Nucleotide and predicted amino acid sequences of Marek's disease virus homologues of herpes simplex virus major tegument proteins

Noboru Yanagida,† Shigeto Yoshida,† Keyvan Nazerian and Lucy F. Lee*

United States Department of Agriculture – Agricultural Research Service, Avian Disease and Oncology Laboratory, 3606 East Mount Hope Road, East Lansing, Michigan 48823, U.S.A.

The DNA sequence of an 8.4-kbp BamHI–EcoRI fragment of Marek's disease virus (MDV) strain GA was determined. Three of the predicted polypeptides are homologous to UL47, UL48 and UL49 encoding the major tegument proteins of herpes simplex virus type 1 (HSV-1), and four are homologous to HSV-1 UL45, UL46, UL49.5 and UL50. These seven genes are found in the long unique region of the MDV genome and are collinear with homologues in HSV-1 and varicella-zoster virus (VZV). Northern blot analysis revealed different transcriptional patterns from those of HSV-1 and VZV. MDV homologues of UL49.5, UL49 and UL47 lack a poly(A) signal immediately downstream of their coding regions. Amino acid conservation between MDV and HSV-1, and between MDV and VZV is as high as that between HSV-1 and VZV. The MDV homologue of UL48 shows 60% similarity to its HSV-1 counterpart. Amino acid sequence comparison reveals that the MDV homologue of UL48 lacks an acidic carboxyl terminus. This homologue, like the VZV homologue of UL48, may be involved in the trans-activation of immediate early genes and may function as an important component of the structural proteins.

Introduction

Marek's disease virus (MDV) is the causative agent of a malignant lymphoma in chickens. MDV was initially classified as a gammaherpesvirus based on its tropism for lymphocytes. However, recent sequence data and studies of its genomic arrangement show it to be more closely resembling an alpha-herpesvirus (Buckmaster et al., 1988). The genome structure of MDV is classified in group E and consists of long (U0) and short (Us) unique sequences flanked by inverted repeats (Roizman et al., 1992). The prototype alpha-herpesvirus with a group E genome structure is herpes simplex virus type 1 (HSV-1).

Regulation of herpesvirus gene expression has been most extensively studied in the case of HSV-1. HSV-1 has three distinct classes of genes, as immediate early (IE), early and late, which are expressed in a cascade fashion (Honess & Roizman, 1974). The five IE genes of HSV-1 are trans-activated by the virion tegument protein VP16 or α-trans-inducing factor (α-TIF), the gene product of UL48 (Campbell et al., 1984; Preston et al., 1984). At least three of these IE proteins are important trans-activators required for maximal expression of early and late genes (O'Hare & Hayward, 1985). In addition to VP16, VP13/14 (the gene product of UL47) and the gene product of HSV-1 UL46 were initially implicated as modulators of the α-TIF trans-activation in transient expression systems (McKnight et al., 1987). A more recent study reported that deletion of UL46, within the context of the virus, had no apparent effect on α-TIF trans-activation but deletion of UL47 had a significant effect (Zhang et al., 1991).

VP16 is important not only as an IE trans-activator but also as a major structural component of the virion. It is found in the tegument region of the virions between the virus envelope and capsid. Other tegument proteins include VP1/2, VP13, VP14 and VP22. The tegument proteins are abundant in the virion and each contributes more than 5% of the mass of the virus particle (Spear & Roizman, 1972; Honess & Roizman, 1973). Recently, the HSV-1 tegument protein VP22 was ascribed to the UL49 gene (Elliott & Meredith, 1992). Barker & Roizman (1992) and Barnett et al. (1992) reported another HSV-1 gene designated UL49.5 located between the genes UL49 and UL50.

Recent studies suggest that VP16 may be essential for virus growth. Virion assembly was shown to be affected...
in an HSV-1 mutant lacking the complete sequence of VP16. This mutant can replicate only in a cell line transformed with a VP16 expression vector (Weinheimer et al., 1992) but produced a lethal phenotype that correlates with a defect in virion assembly. Another mutant of HSV-1 which has an in-frame insertion of four amino acids within VP16 failed to trans-activate IE genes but is capable of growth and resulted in avirulence in mice (Ace et al., 1989). In this paper we report the sequence of the 8.4 kb BamHI-EcoRI subfragment of the BamHI B fragment (Fukuchi et al., 1984) of MDV strain GA and the identification of seven MDV genes. We found the gene arrangement in this region of MDV to be strictly collinear with that found in HSV-1 and choose therefore to name these seven genes after their HSV-1 counterparts. These seven genes have sequences homologous to the UL50, UL49.5, UL49, UL48, UL47, UL46 and UL45 genes of HSV-1 and will be designated ‘h’ for homologue, after the HSV-1 names.

Methods

Viruses and cell culture. The preparation, propagation, and infection of duck embryo fibroblast cell cultures with MDV strain GA (Eidson & Schmitte, 1968) were as previously described (Solomon, 1975).

DNA sequencing. The 8.4 kb BamHI–EcoRI fragment was subcloned from a plasmid containing the BamHI B fragment of MDV strain GA (Fukuchi et al., 1984) into pUC18. The sequencing strategy employed a combination of sequential deletions, using exonuclease III and mung bean nuclease (Henikoff, 1984) and subclones were constructed using appropriate restriction sites. Both strands of the entire 8.4 kb fragment were sequenced by the dideoxynucleotide chain termination method using [32P]dATP (NEN) and Sequenase kits (United States Biochemical Corporation) as suggested by the manufacturer.

Northern blot hybridization. Total cellular RNAs were prepared from duck embryo fibroblast cells infected with MDV strain GA using the guanidine thiocyanate method and analysed on formaldehyde-containing gels followed by transfer onto nylon filters (Sambrook et al., 1989). Four DNA probes, representing open reading frames (ORFs) UL49h, UL48h, UL47h and UL46h, were cut with restriction enzyme or amplified using PCR, and labelled with Prime-It Random Primer labelling kits (United States Biochemical Corporation) as suggested by the manufacturer.

Computer analysis. Homology searches against GenBank release 69.0 (IntelliGenetics), the Protein Identification Resource version 29.0 (National Biomedical Research Foundation), protein–protein homology matrices, and hydrophobicity plots were performed using MacVector (International Biotechnologies). Final alignment of amino acid sequences and calculation of pairwise percentage identities and similarities were performed using the programs GAP and PILEUP of the University of Wisconsin Genetics Computer Group (Devereux et al., 1984) with a mutation matrix reported by Gonnet et al. (1992).

Results and Discussion

Nucleotide sequence of the 8.4 kb BamHI–EcoRI fragment

Based on the close collinearity between MDV and HSV-1 genomes, we estimated MDV genes encoding tegument proteins to be located near the gCh gene which has been mapped in the 18.3 kb BamHI B fragment in the UL region of the MDV genome (Coussens & Velicer, 1988). Accordingly, an 8.4 kb BamHI–EcoRI subfragment of the BamHI B fragment from MDV strain GA was sequenced (Fig. 1). Sequential deletions of the cloned MDV DNA and subclones constructed using appropriate restriction sites were used to sequence the entire fragment in both strands. Six major (encoding more than 200 amino acid residues) and four minor (encoding 90 to 200 amino acid residues) ORFs were identified within the region (Fig. 1 and Fig. 2). Seven of these 10 ORFs, UL50h, UL49.5h, UL49h, UL48h, UL47h, UL46h and UL45h, are homologous to HSV-1 UL50, UL49.5, UL49, UL48, UL47, UL46 and UL45, respectively. The precise locations of the MDV ORFs are listed in Table 1. A 1.3 kb subfragment contiguous to the EcoRI site and comprising UL45h and an ORF of 208 amino acids (ORF 208 in Fig. 1), which has no significant homology to known peptides, was found to be identical in sequence to the fragment previously reported for MDV strains Md5 and BCI (Ihara et al., 1989). Two of the minor ORFs (ORFs 108 and 91), which overlap major ORFs,

![Fig. 1. Genome structure of MDV showing the location of the BamHI B fragment. Upper line shows the genome structure of MDV: hatched boxes indicate terminal and internal repeat regions; bold line shows the BamHI B fragment with restriction sites for BamHI (B) and EcoRI (E); arrows show three homologues of glycoproteins gB, gC and gD (Ross & Binns, 1991). Lower line shows the position and orientation of the ORFs within the 8.4 kb BamHI–EcoRI fragment; bold arrows show the ORFs of HSV-1 homologues; thin arrows show other ORFs.](image-url)
were found to have no significant homology to known peptides. Another minor ORF, which potentially encodes a 95-residue polypeptide, shows a homology of 20% identity and 42% similarity to UL49.5 of HSV-1 (Barker & Roizman, 1992) and 35% identity and 61% similarity to the counterpart ORF (ORF 8.5) of VZV. The fourth and last minor ORF continues beyond the BamHI site. It shows a weak homology to UL50 of HSV-1 (Preston & Fisher, 1984) and ORF 8 of VZV. Each of the seven genes has one or more putative promoter regions. However, putative poly(A) signals are missing between UL49.5h and UL49h, UL49h and UL48h, and UL47h and UL46h. The lack of a poly(A) signal between the UL49.5h and UL49h genes and between the UL47h and UL46h genes in MDV is a feature shared with HSV-1.

Approximately 90 bp upstream from the first major ORF, UL49h, a CAAT box (RRCCAAT; where R is a purine) (Chodosh et al., 1988) was found (AACCAAT). However, there is no TATA-like sequence (Corden et al., 1980) about 30 bp downstream from this signal. It is known that many viral and cellular promoters transcribed in higher eukaryotes by RNA polymerase II lack obvious AT-rich sequences (Wiley et al., 1992; Singer et al., 1990). In addition, some AT-rich regions were present as possible promoters. Approximately 35 and 65 bp upstream from the second ORF (UL48h), two TATA sequences were found as possible promoter regions. An AATAAA sequence was found as a putative poly(A) signal approximately 80 bp downstream from the stop codon. At 30 bp downstream from this signal, a TrTGTtTC sequence was found with a similarity to the consensus sequence, YGTGTTYY (where Y is a pyrimidine), for efficient termination (McLauchlan et al., 1985). A potential promoter sequence (TATAAA) was found 50 bp upstream from the third ORF, UL47h. There are two CAAT box-like sequences (GATCAAT and cATCAAT) immediately and approximately 30 bp upstream of this putative promoter sequence. Approximately 45 bp upstream from the fourth ORF, UL46h, a TATA sequence was found. Putative poly(A) (AATAAA) and termination (TGTGTTTα) sequences were found approximately 100 and 130 bp downstream from this ORF, respectively.

Northern blot hybridization

Total cellular RNAs were prepared from cells infected with MDV strain GA. DNA probes representing four ORFs were prepared by standard procedures (Methods). As shown in Fig. 3(a), both UL49h and UL48h probes detect two major mRNA bands of 7.8 and 2.5 kb; the UL47h probe detects two bands of 7.8 and 4.6 kb; and the UL46h probe detects bands of 7.8, 4.6 and 2.0 kb. Fig. 3(b) illustrates our interpretation of these transcripts in accordance with the sequence data. The detection of the 7.8 kb transcript, which is not detected in mock-infected cells, by all four probes suggests that transcription might start either upstream of the UL49h gene or the UL49.5h gene and terminate after the UL46h gene, and that the poly(A) signal between UL48h and UL47h was not recognized. The 2.5 kb transcript in lanes 1 and 2 was abundant and its transcription might start at the same position as that of the 7.8 kb transcript and terminate after the UL48h gene. In the absence of S1 mapping data, we are unable to conclude whether the 7.8 kb and 2.5 kb transcripts start upstream of the UL49h gene or the UL49.5h gene, or both. The 4.6 kb transcript in lanes 3 and 4 might start upstream of the UL47h gene and terminate after the UL46h gene. The 2.0 kb transcript in lane 4 might start upstream of the UL46h gene and terminate at the UL46h gene. We did not detect a transcript smaller than 2.5 kb in lane 2, suggesting that the promoter for UL48h is much weaker than that for UL49h. Transcription patterns of this region are somewhat different in MDV, HSV-1 and VZV. In the case of HSV-1, UL49, UL48 and UL46 each have their own poly(A) signal and only UL47 shares one with UL46. In the case of VZV, ORFs 9, 10 and 11 (corresponding to UL49, UL48 and UL47 of HSV-1, respectively) have poly(A) signals immediately downstream of each gene coding region, and only ORF 12 (UL46 homologue of HSV-1) shares one with ORF 13. Minor bands, including a 2.5 kb mRNA in lanes 3 and 4, are not easily accounted for.

Analysis of the predicted peptides

Deduced amino acid sequences were compared with the GenBank database. The characteristics of these predicted peptides are listed in Table 1. Four of the six major ORFs show homology to UL49, UL48, UL47 and UL46 of HSV-1 and ORFs 9, 10, 11 and 12 of VZV, respectively. Of these four gene products MDV UL48h has the highest homology to both HSV-1 UL48 (36% identity and 60% similarity) and VZV ORF 10 (37% identity and 62% similarity) (Table 1). The three UL48 homologues have similar hydrophilicity patterns and Mr values (data not shown). Fig. 4 shows an alignment of UL48 homologues from VZV, MDV and HSV-1. High homology is observed throughout the entire region. However, both MDV and VZV lack the acidic carboxy-terminal region of VP16, which has been found to be important for trans-activation in HSV-1 (Triezenberg et al., 1988; Greaves & O’Hare, 1989). It has been reported that the gene product of VZV ORF 10, a homologue of VP16, fails to trans-activate HSV-1 or VZV IE gene
Table 1. Characteristics of MDV proteins

<table>
<thead>
<tr>
<th>Gene</th>
<th>Start* (nt)</th>
<th>Stop† (nt)</th>
<th>Codons</th>
<th>VZV/HSV-1 counterpart‡</th>
<th>Properties of HSV-1 proteins§</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL50h</td>
<td>305</td>
<td>&gt; 101</td>
<td>10159</td>
<td>ORF 8/UL50</td>
<td>dUTPase</td>
</tr>
<tr>
<td>UL49.5h</td>
<td>285</td>
<td>569</td>
<td>95</td>
<td>[ORF 8.5(61)]/UL49.5(42)</td>
<td>Possible transmembrane protein</td>
</tr>
<tr>
<td>UL49h</td>
<td>718</td>
<td>1464</td>
<td>27654</td>
<td>ORF 9(55)/UL49(46)</td>
<td>Tegument protein (VP22)</td>
</tr>
<tr>
<td>UL48h</td>
<td>1575</td>
<td>2855</td>
<td>48655</td>
<td>ORF 10(62)/UL48(60)</td>
<td>Tegument protein (VP16)</td>
</tr>
<tr>
<td>UL47h</td>
<td>3097</td>
<td>5520</td>
<td>91880</td>
<td>ORF 11(44)/UL47(43)</td>
<td>Tegument protein (VP13/14)</td>
</tr>
<tr>
<td>UL46h</td>
<td>5665</td>
<td>7368</td>
<td>63920</td>
<td>ORF 12(48)/UL46(46)</td>
<td></td>
</tr>
<tr>
<td>UL45h</td>
<td>8139</td>
<td>7507</td>
<td>23506</td>
<td>/UL45(38)</td>
<td>Virion protein</td>
</tr>
</tbody>
</table>

* Location of first base or its complement in the first ATG.
† Location of third base or its complement in the last codon before the stop codon. UL50h continues beyond the sequenced region.
‡ Counterparts of VZV and HSV-1 are listed with percentage similarities (in parentheses) of these proteins to MDV proteins. VZV lacks a UL45 counterpart.
§ Gene functions are derived from McGeoch et al. (1988) and Davison & Scott (1986), with additional data for the following HSV-1 genes: UL49.5, Barker & Roizman (1992) and Barnett et al. (1992); UL49, Elliot & Meredith (1992); UL47, McLean et al. (1990); UL45, Telford et al. (1992).

promoters (McKee et al., 1990). However, a more recent study reported that the trans-activation of IE genes of VZV and HSV-1 by the VZV homologue in transient expression systems (Moriuchi et al., 1993). It has also been reported that equine herpesvirus type 1 (EHV-1) possesses a protein functionally equivalent to VP16, although the EHV-1 homologue, like its VZV homologue, lacks an acidic carboxyl terminus (Purewal et al., 1992). These studies suggest that the MDV UL46h gene product may also trans-activate IE genes. VP16 of HSV-1 may have functions other than trans-activation. Mutational studies of VP16 suggest that the protein is also structurally important. HSV-1 with a mutated VP16, which is unable to trans-activate IE genes, is still capable of growth (ACE et al., 1989). The mutation abolished the abilities of VP16 to trans-activate IE genes and to form a protein–DNA complex with cellular proteins and the IE-specific regulatory element. On the other hand, a VP16 deletion mutant cannot replicate in normal cells (Weinheimer et al., 1992). This mutant replicated well in a cell line transformed with a VP16 expression vector. The deletion mutant induced nearly normal levels of viral DNA synthesis and capsid production during infection in normal cells, but did not induce further steps of virion maturation. These studies suggest that MDV UL46h, in a similar manner, may be essential for virus growth.

Of the four MDV ORFs, the UL47h gene product shows the lowest homology with HSV-1 and VZV (19% identity with HSV-1 UL47 and 18% with VZV ORF 11) (Table 1). In spite of weak homology, a hydrophilic amino terminus appears to be a characteristic of these gene products. It is conserved in all three viruses, but with notable differences. The HSV-1 and MDV proteins both possess an arginine-rich region, whereas the VZV protein contains a glutamic acid-rich region, giving it a highly acidic amino terminus. It has been shown that the EHV-4 homologue of UL47 also has an acidic amino terminus similar to that of ORF 11 of VZV (Whittaker et al., 1991). Recently, Whittaker et al. (1991) reported that the UL47 homologue of EHV-4 is phosphorylated and possesses O-linked glycosylation. Furthermore, they
described the gene product of UL47 (VP13/14) of HSV-1, which is also phosphorylated (Lemaster & Roizman, 1980), can bind certain lectins, indicating the presence of carbohydrate. Based on these findings, it is tempting to speculate that MDV UL47h may also be phosphorylated and glycosylated.

MDV UL49h has a moderate homology to its HSV-1 and VZV counterparts (24% and 29% identity, respectively) (Table 1). High homology between UL49h and VZV ORF 9 is mainly in the central region of the protein. UL49h is truncated in the amino-terminal region, which is 87 amino acids shorter than that of HSV-1 and 67 amino acids shorter than that of VZV. The size of the protein encoded by MDV UL49h will therefore be considerably smaller than that of other homologues. The functions of HSV-1 UL49 and VZV ORF 9 are not known.

MDV UL46h is shorter than its HSV-1 and VZV homologues, whereas UL46 of HSV-1 has an extended carboxyl-terminal region. Nonetheless, all three show a high degree of homology from the amino terminus to the central region (data not shown). The role of HSV-1 UL46 was initially thought to be involved in α-TIF-mediated transcriptional induction. Recent studies show that it is not required for modulating transcriptional induction or for growth in cell culture. Analogously, therefore, MDV UL46h may not be essential for viral replication.

The MDV UL49.5h gene identified in this study is found to be homologous to UL49.5 of HSV-1. In recent reports, Barker & Roizman (1992) and Barnett et al. (1992) described the finding of the HSV-1 UL49-5 gene encoding a hydrophobic protein. This protein was initially suggested to be a small capsid protein designated NC-7 (Barker & Roizman, 1992). However, McNabb & Courtney (1992) recently mapped the NC-7 protein to a different HSV-1 ORF, UL35. The function of the protein encoded by HSV-1 UL49.5 is not known at present.

One of the terminal regions of the BamH1–EcoRI subfragment reported in this paper contains a small portion of a sequence homologous to HSV-1 UL50 which encodes the deoxuryridine-5'-triphosphate nucletidylhydrodase (dUTPase) (Preston & Fisher, 1984). The
other terminal region of the sequenced fragment contains a gene homologous to UL45. We found our UL45 sequence in MDV strain GA to be in complete agreement with the published data for those of the Md5 and BC-1 strains (Ihara et al., 1989).

The sequence data of the genes reported here provide basic information for pursuing the identification and characterization of the proteins encoded by them. Of more immediate and practical interest is whether or not MDV UL48h, which lacks the carboxyl-terminal acidic region critical for trans-activation in HSV-1, would trans-activate IE genes and plays any role in MDV infection and transformation, and whether this gene and its protein product can be engineered for disease control.

We acknowledge M. Nonoyama for the generous gift of the BamHI genomic library of MDV strain GA and J. G. Calvert for critical review of the manuscript. These studies were funded in part by an ARS technology transfer agreement with Nippon Zeon Co. Ltd., Tokyo, Japan.

References


MDV major tegument protein genes


(Received 3 March 1993; Accepted 6 May 1993)