Nucleic Acids Research, 1996, Vol. 24, No. 10 **1963–1970**

Use of a pyrimidine nucleoside that functions as a bidentate hydrogen bond donor for the recognition of isolated or contiguous G-C base pairs by oligonucleotide-directed triplex formation

Guobing Xiang, Robert Bogacki and Larry W. McLaughlin*

Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, MA 02167, USA

Received November 13, 1995; Revised and Accepted March 25, 1996

ABSTRACT

Synthesis of the nucleoside building block of the 6-keto derivative of 2′-deoxy-5-methylcytidine ($m^{5ox}C$) as an analog of an N^3-protonated cytosine derivative is described. A series of 15mer oligonucleotides containing either four or six $m^{5ox}C$ residues has been prepared by chemical synthesis. Complexation of the 15 residue oligonucleotides with target 25mer duplexes results in DNA triplexes containing T-A-T and $m^{5ox}C$-G-C base triplets. When the $m^{5ox}C$-G-C base triplets are present in sequence positions that alternate with TAT base triplets, DNA triplexes are formed with T_m values that are pH independent in the range 6.4–8.5. A 25mer DNA duplex containing a series of five contiguous G-C base pairs cannot be effectively targeted with either m^5C or $m^{5ox}C$ in the third strand. In the former case charge–charge repulsion effects likely lead to destabilization of the complex, while in the latter case ineffective base stacking may be to blame. However, if the m^5C and $m^{5ox}C$ residues are present in the third strand in alternate sequence positions, then DNA triplexes can be formed with contiguous G-C targets even at pH 8.0.

INTRODUCTION

A common triple helix structural motif involves the recognition of a polypurine target duplex by a pyrimidine-rich third strand. In this format, the pyrimidine-rich third strand is oriented parallel to the polypurine target and the stability of the complex results from complementary T-A-T and C^+-G-C base triplets (1,2). The T-A-T base triplet, in which a thymidine interacts with a Watson–Crick A-T base pair by making two Hoogsteen hydrogen bonds (3,4), is a neutral, stable and pH-independent base triplet that represents a simple case of bidentate recognition of an A-T base pair embedded within a duplex structure. By comparison, the C^+-G-C base triplet, while offering similar geometry and sites of interaction, requires protonation of the C residue (5,6) of the third strand in order for two Hoogsteen-like hydrogen bonds to form. As expected, the stability of the C^+-G-C base triplet is dependent upon the pH of the solution and this interaction appears to provide only a minimal contribution to triplex stability near physiological pH values (7), although the C residue appears to be protonated at pH values much higher than its intrinsic p*K*a (8). Nevertheless, by lowering the pH slightly below 7.0, very effective triple helix formation takes place with this recognition motif.

A number of base analogs have been employed in attempts to improve the parallel-stranded pyrimidine–purine–pyrimidine recognition motif. Some of the first base analogs employed for triple helix formation attempted to address the 'pH disadvantage' of C^+-G-C base triplets. In 1984 it was shown (9) that poly(5-methylcytosine) (m^5C) formed triplexes that were more stable than corresponding poly(C) triplexes. Subsequently it was reported (7,10) that m^5C produced more stable triplex structures with sequence-defined duplexes. While the substitution of a methyl group at the C_5 position of cytidine clearly enhances the stability of triplexes containing G-C base pair targets, it remains a less than ideal solution for complexes formed near physiological pH values, where protonation is reduced.

Base analogs can be used to eliminate the requirement for protonation of C residues in triplets involving G-C targets. One such analog is the 2′-*O*-methyl derivative of pseudoisocytidine (11,12). This derivative can form two pH-independent hydrogen bonds with guanine, although one interaction depends upon the nature of the tautomeric equilibrium of the N^1 and N^3 nitrogens. In a preliminary communication we have described the use of the 6-keto derivative of m^5C for the pH-independent recognition of G-C base pairs (13) and we expand upon those studies here. The synthesis of a related 2′-*O*-methyl derivative has also been described (14) and a pyrazine C nucleoside has been designed to provide a similar bidentate recognition motif (15). Three classes of purine-like derivatives have also been employed for the interaction with target G-C base pairs in the parallel-stranded recognition motif. In one case, two related pyrazolopyrimidinones (16,17) can be employed to provide the desired Hoogsteen-like hydrogen bonds to guanine and, in similar fashion, N^7-glycosylated guanine will recognize G-C base pairs (18). 8-Oxoadenine (19,20) or its N^6-methyl derivative (21) when present in the *syn* conformation can form the desired base triplets. In the present work we describe the synthesis of the pyrimidine

* To whom correspondence should be addressed

2′-deoxynucleoside derivative $m^{5ox}C$ and its use in targeting DNA duplexes containing isolated and contiguous G-C base pairs.

MATERIALS AND METHODS

Materials

HPLC grade solvents were obtained from Fisher Scientific (Fair Lawn, NJ), other reagents were from Adrich Chemical Co. (Milwaukee, WI). 5′-Dimethoxytrityl nucleoside phosphoramidite monomers as well as all ancillary reagents for nucleic acid synthesis were obtained from Cruachem through Fisher Scientific or from Applied Biosystems, Inc. (Foster City, CA). Oligonucleotides were synthesized using nucleoside phosphoramidite derivatives and an Applied Biosystems 381A DNA synthesizer. High performance liquid chromatography (HPLC) was carried out on an ODS-Hypersil column (0.46 × 25 cm; Shandon Southern, UK), using a Beckman HPLC system. ^{1}H NMR spectra were obtained at 300 or 500 MHz on Varian XL-300 or XL-500 multinuclear spectrometers. ^{31}P NMR spectra were obtained at 121 MHz on the Varian XL-300. Absorption spectra were recorded by a Perkin-Elmer Lambda 3B UV/Vis spectrophotometer. Mass spectra were obtained from the Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois (Urbana, IL). Melting points were obtained on a Thomas Hoover capillary melting point apparatus.

4-Amino-1-[β-2-deoxy-3,5-*O*-bis(toluoyl)-D-erythro-pentofuranosyl]-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (1)

To 4.2 g (30 mmol) 6-aminothymine (22,23) (dried at 100°C overnight) was added 0.25 g $(NH_4)_2SO_4$ and the mixture was refluxed at 140°C under argon in 90 ml HMDS for 2 h (after 30 min the reaction turns clear). After removal of the HMDS and co-evaporation from toluene twice (at 50°C) the residue was further dried under vacuum at 50°C for 30 min. To the viscous residue was added 8.0 g (20.6 mmol) α-1-chloro-3,5-*O*-toluoyl-β-2-deoxy-D-*erythro*-pentofuranose (24), 200 ml 1,2-dichloroethane and 80 ml acetonitrile. After cooling to 0°C (ice–water bath), 1.2 ml TMS triflate in 20 ml 1,2-dichloroethane was added dropwise under argon during a 30 min period and the mixture stirred at ambient temperature for an additional 3 h (the reaction mixture turns clear after 30 min). The solvent was then removed, the residue dissolved in 300 ml dichloromethane and the solution washed with dilute sodium bicarbonate and water, dried over sodium sulfate and the solvent removed by rotary evaporation (*in vacuo*). The resulting mixture (TLC, dichloromethane/methanol, 95:5) contained a major product (R_f 0.28) and a minor product (R_f 0.38). The major product was isolated as a white solid by flash chromatography on silica gel using dichloromethane and a gradient of methanol (0–5%). The resulting 8.5 g (84%) of coupling product was obtained as a roughly 1:1 mixture of the α and β anomers (based upon ^{1}H NMR analysis).

Resolution of the two anomeric nucleosides required further modification of the nucleosides. To 3.21 g (6.5 mmol) of the coupling mixture, co-evaporated twice from anhydrous pyridine, was added 45 ml anhydrous pyridine. This solution was cooled to 0°C (ice–water bath) and 2.84 g (15 mmol) *o*-nitrobenzensulfenyl chloride was added. The reaction was stirred overnight and warmed to ambient temperature. The pyridine was removed by rotary evaporation (*in vacuo*) and the residue was co-evaporated from toluene, dissolved in dichloromethane, washed with dilute sodium bicarbonate, dried over sodium sulfate and the solvent was removed. Two products were evident by TLC (dichloromethane/ethyl acetate, 95:5) with R_f values of 0.75 and 0.58. Resolution of the two products was accomplished by flash chromatography on silica gel using dichloromethane and a gradient of 0–5% ethyl acetate and resulted in 2.3 g (47%) of the faster moving material as a yellow solid. Both the faster and slower moving derivatives appeared to be composed of two different compounds (based upon ^{1}H NMR analysis) and each of these compounds contained two equivalents of the *o*-nitrosulfenyl protecting group. The structures of these materials have not been conclusively identified, but each pair of compounds could be deprotected to yield either the pure α or the pure β isomer as follows.

The materials found in the faster moving TLC spot were deprotected by dissolving the 2.3 g (3 mmol) of mixed derivatives in pyridine and adding 0.9 g *p*-thiocresol (7.25 mmol). After stirring overnight at ambient temperature, the pyridine was removed by rotary evaporation (*in vacuo*), co-evaporated twice from toluene and worked up as described above. After purification by column chromatography (dichloromethane/methanol), 1.4 g (95%) of a single compound, the β isomer of **1**, was obtained as a white solid.

Yield 37% overall from the chloro-sugar derivative. m.p. 238–239°C. R_f (dichloromethane/methanol, 95/5): 0.28. UV (methanol): λ_{max} 233, 268 nm; λ_{min} 209, 254 nm. HRMS (FAB) for $C_{26}H_{28}N_3O_7$ (M+ H) calculated 494.1927, found 494.1941. [α] (methanol) = 0.040° (α isomer, 0.133°). ^{1}H-NMR (DMSO-d_6) δ = 1.67 (s, 3H, CH_3-), 2.36 (s, 3H, CH_3-), 2.39 (s, 3H, CH_3-), 2.50 (DMSO), 2.37 and 3.00 (m, 2H, H2′, H2″), 3.35 (H_2O), 4.35 (m, 1H, H4′), 4.40–4.60 (m, 2H, H5, H5″), 5.76 (m, 1H, H3′), 6.10 (s, 2H, -NH_2), 6.72 (dd, 1H, H1′), 7.20–7.40 (m, 4H, Ar-H), 7.85 (m, 4H, Ar-H), 10.30 (s, 1H, NH) p.p.m. ^{13}C-NMR (DMSO-d_6) δ = 165.5, 165.3, 162.6, 150.5, 149.6, 143.8, 143.6, 129.3 (4C), 129.2 (2C), 129.1 (2C), 126.7, 126.6, 80.8, 80.7, 79.5, 75.2, 64.7, 34.9, 21.1 (2C), 8.2 p.p.m.

Isomer assignments. Assignment of the faster moving material (TLC) as the β isomer was the result of the observation of a NOESY cross-peak between H1′ and H4′, while no cross-peak was observed between H1′ and H3′. The slower moving material, upon deprotection resulted in a single compound which was assigned as the α isomer on the basis of a NOESY cross-peak between H1′ and H3′.

Regiochemistry. Assignment of these derivatives as the N^3-glycosylation products was based on a comparison of the ^{1}H NMR characteristics of the amino proton resonances as described in other work (25). The amino resonances of the N^1-modified heterocycle are shifted downfield in DMSO-d_6 relative to the corresponding N^3-modified pyrimidione. Each of the 3′,5′-bis(toluoyl) nucleoside derivatives were deprotected in 0.1 M sodium methoxide. A small portion of each was purified by preparative TLC to yield the following ^{1}H NMR data (for comparative NMR assignments see Müller; 25): β isomer ^{1}H NMR (DMSO-d_6) δ = 6.11 (s, 2H, -NH_2), 6.55 (t, 1H, H1′), 10.25 (s, 1H, NH); α isomer ^{1}H NMR (DMSO-d_6) δ = 6.17 (s, 2H, -NH_2), 6.44 (t, 1H, H1′), 10.35 (s, 1H, NH)

For comparison, the N^1-glycoslylation products were isolated as an isomeric mixture and deprotected to yield the corresponding data: α + β isomers ^{1}H NMR (DMSO-d_6) δ = 7.09, 7.11 (s,s, 2H,2H, -NH_2).

4-Amino-4-*N*-(*N*-methyl-2-pyrrolidin-2-ylidene)-1-[β-2-deoxy-3,5-*O*-bis(toluoyl)-D-*erythro*-pentofuranosyl]-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (2)

To 1.0 g (2.0 mmol) of **1** in 100 ml freshly distilled dichloromethane was added 4.0 mmol *N*-methyl-2,2-dimethoxypyrrolidine (26,27) and the reaction mixture stirred at ambient temperature for 1 h. The reaction was quenched by the addition of 1 ml water, it was dried over sodium sulfate and the solvent was removed to yield a foam that was purified by flash chromatography on silica gel using dichloromethane and a 0–2.5% gradient of methanol to yield a white solid.

Yield 1.0 g (85%). m.p. 96–97°C. R_f (dichloromethane/methanol, 95/5): 0.50. UV (methanol): λ_{max} = 233, 278 nm; λ_{min} = 209, 257 nm. HRMS (FAB) for $C_{31}H_{35}N_4O_7$ (M+ H) calculated 575.2505, found 575.2500. ^{1}H NMR ($CDCl_3$) δ = 1.69 (s, 3H, CH_3-), 2.02 (m, 2H, -CH_2-), 2.32 (s, 3H, CH_3-), 2.42 (s, 3H, CH_3-) 2.93 (s, 3H, CH_3-N), 2.35 and 3.30 (m, 2H, H2′, H2″), 3.43 (t, 2H, -CH_2-N), 4.48 (m, 1H, H4′), 4.68 (m, 2H, H5′,H5″), 5.82 (m, 1H, H3′), 6.92 (dd, 1H, H1′), 7.13 - 7.29 (m, 4H, Ar-H), 7.93 (m, 4H, Ar-H), 8.61 (s, 1H, NH) p.p.m. ^{13}C NMR ($CDCl_3$) δ = 167.1, 166.6, 165.3, 165.0, 155.0, 151.5, 144.5, 144.1, 130.5 (2C), 130.4 (2C), 129.8 (2C), 129.6 (2C), 128.0, 127.7, 94.6, 82.8, 82.6, 76.5, 65.8, 52.4, 36.0, 32.0, 30.0, 22.34, 22.3, 20.0, 10.7 p.p.m.

4-Amino-4-*N*-(*N*-methyl-2-pyrrolidin-2-ylidene)-1-(β-2-deoxy-D-*erythro*-pentofuranosyl)-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (3)

To 1.0 g (1.74 mmol) of **2** was added 40 ml of freshly prepared 0.1 M sodium methoxide and the reaction stirred at ambient temperature and monitored by TLC (5% methanol in dichloromethane). The reaction was complete after 2 h as judged by TLC. Approximately 1 g of silica gel was added to the solution, the solvent was evaporated and the resulting material was added to the top of a silica gel column packed in dichloromethane. The column was eluted with 5% methanol in dichloromethane to yield 0.56 g (95%) of a white solid assigned the structure **3**.

m.p. 124–125°C. R_f (dichloromethane/methanol, 9/1): 0.63. UV (methanol): λ_{max} = 212, 283 nm, λ_{min} = 247 nm. HRMS (EI) for $C_{15}H_{22}N_4O_5$ (M+) calculated 338.1590, found 338.1582. ^{1}H NMR ($CDCl_3$) δ = 1.68 (s, 3H, CH_3-), 2.01 (m, 2H, -CH_2-), 2.13 and 2.82 (m, 2H, H2′, H2″), 2.42 (m, 2H, -CH_2-), 2.93 (s, 3H, CH_3-N), 3.45 (t, 2H, -CH_2-), 3.69–3.86 (m, 2H, H5′, H5″), 3.93 (m, 1H, H4′), 4.70 (m, 1H, H3′), 6.82 (t, 1H, H1′), 8.58 (s, 1H, NH) p.p.m. ^{13}C NMR ($CDCl_3$) δ =165.8, 165.0, 156.0, 152.3, 95.1, 87.7, 83.0, 72.2, 62.9, 52.4, 38.4, 32.0, 29.8, 19.9, 10.5 p.p.m.

4-Amino-4-*N*-(*N*-methyl-2-pyrrolidin-2-ylidene)-1-[β-2-deoxy-5-*O*-(4,4′-dimethoxytrityl)-D-*erythro*-pentofuranosyl]-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (4)

To 340 mg (1.0 mmol) of **3** that had been co-evaporated from anhydrous pyridine twice was added 15 ml anhydrous pyridine and the solution was cooled in a dry ice–acetone bath (–40°C). To this solution was added 375 mg (1.1 mmol) of 4,4′-dimethoxytrityl chloride and the reaction mixture stirred for 6 h and gradually warmed to ambient temperature. At this point, TLC analysis indicated that the reaction was complete. The pyridine was removed by rotary evaporation (*in vacuo*) and the residue was co-evaporated from toluene twice, dissolved in dichloromethane, washed with dilute sodium bicarbonate, dried over sodium sulfate and the solvents were removed. The product was purified by flash chromatography on silica gel using a gradient of 0–2.5% methanol in dichloromethane to yield 0.55 g (85.8%) of a white solid assigned the structure **4**.

m.p. 133–134°C. R_f (dichloromethane/methanol, 95/5): 0.32. UV (methanol): λ_{max} = 225, 278 nm; λ_{min} = 253 nm. HMRS (FAB) for $C_{36}H_{41}N_4O_7$ (M+ H) calculated 641.2975, found 641.2979. ^{1}H NMR ($CDCl_3$) δ = 1.66 (s, 3H, CH_3-), 1.98 (m, 2H, -CH_2-), 2.18 and 2.82 (m, 2H, H2′, H2″), 2.36 (m, 2H, -CH_2-), 2.91 (s, 3H, CH_3-N), 3.32 and 3.54 (m, 2H, H5′, H5″), 3.42 (m, 2H, -CH_2-), 3.78 (s, 6H, CH_3-O), 3.83 (m, 1H, H4′), 4.69 (m, 1H, H3′), 6.71 (dd, 1H, H1′), 6.92–7.46 (m, 13H, Ar-H), 7.98 (br s, 1H, NH) p.p.m. ^{13}C NMR ($CDCl_3$) δ = 165.3, 164.9, 159.1 (2C), 154.5, 151.3, 145.6, 136.8, 130.8 (4C), 128.9 (2C), 128.4 (2C), 127.4 (2C), 113.8 (4C), 94.7, 87.2, 85.0, 81.7, 74.5, 65.8, 55.9 (2C), 52.3, 38.3, 32.0, 29.9, 20.1, 10.7 p.p.m.

4-Amino-4-*N*-(*N*-methyl-2-pyrrolidin-2-ylidene)-1-{β-2-deoxy-3-*O*-[(2-cyanoethoxy)diisopropylaminophosphino]-5-*O*-(4,4′-dimethoxytrityl)-D-*erythro*-pentofuranosyl}-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (5)

To 321 mg (0.5 mmol) of **4** that had been dried overnight in a vacuum oven and then dissolved in 20 ml of freshly distilled dichloromethane, was added 0.5 ml (2.87 mmol) of *N*,*N*-diisopropylethylamine and 237 mg (1 mmol) of 2-cyanoethyl *N*,*N*-diisopropyl-chlorophosphoramidite and the reaction mixture stirred at ambient temperature for 1.5 h. The reaction was stopped by the addition of 1 ml of methanol and washed with dilute sodium bicarbonate solution, dried over sodium sulfate and the solvents were removed. The product was purified by flash chromatography on silica gel using 5% methanol in dichloromethane containing a trace of triethylamine. The product was precipitated into hexane to yield 380 mg (90%) of **5** as a pale yellow solid.

R_f (dichloromethane/methanol, 95/5): 0.40. ^{31}P NMR ($CDCl_3$) δ = 148.5, 148.6 p.p.m.

DNA synthesis

The 15mers containing $m^{5ox}C$ as well as the native 25mers were prepared by solid phase DNA synthesis under standard conditions. The $m^{5ox}C$ analog could be incorporated into DNA strands with coupling efficiencies that were comparable with those of common nucleoside phosphoramidites. Deprotection of the *N*-methyl-2-pyrrolidine amidine required a 16 h reaction in concentrated aqueous ammonia at 60°C.

Purification of the oligonucleotides was by HPLC (trityl-on) using 50 mM triethylammonium acetate, pH 7.0, and a gradient of acetonitrile (20–65% over 40 min). The collected DMT-protected oligonucleotides were reduced in volume, detritylated with 80% aqueous acetic acid (30 min, 0°C), desalted (Sephadex G-10).

Nucleoside analysis

Oligomers containing $m^{5ox}C$ could not be fully digested with snake venom phosphodiesterase but were effectively digested with S1 nuclease/calf intestinal alkaline phosphatase into monomeric units: a 10 µl reaction mixture containing 0.5 A_{260} units of

oligomer in 200 mM sodium chloride, 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 25 mM sodium acetate, pH 5.5, was incubated for 5 min at room temperature with 267 U S1 nuclease. To this mixture was added 5 µl of 0.1 M Tris–HCl, pH 8.0, and 1 U calf intestinal alkaline phosphatase and the reaction incubated for an additional 60 min at ambient temperature. An aliquot of this mixture was analyzed by HPLC (4.6 × 250 mm column of ODS-Hypersil, 20 mM potassium phosphate, pH 5.5) and resulted in the elution of $m^{5ox}C$ with a retention time of 20.2 min and T with a time of 21.2 min (0–35% methanol over 3 h).

Thermal denaturation studies

Thermal denaturation studies were performed in 10 mM PIPES, pH 6.4, 10 mM PIPES, pH 7.0, 10.5 mM HEPES, pH 7.5, 10.5 mM HEPES, pH 8.0, or 10.5 mM HEPES, pH 8.5, containing 10 mM magnesium chloride and 50 mM sodium chloride at triplex concentrations in the low micromolar range (~1 µM). Absorbance (260 nm) and temperature values were measured with an AVIV 14DS UV/Visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 0.5°C steps (from 0 to 95°C) and when thermal equilibrium was reached, temperature and absorbance data were collected. T_m values were determined both from first order derivatives and by graphical analysis of the absorbance versus temperature plots.

RESULTS AND DISCUSSION

In addition to a straightforward synthetic route from relatively simple starting materials, it appeared to us that an effective nucleoside analog capable of forming stable base triplets with target G-C base pairs using the parallel-stranded recognition motif should maintain three characteristics observed for the T residues present in T-A-T base triplets: (i) it should be a pyrimidine-like ring system, since using isomorphous base triplets for targeting both A-T and G-C base pairs should minimize anomalous conformational changes in the backbone of the triplex strand; (ii) it should function as a bidentate hydrogen bond donor without requiring protonation (see Fig. 1a) in order to interact effectively with the O^6 oxygen and N^7 nitrogen of the target G-C base pair in a pH-independent manner; (iii) it should ideally maintain the 5-methyl group that has been shown to enhance stability in the m^5C^+-G-C triplet. In the presence of neighboring T-A-T base triplets, the presence of a methyl group on each base triplet may contribute to the formation of an important, stabilizing 'hydrophobic spine' (7). One analog that fits these three criteria is 4-amino-1-(β-2′-deoxy-D-*erythro*-pentofuranosyl)-5-methyl-2,6-[1*H*,3*H*]-pyrimidione (abbreviated simply as $m^{5ox}C$), which is the 6-keto derivative of m^5C and is shown in Figure 1b in the putative $m^{5ox}C$-G-C base triplet isomorphous with m^5C^+-G-C. We have described some of the preliminary studies of oligonucleotide-directed triple helix formation in an earlier communication (13). One of the concerns for the effectiveness of this pyrimidine analog in forming two pH-independent Hoogsteen hydrogen bonds with target G-C base pairs is the tautomeric ambiguity present. Of the possible tautomers for this base residue, N^3-H, O^2-H or O^6-H (Fig. 1c), the N^3-H tautomer is the desired structure for triplex formation in the parallel-stranded pyrimidine–purine–pyrimidine motif and it appears to be the preferred tautomeric form based upon single crystal X-ray analysis of the related 4-amino-1-(β-D-ribofuranosyl)-2,6-[1*H*,3*H*]-pyrimidione (28).

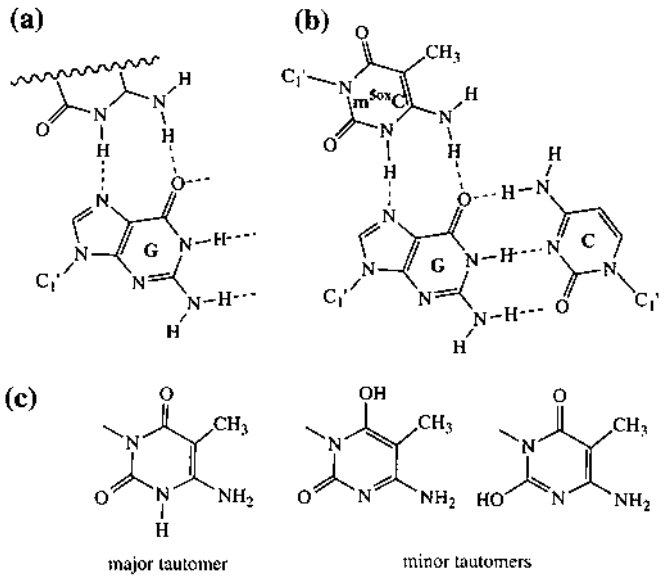
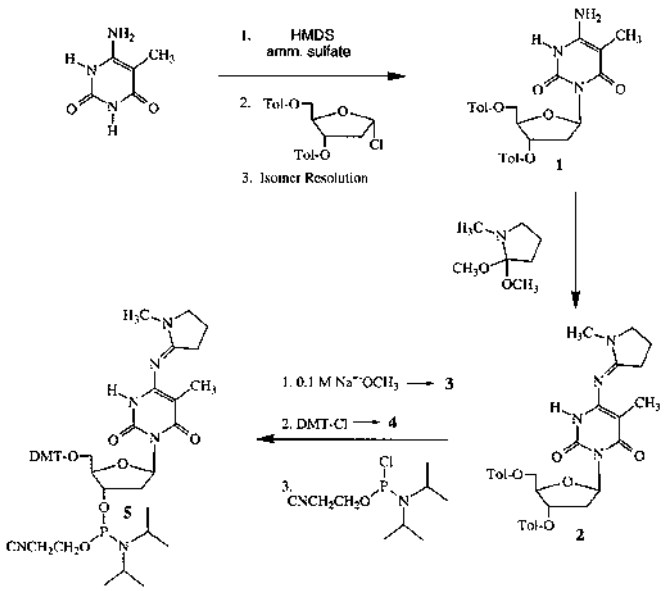


Figure 1. (**a**) General scheme for a bidentate hydrogen bond donor for the recognition of G-C base pairs by the parallel-stranded motif. (**b**) Putative $m^{5ox}C$-G-C base triplet. (**c**) Major and minor tautomeric forms of the $m^{5ox}C$ base analog.

Nucleoside synthesis

The synthesis of the phosphoramidite derivative of $m^{5ox}C$ was accomplished generally with high yields (Scheme 1). The heterocyclic base (6-aminothymine) was prepared by a simple condensation of urea with ethyl 2-cyanopropionate (at a 56 g scale). Glycosylation of the silylated 6-aminothymine resulted in two regioisomeric products. The major nucleoside produced (84%) was the N^3-glycosylation product, $m^{5ox}C$ (α and β anomers), while the minor product (16%) appeared to be the N^1-glycosylation product, 6-aminothymidine (α and β anomers). The N^1- and N^3-glycosylation products could be easily separated by silica gel chromatography and subsequently differentiated on the basis of the NMR characteristics of the amino resonances, as has been described for a number of alkylated 6-aminouracil derivatives (25). The preference for glycosylation at the nitrogen flanked by the two carbonyl residues in 4-amino (or 4-thio) 2,6-pyrimidiones is well documented (29,30).

Resolution of the α and β anomers could not be achieved in our hands for any of the derivatives illustrated in Scheme 1. It was therefore necessary to introduce an additional protection/deprotection step into the synthetic scheme that would permit such separation to occur. After examining the products of a number of protecting group reactions, we discovered that reaction of the anomeric mixture of **1** with *o*-nitrosulfenyl chloride generated two easily resolvable compounds, based upon TLC analysis of the reaction mixture. NMR analysis of the faster migrating material revealed that it was, in fact, two different bis(*o*-nitrosulfenyl)-protected derivatives that could neither be resolved nor adequately identified. However, upon removal of the *o*-nitrosulfenyl groups (thiocresol/pyridine) only the β anomeric nucleoside was recovered. In similar fashion, the slower migrating material also represented two bis(*o*-nitrosulfenyl)-protected derivatives and could be deprotected to produce the α anomeric nucleoside. Both the



Scheme 1.

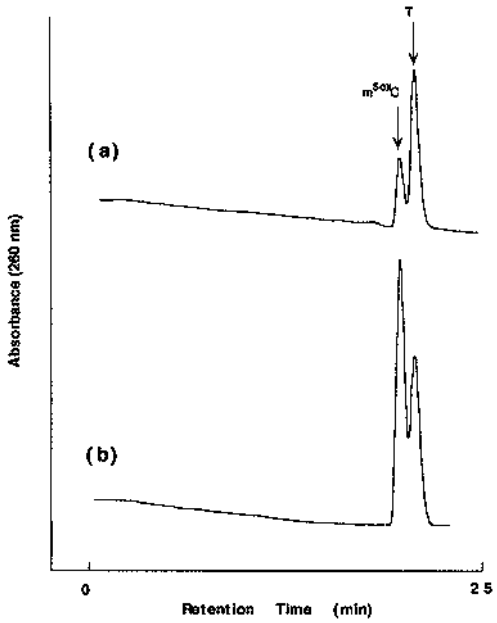


Figure 2. (**a**) HPLC analysis of a 15 residue oligonucleotide containing four $m^{5ox}C$ and eleven T nucleosides after digestion with nuclease P1 and alkaline phosphatase (for digestion and HPLC conditions see Materials and Methods). (**b**) Analysis as described in (a) after the addition of authentic $m^{5ox}C$.

regioisomeric and anomeric character of these two nucleosides were assigned on the basis of NMR experiments. Using these procedures, a 37% overall yield of the β anomeric N^3-glycosylated nucleoside was obtained.

Protection of the exocylic amino group of **1** as the isobutyryl or benzoyl amide failed under all conditions attempted. By comparison, the dimethylformamidine derivative of **1** could be easily prepared, but it was not stable to even the mildest conditions necessary to remove the toluoyl esters (e.g. **2**→**3**, see Scheme 1). The *N*-methyl-2-pyrrolidine amidine derivative (26,27) was the protecting group of choice for the amino group and it was stable during the removal of the toluoyl esters to generate **3** in high yield. Conversion of **3** to the phosphoramidite **5** could be achieved using standard procedures (31).

Oligonucleotide synthesis

The 15 and 25 residue oligonucleotides were prepared by solid-phase phosphite triester synthesis (31,32) on a wide-pore silica support. Incorporation of the $m^{5ox}C$ residue into the growing oligonucleotide chain occurred with yields comparable with those obtained for the common monomers based upon the color of the liberated DMT cation. Deprotection of $m^{5ox}C$-containing sequences required a 16 h reaction in concentrated ammonia at 60°C in order to fully remove the *N*-methyl pyrrolidine amidine protecting group. Isolation of the sequences proceeded by HPLC and PAGE as required. The isolated oligonucleotides were estimated to be >98% pure based upon HPLC analyses. Oligonucleotides containing $m^{5ox}C$ could not be fully digested by snake venom phosphodiesterase, but could be cleaved into the constituent nucleoside components by a combination of S1 nuclease and alkaline phosphatase. Resolution of the nucleoside mixture obtained from such digests could be achieved by HPLC and confirmed the presence of the analog $m^{5ox}C$ residues (see Fig. 2).

Triplex studies

For the present study, we employed a 15 residue purine sequence embedded within a 25 nucleotide target duplex. Examining a 15 residue triplex in the presence of a 25mer duplex permitted effective separation of the triplex and duplex transitions under conditions of thermal denaturation. Samples containing both the duplex target and the triplex strand typically exhibited two transitions (for examples see Fig. 3), while samples without the triplex strand exhibited only the second transition characterizing denaturation of the DNA duplex.

The initial duplex target examined in this study contained four target G-C base pairs and 11 target A-T base pairs, with each of the G-C base pairs isolated by one or more A-T base pairs (see Table 1). Three triplex strands were examined for oligonucleotide-mediated triplex formation. In all cases T residues were employed for the recognition of A-T base pairs using the parallel-stranded T-A-T triplet recognition motif. One of the three complexes employed C residues for recognition of the G-C base pairs, a second employed m^5C residues and the third relied upon the analog $m^{5ox}C$ for recognition of the G-C targets. The results of thermal denaturation studies performed at pH values of 6.4, 7.0, 7.5, 8.0 and 8.5 in the presence and absence of spermine are illustrated in Table 1 [the data obtained in the absence of spermine was previously reported (13) but is included for comparative purposes]. As expected, in the samples where C or m^5C were employed to recognize G-C base pairs, a distinct pH dependence is observed for the triplex T_m values (Table 1). On the other hand there is virtually no variation in the T_m for the $m^{5ox}C$-containing triplex over the pH range 6.4–8.5. At the acidic pH value (6.4) both the C-containing and m^5C-containing triplexes exhibit higher T_m values than does the analog-containing sequence. This

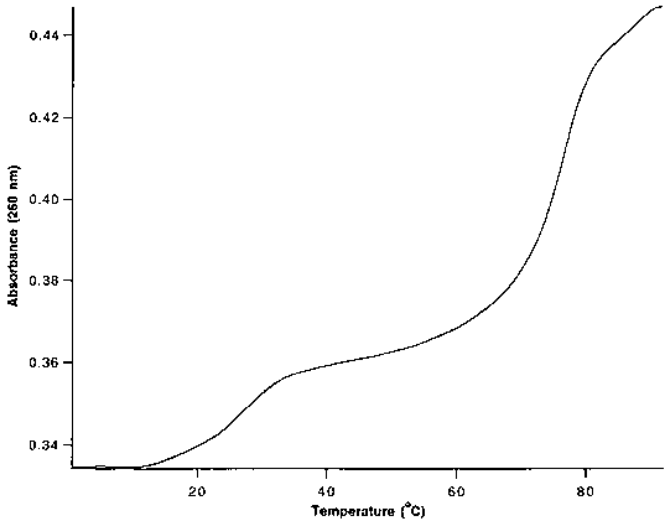


Figure 3. Thermally induced transitions for the triplex containing five contiguous target G-C base pairs and a third strand composed of m^5C and $m^{5ox}C$ residues at pH 6.4 in the presence of spermine.

Table 1. Effect of spermine on the T_m values for C-, m^5C- and $m^{5ox}C$-containing 15mers targeting isolated G-C base pairs

```
5′      TTT X TTTT X T X T X TT      3′
5′  GCGCGAAA G AAAA G A G A G AACCCGG  3′
3′  CGCGCTTT C TTTT C T C T C TTGGCCC  5′
```

X	[Spermine][a]	pH Value 6.4	7.0	7.5	8.0	8.5
C[b]	-	35.4	25.0	19.5	12.1	10.8
C	0.5 mM	45.1	32.6	26.9	15.8	nt[c]
m^5C[b]	-	42.9	34.6	33.0	28.7	26.3
m^5C	0.5 mM	54.6	42.6	37.9	27.2	14.4
$m^{5ox}C$[b]	-	26.9	27.0	27.5	27.3	26.3
$m^{5ox}C$	0.5 mM	39.1	40.7	39.7	38.0	30.8

[a]Buffer consists of 50 mM NaCl, 10 mM $MgCl_2$ and 10 mM PIPES (pHs 6.4 and 7.0) or 10.5 mM HEPES (pHs 7.5, 8.0 and 8.5).
[b]T_m values for this series were reported in Xiang *et al.* (13). These data are provided for comparative purposes only.
[c]nt, no triplex transition observed.

effect may be due to the presence of positive charge at the protonated C (or m^5C) residues that could provide an important shielding effect as the three polyionic strands are assembled; it is well documented that multivalent cations or spermine enhance triplex stability (33–36). As the pH value increases, the C or m^5C complexes would then be destabilized, in part due to the loss of hydrogen bonding functionality at the N^3 nitrogen of the C (m^5C) residues and in part due to the loss of compensating charge effects. The $m^{5ox}C$ analog lacks any compensating positive charge effects, but with the presence of a pH-independent hydrogen bond donor at the pyrimidine N^3 nitrogen, stabilized triple helices are formed equally well at pH 6.4 or 8.5.

The addition of spermine significantly alters the relative stability for the three sets of triple helices. Both the C-containing and m^5C-containing triplexes exhibit an increased T_m value of 3–4°C in the presence of spermine. By comparison, the $m^{5ox}C$ triplex exhibits an 11°C increase in T_m. These observations are consistent with the ability of spermine to shield the charged phosphates and enhance triplex stability, as has been documented (33–36). The spermine effects are less dramatic with the C- and m^5C-containing sequences because at pH 6.4 these residues are largely protonated and already assist in limiting destabilization as a result of charge repulsion effects. The neutral T-A-T and $m^{5ox}C$-G-C-containing triplex benefits more dramatically from the effects of spermine shielding on triplex stabilization.

As noted above, it is possible for the $m^{5ox}C$ analog to adopt more than one tautomeric structure (see Fig. 1c). Although the crystal structure of the related 4-amino-1-(β-D-ribofuranosyl)-2,6-[1*H*,3*H*]-pyrimidione (28) suggests that the N^3-H tautomer is preferred, hydrogen bonding effects can be altered in the solid state. To further judge the tautomeric character of the $m^{5ox}C$ analog, we prepared the Watson–Crick duplex sequence illustrated below, where X = $m^{5ox}C$:

```
5′T  T T X T T T T X T X T X T T3′
3′A  A A G A A A A G A G A G A A5′
```

If the $m^{5ox}C$ residue exists in solution predominately in either the O^2-H or O^6-H tautomer, effective base pairing with G should occur. However, as noted in our earlier communication (13), this complex did not exhibit any cooperative thermally induced transition, suggesting the absence of any duplex structure. This result can be explained if the preferred tautomeric structure of the $m^{5ox}C$ residue is the N^3-H tautomer (see 23), which would result in helix destabilization when paired with G, since the N^3-H of $m^{5ox}C$ and the N^1-H of G sterically oppose one another and thus prevent Watson–Crick base pairing, ultimately disrupting the helical structure.

In order for a base analog such as $m^{5ox}C$ to function effectively, it must be able to discriminate between the cognate and the non-cognate target sequence. To judge the effectiveness of recognizing G-C base pairs over other possibilities, we prepared a 25mer target duplex which differed from the initial complex by a single base pair. In this series of target duplexes one of the G-C base pairs was substituted by A-T, C-G or T-A (see Table 2). T_m measurements were performed at a single pH value of 7.0 with or without the polyamine spermine. In the absence of spermine, the only target base pair (other than G-C) that exhibited an identifiable triplex transition, using $m^{5ox}C$ as the third residue, was that of C-G. The complex containing the $m^{5ox}C$-C-G triplet resulted in a T_m of 13.5°C (by comparison, the complex with the cognate $m^{5ox}C$-G-C triplet exhibited a T_m value of 27°C). The presence of spermine enhanced the stability of each of the three triplexes containing non-cognate base triplets. T_m values of 17.7, 18.9 and 23.6°C were observed when a single target A-T, T-A or C-G base pair, respectively, was targeted by $m^{5ox}C$. With the cognate $m^{5ox}C$-G-C base triplet at this position, a T_m value of 40.2°C was obtained in the presence of spermine. In the presence or absence of spermine there is a minimum difference in T_m values of 13–14°C between complexes containing a single non-G-C pair targeted by the $m^{5ox}C$ residue.

In addition to the noted pH effects observed for C (or m^5C) triple helices, it has been difficult to use the C^+-G-C recognition motif to target continuous G-C base pairs. The necessity for protonation at a series of adjacent base triplets likely leads to destabilizing charge–charge repulsion effects. To examine the ability of the $m^{5ox}C$ analog to target sequences with continuous

Table 2. Effect of single mismatched base triples on the T_m values for $m^{5ox}C$-containing 15mers targeting isolated G-C base pairs

```
5′        TTT C·TTTT C·T X T C·TT        3′
5′  GCGCGAAA G AAAA G A Y A G AACCCGG    3′
3′  CGCGCTTT C TTTT C T Z T C TTGGCCC    5′
```

C′	X	Y	Z	[Spermine][a]	T_m[b]
$m^{5ox}C$	$m^{5ox}C$	G	C	-	27.0[c]
$m^{5ox}C$	$m^{5ox}C$	G	C	0.5 mM	40.7
$m^{5ox}C$	$m^{5ox}C$	C	G	-	13.5
$m^{5ox}C$	$m^{5ox}C$	C	G	0.5 mM	23.6
$m^{5ox}C$	$m^{5ox}C$	A	T	-	nt[d]
$m^{5ox}C$	$m^{5ox}C$	A	T	0.5 mM	17.7
$m^{5ox}C$	$m^{5ox}C$	T	A	-	nt
$m^{5ox}C$	$m^{5ox}C$	T	A	0.5 mM	18.9

[a]Buffer consists of 50 mM NaCl, 10 mM $MgCl_2$ and 10 mM PIPES pH 7.0.
[b]T_m values are the average of at least two measurements.
[c]Reported in Xiang *et al.* (13). This data is provided for comparative purposes only.
[d]nt, no triplex transition observed.

Table 3. Effect of spermine on the T_m values for m^5C- and $m^{5ox}C$-containing 15mers targeting contiguous G-C base pairs

```
5′        TTT X TTTT X Y X Y X TT        3′
5′  GCGCGAAA G AAAA G G G G G AACCCGG    3′
3′  CGCGCTTT C TTTT C C C C C TTGGCCC    5′
```

			pH Value				
X	Y	[Spermine][a]	6.4	7.0	7.5	8.0	8.5
m^5C	m^5C	-	10.8	nt[b]	nt	nt	nt
m^5C	m^5C	0.5 mM	17.7	nt	nt	nt	nt
$m^{5ox}C$	$m^{5ox}C$	-	nt	nt	nt	nt	nt
$m^{5ox}C$	$m^{5ox}C$	0.5 mM	15.2	15.4	15.2	13.6	nt
m^5C	$m^{5ox}C$	-	18.9	13.1	12.3	12.3	nt
m^5C	$m^{5ox}C$	0.5 mM	27.8	15.4	14.7	16.0	nt

[a]Buffer consists of 50 mM NaCl, 10 mM $MgCl_2$ and 10 mM PIPES (pHs 6.4 and 7.0) or 10.5 mM HEPES (pHs 7.5, 8.0 and 8.5).
[b]nt, no triplex transition observed.

G-C base pairs, we prepared a second 25mer duplex with five continuous G-C base pairs and a total of six G-C pairs within the 15mer target (Table 3). A third strand containing six C residues did not exhibit any triplex transition at any pH value tested (data not shown). At pH 6.4, the corresponding sequence containing m^5C residues exhibited a transition at 10.8°C (see Table 1), but similar transitions were not observed at any higher pH values. Both of these results are consistent with early work describing the difficulty in targeting continuous G-C base pairs with m^5C as the third base residue (9).

In spite of the success in using the $m^{5ox}C$ analog in the triplex studies illustrated in Table 1, we were unable to detect any significant triplex transition for the sequence containing six $m^{5ox}C$ residues targeting the contiguous G-C base pair sequence (Table 3) in the absence of spermine. Gel shift assays (data not shown) confirmed that triplex formation under these conditions did not occur. This lack of complex formation may reflect significant charge–charge repulsion coupled with poor base stacking effects between adjacent $m^{5ox}C$ residues, since stable three-stranded complexes are formed when the $m^{5ox}C$-G-C base triplets alternate with T-A-T base triplets. When spermine (0.5 mM) was added to the mixture, a moderate increase in T_m value was observed for the triplex containing five contiguous G-C base pairs targeted by m^5C at pH values of 6.4 (Table 3). In the presence of spermine, the $m^{5ox}C$-containing triplex resulted in measurable T_m values of ~15°C over the range pH 6.4–8.0 (Table 3). To further address the issue of targeting contiguous G-C base pairs, we prepared a sequence in which the third strand contained alternating m^5C and $m^{5ox}C$ residues (Table 3). This sequence resulted in the highest T_m observed in this study for a triplex containing a target of five contiguous G-C base pairs in the absence of spermine. The T_m of 18.9°C at pH 6.4 was 8°C higher than that obtained using m^5C residues. At pH values of 7.0 and 7.5 the triplex transition for this complex decreased to 13.1 and 12.3°C (see Table 3), respectively, consistent with the loss of protonation of the N^3 residues of m^5C as the pH increased. In the presence of spermine, the T_m value at pH 6.4 was increased to 27.8°C, some 10°C higher than that observed for either the sequence containing five contiguous m^5C or $m^{5ox}C$ residues. The T_m value is observed to drop from 27.8 to 15.4°C at pH 7.0, consistent with the need to protonate the m^5C residues. Triplex transitions were still observed at pH values of 7.5 and 8.0, suggesting that the $m^{5ox}C$ residues can still contribute towards the formation of triplex structures even in the absence of low pH conditions.

Although there are a number of base analogs that have been employed as possible replacements of C or m^5C for the pH-independent recognition of G-C base pairs, relatively few have been examined for their ability to function at target sequences containing contiguous G-C base pairs. The purine base analog 8-oxoadenine appears to form triplexes with targets containing as many as four contiguous G-C base pairs that do not result in any hyperchomicity but are relatively pH independent, based upon gel shift assays (19). The purine-like pyrazolopyrimidinone has also been used to target contiguous G-C base pairs (37). This analog was observed to bind a sequence containing six contiguous G-C base pairs with much higher affinity than m^5C. However, this same derivative was less effective than m^5C for the targeting of isolated G-C base pairs. The 2′-*O*-methyl derivative of the pyridine pseudoisocytidine has been reported to form triplexes with a 16mer target sequence containing six contiguous G-C base pairs with a T_m value of 22°C (11,12). The present $m^{5ox}C$ analog functions well for the targeting of isolated G-C base pairs and is reasonably effective when used in combination with m^5C for the targeting of contiguous G-C base pairs. However, it is less effective when used by itself. The low stability of triplexes containing adjacent m^5C residues could result from charge–charge repulsion by the adjacent protonated residues. Similar effects should not be present in adjacent $m^{5ox}C$ residues, but the additional carbonyl at the 6 position may result in undesirable steric effects or poor base stacking between the analog residues. The presence of spermine permits triplex formation with adjacent $m^{5ox}C$ residues that are relatively pH independent but of only moderate stability.

In summary, we have described a short and relatively simple synthesis of a pyrimidine analog that permits the formation of two

pH-independent hydrogen bonds with target G-C base pairs. Triplexes containing the isolated $m^{5ox}C$ analog in place of C or m^5C exhibit pH-independent triplex transitions over the pH range examined (6.4–8.5). Neither m^5C nor $m^{5ox}C$ residues alone can be used effectively to target contiguous sequences of G-C base pairs in the absence of spermine, but when used alternately or in the presence of spermine triplex formation occurs at sites containing multiple G-C residues.

ACKNOWLEDGEMENTS

We thank Walid Soussou for performing the noted gel shift experiments. This work was supported by a grant from the NIH (GM53201). LWM is the recipient of an American Cancer Society Faculty Research Award (FRA-384).

REFERENCES

1 Dervan,P.B. (1989) In Cohen,J.S. (ed.), *Oligonucleotides. Antisense Inhibitors of Gene Expression.* Macmillan, London, UK, pp. 197–210.
2 Thuong,N.T. and Helene,C. (1993) *Angew. Chem. Int. Edn English*, **32**, 666–690.
3 Hoogsteen,K. (1959) *Acta Crystallogr.*, **12**, 822–825.
4 Felsenfeld,G., Davies,D.R. and Rich,A. (1957) *J. Am. Chem. Soc.*, **79**, 2023–2025.
5 Rajagopal,P. and Feigon,J. (1989) *Biochemistry*, **28**, 7859–7870.
6 de los Santos,C., Rosen,M. and Patel,D. (1989) *Biochemistry*, **28**, 7282–7289.
7 Povsic,T.J. and Dervan,P.B. (1989) *J. Am. Chem. Soc.*, **111**, 3059–3061.
8 Lavelle,L. and Fresco,J.R. (1995) *Nucleic Acids Res.*, **23**, 2692–2705.
9 Lee,J.S., Woodsworth,M.L., Latimer,L.J.P. and Morgan,A.R. (1984) *Nucleic Acids Res.*, **12**, 6603–6614.
10 Collier,D.A., Thuong,N.T. and Helene,C. (1991) *J. Am. Chem. Soc.*, **113**, 1457–1458.
11 Ono,A., Ts'o,P.O.P. and Kan,L.-S. (1992) *J. Org. Chem.*, **57**, 3225–3231.
12 Ono,A., Ts'o,P.O.P. and Kan,L.-S. (1991) *J. Am. Chem. Soc.*, **113**, 4032–4033.
13 Xiang,G., Soussou,W. and McLaughlin,L.W. (1994) *J. Am. Chem. Soc.*, **116**, 11155–11156.
14 Berressem,R. and Engels,J.W. (1995) *Nucleic Acids. Res.*, **23**, 3465–3472.
15 von Krosigk,U. and Benner,S.A. (1995) *J. Am. Chem. Soc.*, **117**, 5361–5362.
16 Priestley,E.S. and Dervan,P.B. (1995) *J. Am. Chem. Soc.*, **117**, 4761–4765.
17 Koh,J.S. and Dervan,P.B. (1992) *J. Am. Chem. Soc.*, **114**, 1470–1478.
18 Hunziker,J., Priestley,E.S., Brunar,H. and Dervan,P.B. (1995) *J. Am. Chem. Soc.*, **117**, 2661–2662.
19 Jetter,M.C. and Hobbs,F.W. (1993) *Biochemistry*, **32**, 3249–3254.
20 Miller,P.S., Bhan,P., Cushman,C.D. and Trapane,T.L. (1992) *Biochemistry*, **31**, 6788–6793.
21 Krawczyk,S.H., Milligan,J.F., Wadsani,S., Moulds,C., Froehler,B.C. and Matteucci,M.D. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3761–3764.
22 Cain,C.K., Mallette,M. and Taylor,E.C. (1946) *J. Am. Chem. Soc.*, **68**, 1996–1999.
23 Editorial (1946) *Chem. Abstr.*, **70**, P78006w.
24 Hoffer,M. (1960) *Chem. Ber.*, 2777–2781.
25 Müller,C.E. (1991) *Tetrahedron Lett.*, **32**, 6539–6540.
26 McBride,L.J., Kierzek,R., Beaucage,S.L. and Caruthers,M.H. (1986) *J. Am. Chem. Soc.*, **108**, 2040–2048.
27 McBride,L.J. and Caruthers,M.H. (1983) *Tetrahedron Lett.*, **29**, 2953–2956.
28 Gorski,J. and Tollin,P. (1982) *Cryst. Struct. Commun.*, *11*, 543–546.
29 Noell,C.W. and Robins,R. (1964) *J. Heterocyclic Chem.*, **1**, 34–41.
30 Winkley,M.W. and Robins,R.K. (1969) *J. Chem. Soc.*, 791–796.
31 Beaucage,S.L. and Iyer,R.P. (1993) *Tetrahedron*, **49**, 6123–6194.
32 Matteucci,M.D. and Caruthers,M.H. (1981) *J. Am. Chem. Soc.*, **103**, 3185–3191.
33 Mergny,J.S., Duval-Valentin,G., Nguyen,C.H., Perrouault,L., Faucon,B., Rougee,M., Montenay-Garesteier,T.M., Bisagani,E. and Helene,C. (1992) *Science*, **256**, 1681–1684.
34 Moser,H.E. and Dervan,P.B. (1987) *Science*, **238**, 645–650.
35 Kohwi,Y. and Kohwi-Shigematsu,T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3781–3785.
36 Hampel,K.J., Crosson,P. and Lee,J.S. (1991) *Biochemistry*, **30**, 4455–4459.
37 Priestley,E.S. and Dervan,P.B. (1995) *J. Am. Chem. Soc.*, **117**, 4761–4765.