Long $(dA)_n.(dT)_n$ tracts can form intramolecular triplexes under superhelical stress

Keith R.Fox
Department of Physiology and Pharmacology, University of Southampton, Bassett Crescent East, Southampton SO9 3TU, UK

Received July 9, 1990; Revised and Accepted August 24, 1990

ABSTRACT

Plasmids containing long tracts of $(dA)_n.(dT)_n$ have been prepared and their conformations examined in linear and supercoiled DNA using a series of chemical and enzymic probes which are known to be sensitive to unusual DNA structures. Under superhelical stress and in the presence of magnesium the sequence $T_{69}.A_{69}$ adopts a conformation at pH 8.0 consistent with the formation of an intramolecular DNA triplex. Site specific cleavage of the supercoiled plasmid by single-strand specific nucleases occurs within the A.T insert; the 5'-end of the purine strand is sensitive to reaction with diethylpyrocarbonate while the central 5 – 6 bases of the pyrimidine strand are reactive to osmium tetroxide. By contrast shorter inserts of $A_{33}.T_{33}$ and $A_{23}.T_{23}$ do not appear to form unusual structures.

INTRODUCTION

Homopurine-homopyrimidine tracts are known to adopt unusual structures (H-DNA) under superhelical stress. These are proposed to consist of an intramolecular triplex in which half the pyrimidine strand is unpaired and forms Hoogsteen base pairs with the remaining DNA duplex; the other half of the purine strand is left in an unstructured, single-stranded configuration [1–6]. Various probes have been used to detect the formation of these unusual DNA structures including single-strand specific nucleases (S1 and P1), diethylpyrocarbonate (DEPC) and osmium tetroxide. These react with the long single stranded loop of the purine strand and the short loop of the pyrimidine strand. These structures have previously been detected in DNAs containing 25–100% (G+C) residues [6], and are generally formed at low pH, necessary to form the C^+GC triplet.

Long stretches of homopurine.homopyrimidine commonly occur throughout many eukaryotic genomes [7] and several suggestions have been made concerning their biological relevance. As regards the structure of runs of A or T much attention has previously been given to the occurrence of short runs of A which can cause DNA bending [8,9]; less has been reported concerning longer runs of A or T, although these are also of widespread occurrence. Although polydA.2polydT is known to form a stable triple helix [10] and $(dA)_{10}.2(dT)_{10}$ has been characterised by NMR studies [11] no intramolecular triplexes consisting of only A and T bases have been detected. Indeed Wells and coworkers have demonstrated that $(dA)_{20}.(dT)_{20}$ does not adopt an unusual structure under superhelical stress [6]. In this paper we examine the ability of long runs of (dA).(dT) to adopt similar conformations under superhelical stress. In principle these sequences ought to be able to form intramolecular triplexes at physiological pHs since no extra protonation is required to form the addition A.T Hoogsteen base pair in the T.A.T triplet.

MATERIALS AND METHODS

Plasmids

Plasmids containing long runs of dA.dT or (dA-dG).(dC-dT) were made as follows. PolydA.polydT or poly(dA-dG).poly(dC-dT) were digested with DNAase I generating short fragments which were made blunt-ended by filling in with reverse transcriptase and the appropriate nucleotide triphosphates. These oligonucleotide fragments were cloned into the SmaI site of plasmid pUC19 forming the local sequence GGTACCC(insert)GGGGATC within the EcoR1-HindIII polylinker. The resultant plasmids were transformed into E.coli TG2, isolated and characterised by Maxam-Gilbert sequencing. The four inserts used in this work contained the sequences A_{23}, A_{33}, T_{69} and $A(GA)_{49}$ reading along the coding strand of the plasmid.

Probes of DNA structure

S1 nuclease. 1μg samples of plasmid were cut with 4 units S1 nuclease at 37°C for 5 minutes in 50mM sodium acetate buffer pH 5.0 containing 1mM $MgCl_2$ and 0.5mM $ZnCl_2$. The position of the S1 sensitive sites was mapped by further digestion of these samples with restriction enzyme ScaI which has a unique cutting site on this plasmid. These products were compared with plasmids digested by both EcoRI and ScaI. The products of digestion were resolved on 0.8% agarose gels. For the fine mapping of the S1 sensitive sites 1μg samples of plasmid at native superhelical density were digested by S1 nuclease as described above. The samples were then cut with HindIII (T strand) or EcoR1 (A strand) and labelled with α-^{32}P-dATP using reverse transcriptase before digesting with the second enzyme. This releases the EcoR1-HindIII polylinker containing the insert which is uniquely labelled at one end. The radiolabelled polylinker fragments containing these inserts were eluted from 6% polyacrylamide gels. The cleavage products were resolved on 8% polyacrylamide gels containing 8M urea.

Diethylpyrocarbonate (DEPC). 1μg samples of plasmid, dissolved in 10μl buffer was reacted with 5μl DEPC for 5 minutes at 37°C. The modified plasmids were precipitated with ethanol and cut with HindIII (A_{23} and A_{33}) or EcoR1 (T_{69}), labelled with α-^{32}P-dATP using reverse transcriptase and cut again with the other enzyme. The short radiolabelled polylinker fragments containing the inserts were purified from 6% polyacrylamide gels before boiling in 10% piperidine and lyophilising. The products of reaction were separated on 13% (EcoR1 labelled) or 8% (HindIII labelled) polyacrylamide gels containing 8M urea.

Osmium tetroxide. 1μg samples of plasmid, dissolved in 10μl buffer and 1μl pyridine were reacted with 8μl 2.5mM OsO_4 for 5 minutes. The reaction was stopped by precipitation after which the polylinker containing the insert was labelled and isolated as described above for DEPC. The samples were then treated with hot piperidine, lyophilised and resolved on denaturing polyacrylamide gels.

Electrophoresis

Polyacrylamide gels (8−15% as appropriate) containing 8M urea were run for 1.5−2 hours at 1500V. These were then fixed in acetic acid, dried under vacuum at 80°C and subjected to autoradiography at −70°C with an intensifying screen. Autoradiographs were scanned with a Joyce-Loebl Chromoscan 3 microdensitometer.

RESULTS

S1 nuclease

Plasmids which contain intramolecular triplexes are unusually sensitive to single strand specific nucleases since the purine and pyrimidine strands contain long and short single stranded regions respectively. Figure 1a presents S1 nuclease digestion of plasmids containing A_{23}, A_{33} and T_{69} inserts compared with an $A(GA)_{49}$ insert, a sequence which is known to form an intramolecular triplex under superhelical stress at low pH. Plasmids containing $A(GA)_{49}$ and T_{69} inserts show significant cleavage by this nuclease, suggestive of the presence of an unusual DNA structure, containing single-stranded regions. In contrast plasmids containing shorter poly(dA) inserts show little or no cleavage by S1 nuclease. The position of the S1 sensitive sites was mapped by further digestion with ScaI. Since ScaI cuts the plasmids only

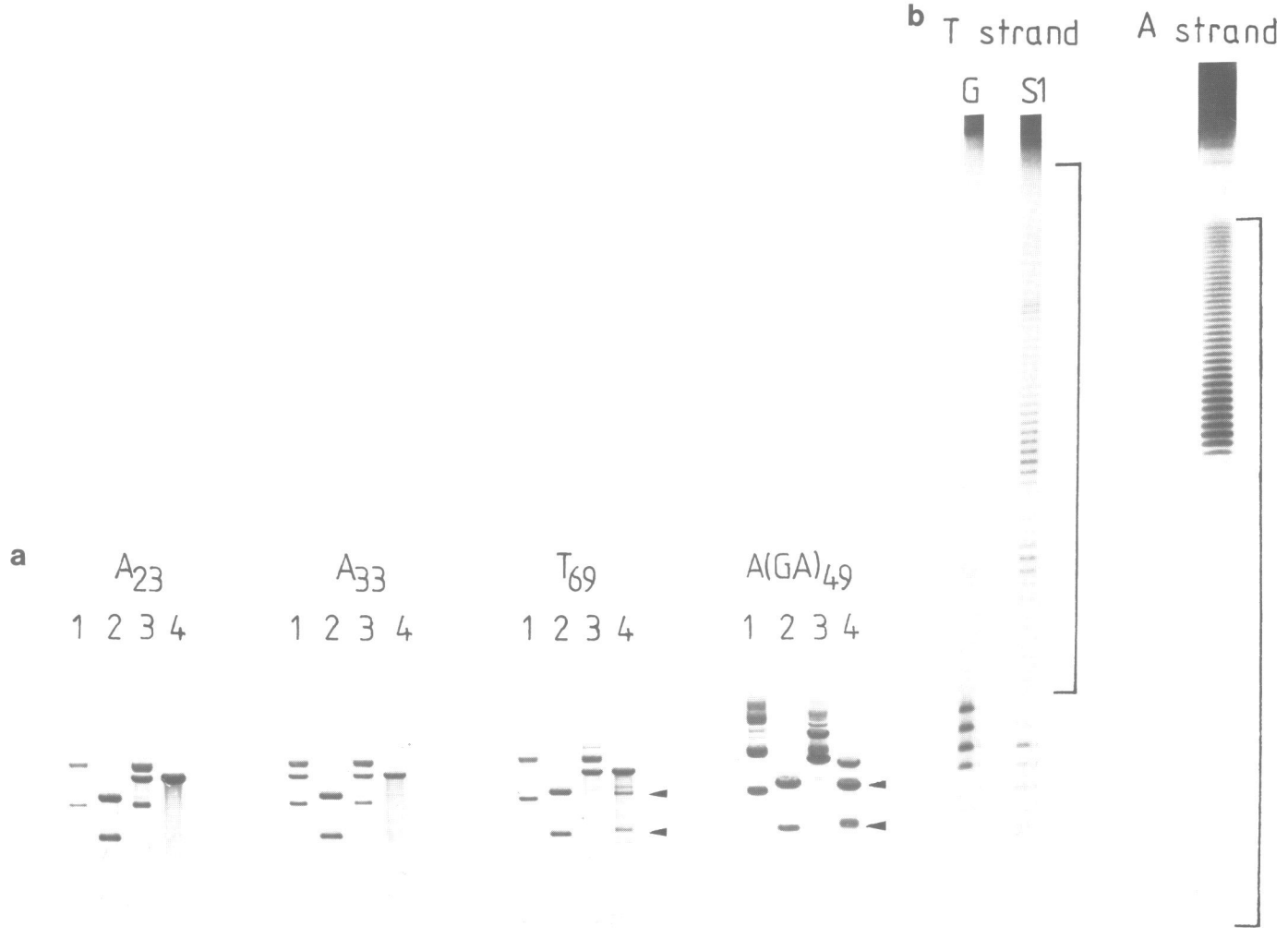

Figure 1. Nuclease S1 cleavage of plasmids containing homopurine.homopyrimidine inserts. a) Coarse mapping of S1 cleavage sites. Lane 1 corresponds to the intact plasmids. In lane 2 the plasmid has been digested by EcoR1 and ScaI each of which has a unique cutting site. Lane 3 corresponds to digestion of the plasmid with S1 nuclease. In lane 4 the S1 treated plasmid has been cut with ScaI. The products of digestion were resolved on 0.8% agarose gels. The two bands corresponding to site specific S1 cleavage, seen after digestion with ScaI, are indicated by arrows. b) Fine mapping of the S1 nuclease cleavage sites on the T_{69} plasmid insert. Square brackets show the position and length of the $T_{69}.A_{69}$ insert. The track labelled 'G; is a Maxam-Gilbert dimethylsulphate-piperidine marker specific for guanine.

once in a region remote from the polylinker, two fragments of lengths 900 and 1800 base pairs will be generated if S1 nuclease has cleaved at specific sites, whereas random S1 cleavage will result in a smear of products of random length. Two such shorter fragments can be seen for the T_{69} and $A(GA)_{49}$ inserts though not with A_{23} and A_{33}. Comparison with EcoR1 and ScaI cleavage confirms that S1 nuclease is cutting close to the EcoR1 site. Similar results are obtained for the T_{69} insert at pH 8.0 using nuclease P1. If the plasmids are first linearised with ScaI, before being treated with S1 nuclease, no site specific nuclease cleavage is detected.

The nature of the S1 cleavage products has been examined by fine mapping studies involving isolation and labelling of short fragments containing the inserts after S1 digestion. The results are presented in Fig. 1b and confirm that the 5′ half of the A-strand is peculiarly sensitive to cleavage, especially towards the centre of the insert. The corresponding 3′-half of the T-strand is also sensitive to this nuclease, though less so than the purine strand.

Diethylpyrocarbonate and osmium tetroxide

These two chemicals are widely used as probes of unusual DNA structures since they react with exposed regions of DNA bases which are normally inaccessible in a B-DNA helix. The former reacts at purine N7 (especially adenine), while the latter detects exposed thymines. These probes react with bases in single-stranded regions but not with the triplex itself [3−6]. Fine mapping studies with these two probes on the inserts are shown in Figs. 2 and 3. It can be seen that the 5′-half of the purine strand of the T_{69} insert is hyperreactive to DEPC (Figure 2); very little reaction occurs at the 3'-end. In contrast, the adenine containing strands of the shorter inserts show an even cleavage throughout their entire length. Similar results are obtained at both pH 5.0 and 8.0. Figure 3 shows the reaction of osmium tetroxide with the thymine-containing strand of the T_{69} insert. It can be seen that the only reactive bases are located at the central 5−6 bases of the sequence, and that this occurs only in the presence of magnesium (see below). This is again consistent with the formation of an intramolecular triplex since the centre of the pyrimidine strand forms a short single stranded loop. Similar

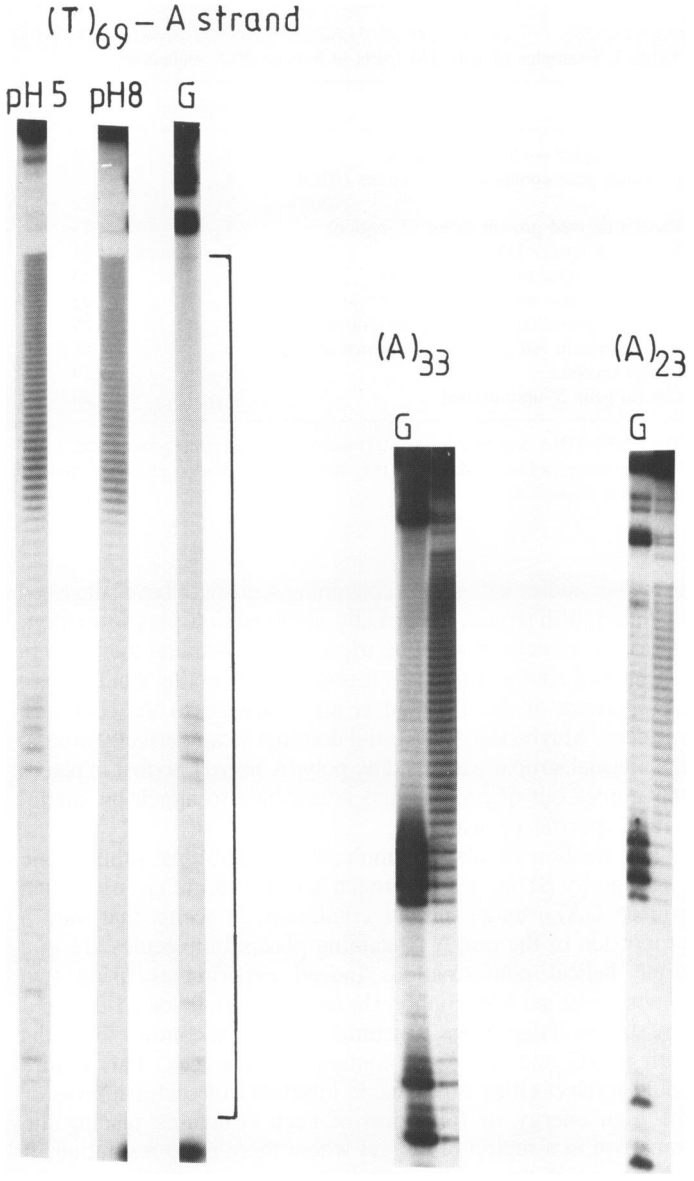

Figure 2. Fine mapping of DEPC on the three (dA).(dT) plasmid inserts. The position and length of the T_{69} insert is indicated by the square bracket. Tracks labelled 'G' correspond to Maxam-Gilbert dimethylsulphate-piperidine markers specific for guanine. The reactions were performed at pH 8.0, except where indicated, in the presence of 1mM $MgCl_2$.

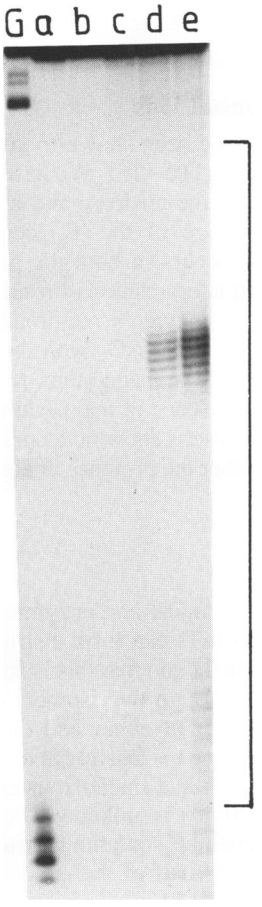

Figure 3.. Modification of the T_{69} insert by OsO_4 in the presence of various concentrations of $MgCl_2$. Lanes a-e correspond to 0, 0.1, 0.2, 0.5 and 1.0mM $MgCl_2$. The track labelled 'G' is a Maxam-Gilbert marker specific for guanine. The positions of the inserts are indicated by square brackets.

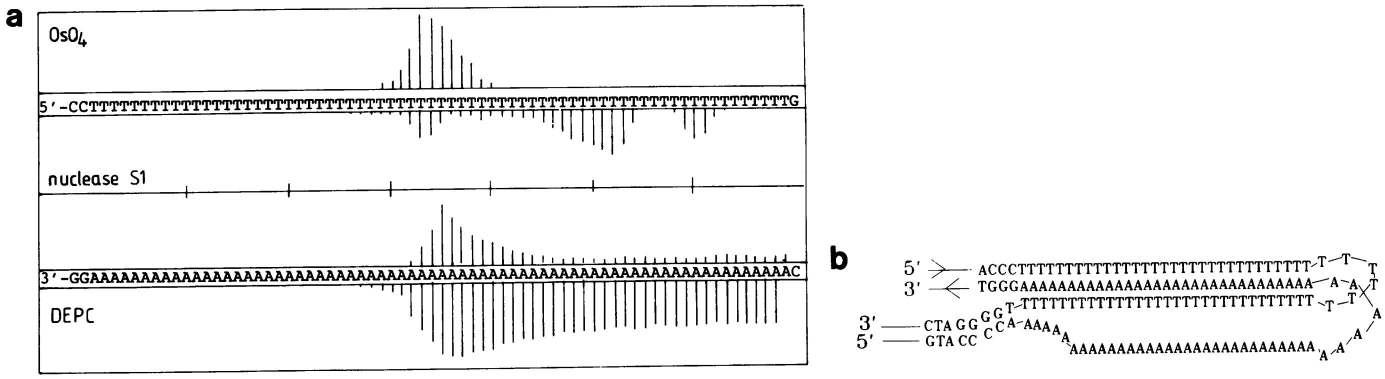

Figure 4. a) Summary of the enzymic and chemical cleavage data for the $T_{69}.A_{69}$ insert. The height of each band represents the relative cleavage at each DNA base assessed by densitometer scans of the data in Figures 1–3. The data for DEPC and OsO_4 were obtained at pH 8.0, those for S1 nuclease were at pH 5.5. All reactions contained 1mM $MgCl_2$. b) Schematic representation of a model for the formation of the intramolecular triple helix in the $T_{69}.A_{69}$ insert.

experiments with the shorter inserts revealed no thymines hyperreactive to OsO_4.

The results of densitometer traces of the data for S1 nuclease, DEPC and osmium tetroxide modification of the T_{69} insert are summarised in Figure 4a. These data are consistent with the formation of an intramolecular triple helix. The reasons why the shorter inserts do not form such a structure under superhelical stress will be considered further in the discussion. A schematic representation of a model for intramolecular helix formation is presented in Figure 4b. With all the probes no site specific modification was detected if the plasmid was first linearised.

Effect of divalent metal ions

Divalent metal ions have been shown to affect the formation of unusual DNA structures [12–15]. We have therefore examined the effect of $MgCl_2$ on the nuclease and chemical reactivity of these plasmids. Cleavage of the T_{69} containing plasmid by nucleases P1 and S1 occurs in both the presence and absence of magnesium, though interpretation of results with the latter may be complicated by the enzyme's requirement for zinc. Figure 3 shows that the reaction of OsO_4 with the central thymines is dependent on the presence of magnesium. Modification by DEPC is also greater in the presence of magnesium chloride. Plasmids containing the shorter inserts show no evidence of intramolecular triplex formation in either the presence or absence of magnesium.

DISCUSSION

The results for both chemical and enzymic probes suggest that long stretches of (dA).(dT) can form a structure resembling an intramolecular triple helix and that this is stable at pH 8.0 under superhelical stress. Although two isomers of this structure are possible, with the triplex at either end of the insert, the data demonstrate that it is found exclusively at the 3′-end of the purine strand with the 5′ end left in a single-stranded conformation. The reason for this selectivity is unknown but is similar to that observed for other guanine containing structures [3,6,16]

Although the entire 5′-half of the A-strand of the T_{69} insert is sensitive to digestion by S1 nuclease reaction is greatest at the centre of the insert. A similar though less pronounced effect is seen with DEPC. This suggests that the long loop of the purine strand may not be purely single-stranded but may be interacting, albeit weakly, with the triplex forming a four-stranded structure.

Table I. Examples of poly(dA) tracts in human DNA sequences:

Source	sequence	reference
T-cell receptor α-chain first intron	T_{29}	21
β-tubulin gene containing Alu repeats (AluH)	T_{34}	22
(AluE)	T_{22}	22
Platelet derived growth factor (3′-region)	T_{31}	23
Ribosomal spacer DNA	T_{29} (2 copies)	24
Follistatin precursor gene (intron1)	T_{32}	25
Arginosuccinate synthetase gene (Alu2)	T_{29}	26
3′-end of pro-α2(I) collagen gene (Alu α2R2)	T_{30}	27
Apolipoprotein AII gene (Alu sequence at 3′-end)	A_{26}	28
Insulin receptor	T_{28}	29
Gastrin gene 5′-untranslated	T_{25}	30

The EMBL DNA data base (release 21) was scanned using the programme 'motif' [31] searching for blocks of A or T. Poly A tails from cDNA copies were removed by visual inspection.

Previous studies with polyG containing sequences have suggested that the fourth strand, not directly involved in triplex formation, is closely associated with the triplex, while leaving the bases in an exposed conformation [6]. It is also worth noting that nuclease S1 cleavage of the T-strand is not restricted to the central T residues. Maybe the third strand does not wrap perfectly around the unusual structure adopted by polydA.polydT so that in places this comes out of register and is sensitive to attack by single-strand specific nucleases.

The fraction of plasmid molecules undergoing site specific cleavage by S1 nuclease is much less for the $(T)_{69}$ insert than with $A(GA)_{49}$ under similar conditions. It seems that only a proportion of the polyA containing plasmid molecules are in a triple helical conformation. Indeed experiments using two dimensional gel electrophoresis (not shown) failed to detected any supercoil dependent structural changes, in contrast to results with polyG and poly(GA) containing sequences. This failure could refelect either the dynanic instability of such triplexes or the high energy of formation of such structures placing the transition in a region of the gel where there is no resolution. It appears as if the polyA containing triplexes are in dynamic equilibrium with other (non-triplex) conformations. This could be important for their biological role; if such regions can adopt several structures with similar energies then perturbation by protein binding or other factors will be relatively easy.

It has been suggested that stretches of $d(G)_n$ as short as 15

bases can form intramolecular triplexes [17]; why are such very long $(A)_n$ tracts required to form similar structures? One explanation may be due to the unusual helical conformation of $(dA)_n.(dT)_n$. Even in linear DNA fragments this adopts a structure which is highly propeller twisted, contains additional bifurcated hydrogen bonds [18,19], and will not wrap around nucleosomes [20]. This structure may be so rigid and stable that short tracts do not participate in supercoiling and act independently of the rest of the plasmid. Only when such regions constitute a large proportion of the total DNA will they be forced to reckon with the superhelical stress. It remains to be seen whether other factors can affect the ability of shorter stretches to adopt this configuration.

Biological Relevance

The data presented here demonstrate that very long runs of (dA).(dT) can form intramolecular triplexes in vitro under the influence of superhelical stress and in the presence of divalent metal ions. Are such structures of any biological relevance? We are unable to assess whether such structures occur in vivo, though the process of wrapping DNA around nucleosome also causes a local unwinding of the helix, similar to that caused by superhelical stress. However it is worth noting that poly(dA) tracts are common in eukaryotes. A simple search of human DNA sequences in the EMBL data revealed several such long tracts; Table I lists a few of the longer examples. These are located in introns and are often in the A-rich region located at the 3'-end of Alu repeats within longer stretches of a repetitive nature. Some of these are known to be in sites which are hypersensitive to S1 nuclease [24]. Although none of these would be long enough to generate an intramolecular triplex in vitro the situation in vivo may be different, especially if there are any cellular factors which could stabilise such a structure. Alternatively longer polyA sequences, capable of forming more stable triplexes, may become genetically unstable and be lost.

ACKNOWLEDGEMENTS

This work was supported by grants from the Medical Research Council the Wellcome Trust and the Lister Institute. KRF is a Lister Institute Research Fellow. I am grateful to Dr. Ian Giles for the EMBL library searches.

REFERENCES

1. Wells, R.D., Collier, D.A., Hanvey, J.C. Shimizu, M. & Wohlrab, F. (1988) FASEB J. 2, 2939–2949.
2. Pulleybank, D.E. (1988) in Nucleic Acids and Molecular Biology, vol. 2, ed. F. Eckstein, & D.M.J. Lilley, Springer-Verlag, Berlin, pp.31–48.
3. Hanvey, J.C., Shimizu, M. & Wells, R.D. (1988) Proc. natl. Acad. Sci. U.S.A. 85, 6292–6296.
4. Johnston, B.H. (1988) Science 241, 1800–1804.
5. Mirkin, S.M., Lyamichev, V.I., Drushiyak, K.N., Dobrynin, V.N., Filippov, S.A. & Frank-Kamenetski, M.D. (1987) Nature 330, 495–497.
6. Hanvey, J.C., Klysik, J. & Wells. R.D. (1988) J. Biol. Chem. 263, 7386–7396.
7. Manor, H., Rao, B.S. & Martin, R.G. (1988) J. Molec. Evol. 27, 96–101.
8. Koo, H.S., Wu, H-M., & Crothers, D.M. (1986) Nature 320, 501–506.
9. Katahira, M., Sugeta, H. & Kyogoku, Y. (1990) Nucl. Acids Res. 18, 613–618.
10. Riley, M., Maling, B. & Chamberlain, M.J. (1966) J. Mol. Biol. 20, 359–389.
11. Pilch, D.S., Levenson, C. & Shafer, R.H. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 1942–1946.
12. Bernues, J., Beltran, R., Casasnovas, J.M., & Azorin, F. (1989) EMBO J. 8, 2087–2094.
13. Kohwi, Y. & Kohwi-Shigematsu, T. (1988) Proc. Natl. Acad. Sci. 85, 3781–3785.
14. Kohwi, Y (1989) Nucl. Acids Res. 17, 4493–4502.
15. Panyutin, I.G., Kovalsky, O.I. & Budowsky, E.I. (1989) Nucl. Acids Res. 17, 8257–8271.
16. Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P. & Frank-Kamenetskii, M.D. (1988) Nature 333, 475–476.
17. Lyamichev, V.I., Mirkin, S.M., Baranova, L.V., Vologodskii, A.V. & Frank-Kamenetskii, M.D. (1989) Nucl. Acids Res. 17, 9417–9423
18. Nelson, H.C.M., Finch, J.T., Bonaventura, F.L. & Klug, A. (1987) Nature 330, 221–226.
19. Arnott, S., Chandrasekaran, R., Hall, I.H. & Puigjaner, L.C. (1983) Nucl. Acids Res. 11, 4141–4155.
20. Kunkel, G.R. & Martinson, H.G. (1981) Nucl. Acids Res. 9, 6859–6888.
21. Yoshikai, Y., Clark, S.P., Taylor, S., Sohn, U., Wilson, B.I., Minden, M.D. & Mak, T.W. (1985) Nature 316, 837–840.
22. Lee, M.G.-S., Loomis, C. & Cowan, N.J. (1984) Nucl. Acids Res. 12 5823–5836.
23. Hoppe, J., Schumacher, L. Eicher, W. & Weich, H.A. (1987) FEBS Lett. 223, 243–246.
24. La Volpe, A., Simeone, A., D'Esposito, M., Scotto, L. Fidanza, V., DeFalco, A. & Boncinelli, E, (1985) J. Mol. Biol. 183, 213–223.
25. Shimasaki, S. et al. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 4218–4222.
26. Jinno, Y., Matuo, S., Nomiyama, H., Shimada, K. & Matsuda, I. (1985) J. Biochem. 98, 1395–1403.
27. Myers, J.C., Dickson, L.A., de Wet, W.J., Bernard, M.P., Chu, M.-L., DiLiberto, M., Pepe, G., Sangiorgi, F.O. & Ramirez, F. (1983) J. Biol. Chem. 258, 10128–10135.
28. Knott, T.J., Wallis, S.C., Robertson, M.C., Priestley, L.M., Urdea, M., Rall, L.B. & Scott, J. (1985) Nucl. Acids Res. 13, 6387–6398.
29. Ebina, Y., Ellis, L., Jarnagin, K., Edrey, M., Graf, L., Clauser, E., Ou, J.-H., Masiarz, F., Kan, Y.W., Goldfine, I.D., Roth, R.A. & Rutter, W.J. (1985) Cell 40 747–758.
30. Kariya, Y., Kato, K., Hayashizaki, Y., Himerio, S., Tarui, S., & Matsubara, K. (1986) Gene 50, 345–352.
31. Cockwell, K.Y. & Giles, I.G. (1989) Cabios 5, 227–232.