Volume 5 Number 5 May 1978 Nucleic Acids Research

Detection of 5-methylcytosine in DNA sequences

Haruo Ohmori, Jun-ichi Tomizawa and Allan M. Maxam*

Laboratory of Molecular Biology, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, MD 20014, USA

Received 13 March 1978

ABSTRACT

Col E1 DNA has methylated cytosine in the sequence 5'-CC*(A/T)GG-3' and methylated adenine in the sequence 5'-GA*TC-3' at the positions indicated by asterisks(*). When the Maxam-Gilbert DNA sequencing method is applied to this DNA, the methylated cytosine (5-methylcytosine) is found to be less reactive to hydrazine than are cytosine and thymine, so that a band corresponding to that base does not appear in the pyrimidine cleavage patterns. The existence of the methylated cytosine can be confirmed by analyzing the complementary strand or unmethylated DNA. In contrast, the methylated adenine (probably N^6-methyladenine) cannot be distinguished from adenine with standard conditions for cleavage at adenine.

INTRODUCTION

In the method of Maxam and Gilbert for determination of the nucleotide sequence of DNA[1], a DNA fragment radioactively labeled at its 5' or 3' end is partially cleaved at one or two of the four normal DNA bases in each of four sets of chemical reactions. In these reactions dimethylsulfate methylates purine rings or hydrazine splits pyrimidine rings, and then displacement and elimination reactions break the DNA. The products are DNA fragments with ^{32}P at one end and the nucleotide immediately preceding the reacted base at the other; when resolved by size on gels and autoradiographed, these products give rise to four series of bands. Since the base-specificities of the four reactions which produced these bands are known, the nucleotide sequence can be deduced from the band patterns.

In this paper we describe an interruption encountered in these patterns where DNA contains 5-methylcytosine. This modified base does not react appreciably with hydrazine in the limited reactions employed for sequencing unmodified pyrimidines, and the DNA then does not break where it occurs when treated with piperidine. The visible result is a gap in the band pattern produced by partial hydrazine/piperidine cleavage of end-labeled DNA. This gap and guanine at the corresponding position in the sequence of

the complementary strand can determine the location of 5-methylcytosine in the DNA sequence.

MATERIALS AND METHODS

1. Preparation of DNA.

Methylated Col E1 DNA was prepared from Escherichia coli C600 (Col E1) as described[2]. Unmethylated Col E1 DNA was prepared from a strain deficient in DNA methylations, GM48 (dam-3 dcm-6[3]; obtained from J. L. Rosner), which had been transformed with Col E1 DNA.

2. Preparation of restriction enzymes.

EcoRII endonuclease was prepared from RY23 (obtained from R. J. Roberts), following the method for EcoRI endonuclease[4]. HaeIII endonuclease was prepared as described[5]. MboI was obtained from New England BioLabs.

3. Determination of nucleotide sequence.

In this paper we describe results derived from the HaeIII-N fragment[6] of Col E1 DNA. This DNA fragment (43 base-pairs in length[6]) was prepared by digesting Col E1 DNA with HaeIII endonuclease and isolating it from others by electrophoresis on a polyacrylamide gel. About 3 ug of HaeIII-N fragment was end-labeled with γ-^{32}P-ATP ($\sim$ 1800 Ci/mmol, ICN) and polynucleotide kinase. ^{32}P-labeled complementary strands were separated by alkali denaturation and electrophoresis on an 8% polyacrylamide slab gel (20 x 40 x 0.3 cm) at 200 volts for 14 hr. Aliquots of F (fast moving) and S (slow moving) strands were degraded with four different sets of reactions[1]. Samples A and G were methylated with 0.07 M dimethylsulfate for 50 and 30 min, respectively, at 20°C. After reacting sample A with 0.1 N HCl at 0°C for 60 min and heating sample G in 0.01 M sodium phosphate (pH 7.0) at 90°C for 15 min, polynucleotides were cleaved by heating at 90°C for 30 min in 0.1 N NaOH. Sample C in 1.0 M NaCl and sample T in H_2O were reacted with 18 M hydrazine at 20°C for 50 and 40 min, respectively, and then with 0.5 M piperidine at 90°C for 60 min. The products were fractionated by electrophoresis on a 30 x 40 x 0.15 cm slab gel polymerized from 20% acrylamide, 0.67% bis acrylamide, 7 M urea, 0.05 M Tris-borate (pH 8.3) and 0.001 M EDTA[1].

RESULTS AND DISCUSSION

Fig. 1 shows sequencing patterns on each strand of the HaeIII-N fragment of Col E1 DNA, which was prepared from the strain C600 (Col E1). As

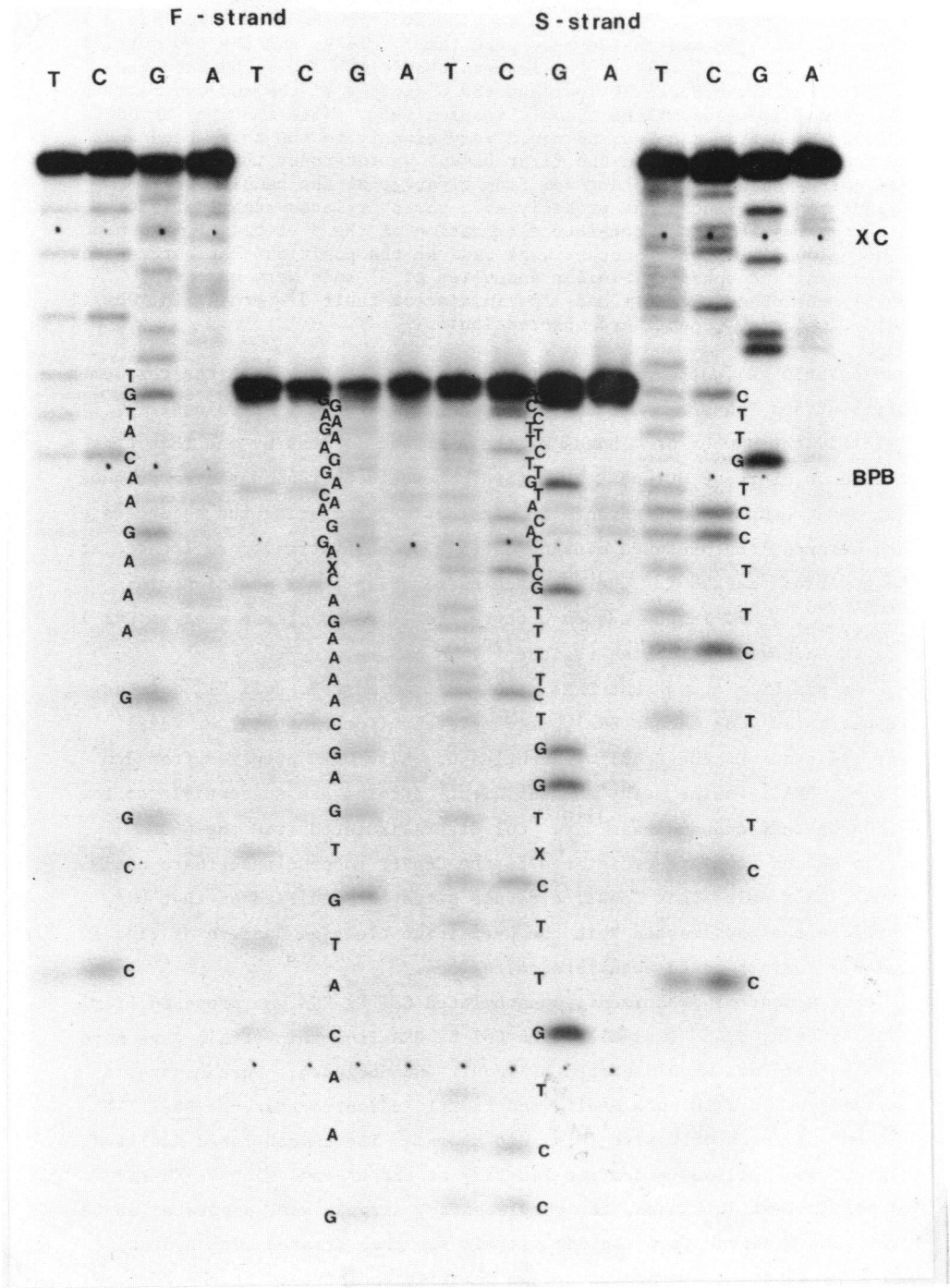
F - strand
S - strand
T C G A T C G A T C G A T C G A
XC
BPB

Fig. 1. Radioautograph for determination of the sequence of the Col E1 HaeIII-N fragment. Both of the complementary strands were run simultaneously. The two inside sets were run for 18 hr and the two outside ones for 6 hr at 800 volts. The letters beside the bands indicate the attacked bases. BPB and XC indicate the positions of the two dye markers bromophenol blue and xylene cyanol, respectively. Note that the second smallest band from each strand moved very closely to the third band and appeared more rounded than the first band. We interpret that the first band corresponds to ^{32}pCp derived from cleavage at the penultimate C residue and the second one probably to a phosphorylated sugar-derivative, which may result from incomplete elimination of the 5'-terminal phosphate. We occasionally observe another weak band at the position that corresponds to inorganic phosphate. Similar anomalies at 5'-ends were observed usually when DNA fragments had a pyrimidine at their 5'-termini (H. Ohmori and J. Tomizawa, unpublished observation).

seen on this gel pattern, the F strand has six C bands while the complementary S strand has seven G bands, including two at the 3'-terminus. Similarly the number of C bands of the S strand is less by one than that of G bands of the F strand. Therefore, it would be reasonable to presume that one C band in both strands is missing. By comparing the sequences of both strands, the presumed missing C bands on both strands are located at the positions marked X. The sequence surrounding the X positions is 5'-CXAGG-3' / 3'-GGTXC-5' , and is the sequence where EcoRII endonuclease cleaves DNA if X is an unmodified cytosine residue[7,8].

It has been shown that DNAs prepared from most E. coli K12 strains are partially (for lambda DNA), or completely (for bacterial or fdRFI DNA) resistant to the EcoRII endonuclease, while DNAs prepared from dcm$^-$ (=mec$^-$, DNA cytosine methylase deficient) strains are susceptible to the action of that endonuclease[9,10]. Col E1 DNA isolated from the C600 strain showed complete resistance to the EcoRII endonuclease (data not shown), indicating that EcoRII cleavage sites are modified on that DNA. Therefore the gaps marked X in the pyrimidine cleavage pattern of Fig. 1 probably correspond to methylated cytosines.

For a control experiment, unmethylated Col E1 DNA was prepared from a dcm$^-$ strain, GM48[3] (Col E1). The Col E1 DNA from this strain gave more than 13 fragments when digested by EcoRII endonuclease. Furthermore, a mixed digestion with both HaeIII and EcoRII indicated that the HaeIII-N fragment has an EcoRII site (data not shown). The unmethylated HaeIII-N fragment was purified and after labeling at the 5'-ends with γ-^{32}P-ATP and polynucleotide kinase, the complementary strands were separated as before. The S strand from the dcm$^-$ strain was then treated with hydrazine in the presence or absence of sodium chloride and cleaved at the reacted pyrimidines by piperidine. The gel pattern obtained was compared with one

obtained with the S strand from the dcm$^+$ strain. Fig. 2 demonstrates that the C band missing in the gel pattern from the methylated DNA (sample A) is present in that from the unmethylated DNA (sample B).

The reaction of hydrazine and pyrimidine nucleotides is not well characterized, but ribothymidylic acid appears to react at least 10-fold slower than uridylic acid[11]. Because these pyrimidines differ only by a methyl group at the 5 position and reactive positions in thymine and cytosine appear to be the same[12,13], by analogy, 5-methylcytosine might be expected to react slower than cytosine with hydrazine. Such a difference in the reaction rates might explain why the limited hydrazine sequencing reaction fails to cleave at 5-methylcytosine.

The above results confirm the previous reports[9,10] that the DNA isolated from wild type E. coli strains are methylated at the second C base of the 5'-CC(A/T)GG-3' sequence. We have determined the sequence of more

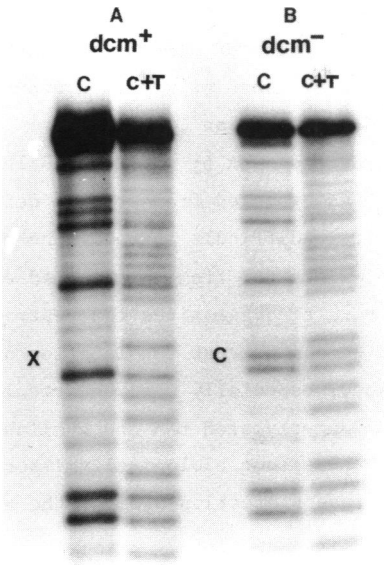

Fig. 2. Comparison of pyrimidine sequencing patterns around a CCTGG sequence in methylated (A) and unmethylated (B) DNAs.
HaeIII-N fragment was prepared from Col E1 DNA propagated in a strain which has (A, dcm$^+$, C600) or has not (B, dcm$^-$, GM48) the host DNA cytosine methylase. The fragment was labeled with ^{32}P at the 5'-end, and the strands were separated. The two S-strands were then partially cleaved at cytosine (C) and cytosine plus thymine (C+T) in standard hydrazine/piperidine reactions[1], electrophoresed for 14 hours, and autoradiographed. A portion of the pyrimidine cleavage patterns obtained is shown. C indicates an expected cytosine band, while X indicates the gap.

than 2,000 bases of Col E1 DNA (H. Ohmori and J. Tomizawa, unpublished data), and have not found the missing C band in sequences other than 5'-CC(A/T)GG-3'. This result supports the proposal[14] that the DNA cytosine methylase of *dcm*$^+$ *E. coli* K12 strains has the same sequence specificity as the *EcoRII* modification methylase[8].

Another type of DNA methylation in *E. coli* is known to occur at adenine, giving rise to N^6-methyladenine. It has been shown that *MboI* and *DpnII* endonucleases, which recognize the 5'-GATC-3' sequence, do not digest DNAs prepared from most *E. coli* strains, since the adenine base in that sequence is N^6-methylated[15-17]. The Col E1 DNA from C600 was completely resistant to *MboI*, but the DNA from GM48, which is also deficient in adenine methylation (*dam*-3)[3], was cleaved by *MboI* to give more than 13 fragments (data not shown). However, N^6-methyladenine could not be distinguished from unmethylated adenine on the gel pattern with standard sequencing reactions[1] for cleavage at adenine (data not shown). This result shows that the pre-existing methyl group at the N^6 position of adenine neither affects the reaction of dimethylsulfate at the N^3 position of the base nor the following cleavage by sodium hydroxide. This is consistent with the previous expectation[1].

For the purpose of sequencing DNA by the Maxam-Gilbert method it would be advisable to use DNAs prepared from strains deficient in DNA methylation, both because it is difficult to recognize the missing C bands, if 5-methylcytosine exists far from the labeled end, and because unmethylated DNA provides more cleavage sites for restriction enzymes.

The results described here show that the location of 5-methylcytosine in DNA can be determined, especially when the results on the complementary strand or the unmethylated DNA are available. 5-methylcytosine in DNA is known to have some biologically important function in at least one case[18], other than protection against the action of restriction enzymes.

ACKNOWLEDGEMENT

We thank J. L. Rosner and R. J. Roberts for providing us bacterial strains, and R. E. Bird and J. McGhee for critical reading of the manuscript.

*Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, USA

REFERENCES

1 Maxam,A.M. and Gilbert,W. (1977) Proc.Nat.Acad.Sci.USA 74, 560-564

2 Clewell,D.B. and Helinski,D.R. (1969) Proc.Nat.Acad.Sci.USA 67, 1159-1166
3 Marinus,M.G. (1973) Mol.Gen.Genet. 127, 47-55
4 Green,P.J., Betlach,M.C., Goodman,H.M. and Boyer,H.W. (1974) in DNA replication, Wickner,R.W., ed., pp. 87-111. Dekker Inc., New York
5 Roberts,R.J., Meyer,P.A., Morrison,A. and Murray,K. (1976) J.Mol.Biol. 103, 199-208
6 Tomizawa,J., Ohmori,H. and Bird,R.E. (1977) Proc.Nat.Acad.Sci.USA 74, 1865-1869
7 Boyer,H.W., Chow,L.T., Dugaiczyk,A., Hedgpeth,J. and Goodman,H.W. (1973) Nature 244, 40-43
8 Bigger,C.H., Murray,K. and Murray,N.E. (1973) Nature 244, 7-10
9 Hughes,S.G. and Hattman,S. (1975) J.Mol.Biol. 98, 645-647
10 Schlagman,S., Hattman,S., May,M.S. and Berger,L. (1976) J.Bacteriol. 126, 990-996
11 Verword,D.W. and Zillig, (1963) Biochim.Biophys.Acta 68, 484-486
12 Temperili,A., Turler,H., Rust,P., Danon,A. and Chargaff,E. (1964) Biochim.Biophys.Acta 91, 462-476
13 Cashmore,A.R. and Petersen,G.B. (1969) Biochim.Biophys.Acta 174, 591-603
14 May,M.S. and Hattman,S. (1975) J.Bacteriol. 122, 129-138
15 Pirrota,V. (1976) Nucleic Acids Res. 3, 1747-1770
16 Lack,S. and Greenberg,B. (1977) J.Mol.Biol. 114, 153-168
17 Gelinas,R.E., Meyers,P.A. and Roberts,R.J. (1977) J.Mol.Biol. 114, 169-179
18 Friedman,F. and Razin,A. (1976) Nucleic Acids Res. 3, 2665-2675