Extension of DNA triple helix formation to a neighbouring $(AT)_n$ site

Simon P. Chandler, Keith R. Fox*

Department of Physiology & Pharmacology, University of Southampton, Bassett Crescent East, Southampton, SO16 7PX, UK

Received 28 November 1994; revised version received 13 January 1995

Abstract We have used DNase I footprinting to examine the formation of intermolecular triple helices at a fragment containing the target sequence $A_{11}(AT)_6\cdot(AT)_6T_{11}$, using oligonucleotides designed to form parallel T·AT and G·TA triplets. We find that, although $(TG)_6$ does not form a complex with $(AT)_6\cdot(AT)_6$, $T_{11}(TG)_6$ forms a stable structure producing a clear footprint which includes the $(AT)_6$ portion of the target site. This complex is not formed in the presence of magnesium, but can be stabilised by either manganese or a triplex-binding ligand.

Key words: DNA triple helix; G·TA triplet; DNA sequence recognition

1. Introduction

The formation of intermolecular DNA triple helices offer a means for achieving artificial DNA sequence recognition [1,2]. In this strategy a synthetic oligonucleotide binds within the duplex major groove forming specific hydrogen bonds to substituents on the base pairs [3]. Two types of triple helices have been described which differ according to the orientation of the third strand. In the most studied motif the third strand is pyrimidine-rich and runs parallel to the duplex purine strand. Within this motif the best characterised triplets are T·AT and C^+·GC [4]. In the second motif the third strand largely consists of purines and is oriented antiparallel to the duplex purine strand, generating G·GC, A·AT and T·AT triplets [5].

Recognition of mixed sequence DNA is hampered by two major restrictions. Firstly the C^+·GC triplet requires conditions of low pH (<6.0). Secondly recognition is generally limited to the duplex purine strand, so that triplexes can only be formed at homopurine·homopyrimidine sequences. Recent work has demonstrated that, within the parallel motif, it is also possible to form G·TA [6–14] and T·CG [7,8,15] triplets, although these each contain only one hydrogen bond and are thought to be less stable than T·AT and C^+·GC. To date most studies with these alternative combinations have used sequences containing only one G·TA (or T·CG) triplet, although two G·TA triplets were present in a complex formed on the 3′ long terminal repeat of HIV DNA, separated by four canonical triplets [6]. It has been suggested that the stability of G·TA is affected by the nature of the surrounding bases and may be stabilized by adjacent T·AT triplets [11,14]. NMR studies on the G·TA triplet have suggested that, as well as containing a hydrogen bond between the guanine 2-amino group and the O4 of thymine, there is potential for an additional weak interaction with a thymine on an adjacent base pair [11,14]. In addition there appears to be a strong stacking interaction with a third strand thymine on the 5′ side.

In this paper we examine whether it is possible to recognise regions of alternating AT using a third strand consisting of alternating GT, generating a structure containing alternating T·AT and G·TA triplets (Fig. 1a,b). We have used a DNA duplex containing the sequence $A_{11}(AT)_6\cdot(AT)_6T_{11}$ and have compared its interaction with $(TG)_6$ and $T_{11}(TG)_6$. The latter should generate a triplex containing a block of T·AT triplets adjacent to the region of alternating G·TA and T·AT (Fig. 1c). We have also examined the ability of a triplex binding ligand (Fig. 1d) [16] to stabilise these structures.

2. Materials and methods

2.1. Chemicals and enzymes

Oligonucleotides were purchased from Genosys Biotechnology Inc. and used without purification. These were dissolved in water and stored at a concentration of 1 mM at −20°C. DNase I was purchased from Sigma and stored at −20°C at a concentration of 7,200 U/ml. Restriction enzymes and reverse transcriptase were purchased from Promega. The triplex-binding naphthoquinoline derivative, shown in Fig. 1d, was a gift from Dr L. Strekowski, Department of Chemistry, Georgia State University. This was stored as a 20 mM stock solution in dimethylsulphoxide at −20°C, and diluted to working concentrations immediately before use.

2.2. DNA fragment

The preparation of plasmid k2 has been previously described [17]. This consists of a *Sau*3A1 fragment of human DNA, cloned into the *Bam*HI site of pUC19, and contains the sequence $(TA)_{11}T_{34}$. The cloned sequence contains an internal *Hin*dIII site so that digestion with *Hin*dIII and *Bso*fI followed by labelling with [α-^{32}P]dATP using reverse transcriptase yields two fragments of 105 and 161 base pairs, the longer of which contains the sequence $(TA)_{11}T_{34}$ [17,18]. This DNA fragment was separated from the rest of the plasmid on a 6% non-denaturing polyacrylamide gel. This was eluted from the polyacrylamide gel slice and dissolved in 10 mM Tris-HCl, pH 7.5, containing 0.1 mM EDTA.

2.3. DNase I footprinting

The radiolabelled fragment (1.5 μl) containing the target site was mixed with 1.5 μl oligonucleotide and 1.5 μl of 30 μM triplex binding ligand (or 1.5 μl buffer), giving final oligonucleotide concentrations of between 100 and 0.1 μM and, where appropriate, a triplex binding ligand concentration of 10 μM. The oligonucleotides and triplex binding ligand were dissolved in 10 mM Tris-HCl, pH 7.5, containing either 5 mM $MgCl_2$ or 5 mM $MnCl_2$. These complexes were left overnight to equilibrate at room temperature. Digestion was initiated by adding 2 μl DNase I (0.01 U/ml, dissolved in 20 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$) and the reaction was stopped after 1 min adding 4.5 μl of 80% formamide containing 10 mM EDTA. The products of reaction were separated on 10% polyacrylamide gels, containing 8 M urea, and run at 1500 V for 2 h. Gels were fixed in 10% acetic acid before drying at 80°C and subjected to autoradiography at −70°C using an intensifying screen. Bands were assigned by comparison with Maxam-Gilbert sequencing lanes specific for guanine and adenine.

*Corresponding author. Fax: (44) (1703) 594 319.

SSDI 0014-5793(95)00069-0

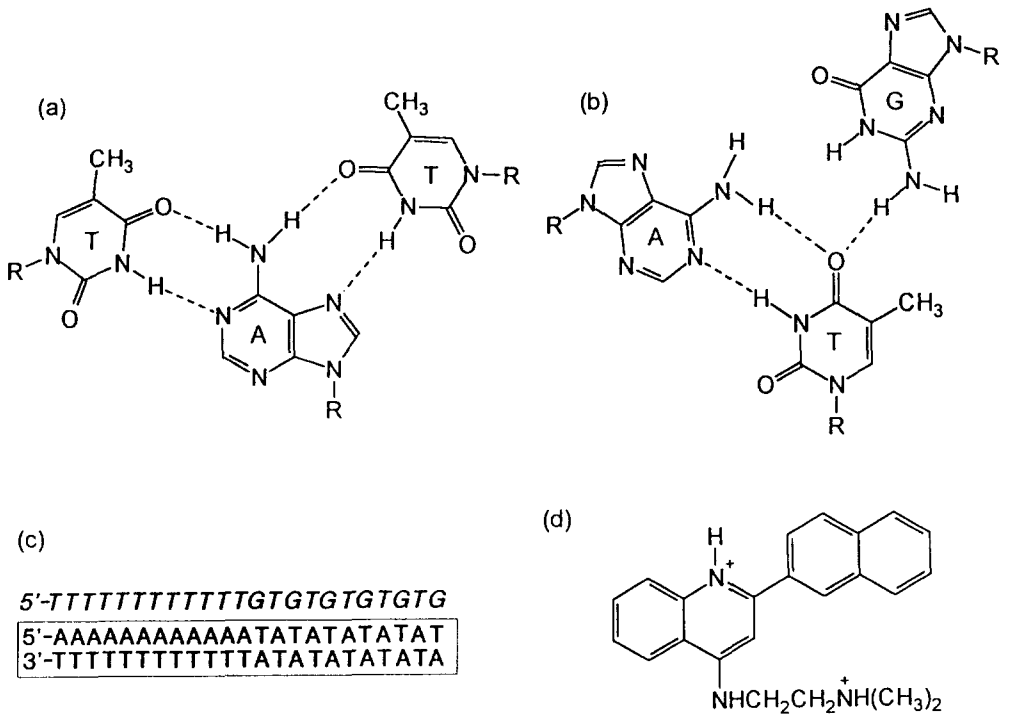

Fig. 1. (a) Structure of the T·AT triplet. (b) Structure of the G·TA triplet. (c) Schematic representation of the triple helix formed between the duplex $A_{11}(AT)_6 \cdot (AT)_6T_{11}$ (boxed) and $T_{11}(TG)_6$. (d) Triple helix-binding ligand **1** [15].

3. Results

Fig. 2 shows the results of DNase I footprinting studies with $(TG)_6$ and $T_{11}(TG)_6$, performed with magnesium as the divalent cation, on a fragment containing the sequence $(TA)_{11}T_{34}$. Looking first at the patterns in the control it can be seen that, as previously noted [17,18], the T_{34} tract is a very poor substrate for the enzyme whereas $(TA)_{11}$ yields clear cleavage bands corresponding to the ApT steps; no cutting is evident for the TpA steps, or at the lowest ApT step. In the presence of $(TG)_6$ the cleavage pattern is hardly affected; all the bands within the $(AT)_n$ tract are still evident (small differences in the cutting pattern are due to variations in the extent of digestion). This suggests that, under these conditions, a block of alternating AT can not be recognised by a third strand consisting of alternating GT.

In an attempt to increase the strength of any weak interaction we repeated the experiment in the presence of a known triplex-binding ligand the structure of which is shown in Fig. 1d [16]. This compound increases the melting temperature of polydA·2polydT [16] and, at a concentration of 10 μM, increases the binding of T_5C_5 to the target sequence $A_6G_6 \cdot C_6T_6$ by at least 100-fold [19]. The results with this compound are also presented in Fig. 2 and reveal that it does not induce the interaction of $(TG)_6$ with its target site. The inability of this triplex-binding ligand to stabilise the interaction of $(TG)_6$ with $(AT)_6 \cdot (AT)_6$ could be due to either the inherent instability of the G·TA-containing triplex, or because it possesses some sequence selectivity and does not bind to structures containing the G·TA triplet.

We have sought further to increase any weak interaction by linking the third strand $(TG)_6$ to a block of thymines, which should be capable of forming a triplex with the adjacent block of $A_n \cdot T_n$, generating the structure shown in Fig. 1c, containing seventeen T·AT and six G·TA triplets. The interaction of $T_{11}(TG)_6$ with the target sequence $A_{11}(AT)_6 \cdot (AT)_6T_{11}$ is shown in the right-hand panel of Fig. 2. Once again we find no evidence for any interaction, suggesting that the six G·TA triplets are sufficient to destabilise the complex with the seventeen T·AT triplets. However, in the presence of 10 μM triplex-binding ligand, a clear DNase I footprint can be seen at the lower end of the $(TA)_n$ tract; the lowest five bands are either attenuated or absent. Since the last ApT step is not cut in the control this reduction in cleavage exactly matches the putative target site. These changes are evident at oligonucleotide concentrations as low as 0.1 μM. Since DNase I cleavage of the T_n tract is extremely poor it is not possible to assess the interaction of the T_{11} portion of the third strand with the $A_n \cdot T_n$ portion of the target.

Although it is clear that the proposed triple helix can form under these conditions, it might be argued that the $(TG)_6$ tail does not contribute to the binding, and that the observed footprint results from non-specific interaction of this tail with DNase I. This seems unlikely since, in the absence of the $(TG)_6$ tail, the T_{11} portion could bind at many different locations within the $(TA)_{11}T_{34}$ target site, and would not be restricted to the region around the $(AT)_{11}$ tract. The specific requirement for the formation of G·TA triplets was checked using $T_{11}(TA)_6$ as the third strand oligonucleotide, replacing the G·TA with A·AT triplets, which are not stable within this parallel triplex. This oligonucleotide did not affect the DNase I cleavage pattern (data not shown), even in the presence of the triplex binding ligand.

Since several studies have suggested that the formation of

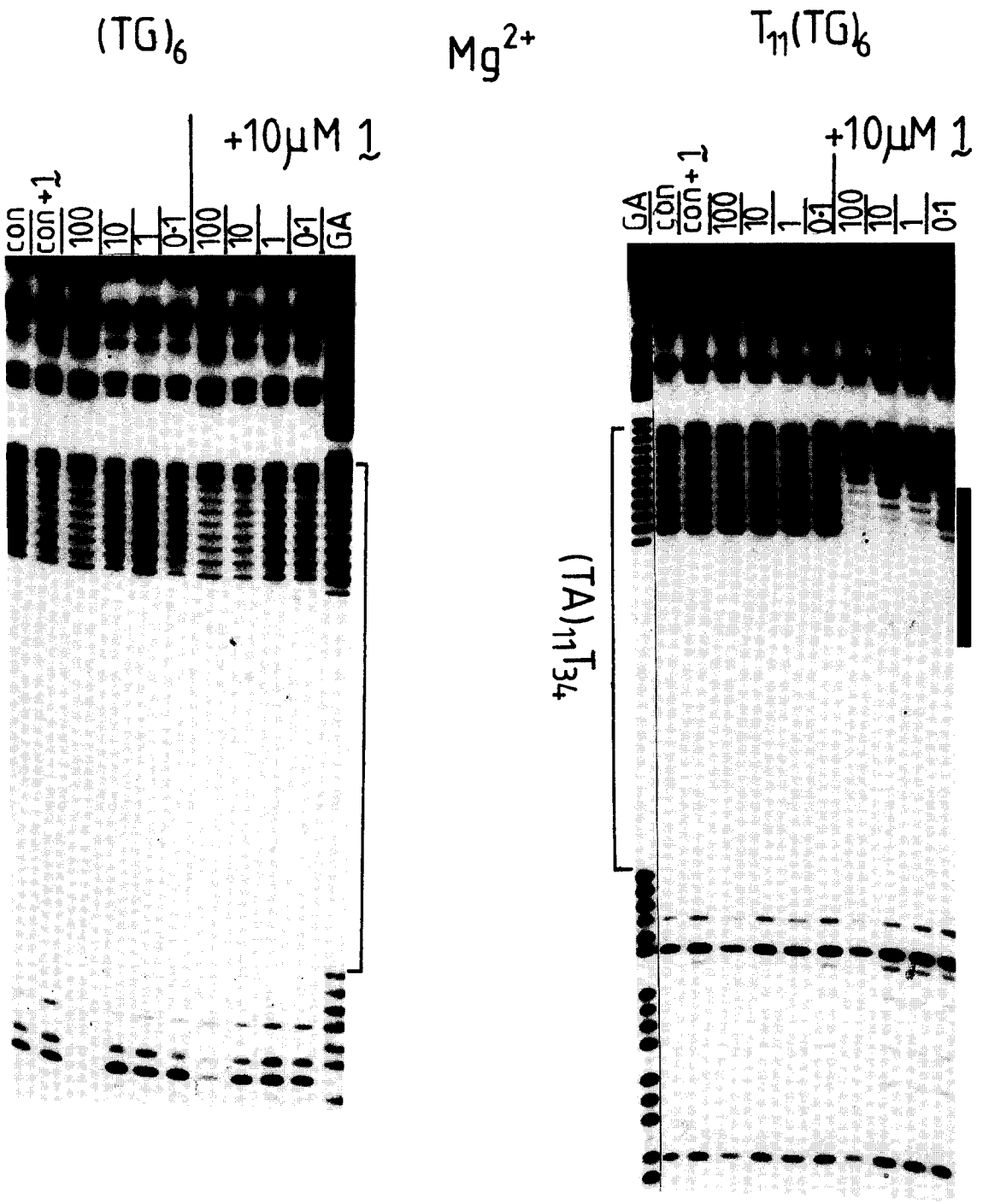

Fig. 2. DNase I digestion of the fragment containing the sequence $(TA)_{11}T_{34}$ in the absence (con) and presence of $(TG)_6$ or $T_{11}(TG)_6$. The reactions were performed in buffer containing 5 mM $MgCl_2$. The right-hand side of each panel was performed in the presence of 10 μM **1**. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The lanes labelled 'con + 1' correspond to cleavage in the presence of 10 μM **1** alone. Tracks labelled 'GA' are Maxam-Gilbert formic acid-piperidine markers specific for purines. The square brackets show the position of the sequence $(TA)_{11}T_{34}$, the filled box shows the position of the $(AT)_6T_{11}$ target site.

DNA triple helices can be affected by the nature of the divalent cation [20,21], we have repeated these experiments in the presence of manganese. The results are presented in Fig. 3. Once again $(TG)_6$ shows no interaction with the target site, both in the presence and absence of the triplex-binding ligand. However, $T_{11}(TG)_6$ generates a clear footprint, even in the absence of the ligand, which persists at concentrations as low as 1 μM. In the absence of the ligand this footprint extends over the first six cleavage products, slightly longer than that in the presence of the ligand. It is also worth noting that at low oligonucleotide concentrations, in the absence of the ligand, each of the bands within the footprint reappears at the same time. In contrast, in the presence of the ligand, bands towards the upper edge of the footprint reappear at the lower oligonucleotide concentrations,

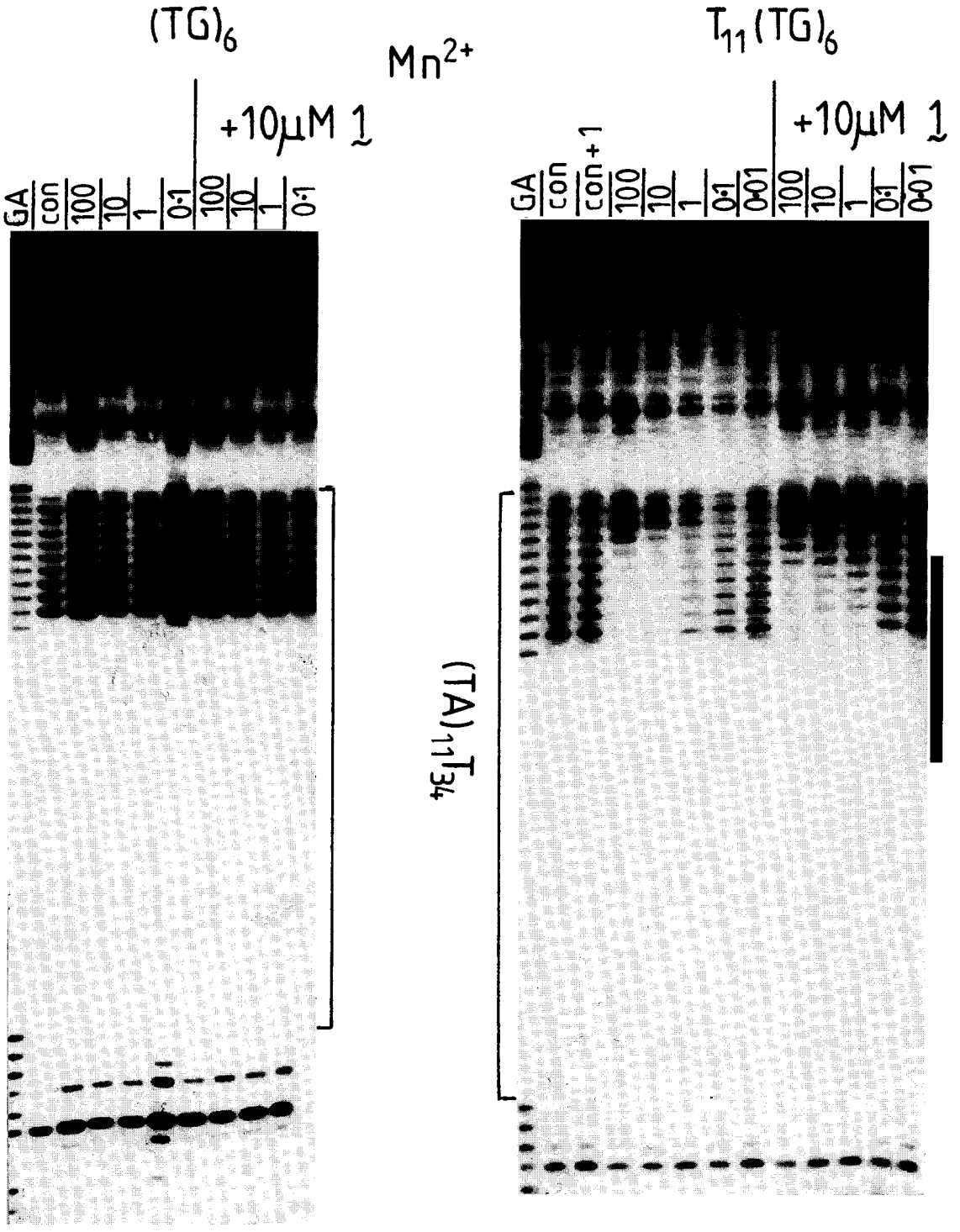

Fig. 3. DNase I digestion of the fragment containing the sequence $(TA)_{11}T_{34}$ in the absence (con) and presence of $(TG)_6$ or $T_{11}(TG)_6$. The reactions were performed in buffer containing 5 mM $MnCl_2$. All other details are as for Fig. 2.

while bands in the lower part of the footprint are still protected. This may indicate some fraying of the 3′ end of the oligonucleotide at low concentrations, suggesting that the ligand has a preferential stabilizing effect on the block of T·AT triplets. Similar experiments with $T_{11}(TA)_6$ did not produce a DNase I footprint (not shown), indicating the requirement for G·TA triplets.

4. Discussion

The results presented in this paper reveal that $(TG)_n$ alone can not be used for generating intermolecular triplexes at regions of $(AT)_n \cdot (AT)_n$. However, when this third strand is tethered to a block of Ts, which form a parallel triple helix with an adjacent block of $A_n \cdot T_n$, it forms a specific interaction. In the

presence of manganese or a triplex binding ligand the adjacent block of T·AT triplets facilitates binding of $(TG)_6$ to its $(AT)_6$ target site. In the complex with $T_{11}(TG)_6$, the position and length of the footprint can not be explained by suggesting that only the T_{11} end of the oligonucleotide interacts with the $A_{34}\cdot T_{34}$ tract, leaving the $(TG)_6$ portion dangling free in solution, but provides clear evidence for the formation of a structure containing G·TA triplets. This is confirmed by the lack of any interaction with $T_{11}(TA_6$, even under the most favourable conditions.

The inability of the triplex-binding ligand to stabilise the interaction with $(TG)_6$ alone suggests that it does not bind to G·TA triplets. This is further evidenced by the slightly shorter footprint observed with $T_{11}(TG)_6$ in the presence of the ligand, suggesting that the compound only stabilises the T·AT triplet, and may even destabilise the G·TA triplets. It is worth noting that $(AT)_n$ tracts may present the most favourable conditions for formation of the G·TA triplet, since it had been suggested that the third strand guanine can form an additional hydrogen bond with the adjacent AT base pair [11]. This might also explain why the triplex binding ligand does not facilitate the interaction with $(TG)_6$ alone, since it would disrupt this weak hydrogen bond.

These results extend the types of sequences which may be targeted by triplex-forming oligonucleotides and demonstrates that specific interactions may be achieved by employing several weaker triplets, provided that these are attached to other, more stable, triplex forming regions. Further experiments are in progress to examine the formation of triplets containing several T·CG triplets. However, oligonucleotides generating these weaker triplets need to be designed with care, since they may target other sequence using other triplet motifs. For example we have previously used $T_{11}(TG)_6$ to form an alternate strand triple helix at the target site $A_{11}(TC)_6\cdot(GA)_6T_{11}$ [22].

Acknowledgements: This work was supported by grants from the Cancer Research Campaign and the Medical Research Council.

References

[1] Chubb, J.M. and Hogan, M.E. (1992) Trends Biotechnol. 10, 132–136.
[2] Moffat, A.S. (1991) Science 252, 1374–1375.
[3] Thuong, N.T. and Hélène, C. (1993) Angewandte Chemie 32, 666–690.
[4] Moser, H.E. and Dervan, P.B. (1987) Science 238, 645–650.
[5] Beal, P.A. and Dervan, P.B. (1991) Science 251, 1360–1363.
[6] Griffin, L.C. and Dervan, P.B. (1989) Science 245, 967–971.
[7] Yoon, K., Hobbs, C.A., Koch, J., Sardaro, M., Kutny, R. and Weis, A. (1992) Proc. Natl. Acad. Sci. USA 89, 3840–3844.
[8] Chandler, S.P. and Fox, K.R. (1993) FEBS Lett. 332, 189–192.
[9] Radhakrishnan I., Gao, X., De los Santos, C., Live, D. and Patel, D.J. (1991) Biochemistry 30, 9022–9030.
[10] Radhakrishnan, I. Patel, D.J. and Gao, X. (1992) Biochemistry 31, 2514–1523.
[11] Radhakrishnan, I., Patel, D.J. Veal, J.M. and Gao, X. (1992) J. Am. Chem. Soc. 114, 6913–6915.
[12] Radhakrishnan, I. and Patel, D.J. (1994) Structure 2, 17–32.
[13] Wang, E., Malek, S. and Feigon, J. (1992) Biochemistry 31, 4838–4846.
[14] Kiessling, L.L., Griffin, L.C. and Dervan, P.B. (1992) Biochemistry 31, 2829–2834.
[15] Radhakrishnan, I. and Patel, D.J. (1994) J. Mol. Biol. 241, 600–619.
[16] Wilson, W.D., Tanious, F.A., Mizan, S., Yao, S., Kiselyov, A.S., Zon, G. and Strekowski, L. (1993) Biochemistry 32, 10614–10621.
[17] Fox, K.R. (1992) Nucleic Acids Res. 20, 1235–1242.
[18] Fox, K.R. and Cons, B.M.G. (1993) Biochemistry 32, 7162–7171.
[19] Cassidy, S.A., Strekowski, L., Wilson, W.D. and Fox, K.R. (1994) Biochemistry, in press.
[20] Malkov, V.A., Voloshin, O.N., Soyfer, V.N. and Frank-Kamenetskii, M.D. (1993) Nucleic Acids Res. 21, 585–591.
[21] Washbrook, E. and Fox, K.R. (1994) Nucleic Acids Res. 22, 3977–3982.
[22] Washbrook, E. and Fox, K.R. (1994) Biochem. J. 301, 569–575.