

NIH Public Access

Author Manuscript

Virology. Author manuscript; available in PMC 2009 October 23.

Published in final edited form as:

Virology. 2008 March 30; 373(1): 181–188. doi:10.1016/j.virol.2007.10.040.

Protein Coding Content of the U_Lb' Region of Wild-Type Rhesus Cytomegalovirus

Kristie L. Oxford, Meghan K. Eberhardt, Kai-Wen Yang, Lisa Strelow, Suzanne Kelly, Shan-Shan Zhou, and Peter A. Barry *

Center for Comparative Medicine; Department of Pathology & Laboratory Medicine; California National Primate Research Center; University of California, Davis; Davis, CA 95616, USA

Abstract

A recent comparison of two rhesus cytomegalovirus (RhCMV) genomes revealed that the region at the right end of the U_L genome component (U_Lb') undergoes genetic alterations similar to those observed in serially passaged human cytomegalovirus (HCMV). To determine the coding content of authentic wild-type RhCMV in this region, the U_Lb' sequence was amplified from virus obtained from naturally infected rhesus macaques without passage *in vitro*. A total of 24 open reading frames (ORFs) potentially encoding >99 amino acid residues were identified, 10 of which are related to HCMV ORFs and 15 to previously listed RhCMV ORFs. In addition, the analysis revealed a cluster of three novel alpha chemokine-like ORFs, bringing the number of predicted alpha chemokine genes in this region to six. Three of these six genes exhibit a high level of sequence diversity, as has been observed for the HCMV alpha chemokine gene UL146.

Keywords

rhesus cytomegalovirus; alpha chemokine; CXC chemokine; ULb'; sequence diversity; human cytomegalovirus; UL146

Rhesus cytomegalovirus (RhCMV) and human cytomegalovirus (HCMV) are ubiquitous throughout their host populations (rhesus macaques and humans, respectively) and cause subclinical, persistent, lifelong infections in healthy individuals (Barry & Chang, 2007; Britt, 2007). In immune compromised individuals, however, CMVs can cause significant morbidity and mortality. RhCMV and HCMV genomes are largely colinear, although species-specific genes are present in each. The region at the right end of the U_L genome component (U_Lb') in HCMV contains multiple open reading frames (ORFs) encoding cell tropism and immune modulating functions that may be mutated or deleted during passage of clinical isolates *in vitro* (Cha et al., 1996; Dolan et al., 2004). Recent work has demonstrated that the corresponding region in RhCMV undergoes similar changes, thus the genetic content of U_Lb' in wild-type virus is presently undefined (Hansen et al., 2003; Rivailler et al., 2006).

^{© 2009} Published by Elsevier Inc.

^{*}Corresponding Author: Center for Comparative Medicine, University of California, Davis, County Rd. 98 & Hutchison Dr., Davis, CA 95616. pabarry@ucdavis.edu, 530-752-6912.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The prototypical RhCMV strain 68-1 (Asher et al., 1974) has been annotated as containing 230 ORFs of >99 amino acid residues (aa) (Hansen et al., 2003) and the 180.92 strain as containing 258 ORFs ranging in size from 21 to 2178 aa (Rivailler et al., 2006). The analytical approaches taken were not the same in both studies, and the number and locations of genuine proteincoding ORFs remains unclear. There are also differences between the two strains, since the genome of 180.92 is approximately 5 kbp smaller than that of 68-1. The majority of the potential coding differences are located in a 15 kbp region that corresponds to $U_{\rm L}b'$ in HCMV, and it appears that, as in HCMV, RhCMV U₁ b' is labile during passage *in vitro*. Comparisons of 68-1 and 180.92 have led to a model in which the sequence of wild-type RhCMV was proposed to contain four contiguous genomic segments designated A, B, C, and D (Fig. 1) (Rivailler et al., 2006). The sequence of 180.92 lacks segments C and D, while that of 68-1 lacks segment A and segments B and C are inverted relative to segment D. In order to characterize the coding capacity of U_L b' in wild-type RhCMV that has not been passaged in vitro and test this proposed model, DNA from RhCMV naturally circulating at the California National Primate Research Center (CNPRC) (RhCMV_{CNPRC}) was purified from seropositive animals and the U_Lb' sequence of RhCMV was determined.

Results

A 13066 bp sequence (46% G+C) of RhCMV_{CNPRC} spanning the region of non-colinearity between the 68-1 and 180.92 genomes was assembled from multiple overlapping amplicons (GenBank accession number **EF990255**). The sequence included 714 and 470 bp of overlap with both genomes at its 5' and 3' ends, respectively (Fig. 1). The 5' end of the sequence begins upstream of UL128, and the 3' end of the sequence maps downstream of the rh167 ORF. The 5' end of the sequence contains a 22 bp sequence (CCGTCTCTCAGACCAAATTTAC) directly repeated 30 times, with 27 copies identical in sequence. This sequence in the overlapping region of the 68-1 genome (CCGTCTCTCAGACTAATTTGAC) is present as 19 imperfect direct repeats, and another variant sequence (CCGGGTCTCAGACCAATTTAC) is present as 15 imperfect direct repeats in 180.92 (underlined nucleotides represent differences from RhCMV_{CNPRC}). Due to the number and degeneracy of the repeats in 68-1 and 180.92, there was some uncertainty about the exact extent of overlap with the 5' end of the RhCMV_{CNPRC} sequence.

Analysis of RhCMV_{CNPRC} largely validated the model for the U_Lb' region (Rivailler et al., 2006). A total of 24 ORFs of >99 aa were identified in both strands, 15 of which (transcribed right to left, Fig, 1) are present in either or both of 68-1 and 180.92. Ten of these 15 ORFs have HCMV sequence homologues. Six of the identified ORFs, oriented from left to right (not shown in Fig. 1) were considered unlikely to represent authentic coding regions since they are not conserved in other primate cytomegalovirus genomes and exhibit no significant similarity to other proteins. The remaining 3 ORFs (RhUL146a, RhUL146b, and rh161.1 in segment E; Fig. 1) are absent from 68-1 and 180.92. Each of these novel genes potentially encodes a protein that is related (31–57% identity) to cellular alpha chemokines (Fig. 2).

Alpha chemokines contain a characteristic motif (CXC), and activation of target cells is dependent on the three residues (ELR) immediately preceding this motif (Clark-Lewis et al., 1991). Two of the novel RhCMV_{CNPRC} ORFs (RhUL146b and rh161.1) contain this (ELRCXC) or a very similar motif (ELYCXC or EQRCXC, respectively). The third ORF (RhUL146a) contains another variant ([H/Y]PRCXC). RhUL146a is proposed to contain an intron since protein sequence homology to alpha chemokines is present in two discontinuous regions separated by appropriately located splice sites (Fig. 2D). The proposed protein product of this gene exhibits strong structural homology to alpha chemokines based on Phyre analysis. Although BLAST analysis did not identify any viral protein sequences homologous to

RhUL146a, CCMV but not HCMV also contains an ORF, UL146A, in the same general region of U_Lb' (Davison et al., 2003). CCMV UL146A is related to the tandemly arrayed UL146 ORF, which is homologous to HCMV UL146, an alpha chemokine (Penfold et al., 1999).

The presence of the three novel RhCMV genes and the genomic arrangement of U_Lb' were confirmed using viral DNA amplified from two additional non-co-housed macaques (Accession numbers **EF990256** and **EU003822**) and a clinical RhCMV isolate (22659) (Accession number **EU130540**) that had undergone a limited but unknown number of passages in culture (Alcendor et al., 1993; Barry et al., 1996). Therefore, the U_Lb' structure presented in Figure 1 likely reflects that of wild-type RhCMV.

Flanking these novel genes in segment E are three other alpha chemokine-like ORFs found only in the 68-1 RhCMV strain: UL146 (accession number <u>AAO40076</u>, not annotated in 68-1; M. Penfold, pers. comm.), UL147 (annotated as rh158 in 68-1), and rh161.2 (Fig. 1). ORF rh161.2, originally described in 68-1 as rh161, shares 39% identity with rh161.1 (Fig. 2B). The conserved regions of rh161.1 and 161.2 include a consensus alpha chemokine motif (EQRCQC in rh161.1 and EKECPC in rh161.2), in addition to 7 other cysteine residues. It is likely that rh161.1 and rh161.2 arose via a duplication event, and the nomenclature assigned herein reflects this. Although sequence identity of rh161.2 with alpha chemokines is limited, Phyre analysis revealed some structural homology to interleukin-8 around the CXC motif. While functional studies remain to be performed, it appears that wild-type RhCMV encodes up to six alpha chemokine-like proteins in the U_Lb' region.

Comparison of wild-type RhCMV U_I b' sequences with those from passaged strains revealed a high sequence conservation in most ORFs (Table 1), mirroring what is generally found between HCMV clinical and tissue culture-passaged isolates. Exceptions to this in RhCMV appear to be limited to the alpha chemokine-like genes. UL146 in wild-type variants were as much as 56% divergent from UL146 in 68-1. UL146 in HCMV clinical isolates can also diverge by as much as 60% (Arav-Boger et al., 2006;Hassan-Walker et al., 2004;Lurain et al., 2006; Prichard et al., 2001). Rh161.2 in wild-type RhCMV was also highly variable from that in 68-1 (76–100% aa identity) in addition to differing in length by 12 to 25 aa at the carboxyl terminus (Fig. 2A). One of the breakpoints for DNA sequence discontinuity between RhCMV₆₈₋₁ and RhCMV_{CNPRC} (nucleotide 167312 in 68-1, Fig. 1) is immediately downstream of the reported stop codon for rh161 in 68-1. Therefore, it appears that the rearrangement of the 68-1 genome eliminated the carboxy terminal 12-25 aa of this particular ORF, and generated a novel stop codon in the process. Two partial UL147 clones also exhibited considerable sequence divergence in the amino terminus (54%) (data not shown), similar to what has been reported for HCMV UL147 (Arav-Boger et al., 2006;Lurain et al., 2006). Further analysis is required to confirm this interpretation. High sequence divergence was similarly prevalent in the 3 novel ORFs in segment E. Wild-type sequences diverged from each other by 10% for rh161.1, 44% for RhUL146b, and 46% for RhUL146a (Fig. 2B, C, and D). RhUL146b was also distinguished by differences in the length of the predicted protein.

The tandemly arrayed rh165 and rh166 ORFs appear to have arisen from a single ancestral gene because the wild-type RhCMV sequences exhibit 39% identity with each other. Both wild-type rh165 and 166 are 100% identical to their 68-1 homologues. Phyre analysis does not indicate structural homology with any viral or cellular proteins. The annotation for RhCMV₆₈₋₁ rh166 indicates that the sequence is "similar to UL133" of HCMV (Hansen et al., 2003), however the wild-type rh166 does not share any evident sequence homology with HCMV UL133. Only a weak homology to chimp CMV (CCMV) UL138 (36% identity to aa 25 - 50 of NP_612763) is present in rh166 (data not shown).

Wild-type RhCMV rh157.4 (no homologue in HCMV) and RhUL130 are both homologous to HCMV UL130 within the carboxy terminal two-thirds and the amino terminal one-third of the two ORFs, respectively. The wild-type rh157.4 encodes an additional 14 aa at the carboxyl terminus that are not present in the rh157.4 ORF of RhCMV_{180.92}. One of the rh157.4 clones also contains an internal 16 aa deletion. Apart from these changes, rh157.4 of RhCMV_{180.92} and wild-type RhCMV are 100% identical. Phyre analysis does not reveal any significant homologies with viral or cellular proteins. Potential splice donor and acceptor sites were identified (GT and AG, respectively), which, if utilized, would generate a fusion of rh157.4 and RhUL130 (Fig. 3). The splice acceptor site is located 13 aa before the putative ATG codon of 157.4, seven of which are identical to CCMV UL130. Utilization of these splice sites, as well as those proposed for RhUL128, RhUL131a, and RhUL146a remain to be confirmed.

DISCUSSION

This report redefines the coding potential of the U_I b' region of RhCMV by the identification of three novel alpha chemokine-like ORFs within multiple RhCMV variants that have never been passed in culture. The genomic structure of U_{I} b' is conserved between RhCMVs currently circulating amongst CNPRC macaques and a viral isolate (22659) first passed in culture almost 20 years ago. This evidence strengthens the interpretation that the order of genes in Figure 1 accurately represents the coding content of U_Lb' in wild-type RhCMV. The U_Lb' region of RhCMV, like the homologous region in HCMV, is uniquely unstable and can accumulate mutations and deletions even after brief passage *in vitro*. The basis for the genomic instability within RhCMV U_Lb' remains unknown. There are no obvious repeat or secondary structures, or distinct changes in base composition that might account for the high propensity to undergo rearrangements. Given the extent of the gene loss in many HCMV and RhCMV strains after passage in culture, it is clear that these ORFs are not required for replication in cultured fibroblasts. However, the absence of $U_{\rm I}$ b' rearrangements in RhCMV 22659, which was propagated initially on MRC-5, human diploid fibroblasts, indicates that limited growth on non-host cells does not necessarily lead to genomic rearrangements. Reported in vitro functions for HCMV ULb'-encoded ORFs strongly indicate that their functions contribute to cell tropism and viral pathogenesis and/or escape from host viral clearance mechanisms during establishment and maintenance of persistence. Thus it would appear that growth on cell types susceptible to productive infection *in vivo* should be used to propagate wild-type virus *in* vitro.

CMV species co-evolved with its cognate host during the radiation and speciation of mammals resulting in similar but not identical repertoires of viral immune modulating genes. Co-evolution of RhCMV with its macaque host appears to have selected for a heavy viral investment in alpha chemokine-like proteins. The presence of six ORFs in RhCMV with homology to alpha chemokines, versus only two in HCMV and three in CCMV, suggest that each virus has distinct evasion strategies against host antiviral mechanisms. Although the six RhCMV ORFs share sequence motifs and or structural characteristics with alpha chemokines, *in vitro* and *in vivo* studies are needed to determine the roles of these ORFs in RhCMV natural history. Depending on which amino acids flank the CXC motif, chemokine engagement of its receptor can be agonistic or antagonistic for signaling (Loetscher & Clark-Lewis, 2001; Moser et al., 1993). The promiscuity of chemokine/receptor interactions allows agonists of certain immune cells, such as T_H1 cells, to be antagonists of others, such as T_H2 cells. The presence of these ORFs in RhCMV may enable a subtle and highly refined modulation of host cell trafficking and activation during the earliest stage of RhCMV infection.

It is especially intriguing that these same ORFs are hyper-variable in circulating strains of both RhCMV and HCMV. Both RhCMV and HCMV UL146 exhibit up to 58% sequence divergence, yet for HCMV UL146, functionality is retained despite the divergence (Prichard

et al., 2001). HCMV UL146 binds to IL-8 receptors and induces neutrophil chemotaxis, calcium flux, and degranulation (Penfold et al., 1999), potentially facilitating viral dissemination. All three wild-type RhCMV UL146 variants maintained the ELRCXC motif required for chemokine receptor binding and activity (Clark-Lewis et al., 1991) and four other cysteine residues, implying preservation of tertiary structure. In addition to UL146, RhCMV_{CNPRC} also encodes the HCMV UL147 homologue (rh158) and contains the conserved CXC motif. However, HCMV UL147 does not bind to the CXC receptor, and no chemokine-like function has been observed (Saederup & Mocarski, 2002). It should be noted that there is no rapid genetic drift of HCMV UL146 and 147 sequences *in vivo*. These sequences remain stable in sequence when HCMV virus is repeatedly isolated from individual patients over time (Lurain et al., 2006). Partial clones suggest a similarly high amount of sequence divergence within RhCMV UL147. Together with the high variability, RhUL146a, RhUL146b, rh161.2, RhUL146 and RhUL147 are linked by an apparent involvement in targeting a similar pathway (s). Why the putative viral-mediated modulation of innate cell trafficking and signaling would be associated with profound sequence diversity remains a mystery.

Sequence analysis of cellular proteins has shown that host defense ligands and their receptors undergo the most rapid divergence compared to other cellular protein families. In addition, this diversity may arise from species-specific molecular mimicry by microbes (Murphy, 1993). Thus, it may be hypothesized that diversity found in both host immune defense proteins and viral immune modulatory proteins arises from a stepwise competition for fitness by each organism, where viral hijacking of cellular genes is a common theme.

Additional chemokine-like ORFs in wild-type RhCMV implies that viral modulation of innate immune responses is an especially critical mechanism during the earliest stage of virus infection Evidence for the presence of multiple chemokines from this and other studies strengthens the hypothesis that RhCMV alters numerous host antiviral responses that cumulatively and ultimately favor viral dissemination and lifelong persistence. Additional studies are required to elucidate specific pathways targeted by these immunomodulatory proteins to subvert host immunity and favor viral persistence in an immune competent host.

Sequence homologues of HCMV and CCMV U_Lb' ORFs that were not identified in any RhCMV isolate (CNPRC, 22659, 68-1, and 180.92) include UL133, 135, 136, 139, 140, 142, 147a, 150, 151, and possibly 138. However, functional homologues may be encoded by the numerous ORFs found only in RhCMV (rh157.4, RhUL146a, RhUL146b, rh161.1, rh161.2, rh165–167). Together with an absence of UL18, the lack of a UL142 homologue means that RhCMV does not appear to encode any of the MHC type I homologues found in HCMV. RhCMV does encode other ORFs implicated in HCMV attenuation of natural killer cell function, including UL40, 141, and a duplication of UL83, although functional studies have not yet been performed. Another HCMV ORF missing from RhCMV, UL138, has recently been implicated in the establishment and maintenance of latency in CD34⁺ myeloid progenitor cells. It is not known at this time whether RhCMV infects the CD34⁺ cell lineage in rhesus macaques and thus whether it encodes a functional homologue to HCMV UL138 or an analogous ORF that promotes latency in a different cellular reservoir. The presence of these additional ORFs in both CCMV and HCMV strongly implies that they conferred some selective advantage to the great ape CMVs.

In future studies of the rhesus macaque/RhCMV model, the use of clinical strains of RhCMV possessing the entire U_Lb' region will enhance the understanding of host-virus dynamics involving tropism, latency, and pathogenesis. Moreover, a thorough comparison between the entire genomes of wildtype and laboratory passaged RhCMV may reveal potentially deleted non-core ORFs in regions other than U_Lb'' .

Materials and methods

Oral swabs were collected from RhCMV-seropositive rhesus macaques by running a Dacron swab along the gumlines and buccal pouch and allowing the swab to absorb a saturating amount of saliva (~0.2 ml). Viral DNA was purified from the unclarified swab samples according to previously published methods (Huff et al., 2003). The $U_{\rm I}$ b' region of RhCMV was amplified in overlapping segments by PCR (Taq or Pfu) using primers corresponding to the sequences published for RhCMV strains 68-1 and 180.92 (accessions AY186194 and DQ120516, respectively) (Hansen et al., 2003; Rivailler et al., 2006). The amplicons were cloned using the pCR2.1 TA cloning vector (Invitrogen Corporation, Carlsbad, CA). Every ORF in Figure 1 was analyzed with amplicons from at least two different animals and a clinical isolate, 22659 (Alcendor et al., 1993; Barry et al., 1996). Multiple independent clones from each of three animals were sequenced for the segment E genes. DNA sequences from overlapping clones were aligned using CHAOS-DIALIGN (Brudno et al., 2003), translation products were located with the ExPASy proteomics server (Gasteiger et al., 2003), protein sequences were aligned using the ClustalW multiple sequence alignment program (www.ebi.ac.uk/clustalw/# (Thompson et al., 1994)), relationships to other proteins were identified using the NCBI BLAST programs (www.ncbi.nlm.nih.gov/BLAST/), signal peptide cleavage sites were identified using SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/ (Emanuelsson et al., 2007)), and the Protein Homology/Analogy Recognition Engine (Phyre) was used to analyze proteins for structural similarities to other proteins (Kelley et al., 2000).

The accession numbers for the RhCMV sequences are **EF990255**, **EF990256**, **EU003822**, and **EU130540**.

ACKNOWLEDGEMENTS

This work was supported by funding from the National Institutes of Health to PAB (AI063356 and AI49342-06) and the California National Primate Research Center (RR000169). We are grateful to Andrew Davison for comments on the manuscript.

REFERENCES

- Alcendor DJ, Barry PA, Pratt-Lowe E, Luciw PA. Analysis of the rhesus cytomegalovirus immediateearly gene promoter. Virol 1993;194:815–821.
- Arav-Boger R, Foster CB, Zong JC, Pass RF. Human cytomegalovirus-encoded alpha -chemokines exhibit high sequence variability in congenitally infected newborns. J. Infect. Dis 2006;193:788–791. [PubMed: 16479512]
- Asher DM, C. J. Gibbs J, Lang DJ, Gadjusek DC, Chanock RM. Persistent shedding of cytomegalovirus in the urine of healthy rhesus monkeys. Proc Soc Exp Biol Med 1974;145:794–801. [PubMed: 4131945]
- Barry PA, Alcendor DJ, Power MD, Kerr H, Luciw PA. Nucleotide sequence and molecular analysis of the rhesus cytomegalovirus immediate-early gene and the UL121–117 open reading frames. Virology 1996;215:61–72. [PubMed: 8553587]
- Barry, PA.; Chang, W-LW. Primate Betaherpesviruses. In: Arvin, A.; Campadielli, G.; Moore, P.; Mocarski, E.; Roizman, B.; Whitley, R.; Yamanishi, K., editors. "Human Herpesviruses: Biology, Therapy and Immunoprophylaxis". Cambridge University Press: Cambridge; 2007. p. 1051-1075.
- Britt, W. Virus entry into host, establishment of infection, spread in host, mechanisms of tissue damage. In: Arvin, A.; Campadielli, G.; Moore, P.; Mocarski, E.; Roizman, B.; Whitley, R.; Yamanishi, K., editors. "Human Herpesviruses: Biology, Therapy and Immunoprophylaxis". Cambridge University Press: Cambridge; 2007. p. 737-764.
- Brudno M, Chapman M, Gottgens B, Batzoglou S, Morgenstern B. Fast and sensitive multiple alignment of large genomic sequences. BMC Bioinform 2003;4:66.

- Cha T, Tom E, Kemble G, Duke G, Mocarski E, Spaete R. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J. Virol 1996;70:78–83. [PubMed: 8523595]
- Clark-Lewis I, Schumacher C, Baggiolini M, Moser B. Structure-activity relationships of interleukin-8 determined using chemically synthesized analogs. Critical role of NH2-terminal residues and evidence for uncoupling of neutrophil chemotaxis, exocytosis, and receptor binding activities. J. Biol. Chem 1991;266:23128–23134. [PubMed: 1744111]
- Davison AJ, Dolan A, Akter P, Addison C, Dargan DJ, Alcendor DJ, McGeoch DJ, Hayward GS. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. The Journal of general virology 2003;84:17–28. [PubMed: 12533697]
- Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, Lee L, Addison C, Dargan DJ, McGeoch DJ, Gatherer D, Emery VC, Griffiths PD, Sinzger C, McSharry BP, Wilkinson GW, Davison AJ. Genetic content of wild-type human cytomegalovirus. The Journal of general virology 2004;85:1301–1312. [PubMed: 15105547]
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc 2007;2:953–971. [PubMed: 17446895]
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucl. Acids Res 2003;31:3784–3788. [PubMed: 12824418]
- Hansen SG, Strelow LI, Franchi DC, Anders DG, Wong SW. Complete sequence and genomic analysis of rhesus cytomegalovirus. J. Virol 2003;77:6620–6636. [PubMed: 12767982]
- Hassan-Walker AF, Okwuadi S, Lee L, Griffiths PD, Emery VC. Sequence variability of the alphachemokine UL146 from clinical strains of human cytomegalovirus. J. Med. Virol 2004;74:573–579. [PubMed: 15484281]
- Huff JE, Eberle R, Capitanio J, Zhou S-S, Barry PA. Differential Detection of B virus and rhesus cytomegalovirus in rhesus macaques. J. Genl. Virol 2003;84:83–92.
- Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Molec. Biol 2000;299:499–520. [PubMed: 10860755]
- Loetscher P, Clark-Lewis I. Agonistic and antagonistic activities of chemokines. J. Leuk. Biol 2001;69:881–884.
- Lurain NS, Fox AM, Lichy HM, Bhorade SM, Ware CF, Huang DD, Kwan SP, Garrity ER, Chou S. Analysis of the human cytomegalovirus genomic region from UL146 through UL147A reveals sequence hypervariability, genotypic stability, and overlapping transcripts. Virol. J 2006;3:4. [PubMed: 16409621]
- Moser B, Dewald B, Barella L, Schumacher C, Baggiolini M, Clark-Lewis I. Interleukin-8 antagonists generated by N-terminal modification. J. Biol. Chem 1993;268:7125–7128. [PubMed: 8463247]
- Murphy PM. Molecular mimicry and the generation of host defense protein diversity. Cell 1993;72:823–826. [PubMed: 8458078]
- Penfold MET, Dairaghi DJ, Duke GM, Saederup N, Mocarski ES, Kemble GW, Schall TJ. Cytomegalovirus encodes a potent a chemokine. Proc. Natl. Acad. Sci. (USA) 1999;96:9839–9844. [PubMed: 10449781]
- Prichard MN, Penfold ME, Duke GM, Spaete RR, Kemble GW. A review of genetic differences between limited and extensively passaged human cytomegalovirus strains. Rev Med Virol 2001;11:191–200. [PubMed: 11376481]
- Rivailler P, Kaur A, Johnson RP, Wang F. Genomic sequence of rhesus cytomegalovirus 180.92: insights into the coding potential of rhesus cytomegalovirus. J. Virol 2006;80:4179–4182. [PubMed: 16571834]
- Saederup N, Mocarski ES Jr. Fatal attraction: cytomegalovirus-encoded chemokine homologs. Curr. Top. Microbiol. Immunol 2002;269:235–256. [PubMed: 12224512]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res 1994;22:4673–4680. [PubMed: 7984417]

Oxford et al.



Fig. 1.

Genomic arrangement of the U_Lb' region of wild-type RhCMV circulating in the rhesus macaque colony at the CNPRC (Accession # EF990255) relative to the U_Lb' regions of RhCMV 68-1 (AY186194) and 180.92 (DQ120516). The regions of overlap with 68-1 and 180.92 are listed by the corresponding numbering system for each RhCMV variant. The proposed structure of the consensus RhCMV $U_{I}b'$ region (segments A – D) proposed by Rivailler et al., (Rivailler et al., 2006) is shown together with the novel segment (E) found in RhCMV CNPRC. Predicted ORF within ULb' region of RhMV CNPRC are presented, all of which are on the complementary strand. Shaded ORF in segment E indicate those ORF not present in either 68-1 or 180.92. ORF homologous to HCMV ORF are designated by 'UL'. ORFs without homologues in HCMV are designated with the naming scheme of Hansen et al (indicated by 'rh'). Boxes joined by a horizontal line represent proposed exons. rh161, annotated in 68-1, is listed here as rh161.2. rh161.1 is named to reflect the sequence identity with the previously named rh161 ORF (see text for details). RhUL146a and 146b are named to reflect their alpha chemokine-like sequence homology, similar to RhUL146. *: The point of overlap with 68-1 and 180.92 is imprecise (see text for details). †: The proposed RhUL130 ORF based on a possible spliced product of the rh157.4 and RhUL130 ORF (annotated in 180.92; see Fig. 3 and text for details).

Oxford et al.

	68-1 CNPRC 1 CNPRC 2 22659	1 1 1	MRVLSNEMNT MNT MNN MRVLSNEMNT	FRVTIATMLL VRVNIDAMLL YRV-ICAILY FRVTIATMLL ** * *	ICLILSEFSG ICLILSGFSD MYILLFGVTN ICLILSEFSG *	SQGSELRCSC SQGSELRCSC TLGSELRCHC SQGSELRCSC ******	VKYYSGIPWT VKYYSGIPWT VNYYSGIPWT VKYYSGIPWT * *******	ATCVYLKPKS ATCVYLKPKS ATCVYLQPMH ATCVYLKPKS ****** *	IECNNYELIV VECNNYELIV AGCNKYELII IECNNYELIV ** ****	YDGSPHKTCV YDGSPHKTCV YDGSPHKTCV YDGSPHKTCV ** * ****
	68-1 CNPRC 1 CNPRC 2 22659	81 74 73 81	RVRNPSVFDR RVRNPSVFDR RVKNPSAFDK RVRNPSVFDR ** *** **	LDKQTWFTVT LDKQTWFTVT INRVTWFTVT LDKQTWFTVT *****	KKP-NRHISL KKP-NRHISL KPTRGKQITL KKP-NRHISL * * *	KPQRTSCAVP KPQRTSCAVP KKSNGSCAVV KPQRTSCAVP * ****	KS 121 KS 114 H- 113 KS 121			
B)	RhUL146a CXCL1/GRO CNPRC1 CNPRC2 CNPRC3 22659	1 1 1 1	MARAALSAAP MQFNK MTFCN MQFNK MKFDK	SNPRFLQVAL LACNIFVVTM INRKLFIFTL LACNIFVVTM LACNIFVVTM	LLLLLVATGR VFMLILS AIILLVS VFMLILS VFMLILS *	RAAGASVVTE GTVFANH HTVFAEY GTVFANH GTVFANH	LRCQCLQTLQ PRCLCPRTMK PRCLCIKTAK PRCLCPRTMK PRCLCPRTMK ** *	GIHPKNIQSV GINATDIQIV GIHPKNIQKV GINATDIQIV GVNASDIQIV ** ** *	NVKAPGPHCA RIKLPSSECD DIKEPNSECD RIKLPSSECD KIKLPSSECH * * *	ETEVIATLKN KTEIIVQRRN KLEIIIQLKN KTEIIVQRRN KTEIIVQRKN * * *
	CXCL1/GRO CNPRC1 CNPRC2 CNPRC3 22659	81 70 70 70 70	GQKACLNPAS GFEVCLDTTS GFQLCLDPAS GFEVCLDTTS GFEVCLDTKS * ** *	PMVQKIIKKM PLGKKLMEKY LLGKRLIEKY PLGKKLMEKY PLGKKLMEKY *	LNCDKSN 10 LKRYEQ- 95 NKLYEQ- 95 LKRYEQ- 95 LKRYEQ- 95	7				
C)	Potential intra atgcagtttaacaa M Q F N F	con aact (]	<u>in RhUll46a</u> ttgcatgtaata L A C N	<u>a</u> atatttgtagt† I F V V	acaatggtat T M V 1	ttatgcttata† F M L I	ttatccggtacc L S G T	cgtttttgcaaa V F A I	accatccacgct N H P R	cgtctatgt C L C
	ccacgtactatgaa P R T M M	agg ((gcattaacgcga G I N A	acagacatccag T D I Q	gatagtcagaa I V R	ttaaactaccaa I K L P	agcagtgaatgt S S E C	gataaaaccga DKTI	aaattat <mark>gtga</mark> g E I I (*)	gtaactcta
	cgtcatgctattag	gcaa	atagttatacaa	atatacagagad	ccaaaatctaa	acctatatttad	cagagttcaaco V Q P	gacgaaacggc R R N G	ttgaagtgtgt F E V C	ttggatac L D T
	T S P L	ggta G	aaaaagttgat K K L M	ggaaaaatacci E K Y 1	LAAAACGTTAC L K R Y	gaacaataa EQ*				
D)	Rhull46b MMMU CXCL1 CNPRC 1 CNPRC 2 CNPRC 3 22659	1 1 1 1	MARAALSAAP MNRAI MN-AS MNRAI MNRAI *	SNPRFLQVAL FNPRVLGVAL SSSRFLGVAL FNPRVLGVAL FNPRVLGVAL * ***	LLLLLVATGR LLMTLIAYHQ LLMTLIAYGH LLMTLIAYHQ LLMTLIAHHQ ** * *	RAAGASVVTE TAAVE SVN-E TAAVE TAA-E	LRCQCLQTLQ LRCQCLQVTQ LYCQCTHVTQ LRCQCLQVTQ LRCQCLQVMK *****	GIHPKNIQSV GINPKNIQSM GISKNVIKTV GINPKNIQSM GIPPSNIQRL ** *	NVKAPGPHCA TITKPNGGCD TITSPTSGCD TITKPNGGCD SITRPNAGCE * *	ETEVIATLKN RREIIATLKN HREIILTLKD RREIIATLKN RREIIATLKN * *****
	MMMU_CXCL1 CNPRC 1 CNPRC 2	81 71 69	GQKACLNPAS GQKVCLNPEA GRQTCLNPHS	PMVQKIIKKM PMMKKILSKF PLGKKLLTTV	LNCDKSN PGGTYSSFWQ TH	HFMTLFTD	107 108 90			

108

71 GQKVCLNPEA PMMKKILSKF PGGTYSSFWQ HFMTLFTD

A) UL146

CNPRC 3

Oxford et al.

	22659	70	GKQVCLDPEA * ** *	PMMKKMLSKI * *	PGGTYPSFWE	HLMTLFRDML	TPQA 114			
E <u>)</u>	rh161.1 68-1 rh161.2 CNPRC 1 CNPRC 2 CNPRC 3	1 1 1 1	MSDHPESLSS	TSSINITAAA 	ARTSVYLTFI	YLIAMRSLVL MRVYVA MRVYVA MRVYVA	CVVIMFLCGT CVLLCLY CVLLCVY CVLLCLY	VQGTAREKEC VHGLVAEQRC AHGLAAEQRC VHGLVAEQRC	PCKGKTKLGF QCIGK-KYNR QCIGK-KYNR QCIGK-KYNR	IPPKSDCLWL IPHRTLCLSI IPRKAICLSI IPHRTLCLSI
	22659	1				MRVYVA ** *	CVLLCLY ** **	VHGLVAEQRC * * * *	QCIGK-KYNR * **	IPHRTLCLSI ** **
	68-1 rh161.2 CNPRC 1 CNPRC 2 CNPRC 3 22659	81 43 43 43 43	HQYGSPCGNE EYAGPRCEVT EHAGPRCEVT EYAGPRCEVT EYAGPRCEVT * *	AIAHFPQTVM EAVASFNPIH EAVASFNPIH EAVASFNPIH EAVASFNPIH	NKNGKPRLPL NRPPI NRPPM NRPPI NRPPI * * *	CLDFQIVNST CLNYENIRNR CLDYSNIRNI CLNYENIRNR CLNYENIRNR *	IPNGNGIYC- FPATPGTWCR FPATPGTWCR FPATPGTWCR *. * *	-VKKTVNGTR VGKSLIKVND VGKSLIKVND VGKSLIKVND *	EYVRNCNQ KNCEICNRFV KNCEICNRFV KNCEICNRFV KNCEICNRFV **	146 TLE 110 TLE 110 TLE 110 TLE 110
F <u>)</u>	rh161.2 68-1 CNPRC 1 CNPRC 2 var1 CNPRC 2 var2 22659	1 1 1 1	MSDHPESLSS MSDHPESLSS MSDHPESLSS MSDHPESLSS MSDHPESLSS ******	TSSINITAAA TSSINITAAA TSSINITAAA TSSITAAA TSSINITAAA *** *****	ARTSVYLTFI ARTSVYLTFI ARTSVYLTFI ARTSVYLTFI ARTSVYLTFI ********	YLIAMRSLVL YLIAMRSLVL YLIAMRSLVL YLIAMRSLVL YLIAMRSLVL ********	-CVVIMFLCGT -CVVIMFLCGT LCIVNMFFVA- LCIINMFFVA- -CVVIMFLCGT * **	YQGTARE VQGTARE VQGIGIAKE VQGIGIAKE VQGIGIAKE YQGTARE *** * *	<pre>< ECPCKGKTKI < ECPCKGKTKI < HCPCRNTNKI < HCPCRNTNKI < ECPCKGKTKI < ***</pre>	GFIPPKSDCL KFIPLKSDCL JGIPPNADCL JGIPPNADCL GFIPPKSDCL
	68-1 CNPRC 1 CNPRC 2 var1 CNPRC 2 var2 22659	79 79 81 79 79	WLHQYGSPCG WLHQYGSPCG WLHRLGSPCG WLHRLGSPCG WLHQYGSPCG *** *****	NEAIAHFPQT NEAIAHFPQT DEAVAHFPPT DEAVAHFPPT NEAIAHFPQT ** **** *	VMNKNGKPRL VMNKNGKPRL FL-RNGKPQT FL-RNGKPQT VMNKNGKPRL ****	PLCLDFQIVN PLCLDFQIVN PLCLDYAFVN PLCLDYAFVN PLCLDFQIVN ***** **	STIPNGNGIY ITIPNGNGIY RTIPNGEGIY RTIPNGEGIY STIPNGNGIY ***** ***	CVKKTVNGTH CVKKTVNGTH CVIRHVNNTH CVIRHVNNTH CVKKTVNGTH ** ** **	R EYVRNCNQ R EYVRNCNQCN R EYVPNCEYCN R EYVPNCEYCN R EYVRNCNQCN * *** **	J KFDLHLRMDK J KRKL-LRAGS J KRKL-LRAGS J KFDLHLRMDK
	68-1 CNPRC 1 CNPRC 2 var1 CNPRC 2 var2 22659	159 158 156 159	IWHANSLWWL T T IWHAKSLWWL	- 146 DWK 171 159 157 DWK 171						

Fig. 2.

Protein alignments of (A) RhUL146, (B) RhUL146a, (D) RhUL146b, (E) rh161.1, and (F) rh161.2 of RhCMV₆₈₋₁, RhCMV_{CNPRC}, and RhCMV₂₂₆₅₉. RhCMV_{CNPRC} proteins isolated from different macaques are indicated as CNPRC1, CNPRC2, or CNPRC3. If two variants were isolated from the same macaque, they are designated as var1 and var2. The proposed exon/intron boundaries for RhUL146a are also shown (C). Predicted signal peptide cleavage sites are illustrated by an arrow; the arrow in (E) refers to the predicted signal peptide cleavage site for rh161.1. Boxed amino acids represent the ELRCXC-like motif. Amino acids conserved in all of the aligned proteins are indicated by an asterisk. RhUL146a is aligned with the chemokine ligand 1 (CXCL1/GRO) of rhesus macaque (*Macaca mulatta*) (Accession # NP 001028050).

Oxford et al.

A) Proposed Exon/Intron boundaries for RhUL130/rh157.4	
atgaaaaactataatcaacaactctatacctttgttttgctgatgctattgcgttatggct	
M K N Y N Q Q L Y T F V L L M L L R Y G	20
ttcggtgatacttcttctatcaaggatcttaaactcggcaatttttgtcccatgtattca	
FGDTSSIKDLKLGNFCPMYS	40
tctcctcgacgaaattttgggttgtttacatcttacaaaactcatccaacgggaccggaa	
S P R R N F G L F T S Y K T H P T G P E	60
tgtggaaacaccggcctttacgtttttcacaataggtataatcaataccttattgaaaga	
CGNTGLYVFHNRYNQYLIER	80
ccatccgcgtggtcaaacaaactggcgttctacctcaggtgagttttgtatcataacgtat	
PSAWSNKLAFYLS(*)	93
aatattcataatacagattataccaatacttttgggtgaacatatatttttataaatcat	
tgtagtgcactacactcgcccgtgttccaaaagttttcgaaaatggctacatcgtctact	
ALHSPVFQKFSK <u>MATSS</u> T	111
acatctatgaatattacggaagaagaagaagaagactttcgctgcgcatatgatacccatg	
ТЅМ № ІТЕЕЕККТҒААНМІРМ	131
cgtagtacaactcttcgttacatagttaaggacggtactgacgtggaacactgtcagatg	
R S T T L R Y I V K D G T D V E H C Q M	151
cgtgtcatcacttgggccaagacagaagctaatttcatctctttcaaagttaaaatagaa	
R V I T W A K T E A N F I S F K V K I E	171
$\verb ctttcaaacgcttacagaagacctagtagcatttgtaccaggcctaatctagttgtccct $	
L S N A Y R R P S S I C T R P N L V V P	191
gaacctgataaaataagcccgacagtaaaaacaccaacagtaacaactcccaagcgaacg	
E P D K I S P T V K T P T V T T P K R T	210
gcgagatcaacggtaaaagcaacgccaaggtcagctccagcaaagcaagtggcaccgaag	
А R S T V K A T P R S A P A K Q V A P K	230
ccagcaacagcggcaaaacaggtagcagcgaaaccagctccagcaaaagccagcaacagcg	
РАТААКОVААКРАРАКРАТА	250
gcaaagccggtagcagccaaaccagcggcaacaaagccagcgactaaaccagcagccgca	
АКРVААКРААТКРАТКРААА	270
aaacctgtagcggtaaagccagcggcagcaaaaccaccaacgtcggcagcaaaatcacaa	
КРVАVКРАААКРРТЅААКЅ Q	290
aacagtgcgcccaaaggccagcgaagccaactgtcacaaaagccaaaacctaaacccaaa	
N S A P K G Q R S Q L S Q K P K P K P K	310
$a \verb caccagtaccaattaggtataaacaatacttccaatccaacggaaaaagaaggtacccg $	
T P V P I R Y K Q Y F Q S N G K R R Y P	330
${\tt ctagcgtcacgatatagttccgcctggctgtttaggcggcatccttccggctaatatgtt}$	
LASRYSSAWLFRRHPSG*	347
B) Alignment of RhU:130/157.4 with CCMV UL130	
Rhull30 35 FCP-MYSSPRRNFGLFTSYKTHPTGPECGNTGLYVFHNRYNQYLI	ERPSAWSNKLAFYLS 93
+CP+ + SP R FT Y P GP C N L + +NR Q LI	ERPS W K+ +YLS
CCMV UL130 84 YCPLTFPSPPRRLSEFTRYHEVPGGPSCNNETLVLLYNRDGQTLI	ERPSPWVKKVTWYLS 143
▼ Rhill.130 94 ALHODVE∩KESKMATS-STTSMNTTEEEKKTEAAHMTDMDSTTD	VIVKDGTDVFHCOMP 152
DALEW DATE DATE TO DAT	+++ DGT E C M+
CCMV III.130 144 GRDHPVFORFAKTASTANDGNVOTRTDDAKVFGSHMVPKOTKLER	EVINDGTRYEMCIMK 203
	I VIIVESTICIENCIAL 205
Rhui.130 153 VITWAKTEANF-ISEKVKIELSNAYRRPSSICTRPNLVV 190	
+ TWA $++$ $++F+$ $+$ A $+$ S CT PNL+V	
CCMV UL130 204 LETWAHVFRDYSVAFOTRLTFTOANNOTYSYCTHPNLLV 242	

Fig. 3.

Proposed splicing of the RhUL130 and rh157.4 ORF. (A) A putative intron for a spliced RhUL130 and rh157.4 fusion is indicated by shading. The first five amino acids of the unspliced rh157.4 ORF are underlined. (B) Alignment of the RhUL130/rh157.2 fusion with chimp CMV UL130 (CCMV; accession # <u>NP_612749</u>) with the site of the proposed exon/exon boundary indicated by the arrow.

ORFs in the ULb' Region of RhCMV _{CNPRC} RhCMV HCMV Length D with D with ORF from to (aa) Exon RhCMV HCMV Comment	None None 1 747 249	None None 2 793 264	None None 3 761 253	None None 9459 9812 118	None None 1141111848 146	None None [12311]12619 102	722 1175 150 3	RhUL128 UL128 1261 1398 46 2 100 44* * Homology	1487 1626 47 1	rh157.4 NH 1628 2359 243 100* *14aa longer	RhUL130 UL130 2481 2762 93 97 44 * Homology 1	Bh111 131.6 I III 131.8 2759 3020 86 2 00 31	MIOLIDIA ULTUTA 3103 3359 86 1 77 31	RhUL132 UL132 3403 4068 221 99 31 (rh160)	RhUL148 UL148 4134 5114 326 99 30* * Homology be (rh159) (rh159) 99 30* * Homology be	RhUL147 UL147 5319 5780 153 96 36 CXC motif (rh158) 96 36	RhUL146 UL146 5831 6175 114 92/58/100*27/33/27/CXC motif; *	RhUL146a ^E NH 6252 6348 31 2 CXC motif	RhUL146b ^E NH 6738 7064 108 CXC motif	rh161.1 ^E NH 7188 7520 110 CXC motif; p	$ \frac{1}{12} \text{ mb} = 1.57 \text{ mb} = 1.51 m$	RhUL145 UL145 8202 8507 101 65 (rh162) (rh162) 65 65 65	RhUL144 UL144 8925 9440 171 98 31 (rh163)	RhUL141 UL141 9643 10935 430 99 44 (rh164)	rh165 NH [11396[11827] 143 100 possible tander	rh166 NH 1187212396 174 100 possible tander	rh167 NH [12527]13030] 167 98
RhCMV ORF	None	None	None	None	None	None		RhUL128		rh157.4	RhUL130	PhIII 131.		RhUL132 (rh160)	RhUL148 (rh159)	RhUL147 (rh158)	RhUL146	RhUL146a ^E	RhUL146b ^E	$rh161.1^{E}$	rh161.2	RhUL145 (rh162)	RhUL144 (rh163)	RhUL141 (rh164)	rh165	rh166	rh167

2 ά 5 Ļ.

NH: no homologue in HCMV

E: ORFs found in segment E of RhCMVCNPRC and RhCMV22659 (see Fig. 1)

Oxford et al.