Identification and transcriptational analysis of the homologues of the herpes simplex virus type 1 UL30 to UL40 genes in the genome of nononcogenic Marek’s disease virus serotype 2

Yoshihiro Izumiya,1 Hyung-Kwan Jang,1 Mie Sugawara,2 Yasuhiro Ikeda,1 Ryuichi Miura,1 Yorihiro Nishimura,1 Kazuya Nakamura,1 Takayuki Miyazawa,1 Chieko Kai1 and Takeshi Mikami1

1Department of Veterinary Microbiology, Faculty of Agriculture, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan
2Biomedical Research Laboratories, Sankyo Co. Ltd, 1-2-58 Hiromachi, Shinagawa-ku, Tokyo 140, Japan

Studies on Marek’s disease virus serotype 2 (MDV2) are important for understanding the natural non-oncogenic phenotype of MDV. This study reports a 27 535 bp nucleotide sequence of part of the MDV2 genome located in the central unique long (UL) region. The analysis revealed 11 complete ORFs with high amino acid sequence identities to the products of other alphaherpesviruses. The MDV2 ORFs were arranged collinearly with the prototype sequence of herpes simplex virus type 1, ranging from the UL30 to UL40 genes. Sequences that were particularly well conserved among alphaherpesviruses were the putative functional domain of the DNA polymerase (UL30) and the ribonucleotide reductase large and small subunits (UL39 and UL40). On the other hand, in contrast to oncogenic MDV1, MDV2 did not contain the conserved proline-repeat region in the UL36 homologue. All the genes identified were confirmed to be transcribed as 3′-coterminal mRNAs and/or unique transcripts in virus-infected cells.

Marek’s disease virus (MDV) is the causative agent of a contagious, malignant T-cell lymphoma in chickens and is divided into three serotypes. MDV1 includes oncogenic strains and their attenuated strains. MDV3 (herpesvirus of turkeys; HVT) includes non-pathogenic strains and has been used for one of the effective vaccines against Marek’s disease. MDV2 includes the naturally non-pathogenic strains isolated from chickens and other birds belonging to the genus Gallus (Lin et al., 1990). In spite of their biological similarity to gammaherpesviruses, the genome structures of all three serotypes of MDV are similar to those of alphaherpesviruses (Ono et al., 1992).

Since MDV2 is a naturally occurring nononcogenic strain of MDV in chickens, comparative studies on the virus genome and its gene products with those of oncogenic MDV1 might be crucial for understanding virus oncogenicity and natural immunity. Thus, we recently reported on the MDV2 UL41–UL51 homologous genes and their transcripts (Izumiya et al., 1998). In this paper, we report 11 newly identified complete ORFs located upstream of the UL41 gene. Based on position and sequence similarities between MDV2 and other alphaherpesviruses (Davison & Scott, 1986; McGeoch et al., 1988; Telford et al., 1992; Schwyzer & Ackermann, 1996), we have named the identified MDV2 ORFs as UL30, UL31, UL32, UL33, UL34, UL35, UL36, UL37, UL38, UL39 and UL40. Furthermore, we confirmed by RNA analysis that all the ORFs were indeed transcribed as 3′-coterminal and/or unique transcripts.

Genomic libraries of restriction enzyme fragments (BamHI and EcoRI) of MDV2 strain HPRS24 were previously constructed in our laboratory (Ono et al., 1992). To obtain subfragments for subsequent DNA sequencing, BamHI fragments A (15-1 kbp), B (13-8 kbp) and I (5-8 kbp) were further digested with restriction enzymes (Fig. 1d). The subfragments were inserted into pBluescript SK(+) vector (Stratagene) and then deleted with exonuclease III (Toyobo) to construct deletion clones. DNA sequencing was performed on both strands by using the ABI Dye Primer cycle sequencing method (Applied Biosystems), and junction sequences between the subfragments were confirmed by the dideoxy chain termination method with the ABI Dye Deoxy terminator cycle sequencing kit. DNA sequences were assembled and analysed with the UWGCG program BESTFIT as described previously (Jang et al., 1998).

The sequence obtained contained 27 535 bp with an overall G+C content of 50.1% and is numbered from left to right with respect to the orientation of the MDV2 genome map (Fig.
As a result of the confirmation of junction sequences, a small 206 bp EcoRI fragment (c) was found between EcoRI fragments K and U1 that was lacking in our previous report (Ono et al., 1992). To identify the corresponding MDV2 ORFs, we were able to take advantage of the known complete sequences of herpes simplex virus type 1 (HSV-1), varicella-zoster virus (VZV), equine herpesvirus-1 (EHV-1) and bovine herpesvirus-1 (BHV-1) (McGeoch et al., 1988; Davison & Scott, 1986; Telford et al., 1992; Schwyzer & Ackermann, 1996) by alignment of predicted amino acid sequences. Thus, the fragments of the MDV2 genome sequenced were found to encode homologues corresponding to the products of HSV-1 genes UL29 to UL40 and their gene arrangement was strictly collinear with that found in HSV-1 (Fig. 1b). However, the UL29 gene was located partially at the 5’ end of the EcoRI-V fragment and its analysis was excluded from this study. The 3’ terminus of the MDV2 UL30 gene overlapped the UL31 gene by 71 nucleotides in a tail-to-tail orientation. Also, the MDV2 UL32 gene overlapped the AT dinucleotide of the ATG initiation codon of the UL33 gene, similar to the other alphaherpesviruses HSV-1, VZV, EHV-1, BHV-1 and MDV1 (McGeoch et al., 1988; Davison & Scott, 1986; Telford et al., 1992; Schwyzer & Ackermann, 1996; Wu et al., 1996). Although the C terminus of the MDV2 UL32 homologue was highly conserved among other alphaherpesviruses, the 3’ terminus of the MDV2 UL32 gene did not overlap the 5’ terminus of the UL31 gene. This feature was not observed in MDV1. Each of the ORFs identified contained a number of potential transcriptional regulatory sites. Potential transcription start signals [TATA box; TATA(A/T)] were found within 220 bp upstream of the initiation codons of all of the ORFs except UL33. In addition, other potential transcriptional elements fitting the consensus sequence of the CAAT box [(A/G)(A/G)GCAAT] (Chodosh et al., 1988) were found at locations ranging from 45 to 414 bp upstream of the UL33 gene, presumably substituting for the TATA box sequence. Potential polyadenylation signals [A(A/T)TAAA] were only detected downstream of the UL33 gene, presumably substituting for the TATA box sequence. The presence of transcripts from the region sequenced was confirmed by Northern blot analysis of total RNAs from
Identification of MDV2 UL30–UL40 homologues

Table 1. Properties of the MDV2 genes identified

Gene | Start | Stop | Amino acids | Molecular mass (kDa) | CCAAT box | TATA box | Poly(A) signal | Identity to homologous proteins
--- | --- | --- | --- | --- | --- | --- | --- | ---
UL29* | 1136 | – | – | – | – | – | – | BHV-1 | EHV-1 | VZV | HSV-1
UL30 | 1386 | 1190 | 133 | 1190 | – | – | – | 55.7 | 59.8 | 54.7 | 54.4
UL31* | 5790 | 4885 | 304 | 342 | 6080–6074 | 5858–5854 | 4973–4788 | 4747–4742
UL32* | 7693 | 5816 | 626 | 693 | 8113–8108 | 7800–7797 | – | 42.7 | 45.7 | 46.5 | 51.5
UL33 | 7692 | 8084 | 131 | 147 | – | – | – | 51.4 | 48.3 | 40.8 | 45.2
UL34 | 8155 | 8699 | 279 | 307 | 8120–8118 | 7800–7797 | – | 48.6 | 52.7 | 54.3 | 50.0
UL35 | 9080 | 9463 | 127 | 140 | 8833–8839 | 8904–8968 | 9466–9491 | 38.7 | 47.2 | 35.6 | 35.6
UL36* | 18725 | 17954 | 3064 | 3333 | 19026–19021 | 18728–18724 | 9488–9483 | 34.0 | 33.9 | 30.3 | 30.5
UL37* | 22035 | 18916 | 1040 | 1155 | – | – | – | 30.1 | 29.9 | 29.3 | 28.3
UL38 | 22405 | 23814 | 470 | 519 | 22300–22365 | 22085–22082 | 22378–22382 | 43.3 | 46.1 | 39.3 | 39.9
UL39 | 23996 | 26386 | 797 | 886 | 23912–23916 | 23638–23689 | 27523–27528 | 54.0 | 52.3 | 53.6 | 50.6
UL40 | 26416 | 27474 | 353 | 398 | 26141–26147 | 26385–26389 | 27523–27528 | 62.7 | 62.7 | 59.3 | 61.3

* Gene located on the reverse DNA strand.

MDV2- or mock-infected chicken embryo fibroblasts (Fig. 2a) with probes I to XI (Fig. 1e). The RNA (10 μg per lane) was separated on a 1:2% agarose–formaldehyde gel and transferred to nylon membranes (Biodyne). The 11 strand- and gene-specific hybridization probes, labelled with α-32PUTP, were transcribed from linearized nested-deletion plasmids or subcloned plasmids by the MAXI transcript kit (Ambion) as recommended by the manufacturer. Northern blots were preincubated and hybridized for 3 h and 14 h, respectively, at 54 °C in 0.6 M NaCl, 0.2 M Tris–HCl (pH 8.0), 0.02 M EDTA, 0.5% SDS and 50% formamide, supplemented with 100 μg/ml denatured yeast RNA (Ambion) for preincubation. After hybridization, the blots were washed under the conditions reported previously (Fuchs & Mettenleiter, 1996). The results of this analysis, showing the putative locations and directions of the viral transcripts, are presented in Fig. 1(c). Probe I reacted with only one viral transcript of 4.0 kb, which probably represents the UL30 mRNA. Probe III (UL32) hybridized to two viral transcripts, of 3.7 and 2.2 kb. Probe II (UL31) hybridized additionally to transcripts of 1.6 and 1.1 kb. All four transcripts (3.7, 2.2, 1.6 and 1.1 kb) presumably terminate at either of two potential polyadenylation signals downstream of the UL31 gene. Based on their sizes, the 2.2 and 1.6 kb transcripts are probably derived from the middle of the UL32 gene and the 3.7 kb transcript is from the antisense strand of the UL33 gene. The 3.7 and 1.6 kb transcripts of the UL32 gene were also reported in MDV1 as transcripts of the same size, but the 2.2 and 1.1 kb transcripts were not observed (Lee et al., 1996). Since the 3.7 kb transcript was detected with probe IV but not V using antisense DNA probes (data not shown), the transcript was indeed derived from the antisense strand of the UL33 gene. Probe VI (UL35) hybridized to a highly abundant transcript of 0.5 kb, along with 2.6, 2.0, 1.5, 1.1 and 0.8 kb transcripts. These 2.6 and 2.0 kb viral transcripts also hybridized to probe V (UL34) and probe IV (UL33). Therefore, the 2.6 and 2.0 kb transcripts were considered to be read-through mRNAs spanning the UL33 to UL35 genes and are possibly 3′-coterminal with the 0.5 kb UL35 mRNA and the 1.5, 1.1 and 0.8 kb mRNAs encoding both the UL34 and UL35 genes. As shown in Fig. 2(b), transcriptional mapping of these six newly identified transcripts was performed by using fine-mapping oligonucleotides. From these results, the 2.6 kb transcript was initiated from upstream of the UL33 gene at a position before nt 7550, the 2.0 kb transcript from between nt 7550 and 8000, the 1.5 kb transcript from between nt 8000 and 8261, the 1.1 and 0.8 kb transcripts from between nt 8261 and 8840 and the 0.5 kb transcript from between nt 8841 and 9410. Also, we performed RT–PCR under conditions reported previously (Jang et al., 1998) and confirmed that these transcripts were not spliced mRNAs (data not shown). Probe VIII (UL37) reacted only with the largest viral transcript of 13.1 kb, which is most likely to represent the UL37 mRNA, whereas probe VII additionally recognized a 10.1 kb UL36 gene-specific transcript. A distinct set of apparently 3′-
Y. Izumiya and others

Fig. 2. Northern analysis of RNA from mock-infected (M) and MDV2-infected (V) chicken embryo fibroblasts. (a) Hybridization between total RNA from MDV2- or mock-infected cells and each of the probes I to XI. Panels I to XI show the results of the hybridization. (b) Hybridization using fine-mapping oligonucleotides. Panels 1 to 5 show the results of the hybridization using probes 1–5 shown in Fig. 1(e). These results indicated that at least six transcripts exist and that they presumably terminate at the same position. Molecular sizes (kb) of the RNA markers (Gibco BRL) are indicated to the left. Open arrowheads denote locations of 28S and 18S rRNA.

coterminal transcripts was detected by probes IX (UL38), X (UL39) and XI (UL40). The 1.5 kb transcript reacting only with probe XI presumably encodes the UL40 protein. The 2–4 and 3–9 kb viral transcripts detected by probes X and XI but not probe IX might encode the UL39 and UL40 proteins. The minor 2–4 kb transcript might be derived from the 5’ end of the UL39 gene, based on the size of the coding region. A 5–6 kb transcript also hybridized to probes IX–XI. Therefore, this 5–6 kb transcript was considered to be a read-through mRNA spanning UL38 to UL40. In contrast, the signals obtained at 4–3 and 1–7 kb are most likely caused by non-specific reactions of degraded viral RNA with the 28S and 18S cellular rRNAs (Fig. 2a, b; open arrowheads).

The location and properties of the MDV2 ORFs identified are summarized in Table 1. The deduced protein encoded by MDV2 UL30 consisted of 1190 amino acids, with 76–5% identity to that of MDV1, and it exhibited extensive amino acid similarity with the DNA polymerase of HSV-1. MDV2 DNA polymerase was longer than that of MDV1 by 10 amino acids and it contained nine highly conserved putative functional regions that were described previously for MDV1 (Sui et al., 1995). As other genes relating to DNA replication, we identified the UL39 and UL40 homologues as ribonucleotide reductase large (RR1) and small (RR2) subunits, respectively. The deduced MDV2 RR1 consisted of 797 amino acids and contained three conserved glycine residues within the sequence G–X–G–X–G at positions 523–528, corresponding to the nucleotide-binding site (Nikas et al., 1986). However, similarly to EHV-4 (Riggio & Onions, 1994), MDV2 lacked the protein kinase domain of RR1. MDV2 RR2 consisted of 353 amino acids and exhibited the highest identity (59–80%) to other alphaherpesviruses among the 11 homologues identified (Table 1). His-144 and Tyr-148 of MDV2 RR2, the putative iron-binding site and tyrosine free radical, were positionally conserved as compared with other herpesviruses (Nikas et al., 1986). Further, the C terminus of RR2 was reported in pseudorabies virus and HSV-1 to have a critical role for subunit association and to be one of the target sites of antivirus drugs (Dutia et al., 1986; Liuzzi et al., 1994). This region was completely conserved in MDV2.

Immature or B capsids of HSV-1 are composed of seven proteins, including the UL35 and UL38 homologues. The deduced 127 and 470 amino acid proteins encoded by the MDV2 gene homologues exhibited only 36–47 and 39–46% identity, respectively, to those of other alphaherpesvirus and these similarities were mainly restricted to the central regions.
The UL33 gene of HSV-1 is necessary for assembly of the DNA-containing capsid (Roizman & Sears, 1996). The overall identity of the deduced 131 amino acid MDV2 UL33 protein to its homologues in other alphaherpesviruses was rather low (40.8–51.4%), and 28 amino acid residues were conserved positionally in all the gene products compared. Although the function of the HSV-1 UL32 protein is not yet clear, it has been identified as a cysteine-rich, zinc-binding, essential cytoplasmic protein (Chang et al., 1996). The MDV2 UL32 protein exhibited extensive identity (51.5%) and contained three completely conserved C–X–X–C motifs as a putative zinc-binding motif. However, the MDV1 UL32 protein was reported to be a membrane glycoprotein, gp82 (Wu et al., 1997). Since the MDV1 UL32 amino acid sequence is not yet available, we could not compare it with that of MDV1. Further studies of this protein will be required.

HSV-1 UL31 protein has been identified as a phosphoprotein and has a hydrophilic N terminus and a nuclear-localization signal (Chang & Roizman, 1993). The deduced 304 amino acid MDV2 UL31 protein exhibited 51.6% identity to that of HSV-1 but no putative nuclear-localization signal was found. However, two peptide sequences in the MDV2 UL31 protein, at residues 89–104 [VADNCL(S/T)LSGMGY(Y/H/LG)] and 283–290 [DG(E/G)LL(M/L)LEY], were conserved with other alphaherpesviruses.

The deduced 279 amino acid MDV2 UL34 protein exhibited 50.0% identity to that of HSV-1 and contained two domains, QNTGVSVL(F/H)(Q/E)GGF and VQL(S/A)FRF(M/V)GP, positionally conserved among alphaherpesviruses. Purves et al. (1991) reported that the HSV-1 viral protein kinase, the US3 gene product, phosphorylated the UL34 protein. The phosphorylation site recognized by the viral protein kinase has the consensus sequence (R)Xn(S/T)XX, where n ≥ 3. Between the R and S/T residues, R, A, V, P or S are preferred, whereas after the S/T residue, the same amino acids are preferred except that acidic residues or proline are unacceptable. (S/T) is the target residue. Although the MDV2 protein kinase motif of the US3 homologue has already been identified (Jang et al., 1998), the MDV2 UL34 protein does not contain the definite consensus target sequence.

Both the UL36 and UL37 genes encode tegument proteins in the HSV-1 genome (Roizman & Sears, 1996). The MDV2 UL36 and UL37 genes were capable of encoding 3064 and 1040 amino acids, respectively. Although the deduced MDV2 UL36 protein had only 30.3–34.0% identity to other alphaherpesvirus counterparts, the MDV2 UL36 protein had a conserved potential lipid membrane-attachment signal (DPHGHHHGGQAC) and two potential ATP-binding motifs, at residues 2227–2234 and 2455–2468, in common with MDV1. Interestingly, MDV1 contains a proline-repeat region in the C terminus of the UL36 homologue (Mao et al., 1996) but MDV2 does not. The UL37 gene homologue was located downstream of the largest gene, UL36. The MDV2 UL37 gene encoded a 1040 amino acid protein with only 28.3–30.1% identity to other alphaherpesvirus UL37 proteins. Further, a characteristic and highly conserved stretch could not be found in the MDV2 UL37 homologue. Therefore, it is not known whether the MDV2 UL37 protein plays the same functional role as other alphaherpesvirus homologues.

Further work is required to determine the properties and functional features of the proteins encoded by the MDV2 genes identified in this study.

This work was supported by grants from the Ministry of Education, Science, Sports and Culture and Ministry of Agriculture, Forestry and Fisheries of Japan. Y. Ikeda, Y. Nishimura and Y. Izumiya are supported by research fellowships for young scientists of the Japan Society for Promotion of Science.

References


Received 12 February 1999; Accepted 30 May 1999