Although GAGA-containing sequences appear to undergo some type of folding process, the resulting electrophoretic mobility of these species is not consistent with an intramolecular tetraplex. Rippe *et al.* (1992) have shown that sequences based on alternating $(A-G)_n$ sequences can adopt a parallel duplex structure. Thus the only sequences that unequivocally adopt the intramolecular tetraplex geometry are based on the repeat units of GGGG and AGGG, which are the sequences that are found in natural telomere sequences (Table I). The natural loop sizes of T_2 and T_4 are both accommodated within the folded structure. All other variants on these sequences that we have studied, including a simple isomerization of the AGGG sequence to GGGA, result in a total or partial loss of the geometry that confers high electrophoretic mobility. While this could simply be a coincidence, it is curious that natural selection seems to have chosen just those sequences that can best adopt the tetraplex geometry for the telomeres of the majority of chromosomes for which information is available. This might indicate a role for tetraplex formation in some aspect of the function of the telomere. What this role might be we cannot say at present. Tetraplex DNA does not serve as a substrate for telomerase (Zahler *et al.*, 1991), leaving potential roles in the protection of the end of the chromosome, and in the association of chromosomes. There have been a number of reports of proteins that appear to recognize the structure of the guanine tetraplex (Walsh and Gualberto, 1992; Fang and Cech, 1993b; Liu *et al.*, 1993; Weismanshomer and Fry, 1993), and it has recently been demonstrated that the *Oxytricha* telomere binding protein promotes the formation of guanine tetraplex structure (Fang and Cech, 1993a). Chromosome telomeres are probably held to the nuclear matrix (de Lange, 1992) by the action of binding proteins, possibly such as the RAP1 protein of yeast (Conrad *et al.*, 1990; Lustig *et al.*, 1990). It is therefore entirely possible that this interaction is mediated by the formation of a four-stranded geometry and that the ability to form the structure therefore dictates the choice of sequences that are tolerated at the telomere.

Materials and methods

Oligonucleotide synthesis

Oligonucleotides were synthesized using β-cyanoethyl phosphoramidite chemistry (Cruachem) (Beaucage and Caruthers, 1981; Sinha *et al.*, 1984), implemented on an Applied Biosystems 394 synthesizer. All 7-deazabase modified oligonucleotides were synthesized with an additional 3′ thymine residue; this had a negligible effect on electrophoretic mobility (RR < 0.01). Deprotected oligonucleotides were purified by electrophoresis in 20% polyacrylamide containing 7 M urea and electroelution. Oligonucleotides were 5′-^{32}P radioactively labelled using T4 polynucleotide kinase and [γ-^{32}P] ATP (Maxam and Gilbert, 1980) and preincubated in 20 ml 90 mM Tris.borate, pH 8.3, 1 mM EDTA, 25 mM KCl at 90°C for 2 min and cooled on ice.

Gel electrophoresis

Intramolecular tetraplex formation was analysed by electrophoresis in native polyacrylamide gels (Williamson *et al.*, 1989; Zahler *et al.*, 1991). Polyacrylamide gels 22–23 cm long (29:1 polyacrylamide/*bis*-acrylamide) of the composition indicated in the text (typically 15% polyacrylamide) were electrophoresed in 90 mM Tris–borate, pH 8.3 and 1 mM EDTA in the presence of the indicated concentration of KCl (10–25 mM) at 4 or 20°C. Electrophoresis buffers were recirculated at >1 l/h. Gels were run at 7 V/cm until the bromophenol blue marker dyes had migrated 18 cm and then subjected to autoradiography.

Chemical probing of folded telomeric DNA

DMS reactions were performed at 15°C under solution conditions identical to those for the gel electrophoretic experiments. Radioactively labelled telomeric oligonucleotides and related sequences were preincubated in 20 μl 90 mM Tris–borate, pH 8.3, 1 mM EDTA, 25 mM KCl at 90°C for 2 min and cooled on ice. The resulting DNA was incubated at 15°C for 5 min and an equal (20 μl) volume of 1% (v/v) DMS (Merck) in 90 mM Tris–borate, pH 8.3, 1 mM EDTA, 25 mM KCl was added. After 4 min the DMS reactions were stopped by the addition of 1.5 M sodium acetate, pH 7.0, 1 M β-mercaptoethanol and 100 μg/ml tR NA, and the DNA was precipitated with ethanol.

Diethyl pyrocarbonate reactions were performed under similar conditions. Radioactively labelled oligonucleotides were preincubated in 20 ml 90 mM Tris–borate, pH 8.3, 1 mM EDTA, 25 mM KCl at 90°C for 2 min and cooled on ice. The DNA was then incubated at 15°C for 5 min and an equal volume of 3% (v/v) diethyl pyrocarbonate (Sigma) in 90 mM Tris–borate, pH 8.3, 1 mM EDTA, 25 mM KCl was added. The reaction was allowed to proceed for 20 min, with vortex mixing at 2 min intervals throughout. Reactions were terminated by the addition of 1.5 M sodium acetate, pH 7.0, 1 M β-mercaptoethanol and 100 μg/ml tRNA, and the DNA was precipitated with ethanol.

Chemically reacted DNA was cleaved at the modified positions by reaction with 1 M piperidine at 90°C for 30 min, followed by extensive lyophilization. Samples were dissolved in deionized formamide and analysed by electrophoresis in 15% polyacrylamide–7 M urea sequencing gels.

Acknowledgements

We thank Professor Alex Rich for communication of information prior to publication, Philip Rawlins for preliminary experiments, Audrey Gough for technical assistance and the CRC for financial support.

References

Aboul-ela,F., Murchie,A.I.H. and Lilley,D.M.J. (1992) *Nature*, **360**, 280–282.

Balagurumoorthy,P., Bramachari,S.K., Mohanty,D., Bansal,M. and Sasizekharan,V. (1992) *Nucleic Acids Res.*, **20**, 4061–4067.

Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Letters*, **22**, 1859–1862.

Blackburn,E.H. (1986) in Bogorad,L. (ed.), *Molecular Structure of Telomeres in Lower Eukaryotes*. Alan R.Liss, Inc., pp. 69–82.

Blackburn,E.H. (1991) *Nature*, **350**, 569–573.

Brown,T., Hunter,W.N., Kneale,G. and Kennard,O. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 2402–2406.

Cheong,C. and Moore,P.B. (1992) *Biochemistry*, **31**, 8406–8414.

Conrad,M.N., Wright,J.H., Wolf,A.J. and Zakian,V.A. (1990) *Cell*, **63**, 739–750.

de Lange,T. (1992) *EMBO J.*, **11**, 717–724.

Fang,G.W. and Cech,T.R. (1993a) *Cell*, **74**, 875–885.

Fang,G.W. and Cech,T.R. (1993b) *Proc. Natl Acad. Sci. USA*, **90**, 6056–6060.

Furlong,J.C. and Lilley,D.M.J. (1986) *Nucleic Acids Res.*, **14**, 3995–4007.

Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.

Greider,C.W. and Blackburn,E.H. (1985) *Cell*, **43**, 405–413.

Guo,Q., Lu,M. and Kallenbach,N.R. (1992) *J. Biol. Chem.*, **267**, 15293–15300.

Hardin,C.C., Henderson,E., Watson,T. and Prosser,J.K. (1991) *Biochemistry*, **30**, 4460–4472.

Harley,C.B., Futcher,A.B. and Greider,C.W. (1990) *Nature*, **345**, 458–460.

Hastie,H.D., Dempster,M., Dunlop,M.G., Thompson,A.M., Green,D.K. and Allshire,R.C. (1990) *Nature*, **346**, 866–868.

Henderson,E., Hardin,C.C., Wolk,S.K., Tinoco,I.,Jr and Blackburn,E. (1987) *Cell*, **51**, 899–908.

Herr,W. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 8009–8013.
Jin,R.Z., Gaffney,B.L., Wang,C., Jones,R.A. and Breslauer,K.J. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8832–8836.
Johnston,B.H. and Rich,A. (1985) *Cell*, **42**, 713–724.
Kang,C., Zhang,X., Ratliff,R., Moyzis,R. and Rich,A. (1992) *Nature*, **356**, 126–131.
Kim,J., Cheong,C. and Moore,P.B. (1991) *Nature*, **351**, 331–332.
Klobutcher,L.A., Swanton,M.T., Donini,P. and Prescott,D.M. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 3015–3019.
Leonard,N.J., McDonald,J.J., Henderson,R.E.L. and Reichmann,M.E. (1971) *Biochemistry*, **10**, 3335–3342.
Liu,Z., Frantz,J.D., Gilbert,W. and Tye,B.-K. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3157–3161.
Lustig,A.J., Kurtz,S. and Shore,D. (1990) *Science*, **250**, 549–552.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Murchie,A.I.H. and Lilley,D.M.J. (1992) *Nucleic Acids Res.*, **20**, 49–53.
Patel,D.J., Kozlowski,S.A., Ikuta,S. and Itakura,K. (1984) *Biochemistry*, **23**, 3207–3217.
Petracek,M.E. and Berman,J. (1992) *Nucleic Acids Res.*, **20**, 89–95.
Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.-S., Kopka,M.L. and Dickerson,R.E. (1987) *Science*, **238**, 498–504.
Rippe,K., Fritsch,V., Westhof,E. and Jovin,T.M. (1992) *EMBO J.*, **11**, 3777–3786.
Sen,D. and Gilbert,W. (1988) *Nature*, **334**, 364–366.
Sen,D. and Gilbert,W. (1990) *Nature*, **344**, 410–414.
Sinha,N.D., Biernat,J., McManus,J. and Koster,H. (1984) *Nucleic Acids Res.*, **12**, 4539–4557.
Smith,F.W. and Feigon,J. (1992) *Nature*, **356**, 164–168.
Sundquist,W.I. and Heaphy,S. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3393–3397.
Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.
Walsh,K. and Gualberto,A. (1992) *J. Biol. Chem.*, **267**, 13714–13718.
Wang,Y. and Patel,D.J. (1992) *Biochemistry*, **31**, 8112–8119.
Wang,Y., de los Santos,C., Gao,X., Greene,K., Live,D. and Patel,D.J. (1991) *J. Mol. Biol.*, **222**, 819–832.
Webster,G.D., Sanderson,M.R., Skelly,J.V., Neidle,S., Swann,P.F., Li,B.F. and Tickle,I.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6693–6697.
Weismanshomer,P. and Fry,M. (1993) *J. Biol. Chem.*, **268**, 3306–3312.
Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell*, **59**, 871–880.
Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) *Nature*, **350**, 718–720.
Zakian,V.A. (1989) *Annu. Rev. Genet.*, **23**, 579–604.

Received on September 9, 1993; revised on November 22, 1993