Nucleic Acids Research, 1994, Vol. 22, No. 15 3233–3240

Binding of T and T analogs to CG base pairs in antiparallel triplexes

Ross H.Durland+, T.Sudhakar Rao, Ganapathi R.Revankar, John H.Tinsley, Melissa A.Myrick, Dale M.Seth, Jamie Rayford, Pradeep Singh and Krishna Jayaraman*
Triplex Pharmaceutical Corporation, 9391 Grogan's Mill Road, The Woodlands, TX 77380, USA

Received December 3, 1993; Revised May 31, 1994 and Accepted June 8, 1994

ABSTRACT

The goal of this study was to address antiparallel triplex formation at duplex targets that do not conform to a strict oligopurine·oligopyrimidine motif. We focused on the ability of natural bases and base analogs incorporated into oligonucleotide third strands to bind to so-called CG inversions. These are sites where a cytosine base is present in an otherwise purine-rich strand of a duplex target. Using a 26-base-triplet test system, we found that of the standard bases, only thymine (T) shows substantial binding to CG inversions. This is qualitatively similar to the report of Beal and Dervan [*Science* (1991), 251, 1360–1363]. Binding to CG inversions was only slightly weaker than binding to AT base pairs. Binding of T to CG inversions was also evaluated in two other sequences, with qualitatively similar results. Six different analogs of thymine were also tested for binding to CG inversions and AT base pairs. Significant changes in affinity were observed. In particular, 5-fluoro-2′-deoxyuridine was found to increase affinity for CG inversions as well as for AT base pairs. Studies with oligonucleotides containing pyridin-2-one or pyridin-4-one suggest that thymine O^4 plays a critical role in the T·CG interaction. Possible models to account for these observations are discussed.

INTRODUCTION

Triple helix formation by oligonucleotides has been an area of intense investigation since it was first demonstrated in 1987 (1). Numerous investigators have shown that under proper conditions, defined oligonucleotides can bind in the major groove of duplex DNA to form a triplex (2; and references therein). Binding is sequence specific and results from the formation of hydrogen-bonded base triplets. While several approaches to triplex formation have been documented, we have chosen to focus on so-called antiparallel triplexes. In this approach, the third strand oligonucleotide binds to the purine strand of a target duplex to form reverse-Hoogsteen G·GC and T·AT (or A·AT) base triplets (3–5). The orientation of the third strand is antiparallel to that of the purine strand of the duplex.

A major limitation of all non-enzymatic triplexes is that bases in the third strand generally hydrogen bond only with purines in one of the two duplex strands. Thus, ideal duplex targets are oligopurine·oligopyrimidine sequences of substantial length (>10 bp). Strict adherence to this requirement would severely limit the range of biologically interesting sequences amenable to triplex formation. Thus, we (and others) have been exploring ways to improve triplex formation at sequences that are not purely oligopurine·oligopyrimidines (6–19). In this report we describe efforts to develop improved antiparallel triplex formation at targets containing isolated cytosine residues in an otherwise homopurine sequence. We refer to such sites as CG inversions.

In describing base triplets, we will employ a convention in which the third strand base is given first, separated from the duplex base pair by a bullet (e.g. T·AT). The first base of the duplex base pair is that of the purine-rich strand, and is generally the base with which the third strand base interacts through hydrogen bonding. For simplicity, we omit the letter d, indicating a deoxyribonucleoside, prior to the base. However, in all cases the oligonucleotides studied here consist entirely of 2′-deoxyribonucleotides.

EXPERIMENTAL

Nucleoside phosphoramidite synthesis

Pyridin-2-one 1-[2-deoxy-5-O-(4,4′-dimethoxytrityl)-β-D-erythro-pentofuranosyl)pyridin-2-one (**1**). 3.3 g (15.62 mmol) of 1-(2-Deoxy-β-D-*erythro*-pentofuranosyl)pyridin-2-one (pyridin-2-one deoxyribonucleoside; Figure 7 and ref. 20) was dried by coevaporation with dry pyridine (2×20 mL). It was then dissolved in dry pyridine (35 mL) followed by addition of 4,4′-dimethoxytritylchloride (6.6 g, 19.5 mmol). The mixture was stirred at room temperature for 3 h, diluted with CH_2Cl_2 (150 mL), and washed with water (25 mL). The organic layer was dried (Na_2SO_4) and evaporated in vacuo. The residue was coevaporated with toluene (2×10 mL) to remove traces of pyridine and purified by silica gel column (2.5×25 cm)

*To whom correspondence should be addressed

+Present address: GeneMedicine Inc., 8080 N. Stadium Drive, Suite 2100, Houston, TX 77054-1823, USA

chromatography using a gradient of 0–3% MeOH in CH_2Cl_2 to yield 7.12 g (88.78%) of pure **1**, mp 86–88°C. Ir (KBr): υ 1655 (C=O), 3260 (OH) cm^{-1}. ^1H nmr (DMSO-d_6): δ 2.04 (m, 1 H, $C_{2'}H$), 2.25 (m, 1 H, $C_{2''}H$), 3.26 (d, 2 H, $C_{5'}H_2$), 3.74 (s, 6 H, 2 OCH_3), 3.97 (m, 1 H, $C_{4'}H$), 4.30 (m, 1 H, $C_{3'}H$), 5.36 (d, 1 H, $C_{3'}OH$), 6.06 (t, 1 H, J = 6.24 Hz, $C_{1'}H$), 6.35 (m, 2 H, C_3H, C_5H), 6.90 (d, 4 H, ArH), 7.23–7.40 (m, 10 H, ArH, C_4H), and 7.77 (dd, 1 H, C_6H). *Anal.* Calc'd. for $C_{31}H_{31}NO_6 \cdot 0.5H_2O$ (522.60): C, 71.25; H, 6.17; N, 2.68. Found: C, 71.11; H, 6.05; N, 2.68.

1-[2-Deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pentofuranosyl]pyridin-2-one-3'-O-(2-cyanoethyl)-N,N-diisopropylphosphoramidite (**2**). 1.03 g (2.0 mmol) of **1** was dissolved in a mixture of anhydrous CH_2Cl_2 (10 mL) and N,N-diisopropylethylamine (1.44 mL, 8.2 mmol). 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (0.60 mL, 2.77 mmol) was added under an argon atmosphere. After stirring the reaction mixture at room temperature for 20 min, it was diluted with ethyl acetate (100 mL) and the solution was washed with saturated, aqueous $NaHCO_3$ solution (30 mL). The organic layer was separated, dried (Na_2SO_4), and evaporated in vacuo. The residue was chromatographed on a silica gel column (2×20 cm) which was packed in a mixture of CH_2Cl_2:EtOAc:Et_3N (68:30:1). The product was eluted using the same solvent system. Fractions containing the product were evaporated and the residue was dissolved in CH_2Cl_2 (5 mL). The solution was added dropwise to cold (< −50°C), stirred pentane. The supernatant was decanted and the precipitate was dried under high vacuum to give 1.2 g (84.6%) of analytically pure **2**. ^{31}P nmr (CD_3CN): δ 149.34. ^1H nmr (CD_3CN): δ 1.10 [m, 12 H, 2 $NCH(CH_3)_2$], 2.20 (m, 1 H, $C_{2'}H$), 2.55 (m, 3 H, $C_{2''}H$, OCH_2CH_2CN), 3.35 [m, 2 H, 2 $NCH(CH_3)_2$], 3.65 (m, 4 H, $C_{5'}H_2$, OCH_2CH_2CN), 3.75, 3.76 (2s, 6 H, 2 OCH_3), 4.15 (m, 1 H, $C_{4'}H$), 4.57 (m, 1 H, $C_{3'}H$), 6.03 (t, 1 H, J = 6.96 Hz, $C_{1'}H$), 6.38 (m, 2 H, C_3H, C_5H), 6.87 (m, 4 H, ArH), 7.20–7.46 (m, 10 H, ArH, C_4H), and 7.77 (dd, 1 H, C_6H). *Anal.* Calc'd. for $C_{40}H_{48}N_3O_7P$ (713.81): C, 67.30; H, 6.78; N, 5.89; P, 4.34. Found: C, 66.93; H, 6.79; N, 5.89; P, 4.44.

Pyridin-4-one 1-[2-deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pentofuranosyl]pyridin-4-one (**3**). In a manner similar to that described for **1**, 0.6 g (2.84 mmol) of 1-(2-Deoxy-β-D-*erythro*-pentofuranosyl)pyridin-4-one (pyridin-4-one deoxyribonucleoside; Figure 7 and ref. 21) was tritylated with 4,4'-dimethoxytritylchloride (1.25 g, 3.7 mmol) in pyridine to give 0.8 g (54.8%) of **3**, mp 110–112 °C. Ir (KBr): υ 1640 (C=O), 3250 (OH) cm^{-1}. ^1H nmr (DMSO-d_6): δ 2.27 (m, 2 H, $C_{2'}H$, $C_{2''}H$), 3.21 (m, 2 H, $C_{5'}H_2$), 3.74 (s, 6 H, 2 OCH_3), 3.95 (m, 1 H, $C_{4'}H$), 4.35 (m, 1 H, $C_{3'}H$), 5.35 (d, 1 H, $C_{3'}OH$), 5.81 (t, 1 H, J = 6.5 Hz, $C_{1'}H$), 5.97 (d, 2 H, C_3H, C_5H), 6.87 (d, 4 H, ArH), 7.23–7.38 (m, 9 H, ArH), and 7.75 (d, 2 H, C_2H, C_6H). *Anal.* Calc'd. for $C_{31}H_{31}NO_6 \cdot 0.5H_2O$ (522.60): C, 71.25; H, 6.17; N, 2.68. Found: C, 71.01; H, 6.15; N, 2.65.

1-[2-Deoxy-5-O-(4,4'-dimethoxytrityl)-β-D-erythro-pentofuranosyl]pyridin-4-one-3'-O-(2-cyanoethyl)-N,N-diisopropylphosphoramidite (**4**). In a manner similar to that described for **2**, phosphitylation of **3** (0.65 g, 1.26 mmol) with 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (0.39 mL, 1.64 mmol) in the presence of N,N-diisopropylethylamine (0.89 mL, 5.04 mmol) in CH_2Cl_2 (10 mL) gave the crude product, which was purified on a silica gel column (2.0×10 cm) packed in a mixture of CH_2Cl_2:EtOAc:Et_3N (45:45:4). The product was eluted using a gradient of a mixture of CH_2Cl_2:EtOAc:Et_3N (45:45:4) to MeOH:CH_2Cl_2:EtOAc:Et_3N (2.5:45:45:4). The appropriate fractions containing the product were evaporated and the residue was dissolved in CH_2Cl_2 (3 mL). The solution was added dropwise to cold (< −50°C), stirred pentane to give 0.79 g (87.49%) of pure **4**. ^{31}P nmr (CD_3CN): δ 149.46. ^1H nmr (CD_3CN): δ 1.06–1.21 [m, 12 H, 2 $NCH(CH_3)_2$], 2.30–2.70 (m, 4 H, $C_{2'}H$, $C_{2''}H$, OCH_2CH_2CN), 3.35 [m, 2 H, 2 $NCH(CH_3)_2$], 3.65 (m, 4 H, $C_{5'}H_2$, OCH_2CH_2CN), 3.76 (s, 6 H, 2 OCH_3), 4.10 (m, 1 H, $C_{4'}H$), 4.65 (m, 1 H, $C_{3'}H$), 5.73 (t, 1 H, J = 6.56 Hz, $C_{1'}H$), 6.03 (m, 2 H, C_3H, C_5H), 6.87 (m, 4 H, ArH), 7.23–7.45 (m, 9 H, ArH), and 7.63 (m, 2 H, C_2H, C_6H). *Anal.* Calc'd. for $C_{40}H_{48}N_3O_7P \cdot 0.5CH_3OH$ (729.83): C, 66.65; H, 6.90; N, 5.76; P, 4.24. Found: C, 66.51; H, 6.63; N, 5.84; P, 4.55.

Oligonucleotide synthesis

DMT-protected 2'-deoxyribonucleoside-3'-phosphoramidites of G, A, T, and C were obtained from MilliGen Corporation. Phosphoramidite derivatives of 2'-deoxyuridine, 2'-deoxy-5-fluorouridine, 2'-deoxy-5-bromouridine, and 2'-deoxy-5-iodouridine were purchased from Glen Research Corporation. Oligonucleotides were synthesized on Applied Biosystems 380B or 394 automated DNA synthesizers on 0.2 or 1 μmole scale. Coupling times were increased to 900 seconds for pyridin-2-one and pyridin-4-one phosphoramidites. Stepwise coupling efficiencies were ≤97% as assessed by dimethoxytrityl cation absorbance. Oligonucleotides were purified by anion exchange HPLC as described by Murphy *et al.* (22), and desalted by membrane filtration or C18 Sep-Pak (Waters). Isolated yields were approximately 20%.

Oligonucleotide purity was evaluated by denaturing polyacrylamide gel electrophoresis and analytical ion-exchange HPLC. In some cases, oligomers were further purified by preparative gel electrophoresis. Oligonucleotides containing pyridin-2-one (Z102-125) or pyridin-4-one (Z102-126) were characterized by nucleoside composition analysis, using P1 nuclease and alkaline phosphatase, followed by reverse phase HPLC. Retention times were compared to purified nucleosides to confirm the identity and relative quantity of all expected nucleosides in the oligomer. For each oligomer, nucleoside peaks corresponding to G, T, and either pyridin-2-one (Z102-125) or pyridin-4-one (Z102-126) were observed (not shown). In addition, digestion of each oligomer resulted in one additional peak. Analysis of the UV absorbance spectra for the additional peaks showed that they exactly matched the spectra obtained by mixing equimolar ratios of 2'-deoxyguanosine and the appropriate pyridine nucleoside. This suggests that the observed peaks are dimers of G and pyridin-2-one or pyridin-4-one, suggesting that P1 nuclease has difficulty in completely digesting oligonucleotides containing these monomers.

Oligonucleotide concentrations were determined spectrophotometrically at 260 nm. Extinction coefficients were estimated according to dinucleotide content, using the formula of Cantor and Tinoco (23), and the dinucleotide extinction coefficients of Alexis (24). For oligonucleotides containing novel nucleosides, extinction coefficients were calculated assuming that the novel nucleoside has the same absorbance properties as T. Since the

novel nucleoside generally accounts for only ~10% of the total bases, this is a reasonable approximation. We estimate that calculated extinction coefficients differ from the true extinction coefficients by no more than 25%.

Binding assays

Triplex formation was assessed using the gel shift assay, essentially as described (3,5). Incubation was in 20 mM Tris−HCl, pH 7.6, 10 mM $MgCl_2$, 10% sucrose, at 37°C for 20−24 h (except as noted). Duplex DNA concentrations were $\sim 2.5\times10^{-11}$ M. Apparent dissociation constants for triplex formation were estimated as the concentration of oligonucleotide at the midpoint in the duplex to triplex transition (5). All values were reproducible within a factor of two.

RESULTS

Unusual stability of T·CG triplets in antiparallel triplexes

We initially chose a 26-bp oligopurine·oligopyrimidine duplex, ZRY100-0, to explore the effects of various triplet mismatches on antiparallel triplex formation. Figure 1A shows that a 26-base oligonucleotide, Z100-50, will bind to ZRY100-0 to form a triplex consisting entirely of G·GC and T·AT base triplets. The relative affinity of Z100-50 for ZRY100-0 can be estimated from the concentration of triplex forming oligonucleotide (TFO) required to bind 50% of the duplex. In this experiment, the midpoint in the duplex to triplex transition occurs at approximately 3×10^{-10} M TFO.

To test the effects of limited pyrimidine substitutions in the purine strand of the duplex, we synthesized ZRY102-0. This duplex is identical to ZRY100-0 except that three non-neighboring GC base pairs have been replaced with CG base pairs. As shown in Figure 1B, Z100-50 is unable to bind to this target, even at concentrations as high as 1μM. Clearly the presence of three CG inversions in ZRY102-0 prevents triplex formation.

We also examined the ability of the other three natural bases, A, C, and T, to bind to CG inversions. Binding of Z102-57 (A), Z102-55 (C), and Z102-56 (T) to ZRY102-0 is shown in Figure 2A−C. Only Z102-56 forms significant amounts of triplex, with a midpoint of approximately 1×10^{-9} M third strand. Z102-55 and Z102-57 show little or no triplex formation even at 10^{-6} M TFO. Under these conditions, the affinity of Z102-56 for ZRY102-0 is only about 3-fold less than that of Z100-50 for ZRY100-0, the ideal triplex. These data indicate that the T·CG interaction is substantially more favorable than G·CG, C·CG, or A·CG, at least in context of this target sequence. This is in general agreement with data reported by Beal and Dervan (19).

In order to compare the strength of the T·CG interaction with that of the standard T·AT triplet, we examined binding of Z102-56 to ZRY102-7. Figure 2D shows that the apparent K_d is $\sim 3\times10^{-10}$ M, similar to the affinity of Z100-50 for ZRY100-0, and clearly better than Z102-56 binding to ZRY102-0. However, the relatively small difference in K_d indicates that the T·CG triplets are only marginally less stable than T·AT or G·GC triplets for this system.

The T·CG interaction is favored in other sequence contexts

To determine whether T·CG is favorable in other sequence contexts, we synthesized several additional duplex targets containing CG inversions. BRY106-1 and BRY108-0 represent naturally occurring sequences in biologically important regions of the Herpes simplex 2 and Herpes simplex 1 viruses, respectively (Fig. 3). Neither sequence is an ideal target for triplex formation, and each contains several CG inversions. Attempts to detect triplex formation between B108-50 and BRY108-0 were unsuccessful (Fig. 3A). This is consistent with the results for Z100-50 and ZRY102-0, indicating that unfavorable interactions associated with trying to form three G·CG triplets (as well as any unfavorable contributions from the two T·TA triplets) prevent triplex formation. In contrast, B108-53 does bind to BRY108-0, although relatively high TFO concentrations are required (Fig. 3B; $K_d \approx 3\times10^{-7}$ M). As expected, using T in the third strand to form T·CG triplets is significantly more favorable, and permits triplex formation at a sequence that contains 20% pyrimidines in the purine-rich strand.

Similar results were obtained for the binding of B106-51 and B106-52 to BRY106-1 (Fig. 3C−D), which contains two CG inversions separated by a single GC pair. B106-51, containing G residues at the positions corresponding to the CG inversions, binds to BRY106-1 with an apparent midpoint of $\sim 2\times10^{-8}$ M.

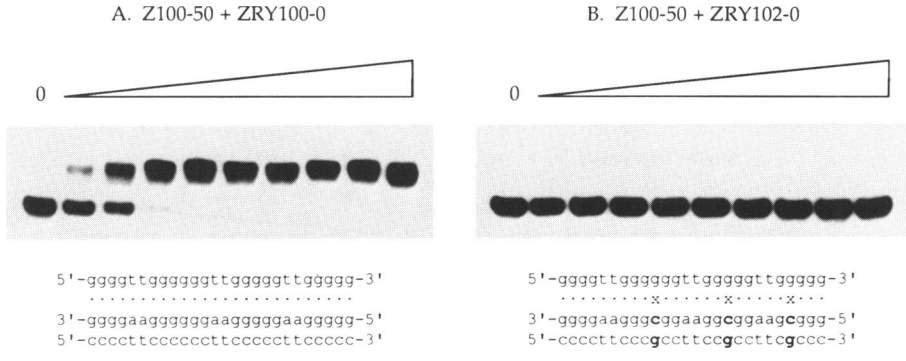

Figure 1. Electrophoretic band shift analysis of triplex formation. (**A**) Binding of the TFO Z100-50 to the duplex target ZRY100-0. (**B**) Binding of Z100-50 to ZRY102-0, containing three CG inversions. Sequences for the triplexes are shown in the lower half of each panel, with the TFO shown above the duplex target. Duplex target concentration in all lanes is $\sim 2.5\times10^{-11}$ M. The first lane of each panel contains radiolabeled duplex in the absence of third strand. Subsequent lanes contain increasing concentrations of TFO: 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000 nM. All samples were incubated for approx. 20 h in 20 mM Tris−HCl, pH 7.6, 10 mM $MgCl_2$, 10% sucrose at 37°C prior to electrophoresis. The appearance of a slower migrating band in panel A indicates the formation of triplex by binding of Z100-50 to ZRY100-0, with an apparent midpoint of $\sim 3\times10^{-10}$ M. No binding of Z100-50 to ZRY102-0 is observed in panel B.

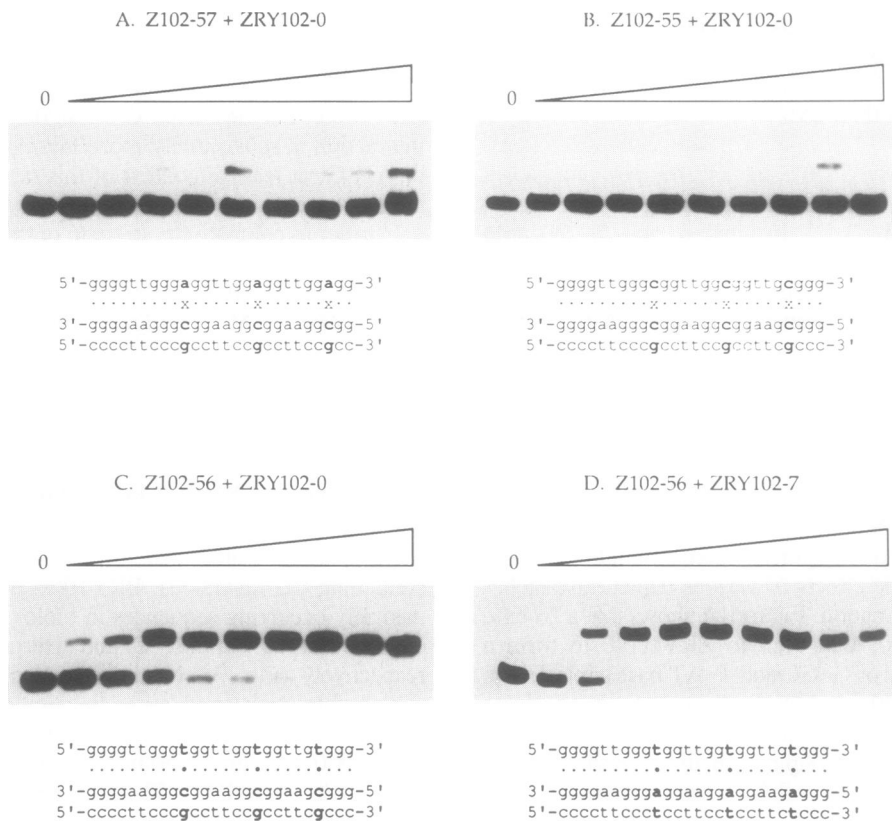

Figure 2. Analysis of binding for the following TFO and duplex combinations. (**A**) Z102-57 plus ZRY102-0. (**B**) Z102-55 plus ZRY102-0. (**C**) Z102-56 plus ZRY102-0. (**D**) Z102-56 plus ZRY102-7. All concentrations and conditions are as described in Figure 1.

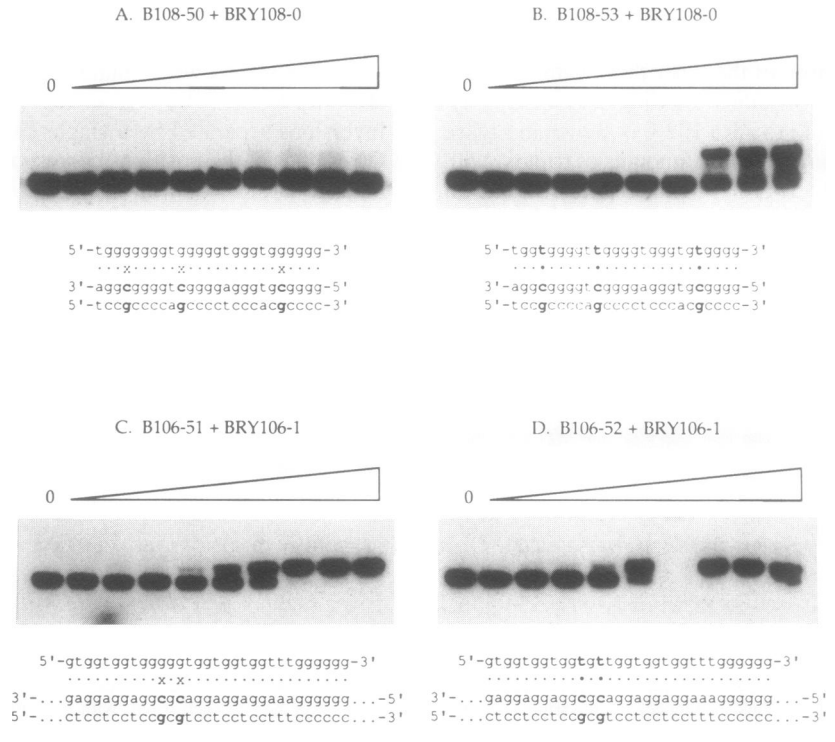

Figure 3. Analysis of triplex formation (as detailed in Figure 1) for: (**A**) B108-50 plus BRY108-0. (**B**) B108-53 plus BRY108-0. (**C**) B106-51 plus BRY106-1. (**D**)B106-52 plus BRY106-1. The sample containing 30nM B106-52 was lost (lane 7).

In this case, the two G·CG triplets are not sufficient to completely disrupt the triplex, probably due to the long stretches of uninterrupted G·GC and T·AT triplets on either side. However, B106-52, which contains T residues at the CG inversions, binds significantly better, with an apparent midpoint in the nanomolar range. These results indicate that T·CG triplets are generally more favorable than other possible interactions at CG inversions (in the context of antiparallel triplexes).

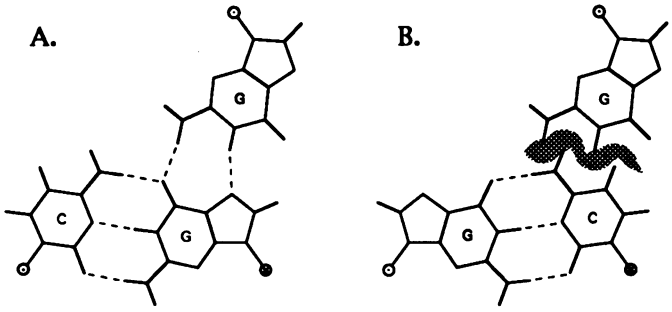

Figure 4. Schematic comparison of a reverse Hoogsteen G·GC triplet (**A**) with a G·CG mismatch (**B**). Circles attached to C-N1 and G-N9 indicate the position of the C1′ atoms of the backbone. Circles containing a cross indicate the 5′−3′ direction of that strand is down (into the page), while circles containing a point indicate the opposite 5′−3′ orientation. Panel A was generated by manually docking a third strand G residue with a GC base pair, to yield appropriate distances for hydrogen bonds (dashed lines) for the reverse Hoogsteen interaction. Panel B was generated by converting the GC base pair to a CG pair, without altering the relative backbone positions of any of the three strands. Note that in panel B, there is substantial van der Waals overlap (shaded area) between the third strand G and C of the duplex. Accommodation of this mismatch would require substantial dislocation of the third strand, the duplex, or both.

Possible explanations for the T·CG interaction

A likely explanation for the observation that T·CG is a more favorable triplet than G·CG or A·CG is that in the latter two cases, the large purine in the third strand would be expected to lead to significant steric clash with the C in the duplex (Fig. 4). In order to accommodate either of these triplets in a stable triplex, a large distortion of the third strand and/or duplex backbones would likely be required. Replacing the third strand purine with a pyrimidine would substantially reduce the steric problem. However, it is not immediately obvious why T is favored over C at CG inversions. Because T and C are similar in size, it seems unlikely that steric effects alone could account for the dramatic binding differences we observe. This suggests that specific hydrophobic interactions, hydrogen bonding, or stacking effects may be involved.

The presence of a carbonyl oxygen at position 2 of both T and C suggests that this group is not critical in binding to CG. However, T and C differ in hydrogen bonding potential at positions 3 and 4. We attempted to determine whether the hydrogen bond donor at N^3 of T plays an important role in the T·CG interaction. At low pH, C becomes protonated at N^3, allowing it to act as a hydrogen bond donor (25). We reasoned

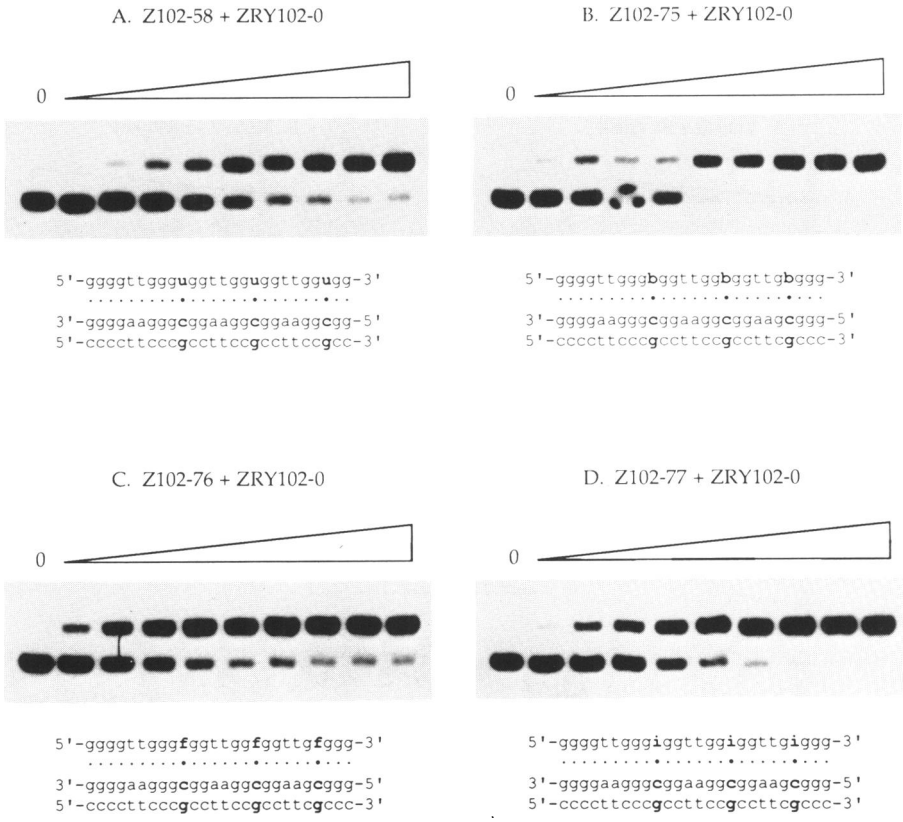

Figure 5. Analysis of triplex formation (as detailed in Figure 1) for: (**A**) Z102-58 plus ZRY102-0. (**B**) Z102-75 plus ZRY102-0. (**B**) Z102-76 plus ZRY102-0. (**D**) Z102-77 plus ZRY102-0. In TFO sequences, u = 2′-deoxyuridine, f = 2′-deoxy-5-fluorouridine, b = 2′-deoxy-5-bromouridine, and i = 2′-deoxy-5-iodouridine.

that if such an interaction was important in the T·CG triplet, a similar C^+bCG triplet might form at low pH. Accordingly, we compared the binding of Z102-55 and Z102-56 to ZRY102-0 at pH 5. The results were essentially indistinguishable from those obtained at pH7.6 (not shown). Z102-55 showed no binding to ZRY102-0 at pH 5, suggesting that a hydrogen bond donor at N^3 is not sufficient to explain the unusual stability of the T·CG interaction. Binding of Z102-56 to ZRY102-0 at pH 5 was essentially identical to the results obtained at pH 7.6.

Table I. Comparison of T and T analogs binding to CG and AT base pairs.

oligo ID	sequence*	apparent K_d
ZRY102-0	5'-ccccttcccgccttccgccttcgccc-3' 3'-ggggaagggcggaaggcggaagcgggg-5'	
Z102-56	5'-ggggttgggtggttggtggttgtggg-3'	$1x10^{-9}$ M
Z102-58	5'-ggggttggguggttgguggttguggg-3'	$1x10^{-8}$ M
Z102-75	5'-ggggttgggbggttggbggttgbggg-3'	$5x10^{-9}$ M
Z102-76	5'-ggggttgggfggttggfggttgfggg-3'	$5x10^{-10}$ M
Z102-77	5'-ggggttgggiggttggiggttgiggg-3'	$3x10^{-9}$ M
Z102-125	5'-ggggttggg2ggttgg2ggttg2ggg-3'	»$1x10^{-6}$ M †
Z102-126	5'-ggggttggg4ggttgg4ggttg4ggg-3'	$1x10^{-7}$ M
ZRY102-7	5'-cccctttccctccttcctccttctccc-3' 3'-ggggaagggaggaaggaggaagaggg-5'	
Z102-56	5'-ggggttgggtggttggtggttgtggg-3'	$3x10^{-10}$ M
Z102-58	5'-ggggttggguggttgguggttguggg-3'	$8x10^{-10}$ M
Z102-75	5'-ggggttgggbggttggbggttgbggg-3'	$6x10^{-10}$ M
Z102-76	5'-ggggttgggfggttggfggttgfggg-3'	$1x10^{-10}$ M
Z102-77	5'-ggggttgggiggttggiggttgiggg-3'	$4x10^{-10}$ M

*Nucleoside designations are u, 2'-deoxyuridine; f, 2'-deoxy-5-fluorouridine; b, 2'-deoxy-5-bromouridine; i, 2'-deoxy-5-iodouridine; 2, pyridin-2-one deoxyribonucleoside; 2, pyridin-4-one deoxyribonucleoside.
†No triplex detected at TFO concentrations up to 10^{-6} M.

T and C also differ by the presence of the methyl group at position 5 of T. In order to determine if the 5-methyl group contributes significantly to the stability of the T·CG interaction, we tested the ability of 2'-deoxyuridine (dU) to bind to CG inversions. Z102-58, containing three dU residues, binds to ZRY102-0 with an apparent K_d of 1×10^{-8} M (Fig. 5A). This is approximately 10-fold lower in affinity than Z102-56 (Fig. 2C), suggesting that the Me^5 group does contribute to the overall stability of the interaction. However, the absence of Me^5 in dU does not abolish binding to CG, indicating that it is not an essential component of the interaction.

Based on these data we hypothesized that the O^4 group of T plays a substantial role in binding to CG inversions. This is in contrast to a model presented by Beal and Dervan, which postulated a hydrogen bond between T-O^2 and C-N^4 (19). Our observation that C is essentially unable to bind CG in this system is inconsistent with this model, as C-O^2 should be comparable to T-O^2 as a hydrogen bond acceptor. It is possible that T-O^4 is hydrogen bonded to C-N^4, but it is also possible that O^4 contributes to the stability of the T·CG triplet in other ways (see discussion).

Effects of T analogs on triplex formation

We next examined binding of several T analogs to CG inversions and to AT base pairs. Initially, we examined analogs commercially available as phosphoramidites. These included dU (described above) and the 5-halogenated deoxyuridines 5-fluoro-dU (5FdU), 5-bromo-dU (5BrdU), and 5-iodo-dU (5IdU). Results

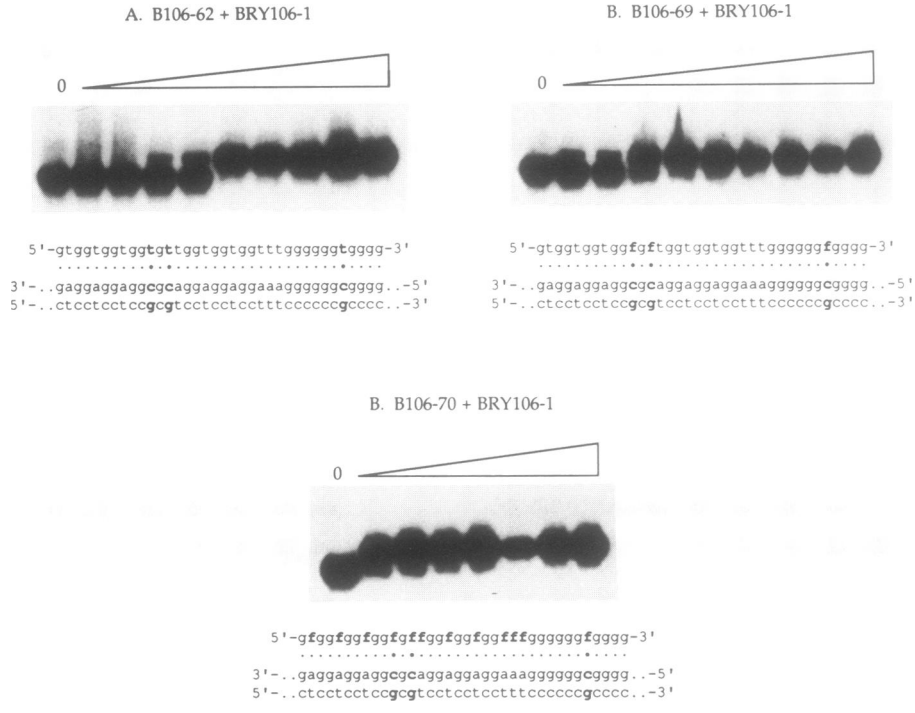

Figure 6. Analysis of triplex formation (as detailed in Figure 1) for: (**A**) B106-62 plus BRY106-1. (**B**) B106-69 plus BRY106-1. (**C**) B106-70 plus BRY106-1. In TFO sequences, f = 2'-deoxy-5-fluorouridine.

with the 5-halogenated-dU derivatives indicated that 5FdU has a slightly higher affinity for CG inversions than T, and a significantly higher affinity than 5BrdU and 5IdU (Fig. 5B–D). We also examined the effects of these analogs on binding to AT base pairs, with similar results (Table I). 5FdU bound slightly better to AT pairs than did T, and significantly better than 5BrdU or 5IdU.

Table II. Binding of 5FdU to CG and AT base pairs in a Herpes simplex 2 sequence.

BRY106-1	5'-gatcgctcctcctccgcgtcctcctcctttccccccgccccgcgcgccccgagcg-3' 3'-ctagcgaggaggaggcgcaggaggaggaaaggggggcggggcgcgcggggctcgc-5' .....•.•.....	
B106-62	5'-gtggtggtggtgttggtggtggtttgggggggtgggg-3'	$6x10^{-9}$ M
B106-69	5'-gtggtggtggfgftggtggtggtttgggggggfgggg-3'	$8x10^{-10}$ M
B106-62	5'-gfggfggfggfgffggfggfggfffgggggggfgggg-3'	$<1x10^{-10}$ M ‡
B106-62	5'-gtggtggtggtgttggtggtggtttgggggggtgggg-3'	$6x10^{-8}$ M §
B106-69	5'-gtggtggtggfgftggtggtggtttgggggggfgggg-3'	$3x10^{-8}$ M §
B106-62	5'-gfggfggfggfgffggfggfggfffgggggggfgggg-3'	$1x10^{-9}$ M §

f = 5-fluoro-dU.
‡≥95% triplex was observed at 1×10^{-10} M third strand, the lowest concentration tested.
§These K_d values were measured in the presence of 1 mM $MgCl_2$, compared to the standard concentration of 10 mM for other assays.

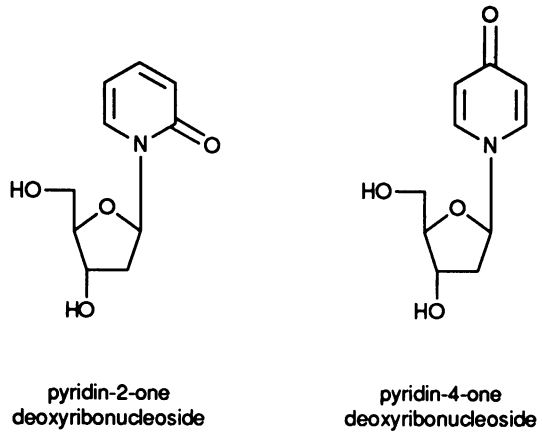

Figure 7. Structures of pyridin-2-one and pyridin-4-one deoxyribonucleosides.

To confirm these results, we examined binding of B106-62, B106-69, and B106-70 to BRY106-1. B106-62 is an extension of B106-52, and is designed to bind to a larger region of the duplex target. B106-69 and B106-70 are 5FdU-containing analogs of B106-62, with 5FdU present only at CG inversions (B106-69), or at CG inversions and AT base pairs (B106-70). Comparison of the binding affinities for these three TFOs indicates that 5FdU substantially enhances triplex formation (Fig. 6; Table II). In the presence of 10mM $MgCl_2$, B106-69 bound with ~10-fold higher affinity than B106-62. B106-70 formed almost 100% triplex at concentrations as low as 10^{-10} M. Due to the limitations of the assay, it was not possible to estimate the K_d under these conditions. In order to better quantify the difference in binding affinity for B106-62, B106-69, and B106-70, we repeated the assays in the presence of 1 mM $MgCl_2$, which reduces the triplex stability enough to permit evaluation of the K_d. Under these conditions, B106-69 and B106-70 bind ~2-fold and ~60-fold more tightly than B106-62 (Table II).

Role of thymidine O^4 in binding to CG inversions

As described above, our data suggests that T-O^4 is a critical element in the unusual stability of T·CG triplets. In order to confirm this, we synthesized oligomers containing nucleoside derivatives of pyridin-2-one and pyridin-4-one (Fig. 7). These compounds were designed to serve as analogs of T containing only a single functional group, namely, a carbonyl at either position 2 or 4. We reasoned that if thymidine O^4 plays a critical role, then pyridin-4-one should also exhibit binding to CG inversions. In contrast, we proposed that thymidine O^2 is not a significant determinant of T·CG triplets, and predicted that pyridin-2-one would not show substantial binding to CG inversions. Accordingly, we tested the binding of Z102-125 and Z102-126 to ZRY102-0 (Fig. 8). The results were essentially as predicted. Z102-126, containing pyridin-4-one nucleosides at CG inversions, bound to the duplex, whereas Z102-125 did not. These results confirm the importance of the O^4 group, and indicate that other groups, including O^2, N^3, and Me^5 are not essential elements in the interaction. It is clear, however, that binding of pyridin-4-one to CG inversions is substantially weaker than that of T (by ~100-fold). Thus, other interactions, such as hydrophobic or stacking effects, are certainly involved, and presumably require additional functional groups.

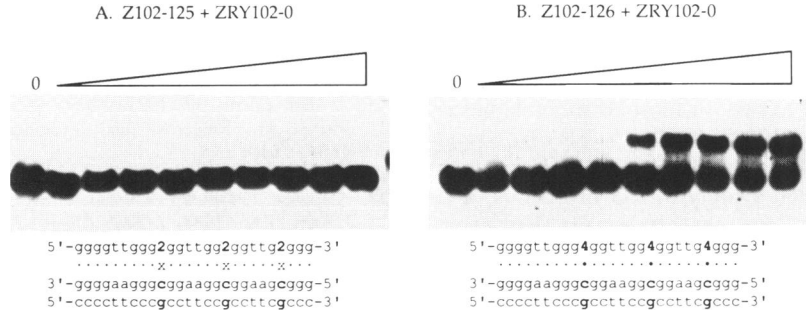

Figure 8. Analysis of triplex formation (as detailed in Figure 1) for: (**A**) Z102-125 plus ZRY102-0. (**B**) Z102-126 plus ZRY102-7. In TFO sequences, 2 = pyridin-2-one, and 4 = pyridin-4-one.

DISCUSSION

The data presented here show that T is unique among the common bases in its ability to interact with CG base pairs in antiparallel triplexes. This confirms and extends the results of Beal and Dervan (19). Although the molecular basis of this interaction is unknown, we may hypothesize that several factors are involved. It is likely that the smaller size of T is a significant advantage relative to A or G. This reduces the potential for unfavorable steric interactions with the cytosines in the duplex. However, C does not permit significant binding in our experiments, so size alone is an insufficient explanation. Data presented here suggest that O^4 of T is a major participant in the T·CG triplet. This is especially supported by the data with pyridin-2-one and -4-one.

Beal and Dervan have proposed a model for the T·CG triplet in which thymidine O^2 forms a single hydrogen bond with cytosine NH^4 (19). However, the data presented here argue against that model. Preliminary work suggests that the thymidine O^2 group and the N^3 proton probably do not play major roles in the T·CG interaction. However, comparison of T and dU does suggest that the 5-methyl group has a stabilizing effect. It is possible that other interactions such as stacking or hydrogen bonding with adjacent bases may play a significant role in the T·CG interaction. Further studies using techniques such as nuclear magnetic resonance or X-ray crystallography will be necessary to fully understand this interaction.

We also studied the effects of the 5-substituent on binding of T analogs to AT and CG base pairs. In both cases, the effect of the 5 substituent on relative binding affinities was approximately $F > CH_3 > I > Br > H$. These results are likely to reflect a number of factors. The high electronegativity of fluorine may be at least partially responsible for its relative superiority. Electron withdrawing effects may alter interactions involving O^4, which may be important in binding to CG inversions. In addition, the pK_a of the imino proton is substantially reduced in 5-halogenated-uridine derivatives (26), suggesting that hydrogen bonding strength may be affected, especially for binding to AT base pairs. Hydrophobic, steric, and stacking effects may also contribute to the observed effects.

Oligonucleotide-based triple helix formation is currently under scrutiny for its potential value in a variety of applications, including transcriptional regulation and targeted DNA damage (reviewed in 2). Such applications are currently limited by an inability to achieve triplex formation with mixed sequence targets. The observation that T exhibits substantial affinity for CG inversions in antiparallel triplexes is important in this regard. In this work, we have demonstrated that T·CG triplets permit triplex formation with targets that would not otherwise be suitable.

A drawback of using T to bind to CG inversions is the potential loss of sequence specificity. The relative affinity of T for AT and CG base pairs in our studies is apparently similar. Thus, a given third strand may bind to two (or more) distinct duplexes with nearly equal affinity (cf. Z102-56 binding to ZRY102-0 and ZRY102-7). Appropriate base modifications may result in a compound that binds with high affinity to CG inversions, and with little or no affinity for AT (or other) base pairs. However, achieving this goal will require a better understanding of the nature of the T·CG interaction. Thus, the observation that pyridin-4-one (but not pyridin-2-one) binds to CG inversions is important. It confirms the importance of O^4, and provides a starting point in the design of other T analogs with increased affinity and specificity for CG inversions.

Although the importance of T-O^4 is clear, the reasons for it are not. As discussed above, one possibility is that O^4 hydrogen bonds to C-N^4. However, recent NMR studies on an intramolecular triplex containing a single T·CG triplet do not appear to support this model (X. Gao, personal communication). Evidence for substantial hydrogen bonding to C-NH^4, by either T-O^4 or T-O^2, was not observed. Instead, evidence suggested a possible close contact between T-NH^3 and C-NH^4. These data suggest that a T analog that is unprotonated at N^3 might be desirable, perhaps permitting hydrogen bonding between T-N^3 and C-NH^4. One such analog is pyrimidin-4-one. We have recently incorporated this analog into oligonucleotides to test the possibility that it may increase both the affinity and the specificity for CG inversions.

In light of the NMR data, it is unclear how T-O^4 stabilizes the T·CG triplet. Interactions with bases above or below the plane of the T·CG triplet may be crucial. Effects on stacking are also possible. Finally, potential interactions between O^4 and solvent or counterions cannot be dismissed. A variety of studies are currently underway to resolve these questions, with the ultimate goal of developing a high-affinity, CG-specific nucleoside for antiparallel triplex formation.

REFERENCES

1. Moser,H.E. and Dervan,P.B. (1987) *Science*, **238**, 645–650.
2. Thuong,N.T. and Helene,C. (1993) *Angew. Chem. Int. Ed. Engl.*, **32**, 666–690.
3. Cooney,M., Czernuszewicz,G., Postel,E.H., Flint,S.J. and Hogan,M.E. (1988) *Science*, **241**, 456–459.
4. Beal,P.A. and Dervan,P.B. (1991) *Science*, **251**, 1360–1363.
5. Durland,R.H., Kessler,D.J., Gunnell,S., Duvic,M., Pettitt,B.M. and Hogen,M.E. (1991) *Biochemistry*, **30**, 9246–9255.
6. Griffin,L.C. and Dervan,P.B. (1989) *Science*, **245**, 967–971.
7. Horne,D.A. and Dervan,P.B. (1990) *J. Am. Chem. Soc.*, **112**, 2435–2437.
8. Ono,A., Chen,C.-N. and Kan,L.-S. (1991) *Biochemistry*, **30**, 9914–9921.
9. Mergny,J.-L., Sun,J.-S., Rougee,M., Montenay-Garestier,T., Barcelo,F., Chomilier,J. and Helene,C. (1991) *Biochemistry*, **30**, 9791–9798.
10. Radhakrishnan,I., Gao,X., de los Santos,C., Live,D. and Patel,D.J. (1991) *Biochemistry*, **30**, 9022–9030.
11. Horne,D.A. and Dervan,P.B. (1991) *Nucleic Acids Res.*, **19**, 4963–4965.
12. Roberts,R.W. and Crothers,D.M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9397–9401.
13. Jayasena,S.D. and Johnston,B.H. (1992) *Biochemistry*, **31**, 320–327.
14. Kiessling,L.L., Griffin,L.C. and Dervan,P.B. (1992) *Biochemistry*, **31**, 2829–2834.
15. Blume,S.W., Gee,J.E., Shrestha,K. and Miller,D.M. (1992) *Nucleic Acids Res.*, **20**, 1777–1784.
16. Yoon,K., Hobbs,C.A., Koch,J., Sardars,M., Kutny,R. and Weis,A.L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3840–3844.
17. Beal,P.A. and Dervan,P.B. (1992) *J. Am. Chem. Soc.*, **114**, 4976–4982.
18. Griffin,L.C., Kiessling,L.L., Beal,P.A., Gillespie,P. and Dervan,P.B. (1992) *J. Am. Chem. Soc.*, **114**, 7976–7982.
19. Beal,P.B. and Dervan,P.B. (1992) *Nucleic Acids Res.*, **20**, 2773–2776.
20. Seela,F. and Bindig,U. (1989) *Liebigs Ann. Chem.*, **9**, 895–901.
21. Woodford,W.J., Swartz,B.A., Pillar,C.J., Kampf,A. and Mertes,M.P. (1974) *J. Med. Chem.*, **17**, 1027–1029.
22. Murphy,M., Reiger,M. and Jayaraman,K. (1993) *BioTechniques*, **15**, 1004–1010.
23. Cantor,C.R. and Tinoco,I. Jr. (1967) *Biopolymers*, **5**, 821–835.
24. Alexis,M. (1978) Ph.D. thesis, University of London.
25. Lipsett,M.N. (1964) *J. Biol. Chem.*, **239**, 1256–1260.
26. Saenger,W. (1984) Principles of Nucleic Acid Structure. Springer-Verlag, New York.