Weissmann, C. (1974) *FEBS Lett. 40*, S10–S18.

Weissmann, C., Borst, P., Burdon, R. H., Billeter, M. A., & Ochoa, S. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 890–897.

Weissmann, C., Feix, G., Slor, H., & Pollet, R. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1870–1877.

Weissmann, C., Feix, G., & Slor, H. (1968) *Cold Spring Harbor Symp. Quant. Biol. 33*, 83–100.

Weissmann, C., Billeter, M. A., Goodman, H. M., Hindley, J., & Weber, H. (1973) *Annu. Rev. Biochem. 42*, 303–329.

Werner, M. (1991) *Biochemistry 30*, 5832–5838.

Winkler-Oswatitsch, R., & Eigen, M. (1979) *Angew. Chem., Int. Ed. Engl. 18*, 20–49.

Witherell, G. W., & Uhlenbeck, O. C. (1989) *Biochemistry 28*, 71–76.

Witherell, G. W., Wu, H. N., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 11051–11057.

Yonesaki, T., & Aoyama, A. (1981) *J. Biochem. (Tokyo) 89*, 741–750.

Yonesaki, T., Furuse, K., Haruna, I., & Watanabe, I. (1982) *Virology 116*, 379–381.

Accelerated Publications

Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase[†]

Lin Chen,[‡] Andrew M. MacMillan,[‡] Wilbur Chang,[‡] Khosro Ezaz-Nikpay,[‡] William S. Lane,[§] and Gregory L. Verdine*[,‡]

Department of Chemistry and Microchemistry Facility, Harvard University, Cambridge, Massachusetts 02138

Received August 29, 1991; Revised Manuscript Received September 27, 1991

ABSTRACT: The overproduction, purification, and determination of the active-site catalytic nucleophile of the DNA (cytosine-5)-methyltransferase (DCMtase) enzyme M.*Hae*III are reported. Incubation of purified M.*Hae*III with an oligodeoxynucleotide specifically modified with the mechanism-based inhibitor 5-fluoro-2′-deoxycytidine [Osterman, D. G., et al. (1988) *Biochemistry 27*, 5204–5210], in the presence of the cofactor *S*-adenosyl-L-methionine (AdoMet), resulted in the formation of a covalent DNA–M.*Hae*III complex, which was purified to homogeneity. Characterization of the intact complex showed it to consist of one molecule of the FdC-containing duplex oligonucleotide, one molecule of M.*Hae*III, and one methyl group derived from AdoMet. Exhaustive proteolysis, reduction, and alkylation of the DNA–M.*Hae*III complex led to the isolation of two DNA-bound peptides—one each from treatment with Pronase or trypsin—which were subjected to peptide sequencing in order to identify the DNA attachment site. Both peptides were derived from the region of M.*Hae*III containing a Pro-Cys sequence that is conserved in all known DCMtases. At the position of this conserved Cys residue (Cys$_{71}$), in the sequence of each peptide, was found an unidentified amino acid residue; all other amino acid residues were in accord with the known sequence. It is thus concluded that Cys$_{71}$ of M.*Hae*III forms a covalent bond to DNA during catalytic methyl transfer. This finding represents a direct experimental verification for the hypothesis that the conserved Cys residue of DCMtases is the catalytic nucleophile [Wu, J. C., & Santi, D. V. (1987) *J. Biol. Chem. 262*, 4778–4786]. Furthermore, the present studies provide ready access to large quantities of a homogeneous, covalent protein–DNA complex that is trapped at an intermediate stage in catalysis.

$\mathbf{T}$he enzymatic addition of methyl groups to DNA is an essential element of genomic function in organisms ranging from bacteria to mammals (Razin et al., 1984; Adams & Burdon, 1985). In prokaryotes, DNA methylation directs the mismatch repair and restriction–modification systems, which correct errors of replication and prevent transformation by non-self-DNA, respectively. In eukaryotes, DNA methylation serves an essential yet poorly understood role in cell differentiation and regulation of gene expression. These complex and intriguing biological effects overlay the fundamental mechanistic questions surrounding how these proteins transfer a methyl group from the cofactor *S*-adenosyl-L-methionine (AdoMet)[1] to duplex DNA. The class of DNA methyltransferase enzymes that direct methylation of the 5-position of C [DNA (cytosine-5)-methyltransferases (DCMtases); eq

[1] Abbreviations: AdoMet, *S*-adenosyl-L-methionine; AdoHcy, *S*-adenosyl-L-homocysteine; BSA, bovine serum albumin; DCMtase, DNA (cytosine-5)-methyltransferase; ds, double stranded; DTT, dithiothreitol; ECPCR, expression-cassette polymerase chain reaction; FdC, 5-fluoro-2′-deoxycytidine; FdU, 5-fluoro-2′-deoxyuridine; IPTG, isopropyl β-D-thiogalactopyranoside; kDa, kilodaltons; M.*Hae*III, DNA (cytosine-5)-methyltransferase enzyme from *Haemophilus aegyptius*; M.*Hha*I, DCMtase from *Haemophilus haemolyticus*; oligo, oligo-2′-deoxynucleotide; PTH, phenylthiohydantoin; R.*Hae*III, restriction endonuclease from *H. aegyptius*; ss, single stranded; TMP-FdU, 4-*O*-(2,4,6-trimethylphenyl)-5-fluoro-2′-deoxyuridine.

1] has received a great deal of attention; this has intensified

(1)
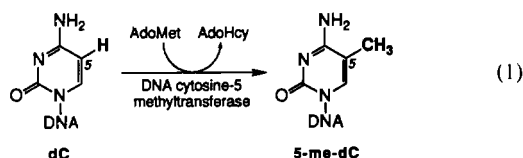
with recent reports that aberrations in cytosine-5-methylation may play a role in human genetic disease (Oberlé et al., 1991). DCMtases have also aroused significant mechanistic interest because they carry out what is prima facie a chemically improbable reaction: substitution of an unactivated vinyl hydrogen with a methyl group, in neutral aqueous solution.

DCMtases are believed to operate by the covalent catalysis mechanism shown in Figure 1a. In this reaction scheme, nucleophilic attack by the protein at the C6-position of cytosine generates a covalent protein–DNA intermediate in which cytosine C5 exists as a resonance-stabilized carbanion; this C-nucleophile attacks the methyl group of AdoMet, resulting in methylation of the covalent protein–DNA intermediate and expulsion of S-adenosyl-L-homocysteine (AdoHcy). Finally, loss of the protein nucleophile by β-elimination—assisted perhaps by a basic residue of the protein—liberates the methylated pyrimidine. Several lines of evidence suggest this mechanism, including the following. (i) Mechanistic conservation: all enzymes known to mediate one-carbon transfers to the C5-position of pyrimidines, including thymidylate synthetase and tRNA (uracil-5)-methyltransferase (Santi & Danenberg, 1984; Santi & Hardy, 1987), appear to operate by covalent addition at C6. (ii) Tritium exchange: incubation of the DCMtase HhaI (M.HhaI) and poly(dG-[5-^{3}H]dC) in the absence of cofactor has been shown to catalyze the release of tritium into the solvent (Wu & Santi, 1987), which is consistent with enzyme-mediated addition–elimination across the pyrimidine 5,6 double bond. (iii) Mechanism-based inactivation by 5-fluoro-2′-deoxycytidine (FdC): Santi and co-workers have shown that FdC residues in DNA, upon incubation with AdoMet and M.HhaI, are involved in the irreversible formation of a catalytically inactive, covalent protein–DNA complex (Osterman et al., 1987). The pathway proposed for this inactivation is shown in Figure 1b; it is analogous to the normal case (Figure 1a) until the β-elimination step, which cannot occur with FdC because of the extremely unfavorable energetics of abstracting F$^+$.

Extensive sequence comparisons of DCMtases (Lauster et al., 1989; Posfai et al., 1989) have revealed the presence of an absolutely conserved Pro-Cys dipeptide, of which the Cys-SH has been suggested (Wu & Santi, 1987) to serve as a catalytic nucleophile (Figure 1). The Pro-Cys sequence is conserved not only in DCMtases but also in thymidylate synthetases, for which its critical role in covalent catalysis has been established conclusively by X-ray crystallography (Matthews et al., 1990a,b). The importance of this Cys in catalysis by DCMtases has also been supported by mutagenesis (Wilke et al., 1988), photo-cross-linking studies (Som & Friedman, 1990), and inactivation by treatment with N-ethylmaleimide (Wu & Santi, 1987).

Despite this wealth of information, direct structural evidence for the involvement of an active-site Cys residue in covalent catalysis by DCMtases have proven elusive. In part, this situation has arisen from the difficulty of preparing a homogeneous duplex oligodeoxynucleotide (hereafter termed oligo) containing FdC at a single site. We have recently succeeded in preparing such an oligo and here report its use in mechanistic studies of a DCMtase. For the following reasons, we
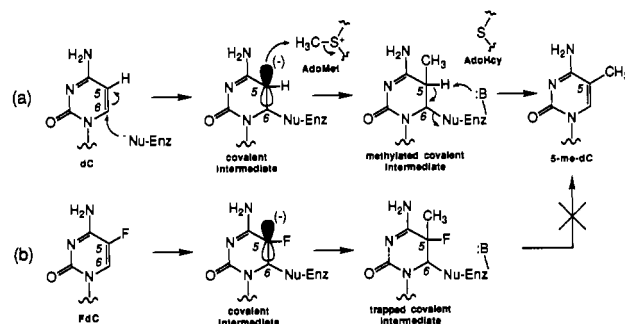

FIGURE 1: Proposed mechanisms for DCMtase-mediated reactions: (a) methylation of dC residues in DNA and (b) mechanism-based inactivation by FdC.
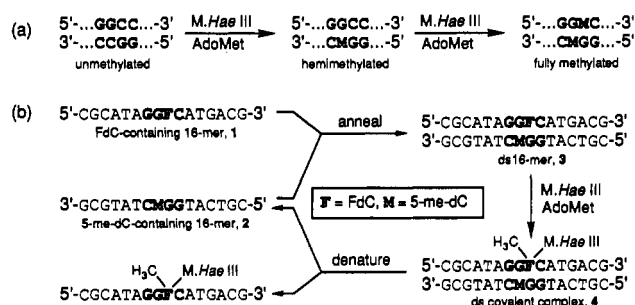

FIGURE 2: (a) Sequence-specific recognition and methylation of DNA catalyzed by M.HaeIII. (b) Formation of a trapped covalent complex between M.HaeIII and an FdC-containing oligonucleotide. In the oligo sequences, F = FdC, and M = 5-Me-dC.

chose to study the DCMtase enzyme from *Haemophilus aegyptius* (M.HaeIII): (i) as a bacterial type II modification methylase, M.HaeIII recognizes a tetranucleotide sequence, 5′-d(GGCC)-3′, within which it methylates specific dC residues (Figure 2a); this enzyme thus provides a system with which to study both recognition and catalysis in protein–DNA interactions; (ii) of the DCMtases for which gene sequences are available, more recognize 5′-d(GGCC)-3′ than any other site (Lauster et al., 1989); therefore, protein sequence homologies should be advantageous in identifying amino acids critical to DNA binding and catalysis; (iii) the gene encoding M.HaeIII has been cloned and sequenced (Slatko et al., 1988); and (iv) with a molecular mass of 37.7 kDa, M.HaeIII is one of the smallest DCMtases known.

Here, we report the overproduction in *Escherichia coli* of M.HaeIII, its mechanism-based inactivation by an FdC-containing oligo, and purification and characterization of the covalent protein–DNA complex. As discussed below, these studies have furnished direct evidence that the Cys residue of the conserved Pro-Cys doublet is indeed the catalytic nucleophile of DCMtase. Moreover, this preparation of a homogeneous, suicidally inactivated protein–DNA complex may provide the basis for high-resolution structural studies on sequence recognition and catalysis by this catalytic DNA-binding protein.

EXPERIMENTAL PROCEDURES

Materials

The M.HaeIII gene in pUC-19 was a gift from Dr. Geoffrey Wilson (New England Biolabs). *E. coli* MC 1061 was obtained from Chaitanya Jain (Harvard University); XL-1 Blue was from Stratagene. The 5-Me-dC phosphoramidite was from Sigma (St. Louis, MO). The GENECLEAN II kit was from Bio101 (La Jolla, CA). *Taq* polymerase was from Promega (Madison, WI) and T4 DNA ligase from BRL

(Gaithersburg, MD). Restriction enzymes and AdoMet were from New England Biolabs (Beverly, MA). Pronase was from Calbiochem (La Jolla, CA) and bovine trypsin from Seravac (Maidenhead, England). Nuclease PI and phosphodiesterase I were from Pharmacia (Milwaukee, WI). [methyl-^{3}H]-AdoMet (11 Ci/mmol) was from New England Nuclear (Boston, MA). Scintillation counting was performed on a Beckman LS 1801 instrument using Fisher Scintiverse II. PCR was carried out using a programmable thermal cycler (MJ Research). Oligos were synthesized on an Applied Biosystems 380A DNA synthesizer. The PRP-1 column was from Hamilton (Reno, NV). A Hewlett-Packard HP1090 system equipped with a diode array detector was used for HPLC. DEAE-cellulose (DE52) was from Whatman (Hillsboro, OR); S-Sepharose Fast Flow, Mono Q HR 5/5, and Superose 12 HR 10/30 columns were from Pharmacia. FPLC was performed on a Pharmacia system. H_2O was twice distilled.

Methods

Construction and Transformation of E. coli LC 1061. LC 1061 F'*lacI^{q}tet^r* is a derivative of MC 1061 (Wertman et al., 1986) that contains the F' plasmid from XL-1 Blue F'*lacI^{q}tet^r* (Bullock et al., 1987). Competent LC 1061 was prepared by the method of Hanahan (1985).

Construction of the M.HaeIII Overproducer. The M.HaeIII overproducer construction was carried out by the ECPCR method (MacFerrin et al., 1990, 1991) as follows: the oligos 5'-d(TAGGGCGAATTCAAGGAGATATAC-CATGAATTTAATTAGTCTTTTTTCAGG)-3' and 5'-d-(TCTAGGCAAAGCTTAATTATTACCTTTACAAAT-TTCCAATGC)-3' were employed in PCR, along with a pUC clone containing the native M.HaeIII gene on a 1.4-kb insert (Slatko et al., 1988), to generate an expression cassette containing the M.HaeIII gene. The M.HaeIII expression cassette was ligated into the transcription vector pHN1 (Schreiber & Verdine, 1991) and the overproducing construct (pM.HaeIII-1) transformed into LC 1061.

Expression and Purification of M.HaeIII. LC 1061-(pM.HaeIII-1) was grown at 37 °C with shaking in 1 L of LB–AST medium (LB plus 50 μg/mL ampicillin, 25 μg/mL streptomycin, and 12.5 μg/mL tetracycline) to an OD_{550} of 0.5–0.6. The cells were then transferred to a 30 °C incubator and induced by the addition of IPTG to a final concentration of 0.3 mM. After 4 h, the culture was centrifuged at 5000g for 30 min, after which time the supernatant was discarded to yield ~3 g of wet cells. All subsequent steps were carried out at 4 °C. The cells were resuspended in 30 mL of lysis buffer (10 mM β-mercaptoethanol, 0.5 mM EDTA, 20 mM phosphate buffer, pH 7.4) and lysed in a French press cell. The lysate was centrifuged at 34000g for 20 min, and the supernatant was then transferred to a fresh tube. The lysis supernatant was loaded onto a 55-mL DE52 column, which had previously been equilibrated with lysis buffer. The column was eluted with a 600-mL linear gradient from 0 to 0.4 M NaCl in lysis buffer (all columns were run at a flow rate of 1 mL/min). Fractions exhibiting M.HaeIII activity (see below) eluted at ~0.2 M NaCl; these were pooled and loaded onto a 31-mL S-Sepharose Fast Flow column preequilibrated in lysis buffer. The column was eluted with a 600-mL linear gradient from 0 to 1.0 M NaCl in lysis buffer. Fractions exhibiting M.HaeIII activity eluted at ~0.45 M NaCl. Active fractions were either used directly in trapping experiments or stored with the addition of glycerol to 30% (v/v) at –20 °C. SDS–PAGE analysis of M.HaeIII prepared by this procedure showed it to be of high purity and to correspond in size to the

expected molecular size (37.7 kDa). The yield as determined by Bradford assay was ~10 mg/L.

M.HaeIII Activity Assay. Fractions were assayed for the ability to protect plasmid DNA from restriction by HaeIII endonuclease (R.HaeIII). In a typical assay, 0.9 μg of pUC-19 was treated with M.HaeIII (1 μL of a column fraction) in 10 μL of buffer consisting of 80 μM AdoMet, 50 mM NaCl, 50 mM Tris-HCl (pH 8.5), 10 mM EDTA, and 1 mM DTT. The methylation mixture was incubated at 37 °C for 30 min and then heated at 70 °C for 10 min. After cooling, the solution was made 20 mM in $MgCl_2$ and one unit of HaeIII endonuclease was added. The reaction was incubated at 37 °C for 30 min and then analyzed by 1% agarose gel electrophoresis.

Synthesis of the FdC-Containing 16-mer 1. Full details of this synthesis will be published elsewhere.[2] Briefly, 4-O-(2,4,6-trimethylphenyl)-5-fluoro-2'-deoxyuridine (TMP-FdU) was converted to the corresponding β-cyanoethyl phosphoramidite (TMP-FdU phosphoramidite). The TMP-FdU phosphoramidite, the four conventional β-cyanoethyl phosphoramidites, and the 5-Me-dC phosphoramidite were employed in the 10-μmol-scale synthesis of the fully protected oligos 5'-d[CGCATAGG(TMP-FdU)CATGACG]-3' and 3'-d[GCGTATC(5-Me-dC)GGTACTGC]-5'. The crude 5'-tritylated oligos were treated with concentrated ammonium hydroxide for 14 h at 55 °C, lyophilized, and purified by reverse-phase HPLC on a Hamilton PRP-1 column. Fractions were pooled, lyophilized, and detritylated (3% aqueous acetic acid, 10 min) to yield 1.65 μmol of the FdC-containing 16-mer 1 and 1.83 μmol of the 5-Me-dC-containing 16-mer 2 (Figure 2b). Nucleoside compositions were confirmed by enzymatic digestion and HPLC analysis in comparison with authentic standards.

Reaction of M.HaeIII with the Duplex 16-mer 3. Reactions were performed directly with active fractions from the S-Sepharose Fast Flow column. A typical 200-μL reaction contained 20 mM potassium phosphate buffer (pH 7.5), 200 mM NaCl, 10 mM β-mercaptoethanol, 0.5 mM EDTA, ~1.5 μM M.HaeIII, 3 μM oligo-3, and 80 μM AdoMet. The reaction was incubated at 37 °C for more than 6 h. This reaction typically yielded 200–250 pmol of the covalent complex 4, after purification (see below). For SDS–PAGE analysis, a 10-μL aliquot of this reaction mixture was mixed with an equal volume of loading dye. The sample was heated at 90 °C for 10 min and run on an SDS–PAGE gel, which was then stained with Coomassie blue. Large-scale complex formation was performed in a volume of 10 mL.

Purification of the Covalent M.HaeIII–DNA Complex 4 and Its Proteolytic Fragments. The reaction mixture from complex formation was loaded onto an FPLC Mono Q HR 5/5 column, equilibrated in buffer A [20 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 10 mM β-mercaptoethanol]. The column was run with a 28-mL linear gradient from 0 to 1 M NaCl in buffer A (flow rate, 1 mL/min). Complex 4 eluted at ~350 mM NaCl. Samples obtained by proteolysis of complex 4 were handled similarly and eluted at a higher salt concentration. Samples were desalted as follows: samples were mixed with 4 volumes of –20 °C acetone and kept at –20 °C for 15 min. Following microcentrifugation at 4 °C for 15 min, the supernatant was removed and the residue lyophilized and redissolved in 50 μL of H_2O.

Size Determination by Gel Filtration. FPLC-purified samples (~3 pmol) were loaded onto a Superose 12 HR 10/30 column equilibrated in buffer B: [200 mM NaCl, 20 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 2 mM DTT], operating

at a flow rate of 0.25 mL/min. Retention volumes for a series of standards, BSA (67 kDa; 12.1 mL), *E. coli* Ada protein (39 kDa; 13.7 mL), cytochrome *c* (12 kDa; 15.1 mL), were compared to those of M.*Hae*III (37.7 kDa; 13.9 mL) and complex **4** (13.1 mL). When molecular mass was plotted vs retention volume, the molecular mass of **4** was determined to be ~50 kDa.

*Formation of Complex **4** Using [methyl-³H]AdoMet.* The ds 16-mer **3** was reacted at 37 °C with a 2-fold excess of M.*Hae*III in 50 μL of buffer containing 68 μM [methyl-³H]AdoMet, 20 mM phosphate buffer (pH 7.5), 200 mM NaCl, 10 mM β-mercaptoethanol, and 0.5 mM EDTA at 37 °C for 8 h. The sample was diluted to 850 μL with H_2O and dialyzed against H_2O for 12 h. Samples (100 μL) from inside and outside the dialysis bags were diluted into 3 mL of scintillation liquid and counted.

*Proteolysis of the Covalent M.HaeIII–DNA Complex **4**.* (*a*) *Pronase.* Pronase was dissolved in degassed 0.1 M potassium phosphate buffer (pH 7.4) to a concentration of 1 mg/mL. A 2.4-nmol pellet of desalted, FPLC-purified **4** was dissolved in 400 μL of this solution, and the reaction mixture was incubated at 37 °C overnight. The proteolysis mixture was fractionated on Mono Q as described above.

(*b*) *Trypsin.* A 1-nmol sample of FPLC-purified, desalted **4** was dissolved in 100 μL of H_2O and then mixed with an equal volume of buffer [4 M urea, 100 mM Tris-HCl (pH 8.0), 2 mM $CaCl_2$]. Trypsin (1 mg) was added to the reaction mixture, which was then incubated at 25 °C for 2 days. The proteolysis mixture was fractionated on Mono Q as described above.

Nuclease Digestion of Purified Peptide–DNA Fragments. The reduced and alkylated peptide–DNA fragment was FPLC purified, desalted, redissolved in 180 μL of H_2O, and mixed with 20 μL of nuclease digestion cocktail: nuclease P1 (~5 μg) and phosphodiesterase I (~5 μg) in 500 mM Tris-HCl (pH 7.5), 20 mM $CaCl_2$, and 80 mM $MgCl_2$. The reaction mixture was incubated at 37 °C overnight.

*Edman sequencing of the M.HaeIII–DNA complex **4** and its proteolytic fragments* was carried out as described (Lane et al., 1991), using an on-line ABI Model 120A HPLC system.

RESULTS

Amine groups on nucleosides must be protected during automated DNA synthesis to prevent their participation in condensation reactions. This requirement has made it very difficult to prepare the reagent required for synthesis of FdC-containing DNA, since attempts to use conventional acyl protecting groups for the amine function of FdC have produced unstable compounds (Osterman et al., 1988; MacMillan et al., unpublished results). On the basis of the demonstration that aryl ethers could serve as latent nucleoside amine groups (Zhou & Chatthopadhyaya, 1986), the trimethylphenyl ether of FdU (TMP-FdU; eq 2) was prepared and converted into the cor-

$$\text{TMP-FdU} \xrightarrow[\text{NH}_4\text{OH}]{\text{conc.}} \text{FdC} \tag{2}$$
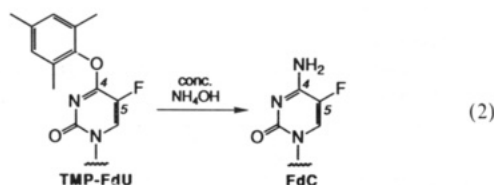
responding phosphoramidite. TMP-FdU phosphoramidite was found to undergo efficient coupling during DNA synthesis and to be converted quantitatively to FdC during oligo deprotection using ammonia. By this chemistry a large quantity (>1.5 μmol) of an oligonucleotide bearing a single FdC moiety,
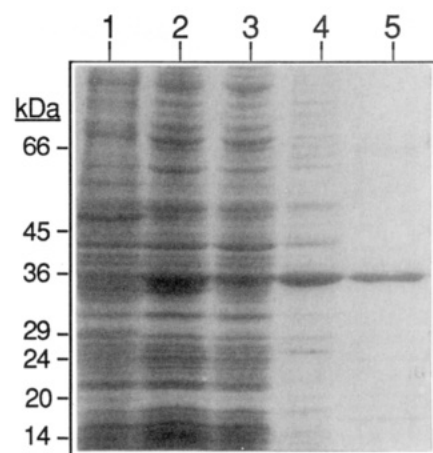

FIGURE 3: Induction and purification of M.*Hae*III: lanes 1 and 2, uninduced and induced *E. coli* LC1061 harboring the M.*Hae*III overproducing phagemid pM.*Hae*III-1; lane 3, supernatant from cell lysis; lanes 4 and 5, pooled fractions from DEAE-cellulose and S-Sepharose chromatography. Denoted on the left are mobilities and sizes (in kDa) of molecular mass markers.

5′-d(CGCATAGGFCATGACG)-3′ (**1**; Figure 2b; F = FdC), was synthesized.[2]

Like all DCMtases that recognize dyad-symmetric sites, M.*Hae*III carries out two sequential methylations of its target site (Figure 2a). For the present studies, it was desirable to carry out the inactivation reaction on a hemimethylated site, so that the methylation state of the non-FdC-containing strand in the duplex substrate would be predetermined. This was accomplished by synthesizing the complementary 16-mer **2** with 5-Me-dC at its M.*Hae*III methylation site (Figure 2b). Equimolar amounts of the single-stranded (ss) oligos **1** and **2** were then mixed to form the double-stranded (ds) suicide substrate **3**.

In order to obtain large quantities of M.*Hae*III for these and future studies, the protein was overproduced in *E. coli* using the expression-cassette polymerase chain reaction (ECPCR) strategy (MacFerrin et al., 1990, 1991; Schreiber & Verdine, 1991). An expression cassette having the M.*Hae*III gene under the translational control of the T7 gene 10 ribosome binding site (5′-AAGGAGATATACC-3′) was inserted into the transcription phagemid vector pHN1+ to give the M.*Hae*III overproduction vector pM.*Hae*III-1. Stable transformation of pM.*Hae*III-1 into *E. coli* required the use of a strain lacking the mcrA and mcrB endonucleases (Raleigh et al., 1988), and repression of the vector's *tac* promoter necessitated a strain with a high titer of the *lac* repressor. To obtain such a strain, conjugation was used to transfer an F′ plasmid containing *lacI*q into the *mcrA⁻B⁻* strain MC 1061, thus yielding *E. coli* LC 1061. Transformation of LC 1061 by pM.*Hae*III-1 yielded an overproducer that expressed M.*Hae*III as the predominant cell protein (Figure 3).

M.*Hae*III overproduced in *E. coli* was present entirely in soluble form; this contrasts with *M.Hha*I overproduced in *E. coli*, in which case a large portion of the protein was found in insoluble, inactive form (Wu & Santi, 1988). Highly purified M.*Hae*III was obtained via a two-step procedure consisting of chromatography on DEAE-cellulose followed by S-Sepharose (Figure 3). The N-terminal sequence of recombinant M.*Hae*III (Met-Asn-Leu-Ile-Ser) and molecular size (~38 kDa by SDS–PAGE vs 37.7 kDa calculated) were in

---

[2] A. M. MacMillan, L. Chen, and G. L. Verdine, submitted for publication.
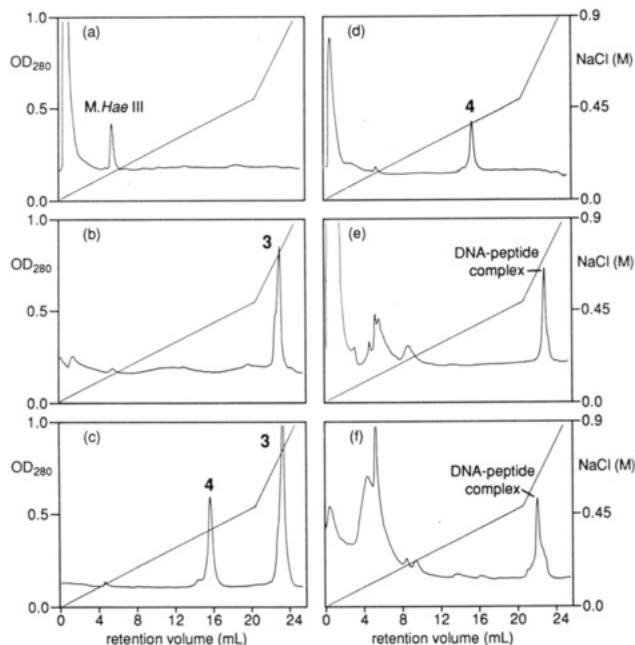
FIGURE 4: FPLC Mono Q elution profiles of various species generated in this study: (a) M.*Hae*III; (b) duplex oligo substrate **3**; (c) products generated from reacting M.*Hae*III with excess **3** and AdoMet; (d) purified protein–DNA complex **4**; e and f products arising from treatment of purified **4** with Pronase and trypsin, respectively. The large peak in the 0–2-mL range of (a), (d), and (e) is due to thiol reducing agents. The line running from lower left to upper right denotes the position of the NaCl gradient with respect to the right-hand scale.

FIGURE 5: SDS–PAGE analysis of the covalent complex formed between M.*Hae*III and oligo **3**. Denoted on the left are mobilities and sizes (in kDA) of molecular mass markers. A small band at the position of M.*Hae*III arose from decomposition of the complex upon extended heating at 100 °C rather than from contamination of complex **4** by M.*Hae*III (data not shown).

accord with predictions based on the known sequence. Moreover, recombinant M.*Hae*III was found to be highly active when assayed for the ability to protect plasmid DNA from digestion by R.*Hae*III and for mechanism-based inactivation by oligo **3** (see below).

During the purification of M.*Hae*III, it was found that this protein binds weakly to anion-exchange matrices; this is illustrated in Figure 4a for chromatography on Mono Q, in which M.*Hae*III eluted at ~0.15 M NaCl (5.8 mL). On the other hand, DNA substrate **3**, being a polyanion, bound strongly to Mono Q (Figure 4b; ~0.7 M NaCl; 24.0 mL). When M.*Hae*III was incubated with excess **3** in the presence of AdoMet, the formation of a new, intermediate mobility peak was observed (Figure 4c; ~0.35 M NaCl; 16.0 mL). This new species, which was purified to homogeneity on Mono Q (Figure 4d), has been assigned structure **4** (Figure 2) on the basis of the studies described below.

The UV spectrum of **4** (not shown) bore features characteristic of both protein and DNA (not shown), with the dominant absorbance being the DNA band at ~260 nm. Protein determination by Bradford assay suggested that **4** contained a significant amount of proteinaceous material (the DNA stained very weakly), and N-terminal sequencing confirmed that the protein component was derived from *M.Hae*III (Met-Asn-Leu-Ile-Ser). Size-exclusion FPLC analysis against known protein standards, under nondenaturing conditions, indicated that **4** had a molecular mass of approximately 50 kDa, which was consistent with its being a 1:1 protein–DNA complex (the DNA molecule **3** alone had a mobility consistent with its size).

To ascertain whether **4** was a covalent complex, a purified sample was denatured vigorously and analyzed by SDS–PAGE, the results of which are presented in Figure 5. The major protein species was found to differ from M.*Hae*III, having an apparent molecular mass of ~43 kDa. This cor-
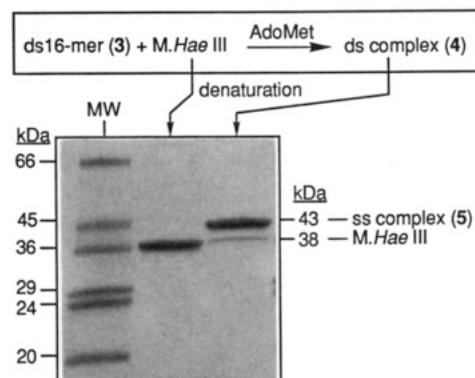
responds to the predicted size of a complex having M.*Hae*III bound to a single-stranded 16-mer—structure **5**—from which the complementary strand had dissociated under the denaturing conditions of the experiment. This difference in gel mobility formed the basis for a convenient activity assay: when freshly prepared M.*Hae*III was titrated with oligo **3** to saturation, greater than 90% of the protein shifted from ~38 to ~43 kDa, indicating that the protein was greater than 90% active (data not shown). Given the strongly denaturing conditions under which these samples were prepared, we concluded that, in forming complexes **4** and **5**, M.*Hae*III had become covalently attached to DNA.

When the incubation of M.*Hae*III with **3** was carried out in the absence of AdoMet, no complex was observed on either Mono Q or SDS–PAGE, suggesting that methyl transfer from AdoMet was an obligate event in the irreversible formation of a covalent protein–DNA complex. This was confirmed by formation of the complex using [*methyl*-^{3}H]AdoMet, in which 0.7 ± 0.3 methyl groups were transferred per mole of ds 16-mer; this radioactivity was not released upon denaturation (data not shown). These results are in accord with the earlier work of Osterman et al. (1988).

Our strategy for identification of the active-site protein–DNA linkage site in complex **4** took advantage of the extremely strong binding of DNA to Mono Q (Figure 4b,c, peak **3**). It was reasoned that DNA with a small attached peptide should also bind strongly, thus allowing it to be separated readily from the early-eluting, non-DNA-bound peptides. Complex **4** was therefore subjected to exhaustive digestion with Pronase, and the digest mixture was fractionated on Mono Q (Figure 4e). In this digest, **4** had been completely converted to other products (compare panels d and e, of Figure 4), one of which eluted late (23.7 mL). UV analysis of peak fractions revealed that only the 23.7-mL peak contained appreciable levels of DNA. The complex was digested in parallel with trypsin under weakly denaturing conditions, and this reaction also produced a late-eluting peak (23.0 mL, Figure 4f), which contained essentially all of the DNA. The 23.7-mL peak (Figure 4e) was subjected directly to Edman degradation, by which a single peptide sequence was found to be present: GGPPXQS (X = an unidentified amino acid). This heptapeptide uniquely corresponds to amino acids 67–73 of M.*Hae*III, thus identifying the X residue as Cys$_{71}$ (Figure 6). The 23.0-mL peak derived from trypsin digestion (Figure 4f) was also subjected to Edman degradation and found to have the sequence XDGIIGGPPXQSWSEGGSLR. This complete tryptic peptide corresponded to residues 62–81 of M.*Hae*III,
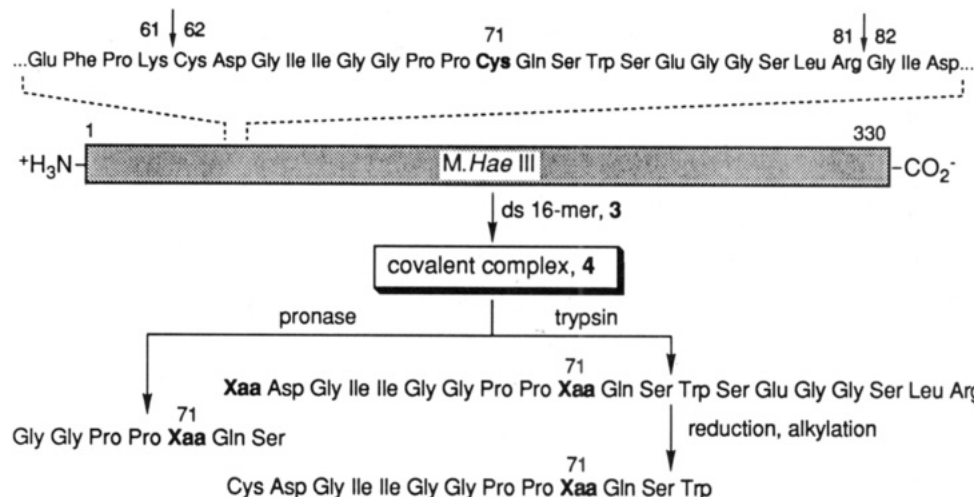
FIGURE 6: Schematic representation of M.*Hae*III (shaded) and, in the expansion above, a peptide segment containing the putative active-site Cys (bold), which is part of a Pro-Cys motif conserved in all known DCMtases. Proteolytic digestion of the trapped covalent M.*Hae*III–DNA complex **4** led to the isolation of peptides whose sequences are shown below. The tryptic peptide was reduced, alkylated, and resubjected to Edman degradation to provide the sequence at the bottom. Arrows denote the predicted positions of trypsin digestion within this region of M.*Hae*III; the termination of the sequence during Edman degradation corresponded to its predicted C-terminus, indicating a complete DNA-attached tryptic fragment had been sequenced. The Pronase-derived fragment was also sequenced in entirety, judging from its behavior during sequencing. Following reduction and alkylation, only a partial sequence of the tryptic fragment could be obtained; all residues shown (except Xaa) were assigned with high confidence. Xaa refers to an unidentified PTH-amino acid residue.

with the X residues being Cys_{62} and Cys_{71}. The detection of X residues in lieu of cysteine could be due either to the inherent inability of this residue, in the free thiol form, to be determined by PTH sequencing or, alternatively, to protection by linkage to DNA. The lack of Cys_{62} in the Pronase peptide ruled out this residue as the linkage site; nevertheless, in the tryptic peptide, Cys_{62} could serve as an internal standard for modification by iodoacetamide. The tryptic peptide was reduced, alkylated, and resubjected to Edman degradation, after which Cys_{62} was unambiguously identified as a PTH-carboxamidomethyl cysteine residue, and Cys_{71} remained unidentifiable. The protection of Cys_{71} but not Cys_{62} from iodoacetamide provides strong evidence that Cys_{71} is not present as the free thiol but rather is attached to (alkylated by; Figure 1) the DNA.

In order to rule out unambiguously the possibility of contamination by peptide artifacts, the DNA component of the reduced and alkylated tryptic peptide–DNA complex was removed by nuclease treatment, and the digest was purified by HPLC. By differential UV detection, five peaks were identified as being nucleoside–peptide conjugates; these differed only in their nucleoside component. The sequence of the major peak was found to be CDGIIGGPPXQS, confirming the position of DNA attachment as Cys_{71}.

Discussion

In this study, we have demonstrated the formation of a stable covalent complex between the DNA (cytosine-5)-methyltransferase M.*Hae*III and a duplex oligo containing the suicide substrate 5-fluoro-2'-deoxycytidine. The developments reported here should facilitate future structural studies, for the following reasons: (i) the M.*Hae*III overproducer affords high levels of native, folded protein in *E. coli*, and this protein is purified to near homogeneity by a two-column procedure; (ii) the FdC-containing oligo, **3**, is synthesized by automated methods on a multimilligram scale; and (iii) the protein–DNA complex is readily formed and purified on a preparative scale.

Our approach relies on the single-site incorporation of FdC into a duplex oligonucleotide through chemical synthesis. This contrasts with an alternative method in which FdC is incorporated into DNA enzymatically, resulting in 80% replacement of dC with FdC (Osterman et al., 1988). We chose the former route, reasoning it to be more suitable for future high-resolution structural studies and detailed enzymological characterization. The incorporation of a single FdC moiety in the oligonucleotide substrate minimizes perturbation of DNA secondary structure and assures that linkage of the protein to DNA will take place at only one site, thereby producing a structurally homogeneous protein–DNA complex. Single-site modification should also increase the precision and accuracy of kinetic measurements; furthermore, since the distance of an FdC moiety from the end of the DNA can be controlled, this synthetic approach should allow the contribution of facilitated diffusion (Jack et al., 1982) to be addressed directly.

To our knowledge, this study represents the first preparation of a homogeneous protein–DNA complex trapped in the midst of a catalytic event. FdC closely resembles the native substrate (dC) and does not appear to undergo rearrangement during inactivation; thus, the structure of complex **4** within the active site is likely to resemble closely the native intermediate in catalysis [cf. Figure 1 and Matthews et al. (1990a)]. X-ray analysis of this complex should therefore provide a detailed glimpse into substrate binding and activation, nucleophilic catalysis, and cofactor binding, in addition to illuminating sequence discrimination and its relationship to DNA structural perturbations that accompany the enzymatic reaction (Lesser et al., 1990). To date, two X-ray structures of catalytic DNA-binding proteins complexed to oligonucleotides have been reported: *Eco*RI endonuclease (McClarin et al., 1986; Kim et al., 1990) and bovine pancreatic deoxyribonuclease I (DNase I; Suck et al., 1988). Owing to particular details of mechanism and experiment in these studies, these structures have poorly assembled active sites and therefore leave many aspects of catalysis obscure. Use of a trapped catalytic intermediate such as that reported here may represent a promising approach for structural studies on catalytic DNA-binding proteins.

An early mechanistic appraisal of DCMtases (Santi et al., 1983) proposed that methylation of cytosine C5 is initiated by attack of a protein nucleophile at C6 (Figure 1), with the

attacking function probably being a Cys thiolate. This suggestion is consistent with findings that DCMtases require a free thiol group for catalytic activity (Cox, 1980). Drawing an analogy with thymidylate synthetase, Santi and Wu (1987) refined this mechanistic hypothesis by pointing to a specific Cys residue in DCMtases that is part of a conserved Pro-Cys dipeptide sequence. Subsequently, extensive sequence comparisons have shown that this dipeptide is indeed conserved in all DCMtases, including those of eukaryotic origin (Lauster et al., 1989). Finally, point mutations in the putative active-site Pro-Cys motif were shown to significantly diminish catalytic activity, although many other mutations throughout the DCMtase gene had similar effects (Wilke et al., 1988). While these results have suggested that the Pro-Cys motif contains a required center for nucleophilic catalysis, this proposal has remained without direct evidence to date. Our studies have provided such direct evidence. Specifically, proteolytic and nucleolytic degradation of the covalent M.*Hae*III–oligonucleotide complex 5 led to the isolation of peptides that contained the proposed active-site dipeptide: $Pro_{70}Cys_{71}$ in M.*Hae*III. Analysis of these peptides by automated sequencing revealed the presence of a single unidentifiable residue, at the position of Cys_{71}. Not only was this Cys protected from alkylation by iodoacetamide, but the release of a modified PTH-amino acid from its position during Edman degradation was accompanied by a significant increase in sequence lag, which is consistent with the attachment of a bulky substituent at that site. These data provide direct and compelling evidence for Cys_{71} as the position of protein–DNA linkage in 5; hence, Cys_{71} is the catalytic nucleophile of M. *Hae*III. By extension, these findings argue strongly that the Cys thiolate of the conserved Pro-Cys motif is the catalytic nucleophile of all known DCMtases.

The role of the Pro-Cys motif at the active site of DCMtases remains obscure. This motif is found not only in DCMtases but also in the active sites of several other proteins, all of which utilize the Cys as a nucleophile: thymidylate synthetase, which catalyzes the transfer of a methyl group to C5 of 2'-deoxyuridine 5'-monophosphate (dUMP), and a number of DNA repair proteins that remove methyl groups from DNA, of which the prototype is the *E. coli* Ada protein (Lindahl et al., 1988). Despite the fact that DCMtases and thymidylate synthetase contain an active-site Pro-Cys sequence, not all-enzymes that carry out one-carbon transfer to the C5-position of pyrimidines contain this motif. For example, the catalytic cysteine nucleophile of tRNA (uracil-5)-methyltransferase, recently identified by Santi and co-workers (Viola, 1991), is not preceded by Pro. Some evidence as to the function of the Pro-Cys motif may be evident in the X-ray structure of a suicidal complex between FdUMP and thymidylate synthetase, in which the amide carbonyls of Pro and the preceding (N-terminal) residue hydrogen-bond with a distal Arg residue. This Arg appears to serve two functions in addition to the aforementioned interactions: it (i) hydrogen-bonds with the charged phosphate of the substrate and (ii) stabilizes the incipient thiolate nucleophile. The bidentate hydrogen bond of Arg with peptide carbonyls preceding the nucleophilic residue may be oriented optimally by Pro due to the unique conformational properties of that residue. Such an interaction could be envisioned between DCMtases (and the Ada protein) and their phosphate-containing substrates.

REFERENCES

Adams, R. L. P., & Burdon, R. H. (1985) *Molecular Biology of DNA Methylation*, Springer-Verlag, New York.

Bullock, W. O., Fernandez, J. M., & Short, J. M. (1987) *BioTechniques 5*, 376–378.

Cox, R. (1980) *Cancer Res. 40*, 61–63.

Hanahan, D. (1985) in *DNA Cloning, Volume 1: A Practical Approach* (Glover, D. M., Ed.) pp 109–135, IRL Press, Washington, DC.

Jack, W. E., Terry, B. J., & Modrich, P. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4010–4014.

Kim, Y., Grable, J. C., Love, R., Greene, P. J., & Rosenberg, J. M. (1990) *Science (Washington, D.C.) 249*, 1307–1309.

Lane, W. S., Galat, A., Harding, M. W., & Schreiber, S. L. (1991) *J. Protein Chem. 10*, 151–160.

Lauster, R., Trautner, T. A., & Noyer-Weidner, M. (1989) *J. Mol. Biol. 206*, 305–312.

Lesser, D. R., Kurpiewski, M. R., & Jen-Jacobson, L. (1990) *Science (Washington, D.C.) 250*, 776–786.

Lindahl, T., Sedgwick, B., Sekiguchi, M., & Nakabeppu, Y. (1988) *Annu. Rev. Biochem. 57*, 133–157.

MacFerrin, K. D., Terranova, M. P., Schreiber, S. L., & Verdine, G. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1937–1941.

MacFerrin, K. D., Chen, L., Terranova, M. P., Schreiber, S. L., & Verdine, G. L. (1991) *Methods Enzymol.* (in press).

Matthews, D. A., Appelt, K., Oatley, S. J., & Xuong, Ng. H. (1990a) *J. Mol. Biol. 214*, 923–936.

Matthews, D. A., Villafranca, J. E., Janson, C. A., Smith, W. W., Welsh, K., & Freer, S. (1990b) *J. Mol. Biol. 214*, 937–948.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science (Washington, D.C.) 234*, 1526–1541.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M. F., & Mandel, J. L. (1991) *Science (Washington, D.C.) 252*, 1097–1102.

Osterman, D. G., DePillis, G. D., Wu, J. C., Matsuda, A., & Santi, D. V. (1988) *Biochemistry 27*, 5204–5210.

Posfai, J., Bhagwat, A. S., Posfai, G., & Roberts, R. J. (1989) *Nucleic Acids Res. 17*, 2421–2435.

Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Reith, A. D., Rigby, P. W. J., Elhai, J., & Hanahan, D. (1988) *Nucleic Acids Res. 16*, 1563–1575.

Razin, A., Cedar, H., & Riggs, A. D., Eds. (1984) *DNA Methylation: Biochemistry and Biological Significance*, Springer-Verlag, Berlin.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Santi, D. V., & Danenberg, P. V. (1984) in *Folates and Pterins* (Blakely, R. L., & Benkovic, S. J., Eds.) Vol. 1, pp 343–396, Wiley-Interscience, New York.

Santi, D. V., & Hardy, L. W. (1987) *Biochemistry 26*, 8599–8606.

Santi, D. V., Garrett, C. E., & Barr, P. J. (1983) *Cell (Cambridge, Mass.) 33*, 9–10.

Schreiber, S. L., & Verdine, G. L. (1991) *Tetrahedron 47*, 2543–2562.

Slatko, B. E., Croft, R., Moran, L. S., & Wilson, G. G. (1988) *Gene 74*, 45–50.

Som, S., & Friedman, S. (1990) *J. Biol. Chem. 265*, 4278–4283.

Suck, D., Lahm, A., & Oefner, C. (1988) *Nature (London)* *332*, 464–468.

Viola, R. E. (1991) *Bioorg. Chem. 19*, 190–224.

Wertman, K. F., Wyman, A. R., & Botstein, D. (1986) *Gene* *49*, 253–262.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., & Trautner, T. A. (1988) *EMBO J. 7*, 2601–2609.

Wu, J. C., & Santi, D. V. (1987) *J. Biol. Chem. 262*, 4778–4786.

Wu, J. C., & Santi, D. V. (1988) *Nucleic Acids Res. 16*, 703–717.

Zhou, X.-X., & Chattopadhyaya, J. (1986) *Tetrahedron 42*, 5149–5156.

Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Lipids: Differential Line Broadening Due to Cross-Correlation Effects as a Probe of Membrane Structure[†]

Eric Oldfield,* Foluso Adebodun, John Chung, Bernard Montez, Ki Deok Park, Hong-biao Le, and Brian Phillips

School of Chemical Sciences, University of Illinois, 505 South Mathews Avenue, Urbana, Illinois 61801

Received August 27, 1991; Revised Manuscript Received October 2, 1991

ABSTRACT: We have obtained proton-coupled carbon-13 nuclear magnetic resonance (NMR) spectra of a variety of lipid–water and lipid–drug–water systems, at 11.7 T, as a function of temperature, using the "magic-angle" sample-spinning (MAS) NMR technique. The resulting spectra show a wide range of line shapes, due to interferences between dipole–dipole and dipole–chemical shielding anisotropy interactions. The differential line-broadening effects observed are particularly large for aromatic and olefinic (sp^2) carbon atom sites. Coupled spectra of the tricyclic antidepressants desipramine and imipramine, in 1,2-di-myristoyl-*sn*-glycero-3-phosphocholine–water mesophases, show well-resolved doublets having different line shapes for each of the four aromatic methine groups, due to selective averaging of the four C–H dipolar interactions due to rapid motion about the director (or drug C_2) axis. ^2H NMR spectra of [2,4,6,8-$^2H_4$]desipramine (and imipramine) in the same 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine–water mesophase exhibit quadrupole splittings of ~0–2 and ~20 kHz, indicating an approximate magic-angle orientation of the C2–^2H(^1H) and C8–^2H(^1H) vectors with respect to an axis of motional averaging, in accord with the ^{13}C NMR results. Selective deuteration of imipramine confirms these ideas. Spectra of digalactosyl diglyceride [primarily 1,2-di[(9Z,12Z,15Z)-octadeca-9,12,15-trienoyl]-3-(α-D-galactopyranosyl-1–6-β-D-galactopyranosyl)-*sn*-glycerol]–H_2O (in the L_α phase) show a large differential line broadening for C9 but a reduced effect for C10, consistent with the results of ^2H NMR of specifically ^2H-labeled phospholipids [Seelig, J., & Waespe-Šaracević, N. (1978) *Biochemistry 17*, 3310–3315]. Thus, both desipramine and imipramine and the glycolipid show magic-angle orientation effects which reduce the amount of differential line broadening observed with other C–H vector orientations. In monogalactosyl diglyceride [primarily 1,2-di[(9Z,12Z,15Z)-octadeca-9,12,15-trienoyl]-3-β-D-galactopyranosyl-*sn*-glycerol]–H_2O (in the H_{II} phase), similar differential line-broadening effects are found for C9,10; C12,13; and C15,16, upon cooling. The resonances of C9 and C10 broaden before those of C12,13, which in turn broaden before those of C15,16. C10 is narrower than C9, and has less differential broadening, consistent with a magic-angle orientation. Computer simulations of the low-temperature spectra of monogalactosyl diglyceride (at −30 °C) using chemical shift and intensity values from the high-temperature spectra permit determination of individual component line widths, even in spectra showing limited overall resolution. Each of the six olefinic carbons (in the mainly linolenoyl chains) exhibits differential line broadening. The good qualitative agreement between ^{13}C and ^2H NMR results suggests that useful orientational (^2H NMR like) information can be deduced from natural-abundance ^{13}C NMR spectra of a variety of mobile solids.

Nuclear magnetic resonance (NMR) spectroscopy has been used for a number of years to study lipid membrane structure, using a variety of natural-abundance as well as isotopic-labeling techniques (Veksli et al., 1969; Seelig, 1977; Rothgeb & Oldfield, 1981; Sefcik et al., 1983; Xu & Cafiso, 1986). Deuterium NMR has been particularly successful (Seelig, 1977; Renou et al., 1989; Vist & Davis, 1990), and more recently ^1H and ^{13}C "magic-angle" sample-spinning (MAS) methods have shown some promise (Forbes et al., 1988). Ideally, a membrane lipid probe would be capable of giving rate information, as well as order parameter information, without use of isotopic enrichment, and at high resolution. At present, however, spectroscopists in general use either low-resolution (^2H) NMR of isotopically labeled lipids or ^1H or ^{13}C MAS, the latter usually taken under conditions of full ^1H decoupling (Sefcik et al., 1983; Forbes et al., 1988). In this paper, we show that *coupled* ^{13}C MAS NMR appears to have potential for combining orientational as well as the more conventional dynamic structural information about the lipid components of membranes that is qualitatively (at present) similar to that deduced by ^2H NMR but does not involve isotopic labeling.

The method used involves conventional magic-angle sample spinning of lipid systems, but without cross-polarization (Pines et al., 1972) or proton decoupling. "J-coupled" spectra are obtained, and in situations where the chemical shift anisotropies are relatively large, interference effects between the