Identification and characterization of the frog virus 3 DNA methyltransferase gene

Kuljeet Kaur,† Jan Rohozinski and Rakesh Goorha*

Department of Virology and Molecular Biology, St Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, USA

Cytosine DNA methyltransferases (MTases) first recognize specific nucleotide sequences and then transfer a methyl group from S-adenosylmethionine to cytosine. This division of function is reflected in five highly conserved motifs shared by cytosine MTases. The region containing the first four motifs is responsible for the catalytic function whereas the region containing the fifth motif V provides specificity of binding to DNA. In at least one case, two separate proteins, one containing the first four motifs and the second containing the last motif combine to provide full functional activity. In the frog virus 3 (FV3) genome we have identified an open reading frame (ORF) whose deduced amino acid (aa) sequence contains motifs characteristic of prokaryotic as well as eukaryotic MTases. The ORF consists of 642 bp which codes for a protein of 214 aa with a predicted molecular mass of 24.8 kDa. This ORF contains the first four highly conserved motifs of cytosine MTases but the fifth motif, responsible for DNA binding specificity, is missing. Presumably, FV3 MTase is composed of two subunits. Northern blot analysis showed that the putative MTase ORF is transcribed into two transcripts belonging to the delayed-early class of FV3 messages. These two transcripts appear to be initiated at two different start sites but terminate in the same 3′ region of the gene. The transcription start sites are not preceded by any known promoter sequences, but two regions of hyphenated dyad symmetry are present at the 3′ end of the message. A protein with a molecular mass of ~28 kDa was synthesized by a rabbit reticulocyte lysate programmed with capped runoff transcripts from the cloned gene, suggesting that the ORF can be transcribed into a message coding for a viral protein. Overall, our results suggest that we have identified a gene for a subunit of MTase in the FV3 genome.

Introduction

DNA methyltransferases (MTases) recognize specific nucleotide sequences and transfer the methyl group from S-adenosylmethionine to cytosine or adenine. In higher eukaryotes, methylation of the genome is mostly restricted to the C-5 position of cytosine in the dinucleotide sequence CpG though cytosine methylation also occurs in the dinucleotides CpA and CpT in vertebrates and trinucleotides CNG in plants. The precise function of DNA methylation is not known. Several studies have suggested that in eukaryotic cells methylation of DNA inhibits gene expression (reviewed in Bestor, 1990).

Frog virus 3 (FV3), genus Ranavirus of the family Iridoviridae, is an icosahedral virus with a linear, circularly permuted and terminally redundant (Goorha & Murti, 1982), double-stranded DNA genome of ~170 kbp in size (Murti et al., 1985). About 20% of the cytosine residues in FV3 DNA are methylated at the CpG sequence (Willis & Granoff, 1980). This is the highest degree of methylation known among animal DNA viruses. Our studies have shown that FV3 induces novel DNA methylation activity in infected cells (Willis et al., 1984). Available evidence also suggests that the MTase is specified by the FV3 genome since a mutant strain of FV3 failed to induce MTase activity (Essani et al., 1987). MTase activity was localized into the cytoplasmic fraction and FV3 DNA is also predominantly methylated in the cytoplasm (Willis et al., 1984). The precise role of DNA methylation in FV3

* Author for correspondence. Fax +1 901 523 2622.
† Present address: Department of Biochemistry, University of Tennessee at Memphis, The Health Science Center, 858 Madison Avenue, Memphis, TN 38163, USA.

The nucleotide sequence data in this paper have been assigned accession number U15575 in the GenBank database.
replication is not known. However, it appears to be required for production of genomic size DNA during FV3 replication (Goorha et al., 1984). In this paper we report identification and partial characterization of the FV3 specified putative DNA MTase gene. The FV3 specified enzyme has a deduced molecular mass of 248 kDa and it has four of the five highly conserved motifs of prokaryotic DNA MTases.

Methods

Virus and cells. Fathead minnow (FHM) cells were propagated as monolayers in tissue culture dishes in Eagle’s MEM supplemented with 5% fetal calf serum (FCS) at 33 °C. A clonal isolate of FV3 was grown in FHM, and purified and assayed as previously described (Naegle & Granoff, 1971). This isolate was used to infect FHM cells at an m.o.i. of 10 p.f.u./cell for all experiments described in this paper.

Bacterial strains and plasmids. A 2.4-kb XbaI K fragment of the FV3 genome was cloned in pBS M13+ (Stratagene). The nucleotide sequence was established using oligonucleotide primers in the dideoxy chain termination method of Sanger employing Sequenase polymerase (USB). The putative MTase gene of 1.1-kb length from this fragment was identified and used for the experiments described in this paper. For in vitro transcription and translation studies, the FV3 MTase ORF was cloned under the T7 promoter in plasmid pTMI (Moss et al., 1990). This construct was named pTMImet and was constructed by introducing suitable EcoRI and BamHI restriction sites in the FV3 MTase gene using PCR.

All plasmids constructed using PCR were sequenced to check orientation, reading frame and correct sequence. Plasmid isolation and other standard molecular biology techniques were done following Sambrook et al. (1989).

Northern blots. FHM cells infected with FV3 (10 p.f.u./cell) were used for extraction of FV3 RNA. To prepare RNA from different classes of FV3 RNA species, FHM cells were infected in the presence of cycloheximide or fluorophenylalanine (Sigma) at a concentration of 100 gg/ml. After infection for 6 h cells were harvested and total RNA was extracted using the guanidine isothiocyanate procedure (Sambrook et al., 1989). FV3 RNA was electrophoresed through a 1.2% agarose gel containing 2.2 M-formaldehyde and 6.6 ng ethidium bromide/ml.

In vitro RNA transcription. The pTMImet plasmid was cleaved with ClaI and BamHI to isolate the MTase ORF with the T7 promoter. This fragment was purified and used for in vitro transcription. Transcription reaction buffer contained 40 mM-Tris–HCl pH 7.5, 6 mM-MgCl2, 2 mM-spermidine, 10 mM-NaCl, 10 mM-DDT, 100 gg/ml acetylated BSA, 1 U/μl ribonuclease inhibitor, 0.5 mM each ATP, CTP, UTP and 0.05 mM-GTP, 0.5 mM-GpppG and 1 U/μl T7 polymerase. The GpppG (NEB) cap analogue was used to increase the stability of the transcript. The reaction mixture was incubated at 37 °C for 60 min. Reaction products were digested with DNase I at 37 °C for 5 min to remove the DNA template and the resulting RNA was extracted with phenol–chloroform–isoamyl alcohol. The resultant RNA preparation was ethanol precipitated and quantified using spectrophotometric analysis.

In vitro translation. FV3 MTase mRNA prepared as described above was used in the translation reaction. Translation was done using the nuclease treated rabbit reticulocyte lysate system (Promega) following instructions provided by the supplier. The translated product was labelled using [35S]methionine. A control reaction was done under the same conditions but without MTase RNA to check for background. Samples were subjected to electrophoresis on 12% SDS–PAGE (100 V, 90 min). After electrophoresis the gel was dried and exposed on Kodak XAR5 film.

Data analysis. Sequence analysis and databank searches were done on a Vax computer using the IntelliGenetics Suite version 5.4.

Results

Identification of the FV3 putative MTase gene

As part of our ongoing search for genes involved in viral DNA replication, the ~2.4-kbp long XbaI K fragment (Fig. 1) was cloned into the multiple cloning site of plasmid pBS M13+ and sequenced in both directions. Four ORFs were identified. The second ORF started at position 1021 with ATG and ended at position 1662 with TGA. This ORF had the characteristic features of other known FV3 genes (Rohozinski & Goorha, 1992). It contained two dyad symmetry sequences at the 3’ end between nucleotides 1701 to 1782, lacked intron sequences and contained no polyadenylation sites (Fig. 2).

The SwissProt and PIR databanks were searched to identify any similarity of the deduced amino acid sequence of the FV3 ORF to any other known protein. The search showed strong amino acid homologies between the protein encoded by the FV3 ORF and prokaryotic as well as eukaryotic cytosine MTases. Cytosine MTases possess ten conserved motifs. However, only five of these motifs are highly conserved whereas the other five possess weak homologies (Posfai et al., 1989). We therefore compared the sequence homologies of the five highly conserved motifs of cytosine MTases (I, IV, VI, VIII and X) with the FV3 ORF (Fig. 3). The FV3 ORF not only possesses sequence similarities to the first four motifs but the order as well as the spacing between these motifs has been maintained (Fig. 3). Motif X is missing from the FV3 ORF. In fact, the FV3 gene has a termination codon soon after motif VIII. Consequently, the deduced molecular mass of the protein encoded by the FV3 ORF is much smaller than that of other known cytosine MTases. The FV3 ORF did not show significant homology to any other gene in the databank.

We have previously described an azacytidine resistant FV3 mutant that lacks MTase activity (Essani et al., 1987). Cells infected with this mutant, in contrast to wild-type FV3, do not exhibit MTase activity and newly synthesized viral DNA remains unmethylated (Essani et al., 1987). We have now sequenced the putative MTase gene of this mutant and found a single base substitution (cytosine to guanosine) at nucleotide 1071 (data not
shown). This mutation changed a cysteine residue to tryptophan in conserved motif I. Based on strong sequence homologies and the azacytidine resistant mutant data, this FV3 ORF was designated as that encoding the cytosine MTase. When the FV3 MTase protein sequence was aligned and compared to other known cytosine MTases, it was observed that the FV3 MTase has greater homology to bacterial enzymes, particularly to phage Phi3T, than to eukaryotic cytosine MTases. For comparison, we have also included the sequence of cytosine MTase from Chlorella virus. Chlorella virus grows in algae and is the only known eukaryotic virus whose MTase gene has been sequenced (Shields et al., 1990). The comparison revealed that the homology between FV3 and Chlorella virus MTases is less than with phage Phi3T. The deduced molecular mass of the putative MTase was 24.8 kDa.

The FV3 MTase is a delayed early gene. To demonstrate that the putative MTase ORF is transcribed into an authentic FV3 transcript, we performed Northern blot analysis of RNA from FV3 infected cells. FV3 RNA synthesis is under temporal control. FV3 transcripts are categorized as immediate early, delayed early and late messages. RNA from each of these stages can be identified using various drugs during FV3 infection (Willis et al., 1984). Using the protein synthesis inhibitor cycloheximide, only immediate early messages are synthesized since they do not require de novo synthesis of viral proteins. In the presence of fluorophenylalanine only immediate early and delayed early messages are synthesized; no late messages are synthesized. Cells infected in the absence of any drug will synthesize all classes of transcripts. Total RNA extracted from FV3 infected cells, under various conditions, was analysed to study FV3 MTase gene regulation. The ORF region of the putative MTase gene was used as a probe for Northern blot analysis (Fig. 4). No FV3 transcript was detected in FV3 infected cells treated with cycloheximide. Two transcripts of ~900 bp and ~700 bp long were detected in untreated and fluorophenylalanine treated RNA preparations. Taken together, these results show that the putative MTase gene is transcribed and its mRNA belongs to the delayed early class of FV3 transcripts. To delineate the relationship between the two transcripts, we used sequences representing different regions of the XbaI K fragment as probes for Northern blot analysis (Fig. 4). Probe PR representing the 5' region of the ORF detected both 900 bp and 700 bp transcripts. However, probes IR and bax representing sequences upstream of the ORF detected only the larger ~900 bp transcript. Probes representing regions upstream of IR did not detect either of the two transcripts (data not shown). These results suggest that the 900 bp transcript originated ~200 bp upstream of the translation start site whereas the 700 bp transcript originated...
Fig. 2. Complete nucleotide sequence and deduced amino acid sequence of the putative MTase gene. The methionine initiation codon and the stop codon are indicated by the symbol ‘>’. In the 3' region, opposing arrows underline the dyad symmetry.

within 10 bp (the distance between the 3' end of probe box and the start site of the ORF initiating methionine codon). A probe representing 102 bp downstream of the termination codon of the ORF (probe 72, Fig. 4) detected both transcripts. Regions immediately downstream of probe 72 when used as a probe did not detect either of the two transcripts in Northern blot analysis (data not shown). These and other data (not shown) suggested that both transcripts terminate in the same 3' region of the MTase gene.

The putative MTase is an approximately 28 kDa protein. To demonstrate that the putative MTase RNA is actually translated into authentic protein, we performed in vitro transcription and translation using the template representing the ORF of the putative MTase gene. As described in Methods, the putative gene was cloned in the plasmid pTMI under control of the T7 promoter. This gene was transcribed and translated in vitro using rabbit reticulocyte lysate and T7 RNA polymerase. A single translation product of ~28 kDa was observed (Fig. 5). This molecular mass is more than that (24.8 kDa) theoretically calculated for the protein. The precise reason for the discrepancy between the calculated and observed molecular masses is not known but this difference may be due to anomalous behaviour of the protein in SDS-PAGE or to post-translational modifications such as phosphorylation of the translated product.
Cytosine DNA methyltransferases possess a common architecture: they contain ten domains or motifs in the same order in their amino acid sequences. Five of these motifs are highly conserved whereas the other five show weaker homologies (Posfai et al., 1989). In this paper, we have used only the five strongly conserved motifs for sequence comparison. Cytosine MTases perform two different functions: first, they recognize a specific nucleotide sequence in the genome and then transfer the methyl group from S-adenosylmethionine to cytosine. The structure of MTases reflects the division of these functions. Experimental evidence suggests that the first four conserved motifs (I, IV, VI and VIII), representing the N-terminal region of the protein, play a role in the enzymatic function whereas the last motif (X) is involved in recognition of the target nucleotide sequence in the genome (Posfai et al., 1989). These two regions of the protein are separated by a long variable region which is 80–120 aa for monospecific MTases and 200–300 aa for multispecific-specific MTases (Lauster et al., 1989; Lange et al., 1991).

Interestingly, the FV3 MTase gene shows strong homologies to the first four motifs but does not seem to possess the long variable region or the fifth motif. The genomic region of the cytosine MTase of Agmenellum quadruplicatum is organized in two parallel overlapping ORFs of 248 and 139 aa, respectively (Kareem & Waard, 1990). The larger ORF contains the first four highly conserved motifs whereas the shorter ORF contains the long variable region and the highly conserved motif X. In vitro complementation data showed that the presence of both proteins is required for
methylase activity. Our data also indicate that two viral proteins of molecular masses 26 kDa and 18 kDa are required for cytosine MTase activity (Essani et al., 1988). It is very likely that, similar to the *A. quadriplicatum* enzyme, the gene identified by us encodes a subunit of the cytosine MTase that carries out the enzymatic function of the protein. We have not yet identified the gene required for the recognition of target DNA sequences but the regions adjacent to the putative cytosine MTase do not contain any ORF that has homology to the highly conserved motif X. Therefore, unlike *A. quadriplicatum*, in the FV3 genome the two ORFs that encode cytosine MTase activity are not overlapping—presumably the ORF that codes for the subunit that recognizes the target DNA sequences is present elsewhere in the genome.

We have previously demonstrated that an MTase−FV3 mutant shows anomalous migration of a 26 kDa protein upon PAGE analysis of the infected cells (Essani et al., 1987). Nucleotide sequence analysis of the putative MTase gene of this MTase− mutant showed a single base substitution that changed a cysteine amino acid residue to tryptophan in the first highly conserved motif. The reasons for the discrepancy in the molecular mass of this protein between the previous report (Essani et al., 1987) and the current paper are not known. This discrepancy may be due to differences in the electrophoretic conditions and/or differences in the post-translational modification of the protein in the infected cells (Essani et al., 1987) and in the *in vitro* translation system (current study). Overall, our results strongly support the notion that we have identified a gene necessary for the FV3 MTase activity.

The putative MTase gene has characteristic features shared by other FV3 genes: it contains no introns and lacks the nucleotide sequence responsible for polyadenylation. The region upstream of the MTase gene does not have any known promoter sequences. This lack of typical promoter organization has been observed in all FV3 genes so far sequenced (Rohozinski & Goorha, 1992). Sequences upstream of FV3 ORFs do function as promoters but only in FV3 infected cells suggesting that FV3 promoters are unique and that they require viral protein(s) for their expression (Willis & Granoff, 1985). Similar to other FV3 genes, the putative MTase gene has two dyad symmetries downstream of the termination codon between nucleotides 1701 and 1781 (Fig. 2). The function(s) of dyad symmetries in FV3 genes is not known but they may play a role in transcription termination.

Two transcripts were detected by Northern blot analysis when the ORF of the MTase gene was used as a probe. Further analysis indicated that both transcripts terminate in the same region of the gene but the 5' ends of these transcripts differ: the larger transcript contains ~200 bp that are missing in the shorter transcript. Known FV3 genes, including the MTase gene, are intronless and thus require no splicing of the primary transcripts. Furthermore, we have found no evidence of processing such as a precursor–product relationship between the two transcripts under a variety of experimental conditions (data not shown). Therefore, the most likely possibility is that these two transcripts have different transcription initiation sites.

Methylation of DNA is a ubiquitous phenomenon in prokaryotic as well as eukaryotic cells (reviewed in Bird, 1992). In prokaryotes, DNA methylation is the basis for the restriction–modification phenomenon whereby certain strains of bacteria recognize and degrade foreign DNA (reviewed in Wilson & Murray, 1991). In vertebrates, the functional importance of DNA methylation remains largely unknown. However, in FV3, DNA methylation protects the viral genome from endonucleolytic cleavage during viral DNA replication and
allows the production of mature progeny DNA (Goorha et al., 1984).

The size of FV3 MTase is quite different from other known cytosine MTases even though the conserved motifs have been retained in the FV3 gene. The size of the FV3 MTase is much smaller, partly because it appears that FV3 MTase activity, similar to A. quadruplicatum, is encoded by two genes. Thus, FV3 protein is missing amino acids that make up highly conserved motif X, as well as an 80–120 aa stretch of the variable region. The mammalian cytosine MTase has a 500 aa C-terminal domain that possesses strong homology to bacterial cytosine MTases and a ~1000 aa N-terminal domain that presumably has a regulatory function (Bestor et al., 1988). The first 300 aa of the N-terminal domain can be proteolytically removed without affecting enzyme activity under in vitro conditions. The N-terminal domain also contains a nuclear localization signal sequence and a cysteine rich region that binds zinc ions (Bestor, 1992). The mammalian enzyme has no de novo methylation activity but efficiently methylates hemi-methylated DNA (Bestor & Ingram, 1983). The FV3 MTase lacks this ~1000 aa regulatory domain and it can efficiently perform de novo methylation of DNA under in vitro conditions (Willis et al., 1984). It appears that the FV3 MTase gene contains minimum core enzyme activity with little or no regulatory domains. Further studies on FV3 MTase enzyme activity may provide a simpler model to elucidate structure–function relationships involved in the enzymatic function of this protein.

We thank Ramona Tirey and Josie Harris-Chambers for excellent technical assistance and Glenith Newberry for typing the manuscript. This work was supported by Public Health Service research grant GM-23638 from the National Institute of General Medical Sciences, Cancer Center Support (CORE) grant CA-21765 from the National Cancer Institute, and the American Lebanese Syrian Associated Charities (ALSAC).

References


(Received 8 November 1994; Accepted 18 April 1995)