Extensive sequence variation exists among isolates of murine cytomegalovirus within members of the *m02* family of genes

Alexandra J. Corbett,[1,2] Catherine A. Forbes,[1,2] Dorian Moro[3]† and Anthony A. Scalzo[1,2]

[1]Immunology and Virology Program, Centre for Ophthalmology and Visual Science, University of Western Australia, Nedlands, WA 6009, Australia

[2]Centre for Experimental Immunology, Lions Eye Institute, 2 Verdun Street, Nedlands, WA 6009, Australia

[3]School of Natural Sciences, Edith Cowan University, Joondalup, WA 6027, Australia

Correspondence
Anthony A. Scalzo
scals@cyllene.uwa.edu.au

Received 5 October 2006
Accepted 16 November 2006

Murine cytomegalovirus (MCMV) is a widely used model for human cytomegalovirus (HCMV) and has facilitated many important discoveries about the biology of CMVs. Most of these studies are conducted using the laboratory MCMV strains Smith and K181. However, wild-derived isolates of MCMV, like HCMV clinical isolates, exhibit genetic variation from laboratory strains, particularly at the ends of their genomes in areas containing known or putative immune-evasion and tropism genes. This study analysed the nucleotide sequence of the *m02–m05* region, within the *m02* gene family, of a number of laboratory and wild-derived MCMV isolates, and found a large degree of variation in both the sequence and arrangement of genes. A new open reading frame (ORF), designated *m03.5*, was found to be present in a number of wild isolates of MCMV in place of *m03*. Two distinct isolates, W8 and W8211, were found to possess both *m03* and *m03.5*. Both *m03* and *m03.5* had early transcription kinetics and the encoded proteins could be detected on the cell surface, consistent with a possible role in immune evasion through binding to host-cell proteins. These data show that gene duplication and sequence variation occur within different isolates of MCMV found in the wild. As this variation among strains may alter the function of genes, these findings should be considered when analysing gene function or host–virus interactions in laboratory models.

INTRODUCTION

Murine cytomegalovirus (MCMV) has been used extensively as a model for human cytomegalovirus (HCMV) infection, as strict species specificity makes *in vivo* HCMV studies difficult. CMVs are betaherpesviruses with slow replication cycles that are capable of persisting for the lifetime of their respective hosts (Ho, 1991). These viruses have large double-stranded DNA genomes ($\sim$230 kb) and encode many immune-modulating proteins. The two viruses share many biological characteristics and considerable genetic homology, and the MCMV model has facilitated many important discoveries about the pathology and the immune control of CMV infections.

Most MCMV studies are conducted using the laboratory strains Smith or K181, both of which have been passaged extensively *in vitro* since their isolation. Variations in the genetic and biological properties of these strains have been reported (Boname & Chantler, 1992; Hudson *et al.*, 1988; Misra & Hudson, 1980). In addition, CMVs have evolved alongside their hosts over millions of years, resulting in significant genetic and biological variation among MCMV wild-derived isolates (Booth *et al.*, 1993; Lyons *et al.*, 1996; Smith *et al.*, 2006; Voigt *et al.*, 2003; Xu *et al.*, 1996), as well as extensive genetic variation among HCMV clinical isolates, which has implications for the immune control and pathogenesis of different HCMV strains (reviewed by Pignatelli *et al.*, 2004).

The MCMV Smith strain is predicted to encode at least 170 open reading frames (ORFs) and, although the sequence of the Smith MCMV genome has been determined (Brocchieri *et al.*, 2005; Rawlinson *et al.*, 1996), many of these have not been characterized. Homologues of approximately 80 HCMV ORFs are encoded in the central region of the MCMV genome,

†Present address: Chevron Australia Pty Ltd, QV1 Building, 250 St George's Terrace, Perth, WA 6000, Australia.

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this study are EF031243–EF031268.

A supplementary table showing primers used for RT-PCR analysis of *m03*, *m03.5* and *m04* is available in JGV Online.

 0008-2623 © 2007 SGM Printed in Great Britain

whereas the left and right ends are not homologous to HCMV. These non-homologous regions of both viruses contain genes that are 'non-essential' for growth *in vitro* and include a number of clearly defined immune-evasion functions (reviewed by Alcami & Koszinowski, 2000; Hengel *et al.*, 1998). Genetic variation in these regions therefore has implications for immune control of these viruses in their respective hosts. The functions of a number of these immune-evasion genes are also likely to be host-genotype specific. This is perhaps not widely appreciated, as most MCMV studies analyse the laboratory strains Smith and K181 in inbred strains of mice. Such model systems may not accurately represent the spectrum of interactions in the natural hosts.

The MCMV genome contains two families of glycoproteins (the *m02* and *m145* gene families) at opposite ends of the genome. The *m02* gene family includes *m02–m16* and forms three subgroups (*m02–m06*, *m07–m10* and *m11–m16*) based on conserved cysteine residues (Rawlinson *et al.*, 1996). This gene family is 'non-essential' for MCMV growth *in vitro*, and members play a role in natural killer (NK) cell-mediated immune escape *in vivo* (Oliveira *et al.*, 2002) and in subverting T-cell responses (reviewed by Hengel *et al.*, 1999). The *m04* gene product (gp34) binds to major histocompatibility complex (MHC) class I molecules in the endoplasmic reticulum (ER) and on the cell surface (Kavanagh *et al.*, 2001; Kleijnen *et al.*, 1997). The *m06* gene encodes a protein that downregulates the surface expression of MHC class I proteins by redirecting them to lysosomal compartments (Reusch *et al.*, 1999). The functions of *m02*, *m03* and *m05* have not been determined. When *m02* alone is knocked out, there is no substantial phenotype *in vivo*, suggesting that this gene has no significant immune-evasion role in the C57BL/6 mouse strain (Oliveira *et al.*, 2002). Thus, some *m02* family members have functions that have been defined by studies in inbred mice, whereas the roles of others in host–virus interactions are not evident based on the limited analyses performed so far.

In the current study, we compared the sequence of the *m02–m05* gene region of seven MCMV strains, comprising the laboratory strains Smith and K181 and five wild-derived isolates. These viruses were found to have three distinct gene arrangements within this region. We have described a new ORF, designated *m03.5*, which is present in a number of distinct MCMV isolates either in place of, or in addition to, *m03*. Individual ORFs were sequenced for additional wild-derived isolates and showed a high degree of variation within some genes. The transcriptional kinetics and surface expression of *m03* and *m03.5* were determined. Our observations raise the possibility that the sequence variation in these genes has been selected for by functional binding to heterogeneous host-cell surface molecules.

METHODS

Viruses and viral DNA preparation. The MCMV laboratory strains used were Smith (obtained from E. S. Mocarski, Stanford University, USA) and K181-Perth (abbreviated to K181). A number of wild-derived isolates (G1F, G2, G3B, G4, G5, G6, K4, K6, K17B, N1, N5, W2, W8 and W8211) have been described previously (Booth *et al.*, 1993). Isolates MI2A, MI2C, MI3A, MI4A, MI6A, MI4B, MI4C, MI6C, TR8A and TR8B were derived by sequential limiting dilution purification from salivary glands from feral house mice trapped on Macquarie Island, Australia (54° 30′ S 158° 57′ E). Tissue culture virus stocks were produced by propagation in mouse embryonic fibroblasts (MEFs) and titres determined by standard plaque assay, as described previously (Farrell & Shellam, 1989). Viral DNA was produced from infected MEFs, as described previously (Xu *et al.*, 1996).

Cell culture. COS-7 cells were maintained in Dulbecco's modified Eagle's medium (Invitrogen) containing 10 % fetal calf serum (Invitrogen), 2 mM glutamine, 1 mM sodium pyruvate, 100 U penicillin ml^{-1} and 40 μg gentamicin ml^{-1}. Primary MEFs were produced by trypsin dispersion of 15–17-day-old embryos from outbred ARC/S mice (Animal Resources Centre, Murdoch, Western Australia), as described previously (Chalmer *et al.*, 1977), and were maintained in minimal essential medium (MEM; Invitrogen) supplemented with 10 % normal calf serum (NCS; Invitrogen). Immediately prior to infection, the culture medium was replaced with MEM supplemented with 2 % NCS.

Sequencing analysis. In order to sequence the complete *m02–m05* region, overlapping ~1–3 kb amplicons were produced using primers based on the published Smith sequence and subsequently on sequences obtained from the wild-derived isolates. PCR amplicons from duplicate reactions were sequenced using the Applied Biosystems Big-Dye Terminator version 3.1 sequencing kit, and at least two exact matches from separate reactions were obtained at each nucleotide position. Sequences obtained from K181 or wild-derived isolates were compared with the published Smith sequence (GenBank accession no. U68299; Rawlinson *et al.*, 1996). Sequence alignments (CLUSTAL W), translations, sequence identity and similarity matrices (BLOSUM62) and phylogenetic analyses (neighbour-joining method) were performed using the BioEdit Sequence Alignment Editor suite of programs (Isis Pharmaceuticals; Hall, 1999). Signal peptide, transmembrane (TM) region and *N*-linked glycosylation site predictions were performed using SignalP version 3.1 (Bendtsen *et al.*, 2004; Nielsen & Krogh, 1998; Nielsen *et al.*, 1997), HMMTOP version 2.0 (Tusnady & Simon, 1998, 2001) and CBS NetNGlyc 1.0 servers, respectively. Annotation of ORFs was based on the published annotation for the Smith MCMV strain (Rawlinson *et al.*, 1996). The initiation ATG codon used for analysis was predicted based on conservation among strains in each case. In all cases, predicted poly(A) motifs, which closely matched the consensus AATAAA, were found downstream of the predicted ORFs.

Analysis of transcription kinetics by RT-PCR. For analysis of the transcription kinetics of *m03*, *m03.5* and *m04*, MEFs were infected with Smith, MI6A or W8211 tissue culture virus stock with centrifugal enhancement, resulting in an m.o.i. of 3. Total RNA was extracted using RNA-Bee (Tel-Test) from MEFs harvested at 2, 4, 8 and 24 h post-infection (p.i.) or uninfected MEFs. Cells harvested at 8 h were infected in the presence of phosphonoacetic acid (PAA; 250 μg ml^{-1}). Cells harvested at 4 and 24 h p.i. were infected in both the presence and absence of cycloheximide (50 μg ml^{-1}) and PAA, respectively. Total RNA was treated with DNA-free (Ambion) to remove contaminating DNA. First-strand cDNA synthesis was performed using an oligo-d$(T)_{17}$ adaptor primer [5′-GACTCGACGTCGACATCGA$(T)_{17}$-3′; Frohman *et al.*, 1988] and avian myeloblastosis virus reverse transcriptase (RT; Promega). RT-PCR was performed on RT^+ and RT^- samples for each gene using specific primers. *ie1*-specific PCR was included as a control for immediate-early (IE) transcription. *β-Actin* RT-PCR was included as a loading control. The primers used are listed in Supplementary Table S1, available in JGV Online.

Construction of FLAG-tagged proteins. In order to construct N-terminal FLAG-tagged proteins, the MI6A *m04* ORF (minus the predicted signal peptide) was amplified from viral DNA using primers containing *Bam*HI and *Nru*I sites and cloned into the pEF-5CMVLFlag vector (Voigt *et al.*, 2003), which contains a human CMV leader peptide. The MI6A CMV leader FLAG–*m04* cassette was then reamplified from this construct with flanking primers containing *Xho*I and *Eco*RI sites to facilitate cloning into the MigR1 vector (Pear *et al.*, 1998). The MI6A *m04* ORF in the MigR1 construct was then replaced with the *m03* (Smith or W8211), *m03.5* (MI6A or W8211) or *m04* (Smith, K181, G4, G1F or W8211) ORF amplified using forward primers containing a *Bam*HI or *Bcl*I site and reverse primers containing an *Eco*RI or *Eco*RV site.

Analysis of expression of FLAG-tagged proteins by flow cytometry. COS-7 cells (2×10^5) were transiently transfected with 1.5 μg plasmid DNA using FuGene 6 transfection reagent (Roche Diagnostics). At 2 days post-transfection, cells were detached with 50 mM EDTA and blocked with 10 % normal goat serum before being stained with anti-FLAG monoclonal antibody (mAb) (mouse M2 clone; Immunex Corporation), followed by Cy5-conjugated anti-mouse IgG (#115-176-062; Jackson Laboratories) and analysed by flow cytometry using a BD FACSCanto (BD Biosciences). Live cells were gated based on exclusion of 7-aminoactinomycin D (Sigma) and green fluorescent protein (GFP) expression was used to identify transfected cells.

Analysis of protein expression by immunofluorescence. For immunofluorescence analyses using FLAG-tagged proteins, COS-7 cells (5×10^4) seeded on glass coverslips were transiently transfected with 0.5 μg plasmid DNA using FuGene 6 transfection reagent. At 2 days post-transfection, cells were washed with PBS, fixed with 4 % paraformaldehyde and permeabilized with 0.1 % Triton X-100 before being stained with anti-FLAG mAb M2, followed by biotinylated anti-mouse IgG (#115-066-071; Jackson Laboratories) and streptavidin–Alexa Fluor 546 (#S-11225; Invitrogen). Cells were analysed by standard epifluorescence and images were captured digitally using a Zeiss Axioskop microscope and Northern Eclipse image analysis software (Empix Imaging).

RESULTS

Sequence variation and gene duplication in the *m02–m05* region

In order to determine the extent of genetic variation among MCMV strains within the *m02–m05* region, overlapping PCR amplicons were produced and sequenced from duplicate PCRs. This region was found to display a high degree of variation in both the DNA and protein sequences. In particular, the *m03–m04* gene region of MCMV shows marked sequence variation. In addition, the gene arrangement in this region was found to differ among strains. In the laboratory strains Smith and K181, which have previously been completely or partially sequenced, the *m02*, *m03*, *m04* and *m05* genes lie in a head-to-tail fashion. In contrast, there was significant variation in some of the wild-derived MCMV isolates (Fig. 1a). Of the five wild-derived isolates for which the complete *m02–m05* region was sequenced, only G4 had the same gene arrangement as the laboratory strains. Two of the isolates analysed, W8 and W8211, possessed *m03* and also contained another ORF in tandem, which we designated *m03.5* because it is most closely related to *m03* based on comparison with the Smith sequence. In contrast,

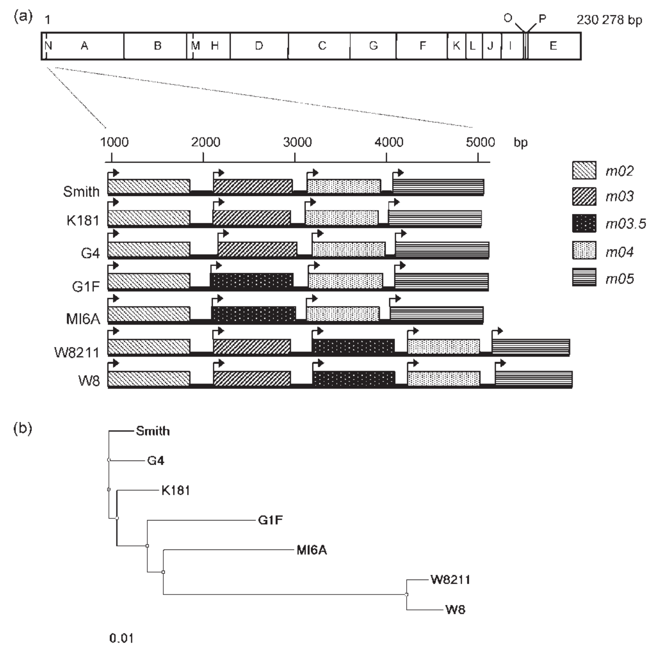

Fig. 1. (a) Schematic of the *m02–m05* region ORF arrangement for the Smith (GenBank accession no. U68299), K181, G4, G1F, MI6A, W8211 and W8 strains of MCMV. The upper part of the figure shows the Smith *Hin*dIII digestion map (Ebeling *et al.*, 1983). Homologous ORFs are indicated by pattern. Arrows show the direction of translation. (b) Phylogenetic tree for the *m02–m05* region nucleotide sequence for seven MCMV strains.

MI6A and G1F did not possess *m03* and only contained *m03.5*. These gene arrangement groupings were reflected in a phylogenetic analysis based on the complete DNA sequence across this region (Fig. 1b).

Conservation of the *m02* ORF

All seven MCMV strains sequenced in this study were found to contain an *m02* ORF. Amino acid sequence identity with the published Smith m02 sequence (Rawlinson *et al.*, 1996) ranged from 92.9 % for the G1F isolate to 97.9 % for the K181 strain (data not shown). The sequences of the predicted signal peptides, TM domains and cysteine motifs were mostly conserved among strains.

Sequence comparison of the *m03* and *m03.5* ORFs

Of the five wild-derived isolates analysed, only G4 had the same *m02*, *m03*, *m04* and *m05* gene arrangement as the two laboratory strains (Fig. 1a). Two isolates, W8 and W8211, had both *m03* and an additional ORF, *m03.5*, which lay between *m03* and *m04*. All other strains analysed had only *m03.5* and the *m03* ORF was absent. Sequence comparisons of m03 showed that amino acid sequence identity with Smith m03 ranged from 86.4 to 93.0 % (Fig. 2a). For all

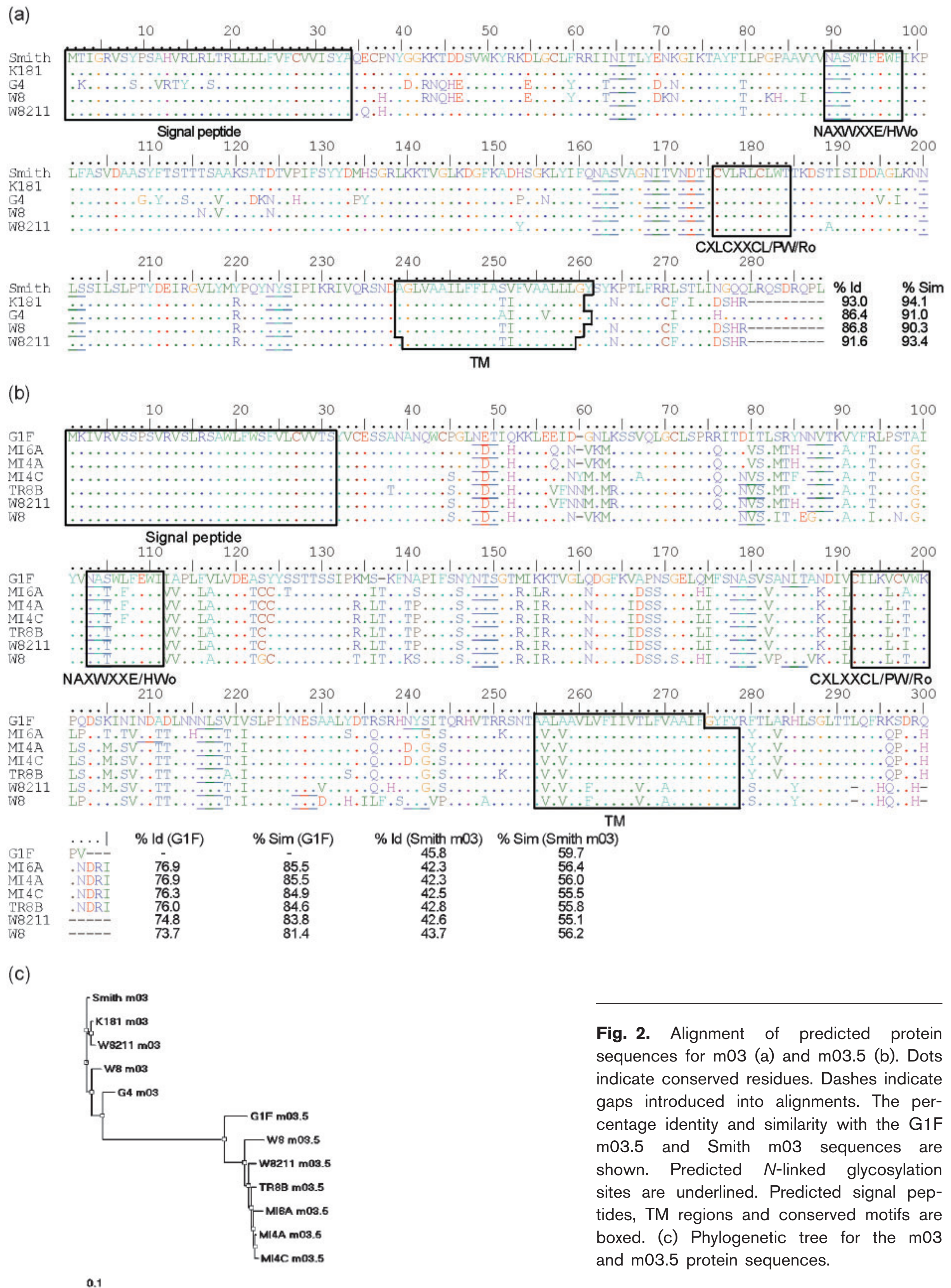

Fig. 2. Alignment of predicted protein sequences for m03 (a) and m03.5 (b). Dots indicate conserved residues. Dashes indicate gaps introduced into alignments. The percentage identity and similarity with the G1F m03.5 and Smith m03 sequences are shown. Predicted *N*-linked glycosylation sites are underlined. Predicted signal peptides, TM regions and conserved motifs are boxed. (c) Phylogenetic tree for the m03 and m03.5 protein sequences.

strains containing this ORF, the encoded protein was predicted to be a type I TM glycoprotein, with an N-terminal signal peptide of 33 aa. The NAXWXXE/HWo and CXLXXCL/PW/Ro motifs described previously (Rawlinson *et al.*, 1996), as well as the seven predicted *N*-linked glycosylation sites in m03, were conserved in all strains analysed.

Isolates W8 and W8211 were found to contain two genes between *m02* and *m04*. We considered this second ORF to be a new gene and not simply a highly variant form of *m03*, as it was present in addition to *m03* in at least two independent MCMV isolates previously shown to have distinct RFLP profiles (Booth *et al.*, 1993), and have given it the designation *m03.5*, as it has closest homology to *m03* based on nucleotide and protein BLAST searches of the GenBank database among all MCMV ORFs. A complete *m03.5* ORF sequence was obtained for seven isolates (Fig. 2b). In addition to those strains for which the complete *m02–m05* region was analysed, 11 more MCMV wild-derived isolates have so far been found to encode an *m03.5* ORF (data not shown). All strains analysed so far encoded at least one of the *m03* or *m03.5* ORFs. The sequence similarity suggests that these have arisen via a gene duplication event and that most strains have subsequently lost one ORF. Although *m03.5* is most closely related to *m03* based on BLAST analysis, these genes are quite distinct (Fig. 2c).

The m03.5 predicted protein contains the NAXWXXE/HWo and CXLXXCL/PW/Ro motifs typical of the *m02* gene family (Rawlinson *et al.*, 1996). Amino acid sequence identity (with G1F m03.5) ranged from 73.7 to 76.9 % (Fig. 2b). Sequence identity with Smith m03 ranged from 42.3 to 45.8 % (and similarity ranged from 55.1 to 59.7 %), again suggesting that *m03.5* is a distinct gene. Two possible initiation ATG codons were present, neither of which matched the Kozak consensus sequence ACC**ATG**G (Kozak, 1986). The second had homology with *m03* and therefore this was used for the alignment. Six *N*-linked glycosylation motifs were conserved among all of the strains, with five more being variably present among these strains. The predicted signal peptide and TM regions were conserved, except for G1F m03.5, which had a shorter predicted TM region. The cytoplasmic tail was truncated in the W8, W8211 (by seven residues) and G1F (by three residues) isolates.

m04/gp34 sequence comparison

The *m04*/gp34 sequence was found to be highly variable among the MCMV strains analysed, with amino acid sequence identity with Smith *m04*/gp34 ranging from 39.4 to 93.6 % (Fig. 3a). Indeed, in order to sequence *m04*, primer sets based on the *m03* and *m05* Smith sequences were used to amplify larger products, as the region directly flanking *m04* was highly variable and Smith-based *m04* primers did not work for many strains. In a recent report, Smith *et al.* (2006) also failed to amplify the *m04* gene in wild-derived MCMV isolates with strain Smith-based primers directly flanking the gene, suggesting that these isolates also contained *m04* sequences that were highly divergent from the Smith strain. The *m04*/gp34 sequences fell into six distinct groups (Fig. 3b). The type I TM structure was conserved with the exception of the MI6A and N1 isolates, which each had two predicted TM regions. However, based on the hydrophobicity plots and comparison with the very similar isolates G3B and MI2A, we believe this prediction of additional TM domains using the HMMTOP version 2.0 server (Tusnady & Simon, 1998, 2001) to be incorrect. Most of the *m04*/gp34 sequence variation occurred in the extracellular region, with the predicted TM region and short cytoplasmic tail being relatively conserved. The position and number of predicted *N*-linked glycosylation sites varied greatly, with different strains containing between one and five sites. Experimental evidence has shown Smith *m04*/gp34 to contain three *N*-linked glycans (Kleijnen *et al.*, 1997). The NAXWXXE/HWo and CXLXXCL/PW/Ro *m02* family motifs described previously (Rawlinson *et al.*, 1996) were conserved in all strains. *m04*/gp34 contained an antigenic peptide, ^{243}YGPSLYRRF^{251}, recognized by cytotoxic T cells (CTLs) (Holtappels *et al.*, 2000b). This was conserved in seven strains, including Smith, G1F, G4 and K181. Five strains, including W8211 and W8, contained the sequence **F**GPSLYRRF, and five strains, including MI6A, had the sequence **F**GPSL**C**RRF. Both of these sequences fit the H-2D^{d} motif XGPXXXXXL/I/F described previously (Holtappels *et al.*, 2000a, b) and are thus likely to be recognized by *m04*/gp34-specific CTLs.

m05 sequence comparison

The m05 sequences analysed formed two distinct groups (Fig. 4a). The first group, containing ten strains, showed a high level of sequence identity with Smith m05 (83.9.1–93.9 %). The second group contained two strains, W8 and W8211, having sequence identity with Smith m05 of 41.6 and 41.0 %, respectively. Interestingly, these were the same strains found to have both *m03* and *m03.5* and which fell into a distinct *m04*/gp34 sequence group. The m05 protein, like other members encoded by the *m02* gene family, is a type I TM glycoprotein with an N-terminal signal peptide, a long extracellular region, a TM region and a short cytoplasmic tail. This structure, as well as the NAXWXXE/HWo and CXLXXCL/PW/Ro motifs described previously (Rawlinson *et al.*, 1996), were conserved in all strains, including W8 and W8211. The m05 predicted protein consisted of 332 aa (including the signal peptide) except for W8 and W8211, where shorter proteins of 282 and 289 aa, respectively, were predicted. The W8 and W8211 m05 sequences contained two predicted *N*-linked glycosylation motifs. These sites were conserved in some of the other strains, which also contained three to five additional sites.

Sequence variation in intragenic regions: clues to genetic events

In addition to the marked variation among MCMV strains within the predicted coding regions, the intragenic regions showed a high level of variation in both length and

(a)

Strain	% Id	% Sim
Smith		
G4	93.6	95.5
K181	83.1	90.2
W2	80.1	87.6
G2	79.7	87.6
K4	76.4	85.3
K17B	75.6	85.4
G1F	61.0	73.9
MI4B	59.9	73.9
N5	59.1	74.6
W8211	57.6	73.4
W8	57.3	73.8
N1	40.9	56.0
G3B	41.3	57.5
MI6A	41.3	57.5
MI2A	39.4	56.4
K6	39.4	56.4

Signal peptide

NAXWXXE/HWo

TM

CXLXXCL/PW/Ro

TM

(b)

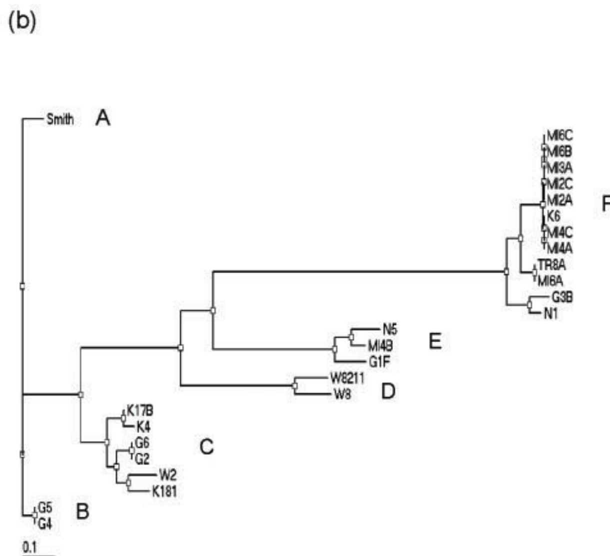

Fig. 3. (a) Alignment of predicted protein sequences for m04. Dots indicate conserved residues. Dashes indicate gaps introduced into alignments. The percentage identity and similarity with the Smith sequence are shown. Predicted *N*-linked glycosylation sites are underlined. Predicted signal peptides, TM regions and conserved motifs are boxed. For strains with identical sequences, only one is shown. (b) Phylogenetic tree for the m04 protein sequences.

nucleotide sequence (data not shown). Homology between the *m02–m03* and *m03–m03.5* regions in both W8 and W8211, along with homology within the genes themselves, suggests that the *m03* and *m03.5* ORFs have arisen from a gene duplication event and subsequently diverged. Similar gene duplication is thought to be a common mechanism for evolution in herpesviruses and other organisms (Arav-Boger *et al.*, 2005; Chee *et al.*, 1990; Lesniewski *et al.*, 2006; Rawlinson *et al.*, 1996; Sahagun-Ruiz *et al.*, 2004). These regions may then have allowed homologous recombination resulting in the loss of one gene in the majority of strains. A 21 nt triple repeat (CGCCTGAGCCTACCTCGAGAG) was found in the *m02–m03* region (nt 995–1015) of the G4 strain. This may be an indication of the capacity of MCMV to evolve through DNA duplication events. Direct repeats are present at the ends of the MCMV genome and are also clustered in the *m02* gene family within the *m08*, *m11* and *m12* ORFs (Rawlinson *et al.*, 1996). Short (1–3 nt) repeats are common in HCMV (Davis *et al.*, 1999) and longer direct repeats are present in other herpesviruses such as Epstein–Barr virus (Dambaugh & Kieff, 1982).

Assessment of the transcriptional kinetics of *m03*, *m03.5* and *m04*

After infection, MCMV (and HCMV) genes are expressed in a temporally controlled manner in the IE, early and late phases of replication. In order to analyse the kinetics of *m03* and *m03.5* expression, RT-PCR was performed on total RNA prepared from infected MEFs at various time points p.i. (Fig. 5). Both *m03* (Smith or W8211 strains) and *m03.5* (MI6A or W8211 strains) were found to be produced as early transcripts, but delayed when compared with *m04* (*m03* and *m03.5* were present at 8 but not 4 h p.i., as for *m04*). Interestingly, in W8211-infected cells, both *m03* and *m03.5* transcripts were produced and these had identical kinetics. This was not due to the amplification of non-specific PCR products, as the W8211 *m03* and *m03.5* product sizes were 840 and 897 bp, respectively. The *m04* gene is transcribed at early times (Holtappels *et al.*, 2000a). As this gene exhibited marked sequence variation among strains, we performed RT-PCR for *m04* from MI6A and W8211 wild isolates, as well as Smith MCMV. In all three viruses, *m04* transcription was evident at early times, with some transcripts present at late times. MI6A *m04* transcripts appeared with slower kinetics than Smith or W8211. Kinetic differences may be due to differences in the promoter regions. These have not been mapped, but intragenic regions likely to contain these sequences differ among these strains (data not shown). However, as these viruses grow differently *in vitro*, a closer analysis of the kinetics is needed using a quantitative approach to determine whether there is a real difference in the transcriptional control of *m04*.

The *m02* family glycoproteins m03, m03.5 and m04 all reach the cell surface

In order to analyse the cellular expression of these molecules, Smith *m03*, W8211 *m03*, MI6A *m03.5* and W8211 *m03.5* ORFs were cloned in frame with an N-terminal leader peptide and FLAG tag in the MigR1 expression vector. Both m03–FLAG and m03.5–FLAG proteins could be detected in fixed COS-7 cells by immunofluorescence after transient transfection (Fig. 6b) and surface expression could be detected by flow cytometry (Fig. 6a) The immunofluorescence staining pattern was identical for m03 and m03.5 from variant strains. The *m04*-encoded protein gp34 binds to MHC class I molecules in the ER and on the cell surface (Kavanagh *et al.*, 2001; Kleijnen *et al.*, 1997). *m04*/gp34 has been expressed previously and detected intracellularly by immunofluorescence using a haemagglutinin tag (Oliveira *et al.*, 2002). As *m04*/gp34 showed a high degree of amino acid sequence variation, we analysed its expression by flow cytometry using FLAG-tagged constructs. The m04–FLAG proteins from strains in each sequence group, A–F (Fig. 3b), could also be detected on the cell surface by anti-FLAG staining after transient transfection in COS-7 cells (Fig. 6a), indicating that *m04*/gp34 from each strain is expressed in a similar fashion, despite sequence variation. In addition, immunofluorescence staining patterns of fixed, transfected COS cells for FLAG-tagged *m04* constructs from all strains tested (Fig. 6b) were similar to those previously reported (Oliveira *et al.*, 2002).

DISCUSSION

In this study, sequence variation was analysed over the *m02–m05* gene region for seven MCMV strains, including two commonly used laboratory strains and five wild-derived isolates, as well as individual ORFs for additional wild-derived isolates. This region was found to exhibit significant genetic variation among MCMV laboratory strains and wild-derived isolates, with most of the variation occurring in the *m03–m04* gene region. The gene arrangement differed among strains, such that three different genotypes were present. We identified the *m03.5* ORF, present in a number of isolates instead of, or in addition to, *m03*. Conserved cysteine motifs, *N*-linked glycosylation sites and protein structure, as well as sequence homology to *m03*, identified *m03.5* as a member of the *m02* gene family. Homology to *m03* in both the gene and intragenic regions suggests that these ORFs have arisen via a gene duplication event and that one ORF has subsequently been lost in some isolates. The conservation of at least one of these ORFs, along with the loss of one in most strains, suggests that if the *m03.5*-encoded protein is expressed in the host after infection, it could share a (presently undefined) redundant function with m03. The W8 and W8211 strains may not have yet had significant evolutionary pressure to lose one ORF or, alternatively, if these proteins do have divergent functions, they may confer survival advantages upon these viruses in certain outbred mouse populations. Transcripts for both *m03* and *m03.5* were expressed early, but with delayed kinetics when compared with *m04*, and this did not vary significantly among the strains. Interestingly, in W8211, which has both *m03* and *m03.5*, both genes were transcribed

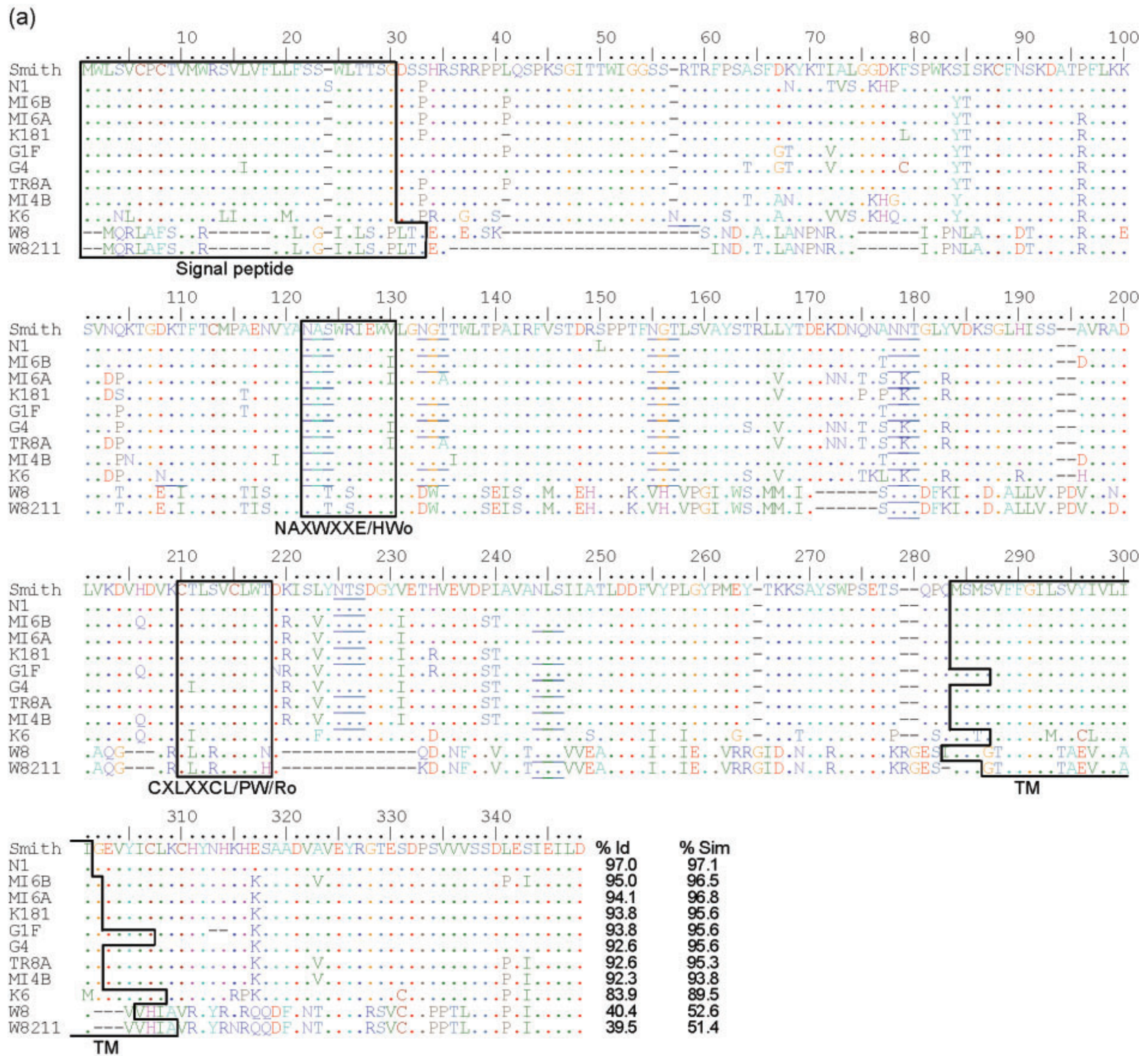

Fig. 4. (a) Alignment of predicted protein sequences for m05. Dots indicate conserved residues. Dashes indicate gaps introduced into alignments. The percentage identity and similarity with the Smith sequence are shown. Predicted *N*-linked glycosylation sites are underlined. Predicted signal peptides, TM regions and conserved motifs are boxed. (b) Phylogenetic tree for m05 protein sequences.

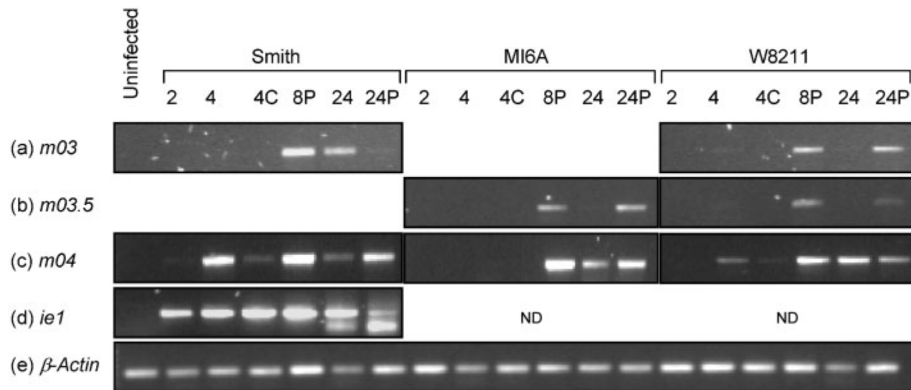

Fig. 5. RT-PCR analysis of MEFs infected with the Smith, MI6A or W8211 strain of MCMV for the *m03* (a), *m03.5* (b), *m04* (c) and *ie1* (d, infection with Smith strain only) genes. RT-PCR of *β-actin* (e) was used as a loading control. No product was present for any of the RT^{-} or H_2O controls (not shown). Numbers indicate h p.i. C, Cycloheximide; P, PAA; ND, not done.

with identical kinetics. This suggests that the two encoded proteins both have the potential to be expressed and supports the argument that *m03* and *m03.5* are separate ORFs and not simply highly variable forms of the same ORF. At this stage however, the potential function of these proteins remains undefined.

ORFs within the *m02* gene family of MCMV lie in a head-to-tail fashion and do not overlap. Multiple alternative initiation ATGs were found to be present for the *m02* (Smith, K181), *m03.5* (G1F, MI6A, W8, W8211), *m04* (G1F, MI6A, W8211) and *m05* (all strains) ORFs. We used those previously annotated for Smith (Rawlinson *et al.*, 1996) for our alignments, with the exception of *m02* where we used the second ATG (in Smith), due to its conservation among the isolates. This second ATG is not within a Kozak consensus sequence. Our analysis of other genes including *m03*, *m04* and *m05* revealed that Kozak sequences are not common in the MCMV genome, at least within this subset of the *m02* gene family. Experimental confirmation is still required for the majority of MCMV ORFs, and annotation of the MCMV genome will continue to be revised as sequence data become available. The shorter *m05* ORFs predicted for the W8 and W8211 strains appear to be true *m05* ORFs based on homology and gene arrangement. We sequenced the *m06* gene for a selection of strains (data not shown) and found this to be highly conserved, consistent with a previous study (Smith *et al.*, 2006). The percentage amino acid identity with Smith m06 ranged from 96.2 (G1F) to 98.2 % (K181). W8211 m06 shared 97.6 % amino acid sequence identity with Smith m06, whilst W8211 m05 shared only 17.6 % sequence identity with Smith m06, compared with 39.5 % identity with Smith m05.

A number of *m02* family members have been found previously to be expressed intracellularly in an over-expression system (Oliveira *et al.*, 2002). The expression of m03 has not been reported previously. FLAG-tagged m03 and m03.5 proteins from all strains tested could be detected on the cell surface of transfected cells with a FLAG-specific mAb, suggesting that, if the native proteins are similarly expressed in infected cells, then these proteins may bind to ligands *in trans* on other cells in a similar fashion to the *m145* family member m157 (Arase *et al.*, 2002; Smith *et al.*, 2002) or to soluble ligands. Alternatively, they may bind host molecules *in cis*, as is the case for *m04*/gp34 (Kavanagh *et al.*, 2001; Kleijnen *et al.*, 1997). If m03 and m03.5 proteins are expressed and interact with host molecules, then the sequence variation observed is likely to be reflected in the binding of these host proteins. In addition to intracellular expression, *m04*/gp34 could be detected on the cell surface by flow cytometry and this did not vary among strains, despite a high degree of sequence variation. This is also the case for m157, which exhibits a similar degree of variation (C. A. Forbes, A. J. Corbett & A. A. Scalzo, unpublished data; Voigt *et al.*, 2003). The *m04*/gp34 protein is expressed in the ER of mouse cells and traffics to the surface only in complex with MHC class I molecules (Kavanagh *et al.*, 2001; Kleijnen *et al.*, 1997). The extensive amino acid sequence variation observed would be consistent with *m04*/gp34 from different MCMV strains being capable of binding different MHC class I molecules and may indicate adaptation to the diverse array of MHC class I molecules in a heterogeneous host population. Alternatively, these variant *m04*/gp34 proteins may bind to host proteins other than MHC class I and therefore may be under different selective pressures. The *m04*/gp34 protein sequences analysed fell into six distinct groups (Fig. 3b), suggesting possible selection by different host *H2* haplotypes. Similarly, the significant variability in the *m03* and *m03.5* ORFs and the probable gene duplication event leading to their evolution may reflect an adaptation of the virus to genetically diverse wild mouse host populations and may give a clue to the function of these genes as potentially interacting with variant host molecules.

Our data are consistent with previous reports that the regions at the extremities of CMV genomes, which in MCMV include the *m02* and *m145* gene families, are

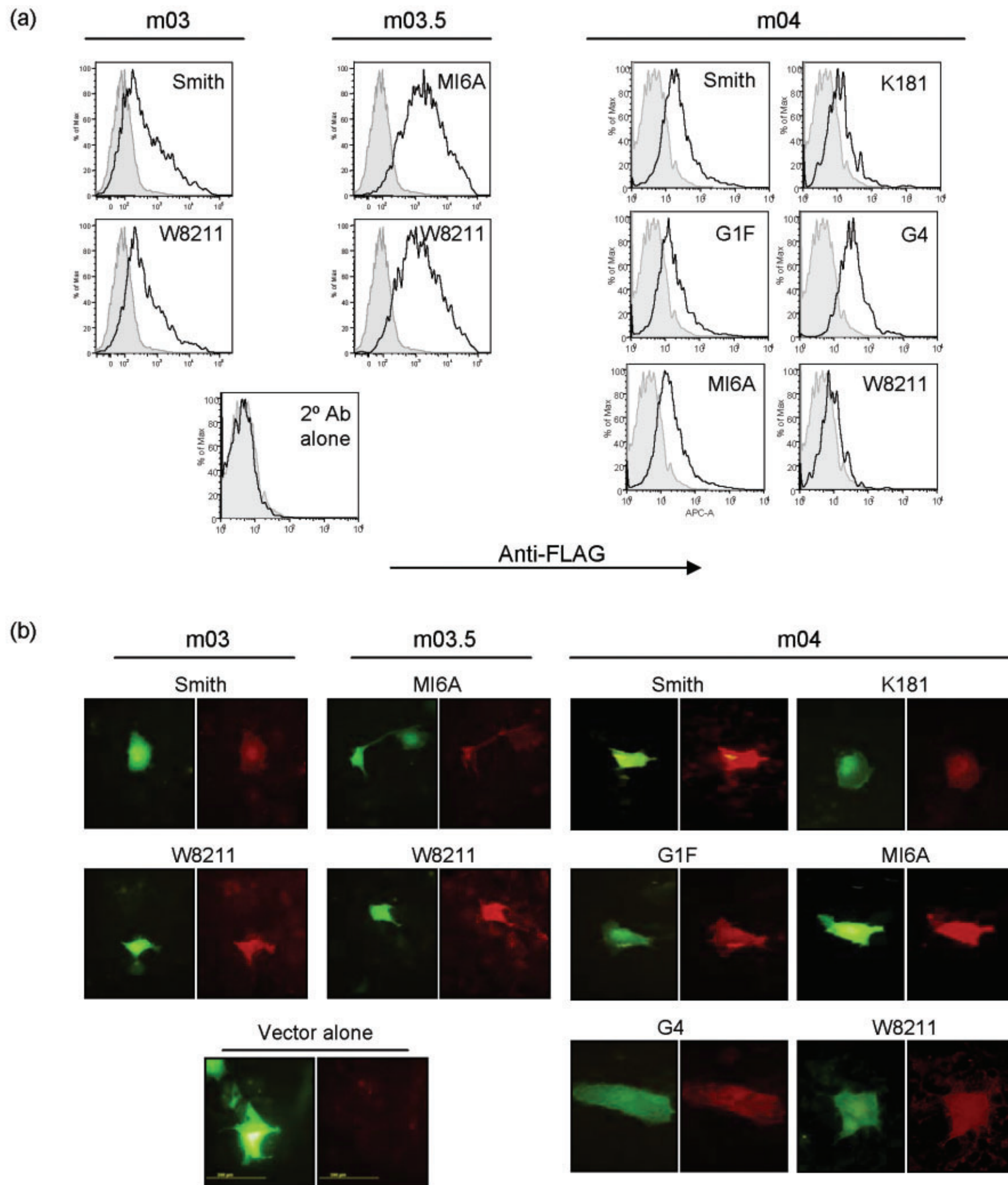

Fig. 6. Expression of m03, m03.5 and m04 proteins. COS-7 cells were transiently transfected with the FLAG-tagged *m03*, *m03.5* or *m04* construct and analysed for expression of FLAG by flow cytometry (a) or immunofluorescence (b) after 48 h. GFP was expressed independently as a marker of transfection and plots were gated on GFP^{+} live cells. Solid lines, transfected cells; filled histograms, empty vector controls. In (b), GFP fluorescence is green and anti-FLAG staining is red.

highly divergent. Phylogenetic analyses produced different groupings of strains for each of the individual *m02* gene family members (Figs 1–4), suggesting that there is limited linkage between these genes. Within genes, regions responsible for maintaining protein structure and function (e.g. the TM domain) were conserved and the extracellular domains were most divergent. ORFs with immune-modulatory function may have developed polymorphisms as a means of evading the host immune system in genetically diverse populations. Polymorphisms in antigenic determinants recognized by CTLs may be useful to the virus in terms of evading this response. However, in the case of *m04*/gp34, the antigenic determinant is conserved. In contrast, the *ie1* epitope was not conserved

in wild MCMV isolates (Lyons *et al.*, 1996). As the *m04*/gp34 peptide is derived from the inner part of the TM region and cytoplasmic domain, it may be essential for function and not readily altered by the virus. A recent report shows that the TM domain of *m04*/gp34 is essential for association with K^b heavy chains (Lu *et al.*, 2006).

MCMV immune evasion genes are likely to be host-strain specific (e.g. *m04*/gp34, which interacts with host MHC class I molecules), and genes from different isolates may have evolved to function in individual host strains. The m157 protein was originally reported to be responsible for a dominant resistant phenotype in C57BL/6 mice by binding to the Ly49H NK cell-activation receptor (Arase *et al.*, 2002; Brown *et al.*, 2001; Scalzo *et al.*, 1990, 1992; Smith *et al.*, 2002), but the binding of m157 to the Ly49I inhibitory receptor in 129/J mice may have a different effect (Bubic *et al.*, 2004). The localized genetic variation in *m03–m05* among relatively conserved *m02* and *m06* genes suggests that amino acid changes have been selected by functional constraints such as interactions with variant host proteins. Recent reports have shown that MCMV can mutate rapidly *in vivo* under functional selective pressure from m157–Ly49H interactions (French *et al.*, 2004; Voigt *et al.*, 2003). However, undertaking studies to test this hypothesis for these *m02* family genes would depend on first characterizing the putative host receptors. Most MCMV studies are conducted using MCMV laboratory strains and inbred mice, which may not reflect the high degree of natural variability in viruses or hosts. The analysis of variation in the immune-evasion genes highlights the need for caution in assigning functions in host–virus combinations that do not fully reflect the spectrum of viral or host variability in nature. Genetic variation in CMVs also has implications for the development of antiviral therapies or vaccines that will be effective against the complete range of HCMV strains, and for PCR-based diagnosis of HCMV infection.

ACKNOWLEDGEMENTS

This work was supported by grants from the NH & MRC and the Australian Antarctic Division. A. J. C. was the recipient of a WA & MG Saw Medical Research Fellowship from the University of Western Australia. A. A. S. was supported by an NH & MRC Senior Research Fellowship. The authors would like to thank Nicole Harvey (University of Western Australia) for technical assistance and Ed Mocarski (Stanford University) for critical reading of the manuscript.

REFERENCES

Alcami, A. & Koszinowski, U. H. (2000). Viral mechanisms of immune evasion. *Immunol Today* **21**, 447–455.

Arase, H., Mocarski, E. S., Campbell, A. E., Hill, A. B. & Lanier, L. L. (2002). Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. *Science* **296**, 1323–1326.

Arav-Boger, R., Zong, J. C. & Foster, C. B. (2005). Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes* **31**, 65–72.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.

Boname, J. M. & Chantler, J. K. (1992). Characterization of a strain of murine cytomegalovirus which fails to grow in the salivary glands of mice. *J Gen Virol* **73**, 2021–2029.

Booth, T. W., Scalzo, A. A., Carrello, C., Lyons, P. A., Farrell, H. E., Singleton, G. R. & Shellam, G. R. (1993). Molecular and biological characterization of new strains of murine cytomegalovirus isolated from wild mice. *Arch Virol* **132**, 209–220.

Brocchieri, L., Kledal, T. N., Karlin, S. & Mocarski, E. S. (2005). Predicting coding potential from genome sequence: application to betaherpesviruses infecting rats and mice. *J Virol* **79**, 7570–7596.

Brown, M. G., Dokun, A. O., Heusel, J. W., Smith, H. R., Beckman, D. L., Blattenberger, E. A., Dubbelde, C. E., Stone, L. R., Scalzo, A. A. & Yokoyama, W. M. (2001). Vital involvement of a natural killer cell activation receptor in resistance to viral infection. *Science* **292**, 934–937.

Bubic, I., Wagner, M., Krmpotic, A., Saulig, T., Kim, S., Yokoyama, W. M., Jonjic, S. & Koszinowski, U. H. (2004). Gain of virulence caused by loss of a gene in murine cytomegalovirus. *J Virol* **78**, 7536–7544.

Chalmer, J. E., Mackenzie, J. S. & Stanley, N. F. (1977). Resistance to murine cytomegalovirus linked to the major histocompatibility complex of the mouse. *J Gen Virol* **37**, 107–114.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., III, Kouzarides, T. & other authors (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125–169.

Dambaugh, T. R. & Kieff, E. (1982). Identification and nucleotide sequences of two similar tandem direct repeats in Epstein–Barr virus DNA. *J Virol* **44**, 823–833.

Davis, C. L., Field, D., Metzgar, D., Saiz, R., Morin, P. A., Smith, I. L., Spector, S. A. & Wills, C. (1999). Numerous length polymorphisms at short tandem repeats in human cytomegalovirus. *J Virol* **73**, 6265–6270.

Ebeling, A., Keil, G. M., Knust, E. & Koszinowski, U. H. (1983). Molecular cloning and physical mapping of murine cytomegalovirus DNA. *J Virol* **47**, 421–433.

Farrell, H. E. & Shellam, G. R. (1989). Immunoblot analysis of the antibody response to murine cytomegalovirus in genetically resistant and susceptible mice. *J Gen Virol* **70**, 2573–2586.

French, A. R., Pingel, J. T., Wagner, M., Bubic, I., Yang, L., Kim, S., Koszinowski, U., Jonjic, S. & Yokoyama, W. M. (2004). Escape of mutant double-stranded DNA virus from innate immune control. *Immunity* **20**, 747–756.

Frohman, M. A., Dush, M. K. & Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* **85**, 8998–9002.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Hengel, H., Brune, W. & Koszinowski, U. H. (1998). Immune evasion by cytomegalovirus – survival strategies of a highly adapted opportunist. *Trends Microbiol* **6**, 190–197.

Hengel, H., Reusch, U., Gutermann, A., Ziegler, H., Jonjic, S., Lucin, P. & Koszinowski, U. H. (1999). Cytomegaloviral control of MHC class I function in the mouse. *Immunol Rev* **168**, 167–176.

Ho, M. (1991). *Cytomegalovirus Biology and Infection*, 2nd edn. New York: Plenum Medical Book Company.

Holtappels, R., Thomas, D., Podlech, J., Geginat, G., Steffens, H.-P. & Reddehase, M. J. (2000a). The putative natural killer decoy early gene *m04* (gp34) of murine cytomegalovirus encodes an antigenic peptide recognized by protective antiviral CD8 T cells. *J Virol* **74**, 1871–1884.

Holtappels, R., Thomas, D. & Reddehase, M. J. (2000b). Identification of a K^d-restricted antigenic peptide encoded by murine cytomegalovirus early gene *M84*. *J Gen Virol* **81**, 3037–3042.

Hudson, J. B., Walker, D. G. & Altamirano, M. (1988). Analysis in vitro of two biologically distinct strains of murine cytomegalovirus. *Arch Virol* **102**, 289–295.

Kavanagh, D. G., Koszinowski, U. H. & Hill, A. B. (2001). The murine cytomegalovirus immune evasion protein m4/gp34 forms biochemically distinct complexes with class I MHC at the cell surface and in a pre-Golgi compartment. *J Immunol* **167**, 3894–3902.

Kleijnen, M. F., Huppa, J. B., Lucin, P., Mukherjee, S., Farrell, H., Campbell, A. E., Koszinowski, U. H., Hill, A. B. & Ploegh, H. L. (1997). A mouse cytomegalovirus glycoprotein, gp34, forms a complex with folded class I MHC molecules in the ER which is not retained but is transported to the cell surface. *EMBO J* **16**, 685–694.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

Lesniewski, M., Das, S., Skomorovska-Prokvolit, Y., Wang, F.-Z. & Pellett, P. E. (2006). Primate cytomegalovirus US12 gene family: a distinct and diverse clade of seven-transmembrane proteins. *Virology* **354**, 286–298.

Lu, X., Kavanagh, D. G. & Hill, A. B. (2006). Cellular and molecular requirements for association of the murine cytomegalovirus protein m4/gp34 with major histocompatibility complex class I molecules. *J Virol* **80**, 6048–6055.

Lyons, P. A., Allan, J. E., Carrello, C., Shellam, G. R. & Scalzo, A. A. (1996). Effect of natural sequence variation at the H-2Ld-restricted CD8+ T cell epitope of the murine cytomegalovirus ie1-encoded pp89 on T cell recognition. *J Gen Virol* **77**, 2615–2623.

Misra, V. & Hudson, J. B. (1980). Minor base sequence differences between the genomes of two strains of murine cytomegalovirus differing in virulence. *Arch Virol* **64**, 1–8.

Nielsen, H. & Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* **6**, 122–130.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.

Oliveira, S. A., Park, S.-H., Lee, P., Bendelac, A. & Shenk, T. E. (2002). Murine cytomegalovirus m02 gene family protects against natural killer cell-mediated immune surveillance. *J Virol* **76**, 885–894.

Pear, W. S., Miller, J. P., Xu, L., Pui, J. C., Soffer, B., Quackenbush, R. C., Pendergast, A. M., Bronson, R., Aster, J. C. & other authors (1998). Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood* **92**, 3780–3792.

Pignatelli, S., Dal Monte, P., Rossini, G. & Landini, M. P. (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev Med Virol* **14**, 383–410.

Rawlinson, W. D., Farrell, H. E. & Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *J Virol* **70**, 8833–8849.

Reusch, U., Muranyi, W., Lucin, P., Burgert, H. G., Hengel, H. & Koszinowski, U. H. (1999). A cytomegalovirus glycoprotein re-routes MHC class I complexes to lysosomes for degradation. *EMBO J* **18**, 1081–1091.

Sahagun-Ruiz, A., Sierra-Honigmann, A. M., Krause, P. & Murphy, P. M. (2004). Simian cytomegalovirus encodes five rapidly evolving chemokine receptor homologues. *Virus Genes* **28**, 71–83.

Scalzo, A. A., Fitzgerald, N. A., Simmons, A., La Vista, A. B. & Shellam, G. R. (1990). Cmv-1, a genetic locus that controls murine cytomegalovirus replication in the spleen. *J Exp Med* **171**, 1469–1483.

Scalzo, A. A., Fitzgerald, N. A., Wallace, C. R., Gibbons, A. E., Smart, Y. C., Burton, R. C. & Shellam, G. R. (1992). The effect of the Cmv-1 resistance gene, which is linked to the natural killer cell gene complex, is mediated by natural killer cells. *J Immunol* **149**, 581–589.

Smith, H. R., Heusel, J. W., Mehta, I. K., Kim, S., Dorner, B. G., Naidenko, O. V., Iizuka, K., Furukawa, H., Beckman, D. L. & other authors (2002). Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc Natl Acad Sci U S A* **99**, 8826–8831.

Smith, L. M., Shellam, G. R. & Redwood, A. J. (2006). Genes of murine cytomegalovirus exist as a number of distinct genotypes. *Virology* **352**, 450–465.

Tusnady, G. E. & Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* **283**, 489–506.

Tusnady, G. E. & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850.

Voigt, V., Forbes, C. A., Tonkin, J. N., Degli-Esposti, M. A., Smith, H. R., Yokoyama, W. M. & Scalzo, A. A. (2003). Murine cytomegalovirus m157 mutation and variation leads to immune evasion of natural killer cells. *Proc Natl Acad Sci U S A* **100**, 13483–13488.

Xu, J., Lyons, P. A., Carter, M. D., Booth, T. W., Davis-Poynter, N. J., Shellam, G. R. & Scalzo, A. A. (1996). Assessment of antigenicity and genetic variation of glycoprotein B of murine cytomegalovirus. *J Gen Virol* **77**, 49–59.