Cloning, Sequencing, in Vivo Promoter Mapping, and Expression in Escherichia coli of the Gene for the HhaI Methyltransferase*

(Received for publication, May 12, 1986)

Micaela Caserta‡§, Wolfgang Zacharias‡, Donald Nwankwo¶, Geoffrey G. Wilson¶, and Robert D. Wells‡

From the Department of Biochemistry, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, Alabama 35294 and ¶ New England Biolabs, Beverly, Massachusetts 01915

A 1476-base pair DNA fragment from Haemophilus haemolyticus containing the HhaI methyltransferase gene was isolated from a cell library and cloned into pBR322. The nucleotide sequence of this fragment was determined. The structural gene is 981 nucleotides in length coding for a protein of 327 amino acids (M, 37,000). The translational start signal (ATG) is preceded by the putative ribosome-binding site (TAAG). Recombinant plasmids containing the 1476-basepair fragment are completely methylated when isolated from Escherichia coli, as judged by their insusceptibility to the HhaI restriction endonuclease. However, the presence of an active HhaI methylase gene in certain E. coli strains results in a very poor yield of transformants and/or in vivo-originated deletions due to the Rgl functions of these hosts. The in vivo transcription initiation sites have been identified by S1 protection and primer-extension experiments using specific probes with total RNA prepared from E. coli cells (HB101 or RR1) which tolerate the expression of MhhaI.

Type II restriction-modification systems consist of two genes, one encoding an endodeoxyribonuclease and the other a methyltransferase. Both enzymes recognize the same specific DNA sequence. The methylase modifies host DNA by the addition of a methyl group at adenines or cytosines in the target site, thus allowing the related endonuclease to selectively cleave unmodified foreign DNA (1–5).

Since their discovery, type II restriction-modification enzymes have been employed for common genetic manipulations (1, 6, 7). Recently, however, interest has developed regarding the possibility of analyzing specific protein–DNA interactions, since the two different activities recognize the same DNA sequence (8). In addition, these enzymes were useful for probing chromatin organization (in terms of methylation-demethylation patterns of gene-flanking regions and accessibility to nuclease (9, 10)) and, more specifically, DNA conformational alterations at certain sequences. For example, cruciform formation in supercoiled molecules can be monitored by the inhibition of endonuclease cleavage at sites contained in the inverted repeats (11, 12). Also, the effect of left-handed Z-DNA conformations can be analyzed; structural alterations near or at the recognition sites can influence the ability of restriction-modification enzymes to utilize them as substrates (13–16).

In this paper, we report the DNA sequence of the gene coding for the HhaI methylase which recognizes the tetranucleotide GCAG and modifies the internal cytosine at position 5 (17). Analysis of the modification pattern of recombinant molecules containing the MHhaI gene demonstrates its expression in Escherichia coli. The transcription initiation sites and the putative ribosome-binding site and promoter sequences were identified. In addition, we find that certain E. coli strains are methylase-sensitive since they are poorly transformed by MHhaI gene-containing plasmids, thus indicating a possible deleterious effect of C-methylation.

Materials and Methods

Plasmids—Plasmid pNW2801 contains the HhaI methylase gene on a 1476-bp fragment cloned into the HindIII site of pBR322 (see below and Fig. 1). pRW951 (not shown) contains the 1476-bp insert cloned into the HindIII site of pBR71 (18), in the opposite orientation relative to pNW2801. pRW959 (Fig. 1) harbors the same Haemophilus haemolyticus fragment in the HindIII site of pSP569 (19) (Promega Biotech). pRW962 (Fig. 1), a derivative of pRW959 obtained as a spontaneous deletion during the course of a cloning procedure, lacked the polynucleotide region of pSP65 from the BamHI to the HindIII site and the first 382 bp of the H. haemolyticus insert. DNA plasmid isolation, after chloroformic amplification, was as described previously (20).

Strains—H. haemolyticus was obtained from the New England Biolabs strain collection; it is also available from the ATCC, strain No. 10014. The E. coli strains, obtained from personal collections, were as follows: R1 was the original recipient of pNW2801; HB101 was used as a host for all the other plasmid DNAs; TG1 (a derivative of JM101, kindly provided by J. Engler, this department) harbored all the M13 recombinant phages; DH1 and W3110 were obtained from A. Shatzman and M. Rosenberg (Smith Kline & French, Inc., Philadelphia). All enzymes were purchased from Boehringer, New England Biolabs, or Bethesda Research Laboratories. Assay conditions were as recommended by the manufacturers.

Purification of H. haemolyticus DNA—10 g of freshly grown H. haemolyticus cells were suspended in 20 ml of 25% sucrose, 50 mM Tris, pH 8.0. 10 ml of 0.25 M EDTA, pH 8.0, and 6 ml of 10 mg/ml lysozyme in 0.25 M Tris, pH 8.0, were added. After 2 h on ice, 24 ml of 1% Triton X-100 in 50 mM Tris, 6% EDTA, pH 8.0, and 5 ml of 10% sodium dodecyl sulfate were mixed in to achieve cell lysis. 70 ml of equilibrated phenol were added and the solution was emulsified by shaking. 70 ml of chloroform were added and the solution was
again unmixed. The mixture was centrifuged at 10,000 rpm for 30 min and the upper layer was re-extracted with equal volumes of phenol and chloroform. The upper layer was dialyzed against four changes of 10 mM Tris, pH 8.0, 1 mM EDTA, and then digested with phenol and chloroform. The upper layer was dialyzed against four
0.55 volume of isopropyl alcohol was layered on top of the solution. The tubes were incubated at 1 h at 57 °C and heated for 15 min at 72 °C to stop the reactions, and 10 μl from each were analyzed by agarose gel electrophoresis. Tubes in which moderate, but incomplete, digestion had occurred were combined. 40 μl (4 μg) of the combined solution were mixed with 10 μl of HindIII-cleaved and dephosphorylated pBR322 (2 μg). After addition of ligase buffer plus 250 μl of T4 DNA ligase and incubation for 4 h at 16 °C, the mixture was transformed into 2 ml of E. coli RR1, made competent by calcium treatment (21). The transformed cells were grown to saturation in 50 ml of Luria broth, centrifuged, and resuspended in a volume of 2.5 ml, and 250-μl aliquots were plated onto Luria agar plates containing 50 μg/ml ampicillin. After incubation at 37 °C, the plates were flooded with 3 ml of 10 mM Tris, pH 7.5, 10 mM MgCl₂, and 0.1 mM S-adenosylmethionine were prepared, digested with HhaI, and retransformed into RR1. The 1476-bp fragment, purified from the pBR322 library was prepared by ligating HindIII fragments of H. haemolyticus DNA into pBR322 (29); it was propagated in E. coli RR1 to allow self-modification of molecules carrying the HhaI modification gene to occur, then the plasmids were purified, digested with HhaI, and rtransformed into RR1. Among the plasmids that survived, many were found to carry a common, 1.5-kilobase pair HindIII fragment. The fragment was judged to encode the HhaI modification gene by three criteria: (a) plasmids carrying the fragment were resistant to digestion by HhaI (GCGC) and HaeIII (RGCGYC), yet sensitive to digestion by FnuDII (CGCG), and Aha11 (GRCGYC), enzymes that recognize sim-
portantly, the labeling of DNA requires the use of a radioactive labeling method. The most common method for labeling DNA is the use of [α-32P]dATP and T4 DNA polynucleotide kinase following dephosphorylation. In addition, it is critical to ensure that the DNA is free of contaminating proteins or nucleases, which can interfere with the labeling process. This is typically achieved by performing a gel retardation assay to verify the absence of high molecular weight contaminants.

The labeling reaction is typically performed at room temperature for 2-4 h with [α-32P]dATP and T4 DNA polynucleotide kinase. The reaction mixture contains the DNA to be labeled, 10-20 mM Tris, 1 mM MgCl₂, 10 mM dithiothreitol, and 1 unit of T4 DNA polynucleotide kinase. After the reaction is complete, the DNA is purified using a column or agarose gel electrophoresis. The labeled DNA is then used for subsequent experiments, such as Northern blot analysis or radiolabeled in situ hybridization.

**RESULTS**

**Isolation of the HhaI Modification Gene**—Recombinants carrying the HhaI modification gene were isolated by selection as survivors from a library of plasmid clones that had been digested with the HhaI restriction endonuclease. A similar approach has been used to isolate clones of the BspRI (35), MspI (36), BsuRI (37), and DdeI (25) restriction endonucleases. The library was prepared by ligating HindIII fragments of H. haemolyticus DNA into pBR322 (39); it was propagated in E. coli RR1 to allow self-modification of molecules carrying the HhaI modification gene to occur, then the plasmids were purified, digested with HhaI, and rtransformed into RR1. Among the plasmids that survived, many were found to carry a common, 5-kilobase pair HindIII fragment. The fragment was judged to encode the HhaI modification gene by three criteria: (a) plasmids carrying the fragment were resistant to digestion by HhaI (GCGC), and BspRI (CCGG), (b) HhaI (GCGC), and Aha11 (GRCGYC), enzymes that recognize similar, but distinct sequences; (b) the chromosomal DNAs of the
cells harboring the plasmids were similarly resistant to HhaI and HaeII yet sensitive to the other enzymes; (c) extracts prepared from cells harboring the plasmids displayed HhaI modification activity in vitro. No restriction endonuclease activity was detected in cell extracts, suggesting that the fragment does not encode the HhaI restriction gene, at least in its entirety. A representative plasmid, pNW2801 (Fig. 1), that carried only the 1.5-kilobase pair fragment, was chosen for analysis.

**Sequence Determination**—The nucleotide sequence of the 2890-bp DNA fragment from pNW2801 (Fig. 1) containing the HhaI DNA methyltransferase gene was determined. The dideoxy-chain termination method (27-29) was employed as well as the Maxam-Gilbert chemical degradation technique (30), when required. More than 90% of the sequence was determined on both strands; portions of the fragment for which data were available from only one of the two strands were sequenced several times.

The sequence strategy is shown in Fig. 2; each arrow corresponds to sequence information derived from at least two independent experiments. The Sau3AI and RsaI sites are indicated since they were utilized to subclone overlapping segments of the HindIII fragment in the phage vectors M13mp10 and M13mp11 (24) (see "Materials and Methods"). A more detailed restriction map is presented in Fig. 1. In the first set of experiments, the M13 single-stranded universal primer was employed for all the subclones obtained. Since some of the DNA inserts expected from the random subcloning were missing, an entire strand of the HindIII fragment was cloned in the phage vector M13mpWB2348 (25). This particular phage is less deletion-prone than the other M13 phages of the mp series, thus allowing larger inserts to be stably propagated.

Six oligonucleotides, ranging in size from 15 to 17 bases, were synthesized and used as specific primers; four were complementary to internal regions of the top strand of the HindIII fragment and the other two to part of the bottom strand of the 635-bp RsaI segment. Dotted lines in Fig. 2 indicate the portion of data obtained by this approach. The chemical degradation method was employed essentially to cover regions of the sequence for which only one strand was available. In addition, this technique was necessary, coupled with the use of a 20% denaturing acrylamide gel, to resolve a region complicated by secondary structure formation. Indeed, the sequence data revealed the presence of a perfect inverted repeat with a 7-bp stem and with 3 nucleotides in the potential loops, situated 5 bases downstream from the termination codon of the HhaI methylase gene (Fig. 3).

**MHhaI Structural Gene and Ribosome-binding Site**—The 1476-bp sequence was searched for all possible reading frames in both strands. Only one of the six frames has a large open region large enough to code for the MHhaI gene. Since some of the genes for restriction-modification systems are organized in pairs with a small untranslatable region in between (37, 40-46), we looked for at least part of a second open reading frame, which could code for the HhaI restriction enzyme. No other suitable open reading frame is present upstream from the methylase gene. It is still possible that the endonuclease gene starts somewhere downstream from the methylase termination signal, although the short segment of *H. haemolyticus* DNA at the 3' end of the gene does not allow any conclusions.

The nucleotide and amino acid sequences corresponding to the HhaI methylase are presented in Fig. 3. The first A of the left HindIII site is position 1 on the sequence. Starting from the ATG at position 437 and ending at position 1418, the gene extends for 981 nucleotides, thus coding for a protein of 327 amino acids. The molecular weight deduced from the predicted amino acid composition is 37,002, in excellent agreement with the size of the purified protein (37,000, as isolated from *E. coli* transformed with pNW2801 (47)). Furthermore, a Pro-Cys doublet, which rarely occurs in proteins, found at position 674-679 on the DNA sequence, is proposed to form...
Sequence and Expression of HhaI Methylase

The putative ribosome-binding site TAAG, indicated in Fig. 3, is 5 nucleotides away from the ATG. Although it appears to be a weak site for translation in E. coli, we showed that pNW2801 does indeed express a methylase activity when E. coli RR1 and HB101 are the recipient strains (see below).

Transcription Signals and Expression in E. coli—MHhaI activity in E. coli was detected by testing purified DNA for insusceptibility to HhaI endonuclease cleavage. pNW2801 was undigested following a 2-h incubation with the HhaI restriction enzyme, whereas an unmethylated internal control DNA (pRW451 (18)) was completely digested (data not shown). pRW951, which is the HindIII insert was inverted relative to pNW2801 (see "Materials and Methods"), shares the same full methylation pattern; this in vivo methylase activity in either orientation was the first indication for the presence of a MHhaI promoter on the 1476-bp fragment. Moreover, this fragment retains methylase activity when cloned in either orientation in both vectors pSP65 and pACYC184. Also, the replicative form of M13mW3248 with the 1476-bp insert in one orientation was found to be methylated.

Another method for testing in vivo expression of MHhaI was used. Purified pRW951 was reacted in vitro with [methyl-3H]AdoMet and MHhaI as described previously (14). No significant incorporation (less than 0.2 sites per molecule) was detected relative to an unmethylated control DNA, thus indicating that all the HhaI sites were indeed modified in vivo (data not shown).

In order to identify sequences required to promote the expression of the MHhaI gene, we carried out primer-extension experiments to map at the nucleotide level the 5' end of the mRNA produced in vivo. For this purpose, two of the six synthetic oligonucleotides utilized for sequencing were selected; the primers, a 17-mer and a 15-mer, are complementary to portions of the sequence located 12 and 229 bases from the ATG (Fig. 3). Timed reaction mixtures were selected; the primers, a 17-mer and a 15-mer, are complementary to portions of the sequence located 12 and 229 bases from the ATG (Fig. 3).

To confirm this result, we also performed S1 nuclease protection experiments. Fig. 5 shows the results obtained using a probe extending from the MadII site (position 606) to the Real site (position 404) and the same total RNA preparation as above. These experiments confirmed that transcription initiates 21-22 nucleotides upstream from the ATG and ruled out the possibility of a premature termination of reverse transcriptase.

Two additional lines of evidence support the finding that the active site of the MHhaI protein (47).

![Fig. 3. Nucleotide and amino acid sequence of the 1476-bp HindIII fragment containing the HhaI methylase gene. The ATG and the TAA are boxed. The ribosome-binding site is indicated by dots. The transcription start sites are shown by two arrowheads. The $^{*}$ indicates the active site of the MHhaI protein (47).](image-url)
Fig. 4. Primer extension analyses for mapping the transcription start sites for the MHhaI gene. The reverse transcriptase products were obtained with the oligonucleotide proximal to the ATG (lanes 1 and 2) and with the distal one (lanes 3 and 4). The two oligomers are shown as filled boxes in Fig. 2. Hybridization temperatures were 30 °C (lanes 1 and 3) and 37 °C (lanes 2 and 4). Dideoxy-sequencing markers were obtained with the corresponding primers. The numbers refer to the positions on the sequence (Fig. 3).

These sequences are indeed required to promote the transcription of the HhaI methylase gene. First, during the course of a subcloning procedure using pRW959 (Fig. 1), we found a spontaneous deletion which lacks most of the sequence preceding the MHhaI ATG but still retains methylase expression in E. coli. This molecule, pRW962 (Fig. 1), was further characterized through restriction analyses and sequencing around the relevant region. The deletion extends from the 3' end of the SmaI site in the poly linker region to position 382 in the MHhaI gene insert (solid triangle in Fig. 3; see also “Materials and Methods”). The sequence of the new -35 element is GGGATT. Therefore, a region of 53 nucleotides upstream from the ATG contains all the necessary information to promote MHhaI expression. Second, in vitro filter binding experiments, using E. coli RNA polymerase and different sets of restriction fragments derived from the 1476-bp HindIII insert, revealed preferential retention of the 130-bp TaqI-RsaI segment (Fig. 1), located just upstream from the ATG of the MHhaI gene and containing the −35 and −10 regions (data not shown).

Base Composition and Codon Usage—An analysis of the base composition for the sequence shown in Fig. 3 reveals a high A+T content (65%), directly influencing the gene codon frequency. A and U are usually preferred bases in the third position of synonymous codons and, when there is a choice, also in the first position. The sequences of the only other two genes from the same organism which were reported, the HhII restriction endonuclease and methyltransferase, also have a high A+T content of 68 and 65%, respectively (43). However, if this feature is not a general characteristic for H. haemolyticus DNA, a regulatory mechanism at the level of translation could be implicated, as suggested in the case of the EcoRI restriction-modification system (42).

Comparison of Protein Sequence with Other Methylases—The predicted amino acid sequence of MHhaI was compared with those of the SPR, BspRI, and BsuRI methylases (37, 49) which are also C-methylases with exclusively GC base pairs in their recognition sequences. The MALIGN program was used to search for segments of identity and optimal alignment among these sequences, taking two residues as the minimum length of identity. The MHhaI protein sequence could thus
be aligned to 27.2% to MSPR, 18.7% to MBspRI, and 21.4% to MBsuRI, whereas the randomly scrambled sequences of MHhaI and MSPR gave an optimal alignment of only 14.4% of MHhaI. With the DFASTP program it was found that MHhaI shares identity with MSPR to 38.9% over a region of 190 residues, with MBspRI to 26.4% over 178 residues, and with MBsuRI to 22.7% over 286 residues. When conservative amino acid replacements were included in this comparison in addition to identical residues, the homology of MHhaI increased to 79.5, 74.2, and 66.8%, respectively.

Several regions are strikingly similar in all four protein sequences. 4 out of 6 residues are identical in a region containing the Pro-Cys unit, which is proposed to form the active site of MHhaI (47). In addition, a region of 4 out of 4 and 7 out of 14 amino acids, located at residues 118-121 and 151-164 of MHhaI, are identical. Altogether, a total of 15 clustered and 10 single invariants were found to be present in all four protein sequences.

**HhaI Methylase Sensitivity of Certain E. coli Strains—** Attempts to transform pNW2801 into a number of laboratory strains of E. coli met with unexpected difficulty. Some strains, notably RR1, MC1061, and K802, were permissive for transformation and accepted pNW2801 at the same frequency that they accepted pBR322 (transformation efficiency = 10^6 transformants/µg of plasmid DNA). Other strains such as DH1, M192, C600, Hfr H, W3110, and wild-type K12, were transformed very poorly by pNW2801 (efficiency <10^4/µg), or other plasmids containing a functional MHhaI gene, even though they accepted pBR322 with normal efficiency. Conversely, plasmids carrying artificially created or spontaneously occurring deletions that affect MHhaI expression could transform those strains with an efficiency equivalent to pBR322 used as a control. E. coli strains HB101 and RR1 were found to tolerate HhaI methylase activity, although the growth rate of the transformants was severalfold slower compared to pBR322 containing cells. TG1 cells were transformed by pNW2801, but the morphology of the colonies appeared to be quite different compared to the control (irregular versus rounded shape).

The “sensitive” behavior of these strains is in agreement with observations reported by other groups (37, 38, 40) working with plasmids which code for cytosine methyletransferases. It has been analyzed in detail using a variety of cloned modification genes and using pBR322 modified in vitro with purified modification enzymes. The analysis indicates that inefficient transformation is the result of restriction of methylation-containing DNA by the E. coli Rgl functions: inefficient transformation occurs only when the transforming DNA is modified at cytosine residues and the bacterial strain is Rgl-proficient (38).

**DISCUSSION**

The DNA sequence of the HhaI methyltransferase gene reveals a coding region of 981 nucleotides which predicts a protein of 327 amino acids (M, 37,002) in excellent agreement with the size of the protein purified from E. coli (M, 37,000 (47)). Each of the restriction-modification systems that have recently been cloned has a characteristic organization of the genes coding for the two related proteins. For the EcoRI and BsuRI systems, both enzymes are coded by the same DNA strand and the endonuclease gene precedes the methylase gene (37, 41, 42). For HhaII and PaeR7, the two genes are colinear, but the methylase is located upstream from the endonuclease (43, 46). However, the PovII, PstI, and EcoRV systems consist of two divergently arranged coding regions (40, 44, 45). The nucleotide sequence flanking the HhaI methylase structural gene does not reveal any other open reading frame suitable for coding for the HhaI restriction endonuclease. The sequence of the 436 nucleotides preceding the MHhaI gene shows several translation termination signals in all possible frames, whereas the region separating the end of the gene from the end of the cloned fragment contains no start codons.

Little is known regarding control mechanisms for the expression of the two related genes inside the cell. A colinear arrangement of the coding regions, with the methylease upstream, appears to be the most simple way to ensure a delay in the endonuclease synthesis. Alternatively, a difference in the strength of promoter or ribosome-binding site could regulate the transcription and/or translation efficiency (37, 44). Folding of the mRNA in a particular secondary structure could also affect protein synthesis (45). Compartmentalization, differential ion requirements, and stability of the protein itself could also be involved in controlling enzymatic activity. In this regard we have observed that E. coli HB101, carrying an active MHhaI gene on pNW2801, contains fully methylated plasmid DNA when grown to saturation, even at higher temperatures (40 and 43°C).

We were interested in defining the sequences promoting transcription and translation of the MHhaI gene in order to obtain regulated expression of this enzyme in E. coli under the control of a different promoter. The nucleotide sequence determination, in combination with transcription mapping experiments, was essential in identifying those regions of interest. We localized the coding region and the orientation of the gene on the basis of the largest open reading frame. The first ATG of this open reading frame appears at position 437 on the HindIII fragment; the putative ribosome binding site (TAAG), is situated 5 nucleotides upstream. An identical sequence exists 7 nucleotides upstream from the start codon of the MHhII gene.

We have searched for other potential fMet codons of the MHhaI gene. A GTG and an ATG triplet are present at position 308 and 381, respectively, in the same open reading frame. We believe that neither of them is the true start codon since the size of the protein would be too large or too small, respectively, compared to the established M, (47), and the transcriptional start sites at positions 415 and 416 automatically rule out the GTG triplet.

The end of the gene is marked by a TAA codon followed by a perfect inverted repeat 7 nucleotides further downstream. This inverted repeat (AAGGGGATAGCCCCTT) predicts a hairpin with 7 nucleotides in the stem and 3 in the loop. A hairpin structure at the 3' end of some bacterial mRNAs has been suggested as a transcription termination signal for RNA polymerase to release the DNA coding strand (48).

The transcriptional start sites were located at positions 415 and 416 through S1 protection and primer extension experiments using total RNA isolated from E. coli RRI harboring pNW2801. The -10 and -35 consensus sequences are indicated in Fig. 3. We do not know whether any of these sequences are utilized by E. haemolyticus for the transcription of the MHhaI gene; nevertheless, they are recognized in E. coli. Moreover, in vitro filter binding experiments, using the E. coli RNA polymerase and pNW2801 digested with different restriction enzymes, have shown preferential retention of the 130 bp Real-TaqI fragment upstream from the mRNA start sites.

The -10 region (TACTGT) close to the Real site at position 404 shows little similarity to the usual Pribnow box found in E. coli promoters (TATATT). The sequence TGATT occurs 2H. O. Smith, personal communication.
around position −35 with a spacing of 17 nucleotides from the −10 region. Only part of this sequence (GATT) is still present in the −35 region (GGGATT) of the deleted plasmid pRW962 that, nevertheless, was isolated as a fully methylated DNA.

An interesting characteristic of the MHhal gene sequence is the absence of Hhal sites (GGCC), frequently found in most other DNAs. An obvious speculation would be to propose the existence of selective pressure during evolution to preserve the integrity of the gene and therefore its essential function in the presence of the related endonuclease. Alternatively, an overall high A+T content (65%) could be responsible for this feature.

The comparison of the predicted amino acid sequence of MHhal with other cytokine methylases revealed extensive homologies among these proteins which are derived from different systems but catalyze the same chemical modification reaction. Sequence homologies between the prokaryotic methylase of BsulI and the phage methylase of SP8 and the potential origin of their similarities have been discussed in detail elsewhere (49 and references therein).

Finally, we find that several E. coli strains do not tolerate expression of the MHhal gene; they are poorly transformed by plasmids carrying an active gene and most of the transformants have lost Hhal methylase activity following extensive deletions. Other groups have observed the same phenomenon in E. coli with other cytokine methylases (37, 40). A more detailed analysis indicates that the function responsible for this sensitive phenotype is located in the RglI box, presumably coding for an unidentified restriction system (38). The RglA and RglB functions were originally identified by plasmids carrying an active gene and most of the transvise deletions. Other groups have observed the same phenomenon in E. coli. It has been generally assumed that Rgl recognizes only 5- hydroxymethylcytosine in their DNA (52, 53). It has been generally assumed that Rgl recognizes only 5-hydroxymethylcytosine-containing DNA but this is evidently not so. In light of this observation, investigators might find it prudent to use only Rgl-deficient strains of E. coli, such as K802, for the primary cloning of prokaryotic and eukaryotic DNA that contains 5-methylcytosine. Moreover, it may be legitimate to speculate that the correlation between the methylation state of the DNA and gene expression, well known in eukaryotic systems, may also apply to prokaryotes.

Acknowledgments—We thank Drs. J. Engler, N. P. Higgins, and T. M. Townes for helpful advice and A. Shatzman and M. Rosenberg for gifts of plasmids and bacterial strains. We also thank Drs. J. C. Wu and D. V. Santi for performing the homology searches with the protein sequences.

REFERENCES
1. Modrich, P., and Roberts, R. J. (1983) in Nucleases (Linn, S. M., and Roberts, R. J., eds) pp. 191-154, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
2. Kruger, D. H., and Bickle, T. A. (1980) Microbial Rev. 47, 345-690
3. Smith, H. O., and Kelly, S. V. (1984) in DNA Methylation: Biochemistry and Biological Significance (Razin, A., Cedar, H., and Riggs, A. D., eds) pp. 30-71, Springer-Verlag, New York
4. Yuan, R., and Smith, H. O. (1984) in DNA Methylation: Biochemistry and Biological Significance (Razin, A., Cedar, H., and Riggs, A. D., eds) pp. 73-80, Springer-Verlag, New York
5. Marinos, M. G. (1984) in DNA Methylation: Biochemistry and Biological Significance (Razin, A., Cedar, H., and Riggs, A. D., eds) pp. 81-108, Springer-Verlag, New York
APPENDIX

Trial Presentation of the MHhal DNA Sequence with Annotations in a Computer-readable Format

The sequence of the 1476-bp HindIII fragment containing the MHhal gene is presented in Softstrip data strips for a Cauzin strip reader. This softstrip was prepared with a Macintosh Plus computer using the Microsoft Word (Version 1.05) and the Cauzin Systems Stripper program. The data in this strip can be read directly into Apple Macintosh, IBM PC, or Apple II desk-top computers that are equipped with a strip reader. Both the strip reader and the necessary software programs can be obtained at nominal cost from Cauzin Systems, Inc., 835 South Main Street, Waterbury, CT 06706.

This figure is being reproduced as an experiment to test the feasibility of publishing nucleotide sequences in computer-readable form. Written comments from readers are invited, and should be sent to the Journal editorial office, 9650 Rockville Pike, Bethesda, MD 20814.