The EMBO Journal vol.12 no.10 pp.4029-4038, 1993

Alternating d(GA)$_n$ DNA sequences form antiparallel stranded homoduplexes stabilized by the formation of G·A base pairs

D.Huertas[1], L.Bellsolell[1,2], J.M.Casasnovas[1], M.Coll[1,2] and F.Azorín[1]

[1]Departamento de Biologia Molecular y Celular, Centro de Investigación y Desarrollo-CSIC, Jorge Girona Salgado 18-26, 08034 Barcelona, and [2]Departamento de Ingeniería Química, Universidad Politécnica de Cataluña, Diagonal 647, 08028 Barcelona, Spain

Communicated by A.Rich

Alternating d(GA)$_n$ DNA sequences form antiparallel stranded homoduplexes which are stabilized by the formation of G·A pairs. Three base pairings are known to occur between adenine and guanine: AH$^+$ (*anti*)·G(*syn*), A(*anti*)·G(*anti*) and A(*syn*)·G(*anti*). Protonation of the adenine residues is not involved in the stabilization of this structure, since it is observed at any pH value from 8.3 to 4.5; at pH ≤ 4.0 antiparallel stranded d(GA·GA) DNA is destabilized. The results reported in this paper strongly suggest that antiparallel stranded d(GA·GA) homoduplexes are stabilized by the formation of alternating A(*anti*)·G(*anti*) and G(*anti*)·A(*syn*) pairs. In this structure, all guanine residues are in the *anti* conformation with their N7 position freely accessible to DMS methylation. On the other hand, adenines in one strand adopt the *anti* conformation, with their N7 position also free for reaction, while those of the opposite strand are in the *syn* conformation, with their N7 position hydrogen bonded to the guanine N1 group of the opposite strand. A regular right-handed helix can be generated using alternating G(*anti*)·A(*syn*) and A(*anti*)·G(*anti*) pairs.

Key words: DNA hairpins/DNA-modification/homopurine sequences/multistranded DNA/non-B DNA structures

Introduction

Homopurine·homopyrimidine DNA sequences of the type d(GA·CT)$_n$, are fairly abundant in eukaryotic genomic DNA, accounting for ~0.3–0.5% of the total mammalian genome (Manor *et al.*, 1988). These sequences are likely to play important biological roles. They are often found in or near transcriptional regulatory regions (reviewed in Wells *et al.*, 1988; Palecek, 1991). Alternating d(GA·CT)$_n$ sequences are also frequent at recombination 'hot-spots' (Sekiya *et al.*, 1981; Hentschel, 1982; Glikin *et al.*, 1983; Mason *et al.*, 1983; Richards *et al.*, 1983; Hunt *et al.*, 1984; Collier *et al.*, 1988; Weinreb *et al.*, 1990) and they have been implicated in the establishment and/or maintenance of large multigene families (Hunt *et al.*, 1984). Furthermore, SV40 viruses carrying a d(GA·CT)$_{22}$ sequence show an increased genomic instability probably as a consequence of an increased rate of 'recombination' (Bernués *et al.*, 1991).

Alternating d(GA·CT)$_n$ sequences show a remarkable degree of structural polymorphism. Several non-B DNA conformations have been described for these sequences (reviewed in Wells *et al.*, 1988; Palecek, 1991), including the formation of pyr-pur-pyr (H-DNA) and pyr-pur-pur (*H-DNA) triplexes (Lyamichev *et al.*, 1986; Bernués *et al.*, 1989). Most of this conformational flexibility resides in the structural properties of the individual strands. Oligo[d(CT)] sequences are known to form antiparallel stranded duplexes at acidic pH, stabilized by the formation of C$^+$·C pairs (Casasnovas *et al.*, in preparation). Similarly, oligo[d(GA)] sequences were proposed to form parallel stranded duplexes (Rippe *et al.*, 1992) as well as multistranded complexes (Lee *et al.*, 1980; Lee, 1990). In this paper, we show that d(GA)$_n$ DNA sequences can also form antiparallel stranded duplexes. This complex is stabilized by the formation of alternate A(*anti*)·G(*anti*) and A(*syn*)·G(*anti*) pairs. Based on the conformational parameters observed in oligonucleotide crystal structures containing G·A mismatches, a helical model for this DNA was built and energy minimized.

Results

Antiparallel stranded d(GA·GA) duplexes are stabilized by the formation of G·A pairs

Alternating d(GA)$_n$ sequences form intramolecular duplexes, reflecting the ability of these sequences to adopt antiparallel stranded conformations. This is particularly clear when the conformational behaviour of alternating d(GA)$_n$ sequences is studied when flanked by autocomplementary bases (Figure 1).The autocomplementary bases flanking the d(GA)$_n$ sequence have a high tendency to pair giving rise to a hairpin. Whether or not the alternating d(GA)$_n$ is forming part of the stem of the hairpin, which will indicate formation of an antiparallel stranded intramolecular d(GA·GA) duplex, can be addressed experimentally through the determination of the patterns of modification obtained with diethylpyrocarbonate (DEPC). DEPC detects unpaired purines. This reagent modifies adenines much more than guanines, at their N7 position. Right-handed B-DNA is not reactive to DEPC. Figure 1A shows the patterns of DEPC modification of a d(GA)$_{15}$ sequence contained within 11 autocomplementary bases, oligo[(GA)$_{15}$]. Reactivity to DEPC is constrained to the centre of the sequence. When the reaction is performed at 4°C, residues A$_{14}$, G$_{15}$, A$_{16}$, G$_{17}$ and A$_{18}$, are hyperreactive to DEPC, at any pH value from 4.5 to 8.3 (Figure 1A, lanes 1). Guanine at position 13 is also slightly reactive to DEPC, particularly at pH ≤ 7.0. The maximum of reactivity corresponds to A$_{16}$. These results show that, as indicated in Figure 1A, the loop of the hairpin extends for six bases. The rest of the alternating GA sequence is contained within the stem of the hairpin, forming an antiparallel stranded duplex. Similar patterns of DEPC modification are obtained at 20°C (Figure 1A, lanes 2), indicating that the hairpin is stable at this temperature. On the other hand, all purines contained within the alternating GA sequence, but not those occurring at the flanking sequences, are hyperreactive to DEPC when the

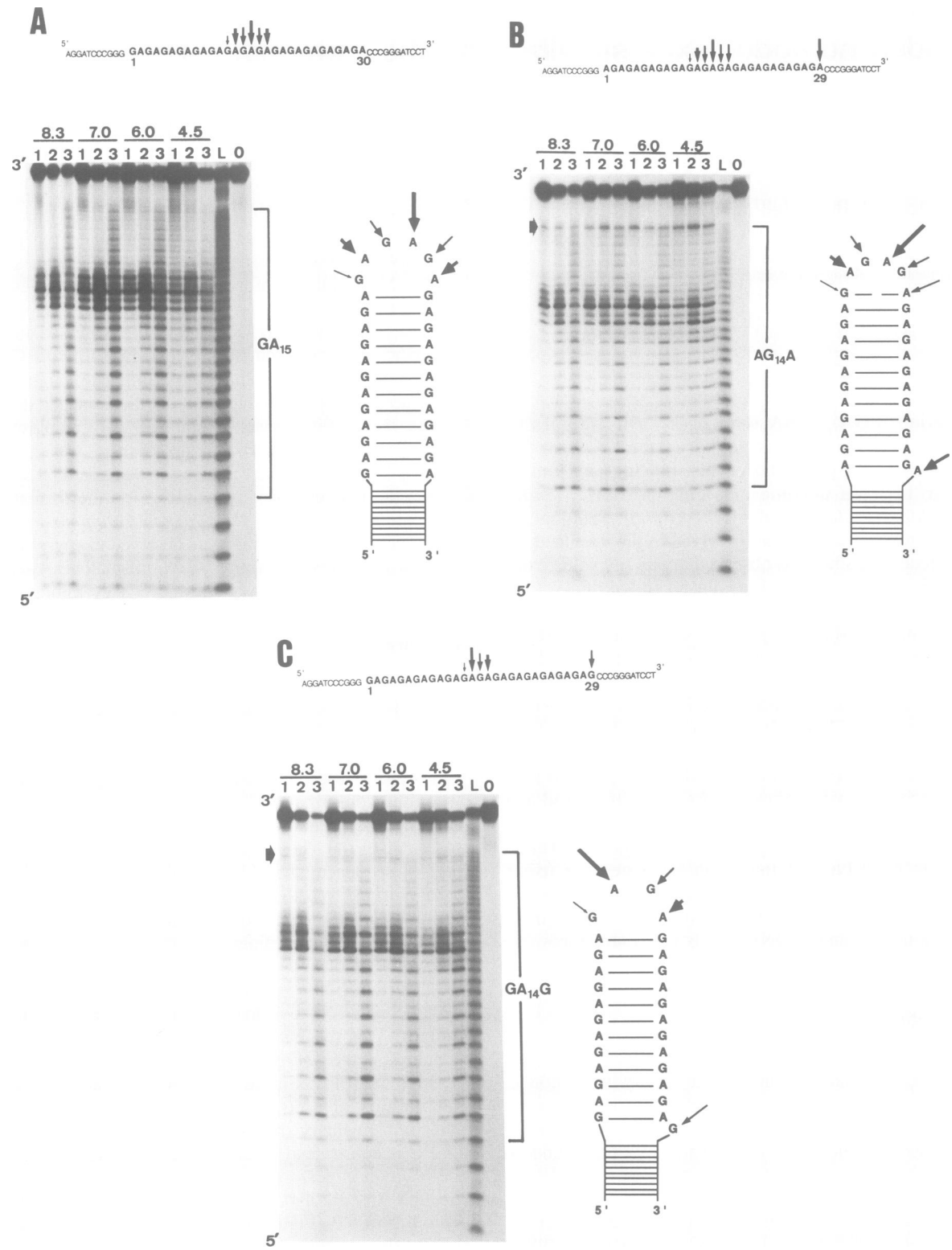

Fig. 1. Antiparallel stranded d(GA·GA) DNA is stabilized by the formation of G·A pairs. The patterns of DEPC modification corresponding to oligo[$(GA)_{15}$] (**A**), oligo[$(AG)_{14}A$] (**B**) and oligo[$(GA)_{14}G$] (**C**) were obtained, at the pH values indicated, and at increasing temperature: 4°C (lane 1); 20°C (lane 2) and 40°C (lane 3). Lanes L correspond to G+A sequencing ladders of the corresponding oligonucleotide. Lanes 0 correspond to the piperidine products obtained from the untreated oligonucleotide. The 5′ to 3′ direction is indicated. Hyperreactive bases are indicated by the arrows. Conformations compatible with the results are schematically indicated on the right of each panel.

modification is carried out at 40°C (Figure 1A, lanes 3), indicating that the antiparallel stranded d(GA·GA) duplex melts at this temperature.

The results shown in Figure 1A suggest that the hairpin form of oligo[$(GA)_{15}$] is stabilized by the formation of G·A base pairs. To investigate further the type of pairing that holds together antiparallel stranded d(GA·GA) DNA, the conformational behaviour of oligonucleotides in which either

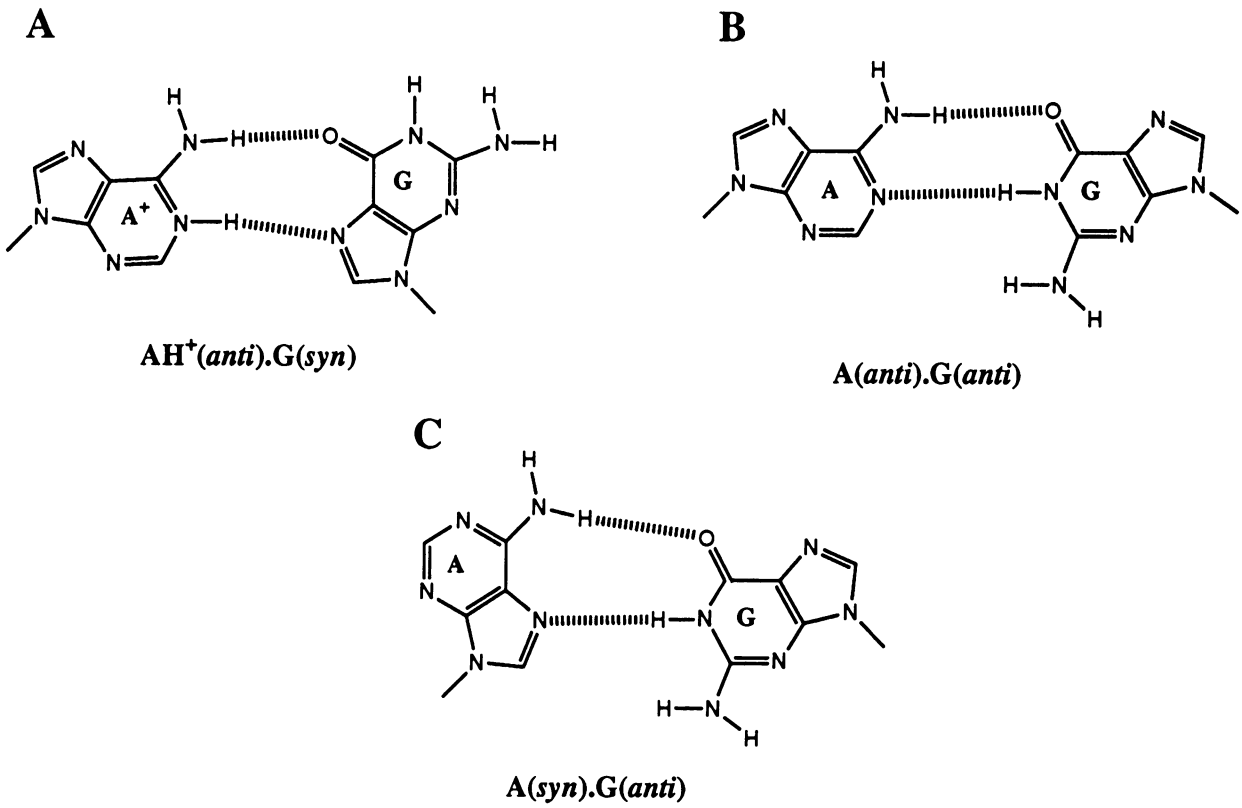

Fig. 2. Schematic representations of the AH$^+$(*anti*)·G(*syn*) (**A**); A(*anti*)·G(*anti* (**B**) and A(*syn*)·G(*anti*) (**C**) pairing schemes.

the first guanine, oligo[(AG)$_{14}$A], or the last adenine, oligo[(GA)$_{14}$G], of the alternating GA sequence were omitted, was studied. The patterns of DEPC modification obtained for oligo[(AG)$_{14}$A] (Figure 1B) are similar to those shown in Figure 1A, corresponding to oligo[(GA)$_{15}$]. Residues at the centre of the sequence are hyperreactive to DEPC but, in addition, A29 (indicated by the arrow in Figure 1B), which is the last adenine of the alternating GA sequence, becomes also hyperreactive to DEPC in this case. These results show that pairing is achieved through the formation of G-A base pairs since, upon removal of the first guanine of the GA tract, the last adenine of the GA sequence becomes unpaired, therefore hyperreactive to DEPC. Similar results were obtained for oligo[(GA)$_{14}$G] (Figure 1C). Also in this case, the last residue of the alternating GA sequence (G$_{29}$) becomes hyperreactive to DEPC. Hyperreactivity in this case is less evident, since the last residue of the alternating GA sequence in oligo[(GA)$_{14}$G] corresponds to a guanine and DEPC modifies guanines much less strongly than adenines.

The structure of the loop corresponding to oligo [(GA)$_{14}$G] appears to be different when compared with oligo[(GA)$_{15}$]. For oligo[(GA)$_{14}$G], the maximum of DEPC reactivity corresponds to A$_{14}$ instead of A$_{16}$, and A$_{18}$ is not hyperreactive. Three bases at the centre (A$_{14}$, G$_{15}$ and A$_{16}$) are hyperreactive to DEPC (Figure 1C) and G$_{13}$ is, also in this case, slightly hyperreactive. These results indicate that the loop in oligo[(GA)$_{14}$G] is formed by four bases instead of six, as observed for oligo[(GA)$_{15}$]. An intermediate situation is observed for oligo[(AG)$_{14}$A]. In this case, hyperreactivity to DEPC extends from residues A$_{13}$ to A$_{17}$, which occupy equivalent positions to those found hyperreactive to DEPC in oligo[(GA)$_{15}$] (Figure 1B). However, the last adenine of the loop, in this case A$_{17}$, is clearly less hyperreactive to DEPC, particularly at low pH values, indicating that it must be paired to some extent to G$_{12}$, which is little reactive in this case. These results suggest that, in the case of oligo[(AG)$_{14}$A], a significant part of the molecules forms a hairpin with a four-member loop. The reason for the different structure of the loops is not clear. Interestingly, the loops start in all cases at an ApG step. It could be possible that ApG steps would be less stable than GpA steps in this antiparallel stranded d(GA·GA) DNA.

Formation of both A(anti)·G(anti) and A(syn)·G(anti) pairs appears to account for the stability of antiparallel stranded d(GA·GA) DNA

The results reported above indicate that antiparallel stranded d(GA·GA) duplexes are stabilized through the formation of G·A base pairs. What pairing scheme holds together the two strands in this duplex? Three different pairing schemes have been observed to occur between adenine and guanine: AH$^+$(*anti*)·G(*syn*) (Brown *et al.*, 1989; Gao and Patel, 1989; Leonard *et al.*, 1990), A(*syn*)·G(*anti*) (Brown *et al.*, 1986; Hunter *et al.*, 1986; Webster *et al.*, 1990) and A(*anti*)·G(*anti*) (Kan *et al.*, 1983; Privé *et al.*, 1987; Nikonowicz and Gorenstein, 1990) (Figure 2).

The pK_a for the adenine N1 group in the free base is ~4.5. In DNA it has been calculated to be slightly higher. Protonation of the adenines is not likely to be involved in the stabilization of the antiparallel stranded d(GA·GA) duplex described above, since this structure is observed at pH values as high as 8.3. As judged from the patterns of DEPC modification obtained at increasing temperature, this structure is melted at 40°C, at any pH ranging from 4.5 to 8.3 (Figure 1). Moreover, antiparallel stranded d(GA·GA)

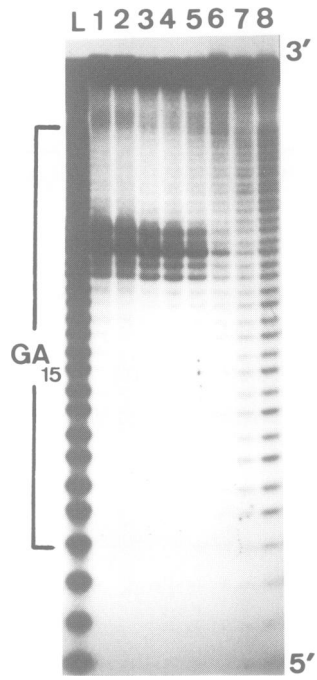

Fig. 3. Antiparallel stranded d(GA·GA) DNA is destabilized at pH ≤ 4.0. Oligo[$(GA)_{15}$] was modified with DEPC, at 4°C, and at decreasing pH: 8.0 (lane 1); 7.0 (lane 2); 6.0 (lane 3); 5.0 (lane 4); 4.5 (lane 5); 4.0 (lane 6); 3.5 (lane 7) and 3.0 (lane 8). All modifications were performed in citric-phosphate buffer at an ionic strength equivalent to 100 mM Na^+. Lane L corresponds to a G+A sequencing ladder. The 5′ to 3′ direction is indicated.

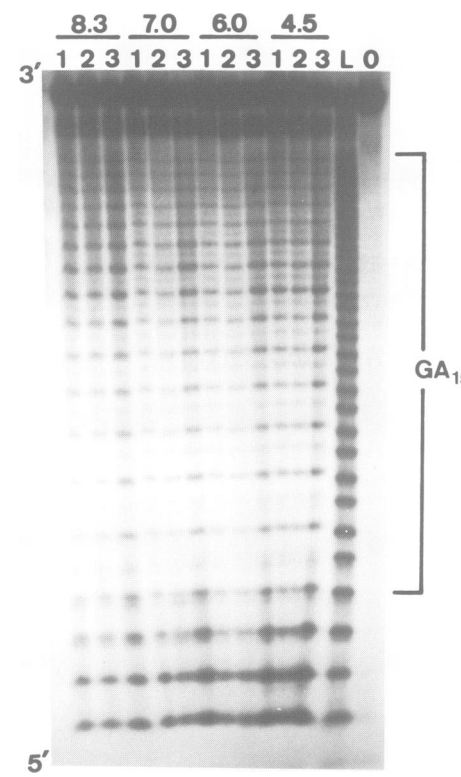

Fig. 4. Patterns of DMS modification of oligo[$(GA)_{15}$] obtained, at the pH values indicated, and at increasing temperature: 4°C (lane 1); 20°C (lane 2) and 40°C (lane 3). Lane L corresponds to a G+A sequencing ladder. Lane 0 corresponds to the piperidine products obtained from the untreated oligonucleotide. The 5′ to 3′ direction is indicated.

DNA is destabilized at pH values lower than the pK_a of the adenine (Figure 3). The pattern of DEPC modification characteristic of the formation of the hairpin, which is observed at 4°C at any pH from 8.3 to 4.5 (Figure 3, lanes 1−5), is no longer observed at pH ≤ 4.0 (Figure 3, lanes 6−8).

The dimethylsulfate (DMS) protection and interference experiments described in Figures 4 and 5, provide additional evidence against the formation of $AH^+(anti)\cdot G(syn)$ pairs. In this pairing scheme, the N7 group of the guanines is involved in hydrogen bonding (Figure 2A). Whether this group is hydrogen bonded or not can be experimentally addressed through the determination of its reactivity with DMS. This reagent methylates guanines at this position and is a good diagnostic tool for the occupancy of this site. It has been used to study the pairing of guanines in triple-stranded and tetra-stranded DNAs (Pulleyblank *et al.*, 1985; Sen and Gilbert, 1988; Voloshin *et al.*, 1988; Sundquist and Klug, 1989; Williamson *et al.*, 1989; Panyutin *et al.*, 1990). As shown in Figure 4, all guanines of the $d(GA)_{15}$ sequence are accessible to DMS methylation at any pH from 4.5 to 8.3. The reactivity at 4°C of the guanines contained within the $d(GA)_{15}$ sequence is similar to that of the guanines located outside of the alternating GA sequence. Furthermore, the patterns of DMS modification obtained at 40°C, at which the hairpin is no longer stable, are identical to those obtained at 4°C (Figure 4, lanes 1 and 3). That the N7 group of the guanines is not involved in hydrogen bonding was also corroborated by the methylation interference experiment described in Figure 5. Oligo[$(GA)_{15}$] was subjected to increasing extents of DMS methylation at 60°C, under conditions at which the hairpin is fully denatured, and its ability to form antiparallel stranded d(GA·GA) duplex analysed by polyacrylamide gel electrophoresis (Figure 5A), as a function of increasing degree of DMS methylation (Figure 5B). The hairpin form of oligo[$(GA)_{15}$] shows an apparent mol. wt of 27 bp (Figure 5A, lane 0), while denatured oligo[$(GA)_{15}$] shows an apparent mol. wt of 52 bp as judged by its electrophoretic migration of 60°C (not shown). According to the patterns of DEPC modification shown in Figure 1A, oligo[$(GA)_{15}$] is partially denatured at 40°C. At this temperature, the complete GA sequence is melted but the rest of the molecule remains double-stranded. This partially denatured form of oligo[$(GA)_{15}$] shows an apparent mol. wt of 42 bp, as judged by its electrophoretic mobility at 40°C (not shown). Methylation decreases slightly the electrophoretic mobility of oligo[$(GA)_{15}$], which appears as a smear showing an apparent mol. wt very close to that of the unmodified hairpin (Figure 5A, lanes 1−7). Increasing methylation does not affect significantly the electrophoretic mobility of oligo[$(GA)_{15}$] which even at very high degrees of methylation (Figure 5B, lane 7) shows an apparent mol. wt in the region corresponding to the hairpin (Figure 5A, lane 7). These results unambiguously rule out the possibility that $AH^+(anti)\cdot G(syn)$ pairs would be involved in the stabilization of antiparallel stranded d(GA·GA) DNA.

Neither of the two alternative base pairing schemes described in Figure 2B and C use the N7 group of the guanines for hydrogen bonding. In the $A(anti)\cdot G(anti)$ scheme the N7 group of the adenines is not involved either in hydrogen bonding (Figure 2B). On the other hand, the

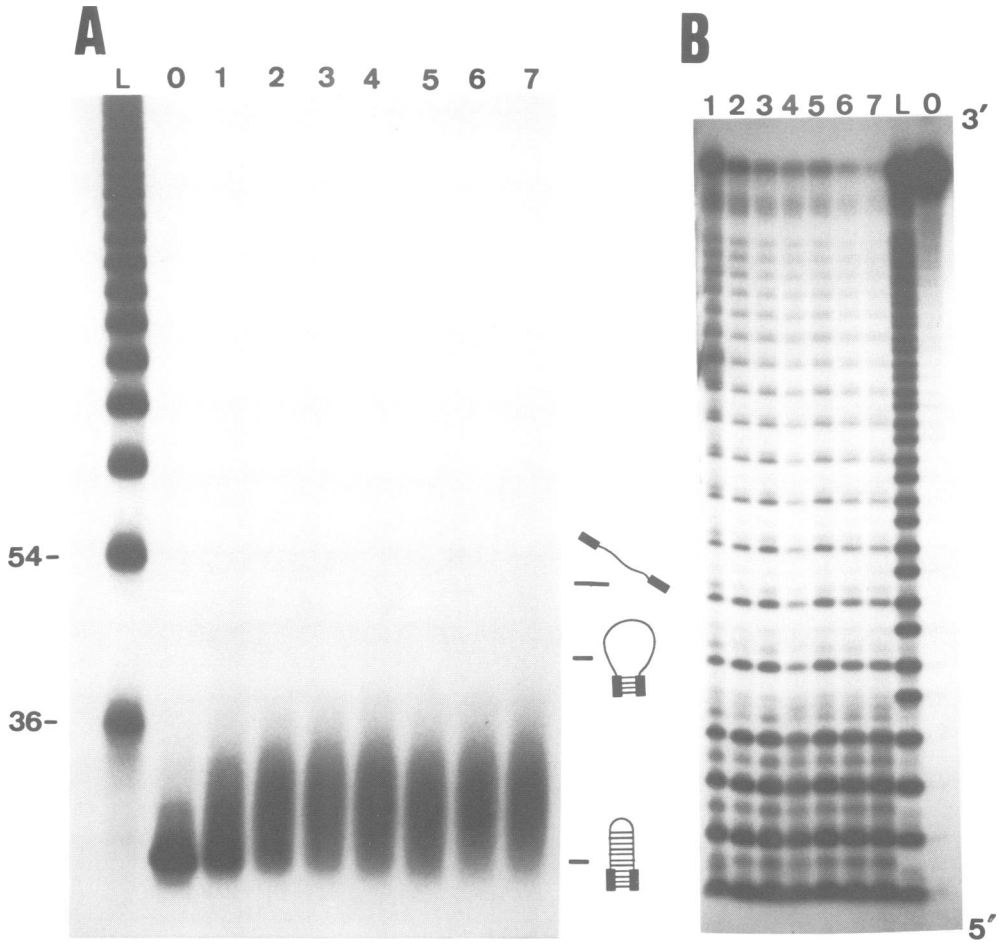

Fig. 5. DMS interference of the formation of antiparallel stranded d(GA·GA) DNA. Oligo[$(GA)_{15}$] was modified with DMS at 60°C for increasing time: 1 min (lane 1); 2 min (lane 2); 3 min (lane 3); 5 min (lane 4); 10 min (lane 5); 15 min (lane 6) and 20 min (lane 7). The ability of each sample to fold into the hairpin structure was then analysed in a 12% native polyacrylamide−TBE native gel run at 4°C (**A**). Lane L in (A) corresponds to the oligomers of an 18mer, used as molecular weight markers. Numbers on the left correspond to the mol. wt in bp of selected oligomers. The electrophoretic migration corresponding to the hairpin, the partially denatured and fully denatured forms of the oligonucleotide are indicated on the right. Lane 0 corresponds to the untreated oligonucleotide. The extent of DMS modification of each sample was determined by cleavage with piperidine and analysis of the cleavage products on a 20% polyacrylamide−7 M urea sequencing gel (**B**). Lane L in (B) corresponds to a G+A sequencing ladder. Lane 0 corresponds to the piperidine cleavage products obtained from the untreated oligonucleotide. The 5′ to 3′ direction is indicated.

A(*syn*)·G(*anti*) pairing scheme is of the Hoogsteen type, the N7 group of the adenines being hydrogen bonded to the N1 group of the guanines (Figure 2C). A detailed analysis of the patterns of DEPC modification shown in Figure 1 reveals that adenine residues located at each side of the loop are likely to be differently paired. Adenines located at the 5′ stem of the hairpin are significantly more reactive to DEPC than those adenines located at the 3′ stem (Figure 1A−C). This differential reactivity of the adenines at either site of the loop is particularly evident when modification is carried out at 20°C (Figure 1, lanes 2). These results suggest that adenine residues located at the 5′ stem of the hairpin use the A(*anti*)·G(*anti*) pairing, in which the adenine N7 group is free for reaction, while those adenines located at the 3′ stem use the A(*syn*)·G(*anti*) scheme, which involves the N7 group of the adenine.

That the adenine N7 groups are to some extent involved in the pairing was corroborated by the DEPC interference experiment described in Figure 6. In contrast to what is observed with DMS (Figure 5), DEPC modification has a strong influence on the ability of the modified molecules to form antiparallel stranded d(GA·GA) duplex. Oligo[$(GA)_{15}$] was subjected to increasing extents of DEPC modification and the ability of the DEPC-modified molecules to adopt the hairpin conformation analysed exactly as described above for DMS. Increasing degrees of DEPC modification result in the unfolding of the hairpin, which is observed only at the lowest degrees of DEPC modification (Figure 6, lanes 1 and 2). At intermediate degrees of DEPC modification, the partially denatured form of the hairpin is observed in which the alternating GA sequence is denatured but the rest of the molecule remains as a duplex (Figure 6, lanes 3−6). The fully denatured form is observed only at very high degrees of modification (Figure 6, lanes 7).

A regular right-handed helical structure can be generated using alternating G(anti)·A(syn) and A(anti)·G(anti) base pairs

The results reported above suggest that antiparallel stranded d(GA·GA) duplexes are stabilized through the formation of alternating G(*anti*)·A(*syn*) and A(*anti*)·G(*anti*) pairs. In this structure, all guanine residues would adopt the *anti*

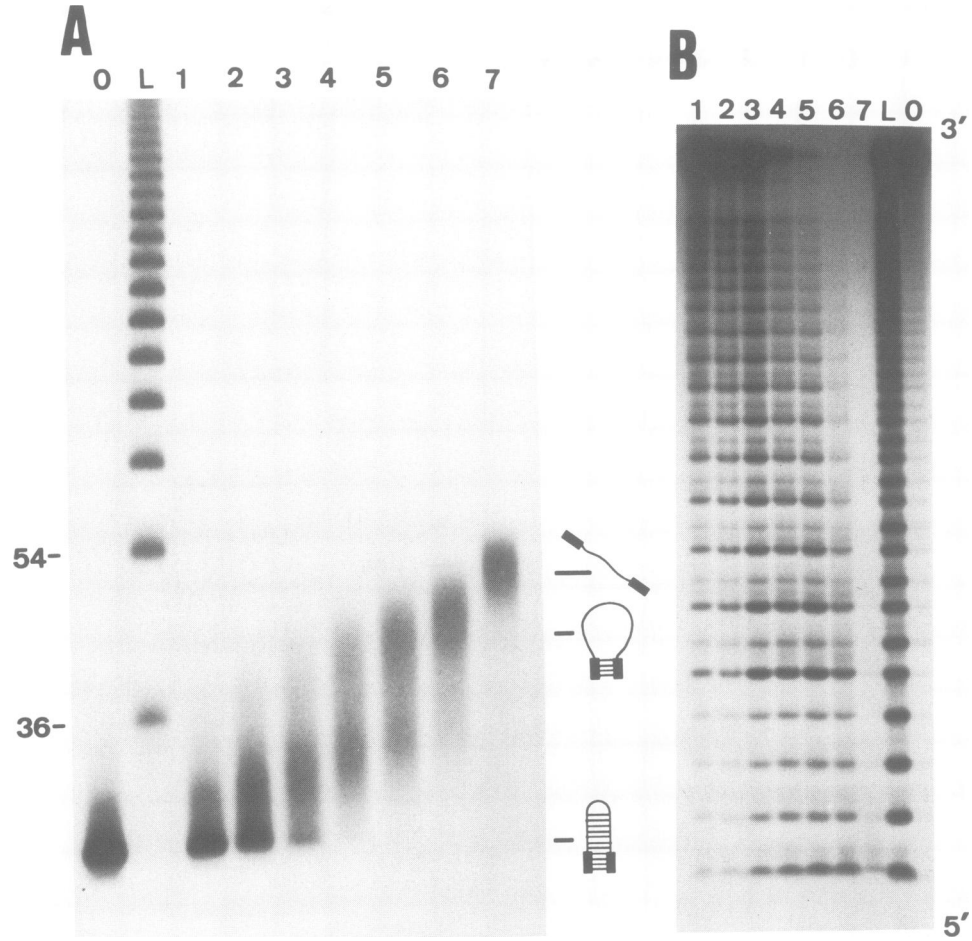

Fig. 6. Interference of DEPC modification on the formation of antiparallel stranded d(GA·GA) DNA. As in Figure 5 but modification of oligo[$(GA)_{15}$] was carried out with DEPC at 60°C for increasing times: 1 min (lane 1); 2 min (lane 2); 5 min (lane 3); 8 min (lane 4); 10 min (lane 5); 15 min (lane 6) and 30 min (lane 7). (**A**) shows the analysis of the ability of the modified oligonucleotide to fold into the hairpin structure. (**B**) shows the extent of modification of each sample in (A).

conformation while adenines would be *syn* in one strand but *anti* in the opposite strand. Based on reported oligonucleotide crystal structures containing G·A mismatches (Privé *et al.*, 1987; Webster *et al.*, 1990) a dinucleotide model was built, one step being a G(*anti*)·A(*anti*) mismatch pair and the other an A(*syn*)·G(*anti*) mismatch pair. A helix was simulated and energy minimized, using the dinucleotide as a subunit and establishing symmetric linkages between subunits. Figure 7A and B are van der Waals representations of the final model, 10 bp long in this plot, facing the major and the minor groove respectively. Figure 7C is a stereo view of the alternating dinucleotide G(*anti*)·A(*anti*) (upper pair) and A(*syn*)·G(*anti*) (lower pair) looking down the helix axis. The DNA helix is right handed and has basically B-DNA parameters, reflecting that it has been built on the ground of G·A mismatches found in B-DNA crystal structures. Some degree of propeller twist (−13°) is observed in the A(*syn*)·G(*anti*) step. This is due to the tilting of the adenine in the *syn* conformation which is a feature already observed in the crystal structure (Webster *et al.*, 1990) that was conserved after minimization. However, the bifurcated hydrogen bond from N7 of A(*syn*) to N1 and N2 of G(*anti*) present in the starting model became a single N7−N1 hydrogen bond after minimization because of the slight displacement of the adenine towards the major groove. The N7 group of the A(*anti*) occupies a rather external position in relation to the helix axis (Figure 7A), which is in agreement with its moderated DEPC reactivity.

The helical twist is lower at the ApG step than at the GpA step, although it is difficult to assess how much this characteristic is biased by the starting ApG dinucleotide model. Minimum phosphate−phosphate distances across the minor groove are constant and ~12 Å; thus the model has a regular and uniform B-like minor groove width. Sugar pucker is C2′-endo for all nucleotides.

Discussion

We have shown in this study that $d(GA)_n$ sequences can form antiparallel stranded duplexes which are stabilized by the formation of G·A pairs. It is known that the base pairing scheme of consecutive G·A mismatches depends strongly on the nature of the neighbouring bases. For instance, formation of A(*anti*)·G(*anti*) base pairs was observed in DNA sequences of the type A-GA-T, containing two consecutive G·A mismatches (Cheng *et al.*, 1992), but no pairing at all was detected in the case of a G-GA-C sequence (Cheng *et al.*, 1992). Our results indicate that two types of base pairing, the A(*anti*)·G(*anti*) and the A(*syn*)·G(*anti*), alternate in antiparallel stranded duplexes containing a large

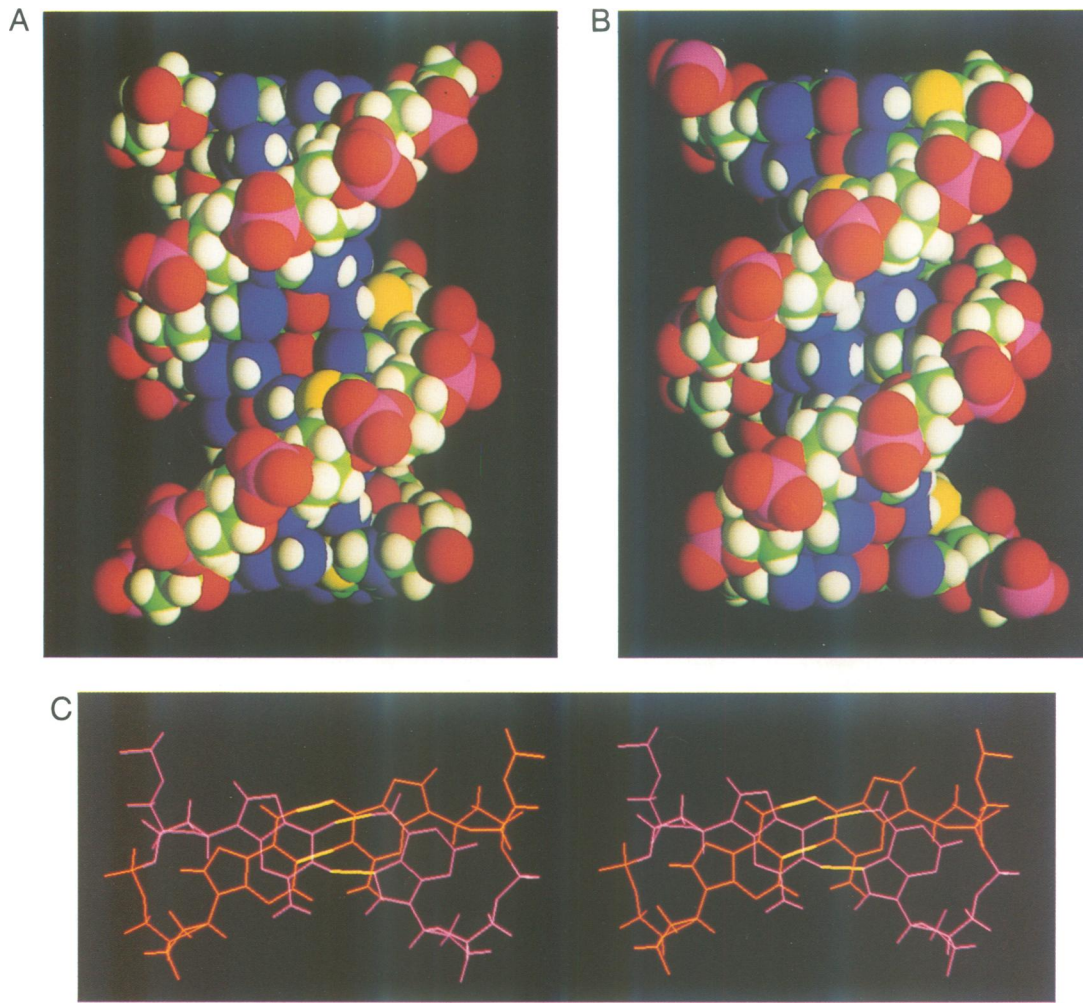

Fig. 7. A structural model for antiparallel d(GA·GA) DNA. Major groove (**A**) and minor groove (**B**) views of the van der Waals representation of the alternating d[G(*anti*)A(*anti*)·G(*anti*)A(*syn*)] model. The double helix is right handed and has B-DNA characteristics. Coloured in yellow are the N7 atoms of A(*anti*) residues. This drawing was generated with the program RIBBONS (Carson and Bugg, 1986). (**C**) Stereo view of the alternating dinucleotide down the helix axis. The upper pair (orange) is G(*anti*)·A(*anti*) whereas the lower pair (pink) is A(*syn*)·G(*anti*). Hydrogen bonds are indicated in yellow.

number of G·A mismatches, all guanine residues adopting the *anti* conformation while adenines are *syn* on one strand but *anti* on the opposite strand. These two types of base pair have been observed to occur both in crystals as well as in solution (Kan *et al.*, 1983; Brown *et al.*, 1986; Hunter *et al.*, 1986; Privé *et al.*, 1987; Nikonowicz and Gorenstein, 1990; Webster *et al.*, 1990). Formation of AH$^+$(*anti*)·G(*syn*) pairs has also been observed at acidic pH (Gao and Patel, 1989; Brown *et al.*, 1989; Leonard *et al.*, 1990). This latter type of pairing uses the N7 group of the guanines which we know is definitely free for reaction with DMS in antiparallel stranded d(GA·GA) DNA.

We also considered the possibility that adenines in both strands will have the same conformation around the glycosylic bond. Our DEPC modification experiments argue against the possibility that all adenines will adopt the *anti* conformation. DEPC modifies adenine residues at their N7 position. This group is not involved in hydrogen bonding in the A(*anti*)·G(*anti*) pairing scheme. The moderated reactivity with DEPC that is observed at the 5′ stem of the hairpin suggests that adenine residues located in this stem are actually adopting the *anti* conformation. If antiparallel stranded d(GA·GA) DNA would be stabilized only through the formation of A(*anti*)·G(*anti*) pairs, all adenine residues should show the same moderated degree of reactivity with DEPC. However, adenines located at the 3′ stem of the hairpin are significantly less reactive with DEPC than those located in the 5′ stem (Figure 1), indicating that, instead of using the A(*anti*)·G(*anti*) pairing, they are likely to use the A(*syn*)·G(*anti*) pairing scheme in which the adenine N7 group is involved in hydrogen bonding. The strong interference of DEPC modification on the formation of antiparallel stranded d(GA·GA) DNA, is also in agreement with this interpretation. It is known that DEPC also carbethoxilates to some extent adenine residues at their N6 position (Vincze *et al.*, 1973). This group is involved in hydrogen bonding in both the A(*anti*)·G(*anti*) and the A(*syn*)·G(*anti*) pairing schemes (Figure 2). Therefore, the DEPC interference experiments described above, could also be interpreted as indicative of the involvement of this group in the pairing. However, DEPC modification has very little effect on the ability to form d(GA·CT)$_n$ duplexes (not shown), although the adenine N6 group is also involved in hydrogen bonding in this case. In good agreement, the

partially denatured form of oligo[$(GA)_{15}$], in which only the GA sequence is denatured, is observed even at high degrees of modification (Figure 6, lane 5 and 6). The fully denatured form is observed only at very high degrees of modification (Figure 6, lanes 7). These results indicate that the strong effect that DEPC modification shows on the stability of antiparallel stranded d(GA·GA) DNA basically reflects the formation of A(*syn*)·G(*anti*) pairs.

The pH dependence of the stability of the hairpin is also in agreement with this interpretation. Protonation of the adenine N1 group is not likely to influence the stability of A(*syn*)·G(*anti*) pairs, since this group is not involved in hydrogen bonding. Therefore, if all adenines would adopt the *syn* conformation, the structure should not be sensitive to protonation. However, as shown in Figure 3, the hairpin conformation of oligo[$(GA)_{15}$] is observed at pH $\geq$ pK_a of the adenine but at lower pH becomes unstable. On the other hand, destabilization of the hairpin form will be expected to occur at a higher pH if only A(*anti*)·G(*anti*) pairs will account for its stability, since protonation of the N1 group of the adenine is incompatible with base pairing formation in this case.

Another type of A(*anti*)·G(*anti*) pairing with completely different hydrogen bonding has been proposed (Li *et al.*, 1991a,b). In this case, pairing involves the N2 amino group of the guanine which is hydrogen bonded to the N7 imino group of the adenine. An additional hydrogen bond is formed between the guanine N3 group and the N6 amino group of the adenine. This type of pairing would show a pattern of chemical reactivity similar to that of a A(*syn*)·G(*anti*) pair. In both cases, the N7 group of the guanines, but not of the adenines, will be accessible for modification by DMS or DEPC, respectively. Formation of this type of pairing has been observed only in sequences containing adjacent G·A mismatches of the type Py-GA-Pu (Cheng *et al.*, 1992; Ebel *et al.*, 1992). From the results presented here we cannot completely exclude the possibility that this type of pairing could be formed in antiparallel stranded d(GA·GA) DNA instead of the A(*syn*)·G(*anti*) base pairing. However, a very important contribution to the stability of this type of base pairing arises from strong G−G and A−A cross-strand stacking interactions which would not take place if this type of pairing alternates with usual A(*anti*)·G(*anti*) pairs. Alternation of these two types of pairing would result in a quite distorted helix.

Formation of parallel stranded d(GA·GA) duplexes has also been reported (Rippe *et al.*, 1992). This type of duplex is stabilized by divalent cations, mainly magnesium, and it involves the formation of G·G and A·A base pairs. A $d(GA)_{20}$ oligonucleotide was found to form intermolecular complexes which could be parallel stranded (Casasnovas *et al.*, in preparation). However, with the oligonucleotides described here and in the range of DNA concentrations used in these experiments, we have never observed formation of such intermolecular complexes. From the available data, it is difficult to ascertain the type of duplex preferred by $d(GA)_n$ sequences since they show similar stabilities. Parallel stranded $d(GA·GA)_{15}$ melts at 25°C in the presence of 10 mM $MgCl_2$ (Rippe *et al.*, 1992). On the other hand, the results reported in Figure 1 indicate that the antiparallel stranded form of oligo[$(GA)_{15}$] is stable at 20°C at relatively low ionic strength (TBE buffer, pH 8.3). A much lower melting temperature of ~4°C was observed in a decamer containing four G-A mismatches (Ebel *et al.*, 1992), suggesting that short antiparallel stranded d(GA·GA) duplexes might use a different pairing scheme.

It has also been proposed that alternating $d(GA)_n$ sequences can form tetra-stranded conformations, which are also stabilized by divalent cations (Lee *et al.*, 1980; Lee, 1990). There was no evidence of the formation of these multistranded complexes in the experiments described in this paper.

Alternating $d(GA·CT)_n$ sequences are known to adopt a variety of conformations depending upon the precise environmental conditions (reviewed in Wells *et al.*, 1988; Palecek, 1991). In particular, formation of intramolecular pyr-pur-pur triplexes (*H-triplex) was observed, at neutral pH and in the presence of particular transition metal ions such as zinc (Bernués *et al.*, 1989). In the *H-triplex, the two purine strands are antiparallel. Upon increasing the metal ion concentration, the pyrimidine strand falls off the triplex and a pur-pur hairpin is formed, in which the purine strands are also antiparallel (Beltrán *et al.*, 1993). The results obtained here suggest that this *H-hairpin is likely to be stabilized by the formation of G·A pairs of the type described here. Preliminary results indicate that zinc does not have any major effect either on the stability of antiparallel stranded d(GA·GA) or in the type of pairing which stabilizes this conformation (Ortiz-Lombardía *et al.*, unpublished results). The situation is not so clear in the case of the *H-triplex. Prevalence of the G·A pairing will imply that T·A·G and C·G·A base triads would be formed, which are known to be only moderately stable (Beal and Dervan, 1992). In addition, the interaction of an adenine residue with a C·G base pair through the major groove must almost necessarily involve the N7 group of the guanine. Experiments are under way to establish the type of pairing which holds together the two purine strands in the *H-triplex, as well as the effect of zinc on this structure.

Biological significance

Alternating $d(GA·CT)_n$ sequences are frequent in eukaryotic genomic DNA and they show a high degree of structural polymorphism. As discussed above, purine−purine interactions appear to predominate at conditions close to physiological. Intramolecular pyr-pur-pur triplexes or pur-pur hairpins, in which the two purine strands are antiparallel, are known to occur at neutral pH in the presence of some metal ions. Formation of parallel stranded $d(GA·GA)_n$ duplexes as well as tetraplexes has also been reported. It is still unclear whether the same high degree of structural polymorphism that these sequences exhibit *in vitro* is also manifested *in vivo*. The characterization of protein factors showing specific binding activities for single-stranded $d(CT)_n$ DNA fragments, but not for $d(GA)_n$ (Yee *et al.*, 1991; Kolluri *et al.*, 1992), provides circumstantial evidence for the *in vivo* existence of pur−pur interactions. Regions of single stranded d(CT) will accompany the formation of pur-pur associations of the type described here. Protein binding at these sites could stabilize such complexes.

The type of purine−purine interaction described here may be relevant in a number of different transactions involving the DNA molecule. Increasing evidence suggests that $d(GA·CT)_n$ sequences are likely to participate in processes of recombination. This type of sequence is found at genomic

locations which are known to undergo frequent recombinatory events. SV40 viruses carrying a d(GA·CT)$_{22}$ sequence become highly unstable, probably due to an increase in the frequency of intramolecular recombination (Bernués *et al.*, 1991). Finally, a 10-fold increase on the rate of excision of plasmid DNA inserted into SV40 was observed when a copy of this type of sequence was present at either end of the inserted plasmid DNA (Beltrán *et al.*, unpublished results). Nothing is known about the molecular mechanisms which could eventually lead to the formation and resolution of a four-way DNA junction at these sites. Purine−purine interactions could play an important role in the alignment of the sites to be involved in recombination. This type of interaction could also take place during the formation and/or resolution of the actual four-way DNA junction, stabilizing intermediate(s) of these processes.

It has also been suggested that d(GA·CT)$_n$ sequences might serve to stop gene amplification events, imposing limits on the extent of the amplified regions (Schimke, 1988). Studies performed in living cells have indicated that d(GA·CT)$_n$ sequences are pause or arrest sites for DNA replication and amplification (Sures *et al.*, 1978; Cheng *et al.*, 1982; Crabtree and Kant, 1982; Hentschel, 1982; Hunt *et al.*, 1984; Manor *et al.*, 1988; Sridhara Rao *et al.*, 1988). Recent studies have shown that pur−pur interactions account for the stop on the elongation of a d(GA)$_n$-containing single stranded DNA fragment by the Klenow enzyme and/or the *Taq* polymerase (Baran *et al.*, 1991).

Formation of A·G pairs might also occur in telomeric DNA. A number of telomeric repeats, including the TTAGGG repeat present in all vertebrate species, contain adenine residues in the vicinity of the characteristic guanine tract. It is known that, in addition to the formation of tetraplexes, these sequences could also form intramolecular folded-back duplexes (Henderson *et al.*, 1987). Formation of A·G pairs could contribute to the stabilization of these telomeric structures.

Materials and methods

DNAs

All oligonucleotides were synthesized on an Applied Biosystems automatic synthesizer and purified by polyacrylamide gel electrophoresis. Oligonucleotides were end-labelled with [γ-^{32}P]ATP and polynucleotide kinase. Before modification, samples were annealed at 40°C.

Conditions for DEPC modification

The patterns of DEPC modification were obtained at different pH values, from 3.0 to 8.3, and at increasing temperatures, from 4 to 40°C. All modifications were performed at a [DNA] = 10^{-4} μg/μl, in a final volume of 20 μl, with 2 μl of DEPC (Sigma) in citric-phosphate buffer at an ionic strength equivalent to 100 mM Na^+. When modification was carried out at pH 8.3, TBE buffer was used instead. Reactions were allowed to proceed for 3 h at 4°C, 2 h at 20°C or 20 min at 40°C. After modification, samples were ethanol precipitated, cleaved with 1 M piperidine and analysed on 20% polyacrylamide−7 M urea sequencing gels.

When the effect of DEPC modification on the formation of antiparallel stranded d(GA·GA) DNA was analysed, samples were treated with DEPC as described before at 60°C for increasing times from 1 to 30 min, at pH 8.3. After modification, samples were annealed as indicated above and their ability to form the hairpin structure was analysed on native 12% polyacrylamide−TBE gels. The extent of DEPC modification was determined for each sample by cleavage with piperidine and analysis of the cleavage products in 20% polyacrylamide−7 M urea sequencing gels.

Conditions for DMS modification

DMS modification was carried out as indicated above for DEPC, but 1 μl of DMS (Sigma) was used. Modification was allowed to proceed for 15 min at 4°C, 3 min at 20°C and 50 s at 40°C, in a final volume of 200 μl, in the presence of 4−8 μg of carrier DNA.

For the DMS interference experiments, oligonucleotides were treated with 1 μl of DMS at 60°C, in a final volume of 400 μl, for increasing times in the same conditions as indicated above for DEPC. After modification samples were analysed as indicated above for DEPC.

Model structure

The dinucleotide d[G(*anti*)A(*anti*)·G(*anti*)A(*syn*)] was built 'manually' on a Cyber 910 graphic station using the program TOM (Cambillau and Horjales, 1987), a modified version of FRODO (Jones, 1982). The d[G(*anti*)A(*anti*)·G(*anti*)A(*anti*)] dinucleotide from the d(CCAAGATTGG) crystal structure (Privé *et al.*, 1987) was superimposed on the d[C(*anti*)A(*syn*)·G(*anti*)G(*anti*)] dinucleotide from the d(CGCAAGC-TGGCG) crystal structure (Webster *et al.*, 1990). After this, the coordinates of G(*anti*)·A(*anti*) base pair from the first dinucleotide and those of the A(*syn*)·G(*anti*) base pair from the second dinucleotide were kept. The dinucleotide generated in this way was energy minimized with the program X-PLOR (Brünger *et al.*, 1985) using modified DNA parameters from the CHARMM force field. A subroutine for fibre refinement, written and implemented in X-PLOR by H.Wang and G.Stubbs, and including helical symmetry and atom linkages between helical subunits was used to generate and energy minimize the DNA helix. Different energy minimization runs were performed systematically for different helical repeat values. The optimized model obtained this way was inspected visually with TOM and geometrically analyzed with the program NEWHEL, written by R.Dickerson.

Acknowledgements

This work was financed by grants from the Ministerio de Educación y Ciencia (PB90-997), the CEC (BIOT-CT91-259) and the Generalitat de Catalunya. M.C. acknowledges support from CICYT, grant SAL90-824. D.H. and L.B. were recipients of doctoral fellowships from the Ministerio de Educación y Ciencia and the Spanish Research Council-CSIC, respectively. J.M.C. was a recipient of a postdoctoral fellowship from the CSIC.

References

Baran,N., Lapidot,A. and Manor,H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 507−511.
Beal,P.A. and Dervan,P.B. (1992) *Nucleic Acids Res.*, **20**, 2773−2776.
Beltrán,R., Martínez-Balbás,A., Bernués,J., Bowater,R. and Azorín,F. (1993) *J. Mol. Biol.* **230**, 966−978.
Bernués,J., Beltrán,R., Casasnovas,J.M. and Azorín,F. (1989) *EMBO J.*, **8**, 2087−2094.
Bernués,J., Beltrán,R. and Azorín,F. (1991) *Gene*, **108**, 269−274.
Brown,T., Hunter,W.N., Kneale,G. and Kennard,O. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 2402−2406.
Brown,T., Leonard,G.A., Booth,E.D. and Chambers,J. (1989) *J. Mol. Biol.*, **207**, 455−457.
Brünger,A.T., Kuriyan,K. and Karplus,M. (1985) *Science*, **235**, 458−460.
Cambillau,C. and Horjales,E. (1987) *J. Mol. Graphics*, **5**, 174−177.
Carson,M. and Bugg,C.E. (1986) *J. Mol. Graphics*, **4**, 121−122.
Cheng,H.L., Blattner,F.R., Fritzmaurice,L., Mushinski,J.F. and Tucker,P.W. (1982) *Nature*, **296**, 410−415.
Cheng,J.-W., Chou,S.-H. and Reid,B.R. (1992) *J. Mol. Biol.*, **228**, 1037−1041.
Collier,D.A., Griffin,J.A. and Wells,R.D. (1988) *J. Biol. Chem.*, **263**, 7397−7405.
Crabtree,G.R. and Kant,J.A. (1982) *Cell*, **31**, 159−166.
Ebel,S., Lane,A.N. and Brown,T. (1992) *Biochemistry*, **31**, 12083−12086.
Gao,X. and Patel,D.J. (1989) *J. Am. Chem. Res.*, **110**, 5178−5182.
Glikin,G.C., Gargiulo,G., Rena-Descalzi,L. and Worcel,A. (1983) *Nature*, **303**, 770−774.
Henderson,E.R., Hardin,C.C., Wolk,S.K., Tinoco,I.,Jr and Blackburn,E.H. (1987) *Cell*, **51**, 899−908.
Hentschel,C.C. (1982) *Nature*, **295**, 714−716.
Hunt,H., Lund,E. and Dalberg,J.E. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 7288−7292.
Hunter,W.N., Brown,T. and Kennard,O. (1986) *J. Biomol. Struct. Dyn.*, **4**, 173−191.
Jones,T.A. (1982) In Sayre,D. (ed.), *Computational Crystallography*. Clarendon Press, Oxford, pp. 303−310.
Kan,L.S., Chandrasegaran,S., Pulford,S.M. and Miller,P.S. (1983) *Proc. Natl Acad. Sci. USA*, **80**, 4263−4265.

Kolluri,R., Torrey,T.A. and Kinniburgh,A.J. (1992) *Nucleic Acids Res.*, **20**, 111–116.
Lee,J.S. (1990) *Nucleic Acids Res.*, **18**, 6057–6060.
Lee,J.S., Evans,D.H. and Morgan,A.R. (1980) *Nucleic Acids Res.*, **8**, 4305–4320.
Leonard,G.A., Booth,E.D. and Brown,T. (1990) *Nucleic Acids Res.*, **18**, 5617–5623.
Li,Y., Zon,G. and Wilson,W.D. (1991a) *Proc. Natl Acad. Sci. USA*, **88**, 26–30.
Li,Y., Zon,G. and Wilson,W.D. (1991b) *Biochemistry*, **30**, 7566–7572.
Lyamichev,V.J., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1986) *J. Biol. Mol. Struct. Dyn.*, **3**, 667–669.
Manor,H., Sridhara Rao,B. and Martin,R.G. (1988) *J. Mol. Evol.*, **27**, 96–101.
Mason,A.J., Evans,B.A., Cox,D.R., Shine,J. and Richards,R.I. (1983) *Nature*, **303**, 300–307.
Nikonowicz,E.P. and Gorenstein,D.G. (1990) *Biochemistry*, **29**, 8845–8858.
Palecek,E. (1991) *CRC Crit. Rev. Biochem. Mol. Biol.*, **26**, 151–226.
Panyutin,I.G., Kovalsky,O.I., Budowsky,E.I., Dickerson,R.E., Rikhirev,M.E. and Lipanov,A.A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 867–870.
Privé,G.G., Heinemann,U., Chandrasegaran,S., Kan,L.-S., Kopka,M.L. and Dickerson,R.E. (1987) *Science*, **238**, 498–504.
Pulleyblank,D.E., Haniford,D.B. and Morgan,A.R. (1985) *Cell*, **42**, 271–280.
Richards,J.E., Gilliam,A.C., Shen,A., Tucker,P.W. and Blattner,F.R. (1983) *Nature*, **306**, 483–487.
Rippe,K., Fritsch,V., Westhof,E. and Jovin,T.M. (1992) *EMBO J.*, **11**, 3777–3786.
Schimke,R.T. (1988) *J. Biol. Chem.*, **263**, 5989–5992.
Sekiya,T., Kuchino,Y. and Nishimura,S. (1981) *Nucleic Acids Res.*, **9**, 2239–2250.
Sen,D. and Gilbert,W. (1988) *Nature*, **334**, 364–366.
Sridhara Rao,B., Manor,H. and Martin,R.G. (1988) *Nucleic Acids Res.*, **16**, 8077–8094.
Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.
Sures,I., Lowry,J. and Kedes,L.H. (1978) *Cell*, **15**, 1033–1044.
Vincze,A., Henderson,R.E.L., McDonald,J.J. and Leonard,N.J. (1973) *J. Am. Chem. Soc.*, **95**, 2677–2682.
Voloshin,O.N., Mirkin,S.M., Lyamichev,V.I., Belotserkovskii,B.P. and Frank-Kamenetskii,M.D. (1988) *Nature*, **333**, 475–479.
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.
Weinreb,A., Collier,D.A., Birshtein,B.K. and Wells,R.D. (1990) *J. Biol. Chem.*, **265**, 1352–1359.
Wells,R.D., Collier,D.A., Hanvey,J.C., Shimizu,M. and Wohlrab,F. (1988) *FASEB J.*, **2**, 2939–2949.
Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell*, **59**, 871–880.
Yee,H.A., Wong,A.K.C., van de Sande,J.H. and Rattner,J.B. (1991) *Nucleic Acids Res.*, **19**, 949–953.

Received on April 13, 1993; revised on May 26, 1993