Molecular cloning and characterization of the gene encoding the adenine methyltransferase M.*Cvi*RI from *Chlorella* virus XZ-6E

Chris Stefan+, Yuannan Xia§ and James L.Van Etten*
Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

Received Octobr 22, 1990; Revised and Accepted December 11, 1990 Genbank accession no. M38173

ABSTRACT

The gene encoding the DNA methyltransferase M.*Cvi*RI from *Chlorella* virus XZ-6E was cloned and expressed in *Escherichia coli*. M.*Cvi*RI methylates adenine in TGCA sequences. DNA containing the M.*Cvi*RI gene was sequenced and a single open reading frame of 1137 bp was identified which could code for a polypeptide of 379 amino acids with a predicted molecular weight of 42,814. Comparison of the M.*Cvi*RI predicted amino acid sequence with another *Chlorella* virus and 14 bacterial adenine methyltransferases revealed extensive similarity to the other *Chlorella* virus enzyme.

INTRODUCTION

Thirty seven large (150 to 190 nm in diameter), polyhedral, dsDNA-containing (>300 kb in size), plaque forming viruses which infect the unicellular, eukaryotic *Chlorella*-like green alga strain NC64A have been isolated and partially characterized (1−3). These viruses were placed into 16 classes on the basis of plaque size, reaction with antibody and the nature and abundance of methylated bases in their genomic DNAs (2; Van Etten unpublished data). Each viral DNA contains 5-methylcytosine (5mC); 5mC varies from 0.1% to 47.5% of the total cytosine. In addition, 25 of the 37 viral DNAs also contain N^6-methyladenine (6mA); 6mA varies from 1.45% to 37% of the total adenine. The finding that methylation was sequence specific led to the discovery that at least some of these viruses direct the synthesis of DNA methyltransferases and DNA site-specific (restriction) endonucleases during infection (4−11).

Cells infected with the virus XZ-6E (21.2% 5mC and 15.2% 6mA) contain two site-specific endonucleases, named *Cvi*RI and *Cvi*RII (Traylor, Xia, Zhang, Roy, and Van Etten, manuscript in preparation). *Cvi*RI is the first restriction endonuclease to recognize the sequence TGCA; the enzyme cleaves between G and C and does not cleave TGC^{m}A sequences. *Cvi*RII recognizes the sequence GTAC. The enzyme cleaves between the G and T and does not cleave GT^{m}AC sequences. Since XZ-6E DNA is resistant to both *Cvi*RI and *Cvi*RII, presumably XZ-6E induces the synthesis of at least two cognate methyltransferases, M.*Cvi*RI and M.*Cvi*RII, which methylate adenine in TGCA and GTAC sequences respectively, or a subset of these sequences.

We are trying to isolate and characterize the genes of some of these *Chlorella* virus encoded DNA methyltransferases and site-specific endonucleases in order to understand their biological significance and to compare them to prokaryotic enzymes. This report describes the successful cloning, characterization, and expression in *Escherichia coli* of the M.*Cvi*RI gene from virus XZ-6E.

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 XZ-6E and the other viruses have been described (1−3, 12).

E. coli strain JH132 [McrA−, McrB−, Mrr−] and RR1 [McrA+, McrB−, Mrr−] served as hosts for plasmids pBR322, pUC18, pUC19, and recombinant plasmids (13, 14). *E. coli* XL-1 Blue (Stratagene) was used for pBluescript KS(−) (Stratagene) cloning and the generation of single stranded DNA for dideoxy sequencing.

Cloning strategy

Virus XZ-6E DNA was isolated as described previously (15). Libraries of *Eco*RI partial digestion products of XZ-6E DNA were prepared in plasmid pBR322 and transformed into *E. coli* strain JH132 by standard procedures (16). Potential clones containing the M.*Cvi*RI and M.*Cvi*RII genes were selected from the XZ-6E DNA library as suggested by Mann *et al.* (17). The restriction endonuclease *Pst*I (CTGCAG) is inhibited by 6mA in the TGCA portion of its recognition sequence (18). Since M.*Cvi*RI methylates adenine in TGCA sequences, recombinant plasmids containing a M.*Cvi*RI gene which is expressed in *E. coli* should be resistant to *Pst*I. Likewise, clones containing

* To whom correspondence should be addressed

Present addresses: +Department of Biology, Washington University, St Louis, MO 63130, USA and §Department of Microbiology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

M.CviRII were screened by treating the library with *Rsa*I; *Rsa*I cleaves GTAC but not GT^{m}AC sequences (19). Resistant plasmids were selected by treating the XZ-6E plasmid library with either *Pst*I or *Rsa*I before transforming *E. coli*.

Enzymes and DNA methyltransferase assays

Methyltransferase activity was assayed either by measuring transfer of methyl-^{3}H from S-adenosyl[methyl-^{3}H] methionine to DNA or by its ability to protect substrate DNAs from digestion with appropriate restriction endonucleases (4). Restriction endonucleases and DNA methyltransferases were purchased from New England Biolabs or Bethesda Research Laboratories and used according to supplier's recommendations.

Nucleic acid electrophoresis and hybridization

DNA was electrophoresed on agarose gels in either TAE buffer (40 mM Tris-acetate, 6 mM EDTA) or TPE buffer (80 mM Tris-phosphate, 8 mM EDTA) (16). Nucleic acids were stained with 0.5 μg/ml ethidium bromide and visualized by mid range ultraviolet illumination. Standard protocols were used for Southern transfer of DNA to nitrocellulose filters (16). For dot blots viral DNAs were denatured and applied to nitrocellulose membranes as described (20). Radioactive DNA probes were prepared by nick translation using a Bethesda Research Laboratories kit.

DNA sequencing

After the M.CviRI gene was localized to plasmid pXZ-6E.14, the entire viral insert DNA was subcloned using restriction sites within the insert DNA and the subclones were inserted in the vector pBluescript KS(−). Single stranded DNA was generated by adding the helper phage MCS-113 (Stratagene) and isolated using common techniques (21). Dideoxy sequencing of both DNA strands using 5′-[α^{35}S]-thiotriphosphate was as previously described (22, 23).

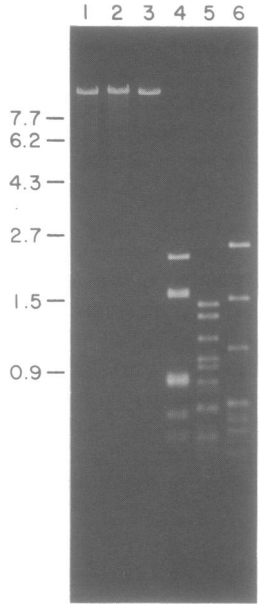

Figure 1. Sensitivity of *Xho*I linearized pXZ-6E5.9 DNA to several restriction endonucleases. Plasmid pXZ-6E5.9 DNA was digested with *Xho*I, *Xho*I/*Pst*I, *Xho*I/CviRI, *Xho*I/*Rsa*I, *Xho*I/*Taq*I, and *Xho*I/*Hin*fI (lanes 1 through 6, respectively). Numbers on the left refer to sizes (kb) of lambda *Sty*I fragments.

RESULTS

Identification and isolation of the M.CviRI gene

About 100 transformants were obtained after *Pst*I digestion of the partial *Eco*RI XZ-6E library. Plasmid DNA was isolated from 30 individual colonies and tested for resistance to *Pst*I. One clone, named pXZ-6E5.9, contained a 5.9 kb insert of XZ-6E DNA and was resistant to *Pst*I and CviRI (Fig. 1, lanes 2 and 3, respectively). Plasmid pXZ-6E5.9 was sensitive to three other restriction endonucleases, *Rsa*I, *Taq*I, and *Hin*fI (Fig. 1, lanes 4, 5, and 6), whereas virus XZ-6E genomic DNA was resistant to each of these enzymes. We conclude that pXZ-6E5.9 contains the M.CviRI gene.

Subclones of plasmid pXZ-6E5.9 were prepared as outlined in Fig. 2A and tested for sensitivity/resistance to *Pst*I and CviRI. These experiments localized the M.CviRI gene to a 2.1 kb region at one end of pXZ-6E5.9; this plasmid was named pXZ-6E.14.

Two additional experiments support the site specificity of the cloned M.CviRI gene. (i) pXZ-6E5.9 DNA was resistant to methylation with M.*Pst*I, but not M.*Taq*I (methylates adenine in TCGA sequences) (data not shown). (ii) The multiple cloning site of pUC19 contains a single M.CviRI site which overlaps the *Hin*dIII site of pUC19. *Hin*dIII is inhibited by methylation of the first adenine in its recognition sequence (18). When the 3.5 kb *Hin*dIII fragment of pXZ-6E5.9 was cloned into the *Hin*dIII site

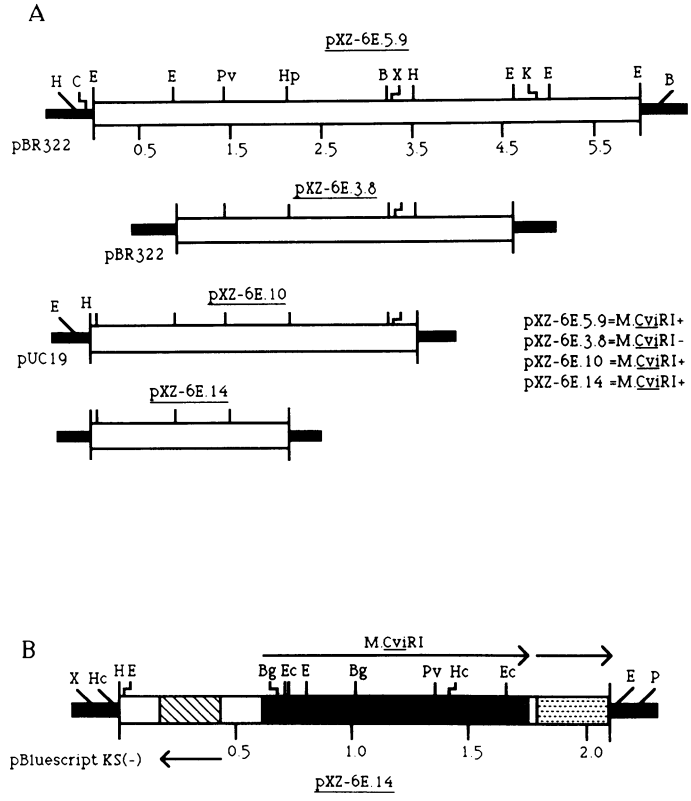

Figure 2. A) The functional domain of the M.CviRI gene in plasmid pXZ-6E5.9 was localized to plasmid pXZ-6E.14 by subcloning. Open region is the insert DNA and the closed end regions the vector. B) Plasmid pXZ-6E.14 was sequenced and the M.CviRI gene is represented by the shaded, cross-hatched region. Another ORF (slashed lines) and a possible ORF (stipled area) at the end of the clone are shown. The direction of the transcripts is indicated by the arrows. Restriction endonucleases are: B (*Bam*HI), Bg (*Bgl*II), C (*Cla*I), E (*Eco*RI), Ec (*Eco*RV), H (*Hin*dIII), Hc (*Hinc*II), Hp (*Hpa*I), P (*Pst*I), Pv (*Pvu*I), and X (*Xho*I).

of pUC19 to construct pXZ-6E.10 (Fig. 2A), pXZ-6E.10 DNA was resistant to *Hin*dIII at this site, as well as to *Pst*I (data not shown). We conclude that M.*Cv*iRI recognizes TGCA sequences or a subset of this sequence.

Attempts to clone the M.*Cv*iRII gene by *Rsa*I digestion of the partial *Eco*RI XZ-6E library were unsuccessful.

DNA sequence analysis of the M.*Cv*iRI gene

The entire viral insert DNA in plasmid pXZ-6E.14 was subcloned and sequenced (Fig. 3). Two open reading frames (ORF) of 243 (ORF-A) and 1137 bases were identified. These two ORFs could code for polypeptides of 81 (predicted molecular weight of 9,504) and 379 (predicted molecular weight of 42,814) amino acids respectively. We believe the larger ORF is the M.*Cv*iRI gene because the amino acid sequence is similar to other adenine methyltransferases (see below).

We assume that the ATG codon designated coordinate 1 is the start codon for the M.*Cv*iRI gene because an ochre (TAA) codon is present two codons prior to ATG. M.*Cv*iRI was also expressed when pXZ-6E.14 was cloned in the opposite orientation with respect to the *lac* promoter in pBluescript KS (data not shown). This suggests that the M.*Cv*iRI gene carries a promoter which can be utilized by *E. coli* RNA polymerase. A putative promoter sequence (ATGACA...15 nucleotides...TACACT) beginning 100 bp upstream from the ATG start codon, corresponds to the *E. coli* promoter consensus sequence, with only one base deviation in the −35 region and two deviations in the −10 region of the RNA polymerase binding sites. Whether this is the functional promoter of the M.*Cv*iRI gene in virus infected *Chlorella* or merely a cryptic promoter fortuitously utilized by *E. coli* RNA polymerase is not known. Transcriptional and sequence studies on ORFs of the prototype *Chlorella* virus PBCV-1 indicate that regions upstream from the ATG translational start site are AT rich (24). The region upstream from the M.*Cv*iRI gene is also AT rich; regions −100 and −50 bases from the ATG start site contain 74% AT and 85% AT, respectively. Finally four direct repeats of the 11 base sequence ACAAAAATGAC occur in tandem 139 to 95 bases upstream of the ATG start codon.

The other ORF (ORF-A) in pXZ-6E.14 is transcribed in the opposite direction from the M.*Cv*iRI gene and begins 215 bases upstream of the M.*Cv*iRI gene. This ORF contains three direct repeats of 33 bases. Translation of these sequences could produce a protein containing three repeats of the amino acid sequence glu-lys-glu-arg-val-lys-ala-glu-lys-leu-ala. The significance of this repeat unit, if any, is unknown. Comparison of the predicted amino acid sequence of ORF-A to proteins in GenBank did not reveal any obvious similarities.

```
GAATTCTACTTATATACCATATCATTTTCCATAACAAATTGAAAGTCGAATGATTTACCACGTCCTCCGATTTGTTCT
ACGCTCTTCAATTTTGTAATATCAATGACATTTGAAATACTTTCTAACAGTCTCTGTTGAACACTTGTATTTTCGTTA
ATATCACGATTATTTAGTGTATCAACTATAATTTTTCTCGCTGCTTATTGTTAATATCGTTGTCTCCGCGAATACCTG
                        * L K E R Q K N N I D N D G R I G
TTACGAAAATATCATCAGGATTATCCCGTTCCTTTTCAGCAAGTTTTTCCGCCTTTACTCGTTCCTTTTCAGCAAGTT
T V F I D D P N D R E K E A L K E A K V R E K E A L
TTTCCGCCTTTACTCGTTCCTTTTCAGCAAGTTTTTCCGCCTTTACTCGTTCCTTTTCGATTTTGCTAACCTTTTTCA
K E A K V R E K E A L K E A K V R E K E I K S V K K
TTTTCATAAGATTGATTATGTTTATAATATTCAGCATATTTATGTTCTGTTCACATATTAATATATATAAATAAAATG
M K M L N I I N I I N L M ⇐ ORF A
ACACAAAAATGACACAAAAATGACACAAAAATGACACAAAAATGACATAGAATTTACACTTGTACACTAGACACGTGT
ACACAATATCATATCAACATACGAAACAACTTAAATTAAAAAAAATGATTGATTTTATAAATT

   1  ATGAAGTTATTCGAAGACAAATCCGTCGAGTTTCATAAACGGTTGAGTAAGAAGGAAAGA   60
      M K L F E D K S V E F H K R L S K K E R
  61  TCTGACGGAGGGGTATTTTTCACTCCGAAAGATATCAGAGATATCGTGTTTGAAGAACTT   120
      S D G G V F F T P K D I R D I V F E E L
 121  GGAGATTTTGAACCTACGAATATATTGGAACCAACATGTGGCACGGGTGAATTCATTTCC   180
      G D F E P T N I L E P T C G T G E F I S
 181  GATTGCAGAAAAGTATACAAAAACTCTCGCATCATTGGCGTAGAAATAGATCCACGTTCT   240
      D C R K V Y K N S R I I G V E I D P R S
 241  GCCGAACTTGCAAGGGACGGTTCGAAAAACGAAATAATTGTTCACGATTTCATGACATGG   300
      A E L A R D G S K N E I I V H D F M T W
 301  GATACTGATGAGAAATTCGACCTTATAATAGGAAATCCACCGTATTTCACAAGACCGACT   360
      D T D E K F D L I I G N P P Y F T R P T
 361  GGTTTCAAACACGACCCTTCCGTAGTCAAATGTAGATCTAACATTTGTATCGAAGTTCTT   420
      G F K H D P S V V K C R S N I C I E V L
 421  CACAAATGTATTACACGCCATTTGGCAGACAATGGTATGTTGGCAATGGTTCTACCGGTA   480
      H K C I T R H L A D N G M L A M V L P V
 481  TCTATTTTGAACAGTAAATTTTACACACCGACTATTGATCTTATCACAGATACTATGGAT   540
      S I L N S K F Y T P T I D L I T D T M D
 541  GTTGTATCTGCCCGCGCTATTAAAAAAAACAATTTTATGGGGACAAACGTGCGCGTAATG   600
      V V S A R A I K K N N F M G T N V R V M
 601  GTATTTATTATTCGTAAGCGTACTCCGGGGTTCGTTAGTAAATATACTTTCAAGACTTCT   660
      V F I I R K R T P G F V S K Y T F K T S
 661  CTTGGGAAGGTCATCATAAACCCCGACGGAGAGCGTCTAGGTTCTATAGTTTCTGGTAAG   720
      L G K V I I N P D G E R L G S I V S G K
 721  AAAACGATCGGTTCGTTAAACGTAAATATTTCTTTCGGTGTAACGTTGGCTAGCGTCAAA   780
      K T I G S L N V N I S F G V T L A S V K
 781  GAATATTTCGTTGACAAAAGTTGTTCAGGTTCGTTCCCACTTATTTGTTATAACAATATT   840
      E Y F V D K S C S G S F P L I C Y N N I
 841  GCTAAGAAAGGTGATCTGCTGTTTGTCAGTGACAAATACTCTAAAAAACGTTTCAACGGA   900
      A K K G D L L F V S D K Y S K K R F N G
 901  AGAGCTATTCTCATACCGAGGGGTTATGCTCATGGGGATTATTCATTCAATTTCATAGAT   960
      R A I L I P R G Y A H G D Y S F N F I D
 961  TACACTAACGATTATTTTATAATAGAAAATCACGTAATTGCAATAACCGGTGAGGATTGT   1020
      Y T N D Y F I I E N H V I A I T G E D C
1021  GTATTAGATATCATAGCAAAATCGTTTGCAGACCATCGGACTCGTGAATTTTGTAGACTA   1080
      V L D I I A K S F A D H R T R E F C R L
1081  TTATGTTCTTCTGGAGATATAAGCAAAGATTATGTCAAAGAAATACCGGTGTTCGGGTGA   1140
      L C S S G D I S K D Y V K E I P V F G *

AATTTAAATATGCCAAACGAATTTATTAAGAACTTATGAAAAATTTACATGATAATTATAATTACGTTTTAATAATGT
CTCTGGTTAAACTTCCCATCATCAATTTCCCGGATGCTATTCTCATAGCAGGAGGTGGTGCTAAGTCGATGAGCGGTC
TCGGAGCTATACACGTTTTAAAACGAAATGGTCAACTAAAAAATTTAAAGGTCGTTGCGGGCACTTCCGCTTTAAGTA
TCGTCGCTGCCGGTGTTGCGTTGAACAAAGACCCTGTGAGTATGTGTAAAGCGTTCATTGATGAGATATACAAACCGT
CGTTTGATATATCTAATTTCGTT
```

Figure 3. DNA sequence of the M.*Cv*iRI gene and its flanking regions. The predicted amino acid sequence of M.*Cv*iRI is indicated. The amino acid sequence asn-pro-pro-tyr which is present in many adenine methyltransferases is double underlined. Sequences bearing homology to *E. coli* consensus promoter sequences in the 5′ flanking region are indicated by the wavy line. A second small ORF (designated ORF-A) and its predicted amino acids are also indicated. A 33 base direct repeat is present three times in ORF-A and the predicted amino acids are underlined.

XZ-6E DNA contains a single M.*Cv*iRI gene which is unique to this virus

The M.*Cv*iRI gene has a single *Pvu*I site 724 bp from the ATG start site and no internal *Bam*HI or *Sty*I sites (Fig. 2B). When Southern transfers of XZ-6E DNA digested with these enzymes were hybridized with the pXZ-6E.14 *Eco*RV 927 bp fragment, two *Pvu*I fragments (Fig. 4, lane 1) and single *Bam*HI (Fig. 4, lane 2) and *Sty*I (Fig. 4, lane 4) fragments were identified as expected. Additionally, two fragments were present in double digests with *Pvu*I plus one of the other two restriction

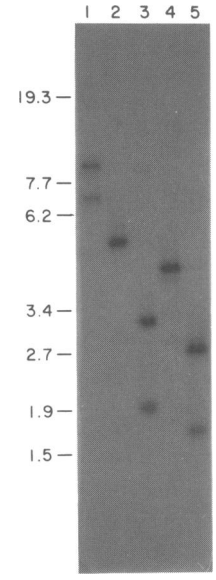

Figure 4. Hybridization of the pXZ-6E.14 *Eco*RV 927 bp restriction fragment to virus XZ-6E DNA. XZ-6E DNA was digested with *Pvu*I, *Bam*HI, *Pvu*I/*Bam*HI, *Sty*I, *Pvu*I/*Sty*I (lanes 1 through 5, repectively). Numbers on the left refer to sizes (kb) of lambda *Sty*I fragments.

endonucleases (Fig. 4, lanes 3 and 5). Thus viral XZ-6E contains a single copy of the M.*Cvi*RI gene.

Plasmid pXZ-6E.14 did not hybridize to dot blots of any of the other 36 *Chlorella* viral DNAs or to the host *Chlorella* DNA (data not shown). Thus the M.*Cvi*RI gene appears to be rare in the *Chlorella* viruses and the host alga does not contain the gene.

Comparison of M.*Cvi*RI to other DNA methyltransferases

The predicted amino acid sequence of M.*Cvi*RI was compared to the *Chlorella* virus cytosine methyltransferase M.*Cvi*JI (11) and to 15 adenine methyltransferases using ALIGN standard deviation scores (Table 1). There was no similarity between M.*Cvi*RI and M.*Cvi*JI. In contrast, M.*Cvi*RI shared similarities with the adenine methyltransferases. Most adenine methyltransferases, including M.*Cvi*RI, contain the sequence (asp/asn)-pro-pro-(tyr/phe) referred to as region IV by Hattman *et al.*, (25). The closest pairing was between M.*Cvi*RI and another *Chlorella* virus encoded adenine methyltransferase M.*Cvi*BIII, followed by the bacterial enzyme M.*Taq*I. M.*Cvi*BIII and M.*Taq*I are isoschizomers whose recognition sequence (TCGA) differs from M.*Cvi*RI in the central base pairs. M.*Pst*I, whose recognition sequence (CTGCAG) contains the subset TGCA, pairs more distantly to M.*Cvi*RI. When M.*Cvi*RI and M.*Cvi*BIII amino acid sequences were aligned, 142 of 347 amino acid residues matched (40% identity) (Fig. 5).

Table 1. Comparison of the M.*Cvi*RI amino acid sequence to fifteen adenine methyltransferases.

Gene	Comparison to M.CviRI[a]	Recognition Sequence	References
E. coli Dam	−0.27	G^{m}ATC	30
M.*Dpn*II	1.06	G^{m}ATC	31
T4 dam	0.19	G^{m}ATC	25, 32
M.*Eco*RV	0.68	G^{m}ATATC	33
M.*Fok*IB	−1.13	G^{m}ATCC	34
M.*Hinf*I	0.53	G^{m}ANTC	35
DpnA	2.69	G^{m}ANTC	36
M.*Rsr*I	−0.28	GA^{m}ATTC	37, 38
M.*Pst*I	1.19	CTGC^{m}AG	39
M.*Pae*R7	2.83	CTCG^{m}AG	40
M.*Taq*I	9.50	TCG^{m}A	26
M.*Cvi*BIII	26.39	TCG^{m}A	10
M.*Eco*RI	0.89	GA^{m}ATTC	41, 42
HsdM	0.65	A^{m}ACN6GTGC	43
M.*Fok*IA	0.72	GG^{m}ATG	34

[a]The comparison was conducted using the NBRF.ALIGN (44) computer program. ALIGN standard deviation scores from available fully sequenced N6-adenine methyltransferases were determined using a gap penalty of 6 and 100 random runs.

```
          1     5     10    15    20    25    30    35    40    45    50    55    60
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII MNVEQYTQVTTDFEKTLTKEKKSKQGIFFTPKTVREKLFGFTEHFQNTPGFSILEPSCGT
M.CviRI   MKL--FEDKSVEFHKRLSKKERSDGGVFFTPKDIRDIVFEELGDFEPT---NILEPTCGT
Common    M                F K L K    S  G FFTPK  R    F       F  T      ILEP CGT

          61    65    70    75    80    85    90    95    100   105   110   115   120
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII GEIISECVERFPLASIKGVELD---NDMSTICSKKYAEYNVDIVNEDFLLWKGG-KFDFI
M.CviRI   GEFISDCRKVYKNSRIIGVEIDPRSAELARDGSKN------EIIVHDFMTWDTDEKFDLI
Common    GE IS C           I GVE D             SK          I   DF  W     KFD I

          121   125   130   135   140   145   150   155   160   165   170   175   180
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII VGNPPYVVRPSGYKNDNRIAKGRSNLYVEFLYKCITEHLKEDGILAFIIPSTIGNSSFYE
M.CviRI   IGNPPYFTRPTGFKHDPSVVKCRSNICIEVLHKCITRHLADNGMLAMVLPVSILNSKFYT
Common     GNPPY  RP G K D     K RSN   E L KCIT HL   G LA   P  I NS FY

          181   185   190   195   200   205   210   215   220   225   230   235   240
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII PIRKLII-TLDILSFEILDKHDFCDTNTRLGSIVI-KNSPG-TGKYTYRDYICDKDIPHH
M.CviRI   PTIDLITDTMDVVSARAIKKNNFMGTNVRVMVFIIRKRTPGFVSKYTFKTSL-GKVIINP
Common    P    LI  T D  S       K  F  TN R       I K  PG   KYT         K I

          241   245   250   255   260   265   270   275   280   285   290   295   300
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII ----------GNSYIGSLDLKFKTGFAWANVNKFFTDKS---EIPF--FTSSNIKLNEIH
M.CviRI   DGERLGSIVSGKKTIGSLNVNISFGVTLASVKEYFVDKSCSGSFPLICYN---------N
Common                 G    IGSL       G   A V   F DKS       P

          301   305   310   315   320   325   330   335   340   345   350   355   360
          '     '     '     '     '     '     '     '     '     '     '     '     '
M.CviBIII IGDK--MKYLTQDTTKF----FTGKALLIKTASAGKRG---GRFEFGFSLYENDKWAVDN
M.CviRI   IAKKGDLLFVSD---KYSKKRFNGRAILIP------RGYAHGDYSFNFIDYTNDYFIIEN
Common    I  K               K       F G A LI         RG   G    F F  Y ND       N

          361   365   370   375   380   385   390   395   400   405   410
          '     '     '     '     '     '     '     '     '     '     '
M.CviBIII DIIVIQGPDTVLSIVQDVLMKDVTNEFINILVNNGHISMKLLKSIPLF-
M.CviRI   HVIAITGEDCVLDIIAKSFADHRTREFCRLLCSSGDISKDYVKEIPVFG
Common      I I G D VL I            T EF   L   G IS     K IP F
```

Figure 5. Maximal alignment between the amino acid sequences for M.*Cvi*BIII and M.*Cvi*RI using the NBRF computer program ALIGN (44).

DISCUSSION

The gene encoding the adenine methyltransferase, M.*Cv*iRI, from *Chlorella* virus XZ-6E has been cloned and expressed in *E. coli*. M.*Cv*iRI is one component of at least two separate restriction-modification systems induced by XZ-6E infection of *Chlorella* NC64A. The virus infected cells contain at least two site-specific endonucleases, *Cv*iRI (recognition site TG/CA) and *Cv*iRII (recognition site G/TAC) (Traylor, Xia, Zhang, Roy, and Van Etten, manuscript in preparation). Since virus XZ-6E DNA contains 21.2% 5mC and 15.2% 6mA, the virus obviously codes additional DNA methyltransferases besides M.*Cv*iRI and presumably M.*Cv*iRII.

The predicted amino acid sequence of M.*Cv*iRI had no similarity to the previously described *Chlorella* virus cytosine methyltransferase M.*Cv*iJI (11). This is not surprising since the lack of similarity between bacterial adenine and cytosine methyltransferases has been noted previously (e.g. 26, 27, but see 28, 29 for some evidence of common motifs). In contrast, the predicted amino acid sequence of M.*Cv*iRI resembles adenine methyltransferases; M.*Cv*iRI contains the sequence (asp/asn)-pro-pro-(tyr/phe) which is present in all adenine methyltransferases. M.*Cv*iRI was most similar to the *Chlorella* virus adenine methyltransferase M.*Cv*iBIII (10); 40% of the amino acids were identical. The recognition sequence of M.*Cv*iRI and M.*Cv*iRII differ only in their central base pairs. Because of their high degree of similarity, the construction of M.*Cv*iRI-M.*Cv*iBIII fusion proteins, if successful, could help to reveal domains important for site-specificity in adenine methyltransferase-DNA interactions.

ACKNOWLEDGMENTS

We thank Dwight Burbank for technical assistance and Les Lane, Roy French and Peter Strasser for helpful discussions. Jack Benner (New England Biolabs) kindly provided the amino acid sequences of the bacterial adenine methyltransferases. This investigation was supported, in part, by Public Health Service grant GM-32441 from the National Institute of General Medicine and grant DE-ACO2-82ER12086 from the Department of Energy. This manuscript has been assigned Journal Series No. 9393, Agriculture Research Division, University of Nebraska.

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. Zhang, Y., Burbank, D.E., and Van Etten, J.L. (1988) Appl. Environ. Microbiol. **54**, 2170–2173.
4. Xia, Y. and Van Etten, J.L. (1986) Mol. Cell. Biol. **6**, 1440–1445.
5. Xia, Y., Burbank, D.E., Uher, L., Rabussay, D., and Van Etten, J.L. (1986) Mol. Cell. Biol. **6**, 1430–1439.
6. Xia, Y., Burbank, D.E., and Van Etten, J.L. (1986) Nucleic Acids Res. **14**, 6017–6030.
7. Xia, Y., Burbank, D.E., Uher, L., Rabussay, D., and Van Etten, J.L. (1987) Nucleic Acids Res. **15**, 6075–6090.
8. Xia, Y., Narva, K.E., and Van Etten, J.L. (1987) Nucleic Acids Res. **15**, 10063.
9. Xia, Y., Morgan, R., Schildkraut, I., and Van Etten, J.L. (1988) Nucleic Acids Res. **16**, 9477–9487.
10. Narva, K.E., Wendell, D.L., Skrdla, M.P., and Van Etten, J.L. (1987) Nucleic Acids Res. **15**, 9807–9823.
11. Shields, S.L., Burbank, D.E., Grabherr, R., and Van Etten, J.L. (1990) Virology **176**, 16–24.
12. Van Etten, J.L., Burbank, D.E., Xia, Y., and Meints, R.H. (1983) Virology **126**, 117–125.
13. Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J., and Hanahan, D. (1988) Nucleic Acids Res. **16**, 1563–1575.
14. Raleigh, E.A., Trimarchi, R., and Revel, H. (1989) Genetics **122**, 279–296.
15. Van Etten, J.L., Meints, R.H., Burbank, D.E., Kuczmarski, D., Cuppels, D.A., and Lane, L.C. (1981) Virology **113**, 704–711.
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd edition). Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Mann, M.B., Rao, R.N., and Smith, H.O. (1978) Gene **3**, 97–112.
18. Nelson, M., and McClelland, M. (1987) Nucleic Acids Res. **15**, r219-r230.
19. Lynn, S.P., Cohen, L.K., Kaplan, S., and Gardner, J.F. (1980) J. Bacteriol. **142**, 380–383.
20. Ross, P.M., Woodley, K., and Baird, M. (1989) BioTechniques **7**, 680–690.
21. Messing, J. (1983) Methods Enzymol. **101**, 20–78, Academic Press, New York.
22. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
23. Tabor, S., and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA **84**, 4767–4771.
24. Schuster, A.M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D., and Meints, R.H. (1990) Virology **176**, 515–523.
25. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., MacDonald, P.M., and Mosig, G. (1985) J. Bacteriol. **164**, 932–937.
26. Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M., and Wilson, G.G. (1987) Nucleic Acids Res. **15**, 9781–9796.
27. Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989) Nucleic Acids Res. **17**, 2421–2435.
28. Lauster, R. (1989) J. Mol. Biol. **206**, 313–321.
29. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. (1989) Nucleic Acids Res. **17**, 9823–9832.
30. Brooks, J.E., Blumenthal, R.M., and Gingeras, R.T. (1983) Nucleic Acids Res. **11**, 837–851.
31. Mannarelli, B.M., Balganesh, T.S., Greenberg, B., Springhorn, S.S., and Lacks, S.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 4468–4472.
32. MacDonald, P.M., and Mosig, G. (1984) EMBO J. **3**, 2863–2871.
33. Bougueleret, L., Schwarzstein, M., Tsugita, A., and Zabeau, M. (1984) Nucleic Acids Res. **12**, 3659–3676.
34. Kita, K., Kotani, H., Sugisaki, H., and Takanami, M. (1989) J. Biol. Chem. **264**, 5751–5756.
35. Chandrasegaran, S., Lunnen, K.D., Smith, H.O., and Wilson, G.G. (1988) Gene **70**, 387–392.
36. Lacks, S.A., Manarelli, B.M., Springhorn, S.S., and Greenberg, B. (1986) Cell **46**, 993–1000.
37. Kaszubska, W., Aiken, C., O'Connor, D., and Gumport, R.I. (1989) Nucleic Acids Res. **17**, 10403–10425.
38. Stephenson, F.H., and Greene, P.J. (1989) Nucleic Acids Res. **17** 10503.
39. Walder, R.Y., Walder, J.A., and Donelson, J.E. (1984) J. Biol. Chem. **259**, 8015–8026.
40. Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F., and Gingeras, T.R. (1985) Nucleic Acids Res. **13**, 8441–8461.
41. Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E., and Rosenberg, J.M. (1981) J. Biol. Chem. **256**, 2143–2153.
42. Newman, A.K., Rubin, R.A., Kim, S.-H., and Modrich, P. (1981) J. Biol. Chem. **256**, 2131–2139.
43. Loenen, W.A.M., Daniel, A.S., Braymer, H.D., and Murray, N.E. (1987) J. Mol. Biol. **198**, 159–170.
44. Dayhoff, M.O., Barker, W.C., and Hunt, L.T. (1983) Methods Enzymol. **91**, 524–545.