© 1997 Oxford University Press
Nucleic Acids Research, 1997, Vol. 25, No. 14 2773–2783

DNA containing 4′-thio-2′-deoxycytidine inhibits methylation by *Hha*I methyltransferase

S. Kumar*, J. R. Horton[1,4], G. D. Jones[2], R. T. Walker[3], R. J. Roberts and X. Cheng[1,4]

New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA, [1]W. M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratories, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA, [2]Welsh School of Pharmacy, University of Wales Cardiff, Redwood Building, King Edward VII Avenue, Cardiff CF1 3XF, UK, [3]School of Chemistry, University of Birmingham, Birmingham B15 2TT, UK and [4]Department of Biochemistry, Room G239, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

Received April 3, 1997; Revised and Accepted May 30, 1997

ABSTRACT

4′-Thio-2′-deoxycytidine was synthesized as a 5′-protected phosphoramidite compatible with solid phase DNA synthesis. When incorporated as the target cytosine (C*) in the GC*GC recognition sequence for the DNA methyltransferase M.*Hha*I, methyl transfer was strongly inhibited. In contrast, these same oligonucleotides were normal substrates for the cognate restriction endonuclease R.*Hha*I and its isoschizomer R.*Hin*P1I. M.*Hha*I was able to bind both 4′-thio-modified DNA and unmodified DNA to equivalent extents under equilibrium conditions. However, the presence of 4′-thio-2′-deoxycytidine decreased the half-life of the complex by >10-fold. The crystal structure of a ternary complex of M.*Hha*I, AdoMet and DNA containing 4′-thio-2′-deoxycytidine was solved at 2.05 Å resolution with a crystallographic R-factor of 0.186 and R-free of 0.231. The structure is not grossly different from previously solved ternary complexes containing M.*Hha*I, DNA and AdoHcy. The difference electron density suggests partial methylation at C5 of the flipped target 4′-thio-2′-deoxycytidine. The inhibitory effect of the 4′ sulfur atom on enzymatic activity may be traced to perturbation of a step in the methylation reaction after DNA binding but prior to methyl transfer. This inhibitory effect can be partially overcome after a considerably long time in the crystal environment where the packing prevents complex dissociation and the target is accurately positioned within the active site.

INTRODUCTION

The DNA cytosine-5 methyltransferases (C5-Mtases) are model systems for enzymes that integrate catalytic activity with molecular recognition of nucleic acid substrates and have been the focus of a renewed interest in DNA methylation (reviewed in 1,2). These enzymes are found in both prokaryotes and eukaryotes. In bacteria, C5-Mtases function as a component of a simple immune system in which cytosine modification is used to distinguish 'self' from foreign DNA. Host sequences that are methylated are resistant to digestion by the host's restriction endonucleases, whereas unmodified infecting viral DNA is left susceptible to degradation (see 3 for a recent review). In eukaryotes, the role of cytosine methylation is more complex and less well understood. DNA methylation has been implicated in a variety of processes including genomic imprinting (4), embryonic development (5) and gene expression (reviewed in 6). 5-Methylcytosine and C5-Mtases themselves appear to be endogenous mutagens and may play a role in disease (7,8).

The bacterial enzyme M.*Hha*I, which recognizes the sequence GCGC and methylates the inner cytosine, is one of the most thoroughly examined C5-Mtases. The gene for this protein has been overexpressed in *Escherichia coli* to allow large-scale production and purification of the protein and subsequent crystallization (9). The structure of this enzyme has been solved in the absence of DNA (10) as well as in complex with both modified and unmodified DNA substrates (11–13). The latter DNA–protein co-crystal structures revealed that binding is accompanied by a novel distortion of the DNA at the binding site: complete extrusion of the base to be methylated out of the DNA helix. This process, termed base flipping, has subsequently been confirmed in the structure of another C5-Mtase, M.*Hae*III (14) as well as two DNA repair enzymes, T4 endonuclease V (15) and uracil–DNA glycosylase (16). Base-flipping may be a fundamental mode of DNA–protein interaction as several other enzymes are likely to use this mechanism (17,18).

An understanding of the catalytic mechanism of C5 methylation used by M.*Hha*I has emerged from the structures cited above as well as a variety of biochemical studies which have identified residues required for catalysis (19,20). Methylation by M.*Hha*I and other C5-Mtases requires the target base to be inserted into the catalytic pocket of the enzyme where the reaction proceeds via the formation of a transient covalent bond between the sulfhydryl group of a conserved cysteine (Cys81 in M.*Hha*I) and C6 of the target cytosine (Fig. 1). The initial covalent bond enhances the reactivity of C5 allowing nucleophilic attack by C5 on the methyl group of the methyl-donating cofactor AdoMet. The high energy carbanion generated at C5 is either resonance stabilized (19,21) or entirely avoided by protonation of N3 by the nearby protonated

*To whom correspondence should be addressed. Tel: +1 508 927 5054; Fax: +1 508 921 1527; Email: kumar@neb.com

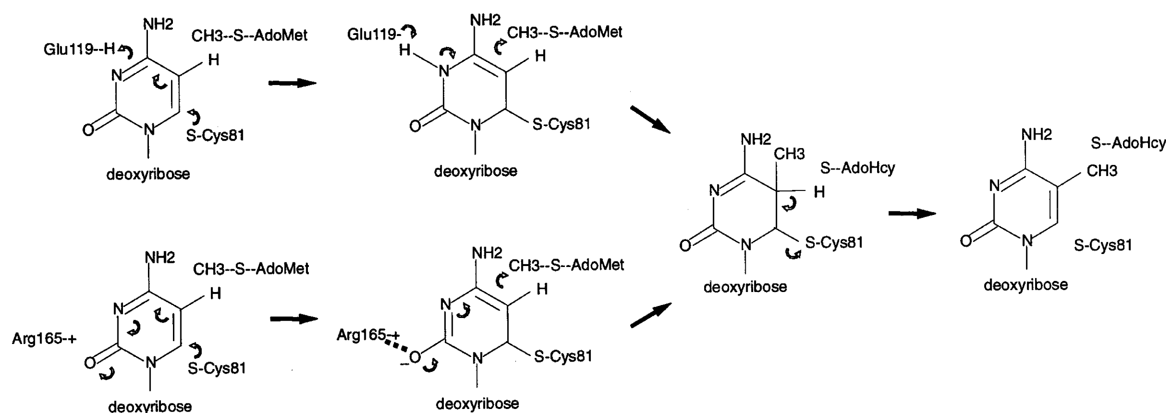


Figure 1. Reaction mechanism of methylation at cytosine-5. The mechanism is adapted from Chen *et al.* (22) (upper path) and Gabbara *et al.* (21) (lower path). Amino acid residue numbers are based on the M.*Hha*I sequence.

glutamic acid residue (Glu119 in M.*Hha*I) (22). Both alternatives allow for the formation of a reactive enamine. AdoMet is converted to AdoHcy by the methyl transfer and elimination of the C5 proton by a water molecule allows dissolution of the covalent complex (12,13).

Nucleotide analogs, when incorporated into synthetic oligonucleotides, provide powerful tools for probing the structural and chemical interactions involved in sequence-specific recognition and catalysis by proteins that act on DNA. Recently, a class of nucleoside analogs, 4′-thionucleosides, has been synthesized that introduce apparently subtle changes in DNA/RNA structure, yet have significant effects on biological function. Introduction of 4′-thiothymidylate into an oligodeoxynucleotide duplex dodecamer resulted in only minor changes in the sugar pucker of the modified nucleoside and small conformational changes in the sugar–phosphate backbone at and adjacent to the nucleoside analog (23). However, when incorporated into the target sequence of the *Eco*RV restriction–modification system, the same analog strongly inhibited the activity of both the endonuclease and methyltransferase *in vitro* (24). Inhibition of the *Eco*RV methyltransferase occurred even though the thio-modified nucleotide was not the target for methylation. Other studies have shown that 4′-thiothymidine is resistant to nucleoside phosphorylase (25). Oligonucleotides incorporating 4′-thiothymidine are resistant to digestion by endonuclease S1 and do not induce RNaseH digestion of bound RNA as strongly as normal DNA (26). At present, the mechanism(s) by which thio-modified nucleosides inhibit enzymatic function is not clear.

In this study we report the synthesis of 4′-thio-2′-deoxycytidine (**1**) (Fig. 2) as a 5′-protected phosphoramidite and show that this 4′-thionucleoside analog, when incorporated into synthetic DNA as a target, inhibits methylation by the bacterial 5mC-DNA methyltransferase M.*Hha*I. We also report the structure of a 13mer DNA oligonucleotide containing 4′-thio-2′-deoxycytidine in complex with M.*Hha*I and AdoMet in which base flipping and partial methylation were observed. Our results show that 4′-thio-2′-deoxycytidine does not disrupt DNA recognition, binding, nor base extrusion, at least under crystallization conditions, but suggest instead that this thio-modified nucleoside interferes with a step between the conformational base-flip and methyl transfer.

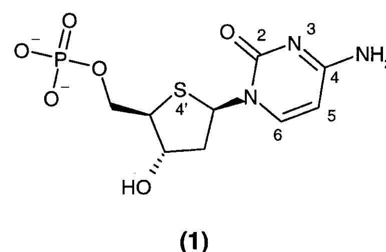


Figure 2. Structure of 4′-thio-2′-deoxycytidine.

MATERIALS AND METHODS

Synthesis

N^4-Benzoyl-3′,5′-di-O-benzyl-4′-thio-2′-deoxycytidine and the α-anomer. To a stirred suspension of N^4-benzoylcytosine (2.50 g, 11.63 mmol) in acetonitrile (50 ml), under a nitrogen atmosphere, was added *bis*-(trimethylsilyl)acetamide (4.91 g, 6.0 ml, 23.26 mmol). The reaction mixture was stirred at room temperature for 1 h, during which time a clear solution was formed. Additional acetonitrile (100 ml) was added, followed by the addition of crushed 4 Å molecular sieves (3.5 g) and a solution of benzyl 3,5-di-*O*-benzyl-2-deoxy-1,4-dithio-D-*erythro*-pentofuranoside (4.23 g, 9.69 mmol) in acetonitrile (25 ml). After 10 min, a solution of *N*-iodosuccinimide (2.40 g, 10.66 mmol) in acetonitrile (25 ml) was added, upon which the reaction mixture changed color to dark brown. After 5 h, crystals of the β-anomer of the product had formed in the reaction mixture, which was filtered and the crystals washed with acetonitrile. The crystals were then dissolved in dichloromethane and filtered, to separate them from the crushed molecular sieves. The dichloromethane was removed *in vacuo* to yield the pure β-anomer of the product (0.86 g, 17%) as a fluffy white solid. The filtrate of the reaction mixture had the solvent removed *in vacuo* and the resulting solid was re-dissolved in dichloromethane. The organic layer was washed with a saturated aqueous solution of sodium thiosulfate (2 × 200 ml) and water (200 ml), dried ($MgSO_4$), filtered and the solvent removed *in vacuo*. The crude product was purified, and the α- and β-anomers separated, by flash column chromatography, using an

eluent of ethyl acetate/n-hexane (3:1). The appropriate fractions were combined to obtain the product as the pure β-anomer (0.82 g, 16%; overall yield of β-anomer: 1.28 g, 33%) and pure α-anomer (2.26 g, 45%). The isolated α/β-anomers of the product (3.54 g, 78%) gave the α/β ratio as 3:2.

β-anomer: R_f [0.36 EtOAc/n-hexane (3:1)]. δ (^{1}H-NMR; 300 MHz, $CDCl_3$): 8.85 (1H, br.s, 4-NHBz), 8.64 (1H, d, ^3J = 7.4 Hz, H-6), 7.90 (2H, d, ^3J = 7.0 Hz, *o*-benzoyl), 7.63–7.17 (14H, m; 10H, PhCH$_2$; 3H, benzoyl aromatic; 1H, H-5), 6.47 (1H, m, H-1′), 4.60–4.43 (4H, m, 2× PhCH$_2$), 4.23 (1H, m, H-3′), 3.84–3.65 (3H, m; 2H, H-5′; 1H, H-4′), 2.69–2.56 and 2.37–2.20 (2H, m, H-2′$_a$ and H-2′$_b$). δ (^{13}C-NMR; 75 MHz, $CDCl_3$): 41.7 (C-2′), 53.6 (C-4′), 63.1 (C-1′), 70.5, 71.7 (2×PhCH$_2$), 73.7 (C-5′), 81.4 (C-3′), 96.9 (*p*-benzoyl aromatic), 127.6, 127.7, 127.9, 128.0, 128.3, 128.5, 128.7, 128.9 (benzyl and benzoyl aromatic), 133.0 (C-5), 133.3 (quaternary benzoyl aromatic), 137.4, 137.6 (quaternary benzyl aromatic), 146.7 (C-6), 155.3 (C-2), 161.8 (C-4), 166.8 (benzoyl carbonyl). Mass spectrum (+ve LSIMS); m/z 528 (17%, $[M+H]^+$), 420 (2%, $[M-BnO]^+$), 216 (100%, $[base+H]^+$). Elemental analysis found: C, 68.27%; H, 5.58%; N, 7.90%; S, 6.00%. $C_{30}H_{29}O_4N_3S$ requires: C, 68.29%; H, 5.54%; N, 7.96%; S, 6.08%.

α-anomer: R_f [0.28 EtOAc/n-hexane (3:1)]. δ (^{1}H-NMR; 300 MHz, $CDCl_3$): 8.78 (1H, br.s, 4-NHBz), 8.61 (1H, d, ^3J = 8.3 Hz, H-6), 7.89 (2H, m, *o*-benzoyl), 7.63–7.17 (14H, m; 10H, PhCH$_2$; 3H, benzoyl aromatics; 1H, H-5), 6.36 (1H, m, H-1′), 4.59–4.34 (4H, m, 2× PhCH$_2$), 4.31 (1H, m, H-3′), 4.02 (1H, m, H-4′), 3.50 and 3.37 (2H, m, H-5′$_a$ and H-5′$_b$), 2.57–2.38 (2H, m, H-2′$_a$ and H-2′$_b$). δ (^{13}C-NMR; 75 MHz, $CDCl_3$): 41.1 (C-2′), 55.1 (C-4′), 65.6 (C-1′), 71.0, 71.9 (2×PhCH$_2$), 73.3 (C-5′), 83.3 (C-3′), 95.9 (*p*-benzoyl aromatic), 127.6, 127.7, 127.8, 127.9, 128.0, 128.5, 128.5, 129.0 (benzyl and benzoyl aromatics), 133.0 (C-5), 133.3 (quaternary benzoyl aromatic), 137.1, 137.7 (quaternary benzyl aromatic), 148.3 (C-6), 155.2 (C-2), 161.9 (C-4), 166.7 (benzoyl carbonyl). Mass spectrum (+ve LSIMS); m/z 660 (6%, $[M+Cs]^+$), 550 (1%, $[M+Na]^+$), 528 (7%, $[M+H]^+$), 420 (1%, $[M-BnO]^+$), 216 (52%, $[base+H]^+$), 105 (42%, $[Bz]^+$), 91 (100%, $[Bn]^+$). Elemental analysis found: C, 68.12%; H, 5.23%; N, 7.90%; S, 5.87%. $C_{30}H_{29}O_4N_3S$ requires: C, 68.29%; H, 5.54%; N, 7.96%; S, 6.08%.

N^4-Benzoyl-4′-thio-2′-deoxycytidine. To a stirred solution of boron trichloride in dichloromethane (1.0 mol/dm^3) (30 ml, 30 mmol), at –85°C under a nitrogen atmosphere, was added dropwise a solution of N^4-benzoyl-3′,5′-di-*O*-benzyl-4′-thio-2′-deoxycytidine (2.00 g, 3.80 mmol) in dichloromethane (75 ml), after which the reaction mixture was left stirring for 4 h. The reaction was then quenched, at –85°C, by the addition of methanol/dichloromethane (1:1) (50 ml). The reaction mixture was then immediately neutralized by the addition of methanolic ammonia (2.0 M) (~40 ml). Then the reaction mixture was allowed to warm to room temperature and was filtered to remove the ammonium chloride formed during neutralization. The solvent was then removed *in vacuo* to yield a white solid. The crude product was purified by flash column chromatography, using an eluent of dichloromethane/methanol (9:1). Combination of the appropriate fractions yielded the product (0.78 g, 59%) as a white solid. R_f [0.40 DCM/MeOH (9:1)]. δ (^{1}H-NMR; 300 MHz, d_6-DMSO): 11.28 (1H, br.s, 4-NHBz), 8.56 (1H, d, ^3J = 7.8 Hz, H-6), 8.00 (2H, d, ^3J = 8.0 Hz, *o*-benzoyl), 7.65–7.47 (3H, m, *m*/*p*-benzoyl), 7.39 (1H, d, ^3J = 7.8 Hz, H-5), 6.31 (1H, m, H-1′), 5.35 (1H, m, 3′-OH), 5.21 (1H, m, 5′-OH), 4.43–4.33 (1H, m, H-3′), 3.74–3.56 (2H, m, H-5′), 3.39–3.31 (1H, m, H-4′), 2.39–2.17 (2H, m, H-2′$_a$ and H-2′$_b$). δ (^{13}C-NMR; 75 MHz, d_6-DMSO): 41.9 (C-2′), 58.3 (C-4′), 62.0 (C-1′), 63.0 (C-5′), 73.4 (C-3′), 97.3 (*p*-benzoyl CH), 128.1, 128.7 (*o*/*m*-benzoyl aromatic), 132.8 (quaternary benzoyl aromatic), 133.1 (C-5), 146.6 (C-6), 155.6 (C-2), 162.4 (C-4), 167.5 (benzoyl carbonyl). Mass spectrum (+ve LSIMS); m/z 370 (13%, $[M+Na]^+$), 348 (47%, $[M+H]^+$), 216 (100%, $[base+H]^+$). Elemental analysis found: C, 55.26%; H, 4.81%; N, 11.82%; S, 9.00. $C_{16}H_{17}O_4N_3S$ requires: C, 55.32%; H, 4.93%; N, 12.09%; S, 9.23%.

N^4-Benzoyl-5′-O-(4,4′-dimethoxytriphenylmethyl)-4′-thio-2′-deoxycytidine. To a stirred solution of N^4-benzoyl-4′-thio-2′-deoxycytidine (508 mg, 1.68 mmol) in pyridine (20 ml), under an argon atmosphere at room temperature, was added triethylamine (0.31 g, 0.42 ml, 3.03 mmol) 4-dimethylaminopyridine (10 mg, 0.08 mmol) and 4,4′-dimethoxytriphenylmethyl chloride (0.97 g, 2.86 mmol). The reaction mixture was left stirring at room temperature for 42 h and was then quenched by the addition of methanol (5 ml) and stirring was continued for a further 10 min. The solvent was removed *in vacuo* and the crude product dissolved in dichloromethane (100 ml). The organic layer was washed with a saturated aqueous solution of sodium hydrogen carbonate (2 × 100 ml) and a saturated aqueous solution of sodium chloride (1 × 100 ml). The organic layer was dried ($MgSO_4$), filtered and the solvent removed *in vacuo*, followed by co-evaporation with toluene and then methanol, to yield the crude product as a yellow foam. The product was purified by silica gel column chromatography, using an initial eluent of dichloromethane/methanol (99:1) (+0.5% TEA), followed by an eluent of dichloromethane/methanol (98:2) (+0.5% TEA). Combination of the appropriate fractions yielded the product (0.84 g, 89%) as a cream foam. R_f [0.46 DCM/MeOH (97:3) +0.5% TEA]. δ (^{1}H-NMR; 300 MHz, $CDCl_3$): 8.37 (1H, d, ^3J = 7.6 Hz, H-6), 7.89 (2H, m, *o*-benzoyl), 7.65–7.24 (14H, m; 9H, DMTr aromatic; 3H, benzoyl aromatic; 1H, H-5; 1H, NHBz), 6.88 (4H, m, DMTr aromatic), 6.34 (1H, m, H-1′), 4.77 (1H, br.s, 3′-OH), 4.35 (1H, m, H-3′), 3.83 and 3.82 (6H, 2s, 2× OMe on DMTr group), 3.54 (2H, m, H-5′), 3.39 (1H, m, H-4′), 2.55 and 2.19 (2H, m, H-2′$_a$ and H-2′$_b$). δ (^{13}C-NMR; 75 MHz, $CDCl_3$): 44.0 (C-2′), 55.3 (2× OMe on DMTr group), 55.6 (C-4′), 62.8 (C-1′), 64.4 (C-5′), 74.5 (C-3′), 87.0 [$C(Ar)_3$ on DMTr group], 97.2 (*p*-benzoyl aromatic), 113.2–130.1 (benzoyl aromatic and DMTr aromatic), 133.0 (quaternary benzoyl aromatic), 133.1 (C-5), 135.6, 135.7, 144.4 (DMTr aromatic), 146.3 (C-6), 155.7 (C-2), 158.7 (DMTr aromatic), 162.1 (C-4), 166.8 (benzoyl carbonyl). Mass spectrum (+ve LSIMS), doped with Na^+; m/z 688 (1%, $[M+K]^+$), 672 (17%, $[M+Na]^+$), 650 (3%, $[M+H]^+$), 303 (100%, $[DMTr]^+$), 238 (38%, $[base+Na]^+$), 216 (30%, $[base+H]^+$). Accurate mass $C_{37}H_{35}N_3O_6NaS$ requires 672.2144. Observed 672.2118.

N^4-Benzoyl-5′-O-(4,4′-dimethoxytriphenylmethyl)-3′-O-[2-cyano-ethyl(diisopropylamino)-phosphinoyloxy]-4′-thio-2′- deoxycytidine. To a stirred solution of N^4-benzoyl-5′-*O*-(4,4′- dimethoxytriphenyl-methyl)-4′-thio-2′-deoxycytidine (1.611g, 2.48 mmol) in tetra-hydrofuran (150 ml), under a nitrogen atmosphere at 0°C, was added *N,N*-diisopropylethylamine (1.60 g, 2.16 ml, 4.96 mmol) followed by cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.17 g, 1.10 ml, 4.96 mmol). The reaction mixture was then allowed to warm to room temperature and was left stirring for 6 h. The solvent was removed *in vacuo* and the resultant white solid was dissolved in ethyl acetate (150 ml). The organic layer was

washed with a saturated aqueous solution of sodium hydrogen carbonate (2 × 150 ml) and a saturated aqueous solution of sodium chloride (1 × 150 ml). The organic layer was dried ($MgSO_4$), filtered and the solvent removed *in vacuo* to yield a yellow/brown oil. The product was purified by flash column chromatography, using an initial eluent of dichloromethane (+0.5% TEA), followed by an eluent of dichloromethane/ethyl acetate (7:3). Combination of the appropriate fractions yielded the product (1.53 g, 73%) as a cream foam. R_f [0.17, 0.22 DCM/EtOAc (7:3) +0.5% TEA]. δ (^{1}H-NMR; 400 MHz, $CDCl_3$): 8.95–8.50 (1H, v.br.s, N$\underline{H}$Bz), 8.42 and 8.29 (1H, 2d, ^3J = 7.5 Hz, H-6), 7.89 (2H, d, ^3J = 7.4 Hz, *o*-benzoyl), 7.62–7.23 (13H, m; 9H, DMTr aromatic; 3H, benzoyl aromatic; 1H, H-5), 6.86 (4H, 2dd, ^3J = 6.0 Hz, 9.0 Hz, DMTr aromatic), 6.43 and 6.41 (1H, 2dd, ^3J = 6.2 Hz, H-1′), 4.56 (1H, m, H-3′), 3.82 and 3.81 (6H, m, 2× OMe on DMTr group), 3.75–3.45 (1H, m, H-4′), 3.65–3.50 (2H, m, $OC\underline{H}_2CH_2CN$), 3.65–3.43 (2H, m, 2× $NC\underline{H}(CH_3)_2$), 3.48–3.34 (2H, m, H-5′), 2.75 and 2.60 (2H, m, $OCH_2\underline{C}H_2CN$), 2.67–2.51 and 2.19–2.07 (2H, m, H-2′$_a$ and H-2′$_b$), 1.18–1.09 (12H, m, 2× $NCH(C\underline{H}_3)_2$). δ (^{13}C-NMR; 100 MHz, $CDCl_3$): 20.2, 20.3 ($OCH_2\underline{C}H_2CN$), 2 × 24.5 ($NCH(\underline{C}H_3)_2$), 2 × 42.9 (C-2′), 43.2, 43.4 ($N\underline{C}H(CH_3)_2$), 2 × 55.2 (2× OMe on DMTr group), 55.7 (C-4′), 58.3, 58.5 ($O\underline{C}H_2CH_2CN$), 2 × 62.5 (C-1′), 63.5, 64.2 (C-5′), 75.0, 76.2 (C-3′), 113.2, 127.1, 127.5, 127.9, 128.3, 128.4, 129.0, 130.1, 130.2, (DMTr aromatic and benzoyl aromatic), 2 × 133.1 (C-5), 135.4, 2 × 135.5 (quaternary DMTr aromatic and quaternary benzoyl aromatic), 2 × 144.2 (DMTr aromatic), 145.2 (C-2), 146.0, 146.2 (C-6), 2 × 158.7 (quaternary DMTr aromatic), 161.8 (C-4), 165.5 (benzoyl carbonyl). δ (^{31}P-NMR; 162 MHz, $CDCl_3$): 134.4, 134.8. Mass spectrum (+ve LSIMS), doped with Na^+; m/z 888 (5%, $[M+K]^+$), 872 (24%, $[M+Na]^+$), 850 (1%, $[M+H]^+$), 634 (2%, $[M\text{-base}]^+$), 632 (3%, $[M-(OP(N(^iPr)_2(OCH_2CH_2CN))]^+$), 303 (100%, $[DMTr]^+$), 201 (5%, $[P(OCH_2CH_2CN)(N(^i(Pr)_2)]^+$). Accurate mass $C_{46}H_{52}N_5O_7NaPS$ requires 872.3202. Observed 872.3223.

Oligonucleotide synthesis and end-labeling

The following oligonucleotides were used in the experiments (*Hha*I target sequence in bold; M, 5-methyl-2′-deoxycytidine; S, 4′-thio-2′-deoxycytidine; F, 5-fluoro-2′-deoxycytidine). Oligonucleotides **A**–**E** were 37mers with the sequence:

```
GACTGGTACAGTATCAGGXGCTGACCCACAACATCCG
  TGACCATGTCATAGTCCGYGACTGGGTGTTGTAGGCT
```

where X and Y varied as follows: **A**, X = S, Y = M; **B**, X = C, Y = M; **D**, X = C, Y = C; and **E**, X = S, Y = C; Oligo **C** contained the unmethylated sequence GGCC in place of GCGC.

The 5′-protected phosphoramidite of 4′-thio-2′-deoxycytidine was synthesized as described above and incorporated into the desired oligonucleotides by standard solid phase synthesis methodology [Roger Knott, New England Biolabs (NEB)] and purified using HPLC and polyacrylamide gel electrophoresis. Individual strands were annealed in 10 mM Tris–HCl (pH 8.0), 1 mM Na_2EDTA (TE) to generate double-stranded DNA. Annealed oligonucleotides were 5′ end-labeled using T4 polynucleotide kinase (NEB) and [γ-^{32}P]ATP [New England Nuclear (NEN)] using the protocol suggested by the manufacturer. Unincorporated nucleotide was removed by spun column chromatography using Quick-Spin G-25 columns (Boehringer Mannheim) or MicroSpin S-200 columns (Pharmacia).

Purification of M.*Hha*I

M.*Hha*I was purified as described previously (9) and exhaustively dialyzed to remove endogenous bound *S*-adenosyl-L-methionine (AdoMet).

DNA binding

DNA binding was studied by gel retardation, essentially as described previously (27). The target oligonucleotide concentration ranged from ~5 to 200 nM. The cofactor *S*-adenosylhomocysteine (AdoHcy) was included in some reactions at a concentration of 1 mM. After autoradiography the portions of each gel lane corresponding to bound and free oligonucleotide were excised and radioactivity measured by scintillation counting in Opti-fluor (Packard). Binding constants were determined by Scatchard analysis (28).

Off-rates for DNA binding were measured by pre-forming complexes in the presence of 1 mM AdoHcy, 6.4 nM normal or 4.8 nM thio-substituted labeled oligonucleotide and 6.4 nM M.*Hha*I. The association reactions were quenched by adding excess unlabelled oligo at 1 µM final concentration. Dissociation of the labeled oligo was allowed to occur for various time intervals. Samples were then loaded onto a running 6% polyacrylamide gel and processed in the same manner as the equilibrium binding experiments. Off-rates were calculated by monitoring the disappearance of the bound labeled DNA.

Methyl transfer

Methylation reactions were performed in [20 mM Tris-acetate (pH 7.9) 10 mM magnesium acetate, 50 mM potasium acetate, 1 mM DTT, 200 µg/ml BSA] at 37°C in the presence of *S*-[methyl-^{3}H]adenosyl-L-methionine (NEN) with a specific activity adjusted to either 9.7 or 15 Ci/mmol. The concentrations of enzyme, DNA and cofactor used are listed in the appropriate figure legends. All components except the DNA were mixed and pre-warmed at 37°C for 3 min, at which point the DNA was added to initiate the reaction. Reactions were stopped by the addition of excess *S*-adenosyl-L-homocysteine (Sigma) to a final concentration between 1 and 10 mM and/or flash freezing in a dry ice–EtOH bath. Methyl transfer to DNA was measured by trapping DNA on Whatman DE-81 filters as previously described (29). Bound radioactivity was determined by scintillation counting in Opti-fluor (Packard).

Restriction digests

Oligonucleotides at a concentration of 1 µM were digested with 14 U restriction endonuclease in 20 µl reactions using the buffer suggested by the enzyme supplier (NEB). Digestion products were separated by electrophoresis on 4% Metaphor agarose (FMC) gels in 1× TBE buffer and visualized by ethidium staining and UV illumination.

X-ray crystallography

The structure of M.*Hha*I in a ternary complex with AdoMet and a short DNA duplex (referred to as the S13 structure) was determined by X-ray crystallography. Two 13mer oligonucleotides (NEB) were annealed to form a 12 bp asymmetric duplex with 5′ single base overhangs:

```
5'-TCCATGMGCTGAC -3'
3'- GGTACGSGACTGT-5'
```

where M = 5-methyl-2′-deoxycytidine and S = 4′-thio-2′-deoxycytidine. Crystals of the S13 complex were grown by mixing the M.*Hha*I–AdoMet–DNA complex, as described previously (12,13), with an equal volume of 1.4–1.8 M ammonium sulfate, 50 mM citrate at pH 5.6, and equilibrating the mixture against 1 ml of the latter solution at 16°C. The S13 crystals formed within days in the same trigonal space group as in the previously reported ternary structures (11–13). One S13 complex was present per asymmetric unit. Three weeks after formation of the crystals, the X-ray diffraction intensities were measured at a wavelength of 1.15 Å on a Mar Research imaging-plate diffractometer at beamline X12C of the National Synchrotron Light Source, Brookhaven National Laboratory (67.5° total rotation with 1.5° increment and 90 s exposure at ~16°C). Data acquisition, similar to that described previously (12,13), yielded 32 880 unique reflections. R-merge = $\Sigma|\langle I\rangle - I|/\Sigma\langle I\rangle = 0.078$ and $\langle I/\sigma\rangle = 12.1$, where I is the measured intensity of each reflection, $\langle I\rangle$ is the intensity averaged from multiple observations of symmetry-related reflections, and σ is the root mean square deviation from the mean.

The structure was solved by the difference Fourier method using the previously solved HM13 ternary structure (M.*Hha*I with AdoHcy and a hemi-methylated DNA identical in sequence to the one used here but containing unmodified cytosine at the target) as the initial model (13). The model was optimized by rigid-body refinement from 20 to 3 Å and was further refined against diffraction data measured from 5.0 to 2.05 Å using X-PLOR (30). After 11 rounds of least square refinement, manual model building, and placement of well-ordered solvent molecules (interpreted as water) by examination of difference electron density, the crystallographic R-factor and R-free ($\Sigma|F_o - F_c|/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively) were reduced to 0.186 and 0.231 for ~90 and 10% of the data, respectively. We used the difference Fourier synthesis ($F_o - F_c$, α_c) to confirm and locate the positions of the sulfur atom of 4′-thio-2′-deoxycytidine and the transferable methyl group of AdoMet.

RESULTS

Synthesis of the 5′-protected phosphoramidite of *N*4-benzoyl-4′-thio-2′-deoxycytidine

The key step in the synthesis of the 5′-protected phosphoramidite of *N*4-benzoyl-4′-thio-2′-deoxycytidine requires either the condensation of a cytosine base with a suitably protected 4-thiosugar compound, or the conversion of the uridine analogue to the corresponding cytidine analogue. Secrist and co-workers have previously reported the synthesis of 4′-thio-2′-deoxycytidine (31). Initially this group attempted to prepare 4′-thio-2′-deoxycytidine via direct condensation of cytosine with a suitably protected 4-thiosugar. However, it was not possible to separate the α/β-anomers at either the protected or deprotected stage and therefore 4′-thio-2′-deoxycytidine was prepared via a uridine to cytidine conversion on the β-anomer of 4′-thio-2′-deoxyuridine. Here we report an improved procedure for the preparation of 4′-thio-2′-deoxycytidine, using *N*4-benzoylcytosine in a direct condensation approach.

*N*4-Benzoylcytosine was synthesized, using the procedure of Brown *et al*. (32), in a yield of 93%. This base was then condensed with benzyl 3,5-di-*O*-benzyl-2-deoxy-1,4-dithio-D-*erythro*-pentofuranoside, synthesized by the method of Dyson *et al*. (33), to give an anomeric mixture (α/β ratio 3:2) of the nucleoside derivative *N*4-benzoyl-3′,5′-di-*O*-benzyl-4′-thio-2′-deoxycytidine. Separation of the anomers either by crystallization of the β-anomer or by silica gel column chromatography proved to be facile, due to the presence of the *N*4-benzoyl protecting group. Removal of the 3′,5′-di-*O*-benzyl protecting groups was achieved using boron trichloride at low temperature, although the yield was lower than expected due to partial debenzoylation of the *N*4-benzoyl protecting group. Dimethoxytritylation of the 5′-hydroxyl group followed by phosphitylation of the 3′-hydroxyl group gave the desired phosphoramidite. The overall yield of the phosphoramidite from the 4-thiosugar was 13%; the poor α/β ratio (3:2) at the condensation step results in a maximum theoretical yield of only 40%.

The 5′-protected phosphoramidite of *N*4-benzoyl-4′-thio-2′-deoxycytidine proved to be completely compatible with standard solid phase synthesis of oligonucleotides via the phosphoramidite approach.

Methylation is inhibited by 4′-thio-2′-deoxycytidine

4′-Thiothymidine is known to inhibit methylation by M.*Eco*RV when present adjacent to the target adenine that becomes methylated (24). We examined the effects of 4′-thio-2′-deoxycytidine as a target nucleotide on catalytic function in a DNA cytosine methyltransferase, M.*Hha*I. M.*Hha*I methylates the inner cytosine of the sequence GCGC to yield G5mCGC. We constructed a 37 base double-stranded DNA oligonucleotide containing a centrally located *Hha*I site (GCGC) in which the inner cytidine was replaced by the thio-modified residue on one strand and by 5-methyl-2′-deoxycytidine in the other strand. Incubation of M.*Hha*I with the hemi-methylated 37mer 4-thio oligonucleotide resulted in very poor transfer of ^{3}H-methyl groups from the cofactor AdoMet relative to a hemi-methylated 37mer substrate identical in sequence but containing normal cytosine (Fig. 3). The level of incorporation of ^{3}H-methyl was comparable to that observed with a non-methylated 37mer containing the non-cognate target sequence GGCC. The turnover rate constant K_{cat} for methyl transfer was >100-fold lower in the presence of the thio-substituted hemi-methylated substrate (0.020 min^{-1}) relative to unmodified hemi-methylated DNA (2.3 min^{-1}). Kinetic analysis of the inhibition suggests that DNA target sites containing 4′-thio-2′-deoxycytidine interact competitively with respect to unmodified target sites (apparent K_i ~2–5 nM) (Fig. 4). This value is comparable to the K_m of 2.3 nM reported for a poly(dGdC) substrate (19).

4′-Thio-2′-deoxycytidine does not interfere with sequence recognition

One obvious explanation for the inhibition of methylation by 4′-thio-2′-deoxycytidine was the potential disruption of the binding of the methyltransferase to the target sequence. To test this, equilibrium binding assays were performed on the same hemi-methylated oligonucleotides that were refractory to methylation. As shown in Figure 5, the 4′-thio modified oligonucleotide bound as well as the unmodified oligonucleotide. Equilibrium dissociation constants for unmodified and 4′-thio-2′-deoxycytidine-containing oligonucleotides were determined by Scatchard analysis to be 18 and 21 nM, respectively. Addition of AdoHcy to the binding reaction increased binding of both oligonucleotides as expected from previous work (13,22,27,34–36) (data not shown).

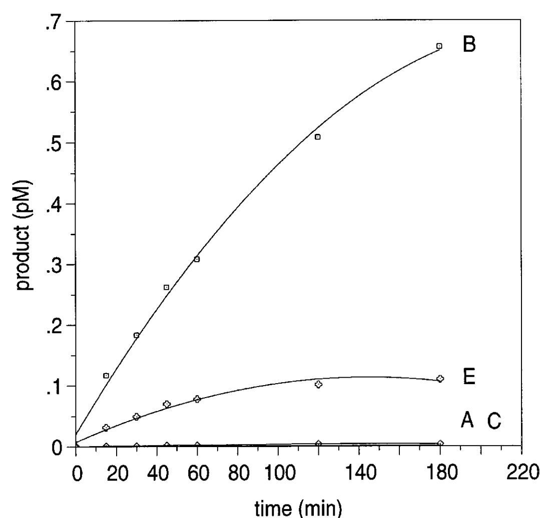


Figure 3. Progress curves for methylation of oligonucleotide substrates. Methylation reactions were performed as described in Materials and Methods. Reaction mixtures contained 1.65 μM AdoMet (9.7 Ci/mmol), 3.2 nM enzyme and 1 μM target oligonucleotide (37mer **A**, **B**, **C** or **E** in Materials and Methods) in 700 μl total volume. At 0, 15, 30, 45, 60, 120 and 180 min, 100 μl aliquots were removed and the reaction quenched by addition of AdoHcy to 2 mM followed by flash freezing in a dry ice/EtOH bath. Incorporation of methyl groups into DNA was measured as described in Materials and Methods. Each data point shown is the average of three samples. Target sequences (S = 4′-thio-2′-deoxycytidine, M = 5-methyl-2′-deoxycytidine): **A**, GSGC/GMGC; **B**, -GCGC/GMGC; **C**, GGCC/GGCC; **E**, GSGC/GCGC. Curves **A** and **C** overlap at the scale shown. K_{cat} values (see text) were estimated from the 15 min time points where the reaction rate was approximately zero-order (data not shown).

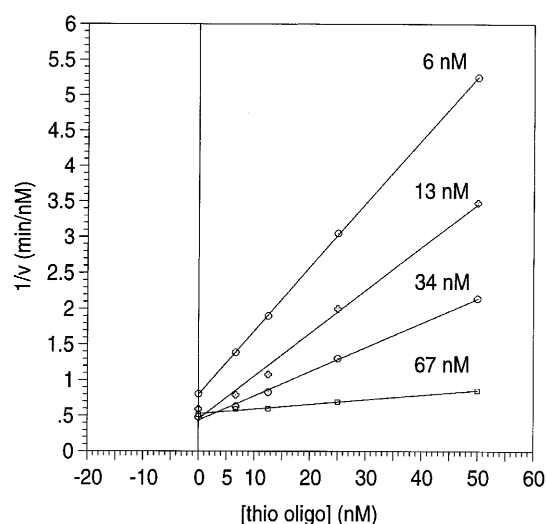


Figure 4. Kinetics of inhibition by 4′-thio-2′-deoxycytidine. Methylation reactions were performed at a series of substrate (phage λDNA) and inhibitor (4′-thio-2′-deoxycytidine DNA, oligo **A**, Materials and Methods) concentrations in the presence of 80 nM AdoMet and 80 pM enzyme. Methyl transfer was measured by trapping DNA on DE-81 filters (Materials and Methods) and scintillation counting. Initial velocities were calculated and displayed as a Dixon plot. The horizontal axis value at the point of intersection of the curves is equal to $-K_i$. The point of intersection is consistent with competitive inhibition.

While the equilibrium binding of the 4′-thio modified oligonucleotide was essentially identical to the unmodified control DNA, the off-rates for the two species were considerably different, as shown in Figure 6. Incorporation of 4′-thio-2′-deoxycytidine into the target sequence substantially increased the off-rate of the DNA relative to unmodified DNA. Half-lives of the 4′-thio and unmodified DNA complexes under the binding conditions used were 60 and 760 min, respectively. In combination with the equilibrium binding studies, this result indicates that the stability of the DNA–methyltransferase complex is substantially altered by the presence of the 4′-thio modification.

4′-Thio-2′-deoxycytidine does not interfere with endonuclease digestion

Oligonucleotides containing 4′-thiothymidine were previously shown to be resistant to cleavage by the endonuclease *Eco*RV when the thio-modification was incorporated within the recognition sequence of the enzyme (24). The conformation of *Eco*RV does not change significantly when complexed with thio-modified DNA as opposed to normal DNA (Kostrewa,D., Hancox,E.L., Walker, R.T., Connolly,B.A. and Winkler,F.K., unpublished results). However, the Mg^{2+} ion required for catalysis was not present in crystals of *Eco*RV bound to a thio-modified substrate, despite being crystallized out of Mg^{2+}-containing solution. In the current study, we constructed non-methylated versions of the oligonucleotides used in the methylation assays and subjected them to digestion by the cognate restriction endonuclease R.*Hha*I and its isoschizomer R.*Hin*P1I. After incubation with R.*Hha*I which cleaves 5′ to the outer cytosine (GCG/C), complete digestion was observed (Fig. 7). To determine if the proximity of the thio- modified cytidine to the actual site of cleavage was a factor in the digestion, the reaction was repeated using R.*Hin*P1I which cleaves 5′ to the inner cytosine (G/CGC), immediately adjacent to the modified residue. As seen in Figure 7, the R.*Hin*P1I digestion also

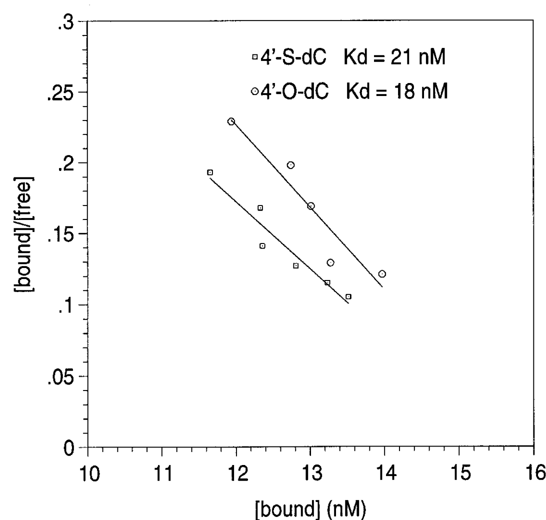


Figure 5. Equilibrium DNA binding. Binding reactions with normal and 4′-thio-2′-deoxycytidine substituted hemi-methylated oligonucleotides were performed in the absence of cofactor as described in Materials and Methods and separated by electrophoresis on a 6% non-denaturing polyacrylamide gel. Binding constants were estimated by Scatchard analysis. K_d values for thio-substituted and normal DNA were 21 and 18 nM, respectively. 4′-S-dC, 4′-thio-2′-deoxycytidine oligo (oligo **A**, Materials and Methods); 4′-O-dC, normal DNA (oligo **B**).

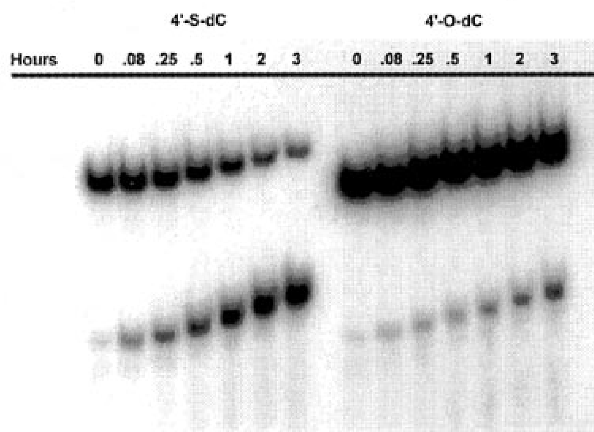


Figure 6. Off-rate of M.*Hha*I from 4′-S-C and normal C oligonucleotides. Each labeled oligonucleotide (6.4 nM normal or 4.8 nM thio-substituted) was pre-incubated with M.*Hha*I (6.4 nM) in the presence of AdoHcy (1 mM) for 30 min at room temperature to form DNA–protein complexes. Excess unlabelled oligonucleotide (1 μM) was then added and the dissociation of labeled oligo was allowed to proceed for the indicated times. Samples were loaded onto running gels, resulting in the pattern of 'half-smiles'. 4′-S-dC, 4′-thio-2′-deoxycytidine oligo (oligo **A**, Materials and Methods); 4′-O-dC, normal cytosine oligo (oligo **B**).

proceeded to completion. The same results were obtained when the inner 2′-deoxycytidine in *both* strands of the DNA was replaced with the thio-modified analog (data not shown). Unlike 4′-thiothymidine, incorporation of 4′-thio-2′-deoxycytidine did not reduce susceptibility to cleavage by restriction endonucleases. Whether this difference is due to structural variations induced by the thionucleoside or differences in the details of the cleavage mechanisms remains to be determined. R.*Eco*RV leaves blunt ends upon cutting, whereas R.*Hha*I and R.*Hin*P1I leave 5′ and 3′ tails, respectively.

Structure of DNA containing 4′-thio-2′-deoxycytidine bound to M.*Hha*I

We have now determined by X-ray crystallography the structure of M.*Hha*I in a ternary complex with AdoMet and a 13mer non-palindromic DNA duplex containing 4′-thio-2′-deoxycytidine as the target nucleotide on one strand, and 5-methyl-2′- deoxycytidine on the complementary strand. The existence of the sulfur atom of 4′-thio-2′-deoxycytidine and the methyl group of the 5-methyl-2′-deoxycytidine on the complementary strand were confirmed by ($F_o - F_c$, α_c) electron density maps in which the sulfur atom and methyl group were omitted from the structure factor calculation, respectively (Fig. 8). The peak height for the sulfur atom was comparable to that for the sulfur atom of Cys-81 and for the sulfur atom of AdoMet. The target 4′-thio-2′-deoxycytidine was flipped out of the DNA helix, while the 5-methyl-2′-deoxycytidine on the complementary strand remained stacked within the DNA helix.

Four other ternary structures containing M.*Hha*I, DNA and cofactor have been solved, each with a different combination of modification at the target cytosine in the two DNA strands (see table 1 of ref. 13). The oligonucleotide complexed with M.*Hha*I in the S13 structure reported here is hemi-methylated and identical in sequence to the DNA in HM13. There are only two differences between the HM13 and S13 complexes: the 4′ atom

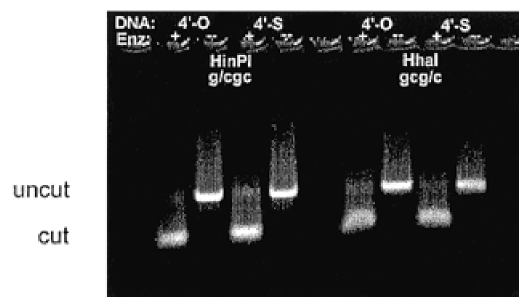


Figure 7. Restriction by R.*Hha*I. Each oligonucleotide (oligos **D** and **E**, Materials and Methods; final concentration 1 μM) was incubated with 14 U restriction endonuclease R.*Hha*I or R.*Hin*P1I at 37°C for 1 h. After the reaction, products were separated on a 4% Metaphor agarose gel and visualized under UV after ethidium bromide staining.

of the deoxyribose ring of the target cytosine (a sulfur atom in S13 versus oxygen in HM13) and the extra methyl group of AdoMet in S13 compared to AdoHcy used in HM13. Because HM13 contains the same DNA sequence as the S13 structure, we focused our comparisons on this structure.

An all-atom superposition of the protein components of S13 and HM13 yielded an rms deviation of 0.4 Å. The nucleic acid components of HM13 and S13 were also very similar, with an rms deviation of 0.3 Å over all atoms, excluding the 4′ ribose atom bearing the thio substitution. The target cytosine was rotated out of the helix and into the catalytic pocket of the enzyme as seen in each of the previous DNA–M.*Hha*I co-crystal structures. The introduction of sulfur into the deoxyribose ring did not grossly distort the structure of the nucleotide. Torsion angles associated with the flipped base in S13 are reasonably close to those in the other structures (data removed), with the greatest differences clustered immediately adjacent to the sulfur and along the DNA backbone. Pseudorotation phase angles indicate that the sugar pucker (C3′-endo) is very similar to that of unmodified cytidine in the N13 structure (12).

Although we expected an intact AdoMet to be present in the structure because AdoMet was initially used in the mixture for the crystallization and methyl transfer was strongly inhibited in solution, we were quite surprised to find that the difference electron density clearly shows two distinguishable locations of the methyl group: one ~1.6 Å away from the donor, the sulfur atom of AdoMet, and a second ~1.6 Å away from the acceptor, the C5 atom of 4′-thio-2′-deoxycytidine (Fig. 9). The first location is most likely the methyl group of AdoMet prior to methyl transfer, consistent with the relative position of the AdoMet methyl group in the M.*Hha*I–AdoMet binary complex (10). The second location is most likely the C5 methyl group. This methyl group is located out of the plane of the cytosine ring, as observed in the M13 structure where the C5 methyl group of the flipped 5-methyl-2′-deoxycytidine is bent out of the plane of the cytosine ring and towards the sulfur atom of AdoHcy (12). Two distinguishable locations of electron density were also associated with the sulfur of Cys81 (Figs 8 and 9). One position 2.7 Å away from cytosine C6 presumably reflects the reactive 'partial-covalent' bond seen in the HM13, N13 and M13 structures (12,13), while the other position 3.6 Å away from C6 (close to the sum of the van der Waals radii of sulfur and an Sp2 carbon atom) may represent a completely unreactive state. This is the first M.*Hha*I structure

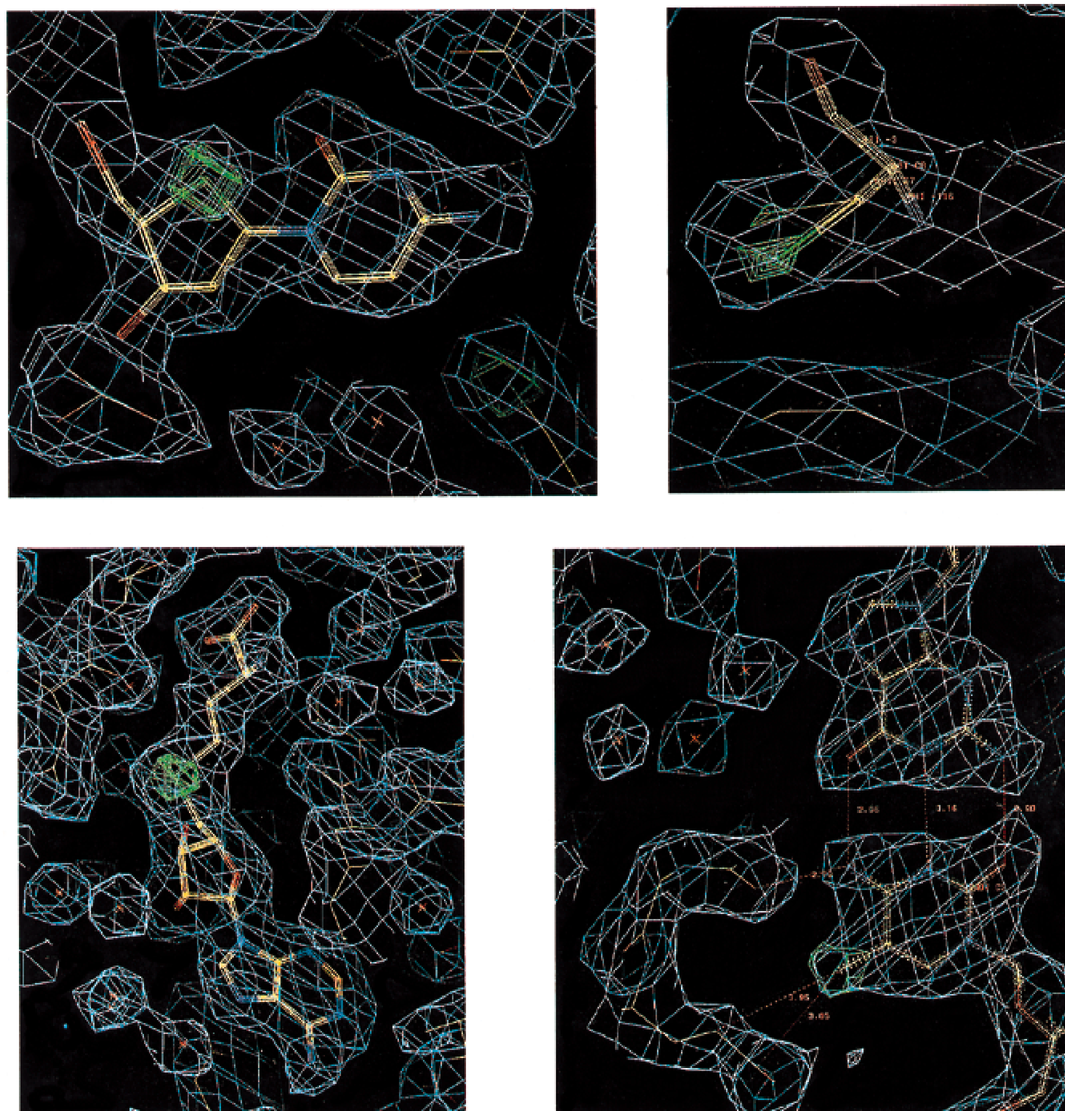


Figure 8. Difference electron density maps showing the position of the sulfur atoms of the modified target, Cys81 and AdoMet. Maps are shown (left to right) in the vicinity of (**a**) target 4′-thio-2′-deoxycytidine, (**b**) Cys81, (**c**) AdoMet and (**d**) 5-methyl-2′-deoxycytidine. The sulfur atom in (a), (b) and (c) and the methyl group in (d) were not included in the structure factor (F_c, α_c) calculation. The ($2F_o - F_c$, α_c) in blue is contoured at 1.0σ. The ($F_o - F_c$, α_c) in green is contoured at 10′σ, 8σ′, 10′σ and 3′σ in (a), (b), (c) and (d), respectively.

in which this alternative extended conformation of the catalytic cysteine has been observed.

In the immediate vicinity of the target cytosine, there is one amino acid which directly contacts the 4′ atom of the deoxyribose ring: Arg165 (Fig. 10). This amino acid is strongly conserved among C5 Mtases (1,10,37). In HM13 and other M.*Hha*I structures, the sidechain of Arg165 has three major interactions with the flipped cytosine. The Arg165 Nη1 (NH_2 group) forms a charge–charge interaction with the phosphate oxygen O1p (3.0 Å); both Nη2 (NH_2 group) and Nε (NH group) form hydrogen bonds with the cytosine O2 (2.9 and 3.0 Å, respectively); and Nη2 forms a weak hydrogen bond with the deoxyribose O4′. The net positive charge carried by the Arg is mainly balanced by the negative charge of the nearby phosphate group. The O4 oxygen has been calculated to carry a total charge of only –0.16e (38), suggesting that it is a weak hydrogen bond acceptor. However, in each of the M.*Hha*I ternary structures, the stereochemical juxtaposition of O4′ and the Arg NH_2 group is favorable and such a hydrogen bond appears feasible. The O2 atom has been calculated to have a negative charge of –0.51e (38) which rendered it a good acceptor for hydrogen bonds.

In S13, where the 4′-deoxyribose atom bears the thio substitution, the conformation of Arg165 remains largely unchanged relative to the HM13 structure. However, there is an indication that the 4′ sulfur sterically ‘bumps’ Arg165. This interaction might explain the observed increase in the distance between the target nucleotide and Arg165, especially near the 5′ phosphate group of the 4′-thio-2′-deoxycytidine. The separation between Arg165 Nη1 and the phosphate oxygen O1p (3.4 Å) represents an extreme value in the range of observed distances (Fig. 10). As a result, the normal charge–charge interaction between these two atoms may be disrupted.

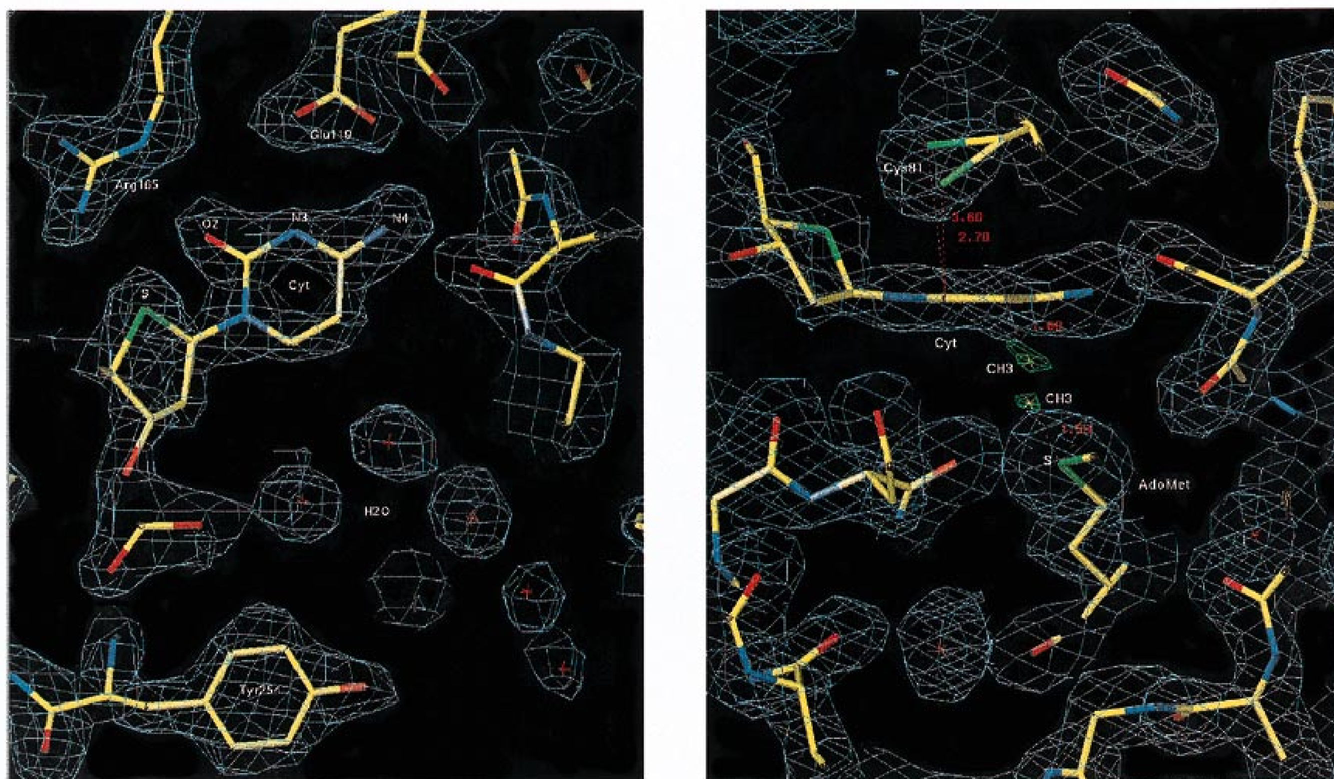
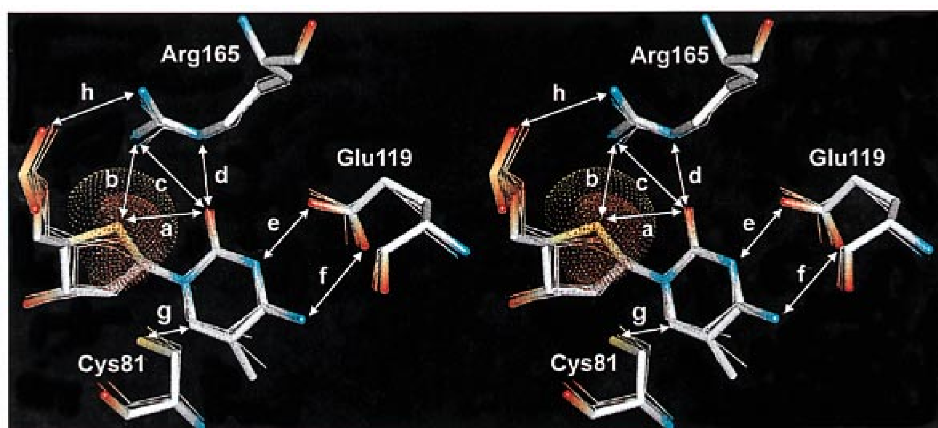


Figure 9. Electron density maps showing the two detected methyl groups. The difference electron density maps, ($2F_o - F_c$, α_c) in blue and ($F_o - F_c$, α) in green, contoured at 2.0σ and 3.0σ, respectively, superimposed on the refined coordinates illustrating the position of the methyl group. The methyl group was not included in the structure factor (F_c, α_c) calculation. There are two density peaks above the 3.0σ level for the methyl group, an indication that the methyl group has two partially occupied positions: one is 1.6 Å away from the sulfur atom of AdoMet and the other is 1.6 Å away from the C5 atom of cytosine. The two views are related by a 90° rotation. ($F_o - F_c$, α_c) density is only shown in one view.

DISCUSSION

Oligonucleotides bearing 4′-thio modifications display biological properties that differ from their unmodified counterparts, even though their structure appears to be only subtly altered. The structure of DNA containing 4′-thiothymidine has previously been characterized by X-ray crystallography (23) and that of free 4′-thiothymidine by NMR and X-ray crystallography (39). In these studies, the conformational changes induced by the thio substitution have been minor, short-range, and limited to the sugar–phosphate backbone in the immediate vicinity of the modified nucleotide. However, incorporation of 4′-thiothymidine adjacent to the target adenine was sufficient to interfere with methyl transfer by M.*Eco*RV (24). In the present work, 4′-thio-2′-deoxycytidine was used to replace the target 2′-deoxycytidine for methylation by M.*Hha*I. As with M.*Eco*RV, in solution the methylation reaction was strongly inhibited. This result is somewhat surprising since the presence of sulfur at the 4′ position of the sugar is not obviously incompatible chemically with the known mechanism for C5-cytosine methylation. The strong inhibition observed in the solution studies can be overcome over a long period of time in the crystal environment, as evidenced by the substantial, albeit partial, methylation of the cytosine C5 position in the crystal structure of the ternary complex of M.*Hha*I, 4′-thio-substituted DNA, and AdoMet. These results attest to a greatly-slowed reaction, even in the crystal environment where packing forces stabilize the complex and position the substrate in an optimal orientation for the reaction to proceed.

Why the reaction is slowed is difficult to discern, even from the high resolution crystal structure. It is formally possible that in the absence of crystal packing forces the minor distortions evident in the structure might be amplified sufficiently to dislocate the target cytosine such that it is no longer optimally oriented, leading to the inhibition of methylation seen in solution. However, given the overall similarity in the S13 and HM13 structures, we believe that analysis of the current structure is relevant to understanding this inhibition (especially that seen in the crystal environment). In principle, there are three points at which the methylation reaction could be inhibited: recognition of the target sequence and flipping out the target base, formation of the initial covalent bond at C6, and methyl transfer to C5. In DNA binding studies, the thio modification did not significantly reduce the level of binding as measured under equilibrium conditions, but did significantly increase the off-rate of the bound oligonucleotide, suggesting that the complex forms but fails to become fully stabilized. Although 4′-thio substitution might introduce backbone distortions that raise the thermodynamic barrier to base-flipping, no substantial structural deformations that would be incompatible with base extrusion were observed in the S13 crystal structure. At present, we are unable to assess definitively whether the rate of base flipping is reduced by 4′-thio-2′-deoxycytidine in solution as we lack a reliable assay for this process in the absence of methylation. However, given the equilibrium binding and structural data, as well as the observation that methylation is slowed in the crystal where flipping has definitely taken place, it is unlikely that base



	S13	HM13	N13	M13	F13	H3a	H3b
a	**2.9**	3.1	2.9	2.9	2.8	3.1	3.0
b	**2.9**	3.3	3.0	3.3	3.2	2.6	2.8
c	**3.4**	2.9	3.1	3.4	3.2	2.8	3.1
d	**3.1**	3.0	2.7	2.7	2.5	2.7	2.7
e	**2.9**	3.0	3.2	3.0	2.7	3.4	2.9
f	**3.1**	2.8	2.9	3.0	3.2	2.9	2.4
g	**2.7/3.6***	2.6	2.6	2.8	1.8	1.8	1.8
h	**3.4**	3.0	3.0	2.9	2.8	2.8	2.5

Figure 10. Interatomic distances within the catalytic pocket. A comparison of key interatomic distances among residues that interact with the target cytosine. S13 is the M.*Hha*I structure containing 4′-thio-2′-deoxycytidine. HM13 contains the identical DNA sequence without the thio substitution. N13, M13 (12) and F13 (11) contain a different DNA sequence with unmethylated, fully methylated and 5-fluoro-substited target cytosines, respectively. H3A and H3B are the two protein–DNA complexes in the M.*Hae*III crystal structure (14). F13, H3A and H3B are covalent complexes in which the enzyme and DNA are covalently bonded between C6 and the catalytic cysteine. The stereo view shown above the table displays S13 structure elements in thick lines. Thin lines represent conformations in the four other M.*Hha*I ternary structures discussed in the text. The van der Waals radii of sulfur and oxygen are represented by yellow and red dotted spheres, respectively. (*) The sulfur atom of Cys81 has two detectable positions. The position shown corresponds to 2.9 Å, the weighted average of the two positions during the refinement.

flipping is the key step affected. The observed differences in complex stability, as measured by the off-rate of DNA from the complex, occurred in the presence of AdoHcy and not AdoMet, which suggests that the primary effects of 4′-thio substitution occur prior to methyl transfer. An indication that the initial nucleophilic attack may be unfavorable comes from the inhibition seen in solution and the van der Waals length contacts observed between Cys81 and C6 in the crystal. The instability of the DNA–protein complex may also be related to the inability to form the initial transient covalent bond. For these reasons, the step most likely inhibited in solution by the introduction of a sulfur atom at the 4′-deoxyribose position is the formation of the covalent bond between the catalytic cysteine and cytosine C6.

Given the close similarity between the S13 structure and previous M.*Hha*I structures and the lack of a culpable change in conformation around the reactive C5 or C6 atoms, we propose that the observed inhibition is the result of perturbations in the electronic environment of the cytosine ring. The lower electronegativity of sulfur relative to oxygen (2.5 versus 3.5) and the separation of the 4′ atom from the ring atoms lowers the likelihood of direct inductive or resonance effects by sulfur on the electron distribution within the pyrimidine ring. However, it is worth noting that the related compounds 5-fluoro-4′-thio-2′-deoxycytosine and 5-fluoro-4′-thio-2′-deoxyuridine have absorption profiles that differ from the analogous compounds bearing a 4′-oxygen (R.T.W., unpublished results), so some influence may occur. The 4′-sulfur atom may indirectly affect the electronic environment of the pyrimidine ring by perturbing the charge interactions of other groups with the ring atoms. By virtue of its larger van der Waals radius (1.85 versus 1.40 Å for O) sulfur appears to 'bump' the nearby Arg165 residue and thereby increase the separation between the terminal amino groups of Arg165 and the nearby charged phosphate group which normally participate in a charge–charge interaction. The separation between the charged functional groups of Arg165 and the O2 carbonyl of cytosine is also slightly increased. It is the latter ring attached atom that may be the focus of the electronic disturbance.

The influence of the pyrimidine ring's delocalized electronic structure on the rate of methylation has been noted previously. Baker *et al.* (40) have suggested that cytosines which are protonated at N3 should be preferentially methylated because protonation substantially increases the charge on C6 (40,41), predisposing the ring to nucleophilic attack at that position (42). Liu and Santi (43) have shown that protonation of N3 is required to initiate catalysis of dCMP methylation by the N229D mutant of thymidylate synthase which has a catalytic mechanism similar to that of C5-Mtases. Since the π molecular orbital system of cytosine encompasses the O2 carbonyl, the charge on C6 will be affected by the delocalized charge distribution on O2 as well as N3. Gabbara *et al.* (21) have suggested that delocalized charge may be stabilized on the O2 carbonyl by interaction with the nearby arginine (Arg165) as an alternative to protonation at N3. Their model presupposes that nucleophilic attack has already taken place; however, the influence of Arg165 on the O2 carbonyl that they propose is still of significance. Changes in the electronic environment around the O2 carbonyl may also destabilize the charge distribution on this atom, feeding excess electrons back into the ring. Decreasing the charge on C6 would reduce the propensity for nucleophilic attack on this atom. Alternatively,

excessive polarization or protonation of the O2 oxygen may sequester delocalized electrons and inhibit later stages of the reaction. Changes in the electron distribution around the ring may also perturb the pKa of the N3 atom and alter its protonation. The precise nature of the electronic effects occurring in the S13 structure are too complicated to predict accurately based on the current structure, but the importance of the O2 carbonyl in influencing the rate of the methylation reaction is worth noting.

In conclusion, our work suggests that 4′-thio-2′-deoxycytidine may be an excellent compound for inhibiting DNA cytosine methylation. This result may help to explain why incorporation of 4′-thio-2′-deoxynucleotides into DNA results in toxicity to the organism. Thus, 4′-thiothymidine which is a substrate for cellular and viral kinases is toxic to both cells and herpes virus (44), whereas 4′-thio-5-ethyl-2′-deoxyuridine which is only a substrate for herpes virus kinase only shows overt toxicity to the virus (45). It is probable that the inhibition of methylation caused by the presence of 4′-thio-2′-deoxycytidine in the DNA, seen here, is only one example of many necessary metabolic steps which require precise nucleic acid–protein recognition and which are disrupted by the modification. In the case of cytosine methylation by M.*Hha*I in solution, the modification probably acts by disrupting the electronic configuration of cytosine required for methyl transfer. The thio-modification does not appear to distort the structure of the DNA grossly nor hinder its ability to be recognized by DNA-binding enzymes. The work also suggests that the currently accepted reaction pathway may require modification in order to include the influence of Arg165 on the reactivity of the cytosine base and a more significant role for the O2 carbonyl in determining the reaction rate. Clearly, the reaction mechanism of cytosine methylation is more sensitive to subtle changes in its environment than once believed and merits further study.

ACKNOWLEDGEMENTS

We thank K. McCloy for protein purification and initial crystallization studies of the DNA–protein complex. We are grateful to R. Knott, J. Buswell and the organic synthesis division of NEB for synthesis and purification of oligonucleotides containing 4-thio and 5-fluoro modified nucleosides. We also thank M. O'Gara, A. Mackerel, D. Landry, C. Noren and A. Kumar for helpful discussions and critical comments on the manuscript. The authors wish to thank the Leverhulme Trust and Amgen Inc. for financial support. This work was also supported by grants from the National Institutes of Health (GM49245, GM/OD52117 to X.C. and GM46127 to R.J.R.).

REFERENCES

1 Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994) *Nucleic Acids Res.* **22**, 1–10.
2 Bestor, T.H., and Verdine, G.L. (1994) *Curr. Opin. Cell Biol.* **6**, 380–389.
3 Roberts, R.J., and Halford, S.S. (1993) In Linn, S.M., Loyd, R.S., and Roberts, R.J. (eds) *Nucleases.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 35–88.
4 Barlow, D.P. (1995) *Science* **270**, 1610–1613.
5 Li, E., Bestor, T.H., and Jaenisch, R. (1992) *Cell* **69**, 915–926.
6 Jost, J.P., and Saluz, H.P. (eds) (1990) *DNA Methylation: Molecular Biology and Biological Significance*. Birkhauser Verlag, Basel, Switzerland.
7 Rideout, W.M. 3d, Coetzee, G.A., Olumi, A.F., and Jones, P.A. (1990) *Science* **249**, 1288–1290.
8 Laird, P.W., and Jaenisch, R. (1994) *Hum. Mol. Genet.* **3**, 1487–1495.
9 Kumar, S., Cheng, X., Pflugrath, J.W., and Roberts, R.J. (1992) *Biochemistry* **31**, 8648–8653.
10 Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J. (1993) *Cell* **74**, 299–307.
11 Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994) *Cell* **76**, 357–369.
12 O'Gara, M., Klimasauskas, S., Roberts, R.J., and Cheng, X. (1996) *J. Mol. Biol.* **261**, 634–645.
13 O'Gara, M., Roberts, R.J., and Cheng, X. (1996) *J. Mol. Biol.* **263**, 597–606.
14 Reinisch, K.M., Chen, L., Verdine, G.L., and Lipscomb, W.N. (1995) *Cell* **82**, 143–153.
15 Vassylyev, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) *Cell* **83**, 773–782.
16 Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E., and Tainer, J.A. (1996). *Nature* **384**, 87–92.
17 Roberts, R.J. (1995) *Cell* **82**, 9–12.
18 Cheng, X., and Blumenthal, R. (1996) *Structure* **4**, 639–645.
19 Wu, J.C., and Santi, D.V. (1987) *J. Biol. Chem.* **262**, 4778–4786.
20 Mi, S., and Roberts, R.J. (1993) *Nucleic Acids Res.* **21**, 2459–2464.
21 Gabbara, S., Sheluho, D., and Bhagwat, A.S. (1995) *Biochemistry* **34**, 8914–8923.
22 Chen, L., MacMillan, A.M., and Verdine, G.L. (1993) *J. Am. Chem. Soc.* **115**, 5318–5319.
23 Boggon, T.J., Hancox, E.L., McAuley-Hecht, K.E., Connolly, B.A., Hunter, W.N., Brown, T., Walker, R.T., and Leonard, G.A. (1996) *Nucleic Acids Res.* **24**, 951–961.
24 Hancox, E.L., Connolly, B.A., and Walker, R.T. (1993) *Nucleic Acids Res.* **21**, 3485–3491.
25 Van Draanen, N.A., Freeman, G.A., Short, S.A., Harvey, R., Jansen, R., Szczech, G., and Koszalka, G.W. (1996) *J. Med. Chem.* **39**, 538–542.
26 Jones, G.D., Lesnik, E.A., Owens, S.R., Risen, L.M., and Walker, R.T. (1996) *Nucleic Acids Res.* **24**, 4117–4122.
27 Klimasauskas, S., and Roberts, R.J. (1995) *Nucleic Acids Res.* **23**, 1388–1395.
28 Segal, I.H. (1975) *Enzyme Kinetics*. John Wiley & Sons, Inc. New York, NY.
29 Osterman, D.G., DePillis, G.D., Wu, J.C., Matsuda, A., and Santi, D.V. (1988) *Biochemistry* **27**, 5204–5210.
30 Brunger, A.T. (1992) *X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR.* Yale University Press, New Haven, CT.
31 Secrist, J.A. 3d, Tiwari, K.N., Riordan, J.M., and Montgomery, J.A. (1991) *J. Med. Chem.* **34**, 2361–2366.
32 Brown, D.M., Todd, A., and Varadarajan, S. (1956) *J. Chem. Soc*. 2384.
33 Dyson, M.R., Coe, P.L., and Walker, R.T. (1991) *Carbohydrate Res.* **216**, 237–248.
34 Dubey, A.K., and Roberts, R.J. (1992) *Nucleic Acids Res.* **20**, 3167–3173.
35 Wyszynski, M.W., Gabbara, S., Kubareva, E.A., Romanova, E.A., Oretskaya, T.S., Gromova, E.S., Shabarova, Z.A., and Bhagwat, A.S. (1993) *Nucleic Acids Res.* **21**, 295–301.
36 Yang, A.S., Shen, J.C., Zingg, J.M., Mi, S. and Jones, P.A. (1995) *Nucleic Acids Res.* **23**, 1380–1387.
37 Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
38 Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, NY.
39 Koole, L.H., Plavec, J., Liu, H., Vincent, B.R., Dyson, M.R., Coe, P.L., Walker, R.T., Hardy, G.W., Rahim, S.G., and Chattopadhyaya, J. (1992) *J. Am. Chem. Soc*. **114**, 9936–9943.
40 Baker, D.J., Kan, J.L.C., and Smith, S.S. (1988) *Gene* **74**, 207–210.
41 Jordan, F., and Sostman, H.D. (1973) *J. Am. Chem. Soc*. **95**, 6544–6554.
42 Hayatsu, H. (1976) *Biochemistry* **15**, 2677–2682.
43 Liu, L. and Santi, D.V. (1992) *Biochemistry* **31**, 5100–5104.
44 Dyson, M.R., Coe, P.L., and Walker, R.T. (1991) *J. Med. Chem.* **34**, 2782–2786.
45 Rahim, S.G., Trivedi, N., Bogunovic-Batchelor, M.V., Hardy, G.W., Mills, G., Selway, J.W., Snowden, W., Littler, E., Coe, P.L., Basnak, I., Whale, R.F., and Walker, R.T. (1996) *J. Med. Chem.* **39**, 789–795.