JOURNAL OF VIROLOGY, Jan. 1996, p. 78–83
0022-538X/96/$04.00+0

Vol. 70, No. 1

Human Cytomegalovirus Clinical Isolates Carry at Least 19 Genes Not Found in Laboratory Strains

TAI-AN CHA,[1]* EDWARD TOM,[1] GEORGE W. KEMBLE,[1] GREGORY M. DUKE,[1] EDWARD S. MOCARSKI,[2] AND RICHARD R. SPAETE[1]

Aviron, Burlingame, California 94010,[1] and Department of Microbiology & Immunology, Stanford University School of Medicine, Stanford, California 94305-5402[2]

Received 9 August 1995/Accepted 26 September 1995

Nucleotide sequence comparisons were performed on a highly heterogeneous region of three human cytomegalovirus strains, Toledo, Towne, and AD169. The low-passage, virulent Toledo genome contained a DNA segment of approximately 13 kbp that was not found in the Towne genome and a segment of approximately 15 kbp that was not found in the AD169 genome. The Towne strain contained approximately 4.7 kbp of DNA that was absent from the AD169 genome, and only about half of this segment was present, arranged in an inverted orientation, in the Toledo genome. These additional sequences were located at the unique long (U_L)/*b′* (IR_L) boundary within the L component of the viral genome. A region representing nucleotides 175082 to 178221 of the AD169 genome was conserved in all three strains; however, substantial reduction in the size of the adjacent *b′* sequence was found. The additional DNA segment within the Toledo genome contained 19 open reading frames not present in the AD169 genome. The additional DNA segment within the Towne genome contained four new open reading frames, only one of which shared homology with the Toledo genome. This comparison was extended to five additional clinical isolates, and the additional Toledo sequence was conserved in all. These findings reveal a dramatic level of genome sequence complexity that may explain the differences that these strains exhibit in virulence and tissue tropism. Although the additional sequences have not altered the predicted size of the viral genome (230 to 235 kbp), a total of 22 new open reading frames (denoted UL133 to UL154), many of which have sequence characteristics of glycoproteins, are now defined as cytomegalovirus specific. Our work suggests that wild-type virus carries more than 220 genes, some of which are lost by large-scale deletion and rearrangement of the U_L/*b′* region during laboratory passage.

Human cytomegalovirus (HCMV) is a widespread herpesvirus that usually produces asymptomatic infections in immunocompetent hosts. Serious disease can occur in immunocompromised individuals and in congenitally infected newborns (1). The attenuated HCMV strains Towne and AD169, now widely used in laboratory studies, were developed as vaccine candidates and used in clinical trials. The AD169 strain was shown to be well tolerated and immunogenic as judged by seroconversion without apparent side effects (10), and the Towne strain induced seroconversion without producing disease (27, 31). The Towne vaccine protected immunosuppressed renal transplant recipients against disease and was not shed (28, 29). In contrast, a low-passaged HCMV strain, Toledo, has produced clinically apparent disease when administered to healthy adult volunteers (31). These clinical data indicate that HCMV strains exhibit different levels of virulence depending on their passage history in cell culture.

High- and low-passaged HCMV strains also exhibit different competencies in tropism for epithelial cells. When the growth of Toledo, Towne, and AD169 was compared in the SCID-hu (thymus plus liver) mouse, Toledo replicated to levels at least 3 orders of magnitude higher than did either of the laboratory strains. The differences among these viruses were attributed to the possible loss of genetic information by Towne and AD169 during propagation in cell culture (4).

In this study, we hypothesize that a substantial amount of genetic information encoded by the viral genome must have been lost during long-term passage in cell culture. The HCMV genome, the largest of the herpesviruses thus far described, is a linear, double-stranded DNA consisting of approximately 230 kbp that is organized as two regions of unique sequence (unique long [U_L] and unique short [U_S]) flanked by two sets of inverted repeats, *ab* (TR_L/IR_L) and *ac* (IR_S/TR_S), in an arrangement that can be depicted as $5'\text{-}a_L a_n b\text{-}U_L\text{-}b'a'_m c'\ U_S\text{-}ca_n a_S\text{-}3'$ (23). The genomes of HCMV strains are highly homologous as judged by reassociation kinetics (15, 30), Southern hybridization (20, 30, 48), and restriction site polymorphism (5, 12, 47). Only the AD169 genome has been completely sequenced (EMBL accession number X17403) (6). Additional nucleotide sequence information is available for only a few loci or genes of other strains (for example, see references 8, 9, 13, 22, 25, 26, 40, 41, and 43). Although these lines of evidence suggest a high degree of similarity among HCMV strains, there have been reports of substantial variability near the U_L and *b*-repeat boundaries (20, 42, 48). We performed nucleotide sequence analysis of the U_L/*b′* region, comparing the genomes of the low-passage strain Toledo and the two high-passage strains Towne and AD169. Our results highlight the discovery of a large region of Toledo that is missing in both Towne and AD169.

MATERIALS AND METHODS

Cells, virus, and viral DNAs. HCMV Toledo (passage 8, from S. Plotkin, Pasteur Merieux, Paris, France), Towne (passage 134, from S. Plotkin), and AD169 (ATCC VR-538) were propagated in human foreskin fibroblasts cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (JRH Biosciences, Lenexa, Kans.) (38). Viral DNA was purified from infected-cell lysates by equilibrium centrifugation in NaI gradients (39). Viral DNA of clinical isolates (HCMVF, C128, C354, C793, and C980), obtained from M. Fiala, University of California Los Angeles, and S. Chou, Oregon Health Sciences University, was purified as previously described (36).

* Corresponding author. Mailing address: Aviron, 1450 Rollins Rd., Burlingame, CA 94010. Phone: (415) 696-9129. Fax: (415) 347-6790. Electronic mail address: cha@aviron.com.

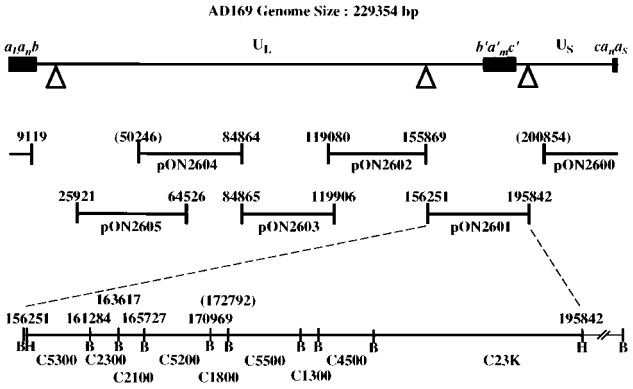

FIG. 1. Schematic illustration of Toledo cosmid and plasmid subclone maps, with the AD169 genome diagram shown on top. A set of cosmids representing the *Hin*dIII partially digested Toledo genome is shown under the diagram of the AD169 genome. The AD169 genome is displayed in the prototype orientation, with unique sequence denoted by a thin line and inverted repeats flanking the U_L and U_S components denoted by boxes, $a_L a_n b$, $b'a'_m c'$, and $ca_n a_s$. Open triangles denote the corresponding AD169 regions of the Toledo genome not represented by the cosmids. An expanded map of pON2601 indicating restriction sites B (*Bam*HI) and H (*Hin*dIII) is shown on the bottom line. The numbers above the diagrams denote the corresponding nucleotide positions on the AD169 genome. The numbers within parentheses denote the presence of nucleotide changes at these restriction sites in AD169. Subclones are labeled below the diagrams.

Clones. Purified Toledo genomic DNA was partially digested by *Hin*dIII, and the fragments generated were cloned into a cosmid vector, pHC79 (14). By using the published AD169 sequence as a reference (6), a set of six cosmids (pON2600, pON2601, pON2602, pON2603, pON2604, and pON2605) representing greater than 90% of the Toledo genome were identified by sequencing the ends of the viral DNA inserts (Fig. 1). pON2601 (~50 kbp), which spanned the U_L/b' region (corresponding to nucleotide positions 156251 to 195842 of the AD169 genome), was digested into nine *Bam*HI fragments, and eight (C5300, C2300, C2100, C5200, C1800, C5500, C1300, and C4500) were subcloned into pBluescript M13+ plasmid vector (Stratagene, La Jolla, Calif.) (Fig. 1). The remaining fragment containing the cosmid vector and a Toledo sequence was recircularized to generate C23K (Fig. 1). Plasmids harboring the *Xba*I E, Q, and T fragments of Towne were obtained from M. Stinski, University of Iowa (44).

Nucleic acid sequence determination and analysis. Both strands of the U_L/b' region of Toledo and Towne were sequenced by single-cycle and multicycle dideoxynucleotide-chain termination methods (35). Primers were synthesized in house with an ABI 392 oligonucleotide synthesizer (Applied Biosystems, Foster City, Calif.). To expedite sequencing, a variety of DNA templates (single-stranded phage DNA, double-stranded plasmids and cosmids, viral genomic DNA, and PCR products) were used. Sequencing ladders were analyzed manually and with an ABI 373 automated sequencer (Applied Biosystems). The contiguity and analysis of the sequences including homology and motif searches were performed with MacDNASIS (Hitachi Software, San Bruno, Calif.). Homology searches were also performed with BLAST (2) through NCBI services.

Southern blot analysis. Viral DNAs were restriction enzyme digested, electrophoresed in 1% agarose gels under standard conditions (34), transferred to Hybond-N+ nylon membranes (Amersham Corp., Arlington Heights, Ill.), and immobilized with a UV Crosslinker 1000 (Hoefer Scientific Instruments, San Francisco, Calif.). Radiolabeled probes were prepared by nick translation of either plasmid or viral DNA in the presence of [α-^{32}P]dCTP (Amersham Corp., Arlington Heights, Ill.) (33). Hybridizations were performed essentially as previously described (39). Membranes were probed, washed three times with 2× SSC (0.3 M NaCl, 0.03 M sodium citrate) at 68°C and twice with 2× SSC at room temperature, air dried, and autoradiographed with Kodak XAR film.

Nucleotide sequence accession numbers. The sequences of Toledo and Towne determined in this study have been submitted to the GenBank (accession numbers U33331 and U33332, respectively).

RESULTS

Mapping large DNA sequence differences in Toledo, Towne, and AD169 strains of HCMV. The report by LaFemina and Hayward (20) that ~2 kbp was missing from the AD169 genome relative to Towne led to our initial analysis of the *Xba*I T fragment of Towne DNA (19, 37). We sequenced over 6 kbp of Towne DNA including the entire *Xba*I T fragment and portions of the flanking *Xba*I fragments, E and Q (Fig. 2C). Contiguity between *Xba*I-E/T and *Xba*I-T/Q was established by sequencing PCR products generated with Towne genomic DNA and primers that annealed to opposite sides of the *Xba*I sites. As indicated in Fig. 2C and D, the sequence colinearity between Towne and AD169 broke at positions corresponding to positions 178221 and 180029 of AD169, within the *Xba* I E and Q fragments, respectively, of Towne. Sequence analysis demonstrated that the 1,808-bp sequence of AD169, at the left end of the IR_L (b'), has been replaced in Towne by 4,706 bp of novel sequence.

To compare Toledo and the two laboratory strains, we per-

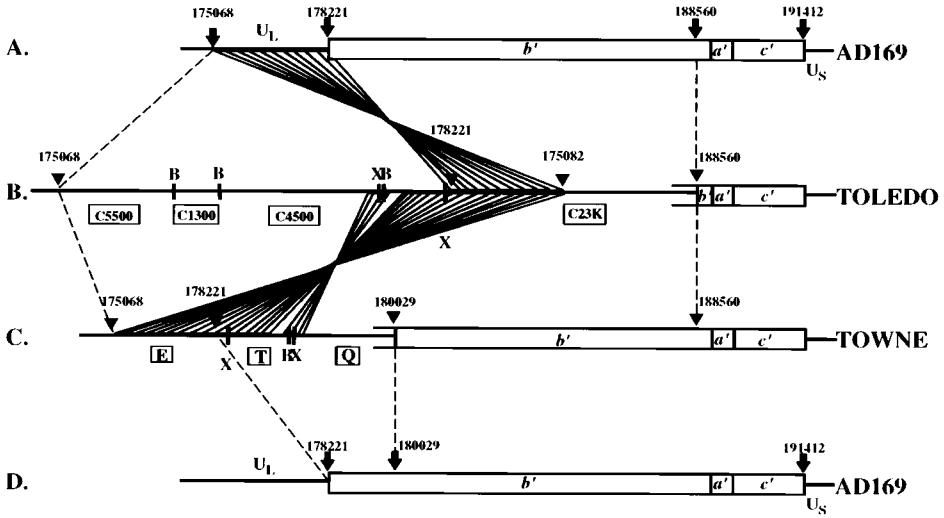

FIG. 2. Schematic illustration of the U_L/b' region in Toledo and Towne genomic DNA in comparison with the corresponding region in AD169. Dashed lines bracket the region of different sequence when pairwise comparisons are made. U_L, U_S, and internal repeat regions (b', a', c') are indicated. B and X represent *Bam*HI and *Xba*I sites, respectively. The four *Bam*HI-generated subclones (C5500, C1300, C4500, and C23K) of pON2601 are labeled under the Toledo diagram. *Xba*I fragments (E, T, and Q) are indicated under the Towne diagram. Solid triangles mark the positions where sequence similarity breaks between pairwise comparisons of Toledo, Towne, and AD169. The skewed-shaded areas represent regions where the sequence is in an inverted orientation in Toledo relative to the others. Inverted arrows represent reference points on the AD169 genome. The precise position of the U_L/b' junction of Toledo and Towne is not determined. Nucleotide positions are denoted as based on the AD169 sequence.

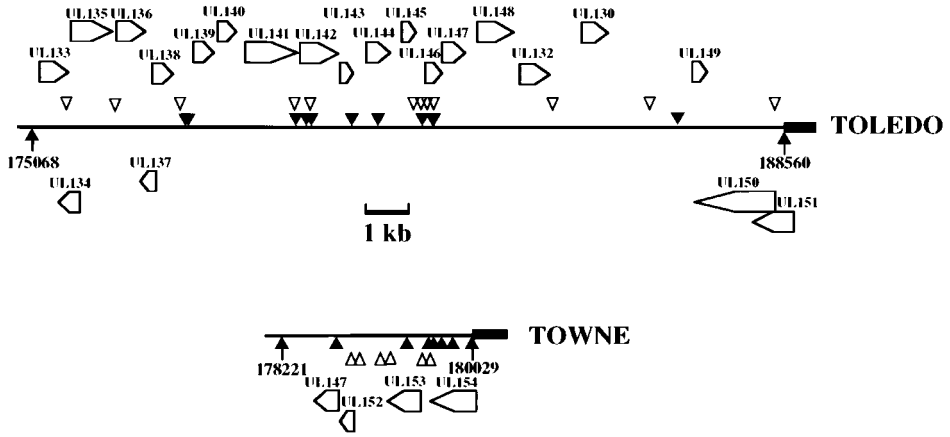

FIG. 3. Relative position of the predicted ORFs in the U_L/b' region of Toledo and Towne. Thin horizontal lines represent U_L, and thick horizontal lines represent an internal repeat region. Vertical arrows bracket the regions of different sequence in Toledo and Towne when compared with AD169. Corresponding nucleotide positions in AD169 are denoted below the lines. The position, length, and orientation of the ORFs are illustrated by open arrows. The six translation frames of predicted ORFs are denoted by a staggered pattern above and below the genome diagram. Solid triangles mark the position of TATAA sequence, and open triangles denote the position of AATAAA sequence.

formed interstrain cross-hybridization experiments. DNA probes of Toledo and Towne DNA with an average size of 500 bp were prepared by sonication followed by multiple restriction enzyme digestions (*Ava*I, *Ban*I, and *Bfa*I). The probes were radioactively labeled and hybridized to immobilized restriction enzyme-digested Toledo, Towne, and AD169 DNAs. This technique identified a *Bam*HI fragment (~4.5 kbp) of Toledo that did not hybridize to probes generated from Towne or AD169 DNA (data not shown). This *Bam*HI fragment was subsequently cloned (Clone65) and sequenced in its entirety (4,663 bp). Our analysis revealed a 327-bp sequence identity between Clone65 and Towne DNA but no sequence similarity to AD169. These data mapped the 327-bp sequence to the *Xba* I T/Q junction of Towne, placing Clone65 in the U_L/b' region, and indicated the existence of a much larger sequence difference between Toledo and the laboratory strains at this location.

To facilitate sequencing, we generated Toledo DNA subclones of the entire U_L/b' region by constructing a cosmid library with *Hin*dIII partially digested Toledo genomic DNA. The cosmids were mapped by sequencing the ends of each insert and guided by referring to the published AD169 sequence (6). As shown in Fig. 1, we identified six cosmids that constituted one set of the clones representing greater than 90% of the Toledo genome with respect to AD169. One of the cosmids (pON2601, ~50 kbp) contained the entire internal repeats and portions of the neighboring U_L and U_S regions (corresponding to a region encompassing nucleotides 156251 to 195842 of the AD169 genome). Nine subclones representing *Bam*HI fragments of pON2601 were generated and mapped by sequencing the ends of each insert (Fig. 1). Five (C5300, C2300, C2100, C5200, and C1800) of the nine subclones could be mapped to the AD169 genome. Two of the remaining four subclones, C1300 and C4500, contained no sequence similarity to AD169, and the other two subclones, C5500 and C23K, were similar to AD169 at only one of their ends. Subclone C4500 was shown to be identical to Clone65. The arrangement of these four *Bam*HI fragments (Fig. 1 and 2B) was established by sequencing across the *Bam*HI sites of pON2601.

As shown in Fig. 2B, approximately 19 kbp of Toledo sequence from clones C1300, C4500, and portions of C5500 and C23K was determined. The sequence colinearity between Toledo and AD169 broke at the corresponding nucleotide positions 175068 and 188560 of AD169 located within C5500 and C23K of Toledo, respectively. A portion of this region (nucleotides 175082 to 178221) from the right end of AD169 U_L was conserved in Toledo but was arranged in an inverted orientation (Fig. 2A and B). Sequence analysis also revealed that ~5 kbp of Towne sequence (including 3,140 bp shared by all three strains) was conserved in Toledo in an inverted orientation (Fig. 2B and C). Significantly, our data revealed a very large DNA sequence of Toledo that was absent from both laboratory strains, i.e., ~13 kbp relative to Towne and ~15 kbp relative to AD169. Although the exact junction between U_L and b' in Toledo was not mapped, Southern blot experiments (described below) indicated that this additional sequence extended the U_L region in Toledo. The shortened b' repeat sequence of the Toledo genome relative to the laboratory strains serves to compensate for the small change in length over the entire genome.

ORF analysis of the Toledo and Towne U_L/b' region. With an arbitrarily chosen parameter of 10 kDa as the minimum calculated protein molecular mass, the additional U_L/b' sequences in Toledo and Towne contained 30 and 4 open reading frames (ORFs), respectively. By eliminating ORFs that were smaller or that overlapped other ORFs by more than 60% of their lengths, these numbers could be reduced to 19 (Toledo) and 4 (Towne). These regions did not contain any consensus splice donor or acceptor sequences (24). In Fig. 3, the arrangements of predicted ORFs of Toledo and Towne, along with potential promoter element (TATAA) and poly(A) signal (AATAAA) sequences, are diagrammed. The *cis* signals identified on the complementary strand of Toledo DNA appeared to have no relevance to the identified ORFs and are not shown in Fig. 3. The nucleotide positions of the ATG and stop codon of each ORF, along with the number of amino acids, are tabulated in Table 1. We extended the ORF nomenclature of AD169 and denoted the new ORFs of Toledo UL133 to UL151. The new ORFs in Towne were denoted UL147 (because of its homology to the corresponding ORF in Toledo) and UL152 to UL154. As predicted by nucleotide sequence comparisons, UL130 and UL132 were present in all three strains, although they were arranged in an inverted orientation in the Toledo genome. UL131 as reported by Chee et al. (6) is not shown in Fig. 3, because it falls below the minimum size criteria established for inclusion as an ORF in our analysis.

TABLE 1. Features ORFs in the U_L/b' region of Toledo and Towne[a]

Virus	ORF	Strand[b]	Position[c] of:		No. of amino acids	Comment[d]
			Start	Stop		
Toledo	UL133		51	824	258	M(1)
	UL134	C	1068	541	176	
	UL135		*941	1927	329	G(2), P(1), M(4)
	UL136		2018	2740	241	P(1), M(4)
	UL137	C	2890	2600	97	P(2), M(1)
	UL138		2823	3332	170	P(3), M(1)
	UL139		3895	4302	136	G(5), M(3)
	UL140		4484	4828	115	G(2), M(1)
	UL141		5098	6375	426	G(3), P(4), M(4)
	UL142		6448	7368	307	G(20), M(5)
	UL143		7353	7631	93	G(2), M(1)
	UL144		8008	8538	177	G(7), M(3), cysteine block motif(1)
	UL145		8867	9169	101	
	UL146		9450	9803	118	G(3), P(1), M(1)
	UL147		9868	10347	160	
	UL148		10646	11596	317	G(2), P(1), M(2), RGD motif(1)
	UL132		11673	12485	271	G(4), P(2), M(4)
	UL130		13109	13753	215	G(3)
	UL149		15756	16124	123	P(1)
	UL150	C	*17802	15874	643	G(1), P(5), M(8)
	UL151	C	18299	17289	337	G(2), P(2), M(17)
Towne	UL147	C	1321	842	160	
	UL152	C	1721	1365	119	M(1)
	UL153	C	3337	2501	279	G(10), M(4)
	UL154	C	4711	3512	400	G(21), P(1), M(3)

[a] The +1 nucleotide position in Toledo is assigned to the first base after the place where the sequence similarity breaks between Toledo and AD169 (corresponding to position 175068 of AD169). The +1 nucleotide position in Towne is assigned to the first base after the place where the sequence similarity breaks between Towne and AD169 (corresponding to position 178221 of AD169).

[b] C refers to the complementary strand.

[c] Start is the first base in the ATG codon. Stop is the third base in the stop codon. Asterisks denote the initiator codons whose context sequences do not resemble RNNATG or NNNATGG (18).

[d] G denotes N-linked glycosylation site. P denotes phosphorylation site. M denotes myristoylation site. The numbers of sites are enclosed in parentheses.

Parenthetically, our sequence analysis of Toledo revealed a nucleotide change with respect to the AD169 sequence at this locus, resulting in a 14-amino-acid addition to the C terminus of the previously reported ORF131.

Each ORF was used to search the current nonredundant NCBI databases with BLAST (2). Also searched were a recently published sequence of human herpesvirus 6 (HHV-6) A variant (11) and unpublished DNA sequences of murine CMV (MCMV) and HHV-6 B variant (32, 49). In most searches, no striking homologs were identified, except that UL132 is 76% identical to a previously described mtrIII region of Towne (43). In addition, UL153 and UL154 show regions of homology to IRL14 and IRL12 of AD169, respectively. The presence of motifs in the amino acid sequence of each ORF was detected by searching the PROSITE database (3). As listed in Table 1, most of the ORFs contained various numbers of posttranslational modification sites: N-linked glycosylation, phosphorylation, and N-myristoylation. As judged by potential N-linked glycosylation sites, 11 and 2 new glycoproteins may exist in the Toledo and Towne sequences, respectively. Also found were a cysteine block motif in UL144 that is often present in growth factor receptors and an RGD motif in UL148 that may be implicated in cell attachment (Table 1).

Presence of the Toledo U_L/b' sequence in other clinical isolates. Southern analysis with radioactively labeled plasmid C4500 (positions 4665 to 9327; see Table 1 for nucleotide numbering) to probe restriction enzyme-digested viral DNAs was performed to determine the prevalence of the Toledo U_L/b' sequence in other clinical isolates. Viral DNAs of five clinical isolates (HCMVF, C128, C354, C793, and C980), Toledo, Towne, and AD169 were digested with either *Bam*HI or *Xba*I and probed with [α-^{32}P]C4500. As shown in Fig. 4, C4500 hybridized to a 4.5- to 4.7-kbp fragment in *Bam*HI-digested DNAs of Toledo and each clinical isolate. C4500 probe did not hybridize to AD169 and hybridized only weakly to Towne. This was consistent with the nucleotide sequence comparisons, which showed a small region (327 bp) of identity between Towne *Bam*HI-U (16) and Toledo C4500 (Fig. 2). C4500 hybridized to high-molecular-mass (>15-kbp) fragments of *Xba*I-digested DNA of Toledo and each clinical isolate but did not hybridize to Towne and AD169 DNA at the same area of the blot (Fig. 4). The existence of multiple bands hybridizing to the C4500 probe on the *Xba*I digest is indicative of linkage of these sequences to junction fragments in these isolates. As predicted from the nucleotide sequence data, the probe hybridized to the *Xba*I Q fragment (~4.2 kbp) of Towne DNA (Fig. 4, lane 10). The probe also recognized a smaller band (~1.8 kbp) of Toledo and the clinical isolates that comigrated with the *Xba*I T fragment of Towne (Fig. 4, lanes 9, 10, and 12 through 16). These results establish that the region of Toledo genomic DNA represented by C4500 is present in each of the five clinical isolates. Additional probes at opposite sides of the C4500 (Srf/Bam, positions 1490 to 3273, and Pst/Pst, positions 16011 to 17185 [see Table 1 for nucleotide numbering]) and additional clinical isolates were similarly analyzed, confirming the presence of Toledo U_L/b' sequence in all clinical isolates examined (blots not shown).

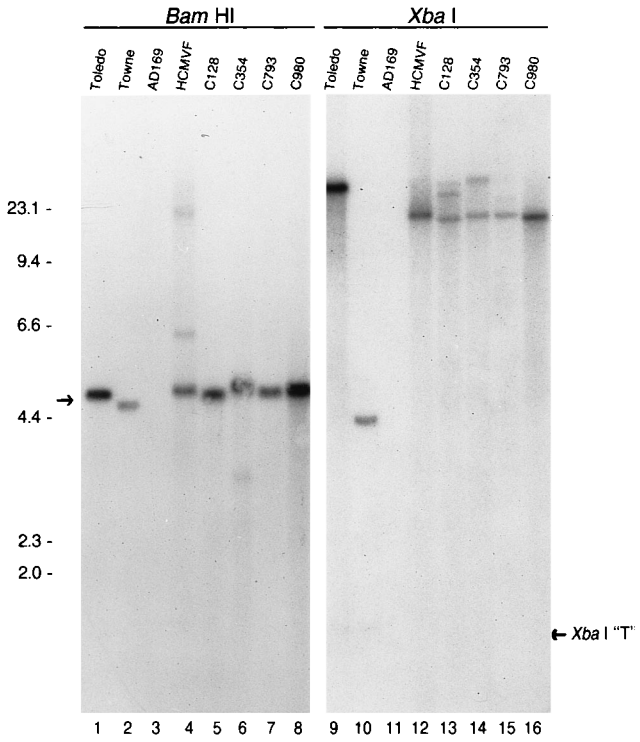

FIG. 4. Southern hybridization analysis of HCMV strains. Purified viral DNAs from Toledo (lanes 1 and 9), Towne (lanes 2 and 10), AD169 (lanes 3 and 11), and five clinical isolates (lanes 4 to 8 and 12 to 16) were digested with either *Bam*HI or *Xba*I, electrophoresed through agarose, transferred to a Hybond-N+ nylon membrane, and probed with nick-translated [α-^{32}P]C4500 plasmid DNA, as described in Materials and Methods. The arrow on the left indicates the position of the ~4.6-kbp *Bam*HI fragment (C4500) of Toledo DNA. The arrow on the right marks the position of the ~1.8-kbp *Xba*I T fragment of Towne DNA. DNA molecular size markers (in kilobases) are indicated.

DISCUSSION

In this study, we report the existence of a large DNA sequence difference between clinical isolates and laboratory strains of HCMV. Depending on the laboratory strain compared, our data reveal that differences of 13 to 15 kbp of DNA exist between clinical isolates and laboratory strains which have been extensively passaged in cell culture. These findings extend our understanding of the sequence complexity that defines a wild-type HCMV genome and provide additional insights into the evolution of complex herpesvirus genomes.

The significance of the identification of such a large segment of DNA is strengthened by its presence in all clinical isolates that have been examined. An even greater significance awaits the identification of gene products. Although dispensable for growth in fibroblasts, these ORFs may play important roles in the replication, latency, or pathobiology of this virus in the human host. At present, the functions of the putative proteins encoded in this region are unknown. A search of the NCBI database failed to reveal any close homology with currently deposited protein sequences. In addition, none of the Toledo ORFs exhibits significant homology to ORFs of other betaherpesviruses. Nevertheless, several of the ORFs have the potential to encode glycoproteins that may have an important role in virus adsorption to cellular receptors in vivo. Tissue tropism could also be determined by structural proteins, because they influence virion assembly, egress of progeny virus, fusion with plasma membranes, and penetration into the cytoplasm in a cell-type-specific manner. In this regard, the CMVs are characterized in part by a particular tropism for salivary gland tissue. The *sgg1* gene, which encodes an integral membrane protein in MCMV, has been shown to play a role in the spread of virus from infected monocytes to the salivary gland (21). Although a homolog for this protein was not identified in the HCMV sequences, functionally related gene products could be encoded within the newly identified sequences. The possibility also exists that nonstructural proteins are encoded by this sequence, affecting the replication of virus in particular cell types. In cases of disseminated disease, HCMV exhibits wide organ system involvement (1). The large number of glycoproteins are likely to play a significant role in the pathogenesis of the virus and are undoubtedly important in causing the "protean clinical manifestations" exhibited by HCMV infections as first described by Weller (45, 46).

We are currently analyzing transcription patterns in the sequenced region in an initial attempt to catalog the mRNA content and kinetics of expression. The lack of a one-to-one correspondence between the presence of transcription signal sequences and an ORF (Fig. 3) indicates the possibility that an unconventional transcription unit is present. A TATAA-less gene has been observed for herpes simplex virus (7). As has been found in other regions of the viral genome, several genes may have common polyadenylation signals (23).

Our examination at the nucleotide level of one particular low-passaged strain, Toledo, revealed the presence of a sequence rearrangement relative to Towne and AD169 DNA (Fig. 2). This simple inversion is not an experimental artifact resulting from the construction of cosmids, since the same rearrangement was also observed in another independently constructed Toledo cosmid library. Initial Southern hybridization hinted at a larger sequence rearrangement, extending over the entire newly described Toledo DNA segment. For example, the *Xba*I digest (Fig. 4) showed multiple high-molecular-weight bands with the C4500 probe, which are best explained by the lack of an *Xba*I site between the probe sequence and the *a* sequence, yielding both terminal and junction segments. We have used multiple probes to analyze restriction enzyme-digested genomic DNAs of several clinical isolates in addition to those shown in Fig. 4. These analyses have suggested that the Toledo sequence from corresponding nucleotide positions 175068 to 175082 of AD169 (Fig. 2B) was flipped relative to that in all other clinical isolates examined thus far. PCR experiments with two sets of primers that would generate products across these two inversion points confirmed that the Toledo arrangement is not found in other clinical isolates. Nonetheless, regardless of orientation, the presence of this unique sequence serves as a diagnostic marker for HCMV clinical isolates. The biological significance of this sequence rearrangement is not yet clear, but it is possible that it is acquired during its passage in cell culture.

Recently, Gompels et al. (11) have renewed the discussion of relationships between strains of herpesviruses with respect to HHV-6. They point to the pattern of close similarity from strain to strain in the center of the genome with increasing divergence toward the ends. For HCMV, our comparison of the divergence in the genomes of Toledo, Towne, and AD169 reinforces the theme of differences observed at or near repeat regions of herpesvirus genomes. Intriguingly, one of the breakpoints in both Towne and Toledo DNA occurs precisely at the nucleotide identified as the U_L/b' boundary of the AD169 strain. For HCMV, examples of differences near the unique *b*-sequence boundaries have been described (20, 42, 48). Specifically, the Tanaka strain has a sequence difference of ~7.8 kbp relative to Towne near the b/U_L boundary (42). Deletions and duplications in the U_S component which have resulted in expansions in the length of the U_S component by nearly 3%

relative to the wild-type virus have also been observed (17). However, on the basis of the number of clinical isolates examined with respect to the presence of the new sequence we have described in the Toledo genome, we speculate that some genetic information may have been lost from laboratory strains during extensive passage in cell culture. This information may be critical in identifying those genes that are responsible for HCMV virulence. Therefore, to study the function of the proteins encoded in this region it will be useful to construct interstrain recombinants between Toledo and high-passaged laboratory strains. Such recombinant viruses may show enhanced replicative ability and increased immunogenicity without exhibiting the full virulence of the clinical isolates. In addition to providing new opportunities for study, these recombinant viruses may prove useful as vaccine candidates for preventing the disease associated with this clinically important human pathogen.

ACKNOWLEDGMENTS

We thank Meggen Chadsey for constructing the Toledo cosmids, Mark Stinski for providing the Towne plasmids, Bill Rawlinson and Koichi Yamanishi for sharing the MCMV and HHV-6 sequence information in advance of publication, Jackie Zhao and Ruth Winter for technical assistance, and Bernard Roizman for encouragement.

REFERENCES

1. **Alford, C. A., and W. J. Britt.** 1993. Cytomegalovirus, p. 227–255. *In* B. Roizman, R. J. Whitley, and C. Lopez (ed.), The human herpesviruses. Raven Press, New York.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bairoch, A., and P. Bucher.** 1994. PROSITE: recent developments. Nucleic Acids Res. **22:**3583–3589.
4. **Brown, J. M., H. Kaneshima, and E. S. Mocarski.** 1995. Dramatic interstrain differences in the replication of human cytomegalovirus in SCID-hu mice. J. Infect. Dis. **171:**1599–1603.
5. **Chandler, S. H., and J. K. McDougall.** 1986. Comparison of restriction site polymorphisms among clinical isolates and laboratory strains of human cytomegalovirus. J. Infect. Virol. **67:**2179–2192.
6. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison III, T. Kouzarides, J. A. Martignetti, E. Preddie, 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.
7. **Chou, J., and B. Roizman.** 1986. The terminal *a* sequence of the herpes simplex virus genome contains the promoter of a gene located in the repeat sequence of the L component. J. Virol. **57:**629–637.
8. **Chou, S.** 1992. Comparative analysis of sequence variation in gp116 and gp55 components of glycoprotein B of human cytomegalovirus. Virology **188:**388–390.
9. **Chou, S.** 1992. Molecular epidemiology of envelope glycoprotein H of human cytomegalovirus. J. Infect. Dis. **166:**604–607.
10. **Elek, S. D., and H. Stern.** 1974. Development of a vaccine against mental retardation caused by cytomegalovirus infection in utero. Lancet **i:**1–5.
11. **Gompels, U. A., J. Nicholas, G. Lawrence, M. Jones, B. J. Thomson, M. E. D. 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.
12. **Hayward, G. S., R. Ambinder, D. Ciufo, S. D. Hayward, and R. L. LaFemina.** 1984. Structural organization of human herpesvirus DNA molecules. J. Infect. Dermatol. **83**(Suppl.)**:**29–41.
13. **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.
14. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11:**291–298.
15. **Huang, E.-S., B. A. Kilpatrick, Y.-T. Huang, and J. S. Pagano.** 1976. Detection of human cytomegalovirus and analysis of strain variation. Yale J. Biol. Med. **49:**29–43.
16. **Kemble, G. W., A. L. McCormick, L. Pereira, and E. S. Mocarski.** 1987. A cytomegalovirus protein with properties of herpes simplex virus ICP8: partial purification of the polypeptide and map position of the gene. J. Virol. **61:**3143–3151.
17. **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:**5184–5199.
18. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108:**229–241.
19. **LaFemina, R. L., and G. S. Hayward.** Personal communication.
20. **LaFemina, R. L., and G. S. Hayward.** 1980. Structural organization of the DNA molecules from human cytomegalovirus, p. 39–55. *In* B. N. Fields and R. Jaenisch (ed.), Animal virus genetics. Academic Press, Inc., New York.
21. **Lagenaur, L. A., W. C. Manning, J. Vieira, 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.
22. **Lehner, R., T. Stamminger, and M. Mach.** 1991. Comparative sequence analysis of human cytomegalovirus strains. J. Clin. Microbiol. **29:**2494–2502.
23. **Mocarski, E. S. J.** 1993. Cytomegalovirus biology and replication, p. 173–226. *In* B. Roizman, R. J. Whitley, and C. Lopez (ed.), The human herpesviruses. Raven Press, New York.
24. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10:**459–472.
25. **Pachl, C., W. S. Probert, K. M. Hermsen, F. R. Masiarz, L. Rasmussen, T. C. Merigan, and R. R. Spaete.** 1989. The human cytomegalovirus strain Towne glycoprotein H gene encodes glycoprotein p86. Virology **169:**418–426.
26. **Pande, H., K. Campo, B. Tanamachi, and J. A. Zaia.** 1991. Human cytomegalovirus strain Towne pp65 gene: nucleotide sequence and expression in *Escherichia coli*. Virology **182:**220–228.
27. **Plotkin, S. A., J. Farquhar, and E. Hornberger.** 1976. Clinical trials of immunization with the Towne 125 strain of human cytomegalovirus. J. Infect. Dis. **134:**470–475.
28. **Plotkin, S. A., H. M. Friedman, G. R. Fleisher, D. C. Dafoe, R. A. Grossman, M. L. Smiley, S. E. Starr, C. Wlodaver, A. D. Friedman, and C. F. Barker.** 1984. Towne-vaccine-induced prevention of cytomegalovirus disease after renal transplants. Lancet **i:**528–530.
29. **Plotkin, S. A., and E.-S. Huang.** 1985. Cytomegalovirus vaccine virus (Towne strain) does not induce latency. J. Infect. Dis. **152:**395–397.
30. **Pritchett, R. F.** 1980. DNA nucleotide sequence heterogeneity between the Towne and AD169 strains of cytomegalovirus. J. Virol. **36:**152–161.
31. **Quinnan, G. V., M. Delery, A. H. Rook, W. R. Frederick, J. S. Epstein, J. F. Manischewitz, L. Jackson, K. M. Ramsey, K. Mittal, S. A. Plotkin, and M. R. Hilleman.** 1984. Comparative virulence and immunogenicity of the Towne strain and a nonattenuated strain of cytomegalovirus. Ann. Intern. Med. **101:**478–483.
32. **Rawlinson, B.** Personal communication.
33. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:**237–251.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
36. **Spaete, R. R., and N. Frenkel.** 1982. The herpes simplex virus amplicon: a new eucaryotic defective-virus cloning-amplifying vector. Cell **30:**295–304.
37. **Spaete, R. R., and E. S. Mocarski.** Unpublished data.
38. **Spaete, R. R., and E. S. Mocarski.** 1985. Regulation of cytomegalovirus gene expression: α and β promoters are *trans*-activated by viral functions in permissive human fibroblasts. J. Virol. **56:**135–143.
39. **Spaete, R. R., and E. S. Mocarski.** 1985. The *a* sequence of the cytomegalovirus genome functions as a cleavage/packaging signal for herpes simplex virus defective genomes. J. Virol. **54:**817–824.
40. **Spaete, R. R., R. M. Thayer, W. S. Probert, F. R. Masiarz, S. H. Chamberlain, L. Rasmussen, T. C. Merigan, and C. Pachl.** 1988. Human cytomegalovirus strain Towne glycoprotein B is processed by proteolytic cleavage. Virology **167:**207–225.
41. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate-early gene of human cytomegalovirus. J. Virol. **49:**190–199.
42. **Takekoshi, M., S. Ihara, S. Tanaka, F. Maeda-Takekoshi, and Y. Watanabe.** 1987. A new human cytomegalovirus has an invertable sub-segment within its L component producing eight genome isomers. J. Gen. Virol. **68:**765–776.
43. **Thompson, J., A. Inamdar, N. Jahan, J. Doniger, and L. J. Rosenthal.** 1993. Localization and sequence analysis of morphological transforming region III within human cytomegalovirus strain Towne. Intervirology **36:**121–127.
44. **Thomsen, D. R., and M. F. Stinski.** 1981. Cloning of the human cytomegalovirus genome as endonuclease *Xba* I fragments. Gene **16:**207–216.
45. **Weller, T. H.** 1971. The cytomegaloviruses: ubiquitous agents with protean clinical manifestations. I. N. Engl. J. Med. **285:**203–214.
46. **Weller, T. H.** 1971. The cytomegaloviruses: ubiquitous agents with protean clinical manifestations. II. N. Engl. J. Med. **285:**267–274.
47. **Weststrate, M. W., J. L. Geelen, and J. van der Noordaa.** 1980. Human cytomegalovirus DNA: physical maps for restriction endonucleases *Bgl* II, *Hin*d III and *Xba* I. J. Gen. Virol. **49:**1–21.
48. **Weststrate, M. W., J. L. Geelen, P. M. E. Wertheim, and J. van der Noordaa.** 1983. Comparison of the physical maps of the DNAs of two cytomegalovirus strains. J. Gen. Virol. **64:**47–55.
49. **Yamanishi, K.** Personal communication.