JOURNAL OF VIROLOGY, Nov. 1984, p. 541–548
0022-538X/84/110541-08$02.00/0
Copyright © 1984, American Society for Microbiology
Vol. 52, No. 2

Structure of the Heterogeneous L-S Junction Region of Human Cytomegalovirus Strain AD169 DNA

JOYCE C. TAMASHIRO,[1] DAVID FILPULA,[2]† THEODORE FRIEDMANN,[2] AND DEBORAH H. SPECTOR[1]*
Departments of Biology[1] and Pediatrics,[2] University of California, San Diego, La Jolla, California 92093

Received 12 April 1984/Accepted 18 July 1984

The genome of human cytomegalovirus strain AD169 contains a region of heterogeneity located at the junction between the long (L) and short (S) components of the viral DNA. Twelve cloned L-S junction fragments were studied by using the restriction enzymes *Hae*II and *Xho*I. The region of heterogeneity was localized within a single *Hae*II restriction fragment. The enzyme *Xho*I was used to subdivide this region and revealed the presence of three types of heterogeneity within the junction fragments. Each of the cloned junction fragments contained one of the following fragments: 0.553, 0.95, or 1.35 kilobase pairs (referred to as class I heterogeneity). Class II heterogeneity was defined as the presence of tandem duplications of class I fragments. In addition, a variable number (0 to 5) of a 0.2-kbp fragment (class III heterogeneity) was observed. Mapping of these fragments with partial *Xho*I digestions revealed that the class I and class III heterogeneous fragments were adjacent. The DNA sequence of the smallest cloned L-S junction fragment was determined and analyzed. This junction fragment contained a single 0.553-kbp *Xho*I fragment and no copies of the 0.2-kbp fragment. The 0.553-kbp *Xho*I fragment was similar in structure to the *a*-sequences of herpes simplex virus types 1 and 2. In addition, a region of homology was found between the *a* sequences of herpes simplex virus types 1 and 2 and the 0.553-kbp *Xho*I fragment from the human cytomegalovirus junction.

Human cytomegalovirus (HCMV), a member of the herpesvirus group, has a double-stranded DNA genome approximately 240 kilobase pairs (kbp) in length (5, 7, 12, 13, 28, 31). The genome consists of two components, termed L (long) and S (short), which are composed of unique sequences (U_S or U_L) bounded by inverted repeats (6, 12, 23, 28, 35). In the AD169 strain, the repeats bordering the U_L segment are approximately 11 to 12 kbp, whereas those flanking the U_S component are approximately 2 kbp (28). The region where the L and S components meet is termed the L-S junction, and there is evidence that the L and S segments can invert relative to the junction point so that four genome arrangements are possible (4, 6, 11, 12, 23, 28).

Heterogeneity has been observed at the L-S junction and termini of HCMV. In strain AD169, terminal fragment *Eco*RI-W from the long repeat (see Fig. 1 for *Eco*RI restriction map) varies by increments of approximately 0.2 kbp; terminal fragments *Eco*RI-N and -L from the short segment vary by approximately 0.6 kbp (28). The Towne and Davis strains of HCMV contain heterogeneity at the S terminus, but heterogeneity at the L terminus has only been observed in 10 to 20% of the DNA molecules (4, 12; G. S. Hayward, personal communication).

We have cloned a set of heterogeneous *Eco*RI-WL (joint F) and -WN (joint H) junction fragments from virion DNA (32). Due to the inversion properties of the genome, these junction fragments should contain the heterogeneous regions from both the L and S termini. Cleavage of these recombinant plasmids with various restriction enzymes indicated that the region of heterogeneity was contained within a single *Bam*HI-*Pvu*II fragment. This fragment mapped at the L-S junction and, except for the heterogeneity, was identical for both junction fragments *Eco*RI-WL and -WN.

In this paper we used restriction endonucleases to characterize the nature of the heterogeneity. The endonuclease *Xho*I, which cleaved within the region of heterogeneity, was used to characterize the size and distribution of heterogeneous fragments within the cloned junctions. We also determined the nucleotide sequence of one of the cloned L-S junction fragments.

MATERIALS AND METHODS

Construction of recombinant plasmids. Recombinant plasmids, containing *Eco*RI fragments of HCMV (strain AD169) DNA and the plasmid vector pACYC184, were constructed and isolated as previously described (32). Subcloning of the heterogeneous region of the junction fragments was accomplished with *Pvu*II-*Eco*RI subfragments of *Eco*RI junction fragments WN and WL (28) and the 3.8-kbp *Pvu*II-*Eco*RI fragment of pACYC184. Both the junction subfragments and the plasmid were gel purified with 0.8% Seaplaque agarose (Marine Colloids) gels and benzoyl-napthol-DEAE (BND)-cellulose chromatography (32). Ligation and transfection conditions were identical to those described previously for cloning of the *Eco*RI viral DNA fragments (32), except that the plasmid was not treated with bacterial alkaline phosphatase. Recombinant plasmids were screened by a rapid plasmid isolation procedure and gel analysis (8). Subclones containing selected *Xho*I fragments were generated by ligation of gel-purified fragments into the *Xho*I site of the plasmid vector pMK16 (9). The vector was treated with calf intestinal phosphatase (Boehringer Mannheim Biochemicals) before ligation. Tetracycline-resistant, kanamycin-sensitive colonies were screened as described above.

Restriction mapping. Restriction endonucleases *Eco*RI, *Pvu*II, *Bam*HI, *Hae*II, and *Xho*I were obtained from Bethesda Research Laboratories and were used in the buffer recommended by the supplier. Agarose gel electrophoresis was carried out as previously described (32) with 1.2 and 1.5% agarose gels to analyze *Hae*II and *Xho*I digestion products, respectively.

For partial digestion mapping (27), gel-purified fragments were treated with bacterial alkaline phosphatase (Worthington Diagnostics) and end labeled with [γ-^{32}P]ATP (ICN Pharmaceuticals) and T4 polynucleotide kinase (Bethesda

* Corresponding author.
† Present address: Genex Corp., Gaithersburg, MD 20877.

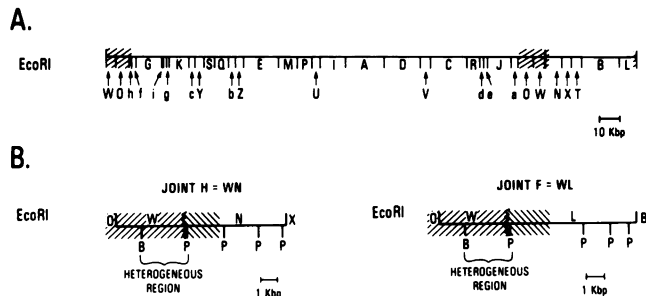


FIG. 1. (A) *Eco*RI restriction map of the HCMV genome strain AD169 shown in one orientation. The long and short inverted repeats are indicated by the slanting lines. (B) Expanded map of the two junction fragments *Eco*RI-WL and -WN showing the cleavage sites for the enzymes *Pvu*II (P) and *Bam*HI (B). The portion of the *Eco*RI junction fragment containing sequences from the long and short repeat are indicated by the slanting lines.

Research Laboratories) as previously described (32). In some cases 3′ end labeling was performed by the method of Challberg and Englund (1). Sephadex G50 column chromatography was used to separate the DNA from unincorporated label (15), and yeast tRNA was added as carrier. The labeled ends were separated by gel electrophoresis after an asymmetric cleavage. The isolated fragments were partially digested with *Xho*I, and the products of the reaction were subjected to electrophoresis on 1.5% agarose gels. The gels were dried, and the labeled fragments were visualized by autoradiography at −70°C in the presence of Lightning Plus intensifying screens.

Hybridization studies. Purified virion DNA (32) and recombinant plasmids containing the *Pvu*II-*Eco*RI heterogeneous fragment (5 ng per sample) were cleaved with *Xho*I, subjected to electrophoresis on 1.5% agarose gels, and transferred to diazobenzyloxymethyl paper (Schleicher & Schuell Co.) as recommended by the supplier. The filters were prehybridized at 37°C for 12 to 16 h in 50% formamide, 50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (pH 7.0), 5× Denhardt solution (1× Denhardt solution is 0.02% each of polyvinylpyrrolidone, Ficoll, and bovine serum albumin), 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.25 mg of denatured salmon DNA per ml, 5 mM EDTA, 0.1% sodium dodecyl sulfate, and 1% glycine. Cloned restriction fragments to be used as probes were separated from their respective plasmids by digestion with the appropriate restriction endonuclease and gel purification. These fragments were labeled with [α-^{32}P]dCTP by nick translation as previously described (32). Hybridization was performed for 3 days at 37°C in prehybridization buffer minus glycine. After hybridization, filters were washed as follows: four washes in 2× SSC–0.1% sodium dodecyl sulfate at room temperature, 5 min each; two washes in 0.1× SSC–0.1% sodium dodecyl sulfate at 50°C, 30 min each; and several rinses with 0.1× SSC at room temperature. Hybridization was detected by autoradiography.

DNA sequence analysis. The 2.5-kbp *Pvu*II-*Bam*HI subfragment from the *Pvu*II-*Eco*RI heterogeneous fragment of clone 310 was gel purified. Subclones of this fragment were generated with M13mp8 and the methodology of Deininger (3). Briefly, the 2.5-kbp fragment from clone 310 was sonicated to yield randomly sheared fragments of approximately 100 to 1,000 base pairs (bp). The ends of the fragments were repaired by treatment with T4 DNA polymerase, and fragments of 400 to 700 bp were purified by agarose gel electrophoresis. The gel-purified fragments were then ligated with the replicative form of the M13mp8 vector (17), which had been cleaved with the restriction endonuclease *Sma*I and treated with calf intestinal alkaline phosphatase. With this procedure the subclones obtained contained mainly the end fragments of the 2.5-kbp segment. To obtain subclones with a higher representation of the internal fragments, two modified procedures were used. In the first, the 2.5-kbp segment was treated with bacterial alkaline phosphatase before sonication to prevent ligation of the *Pvu*II or *Bam*HI end fragments to the alkaline phosphatase-treated vector. In the second method, the original protocol was followed, except that the sonicated fragments were treated with S1 nuclease after size fractionation to increase the number of internal fragments containing blunt ends. These modified protocols yielded cloned subfragments representative of the entire 2.5-kbp segment.

The transformation and selection of clones in the host strain *Escherichia coli* JM101 were performed as previously described (16). A single-stranded template was prepared from white M13 plaques and used for sequence analysis by the dideoxy chain-termination method (16, 25) with an M13 pentadecamer primer (New England Biolabs). Computer programs described by Staden (29, 30) and Kanehisa (10) were used to analyze the DNA sequence.

RESULTS

Localization of the region of heterogeneity within the L-S junction fragments. In our earlier studies, we presented evidence for the presence of heterogeneity within the L-S junction region of HCMV. The L-S junction fragments generated by an *Eco*RI digest of virion DNA, fragments WL and WN, contain part of the long repeat, all of the short repeat, and portions of the short unique segment. An *Eco*RI cleavage map for HCMV strain AD169 and expanded maps of the junction fragments are shown in Fig. 1. Within the junction fragments, the heterogeneity was localized to a

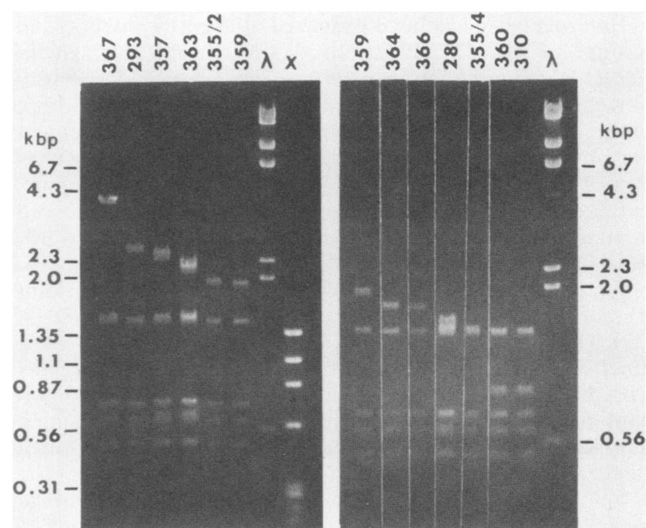


FIG. 2. *Hae*II digestion of cloned L-S junction subfragments containing the region of heterogeneity. Gel-purified *Pvu*II-*Eco*RI heterogeneous fragments were cleaved with *Hae*II and subjected to electrophoresis on 1.2% agarose gels. The number designating each clone is indicated at the top of the lane. The sizes of the *Hin*dIII-cleaved lambda DNA fragments and *Hae*III-cleaved ϕX174 RF DNA fragments used as size markers are indicated along the sides.

single *Pvu*II-*Eco*RI fragment (28, 32). The enzyme *Bam*HI cleaved the *Pvu*II-*Eco*RI heterogeneous fragment once, 1.6 kbp from the *Eco*RI site, yielding a single variable fragment (28). In this expanded study, we have characterized 12 cloned L-S junction fragments. The heterogeneous fragments from *Eco*RI-WL and -WN were treated as identical fragments since they were derived from the repeated region of the junctions. A number of other restriction enzymes were tested to determine their usefulness for characterizing the region of heterogeneity. *Hae*II cleaved close to, but not within, the region of heterogeneity. *Xho*I was used to subdivide the heterogeneous region. Studies on the heterogeneity were carried out with *Pvu*II-*Eco*RI heterogeneous fragments that were either subcloned or separated from the remainder of the junction fragment by gel purification. *Hae*II digestion of the *Pvu*II-*Eco*RI fragment produced four common bands (1.4, 0.7, 0.605, and 0.49 kbp) and a single heterogeneous band that ranged in size from 0.866 to 3.8 kbp (Fig. 2). Whenever possible, the fragment sizes given are those obtained from the DNA sequence of the *Bam*HI-*Pvu*II fragment of clone 310.

Identification of three classes of heterogeneity within the L-S junction fragments. The restriction enzyme *Xho*I produced three common bands (2.3, 1.039, and 0.208 kbp) and two classes of heterogeneous fragments (class I and class III) (Fig. 3). Individual clones contained one of the following class I heterogeneous fragments: 0.553, 0.95, or 1.35 kbp. Out of the 12 clones examined, 5 contained a fragment equivalent to the 0.553-kbp fragment of clone 310, 6 had a 0.95-kbp fragment, and 1 contained a 1.35-kbp fragment. Variation in the intensities of these fragments in different clones indicated that multiple copies of these sequences might be present. This seemed likely, since the presence of a single 0.553-, 0.95-, or 1.35-kbp fragment could not account for the size of the heterogeneous region observed in the larger cloned junction fragments. This variation in copy number of class I heterogeneous fragments was designated class II heterogeneity. The variation in intensity of the 0.2-kbp fragment indicated that the third class of heterogeneity was probably due to variation in copy number of this fragment.

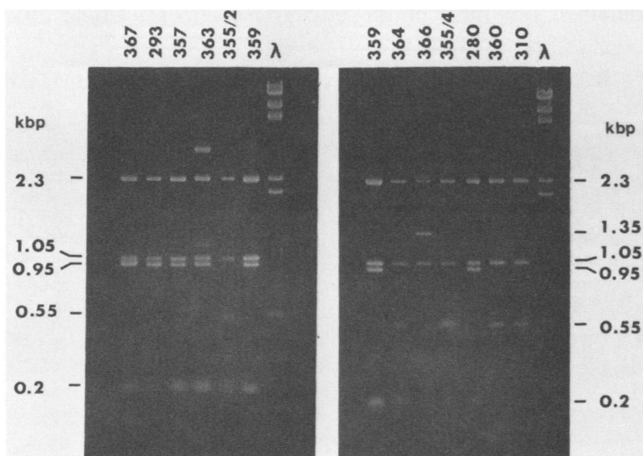


FIG. 3. *Xho*I digestion of L-S junction subfragments containing the region of heterogeneity. Gel-purified *Pvu*II-*Eco*RI heterogeneous fragments were digested with *Xho*I and subjected to electrophoresis on 1.5% agarose gels. The number designating each clone is indicated at the top of the lane. *Hin*dIII-digested lambda DNA fragments served as size markers. The numbers at the sides of the gels indicate the lengths (kbp) of the *Xho*I fragments. The 1.45-kbp fragment of clone 363 and the bands above 2.35 kbp, visible in some of the lanes, are partial digestion products.

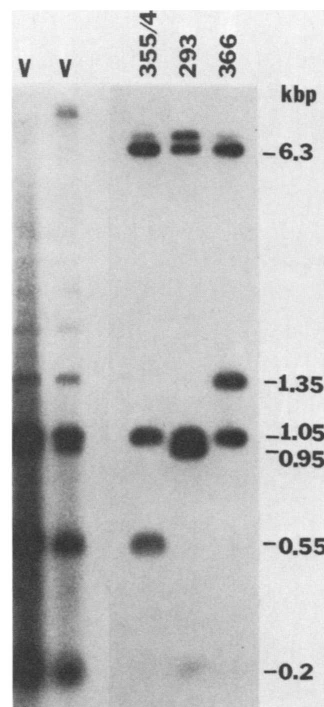


FIG. 4. *Xho*I digestion of virion DNA. Virion DNA and recombinant plasmids containing *Pvu*II-*Eco*RI heterogeneous junction fragments were digested with *Xho*I and subjected to electrophoresis on 1.5% agarose gels. The DNA was transferred to diazobenzyloxymethyl paper and hybridized with a ^{32}P-labeled probe consisting of a cloned *Pvu*II-*Eco*RI heterogeneous fragment and the *Xho*I 0.2-kbp heterogeneous fragment. The lanes labeled V contain samples of the virion DNA preparations originally used to clone the L-S junction fragments. The *Pvu*II-*Eco*RI junction fragments are indicated by clone numbers at the top of the lanes.

Viral origin of heterogeneous *Xho*I fragments. To verify that the 0.2-, 0.553-, 0.95-, and 1.35-kbp *Xho*I heterogeneous fragments were of viral origin and not a cloning artifact, the virion DNA originally used to clone the *Eco*RI junction fragments was cleaved with *Xho*I, electrophoresed on an agarose gel, and transferred to diazobenzyloxymethyl paper. This filter was hybridized with a mixture of the *Pvu*II-*Eco*RI heterogeneous fragment of clone 363 and the *Xho*I 0.2-kbp heterogeneous fragment. Virion DNA bands were observed that comigrated with the 0.2-, 0.553-, 0.95-, and 1.35-kbp heterogeneous fragments detected in the cloned junction fragments (Fig. 4). The 0.2-kbp virion fragment was also visible when only the *Xho*I 0.2-kbp heterogeneous fragment was used as a probe (data not shown). Higher-molecular-weight bands were also observed which varied in size by approximately 0.4 to 0.5 kbp, but these fragments were not represented in our cloned junctions. Rehybridization of the filter with a fragment from the unique region of the genome verified that these multiple bands were not due to partial digestion products (data not shown).

Arrangement of heterogeneous *Xho*I fragments within the L-S junction fragments. To determine the arrangement and number of class I and III heterogeneous fragments within individual junction fragments, we mapped the *Xho*I fragments by partial *Xho*I digestion of fragments ^{32}P labeled at

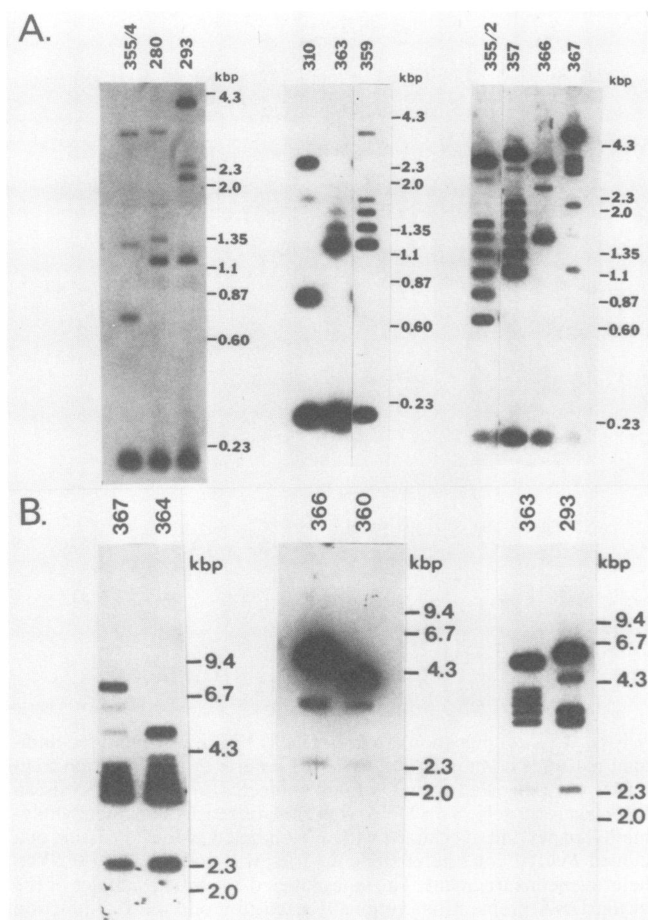


FIG. 5. Partial *XhoI* restriction digestion of fragments containing the region of heterogeneity. Partial digestion products were subjected to electrophoresis on 1.5% agarose gels. The gels were dried, and the labeled fragments were detected by autoradiography. Clone numbers are indicated at the top of each lane. The numbers to the right of the lanes indicate the sizes (kbp) of selected *Hin*dIII-cleaved lambda and *Hae*III-cleaved ϕX174 RF restriction fragments used as markers. (A) *Pvu*II-*Eco*RI heterogeneous fragments from cloned junction fragments were end labeled and cleaved with *Bam*HI. The *Bam*HI-*Pvu*II fragment was gel purified before partial digestion with *Xho*I. (B) *Eco*RI-WL and -WN were end labeled and cleaved with *Pvu*II. The *Pvu*II-*Eco*RI heterogeneous fragments were isolated by gel purification before partial digestion with *Xho*I.

only one end of the molecule. The *Pvu*II-*Bam*HI heterogeneous fragment, labeled at the *Pvu*II site, was used to order the *Xho*I fragments in one direction; the *Pvu*II-*Eco*RI heterogeneous fragment, labeled at the *Eco*RI site, was used to order the *Xho*I fragments in the opposite direction. Figure 5 shows the results obtained with a representative group of clones. Figure 5A shows the result of partial *Xho*I digestion of fragments labeled at the *Pvu*II site; Fig. 5B shows partial digestion products of fragments labeled at the *Eco*RI site. From the results shown in Fig. 5A and B, it was determined that the 0.208-kbp fragment was located at the *Pvu*II end of the molecule. A 0.553-, 0.95-, or 1.35-kbp fragment (class I heterogeneity) or multiples of these fragments (class II heterogeneity) were adjacent to the 0.208 kbp fragment, followed by a variable number of the class III 0.2-kbp fragments. The heterogeneous fragments were followed, in order, by the 1.039-kbp *Xho*I fragment and the 0.738-kbp *Xho*I-*Bam*HI fragment. This fragment order, summarized in

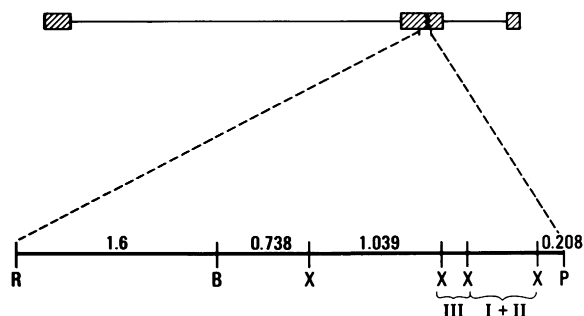


FIG. 6. Diagram of the HCMV genome with an expanded restriction endonuclease cleavage map of the *Eco*RI-*Pvu*II fragment containing the heterogeneous region of the L-S junction. The long and short inverted repeats are indicated by slanting lines. The following letters indicating restriction enzyme cleavage sites appear below the line: R, EcoRI; B, BamHI; X, XhoI; P, PvuII. The lengths (kbp) of the fragments are indicated above the line. Roman numerals indicate the location of the three types of heterogeneity. In class I heterogeneity, three fragment sizes were observed in the cloned junctions (0.553, 0.95, and 1.35 kbp). The results shown in Fig. 4 suggested that other fragment sizes exist in the virion population. In class II heterogeneity, some of the cloned junctions contained multiple copies of a class I fragment. Mixtures of class I fragments (0.55- and 0.95-kbp fragments within a single junction) were never observed. Class III heterogeneity was represented by a variable number (0 to 5) of tandemly arranged 0.2-kbp fragments.

Fig. 6, was consistent with the DNA sequence obtained from clone 310.

The three classes of heterogeneity seemed to vary independently (Table 1). Cloned junctions containing one copy of the 0.553-kbp fragment (clones 310, 360, 364, and 355/2) had from 0 to 5 copies of the 0.2-kbp fragment (0, 0, 4, and 5, respectively). One clone, 355/4, contained two copies of the 0.553-kbp fragment and no copies of the 0.2-kbp fragment. Half of the 12 clones studied contained the 0.95-kbp heterogeneous fragment. Clones 280, 359, 363, and 357 contained a single copy of this fragment and 1, 3, 4, and 5 copies of the class III 0.2-kbp fragment, respectively. Clones 293 and 367 contained multiples of the 0.95-kbp fragment. Clone 293

TABLE 1. Composition of heterogeneous regions from individual cloned L-S junction fragments

Clone	Length of *Pvu*II-*Eco*RI fragment (kbp)[a]	Length of *Hae*II heterogeneous fragment (kbp)[b]	*Xho*I heterogeneous fragments[c]		
			Class I (kbp)	Class II (no. of class I fragments)	Class III (no. of 0.2 kbp fragments)
367	7.0	3.8	0.95	3	3
293	5.65	2.45	0.95	2	1
357	5.5	2.3	0.95	1	5
363	5.3	2.1	0.95	1	4
355/2	5.1	1.9	0.553[d]	1	5
359	5.1	1.9	0.95	1	3
364	4.9	1.7	0.553[d]	1	4
366	4.9	1.7	1.35	1	0
280	4.7	1.5	0.95	1	1
355/4	4.65	1.45	0.553[d]	2	0
360	4.1	0.87	0.553[d]	1	0
310	4.1	0.866[d]	0.553[d]	1	0

[a] The data were compiled from references 28 and 32 and our unpublished results.

[b] The data were compiled from analysis of the gels shown in Fig. 2.

[c] The data were compiled from analysis of the gels shown in Fig. 3 and 5.

[d] The lengths of the *Hae*II and *Xho*I fragments were obtained from the sequence of the heterogeneous region of clone 310.

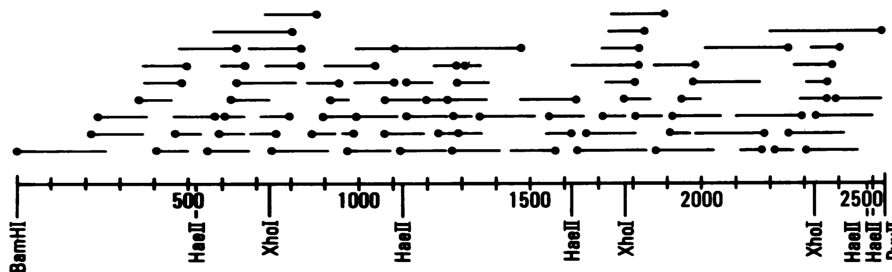


FIG. 7. Strategy for determination of the DNA sequence of the *BamHI-PvuII* subfragment of clone 310. The distribution and orientation of the M13 mp8 subclones sequenced are indicated. The axis represents the nucleotide map of the restriction fragment (bp). The positions of several restriction endonuclease sites are also indicated.

contained two copies of the 0.95-kbp fragment and one copy of the 0.2-kbp fragment; clone 367, the largest junction fragment studied, contained three copies of the 0.95-kbp fragment and three copies of the 0.2-kbp fragment. Only one clone, 366, contained a 1.35-kbp fragment, and this clone had no copies of the 0.2-kbp fragment. Whereas three of the clones studied contained multiples of the 0.553- or 0.95-kbp fragment, none of the clones contained a combination of these fragments. In general, junction fragments of equal size did not contain the same distribution of *XhoI* heterogeneous fragments (Table 1).

DNA sequence analysis. To characterize further the L-S junction region of HCMV, we determined the nucleotide sequence of the 2,538-bp *BamHI-PvuII* fragment of clone 310, which spanned the heterogeneous region. This clone contained the smallest region of heterogeneity analyzed and, from restriction endonuclease data, appeared to consist of one copy of the 0.553-kbp *XhoI* fragment and no copies of the 0.2-kbp *XhoI* fragment. Randomly sheared fragments from this region were cloned into the M13mp8 phage vector (17) and were subjected to DNA sequence analysis by the dideoxy chain-termination method (25). Figure 7 shows the distribution and orientation of the subclones sequenced. The entire sequence of this region was generated with computer-assisted programs (3) (Fig. 8). All portions of the reported sequence, with the exception of 210 nucleotides adjacent to the *BamHI* site, were obtained from at least two independent, overlapping clones or from both strands of the DNA or, in most cases, from both types of analysis.

The entire region of heterogeneity was included within the *HaeII* sites centered at nucleotides 1621 and 2487. This region was G+C rich (70%) and contained multiple direct repeats. For example, the *XhoI* sites (centered at nucleotides 1779 and 2332) bounding the 0.553-kbp fragment were contained within a 25-nucleotide direct repeat (designated DR1). Within the 0.553-kbp *XhoI* fragment, we found an 8-bp sequence, GTGTGNNG (N is T, C, or G), which was repeated 16 times in the region bounded by nucleotides 1819 and 2051. These octamers could be arranged into longer direct repeats of at least 15 nucleotides as indicated in Fig. 8. Several inverted repeats were also scattered throughout the entire sequenced segment, but none of these was greater than 12 nucleotides in length. Within the 0.553-kbp *XhoI* fragment, we have noted the presence of the octamer GGTGTTTT (at nucleotide 1809) and its complement (at nucleotide 2313) near the inner boundaries of the 25-bp direct repeat DR1 (indicated by arrows in Fig. 8). The alternating dinucleotides G-T and C-A appeared in short stretches (3 to 9 bp) throughout the sequence. Although the G-T dinucleotide was more common, there was a tendency for each pair of nucleotides to predominate in different regions. Alternating G-T dinucleotides were concentrated in the regions from positions 1080 to 1434 and from 1810 to 2050. Alternating C-A sequences were predominant from positions 1538 to 1804 and from 2179 to 2345.

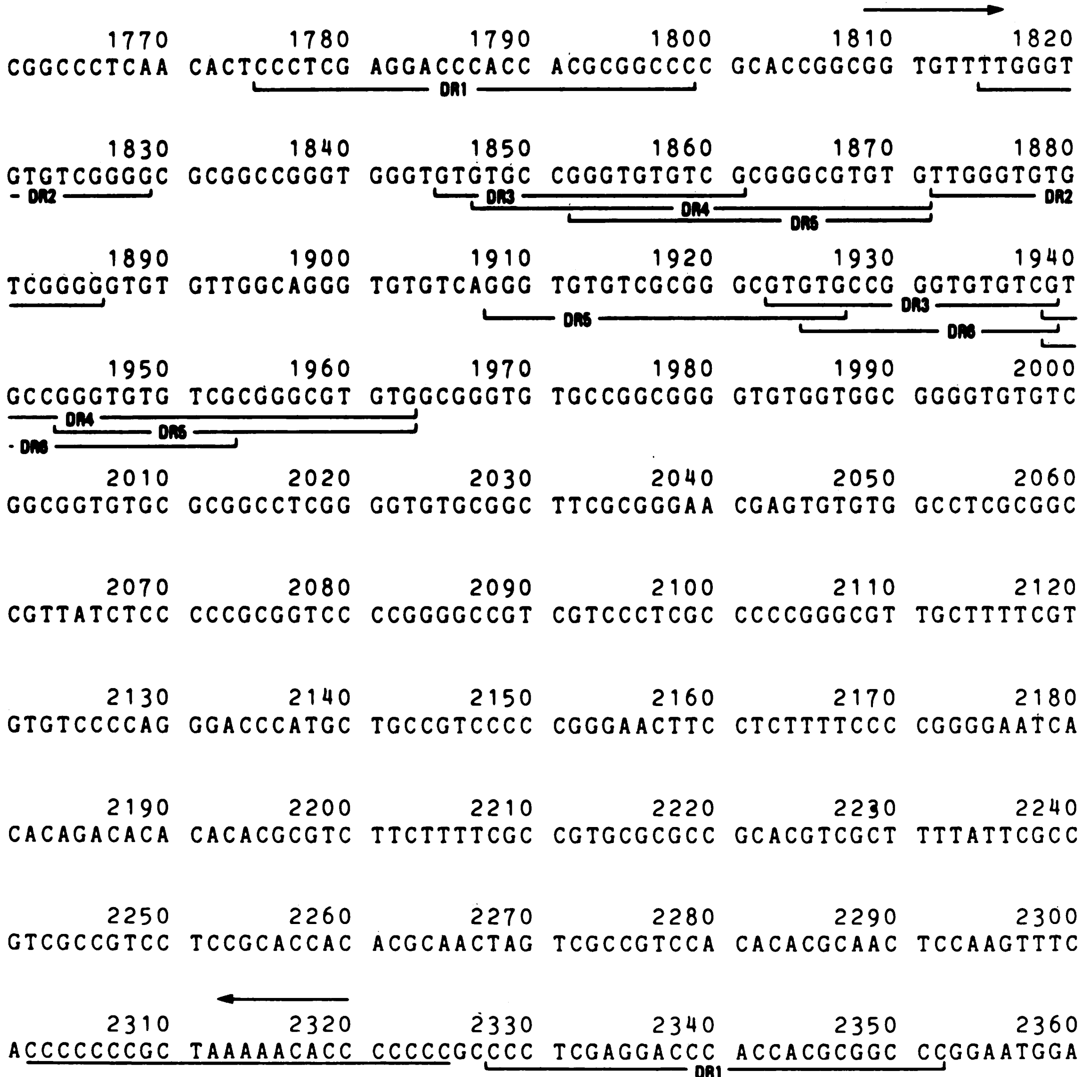
DISCUSSION

In our previous studies of the HCMV (strain AD169) genome, we found evidence for the presence of heterogeneity in the L-S junction region. This heterogeneity was also represented in our cloned *EcoRI* junction fragments (28, 32). Characterization of the cloned junction fragments with the restriction enzymes *PvuII* (28, 32), *BamHI* (28), and *HaeII* enabled us to localize the region that contained the heterogeneity to the L-S junction itself. The restriction endonuclease *XhoI* cleaved within the heterogeneous region and revealed three classes of heterogeneity. Class I heterogeneous fragments varied in size; fragments of 0.553, 0.95, or 1.35 kbp were observed in the cloned junctions. Class II heterogeneity was generated by tandem duplications of class I fragments. We did not observe a junction fragment containing a combination of 0.553-, 0.95-, and 1.35-kbp fragments. Class III heterogeneity resulted from a variable number of tandem 0.2-kbp fragments. The order of the *XhoI* fragments was determined by partial *XhoI* digestion of heterogeneous junction fragments ^{32}P labeled at one end of the molecule. The 0.2-kbp class III heterogeneous fragments were located on the W side of the junction. The size variation of heterogeneous *EcoRI* W fragments within virion DNA (28) was probably due to the 0.2-kbp class III heterogeneous fragments. Class I heterogeneous fragments were located at the L-S junction adjacent to the region of class III heterogeneity.

We have sequenced the *BamHI-PvuII* subfragment of clone 310. This clone contained a single copy of the *XhoI* 0.553-kbp class I fragment and no copies of the 0.2-kbp class III heterogeneous fragment. The results obtained from our partial and complete *XhoI* digestions correspond with the order and size of *XhoI* restriction fragments predicted from the sequence. Examination of the sequence revealed that the *XhoI* sites bounding the 0.553-kbp heterogeneous fragment lay within 25-bp direct repeats (DR1). The sequence contained many other direct and inverted repeats, clustered mainly in and around the 0.553-kbp *XhoI* fragment. One inverted repeat (GGTGTTTT and its complement) was located near the inner boundaries of the 25-bp direct repeat DR1. The alternating dinucleotides G-T and C-A were observed in short stretches (3 to 9 bp) throughout the sequenced fragment.

All heterogeneous fragments present in the cloned junction fragments were observed in *XhoI*-cleaved virion DNA. In addition, higher-molecular-weight bands that varied by 0.4 to 0.5 kbp were observed. By intensity of hybridization these heterogeneous fragments appeared to be less prevalent in the virion population than the 0.553-, 0.95-, and 1.35-kbp

```
        10         20         30         40         50         60         70         80         90        100        110        120
GATCCACAAA AACAACAACC TCTGTATGGA AAATGCGCTG TTTTATCTCA GCTTTTCTCC CAAACCTCGG TTTCTTCCTA TTCTTATGTT TTCCCTAGTA TATTTGCCTC CTTATAAGAA
       130        140        150        160        170        180        190        200        210        220        230        240
AAGAAGCACA AGCTCGGTCG CACGGATTAT TCCTTCTGCT AATCTATTAT TTTGTTCCTT TTTTTTTTCT TTGCCTTCAC CCTCTTCACT CCCTGTAGCA ACACAGAGTA GTAGACACAA
       250        260        270        280        290        300        310        320        330        340        350        360
TAAATGAGAA GTTTGCATGC ATTTGTCGTG TCCGTGGTTT GTTATGGCGT GTGGAGTGCT CGGGATGGGT GGACGTGGGG ACGGATTCTT GAGGCTACAA AGATACGCGG AGACGTCGTG
       370        380        390        400        410        420        430        440        450        460        470        480
GCGAGGGGAT GGGTTTATTG GATATCGGTG AAGCAGCGTG GCGGCGAAAG ACGCGATCCC TGGGCTGGTA GATCCCCCTA CCCCGTCTAC CGGGGACGTT TATCCTTTGG ACACGTAAAT
       490        500        510        520        530        540        550        560        570        580        590        600
GTCTCGGCCG GCATCCACGC GCCACGTTCA CCGCGTTGTG CCCAGCGCCA TGTGCGGGTC GTTTCGGCGT GAAGTTGGAC GGCGTAGTTT CGGGGATTGT GAACCGTGGC TGAGGGTGTA
       610        620        630        640        650        660        670        680        690        700        710        720
GATGGGACAG GAAAAAGCGT GTGATCTCAC CGAGGCGAAG CATGTGGGTG GTGCGATGCG GTGGATGTGG CCGGGGTGCG GCGGTTTCCG ACGTGGAGAT GTGGACATGG GGGTGATCCG
       730        740        750        760        770        780        790        800        810        820        830        840
GATGCGTGGC AAGAGGCCTC GAGCTTGGGC TTCTCCCGCG GATGGACGTT CTAACTGTAC ACGGCGGCCG TGGCCTCCGA GTAAAAAAAC CAGGTGCTGA CGCCAGACAG AGACGCCGTC
       850        860        870        880        890        900        910        920        930        940        950        960
CTCGGAATCG TGTGCGCGAA AGCCTGTGCC GCGGCAGCGT ACGACGTTCC AGTCAGCGAG GCCGTCGCGT TGGCGCGCCA ACAGTAAGGT GACGACAGGT TGGCGGCCCA TGGTTCCGGA
       970        980        990       1000       1010       1020       1030       1040       1050       1060       1070       1080
GCGTCCCCAC ATGCGCCAGC AGTCGGCGTC AAAGTCGCTT GCGCTGTCGG CCCAGTCGCC ACCGCCGCGG CGGATTTCCG CGCGGGGGAC GGGGTAGCCG AGTGCTGCGC CCTCGCCAAT
      1090       1100       1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
GTTGTGAAGT GGATGCGTGA GTTGATGTTG ATTCTCTGTG GGAAAATGAG CGCTGTCCTG TGGGTTGGTG TTGGGGTATG CGAGTAGTAG GGGTTGTGTT TGATCGTAGA GGTGTTGGCG
      1210       1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320
GGCCTGTGCG CGAGCAGCGT AGTCTGCGGC GTCGAGCTCC ATCTGTGTGC GGTGTTCTTC GTCGGCGTGT TTGTCCGAGG TTCGGACATG CGGTTGTGTG TTGCTGTGGT GTAAGGGTGA
      1330       1340       1350       1360       1370       1380       1390       1400       1410       1420       1430       1440
CGTGTGTTGG GCGTCTGGGT GAAGCGGCGT GGTGTGGGTG CTGTTTGTGT CTGTGGCTGG CATGATTGTG CGGCATGTGT GTGTTGTAGT GGGTGGAGGT TAAATAGGTG AGGTGGGTTC
      1450       1460       1470       1480       1490       1500       1510       1520       1530       1540       1550       1560
CCTGGTCCGC GCCGCAAACT GTCCCCGTAC CCAACGTAAC CTCCCCTACG CGGCGCGAAC AGCCCCGGAC CCAGCGCAAC CCCCGTCCCC GGCCCCGACA CCGTCCCGCA CACCCCCCGT
      1570       1580       1590       1600       1610       1620       1630       1640       1650       1660       1670       1680
CTCCGCAACA CCCCGGCATC GCCGGCGGCC AGAACGCTCG AAAACCCCCG ACAAGCGCAG CGCCGAAACG ACACAGGCAA GGACCGTGGA ACGCACCGGC AGCGCGCCGA AACACCGTCC
      1690       1700       1710       1720       1730       1740       1750       1760
CGAAGCCCGG TGCCGACAAC AAATACCGTG GGACGACACG CACCGGCAGT GCGCAGGCAG CGTCGGACAC AACACGCTTA
```




```
      2370       2380       2390       2400       2410       2420       2430       2440       2450       2460       2470       2480
TGTCGGGCGT CCACCTAGAT GGGTGCGCGC CCGGGAGGCG GCTGTGCGCT CCAGTGGTAC GCGCCTGCCG CGCGTCTTCC TTCGGGTAGC TGCCTTTCCC AGTCCACGGC CTTCCAGACT
      2490       2500       2510       2520       2530
GCGTGGCGCC AAGGCGGCGC CAGCACGCGC CGTGCACGTC GCTGCCTATA AAAGCCAG
```

FIG. 8. Nucleotide sequence of the *Bam*HI-*Pvu*II subfragment of clone 310. Direct repeats of 15 bp or more are indicated by DR. The arrows designate the inverted repeat near DR1, and the region that shows homology to the HSV *a* sequences is underlined.

fragments represented in our cloned junctions. The 0.95- and 1.35-kbp fragments and other class I *Xho*I fragments may consist of multiple copies of the 0.553-kbp fragment that have lost the interior *Xho*I site(s), perhaps by a deletion within the DR1 region. Further studies, including DNA sequence analysis, will be required to determine the precise structure and relationship of these *Xho*I fragments.

The HCMV strains Towne and Davis have heterogeneity predominantly at the S terminus, but some heterogeneity has been observed at the L terminus. In the Towne strain an additional 0.9 kbp is observed in 10% of the L termini (12; G. S. Hayward, personal communication), and a terminal fragment 0.45 to 0.6 kbp larger is observed in 10 to 20% of the L termini of the Davis strain (4). In contrast with other strains of HCMV, heterogeneity is commonly observed at both termini of strain AD169 (28; G. S. Hayward, personal communication). The heterogeneity at the S terminus of AD169 varies by increments of approximately 0.5 to 0.6 kbp and is probably produced by the class I and class II heterogeneity. The S-terminal heterogeneity is similar, although not identical, for the Towne, Davis, and AD169 strains. The S-terminal heterogeneity in the Davis strain takes the form of two equimolar terminal fragments that vary by 0.45 to 0.75 kbp (4). The heterogeneity that appears as a ladder of bands differing by 200 bp at the L terminus is unique to the AD169 strain and is probably generated by the 0.2-kbp class III heterogeneous fragment. Determination of the precise relationship of the two classes of heterogeneity will require direct DNA sequence analysis; these studies are in progress.

The herpes simplex virus type 1 (HSV-1) and HSV-2 genomes consist of covalently linked long (L) and short (S) segments, each bounded by inverted repeats (26), that are analogous to those in the HCMV genome. The repeats bordering the L segment have the form *ab* and *b'a'*, whereas those flanking the S segment have the structure *a'c'* and *ca* (33). The *a* region of HSV-1 is 0.25 to 0.5 kbp in length (2, 14, 19, 34); the length of the HSV-2 *a* region is approximately 0.25 kbp (2). Molecular analyses of the genomes of HSV-1 and HSV-2 have revealed that the L-S junction and L terminus are heterogeneous. The heterogeneity is due to a variation in the number of *a* sequences at the L-S junction and the L terminus and also to small 10 to 50-bp insertions or deletions within the *a* and *c* sequences (2, 14, 19, 34). In the majority of HSV-1 molecules, only a single copy of the *a* sequence has been observed at the S terminus. Sequence analyses of the L-S junction regions of HSV-1 (2, 19) and HSV-2 (2) have shown that the *a* sequences of both HSV-1 and HSV-2 are bounded by 17 to 21-bp direct repeats. In junction segments with two *a* sequences, only three copies of the direct repeats are found; the *a* sequences are tandemly arranged and share the direct repeat that lies between them.

Comparison of the L-S junction of HCMV with that of HSV-1 and HSV-2 suggested that the 0.553-kbp *Xho*I heterogeneous fragment corresponded to the *a* sequence of the AD169 strain of HCMV (Fig. 9). This fragment resembled both the HSV-1 and HSV-2 *a* sequences in location and structure; it was located at the L-S junction and was bounded by direct repeats. Tandem copies of this fragment were observed in clone 355/4, similar to the occurrence in HSV of multiple *a* sequences. Like the HSV-1 *a* sequence, the 0.553-kbp *Xho*I fragment contained reiterated sequences. However, the HSV reiterated sequences are tandem direct repeats, whereas the reiteration in the HCMV 0.553-kbp *Xho*I fragment occurred as a cluster of dispersed direct repeats. A direct comparison of the nucleotide sequences of the HSV-1 and HSV-2 *a* sequences (2, 19) with

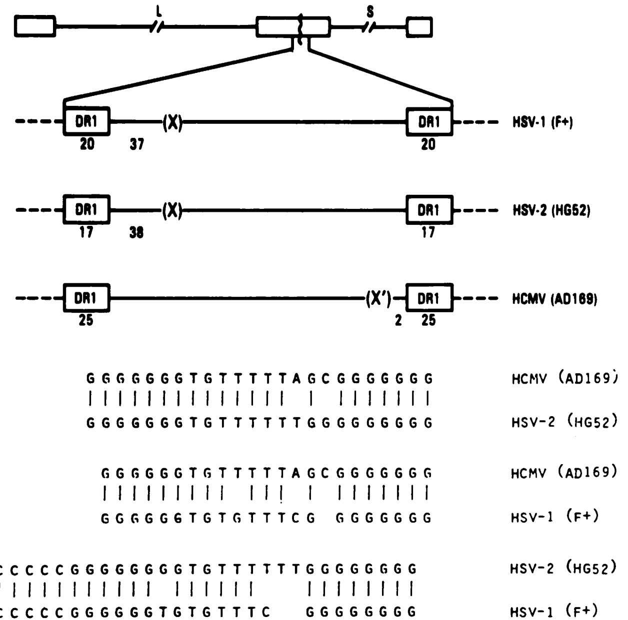


FIG. 9. Comparison of the conserved sequence (X) found in the 0.553-kbp *Xho*I fragment of HCMV (strain AD169) and the *a* sequences of HSV-1 (strain F$^+$) and HSV-2 (strain HG52). A schematic drawing of the linear genome is shown. The direct repeats (DR1) flanking the HSV *a* sequences and the HCMV 0.553-kbp *Xho*I fragment are indicated by the open boxes in the expanded diagram. The numbers below the DR1 indicate the length of the repeat. The position of the X sequence in HSV-1 and HSV-2 and the inverted sequence in HCMV are indicated; the numbers indicate the distance in bp between the X sequence common to the three viruses and the DR1. At the bottom are shown the computer-generated alignments of the X sequence. DNA sequence information for HSV-1 and HSV-2 was obtained from Mocarski and Roizman (19) and from Davison and Wilkie (2).

that of the HCMV 0.553-kbp *Xho*I fragment revealed one striking region of homology (designated X) shown schematically in Fig. 9. This sequence was present in the inverted orientation in the HCMV 0.553-kbp *Xho*I fragment at nucleotides 2302 through 2325.

The *a* sequence appears to have several important roles in viral replication. Experiments on HSV-1 (18, 20, 21, 24) suggested that inverted copies of the *a* sequence are necessary for L and S segment inversions, thus indicating that the *a* sequences contain a signal for recombination. The *a* sequences of HSV-1 also have a role in DNA replication; concatemers formed during viral replication are cleaved to genome-length fragments within the direct repeat (DR1) bounding the *a* sequence. The cleavage occurs such that the majority of the molecules contain a single copy of the *a* sequence at the S terminus. The role of the conserved X sequence remains to be determined. It could function as a site of recombination, or it might play a role in the packaging of the viral DNA either by providing a signal for the cleavage of unit-length DNA or by orienting the direction of packaging.

The short stretches of alternating G-T and C-A dinucleotides observed in the junction region may be important for *a* sequence functions. Alternating purine and pyrimidine residues favor the formation of left-handed helices (Z-DNA), which might influence enzyme-DNA interactions. Nordheim and Rich (22) have shown that 8-bp segments of alternating purines and pyrimidines within the simian virus 40 enhancer

region can form Z-DNA. They also reported that the sequences of other viral enhancers contain potential Z-DNA-forming regions. These viral enhancers have a pair of potential Z-DNA segments (8 to 13 bp) separated by 50 to 80 bp of DNA. Although the DNA sequence of the L-S junction region of HCMV did not fit the pattern described for viral enhancers, the presence of potential Z-DNA-forming segments raises the possibility that the secondary structure of the DNA in this region may play a role in recombination, replication, or packaging of viral DNA.

ACKNOWLEDGMENTS

We thank S. Shaw, S. McDonough, S. Staprans, J. Mercer, J. Marks, and R. Rasmussen for helpful discussions and T. Schmidhauser for providing pMK16. We also thank F. Sanger, B. Barrell, and P. Deininger for providing laboratory space and expertise during the initial sequencing work.

This research was supported by Public Heaalth Service grant CA 34729 and training grant CA 09345 from the National Institutes of Health. J.C.T. was a predoctoral fellow of the National Science Foundation.

LITERATURE CITED

1. **Challberg, M. D., and P. T. Englund.** 1980. Specific labeling of 3' termini with T4 DNA polymerase. Methods Enzymol. **65**:39–43.
2. **Davison, A. J., and N. M. Wilkie.** 1981. Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. J. Gen. Virol. **55**:315-331.
3. **Deininger, P. L.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. **129**:216–223.
4. **DeMarchi, J. M.** 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early, and late RNAs. Virology **114**:23-38.
5. **DeMarchi, J. M., M. L. Blankenship, G. D. Brown, and A. S. Kaplan.** 1978. Size and complexity of human cytomegalovirus DNA. Virology **89**:643–646.
6. **Fleckenstein, B., I. Muller, and J. Collins.** 1982. Cloning of the complete human cytomegalovirus genome in cosmids. Gene **18**:39–46.
7. **Geelen, J. L. M. C., C. Walig, P. Wertheim, and J. van der Noordaa.** 1978. Human cytomegalovirus DNA. I. Molecular weight and infectivity. J. Virol. **26**:813–816.
8. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
9. **Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remaut, and D. R. Helinski.** 1980. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K and RK2. Methods Enzymol. **68**:268–280.
10. **Kanehisa, M. I.** 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–196.
11. **Kilpatrick, B. A., and E. S. Huang.** 1977. Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. J. Virol. **24**:261–276.
12. **LaFemina, R. L., and G. S. Hayward.** 1980. Structural organization of the DNA molecules from human cytomegalovirus, p. 39–55. *In* B. N. Fields, R. Jaenisch, and C. F. Fox (ed.), Animal virus genetics, ICN-UCLA Symposia on Molecular and Cellular Biology, vol. 18. Academic Press, Inc., New York.
13. **Lakeman, A. D., and J. E. Osborn.** 1979. Size of infectious DNA from human and murine cytomegaloviruses. J. Virol. **30**:414–416.
14. **Locker, H., and N. Frenkel.** 1979. *Bam*I, *Kpn*I, and *Sal*I restriction enzyme maps of the DNAs of herpes simplex virus strains Justin and F: occurrence of heterogeneities in defined regions of the viral DNA. J. Virol. **32**:429–441.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual, p. 466–467. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
17. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double digestion restriction fragments. Gene **19**:269–276.
18. **Mocarski, E. S., L. E. Post, and B. Roizman.** 1980. Molecular engineering of the herpes simplex virus genome: insertion of a second L-S junction into the genome causes additional genome inversions. Cell **22**:243–255.
19. **Mocarski, E. S., and B. Roizman.** 1981. Site-specific inversion sequence of the herpes simplex virus genome: domain and structural features. Proc. Natl. Acad. Sci. U.S.A. **78**:7047–7051.
20. **Mocarski, E. S., and B. Roizman.** 1982. Herpesvirus-dependent amplification and inversion of cell-associated viral thymidine kinase gene flanked by viral *a* sequences and linked to an origin of viral DNA replication. Proc. Natl. Acad. Sci. U.S.A. **79**:5626–5630.
21. **Mocarski, E. S., and B. Roizman.** 1982. Structure and role of the herpes simplex virus DNA termini in inversion, circularization and generation of virion DNA. Cell **31**:89–97.
22. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
23. **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.
24. **Post, L. E., A. J. Conley, E. S. Mocarski, and B. Roizman.** 1980. Cloning of reiterated and non-reiterated herpes simplex virus 1 sequences as BamHI fragments. Proc. Natl. Acad. Sci. U.S.A. **77**:4201–4205.
25. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
26. **Sheldrick, P., and N. Berthelot.** 1975. Inverted repetitions in the chromosome of herpes simplex virus. Cold Spring Harbor Symp. Quant. Biol. **39**:667–678.
27. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids Res. **3**:2387–2398.
28. **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.
29. **Staden, R.** 1977. Sequence data handling by computer. Nucleic Acids Res. **4**:4037–4051.
30. **Staden, R.** 1979. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. **6**:2601–2610.
31. **Stinski, M. F., E. S. Mocarski, and D. R. Thomsen.** 1979. DNA of human cytomegalovirus: size heterogeneity and defectiveness resulting from serial undiluted passage. J. Virol. **31**:231–239.
32. **Tamashiro, J. C., L. J. Hock, and D. H. Spector.** 1982. Construction of a cloned library of the *Eco*RI fragments from the human cytomegalovirus genome (strain AD169). J. Virol. **42**:547–557.
33. **Wadsworth, S., R. J. Jacob, and B. Roizman.** 1975. Anatomy of herpesvirus DNA. II. Size, composition and arrangement of inverted terminal repetitions. J. Virol. **15**:1487–1497.
34. **Wagner, M. J., and W. C. Summers.** 1978. Structure of the joint region and the termini of the DNA of herpes simplex virus type 1. J. Virol. **27**:374–387.
35. **Weststrate, M. W., J. L. M. C. Geelen, and J. van der Noordaa.** 1980. Human cytomegalovirus DNA: physical maps for the restriction endonucleases BglII, HindIII, and XbaI. J. Gen. Virol. **49**:1–21.