JOURNAL OF VIROLOGY, Dec. 1996, p. 8833–8849
0022-538X/96/$04.00+0
Copyright © 1996, American Society for Microbiology
Vol. 70, No. 10

Analysis of the Complete DNA Sequence of Murine Cytomegalovirus

WILLIAM D. RAWLINSON,[1]* HELEN E. FARRELL,[2]† AND BARCLAY G. BARRELL[3]

Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH,[1] Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge,[2] and The Sanger Centre, Hinxton Hall, Hinxton, Cambridge CB10 1RQ,[3] United Kingdom

Received 26 March 1996/Accepted 11 September 1996

The complete DNA sequence of the Smith strain of murine cytomegalovirus (MCMV) was determined from virion DNA by using a whole-genome shotgun approach. The genome has an overall G+C content of 58.7%, consists of 230,278 bp, and is arranged as a single unique sequence with short (31-bp) terminal direct repeats and several short internal repeats. Significant similarity to the genome of the sequenced human cytomegalovirus (HCMV) strain AD169 is evident, particularly for 78 open reading frames encoded by the central part of the genome. There is a very similar distribution of G+C content across the two genomes. Sequences toward the ends of the MCMV genome encode tandem arrays of homologous glycoproteins (gps) arranged as two gene families. The left end encodes 15 gps that represent one family, and the right end encodes a different family of 11 gps. A homolog (m144) of cellular major histocompatibility complex (MHC) class I genes is located at the end of the genome opposite the HCMV MHC class I homolog (UL18). G protein-coupled receptor (GCR) homologs (M33 and M78) occur in positions congruent with two (UL33 and UL78) of the four putative HCMV GCR homologs. Counterparts of all of the known enzyme homologs in HCMV are present in the MCMV genome, including the phosphotransferase gene (M97), whose product phosphorylates ganciclovir in HCMV-infected cells, and the assembly protein (M80).

A more complete understanding of the protein coding in many herpesviruses has been assisted by the availability of the complete DNA sequences of their genomes (2, 6, 22, 37, 38, 51, 94, 107a, 140, 141). Although previous analyses of the complete nucleotide sequences of several herpesviruses have largely confirmed their biological classification, there have been notable exceptions, including human herpesvirus 6 (HHV-6) (51, 75), channel catfish virus (CCV) (37), herpesvirus of turkeys, and Marek's disease virus (16). Analysis of short regions of the murine cytomegalovirus (MCMV) genome has confirmed its biological classification as a betaherpesvirus. A number of MCMV genes with counterparts in human cytomegalovirus (HCMV) have been described, although the precise location and order of the genes have not been determined (17, 31, 44, 70, 87, 91, 97, 98, 119, 154).

MCMV is a betaherpesvirus that causes acute, latent, and persistent infection of the natural host (57, 115, 121). It has been distinguished from other murine viruses on the basis of cytopathic effect, tissue tropism, and morphology (111). Both MCMV and HCMV have a highly restricted host range. As a consequence of the difficulties associated with studies of HCMV pathogenesis in humans, infection of mice with MCMV has been used as a model for HCMV infection. The major reason for using MCMV as a model is the matching biological characteristics of these virus infections in their natural settings. Both MCMV and HCMV cause severe infections in the immunocompromised or immunologically immature host, resulting in similar clinical syndromes (28, 104, 111, 128). In addition, both HCMV and MCMV are susceptible to the antiviral agent 9-(1,3-dihydroxy-2-propoxymethyl)guanine (DHPG, ganciclovir, or GCV) (127). However, MCMV and HCMV have some biological differences, as transplacental transmission of MCMV has not been demonstrated and mouse models of fetal infection involve direct inoculation of MCMV into the central nervous system or uterus (66, 144).

The genome of MCMV is known to comprise a single unique sequence (type F) with short direct repeats (DRs) at either end (43). The genome has a relatively high G+C content (59%), like HCMV, herpes simplex virus (HSV), and Epstein-Barr virus (EBV) but unlike varicella-zoster virus (VZV), human herpesvirus 6 (HHV-6), and herpesvirus saimiri (HVS). An analysis of the complete 230-kb DNA sequence of the Smith strain of MCMV is presented here. The genome is predicted to contain 170 genes and was found to be essentially colinear with the HCMV genome over the central 180 kb. Counterparts of the G protein-coupled receptors (GCR), UL25, UL82, and US22 gene families of HCMV were identified. There are 24- and 27-kb regions at the left and right ends of the genome, respectively, that encode genes currently only found in MCMV. Within these regions, families of tandemly arranged genes are predicted to encode membrane glycoproteins, a feature that resembles the RL11 and US6 families identified in HCMV. There are three open reading frames (ORFs) with homology to eukaryotic cellular genes. The M33 and M78 ORFs have sequence similarity to GCRs, and the predicted protein product of m144 is similar to major histocompatibility complex (MHC) class I proteins. Unlike the MCMV GCR homologs which resemble the HCMV counterparts, the m144 ORF does not share amino acid sequence similarity or colinearity with the HCMV MHC class I homolog (UL18). A discussion of the proteins predicted to be encoded by the ORFs defined in this analysis is presented.

* Corresponding author. Mailing address: Department of Virology, ICPMR, Westmead NSW 2145, Australia. Phone: 61-2-98456225. Fax: 61-2-96335314. Electronic mail address: WDR@westmed.wh.su.edu.au

† Present address: Department of Microbiology, University of Western Australia, Nedlands, Western Australia, 6009 Australia.

MATERIALS AND METHODS

This study used a whole-virus-genome shotgun approach to determine the DNA sequence (36), that is, sequencing templates were made directly from

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

purified virion DNA, without the use of restriction fragment libraries. Regions of ambiguity were then resolved by directed techniques (32).

Virus culture and template DNA preparation. The Smith strain of MCMV (ATCC VR-194) was grown in mouse embryonic fibroblasts until the cultures exhibited 90% cytopathic effect. Virions were purified from the culture supernatant by two rounds of sucrose density gradient centrifugation. The virus was pelleted and DNA was extracted with phenol-chloroform (8). The random cloning and sequencing of whole virion DNA were performed by methods previously described for sequencing of restriction fragments (7). Random fragments of virion DNA of approximately 2 kb were generated by sonication, end repaired with Klenow fragment of DNA polymerase to produce blunt ends, and used to construct an M13 library (8). Single-stranded sequencing templates were prepared using *Escherichia coli* TG1, by either previously described techniques (8) or a modified method in microtiter trays (130) employing sodium iodide as the chaotropic agent instead of sodium dodecyl sulfate for cell lysis (3). Double-stranded DNA templates were generated by PCR from the M13 single-stranded template with M13 forward and reverse PCR primers. To obtain sequence from areas underrepresented in the M13 library, double-stranded DNA templates were prepared by PCR of virion DNA. In several instances, custom oligonucleotide primers were used to generate sequence from single-stranded M13 templates containing inserts larger than 2 kb. Double-stranded templates were purified by precipitation with polyethylene glycol (32).

DNA sequencing. All sequences were obtained by modified dideoxy chain termination techniques (125). The majority of sequences (85%) were determined with an on-line fluorescent sequencing machine from Applied Biosystems (the 373A DNA sequencer), with version 2.0 of the analysis software. Dye-labelled primers and Taq cycle-sequencing reactions were used to generate fluorescent sequence data from templates (32). Sequence information from templates was also obtained by standard radioactive sequencing techniques (8), and the data from the autoradiographs were read with an Amersham film scanner and Base-Scanner software versions 2.0 to 2.3. Data were generated from regions containing compressions or single-stranded sequences by using double-stranded templates, custom oligonucleotide primers, and Taq cycle sequencing with fluorescence-labelled dideoxynucleotides (59). The final sequence was well determined (39) on both strands over 99.19% of the genome, and each base was represented in gel readings an average of 10.6 times.

Sequence assembly and screening of gel readings. Sequence data were compiled by using the program BAP (39, 134), and the sequences were edited on a SUN Microsystems computer. The output from all gel readings was standardized to allow trace files of sequence data derived from autoradiographs and those from the fluorescence-based sequencer to be assembled and handled in an identical manner (40, 134). All gel readings were screened against files containing the sequences of M13mp18, a 5-kb LINE1 murine repeat sequence, and the 300-bp murine alu equivalent (SINE b1) sequence (60, 147). This initial screen allowed removal of M13 and some murine sequences, as the murine repeats may constitute approximately 10% of the mouse genome.

Analysis of homologies, protein coding content, and motif searches. Homology searches with the local alignment algorithm BLAST (64) and the global alignment algorithm FastA (81) were performed against a nonredundant combined database containing the SwissProt28 (March 1994) and PIR 39 (1994) protein libraries. Examination of the DNA sequence for transcription signals was performed by using the search features of the computer programs acedb (42b) and DIANA (56a). Nucleotide motif searches were performed against a subset of motifs derived from the EMBL 38 nucleotide library (March 1994) (48). The sequence was analyzed for tandem and inverted repeats by using the programs xnip (134), repeatsearch (112a), and quicktandem, tandem, and inverted (42a). The strategies used to identify genes of MCMV likely to be coding were based on those used in the sequence analysis of other herpesviruses. The major criteria for identifying a coding sequence were the presence of an ORF with a minimum length of 300 bp and less than 60% overlap with adjacent ORFs. In the ORFs m29, m29.1; m48.1, m48.2; m107, m108; m117, m117.1; m119.4, m119.5; m134, m135; and m147, m148 (Table 1), the overlap was greater than 60%, as the ORFs were of similar length, and none of the other ORF identification criteria distinguished between them satisfactorily. In addition, analysis of third-position bias, the presence of consensus promoter and translation initiation sequences, and similarity to known genes were used to support initial ORF selections. Gaps between ORFs were inspected for smaller potential reading frames and were included in the final coding map if they satisfied the other criteria used for ORF selection. Multiple alignments were done by using the program clustal V (54), with gap penalties of 10 (fixed) and 10 (floating), and the PAM250 protein weight matrix.

Nomenclature. The naming system used for genes of MCMV numbers them from the left end of the prototype genome (69). As the MCMV and HCMV genomes were found to be essentially colinear over the central parts of both genomes, the numbering system for these MCMV genes is congruent with the HCMV numbering system for the long unique (UL) region. MCMV genes with homologs in HCMV are indicated by uppercase prefixes (M25), whereas ORFs without sequence similarity with HCMV genes are indicated by lowercase prefixes (m02). In order to maintain the correlation between the two numbering systems, suffixes (as in m25.1 and m25.2, etc.) were introduced when additional unique MCMV genes were found between homologs of HCMV genes. These suffixes do not indicate any similarity between these genes. When MCMV genes were found to be homologous with gene families of HCMV (and also with other members of the gene families of MCMV and HCMV), this is indicated in Table 1. No attempt was made to correlate MCMV homologs with the numbering system of the HCMV short unique (US) and repeat (TRL, IRL, IRS, and TRS) regions.

Nucleotide sequence accession number. The complete DNA sequence has been submitted to the GenBank database, accession number U68299.

RESULTS

Features of the nucleotide sequence of MCMV. The genome of MCMV Smith is 230,278 bp in length, which is around 5 kb less than the size estimated from restriction fragment analysis of the Smith strain (43). The overall G+C content is 58.7%, with markedly reduced G+C content between 155 to 163 kb, within *Hin*dIII fragment F (Fig. 1). The regions between 1 and 17 kb (*Hin*dIII fragment A) and 204 and 221 kb (*Hin*dIII fragments O, P, and E) also show a less prominent reduction in G+C content. These regions of reduced G+C content are congruent with the positions of members of the MCMV membrane glycoprotein families m02 and m145, respectively (Fig. 1), and are colinear with regions of reduced G+C content in the HCMV genome. There is no evidence of generalized CpG suppression. The arrangement of the CpG dinucleotides across the genome shows localized CpG suppression corresponding to the major immediate-early (MIE) region (Fig. 1), as was found for HCMV (56).

The viral genome is a single long unique sequence, with short (31-bp) DRs at either end (Fig. 2). These two DRs are not represented elsewhere in the genome. The termini differ from those published for strain K181 (89). The Smith termini have one less C residue at position 34 (outside the left DR) and one additional residue at position 1 (within the left DR) and position 230,268 (within the right terminal DR). Other DRs are clustered within the m02 membrane glycoprotein family at the left end of the genome (within the regions covered by genes m08, m11, and m12), within M83, within the A+T-rich region between 154.5 and 162 kb, within M117, at the start of M122(ie3), and in the m123(ie1) promoter (Fig. 2). The MIE region has five 94-bp DRs between base 183312 and base 183778 (42). There are also short inverted repeat sequences located throughout the genome (Fig. 2). They are present upstream of m59, in the A+T-rich origin of replication region between 154.5 and 162 kb (92), in M112, and in the m123(ie1) promoter region (42).

Protein coding content. The arrangement of the MCMV genes homologous with those of other herpesviruses is shown in Fig. 3 as unfilled blocks and explained in the legend. All of the conserved herpesvirus gene blocks are retained in MCMV, and the position and orientation of these blocks are identical with those in HCMV (22) and HHV-6 (51). It is predicted that the genome encodes 170 ORFs, of which 78 have significant amino acid identity with genes of HCMV (Table 2). These homologs are located within the central part of the genome, interspersed with a small number of ORFs predicted to encode genes unique to MCMV. Homologs of the HCMV gene families UL25, UL82, and US22 and the GCRs are present, but MCMV contains no amino acid sequence homologs of the RL11, US1, US2, US6, or US12 gene families of HCMV. Not present are MCMV homologs of the extra 19 ORFs described for HCMV strain Toledo or the extra three ORFs found in HCMV strain Towne, additional to those described for HCMV strain AD169 (20). A number of HCMV genes with as yet poorly understood functions are moderately to strongly conserved and correspond to the MCMV genes M26, M27, M31, M34, M47, M49, M50, M51, M71, M76, M87, M88, M91, M92, M93, M94, M95, and M103.

The genome and individual ORFs. (i) Spliced genes. MCMV genes that are known to be spliced in the Smith strain are

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 1. Summary of the map locations and features of the 170 predicted ORFs of the MCMV (strain Smith) genome[a]

ORF	Strand	Position[b]		Length (aa)	MW (kDa)	HCMV name[c]	Comments (reference[s])[d]
		From:	To:				
m01	C	483	836	118	13.3		
m02		999	1976	326	36.6		Glycoprotein family m02
m03		2236	3099	288	32.4		Glycoprotein family m02
m04		3270	4067	266	30.2		Glycoprotein family m02
m05		4185	5207	341	38.0		Glycoprotein family m02
m06		5300	6334	345	38.7		Glycoprotein family m02
m07		6463	7404	314	34.2		Glycoprotein family m02
m08		7459	8526	356	39.2		Glycoprotein family m02
m09		8632	9510	293	32.1		Glycoprotein family m02
m10		9624	10496	291	32.0		Glycoprotein family m02
m11		10715	11611	299	33.2		Glycoprotein family m02
m12		11686	12501	272	30.4		Glycoprotein family m02
m13		12599	12997	133	14.4		Glycoprotein family m02
m14		13085	13987	301	33.8		Glycoprotein family m02
m15		14085	15062	326	35.6		Glycoprotein family m02
m16		15044	15673	210	23.1		Glycoprotein family m02
m17	C	15752	16951	400	45.9		Member of MGP family m145 (right end)
m18	C	17074	20193	1040	108.5		
m19		20338	20778	147	17.3		
m20	C	20805	23045	747	82.5		
m21·		22645	23331	229	25.3		
m22		23586	23897	104	11.1		
M23	C	23781	24953	391	43.1	UL23 (GF2) P	US22 family homolog
m23.1		24826	25158	111	12.1		
M24	C	25151	26119	323	35.8	UL24 (GF2)	US22 family homolog
M25		26015	28810	932	103.2	UL25 (GF1)	UL25 family homolog (34)
m25.1**	C	29001	30281	427	47.5	UL23 (GF2)	US22 family homolog
	C	29001	30602	534	60.4	UL23 (GF2)	
	C	30248	31216	323	35.4	US22 (GF2)	
m25.2**	C	30248	31657	470	51.8	US22 (GF2)	US22 family homolog
M26	C	31350	31925	192	21.7	UL26	
M27	C	32250	34295	682	78.7	UL27	
M28	C	34489	35778	430	47.3	UL28	
m29·		35747	36727	327	35.7		
m29.1·	C	36112	36660	183	19.8		
m30		36884	37726	281	30.8		
M31		37279	38826	516	56.7	UL31	
M32	C	39283	41436	718	78.6	UL32 (pp150)	Tegument protein of HCMV (61, 62)
M33		41486	42777	377	42.2	UL33 (GCR)	Spliced GCR (21, 35)
M34		43083	45644	854	94.5	UL34	
M35		45909	47465	519	58.1	UL35 (GF1)	UL25 family homolog
M36	C	47621	49267	507	56.9	UL36 (GF2)	M36 E1+E2, US22 family homolog (72)
M36 Ex2	C	47621	48909	429.7	48.2	UL36 Exon 2	US22 family homolog (72)
M36 Ex1**	C	49036	49267	77.3	8.7	UL36 Exon 1	US22 family homolog (72)
M37	C	49444	50478	345	38.0	UL37	Glycoprotein present at IE times (72)
M38	C	50465	51955	497	54.8	UL38	(72)
m39	C	52487	53200	238	25.6		
m40	C	53268	53630	121	13.3		
m41	C	53786	54199	138	14.6		
m42	C	54355	54843	163	17.7		Possible glycoprotein
M43	C	55354	57144	597	67.0	UL43 P	US22 family homolog
M44	C	57888	59120	411	44.6	UL44 (DPAP)	DNA binding protein (46)
M45[e]	C	59518	62145	876	96.9	UL45 (RRL)	Homolog of the large subunit of ribonucleotide reductase (117)
m45.1	C	61767	63038	424	42.0		
M46	C	63043	63924	294	33.2	UL46	Minor capsid protein (114)
M47		63923	67042	1,040	118.1	UL47	
M48		67042	73488	2,149	238.5	UL48 (Teg)	Large tegument protein (131)
m48.1·		73562	73870	103	11.1		
m48.2·	C	73574	73867	98	9.8		
M49	C	73922	75529	536	60.8	UL49	
M50	C	75504	76451	316	34.7	UL50	
M51**	C	76518	77216	233	25.1	UL51	
M52		76915	78465	517	59.2	UL52	
M53		78461	79459	333	38.3	UL53	
M54	C	79704	82994	1,097	123.8	UL54 (DNApol)	DNA polymerase delta subtype (73)
M55	C	83006	85816	937	104.9	UL55 (gB)	Glycoprotein B (29, 119)
M56	C	85719	88112	798	89.0	UL56 (NM)	HSV ICP18.5 homolog (12, 51)
M57	C	88322	91894	1,191	131.4	UL57 (MDBP)	Major DNA binding protein (4, 5)
m58		91761	92462	234	25.4		
m59		93241	94260	340	35.5		

Continued on following page

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 1—*Continued*

ORF	Strand	Position[b]		Length (aa)	MW (kDa)	HCMV name[c]	Comments (reference[s])[d]
		From:	To:				
M69	C	96196	98721	842	93.0	UL69	Similar to HCMV transactivator (151)
m69.1·		98530	98886	119	12.7		
M70	C	99013	101904	964	109.8	UL70 (HP)	(34)
M71		101903	102799	299	32.9	UL71	
M72	C	103034	104236	401	45.0	UL72 (dUTPase)	(118)
M73		104100	104516	139	14.9	UL73	
m74	C	104499	105812	438	49.1	UL74 P	Possible glycoprotein
M75	C	106113	108287	725	81.3	UL75 (gH)	Glycoprotein H (30, 154)
M76		108384	109145	254	28.9	UL76	
M77		108931	110814	628	68.6	UL77	Pyruvoyl decarboxylase homolog (156)
M78		110989	112401	471	51.5	UL78	GCR homolog (51)
M79	C	112642	113415	258	29.4	UL79	
M80		113414	115504	697	74.0	UL80 (AP)	Assembly protein, protease (9, 84)
M82	C	115714	117507	598	67.4	UL82 (pp71)	Upper matrix phosphoprotein (123)
M83	C	117617	120043	809	90.9	UL83 (pp65)	Lower matrix phosphoprotein (123)
M84	C	120085	121845	587	64.5	UL84	Early nuclear nonstructural protein (53)
M85	C	122192	123124	311	34.6	UL85	
M86	C	123202	127260	1,353	151.4	UL86 (MCP)	Major capsid protein (23)
M87		127383	130160	926	102.4	UL87	
M88		130243	131520	426	47.3	UL88	
M89 Ex2		131649	132771	374.3	42.8	UL89 (CHS)	Conserved herpesvirus spliced gene (27)
m90	C	132923	133876	318	35.8		Potential MGP
M91		133668	134069	134	14.1	UL91	
M92		134075	134764	230	25.4	UL92	
M93		134733	136277	515	58.1	UL93	
M94		136234	137268	345	37.7	UL94	
M89 Ex1	C	137393	138283	296.7	34.0	UL89 (CHS)	Conserved herpesvirus spliced gene (27)
M95		138282	139532	417	45.8	UL95	
M96		139535	139921	129	14.2	UL96	
M97		140141	142069	643	71.1	UL97 (PK)	Phosphotransferase, HCMV UL97 phosphorylates GCV (82, 139)
M98		142101	143783	561	62.0	UL98 (DNase)	Exonuclease (96)
M99		143723	144058	112	11.9	UL99 (pp28)	Phosphoprotein (100)
M100	C	144299	145411	371	42.3	UL100 (gM)	Glycoprotein M or integral membrane protein (79, 126)
M102		145596	148031	812	91.0	UL102 (HP)	Helicase-primase component (85, 129)
M103	C	148185	149135	317	35.4	UL103	(85)
M104	C	149116	151227	704	80.6	UL104	Structural protein (131)
M105		151028	153871	948	106.4	UL105 (Hel)	DNA helicase (90)
m106	C	153916	154356	147	15.9		
m107·		161983	162675	231	24.6		
m108·	C	162213	162770	186	19.8		
M112 Ex1		163097	163889	264.3	27.7	UL112	Exon 1 of e1 (17); total e1 length is 330 aa, MW is 36.4 kDa
M113		163983	165076	364.7	38.3	UL113 P	Potential alternative splice to M112 Ex1 (17), as found for HCMV UL112/UL113 (153)
M112 Ex2		163983	164159	59	6.2	UL112	Exon 2 of e1 (17)
M112 Ex3		164486	164505	6.7	0.7	UL112	Exon 2 of e1 (17)
M114	C	165599	166384	262	29.6	UL114 (UNG)	Uracil DNA glycosylase homolog (152)
M115	C	166387	167208	274	31.2	UL115 (gL)	Glycoprotein L (67, 155)
M116	C	167208	169142	645	66.1	UL116	Serine-alanine-rich glycoprotein
m117·	C	169216	170910	565	61.4		
m117.1·		169541	170953	471	45.3		
M118	C	170983	171633	217	24.8	UL118	Possible splice to m119 as for HCMV UL118 (77, 120)
m119	C	171646	171945	100	11.6		Conserved UL119 position
m119.1	C	172059	172991	311	35.2		Potential glycoprotein
m119.2	C	173025	173390	122	13.5		
m119.3	C	173413	173721	103	10.8		
m119.4·	C	174057	174335	93	10.8		
m119.5·		174154	174486	111	11.9		
m120	C	174302	174574	91	10.1		
M121	C	175682	177775	698	73.2	UL121 P	Serine-alanine-rich glycoprotein with low homology to HSV-1 ICP0
M122 Ex5	C	177983	179517	511.7		UL122 (IE2)	MCMV ie3, with mRNA terminating at base 177817 (98); total length of Ex2 plus Ex3 plus and Ex5 is 611 aa, MW is 68.1 kDa
m123 Ex4	C	179763	181249	495.7			ie1 exon 4, with mRNA terminating at base 179544 (68–70); total length of Ex2 plus Ex3 plus Ex4 is 595 aa, total MW is 66.7 kDa
m123 Ex3	C	181372	181562	63.7			ie1/ie3 exon 3
m123 Ex2	C	181660	181766	35.7			ie1/ie3 exon 2; initiating ATG for ie1/ie3 at nucleotides 181764 to 181766
m124		181933	182277	115	12.4		
m124.1·	C	182014	182418	135	15.1		
m125		183436	183762	109	12.8		
m126		184535	184807	91	10.4		

Continued on following page

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 1—*Continued*

ORF	Strand	Position[b]		Length (aa)	MW (kDa)	HCMV name[c]	Comments (reference[s])[d]
		From:	To:				
m127	C	185193	185591	133	14.9		
m128 Ex3		186085	187296	404	45.5	US22 (GF2)	ie2 exon 3, with mRNA terminating at base 187353 (98); additional C residue at base 187166 compared to that in reference 98
m129	C	187350	187847	166	18.8		
m130•		187807	188277	157	17.5		
m131•	C	188029	188376	116	13.0		
m132 Ex2	C	188382	188601	7	37.0		Exon 2 of M133(sgg1) (74, 87); total MW of Ex1 plus Ex2 shown
m133 Ex1	C	188881	189795	305			Exon 1 of M133(sgg1) (74, 87); alternative donor present in Smith strain at position 188840
m134•		189868	190278	137	14.9		
m135•	C	189898	190221	108	11.6		3′ coterminal with M133(sgg1) in MCMV strain K181 and shares splice acceptor sequence with exon 2 (74)
m136	C	190313	191071	253	28.8		
m137	C	191091	192092	334	37.1		Possible glycoprotein
m138	C	192236	193942	569	63.1	fcr1	Fc receptor glycoprotein (142)
m139	C	194085	196016	644	71.8	US22 (GF2)	US22 family homolog
m140	C	196065	197516	484	55.8	US23 (GF2)	US22 family homolog
m141	C	197708	199231	508	57.2	US24 (GF2)	US22 family homolog
m142	C	199444	200748	435	49.0	US26 (GF2)	US22 family homolog
m143	C	200923	202593	557	63.6	US23 (GF2)	US22 family homolog
m144	C	202745	203893	383	42.7		MHC class I homolog (10, 14)
m145	C	204032	205492	487	54.6		Member of MGP family m145
m146	C	205645	206775	377	42.2		Member of MGP family m145
m147•	C	206865	207299	145	16.9		Possible multiple membrane spanning protein
m148•		206928	207284	119	13.6		
m149		207326	208012	229	25.6		
m150	C	207626	208789	388	42.8		Member of MGP family m145
m151	C	208817	209983	387	42.4		Member of MGP family m145
m152	C	210244	211377	378	43.5		Member of MGP family m145
m153	C	211590	212804	405	45.5		Member of MGP family m145
m154	C	212945	214048	368	37.7		Threonine-serine-rich glycoprotein of MGP family m145, some homology to EHV1 g(X) (141)
m155	C	214437	215567	377	42.8		Member of MGP family m145
m156	C	215537	215977	147	16.8		
m157	C	215898	216884	329	37.1		Member of MGP family m145
m158	C	216935	218002	356	40.0		Member of MGP family m145
m159	C	218173	219366	398	44.1		Putative membrane glycoprotein
m160	C	219601	220524	308	34.3		Putative membrane glycoprotein
m161	C	220475	221149	225	25.5		Putative membrane glycoprotein
m162	C	221189	221665	159	16.7		Putative membrane glycoprotein anchor sequence only
m163	C	221878	222414	179	19.1		Putative membrane glycoprotein
m164	C	222369	223649	427	46.6		Putative membrane glycoprotein
m165	C	223283	224278	332	35.8		Putative membrane glycoprotein
m166	C	224416	225561	382	42.4		Putative membrane glycoprotein
m167	C	225782	227089	436	47.7		Putative membrane glycoprotein
m168		227920	228462	181	19.9		
m169	C	228313	228708	132	15.3		
m170	C	229342	230046	235	26.5		

[a] Abbreviations and symbols: MW, molecular mass; Ex1, Ex2, and Ex3, exons 1, 2, and 3 of spliced genes; MGP, potential membrane glycoprotein; P, positional homolog; •, ORFs that overlap another gene for greater than 60% of their length, that have no homologs in herpesviruses or cellular proteins, and that are less likely to be coding; **, ORFs trimmed to a second methionine downstream of the 5′ end, on the basis of alignment with HCMV, HHV-6, and/or overlap with adjacent ORFs; gp, glycoproteins; GF2, members of the US22 gene family; GF1, members of the UL25 gene family; DPAP, DNA polymerase accessory protein; RRL, large subunit of ribonucleotide reductase; DNApol, DNA-dependent DNA polymerase; NM, homolog of HSV ICP 18.5 involved in nucleocapsid maturation; HP, component of the helicase-primase complex.

[b] "From" and "To" indicate the limits of the MCMV genes on the prototype genome arrangement. References are given to published data for the HCMV homolog or the MCMV gene.

[c] HCMV homologs or alternative (descriptive) names for MCMV-specific ORFs.

[d] Known exon length where the exon length has been determined from transcription data and differs from the predicted ORF length.

[e] The ORF M45 has not been trimmed, as there is similarity between the MCMV and HCMV homologs extending upstream of the first methionine, which would be consistent with splicing of this ORF in MCMV and HCMV.

M112-M113(e1) and the IE genes M122(ie3), m123(ie1), and M128(ie2), as indicated in Fig. 3 and in Table 1. An alternative splice donor for M112 predicted from sequence analysis is shown in Table 3. The M112-M113 gene would potentially encode a much longer carboxy tail if the donor represented in Table 3 rather than the published M112 site were used (17). The major protein translated from the e1 gene is approximately 36 kDa in size, although four proteins in the size range 33 to 38 kDa were recognized by immunoprecipitation of viral proteins using murine monoclonal antibodies. If the alternative

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

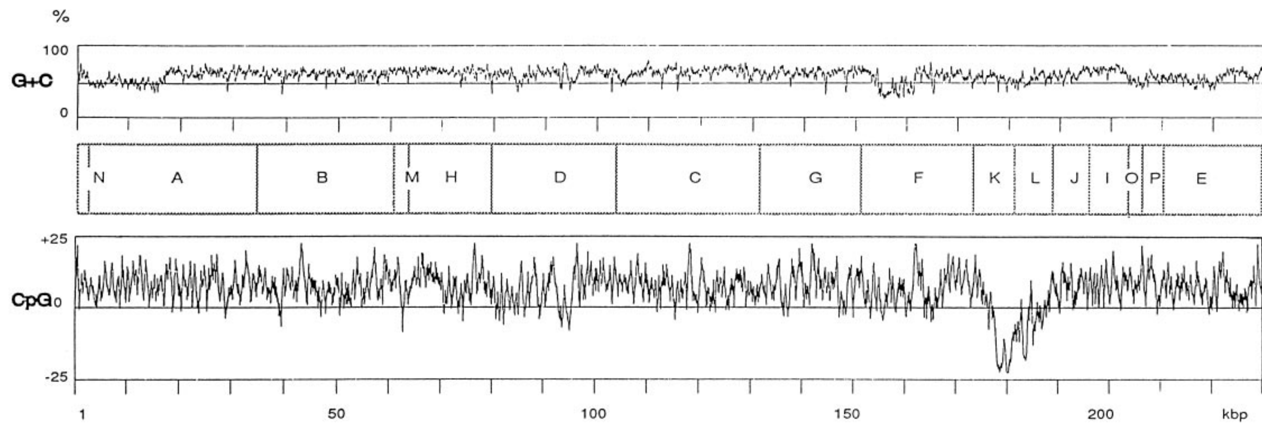

FIG. 1. The G+C content of the entire MCMV (Smith) genome is shown on the upper plot, with the horizontal line representing 50% G+C content. The corresponding CpG content is shown on the lower plot, with the lower horizontal line representing zero CpG content. The lowest horizontal scale is divided into 10-kbp segments. Both figures were generated with xnip software, with a window length of 20 (133). The *Hin*dIII restriction fragments are shown in the central part of the figure (43).

donor site present in M112 were used (see Table 3), a primary translation product of around 66 kDa would be predicted. Additional transcription data are required to decide which of these two splice donor sites are used by the virus. The donor and acceptor sites shown in Table 3 that have not been mapped were predicted on the basis of homology with consensus herpesvirus splice sites. These include the MCMV ORFs downstream of m123(ie1) and M122(ie3). These ORFs are similar to HCMV UL118, UL117, UL116, and UL115, which have been shown to undergo differential splicing, resulting in 3′ coterminal transcripts at IE and late times after infection (77, 120). Positional or amino acid sequence homologs of HCMV UL115-119 were identified in MCMV, and there were predicted splice sites for the homologs of UL119/118 identified (Table 3). Consensus splice sites were found within MCMV M36 but not in M37. These ORFs are the homologs of known spliced HCMV IE genes UL36 and UL37 (72). Predicted splice sites within the MCMV homolog of the conserved late spliced gene of herpesviruses (M89) are consistent with known splicing occurring in the homologous gene of HCMV (120).

In the more virulent K181 strain of MCMV, the m133 (sgg1) gene has been shown to be spliced (74, 87). The ORFs encod-

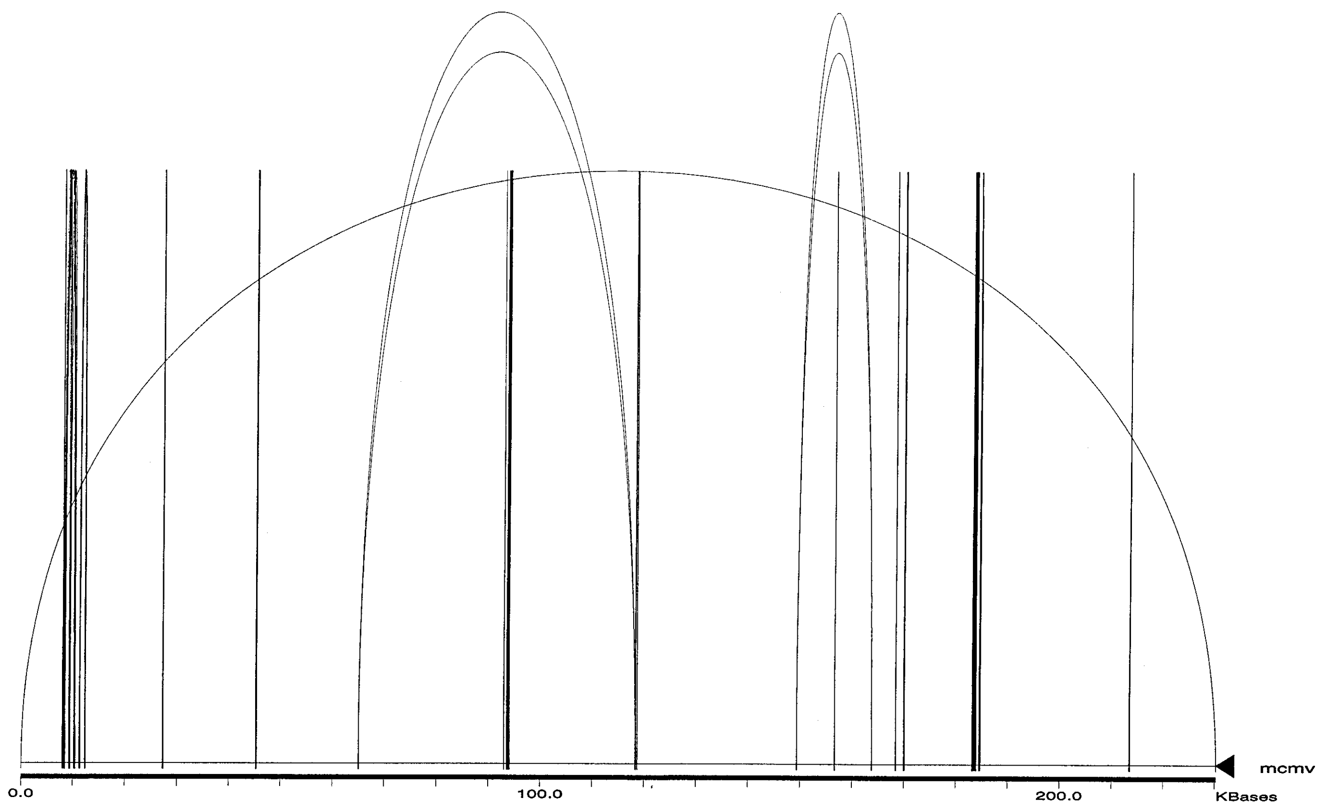

FIG. 2. DR sequences within the MCMV genome (of more than 30 contiguous nucleotides) are joined by shorter curved or vertical lines. Inverted repeats are joined by the taller curved lines.

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

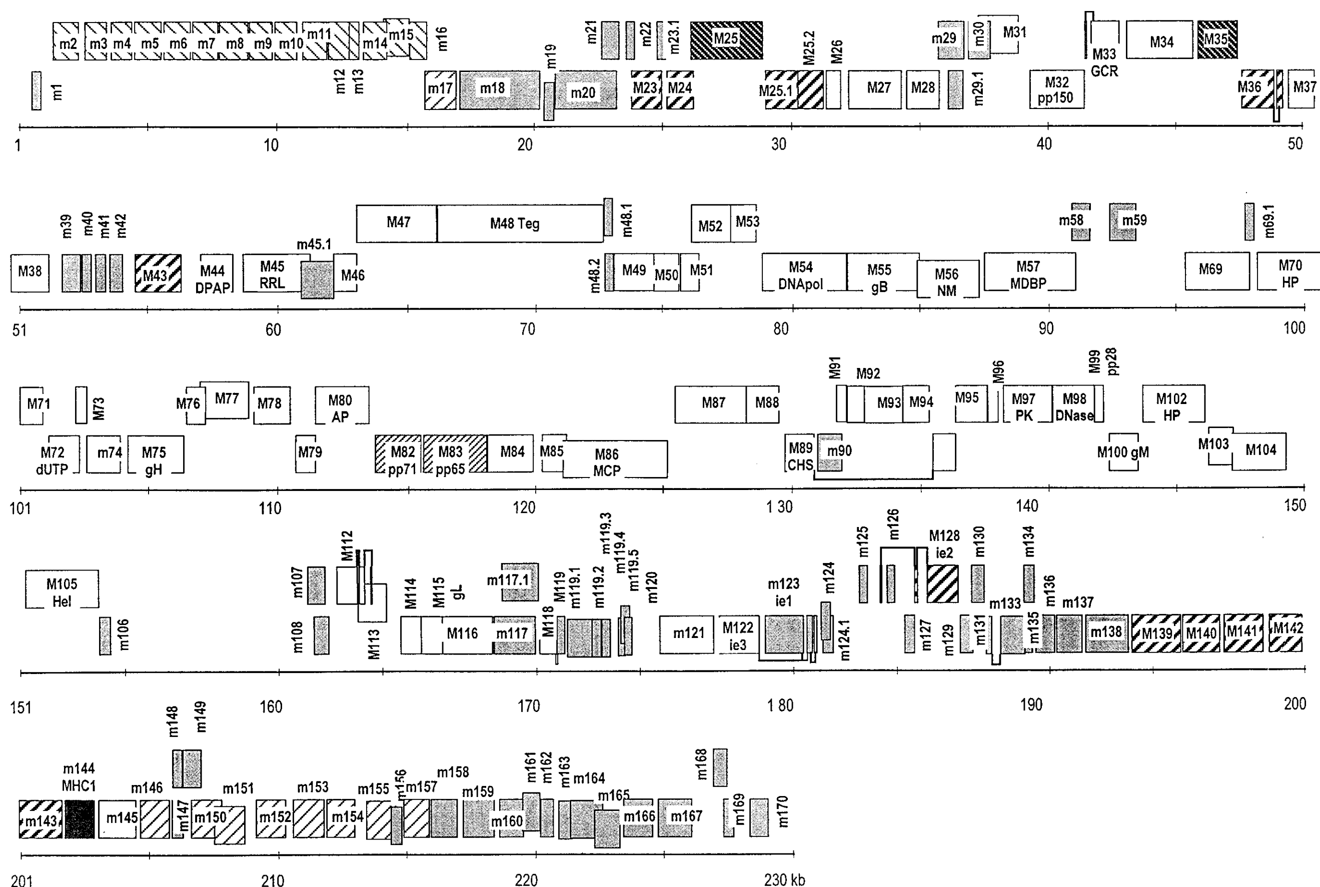

FIG. 3. Map of MCMV (strain Smith) showing the ORFs on the top strand (coding left to right) above those on the bottom strand (coding right to left) from the start (base 1) to the end (base 230278) of the prototype arrangement of the genome. MCMV predicted genes are numbered (as in Table 1) from m01 to m170, with a suffix (such as 25.1) introduced to keep the MCMV and HCMV gene numbering systems congruent. Established HCMV homolog names are shown (131), and ORFs encoding known or predicted exons are joined by lines. Shading is used to distinguish genes unique to MCMV from those with HCMV homologs (no shading unless a member of a gene family). The gene families are shaded, with the m02 (▧), M25 (▩), US22 homologs (▨), matrix proteins M82 and M83 (▨), and m145 (▨) families distinguished. Abbreviations for homologs and/or corresponding genes of HCMV (shown in brackets): AP, assembly protein (UL80); CHS, conserved herpesvirus spliced gene (UL89); DNase, alkaline exonuclease (UL98); DNApol, DNA-dependent DNA polymerase (UL54); DPAP, DNA polymerase accessory protein (UL44); dUTPase (UL72); GCR (UL33); gB (UL55); gH (UL75); gL (UL115); GF1, members of the UL25 gene family (UL25, UL35); GF2, members of the US22 gene family; Hel, helicase-primase subunit containing conserved motifs (UL105); MHCI, MHC class I homolog (UL18); HP, component of the helicase-primase complex (UL70); ie, spliced immediate-early glycoprotein (UL36); ie1, immediate-early gene 1 (UL123); ie2, immediate-early gene 2; ie3, immediate-early gene 3 (UL122); MCP, major capsid protein (UL86); MDBP, major single-stranded DNA binding protein (UL57); NM, homolog of HSV ICP 18.5 involved in nucleocapsid maturation (UL56); PK, phosphotransferase, protein kinase, or ganciclovir kinase (UL97); pp28, 28-kDa phosphoprotein (UL99); pp65, phosphoprotein 65 (UL83); pp71, phosphoprotein 71 (UL82); pp150, large structural phosphoprotein (UL32); RRL, large subunit of ribonucleotide reductase (UL45); UNG, uracil-DNA glycosylase (UL114).

ing the Smith strain sgg1 gene (m133 and m132) lie on the complementary strand near m128(ie2) (Fig. 3) between nucleotides 189795 and 188382. In the K181 strain, these encode a major 1.5- and a minor 1.8-kb transcript at early times postinfection (87). Use of an additional splice donor site in Smith strain (at nucleotide 188840) would produce a 392- rather than a 312-residue protein. The published data for strain K181 suggest that this gene may encode proteins of 30 to 70 kDa in size, and the longer putative protein here would be predicted to be 44 kDa in size. Expression of the gene in cell culture produces a 37-kDa product (74). Recent evidence has shown that both HCMV UL33 and MCMV strain K181 M33 are spliced at the 5′ end, and the sequences at the splice sites are shown in Table 3. MCMV (M33) utilizes consensus donor and acceptor sequences located near the predicted amino terminus of the gene (35). The HCMV homolog (UL33) has also been shown to be spliced, and the HHV-6 homolog (U12 or P1RF1) is predicted to be spliced at a similar position in the gene (35).

(ii) Proteins essential for origin-dependent replication. MCMV homologs of six of the seven proteins necessary for HSV-1 replication were identified, namely, the DNA polymerase (M54), a polymerase accessory protein (M44), the single-stranded DNA binding protein (M57), and the helicase-primase complex of three proteins (M70, M102, and M105). The six motifs described in DNA helicases, including those of HSV (157) and HCMV, are found in congruent positions in MCMV M105(Helicase). Like EBV and HCMV, MCMV does not encode a gene with sequence similarity to that of HSV-1 UL9 (origin binding protein). HHV-6 is known to encode an origin binding protein (U73) and therefore differs from the other sequenced betaherpesviruses in this respect (51). MCMV does not possess a sequence or positional homolog of HCMV UL101. At least part of UL101 lies within a region that encodes *trans*-acting factors that function in HCMV DNA replication in transient transfection assays (112). Five additional loci (comprising nine ORFs) that are required for transient complementation of the HCMV lytic origin of replication (112) are present in MCMV, including homologs of HCMV UL112/113 [M112/113(e1)] and UL84 (M84) (17, 53, 136). HCMV UL84 is known to interact with UL122(IE2), possibly

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 2. Similarity between the 78 MCMV genes with amino acid sequence homology to HCMV[a]

MCMV gene	Length (aa)	HCMV homolog:	Homolog length (aa)	FastA identity for overlap (%)	Overlap length (aa)	FastA score[b]
M23	391	UL23 (GF2) P[c]	342	39	54	110
M24	323	UL24 (GF2)	358	23	282	214
M25	932	UL25 (GF1)	656	22	531	378
m25.1**	427	UL23 (GF2)	342	21	270	148
m25.2**	470	US22 (GF2)	358	20	337	104
M26	192	UL26	188	37	126	284
M27	682	UL27	608	23	416	392
M28	430	UL28	379	19	246	108
M31	516	UL31	694	24	313	191
M32	718	UL32 (pp150)	1,048	25	285	329
M33	377	UL33 (GCR)	390	47	324	924
M34	854	UL34	504	27	312	470
M35	519	UL35 (GF1)	640	24	499	490
M36 Ex2	429.7	UL36 Exon 2	408.7	25	302	262
M36 Ex1**	77.3	UL36 Exon 1	67.3			
M37	345	UL37	487	20	284	162
M38	497	UL38	331	25	311	407
M43	597	UL43 P	187	25	61	95
M44	411	UL44 (DPAP)	433	59	351	1,035
M45	876	UL45 (RRL)	915	28	662	684
M46	294	UL46	290	39	268	600
M47	1,040	UL47	982	26	620	754
M48	2,149	UL48 (Teg)	2,241	22	1,595	1,216
M49	536	UL49	570	40	470	1,104
M50	316	UL50	397	46	252	610
M51**	233	UL51	157	55	78	278
M52	517	UL52	668	44	250	637
M53	333	UL53	376	49	280	788
M54	1,097	UL54 (DNApol)	1,242	50	415	1,202
M55	937	UL55 (gB)	906	44	550	1,220
M56	798	UL56 (NM)	850	42	696	1,373
M57	1,191	UL57 (MDBP)	1,235	51	597	1,717
M69	842	UL69	744	25	514	593
M70	964	UL70 (HP)	1,062	36	982	1,962
M71	299	UL71	411	41	201	492
M72	401	UL72 (dUTPase)	388	23	281	259
M73	139	UL73	138	34	95	195
M75	725	UL75 (gH)	743	23	693	783
M76	254	UL76	325	38	242	434
M77	628	UL77	642	49	478	1,191
M78	471	UL78	431	20	249	240
M79	258	UL79	295	49	246	728
M80	697	UL80 (AP)	708	30	736	478
M82	598	UL82 (pp71)	559	19	360	182
M83	809	UL83 (pp65)	561	17	323	192
M84	587	UL84	586	22	189	130
M85	311	UL85	306	55	313	931
M86	1,353	UL86 (MCP)	1,370	53	896	2,587
M87	926	UL87	941	59	408	1,266
M88	426	UL88	429	28	432	379
M89 Ex2	374	UL89 (CHS)	378	65	673	2,400
M91	134	UL91	111	39	65	147
M92	230	UL92	201	50	199	656
M93	515	UL93	594	33	89	196
M94	345	UL94	345	32	343	612
M89 Ex1	297	UL89 (CHS)	296			
M95	417	UL95	531	59	108	386
M96	129	UL96	115	26	94	106
M97	643	UL97 (PK)	707	37	441	844
M98	561	UL98 (DNase)	584	51	144	439
M99	112	UL99 (pp28)	190	35	66	95
M100	371	UL100 (gM)	372	47	350	1,026
M102	812	UL102 (HP)	798	35	220	345
M103	317	UL103	249	33	252	494
M104	704	UL104	697	42	406	763
M105	948	UL105 (Hel)	956	43	861	1,955
M112 Ex1	264	UL112	252.3	34	274	389

Continued on following page

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 2—*Continued*

MCMV gene	Length (aa)	HCMV homolog:	Homolog length (aa)	FastA identity for overlap (%)	Overlap length (aa)	FastA score[b]
M113	365	UL113 P	499			
M112 Ex2	59	UL112	252	34	274	389
M112 Ex3	7	UL112	252	34	274	389
M114	262	UL114 (UNG)	250	52	262	725
M115	274	UL115 (gL)	306	32	228	384
M116	645	UL116	344	15	172	98
M118	217	UL118	209	21	196	106
M121	698	UL121 P	180	47	19	55
M122 Ex5	512	UL122 (IE2)	495	33	346	531
m128 Ex3	404	US22 (GF2)	593	20	173	135
m139	644	US22 (GF2)	593	30	357	472
m140	484	US23 (GF2)	592	29	410	611
m141	508	US24 (GF2)	500	29	400	538
m142	435	US26 (GF2)	603	28	276	376
m143	557	US23 (GF2)	592	27	372	411

[a] For FastA calculations, the length of overlap in amino acids, percent identity of the two proteins, and raw FastA scores for the alignments are shown. See Table 1, footnote *a*, for abbreviations and symbols.
[b] In some cases, comparison of MCMV with another herpesvirus or eukaryotic gene has produced a higher FastA score; this is discussed in the text.
[c] P, positional homology, but minimal sequence homology over short regions only, for the genes M23, M43, and M121.

mediating transcriptional control (132). The alkaline exonuclease (DNase) gene (M98) that is required for HSV-1 growth in cultured cells was highly conserved in MCMV (Table 2) (96). There is a short (22-base) inverted repeat close to the start of the MCMV gene, unlike with HCMV UL98 gene.

The MCMV origin of replication has been mapped in MCMV strain K181 to a region corresponding to *Hin*dIII D of MCMV strain Smith (92). The minimum sequence functioning as the origin of replication in lytic infection has previously been shown to be 1.5 kb in length and to contain 32-bp repeats (six in strain K181 [92] and four in strain Smith sequenced here). The origin is located upstream of the single-stranded DNA binding protein (M57) within an A+T-rich region of the genome, lying approximately between nucleotides 93870 and 94500 (Fig. 1). This region in strain Smith contains two potential ORFs (m58 and m59), although they have no similarity to known genes of other herpesviruses.

(iii) Other genes involved in DNA and nucleotide metabolism. Homologs of enzymes found in other herpesviruses, which are required for nucleotide metabolism, replication, and repair,

TABLE 3. Location and nucleotide sequences of the donor and acceptor sites of predicted MCMV homologs of HCMV spliced ORFs and known MCMV ORFs[a]

MCMV ORF	HCMV homolog	Splice position		MCMV sequence	
		Donor	Acceptor	Donor	Acceptor
Consensus				NAGgtaagt	nttntntttttttncagG
M33 (GCR)	UL33	41518	41680	ATGgtaagc	tcgccctctcgcccgcagG
M36	UL36	49036	48909	AAGgtgagt	cgacgtcctccctaaaagG
M89	UL89	137393	132770	AACgtgagt	cgtctcttcctcacgcagA
M112 (don)	UL112	163889		AATgtaagt	
M113 (aac)	UL113		163983		gttactctgttcggaaagG
M113[b]	UL113	164159	164486	ATGgtgcgt	cttcctccgaattcacagG
m119 (don)[c]	UL119	171684		ACAgtaagt	
M118 (acc)	UL118		171581		tgtctttctctcccacagC
m123 (ie1)	UL123 (IE1)[d]				
Exon 1			182596	AATgtaagt	
Exon 2		181660	181766	CAAgtaagt	ttatattttttagagagA
Exon 3		181372	181562	CAGgtgagt	ccactcttgttctctcagG
Exon 4			181249		accactcttatattccagT
M122 (ie3)	UL122 (IE2)		179517		actctttatggttcacagG
Exon5	Exon5				
m128 (ie2)	Nil				
Exon 1		184347		CGGgtaagg	
Exon 2		185717	185594	CTAgtgagt	ctctcttgttctccatagG
Exon 3			186083		ggttctgtttgtctgtagA
m132	Nil	188881		GCGgtacgt	
		188840[e]		GTGgtacgt	
m133 (sgg1)	Nil		188601		cccgtttttgtctttcagG

[a] The locations of the last base of exon 1 and the first base of exon 2 (underlined) are shown for the prototype arrangement of the MCMV genome. When splicing is between two ORFs, these ORFs are indicated as don or acc. Additional transcription data from references 70, 97, and 98.
[b] Transcription data from reference 17 describes an additional intron, resulting in a frameshift and a stop at position 164505 rather than 165076 as predicted here.
[c] MCMV m119 conserves position but not amino acid sequence with HCMV UL119.
[d] MCMV ie1 conserves position and function but not amino acid sequence with HCMV UL123(IE1).
[e] The second donor present in strain Smith (but not in the published strain K181 sequence [74]) would produce a larger ORF if spliced to the same acceptor.

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

TABLE 4. Members of the gene families of MCMV and homologs of HCMV gene families

Index ORF of gene family	Members of gene family	Characteristic motifs	HCMV homolog(s)	Reference(s)
m02	m02 m03 m04 m05 m06 m07 m08 m09 m10 m11 m12 m13 m14 m15 m16	NAXWXX E/H Wo, CXLXXC L/P W/R o, CXXXXC, C(X)$_5$CIXA T/S, WrXXW, CXXXoC, oE M/R Q (for m07 to m10 genes), CXX T/I XC (for m11 to m16 genes)	Nil	
M25	M25 M35		UL25 family UL25, UL35	22
M33	M33, M78	G protein-coupled receptor motif (M33 only), multiple transmembrane domains, S-T-rich carboxy terminus	UL33 US27 US28, UL78	21, 35, 51, 150
M23	M23 M24 m25.1 m25.2 M36 M43 m128(ie2) m139 m140 m141 m142 m143	oCCXD/EX$_{1-4}$oXXoG, two runs of hydrophobic residues, ooooDXXGXo, carboxy-terminal acidic residues	US22 family	72, 150
M82	M82(pp71) M83(pp65)		UL82 family	123
m145	m145 m146 m150 m151 m152 m153 m154 m155 m157 m158 m17	—[a]	Nil	

[a] See text.

including uracil-DNA glycosylase (M114), ribonucleotide reductase (M45), and dUTPase (M72), are found in MCMV. The ribonucleotide reductase homologs of both MCMV and HCMV are smaller than the large subunit of the HSV enzyme (97 kDa compared with 124 kDa for HSV), both MCMV and HCMV lack a homolog of the small subunit of the enzyme, and both may not be functional as ribonucleotide reductases. In MCMV the position analogous to the location of the small subunit is occupied by M44, which is a homolog of the essential DNA polymerase accessory protein of HCMV (45). The dUTPase homolog in MCMV does not contain motifs usually found in functional dUTPase enzymes (116).

As was found for HCMV, MCMV does not encode a thymidylate synthetase or a thymidine kinase gene. An MCMV homolog (M97) of the herpesvirus phosphotransferases including HCMV (UL97), HHV-6 (U69), HHV-7 (107a), and HSV (UL13) was identified. This gene conserves the domain structure of eukaryotic protein kinases, and the strongest similarity between the genes is in the carboxy-terminal region. In HCMV, UL97 has been shown to phosphorylate GCV (82), and MCMV has also been shown to be susceptible to this drug in infection of mice (107). MCMV, like HCMV, does not encode additional protein kinases, as are present in some alphaherpesviruses, such as HSV-1 US3 and CCV ORFs 73, 74, 14, 15, and 16 (37). HCMV is predicted to contain a putative pyruvoyl decarboxylase gene (UL77). The prediction was based upon the presence of a single pyruvoyl decarboxylase enzyme prosthetic group (PDEPG) site, with the sequence TLGSSLFN (156). MCMV encodes a strong homolog of UL77 (M77), with the sequence TLGQQFFN in the region of the gene congruent with the potential PDEPG site. HHV-6 contains a homolog of UL77 (U50), which is shorter by 73 residues and does not contain a PDEPG site (51).

(iv) Regulatory genes. MCMV homologs of HCMV regulatory genes, whose products are important in the temporal cascade of viral gene expression, include genes within the MIE region (138), the UL36 to UL38 region (72), the TRS1-IRS1 genes (137), and UL69 (151). However, no homolog of the differentially spliced HCMV US3 transcription unit was found (26, 150). The organization, sequencing, and mapping of the 5′ coterminal transcripts of the MCMV MIE gene locus have been previously described (68). The MIE region of MCMV is roughly colinear with that of HCMV, and matching splicing patterns between HCMV IE1-IE2 and MCMV ie1-ie3 have been characterized (68–70, 97). In both viruses transcription from the MIE genes is controlled by a strong enhancer (42). The spliced m128(ie2) gene of MCMV has been previously sequenced and has sequence similarity to members of the US22 family of HCMV (98) and MCMV. The sequence determined here has an additional C nucleotide at position 187166, resulting in a frameshift at the carboxy terminus of m128(ie2). The m128(ie2) gene is therefore predicted to encode a slightly longer peptide of 404 (rather than 391) amino acids (Table 1).

MCMV (and HHV-6 [109]) do not encode homologs of UL37 exons 1 and 2. HCMV TRS1-IRS1 are members of the US22 gene family and encode a transactivator(s) of HCMV UL44. m143 is the US22 gene family homolog that is most similar to TRS1-IRS1 (Table 4). However, the level of amino acid similarity between m143 and TRS1-IRS1 is considerably lower than that between other US22 members, such as US23, US22, US26, and US24. MCMV contains a homolog (M44) of HCMV UL44 (Tables 1 and 2). HCMV UL44 is the putative target gene that is transactivated by HCMV TRS1. MCMV M44 has a distinctive motif repeated at nine locations throughout the gene, that of a lysine followed by two charged residues. Although similar motifs are observed in HCMV UL44 (45), they are restricted to the carboxy terminus and are of uncertain functional significance.

HCMV UL69 is a transactivator of early genes, and an MCMV homolog (M69) was found. Homologs of these genes have also been detected in the alpha- and gammaherpesviruses, where they have also been defined as *trans*-regulators of gene expression (151). The MCMV homolog contains a signature sequence present in mitochondrial energy transfer proteins (105), but this sequence is not present in HCMV UL69, and it is very unlikely to be biologically significant. Like HCMV UL69, MCMV M69 differs from the homologous genes of other herpesvirus subgroups in having a unique carboxy-terminal amino acid sequence. MCMV ORFs M121 and M116 were found to have sequences similar to the HSV-1 IE transactivator ICP0 (Table 1). This similarity may result from compositional bias, given the high content of serine, threonine, and alanine residues in HSV-1 ICP0 and the MCMV proteins.

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

Comparison of M121 and HSV-1 ICP0 shows a FastA score of 209 and 20% amino acid identity over a 435-amino-acid overlap, with alignment beginning around position 230 in both proteins and extending to the carboxy-terminus. MCMV M116 also has a FastA score of 215, with similar levels of homology, but the alignment begins between M116 position 20 and ICP0 position 385.

(v) Structural proteins. MCMV encodes homologs of the herpesvirus major capsid protein (M86), the large tegument protein (M48), and a minor capsid protein (M46). Homologs of the HSV-1 capsid components VP19C (HSV UL38), VP26 (HSV UL35), and the tegument protein Vmw65 (HSV UL48) were not found in MCMV by using amino acid identity searches. Found were homologs of genes encoding structural proteins unique to the beta-herpesvirus group, which includes the upper (M82) and lower (M83) matrix proteins as well as the large (M32) and small (M99) phosphoproteins. The product of the M99 gene has been shown to be around 16 kDa in size and to be associated with the virion (31). The HCMV UL80 assembly protein is presumed to function in packaging DNA. A protease is encoded by the N-terminal region of UL80 that cleaves the assembly protein precursor at a site near the C terminus. The MCMV homolog (M80) of the assembly protein of HCMV (UL80) conserves the domain structure and cleavage sites present in HCMV UL80. MCMV is different in that the conserved CD1 and CD2 domains are separated by 100 amino acids, whereas in all other sequenced herpesviruses the two domains are separated by 80 to 84 residues (84). MCMV M56 is homologous to HCMV UL56, HSV-1 UL28 (icp18.5), and the processing proteins of equine herpesvirus 1 (EHV-1), pseudorabies virus (PRV), EBV, and VZV (94). The 130-kDa product of HCMV UL56 is associated with the nucleocapsid-tegument fraction and has been shown to be nonglycosylated (12). The gene product may be involved in capsid maturation, by analogy with the HSV-1 gene product (94). The sequence determined here for M56 differs from the published sequence of strain Smith (99) by a single conservative change from an alanine to a threonine at position 235.

(vi) MCMV glycoproteins with homologs in HCMV. A total of 64 genes that could possibly encode glycoproteins were identified by similarity to known herpesvirus glycoproteins or by the presence of likely *trans*-membrane hydrophobic sequences and potential N-linked glycosylation sites (of the form NXT/S, where X represents any amino acid) (93). Potential MCMV glycoproteins include homologs of the conserved herpesvirus glycoproteins gB (M55), gH (M75), gM (M100), and gL (M115). MCMV M115(gL) has significant amino acid similarity to the gL homologs present in HCMV and HHV-6. No homologs of the HCMV glycoproteins gpUL4(gp48), gpUS10, gpUS11, or gpUL16 were found in MCMV (131). The amino acid sequence of gB (M55) determined here differs from that previously found for strain Smith by having 10 additional amino acid residues and 22 amino acid transversions (44, 119). The gH (M75) sequence found here has an additional amino acid (C604) and 10 nonconservative and 22 conservative amino acid substitutions, compared with the published sequence of strain K181 (154).

Previous studies have established that gL from the alpha-, beta-, and gammaherpesviruses associates with gH (67). MCMV M115(gL) contains five potential glycosylation sites, and it has previously been shown to be glycosylated in virions (155). The MCMV gM homolog (M100) has seven hydrophobic stretches that are potential membrane-spanning regions, and the hydropathicity plot resembles that of HCMV UL100. The derived amino acid sequence contains four potential N-linked glycosylation sites, two of which are congruent with the sites present in HCMV. The MCMV homolog lacks the highly acidic carboxy-terminal residues present in UL100 of HCMV. Also, the MCMV M100 gene in strain Smith may have an additional 24 residues at the amino terminus, as there is an additional Kozak consensus methionine upstream of the homologous methionine found in HCMV. However, transcription data in strain K181 suggest that the M100 gene does not contain these additional 24 residues, as the mRNA start site is at nucleotide 145373 or 145368, with the translation start at 145357 (79, 126).

The m138 gene has previously been described as encoding a receptor for the Fc domain of murine immunoglobulin G molecules (142). The gene lies on the complementary strand, between nucleotides 192236 and 193942, with the mRNA start site at 193998 (Table 1). A 1.9-kb transcript has been described from 2 h postinfection, with maximal transcription between 8 and 24 h postinfection (142). The other membrane glycoproteins, which occur as gene families, are described below.

Families of homologous genes of MCMV. MCMV encodes six groups of gene families. Two of these families are unique to MCMV, and four have homologs in HCMV (Table 4).

(i) The US22 family. Members of the US22 gene family have previously been found in HCMV (72, 150), HHV-6 (109), and HHV-7 (107a). Twelve ORFs with homology to members of the US22 family are present in MCMV. The homologs are arranged in a similar fashion along the MCMV and the HCMV genomes, as two clusters of tandemly repeated homologs at either end of the genome, with additional single genes located elsewhere. They are characterized by short stretches of hydrophobic and charged residues, as well as one or more conserved sequence motifs (Table 4). Four sequence motifs have been identified previously in the members of this family in HCMV, and these four motifs as well as an additional motif were found in the MCMV family members. The most strongly conserved motif is ooCCXD/E(X)$_{1-4}$oXXoG (where o represents any hydrophobic amino acid and X represents any residue). This motif occurs first in the predicted peptide sequence of all members of the US22 gene family, with the exception of M43 and m128, and is only partially conserved in m25.1. This motif has also been identified in members of the US22 family in HCMV (72) and in HHV-6 (109). The acidic domain present in the carboxy-terminal region is most prominent in M23, m139, and m142.

The functions of the members of the US22 family in HCMV and MCMV are unclear, although mutagenesis of at least two members of the US22 family [m128(ie2) and m139] in MCMV does not result in a phenotype different from the wild type when it is grown in cell cultures (86, 146). In other beta-herpesviruses some members of this family are transactivators [HCMV TRS1-IRS1 participates in transactivation of HCMV UL44(icp36) (137)], and HHV-6 EPLF3 is able to transactivate the HIV-1 LTR (109). Multiple alignment of the 12 members of the family in MCMV showed that ORFs m139 to m143 at the righthand side of the genome were more similar to each other and to the colinear US22 to US26 members of HCMV than to the US22 family members located at the lefthand side of the MCMV genome. The converse was also true, in that MCMV US22 members at the lefthand side of the genome were found to be more similar to each other than to members at the righthand side of the genome. In addition, an amino-terminal motif (GXXoXoXWP) that is conserved in HCMV UL23, UL24, and UL36 (exon 1) and in HHV-6 ORFs U2, U3, U17, and U25 is present only in MCMV M24 and M36 (exon 1). The multiple alignments suggest that US22 members may be subdivided to describe ORFs which are UL-like (i.e., similar

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

to US22 members in the long unique region of HCMV) or US-like (most similar to US22 to US26).

(ii) The MCMV m02 family of genes. The MCMV family m02 contains 15 members. The members differ in length from 133 amino acids (m13) to 356 amino acids (m08). Hydropathicity plots (not shown) demonstrate the presence of a hydrophobic carboxy-terminal (anchor) sequence and a shorter amino-terminal signal sequence in all members of the family. All members of the m02 gene family lie in tandem on the top strand at the left end of the prototype MCMV genome. They all contain the sequence CXXXXC, contain potential N-linked glycosylation sites (of the form NXT/S), and resemble each other in global sequence alignments (with FastA scores of more than 100 to other members of the m02 family but not to other MCMV genes). Alignment of the 15 sequences demonstrates that the genes cluster into three groups (m02 to m06, m07 to m10, and m11 to m16) if conserved cysteine residues are used as a primary criterion for alignment.

The m02, m03, m04, m05, and m06 genes all contain the sequences CXLXXCL/PW/Ro and a proline-rich region in the carboxy third of the gene. In addition, the m03, m04, m05, and m06 genes all contain the sequence NAXWXXE/HWo toward the start of the predicted protein sequence. The m07 to m10 genes all contain the motifs $C(X)_5$CIXAT/SWrXXW (where r is a charged or polar amino acid), CXXXoC, and oEM/RQ. The m07, m08, and m09 sequences align more closely with each other than with m10 and contain several other highly conserved motifs that are not present in m10. The m11 to m16 genes vary in length more than the other two subgroups (Table 1), but all have similarly located cysteine residues and conserve a CXXT/IXC motif.

(iii) The MCMV m145 family of genes. Members of the m145 gene family are found on the complementary strand of the prototype genome, arranged in tandem toward the right end, except for m17, which lies at the other end of the genome immediately at the end of the gene block containing the m02 gene family (Fig. 3). The genes vary in length from 329 to 487 amino acids, and all contain potential signal sequences, anchor sequences, and potential N-linked glycosylation sites (Table 1). None of the predicted genes are homologous to known herpesvirus proteins. The genes m145, m146, m150 to m154, and m17 all conserve three cysteine residues located in the middle third of the protein.

MCMV ORFs with homology to cellular genes. (i) MHC class I homolog. The predicted protein encoded by m144 is similar to class I MHC proteins. However, the MCMV MHC class I homolog does not possess positional homology or significant sequence similarity to the MHC class I homolog of HCMV encoded by UL18 (FastA score of 61 for an overlap of 40 amino acids). MCMV m144 contains four potential N-linked glycosylation sites (of the form NXT/S), which is similar to the number of potential sites present in many MHC class I molecules but many fewer than the 13 potential sites present in HCMV UL18. Nevertheless, both MCMV m144 and HCMV UL18 possess three extracellular regions that resemble the MHC class I domains α1, α2, and α3, and putative leader, transmembrane, and cytoplasmic sequences. The most strongly conserved section is the α3 domain, which contains the beta-2-microglobulin (β-2m) binding region in MHC class I molecules. The predicted MCMV MHC homolog possesses cysteine residues at three out of four sites that are conserved with MHC class I proteins, including two cysteine residues within the analogous α3 domain, which form internal disulfide bonds flanking an evenly spaced globular domain. However, the MCMV homolog is considerably shorter in the α2 domain than in MHC class I proteins or HCMV UL18, which would affect the conformation of the classical antigen-binding cleft. The presumptive CD8 binding site (the motif RFDS) within the α1 domain of MHC class I proteins is not found in either m144 or HCMV UL18 (123a).

(ii) GCR homologs. Homologs of the eukaryotic GCRs are encoded by HCMV UL33, US27, and US28 (21) and possibly UL78 (51). To date, UL33, US27, and US28 have been shown to be transcribed (148). The MCMV GCR homolog (M33) most closely resembles HCMV UL33 and is located in a colinear position. The two cysteine residues that are highly conserved between members of the GCR family (and essential for the structure of bovine rhodopsin [65]) are conserved in the M33 protein (positions C118 and C201). Two additional cysteines located toward the carboxy terminus of the proteins are conserved between M33, HCMV UL33, HCMV US28, and the human RDC1-like GCRs (80). The biological functions of the HCMV GCR proteins are as yet unknown, and in vitro studies of the MCMV GCR homologs will be of particular interest. The MCMV M33 and HCMV UL33 genes have recently been shown to undergo mRNA splicing near the amino end of the protein (35). Recently the HHV-6 U51 gene has been described as a GCR homolog, and the HCMV UL78 gene has been predicted to be a member of the GCR family (51). MCMV M78 is homologous to HCMV UL78, HCMV US28, and HHV-6 U51 and to known GCRs. The M78 gene is more closely related to the viral counterparts in HCMV (UL78) and EBV than to the cellular GCRs. The opposite is true of the HHV-6 U51 gene, which more closely resembles the cellular GCRs than the viral homologs (51).

DISCUSSION

The two strains of MCMV studied in most detail are the less virulent, laboratory-passaged Smith strain and the more virulent K181 strain (57, 58), which have different phenotypes in an infected animal. Genotypic differences between K181 and Smith strains have not been extensively studied. We have shown relatively minor differences between the DNA sequences and predicted amino acid sequences of the Smith and published K181 genes m133(sgg1) (74), M55(gB) (44, 119), and M75(gH) (154).

The complete DNA sequence of MCMV was checked against databases containing the complete sequences of HSV-1, VZV, HCMV, EHV-1, EBV, and CCV (2, 6, 22, 37, 38, 94, 141). Despite significant differences in the overall arrangement of the genomes of MCMV and HCMV (103), we have demonstrated that as previously predicted, the genomes are very similar at the genetic and nucleotide compositional levels (55). Indeed, plots of the G+C content of the genomes show similarities when both 230-kb sequences are aligned. The G+C content is reduced at the ends of the MCMV genome, regions which are predicted to encode genes of the glycoprotein families m02 (left end) and m145 (right end). The left end of the genome of HCMV AD169 has a low G+C content, but the right end has a higher G+C content of around 56%. Both MCMV and HCMV have a high G+C content immediately adjacent to the right and left ends of the genome. As the predicted coding regions of the genome have an overall higher G+C content than the noncoding regions, the relationship between the reduced G+C content in these regions and the occurrence of these gene families is interesting. It is uncertain whether the occurrence of these gene families results from evolutionary pressures applied by the host immune response to virally encoded proteins or from other causes (149). It is noteworthy that abundant early transcripts are encoded by genes near both termini (88), as these may feature in the viral charges

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

to the host immune response. As expected, MCMV genes with significant amino acid homology to HCMV genes are all located in the central parts (around 78%) of both genomes.

Transcriptional data are available for differentially spliced transcripts from ORFs M122, m123, and m128 within the MCMV MIE region (42, 69, 70, 88, 97). By analogy with HCMV, the MCMV MIE regions ie1, ie2, and ie3 (135) are predicted to extend between positions 165599 (M114) (77, 120) and 187296 [M128(ie2)]. The sequence around 700 bp upstream of the m123(ie1) gene (like the homologous region in HCMV) contains a strong enhancer, made up (at least in part) of five 94-bp tandem repeats (Fig. 2) containing many transcriptional regulation sites (42).

This study and previous work (69) have shown MCMV encodes ORFs (MIE, M36-M38, and m143) similar to three of the four regions of HCMV that are transcribed at IE times (MIE, UL36-38, TRS1, and US3). MCMV does not contain a homolog of the HCMV US3 ORF, which is nonessential for growth in cell culture (71) and nonessential for growth in SCID-hu mice (52). Unlike HCMV UL37, the MCMV M37 gene has no obvious second exon and no consensus splice donor or acceptor sites (72). The arrangement of the UL37 homolog in this region of MCMV (and that of HHV-6 [109]) therefore differs from HCMV at the nucleotide sequence level. MCMV m143 is similar to TRS1 (a member of the US22 gene family). HCMV TRS1 is expressed at IE times and transactivates the UL44(icp36) promoter (137). However, MCMV m143 is more closely homologous to HCMV US22 and US23 than to TRS1 (Table 2), so whether M143 and TRS1 are functionally related remains very uncertain.

MCMV homologs of other genes with regulatory functions in HCMV and HSV include M121, M44(DPAP), and M69. MCMV M121 has low-level sequence similarity and positional homology with HCMV UL121 (Table 3), and it is located immediately downstream of MCMV ie3 (M122). The M121 gene has low-to-moderate global similarity (19.8% identity over a 435-residue overlap) with HSV ICP0 (IE110 or Vmw110), possibly as a result of compositional bias rather than genetic similarity (94). ICP0 has a role in virus switching from latent to lytic infection and has been shown to transactivate promoters from other herpesviruses, including HCMV (46). The ICP0 protein contains a functional zinc finger sequence close to the amino terminus (47, 145) and has a carboxy-terminal region that allows heterocomplex formation (24). MCMV M121 does not contain a consensus zinc finger motif (11). Immediately before the carboxy terminus of M121 is a short stretch of charged amino acids (RRTRSVY), and the gene contains one highly charged region (of 10 acidic residues), which are features of some *trans*-acting transcriptional activators (101).

MCMV encodes proteins with amino acid similarity to all of the known enzyme homologs encoded by HCMV. To date, M54(DNA polymerase) is the only MCMV enzyme for nucleotide metabolism to have been studied in any detail (44, 110). Neither MCMV nor HCMV encodes a thymidine kinase, and neither virus has been shown to enhance the activity of cellular thymidine kinase (110). The HCMV UL97 protein kinase homolog phosphorylates the nucleoside analog GCV in cell culture (82, 139), and MCMV is susceptible to GCV (124). MCMV is more sensitive than HCMV to acyclovir, and this enhanced susceptibility may reside in the DNA polymerase (M54), rather than the phosphotransferase (M97), as mutants resistant to acyclovir also show less susceptibility to chemicals that interact with DNA polymerase such as phosphonoacetic acid (44).

Infection with HCMV and MCMV results in increased synthesis of cellular polyamines (spermine and spermidine) in cultured cells. These amine products are known to be incorporated into AD169 virions (25). Pyruvoyl decarboxylase enzymes form a PDEPG that is part of the active site of the enzyme during synthesis of cellular polyamines (156). HCMV UL77 is a pyruvoyl decarboxylase homolog and contains a PDEPG (156). MCMV M77 has strong homology to HCMV UL77, and M77 contains a single PDEPG site congruent with that of HCMV. It will be of interest to investigate whether the HCMV or MCMV pyruvoyl decarboxylase homolog can be shown to be functional in polyamine biosynthesis, particularly as inhibitors of ornithine decarboxylase (such as D,L-α-difluoromethylornithine) that reduce polyamine biosynthesis also reduce the number of HCMV particles released from infected cells in cell culture (50).

HCMV has been predicted to encode at least 35 structural proteins (22, 31, 131). MCMV homologs of the structural genes present in pre-B capsids of HCMV are M86 (the major capsid protein), M46 (a minor capsid protein), and M80; the last is a homolog of the capsid assembly proteins of simian cytomegalovirus and HCMV (UL80) (84, 122). It was not possible to determine if a homolog of the 12-kDa protein of HCMV pre-B capsids was present, as the gene encoding this has not yet been identified (131). HCMV UL80 encodes several phosphoproteins. A protease of 28 kDa (UL80a) is transcribed from the amino-terminal region. It cleaves the UL80 phosphoprotein to produce the final proteins (9, 18) in a fashion similar to the protease of HSV-1 (UL26.5) (83). There is considerable potential for the M80 gene to be used as a model to study the characteristics of the assembly protease of HCMV in the mouse system.

The MCMV genome contains sequence homologs of the HCMV tegument proteins pUL32, pUL48, pUL82, pUL83, and pUL99 but not pUL65 (131). The MCMV tegument protein homologs are predicted to be similar in length and molecular weight to the HCMV proteins, except for M32 and M99, both of which are around half the predicted size of their HCMV counterparts, pUL32 and ppUL99, respectively (131). Recent studies have shown the M99 (pp16) gene product is phosphorylated, is 15.2 or 16.3 kDa in size, is recognized by mouse hyperimmune serum, and migrates in gels with apparent molecular weights different from those predicted from the primary amino acid sequence (31). The MCMV homolog (M83) of the HCMV upper matrix protein is predicted to encode a protein of 809 residues and a molecular mass of 90.9 kDa, which is around 50% larger than HCMV ppUL83(pp65) (62). As no homolog of HCMV UL65 was found on primary sequence analysis, it will be of interest to investigate whether splicing occurs in the region between M57 and M69 (Table 1) to produce a UL65 homolog. MCMV M57 (MDBP) and M69 are ORFs that are homologous with HCMV transactivators of early genes, and they lie in an area of low G+C content (Fig. 1) that also contains the origin of lytic replication (92).

Extensive families of homologous genes are not a general feature of herpesviruses, but they have been described for the beta-herpesviruses HCMV, HHV-6, HHV-7 (107a), and now MCMV. HCMV encodes nine gene families, with the six largest clustered at each end of the genome (22). Examination of the MCMV sequence has revealed homologs of the HCMV gene families US22, UL25, and UL82 and the GCRs. Weak homology with a member of the RL11 family (m11) most likely results from compositional bias. No members of the HCMV gene families US1, US2, US6, and US12 were found in MCMV. The product of the HCMV US22 gene is a nuclear protein (ICP22) which is expressed at early and late times (102, 113). No other members of the US22 gene family were identified in HCMV by the MAb used to identify the US22(ICP22)

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

gene product (102). The US22 gene family member TRS1 acts as a transactivator (137), but no functions have been assigned to the other members of the family. The HHV-6 homologs of the US22 family (U25, U17, and U16) have been shown to act as transactivators in cotransfection assays (109). As the US22 gene family has not been identified in alpha- or gamma-herpesviruses, the function of these genes may relate to the biological characteristics of beta-herpesviruses, although it is unknown why HCMV, MCMV, and HHV-6 have evolved duplicated members. It is reasonable to assume that their function is important to beta-herpesvirus infection of the natural host, given the conservation of large numbers of the gene family in all three mammalian beta-herpesviruses.

The occurrence of two gene families unique to MCMV (m02 and m145) is particularly interesting, as they (plus the US22 gene family) occupy the region congruent with the short unique and long terminal repeat of HCMV. A single member (m17) of the righthand-end (m145) family occurs downstream of the cluster of the lefthand-end (m02) family (Fig. 3). It has been hypothesized that duplication produces the gene families and that this may account for the larger size of beta-herpesviruses (150). It seems likely that these gene families were produced by duplication and that a subsequent recombination event produced this movement of the m17 gene. This would appear more likely if recombination followed divergence of HCMV and MCMV, as the MHC class I homolog is located just before the first member of the righthand-end family in MCMV (m145) but at the opposite end of the genome in HCMV (UL18). Further studies of these two families will be of particular interest, as significant homology with other herpesviruses was found for only one member of the m145 family (m154 and EHV-1 gX), and this is most likely a result of compositional bias (141).

In light of the fact that the HCMV and MCMV MHC class I genes do not share strongly similar amino acid sequences, it is intriguing that the class I MHC gene is conserved in both beta-herpesviruses. The nature of the cellular gene product suggests that homologs in HCMV and MCMV may play a role in subverting the host's immune response during virus infection, particularly as both viruses have been shown to reduce synthesis of MHC class I proteins (19). It has been demonstrated that early gene expression results in a reduction in IE antigen presentation, with retention of the MHC class I-MCMV peptide complexes in the endoplasmic reticulum (41). Furthermore, recent evidence has shown that the reduction in MHC class I synthesis that follows MCMV infection is due at least in part to a gene product from within *Hin*dIII E (143). Unlike MCMV, the HCMV-induced MHC class I shutdown has been associated with an increased turnover of the MHC class I protein, rather than retention in the endoplasmic reticulum-*cis*-Golgi apparatus. The downregulation of class I expression in HCMV-infected cells has been associated (at least in part) with the product of HCMV US11 (63). Previous studies aimed at determining the function of the HCMV MHC class I homolog have shown that when expressed in vaccinia virus, the UL18 gene product can bind β-2m (15). However, in the context of HCMV infection, it was later demonstrated that the MHC class I homolog is not involved in sequestering cellular β-2m and is not directly responsible for the downregulation of MHC class I typically observed in infected cells (15). As the recombinant HCMV lacking the UL18 gene is indistinguishable from wild-type HCMV in cell culture, it is probable that it plays a role either in the interaction with the host's immune system or in the infection of particular cell types in vivo (14). This may be also true for the MCMV m144, the MHC class I homolog.

The MCMV M33 ORF is homologous with the GCRs, which are cell surface receptors that are coupled to G proteins and are activated by extracellular ligands or photons (13). The G proteins act by transducing signals from these surface receptors to a small number of effector molecules (such as adenylyl cyclase), which then act upon intracellular functions. Homologs of the eukaryotic GCRs are encoded by the beta-herpesviruses HCMV [UL33, US27, US28 [21], and possibly UL78 [51]), HHV-6 (109), HHV-7 (107a), and the gamma-herpesviruses HVS (108) and EHV-2 (140). The HCMV GCR homologs have been shown to be transcribed (148), and the M33 and the HCMV UL33 genes are spliced at the 5′ end to an upstream exon (35). The M33 protein most closely resembles HCMV UL33 and is located in a colinear position. The two cysteines that are highly conserved between members of the GCR family (and essential for the structure of bovine rhodopsin [65]) are conserved in M33. Two additional cysteines (located toward the carboxy terminus) are conserved between M33, UL33, US28, and the human RDC1-like GCRs (80). The HCMV GCR proteins are of unknown function, although US28 is capable of acting as a chemokine receptor in vitro (106). In vitro and in vivo studies of the MCMV GCR homolog will be of particular interest.

The provision of a complete genomic sequence of another beta-herpesvirus has allowed assessment of genes common to this group of herpesviruses. Study of the MCMV genes important in the pathogenesis of infection will provide further information about the role of these genes in infection of the host with this virus. Furthermore, the information gained from experiments with MCMV in the natural host should provide information about homologous genes in HCMV. As many of the MCMV genes with homologs in HCMV and HHV-6 are not found in the alpha- or gammaherpesviruses, MCMV should continue to provide useful information about the role of these genes in the betaherpesviruses.

ACKNOWLEDGMENTS

All authors contributed equally to this study, and we are indebted to a great many people. We thank James Bonfield, Mark Chee, Andrew Davison, Nicholas Davis-Poynter, Simon Dear, Duncan McGeoch, Tony Minson, Ed Mocarski, Brian Neal, Ilaria Portilicchio, Gillian Scott, and Rodger Staden for their excellent help during this project.

The financial support of the University of Cambridge Division of Virology, the Medical Research Council, and the Sanger Centre is gratefully acknowledged. W.D.R. was supported by a Sir Robert Menzies Memorial Scholarship and a University of Cambridge research associateship, and H.E.F. was supported by a Royal Society Florey Fellowship, MRC research fellowship, and AMRAD Corporation Ltd.

REFERENCES

1. **Addison, C., F. J. Rixon, J. W. Palfreyman, M. O'Hara, and V. G. Preston.** 1984. Characterisation of a herpes simplex virus type 1 mutant which has a temperature-sensitive defect in penetration of cells and assembly of capsids. Virology **138:**246–259.
2. **Albrecht, J. C., J. Nicholas, D. Biller, K. R. Cameron, B. Biesinger, C. Newman, S. Wittmann, M. A. Craxton, H. Coleman, B. Fleckenstein, and R. W. Honess.** 1992. Primary structure of the herpesvirus saimiri genome. J. Virol. **66:**5047–5058.
3. **Alderton, R. P., L. M. Eccleston, R. P. Howe, C. A. Read, M. A. Reeve, and S. Beck.** 1991. Magnetic bead purification of M13 DNA sequencing templates. Anal. Biochem. **201:**433–441.
4. **Anders, D. G.** 1990. Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding protein gene: comparison with alpha- and gamma-herpesvirus counterparts reveals conserved segments. J. Gen. Virol. **71:** 2451–2456.
5. **Anders, D. G., and W. Gibson.** 1988. Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1. J. Virol. **62:**1364–1372.
6. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J.**

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310:**207–211.
7. **Bankier, A. T., and B. G. Barrell.** 1989. Sequencing single strand DNA using the chain-termination method, p. 37–78. *In* C. J. Howe and E. S. Ward (ed.), Nucleic acids sequencing: a practical approach. IRL Press, Oxford.
8. **Bankier, A. T., K. M. Weston, and B. G. Barrell.** 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. Methods Enzymol. **155:**51–93.
9. **Baum, E. Z., G. A. Bebernitz, J. D. Hulmes, V. P. Muzithras, T. R. Jones, and Y. Gluzman.** 1993. Expression and analysis of the human cytomegalovirus UL80-encoded protease: identification of autoproteolytic sites. J. Virol. **67:**497–506.
10. **Beck, S., and B. G. Barrell.** 1988. Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens. Nature (London) **331:**269–272.
11. **Berg, J. M.** 1988. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. USA **85:**99–102.
12. **Bogner, E., M. Reschke, B. Reis, T. Mockenhaupt, and K. Radsak.** 1993. Identification of the gene product encoded by ORF UL56 of the human cytomegalovirus genome. Virology **196:**290–293.
13. **Bourne, H. R., D. A. Sanders, and F. McCormick.** 1990. The GTPase superfamily: a conserved switch for diverse cell functions. Nature (London) **348:**125–132.
14. **Browne, H., M. Churcher, and T. Minson.** 1992. Construction and characterization of a human cytomegalovirus mutant with the UL18 (class I homolog) gene deleted. J. Virol. **66:**6784–6787.
15. **Browne, H., G. Smith, S. Beck, and A. C. Minson.** 1990. A complex between the MHC class I homologue encoded by human cytomegalovirus and beta 2 microglobulin. Nature (London) **347:**770–772.
16. **Buckmaster, A. E., S. D. Scott, M. J. Sanderson, M. E. Boursnell, N. L. Ross, and M. M. Binns.** 1988. Gene sequence and mapping data from Marek's disease virus and herpesvirus of turkeys: implications for herpesvirus classification. J. Gen. Virol. **69:**2033–2042.
17. **Buhler, B., G. M. Keil, F. Weiland, and U. H. Koszinowski.** 1990. Characterization of the murine cytomegalovirus early transcription unit e1 that is induced by immediate early proteins. J. Virol. **64:**1907–1919.
18. **Burck, P. J., D. H. Berg, T. P. Luk, L. M. Sassmannshausen, M. Wakulchik, D. P. Smith, H. M. Hsiung, G. W. Becker, W. Gibson, and E. C. Villarreal.** 1994. Human cytomegalovirus maturational proteinase: expression in *Escherichia coli*, purification, and enzymatic characterization by using peptide substrate mimics of natural cleavage sites. J. Virol. **68:**2937–2946.
19. **Campbell, A. E., and J. S. Slater.** 1994. Down-regulation of major histocompatibility complex class I synthesis by murine cytomegalovirus early gene expression. J. Virol. **68:**1805–1811.
20. **Cha, T. A., E. Tom, G. W. Kemble, G. M. Duke, E. S. Mocarski, and R. R. Spaete.** 1995. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J. Virol. **70:**78–83.
21. **Chee, M., S. Satchwell, E. Preddie, K. Weston, and B. G. Barrell.** 1990. Human cytomegalovirus encodes three G protein-coupled receptor homologues. Nature (London) **344:**774–777.
22. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison, T. Kouzarides, J. A. Martignetti, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell.** 1990. Analysis of the protein coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**125–169.
23. **Chee, M. S., S. A. Rudolph, B. Plachter, B. G. Barrell, and G. Jahn.** 1989. Identification of the major capsid protein gene of human cytomegalovirus. J. Virol. **63:**1345–1353.
24. **Ciufo, D. M., M. A. Mullen, and G. S. Hayward.** 1994. Identification of a dimerization domain in the C-terminal segment of the IE110 transactivator protein from herpes simplex virus. J. Virol. **68:**3267–3282.
25. **Clarke, J. R., and A. S. Tyms.** 1991. Polyamine biosynthesis in cells infected with different clinical isolates of human cytomegalovirus. J. Med. Virol. **34:**212–216.
26. **Colberg-Poley, A. M., L. D. Santomenna, P. P. Harlow, P. A. Benfield, and D. J. Tenney.** 1992. Human cytomegalovirus US3 and UL 36-38 immediate early proteins regulate gene expression. J. Virol. **66:**95–105.
27. **Costa, R. H., K. G. Draper, T. J. Kelly, and E. K. Wagner.** 1985. An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. J. Virol. **54:**317–328.
28. **Craighead, J. E., W. B. Martin, and S. A. Huber.** 1992. Role of CD4+ (helper) T cells in the pathogenesis of murine cytomegalovirus myocarditis. Lab. Invest. **66:**755–761.
29. **Cranage, M. P., T. Kouzarides, A. T. Bankier, S. C. Satchwell, K. W. Weston, P. Tomlinson, B. G. Barrell, H. Hart, A. C. Minson, and G. L. Smith.** 1986. Identification of the human cytomegalovirus glycoprotein B gene and induction of neutralizing antibodies via its expression in recombinant vaccinia virus. EMBO J. **5:**3057–3063.
30. **Cranage, M. P., G. L. Smith, S. E. Bell, H. Hart, C. M. Brown, A. T. Bankier, P. Tomlinson, B. G. Barrell, and A. C. Minson.** 1988. Identification and expression of a human cytomegalovirus glycoprotein with homology to the Epstein-Barr virus BXLF2 product, varicella zoster virus gp III and herpes simplex virus type I glycoprotein H. J. Virol. **62:**1416–1422.
31. **Cranmer, L. D., C. Clark, and D. H. Spector.** 1994. Cloning, characterization, and expression of the murine cytomegalovirus homologue of the human cytomegalovirus 28-kDa matrix phosphoprotein (UL99). Virology **205:**417–429.
32. **Craxton, M.** 1991. Linear amplification sequencing, a powerful method for sequencing DNA, p. 20–26. Methods: a companion to methods in enzymology, vol. 3. **3:**20–26.
33. **Crute, J. J., T. Tsurumi, L. Zhu, S. K. Weller, P. D. Olivo, M. D. Challberg, E. S. Mocarski, and I. R. Lehman.** 1989. Herpes simplex virus 1 helicase-primase: a complex of three herpes-encoded gene products. Proc. Natl. Acad. Sci. USA **86:**2186–2189.
34. **Dallas, P. B., P. A. Lyons, J. B. Hudson, A. A. Scalzo, and G. R. Shellam.** 1994. Identification and characterization of a murine cytomegalovirus gene with homology to the UL25 open reading frame of human cytomegalovirus. Virology **200:**643–650.
35. **Davis-Poynter, N., W. D. Rawlinson, B. G. Barrell, G. R. Shellam, and H. E. Farrell.** 1995. Identification and characterisation of a G protein-coupled receptor homologue encoded by murine cytomegalovirus, p. 88. Abstracts of the International Herpes Virus Workshop. Groningen, The Netherlands.
36. **Davison, A. J.** 1991. Experience in shotgun sequencing a 134 kilobase pair DNA molecule. DNA Seq. **1:**389–394.
37. **Davison, A. J.** 1992. Channel catfish virus: a new type of herpes virus. Virology **186:**9–14.
38. **Davison, A. J., and J. E. Scott.** 1986. The complete DNA sequence of varicella-zoster virus. J. Gen. Virol. **67:**1759–1816.
39. **Dear, S., and R. Staden.** 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. **19:**3907–3911.
40. **Dear, S., and R. Staden.** 1992. A standard file format for data from DNA sequencing instruments. DNA Seq. **3:**107–110.
41. **Del Val, M., K. Munch, M. J. Reddehase, and U. H. Koszinowski.** 1989. Presentation of CMV immediate-early antigen to cytolytic T lymphocytes is selectively prevented by viral genes expressed in the early phase. Cell **58:**305–315.
42. **Dorsch-Hasler, K., G. M. Keil, F. Weber, M. Jasin, W. Schaffner, and U. H. Koszinowski.** 1985. A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. Proc. Natl. Acad. Sci. USA **82:**8325–8329.
42a. **Durbin, R.** Unpublished data.
42b. **Durbin, R., and J. T. Mieg.** Unpublished data.
43. **Ebeling, A., G. M. Keil, E. Knust, and U. H. Koszinowski.** 1983. Molecular cloning and physical mapping of murine cytomegalovirus DNA. J. Virol. **47:**421–433.
44. **Elliott, R., C. Clark, D. Jaquish, and D. H. Spector.** 1991. Transcription analysis and sequence of the putative murine cytomegalovirus DNA polymerase gene. Virology **185:**169–186.
45. **Ertl, P. F., and K. L. Powell.** 1992. Physical and functional interaction of human cytomegalovirus DNA polymerase and its accessory protein (ICP36) expressed in insect cells. J. Virol. **66:**4126–4133.
46. **Everett, R. D.** 1984. Trans-activation of transcription by herpes virus products: requirement for two HSV-1 immediate-early polypeptides for maximum activity. EMBO J. **3:**3135–3141.
47. **Everett, R. D.** 1989. Construction and characterization of herpes simplex virus type 1 mutants with defined lesions in immediate early gene 1. J. Gen. Virol. **70:**1185–1202.
48. **Faisst, S., and S. Meyer.** 1992. Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res. **20:**3–26.
49. **Foung, S. K. H., S. Perkins, P. Bradshaw, J. Rowe, L. B. Rabin, G. R. Reyes, and E. T. Lennette.** 1989. Human monoclonal antibodies to human cytomegalovirus induces *in vitro* immune responses similar to those of whole viral antigen. J. Infect. Dis. **159:**436–443.
50. **Gibson, W., B. R. Van, A. Fields, R. LaFemina, and A. Irmiere.** 1984. D,L-α-Difluoromethylornithine inhibits human cytomegalovirus replication. J. Virol. **50:**145–154.
51. **Gompels, U. A., J. Nicholas, G. Lawrence, M. Jones, B. J. Thomson, M. E. Martin, S. Efstathiou, M. Craxton, and H. A. Macaulay.** 1995. The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. Virology **209:**29–51.
52. **Greaves, R. F., J. M. Brown, J. Vieira, and E. S. Mocarski.** 1995. Selectable insertion and deletion mutagenesis of the human cytomegalovirus genome using the *E. coli* guanosine phosphoribosyl transferase (gpt) gene. J. Gen. Virol. **76:**2151–2160.
53. **He, Y. S., L. Xu, and E. S. Huang.** 1992. Characterization of human cytomegalovirus UL84 early gene and identification of its putative protein product. J. Virol. **66:**1098–1108.
54. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. Clustal V: improved software for multiple sequence alignment. CABIOS **8:**189–191.
55. **Honess, R. W.** 1984. Herpes simplex and the 'herpes complex': diverse observations and a unifying hypothesis. J. Gen. Virol. **65:**2077–2107.
56. **Honess, R. W., U. A. Gompels, B. G. Barrell, M. Craxton, K. R. Cameron, R. Staden, Y. N. Chang, and G. S. Hayward.** 1989. Deviations from ex-

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

pected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. J. Gen. Virol. **70:**837–855.

56a. **Horsnell, T., J. Crooke, M. A. Rajeandram, and B. G. Barrell.** Unpublished data.

57. **Hudson, J. B.** 1979. The murine cytomegalovirus as a model for the study of viral pathogenesis and persistent infections. Arch. Virol. **62:**1–29.

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

59. **Hunkapillar, T., R. J. Kaiser, B. F. Koop, and L. Hood.** 1991. Large scale and automated DNA sequence determination. Science **254:**59–67.

60. **Hutchison, C. A., S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell.** 1989. LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, p. 593–617. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.

61. **Jahn, G., T. Kouzarides, M. Mach, B.-C. Scholl, B. Plachter, B. Traupe, E. Preddie, S. C. Satchwell, B. Fleckenstein, and B. G. Barrell.** 1987. Map position and nucleotide sequence of the gene for the large structural phosphoprotein of human cytomegalovirus. J. Virol. **61:**1358–1367.

62. **Jahn, G., B.-C. Scholl, B. Traupe, and B. Fleckenstein.** 1987. The two major structural phosphoproteins (pp65 and pp150) of human cytomegalovirus and their antigenic properties. J. Gen. Virol. **68:**1327–1337.

63. **Jones, T. R., L. K. Hanson, L. Sun, J. S. Slater, R. M. Stenberg, and A. E. Campbell.** 1995. Multiple independent loci within the human cytomegalovirus unique short region down-regulate expression of major histocompatibility complex class I heavy chains. J. Virol. **69:**4831–4841.

64. **Karlin, S., and S. F. Altschul.** 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA **87:**2264–2268.

65. **Karnik, S. S., T. P. Sakmar, H.-B. Chen, and H. G. Khorana.** 1988. Cysteine residues 110 and 187 are essential for the formation of correct structure in bovine rhodopsin. Proc. Natl. Acad. Sci. USA **85:**8459–8463.

66. **Kashiwai, A., N. Kawamura, C. Kadota, and Y. Tsutsui.** 1992. Susceptibility of mouse embryo to murine cytomegalovirus infection in early and midgestation stages. Arch. Virol. **127:**37–48.

67. **Kaye, J. F., U. A. Gompels, and A. C. Minson.** 1992. Glycoprotein H of human cytomegalovirus (HCMV) forms a stable complex with the HCMV UL115 gene product. J. Gen. Virol. **73:**2693–2698.

68. **Keil, G. M., A. Ebeling-Keil, and U. Koszinowski.** 1987. Immediate-early genes of murine cytomegalovirus: location, transcripts, and translation products. J. Virol. **61:**526–533.

69. **Keil, G. M., A. Ebeling-Keil, and U. H. Koszinowski.** 1984. Temporal regulation of murine cytomegalovirus transcription and mapping of viral RNA synthesized at immediate early times after infection. J. Virol. **50:**784–795.

70. **Keil, G. M., A. Ebeling-Keil, and U. H. Koszinowski.** 1987. Sequence and structural organization of murine cytomegalovirus immediate-early gene 1. J. Virol. **61:**1901–1908.

71. **Kollert-Jons, A., E. Bogner, and K. Radsak.** 1991. A 15 kilobase pair region of the human cytomegalovirus genome which includes US1 through US13 is dispensable for growth in cell culture. J. Virol. **65:**5184–5189.

72. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, E. Preddie, and B. G. Barrell.** 1988. An immediate early gene of human cytomegalovirus encodes a potential membrane glycoprotein. Virology **165:**151–164.

73. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell.** 1987. Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene. J. Virol. **61:**125–133.

74. **Lagenaur, L. A., W. C. Manning, J. Viera, C. L. Martens, and E. S. Mocarski.** 1994. Structure and function of the murine cytomegalovirus sgg1 gene: a determinant of viral growth in salivary gland acinar cells. J. Virol. **68:**7717–7727.

75. **Lawrence, G. L., M. Chee, M. A. Craxton, U. A. Gompels, R. W. Honess, and B. G. Barrell.** 1990. Human herpesvirus-6 is closely related to human cytomegalovirus. J. Virol. **64:**287–299.

76. **Leach, F. S., and E. S. Mocarski.** 1989. Regulation of cytomegalovirus late-gene expression: differential use of three start sites in the transcriptional activation of ICP36 gene expression. J. Virol. **63:**1783–1791.

77. **Leatham, M. P., P. R. Witte, and M. F. Stinski.** 1991. Alternate promoter selection within a human cytomegalovirus immediate early and early transcription unit (UL119-115) defines true late transcripts containing open reading frames for putative viral glycoproteins. J. Virol. **65:**6144–6153.

78. **Lehner, R., H. Meyer, and M. Mach.** 1989. Identification and characterization of a human cytomegalovirus gene coding for a membrane protein that is conserved among human herpesviruses. J. Virol. **63:**3792–3800.

79. **Li, W., K. Eidman, R. C. Gehrz, and B. Kari.** 1995. Identification and molecular characterization of the murine cytomegalovirus homolog of the human cytomegalovirus UL100 gene. Virus Res. **36:**163–175.

80. **Libert, F., M. Parmentier, A. Lefort, J. E. Dumont, and G. Vassart.** 1990. Complete nucleotide sequence of a putative G protein-coupled receptor: RDC1. Nucleic Acids Res. **18:**1917.

81. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.

82. **Littler, E., A. D. Stuart, and M. S. Chee.** 1992. Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir. Nature (London) **358:**160–162.

83. **Liu, F., and B. Roizman.** 1991. The herpes virus simplex virus 1 gene encoding a protease also contains within its coding domain the gene encoding the more abundant substrate. J. Virol. **65:**5149–5156.

84. **Loutsch, J. M., N. J. Galvin, M. L. Bryant, and B. C. Holwerda.** 1994. Cloning and sequence analysis of murine cytomegalovirus protease and capsid assembly protein genes. Biochem. Biophys. Res. Commun. **203:**472–478.

85. **Lyons, P. A., P. B. Dallas, C. Carrello, G. R. Shellam, and A. A. Scalzo.** 1994. Mapping and transcriptional analysis of the murine cytomegalovirus homologue of the human cytomegalovirus UL103 open reading frame. Virology **204:**835–839.

86. **Manning, W. C., and E. S. Mocarski.** 1988. Insertional mutagenesis of the murine cytomegalovirus genome: one prominent alpha gene (ie2) is dispensable for growth. Virology **167:**477–484.

87. **Manning, W. C., C. A. Stoddart, L. A. Lagenaur, G. B. Abenes, and E. S. Mocarski.** 1992. Cytomegalovirus determinant of replication in salivary glands. J. Virol. **66:**3794–3802.

88. **Marks, J. R., J. A. Mercer, and D. H. Spector.** 1983. Transcription in mouse embryo cells permissively infected by murine cytomegalovirus. Virology **131:**247–254.

89. **Marks, J. R., and D. H. Spector.** 1984. Fusion of the termini of the murine cytomegalovirus genome after infection. J. Virol. **52:**24–28.

90. **Martignetti, J. A., and B. G. Barrell.** 1991. Sequence of the *Hin*dIII T fragment of human cytomegalovirus, which encodes a DNA helicase. J. Gen. Virol. **72:**1113–1121.

91. **Masse, M. J., S. Karlin, G. A. Schachtel, and E. S. Mocarski.** 1992. Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region. Proc. Natl. Acad. Sci. USA **89:**5246–5250.

92. **Masse, M. J., M. Messerle, and E. S. Mocarski.** 1992. Structure of the murine cytomegalovirus (CMV) DNA lytic origin replication (oriLyt). Relation to other CMV oriLyt, p. 9. *In* J. Subak-Sharpe (ed.), Abstracts of the 17th International Herpesvirus Workshop, Edinburgh.

93. **McGeoch, D. J.** 1985. On the predictive recognition of signal peptide sequences. Virus Res. **3:**271–286.

94. **McGeoch, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, J. E. Scott, and P. Taylor.** 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69:**1531–1574.

95. **McGeoch, D. J., and A. J. Davison.** 1986. Alphaherpesviruses possess a gene homologous to the protein kinase family of eukaryotes and retroviruses. Nucleic Acids Res. **14:**1765–1777.

96. **McGeoch, D. J., A. Dolan, and M. C. Frame.** 1986. DNA sequence of the region in the genome of herpes simplex virus type 1 containing the exonuclease gene and neighbouring genes. Nucleic Acids Res. **14:**3435–3448.

97. **Messerle, M., B. Buhler, G. M. Keil, and U. H. Koszinowski.** 1992. Structural organization, expression and functional characterization of the murine cytomegalovirus immediate-early gene 3. J. Virol. **66:**27–36.

98. **Messerle, M., G. M. Keil, and U. H. Koszinowski.** 1991. Structure and expression of murine cytomegalovirus immediate-early gene 2. J. Virol. **65:**1638–1643.

99. **Messerle, M., G. M. Keil, K. Schneider, and U. H. Koszinowski.** 1992. Characterization of the murine cytomegalovirus genes encoding the major DNA binding protein and the ICP18.5 homolog. Virology **191:**355–367.

100. **Meyer, H., A. T. Bankier, M. P. Landini, C. M. Brown, B. G. Barrell, B. Ruger, and M. Mach.** 1988. Identification and procaryotic expression of the gene coding for the highly immunogenic 28-kilodalton structural phosphoprotein (pp28) of human cytomegalovirus. J. Virol. **62:**2243–2250.

101. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371–378.

102. **Mocarski, E. S., L. Pereira, and A. L. McCormick.** 1988. Human cytomegalovirus ICP22, the product of the HWLF1 reading frame, is an early nuclear protein that is released from cells. J. Gen. Virol. **69:**2613–2621.

103. **Mosmann, T. R., and J. B. Hudson.** 1973. Some properties of the genome of murine cytomegalovirus (MCV). Virology **54:**135–149.

104. **Mutter, W., M. J. Reddehase, F. W. Busch, H. J. Buhring, and U. H. Koszinowski.** 1988. Failure in generating hemopoietic stem cells is the primary cause of death in the immunocompromised host. J. Exp. Med. **167:**1645–1658.

105. **Nelson, D. R., J. E. Lawson, M. Klingenberg, and M. G. Douglas.** 1993. Site-directed mutagenesis of the yeast mitochondrial ADP/ATP translocator. J. Mol. Biol. **230:**1159–1170.

106. **Neote, K., D. DiGregorio, J. Y. Mak, R. Horuk, and T. J. Schall.** 1993. Molecular cloning, functional expression, and signaling characteristics of a C-C chemokine receptor. Cell **72:**415–425.

107. **Neyts, J., J. Balzarini, L. Naesens, and C. E. De.** 1992. Efficacy of (S)-1-(3-hydroxy-2-phosphonylmethoxypropyl)cytosine and 9-(1,3-dihydroxy-2-propoxymethyl) guanine for the treatment of murine cytomegalovirus infection in severe combined immunodeficiency mice. J. Med. Virol. **37:**67–71.

107a. **Nicholas, J.** 1996. Determination and analysis of the complete nucleotide

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest

sequence of human herpesvirus 7. J. Virol. **70:**5975–5989.

108. **Nicholas, J., K. R. Cameron, and R. W. Honess.** 1992. Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cyclins. Nature (London) **355:**362–365.
109. **Nicholas, J., and M. E. Martin.** 1994. Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. J. Virol. **68:**597–610.
110. **Ochiai, H., K. Kumura, and Y. Minamishima.** 1992. Murine cytomegalovirus DNA polymerase: purification, characterization and role in the antiviral activity of acyclovir. Antiviral Res. **17:**1–16.
111. **Osborn, J. E.** 1986. Cytomegalovirus and other herpesviruses of mice and rats, p. 420–451. *In* P. N. Bhatt, R. O. Jacoby, H. C. Morse, and A. E. New (ed.), Viral and mycoplasmal infections of laboratory rodents. Academic Press, London.
112. **Pari, G. S.** 1993. Eleven loci encoding trans-acting factors are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA replication. J. Virol. **67:**6979–6988.

112a. **Parsons, J.** Unpublished data.

113. **Pereira, L., and M. Hoffman.** 1986. Immunology of human cytomegalovirus proteins, p. 69–92. *In* C. Lopez and B. Roizman (ed.), Human herpesvirus infections. Raven Press, New York.
114. **Pertuiset, B., M. Boccara, J. Cerbrian, N. Berthelot, S. Chousterman, F. Puvion-Dutilleul, J. Sisman, and P. Sheldrick.** 1989. Physical mapping and nucleotide sequence of a herpes simplex virus type 1 gene required for capsid assembly. J. Virol. **63:**2169–2179.
115. **Pomeroy, C., P. J. Hilleren, and M. C. Jordan.** 1991. Latent murine cytomegalovirus DNA in splenic stromal cells of mice. J. Virol. **65:**3330–3334.
116. **Preston, V. G., and F. B. Fisher.** 1984. Identification of the herpes simplex virus type 1 gene encoding the dUTPase. Virology **138:**58–68.
117. **Preston, V. G., J. W. Palfreyman, and B. M. Dutia.** 1984. Identification of a herpes simplex virus type 1 polypeptide which is a component of the virus-induced ribonucleotide reductase. J. Gen. Virol. **65:**1457–1466.
118. **Pyles, R. B., N. M. Sawtell, and R. L. Thompson.** 1992. Herpes simplex virus type 1 dUTPase mutants are attenuated for neurovirulence, neuroinvasiveness, and reactivation from latency. J. Virol. **66:**6706–6713.
119. **Rapp, M., M. Messerle, B. Buhler, M. Tannheimer, G. M. Keil, and U. H. Koszinowski.** 1992. Identification of the murine cytomegalovirus glycoprotein B gene and its expression by recombinant vaccinia virus. J. Virol. **66:** 4399–4406.
120. **Rawlinson, W. D., and B. G. Barrell.** 1993. Spliced transcripts of human cytomegalovirus (HCMV). J. Virol. **67:**5502–5513.
121. **Reddehase, M. J., M. Balthesen, M. Rapp, S. Jonjic, I. Pavic, and U. Koszinowski.** 1994. The conditions of primary infection define the load of latent viral genome in organs and the risk of recurrent cytomegalovirus disease. J. Exp. Med. **179:**185–193.
122. **Robson, L., and W. Gibson.** 1989. Primate cytomegalovirus assembly protein: genome location and nucleotide sequence. J. Virol. **63:**669–676.
123. **Ruger, B., S. Klages, B. Walla, J. Albrecht, B. Fleckenstein, P. Tomlinson, and B. G. Barrell.** 1987. Primary structure and transcription of the genes coding for the two virion phosphoproteins pp65 and pp71 of human cytomegalovirus. J. Virol. **61:**446–453.

123a. **Salter, R. D., R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, and P. Parham.** 1990. A binding site for the T-cell co-receptor CD8 on the alpha 3 domain of HLA-A2. Nature (London) **345:**41–46.

124. **Sandford, G. R., J. R. Wingard, J. W. Simons, S. P. Staal, R. Saral, and W. H. Burns.** 1985. Genetic analysis of the susceptibility of mouse cytomegalovirus to acyclovir. J. Virol. **54:**104–113.
125. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
126. **Scalzo, A. A., C. A. Forbes, N. J. Davis-Poynter, H. E. Farrell, and P. E. Lyons.** 1995. DNA sequence and transcriptional analysis of the glycoprotein M gene of murine cytomegalovirus. J. Gen. Virol. **76:**2895–2901.
127. **Shanley, J. D., J. Morningstar, and M. C. Jordan.** 1985. Inhibition of murine cytomegalovirus lung infection and interstitial pneumonitis by acyclovir and 9-(1,3-dihydroxy-2-propoxymethyl)guanine. Antimicrob. Agents Chemother. **28:**172–175.
128. **Shellam, G. R., J. P. Flexman, H. E. Farrell, and J. M. Papadimitriou.** 1985. The genetic background modulates the effect of the beige gene on susceptibility to cytomegalovirus infection in mice. Scand. J. Immunol. **22:**147–155.
129. **Sherman, G., J. Gottlieb, and M. Challberg.** 1992. The UL8 subunit of the herpes simplex virus helicase-primase complex is required for efficient primer utilization. J. Virol. **66:**4884–4892.
130. **Smith, V., M. Craxton, A. T. Bankier, C. M. Brown, W. D. Rawlinson, M. S. Chee, and B. G. Barrell.** 1993. Microtitre methods for preparation and fluorescent sequencing of M13 clones. Methods Enzymol. **218:**173–187.
131. **Spaete, R. R., R. C. Gehrz, and M. P. Landini.** 1994. Human cytomegalovirus structural proteins. J. Gen. Virol. **75:**3287–3308.
132. **Spector, D. J., and M. J. Tevethia.** 1994. Protein-protein interactions between human cytomegalovirus IE2-580aa and pUL84 in lytically infected cells. J. Virol. **68:**7549–7553.
133. **Staden, R.** 1984. Graphic methods to determine the function of nucleic acid sequences. Nucleic Acids Res. **12:**521–538.
134. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14:**217–231.
135. **Stamminger, T., and B. Fleckenstein.** 1990. Immediate-early transcription regulation of human cytomegalovirus. Curr. Top. Microbiol. Immunol. **154:** 3–19.
136. **Staprans, S. I., and D. H. Spector.** 1988. 2.2 kilobase class of early transcripts encoded by cell related sequences in human cytomegalovirus strain AD169. J. Virol. **57:**591–602.
137. **Stasiak, P. C., and E. S. Mocarski.** 1992. Transactivation of the cytomegalovirus ICP36 gene promoter requires the alpha gene product TRS1 in addition to IE1 and IE2. J. Virol. **66:**1050–1058.
138. **Stenberg, R. M., A. S. Depto, J. Fortney, and J. Nelson.** 1989. Regulated expression of early and late RNAs and proteins from the human cytomegalovirus immediate early gene region. J. Virol. **63:**2699–2708.
139. **Sullivan, V., C. L. Talarico, S. C. Stanat, M. Davis, D. M. Coen, and K. K. Biron.** 1992. A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus infected cells. Nature (London) **358:**162–164.
140. **Telford, E. A. R., M. S. Watson, H. C. Aird, J. Perry, and A. J. Davison.** 1995. The DNA sequence of equine herpesvirus-2. J. Mol. Biol. **249:**520–528.
141. **Telford, E. A. R., M. S. Watson, K. McBride, and A. J. Davison.** 1992. The DNA sequence of equine herpesvirus-1. Virology **189:**304–316.
142. **Thale, R., P. Lucin, K. Schneider, M. Eggers, and U. H. Koszinowski.** 1994. Identification and expression of a murine cytomegalovirus early gene coding for an Fc receptor. J. Virol. **68:**7757–7765.
143. **Thale, R., U. Szepan, H. Hengel, G. Geginat, H. E. Farrell, W. D. Rawlinson, and U. Koszinowski.** 1995. A mouse cytomegalovirus glycoprotein blocks the transport of nascent MHC class I molecules to the plasma membrane, p. 218. Abstracts of the International Herpes Virus Workshop, Groningen, The Netherlands.
144. **Tsutsui, Y., A. Kashiwai, N. Kawamura, and C. Kadota.** 1993. Microphthalmia and cerebral atrophy induced in mouse embryos by infection with murine cytomegalovirus in midgestation. Am. J. Pathol. **143:**804–813.
145. **Vaughan, P. J., K. J. Thibault, M. A. Hardwicke, and G. R. Sandri.** 1992. The herpes simplex virus immediate early protein ICP27 encodes a potential metal binding domain and binds zinc in vitro. Virology **189:**377–384.
146. **Vieira, J., H. E. Farrell, W. D. Rawlinson, and E. S. Mocarski.** 1994. Genes in the *Hin*dIII fragment of the murine cytomegalovirus genome are dispensable for growth in cultured cell: insertion mutagenesis with a *lac*Zlgpt cassette. J. Virol. **68:**4837–4846.
147. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55:**631–661.
148. **Welch, A. R., L. M. McGregor, and W. Gibson.** 1991. Cytomegalovirus homologs of cellular G protein-coupled receptor genes are transcribed. J. Virol. **65:**3915–3918.
149. **Weston, K.** 1988. An enhancer element in the short unique region of human cytomegalovirus regulates the production of a group of abundant immediate early transcripts. Virology **162:**406–416.
150. **Weston, K., and B. G. Barrell.** 1986. Sequence of the short unique region, short repeats and part of the long repeat of human cytomegalovirus. J. Mol. Biol. **192:**177–208.
151. **Winkler, M., S. A. Rice, and T. Stamminger.** 1994. UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. J. Virol. **68:** 3943–3954.
152. **Worrad, D. M., and S. Caradonna.** 1988. Identification of the coding sequence for herpes simplex virus uracil-DNA glycosylase. J. Virol. **62:**4774–4777.
153. **Wright, D. A., S. I. Staprans, and D. H. Spector.** 1988. Four phosphoproteins with common amino termini are encoded by human cytomegalovirus AD169. J. Virol. **62:**331–340.
154. **Xu, J., P. B. Dallas, P. A. Lyons, G. R. Shellam, and A. A. Scalzo.** 1992. Identification of the glycoprotein H (gH) gene of murine cytomegalovirus. J. Gen. Virol. **73:**1849–1854.
155. **Xu, J., A. A. Scalzo, P. A. Lyons, H. E. Farrell, W. D. Rawlinson, and G. R. Shellam.** 1994. Identification, sequencing and expression of the glycoprotein L (gL) gene of murine cytomegalovirus. J. Gen. Virol. **75:**3235–3240.
156. **Yoakum, G. H.** 1993. Mapping a putative pyruvoyl decarboxylase active site to human cytomegalovirus open reading frame UL77. Biochem. Biophys. Res. Commun. **194:**1207–1215.
157. **Zhu, L., and S. K. Weller.** 1992. The six conserved helicase motifs of the UL5 gene product, a component of the herpes simplex virus type 1 helicase-primase are essential for its function. J. Virol. **66:**469–479.

Downloaded from http://jvi.asm.org/ on November 13, 2019 by guest