Human Cytomegalovirus–Encoded α-Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns

Ravit Arav-Boger,1,2 Charles B. Foster,1 Jian-Chao Zong,2 and Robert F. Pass3
1Division of Infectious Diseases, Department of Pediatrics, Johns Hopkins Hospital, and 2Molecular Virology Laboratories, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland; 3Department of Pediatrics, University of Alabama at Birmingham, Birmingham

Most congenital human cytomegalovirus (HCMV) infections are asymptomatic, but some lead to severe disease. We hypothesized that differences in disease manifestations may be partially explained by differences in viral strains. We recently reported an association between unique long (UL) 144 gene polymorphisms and clinical disease. We now report on the sequence heterogeneity of 2 potential HCMV virulence genes that encode α-chemokines: UL146 and UL147. These 2 genes were highly divergent in cultured isolates obtained from 23 newborns with congenital HCMV infection and were difficult to categorize. Unlike our findings for the contiguous UL144 gene, no specific UL146 or UL147 genotype was associated with disease outcome.

Human cytomegalovirus (HCMV) is the most common cause of congenital infection in the United States, infecting ~1% of newborns. Of infected newborns, ~10% are symptomatic and have fatal or multisystem disease, and ~90% show no clinical evidence of infection during the neonatal period [1]. The majority of symptomatic newborns will have central nervous system sequelae, such as mental retardation, cerebral palsy, and hearing loss, whereas only ~15% of those who are asymptomatic as newborns will have such sequelae [1]. Knowledge of either host or viral factors that could account for the variability in the severity of disease and the outcome of congenital HCMV infection in newborns is very limited. The role played by putative HCMV virulence genes as determinants of disease outcome during congenital HCMV infection is gaining interest, as sequence information from laboratory strains and human isolates is accumulating [2].

A substantial portion of the HCMV genome encodes molecules that have the potential to affect virulence through immune evasion, molecular mimicry, or interference with host chemokines [3]. Potential virulence determinants have been located in both the unique long (UL) region and the unique short region of the HCMV genome [4, 5]. Clinical isolates of HCMV—but not the laboratory-adapted AD169 and late Towne-vaccine strains [6]—contain a tumor necrosis factor (TNF)–α receptor gene (UL144) and 2 adjacent α-chemokine genes (UL146 and UL147). The UL144 sequences can be divided into 3 distinct phylogenetic variants. We have reported previously that uncommon UL144 variants correlate with the severity of disease caused by congenital HCMV infection [7]. UL146 encodes a protein that induces calcium mobilization, chemotaxis, and degranulation of neutrophils [4]. The function of the closely related UL147 gene has not been previously studied. Chemokine activation of infected neutrophils, which leads to extravasation, cell trafficking, and dissemination of HCMV-infected neutrophils to tissues, could contribute to the outcome of congenital HCMV infection. To determine whether sequence variability in these α-chemokine–encoding genes among strains is associated with disease outcome and to characterize the relationship between previously identified variants in UL144, we sequenced the UL146 and UL147 genes from cultured isolates of 23 newborns with congenital HCMV infection, including 10 who were asymptomatic and 13 who were symptomatic.

Subjects, materials, and methods. We studied viral isolates from 23 HCMV-infected newborns who were identified during the first 3 weeks of life. Thirteen viral isolates (8 cultured from urine samples, 4 cultured from saliva samples, and 1 cultured from a blood sample) were obtained from symptomatic newborns in Birmingham, Alabama, and Baltimore, Maryland; 10 other isolates (all cultured from saliva samples) were obtained from asymptomatic newborns identified in Birmingham through an HCMV screening program. Viral isolates were propagated in human foreskin fibroblast cultures until cytopathic effect (CPE) was observed; no differences in growth or CPE were observed among the isolates.

After CPE was observed, total genomic cell and viral DNA
was extracted from infected cells by use of a capture-column kit (Gentra Systems). Extracted DNA preparations were diluted in water for use as templates for polymerase chain reaction (PCR) amplification.

All PCRs were performed in a PCR-dedicated room with clean, PCR-dedicated equipment. Filter tips were used for all pipetting. Positive and negative controls were run with each PCR. Forward and reverse primers flanking the contiguous UL146 and UL147 genes were used for first-round PCR amplification, yielding a size of 1130 bp (table 1). Because of the unexpected hypervariability of these genes, 2 sets of nested primers were used for amplification of UL146 and UL147 when the product of the outside primers was negative or weak (table 1). To improve annealing of the nested primers to hypervariable regions, >1 nt was designed at a specific position; in table 1, these appear as superscript and subscript characters. The conditions for amplification with all primer sets were as follows: 5 min at 94°C; 34 cycles of 1 min at 94°C, 1 min at 57°C, and 1 min at 72°C; and a single extension step of 7 min at 72°C.

For DNA sequencing, PCR products were gel purified by use of a Qiagen extraction kit and were then sequenced directly by use of the BigDye Terminator Cycle Sequencing Kit (Perkin-Elmer/Applied Biosystems), using the same primers. The sequencing products were analyzed on an ABI 310 automated sequencer.

Protein sequences were aligned in ClustalW (version 1.8; available at: http://www.ebi.ac.uk/clustalw/), and DNA sequences were aligned on the basis of the protein sequence alignment by use of RevTrans (version 1.4; available at: http://www.cbs.dtu.dk/services/RevTrans/) [8]. Phylogenetic trees were inferred from nucleotide sequences by use of PAUP (version 4.0; Sinauer Associates). The HKY-85 model of evolution was used. Trees were constructed using the neighbor-joining method, and internal node support was verified using the bootstrap method, with 1000 replicates [9].

We initially categorized the sequences without knowledge of disease outcome, and then the ordered distribution of UL146 and UL147 polymorphisms was compared for asymptomatic and symptomatic newborns. We also determined whether there was phylogenetic linkage between UL144 genotypes and UL146 and UL147 genotypes.

Results. We amplified 21 UL146 sequences and 23 UL147 sequences of isolates from 23 newborns with congenital HCMV infection. To provide reference sequences, we sequenced the laboratory strains Toledo, Towne, Fiala B, and Fiala F as well as 5 HCMV isolates from 3 bone marrow transplant (BMT) recipients and 1 renal transplant recipient; for one of the BMT recipients (BMT2), HCMV was isolated from both the blood and the cerebrospinal fluid (CSF). We found that UL146 varied in length and amino acid composition among isolates from the different newborns. Each of the 21 neonatal UL146 sequences and the 9 reference UL146 sequences (GenBank accession numbers AY681088–AY681116) showed conservation in the CXC chemokine motif, and all retained the 4 cysteine residues that are considered to be necessary for the functional activity of chemokines. Twenty-eight sequences included the glutamic acid–leucine–arginine (ELR) CXC motif, whereas 2 extremely variant subtypes (from newborns A2 and BMT3) did not include this important motif (figure 1). A high level of divergence (up to 70%) was observed at the amino acid level. Because of the high sequence variation and the small number of isolates in each group, we have not assigned a nomenclature to these groups.

We also found UL147 to be highly variable in the 23 neonatal and the 9 reference sequences analyzed (GenBank accession numbers AY689094–AY689124). There was up to 20% variability in the predicted amino acid sequences. Variability was mainly observed in the N-terminal region of UL147, similar to the variation observed in UL144. The cysteine motifs were conserved (as in UL146) in all 32 sequences. For BMT2, the isolate cultured from the blood and the isolate cultured from the CSF exhibited different sequences in the UL146 and the UL147 locus, suggesting the presence of multiple genotypes. The adjacent UL144 gene clusters into 3 major genotypes, as previously reported [7]. Some phylogenetic linkage was found among UL146 and UL147 genotypes (figure 1); however, no phylogenetic linkage was observed among UL144 and UL146–UL147 sequenced from the same isolates.

The analysis to determine whether an association existed between the UL146 and UL147 genotypes and the outcome of congenital HCMV infection was underpowered, because of the unexpectedly high sequence variation in both genes. No specific UL146 or UL147 genotype was associated with disease outcome in either the asymptomatic or the symptomatic newborns. Small numbers of each genotype were noted among both the asymptomatic and the symptomatic newborns.

Discussion. HCMV has a large and complex genome, containing ~235 kb of DNA and open-reading frames (ORFs) for ~200 proteins. Significant variability has been observed in the UL/b' region among strains [6, 7, 10, 11], an unusual finding.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL146-7 forward</td>
<td>GTCATGGACCGAGTTTG</td>
</tr>
<tr>
<td>UL146-7 reverse</td>
<td>GAACGATCTCGCTCCTGTT</td>
</tr>
<tr>
<td>UL146 heminested reverse 1</td>
<td>CTAAAAATGAGGCTAGG</td>
</tr>
<tr>
<td>UL146 heminested reverse 2</td>
<td>GTCTGAATTCATCCAATTC</td>
</tr>
<tr>
<td>UL147 heminested reverse 1</td>
<td>CCACTATGTCCCCCCTAGGTC</td>
</tr>
<tr>
<td>UL147 heminested reverse 2</td>
<td>TTATCGGATGATAATG</td>
</tr>
</tbody>
</table>

NOTE. To improve annealing of the nested primers to hypervariable regions, >1 nt was designed at a specific position; these appear as superscript and subscript characters.
for a DNA virus. Therefore, considerable interest in the biological and evolutionary roles played by these nucleotide variations exists. We here report extraordinary sequence diversity of 2 HCMV-encoded \( \alpha \)-chemokine genes among 9 reference strains and strains isolated from 23 congenitally infected newborns.

HCMV uses multiple strategies to avoid host immune detection, one of which is the synthesis of molecules that manipulate host immune/inflammatory responses. Both murine CMV and HCMV contain ORFs that encode G protein–coupled receptor homologues, which bind to host chemokines and may sequester chemokines from the extracellular environment [3]. HCMV also encodes divergent proteins with the sequence characteristics and activities of chemokines [12]. Murine CMV encodes a CC chemokine homologue (MCK-2) that confers increased inflammation, higher levels of viremia, and higher titers of virus in salivary glands [13].

The 2 highly variant \( \alpha \)-chemokine genes in HCMV, UL146 and UL147, share size and sequence similarity with those of human \( \alpha \)-chemokines. Chemokines belonging to this family often function as chemoattractants that recruit leukocytes to sites of infection or inflammation [14]. The HCMV-encoded \( \alpha \)-chemokine genes were identified among 19 ORFs in the UL/b’ region of low-passage clinical isolates, ORFs that were lost after extensive serial passage of Towne and AD169 in tissue culture [6]. Given that the absence of these \( \alpha \)-chemokine genes and contiguous genes is associated with an attenuated phenotype, it is reasonable to infer that the gene products may confer a selective advantage to the virus in human hosts. UL146 encodes the \( \alpha \)-chemokine vCXC1, which induces calcium mobilization, chemotaxis, and degranulation of neutrophils [4]. This protein has also been shown to bind to CXCR2 with a biological potency similar to that of other known \( \alpha \)-chemokines. UL147 encodes a less-characterized \( \alpha \)-chemokine homologue—designated “vCXC2”—that has a less-specific CXC motif. The predicted gene product for UL146 was highly divergent yet maintained the amino acids required for activity. This suggests that divergence occurred in the context of a selective advantage with respect to the expression of a functional chemokine. Our results agree with those of Dolan et al. [10]
in that the ELRCXC motif was conserved in most genotypes but the asparagine-glycine-arginine (NGR) CXC motif appeared in few. Chemokines with the ELRCXC motif are chemoattractic to neutrophils, whereas chemokines lacking this motif may act on T and B cells [15].

The higher variation in UL146 and UL147 relative to UL144 could be explained by the fact that the gene products for UL146 and UL147 are secreted and, therefore, are subject to selective pressure by the immune system, whereas UL144 encodes an integral membrane protein of the TNF superfamily. However, we found no evidence to support the hypothesis that protein sequence variation is a result of positive selective pressure. Indeed, the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site across both UL146 and UL147 was <1, suggesting that both genes are under purifying selection. A random distribution of sequence variants and extensive loss of linkage disequilibrium at closely related loci suggested that the high nucleotide diversity is due to homologous recombination between UL146 and UL147 [12].

We and others have noted the occurrence of infection with multiple HCMV strains [7, 16]. In the present study, we detected 2 different UL146 and UL147 genotypes in a BMT recipient from whom we had both a blood and a CSF sample. However, for all congenitally infected newborns, we had sequence information for virus obtained from only 1 site (urine, saliva, or blood). Although we did not detect multiple genotypes in this neonatal cohort, we cannot exclude the possibility that some of the newborns may have harbored additional variants. A limitation associated with PCR is that genotypes may be missed if the primer binding domains in the viral strains do not exactly match the primers used, but our use of degenerate primers minimized this possibility (table 1) [17].

We previously found an association between UL144 polymorphisms and clinical disease in congenitally infected newborns [7]. We therefore expected to find a nonrandom segregation of the α-chemokine genes between asymptomatic and symptomatic newborns. However, there was no linkage of UL146-UL147 with UL144, nor did these 2 α-chemokine genes segregate by disease severity. A technical explanation for this is that, because of the highly variant sequences, our study was underpowered to detect an association in a sample of 23 isolates. The findings of our study agree with those of a recent study describing the sequence variability of UL146 among clinical strains of HCMV [18]. We show here extraordinary hypervariability of UL146 and UL147 in a cohort of newborns with congenital HCMV infection, and this hypervariability did not allow us to detect an association between 2 α-chemokine gene polymorphisms and disease outcome. These genes should be further studied, to better understand both the mechanisms that contribute to their hypervariability and their differing abilities to affect neutrophil recruitment and to interfere with host immune responses.

References