DNA sequence and transcriptional analysis of the glycoprotein M gene of murine cytomegalovirus

Anthony A. Scaizo,* Catherine A. Forbes, Nicholas J. Davis-Poynter, Helen E. Farrell and Paul A. Lyons†

Department of Microbiology, University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, WA 6907, Australia

We have characterized the gene encoding the murine cytomegalovirus (MCMV) homologue of the human cytomegalovirus (HCMV) UL100 open reading frame (ORF) that encodes the HCMV glycoprotein M (gM) molecule. It was identified based on its collinearity with MCMV homologues of the HCMV UL99, UL102, UL103 and UL104 ORFs which lie in the HindIII G fragment of the K181 strain of MCMV. Sequencing of a 2.3 kb EcoRI-BamHI subfragment of the EcoRI G fragment adjacent to the EcoRI A fragment revealed the presence of the complete MCMV gM ORF and two incomplete ORFs, which corresponded to homologues of HCMV UL99 and UL102. The MCMV gM ORF consists of 1059 nucleotides and is expressed as a 1-2 kb transcript at late times post-infection. To precisely characterize the gM transcript, the 5’ and 3’ ends were mapped. It was found that the transcript initiates at nucleotides 740 or 745, and that the site of polyadenylation at nucleotide 1961 occurs downstream of the second potential polyadenylation signal located at nucleotide 1934. Based on these findings the MCMV gM is predicted to consist of 353 residues and when compared with HCMV gM has a 47% level of identity. Of great interest is the finding that the MCMV gM amino acid sequence is completely conserved among six isolates of MCMV that had been shown to exhibit considerable variation both in the MCMV glycoprotein B and the immediate-early 1 gene-encoded pp89 molecule. Thus, this glycoprotein appears to be antigenically conserved.

Herpesviruses encode a number of glycoproteins which function in the processes of viral attachment, entry, membrane fusion, cell-to-cell spread and egress of mature virions from infected cells (for review see Spear, 1993). Among the three herpesvirus subfamilies there are four glycoprotein species which appear to be conserved in all members investigated thus far, and these are glycoprotein B (gB), glycoprotein H (gH), glycoprotein L (gL) and glycoprotein M (gM). The conservation of these glycoproteins indicates that they carry out important functions in the process of herpesvirus replication.

The herpesvirus gM homologues are characterized by the presence of multiple hydrophobic membrane-spanning domains (Lehner et al., 1989; MacLean et al., 1991; Baines & Roizman, 1993; Pilling et al., 1994) and while the functions of these proteins are not known, it has been speculated that they may form membrane ion channels. Insertion mutagenesis of the herpes simplex virus type 1 (HSV-1) UL10 gene, which encodes gM, has revealed that this protein is not essential for virus replication in vitro, although there is a 10-fold reduction in virus replication in susceptible cell lines (Baines & Roizman, 1991). Disruption of the HSV-1 gM gene has also been shown to result in a decrease in virus-induced syncytium formation (Davis-Poynter et al., 1994). However, the effect of deletion of HSV-1 gM on viral pathogenesis in vivo and on growth in various tissues was not determined. The human cytomegalovirus (HCMV) gM homologue, which is encoded by the UL100 ORF (Chee et al., 1990), has been shown to be a 45 kDa viral structural protein (Lehner et al., 1989). It encodes a component of the gC-II glycoprotein complex recognized by Group 2 monoclonal antibodies (MAbs; Kari et al., 1994). This class of glycoproteins defined by the Group 2 MAbs has been shown to possess heparin-binding activity which led to the suggestion that they play a role in viral attachment (Kari & Gehrz, 1993). However, the precise role of HCMV gM in determining infectivity is not known.

While a number of recent studies have described the
Short communication

identification and characterization of the murine cytomegalovirus (MCMV) genes encoding gB (Rapp et al., 1992), gH (Xu et al., 1992, 1994a; Rapp et al., 1994) and gL (Xu et al., 1994a), the gene encoding the MCMV gM homologue has not been characterized. In this study we describe its identification, based on its positional conservation with the MCMV homologues of UL99 (Cranmer et al., 1994), and UL102, UL103 and UL104 (Lyons et al., 1994), together with an analysis of the MCMV gM transcript.

The viral glycoproteins, because of their surface expression, have been considered to be useful candidates for inclusion in vaccines (Rasmussen, 1990; Spaete, 1991). However gB, which is known to elicit a protective immune response, exhibits sequence variation in both expression, have been considered to be useful candidates for inclusion in vaccines (Rasmussen, 1990; Spaete, 1991). However gB, which is known to elicit a protective immune response, exhibits sequence variation in both...
vaccine candidate if variants escape the immune surveillance elicited by the immunizing gB sequence. Similarly, the MCMV ial-1-encoded pp89, which elicits an immunodominant cytotoxic T lymphocyte (CTL) response, also exhibits amino acid sequence heterogeneity in the minimal CTL epitope leading to reductions in CTL recognition (P.A. Lyons, C. Carrello, G.R. Shellam & A.A. Scalzo, unpublished). In the present study we have investigated the extent of sequence heterogeneity in MCMV gM and found complete amino acid sequence conservation between MCMV strains.

Previous studies have localized the MCMV homologues of HCMV UL99, UL102, UL103 and UL104 to the MCMV HindIII G fragment (Lyons et al., 1994; Cranmer et al., 1994), suggesting that this region was collinear with HCMV. Hence, to identify and characterize the UL100 or gM ORF of the pathogenic MCMV strain K181, a 2.3 kb EcoRI–BamHI subfragment of the EcoRI G fragment adjacent to the EcoRI A fragment was targeted for sequence analysis. Size-fractionated, sonicated fragments of the EcoRI–BamHI subfragment were shotgun ligated into the Smal site of Bluescript SK(−) (Stratagene) using the method of Bankier et al. (1987). Following sequencing of the clones, using an ABI PRISM dye-deoxy kit and an ABI 373A automated sequencer (Perkin-Elmer Cetus), the complete sequence was assembled using the MacVector V3.5 sequence analysis program (IBI). A total of 37 clones of average length of 400 bp were sequenced in both orientations giving an average depth of sequence determination of approximately 12-fold. The EcoRI–BamHI subfragment was 2291 bp in length and contained one complete ORF and two partial ORFs (Fig. 1; Genome Sequence Data Base accession number L41088).

Comparison of the translated protein products of these ORFs with those present in the sequence databases using FASTA (Pearson & Lipman, 1988) revealed that the complete ORF shows high levels of identity with HCMV gM (optimized FASTA score 1026, 47% identity) and human herpesvirus 6 (HHV-6) gM (optimized FASTA score 979, 42% identity), and lower levels of identity with the gM homologues of herpesvirus saimiri, Epstein–Barr virus, equine herpesvirus type 1 and HSV-1 (FASTA scores 495, 461, 325 and 211, respectively). The incomplete ORF, spanning nucleotides 2291–2055, which is in a tail-to-tail arrangement with the gM homologue, has been previously identified by Cranmer et al. (1994) as representing the MCMV homologue of HCMV UL99, a finding which is also consistent with its positional homology (Cranmer et al., 1990). The other incomplete ORF, spanning nucleotides 507–1, shows a low level of identity with the HCMV UL102 ORF (FASTA score 80), and the sequence shown represents the 5’ end of the previously identified MCMV homologue (Lyons et al., 1994). The MCMV UL102 homologue (designated m102) extends further upstream than the originally proposed HCMV UL102 ORF. A Clustal V alignment of the deduced m102 amino acid sequence determined in this study (172 residues) with the deduced amino acid sequence of the HCMV nucleotide sequence in the region upstream of the previously identified first methionine of UL102 (Chee et al., 1990) is shown in Fig. 2. It reveals that while there is an in-frame stop codon at nucleotide 146744, conservation exists between the two deduced sequences for a region encoding 158 HCMV codons that are in-frame with the previously identified UL102. This suggests that the HCMV UL102 ORF may initiate from an ATG upstream of that previously identified. The finding that the HCMV sequence around the upstream methionine codon AGGATGA more closely resembles a Kozak ATG than that of the downstream codon GCCATGTT (Kozak, 1986) supports this concept. No evidence could be found for the existence of a MCMV homologue of the putative HCMV UL101 ORF.

The predicted gM ORF extends from nucleotides 756–1814 and is thus 1059 nucleotides in length. Although there is an upstream in-frame methionine at nucleotide 702, the nucleotide sequence around the methionine at base 756 (CACCTGCACGATGGC) most closely matches the consensus translation initiation site for vertebrate viruses (Cavener & Ray, 1991; Kozak, 1986). Furthermore, alignment between the gM amino acid sequences of MCMV, HCMV and HHV-6 is more consistent with an initiation site located at base 756 (data not shown). A potential TATA-box with the sequence TATTGA exists at nucleotides 712–718. In addition, there are three potential polyadenylation signals at nucleotide positions 1920, 1934 and 1990.

The primary translation product of the MCMV gM homologue is predicted to be 335 residues in length with an unprocessed protein size of 40.1 kDa. Hydrophobicity analysis of the predicted protein product of MCMV gM (data not shown) revealed that like HCMV gM (Lehner et al., 1989), the MCMV homologue has eight hydrophobic regions separated by small stretches of hydrophilic residues, and therefore has the characteristics of being a multiply-membrane spanning protein. Studies on the glycosylation of the HCMV gCII glycoproteins as defined by Group 2 MAbss indicated that these UL100-encoded proteins (Kari et al., 1994) express N-linked, but not O-linked carbohydrate side-chains (Kari & Gehrz, 1988, 1993). The MCMV gM homologue encodes four potential N-linked glycosylation sites. Two of the potential MCMV gM glycosylation sites at residues 62–64 (NLT) and 119–121 (NQS) are conserved in terms of location with sites present in HCMV gM (Lehner et al., 1989). Moreover, by inspection of the aligned
Fig. 2. Comparison of the deduced amino acid sequences of the 5′ region of the HCMV and MCMV UL102 proteins. The amino acid sequence shown for HCMV represents a translation of the HCMV strain AD169 nucleotide sequence upstream of the originally proposed UL102 ATG which is located at nucleotide 147128 (Chee et al., 1990). The translation shown represents nucleotides 146522–146995 from the HCMV sequence and includes an in-frame stop codon at nucleotide 146744. It is represented in the HCMV amino acid sequence shown above by a hash sign (#). For the comparison of the two sequences above identical amino acids are indicated by an asterisk (*) and conserved amino acids are indicated by a dot (.). The first methionine for the m102 ORF represents the ATG codon at nucleotide 507 (see Fig. 1). The first methionine for the HCMV UL102 translation represents the ATG codon at nucleotide 146552 (Chee et al., 1990).

Fig. 3. Expression kinetics of the MCMV gM and m102 genes. (a) Northern blot analysis using a 32P-labelled HincII fragment (nucleotides 729–1117; Fig. 1) as a probe. RNA was extracted from uninfected mouse embryo fibroblasts (MEF), or from cells at 2, 4, 8, 12, 16, 20 or 24 h p.i. with the K181 strain of MCMV, or from cells at 24 h p.i. in the presence of the metabolic inhibitor PAA. (b) Northern blot analysis using a 32P-labelled 3.3 kb XbaI fragment from nucleotide 1659 in the gM gene through into the m102 ORF. RNA was extracted from uninfected MEF, or from cells at 2, 4, 8, 16 or 24 h p.i., or from cells at 24 h p.i. in the presence of PAA.

sequences (data not shown), the site-spanning residues 62–64 are conserved in position in all of the herpesviral gM homologues sequenced so far, a finding consistent with previous observations (Baines & Roizman, 1993).

The transcription of the MCMV gM homologue was assessed by Northern blot analysis using a 388 bp HincII fragment that represents the 5′ region of the ORF (nucleotides 729–1117; See Fig. 1). This probe detected a unique transcript of approximately 1.2 kb expressed at 20 and 24 h post-infection (p.i.; Fig. 3a), indicating that there are no overlapping transcripts in this region. The transcript was not detected in RNA extracted from cells
Fig. 4. Primer extension analysis of the 1.2 kb gM transcript. Primer extension analysis was performed using avian myeloblastosis virus reverse transcriptase (Promega) with RNA isolated from uninfected MEF or MCMV-infected MEF at 24 h p.i. and using a 32P-labelled oligonucleotide primer complementary to the sense strand (5' CATCGTCAGGTTTCATGTTGTCG 3', nucleotides 950–929; Fig. 1). The extension products were electrophoresed on a 6% denaturing polyacrylamide gel. The extension products align with the nucleotides 740 and 745 (marked by arrows) indicated by the DNA sequence analysis performed using the same oligonucleotide primer and a MCMV genomic clone spanning nucleotides 1–1011 (Fig. 1). End-labelled HpaII fragments of the plasmid vectors pBK-CMV (Stratagene) and pGEM-11Zf(+) (Promega) were included as size standards.

infected in the presence of the DNA polymerase inhibitor phosphonoacetic acid (PAA; Fig. 3a), indicating that the gM ORF is expressed as a true late transcript. This finding was confirmed using a larger XhoI fragment as a probe. This probe spans from nucleotide 1659 in the gM ORF to the m102 ORF (total probe length 3.3 kb). Fig. 3(b) again shows that the 1.2 kb gM transcript is expressed as a true late protein. The 2.6 kb band probably represents the m102 transcript, and the higher molecular mass transcript, present only at 24 h p.i., is likely to represent the MCMV UL104 homologue, since this transcript has been previously shown to overlap with m102 (Lyons et al., 1994). In contrast to our previous observations (Lyons et al., 1994) suggesting that the m102 transcript is expressed as a leaky late gene, its detection at early times p.i. (4 h) and at 24 h p.i. in the presence of PAA (Fig. 3b) indicates that it should be reclassified as an early gene.

To determine the location of the 5' initiation site of the MCMV gM transcript, we employed primer extension analysis using a 32P-labelled oligonucleotide primer complementary to the sense strand (5' CATCGTCAGGTTTCATGTTGTCG 3', nucleotides 950–929; Fig. 1). RNA from cells late after MCMV infection (24 h p.i.) or from uninfected cells, was reverse transcribed and the products were electrophoresed on a 6% denaturing polyacrylamide gel, together with a sequencing ladder that was generated using the same primer as that used for the primer extension assay (Fig. 4). Two major extension products were obtained with RNA from infected cells (Fig. 4, lane marked ‘Infected’), corresponding to transcriptional start sites occurring at nucleotides 740
and 745. These results confirmed the conclusions made after RNAase protection experiments using a riboprobe from nucleotides 638–1042 (Fig. 1), for which an approximately 120 base reduction in probe length was observed following protection with RNA from cells at late times p.i., indicating a start site at approximately nucleotide 760 (data not shown). The transcription initiation sites identified by the primer extension analysis at nucleotides 740 and 745 are downstream of the first in-frame methionine at nucleotide 702, and upstream of the second in-frame methionine at nucleotide 756 (Fig. 1). These sites are approximately 30 bp downstream of the potential TATA-box sequence TATTTGA (nucleotides 712–718), which is consistent with the location of transcription start sites relative to the TATA-box (Bucher, 1990). Hence, the primer extension data support the prediction that the putative MCMV gM protein commences translation from the ATG at nucleotide 756.

To map the 3’ end of the transcript the technique described by Frohman et al. (1988) for rapid amplification of cDNA ends (RACE) was employed. A NotI-oligo(dT)$_{15}$ primer adaptor (Promega) was used to generate first strand cDNA from RNA from cells harvested at late times p.i. The products of the first-strand cDNA synthesis reaction were then amplified using the NotI-oligo(dT)$_{15}$ primer adaptor and a MCMV gM-specific oligonucleotide primer (5′ GACA-TCAACATCGGCATCACGGT 3′) corresponding to nucleotides 1665–1687 (Fig. 1). A single RT–PCR product of approximately 300 bp was obtained (data not shown). This product was then cloned into the pGEM-T vector (Promega), and the inserts were sequenced using SP6 and T7 primers with an ABI PRISM dyedeoxy kit and an ABI 373A automated sequencer. Four independent clones were sequenced on both strands and all had identical sequences to that shown in Fig. 1 from nucleotides 1687–1691. However, each clone differed in the length of the poly(A) tail, which varied from 16–90 bp. The point of addition of the poly(A) tail at nucleotide 1961 is immediately upstream of the G/T-rich sequence CATGTTGT (nucleotides 1961–1968; Fig. 1), which closely resembles the consensus transcription termination motif YGTGTTYY (Birnstiel et al., 1985). This site is 21 bp downstream of the second potential polyadenylation motif AATAAA (nucleotides 1934–1939). The locations of the sites of initiation and termination of transcription at nucleotides 740/745 and 1961, respectively, are consistent with the observed transcript size of approximately 1.2 kb obtained by Northern blot analysis (Fig. 3).

To assess the level of genetic heterogeneity in the gM gene, we selected six MCMV field isolates – G4, G6, K17B, K17E, K29 and N1 (Booth et al., 1993) – for gM sequence analysis. The strains selected were those found to display sequence variation in gB and iel (J. Xu and others, unpublished; P. A. Lyons and others, unpublished). The gM gene from each of these isolates was amplified by PCR with the specific primers (5′ GGACGGTCTCTCTAAGG 3′; nucleotides 685–701) and (5′ ACATGGTGACAAACAC 3′; nucleotides 1979–1967). The PCR products of each strain were then used as templates for direct automated cycle sequencing using the ABI PRISM kit. The gM ORF of each isolate was sequenced in its entirety by using six specific oligonucleotide primers, to facilitate sequencing of both strands. These were (5′ CGTGAGGACCATGAAACC 3′; nucleotides 788–804), (5′ GATCGTGGCCGTGCAGTAC 3′; 1508–1526), (5′ GACATCAACATCGCACCTACACGGT 3′; nucleotides 1665–1687), (5′ CATGCATTCTACATGTGTCG 3′; nucleotides 950–929), (5′ CGACACCTTCATGTGACG 3′; nucleotides 1571–1555) and (5′ ACATGGTGACAAACAC 3′; nucleotides 1979–1967). Only two nucleotide substitutions were observed in the entire gM ORF (Genome Sequence Data Base accession numbers: isolate K29, L41089; G4, L41090; G6, L41091; K17B, L41092; K17E, L41093; and N1, L41094). The strains G4 and G6 had a C → T substitution at nucleotide 926 and strains G4, G6, K17B, K17E and N1 had a T → C substitution at nucleotide 1760. Both of these substitutions were silent and did not lead to changes in the encoded amino acids. Hence, the MCMV gM protein sequence is completely conserved among strains that have been previously shown to vary in gB and iel.

As HCMV glycoproteins have been proposed as candidates for inclusion in HCMV subunit vaccines (Rasmussen, 1990; Spaete, 1991), the evaluation of the value of specific glycoproteins in inducing protective responses, mediated by either neutralizing antibodies or CTLs, is an important component of this process. The MCMV mouse model provides a useful model for studying the ability of a range of potential target proteins to induce protection. The finding in the present study that MCMV gM is completely conserved among viral strains, which have been shown to differ considerably in gB (J. Xu and others, unpublished) and pp89 (P. A. Lyons et al., unpublished), indicates that further studies to investigate the extent of HCMV UL100 sequence heterogeneity among clinical isolates and the ability of MCMV gM to induce protective neutralizing antibodies or CTLs are warranted. The results of these studies may have important implications for HCMV vaccine design.

We would like to thank Dr Jane E. Allan for her critical comments on this manuscript. This work was supported by a National Health and MRC Program grant no. 91 3205.

Note added in proof. Since the submission of this paper the sequence of MCMV gM (Smith strain) has been published (Li et al., Virus Research 36, 163–175, 1995).
References

Baines, J. D. & Roizman, B. (1991). The open reading frames UL3, U1,0, U1,10, and U1,16 are dispensable for the replication of herpes simplex virus 1 in cell culture. *Journal of Virology*, 65, 938–944.


(Received 4 April 1995; Accepted 22 June 1995)