Secondary Binding Sites for Triplex-Forming Oligonucleotides Containing Bulges, Loops, and Mismatches in the Third Strand†

Keith R. Fox,* Emily Flashman, and Darren Gowers‡

Division of Biochemistry and Molecular Biology, School of Biological Sciences, Southampton University, Bassett Crescent East, Southampton SO16 7PX, U.K.

Received December 3, 1999; Revised Manuscript Received March 23, 2000

ABSTRACT: We have used DNase I footprinting to examine the binding of five different 17-mer oligonucleotides to a 53-base oligopurine tract containing four pyrimidine interruptions. Although all the expected triplexes formed with high affinity ($K_d \approx 10-50$ nM), one oligonucleotide produced a footprint at a second site with about 20-fold lower affinity. We have explored the nature of this secondary binding site and suggest that it arises when each end of the third strand forms a 7-mer triplex with adjacent regions on the duplex, generating a contiguous 14-base triplex with a bulge in the center of the third strand oligonucleotide. This unusual binding mode was examined by use of oligonucleotides that were designed with the potential to form different length third-strand loops of various base composition. We find that triplexes containing single-base bulges are generally more stable than those with dinucleotide loops, though triplexes can be formed with loops of up to nine thymines, generating complexes with submicromolar dissociation constants. These structures are much more stable than those formed by adding two separate 7-mer oligonucleotides, which do not generate DNase I footprints, though a stable complex is generated when the two halves are covalently joined by a hexa(ethylene glycol) linker. MPE produces less clear footprints, presumably because this cleavage agent binds to triplex DNA, but confirms that the oligonucleotides can bind in unexpected places. These results suggest that extra care needs to be taken when designing long triplex-forming oligonucleotides so as to avoid triplex formation at shorter secondary sites.

The formation of intermolecular DNA triple helices offers the possibility of designing compounds with extensive sequence recognition properties, which be used as antigene agents or experimental tools (*1*−*4*). These structures are formed by the sequence-specific binding of synthetic oligonucleotides in the major groove of duplex DNA, making specific hydrogen contacts to substituents on the exposed faces of the duplex purine bases (*5*−*7*). Two types of triplex motif have been described, in which the third strand binds in a parallel or antiparallel direction in relation to the duplex purine strand. The parallel form is typified by pyrimidine-rich third strands and formation of T·AT and C^+·GC triplets (*8*−*10*), while the antiparallel form uses G·GC, A·AT, or T·AT triplets (*11*−*13*).

An outstanding problem in the use of triplexes for achieving sequence-specific recognition of DNA is that these complexes are usually restricted to homopurine tracts, since there is no simple means for recognizing pyrimidine bases (*14*, *15*). Within the parallel motif TA base-pair inversions can be successfully targeted by forming G·TA triplets (*14*−*19*). Although this triplet is specific, it is less stable than T·AT or C^+·GC, since the third strand base only makes one hydrogen bond contact with the duplex pyrimidine that is accompanied by local distortions in backbone geometry. The G·TA triplet is more stable when flanked by T·AT than C^+·GC (*17*), an effect that arises from the presence of an additional hydrogen bond to thymine in the adjacent T·AT triplet (*18*).

To be selective for a unique DNA sequence within the human genome of 3×10^9 base pairs, a DNA binding agent needs to recognize at least 16−17 consecutive bases. Triple helix formation is able to meet this requirement. Long triplexes also form more stable complexes than shorter ones, though some high-affinity structures have been demonstrated with oligonucleotides as short as nine bases (*20*, *21*). This strategy also achieves stringent binding, and single-base mismatches destabilize complexes in much the same way as mismatches destabilize DNA duplexes (*22*). However, although long (>17-base) oligonucleotides should bind with greater affinity than shorter ones, the increase in length may also increase the affinity for secondary sites, generating mismatches, loops, or other structures. Several strategies have been suggested for increasing triplex affinity, such as the use of novel base analogues (*7*), triplex-binding ligands (*23*−*26*), or tethered DNA binding ligands (*27*−*29*). However, as triplex affinity is increased, we would expect the binding to related secondary sites to increase as well. Although several studies have explored the stringency of triplex formation, these have examined how changes to the third strand affect the affinity for the duplex target. There

† This work was supported by grants from the Cancer Research Campaign.

* Corresponding author: tel +44 2380 594374; fax +44 2380 594459; e-mail K. R.Fox@soton.ac.uk.

‡ Present address: Department of Biochemistry, University of Bristol, School of Medical Sciences, University Walk, Bristol BS8 1TD, U.K.

10.1021/bi992773+ CCC: $19.00
Published on Web 05/12/2000

Table 1: Sequence of Triplex-Forming Oligonucleotides Used in This Work and Their Abbreviations[a]

abbreviation	sequence	abbreviation	sequence
oligo 1	TTCTTTTCGTTCTTTCT	+A	TTCTTTCATTTTCTT
oligo 2	TTCTTTCTTTTCTTCTT	+C	TTCTTTCCTTTTCTT
oligo 3	TTCTTCTTGCTTTTCTT	+G	TTCTTTCGTTTTCTT
oligo 4	CTTTTCTTGTTTCTTTC	HEX	TTCTTTCHTTTTCTT
oligo 5	TTTCTTTCGCTTTTCTT	+T	TTCTTTCTTTTTCTT
oligo 6	TTCTTTCTTTTCTT	2T	TTCTTTCTTTTTTCTT
short 1	TTTTCTT	3T	TTCTTTCTTTTTTTCTT
short 2	TTCTTTC	5T	TTCTTTCTTTTTTTTTTCTT
+CG	TTCTTTCCGTTTTCTT	9T	TTCTTTCTTTTTTTTTTTTTCTT
+AT	TTCTTTCATTTTTCTT		
+AA	TTCTTTCAATTTTCTT		
+TA	TTTCTTTCTATTTTCTT		

[a] For oligonucleotides that might bind by loop formation, the looped bases are underlined. Hex (H) = hexa(ethylene glycol).

have been fewer studies examining the binding of a given third-strand oligonucleotide to a range of related targets (*30*−*33*). Combinatorial methods, such as REPSA (restriction enzyme protection selection and amplification), give valuable information on the best binding sites but may not reveal the presence of weaker, secondary binding sites (*31*).

The formation of intermolecular DNA triplexes is usually studied on simple linear target sites, though there are several examples of triplexes containing other structures. A number of studies have shown that DNA triplex formation can tolerate the presence of bulges in either the third strand or duplex (*30*, *34*). These studies have shown that single adenine bulges in the third strand are more destabilizing than in the Watson−Crick pyrimidine strand (*34*) and that the relative stability follows the trend perfect triplex > single bulge > single mismatch > double bulge > double mismatch (*30*). In another study a triplex binding ligand has been shown to facilitate the binding of a 28-mer oligonucleotide to a 27-mer target site in a structure that generates a single looped thymine (*35*). Two triplex-forming oligonucleotides have also been induced to bind by incorporating a dimerization domain within the third strands, which formed a looped Watson−Crick duplex within the triplex structure (*36*, *37*), and which could also be stabilized by a sequence-selective ligand (*38*). Triplex formation at noncontiguous sites by third-strand oligonucleotides that are tethered by a flexible linker can also be used to induce DNA bends (*39*, *40*). These studies demonstrate that triplex-forming oligonucleotides can form a variety of related structures, which might be adopted by long third strands and which would compete for binding to the intended target sites.

In this paper we examine the interaction of several 17-mer oligonucleotides with different portions of a 53-base oligopurine tract, which is interrupted by TA base pairs at four different locations. We find that one of these triplex-forming oligonucleotides shows significant binding to a secondary site by extruding two bases from the third strand, generating a shorter contiguous triplex. These studies, which explore the sequence dependence of this interaction, demonstrate that a nine-base loop can be tolerated at the center of a 14-mer triplex.

MATERIALS AND METHODS

Chemicals and Enzymes. Oligodeoxynucleotides were purchased from Oswel DNA Service. The sequences of the third-strand oligonucleotides used in this work together with their short names are shown in Table 1. *Bam*HI-cut alkaline phosphatase-treated pUC18, DNA ligase, and radiochemicals were from Amersham Pharmacia Biotech Ltd. Bovine DNase I was purchased from Sigma and stored frozen at 7200 units/mL. Methidiumpropyl−EDTA was purchased from Sigma and stored frozen at a concentration of 100 μM in water. Restriction enzymes and reverse transcriptase were purchased from Promega.

DNA Sequences. Complementary oligonucleotides generating the long oligopurine tract shown in Figure 1A were treated with polynucleotide kinase and annealed prior to ligation into *Bam*HI-cut pUC18. Following transformation of calcium-permeabilized *Escherichia coli* TG2, successful clones were picked from agar plates containing ampicillin IPTG and X-gal as white colonies. Clones were sequenced by use of a T7 dideoxy sequencing kit (Amersham Pharmacia Biotech Ltd). Two clones (DMG60R and DMG60Y) were obtained in which the insert was oriented in opposite orientations. Labeling the 3′-end of the *Hin*dIII site visualized the purine-rich strand of DMG60R and the pyrimidine-rich strand of DMG60Y.

DNA Fragments. Plasmids DMG60R or DMG60Y, containing the cloned insert in opposite orientations, were digested with *Hin*dIII and *Sac*I before labeling at the 3′-end of the *Hin*dIII site with [α-^{32}P]dATP and reverse transcriptase. The fragments were separated from the remainder of the plasmid on 6% (w/v) nondenaturing polyacrylamide gels. The labeled DNA was eluted from the gel and dissolved in 10 mM Tris-HCl, pH 7.5, containing 0.1 mM EDTA at a concentration of 10−20 cps/μL as measured on a hand-held Geiger counter (approximately 10 nM).

DNase I Footprinting. Radiolabeled DNA (1.5 μL) was mixed with oligonucleotides (3 μL) diluted in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$ to give final third strand concentrations between 10 μM and 10 nM. The complexes were left to equilibrate overnight at 20 °C. Digestion was started by adding 2 μL of DNase I (approximately 0.01 unit/mL), dissolved in 20 mM NaCl, 2 mM $MnCl_2$, and 2 mM $MgCl_2$ and terminated after 1 min by adding 4 μL of 80% formamide containing 10 mM EDTA, 1 mM NaOH, and 0.1% (w/v) bromophenol blue.

Footprinting with Methidiumpropyl−EDTA−Fe(II). Radiolabeled DNA (2 μL) was mixed with oligonucleotides (4 μL) diluted in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$ to give final third-strand concentrations of 2 μM. The complexes were left to equilibrate overnight at 20

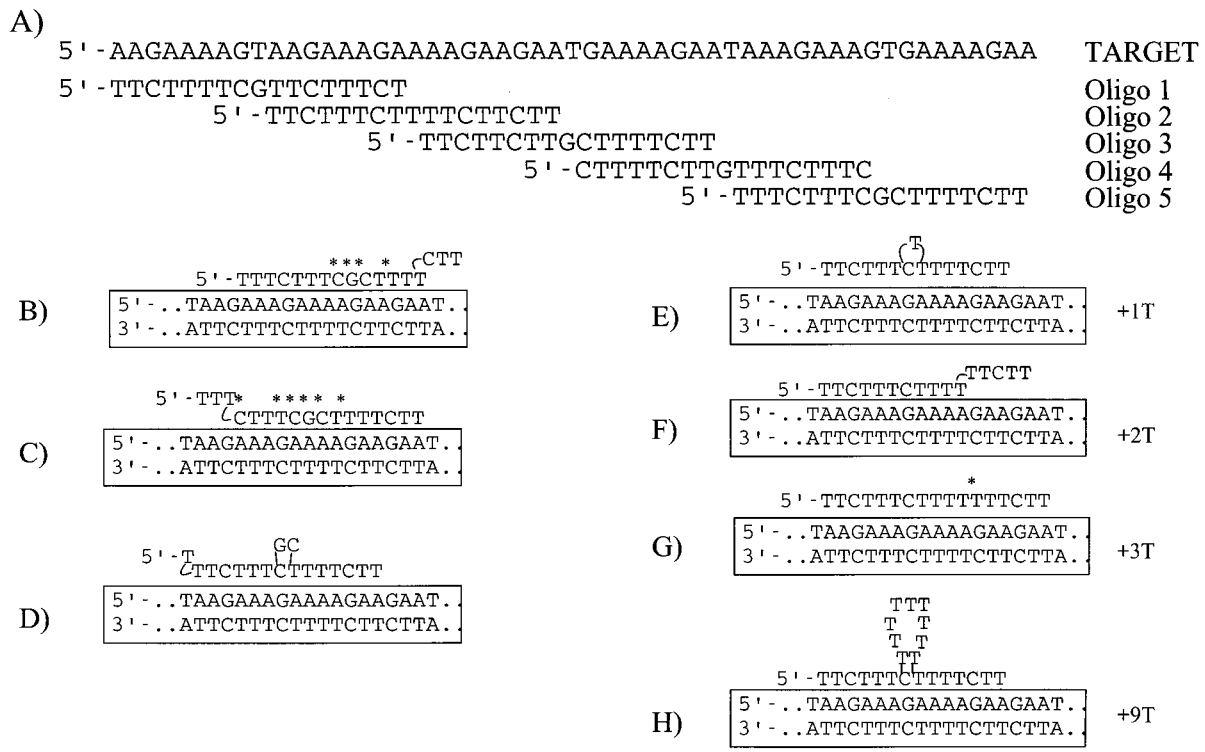

FIGURE 1: (A) Sequence of the purine-rich strand of the insert in plasmid DMG60R, together with the five 17-mer oligonucleotides designed to bind at different locations. The sequence was cloned into the *Bam*HI site of pUC19. (B−D) Possible schemes for the interaction of oligo 5 with its secondary binding site: (B) three bases fraying at the 3′-end; (C) three bases fraying at the 5′-end; (D) with a two-base loop in the center. (E−H) Possible schemes for the interaction of oligonucleotides +T, +2T, +3T, and +9T with their binding sites. The asterisks correspond to mismatched triplets.

°C. To this was added 5 μL of a solution containing 20 μM MPE and 20 μM ferrous ammonium sulfate, and the mixture was equilibrated for a further 5 min. The cleavage reaction was started by adding 3 μL of 10 mM dithiothreitol and stopped after 30 min by ethanol precipitation. Samples were redissolved in 8 μL of 80% formamide, containing 10 mM EDTA, 1 mM NaOH, and 0.1% (w/v) bromophenol blue.

Electrophoresis. Products of digestion were separated on 10% (w/v) denaturing polyacrylamide gels (National Diagnostics) containing 8 M urea. Electrophoresis conditions were 1500 V for about 2 h. Gels were fixed in 10% (v/v) acetic acid before being dried at 80 °C for 1 h and exposed overnight either to autoradiography film at −70 °C with an intensifying screen or to a storage-phosphor screen. Bands in each digestion pattern were assigned by comparison with Maxam−Gilbert markers specific for purines.

Quantitative Analysis. Gels were exposed to a storage-phosphor screen and analyzed with a Molecular Dynamics Storm 860 phosphorimager. The intensity of bands in the footprint was measured by use of ImageQuant software. In each case the values were normalized with respect to bands outside the footprint to correct for differences in gel loading and DNase I digestion. Footprinting plots (*41*), describing band intensity as a function of oligonucleotide concentration, were analyzed with FigP for Windows (Biosoft) according to the equation $I/I_0 = C_{50}/(L + C_{50})$, where I and I_0 are the band intensities in the presence and absence of the oligonucleotide, respectively, L is the oligonucleotide concentration, and C_{50} is the oligonucleotide concentration that reduces the intensity of bands in the footprint by 50%. The use of this equation to analyze footprinting data assumes that the DNA concentration is low (lower than the dissociation constant of the ligand). Under these conditions, for which the absolute DNA concentration need not be known, C_{50} is equal to the thermodynamic dissociation constant.

MPE footprinting data are presented as differential cleavage plots in the form f_t/f_c, where f_t is the fractional cleavage of a bond in the presence of oligonucleotide and f_c is the fractional cleavage of the same bond in the control (*35*). In these representations, differential cleavage values of less than 1 correspond to regions that are protected from cleavage.

RESULTS

As part of a continuing series of studies on the stability of DNA triplexes containing G·TA triplets, we prepared a fragment containing a long polypurine tract that is interrupted at four position by thymines (Figure 1A). Each of these thymines is located in a different sequence context, i.e., GTG, ATG, GTA, and ATA, which are arranged so that four overlapping 17-mer triplexes can be generated with oligos 1−5, as shown in Figure 1A. Each of these triplexes (except the one formed with oligo 2, which binds to a 17-mer homopurine tract) contains four C^+·GC and 12 T·AT triplets with one G·TA triplet positioned in the center. The results of DNase I footprinting experiments with these triplex-forming oligonucleotides are shown in Figure 2. It can be seen that each oligonucleotide produces a clear footprint at its predicted target site. These footprints persist to oligonucleotide concentrations of between 10 and 50 nM, confirming the possibility of forming stable triplexes at target sites containing a single pyrimidine interruption. However it can also be seen that oligo 5, which is designed to generate a complex containing a G·TA triplet flanked by C^+·GC on either side, generates an additional footprint at higher

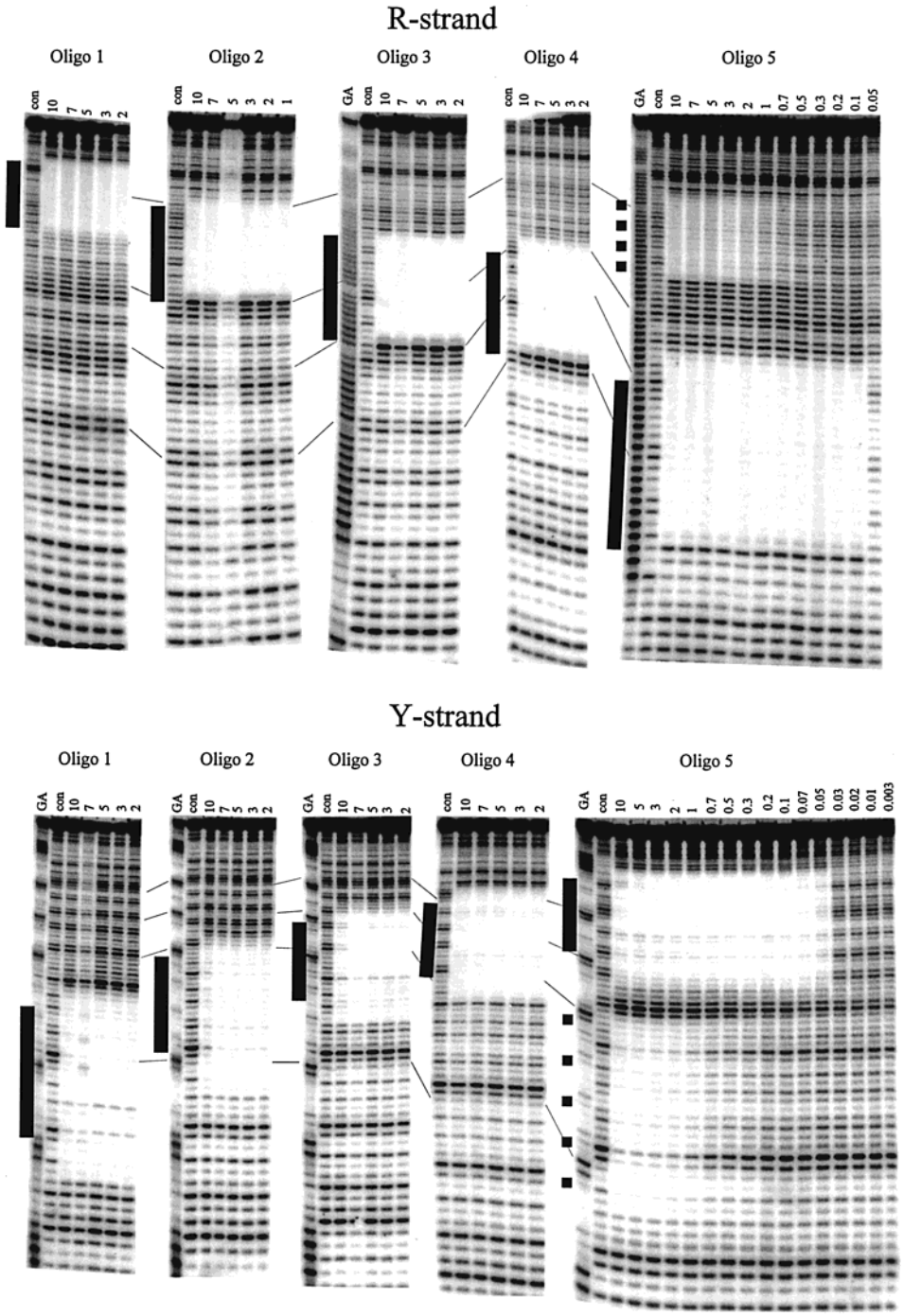

FIGURE 2: DNase I footprints showing the interaction of oligos 1−5 with their target sites in fragment DMG60R (top) and DMG60Y (bottom). The fragments were labeled at the 3′-end of the *Hin*dIII site visualizing the purine-containing strand for DMG60R and the pyrimidine strand for DMG60Y. Reactions were performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Oligonucleotide concentrations (micromolar) are shown at the top of each lane. Lanes labeled GA are Maxam−Gilbert markers specific for purines. Lanes labeled con show digestion in the absence of added oligonucleotide. The thin lines join equivalent positions on the different panels. The solid bars show the expected target sites for each of the oligonucleotides. The dashed bars shows the position of the secondary footprint seen with oligo 5.

oligonucleotide concentrations. This footprint, which is indicated by the dashed bars in Figure 2, is evident on both labeled strands and is in a similar location to that generated by oligo 2. Quantitative analysis of the footprints produced by oligo 5 gave C_{50} values of 51 ± 16 nM and 1.29 ± 0.45 μM at the strong and weak sites, respectively, on the labeled purine strand and 29 ± 10 nM and 0.43 ± 0.12 μM on the labeled pyrimidine strand (Table 2). We attempted to determine the exact location of this secondary binding site by comparing the position and length of its footprint with the other five intended target sites. On the labeled purine strand (Figure 2A), the 3′- (lower) edge of the footprints for oligos 1−5 is located at the triplex−duplex junction, where there is enhanced cleavage at the position corresponding to the terminal purine in the target site (except with oligo 1). This enhanced cleavage is frequently observed on the purine strand at the triplex−duplex junction and is assumed to represent a structural distortion at this position that renders the DNA more susceptible to DNase I cleavage. The precise origin of this effect is not clear and there is no simple

Table 2: C_{50} Values for the Footprints Produced by Each of the Oligonucleotides[a]

oligonucleotide	R-strand	Y-strand
oligo 5	0.051 ± 0.016	0.029 ± 0.010
(secondary site)	1.29 ± 0.45	0.43 ± 0.12
oligo 6	<0.1	<0.1
+C	<0.1	<0.1
+T	<0.1	<0.1
+A	0.064 ± 0.036	0.046 ± 0.018
+G	0.15 ± 0.07	0.062 ± 0.009
+HEX	0.22 ± 0.12	0.18 ± 0.05
+CG	0.33 ± 0.11	0.36 ± 0.03
+TA	0.28 ± 0.09	0.16 ± 0.02
(secondary site)	1.29 ± 0.45	0.42 ± 0.05
+AT	0.23 ± 0.06	0.52 ± 0.05
+AA	no binding	no binding
+2T	0.23 ± 0.14	0.20 ± 0.11
+3T	0.13 ± 0.07	0.09 ± 0.05
+5T	0.29 ± 0.10	0.22 ± 0.07
+9T	0.27 ± 0.09	0.34 ± 0.16

[a] For oligo 5 and +TA the C_{50} values (micromolar) for each of the two footprints produced by these oligonucleotides is shown. For some of the tightest binding oligonucleotides, reliable C_{50} values could not be determined from the footprinting plots. In these cases the value is shown as <0.1 μM.

explanation for its absence with oligo 1. On the labeled purine strand (Figure 2A) these footprints continue for two bases beyond the upper (5′) edge of the target sites, and each footprint covers a total of about 18 bases. Upon looking at the labeled pyrimidine strand (Figure 2B), it can be seen that the footprints continue for three to four bases beyond both the 3′- (lower) and 5′- (upper) edges of the target site. We can now use this information to determine the location of the secondary footprints, since we would expect the 3′- (lower) edge of the footprint on the labeled purine strand to correspond to the triplex−duplex junction, while the upper edge should proceed for two bases beyond the binding site. Inspection of Figure 2 shows that the lower (3′-) edge of this secondary footprint is three bases higher than that seen with oligo 2, suggesting that the secondary binding site for oligo 5 ends three bases to the 5′-side of the target for oligo 2. The upper edge of this secondary footprint is less distinct but appears to be in the same location as that produced by oligo 2, so that 15−16 bases are protected from cleavage. On the basis of these observations, we estimate that the 17-mer oligonucleotide interacts with 14−15 base pairs and propose that the secondary site for oligo 5 covers the sequence 5′-AAGAAAGAAAAGAA. Upon looking at the labeled pyrimidine strand (Figure 2B), the 3′- (lower) edge of the footprint is in the same position as that with oligo 2, while the upper end is about three bases lower. Again it is clear that the 5′-end of the oligonucleotide is located in the same position as oligo 2. Although it is harder to precisely locate the binding site from the footprints on the labeled pyrimidine strand, these data are consistent with 5′-AAGAAAGAAAAGAA as the proposed secondary binding site.

How can a 17-mer oligonucleotide occupy a binding site of only 14 bases? We first considered the possibility that three bases were fraying from one or the other end of the third strand. If three bases at the 3′-end of oligo 5 fray from the target, then the resulting complex will cover 14 base pairs but involves the formation of four unusual (mismatched) triplets as shown in Figure 1B. Allowing three bases to fray at the 5′-end of the oligonucleotide produces an even less likely complex that contains 6 triplet mismatches (Figure 1C). A more likely possibility (which is consistent with other results described below) is shown is Figure 1D. In this complex the 5′-terminal T and two bases at the center of the third strand are not involved in triplet formation. This structure produces 14 contiguous T·AT and C^+·GC triplets on the proposed target site, though it contains an unusual short loop in the center of the third-strand oligonucleotide. We have tested this proposal by determining the binding sites for several oligonucleotides that have been designed to place loops in this position.

Oligonucleotides with One Bulged Base. Figure 3 shows the results of experiments with 15-mer oligonucleotides that have been designed to generate 14-mer triplexes with a single-base bulge at the center of the third strand. The proposed structure for the complex with one additional T is shown in Figure 1E. We tested oligonucleotides containing each base (A, G, T, and C) in turn in the bulge. Each of these 15-mer oligonucleotides produces a clear footprint at the secondary target site, which persist to oligonucleotide concentrations below 0.3 μM. These four oligonucleotides produce very similar footprinting patterns at this position, which are identical to that produced by oligo 5. Examination of the concentration dependence of these footprints shows that +C binds the tightest (with a footprint still evident at 30 nM oligonucleotide); +T binds slightly less well (footprinting to 0.1 μM), while +A and +G require oligonucleotide concentrations of 0.3 μM and above. C_{50} values determined for these interactions are presented in Table 2. All these oligonucleotides bind more strongly to this region than oligo 5. Upon comparing these footprints with those in Figure 2, it can be seen that, as expected, the original footprints at the intended target site for oligo 5 are no longer evident with oligos +A, +T, and +G. However, a weak footprint at this site can be seen with the highest concentrations of +C. This appears to cover the entire region originally occupied by the 17-mer oligo 5 (Figure 8A). The simplest explanation that we can offer for this footprint is that it consists of the sum of two 7-mer triplexes, corresponding to interaction of the third strand with each end of the target site, as shown in Figure 8B. Although the upper of these two structures can also be formed with oligonucleotides +A, +G, and +T, the lower will not be possible as the central C is changed to A, G, or T.

The suggestion that oligonucleotide +C might bind to each half of the original target site raises the possibility that all four oligonucleotides (+A, +G, +T, and +C) bind to the secondary site by the simultaneous interaction of two third-strand molecules with each half of the target site, rather than requiring the proposed bulge formation. This seems unlikely since these footprints persist to much lower oligonucleotide concentrations than the weak interaction of +C with the original target site for oligo 5. However, we tested this possibility by examining the binding of each half of the oligonucleotide in turn, using the 7-mer oligonucleotides short 1 and short 2, which correspond to the 5′- and 3′-ends of the 15-mers. The results are presented in Figure 3 and show that neither of the short oligonucleotides nor the combination of both affects DNase I cleavage in the target region. We can therefore discount simultaneous half-site recognition as an explanation for the secondary footprints. We also examined the binding of an oligonucleotide in which

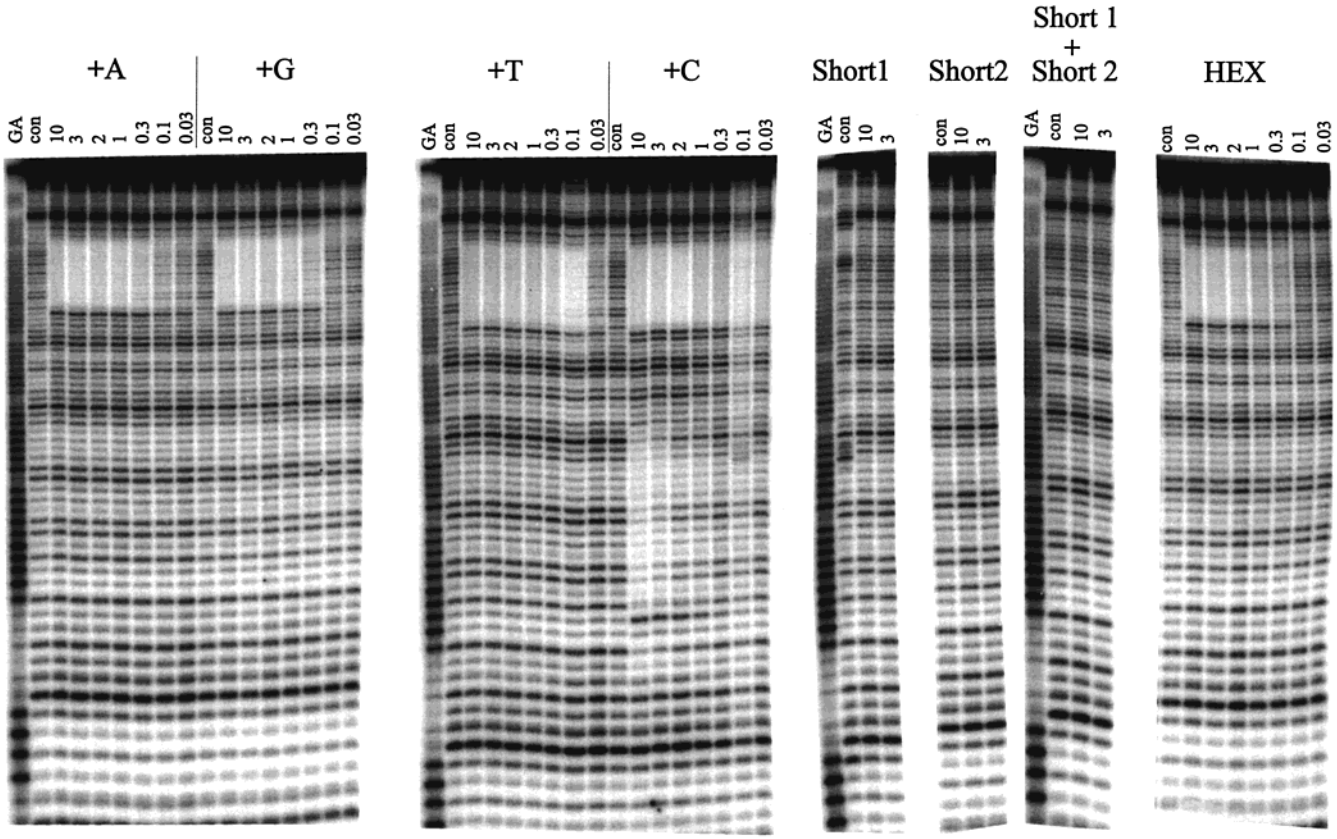

FIGURE 3: DNase I footprints showing the interaction of various oligonucleotides designed to generate single-base bulges with fragment DMG60R. The fragment was labeled on the purine-rich strand. Reactions were performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Oligonucleotide concentrations (micromolar) are shown at the top of each lane. For the panel labeled short 1 + short 2, both oligonucleotides were added at the concentration indicated. Lanes labeled GA are Maxam−Gilbert markers specific for purines. Lanes labeled con show digestion in the absence of added oligonucleotide.

the two 7-mers were joined by a hexa(ethylene glycol) moiety, i.e., in which the bulged nucleotide is replaced by hexa(ethylene glycol). The results are shown in the final panel of Figure 3 and show that this oligonucleotide produces a clear footprint at the same position as the oligonucleotide containing a single bulged base and that this persists to a concentration of about 0.3 μM, yielding C_{50} values of 0.22 $\pm$ 0.12 μM and 0.15 $\pm$ 0.18 μM on the labeled purine and pyrimidine strands, respectively (Table 2). This suggests that the bulged base is not directly affecting the interaction but is merely acting as a linker holding the two 7-mer triplexes together, enabling their simultaneous binding.

Oligonucleotides with Loops of Variable Length. We have extended these studies by examining the binding of oligonucleotides that have been designed to contain loops of between two and nine thymines at the same position, joining the two 7-mer oligonucleotides, generating a contiguous 14-mer triplex with a central short loop. The results of these experiments are presented in Figure 4. It can be seen that, as expected, none of these oligonucleotides binds to the intended site for oligo 5, but they all produce clear footprints around its secondary binding site, close to the intended target for oligo 2. Looking first at the results for +9T, which was designed to form a triplex with the longest loop, it can be seen that this oligonucleotide generates footprints on both purine- and pyrimidine-labeled strands that are identical to those seen in Figure 3 with oligonucleotides containing a single base bulge or with a hexa(ethylene glycol) linker. This footprint persist to an oligonucleotide concentration of about 0.3 μM, suggesting that a long loop can be accommodated at the center of a 14-mer triplex. A scheme for the proposed structure is shown in Figure 1H. It should be noted that the loop could also be located at other positions within the T-tract, slipping in the 3′-direction. However, in these cases the two halves of the contiguous triplex would be of unequal lengths, and we would expect this to produce a less stable complex. Oligonucleotides +2T, +3T, and +5T also produce DNase I footprints at the same location but with significant differences in their lengths. Oligonucleotide +2T produces the shortest footprint; on the labeled purine strand the lower (3′-) edge of the footprint is situated three bases higher than that seen with +T or +9T, though the upper (5′-) end of the footprint is located in the same position. A similar effect is seen on the labeled pyrimidine strand for which the footprint with this oligonucleotide is shorter at the upper (5′-) end than with the other oligonucleotides. One possible explanation for this footprint is shown in Figure 1F. It appears that this oligonucleotide does not bind by forming a dinucleotide loop at the center of the triplex but instead forms a contiguous 11-mer triplex with five bases frayed at the 3′-end of the third strand oligonucleotide. This is consistent with the results with other oligonucleotides described below, which suggest that triplexes containing two looped bases are less stable than their single-base counterparts. In contrast, oligonucleotide +3T produces the longest footprint. On the labeled purine strand this oligonucleotide produces enhanced cleavage at the 3′- (lower) end of the footprint in a similar location to that seen with oligo 2, six bases below that with +2T and

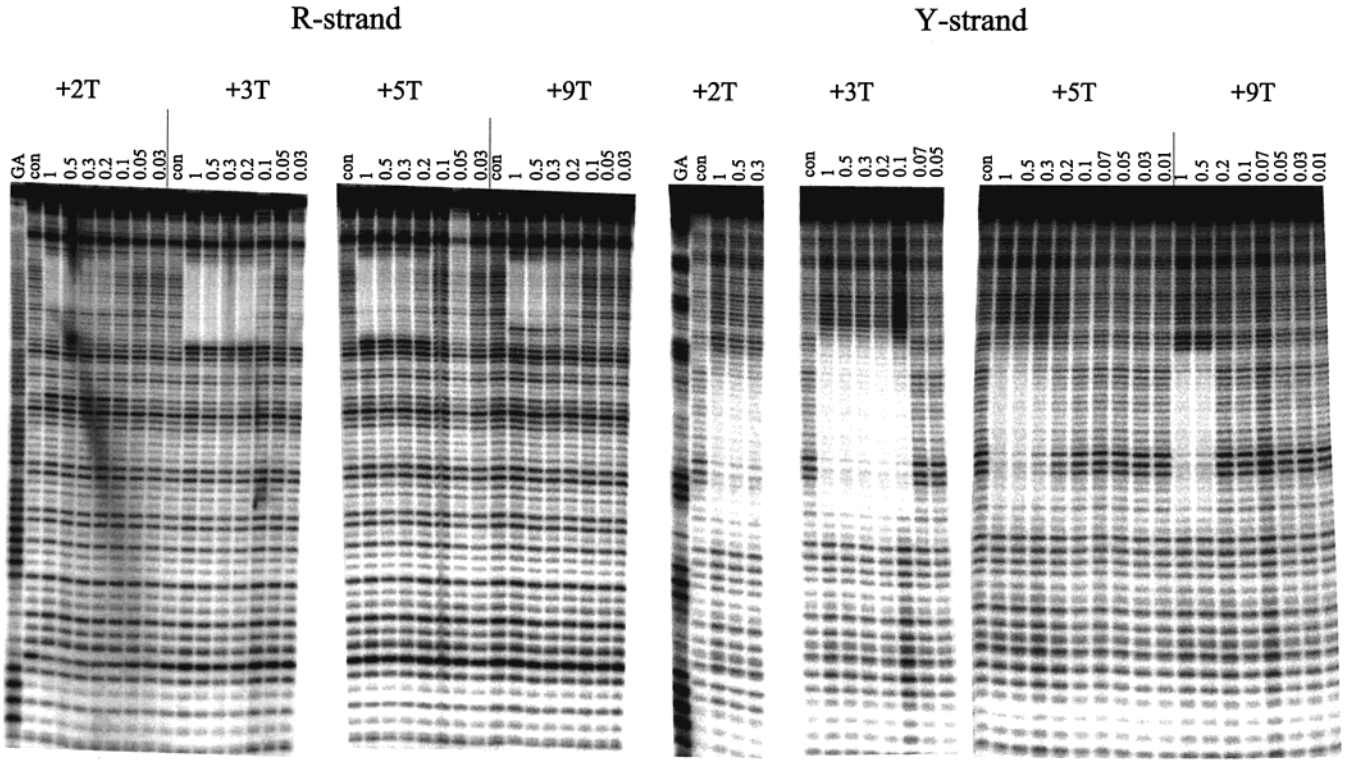

FIGURE 4: DNase I footprints showing the interaction of oligonucleotides, designed to form different length third-strand loops, with fragments DMG60R and DMG60Y. The first two panels show the results with DMG60R, revealing the purine-rich (R) strand, while the remaining panels show the results for DMG60Y, revealing the pyrimidine-rich (Y) strand. Reactions were performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Oligonucleotide concentrations (micromolar) are shown at the top of each lane. Lanes labeled GA are Maxam−Gilbert markers specific for purines. Lanes labeled con show digestion in the absence of added oligonucleotide.

three bases below +T or oligo 5. On the labeled pyrimidine strand this oligonucleotide produces a footprint that extends further at the 5′- (upper) end than with the other oligonucleotides. It appears that this 17-mer covers the same site as oligo 2, suggesting that its binding does not involve loop formation. The simplest explanation for the size and location of this footprint is that the oligonucleotide forms a 17-mer triplex containing a single T·GC mismatch toward its 3′-end as shown in Figure 1G. The formation of a 17-mer triplex containing a single mismatch therefore appears to be more stable than a 14-mer triplex with a three-base loop in the center. Surprisingly, we find that oligonucleotide +5T produces a similar footprint to +3T, even though it is two bases longer. It therefore appears that this oligonucleotide does not bind with the formation of a five-base loop. Instead we suggest that this 19-mer binds in a similar position to +3T, forming a 17-mer triplex containing a single T·GC mismatch, with the two additional thymines looped out. Since the results with +2T and the other dinucleotide loops described below show that these triplexes are less stable than those with single-base bulges, it is possible that the two additional thymines are looped out at different positions from the tract of nine Ts. C_{50} values determined for the binding of each of these oligonucleotides are presented in Table 2.

Footprinting with Methidiumpropyl−EDTA−Fe(II). Since DNase I produces uneven cleavage patterns and overestimates ligand binding site sizes, it is often not easy to determine rigorously the exact binding site location and size. We therefore attempted to map these oligonucleotide binding sites with greater accuracy using other footprinting probes. Hydroxyl radicals, which generate high-resolution cleavage maps, do not produce triplex footprints since cleavage occurs from the minor groove and is not affected by interaction of the oligonucleotide in the major groove. MPE−Fe(II) produces an even cleavage pattern that has been widely used to determine protein and small molecule binding sites. However, it has only occasionally been used for footprinting triplexes (*35*, *42*) and produces less clear footprints than DNase I (*35*), presumably because the methidium moiety can also intercalate into DNA triplexes (*43*−*45*). MPE footprinting patterns for several of these oligonucleotides on both DMG60Y and DMG60R are shown in Figure 5, and differential cleavage plots derived from these and other data are presented in Figure 6. It can be seen that oligo 2 produces attenuated cleavage over most of the expected target site. Oligo 5 does not affect the cleavage in this region but produces a footprint around its intended target site (Figure 5). The failure of oligo 5 to produce a footprint is probably because it binds less tightly than the other oligonucleotides. All the other oligonucleotides affect MPE cleavage in the vicinity of the binding site for oligo 2 but produce footprints of different sizes. These footprints are similar at the 5′-end of the target purine strand and the 3′-end of the pyrimidine strand, suggesting that the 5′-ends of the oligonucleotides are bound to the same region as oligo 2. However, these footprints are shorter than oligo 2 at the other end of the sequence. Although the boundaries of these footprints are less distinct than those produced by DNase I, it appears that oligonucleotide +T generates a footprint that is about three to four bases shorter than oligo 2, consistent with the model proposed in Figure 1E. Oligonucleotide +2T produces a footprint that is shorter than +T, consistent with the suggestion that it has an altered binding mode. Oligonucleotides +3T, +5T, and +9T produce footprints of similar

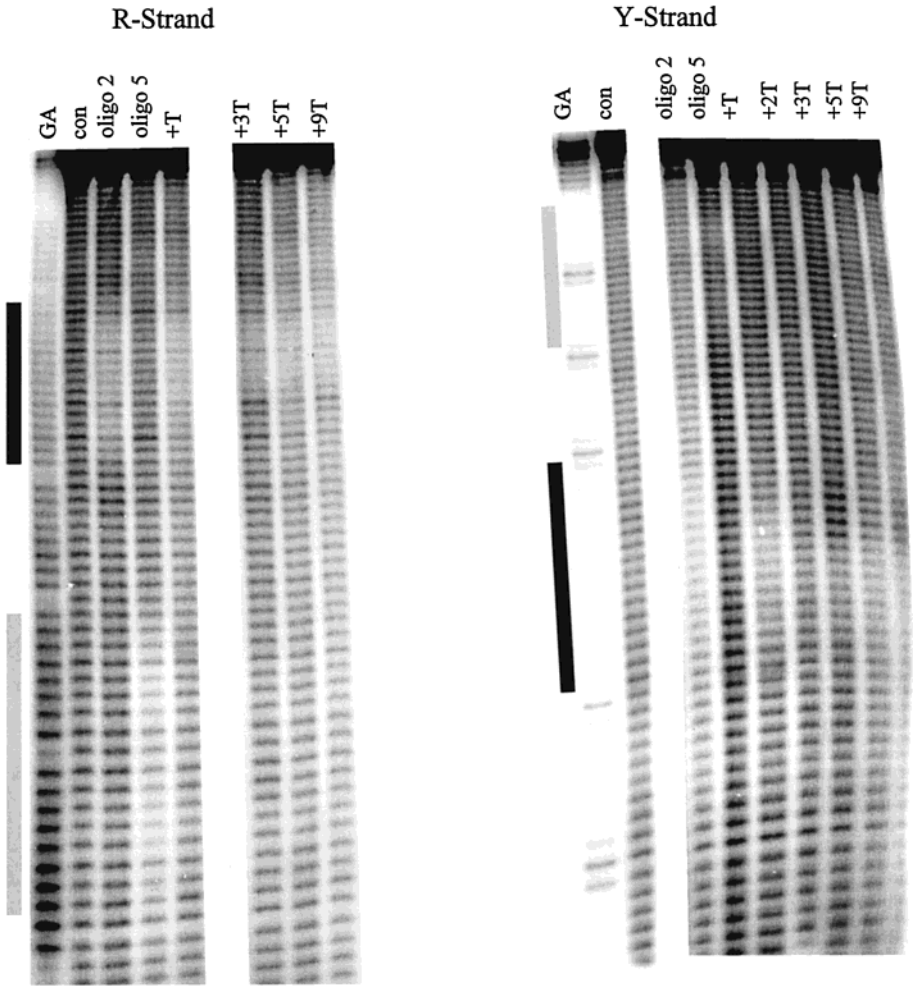

FIGURE 5: MPE−Fe(II) digestion of fragments DMG60R (left-hand panels) and DMG60Y (right-hand panels) in the presence of 2 μM of various oligonucleotides. Reactions were performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Lanes labeled GA are Maxam−Gilbert markers specific for purines. Lanes labeled con show digestion in the absence of added oligonucleotide. The dark bars indicate the binding site for oligo 2, while the light bars show the intended binding site for oligo 5.

size, which are shorter than those produced by the other oligonucleotides and which terminate about 5−6 bases from the end of the target site for oligo 2. These data suggest that the 3′-ends of oligonucleotides +3T, +5T, and +9T are not interacting with the DNA target site and are consistent with the binding model proposed for oligonucleotide +2T (Figure 1F). The DNase I and MPE footprinting data therefore suggest conflicting binding modes for oligonucleotides +3T, +5T, and +9T. These differences are considered further in the Discussion and are likely to arise from the interaction of MPE with triplex DNA, thereby perturbing the triplex binding equilibrium.

Oligonucleotides with Two Bases in the Loop. The results presented above suggest that oligo 5, with a central GC step, binds by loop formation with a C_{50} of about 1 μM, while the oligonucleotide with a central TT step (+2T) binds in an alternative configuration, generating an 11-mer triplex with five fraying bases. We have therefore examined the binding of other oligonucleotides that have the potential to generate triplexes containing two looped bases (+AT, +AA, +TA, and +CG) at the center of the 14-mer triplex. The results of these experiments are presented in Figure 7, along with those for the 14-mer oligonucleotide that should bind to this site by forming a simple triplex with no internal bulges (oligo 6). It can be seen that this perfectly matched oligonucleotide generates a stable triplex at the expected target site that persists to a concentration of less than 0.1 μM. Although the other oligonucleotides generate footprints at the same site, these require higher concentrations. Upon looking at each oligonucleotide in turn, it can be seen that +TA, +CG, and +AT produce footprints at the same site as oligo 6 but require higher concentrations. C_{50} values determined for the binding of these oligonucleotides are presented in Table 2. In contrast, +AA hardly affects the cleavage pattern, even at concentrations as high as 10 μM. In addition it can be seen that +TA also binds to another site, which is lower down the gel for the labeled purine strand and higher up for the labeled pyrimidine strand. The footprinting patterns for +TA are identical to those for oligo 5. It is not clear why +TA is able to form a 17-mer triplex at this site while +CG, +AA, and +AT do not. All of these 17-mer triplexes must generate at least two mismatched triplets at this site: with +TA these will be T·TA and A·GC, as illustrated in Figure 8C; +AT will generate A·TA and T·GC triplets; +AA; will generate A·TA while A·GC and +CG will form C·TA and G·GC.

DISCUSSION

The results presented in this paper demonstrate that triplex-forming oligonucleotides can bind at unexpected sites,

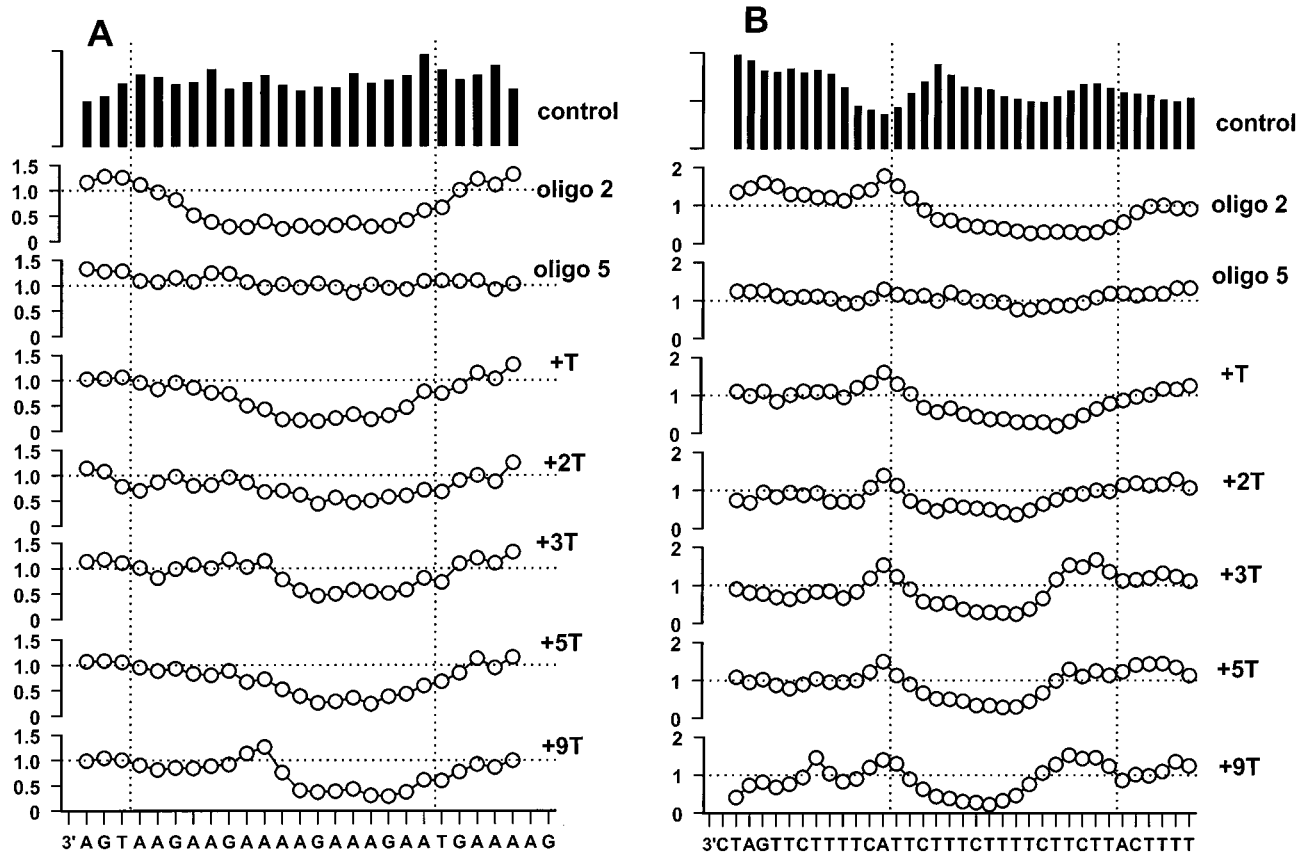

FIGURE 6: Differential cleavage plots showing the interaction of different oligonucleotides with the region in the vicinity of the binding site for oligo 2. Panel A shows data for the labeled purine strand (DMG60R), while panel B corresponds to the labeled pyrimidine strand. Note that in each case the DNA sequence in written is the direction 3′ → 5′ so that the right-hand end corresponds to the top of the gels. In each case the upper histogram shows the relative cleavage pattern of bands in the control DNA. The other plots show the differential relative cleavage of bands in the oligonucleotide-treated lanes compared with that in the control as described under Materials and Methods, so that values less than 1 indicate protection from cleavage. The horizontal dotted lines indicate differential cleavage of 1, i.e., where there is neither protection nor relative enhancement. The vertical dotted lines show the boundaries of the intended binding site for oligo 2.

generating complexes in which the third strand is not properly aligned with its target sequence. These secondary binding modes, which can involve one or two mismatched triplets, fraying of the third strand, or loop formation within the third strand, generate complexes with submicromolar dissociation constants.

Secondary Sites. At first sight it is surprising that, of the five oligonucleotides that were designed to bind to different regions of the oligopurine tract in DMG60, only oligo 5 produces a secondary footprint and that the secondary sites are always located in the vicinity of the target site for oligo 2. It seems likely that secondary binding occurs at this position as it contains the longest uninterrupted oligopurine tract in the sequence (17 bases). All the other oligopurine tracts are only eight bases long. The ability to form these secondary footprints is therefore related to the presence of a relatively long block of T·AT and C^+·GC triplets. The generation of a secondary footprint at any other site on this fragment would involve a pyrimidine interruption in the target region and would therefore produce a less stable complex. We attempted to align oligos 1−4 with different regions of the target sequence for oligo 2 but could find no matches that involved either one or two triplex mismatches or that generated a contiguous tract of T·AT and C^+·GC triplets by looping out a region of the third strand.

Loop Length. The model that we propose to account for the secondary binding of oligo 5 involves the formation of two adjacent 7-mer triplexes. This is made possible by looping out two bases (GC) in the center of the third strand. It should be remembered that in these experiments the DNA target concentration is much lower than that of the third-strand oligonucleotide. It is therefore not possible to use single-strand-specific cleavage agents to detect the presence of the single-stranded loops or bulges, as most of the added oligonucleotide, which is present in stoichiometric excess, is not bound to DNA. We attempted experiments with low concentrations (10 nM) of 5′-end-labeled oligonucleotides and high concentrations (2 μM) of unlabeled 53-mer duplex target sequence, comparing the accessibility of thymines to reaction with potassium permanganate (not shown). However, because of the short length of these oligonucleotides it was not possible to assess properly the reactivity in the region of the proposed loops and the results were equivocal. In general we found that, although oligos 2 and 5 were protected from modification when bound to the target, all the thymines in oligonucleotides containing T-loops were susceptible to cleavage. This may reflect the inability of permanganate to detect the interaction with these weaker-binding species.

We might anticipate that this interaction would not be dependent on the nature of the bases in the loop but that these merely act as a flexible linker enabling the simultaneous formation of the two triplexes. It is clearly essential that the two halves are covalently joined since the two 7-mer oligonucleotides (short 1 and short 2) do not produce a

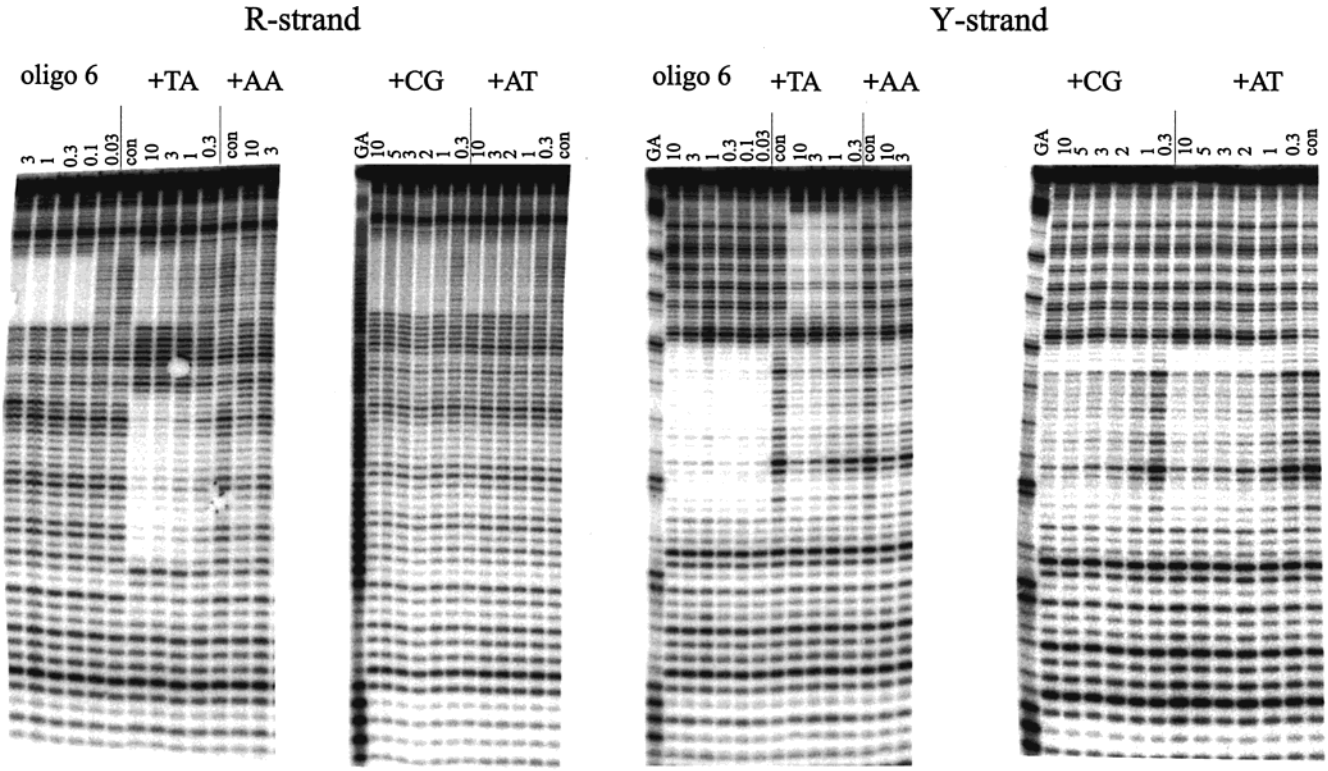

FIGURE 7: DNase I footprints showing the interaction of oligonucleotides, designed to form different dinucleotide bulges, with fragments DMG60R and DMG60Y. The fragments were labeled on either the purine-rich (R) strand (first two panels) or the pyrimidine-rich (Y) strand (last two panels). Reactions were performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Oligonucleotide concentrations (micromolar) are shown at the top of each lane. Lanes labeled GA are Maxam−Gilbert markers specific for purines. Lanes labeled con show digestion in the absence of added oligonucleotide.

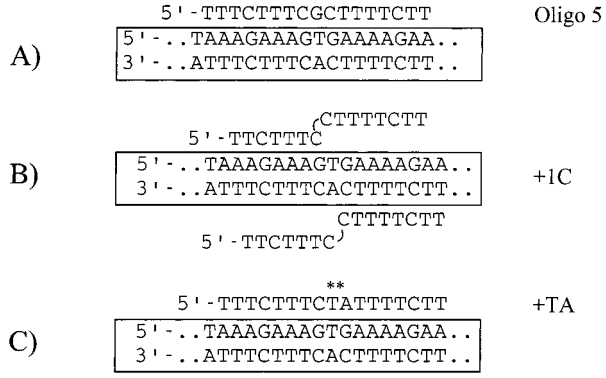

FIGURE 8: (A) Interaction of oligo 5 with its intended binding site. (B) Schematic for the possible two binding modes for oligonucleotide +1C at its secondary binding site. (C) Schematic for the binding of oligonucleotide +TA to its secondary binding site. The asterisks indicate the position of the mismatched triplets.

footprint either alone or when used together. Previous studies have shown cooperative binding of oligonucleotides to adjacent triplex sites (*48*, *49*). This is promoted by the continuous base stacking at the end of the two triplexes and is not observed when the two binding sites are separated by one base pair. The 7-mer oligonucleotides used in the present study bind too weakly to show such a 20-fold enhancement in equilibrium constant. However, cooperative binding of the two 7-mers can be induced by joining the two halves with a nonnucleosidic linker. Although our first observations with oligo 5 predicted the formation of a two-base loop, it appears that complexes with a single-base bulge form more stable complexes. This is consistent with NMR experiments on intramolecular triplexes, for which a complex containing a two-base AG bulge in the third strand gave a spectrum of poor quality, in contrast to a single adenine bulge, which produced a stable complex (*31*). However, the oligonucleotide designed to generate a nine-base loop (+9T) generates a complex that has a similar stability to that of other complexes with two bases in the loop.

Comparison of DNase I and MPE−Fe(II) Footprinting Data. Although MPE−Fe(II) is widely used as a footprinting probe for studying the binding of proteins and small molecules to DNA (*46*, *47*) there have been surprisingly few studies in which it has been used for examining triplex formation (*35*, *42*). We hoped that MPE would provide higher resolution footprinting data than DNase I, which in many cases does not produce reliable estimates of the absolute binding site size. We find that oligos 2 and 5 attenuate MPE cleavage in the vicinity of their intended target sites and that the looped oligonucleotides +T and +2T generate patterns consistent with the DNase I data. However, oligonucleotides +3T, +5T, and +9T produced shorter MPE footprints than expected. We suggest that this footprinting tool does not produce clear triplex footprints, as previously noted (*35*), because it binds to triplex DNA and perturbs the triplex−duplex equilibrium, as does the parent molecule ethidium (*43*−*45*). It appears that in the presence of this intercalator the formation of triplexes containing third-strand loops is not favored and the oligonucleotides adopt a binding mode in which the 3′-ends fray from the target site. The different site sizes indicated by the two footprinting probes emphasize that several secondary binding modes are possible and that these may be in a dynamic equilibrium that can be perturbed by factors such as ligand binding.

Base Composition. If the bases in the bulge were merely acting as a flexible linker, then we would not expect triplex formation to be affected by its sequence composition. However, the C_{50} values, which approximate to the triplex dissociation constants, show that the stability is influenced by the nature of the looped bases. For single-base additions, triplex stability decreases in the order C > T > G = A, suggesting that pyrimidine additions generate more stable complexes than purines. This might be caused by the larger aromatic surface area of purines, which is likely to be exposed to solvent. The greater stability of complexes with bulged C than T may be a consequence of the positive charge on the pronated cytosine and is consistent with previous studies using gel competition and melting curves (*30*). The greater stability of secondary complexes containing central pyrimidines is disappointing from a therapeutic perspective, as these are the bases that are required for generating parallel triplexes. It therefore appears that an *N*-mer triplex-forming oligonucleotide may show significant interaction with a related $N - 1$ site by extruding a single base from the center of the third strand.

The complexes with two additional bases also show some dependence on the nature of the looped bases. The complexes formed with central TA, AT, and CG have lower C_{50} values than GC (oligo 5). This cannot be simply due to the additional thymine at the 5′-end of oligo 5, as +TA also possesses an additional T at the 5′-end. The effect of the central dinucleotide is most apparent with +AA, which did not produce a DNase I footprint. This may also reflect the lower stability of complexes with central purines than pyrimidines but might also be due to stacking of the As, preventing the formation of a short loop.

Loops or Other Structures. The results with oligonucleotides containing different numbers of additional thymines show that, although these all bind in a similar location, only +T and +9T bind by loop formation. It appears that for +2T, +3T, and +5T other triplex structures are more stable than those containing the proposed thymine loops. This emphasizes that secondary triplex binding may occur by a variety of different mechanisms, including base mismatches and fraying ends, making it hard to predict. Any one oligonucleotide may also be able to bind in several different secondary modes, which may be in dynamic equilibrium

Implications. The formation of these unusual and unexpected triplex structures means that extra care will be needed if triplex-forming oligonucleotides are to be used as gene-specific agents. To target a unique sequence within the human genome it is necessary to recognize at least 17 base pairs; shorter sequences are likely to occur more than once and will therefore target more than one gene. However, on the basis of the results presented in this paper, it appears that in addition to forming a 17-mer triplex, micromolar concentrations of a 17-mer would also form complexes with several other sites. These would include shorter targets (11 bases and longer) to which one end of the oligonucleotide might bind, 17-mer tracts generating complexes with one triplet mismatch and 14-, 15-, and 16-mer oligopurine tracts to which it might bind, forming three-, two-, and one-base third-strand loops. Indeed, it may be that there will be fewer genomic target sites for shorter oligonucleotides.

REFERENCES

1. Vasquez, K. M., and Wilson, K. H. (1998) *Trends Biochem. Sci. 23*, 4−9.
2. Chubb, J. M., and Hogan, M. E. (1992) *Trends Biotechnol. 10*, 132−136.
3. Chan, P. P., and Glazer, P. M. (1997) *J. Mol. Med. 75*, 267−282.
4. Neidle, S. (1997) *Anti-Cancer Drug Des. 12*, 433−442.
5. Soyfer, V. N., and Potaman, V. N. (1996) *Triple Helical Nucleic Acids*, Springer-Verlag, New York and Berlin.
6. Thuong, N. T., and Hélène, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 666−690.
7. Fox, K. R. (2000) *Curr. Med. Chem. 7*, 17−37.
8. Moser, H. E., and Dervan, P. B. (1987) *Science 238*, 645−650.
9. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., and Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749−7760.
10. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 17−32.
11. Beal, P. A., and Dervan, P. B. (1991) *Science 251*, 1360−1363.
12. Chen, F.-M. (1991) *Biochemistry 30*, 4472−4479.
13. Radhakrishnan, I., and Patel, D. J. (1993) *Structure 1*, 135−152.
14. Doronina, S. O., and Behr, J. P. (1997) *Chem. Soc. Rev. 26*, 63−71.
15. Gowers, D. M., and Fox, K. R. (1999) *Nucleic Acids Res. 27*, 1569−1577.
16. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967−971.
17. Kiessling, L. L., Griffin, L. C., and Dervan, P. B. (1992) *Biochemistry 31*, 2829−2834.
18. Radhakrishnan, I. Patel, D. J., and Gao, X. (1992) *Biochemistry 31*, 2514−2523.
19. Gowers, D. M., and Fox, K. R. (1998) *Nucleic Acids Res. 26*, 3626−3633.
20. Svinarchuk, F., Bertrand, J.-R. and Malvy, C. (1994) *Nucleic Acids Res. 22*, 3742−3747.
21. Svinarchuk, F., Paoletti, J., and Malvy, C. (1995) *J. Biol. Chem. 270*, 14068−14071.
22. Mergny, J. L., Sun, J. S., Rougee, M., Montenay-Garestier, T., Barcelo, B., Chomilier, J., and Hélène, C. (1991) *Biochemistry 30*, 9792−9798.
23. Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougée, M., Montenay-Garestier, T., Nisagni, E., and Hélène, C. (1992) *Science 256*, 1681−1684.
24. Wilson, W. D., Tanious, F. A., Mizan, S., Yao, S., Kiselyov, A. S., Zon, G., and Strekowski, L. (1993) *Biochemistry 32*, 10614−10621.
25. Escudé, C., Nguyen, C. H., Kukreti, S., Janin, Y., Sun, J.-S., Bisagni, E., Garestier, T., and Hélène, C. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 3591−3596.
26. Keppler, M. D., Read, M. A., Perry, P. J., Trent, J. O., Jenkins, T. C., Reszka, A. P., Neidle, S., and Fox, K. R. (1999) *Eur J. Biochem. 263*, 817−825.
27. Sun, J.-S., Francois, J.-C., Montenay-Garestier, T., Saison-Behmoaras, T., Roig, V., Thuong, N. T., and Hélène, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9198−0202.
28. Giovannangeli, C., Thuong, N. T., and Hélène, C. (1992) *Nucleic Acids Res. 20*, 4275−4281.
29. Keppler, M. D., McKeen, C. M., Zegrocka, O., Strekowski, L., Brown, T., and Fox, K. R. (1999) *Biochim. Biophys. Acta 1447*, 137−145.
30. Roberts, R. W., and Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397−9410.
31. Hardenbol, P., and Van Dyke, M. W. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7811−7816.
32. Chandler, S. P., and Fox, K. R. (1993) *FEBS Lett. 332*, 189−192.
33. Fossella, J. A., Kim, Y. J., Shih, H., Richard, E. G., and Fresco, J. R. (1993) *Nucleic Acids Res. 21*, 4511−4515.
34. Wang, Y., and Patel, D. J. (1995) *Biochemistry 34*, 5696−5704.

35. Marchand, C., Bailly, C., Nguyen, C. H., Bisagni, E., Garestier, T., Hélène, C., and Waring, M. J. (1996) *Biochemistry 35*, 5022−5032.
36. Distefano, M. D., Shin, J. A., and Dervan, P. B. (1991) *J. Am. Chem. Soc. 113*, 5901−5902.
37. Distefano, M. D., Shin, J. A., and Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1179−1183.
38. Distefano, M. D., and Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 11006−11007.
39. Liberles, D. A., and Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 9510−9514.
40. Akiyama, T., and Hogan, M. E. (1997) *Biochemistry 36*, 2307−2315.
41. Dabrowiak, J. C., Goodisman, J. and Ward, B. (1997) *Methods Mol. Biol. 90*, 23−42.
42. Beal, P. A., and Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 4976−4982.
43. Scaria, P. V., and Shafer, R. H. (1991) *J. Biol. Chem. 266*, 5417−5423.
44. Mergny, J. L., Collier, D., Rougée, M., Montenay-Garestier, T., and Hélène, C. (1991) *Nucleic Acids Res. 19*, 1521−1526.
45. Pilch, D. S., and Breslauer, K. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9332−9336.
46. Hertzberg, R. P., and Dervan, P. B. (1984) *Biochemistry 23*, 3984−3945.
47. Van Dyke, M. W., and Dervan, P. B. (1982) *Biochemistry 22*, 2372−2377.
48. Colocci, N., Distefano, M. D., and Dervan, P. B. (1993) *J. Am. Chem. Soc. 115*, 4468−4473.
49. Colocci, N., and Dervan, P. B. (1994) *J. Am. Chem. Soc. 116*, 785−786.

BI992773+