Specificity of Antiparallel DNA Triple Helix Formation†

Simon P. Chandler and Keith R. Fox*

Department of Physiology and Pharmacology, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K.

Received April 22, 1996; Revised Manuscript Received September 10, 1996⊗

ABSTRACT: We have used DNase I footprinting to examine the formation of antiparallel DNA triple helices on DNA fragments containing the homopurine target sites $(GGA)_2GGX(GGA)_2GG{\cdot}(CCT)_2CCZ(CCT)_2CC$ (where X·Z is each base pair in turn), with the GA- and GT-rich oligonucleotides, $(GGA)_2GGN(GGA)_2GG$ and $(GGT)_2GGN(GGT)_2GG$ (N = each base in turn). These were designed to form G·GC and A·AT or T·AT triplets with a central N·XZ mismatch, which should bind in an antiparallel orientation. We find that almost all combinations generate DNase I footprints at low micromolar concentrations. At each target site, the relative binding of the GA- and GT-containing oligonucleotides was not the same, suggesting that these two triplexes adopt different conformations. For a central GC base pair, the most stable complex is observed with a third strand generating a G·GC triplet as expected. A·GC is also stable, especially in the GT oligonucleotides. For a central AT base pair, all four bases form stable complexes though T·AT is favored for the GA-rich thirds strands and A·AT for the GT-rich strands. For a central CG base pair, the stable complexes are seen with third strands generating T·CG triplets, though A·CG and C·CG are stable with GT- and GA-containing oligonucleotides, respectively. C·TA is the best triplet at a central TA base pair. The third strands with central guanines avoided the formation of G·YR triplets on the fragments containing central pyrimidines, producing DNase I footprints which had slipped relative to the target site. These oligonucleotides bound at a different location, generating complexes containing 11 contiguous stable triplets at the 3′-end of the third strand. The results suggest rules for designing the best third strand oligonucleotides for targeting sequences in which homopurine tracts are interrupted by pyrimidines.

Intermolecular DNA triple helix formation has the potential for designing agents endowed with considerable sequence recognition properties, which may be used for controlling gene expression (Hélène, 1991a,b; Moffat, 1991; Chubb & Hogan, 1992). The discovery of triple-stranded structures was made nearly 40 years ago using synthetic polyribo- and polydeoxyribonucleotides (Felsenfeld *et al*., 1957; Arnott & Selsing, 1974). Since then, studies have shown that the third strand nucleic acid resides within the major groove of the target duplex and that this binding is sequence-specific (Moser & Dervan, 1987; Le Doan *et al*., 1987). The third strands bind by forming specific hydrogen bonds to substituents on the purines of the target DNA (Thuong & Hélène, 1993; Hélène, 1993; Sun & Hélène, 1993). The orientation of the third strand depends on the nature of its sequence composition.

Pyrimidine-rich third strand oligonucleotides are aligned in a parallel orientation with respect to the homopurine strand of the target. Specific interactions are attained through Hoogsteen-type base pairing, generating T·AT and C^+·GC triplets (Moser & Dervan, 1987; Le Doan *et al*., 1987; Radhakrishnan & Patel, 1994). Within this structure, the C^+·GC is only stable at low pHs (<6.0) as a result of the requirement for protonation of the third strand cytosine. Purine-rich third strand oligonucleotides bind in an antiparallel orientation with respect to the homopurine target strand, forming reverse Hoogsteen interactions, generating G·GC, A·AT, and T·AT base triplets (Beal & Dervan, 1991; Durland *et al*., 1991; Radhakrishnan & Patel, 1993; Radhakrishnan *et al*., 1993). The formation of this triplex motif is not dependent on pH, though both structures require the presence of divalent cations, especially magnesium.

The parallel motif has been the most well studied, though its pH dependence must limit its therapeutic use. A major effort has been made to discover a universal recognition code allowing the specific binding of mixed sequence DNA, and other weaker triplets have been described including G·TA and T·CG (Chandler & Fox, 1993; Griffin & Dervan, 1989; Yoon *et al*., 1992; Kiessling *et al*., 1992). Structural studies show that, in these triplets, the third strand is held in place by the formation of only one hydrogen bond, and suggest that flanking sequences can have an important effect on their stability. In theory, these different triplets could permit the recognition of all four DNA bases, generating T·AT, C^+·GC, G·TA, and T·CG triplets. Similar studies on antiparallel triplexes have shown the possibility of forming A·GC, T·CG, and C·AT as weaker triplets (Beal & Dervan, 1991, 1992; Dittrich *et al*., 1994; Durland *et al*., 1994). However, it should be noted that, in this motif, there is no documented means of recognizing a TA base pair.

We have previously explored the stringency of parallel triplex formation at homopurine target sites interrupted by a central pyrimidine residue (Chandler & Fox, 1993), using DNA fragments with $A_8NA_8{\cdot}T_8NT_8$ target sites, recognizing them with third strands of the type T_8NT_8. In this paper,

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

* Correspondence should be addressed to this author at the Department of Physiology and Pharmacology, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K. Telephone: +1703-594374. Fax: +1703−594319.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0006-2960(96)00967-1 CCC: $12.00

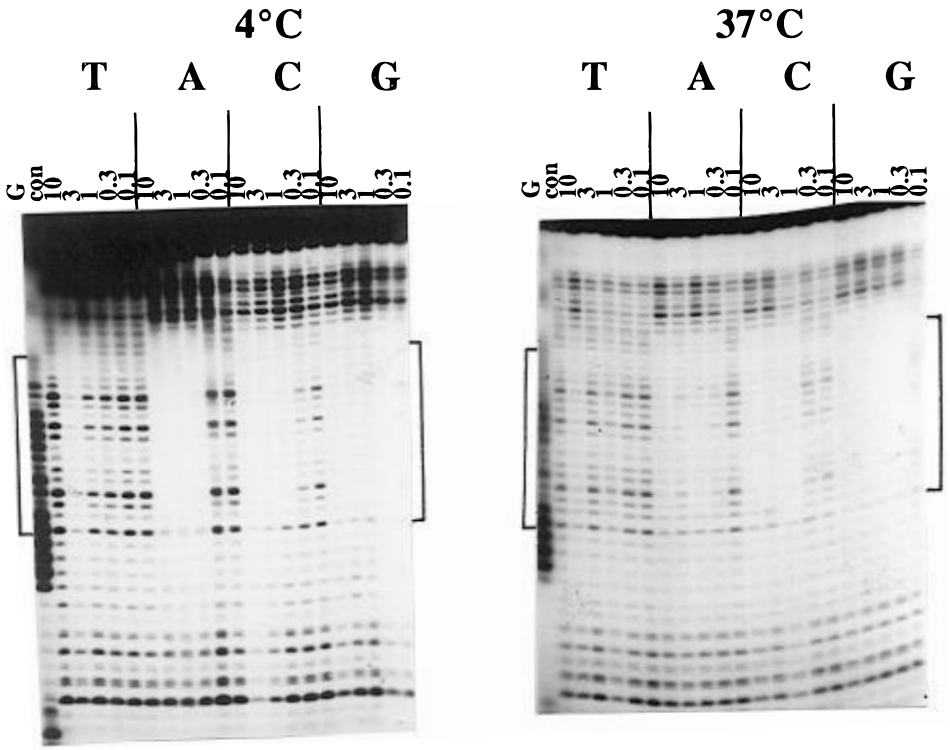

FIGURE 1: DNase I digestion of the fragment containing the target site $(GGA)_2GGG(GGA)_2GG$ in the absence (con) and presence of N_A oligonucleotides at 4 and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square bracket indicates the position of the 17 base triplex target site. The track labeled "G" represents a Maxam−Gilbert dimethyl sulfate−piperidine marker specific for guanines.

we report the results of similar footprinting studies on the formation of antiparallel triplexes. Since G·GC appears to be the strongest antiparallel triplet, we used GC-rich target sites. However, G-rich oligonucleotides can adopt stable structures, and GC-rich DNAs often suffer from band compressions during electrophoresis. We therefore chose to work with a target site containing both A and G residues, using sequences of the type $(GGA)_2GGX(GGA)_2GG$·-$(CCT)_2CCZ(CCT)_2CC$ (where X·Z = each base pair in turn). Within this motif, AT base pairs can be recognized by either A or T, generating A·AT and T·AT triplets, respectively. We therefore used two sets of oligonucleotides to recognize these targets of the type $(GGA)_2GGN(GGA)_2GG$ and $(GGT)_2GGN(GGT)_2GG$ (where N = each base in turn).

MATERIALS AND METHODS

Chemicals and Enzymes. Deoxyoligonucleotides were purchased from Genosys Biotechnologies, Inc., or Oswel DNA Service, dissolved in water to a concentration of 1 mM, and used without further purification. These stock solutions were stored at −20 °C.

Bovine pancreatic DNase I was purchased from Sigma and stored at −20 °C at a concentration of 7200 units/mL. Restriction enzymes and AMV reverse transcriptase were purchased from Promega. *Sma*I-cut, alkaline phosphatase treated pUC18 plasmid was purchased from Pharmacia. All other chemicals were purchased from Sigma.

DNA Plasmids. Oligonucleotides $(GAA)_2GGN(GGA)_2GG$ and $(CCT)_2CCN(CCT)_2CC$, containing a redundant central base, were treated with polynucleotide kinase and ATP, annealed, and ligated into *Sma*I-cut, alkaline phosphatase-treated pUC18. The mixture was transformed into *E. coli* TG2 and grown on agar plates containing ampicillin and X-gal. Successful clones, detected as white colonies in the usual way, were sequenced using a T7 sequencing kit (Pharmacia). Plasmids containing each of the four bases were obtained in this way, and are denoted by the names pSCAP-A, pSCAP-T, pSCAP-C, and pSCAP-G, where the final letter refers to the identity of the central base on the purine strand. The inserts were oriented so that labeling the 3′-end of the *Hin*dIII site visualized the purine-rich strands of pSCAP-A, pSCAP-G, and pSCAP-C, but the pyrimidine-rich strand of pSCAP-T.

Labeled DNA fragments containing these inserts were obtained by cutting the plasmids with *Hin*dIII, labeling at the 3′-end with [α-^{32}P]dATP using AMV reverse transcriptase, and cutting again with *Eco*RI. The radiolabeled fragments were separated from the rest of the plasmid on 8% polyacrylamide gels. In some instances, the fragments were labeled at the other end by reversing the order of addition of *Eco*RI and *Hin*dIII. The isolated DNA fragments were dissolved in 10 mM Tris-HCl, pH 7.5, containing 0.1 mM EDTA at a strand concentration of approximately 10 nM.

Oligonucleotides. Third strand oligonucleotides are referred to by the symbols N_A and N_T, where N denotes the identity of the central base and the subscript A or T indicates whether the oligonucleotide contains GA or GT bases. The

Table 1: Abbreviations Used for the Various Third Strand Oligonucleotides

sequence	symbol	sequence	symbol
$(GGA)_2GGA(GGA)_2GG$	A_A	$(GGT)_2GGA(GGT)_2GG$	A_T
$(GGA)_2GGG(GGA)_2GG$	G_A	$(GGT)_2GGG(GGT)_2GG$	G_T
$(GGA)_2GGC(GGA)_2GG$	C_A	$(GGT)_2GGC(GGT)_2GG$	C_T
$(GGA)_2GGT(GGA)_2GG$	T_A	$(GGT)_2GGT(GGT)_2GG$	T_T

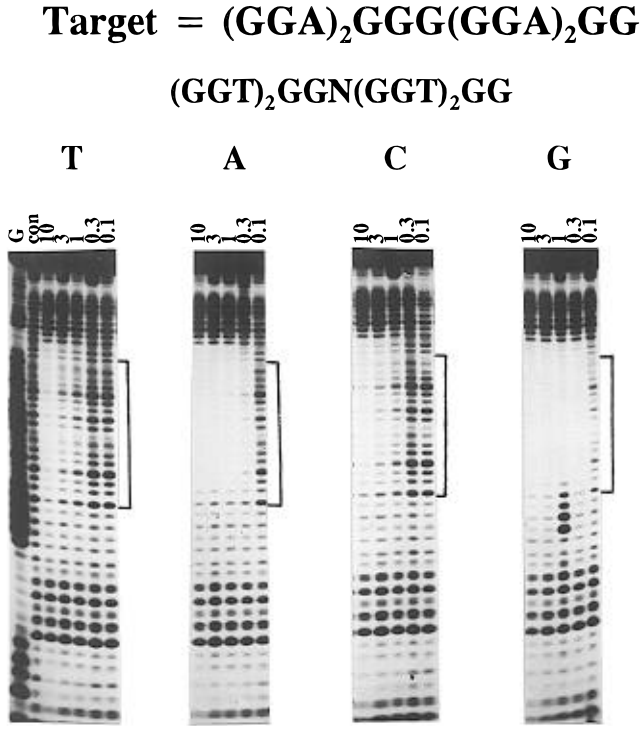

FIGURE 2: DNase I digestion of the fragment containing the target site $(GGA)_2GGG(GGA)_2GG$ in the absence (con) and presence of N_T oligonucleotides. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The experiment was performed at 20 °C. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. The track labeled "G" represents a Maxam−Gilbert dimethyl sulfate−piperidine marker specific for guanines.

sequences of these third strand oligonucleotides are summarized in Table 1.

DNase I Footprinting. The radiolabeled DNA fragments (1.5 μL) containing the cloned target sites were mixed with 3.0 μL of oligonucleotide, dissolved in 10 mM Tris-HCl, pH 7.5, containing 5 mM $MgCl_2$. It should be noted that spermine was not included in any of the incubation mixtures, since we find that it affects the efficiency of DNase I cleavage. All concentrations in the text refer to the final concentrations in the complex mixture. The complexes were then heated to 65 °C for 5 min to disrupt any internal structures adopted by the G-rich oligonucleotides, cooled to the appropriate temperature, and equilibrated overnight. We found that the heating step was essential for generating successful complexes with several of the oligonucleotides. Each complex was studied at several temperatures (4, 20, and 37 °C), though, for the sake of clarity, only representative examples are presented. The rate of formation of triplexes is known to be much slower than that of duplexes with bimolecular rate constants of around 10^3 M^{-1} s^{-1} (Shindo *et al.*, 1993; Rougée *et al.*, 1992; Vasquez *et al.*, 1995; Fox, 1995). We estimate that under the conditions used in

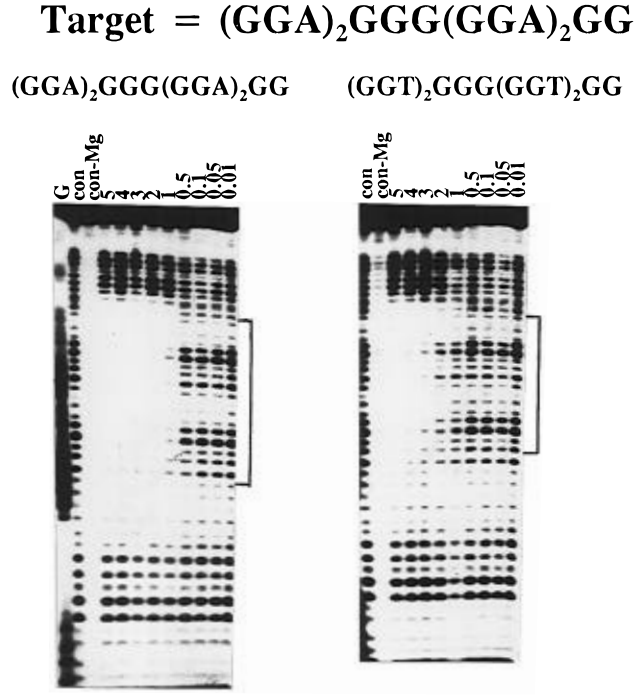

FIGURE 3: DNase I digestion of the fragment containing the insert $(GGA)_2GGG(GGA)_2GG$ with 1 μM G_A and G_T in the presence of varying concentrations of magnesium. The experiment was performed at 20 °C in buffer containing 10 mM Tris-HCl, pH 7.5. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Magnesium concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. The track labeled "G" represents a Maxam−Gilbert dimethyl sulfate−piperidine marker specific for guanines.

footprinting experiments, in which the concentration of the target DNA is much less than that of the third strand, the association half-life should therefore be about 2 h with 0.1 μM oligonucleotide. The overnight incubation should therefore be sufficient to attain equilibrium at the higher temperatures, though it should be noted that this may not be so at the lowest temperatures. This is considered further under Discussion. The samples were digested with 2 μL of DNase I, dissolved in 2 mM $MgCl_2$, 2 mM $MnCl_2$, and 20 mM NaCl. The enzyme concentration was varied according to the reaction conditions, so as to ensure that about 70% of the DNA remained uncut. At 20 °C, the enzyme concentration was typically 0.01 unit/mL. The digestion was stopped after 1 min by adding 4.5 μL of DNase I stop solution (80% formamide containing 10 mM EDTA). The products of digestion were separated on 40 cm long polyacrylamide gels, 10% (for *Hin*dIII-labeled fragments) or 13% (for *Eco*RI-labeled fragments), containing 8 M urea and run at 1500 V for 2 h. The gels were then fixed in 10% acetic acid before drying at 80 °C and subjecting to autoradiography at −70 °C using an intensifying screen. Bands were assigned by comparison with Maxam−Gilbert sequencing markers specific for guanine or purines. When comparing the oligonucleotide concentrations required to generate DNase I footprints on each fragment, care was taken to ensure that each sample had been digested to a similar extent, as assessed by the intensity of bands outside the triplex target site.

RESULTS

pSCAP-G Target Sequence. Figure 1 shows DNase I digestion of the fragment derived from pSCAP-G in the

presence of the GA-containing oligonucleotides (N_A). These oligonucleotides were designed to generate G·GC and A·AT triplets on either side of the central base pair. This fragment contains a guanine at the center of the purine strand, and should therefore represent a good triplex binding site for the oligonucleotide G_A, generating a central G·GC triplet. It can be seen that, at 4 °C, G_A does indeed produce a clear footprint which extends to concentrations as low as 0.1−0.3 μM. Although the exact location of the footprint is difficult to assess by DNase I footprinting, it appears to be centered around the target site. G_A is the only oligonucleotide able to form a standard (G·GC) antiparallel triplet at the central guanine; however, it can be seen that the other oligonucleotides also produce footprints around this target site, though at higher concentrations. Both A_A and C_A, which would generate central A·GC and C·GC triplets, respectively, produce footprints which are still evident at oligonucleotide concentrations as low as 1 μM, but not 0.3 μM. The least stable interaction is seen with T_A which only produces a footprint at the highest concentration used (10 μM). Similar results were obtained when this experiment was repeated at 37 °C, confirming G_A and T_A as the best and worst third strands, respectively, and giving the order of stability as G·GC > A·GC = C·GC > T·GC, which is similar to that previously reported (Beal & Dervan, 1992).

Figure 2 shows the results of similar experiments using the N_T third strands, designed to generate G·GC and T·AT triplets. The results presented correspond to 20 °C; similar patterns were evident at both 4 °C and 37 °C. Once again, a clear footprint is produced by G_T, as expected since this generates a central G·GC triplet, which persists down to an oligonucleotide concentration of 0.3 μM. A_T also affects the cleavage pattern at low concentrations, producing a clear footprint at 1 μM with strongly attenuated cleavage at 0.3 μM. In contrast, both oligonucleotides with central pyrimidines (C_T and T_T) bind less well and only show clear footprints at concentrations around 10 μM. In this sequence context, the order of stabilities is G·GC ≥ A·GC > C·GC = T·GC

It is clear that, for both GT- and GA-containing oligonucleotides, third strands with a central G (G_A and G_T) form the most stable complexes at this target site as expected. However, the rank order of binding of the N_A and N_T oligonucleotides is not the same. A_A binds less well than G_A, while A_T and G_T have similar affinities; C_A is very similar to A_A, while C_T binds less well than A_T. These differences suggest that the relative affinity of different triplex "mismatches" depends on the sequence context. Since different triplets have been shown to be selectively stabilized by various divalent metal ions (Malkov *et al.*, 1993), we compared the magnesium ion dependency of the complexes formed by G_T and G_A at this target site. The results of this experiment are presented in Figure 3. In these experiments, the concentration of each oligonucleotide (G_T and G_A) was maintained at 1 μM, and the divalent magnesium ion concentration varied between 10 μM and 5mM (the experiments presented in Figures 1 and 2 were performed in the presence of 5 mM $MgCl_2$). It can be seen that 1 μM G_A produces a footprint at magnesium concentrations as low as 1 mM, while 1 μM G_T requires slightly higher concentrations, but still induces a clear footprint at 3 mM $MgCl_2$.

pSCAP-A Target Sequence. Figure 4 shows DNase I digestion of the fragment derived from pSCAP-A in the presence of the GA-containing oligonucleotides (N_A). This

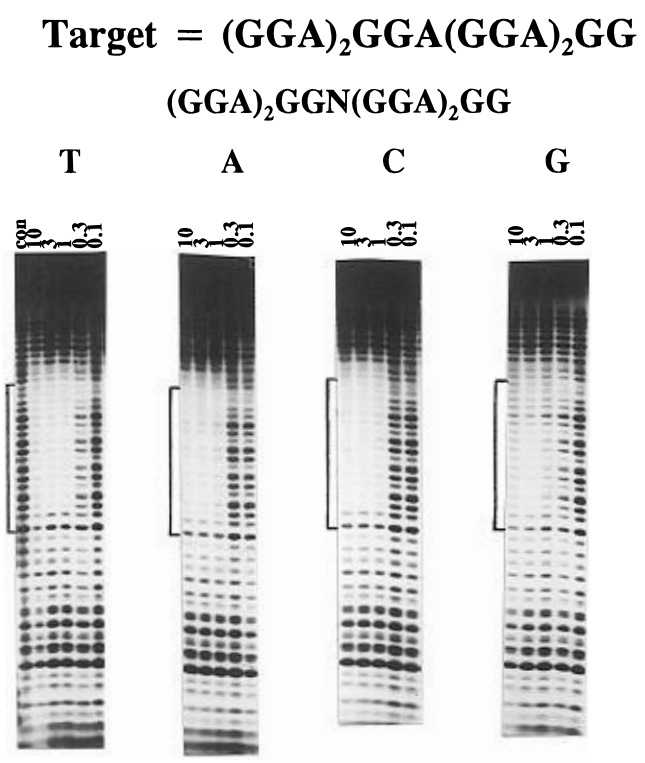

FIGURE 4: DNase I digestion of the fragment containing the target site $(GGA)_2GGA(GGA)_2GG$ in the absence (con) and presence of N_A oligonucleotides. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The experiment was performed at 4 °C. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site.

fragment has an adenine at the center and so presents an uninterrupted run of purines, which should be a good triplex target site. Since A can be recognized by A or T, within an antiparallel triplex, generating A·AT and T·AT triplets, we might expect that A_A and T_A should form the most stable complexes. It can be seen that, at 4 °C, all four oligonucleotides generate DNase I footprints which persist to similar concentrations (about 1 μM), though G_A and T_A appear to be slightly better, producing attenuated cleavage at 0.3 μM. A similar trend is evident at 37 °C (not shown), with T_A forming the most stable triplex at a concentration of 0.1 μM, while A_A, C_A, and G_A all produce footprints at 0.3 μM. In each case, the footprints are centered around the target site; the footprint is coterminal with the 3′ (lower) edge of the target and extends by a few bases in the upper (5′) direction. These results suggest that, in the context of antiparallel A·AT and G·GC triplets, an AT base pair can form reasonably stable complexes with all four bases, though T·AT appears to be the slightly better at higher temperatures.

Figure 5 shows the results of similar experiments using the oligonucleotides designed to form T·AT and G·GC triplets (N_T). At 4 °C, G_T forms the most stable interaction, producing a footprint at 1.0 μM and attenuated cleavage at 0.3 μM. The least stable complexes are formed by the oligonucleotides with central pyrimidines (C_T and T_T), which both yield footprints only at 10 μM. A_T forms a triplex of intermediate stability producing a footprint at 3.0 μM, but not 1 μM. This order of stability is retained at 20 °C (not shown), at which slightly lower concentrations of each oligonucleotide are required. G_T is the most stable, giving a clear footprint at 0.3 μM; C_T and T_T require 3 μM, and A_T

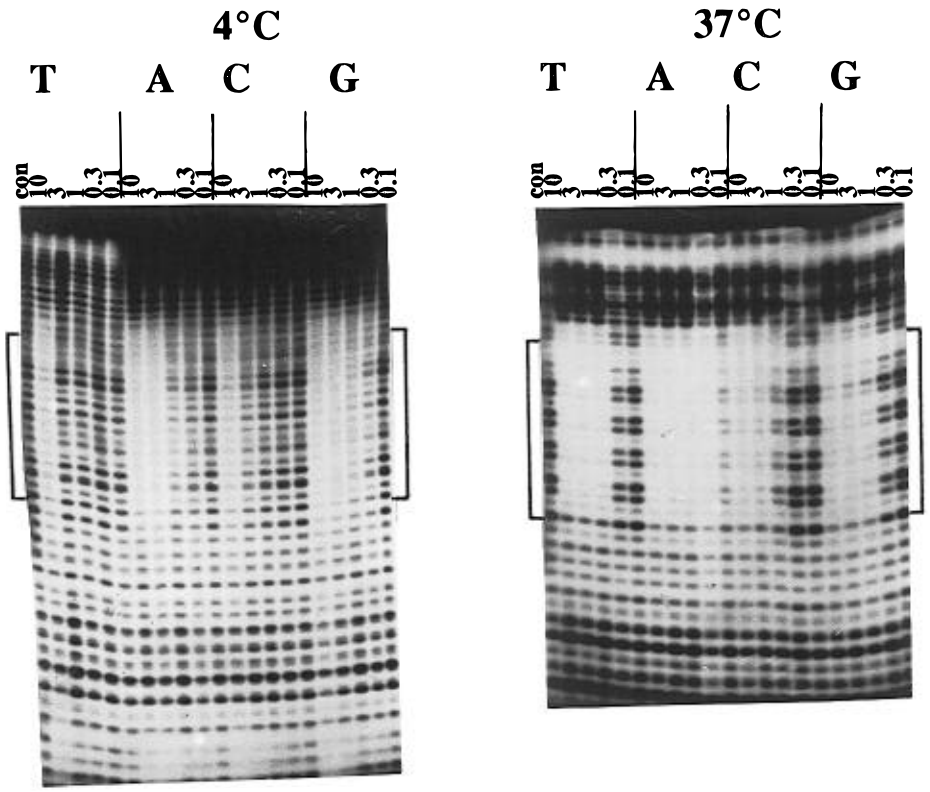

FIGURE 5: DNase I digestion of the fragment containing the target site $(GGA)_2GGA(GGA)_2GG$ in the absence (con) and presence of N_T oligonucleotides at 4 and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site.

displays intermediate stability, protecting at 1.0 μM. Under these conditions, it appears that the G·AT triplet is favored over A·AT, while T·AT and C·AT appear to be weaker. The poor binding of C_T and C_A is surprising since the formation of a C·AT triplet has been suggested in previous studies (Beal & Dervan, 1992). In contrast, G·AT has not previously been noted as a stable triplet. Surprisingly, recognition of the central AT base pair with T does not generate a particularly stable triplex, even though it is formed in the context of other T·AT triplets. At 37 °C, the order of stability is modified with the most stable interaction evident with A_T (0.3 μM). T_T, G_T, and C_T all footprint at 1.0 μM, while the least stable triplex is seen with C_T, which only produces a footprint at 3 μM. As shown with the N_A oligonucleotides, at 37 °C, all four bases in the third strand produce a stable complex opposite an AT base pair. In view of the temperature dependence of the stringency seen for the N_T oligonucleotides at this target site, we checked whether these differences might be affected by the slow rate of triplex formation. We find that after 6 days incubation the patterns produced at 37 °C are the same as those shown in Figure 5 (overnight incubation). After 6 days incubation at 4 °C, T_T still showed the weakest interaction, requiring 3 μM oligonuceotide to produce a footprint. However, the rank order of the other oligonucleotides was closer to that observed at 37 °C: A_T was the best (0.1 μM), while C_T and G_T showed intermmediate affinity (0.3−1 μM). These temperature-dependent effects are considered further under Discussion.

pSCAP-C Target Sequence. Figure 6 shows DNase I digestion of the fragment derived from pSCAP-C in the presence of the GA-containing oligonucleotides (N_A) at three temperatures. This fragment has a cytosine at the center, interrupting the block of purines. Previous studies have suggested that in an antiparallel triplex C can be recognized by T forming the T·CG triplet (Beal & Dervan, 1992; Dittrich *et al.*, 1994; Durland *et al.*, 1994). In view of the unusual interaction with G_A and G_T (see below), the results are presented for three different temperatures. At 4 °C, both strands containing a central pyrimidine (C_A and T_A) produce clear footprints at 3 μM. The 3′ (lower) boundary of these footprints is coincident with the edge of the target site, while the 5′ (upper) boundary extends a few bases beyond the target site, as noted above for other complexes. No interaction is seen with A_A. The third strand with a central guanine (G_A) also produces a footprint which is evident at concentrations as low as 1−3 μM. However, the location of this footprint is not the same as that observed with the other oligonucleotides; cleavage products are clearly evident within the upper (5′) half of the target site, while the footprint extends further in the lower (3′) direction. It appears that this oligonucleotide has bound in a different position, covering only the 3′-end of the purine target strand. A similar result is seen at 20 °C; the oligonucleotides containing central pyrimidines (C_A and T_A) produce footprints at similar concentrations (1 μM), while the A-containing strand A_A only protects from cleavage at the highest concentration (10 μM). Each of these

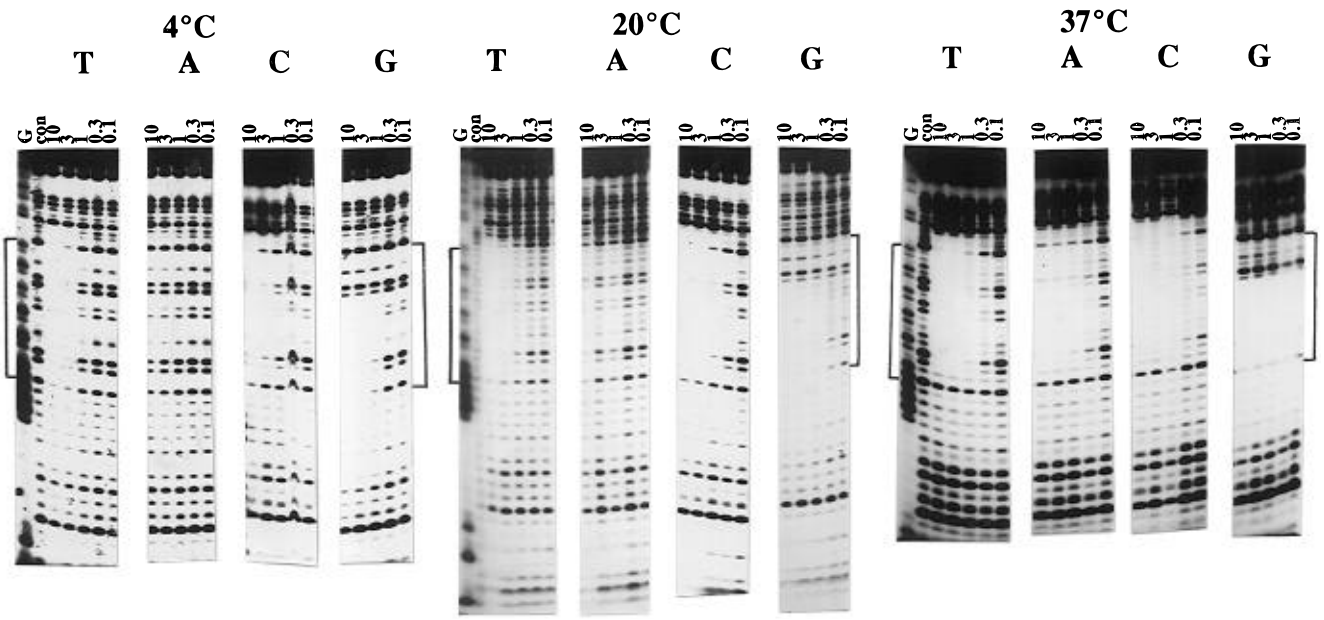

FIGURE 6: DNase I digestion of the fragment containing the target site $(GGA)_2GGC(GGA)_2GG$ in the absence (con) and presence of N_A oligonucleotides at 4, 20, and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. Tracks labeled "G" represent Maxam−Gilbert dimethyl sulfate−piperidine markers specific for guanine.

footprints is centered around the intended target site. Once again, G_A produces a clear footprint with oligonucleotide concentrations as low as 1 μM, but which only covers the 3′ (lower) portion of the target site; DNase I cleavage products are clearly visible in the upper part of the target site. For those oligonucleotides which generate footprints at the correct site, it appears that T·CG and C·CG are more stable than A·CG. At 37 °C, T_A and C_A again give equally stable complexes, footprinting at concentrations as low as 0.3−1 μM. At this temperature, A_A also appears to form a stable complex, with a similar concentration dependence. Once again, the complex with G_A produces a footprint, stable at concentrations of less than 0.1 μM, which does not include the upper (5′) portion of the target site and which extends for several bases in the lower (3′) direction.

Figure 7 shows the results of similar experiments performed with the GT-containing oligonucleotides (N_T). At 4 °C, clear footprints are produced by T_T, A_T, and G_T, while C_T only attenuates cleavage at the higher concentrations. This is in contrast to the results with the N_A oligonucleotides, for which A_A gave the weakest interaction at this temperature. T_T and A_A produce footprints at 3 μM which extend by a few bases above (5′) the target site. The oligonucleotide with a central guanine (G_T) produces a footprint at 10 μM, which has slipped off the target site in a similar way to that observed above with G_A. Bands are still evident within the upper portion of the target site, and the footprint extends by several bases in the lower (3′) direction. A similar pattern is seen at 37 °C. C_T has little effect on the cleavage pattern, while A_T and T_T produce footprints down to about 1 μM, which cover the entire target site. Once again the footprint with G_T, evident at 3 μM oligonucleotide, appears to have slipped off the target site, leaving cleavage products within the upper (5′) end of the target site. For the oligonucleotides which footprint at the correct site, the relative order of stability is A·CG = T·CG > C·CG.

pSCAP-T Target Sequence. Figure 8 shows DNase I digestion of the fragment derived from pSCAP-T in the presence of the GA-containing oligonucleotides (N_A). This fragment has a thymidine in the center, interrupting the block of purines. Since there have been no previous claims for recognition of T within an antiparallel triplex by natural bases, we expect this to be the poorest target site. At 4 °C, C_A shows the best interaction, producing a footprint at a concentration of 1.0 μM. No interaction is seen with T_A, while A_A only produces attenuated cleavage at the highest oligonucleotide concentration (10 μM). G_A also protects from DNase I cleavage, producing a footprint at 3 μM, though once again this seems to lie off the target site. With this oligonucleotide, cleavage products are still evident in the lower (3′) portion of the target site, and the footprint extends further at the upper (5′) end. It should be noted that this target site is positioned in the opposite orientation to the other sequences, labeling the pyrimidine, rather than the purine, strand. This footprint has therefore slipped in the same direction, relative to the purine strand of the target, as those described for G_A and G_T with pSCAP-C. Similar patterns of protection are observed at 20 °C (not shown), confirming the shifted position of the footprint with G_A. At 37 °C, the most stable interaction is seen with C_A, producing a footprint down to 0.3 μM oligonucleotide; T_A only protects at 3 μM, while A_A has intermediate stability, protecting at 1 μM. With G_A, which produces clear footprints down to 0.3 μM, bands are once again visible within the lower (3′) half of the target site, though in this case the footprint does not seem to extend any further in 5′ (upper) direction. Some bands are also present in the lower part of the target site with the other oligonucleotides (especially T_A), though these

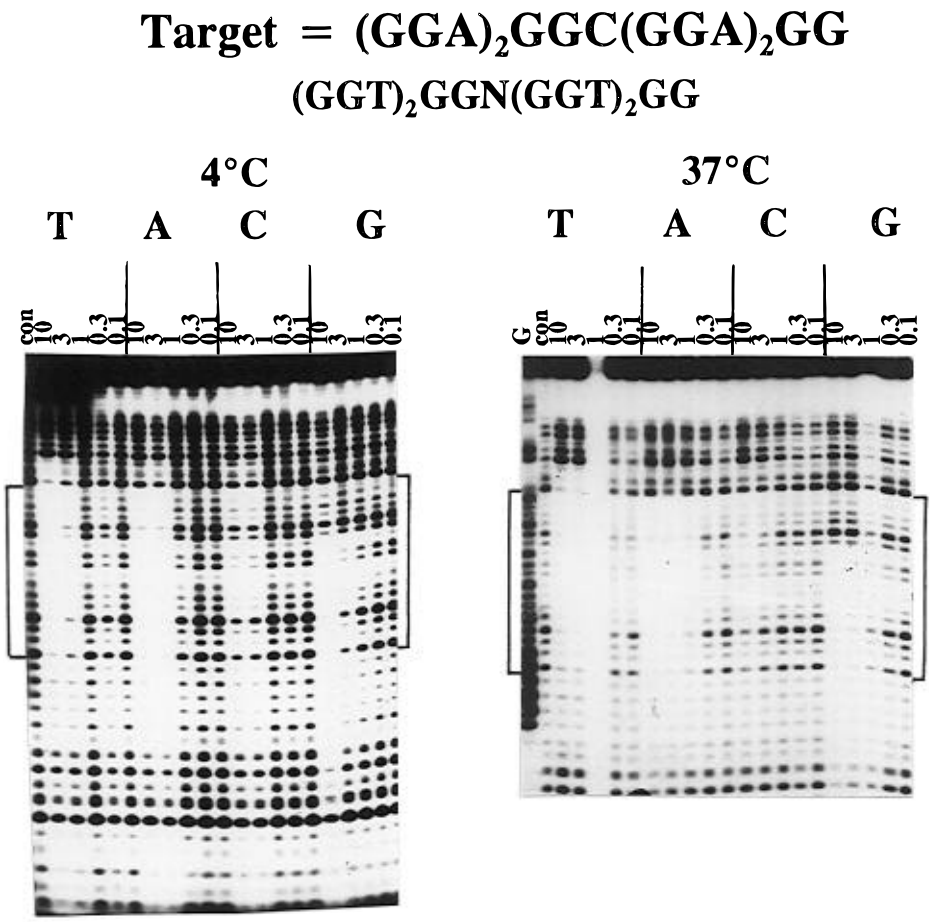

FIGURE 7: DNase I digestion of the fragment containing the target site $(GGA)_2GGC(GGA)_2GG$ in the absence (con) and presence of N_T oligonucleotides at 4 and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the purine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. Tracks labeled "G" represent Maxam−Gilbert dimethyl sulfate−piperidine markers specific for guanine.

are less clear than with G_A. For those oligonucleotides which footprint at the correct target site, the relative order of stability is C·TA > A·TA > T·TA.

Figure 9 shows the results of similar experiments performed with the GT-containing oligonucleotides (N_T) at this target site. Few changes in the DNase I cleavage patterns are evident at 4 °C with any of the oligonucleotides, though there is some attenuation at the highest concentrations. At 37 °C, no interaction is seen with T_T, while A_T and C_T produce clear footprints at 3−10 μM, which are centered around the intended target site. G_T produces the most stable complex, producing a clear footprint at 0.3 μM. Once again, this footprint is shifted off the target site, protecting more bands above the insert and showing clear cleavage products in the lower portion of the target site. For those oligonucleotides which footprint at the correct target site, the relative order of stability is C·TA = A·TA ≫ T·TA.

We further confirmed the position of these footprints using the same fragment, labeled at the 3′-end of the *Eco*RI site (not shown), labeling the purine-containing strand. We find that A_T, C_T, and T_T produce footprints at similar concentrations to those observed on the pyrimidine-labeled strand, which terminate at the 3′-boundary of the target site and which extend by 2−3 bases in the 5′ (upper) direction. In contrast, G_T generates a footprint which extends below (3′) the target site, and in which 5−6 bands are evident within the upper (5′) end of the target site.

DISCUSSION

Binding Affinity. Before discussion of the effect of various central triplet combinations on the antiparallel triplexes studied in this paper, it should be noted that all the oligonucleotides produce footprints at each target at concentrations of 10 μM, and many complexes are still stable at 1.0−3.0 μM, even for those containing non-standard central triplets. This is in marked contrast to the results for parallel triple helix formation at A_8XA_8 sites using T_8XT_8 oligonucleotides which showed that a mismatch in the center of the third strand could totally abolish triplex formation, even at oligonucleotide concentrations as high as 100 μM (Chandler & Fox, 1993). This lower stringency and tighter binding could be a general property of antiparallel triplexes, but may reflect the greater stability of the antiparallel G·-GC over parallel T·AT triplets. It is also worth noting that the lowest oligonucleotide concentrations required to generate footprints in this work are about 0.1 μM, suggesting a binding constant of around 10^{-7} M, which is a fairly low affinity for purine-rich third strands. There are several possible explanations for this apparently lower affinity. First, our reaction conditions do not include spermine, though we do have a relatively high concentration (5 mM) of magnesium. Second, with the exception of the interaction of G_A and G_T with pSCAP-G, none of the complexes used in this work contain more than two adjacent G·GC triplets, in contrast to other studies which generally contain longer G-tracts. If antiparallel triplex formation is dominated by the formation of G·-

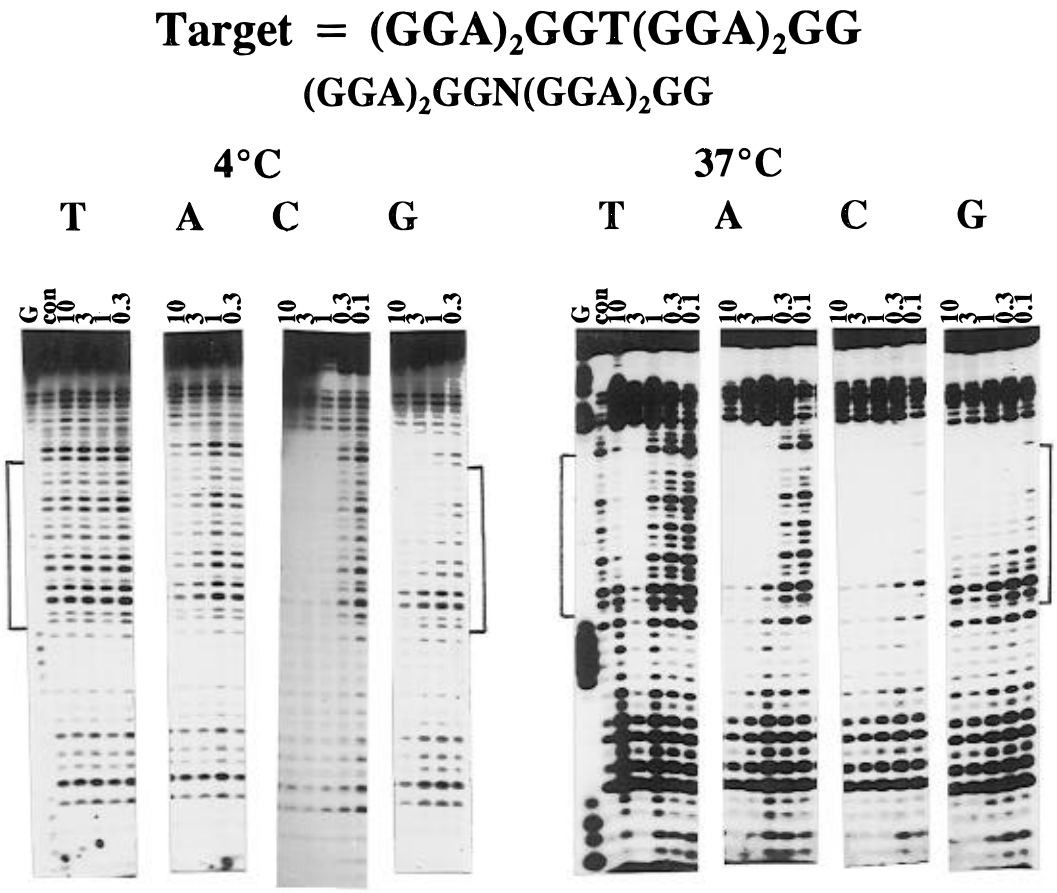

FIGURE 8: DNase I digestion of the fragment containing the target site $(GGA)_2GGT(GGA)_2GG$ in the absence (con) and presence of N_A oligonucleotides at 4 and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the pyrimidine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. Tracks labeled "G" represent Maxam−Gilbert dimethyl sulfate−piperidine markers specific for guanine.

GC triplets, with other triplets such as A·AT and T·AT playing a lesser role, then our complexes will be weaker. This weaker binding means that these other triplets are not overwhelmed by the G·GC triplets, enabling us to study their properties in greater detail. Third, it is possible that DNase I footprinting underestimates binding constants compared with other techniques such as affinity cleavage or gel retardation.

It might be argued that footprints generated by the oligonucleotides used in the present study are not due to recognition of the entire 17 base pair target site, but represent interaction of half of the oligonucleotide with a part of the target sequence on either side of the variable central base pair. However, this seems unlikely since it does not explain the range of stabilities observed between oligonucleotides which differ in the nature of the central base, which would not be involved in such half-site recognition.

Sequence Recognition. These results clearly demonstrate that, within these sequences, recognition of a GC base pair at the center of the purine tract is best using a third strand guanine, generating the standard antiparallel G·GC triplet. This is seen with both G_A and G_T oligonucleotides. Within the N_A strands, the relative order of stability is G·GC > A·GC = C·GC > T·GC, while the order for N_T strands is G·GC ≥ A·GC > C·GC = T·GC. The apparent stability of A·GC may be the result of the formation of a specific triplet, as previously suggested (Beal & Dervan, 1992; Malkov *et al*., 1993), but could also result from greater stacking of purines with the adjacent third strand guanines. The weakest interaction is seen with the oligonucleotide containing a central T. The slightly different order of stability for the complexes formed with N_A and N_T oligonucleotides at this and other target sites is surprising since the central base mismatch is flanked on each side by two G·GC triplets in each case. This may suggest that the conformation of antiparallel triplexes formed with GA-containing oligonucleotides is not the same as with GT. It has been suggested that GT-rich oligonucleotides can bind either parallel or antiparallel to the duplex purine strand, depending on the number of GpT and TpG steps (Giovannangeli *et al*., 1992). However, this seems unlikely in the present sequences for the following reasons. First, Giovannangeli *et al*. (1992) reasoned that, since G·GC and T·AT triplets are not isohelical, the antiparallel configuration should be preferred with increasing numbers of GpA and ApG steps in the target. The sequences used in this work contain four GpA and four ApG steps, which should strongly favor an antiparallel orientation for GT-containing oligonucleotides. Other recent work has shown that GT-containing oligonucleotide can bind in an antiparallel but not parallel orientation (Escudé *et al.*, 1996). Second, GA-containing oligonucleotides must bind in an antiparallel orientation, since it is not possible to generate a parallel A·AT triplet; the observation that GA- and GT-rich oligonucleotides produce very similar footprints at each target site provides circumstantial evidence that the latter also bind in an antiparallel configuration. Third, if the N_T oligonucleotides bound in a parallel configuration, then C_T and T_T would be expected to produce unusually

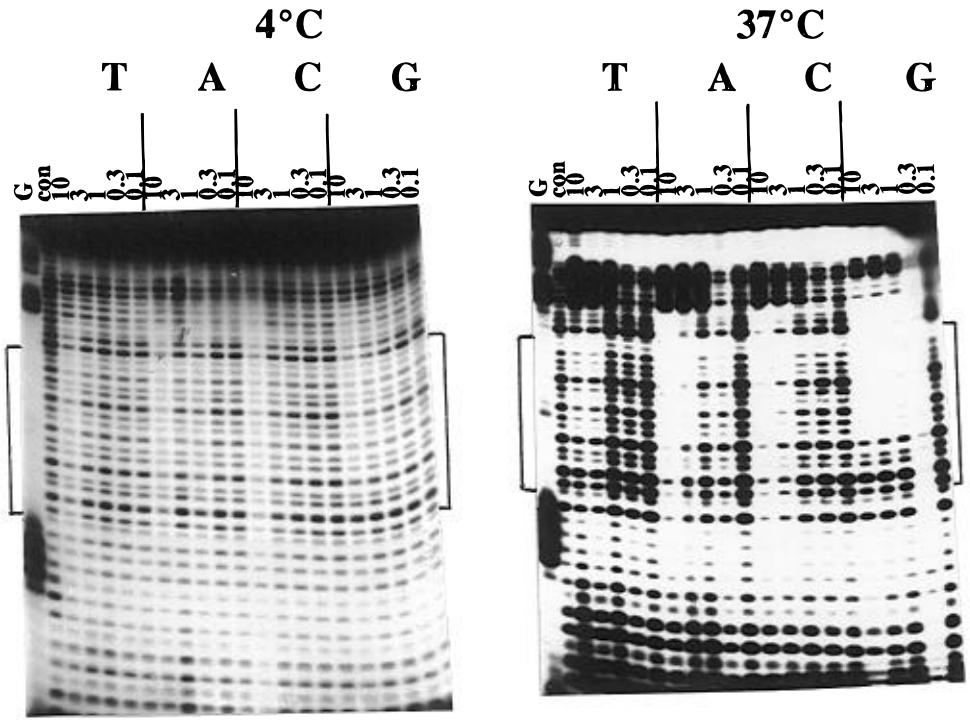

FIGURE 9: DNase I digestion of the fragment containing the target site $(GGA)_2GGT(GGA)_2GG$ in the absence (con) and presence of N_T oligonucleotides at 4 and 37 °C. The identity of the base in the center of the third strand (T, A, C, G) is indicated above each set of lanes. The DNA was labeled at the 3′-end of the *Hin*dIII site, visualizing the pyrimidine-containing strand. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. The square brackets indicate the position of the triplex target site. Tracks labeled "G" represent Maxam−Gilbert dimethyl sulfate−piperidine markers specific for guanine.

stable complexes with pSCAP-G and pSCAP-A, since they would contain standard C·GC and T·AT central triplets. Although we cannot prove the orientation of these oligonucleotides, the data strongly suggest the predicted antiparallel configuration.

The results suggest that all four N_A oligonucleotides can form a stable complex with the target containing a central AT base pair, though T_A is slightly better at higher temperatures. This apparent lack of stringency is surprising, but suggests that formation of G·GC triplets is the strongest contributory factor to the stability of these complexes. Other bases at the center do not contribute much to the overall binding energy. In this context, N·AT neither contributes nor hinders the interaction and might be regarded as a null-triplet. Recognition of a central adenine with N_T oligonucleotides is less straightforward. At low temperatures, T_T produces the least stable complex while the rank order of the other oligonucleotides alters with prolonged incubation times. The results at 4 °C should therefore be interpreted with caution (see below). The stronger affinity of A_T than T_T might be explained by invoking the importance of base stacking in complex stability, while the relatively strong interaction with guanines might arise from some intrastrand interaction between the block of five adjacent third strand guanines. At 37 °C, A·AT and T·AT both form stable complexes, with A·AT being slightly better. Once again, G·AT is an unusually stable interaction. Since stacking interactions should be less important at elevated temperatures, the increased stability of T·AT relative to G·AT is consistent with the suggestion that only the former has a significant contribution from the formation of specific hydrogen bonds. It should be noted that, under these conditions, all four oligonucleotides produce stable complexes as noted above for N_A. The ability of all four bases to form stable complexes at a central AT base pair is consistent with previous observations (Beal & Dervan, 1991) which showed that the relative stability of complexes containing three N·AT triplets was 1:0.56:0.16 for T·AT:A·AT:C·AT.

For recognition of a central CG base pair, the order of stability is T·CG = C·CG > A·CG for N_A strands and T·CG = A·CG ≫ C·CG for N_T. Since both G_A and G_T produce footprints which have slipped relative to the intended target site, we can assume that a central G·CG is so unstable that the oligonucleotide avoids this mismatch and binds in an alternative location (see below). The formation of an antiparallel T·CG triplet has been previously suggested (Beal & Dervan, 1992; Durland *et al.*, 1994; Dittrich *et al.*, 1994; Ji *et al.*, 1996). In most of these studies, the T·CG was located at the center of at least four G·GC triplets, in a similar position to that used in this paper. Although Durland *et al.* (1994) suggested that the T·CG triplet was only slightly weaker than the standard antiparallel T·AT triplet, we find that lower concentrations of both T_A and T_T are required to generate footprints on the fragment containing a central A (Figures 4 and 5) than a central C (Figures 6 and 7). However, since T·CG is only marginally more stable than A·CG (especially for N_A), it is an open question as to whether this triplet involves the formation of specific contacts

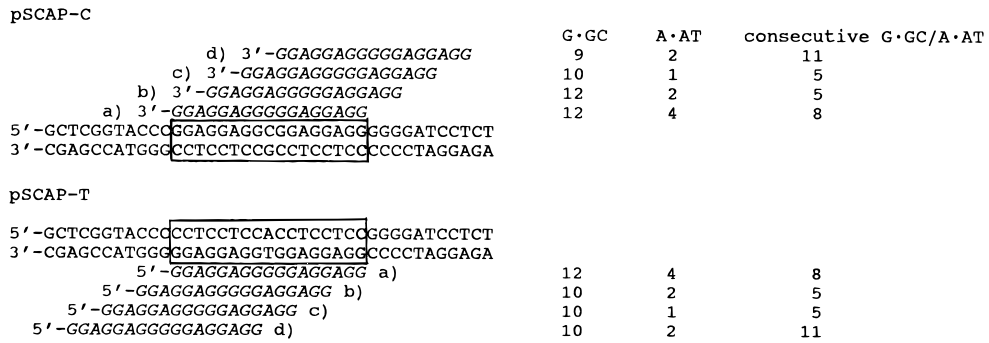

FIGURE 10: Possible locations for G_A oligonucleotides on the fragments derived from pSCAP-C and pSCAP-T. The target sites are boxed, and the third strands are shown in italics. Locations a)−d) represent slippage of the third strand by successive three base steps in the 3′-direction relative to the duplex purine strand. Position a) shows the expected location, while position d) is consistent with the footprinting data. Also shown are the number of antiparallel G·GC and A·AT triplets that would be formed in each location, together with the longest contiguous stretch of such triplets.

between the third strand T and the CG base pair, or is merely the best null base at this position. NMR studies by Dittrich *et al*. (1994) showed that the T was in equilibrium between a nonbonded form and a conformation involving a single hydrogen bond between T(O4) and C(NH_2). In addition, the G·GC triplet on the 3′ (but not 5′) side of the third strand T was distorted, confirming that T·CG does not fit easily within a standard antiparallel structure.

For recognition of a TA base pair at the center of the homopurine site, both N_T and N_A give the relative order of stability as C·TA > A·TA > T·TA. As noted for the central CG base pair, G_T and G_A generate footprints at unusual locations, suggesting that the formation of G·TA is avoided; it is therefore not possible to comment on the relative stability of the G·TA triplet. There have been no previous reports of successful antiparallel triplex formation across a TA inversion using natural bases. The present results suggest that the most stable complex is generated with a third strand C, while T is the least stable.

In summary, although the exact binding properties depend on the type of oligonucleotide N_A or N_T, the most stable interactions in this sequence context involve recognition of GC by G or A, recognition of AT by T, A, or G, recognition of CG by T or A, and recognition of TA by C.

Comparison of N_A and N_T. In general, N_A third strands require lower concentrations to generate footprints than N_T, though the perfect match third strands show equivalent stability for targets consisting of an uninterrupted run of purines. N_A strands also appear to be more stringent than N_T, showing a wider range of oligonucleotide concentrations required to produce a footprint. It is worth noting that previous studies on antiparallel mismatches have only used GT-containing oligonucleotides (Beal & Dervan, 1992; Durland *et al*., 1994; Dittrich *et al*., 1994). This may be due to differences in deformation of the third strand backbone, which is less for a GpA step than GpT (Sun *et al*., 1991), and that G·GC is more closely isomorphous to A·AT than reverse Hoogsteen T·AT (Radhakrishnan *et al*., 1993).

Effect of Temperature. The results presented in this paper show that these antiparallel complexes are generally more stable at higher temperatures, requiring lower oligonucleotide concentrations to produce footprinting patterns. This is opposite to the effect of temperature on the parallel motif (Chandler & Fox, 1993) and is counter to our expectation. By comparison with duplex stability, we would expect lower temperatures to decrease stringency, and increase binding strength. A possible explanation for this phenomenon is that higher temperatures destabilize any secondary structures in the third strand, increasing its effective concentration, thereby pushing the equilibrium toward triplex formation. Since the rate of triplex formation is known to be slow, with bimolecular association rate constants of about 10^3 M^{-1} s^{-1} (Blake & Fresco, 1966; Rougée *et al*., 1992; Shindo *et al*., 1993), it is worth considering whether the complexes have attained equilibrium. If we assume that the association rate constant is 10^3 M^{-1} s^{-1}, then, with 0.1 μM oligonucleotide, the half-life for association will be about 2 h, sufficient for complete equilibration during the overnight incubation. It might also be argued that the rate of association will be much slower at lower temperatures (4 °C) and that the complexes may have not reached equilibrium. Indeed, the results obtained with pSCAP-G and N_T at long incubation times confirm a very slow rate of reaction. Several studies have shown that the formation of parallel triple helices displays a negative energy of activation; i.e., the reaction is faster at lower temperatures (Rougée *et al*., 1992; Shindo *et al*., 1993). This is explained by the nucleation−zipper model for triplex formation in which complexes are initiated from a short (4−5 base) nucleation zone; lower temperatures stabilize this intermediate and thereby speed up the rate of complex formation. However, one report suggests that antiparallel complexes do not show the same temperature dependence and are slower at lower temperatures (Svinarchuk *et al*., 1994). As a result, the unusual differences at 4 °C should be interpreted with caution.

In some cases, the stringency of interaction is also affected by temperature. For example, at 4 °C the target site with a central AT (Figure 5) shows the weakest interaction with T_T whereas on increasing the temperature to 37 °C this forms one of the more stable complexes. However, this change in relative stability may also reflect the different stabilities of secondary structures adopted by the various oligonucleotides. In general, stringent recognition of homopurine target sites is less temperature-sensitive than those targets containing a central pyrimidine. This may reflect the lower stability of antiparallel X·YR triplets. N_A strands are also less affected by the changes in temperature than N_T.

Slippage. Both G_A and G_T produce footprints, on the fragments containing central pyrimidines, which have slipped with respect to the intended target site by about 6−9 base pairs. In each case, DNase I cleavage products are still evident within one end of the target site. This slippage is in the 3′-direction for the CG-containing fragment and in the

5′-direction for TA. Since these two target sites are inserted in opposite orientations, visualizing the purine strand for CG, but the pyrimidine strand for TA, the slippage is in the same direction with respect to the purine strands. Whatever may be the nature of the slipped structure, it must generate a more stable complex than the intended one with a single central G·YR mismatch.

DNase I is not a good probe for assessing ligand binding site sizes, on account of its size and because its cleavage pattern is staggered by about three bonds in the 3′-direction across the two strands (Drew & Travers, 1984). In addition, interpreting DNase I footprints of triple helix formation is not simple, since the enzyme cuts from the DNA minor groove, while the third strand is located in the major groove. The protections must arise from changes in DNA structure and/or rigidity. However, it is clear that the oligonucleotides containing a central guanine produce footprints at very different positions to the other oligonucleotides. Any explanation for this phenomenon must justify why it only occurs with the oligonucleotides with central guanines (not A, C, or T) and why it is only found at the target sites with central pyrimidines (not G or A).

In order to explore the origin of the slipped footprints, we have examined the features of various other locations for the third strand. These are illustrated in Figure 10 in which the target site is boxed and the third strands are shown in italics. Position a) shows the predicted location containing 16×G·GC and A·AT triplets with a G·YR mismatch in the center. In the other three configurations, the position of GGA opposite (GGA·TCC) is retained, and the third strand is moved by three base pairs in the 3′-direction along the duplex purine strand, overlapping with the remainder of the polylinker fragment. Position b) contains three mismatched triplets (A·CG, G·AT, and A·GC) located such that there are no more than five consecutive G·GC and A·AT triplets. Similarly, position c) contains six mismatches and has no more than five consecutive G·GC and A·AT triplets. In contrast, although position d) also has 6 mismatches, these are located at the 5′-end of the oligonucleotide, and the complex contains a run of 11 contiguous G·GC and A·AT triplets. We favor this complex d) as the most likely, which is consistent with the observed cleavage patterns. It appears that this slipped triplex is stabilized because cloning the target sequence into the *Sma*I site (CCC/GGG) has generated a run of five guanines, which exactly matches the sequence in the center of the G_A and G_T oligonucleotides. It therefore appears that the slipped structure, containing 11 contiguous triplets, is more stable than the intended structure containing a central G·YR mismatch with 8 contiguous triplets on either side. This interaction can only occur with oligonucleotides containing a central guanine, since these are the only ones that possess five contiguous guanines. The observation that no such slippage is observed with the target sites containing central G or A confirms that G·GC and G·AT triplets are more stable than G·CG and G·TA.

Summary. Successful triplex formation has been demonstrated at homopurine target sites interrupted at the central position by each of the natural base pairs, using GT- and GA-rich oligonucleotides. As well as the standard G·GC, A·AT, and T·AT triplets, we have observed the formation of A·GC and G·AT. The CG base pair is recognized by T but can also form stable complexes with A or C in G_T or G_A oligonucleotides, respectively. The TA base pair forms the most stable complexes with C. Nevertheless, each oligonucleotide produced a footprint on all the target sites at a concentration of 10 μM at 37 °C. A G·YR triplet at the center of the complex is unstable, and the oligonucleotide binds to an alternative site, forming fewer triplets, but containing a longer stretch of contiguous canonical triplets.

REFERENCES

Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509−521.

Beal, P. A., & Dervan, P. B. (1991) *Science 251*, 1360−1363.

Beal, P. A., & Dervan, P. B. (1992) *Nucleic Acids Res. 20*, 2773−2776.

Blake, R. D., & Fresco, J. R. (1966) *J. Mol. Biol. 19*, 145−160.

Chandler, S. P., & Fox, K. R. (1993) *FEBS Lett. 332*, 189−192.

Chubb, J. M., & Hogan, M. E. (1992) *Trends Biotechnol. 10*, 132−136.

Dittrich, K., Gu, J., Tinder, R., Hogan, M., & Gao, X. (1994) *Biochemistry 33*, 4111−4120.

Drew, H. R., & Travers, A. A. (1984) *Cell 37*, 491−502.

Durland, R. H., Kessler, D. J., Gunnell, S., Duvic, M., Petitt, B. M., & Hogan, M. E. (1991) *Biochemistry 30*, 9246−9255.

Durland, R. H., Rao, T. S., Revankar, G. R., Tinsley, J. H., Myrick, M. A., Seth, D. M., Rayford, J., Singh, P., & Jayaraman, K. (1994) *Nucleic Acids Res. 22*, 3233−3240.

Escudé, C., Sun, J. S., Nguyen, C. H., Bisagni, E., Garestier, T., & Hélène, C. (1996) *Biochemistry 35*, 5735−5740.

Felsenfeld, G., Davies, D., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023−2024.

Fox, K. R. (1995) *FEBS Lett. 357*, 312−316.

Giovannangeli, C., Rougée, M., Garestier, T., Thuong, N. T., & Hélène, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8631−8635.

Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967−971.

Hélène, C. (1991a) *Anti-Cancer Drug Des. 6*, 569−584.

Hélène, C. (1991b) *Eur. J. Cancer 27*, 1466−1471.

Hélène, C. (1993) *Curr. Opin. Biotechnol. 4*, 29−36.

Ji, J., Hogan, M. E., & Gao, X. L. (1996) *Structure 4*, 425−435.

Kiessling, L. L., Griffin, L. C., & Dervan, P. B. (1992) *Biochemistry 31*, 2829−2834.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids. Res. 15*, 7749−7760.

Malkov, V. A., Voloshin, O. N., Veselkov, A. G., Rostapshov, V. M., Jansen, I., Soyfer, V. N., & Frank-Kamenetskii, M. D. (1993) *Nucleic Acids Res. 21*, 105−111.

Moffat, A. S. (1991) *Science 252*, 1374−1375.

Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645−650.

Radhakrishnan, I., & Patel, D. J. (1993) *Structure 1*, 135−152.

Radhakrishnan, I., & Patel, D. J. (1994) *Structure 2*, 17−32.

Radhakrishnan, I., de los Santos, C., & Patel, D. J. (1993) *J. Mol. Biol. 234*, 188−197.

Rougée, M., Faucon, B., Mergny, J. L., Barcelo, F., Giovannangeli, C., Garestier, T., & Hélène, C. (1992) *Biochemistry 31*, 9269−9278.

Shindo, H., Torigoe, H., & Sarai, A. (1993) *Biochemistry 32*, 8963−8969.

Sun, J.-S., & Hélène, C. (1993) *Curr. Opin. Struct. Biol. 3*, 345−356.

Sun, J. -S., De Bizemont, T., Duval-Valentin, G., Montenay-Garestier, T., & Hélène, C. (1991) *C. R. Acad. Sci. Paris, Ser. III 313*, 585−590.

Svinarchuk, F., Bertrand, J.-R., & Malvy, C. (1994) *Nucleic Acids Res. 22*, 3742−3747.

Thuong, N. T., & Hélène, C. (1993) *Angew. Chem. 32*, 666−690.

Vasquez, K. M., Wensel, T. G., Hogan, M. E., & Wilson, J. H. (1995) *Biochemistry 34*, 7243−7251.

Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., & Weis, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3840−3844.

BI9609679