DNA sequence and organization of genes in a 5.5 kbp EcoRI fragment mapping in the short unique segment of Marek's disease virus (strain RB1B)

L. J. N. Ross,* M. M. Binns and J. Pastorek

AFRC Institute for Animal Health, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, U.K.

The DNA sequence of a 5.5 kbp EcoRI fragment located in the short unique region (Us) of the 'highly oncogenic' strain RB1B of Marek's disease virus (MDV) was determined. The sequence contained six open reading frames (ORFs), four of which were homologous to proteins mapping in the Us region of herpes simplex virus type 1 (HSV-1). These include the homologues of HSV-1 protein kinase, glycoprotein D (gD), glycoprotein I (gI) and US2 which is of unknown function. The MDV ORFs had a marked bias for A or T in the third codon position and analysis of the dinucleotide frequencies showed a marginal deficit in ApG/CpT but no overall deviation of CpG from random expectations. Comparison of genes in the Us region of MDV to herpesvirus proteins confirmed and extended our previous observation that MDV is more closely related to alphaherpesviruses than to gammaherpesviruses. We also showed that MDV possessed a homologue of HSV-1 gD which is lacking in varicella-zoster virus (VZV) but that MDV probably lacked homologues of US4 and US5 of HSV-1. These results show that in contrast to the genes in the long unique region which were grossly collinear in HSV, VZV and MDV, those mapping in Us show greater diversity.

Marek's disease virus (MDV) DNA is a linear, 175 kbp double-stranded molecule consisting of unique long (UL) and unique short (Us) segments flanked by inverted repeats (IRs) (Cebrian et al., 1982). Recently, Buckmaster et al. (1988) identified 35 MDV genes by comparing the translated sequence of random fragments of MDV DNA with the amino acid sequences of known herpesvirus proteins. Most of the genes identified mapped in the UL region and were collinear with genes of the alphaherpesviruses herpes simplex virus (HSV) and varicella-zoster virus (VZV). Moreover, the random sequencing study showed that MDV was more closely related to alphaherpesviruses than to beta- and gammaherpesviruses on the basis of amino acid sequence conservation. Genes of MDV that have been sequenced fully so far map in UL (Coussens & Velicer, 1988; Binns & Ross, 1989; Ross et al., 1989; Scott et al., 1989). In this paper we report on the DNA sequence and organization of predicted open reading frames (ORFs) in approximately half of the Us segment of MDV strain RB1B (Schat et al., 1982). We were particularly interested in determining whether MDV had a homologue of the HSV glycoprotein D (gD) and the nature and arrangement of genes adjacent to gD.

Initially, random fragments of MDV DNA derived from a 23 kbp BamHI A fragment of the GA strain of MDV (Fukuchi et al., 1984) containing part of Us and most of the IRs (Fig. 1) were sequenced and the predicted amino acid sequences compared to the sequences of known herpesvirus genes using FASTPM as described previously (Buckmaster et al., 1988). MDV homologues of several HSV type 1 (HSV-1) genes including immediate early (IE) 175, IE 68, protein kinase and gD were identified. The M13 clone that showed homology to HSV gD was labelled and hybridized to a library of EcoRI fragments of the RB1B strain of MDV (Binns & Ross, 1989) to identify a clone which might contain the entire gD gene. A 5.5 kbp EcoRI fragment (Fig. 1) which hybridized to the M13 clone was selected for further study.

Fig. 1. Genome structure of MDV showing the location of the BamHI A fragment in IRs and Us and of the 5.5 kbp EcoRI fragment that has been sequenced. The latter has been expanded in the lower part of the figure to show the position and orientation of the ORFs labelled 1 to 6.

B, BamHI and E, EcoRI restriction enzyme sites.
Fig. 2. Sequence of 5255 bp located within the 5.5 kbp EcoRI fragment mapping in Us. The sequence is given as the rightward 5' to 3' strand only and is numbered from the 5' end which is proximal to IRs. The deduced amino acid sequence of ORFs 2, 3, 4, 5 and 6 encoded in a rightward direction are shown above the sequence and the leftward encoded protein (ORFIC) is shown below the sequence. Note that the sequence shown does not have the EcoRI site at the 5' end. The locations of BamHI and EcoRI sites are underlined and labelled.

The base composition of the sequence determined is 40.1% C + G. A study of the codon catalogue of the ORFs sequenced here and of the gB gene of MDV (Ross et al., 1989) showed that there was a marked bias for adenine or thymine in the third codon position. The mean percentages of preferred codons containing adenine or thymine in the first, second and third positions were 45%, 57% and 89%, respectively. Our analysis (not shown) indicated that the codon preference for the ORFs in Us was similar but not identical to that of MDV gB. The differences in codon preference noted could be due to selective processes acting on DNA, to different origins of the genes, or could reflect differences in the degree of gene expression (Grantham et al., 1981).

A study of dinucleotide frequencies using the ANALYSEQ program of Staden (1982, 1984) showed that the frequency of CpG in the entire fragment did not deviate from random expectations but that ApG/CpT were under-represented (an approximately 1-3-fold deficit from random expectations calculated according to Honess et al., 1989). Our results are in broad agreement with those of Honess et al. (1989) who analysed MDV sequences derived from other regions of the MDV genome and concluded that in contrast to Epstein–Barr virus (EBV) and herpesvirus saimiri (HVS), MDV DNA does not show a deficit in CpG. It was suggested that a deficiency in CpG and an increase in TpG/CpA in the genomes of EBV and HVS is probably a reflection of mutations arising from methylation of viral DNA in latently infected and transformed lymphocytes and selection of mutated virus during natural transmission via a latent virus intermediate. Whether the apparent lack of CpG deficiency in MDV DNA reflects a different form of latency and natural transmission of MDV is not clear because the reasons for the deficiency in ApG/CpT are not known. Moreover we have noted marginal deficits in the frequency of CpG in ORFs 1 and 2 (1.2-fold and 1.45-fold, respectively).

Analysis of the sequence for coding regions in all six reading frames using ORFSCAN (M. Boursnell, IAH Houghton) revealed six potential ORFs greater than 400 bp in size (Fig. 1). There is strong indirect evidence that at least four of these encode proteins since the predicted amino acid sequences are homologous to known HSV proteins (Table 1). These are ORF1, ORF3, ORF5 and ORF6 which are the MDV homologues of HSV-1 Us2, Us3 (protein kinase), Us6 (gD) and Us7 (gI) respectively. The most conserved genes were the MDV homologues of HSV-1 Us2 and protein kinase which gave FASTP scores of 612 and 337, respectively. Of the

### Table 1. Location and properties of MDV ORFs

<table>
<thead>
<tr>
<th>ORF</th>
<th>Position*</th>
<th>M,†</th>
<th>Gene Identity (§)</th>
<th>Score‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>(C) 1134</td>
<td>325</td>
<td>29.8K</td>
<td>US2</td>
</tr>
<tr>
<td>ORF2</td>
<td>623</td>
<td>1163</td>
<td>19.3K</td>
<td></td>
</tr>
<tr>
<td>ORF3</td>
<td>1246</td>
<td>2451</td>
<td>44.7K</td>
<td>Protein kinase (US3)</td>
</tr>
<tr>
<td>ORF4</td>
<td>2564</td>
<td>3004</td>
<td>16.8K</td>
<td></td>
</tr>
<tr>
<td>ORF5</td>
<td>3191</td>
<td>4384</td>
<td>45.4K</td>
<td>gD (US6)</td>
</tr>
<tr>
<td>ORF6</td>
<td>4495</td>
<td>5255</td>
<td>ND§</td>
<td>gI (US7)</td>
</tr>
</tbody>
</table>

* The positions of potential protein-coding regions (>400 bp) are shown from the first residue of the translation initiation codon to the first residue of the stop codon as numbered in Fig. 2.† The M, s are those of primary polypeptides.‡ Similarities between MDV proteins and known HSV-1 proteins are indicated by scores (Lipman & Pearson, 1985); percentage identity and length of amino acid (aa) overlap were determined using FASTPM (Binnns et al., 1987).§ (C) indicates that ORF1 is oriented leftward.** ND, Not done.
Apart from ORF1 which is transcribed in a leftward direction, all the remaining ORFs are transcribed in a rightward direction (Fig. 1 and 2). Since ORF1 is homologous to a gene of HSV-1 (US2) and overlaps ORF2 extensively, it is unlikely that ORF2 encodes a protein. However, definite conclusions on this await an analysis of RNA transcripts. Codon usage analysis of ORF5 (gD) showed that it could encode a small protein.

Inspection of the sequence (Fig. 2) showed that none of the translation initiation sites conform to the consensus of Kozak (1986) except for the AUG of ORF1 at position 1134. However, deviations from Kozak's consensus have been noted for many known genes of MDV (Coussens & Velicer, 1988; Binns & Ross, 1989; Scott et al., 1989). Because of the AT-rich sequence, it is not possible to predict potential transcription initiation and termination sites with confidence in the absence of RNA transcription mapping data. However, 'TATATA' sequences at positions 2493, 3123 and 4453 could be potential TATA box elements for ORFs 4, 5 and 6 respectively and 'AATAAA' sequences at positions 2761 and 3115 could be potential polyadenylation signals for ORFs 3 and 4. As there is no obvious polyadenylation signal for ORF5, it is possible that the mRNA for ORFs 5 and 6 are 3' co-terminal as is the case for the HSV-1 homologues (McGeoch et al., 1988). We are unable to detect this because the EcoRI fragment sequenced here does not contain the entire ORF6 and flanking 3' sequences. The promoter region of ORF5 contained the motif ATCGTC (position 3048) and downstream sequences which are similar to the consensus recognition sequence for binding of the IE 175 protein to the gD promoter of HSV-1 (Tedder et al., 1989). It should be possible to determine whether the interaction between these sequences and MDV IE proteins contribute to the regulation of expression of MDV gD.

The arrangement of genes in the Us region of MDV and in the alphaherpesviruses HSV-1 (McGeoch et al., 1988), VZV (Davison & Scott, 1986) and pseudorabies virus (PRV) (van Zijl et al., 1990) is shown in Fig. 3. It is clear from the figure that MDV has a homologue of HSV gD which is also shared with PRV (gp50) but is lacking in VZV. The gI homologue however is conserved in all three viruses (gp4 or gene 67 of VZV and gp63 of PRV). Similarly, homologues of HSV US3 (protein kinase) are present in all the viruses. However, MDV has no counterpart for HSV US5 and ORF4 appears to be unique to MDV. Likewise, gX of PRV does not appear to have a homologue in HSV-1 but is related to HSV-2 gG (US4 of HSV-2) (McGeoch et al., 1987). MDV ORF1 is collinear with its homologue in HSV-1 (US2) which is lacking in VZV and is rearranged in PRV. Thus in contrast to the Ul region in which genes of MDV and the alphaherpesviruses are mainly collinear, the Us region shows greater diversity in organization and content. Whitton & Clements (1984) proposed that genes in Us have been deleted or rearranged during evolution as a result of contraction or expansion of the IRs flanking Us and recombination. Recently, McGeoch (1990) postulated that glycoproteins in Us may have arisen by a process of gene duplication and that the glycoproteins subsequently evolved separately to perform novel functions. In the accompanying paper (Ross & Binns, 1991) we examine the evolutionary relationships between the MDV genes reported here and their homologues in other herpesviruses in greater detail. The results reported here support our previous observation that MDV appears to have little in common with gammaherpesviruses other than their capacity to cause latent infection and transformation of lymphocytes. We have shown further that MDV is more similar to HSV-1 than to other alphaherpesviruses on the basis of gene conservation and arrangement in the part of Us that has been sequenced.

We are grateful to Mrs J. Simpson and Miss B. Britton for excellent technical assistance.

References

SHORT COMMUNICATION


(Received 1 October 1990; Accepted 19 December 1990)