Conversion of a Fraction of the Unique Sequence to Part of the Inverted Repeats in the S Component of the Herpes Simplex Virus Type 1 Genome

By KENICHI UMENE

Department of Virology, Faculty of Medicine, Kyushu University 60, Fukuoka 812, Japan

(Accepted 21 February 1986)

SUMMARY

A novel genome variant of herpes simplex virus type 1 (HSV-1) was isolated, in the S component of which a fraction of the unique sequence (map units 0.865 to 0.880) was converted to part of the diploid inverted repeats and another fraction of the unique sequence (map units 0.937 to 0.955) had been deleted. The S component of the variant consisted of a shortened unique sequence (map units 0.880 to 0.937) and a pair of elongated inverted repeats (map units 0.820 to 0.880, and 0.937 to 1.000). The conversion occurred as a result of a recombination event between two points (map units 0.880 and 0.937) having a 5 base pair stretch (5'-CCCCG-3') of homology, in an inverted direction in the unique sequence. An involvement of the mechanism causing L-S inversion was inferred to account for the generation of the variant, as there are multiple copies of the 5'-CCCCG-3' stretch in the ‘a’ sequences. The occurrence of the variant indicates that the products of HSV-1 genes US9, US10, US11 and US12 are unnecessary for an HSV-1 productive infection in tissue culture cells, and also suggests the presence of a mechanism whereby expansion of the inverted repeats could occur.

INTRODUCTION

The genome of herpes simplex virus type 1 (HSV-1) is a linear duplex DNA molecule of about 160 kilobase pairs (kb), and is composed of two linked components, L and S. Each component consists of unique sequences (UL and US) flanked by inverted repeats (RL and Rs) (Fig. 1a). The two components, L and S, invert relative to each other during viral replication, and the DNA extracted from HSV-1 virions consists of an equimolar mixture of four isomers, differing in the relative orientations of the L and S components (Roizman, 1979).

Although the mechanism resulting in the inversion of the two components remains unclear, evidence has been presented that inversion is due to site-specific recombination between inverted copies of the reiterated ‘a’ sequences and is mediated by trans-acting viral gene products (Davison & Wilkie, 1981; Mocarski et al., 1980; Mocarski & Roizman, 1981, 1982a, b; Smiley et al., 1981). The inverted repeat sequences of HSV-1 remain identical as a consequence of high-frequency recombination events initiated within the ‘a’ sequence, and this may also result in the inversion of the L and S components (Hubenthal-Voss & Roizman, 1985; Knipe et al., 1978; Varmuza & Smiley, 1984). Pogue-Geile et al. (1985) showed that an inverted duplication of a sequence from UL (BamHI-L, map units 0-706 to 0-744) resulted in inversion of the DNA sequences flanked by the inverted copies of the duplication, independently of the ‘a’ sequences. If both of these genome regions, i.e. ‘a’ and BamHI-L, contain signals for site-specific recombination, then presumably the signals are for different systems. The physiological function of inversion is unknown, and a viable, non-inverting HSV-1 genome derived by deletion of sequences at the L–S junction has been isolated (Poffenberger et al., 1983). If there is a physiological requirement for inversion, it would be expected to occur, perhaps in other ways, in many herpesvirus DNAs, but only two of the five known herpesvirus DNA structures contain components that invert (Roizman et al., 1981).
I now report a novel HSV-1 variant, in the S component of which a fraction of the unique sequence has been converted to part of the inverted repeat sequences, after a recombination event between a pair of 5 base pair (bp) stretches of homology, arranged in opposing directions. This may have been caused by a mechanism similar to that involved in the L–S inversion.

METHODS

Cells and viruses. Vero cells were grown in Eagle's MEM supplemented with 5% calf serum. SP23, used as the standard HSV-1 in this work, is a single plaque isolate from HSV-1 strain Patton (Umene & Enquist, 1985). N38 is the HSV-1 variant characterized in this work and was derived from SP23. Working stocks of HSV-1 were made on Vero cells in MEM with 2% foetal bovine serum (FBS) at a low m.o.i. (0.01 p.f.u./cell).

HSV-1 DNA was prepared from viral particles obtained after glycerol gradient centrifugation, as described by Denniston et al. (1981).

One-step virus growth curves. Vero cells were infected at a m.o.i. of 5 p.f.u./cell in Petri dishes (Falcon no. 3001) (Dargan & Subak-Sharpe, 1985). After 2 h adsorption at 37 °C, the cells were washed with MEM three times and overlaid with MEM containing 2% FBS. Incubation at 37 °C was continued until the infected cells were harvested by scraping into the growth medium at 2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 28, 36 and 48 h after adsorption. After ultrasonic disruption of the cells, the infectious virus yields from each sample were titrated in Vero cells at 37 °C.

Phages and cloning of a DNA fragment in λgtWES. As a standard, a hybrid lambda phage, Dec36, which carries the EcoRI-H fragment of HSV-1 strain Patton was used (Umene & Enquist, 1981a). The EcoRI-H fragment of HSV-1 N38 was cloned in λgtWES, as described by Enquist et al. (1979), and the hybrid phage was named λNTE-1. The DNAs of hybrid lambda phages were extracted from lambda phage particles purified through glycerol gradients, according to Umene & Enquist (1981a).

Restriction enzyme digestion, gel electrophoresis and Southern hybridization. Restriction endonucleases were purchased from Takara Shuzo Company (Kyoto, Japan) and Nippon Gene Company (Toyama, Japan), and digestion conditions were as recommended by the manufacturers. DNA markers and gel electrophoresis procedures were as described by Denniston et al. (1981). Southern hybridization was carried out on a Biodyne A transfer membrane (Pall Ultrafine Filtration Corp.) as described by Umene (1985a).

DNA sequencing. An appropriate restriction fragment was subcloned into pUC18 (Norrander et al., 1983), and was sequenced by the dideoxynucleotide chain termination procedure described by Sanger et al. (1977).

RESULTS

Isolation of an HSV-1 variant, N38, lacking the BamHI-Z fragment (map units 0.936 to 0.949)

To study intermolecular recombination of the HSV-1 genome, a viral suspension was prepared by co-infecting two HSV-1 strains, SP23 and Isolate No. 3 (Umene et al., 1984), on Vero cells, each at an input multiplicity of 20 p.f.u./cell (Umene, 1985b). For clonal isolation of the progeny viruses, the virus stock was titrated, diluted, added to Vero cell suspensions, and distributed in 96-well microtest plates (Falcon no. 3072). Each well received a volume of 0.2 ml containing 5 × 10⁴ cells and 0.3 p.f.u. of virus (Manservigi, 1974). In the course of the study of the genome structures of 100 single-clonal isolates from the viral suspension prepared after co-infection, an HSV-1 variant named N38 was isolated. This variant was distinguished from other HSV-1 isolates by the striking feature that the BamHI digest of N38 DNA did not contain an obvious fragment which hybridized with the BamHI-Z fragment (map units 0.936 to 0.949) of HSV-1. This observation suggested the presence of a variation in the S component of the N38 genome. There were no evident variations in the L component of the N38 genome. The two HSV-1 strains, SP23 and Isolate No. 3, can be differentiated with respect to seven restriction enzyme cleavage sites in the L component, detectable in the form of a restriction fragment length polymorphism (Umene et al., 1984). The variant N38 was not distinguishable from SP23 by this criterion, and was assumed to have been generated from SP23 after a genetic rearrangement of the S component of SP23.

Twelve secondary single-plaque clones of the variant N38 were isolated by the method described above, and the DNAs of these clones were propagated and analysed. All 12 clones retained the same BamHI cleavage pattern as in the primary N38 clone. Thus, the variation in the N38 genome was stable, and the possibility of the presence of a wild-type virus acting as a helper to propagate the variant genome was excluded.
Conversion of $U_s$ to $R_s$ in HSV-1 genome

Fig. 1. Maps of HSV-1 DNA. (a) Structure of the HSV-1 genome. HSV-1 DNA consists of two covalently linked components, L and S, that constitute 82 and 18% of the genome, respectively. Each component is composed of unique sequences ($U_L$ and $U_S$) (thin lines) bracketed by inverted repeat sequences ($TR_L$, $IR_L$, $IR_S$ and $TR_S$) (thick lines). A short sequence 'a' is repeated directly at the termini of the genome and is also present in inverted orientation at the L-S junction (Roizman, 1979). The $EcoRI-H$ fragment, containing the entire $U_S$ region and about 1 kb of $IR_S$ and $TR_S$, is indicated. (b) Map of the $EcoRI-H$ fragment in the P (prototype) arrangement (Umene & Enquist, 1981a). Restriction sites are abbreviated as follows: $B_1$ to $B_5$, $BamHI$; $Bg$, $BglII$; $E_1$ and $E_2$, $EcoRI$; $H_1$ and $H_2$, $HindIII$; $K_1$ and $K_2$, $KpnI$; $P_1$ to $P_3$, $PvuI$. The $BamHI-Z$ fragment (map units 0.936 to 0.949) is indicated ($B_3$ to $B_2$). The unique sequence and the inverted repeat sequences are represented by a thin line and thick lines, respectively. (c) $SmaI$ restriction map of the $EcoRI-H$ fragment. The numerals on the map represent the size of each $SmaI$ fragment in kb, as determined in previous work (Umene & Enquist, 1981 b). The thick line indicates the region of $SmaI$ fragments duplicated in the N38 genome, and the wavy line indicates the region of the $SmaI$ fragments absent from the N38 genome.

Structure of the S component of the N38 genome

The DNA of N38 was analysed by Southern hybridization procedures to determine the nature of the variation in the S component. The DNA fragments of N38 digested with $BamHI$ were transferred to a Biodyne A transfer membrane after separation on a 1% agarose gel, and were hybridized with $^{32}P$-labelled cloned $EcoRI-H$ probe (Dec36) (Fig. 2a). The $BamHI$ digest of N38 DNA lacked two fragments, one of 1.9 kb ($BamHI-Z$) and another of 1.85 kb (which includes the region from $B_1$ to $E_1$ shown in Fig. 1b), and contained an extra fragment of 3.6 kb, in comparison with the digest of SP23 DNA (Fig. 2a). Two $BamHI$ fragments of 6.7 kb ($B_4$ to $B_5$ in Fig. 1b) and 5.2 kb (which includes the region from $E_2$ to $B_5$ in Fig. 1b) were present in digests of both N38 and SP23. No significant difference was found between N38 and SP23 in the total length of $BamHI$ fragments hybridizing with the $EcoRI-H$ probe. From these observations, it was conjectured that the variation in the N38 genome was due to a replacement of a region between $B_4$ and $E_1$ (Fig. 1b) by other sequences.

The $EcoRI-H$ fragment of HSV-1 can be cleaved into about 30 $SmaI$ fragments (Umene & Enquist, 1981 b), and the conservation of any region in the $EcoRI-H$ fragment can be deduced from the presence of the corresponding $SmaI$ fragments. If a region in the $EcoRI-H$ fragment is deleted, the $SmaI$ fragments that derive from the region will be absent. The Southern hybridization profile of N38 DNA digested with $SmaI$, using Dec36 DNA as a probe, was compared with that of SP23 DNA after separation on a 1% agarose gel (Fig. 2b) and on a 5% acrylamide gel (Fig. 3). The nomenclature of the $SmaI$ fragments is based on the sizes of the
Fig. 2. Southern hybridization analysis of genome structure of N38, on agarose gels. The DNAs of SP23 (standard; lane S) and N38 (variant; lane N) were digested with (a) BamHI, (b) SmaI, (c) SalI, (d) KpnI and (e) BglII, and were electrophoresed on 1% (a, b, c) and 0.5% (d, e) agarose gels. After transfer to Biodyne A membranes, the blots were hybridized with 32P-labelled Dec36 (EcoRI-H) (a to d) or 32P-labelled pRWA-I (e). The plasmid pRWA-I carries the BamHI-(T+R) fragment that spans the L-S junction of HSV-1 (Umene et al., 1984). The sizes of the fragments are shown in kb (Locker & Frenkel, 1979; Umene & Enquist, 1981 a, b). The sizes of the fragments that were present in digests of SP23 and absent in digests of N38 are underlined. The fragments which were detected only in the digests of N38 are indicated by a horizontal arrow.

fragments given in kb, as determined previously on a 10% acrylamide gel containing 25% glycerol (Umene & Enquist, 1981 b). In previous work, the 0.66 and 0.52 kb bands were observed to be doublets (Umene & Enquist, 1981 b). In the present work, using a 5% acrylamide gel, each of the 0.66 and 0.52 kb bands was discerned to be separate; these were named 0.66F and 0.66N, and 0.52H and 0.52M, respectively (Fig. 1 and 3). The letter attached as a suffix was derived from the names of the groups of SmaI fragments, determined previously (Umene & Enquist, 1981 b). The relative mobility of the 0.74 and 0.72 kb SmaI fragments depends on the conditions of electrophoresis. The 0.72 kb SmaI fragment ran faster on the 10% acrylamide gel containing 25% glycerol, but slower on the 1% agarose gel, than the 0.74 kb SmaI fragment, as reported previously (Umene & Enquist, 1981 b) (Fig. 2b). The 0.74 kb SmaI fragment ran faster than the 0.72 kb SmaI fragment on the 5% acrylamide gel (Fig. 3 and 4).

The SmaI digest of N38 DNA lacked seven fragments of 1.7, 0.72, 0.44, 0.32, 0.135, 0.125 and 0.105 kb (these sizes are underlined in Fig. 2b and in Fig. 3), and it contained an extra fragment of 1.3 kb (indicated by a horizontal arrow in Fig. 2b and in Fig. 3), in comparison with that of SP23 DNA. The absence of the three SmaI fragments of 1.7, 0.72 and 0.135 kb from the SmaI digest of N38 coincided with the former result, i.e. the absence of the BamHI-Z fragment from the N38 genome (see Fig. 1 b, c). The seven SmaI fragments missing from the N38 digest are normally located in the right-hand part of US, as shown by the wavy line in Fig. 1 (c). These results indicate the occurrence of a deletion in the right part of US in the N38 genome, and specifically that the region from a point in the 0.72 kb SmaI fragment to a point in the 0.44 kb SmaI fragment had been deleted. The total length of these seven SmaI fragments was 3.5 kb, and the length of the S component of the N38 genome was estimated to have decreased by 2.2 (3.5 - 1.3) kb, compared to that of the SP23 genome. However, the difference in the total length of the BamHI fragments hybridizing with the EcoRI-H probe was only 0.15 (1.9 + 1.85 - 3.6) kb between the BamHI digests of N38 and SP23 (Fig. 2a). The discrepancy between the total lengths of the SmaI and BamHI fragments that hybridized with EcoRI-H can be explained if it is assumed that there was a duplication in the S component of the N38 genome. If the deleted region had been replaced by the duplicated region there should have been no remarkable change in the total length.
Conversion of Us to Rs in HSV-1 genome

To analyse variations in the S component of the N38 genome in detail, the EcoRI-H fragment of N38 was cloned in λgtWES, and the hybrid phage was named λNTE-1. The SmaI digest of λNTE-1 was compared with that of Dec36, which carries the EcoRI-H fragment of strain Patton, after separation on a 5% acrylamide gel (lanes 3 and 4 of Fig. 4). Similar to the results obtained from Southern hybridization analyses in Fig. 3, the SmaI digest of λNTE-1 lacked seven fragments of 1.7, 0.72, 0.44, 0.32, 0.135, 0.125 and 0.105 kb, and it contained an extra fragment of 1.3 kb, in comparison with the digest of Dec36. Another characteristic feature of the SmaI digest of λNTE-1 was that the staining intensities of at least three SmaI fragments of λNTE-1, i.e. 0.66N, 0.52M and 0.48 kb (lane 3 of Fig. 4a), were increased over the corresponding fragments of Dec36. The staining intensity of the 0.74 kb SmaI fragment of λNTE-1 was similar to that of Dec36. The results suggested that the region from a point in the 0.66N kb SmaI fragment (probably the left Rs/Us junction) to a point in the 0.74 kb SmaI fragment, as indicated by a thick line in Fig. 1(c), had been duplicated in the N38 genome. If this region had indeed been duplicated, the five SmaI fragments of 0.66N, 0.52M, 0.48, 0.145...
Fig. 4. Analysis of Smal digests of hybrid phages carrying the EcoRI-H fragments of HSV-1, on an acrylamide gel. A photograph (a) and a schematic representation (b) of the DNAs stained with ethidium bromide after separation on the 5% acrylamide gel are shown. Lane 1, Smal digest of the 5.0 kb EcoRI–BamHI fragment of λNTE-1 corresponding to the E2 to B5 fragment in Fig. 1(b) or Fig. 5(1). The Smal–B5 fragment of 0.66 kb, characteristic of the E2 to B5 fragment, is indicated as a wavy line in lane 1 of (b). Lane 2, Smal digest of the 3.5 kb EcoRI–BamHI fragment of λNTE-1 corresponding to the B3 to E2* fragment in Fig. 5(2), which is characteristic of the N38 genome. The B3–Smal fragment of 0.64 kb, containing the novel recombination point of N38, is indicated as a wavy line in lane 2 of (b). The 0.225 kb fragment in lanes 1 and 2 is the EcoRI–Smal fragment (see Fig. 1c). Lane 3, Smal digest of λNTE-1 carrying the EcoRI-H fragment of N38. The bands indicated by arrows in lane 3 of (b) are doubled in the Smal digest of λNTE-1. The Smal fragment of 1.3 kb, containing the recombination point, is indicated as a wavy line in lane 3 of (b). Lane 4, Smal digest of Dec36. The sizes of the fragments are shown in kb, as determined previously (Umene & Enquist, 1981b). Lane 5, HaeIII digest of φX174 DNA.

and 0.084 kb would be expected to have been doubled in the Smal digest of λNTE-1. For a closer examination of the duplication in λNTE-1, two EcoRI–BamHI fragments, i.e. the 5.0 kb fragment from E2 to B5 (Fig. 1b) and the 3.5 kb fragment from B3 to E2* [see Fig. 5(2) or Fig. 6a(2)], were isolated from λNTE-1 DNA. The two isolated EcoRI–BamHI fragments were electrophoresed on an acrylamide gel after cleavage with Smal (lanes 1 and 2 of Fig. 4). The five
Conversion of \( \scriptsize{U_S} \) to \( \scriptsize{R_S} \) in HSV-1 genome

---

**Fig. 5.** Comparison of the genome structure of N38 with that of SP23. Restriction maps of the S components and the flanking fractions of L components of SP23 (line (1)) and N38 (line (2)) are presented in the middle part (Denniston et al., 1981; Locker & Frenkel, 1979; Umene & Enquist, 1981). BamHI sites (B21, B43, B5 and B6), EcoRI sites (E1, E2 and E2*), KpnI sites (K1 to K3), and Sall sites (S) are indicated as in Fig. 1 (b). E2* is the EcoRI site in the right Rs of N38, and is presumed to be derived from the EcoRI site in the left Rs of SP23 (E2) after the recombination event generating the N38 genome. The BamHI sites B21 and B43 are B2 plus B1, and B4 plus B3, respectively. The S component is arranged in the P (prototype) arrangement (Roizman, 1979). The junction between L and S components is represented by an open circle (O). The regions duplicated and deleted in the N38 genome are indicated by thick lines and a wavy line, respectively. The closed circles (●) in the SP23 genome indicate the sites of recombination, and the cross (×) in the N38 genome is the recombination point. Restriction enzyme cleavage products of SP23 hybridizing to the EcoRI-H probe (Dec36) are delineated in the upper part. Sizes are indicated in kb. Restriction enzyme cleavage products of N38 hybridizing to the EcoRI-H probe, which can be deduced from the structure of N38 in line (2), are delineated in the lower part.

Sall fragments expected to have been doubled were present in Sall digests of both the 5.0 and 3.5 kb EcoRI–BamHI fragments, as predicted. The results of analyses of Sall digests of N38 and \( \alpha \)NTE-1 DNAs suggested that in the N38 genome, the region from the left Rs/Us junction within the 0.66N kb Sall fragment to a point in the 0.74 kb Sall fragment was duplicated (thick line in Fig. 1c), and the region from a point in the 0.72 kb Sall fragment to the right Rs/Us junction within the 0.44 kb Sall fragment was deleted (wavy line in Fig. 1c).

The deduced structure of the S component of N38 is shown in Fig. 5 (2). The variation in the N38 genome can be assumed to have been generated by a recombination event between a point in the 0.74 kb Sall fragment [closed circle in the region from E2 to K2 in Fig. 5 (1)] and a point in the 0.72 kb Sall fragment [closed circle in the region from B43 to S in Fig. 5 (1)]. The novel recombination point in the N38 genome is indicated by a cross symbol in Fig. 5 (2). This recombination caused duplication of the region from the left Rs/Us junction to the novel recombination point in the 0.74 kb Sall fragment [thick lines in Fig. 5 (1) and (2)], and also led to deletion of the region from the novel recombination point in the 0.72 kb Sall fragment to the
Fig. 6. Nucleotide sequence analysis of the novel recombination point in the N38 genome. (a) Locations of the regions involved in the recombination event. (1) S component of standard HSV-1. Restriction enzyme cleavage sites are abbreviated as in Fig. 1(b). Two closed circles (●) represent the sites of the recombination. (2) S component of N38. The novel recombination point is represented by a cross (×). The roman numerals accompanied by an arrow in lines (1) and (2) indicate the regions whose nucleotide sequences are presented in (b). (3) Elongated Rs of N38 represented by open arrows. (b) Nucleotide sequences of the regions indicated by the roman numerals in (a). The arrows indicate the orientation of the sequences. (I) A portion of the 0.74 kb SmaI fragment of standard HSV-1. (II) Sequence of the region containing the recombination point in the N38 genome. (III) A portion of the 0.72 kb SmaI fragment of standard HSV-1. The numbering system is that of McGeoch et al. (1985) and the left Rs/U5 junction is the reference point.
Conversion of Us to Rs in HSV-1 genome

Fig. 7. Autoradiographs showing nucleotide sequences of the novel recombination point in the N38 genome. The DNA was sequenced using the dideoxynucleotide chain termination method primed by (a) the M13 single-strand primer (17 bases) (P-L Biochemicals, cat. no. 27-1534), and (b) the M13 reverse sequence primer (17 bases) (P-L Biochemicals, cat. no. 27-1532), and the products were separated in a thin 6\% polyacrylamide-urea gel. The nucleotide numbers were derived by comparison with the sequence published by McGeoch et al. (1985) with the reference point at the left Rs/Us junction. The five nucleotides at the novel recombination point are shown in the boxed area.

25.3 kb (J + L), 20.8 kb (F, from the right terminus of the L component), 17.3 kb (H, from the left terminus of the S component), 15.5 kb (J, from the left terminus of the L component) and 9.8 kb (L, from the right terminus of the S component) (Locker & Frenkel, 1979). Of the eight BglII fragments that should hybridize with the BamHI-(T + R) probe spanning the L–S junction, five were detected as separated bands in BglII digests of both N38 and SP23, and three fragments (38.1, 32.8 and 30.6 kb) could not be differentiated (Fig. 2e). The BglII-L fragment of N38 was expected to be shorter than that of SP23 by 0.4 kb, because of the differences in the S component. As shown in Fig. 2(e), the two fragments BglII-L (9.4 kb) and BglII-(J + L) (24.9 kb) from N38 were indeed shorter than the corresponding fragments from SP23 by 0.4 kb. In summary, the results in Fig. 2(d) and (e) show the occurrence of L–S inversion in the N38 genome, as in SP23.

**Nucleotide sequence of the novel recombination point in the N38 genome**

A PvuI fragment of about 0.2 kb [P1 to P2 in Fig. 6a (2)], containing the novel recombination point, was isolated from λNTE-1, cloned in pUC18, and sequenced by the dideoxynucleotide chain termination procedure (Fig. 7). The nucleotide sequence of N38 around the recombination point was compared with the corresponding sequences of the standard HSV-1 genome (Fig. 6). The numbering system was that of McGeoch et al. (1985) with the reference point at the left Rs/Us junction. The results showed that the novel recombination had occurred between a pair of common sequences of five nucleotides (5′-CCCCG-3′) arranged in opposing
One-step growth curves

To investigate the effect of the changes in the N38 genome on the growth of the virus, one-step growth curves of N38 and SP23 were constructed (Fig. 8). As the growth curve of N38 was comparable to that of SP23, the genome differences between the strains probably had no deleterious effect on the growth of N38 in Vero cells.

DISCUSSION

I have isolated a novel HSV-1 variant named N38. In the S component of this variant, a fraction of the unique sequence (map units 0.865 to 0.880) has been converted to part of the diploid inverted repeat sequences while another fraction of the unique sequence (map units 0.937 to 0.955) has been deleted. In the S component of N38, the Rs/Us junctions are moved towards the centre of the S component (from map unit 0.865 to 0.880, and from 0.955 to 0.937), and the inverted repeats are elongated (Fig. 6 and 9).

The deleted region of N38 (map units 0.937 to 0.955) contains at least four genes: US9, US10, US11 and US12 (McGeoch et al., 1985; Rixon & McGeoch, 1985). This indicates that the products of these genes are unnecessary for an HSV-1 productive infection in tissue culture cells, at least in Vero cells (Fig. 8). Gene US9 is the distal member of the pair US8 (encoding glycoprotein E) and US9, which have 3' co-terminal mRNAs. The function of the gene product encoded by US9 is unknown. The three genes US10, US11 and US12 specify a 3' co-terminal family of mRNAs. The gene US12 specifies immediate-early mRNA-5; the promoter region,
most of the 5' non-coding region and an intron lie within the flanking Rs (Murchie & McGeoch, 1982; Watson et al., 1981; Watson & Vande Woude, 1982). The gene US11 encodes a 21K polypeptide which is thought to be a DNA-binding protein and is predicted to contain 24 copies of the amino acid sequence X-Pro--Arg (Rixon & McGeoch, 1984). A 21K (and also a 22K) protein, which is induced by HSV-1 infection, has been shown to interact specifically with the 'a' sequence of HSV-1 (Dalziel & Marsden, 1984). Whether the 21K (and 22K) protein corresponds to the 21K polypeptide encoded by the gene US11 is unknown. If the 21K (and 22K) protein is essential for inversion of the HSV-1 genome, the protein(s) detected by Dalziel & Marsden should be different from the 21K polypeptide encoded by gene US11, because inversion was observed in the N38 genome (Fig. 2d, e).

The novel genome structure in the S component of N38 has been generated by a recombination event between two 5 bp stretches of homology (5'-CCCGG-3') arranged in inverted directions in Us (Fig. 6, 7 and 9). Studies on the illegitimate recombination of viruses or between virus and chromosome have shown an involvement of short stretches of homology at the regions of strand exchange (Bullock et al., 1984; Ruley & Fried, 1983; Stringer, 1982; Woodworth-Gutai, 1981). A homology of three nucleotides was observed at the recombination point in the genome of an HSV-1 variant with tandem duplications (Umene & Enquist, 1985). The pair of 5 bp stretches of homology in the S component probably played an important role in the recombination event found in the N38 genome. The 'a' sequence, which can function as a specific inversion site, is present as a direct repeat at each terminus and also in inverted
orientation at the L–S junction. The ‘a’ sequence is GC-rich, and many copies of the sequence 5'-CCCG-3' exist in it (Davison & Wilkie, 1981; Mocarski & Roizman, 1981). A pair of homologous 5 bp stretches of 5'-CCCG-3' existing in inverted directions may be a potential site of action for the activity responsible for L–S inversion (Davison & Wilkie, 1981; Mocarski et al., 1980; Mocarski & Roizman, 1981, 1982a, b; Pogue-Geile et al., 1985; Smiley et al., 1981).

Intermolecular or intramolecular recombination in Us, which may ensue from an aberrant inversion, results in the elongation of Rs and the shortening of Us observed in the genome of N38 (Fig. 9). The inversion activity may have had an influence in the formation of the present HSV-1 genome structure, and I propose a pre-S component, in light of the peculiar relationship between the genes for immediate-early mRNAs 4 (IE-4) and 5 (IE-5) (Murchie & McGeoch, 1982; Watson et al., 1981; Watson & Vande Woude, 1982). The 5' non-coding regions of IE-4 and IE-5 share a common sequence in Rs, and the coding regions of the two genes differ in Us.

The common sequence in Rs may have been derived from the 5' non-coding region of either the IE-4 or the IE-5 gene of the hypothetical pre-S component, after a recombination event at the point corresponding to the Rs/Us junction of standard HSV-1, as noted in the formation of the elongated Rs in the N38 genome (Fig. 9). Whitton & Clements (1984) found that the first translational initiation codon of IE-5 of HSV type 2 (HSV-2) has the A of ATG as the first base in Us and this defines the Rs/Us junction. By comparing it with the Rs/Us junction of HSV-1, they speculated that the Rs of HSV expands, and that the ATG acts as a buffer to prevent this expansion. They proposed that the expansion of Rs might come about by non-homologous recombination. The model accords with that in Fig. 9. They also predicted the possibility of a recombinant in which only one of the genes of IE-4 or IE-5 was retained, if the crossover occurred beyond the 3' termini of the genes. The variant N38 is this predicted recombinant; only the gene of IE-4 is retained and that of IE-5 is lost. The existence of N38 supports the possibility that there is a mechanism whereby the Rs expands (Whitton & Clements, 1984).

It may be possible to isolate other types of HSV-1 variants with rearranged genome structures, and for example several deletion mutants of HSV-1 and HSV-2 have been reported (Harland & Brown, 1985; Schröder et al., 1985). Viable HSV-2 variants with deletions in IRE, TRL or both have been isolated (HG52XD series) during screening for HSV-2 genomes lacking XbaI cleavage sites (Harland & Brown, 1985; Brown et al., 1984). These variants, like N38, were isolated unexpectedly in the course of analysing genome structures of viruses with restriction endonucleases. They were not selected by a distinguishable biological phenotype. The variants with deletions, i.e. the HG52XD series and N38, grow as well as, or better than parental strains (Harland & Brown, 1985) (Fig. 8). In contrast, an HSV-1 mutant with a deletion in the ICP4 gene was selected by its phenotype of small plaque size, and the replication capacity of the mutant was found to be reduced by a factor of 1000 (Schröder et al., 1985). Restriction enzyme analysis of virus clones obtained after transfection with viral DNAs digested with restriction endonucleases, or infection at a high m.o.i., may be one method of detecting viruses with a variant genome structure, as an alternative to the selection of a virus clone with a distinguishable phenotype.

Gratitude is extended to M. Ohara for comments on the manuscript. This work was supported by grants from the Ministry of Education, Science and Culture of Japan and from the Fukuoka Cancer Society.

**REFERENCES**


(Received 17 October 1985)