Nucleotide Sequences of the Joint between the L and S Segments of Herpes Simplex Virus Types 1 and 2

By A. J. DAVison AND N. M. WILKIE*

Medical Research Council Virology Unit, University of Glasgow, Church Street, Glasgow G11 5JR, U.K.

(Accepted 27 May 1981)

SUMMARY

The a sequence of herpes simplex virus (HSV) is present as a direct repeat at the genomic termini and also in inverted orientation at the joint between the L and S segments. DNA sequences have been determined for the joint regions of the genomes of HSV-1 and HSV-2, and relative to these sequences the genomic termini are in both cases located close to a short direct repeat of 17 to 21 base pairs (bp) at the b–a and a–c junctions. The HSV-1 joint region contains three separate tandem direct reiterations of short sequences, (12, 16 and 17 bp in strain 17) and we conclude that size heterogeneity in the a and c sequences is due to variable copy numbers of these repeated units. It is likely that a considerable part of the HSV-1 joint region does not code for polypeptide.

INTRODUCTION

The genome of herpes simplex virus type 1 (HSV-1) is a linear DNA duplex of mol. wt. 96 x 10^6 (Kieff et al., 1971; Clements et al., 1976). The genome structure is shown in Fig. 1(a). It consists of a long (L) and a short (S) segment, each of which comprises a unique region (U_L, U_S) bounded by inverted repetitions (TR_L, IR_L, TR_S, IR_S). Virion DNA populations contain four types of molecule which rise by inversion of the L and S segments about their mutual joint (Sheldrick & Berthelot, 1974; Hayward et al., 1975; Clements et al., 1976; Delius & Clements, 1976; Wilkie & Cortini, 1976; Skare & Summers, 1977).

Neither the mechanism of inversion nor its biological significance are known, but some detail is known of the structural organization of the DNA sequences around the L–S joint. The a sequence is present as a direct repeat at each terminus and also in inverted orientation at the joint (Sheldrick & Berthelot, 1974; Grafstrom et al., 1974, 1975). Its size has been estimated in different strains of HSV-1 to be 400 to 1600 base pairs (bp) by electron microscopy (Grafstrom et al., 1975; Wadsworth et al., 1976; Kudler & Hyman, 1979) and 265 bp by restriction endonuclease mapping (Wagner & Summers, 1978). Two types of size heterogeneity in the joint and terminal regions have been described, one consisting of insertions closely corresponding to the size of, and thought to contain, a sequences (Wilkie, 1976; Wilkie et al., 1977; Wagner & Summers, 1978; Locker & Frenkel, 1979), and the other of smaller insertions of 10 to 50 bp within the a sequence and adjacent c sequence (Wagner & Summers, 1978).

In this paper we report the nucleotide sequences of the joint regions of HSV-1 and HSV-2, and identify the positions of the 5' termini of the genome relative to them. We describe the nature of size heterogeneity at the level of nucleotide sequence and locate several features which may be of functional significance.
Fig. 1. (a) Structural arrangement of the HSV-1 genome. The L and S segments are shortened in order to emphasize the locations and orientations (arrowed) of the a sequences, or terminal repetitions (tr), at the genomic termini and at the L–S joint. A single DNA molecule contains each segment in either orientation with equal probability, as shown, allowing four permutations of sequence arrangement. Sizes of regions of the genome are given in kilobase pairs (kbp). The genome structure of HSV-2 is identical, but is about 3 kbp larger (Morse et al., 1978) in the S segment (Reyes et al., 1979; A. J. Davison, unpublished observations). (b) BamHI restriction maps of HSV-1 strain 17 DNA (sites above the represented genome) and HSV-2 strain HG52 DNA (sites below) (Wilkie et al., 1978).

**METHODS**

*Growth of virus.* Viruses used in this study were HSV-1 strain 17 (Glasgow), HSV-1 strain USA-8 (Lonsdale et al., 1979) and HSV-2 strain HG52 (Timbury, 1971). Stocks were grown and titrated at 37 °C as described previously (Brown et al., 1973; Marsden et al., 1976) in baby hamster kidney cells (BHK C13) cultured in Eagle's medium containing twice the normal concentrations of vitamins and amino acids, 10% (v/v) calf serum, and 10% (v/v) tryptose phosphate broth. Virus plaques were picked from infected cell monolayers which had been incubated under Eagle's medium containing 2% (v/v) human serum.

*Preparation of virus DNA.* Virus DNA was prepared by the method of Wilkie (1973). Eighty oz roller bottles of BHK cells were infected with stock virus at a multiplicity of infection of 1:300 and incubated at 37 °C for 2 to 3 days. Infected cells were harvested by centrifugation and cytoplasmic virus extracted by treatment with Nonidet P40 (BDH). Virus was centrifuged from the cell medium and cytoplasmic extract, and DNA was prepared by extraction with SDS and phenol and purified by isopycnic banding on caesium chloride gradients. DNA concentrations were estimated from absorbance at 260 nm.

*Enzymes.* All enzymes were obtained from Bethesda Research Laboratories or New England Biolabs, with the exceptions of T4 polynucleotide kinase (P-L Biochemicals) and
SmaI (the gift of Janet Arrand). DNA was digested with restriction endonucleases at 37 °C in 50 μl 0-006 M-tris-HCl pH 7-5, 0-006 M-MgCl₂, 0-006 M-mercaptoethanol, 0-02 M-KCl and 0-2 mg/ml bovine serum albumin.

Gel electrophoresis. DNA fragments were separated by either agarose gel (Wilkie & Cortini, 1976) or polyacrylamide gel electrophoresis (Maniatis et al., 1975) and were visualized by ethidium bromide fluorescence under long-wave u.v. irradiation or by autoradiography. Fragments were isolated from agarose gels by hydroxylapatite chromatography (Wilkie, 1976), dialysed against 0-0015 M-sodium citrate pH 7-5, 0-015 M-NaCl, and ethanol-precipitated in some instances in the presence of 10 μg/ml yeast RNA.

32P-labelling of virus DNA. Virus DNA was labelled in vivo with [32P]orthophosphate (PBS 11 from The Radiochemical Centre, Amersham) as described by Preston et al. (1978), and extracted from combined cell-released virus with SDS and phenol (Wilkie, 1973). Restriction endonuclease digestion of purified 32P-labelled DNA was done in the presence of 4 μg/ml lambda DNA (Bethesda Research Laboratories).

Autoradiography. Wet or dried agarose or polyacrylamide gels were autoradiographed using Kodak X-Omat H film. In most cases gels were exposed to preflashed film at -70 °C with phosphotungstate screens (Dupont Cronex).

Construction of recombinant plasmid DNA molecules. DNA (100 μg) of HSV-1 strains 17 or USA-8 or HSV-2 strain HG52 was digested with BamHI and the fragments separated by agarose gel electrophoresis in the absence of ethidium bromide. Bands were located by staining strips from the edges of the gel in electrophoresis buffer containing 1 μg/ml ethidium bromide. Fragments containing the joint between the L and S segments of the genome were isolated and ligated into the BamHI site of pAT153 (Twigg & Sherratt, 1980) essentially according to the method of Tanaka & Weisblum (1975). Ligation was carried out with a total DNA concentration of 50 μg/ml, using a 10-fold molar excess of HSV DNA fragment. Ligated DNA was transfected into Escherichia coli K12 strain HB101 (Boyer & Roulland-Dussoix, 1969) by the method of Cohen et al. (1972), and bacteria were plated on agar plates containing L-broth (0-17 M-NaCl, 10 g/l Difco bactotryptone, 5 g/l yeast extract) and 100 μg/ml ampicillin (Pfizer). Resulting colonies were grown on gridded nitrocellulose filters (Schleicher & Schull) situated on L-broth agar plates containing 100 μg/ml ampicillin, and colonies containing plasmids with HSV DNA inserts were identified by colony hybridization (Grunstein & Hogness, 1975), using as a probe the virus DNA fragment which has been nick translated by the method of Rigby et al. (1977). Bacterial stocks were prepared from single colonies and stored at -20 °C in 10 g/l Difco bactopeptone in 40% (v/v) sterile glycerol. Procedures involving living bacteria subsequent to ligation of DNA were carried out under Category II containment conditions, having obtained the approval of the Genetic Manipulation Advisory Group.

Preparation of plasmid DNA. One to two l of LB-broth containing 100 μg/ml ampicillin was inoculated with a 20 ml culture of bacteria in the same medium, and grown with vigorous aeration overnight at 37 °C. Bacteria were harvested by centrifugation and lysed using lysozyme treatment and Nonidet P40 extraction by the method of Komano & Sinsheimer (1968), and plasmid DNA isolated essentially as described by Clewell & Helinski (1970). Supercoiled DNA was purified by isopycnic banding on caesium chloride gradients containing 500 μg/ml ethidium bromide (Wilkie et al., 1979). The numbers of positive clones from which DNA was prepared were eight for HSV-1 strain 17, three for HSV-1 strain USA-8, and nine for HSV-2 strain HG52. The size of each HSV DNA insert was ascertained by comparative agarose gel electrophoresis of plasmid DNA and virus DNA digested with BamHI.

Construction of restriction endonuclease maps. Nick-translated DNA representing the region of the joint between the L and S segments was prepared from both virus and plasmid
DNA, and restriction maps deduced from experiments involving digestion with restriction endonucleases in combination or recleavage of isolation restriction fragments. 

_BamHI_ fragments from the genomic termini of purified virus DNA were mapped similarly. Smaller fragments containing the joint were then isolated from plasmid DNA and restriction sites mapped prior to determination of nucleotide sequences. In mapping experiments, fragments of _pBR322_ DNA of known size (Sutcliffe, 1978) were used to calibrate mol. wt.

**Determination of nucleotide sequences.** The chemical degradation method was used (Maxam & Gilbert, 1980). Specific restriction sites were labelled at 5' termini using T4 polynucleotide kinase and γ-[32P]ATP (PB 218, The Radiochemical Centre, Amersham). DNA was digested with a second endonuclease to produce uniquely labelled fragments, which were then isolated from polyacrylamide gels and subjected to the chemical degradation reactions specific for G, G + A, A > C, C + T and C residues. Products were resolved on 43 × 26 cm sequencing gels containing 8 M-urea (19% polyacrylamide) or 9.3 M-urea (8% and 6% polyacrylamide). Electrophoresis was carried out at 40 W.

**RESULTS**

_Size variability in the joint region_

Wagner & Summers (1978) have shown that the _a_ and adjacent _c_ sequences of the DNA of clonally related stocks of HSV-1 strain KOS are heterogeneous in size. We complemented and extended these observations by comparing the variable regions in the DNA of HSV-1 strain 17 and HSV-2 strain HG52.

Recently plaque-purified stocks of HSV-1 strain 17 and HSV-2 strain HG52 were plated on BHK cells, grown under Eagle's medium containing 2% human serum, and a single plaque of each picked. Ten subplaques of each were isolated, plaque-purified twice more, and a virus stock prepared at 37 °C for each parent and progeny plaque. Virion DNA was 32P-labelled _in vivo_ and _BamHI_ restriction profiles compared (Fig. 2). Terminal and joint fragments of HSV-1 DNA were the most variable in size (_BamHI k, q, s_) and variations in the joint and _S_ terminal fragments (_BamHI k_ and _q_ respectively) seemed correlated. Variation was shown to be in _q_ and not _p_, which co-migrated in some cases, by digestion with other restriction endonucleases. Less obvious variation was observed in the mobilities of HSV-1 _BamHI n_ and _x_, which span the junctions between _U_s_ and _TR_s/IR_s_, and in _BamHI z_, which maps in _U_s_. Inter- and intrastrain variability at all these loci has previously been reported for HSV-1 (Lonsdale _et al._, 1980), and we have routinely observed variation in these HSV-1 loci when present in HSV-1/HSV-2 intertypic recombinants. Fig. 2 shows, in contrast, that HSV-2 isolates varied only in _BamHI z_ and _a'_, which span the junctions between _U_s_ and _TR_s/IR_s_, and no variation in HSV-2 joint (_BamHI g_) or terminal fragments (_BamHI u_ and _v_) was detected. Variability was shown to reside in _z_ and not _y_ by digestion with other restriction endonucleases. HSV-2 _BamHI g_ does vary in size, however, between different clinical isolates of HSV-2 (S. Chaney, personal communication). We have observed occasional variation in the _c_ sequence of HSV-2 _BamHI g_ when present in intertypic recombinants, but not in the _a_ sequence (A. J. Davison, unpublished observations).

The region of size variability in the HSV-1 joint region was located more precisely by _SmaI_ digestion of _BamHI_ joint fragments isolated from _in vivo_ 32P-labelled virion DNA. The joint region of clonally related HSV-2 isolates was analysed similarly. Results for HSV-1 are shown in Fig. 3, and _SmaI_ restriction maps are shown in Fig. 4. _SmaI G_, which is common to the joint and both termini of HSV-1 DNA and therefore is contained within the _a_ sequence, varied in size between 275 and 400 bp seemingly by the insertion or deletion of multiples of approx. 11 to 12 bp. HSV-1 _SmaI A_ exhibited size variability of up to 400 bp which was independent of that in _SmaI G_. Recleavage of _BamHI_ fragments containing the genomic
Nucleotide sequences of HSV joint regions

Fig. 2. Autoradiograph showing BamHI restriction profiles of in vivo 32P-labelled virion DNA from ten stocks of HSV-1 strain 17 and their parent stock (P), and from ten stocks of HSV-2 strain HG52 and their parent stock (P). Gel electrophoresis was carried out on 0.7% (HSV-1) and 1% agarose (HSV-2). Joint and terminal fragments, and fragments exhibiting size variability, are indicated. Restriction maps are shown in Fig. 4.

termini showed, for each isolate, the same size distribution of variable fragments as in the joint region (data not shown). No variation in the BamHI joint fragments of clonally related HSV-2 strain HG52 isolates was detected by cleavage with other restriction endonucleases (data not shown).

HSV-1 BamHI joint fragments cloned in E. coli, like those isolated directly from virion DNA, showed size variation in SmaI A and G (Fig. 5), whereas no variation was observed in cloned HSV-2 joint fragments (data not shown) except that due to the presence of an extra a sequence as described below. One of the HSV-1 clones, k2, had a smaller SmaII fragment, although variability in this fragment was not observed in virion DNA.

These results show that the genome of HSV-1 strain 17 possesses variable regions in the a and e sequences similar to, but more pronounced than those of HSV-1 strain KOS (Wagner & Summers, 1978), whereas no variation was detected in HSV-2 strain HG52 in these regions. Although HSV-1 and HSV-2 differ in this respect, they both possess the second type of heterogeneity in the joint region and L terminus which consists of insertions of approx. 300 bp thought to contain the a sequence (Wilkie, 1976; Wilkie et al., 1977; Wagner & Summers, 1978; Locker & Frenkel, 1979; A. J. Davison, unpublished observations). We have not detected similar insertions at the S terminus of HSV-1 strain 17 or HSV-2 strain HG52 (data not shown). Therefore, the majority of HSV DNA molecules has a single a sequence at the joint and long terminus, while a smaller proportion has two or more. This is reflected in cloned HSV-1 joint fragments, most of which had a single a sequence (k1, k2, k3 in Fig. 5) but some of which had two (k1, k2). Plasmids k1 and k2 each had two SmaI G fragments, the latter with two of the same size but the former with two of different sizes, and SmaI profiles were consistent with the presence of the two a sequences as tandem direct repeats (see Fig. 4 for restriction map). Similar results were obtained with clones HSV-2 joint fragments, and an isolate with two a sequences (g1) is included in Fig. 5.

Definition of the genomic termini and location of the a sequence

Fig. 6 shows the strategies adopted to determine the nucleotide sequences shown in Fig. 8. Sequences were determined for the joint regions of HSV-1 strains 17 and USA-8 and HSV-2
Fig. 3. Autoradiograph showing *Sma*I restriction profiles of *Bam*H1 joint fragments isolated from *in vivo* 32P-labelled DNA from ten HSV-1 strain 17 stocks, their parent (P), and a mixture of the ten fragments (M). Gel electrophoresis was carried out on a 7.5% polyacrylamide gel, and the portion below J represents a longer exposure of the gel. Fragment sizes (bp) were calibrated with *Hae*III and *Hin*I digests of pBR322 DNA (Sutcliffe, 1978), and size ranges of the variable fragments *Sma*I A and G are indicated. See Fig. 4 for map.
Nucleotide sequences of HSV joint regions

Fig. 4. Smal restriction maps of the BamHI fragments of HSV-1 DNA which contain the joint (6 kbp) and terminal regions. The two termini are aligned with the joint so that the polarities of the three regions, including the a sequences (rectangles), are equivalent. The remainder of the genome is indicated by a dotted line. X₁ and Xₛ denote the Smal terminal fragments of the genome. The presence of two tandem directly repeated a sequences at the joint, as shown in the lower illustration, introduces a second Smal G fragment, and a fragment P' of 43 bp which is the same size as Smal P.

strain HG52, and also for the region spanning the junction between two directly repeated a sequences in HSV-1 strain 17 and HSV-2 strain HG52.

The a sequence is defined as the sequence present as direct repeats at the genomic termini and as an inverted repeat at the joint between the L and S segments. Hence the termini may be aligned with the joint region as shown in Fig. 4. The position of the a sequence at the joint cannot be determined directly from the joint nucleotide sequence, but it can be determined by locating the termini with respect to the joint nucleotide sequence, using their common restriction sites. It is assumed, although it has not yet been proved, that the terminal and joint a sequences are identical.

HSV-1 DNA was 3²P-labelled specifically at the 5' termini and sizes of genomic terminal restriction fragments determined (Fig. 9). It is noted that this method locates the 5' termini of the genome, and not the 3' termini, with respect to the joint sequence. The sizes in nucleotides for the S terminus of strain 17 were: Smal 36-37, HaeIII 23-24. The difference in size of these two fragments is two nucleotides less than would be predicted from the joint nucleotide sequence in Fig. 8 (a), in which the Smal and HaeIII sites are located after nucleotides 367 and 352 respectively. This discrepancy is probably a result of a high proportion of C moieties (10/15) in the labelled strand between the two sites, which increases the overall charge of mass ratio of the larger fragment, causing it to migrate faster in the gel. If this is so, the smaller HaeIII terminal fragment provides the more accurate estimate, placing the S terminus at 329-330 on the joint sequence. The sizes in nucleotides for the L terminus in strain 17 were: SsrII 15-16, HaeIII 9-10, Smal 6-7. These restriction sites are located in the joint sequence after nucleotides 711, 719 and 722 respectively, bearing in mind that in this case the labelled strand is complementary to that shown in Fig. 8 (a). For the reason described above the smaller fragments might provide a better estimate of the nucleotides defining the L terminus (728-729). The HSV-1 strain 17 a sequence in plasmid k1 therefore has a size of
Fig. 5. *SmaI/BamHI* restriction profiles of recombinant plasmids bearing the *BamHI* joint fragment of HSV-1 strain 17 or HSV-2 strain HG52 DNA. Two μg DNA was applied to each track of a 4% polyacrylamide gel, which after electrophoresis was stained with ethidium bromide and photographed under long-wave u.v. illumination. The plasmid vector *pAT153* contains no *SmaI* sites and therefore migrates as a large fragment in each track. Clones k1, k2 and k3 contain the HSV-1 joint fragment with a single a sequence, k1 and k2 contain the same fragment with two a sequences, and g1 contains the HSV-2 joint fragment with two a sequences. HSV-1 *SmaI* fragments are indicated on the left (see Fig. 4 for map), with size ranges of A and G, and sizes in bp of φX174 DNA *HaeII* markers, *M* (New England Biolabs) on the right. Standard reference numbers of the clones are: k1, pGX2; k2, pGX7; k3, pGX5; k1, pGX4; k2, pGX8; g1, pGZ1. The reference number of the HSV-1 strain USA-8 *BamHI* k clone is pGX10, and that of the HSV-2 strain HG52 *BamHI* g clone, for which part of the sequence is shown in Fig. 8 (b), is pGZ6.

398 to 400 bp (between 329–330 and 728–729) but, as is explained below, differed in size in other plasmids.

The genomic termini of HSV-1 strain USA-8 were defined by the same criteria (Fig. 9) and are shown in Fig. 8 (a) at 330–331 and 728–729 with respect to the strain 17 sequence. From the autoradiograph reproduced in Fig. 9 it could be seen that S terminal fragments from strain USA-8 DNA were two nucleotides smaller than those from strain 17 DNA, whereas L terminal fragments were the same size in both strains. In an analogous manner the S and L termini of HSV-2 strain HG52 were located at 62–64 and 309–319 respectively (Fig. 8 b). The HSV-2 a sequence therefore has a size of 245 to 257 bp.

In both HSV-1 strains and the HSV-2 strain the genomic termini are located close to the right-hand end of direct repeats in the joint sequences. The direct repeat in strain USA-8 comprises the first 17 nucleotides of the 21 nucleotide direct repeat in strain 17 at 311–331
Fig. 6. Restriction endonuclease maps of the joint regions of HSV-1 strain 17 (left) and HSV-2 strain HG52 (right) and the nucleotide sequencing strategies employed. The HSV-1 fragment contains three reiterations shown in black (I, II, III), and the locations of the direct repeats at the b–a and a–c junctions are indicated. Restriction sites in this region of the DNA of HSV-1 strain USA-8 were identical to those in strain 17, except for an extra HindIII site in the c sequence 50 bp from the a sequence.

and 710–730. However, the primary sequences of the HSV-1 direct repeats are different from that of HSV-2.

The nucleotide sequence of plasmids containing joint fragments of HSV-1 strain 17 or HSV-2 strain HG52 with two a sequences showed that the a sequences are tandem direct repeats with a single copy of the appropriate 21 or 17 nucleotide direct repeat at the a–a junction (Fig. 7 b, 10).

Tandem reiterations in the joint region

The HSV-1 strain 17 sequence shown in Fig. 8 a was determined using plasmid k1, but k2 was also sequenced in the regions where it differed in size from k1. Three regions of reiteration were found in the HSV-1 strain 17 joint region, one located in the a sequence and two in adjacent c sequences (reiterations I, II and III in Fig. 6, 8 a). The three repeated units of 12, 16 and 17 nucleotides of which the reiterations consist have different primary sequences, although they all contain a predominance of C residues in the same strand of the duplex. The sequence of plasmid k2 in two of the reiterations (I and III) differed from that of k1 only in the number of repeated units. Reiteration I (408–623) in k1 contained 18 copies of CCGCTCCTCCCC whereas k2 contained 21 copies. Similarly, reiteration III (1167–1489) consisted of 19 copies of CCGCCCCTCGCCCCCTC in k1 and 11 copies in k2 (Fig. 7 a). These two reiterations are contained within SmaI G and SmaI A respectively, and provide an explanation for the observed size variability in the a and c sequences of different plaque isolates of HSV-1 strain 17.

Reiteration II (928–1055) did not contain a simple repeated unit. Furthermore, the difference in sequence between plasmids k1 and k2 was not reflected in different copy numbers but in the organization of repeated units, thus:

\[
k1 \quad (\text{CCCTCCCCAGCCCCAG})_2 \ (\text{CCCTCCCCGGCCCCAG})_6 \\
k2 \quad (\text{CCCTCCCCAGCCCCAG})_7 \ (\text{CCCTCCCCGGCCCCAG})_2
\]

The two 16 nucleotide repeated units differ in a single nucleotide.

Reiterations I and II were also identified in the joint region of HSV-1 strain USA-8. Reiteration II comprised at least 15 copies of the first 16 nucleotide variant of the repeated unit of strain 17 (shown above and in Fig. 8 a). Reiteration I is related to that of strain 17 in a...
Fig. 7. (a) Autoradiograph of 8% polyacylamide sequencing gels showing the 17 nucleotide repeated unit of reiteration III in the c sequence of HSV-1 strain 17. The sequence is from the Sau3A I site of plasmids k1 and k2 in a rightwards direction (see Fig. 6). The copy numbers of repeated units are indicated. The chemical degradation reactions identify G, G + A, A > C, C + T, and C residues. (b) Autoradiograph of 8% polyacylamide sequencing gels showing the 21 nucleotide direct repeat at the a–c junction is plasmid k1 and at the a–a junction is plasmid k2, which has two tandem repeated a sequences. The sequences are from the AvaI site in a rightwards direction (see Fig. 6).

more complex way. Fig. 8(a) shows that it contains eight copies of a unit which consists of the first 11 nucleotides of the strain 17 12 nucleotide repeated unit. Adjacent to this are six copies of a unit which comprises the 11 nucleotides with the sequence TCTGTGGGTGGGG inserted after the third nucleotide. Reiteration III was also identified in strain USA-8 but the sequence was not analysed in detail.
A partial overlap of the repeated unit of reiterations I, II and III is present immediately to the right of each reiteration defined by the brackets in Fig. 8 (a). (The location of each partial overlap is arbitrary and depends on the definition of each repeated unit.) The partial overlaps consist of the first 7, 11 and 10 nucleotides of the repeated units of reiterations I, II and III respectively.

No tandem reiterations equivalent to those present in the a sequence of HSV-1 were found in the HSV-2 sequence (Fig. 8 b). We do not know whether reiterations equivalent to II and III exist in HSV-2 DNA.

**Comparison of nucleotide sequences**

The region of greatest difference in the joint sequences of HSV-1 strains 17 and strain USA-8 is in the c sequences immediately adjacent to the a sequence. The differences in this region, which is AC rich on the strand shown in Fig. 8 (a), represent both substitution of nucleotides and deletion or insertion of short sequences. Only one nucleotide differs between the 310 analysed nucleotides of the b sequences to the left of the first 17 nucleotide repeat at 311–327 and strains 17 and USA-8 are identical in 166 comparable nucleotides at the right-hand end of the c sequence shown (1601–1766). Reiteration I is the region of greatest difference between the a sequences of the two HSV-1 strains and there are ten base substitutions and one deleted nucleotide in the remainder of the a sequence. Since reiterations I, II and III were identified in both strains, although these differ widely in time and geography of isolation, it seems likely that they will be a general feature of HSV-1 isolates.

The a sequences of HSV-1 and HSV-2 both have a high GC content (83% and 84% respectively), but the sequences show little similarity. Within the a sequences the longest similar sequence, which is also highly asymmetric between the two strands of the duplex, is 362–389 of HSV-1 strain 17 and 94–123 of HSV-2 strain HG52. Unlike HSV-1, HSV-2 does not possess reiteration I, but both serotypes have a direct repeat at the b–a and a–c junctions.

**Expression of the HSV-1 joint region**

There is as yet no evidence that transcription starts in the nucleotide sequence of the HSV-1 joint region shown in Fig. 8 (a), and no sequences consistent with the consensus 'Hogness box' (Gannon et al., 1979) are present. The hexanucleotide AAUAAA has been found in close proximity to the 3' termini of the majority of polyadenylated eukaryotic mRNAs so far analysed (Benoist et al., 1980). A few exceptions, which possess a sequence closely related to AAUAAA, have been noted (MacDonald et al., 1980; Hobart et al., 1980; Goeddel et al., 1981). The DNA sequence AATAAA is located at two positions in the strain 17 joint region, at 1500–1505 and on the opposite strand at 1546–1541. Benoist et al. (1980) have observed a conserved sequence related to TTTTCACTGC (more accurately \( T^T, T^C, G^G, T^A, A^G \)) located 3' to the AATAAA close to the polyadenylation site of several, but not all, mRNAs, and we find the related sequence TTTGCAGTAG at 1520–1511 3' to the AATAAA at 1546–1541 (considering the other strand of the duplex from that shown). No such sequence is present 3' to the other AATAAA. These results suggest that a mRNA is transcribed from right to left and is polyadenylated in the region 1530–1520. F. Rixon et al. (unpublished results) have shown that the mRNA coding for immediate early polypeptide \( V_{mw} \) IE 175 indeed is transcribed in this direction, and that the 3' terminal-coded nucleotide of this RNA is located in the region 1530–1525. They were unable to detect transcripts from the other strand terminating close to the AATAAA at 1500–1505 either in immediate early mRNA or in mRNA from cells harvested 6 h post-infection at 37 °C.

It is not yet known whether joint sequences other than those coding for \( V_{mw} \) IE 175 mRNA are represented in stable transcripts, or where primary transcription terminates. The
region contains several potential splice donor (PuGTXXG) and acceptor (PyPyXPYAG) signals (Seif et al., 1979), but these cannot be interpreted in the absence of transcript mapping data. Three features of the DNA sequence make it likely that a considerable part of the joint region does not code for polypeptide, notably the a and adjacent c sequences. First, translation is unlikely to occur through reiteration I in the a sequence. Variation in the copy number of the 11 nucleotide repeated unit of this reiteration in strain USA-8 could result in translation of each of the three possible downstream reading frames. Similar arguments apply to the c sequence adjacent to a, which contains reiterations II and III. Second, if translation of mRNA representing reiteration I of strain 17 occurred, the polypeptide would contain the unusual structure of 18 tandem repeats of four amino acids (plasmid k1). Similarly, reiteration III would code for a polypeptide containing six tandem repeats of 17 amino acids. Third, the absence of nucleotides from one strain compared with the other would generate translational frameshifts. Strain USA-8 has one nucleotide inserted between 328 and 329 of strain 17, one nucleotide deleted at 340, and five nucleotides inserted between 740 and 741 which would produce this effect.

**DISCUSSION**

Determination of the DNA sequences at the joint region between the L and S segments of the genomes of HSV-1 and HSV-2 has revealed several features. The presence of reiterations
in HSV-1 provides a ready explanation for the widely observed size variability in the $a$ and $c$ sequences. The HSV genome is known readily to undergo recombination during replication, and unequal crossover events between tandem reiterations would give rise to size variability in these regions. Similar tandem reiterations might be responsible for such variability elsewhere in the genome. A functional role for the reiterations has not yet been identified, and it is possible that they encode no essential functions but represent regions in which errors in DNA replication can accumulate. The absence of reiteration I in the $a$ sequence of HSV-2 strain HG52 argues against an essential role for that tandem reiteration in virus replication. The 11 nucleotides CCCGCTCCTCC which form the first repeated unit of reiteration I in HSV-1 strain USA-8 [in this case redefining the unit shown in Fig. 8 (a) by moving the brackets one nucleotide to the left] are also present at 871–881 (with respect to the strain 17 sequence), showing that this sequence is not amplified of necessity in the HSV genome. The repeated units of the reiterations, although different, have similarities such as a preponderance of C residues on one strand of the duplex and G residues on the other, and it is possible that such sequences are more readily amplified than others. It is of interest to note that the sequence CPyCCXCCC is present in reiterations I and III of both HSV-1 strains; indeed it is present on both strands of the 24 nucleotide repeated unit of reiteration I in strain USA-8, and GGCTCCTCC is present in reiteration II. The SV40 genome possesses six copies of PyPyCCGCCC close to the origin of DNA replication (Dhar et al., 1977), two of which are within T-antigen binding site III (Tjian, 1978). However, the presence of this binding site is not essential for SV40 DNA replication (Subramanian & Shenk, 1978). T-antigen binding site II, the presence of which is required for DNA replication (Shenk, 1978; Gluzman et al., 1980), possesses two palindromes of GCCTC (Tjian, 1978), and the HSV-1 reiterations contain PyCCTC. Such sequences however, are sufficiently short to be present commonly in HSV-1 DNA (e.g. PyPyCCGCCC at 1688–1695 and GCCTC at 1686–1690), and so it remains to be shown that the sequences of the repeated units are specific for preferential amplification.

In both serotypes the S terminus is equivalent to a position close to the right-hand end of the 17 and 21 nucleotide direct repeat at the $b$–$a$ junction, and the L terminus is equivalent to the right-hand end of the other direct repeat at the $a$–$c$ junction. Although the locations of the termini with respect to the joint sequence have not been defined to single nucleotides, Fig. 8 (a) shows that the HSV-1 strain USA-8 S terminus is one nucleotide to the right of that in HSV-1 strain 17, whereas the L termini coincide. A minority of virion DNA molecules contain more than one copy of the $a$ sequence at the L terminus and at the joint. Determination of the nucleotide sequences between two tandem HSV-1 or HSV-2 $a$ sequences has shown that the $a$ sequences are arranged as head-to-tail dimers which have one copy of the direct repeat at their point of fusion (Fig. 10). This arrangement could arise either by misaligned genetic recombination at the direct repeats, or by direct ligation of the L and S
Nucleotide sequences of HSV joint regions

The fact that multiple a sequences have not been observed at the S terminus of HSV-1 strain 17 or HSV-2 strain HG52 presumably reflects the mechanism of DNA replication and maturation. Our observations raise the possibility that the direct repeats serve as recognition sites for the maturation of unit length DNA molecules from concatemers. It is likely that concatemeric DNA is specifically cleaved at alternate joint regions to produce DNA of genome length, but there is as yet insufficient evidence upon which to base detailed models. Nevertheless, the consequence of the maturation mechanism is that the L terminus essentially contains both copies of the direct repeat, whereas the S terminus contains only the a-c copy (Fig. 10).

A role for the a sequence in segment inversion is suggested by two observations. Firstly, intertypic recombinants which do not invert normally in L or L and S have totally heterotypic b and a or c and a sequences bounding the affected unique region(s). However, recombinants with otherwise identical structure, but which have homologous a sequences, invert normally (A. J. Davison & N. M. Wilkie, unpublished results). This strongly suggests that a sequence homology alone is sufficient to permit normally observed inversion. The second observation stems from experiments in which a joint fragment of 4-2 kbp was inserted into Ul (Mocarski et al., 1980; J. Smiley, personal communication). This resulted in additional rearrangements about the novel joint, and was interpreted by Mocarski et al. (1980) as indicating site-specific recombination. Precisely which sequences are involved in the proposed site-specific recombination events is unknown. One possibility is that the region which is conserved between the HSV-1 and HSV-2 a sequences [362-389 in Fig. 8(a) and 94-123 in Fig. 8(b)] is a recombinator sequence analogous to chi in prokaryotes (Smith et al., 1980) and cog in fungi (Catcheside & Corcoran, 1973), promoting high local rates of recombination, or that it is equivalent to att in lambda (Landy & Ross, 1977), promoting recombination in a specific sequence. These hypotheses could be tested by transferring this sequence to any other readily studied genome.

Neither the number of modes nor the mechanism of HSV DNA replication and recombination are known. Roizman (1979) has proposed a rolling circle model in which the resulting concatemers are cleaved at one side of a single a sequence to produce linear molecules lacking an a sequence at one terminus. The a sequence is regenerated by recombination and replication at the joint with concurrent inversion of the segments. This model predicts obligatory identity of the a sequences at joint and termini. The prediction found apparent support in the claimed location by marker rescue of the mutation tsc75 (HSV-1 strain 13vB4) in the a sequences at the joint and both termini, and led to the conclusion that the a sequence is part of the structural gene for V_mw IE 175 (ICP4; Knipe et al., 1979). These results and conclusions have since been discounted by the work of V. Preston, who has located this mutation by marker rescue with cloned DNA fragments about 2 kbp from the a sequence in the c sequence between the tsD and tsK mutations (in press), by the work described here, and by location of the 3' terminus of V_mw IE 175 mRNA in the c sequence by F. Rixon et al. (unpublished results). Nevertheless, our observation that ten HSV-1 subclones with varying copy numbers in reiteration I at the joint have identical copy numbers at the termini could be interpreted as supporting a model involving facultative identity of a sequences. Any model invoking obligatory identity of a sequences is disproved by the existence of intertypic recombinants with heterologous a sequences bounding the L or the L and S segments (A. J. Davison & N. M. Wilkie, unpublished results). Further investigation of the functions of the a sequence will distinguish between essential and optional events related to it in DNA replication and maturation.

We are grateful to Professor J. H. Subak-Sharpe, Duncan McGeoch, Russell Thompson and Nigel Stow for critical reading of the manuscript, and to Frazer Rixon for allowing us to quote his
unpublished data on the mapping of 3' terminus or \( V_{mw} \) IE 175 mRNA. We wish to thank Mrs Cathie Adair for expert technical assistance.

**REFERENCES**


Nucleotide sequences of HSV joint regions


(Received 30 March 1981)