JOURNAL OF VIROLOGY, Dec. 1997, p. 9833–9836 Vol. 71, No. 12
0022-538X/97/$04.00+0

The Published DNA Sequence of Human Cytomegalovirus Strain AD169 Lacks 929 Base Pairs Affecting Genes UL42 and UL43

D. J. DARGAN,* F. E. JAMIESON, J. MACLEAN, A. DOLAN, C. ADDISON, AND D. J. MCGEOCH

MRC Virology Unit, University of Glasgow, Glasgow G11 5JR, United Kingdom

Received 19 June 1997/Accepted 2 September 1997

Compared with the published DNA sequence (M. S. Chee, et al. Curr. Top. Microbiol. Immunol. 154:125–170, 1990), most isolates of human cytomegalovirus strain AD169 contain an additional 929 bp after nucleotide 54612. This results in a changed reading frame for the 5′-terminal 50 codons of gene UL42 and expansion of gene UL43 (a US22 family member) from 187 (3′-truncated) to 423 (full-length) codons. The UL42 and UL43 gene products are nonessential for growth in culture.

The published DNA sequence of the AD169 strain of human cytomegalovirus (HCMV) (4) is a key resource for molecular studies with HCMV. Here, we report a 929-bp segment of AD169 DNA absent from the published sequence. This sequence is unrelated to the 15-kbp fragment of HCMV DNA present in clinical isolates and the Toledo strain of HCMV but not in AD169 (3). Unlike this 15-kbp deletion, the 929-bp fragment is present in most stocks of AD169, and an adjustment to the published DNA sequence is therefore proposed.

Detection and mapping of a novel DNA sequence in the HCMV AD169 genome. The following four distinct stocks of HCMV AD169 were obtained: AD169 (ATCC) from the American Type Culture Collection (ATCC), AD169 (London) from J. Booth (St. George's Hospital Medical School, University of London, United Kingdom), AD169 (Cambridge) from H. Browne (Division of Virology, Department of Pathology, University of Cambridge, United Kingdom), and AD169 (Glasgow) from J. Macnab of this unit, who received it from J. Oram (Genetic Manipulation Laboratory, Centre for Applied Microbiology and Research, Porton Down, United Kingdom). The Towne and Davis strains of HCMV were obtained from the ATCC. Virus stocks were propagated in HFFF-2 cells, and DNA was extracted from virions pelleted from cell supernatant medium.

Digestion of AD169 (Cambridge) DNA produced a 10.3-kbp *Hin*dIII M fragment (17) in agreement with the size expected from the published sequence (4), while AD169 (Glasgow) gave a novel 11.4-kbp (M′) fragment that comigrated with *Hin*dIII L (11.4 kbp) (Fig. 1A; compare M with P in lane 2 and M′ and L with P in lane 3). This variability in the size of the *Hin*dIII M fragment has also been observed by others (2, 17).

An AD169 (Glasgow) cosmid (Cos 15) containing *Hin*dIII M′ was prepared as described elsewhere (6). Figure 2A shows restriction site cleavage maps for Cos 15 and the corresponding region of the sequenced virus genome (12, 20). *Bam*HI digestion of Cos 15 (Fig. 1B) revealed that the 14.3-kbp *Bam*HI A fragment was absent while two novel *Bam*HI fragments of approximately 12.8 and 2.5 kbp were generated. All other fragments were of the expected sizes. Because *Hin*dIII M is located entirely within *Bam*HI A (Fig. 2A), it is evident that the additional sequence in *Hin*dIII M′ contains a *Bam*HI site allowing cleavage of *Bam*HI A into 2.5 (A′)- and 12.8 (A″)-kbp components.

*Eco*RI digestion of Cos 15 (Fig. 1B) revealed that the 1.9-kbp *Eco*RI c fragment was absent while a novel fragment of 2.8 kbp was generated. All other fragments were of the expected sizes. The 1.9-kbp (*Eco*RI c) fragment is located within *Bam*HI A, with most of the sequence also contained within the left end of *Hin*dIII M (Fig. 2A). Thus, the additional sequence in AD169 (Glasgow) is contained entirely within the novel 2.8-kbp (*Eco*RI c′) fragment (Fig. 2A).

The novel DNA sequence is present in other stocks of AD169 and in other strains of HCMV. *Eco*RI fragments from AD169 (London), AD169 (Cambridge), AD169 (Glasgow), and AD169 (ATCC) and from the Towne and Davis strains of HCMV were investigated by Southern blot hybridization at 65°C (19) with an [α-^{32}P]dCTP-labeled *Eco*RI c′ fragment as the probe (8) (Fig. 1C).

The *Eco*RI c′ probe hybridized only to 2.8-kbp fragments from AD169 (Glasgow), AD169 (ATCC), and Towne and Davis virus DNAs, indicating that the novel sequence was of viral origin. The probe hybridized to fragments of 2.8 and 1.9-kbp from AD169 (Cambridge) and AD169 (London) virus DNAs suggesting that these contained a mixed population of viruses having the *Eco*RI c and *Eco*RI c′ genotypes. Since the mixed-population AD169 (London) stock (fourth passage from vaccine [1]) is related by linear passage history (1) to the stock from which the *Hin*dIII fragments were cloned (17), it is likely that this stock too contained a mixed population. Evidently, the *Hin*dIII M fragment was cloned (17) from a virus with a deletion mutation and sequenced (4).

Determination and interpretation of the novel sequence in HCMV *Eco*RI c′. The complete sequence of the *Eco*RI c′ fragment was determined by using M13mp18 clones with the insert in each orientation and specific oligonucleotide primers. The sequence was compiled with the programs of Staden (21). Interpretation used the Genetics Computer Group program set (9). The *Eco*RI c′ sequence was identical to that of *Eco*RI c as determined by Chee et al. (4), except that it contained an additional internal section of 929 bp following nucleotide 54612 (Fig. 2B and C). This novel sequence was completely determined for both strands.

The 929-bp sequence had a base composition of 65.0% G+C and contained the expected *Bam*HI site. Nearest neighbor dinucleotide frequencies were close to those of bulk HCMV DNA, with no CpG shortage. Within *Eco*RI c′, the novel

* Corresponding author. Phone: 44 141 339 8855 ext. 4641. Fax: 44 141 330 3520. E-mail: d.dargan@vir.gla.ac.uk.

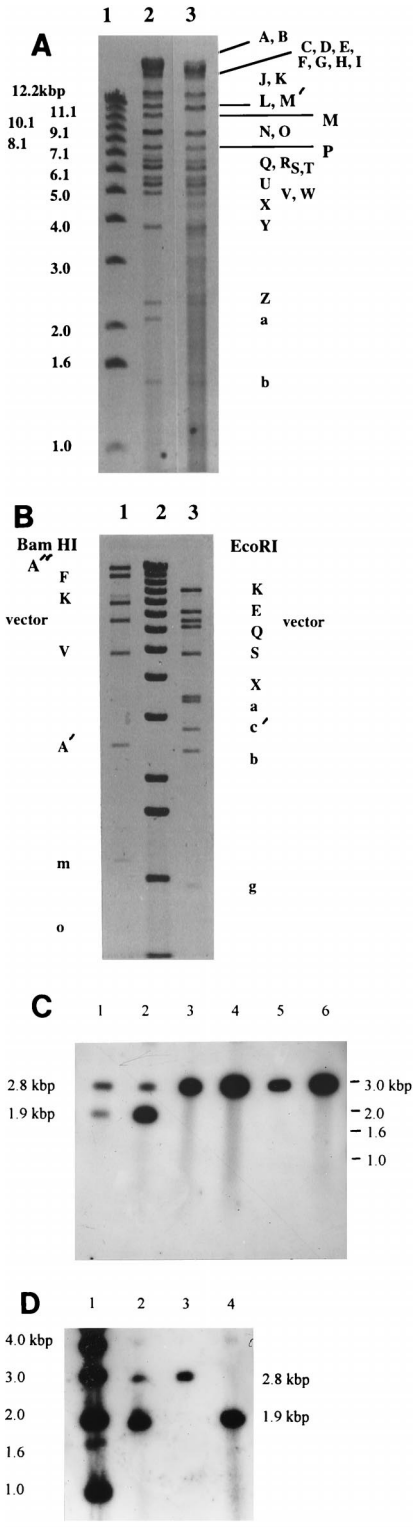

FIG. 1. (A) Electrophoresis of *Hin*dIII DNA fragments obtained from two stocks of HCMV AD169. The HCMV *Hin*dIII fragments (17) resolved in a 0.6% agarose gel are identified. Lanes: 1, 1-kbp DNA ladder (Gibco-BRL) (DNA fragment sizes are shown to the left); 2, AD169 (Cambridge); and 3, AD169 (Glasgow). (B) Electrophoresis of *Bam*HI and *Eco*RI fragments generated by digestion of Cos 15 DNA. The Cos 15 HCMV DNA *Bam*HI and *Eco*RI fragments separated on a 0.6% agarose gel were identified (12), except for *Eco*RI i (184 bp), which was not resolved. Lanes: 1, Cos 15 *Bam*HI fragments; 2, 1-kbp DNA ladder (DNA fragment sizes are as indicated in panel A); and 3, Cos 15 *Eco*RI fragments. The 6.5-kbp cosmid vector fragment is liberated upon digestion with either enzyme. (C) Investigation of AD169 stocks and other strains of HCMV for the presence of the novel DNA sequence contained in AD169 (Glasgow); Southern blot showing hybridization of probe (*Eco*RI c′) to HCMV *Eco*RI DNA fragments. Lanes: 1, AD169 (London); 2, AD169 (Cambridge); 3, AD169 (Glasgow); 4, AD169 (ATCC); 5, HCMV strain Towne; and 6, HCMV strain Davis. DNA size marker locations are indicated to the right. The 2.8-kbp (*Eco*RI c′) and 1.9-kbp (*Eco*RI c) fragments are indicated. (D) Purification of WT (*Eco*RI c′) and ΔUL42/UL43 (*Eco*RI c) viruses from the mixed population AD169 (Cambridge) virus stock; hybridization of *Eco*RI c′ to HCMV *Eco*RI fragments. Lanes: 1, 1-kbp DNA ladder (DNA fragment sizes are to the left); 2, AD169 (Cambridge); 3, WT virus; and 4, ΔUL42/UL43 virus.

sequence was flanked by a tetranucleotide repeated sequence (5′-GCAG), with only one copy present in the shorter, *Eco*RI c version.

The deletion point lies within the coding region of gene UL43 (4). In the deleted version, this open reading frame (ORF) terminates three codons downstream of the deletion point giving an ORF length of 187 codons, while in the nondeleted version the ORF is extended to 423 codons and terminates within the novel sequence (Fig. 2B and C). UL43 was previously identified as an incomplete member of the US22 multigene family on the basis of its encoding N-proximal motifs from the set of amino acid sequence motifs conserved in the family (4, 7, 14) and is now seen to be a complete member of the family, with further conserved motifs encoded within the deletion region (Fig. 2C).

Codon preference tests suggest that UL42 is also oriented leftward, with its coding region starting within the novel sequence, separated by a short noncoding sequence from the termination codon of UL43 and then running across the left boundary of the novel sequence (Fig. 2B and C). From codon 50, the revised UL42 ORF is equivalent to the distal portion of gene UL42 as proposed by Chee et al. (4), but the 5′-terminal coding sequences have a different reading frame. The only distinctive feature of the amino acid sequence of revised UL42 is an uncharged and highly hydrophobic segment of 25 residues proximal to the C terminus, which might form a transmembrane segment.

The coding content of the novel 929-bp sequence of HCMV AD169 thus represents the upstream portion of gene UL42 and the adjacent downstream portion of gene UL43. As a comment in passing, we consider that on the basis of codon usage pattern, the leftmost functional ORF in *Eco*RI c is likely to be the leftward ORF (UL41alt) rather than the larger leftward ORF UL41 proposed by Chee et al. (4) in the same region (Fig. 2B; Table 1).

We compared our revised gene content for *Eco*RI c′ of HCMV with the corresponding regions in murine cytomegalovirus (MCMV), human herpesvirus types 6 (HHV-6) and 7 (HHV-7). All have convincing homologs of the revised UL43 gene (11, 16, 18). Optimized FastA scores for the revised HCMV UL43 protein sequence compared with those of the homologs were as follows: MCMV UL43, 349; HHV-6 U25, 268; and HHV-7 U25, 285. The other viruses do not have homologs that show extended amino acid sequence similarities to HCMV UL42, but each has a correspondingly located ORF which encodes, like revised UL42, a short protein sequence with a similar C-proximal uncharged and hydrophobic segment; thus, these genes probably constitute a homologous but strongly diverged set. In the region corresponding to sequences lying between HCMV genes UL38 and UL42, there appears to be no homologous set of genes common to all four available genome sequences.

Because the novel 929-bp sequence is present in most isolates of HCMV AD169, it is appropriate that a modified version of the complete genome sequence for HCMV AD169

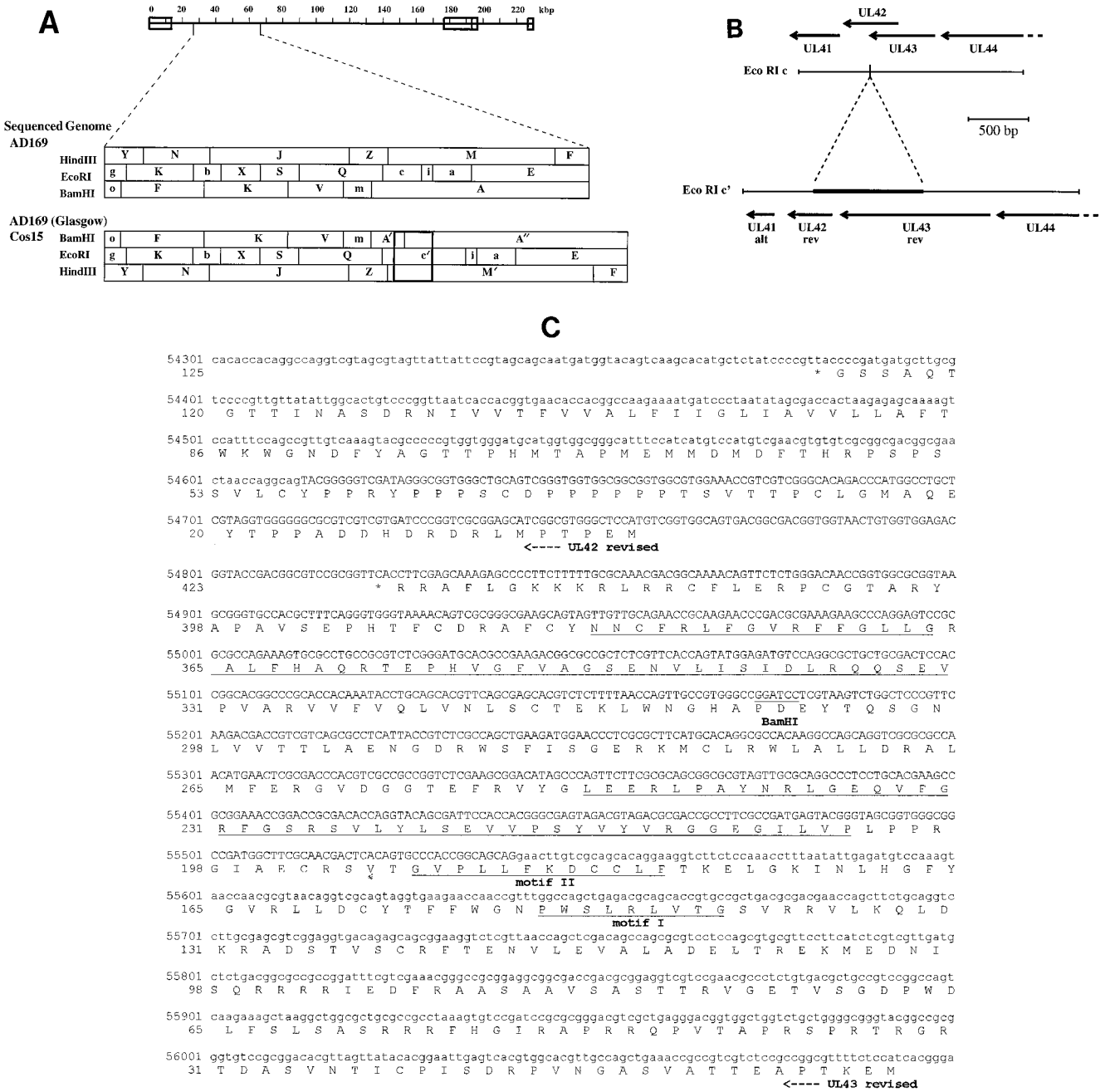

FIG. 2. (A) Mapping the location of the novel DNA sequences in AD169 (Glasgow). The upper part represents the HCMV genome with long and short unique DNA sequences flanked by terminal and internal repeat sequences (indicated by the large and small rectangles). Cos 15 contains HCMV DNA sequences located between the *Bam*HI sites at nucleotides 26191 and 66934, indicated here by the expanded region. The *Bam*HI, *Eco*RI, and *Hin*dIII cleavage sites contained within this DNA fragment for both the AD169 (Glasgow) Cos 15 DNA sequence and the corresponding region of the sequenced virus genome are shown. The location of the additional DNA sequences in the AD169 (Glasgow) genome is indicated by the box outlined in heavy lines. (B) Genome organization of the AD169 *Eco*RI c and c′ fragments. Comparison of proposed gene contents of the *Eco*RI c and c′ fragments. The upper part represents *Eco*RI c with ORFs UL41, UL42, and UL43 and part of UL44, as proposed by Chee et al. (4). The lower part represents *Eco*RI c′, with the novel 929-bp sequence indicated by a heavy line. Revised (rev) versions of ORFs UL42 and UL43, along with the alternative (alt) UL41 ORF, are shown. (C) Listing of the novel 929-bp sequence in *Eco*RI c′. The DNA sequence shown represents nucleotides 54301 to 56100 of the rightward 5′-to-3′ strand in the HCMV AD169 genome sequence as obtained after insertion of the 929-nucleotide sequence into the version described by Chee et al. (4). Uppercase, 929-nucleotide sequence; lowercase, adjacent parts of *Eco*RI c′. Proposed encoded amino acid sequences of revised genes UL42 and UL43 are shown. US22 gene family amino acid motifs I (7) and II (4, 14) in UL43 are indicated, while regions of hydrophobic and charged amino acid residues are underlined. The *Bam*HI site predicted from restriction mapping studies is also indicated.

should include this section. The HCMV AD169 sequence thus revised has a total genome length of 230,283 bp. A revised listing for HCMV AD169 genes UL41, UL42, and UL43 is shown in Table 1.

Purification and one-step virus growth kinetics of the deletion mutant. Mutant (ΔUL42/UL43) and wild-type (WT) viruses were isolated from the AD169 (Cambridge) stock and cloned by three rounds of plaque purification, and their ho-

TABLE 1. Relationship of ORFs contained within the *Eco*RI c′ DNA fragment

ORF[a]	Nucleotide position	
	Start	Stop
UL41 (Chee et al. [4])	54358	53936
UL41alt	54276	54043
UL42 revised	54758	54384
UL43 revised	56093	54825

[a] All genes were leftward oriented. Sequence numbering was obtained by insertion of a novel 929-bp segment after position 54612 in the published sequence.

mogeneity was confirmed by Southern blot hybridization with the *Eco*RI c′ probe (Fig. 1D). One-step growth curves (Fig. 3) show that WT and ΔUL42/UL43 virus replication kinetics are similar, indicating that the UL42 and UL43 gene products are nonessential for growth in cultured HFFF-2 cells.

In the ΔUL42/UL43 mutant, expression of the UL42 gene is precluded due to deletion of 5′-terminal coding and upstream sequences. Partial deletion of UL43 coding sequences could yield a 3′-truncated product of 187 amino acids (4) (28% of the full-length sequence); however, we consider it unlikely that this would be biologically active, since it lacks both a complete copy of US22 motif 2 (4) and the downstream regions of hydrophobic and charged residues present in UL43 homologs (11, 16, 18). While the function provided by HCMV UL43 has yet to be elucidated, US22 family genes generally encode transcriptional activators (5, 10, 13, 22), although these functions may be nonessential for growth in culture (15, 23).

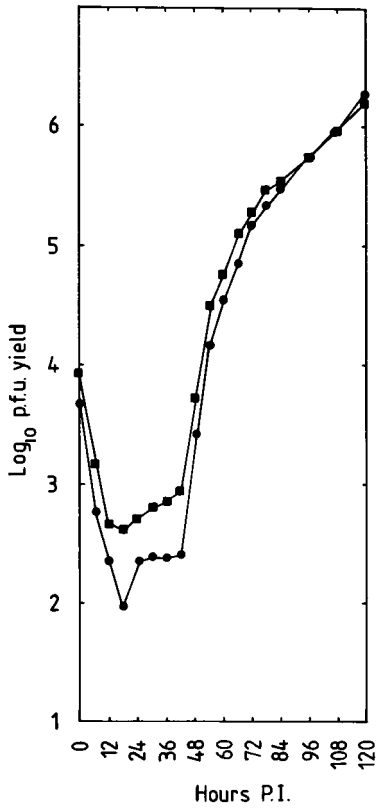

FIG. 3. One-step virus growth curve obtained for the WT and deletion mutant viruses. HFFF-2 cells were infected at a multiplicity of infection of 5 PFU/cell with AD169 WT virus (●) or ΔUL42/UL43 mutant virus (■). P.I., postinfection.

Nucleotide sequence accession number. The sequence of *Eco*RI c′ has been deposited with the EMBL data library under accession no. Y13735.

It is our pleasure to thank J. Oram for supplying the HCMV *Hin*dIII clones and J. Booth, H. Brown, and J. Macnab, who supplied the AD169 (London), AD169 (Cambridge), and AD169 (Glasgow) viruses, respectively. We also thank A. Davison and C. Cunningham for advice and assistance in the preparation of HCMV cosmids. Our thanks are also due to A. Davison for critical reading of the manuscript.

REFERENCES

1. **Booth, J.** Personal communication.
2. **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.
3. **Cha, T.-A., E. Tom, G. Kemble, G. M. Duke, E. S. Mocarski, and R. R. Spaete.** 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J. Virol. **70:**78–83.
4. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchinson, 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 the human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**125–170.
5. **Colberg-Poley, A. M., L. D. Santomenna, P. P. Harlow, P. A. Benfield, and D. J. Tenney.** 1992. Human cytomegalovirus US3 and UL36-38 immediate early proteins regulate gene expression. J. Virol. **66:**95–105.
6. **Cunningham, C., and A. J. Davison.** 1993. A cosmid based system for constructing mutants of herpes simplex virus type 1. Virology **197:**116–124.
7. **Efstathiou, S., G. L. Lawrence, C. M. Brown, and B. C. Barrell.** 1992. Identification of the human cytomegalovirus US22 gene family in human herpesvirus 6. J. Gen. Virol. **73:**1661–1671.
8. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
9. **Genetics Computer Group.** 1991. Program manual for the GCG package, version 7, April 1991. Genetics Computer Group, Madison, Wis.
10. **Geng, Y., B. Chandran, F. Josephs, and C. Wood.** 1992. Identification and characterization of a human herpesvirus 6 gene segment that transactivates the human immunodeficiency virus type 1 promoter. J. Virol. **66:**1564–1570.
11. **Gompels, U. A., J. Nicholas, G. Lawrence, M. Jones, B. J. Thompson, M. E. D. Martin, S. Efstathiou, M. Craxton, and H. A. Macauley.** 1995. The DNA sequence of human herpesvirus-6: structure, coding content and evolution. Virology **209:**29–51.
12. **Greenaway, P. J., J. D. Oram, R. G. Downing, and K. Patel.** 1982. Human cytomegalovirus DNA: BamHI, EcoRI and PstI restriction endonuclease cleavage maps. Gene **18:**355–360.
13. **Kashanchi, F., J. Thompson, M. R. Sadaie, J. Doniger, J. Duvall, J. N. Brady, and L. J. Rosenthal.** 1994. Transcriptional activation of minimal HIV-1 promoter by ORF-1 protein expressed from the Sal1 L fragment of human herpesvirus 6. Virology **201:**95–106.
14. **Kouzarides, T., A. T. Bankier, S. C. Satchwell, E. Preddy, and B. G. Barrell.** 1988. An immediate-early gene of human cytomegalovirus encodes a potential membrane glycoprotein. Virology **165:**151–164.
15. **Manning, W. C., and E. S. Mocarski.** 1988. Insertional mutagenesis of the murine cytomegalovirus genome. One prominent α gene (ie2) is dispensable for growth. Virology **167:**477–484.
16. **Nicholas, J.** 1996. Determination and analysis of the complete nucleotide sequence of human herpesvirus type 7. J. Virol. **70:**5975–5989.
17. **Oram, J. D., R. G. Downing, A. Akrigg, A. A. Dollery, C. J. Duggleby, G. W. G. Wilkinson, and P. J. Greenaway.** 1982. Use of recombinant plasmids to investigate the structure of the human cytomegalovirus genome. J. Gen. Virol. **59:**111–129.
18. **Rawlinson, W. D., H. E. Farrell, and B. G. Barrell.** 1996. Analysis of the complete DNA sequence of murine cytomegalovirus. J. Virol. **70:**8833–8849.
19. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
20. **Spector, D. H., L. Hock, and J. C. Tamashiro.** 1982. Cleavage maps for human cytomegalovirus DNA strain AD169 for restriction endonucleases *Eco*RI, *Bgl*II, and *Hin*dIII. J. Virol. **42:**558–582.
21. **Staden, R.** 1987. Computer handling of DNA sequencing projects, p. 173–217. *In* M. J. Bishop and C. J. Rawlings (ed.), Nucleic acid and protein sequence analysis: a practical approach. IRL Press, Oxford, United Kingdom.
22. **Stasiak, P. C., and E. S. Mocarski.** 1992. Transactivation of the cytomegalovirus ICP36 gene promoter requires the α gene product TRS1 in addition to IE1 and IE2. J. Virol. **66:**1050–1058.
23. **Vieira, J., H. E. Farrell, W. D. Rawlinson, and E. S. Mocarski.** 1994. Genes in the *Hin*dIII J fragment of the murine cytomegalovirus genome are dispensable for growth in cultured cells: insertion mutagenesis with a *lacZ*/*gpt* cassette. J. Virol. **68:**4837–4846.