Characterization and transcript mapping of a bovine herpesvirus type 1 gene encoding a polypeptide homologous to the herpes simplex virus type 1 major tegument proteins VP13/14

Sylvie LaBoissière, Michel Trudel and Claire Simard*

Centre de Recherche en Virologie, Institut Armand-Frappier, Université du Québec, 531 Boulevard des Prairies, Laval, Québec, Canada H7V 1B7

Using in vitro translation of hybrid-selected mRNA, we have previously shown that bovine herpesvirus type 1 HindIII fragment M encodes an abundant 94K polypeptide. Using immunoprecipitation and sequencing analyses, it has now been shown that the polypeptide is related to the major tegument protein VP8 and is homologous to the herpes simplex virus type 1 major tegument proteins VP13/14. The sequence of the VP8 gene (field isolate 34) is reported and compared to published data. Several differences between the sequences were detected, resulting particularly from base insertions/deletions generating three major frameshifts affecting an area of 87 amino acid residues of the encoded protein. In addition, sequence comparison revealed 29 single base alterations, excluding frameshift regions, producing 17 amino acid substitutions. Overall, 14.1% of the deduced amino acid sequences were divergent. We have also established that the last 152 nucleotides of the previously reported sequence correspond to the sequence of the minus not the sense strand. Finally, we report that the 4.4 kb transcript of the VP8 gene is initiated 39 nucleotides upstream from the translation start codon.

Introduction

Bovine herpesvirus type 1 (BHV-1), the aetiological agent of infectious bovine rhinotracheitis and infectious pustular vulvovaginitis, is a virus of cattle distributed world-wide. The virus, a member of the Alphaherpesvirinae subfamily (Roizman et al., 1982), has also been associated with conjunctivitis, encephalitis, enteritis, dermatitis and abortion (Gibbs & Rweyemamu, 1977).

The BHV-1 genome is a linear dsDNA molecule of approximately 140 kbp which is divided into unique long (UL) and short (US) regions. The latter is bounded by inverted repeat sequences (RS) and can invert its orientation relative to UL, resulting in the existence of two isomeric forms (Farley et al., 1981). The genome encodes more than 40 polypeptides (Metzler et al., 1985; Misra et al., 1981), 20 of which have been located in discrete genomic regions using translation of mRNA hybrid-selected to individual HindIII DNA fragments (Simard et al., 1990). In particular, coding sequences of an abundant polypeptide (hereafter named VP8) with an apparent Mr of 94K have been assigned to the small 3.7 kbp HindIII fragment M. At present, gene coding sequences for the gI, gIII and gIV major glycoproteins, gH glycoprotein, thymidine kinase, part of the DNA polymerase and the VP8 major tegument protein have been reported (Carpenter & Misra, 1991; Fitzpatrick et al., 1989; Kit & Kit, 1986; Meyer et al., 1991; Misra et al., 1988; Mittal & Field, 1989; Owen & Field, 1988; Smith et al., 1990; Tikoo et al., 1990; Whitbeck et al., 1988).

In this study, the sequence of the gene encoding the VP8 tegument protein of a Canadian field isolate (34) is reported and compared to data obtained by Carpenter & Misra (1991) using the American isolate P8-2. This comparison identified several areas of divergence resulting from base insertions/deletions as well as single base alterations. In particular, three distinct frameshifts affecting a total of 87 amino acid residues of the encoded protein were found in the previously reported sequence. Finally, the precise mapping of the 5' end and the kinetics of expression of the VP8 gene transcript are presented.

Methods

Cells, virus and viral DNA. The BHV-1 isolate (strain 34) was provided by Dr D. Mitchell (Gouvernement du Canada, Lethbridge, Alberta) and has been shown to be related to the Colorado strain.
(subtype 1.1, Metzler et al., 1985) based on restriction endonuclease (Simard et al., 1991) and serological analyses. Virus was grown in confluent monolayers of an ovine kidney (OK) cell line as previously described (Trudel et al., 1987). Extracellular virions were concentrated by ultracentrifugation (Trudel & Payment, 1980) and viral DNA was purified as reported (Simard et al., 1990).

Northern blot hybridization and primer extension analyses. Poly(A)* mRNA was isolated as described (Simard et al., 1990) from cells at 0, 6, 12, 18 and 24 h post-infection (p.i.) with BHV-1. mRNA (2 μg) was fractionated in an agarose gel, blotted onto a Hybond-N membrane and hybridized with a radiolabelled 387 bp Smal–EcoRI fragment, according to the method of Simard et al. (1992).

Primer extension analyses were performed essentially as described by Sambrook et al. (1989) using the oligonucleotide 5' GCCCATCCC-TAGCCGCGTCCATGGC 3', encompassing the translation initiation codon of the VP8 gene coding sequences. Briefly, the oligonucleotide was radiolabelled with [γ-32P]ATP and T4 kinase, and then annealed with 10 μg of total RNA extracted either from BHV-1-infected (12 h p.i.) or mock-infected cells. Reaction mixtures were heated at 80 °C for 10 min and incubated at 37 °C for 12 h to allow annealing. After reverse transcription, synthesized products were analysed on 8% sequencing gels.

DNA sequencing. The cloning of the 3.7 kbp BHV-1 HindIII fragment M into the prokaryotic pKS vector (clone pKS/Mhd) has been described elsewhere (Simard et al., 1990). Caesium chloride-purified pKS/Mhd plasmid DNA was mapped using combinations of single and double digestions with several restriction enzymes. Both strands of the inserted fragment were then sequenced by the chemical method of Maxam & Gilbert (1977). Sequence data were assembled and analysed using the MacVector sequence analysis software (International Biotechnologies).

Results and Discussion

Individual BHV-1 HindIII fragments have previously been analysed for their ability to select complementary mRNA. Upon in vitro translation, a gene encoding an abundant 94K polypeptide has been assigned to the small 3.7 kbp fragment M (Simard et al., 1990). Owing to the abundance of the translated polypeptide, we planned to express the gene in a viral vector to define the role of the protein in cellular and humoral immunity. To this end, the characterization of the gene coding sequences has been undertaken, and the restriction map and sequencing strategies employed are shown in Fig. 1.

The nucleotide sequence of fragment M has a G + C content of 72%, which agrees with previous estimates based on the whole viral genome (Graham et al., 1972). This high G + C content generates occasional nucleotide compressions in sequencing gels, particularly in regions containing long stretches of 20 to 40 G and C residues. Therefore, the regions involved were sequenced several times.

Open reading frame (ORF) analyses identified a single long and complete ORF (Fig. 1) with considerable similarity to the VP8 gene coding sequences reported previously (Carpenter & Misra, 1991). In addition, two partial ORFs located at the extremities of fragment M were found in the same orientation as the VP8 gene. The nucleotide sequence of the latter gene and of proximal regions can be seen in Fig. 2(a).

The VP8 gene coding sequence initiates with an ATG codon in a good context for the initiation of translation, since it is flanked by GACGCCATGG (Kozak, 1986, 1987). The latter feature is of interest because it may partially explain the abundance of the polypeptide in vivo (Carpenter & Misra, 1991) and in vitro (Simard et al., 1990). The VP8 polypeptide is composed of 739 amino acid residues with an Mr of 80-7K (Fig. 2a), which differs significantly from the Mr estimated from the polypeptide translated in vitro (94K). However, the deduced primary structure of the protein indicated that it contains 8% proline residues and it has previously been shown that proline-rich polypeptides exhibit anomalous migration in SDS–PAGE (Ferguson et al., 1984). Finally, analysis of the VP8 amino acid sequence did not reveal putative sites for N-glycosylation (Kornfeld & Kornfeld, 1985), or a hydrophobic region characteristic of a signal peptide. However, recent studies have provided evidence that some herpesviral tegument proteins may be glycosylated. It has been reported that herpes simplex virus type 1 (HSV-1) VP13/14 can bind lectins (Meredith et al., 1991), whereas equine herpesvirus type 4 (EHV-4) gp10 (Whittaker et al., 1991) and human cytomegalovirus (Benko et al., 1988) tegument proteins can be radio labelled with [3H]glucosamine and [3H]galactose, respectively. However, it remains to be determined whether the VP8 protein is glycosylated. Actual results indicate that such processing may occur only at a low level because the apparent Mr of the VP8 protein synthesized either in vivo (92K; Carpenter & Misra, 1991) or in vitro (94K; Simard et al., 1990) is very similar.

Upstream from the coding region, a CAAT box motif and a TATA box are found at positions 17 and 190, respectively. These sequences may function as a promoter for the expression of the VP8 transcript (Jones & Yamamoto, 1985; McKnight & Kingsbury, 1982). However, the CAAT box motif identified here may not be significant because it has recently been demonstrated that sequences upstream from the TATA box are dispensable for the transcription of true HSV-1 late genes (Steffy & Weir, 1991; Kibler et al., 1991). Finally, no polyadenylation consensus sequence was found downstream from the stop codon.

Sequence variations of Canadian versus American isolates

The sequence presented in Fig. 2(a) has been conscientiously compared with previously published data (Carpenter & Misra, 1991). However, several divergences were identified, even though both isolates belong to BHV-1 subtype 1.1. In particular, three framehifts
affecting 14, 55 and 18 amino acid residues of the encoded protein were created in the reported sequence by the addition of base residues. Specifically, the first frameshift (FS-1) extends from positions 357 to 399, being generated by the insertion of two pyrimidines (CT) and terminated by an additional C residue. With the exception of the latter three differences, the sequences are identical in the region involved. As shown by a sequencing gel representing the region encompassing the beginning of FS-1 (Fig. 3a), no additional C or T residues are contained within our sequence. The second frameshift was generated by the insertion of an A residue at position 1353, an additional T at position 1403, two G residues at positions 1457 and 1458 and a C at position 1467. The frameshift ended with the deletion of two C residues at positions 1468 and 1517. Confirming our results, Fig. 3(b) and (c) present sequencing gels representing the beginning and end of FS-2, respectively. In addition, two other differences were found in the FS-2 region at positions 1385 and 1386, where T and G were replaced by G and C residues. These changes could either represent individual transversions (T→G, G→C) or dinucleotide inversions (TG→GC). Finally, a third frameshift (FS-3) was found between positions 2092 and 2147, being created by the addition of a C residue and terminated by the insertion of two C residues. With the exception of these three differences, the sequences were identical in the FS-3 region. Sequencing gels representing the beginning and end of FS-3 revealed that our sequence did not contain additional residues in these regions (Fig. 3d and e, respectively). All these changes show that the previously reported VP8 amino acid sequence is three amino acid residues longer than ours.

Other differences in the coding region were found. These resulted from 29 base alterations involving eight C→T and four G→A transitions, three G→C transversions, and five GC→CG, one GT→TG and one GT→CG dinucleotide inversions. As a consequence, 17 amino acid substitutions were identified in the deduced polypeptide sequences. Finally, four local base additions, two base deletions and one T→A transversion were found in the 3' non-coding region. Furthermore, in the region between positions 2538 to 2607, the previously reported nucleotide sequence was found to be perfectly complementary (with the exception of an A→C) to our sequence. We have used four distinct strategies in the sequencing of this region, as compared to the single strand sequencing used for the P8-2 sequence. It is thus probable that the latter sequence may have been incorrect in this region.
denoting frameshifts, the latter are represented only where nucleotide mismatches and amino acid substitutions were observed.

Fig. 2. (a) Comparison of the VP8 gene sequences from the BHV-1 isolates 34 and P8-2. The nucleic and deduced amino acid sequences from isolate 34 are numbered to the right and shown above those of P8-2 (Carpenter & Misra, 1991). With the exception of three regions denoting frameshifts, the latter are represented only where nucleotide mismatches and amino acid substitutions were observed.
Fig. 3. Sequencing gels of the BHV-1 34 VP8 gene representing the beginning and end of frameshifts found in isolate P8-2. Areas of sequence correspond to positions 349 to 363 (a), 1346 to 1356 (b), 1511 to 1525 (c), 2084 to 2098 (d) and 2153 to 2139 (e) in Fig. 2(a). Arrows point at regions of difference in the two sequences. In strain P8-2, extra residues were as follows: C and T at positions 357 and 358, A at residue 1353, C at residue 2093, and C and C at residues 2148 and 2149. Finally, a C residue was deleted at position 1517.

Significance of sequence variations

The nucleic acid sequence divergences described above have major effects on the predicted amino acid sequence, especially in the three frameshift regions. Overall, 14.1% of the deduced amino acid sequences were divergent. Nevertheless, it will be essential to establish whether these differences arose from strain variations or from sequencing errors.

Both sequences were compared with homologous polypeptide sequences derived from the HSV-1 UL47 (McGeoch et al., 1988) and EHV-4 ORF B6 (Whittaker et al., 1991) genes. Regions included in the three frameshifts yielded information relevant to the divergence observed. As shown in Fig. 2(b), the sequence derived from isolate 34 is more similar than that of P8-2 to those of the HSV-1 and EHV-4 counterparts. This was especially true for the third frameshift, in which 14 of 17 residues were conserved in strain 34, compared to a single one in P8-2. These findings may suggest that the P8-2 sequence is incorrect in these regions.

Transcript identification and mapping

The kinetics of expression of the transcript encoding the VP8 polypeptide were analysed by Northern blot hybridization using a 387 bp Smal–EcoRI fragment, positions 822 to 1209, as a probe. The probe detected an abundant 4.4 kb transcript which was present at a maximum between 18 and 24 h p.i. (Fig. 4a). The mRNA began to appear at 6 h p.i., as observed after prolonged exposure of the gel (data not shown). Our observations confirm that the VP8 transcript is expressed late during infection and indicate that it represents the true late (γ2) 4.5 kb transcript previously mapped to HindIII fragment M (Wirth et al., 1989). Finally, the size of the transcript confirms that no polyadenylation site is found downstream from the stop codon.

The precise location of the 5′ end of the VP8 mRNA was determined by primer extension analyses using a 25-mer oligonucleotide complementary to nucleotides 258 to 282, encompassing the translation initiation codon.
of the VP8 gene coding sequences (Fig. 4b). Reverse transcription of RNA isolated from BHV-1-infected cells resulted in the synthesis of a 61 base DNA fragment (lane I). In contrast, no DNA was synthesized from RNA isolated from mock-infected cells (lane M). These results indicate that the 5' extremity of the VP8 transcript is located 39 nucleotides upstream from the translation initiation codon or 25 nucleotides downstream from the putative TATA box. These findings expand upon those obtained by Carpenter & Misra (1991), in which the 5' end was located approximately 1100 bp upstream from position 1209, corresponding to the unique EcoRI site.

In conclusion, the major sequence divergences found in the VP8 protein genes from the two different BHV-1 isolates could provide further insight into the VP8 gene stability of BHV-1 isolates. Indeed, when the BHV-1 VP8 protein was compared with homologous polypeptides from HSV-1, EHV-4 and varicella-zoster virus (ORF 11; Davison & Scott, 1986), little conservation of sequences was observed. This suggests that genes encoding this tegument protein are subject to frequent mutation. This hypothesis is strengthened by the fact that no homology was found between these alphaherpesvirus proteins and those of either beta- or gammaherpesviruses. Thus, it will be of great interest to confirm the data for BHV-1 strain P8-2, and subsequently to clone and sequence VP8 genes from other isolates.

The authors thank Ms Janet Stewart for proofreading the manuscript. S.L. was supported by a studentship from the Natural Science and Engineering Research Council (NSERC), Canada. This work was funded by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (90-EQ-3021) and NSERC OGP0106053.

References


(Received 13 July 1992; Accepted 23 July 1992)