Molecular cloning and characterization of the gene encoding the DNA methyltransferase, M.CviBIII, from *Chlorella* virus NC-1A

Kenneth E.Narva¹, Douglas L.Wendell², Merri P.Skrdla and James L.Van Etten

Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

Received July 28, 1987; Revised and Accepted October 29, 1987

ABSTRACT

The gene encoding the DNA methyltransferase, M.CviBIII, from *Chlorella* virus NC-1A was cloned and expressed in *E. coli* plasmid pUC8. Plasmid (pNC-1A.14.8) encoded M.CviBIII methylates adenine in TCGA sequences both *in vivo* in *E. coli* and *in vitro*. Transposon Tn5 mutagenesis localized the M.CviBIII functional domain to a 1.5 kbp region of pNC-1A.14.8 and also indicated that a virus promoter directs transcription of the gene in *E. coli*. The 2.1 kbp insert containing the M.CviBIII gene was sequenced and a single open reading frame of 1131 bp was identified within the domain determined by Tn5 mutagenesis. When the M.CviBIII gene was fused in-frame with the 19 amino-terminal codons of *lacZ* a 45 kD polypeptide was identified in maxicells as predicted by the DNA sequence. The M.CviBIII gene was not essential for virus replication since a virus M.CviBIII deletion mutant also replicated in *Chlorella*.

INTRODUCTION

We have isolated and partially characterized 30 large (150 to 190 nm in diameter), polyhedral, dsDNA containing (>300 kbp in size), plaque forming viruses which infect certain unicellular, eukaryotic *Chlorella*-like green algae (1,2). These viruses were placed into 11 classes on the basis of plaque size, reaction with antibody and the nature and abundance of methylated bases in their genomic DNAs (2). Each of the virus DNAs contains 5-methyldeoxycytosine (m⁵dC); the concentration of m⁵dC as a percentage of deoxycytosine, varies from 0.1% to 47.5%. In addition, 18 of the 30 virus DNAs also contain N⁶-methyldeoxyadenosine (m⁶dA); the concentration of m⁶dA, as a percentage of deoxyadenosine, varies from 1.45% to 37%. The finding that methylation was sequence specific led to the discovery that *Chlorella* sp. infected with at least some of these viruses contain DNA methyltransferases and DNA restriction endonucleases (3,4,5). For example, virus NC-1A

infected Chlorella NC64A cells contain a type II DNA restriction endonuclease, named CviBI, which recognizes the sequence GANTC and cleaves between G and A; CviBI does not cleave G^mANTC sequences (5). NC-1A infected Chlorella cells contain the cognate methyltransferase, named M.CviBI, which specifically methylates adenine in GANTC sequences and at least two other distinct methyltransferases M.CviBII and M.CviBIII. M.CviBII and M.CviBIII specifically methylate adenine in GATC and TCGA sequences, respectively (Xia, Narva and Van Etten, unpublished data). Their putative cognate restriction endonucleases, CviBII and CviBIII, have not been detected in extracts of NC-1A infected Chlorella NC64A.

We are trying to isolate and characterize the genes encoding some of these DNA methyltransferases and restriction endonucleases to understand their biological significance and to compare them to prokaryotic systems. This report describes the successful cloning, characterization and expression of the methyltransferase gene encoding M.CviBIII from virus NC-1A in Escherichia coli. To our knowledge this is the first virus encoded methyltransferase gene to be isolated from a virus infecting a eukaryotic organism.

MATERIALS AND METHODS

Viruses, vectors and host strains.

The growth of the host alga, Chlorella NC64A, on MBBM medium and the production and purification of NC-1A and the other viruses have been described (1,6).

E. coli strains LE392 (7), JM83 (8) and GM2163 (dam⁻, dcm⁻) (9) were used as hosts for plasmids pUC8, pUC9 and recombinants thereof (8). E. coli CSR 603 (10) was used for analysis of plasmid encoded proteins. Bacteriophage lambda 467 was the source of transposon Tn5 (11).

Enzymes. All enzymes for recombinant DNA experiments were purchased from Bethesda Research Laboratories or New England Biolabs and used according to the suppliers recommendations.

Cloning strategy. Virus NC-1A DNA was purified as described previously (12). A library of Sau3A partial digestion products of NC-1A DNA was prepared in E. coli plasmid pUC8 by standard procedures (7).

A clone containing the M.CviBIII gene, plasmid pNC-1A.14.8, was selected from the NC-1A DNA library by a

procedure originally suggested by Mann et al. (13) to clone bacterial methyltransferase genes. Briefly, the restriction endonucleases SalI (GTCGAC) and TaqI (TCGA) are inhibited by the presence of m⁶da in their recognition sequences. Since TCGA sequences are methylated by M.CviBIII, recombinant plasmids containing a M.CviBIII gene functional in E. coli should be resistant to SalI and TaqI. These plasmids were selected by treating the plasmid library with SalI before transforming E. coli.

Enzyme extracts and methyltransferase assays.

E. coli LE392 harboring pNC-1A.14.8 was grown to late log phase in LB broth (7) containing 100 µg/ml ampicillin and harvested by centrifugation (5,000 x g for 5 min at 4°C). The cells were resuspended in 10 mM Tris-HCl, pH 8.0, 10 mM 2-mercaptoethanol, 50 µg/ml phenylmethylsulfonyl fluoride and 1 mg/ml lysozyme for 30 min at room temperature. The suspension was frozen at -80°C, thawed at room temperature and then sonicated for three 30 second intervals at 100 watts with a Branson cell disrupter (model W1850). Cell debris was removed by centrifugation at 10,000 x g for 20 min at 4°C and glycerol was added to the supernatant to a final concentration of 10% (v/v).

Enzyme extracts (Fraction II) were prepared from virus infected Chlorella NC64A cells as described (5).

Methyltransferase activity was assayed either by measuring transfer of methyl-³H from S-adenosyl[methyl-³H]methionine to DNA or by its ability to protect substrate DNAs from digestion with appropriate restriction endonucleases (3).

Tn5 mutagenesis.

Tn5 mutagenesis of recombinant plasmids was performed as previously described (11). Sites of Tn5 insertion within the M.CviBIII gene were mapped with HindIII and HindIII plus EcoRI. Insertional inactivation of the M.CviBIII gene was determined by testing the TaqI sensitivity of individual plasmids containing unique Tn5 elements.

Nucleic acid electrophoresis and hybridization.

DNA was electrophoresed on agarose gels in either TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) or TPE buffer (80 mM Tris-phosphate, 8 mM EDTA) (7). Nucleic acids were stained with 0.5 µg/ml ethidium bromide and visualized by ultraviolet illumination. Standard protocols were used for

Southern transfer of DNA to nitrocellulose filters and for dot blot experiments (7). Radioactive DNA probes were prepared by nick translation using a kit from Bethesda Research Laboratories.

DNA sequencing.

Nested deletions of the DNA fragment containing the *M.CviBIII* gene in the single-stranded DNA phage vectors M13mp18

Figure 1. Resistance of pNC-1A.14.8 DNA, which encodes *M.CviBIII*, to *SalI* and *TaqI* digestion. Agarose gel (0.8%) electrophoresis of pNC-1A.14.8 digested with *HindIII*, *HindIII* plus *SalI*, *HindIII* plus *TaqI*, *HinfI* and *MboI* (lanes 1 through 5, respectively). Numbers on the left refer to sizes (kbp) of lambda *HindIII* fragments. pNC-1A.14.8 was propagated in *E. coli* GM2163 (*dam*⁻, *dcm*⁻).

or M13mp19 (14) were made using the Cyclone rapid deletion subcloning system marketed by International Biotechnologies Inc. M13 cloning, propagation and dideoxy sequencing were as described previously using deoxyadenosine 5'-[α - 35 S]-thiotriphosphate and buffer gradient gels (15,16,17).

In vivo analysis of plasmid encoded proteins.

E. coli strain CSR 603 (maxicells) was used to identify plasmid encoded proteins as previously described (10). Expression of M.CviBIII was increased for this purpose by placing the M.CviBIII gene under control of the *lac* promoter. A 1.2 kbp *Dra*I-*Hind*III fragment containing the gene was isolated from pNC-1A.14.8, the *Hind*III site was filled in with the Klenow fragment of DNA polymerase I and the fragment was ligated into the *Sma*I site of pUC19 to create an in-frame fusion of M.CviBIII to the 19 amino-terminal codons of *lacZ*. The resulting 3.9 kbp plasmid was designated pNC-1A.14-lac.

RESULTS

Isolation of the M.CviBIII gene.

Several transformants of two *Sal*I resistant clones were obtained when 10^5 transforming units (ca. 5 times the number of individual clones in the library) were digested with *Sal*I before transformation of *E. coli* LE392. One clone, pNC-1A.14.8, contained a 2.1 kbp insert of NC-1A DNA and was resistant to *Sal*I and *Taq*I when propagated in *E. coli* (Fig. 1, lanes 2 and 3). Plasmid pNC-1A.14.8 DNA was digested by *Hinf*I and *Mbo*I,

Table 1. In vivo DNA methylation by M.CviBIII in *E. coli* prevents methylation by M.TaqI.

DNA substrate	cpm of 3 H-methylated DNA ^a
pUC8	5185
pNC-1A.14.8	294
pNC-1A.14.8 (<i>dam</i> ⁻ , <i>dcm</i> ⁻) ^b	282
minus DNA	219
pUC8 (minus M.TaqI) ^c	199

^aValues represent the average of duplicate assays. Reaction mixtures (20 μ l) containing 2 μ Ci of S-adenosyl [methyl- 3 H] methionine, 1 μ g of DNA and 10 units of M.TaqI were incubated at 65°C for 1 hr.

^bpNC-1A.14.8 was propagated in *dam*⁻, *dcm*⁻ *E. coli* GM2163.

^cM.TaqI was omitted from the reaction.

Table 2. TaqI digestion of DNA destroys sequences (TCGA) methylated by M.CviBIII.

DNA Substrate	Methyltransferase ^a	³ H-cpm of methylated DNA ^b
<u>TaqI</u> - digested lambda	M. <u>CviBIII</u>	317
	<u>lacZ</u> - M. <u>CviBIII</u>	500
	M. <u>TaqI</u>	292
Undigested lambda	M. <u>CviBIII</u>	6,558
	<u>lacZ</u> - M. <u>CviBIII</u>	55,913
	M. <u>TaqI</u>	159,325
<u>HaeIII</u> - digested lambda	M. <u>CviBIII</u>	10,370
	<u>lacZ</u> - M. <u>CviBIII</u>	51,122

^aCrude extracts of M.CviBIII and lacZ - M.CviBIII were prepared from cell-free sonicates of dam⁻, dcm⁻ E. coli harboring pNC-1A.14.8 and pNC-1A.14.lac (an M.CviBIII-overproducing plasmid), respectively. Commercial M.TaqI was used as a control.

^bValues represent the average of two or three assays. Reaction mixtures (50 μ l) contained 2 μ Ci S-adenosyl [methyl-³H] methionine, 3 μ g DNA and either 10 μ l crude cell extract or 10 U commercial M.TaqI. M.CviBIII and lacZ-M.CviBIII values were normalized to equivalent amounts of protein.

which recognize the same base sequences as the NC-1A methyltransferases M.CviBI and M.CviBII, respectively (Fig. 1, lanes 4 and 5). These results indicate that pNC-1A.14.8 encodes only M.CviBIII and not M.CviBI or M.CviBII.

Additional experiments were conducted to confirm the site specificity of the cloned M.CviBIII gene. Plasmid pNC-1A.14.8 DNA was resistant to M.TaqI methylation (Table 1). The E. coli dam and dcm methyltransferases were not responsible for adenine methylation in TCGA sequences since identical results were obtained when pNC-1A.14.8 was propagated in the dam⁻, dcm⁻ E. coli strain GM2163. These results indicate that adenine in TCGA sequences was methylated by cloned M.CviBIII in vivo in E. coli.

The site specificity of the cloned M.CviBIII gene product was also demonstrated *in vitro*. When TaqI-digested lambda DNA was used as the substrate for M.CviBIII prepared from dam⁻, dcm⁻ E. coli harboring pNC-1A.14.8 or the overproducing plasmid, pNC-1A.14-lac (described in Materials and Methods), the level of incorporation of ³H-methyl groups from S-adenosyl-[methyl] methionine was reduced to control levels (Table 2). This result demonstrates that M.CviBIII methylates

Figure 2. Hybridization of pNC-1A.14.8 to NC-1A DNA. NC-1A DNA (1 μ g) was digested with NdeI, BamHI, NdeI and BamHI, BglIII, NdeI and BglIII, SstI, NdeI and SstI, HindIII, and NdeI and HindIII (lanes 1 through 9, respectively). The restriction digests were electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose and probed with 32 P-pNC-1A.14.8.

adenine in TCGA sequences since TaqI cleavage of substrate DNA destroys potential M.CviBIII methylation sites.

M.CviBIII is encoded by NC-1A DNA.

The 2.1 kbp fragment containing M.CviBIII has a single NdeI site about 1.2 kbp from one end and no internal BamHI, BglIII, SstI and HindIII sites (Fig. 3B). When Southern blots of NC-1A DNA digested with these restriction endonucleases were probed with nick-translated pNC-1A.14.8, two NdeI fragments and single BamHI, BglIII, SstI and HindIII fragments were identified as expected (Fig. 2). Two fragments were also present in double digests of NdeI plus one of each of the other four restriction endonucleases. No hybridization to dot blots of Chlorella NC64A DNA was observed (data not shown). These results indicate that NC-1A contains a single copy of the M.CviBIII gene and that the host Chlorella does not contain the gene.

A

B

Figure 3. (A) Functional domain of the M.CviBIII gene as determined by Tn5 mutagenesis. Symbols represent unique Tn5 insertions for which M.CviBIII activity was lost (open symbols) or retained (closed symbols).

(B) Schematic representation of pNC-1A.14.8. Lines and enclosed regions indicate pUC8 and virus NC-1A sequences, respectively. The shaded region indicates the location of the M.CviBIII gene as determined by Tn5 mutagenesis.

Analysis of the M.CviBIII gene by Tn5 mutagenesis.

A physical map of the M.CviBIII gene was determined by mutagenesis of pNC-1A.14.8 with transposon Tn5. Plasmids containing Tn5 insertions were assayed qualitatively for M.CviBIII directed methylation by testing the sensitivity of the plasmid DNA to TaqI. Independent insertions localized M.CviBIII to a 1.5 kbp region on pNC-1A.14.8 (Fig. 3A).

Tn5 insertions which had no effect on M.CviBIII activity

ATCAGTAATGTGTTAATTCBAACTTGTATGTGTAACTCAATTCBAAATCAGTCTACTACAAATATTCTTTBACACCTTGT
 TTACAGCAGCATCAGAAATTTGATATTCACAACTCTGTATAAATATTGTTTTATCTGTTTCAGATCATATGAAGTCTACTGTTTTA
 GAGTGTCAAAATATCATATCAGATATATATTTATBACAACTCTCAGTCTCATAGTTATTTTTCAGCAATGAGTCTCTGT
 CATCATATCAATTTGACCAATGTTGTTCTGTATACAAATATACAGATGATTTTATTTTBCBAAATCAGATCTCTTCTATGAGAGAA
 CCAGTTAATACAAATCBAAGAGTTTCATACAGCTTCAATTTATATBACATAGATGACBAACATTGTACAAATTTACACAGAGTG
 GTCTCATACATACAAATGATTTGCAATATATTCTCAATGATACACACCAAGTTTGTGAGAAATTTTGTACAGTACAAATTA
 TCAGAGGAGCTTACBAGATATGATTTTATTTGCGCCACATACAGAGTAAATTTGATAGAGAAACCAATGATGATTAATTTT
 TCBAAATTAATATAGAACTTCBTAAGAAATATCTGTTTACTGCAAAACATTCATATGATATTAATCTCTCATGATATTCTGT
 TGTTTGAAGTAAAGTTGAAGCTTTCTGTTTGTAAATACAAATGATATGATGATTTGTTGTGCAATGTCATATCAAAATGTTG
 TGATATATGTGTAACTAAATACATATATATTAA

1 27 54 81
 ATGACGTAGAACATATACAAATTAACCACTGATTTTGAAGAACTTACAAAGAAAGAAATCAAAACAGATATTTTTCCT
 H N V E Q Y T Q V T T B F E K T L T K E K S K Q G I F F T

108 135 162
 CCAAAACAGTGGAGAAATTTATTTGGTTACAGAACATTTTCAAAATACCCCTGTTTATGATTTCTAGAACAGTTGTGGACG
 P K T V R E K L F G F T E N F Q N T P B F S I L E P S C G T

189 216 243 270
 GGGAAATCTCAGAAATGTTGTGACGTTTCCACTTGCAGCATCAAGGAGTAGAACTTGATATGATATGTCATACATTTGTTCA
 G E I I S E C V E R F P L A S I K G V E L D N D N S T I C S

297 324 351
 AAAAAATACBAGATATATGATGACATTTGTAATGAAGCTTTTGTCTTGGAAAGTGGAAATTTGATTTTCATAGTGGAAATCCA
 K K Y A E Y N V D I V N E D F L L M K G G K F D F I V G N P

378 405 432
 CCATATGTAGTGAACATCTGATATAAACGATACAGAAATGCAAAAGGAGGTCAATCTTTACGTAGAAATTTCTGTATAATGT
 P Y V V R P S G Y K N D N R I A K G R S N L Y V E F L Y K C

459 486 513 540
 ATACAGAACTCTGAAGAGGATGAACTCTGTTTTATCATACCAAGCATTTGAAATAGTAGTTTCTACBAAACCCATCCGAA
 I T E N L K E D G I L A F I I P S T I G N S S F Y E P I R K

567 594 621
 CTATTAATTAATCTGATATCTATGTTGATATGATAAACATGATTTCTGTATACCAATACGCGTTGTGTTCTGTATGATC
 L I I T L D I L S F E I L D K N D F C D T N T R L C S I V I

648 675 702
 AAAAACTCCCTGAACTGAAATATACATATAGGATATATTTCBATAAAGACATCCACATCATGAAATAGTTATATAGGTTG
 K N S P G T G K Y T Y R D Y I C D K D I P H N G N S Y I G S

729 756 783 810
 TTGATCTAAATTCAGAACTGTTTGTCTGTTGGCAATGTCATTAATTTTCTGATAGAAAGTGAAATCCATTTTTCAGAGTAGT
 L D L K F K T G F A M A M V N K F F T D K S E I P F F T S S

837 864 891
 AATATCAATTAATGAATTCATATTGATATAAATGAATACTTGACAGAGATACGACGAATTTTACAGGAAGGCCCTATTG
 N I K L N E I N I G D K N K Y L T Q D T T K F F T G K A L L

918 945 972
 ATAAAGACAGCTCTGCCGTAAAGAGTGGAGATTTGAGTTTGGTTTCTCTTACGAAATGATAAATGGCCGTGATATGAC
 I K T A S A G K R G G R F E F G F S L Y E N D K M A V D N D

999 1026 1053 1080
 ATATCTAAATTCAGAGCTGATACCTTCTGTATAGTTCAAGATGTTGATGAGGATGTTACAAACGAATTCATTAATATCTT
 I I V I Q G P D T V L S I V Q D V L N K D V T N E F I N I L

1107
 GTCAATACGACATCAGATGAACATTAATATCTATCCATGTTTTAACTGAATATCAGTTGTGTATAGAGTTGTCAATGT
 V N N G H I S K L L K S I P L F .

CAATAGTATCAAACTGCTACCCATGCAAGTACATAGTACATTCACACCATGATC

Figure 4. DNA sequence of the 2.1 kbp fragment containing the *M. QviBIII* gene. The predicted amino acid sequence is below the only long open reading frame found within the nucleotide sequence. Sequences bearing homology to *E. coli* consensus promoter sequences are underlined in the 5'-flanking region. A potential polyadenylation signal, AATAAA, is underlined in the 3'-flanking region.

were all located between the lac promoter of pUC8 and the putative gene (Fig. 3B). Thus M.CviBIII transcription is apparently initiated from a promoter in the virus DNA because Tn5 contains both transcription and translation termination signals (11).

M.CviBIII was also expressed when the gene was cloned in the opposite orientation with respect to the lac promoter in pUC9 (data not shown); this provides additional evidence that M.CviBIII transcription in E. coli utilizes a promoter in the virus DNA.

DNA sequence analysis of the M.CviBIII gene.

The DNA sequence of the entire 2.1 kbp fragment containing the M.CviBIII gene was determined by dideoxynucleotide chain termination (Fig. 4). A single open reading frame (ORF) of 1131 bp was identified within the functional domain predicted by Tn5 mutagenesis which could encode a polypeptide of 377 amino acids with a molecular weight of 42,828. We assume that the ATG codon designated coordinate 1 is the initiation codon for the M.CviBIII gene for two reasons: (i) An ochre (TAA) codon immediately precedes ATG and (ii) Fusion of the lacZ amino terminus at this site results in overproduction of M.CviBIII (see below).

The DNA sequence of the M.CviBIII gene displays additional interesting characteristics. The Tn5 map suggested that the M.CviBIII gene contains a transcriptional promoter, utilized by E. coli RNA polymerase, located within about 200 bp 5' to the translational start site. A putative promoter sequence (TTGAAA...19 nucleotides...AAAAAT) beginning 113 bp upstream from the ATG start codon 1, corresponds to the E. coli promoter consensus sequence (18) with only one base deviation in the -35 region and two deviations in the -10 region of the RNA polymerase binding sites.

The gene also has a potential polyadenylation signal (AATAAA) 19 bases downstream from the end of the ORF. Although polyadenylation of M.CviBIII mRNA from infected Chlorella cells has not been demonstrated by sequencing cDNA, the mRNA adheres to oligo (dT)-cellulose and is approximately 270 bases larger than the ORF (manuscript in preparation).

Identification of a lacZ-M.CviBIII fusion protein in maxicells.

The M.CviBIII ORF predicts a polypeptide with a molecular weight of 42,828. Attempts to identify the M.CviBIII gene

Figure 5. Plasmid-directed protein synthesis in maxicells. 35 S-methionine-labeled maxicell lysates prepared from strains without plasmids (lane 1), harboring pUC19 (lane 2) or harboring pNC-1A.14-lac (lane 3) were electrophoresed on a 12.5% SDS-PAGE gel, dried and autoradiographed. The arrow indicates the new polypeptide encoded by pNC-1A.14-lac. The two forms of beta-lactamase, bla* and bla are indicated. The numbers are the sizes (kd) of molecular weight markers.

product in maxicells harboring the original clone, pNC-1A.14.8 were unsuccessful, probably due to relatively poor expression from the virus promoter. Therefore, the M.CviBIII gene was placed under control of the *lac* promoter by creating an in-frame fusion with the 19 amino-terminal codons of *lacZ* in pUC19 to generate pNC-1A.14-lac. The specific activity of M.CviBIII in sonicated crude cell extracts from strains

Figure 6. The *M.CviBIII* gene is deleted in NC-1A.S1c.

(A) Agarose gel (0.8%) electrophoresis of NC-1A DNA (lane 1) and NC-1A.S1c DNA (lane 2) digested with *Bam*HI.

(B) Hybridization of 32 P-pNC-1A.14.8 to a Southern blot of the gel in A.

harboring pNC-1A.14-lac was approximately nine times that of the original clone (Table 2). The DNA sequence of the lacZ-*M.CviBIII* fusion predicts a polypeptide of about 45 kD. A protein of this size was synthesized in maxicells containing pNC-1A.14-lac, but not in vector-containing strains (Fig. 5).

Mutant of NC-1A lacks *M.CviBIII*.

DNA isolated from a spontaneously derived small-plaque-forming NC-1A mutant, designated NC-1A.S1c, was sensitive to *Sal*I and *Taq*I and resistant to *Hinf*I and *Mbo*I (data not shown). These results suggest that the mutant virus is defective for *M.CviBIII* but not *M.CviBI* and *M.CviBII*.

The *Bam*HI fragment pattern of NC-1A.S1c DNA differed

Figure 7. Six other *Chlorella* viruses contain DNA sequences homologous to the *M.ϕviBIII* gene. DNA from *Chlorella* viruses NC-1A, NE-8A, Al-2C, MA-1E, NY-2F, CA-1D and NC-1B (lanes 1 through 7, respectively) was digested with *DpnI* and electrophoresed on a 2% agarose gel. The restriction fragments were transferred to nitrocellulose and probed with ^{32}P -pNC-1A.14.8. Sizes (kbp) of pUC8 *HinfI* fragments are indicated on the left.

slightly from those of wild type NC-1A DNA (Fig. 6A). pNC-1A.14.8 DNA did not hybridize to Southern blots of NC-1A.S1c DNA (Fig. 6B); this suggests that the entire *M.ϕviBIII* gene is deleted from the mutant. This was confirmed by probing dot blots of DNA with pNC-1A.14.8 (data not shown). Preliminary mapping data indicates that a region of about 40 kbp of the NC-1A genome, including the *M.ϕviBIII* gene, is deleted in NC-1A.S1c (Skrdla and Van Etten, unpublished data). This indicates that *M.ϕviBIII* is not essential for NC-1A replication.

Occurrence of M.CviBIII gene in other Chlorella viruses.

We have partially characterized 30 viruses which infect the same host, Chlorella NC64A, and placed these viruses into 11 classes and 9 restriction groups (2). One classification criterion was the resistance or susceptibility of the virus DNAs to 13 restriction endonucleases; DNA sequences containing m⁵dC and/or m⁶dA can be implied from the resistance of a DNA to certain restriction endonucleases. NC-1A belongs to DNA restriction group D and is the sole member placed in class 6. To determine if any of the other 29 viruses contain a homologous M.CviBIII gene, dot blots of each of their DNAs were probed with pNC-1A.14.8. DNA from the six viruses included in class 7, restriction group E, hybridized to pNC-1A.14.8 (data not shown). DNA from each of these viruses was also resistant to SalI and TaqI. Presumably a methyltransferase homologous to M.CviBIII methylates adenine in TCGA sequences in each of these viruses. To examine the relatedness of these genes Southern blots of DpnI digests of these 6 virus DNAs and NC-1A DNA were probed with pNC-1A.14.8; the hybridization patterns were identical to that of NC-1A (Fig. 7). Identical patterns of hybridization were also observed in MspI digests (data not shown). The lack of restriction fragment length polymorphisms indicates that the gross structure of the M.CviBIII gene has not diverged in this group of Chlorella viruses.

Interestingly, pNC-1A.14.8 did not hybridize to DNA from class 5 viruses, SC-1A and SC-1B, which are also members of restriction group D. SC-1A and SC-1B DNAs, like NC-1A, are resistant to SalI and TaqI and contain similar levels of m⁶dA. This suggests that SC-1A and SC-1B may encode a methyltransferase different from M.CviBIII which methylates adenine in TCGA sequences.

DISCUSSION

The M.CviBIII gene is the first Chlorella virus gene to be cloned and expressed in E. coli; adenine residues in TCGA sequences are methylated in vivo and in vitro with crude cell extracts from E. coli strains containing M.CviBIII. M.CviBIII is one component of a complex of restriction and modification enzymes synthesized after NC-1A infection of Chlorella NC64A. Two other DNA methyltransferases, M.CviBI and M.CviBII, which methylate adenine in GANTC and GATC sequences respectively

(Xia, Narva, and Van Etten unpublished data), in addition to one cognate type II restriction endonuclease CviBI (5) have been identified in extracts from NC-1A infected Chlorella cells. Since NC-1A DNA contains 7.1% m⁵dC in addition to 7.3% m⁶dA (19) additional unidentified cytosine and possibly adenosine methyltransferases must be synthesized after NC-1A infection. Presumably all of these enzymes are encoded in the virus DNA since their activities only appear after NC-1A infection of the alga. The hybridization of the M.CviBIII gene to NC-1A DNA but not to Chlorella NC64A DNA conclusively establishes the viral origin of this methyltransferase.

Interestingly, both Tn5 mutagenesis and cloning the gene in the opposite orientation indicate that M.CviBIII expression in E. coli is independent of the lac promoter in pUC plasmids. Sequences bearing considerable homology to the -35 and -10 regions of the consensus sequence for E. coli promoters (18) are located beginning 113 bp upstream from the translational start site. Whether this is the functional promoter for the M.CviBIII gene in virus infected Chlorella or merely a cryptic promoter fortuitously utilized by E. coli RNA polymerase awaits further experimentation.

The function of these Chlorella virus restriction-modification systems has not been established. However, one hypothesis is that the DNA restriction endonucleases participate in nucleic acid metabolism after infection by selectively cleaving host DNA. Methylation of the virus DNAs by methyltransferases may protect nascent virus DNA. In some cases methyltransferases, including M.CviBIII, exist without an apparent corresponding restriction endonuclease. Perhaps these DNA methyltransferases influence gene regulation, DNA repair, or control of replication. It is noteworthy that M.CviBIII expression parallels the onset of virus DNA synthesis; both events begin about 45 to 60 min after infection (unpublished data).

Only 7 of 30 Chlorella viruses contained a gene homologous to M.CviBIII. Finding a NC-1A mutant deleted for M.CviBIII infers that M.CviBIII is not essential for replication or control of gene expression in NC-1A. However, since the deletion spans about 40 kbp, one can not rule out compensatory deletion of other genes which restore viability.

The restriction-modification systems encoded by these

Chlorella viruses may provide basic information pertaining to site-specific enzyme-DNA interactions at the molecular level. For instance, in the case of M.QviBIII comparison of the predicted amino acid sequence with that of other cloned adenine bacterial methyltransferases revealed that M.QviBIII shares regions of amino acid sequences homologous to those in M.TaqI; M.TaqI also methylates adenine in TCGA sequences (Jack Benner, New England Biolabs, personal communication; manuscript in preparation). Since Chlorella viruses are evolutionarily distant from bacteria, examining conserved polypeptide sequences should reveal structural and/or functional domains important to site specific DNA methylation.

ACKNOWLEDGMENTS

We thank M.G. Marinus for kindly providing E. coli strain GM2163, Paul Staswick and Les Lane for many helpful discussions and Dwight Burbank for technical assistance. This investigation was supported, in part, by Public Health Service grant GM-32441 from the National Institute of General Medicine and grant DE-AC02-82ER12086 from the Department of Energy.

Published with the approval of the Director as paper no. 8386, journal series, Nebraska Agricultural Research Division.

Present addresses: ¹American Cyanamid Co., Lederle Laboratories 96A/207, Pearl River, NY 10965 and ²Department of Genetics, University of California, Davis, CA 95616, USA

REFERENCES

1. Van Etten, J.L., Burbank, D.E., Schuster, A.M. and Meints, R.H. (1985) *Virology* 140,135-143.
2. Schuster, A.M., Burbank, D.E., Meister, B., Skrdla, M.P., Meints, R.H., Hattman, S., Swinton, D. and Van Etten, J.L. (1986) *Virology* 150,170-177.
3. Xia, Y. and Van Etten, J.L. (1986) *Mol. Cell. Biol.* 6,1440-1445.
4. Xia, Y., Burbank, D.E., Uher, L., Rabussay, D. and Van Etten, J.L. (1986) *Mol. Cell. Biol.* 6,1430-1439.
5. Xia, Y., Burbank, D.E. and Van Etten, J.L. (1986) *Nucleic Acids Res.* 14,6017-6030.
6. Van Etten, J.L., Burbank, D.E., Xia, Y. and Meints, R.H. (1983) *Virology* 126,117-125.
7. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. Vieira, J. and Messing, J. (1982) *Gene* 19,259-268.
9. Marinus, M.G., Carraway, M., Frey, A.Z., Brown, L. and Arraj, J.A. (1983) *Mol. Gen. Genet.* 192,288-289.

10. Sancar, A., Hack, A. and Rupp, W.D. (1979) J. Bacteriol. 137,692-693.
11. de Bruijn, F.J. and Lupski, J.R. (1984) Gene 27,131-149.
12. Van Etten, J.L., Meints, R.H., Burbank, D.E., Kuczmariski, D., Cuppels, D.A. and Lane, L.C. (1981) Virology 113,704-711.
13. Mann, M.B., Rao, R.N. and Smith, H.O. (1978) Gene 3,97-112.
14. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33,103-119.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74,5463-5467.
16. Messing, J. (1983) Methods in Enzymology, Wu, R., Grossman, L. and Moldave, K. Eds., Vol. 101, pp. 20-78, Academic Press, New York.
17. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80,3963-3965.
18. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13,319-353.
19. Van Etten, J.L., Schuster, A.M., Girton, L., Burbank, D.E., Swinton, D. and Hattmann, S. (1985) Nucleic Acids Res. 13,3471-3478.