Structure of the Genome Termini of Varicella-Zoster Virus

By A. J. DAVISON

MRC Virology Unit, Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, U.K.

(Accepted 1 August 1984)

SUMMARY

The DNA molecule of varicella-zoster virus (VZV) is represented structurally as L–S, where L is a unique sequence and S is a unique sequence flanked by an inverted repeat. S may be present in either orientation in virion DNA molecules, but, to date, L has been found in only one orientation. DNA sequences were determined at the L–S joint and genome termini, which were cloned using methods designed to conserve either the 3' or 5' terminal nucleotide. Molecular hybridization experiments and analysis of the sequences showed that: (i) the genome is not terminally redundant; (ii) the unique sequence in L is flanked by an inverted repeat of 88.5 base pairs; (iii) a single unpaired nucleotide is located at each 3' terminus of the genome, such that fusion of the termini would produce a sequence identical to that at the L–S joint; (iv) approx. 5% of virions contain genomes with L inverted. The implications of these results in possible mechanisms for VZV DNA replication are discussed.

INTRODUCTION

Chickenpox and shingles are caused by a herpesvirus, varicella-zoster virus (VZV). The structure of the large linear double-stranded DNA molecule of this virus, shown in Fig. 1, has been elucidated by electron microscopy, restriction endonuclease analysis, molecular cloning and DNA sequencing (Dumas et al., 1981; Ecker & Hyman, 1982; Straus et al., 1982; Gilden et al., 1982; Davison & Scott, 1983; Davison, 1983). The genome consists of two covalently linked segments, L and S. L comprises a unique sequence [UL, approx. 100000 base pairs (bp)], and S a unique sequence (Us, 5232 bp) flanked by an inverted repeat (IRs/TRs, 7320 bp). S is present in either orientation in virion DNA molecules, giving rise to two equimolar genome arrangements defined arbitrarily as P (prototype) and Is (S inverted).

![Fig. 1. VZV genome structure and restriction endonuclease maps for SalI and KpnI (Davison & Scott, 1983). The orientations of the two genome segments (L and S) in the P (prototype) and Is (S inverted) genome arrangements are indicated.](image-url)
In this paper, DNA sequences at the genome termini and L–S joint are described, and a model for VZV DNA replication is proposed.

METHODS

**Purification of VZV DNA.** The VZV DNA strain used by Dumas et al. (1981) was grown in a line of human foetal lung cells established by Dr B. Carritt at the Institute of Genetics, Glasgow. Infected cells were trypsinized when half showed cytopathic effect and added at a ratio of approx. 1:10 to fresh cell monolayers. DNA was prepared by phenol extraction of virions isolated from infected cells and purified on sucrose gradients (Dumas et al., 1980). DNA was isolated from virus at passages 13 to 20.

**DNA–DNA hybridization.** VZV DNA fragments were transferred from agarose gels to nitrocellulose sheets using the method of Southern (1975). Nick-translated probes were prepared essentially according to the method of Rigby et al. (1977) and hybridized to nitrocellulose strips at 45 °C in 3 x SSC (0-45 M-NaCl, 0-045 M-trisodium citrate pH 7.5), 5 x Denhardt's solution (Denhardt, 1966), 0-1 mg/ml single-stranded salmon sperm DNA and 50% formamide. The strips were washed extensively in 2 x SSC at 60 °C, dried, applied to a glass plate and autoradiographed.

**Recombinant plasmids.** Davison & Scott (1983) have described plasmids constructed by annealing VZV KpnI fragments, which had been extended at 3' termini with deoxycytidine residues using terminal deoxynucleotidyl transferase, to PstI-cleaved plasmid vector pAT153 which had been extended with deoxyguanosine residues. Additional plasmids were obtained in three ways.

1. As described above, except that plasmid vector pUC9 and host Escherichia coli JM83 were used (Messing & Vieira, 1982).
2. As described in (1), except that VZV KpnI fragments were extended with deoxyguanosine residues and linearized pUC9 was extended with deoxycytidine residues.
3. VZV KpnI fragments were treated with T4 DNA polymerase in the presence of the four deoxynucleoside triphosphates to produce blunt ends and ligated into the SmaI site of pUC9.

Clones bearing plasmids containing the genome termini and L–S joint were identified by hybridizing nick-translated KpnI m and t and by restriction endonuclease analysis. Rare clones from (2) containing the novel L terminus and L–S joint were identified by hybridizing nick-translated KpnI m, l and t and by restriction endonuclease analysis. There was no possibility of plasmids constructed in (1) and (2) arising from ligation of non-contiguous fragments because these methods did not involve a ligation step in vitro. Manipulations were carried out under conditions of good microbiological practice, and bacterial stocks and plasmid DNA were prepared as described previously (Davison & Wilkie, 1981).

**DNA sequencing.** Plasmids containing SsrI/S terminal fragment of 6800 bp, KpnI l (L–S joint fragment of 2841 bp) or KpnI t (L terminal fragment of 893 bp) described by Davison & Scott (1983) were transferred from the original host (E. coli HB101) to a modification-plus host (E. coli DH1). Restriction endonuclease fragments or random fragments (400 to 1000 bp) generated by sonication were ligated into M13 mp8 (Messing & Vieira, 1982). Recombinant phage DNA was prepared under conditions of good microbiological practice from phage released from infected E. coli JM101, and templates for sequencing were identified by hybridization of the appropriate nick-translated restriction fragment. DNA was sequenced using the dideoxynucleotide technology (Sanger et al., 1977, 1980), and products were separated in thin 6% polyacrylamide–urea gels containing a buffer concentration gradient (A. T. Bankier & B. G. Barrell, personal communication). Each gel was bonded to one glass plate prior to electrophoresis and then dried prior to autoradiography (Garoff & Ansorge, 1981). DNA sequence data were collated using the programs of Staden (1982) in a DEC PDP-11 computer operating under the RSX11M system. The entire sequence of each of the three fragments was obtained. Specific restriction fragments were excised from plasmids and ligated into cleaved M13 mp8 or mp9 in order to determine sequences at the genome termini.

RESULTS AND DISCUSSION

**Structure of the genome termini and L–S joint**

The structure of the genome termini could not be determined easily from intact virion DNA, so two sets of plasmids were constructed to preserve either the 3' or 5' terminal nucleotide. The first set comprised terminal fragments linked to the vector via short homopolymer tracts added at the 3' termini. Separate clones of terminal fragments extended by either deoxycytidine or deoxyguanosine residues were used for unambiguous identification of the 3' nucleotide. The second set of plasmids was constructed by ligation of the vector to terminal fragments modified by the action of T4 DNA polymerase in the presence of the four deoxyribonucleoside triphosphates. Under these conditions 3' termini are either extended or degraded until they are flush with 5' termini, thus conserving the 5' terminal nucleotide. Terminal fragments from the
two sets of plasmids were transferred into an M13 vector for sequencing. The possibility that single plasmid clones contained unrepresentative termini was excluded by sequencing more than one of each type. The results for representative clones (Fig. 2) show that plasmids constructed by homopolymer extension of 3' termini (panels $2, 3, 2, 3$) have an additional base pair at the genome termini when compared with plasmids constructed by treatment with T4 DNA polymerase (panels $1, 1$). This is most simply explained by the presence of a single unpaired G residue at the 3' terminus of S and a complementary unpaired C residue at the 3' terminus of L. The observation that terminal and internal fragments of equivalent size were cloned in approximately equal numbers using both methods indicates that most virion DNA molecules do not have a covalently bound blocking group at either terminus.

Previous examination of restriction endonuclease sites close to the genome termini excluded the presence of a terminal redundancy (a sequence at one terminus directly repeated at the other) greater in size than 20 bp (Davison & Scott, 1983). More recently, Ecker et al. (1984) were unable to detect a terminal redundancy greater in size than 100 bp using electron microscopy and molecular hybridization. Unfortunately, the conclusion from the latter experiments must be discounted since neither of their probes contained a terminus. Although Fig. 3 shows that the sequence AGG on the upper strand (CTC on the lower strand) is present at both termini, the VZV genome is not terminally redundant in any acceptable sense. However, 88.5 bp at the L terminus, the fraction resulting from the unpaired terminal nucleotide, are present in inverse orientation at the L–S joint, so that L comprises a unique sequence ($U_L$) flanked by a short inverted repeat ($IR_L/IR_L$).

**A minor DNA population with L inverted**

Results of hybridizing VZV plasmids to SalI and KpnI fragments of virion DNA are shown in Fig. 4, and locations of SalI and KpnI fragments are shown in Fig. 1. KpnI $r, i$, $l$ and $m$ hybridized strongly to SalI fragments mapping in equivalent locations. KpnI $t$ hybridized weakly to SalI $f$
Fig. 3. Sequences of the VZV termini and L-S joint. The sequences shown are those of *KpnI* (L terminus) and appropriate parts of *KpnI* (L-S joint) and *SstI* (S terminus). The precise structure of each genome terminus was derived from Fig. 2. The *SmaI* and *XhoI* sites relevant to Fig. 2 and 4 are indicated.
VZV genome termini

Fig. 4. Autoradiograph showing results of hybridizing VZV DNA (V) and plasmids containing VZV KpnI t, i, l, m, LS or L to SalI and KpnI fragments of virion DNA. Normal SalI and KpnI fragments are shown to the left of each panel and the minor fragments L and LS are indicated on the right.

and KpnI l to SalI o because the L terminus and L–S joint share an 88.5 bp sequence, but hybridization between the termini was not detected because the genome is not terminally redundant. KpnI t, i, l, and m also hybridized to the minor fragments SalI LS and KpnI i and l to the minor fragment SalI L. The sizes and hybridization characteristics of the two minor fragments are consistent with those of a novel L terminus (SalI L) and L–S joint (SalI LS) from DNA molecules with L inverted. The plasmids hybridized to the expected KpnI fragments (Fig. 4), and KpnI t and l showed weak hybridization to each other owing to the shared 88.5 bp sequence. KpnI t, l, and m also hybridized to the novel L–S joint (KpnI LS). The expected fragment from the novel L terminus (KpnI L) would have migrated off the gel, and therefore KpnI l did not identify a second minor fragment. The results of hybridizing VZV plasmids to PstI, BglII, XbaI, EcoRI, SstI, XhoI, and PvuII fragments of VZV DNA (data not shown) fully supported the conclusion that L is inverted in some virion DNA molecules, and excluded the possibility that the minor fragments resulted from partial restriction endonuclease digestion.

The identities of KpnI L and LS were confirmed by cloning them and hybridizing to SalI and KpnI fragments of virion DNA (Fig. 4). Finally, appropriate fragments from the clones were transferred to M13 mp8 or mp9 and sequenced (Fig. 5). Identical sequences were observed at the novel and normal L termini (compare Fig. 5 panel 1 with Fig. 2 panel L3), the novel and normal L–S joints (compare Fig. 5 panel 2 with 3), the region of UL adjacent to the novel L terminus and normal L–S joint (compare Fig. 5 panel 4 with 6), and the region of UL adjacent to the novel L–S joint and normal L terminus (compare Fig. 5 panel 5 with the sequence of the L terminus shown in Fig. 3). The results show that the novel L terminus and L–S joint have precisely the structure expected of these regions in virion DNA molecules with L inverted about the L–S joint. The proportion of these genomes in the DNA population was estimated as approx. 5% by considering the intensities of the approximately equimolar minor bands shown in Fig. 4 and the numbers of clones obtained.

Ecker et al. (1984) detected two minor fragments in HindIII digests of the DNA of different VZV strains. They suggested that one represents a heterogeneity at the S terminus and the other a fusion of the genome termini, perhaps originating from 'aberrantly processed' unit length linear
Fig. 5. Autoradiograph showing sequences at the VZV novel L terminus and L–S joint. Each sequence proceeds from the XhoI site shown in Fig. 3 and shows: 1, novel L terminus (extended with deoxyguanosine residues) across the terminus into vector sequences; 2, normal L–S joint towards U₅; 3, novel L–S joint towards U₅; 4, novel L terminus towards U₅; 5, novel L–S joint towards U₅; 6, normal L–S joint towards U₅. Arrows indicate the terminal nucleotide in 1, the L–S junction in 2 and 3, and the junction between IR₅/TRL and Ul in 4, 5 and 6.

genomes or from circular DNA. However, heterogeneity at the S terminus was not detected in hybridization experiments (Fig. 4). Moreover, since the two fragments correspond in size to expected HindIII fragments containing the novel L terminus and L–S joint, both would result from inversion of L. This explanation makes it unnecessary to propose that some virions contain circular genomes. Straus et al. (1982) detected hybridization of EcoRI E, which contains the L–S joint, to EcoRI N, which maps in U₅. However, it is probable that the probe actually hybridized to the novel L terminus, which in this digest would co-migrate with EcoRI N. No hybridization of sequences close to the L–S joint to any region in U₅ was detected in Fig. 4.

Two observations counter the possibility that the novel L terminus and L–S joint originated from defective DNA molecules, which arise from repeated in vitro passage of virus at high multi-
P  
VZV genome termini  

Fig. 6. Model for VZV DNA replication.

plicity and, in other herpesviruses, comprise tandem repeats of a small part of the complete genome. Firstly, these species were detected in all DNA preparations examined, even that isolated from virus at the lowest available passage number (passage 13). Ecker et al. (1984) detected them in several virus isolates and in virus obtained directly from vesicle fluid (passage 0). Secondly, the sizes of even the largest fragments containing the novel L terminus (an SstI fragment of 9000 bp) and L–S joint (a PstI fragment of 24 000 bp) are consistent with the generation of these fragments from complete genomes with L inverted.

Model for DNA replication

VZV is genetically related to other alphaherpesvirinae such as the two serotypes of herpes simplex virus (HSV-1 and HSV-2) and pseudorabies virus (PRV) (Davison & Wilkie, 1983b; A. J. Davison, unpublished results). The PRV genome is similar in structure to that of VZV (Powell, 1979; Ben-Porat et al., 1979). However, an inverted repeat flanking U_L and inversion of L have not yet been detected. The HSV genome differs from that of VZV and PRV in possessing a terminal redundancy of approx. 400 bp which is also present in inverse orientation at the L–S joint (Sheldrick & Berthelot, 1974; Davison & Wilkie, 1981). Moreover, IR_L/TR_L is much larger
in HSV than VZV (approx. 9000 bp), and both segments invert, giving rise to four equimolar genome arrangements defined as P (prototype), Iₘ (S inverted), Iₙ (L inverted) and Iₛₘ (S and L inverted). Parenthetically, the arbitrary way in which the VZV and HSV genome arrangements were defined results in the following correspondence between the two genomes on the basis of gene order: P = Iₛₘ, Iₘ = Iₙ, Iₙ = Iₛ, Iₛₘ = P (Davison & Wilkie, 1983b; A. J. Davison & D. J. McGeoch, unpublished results).

Fig. 6 shows a model for VZV DNA replication based upon the structural features described above and experimentally supported views of PRV and HSV-1 DNA replication.

In step A, an input linear VZV DNA molecule in one of the two major (P and Iₛ) or two minor arrangements (Iₙ and Iₛₘ) circularizes. Ben-Porat & Veach (1980) showed that the PRV genome circularizes prior to DNA replication. Since the VZV genome is not terminally redundant, the simplest mechanism for circularization involves direct ligation of the termini to produce a novel L–S joint identical in sequence to the normal L–S joint. Rapid circularization could be facilitated by close proximity of the termini in the virion. Mocarski & Roizman (1982a) claimed that the HSV-1 genome, like that of VZV, has a single unpaired nucleotide at each 3' terminus, and Davison & Wilkie (1983a) concluded from a study of recombinants between HSV-1 and HSV-2 that the HSV genome may also circularize by direct ligation of termini.

In step B, limited replication of circular molecules occurs, and segment inversion takes place by intramolecular recombination between inverted repeats. Inversion in HSV-1 is mediated by the action of virus-specified factors upon the terminally redundant sequences common to the termini and L–S joint (Mocarski & Roizman, 1982b). There is no evidence yet that segment inversion in VZV is due to site-specific rather than to general recombination. However, a sequence of about 30 bp commencing approx. 30 bp from the S terminus, comprising A and T residues flanked by G + C-rich sequences, is conserved in HSV-1, HSV-2, PRV and VZV, and could be a signal involved in site-specific segment inversion or in genome cleavage and encapsidation (Davison & Rixon, 1984).

In step C, DNA replication occurs by the generation of head-to-tail concatemers, perhaps by a rolling circle mechanism, as has been shown for PRV and HSV-1 (Ben-Porat & Rixon, 1979; Jacob et al., 1979).

In step D, concatemers are cleaved specifically to generate unit length genomes for encapsidation. Cleavage usually occurs at the novel L–S joint but occasionally at the normal L–S joint, generating the major and minor arrangements. The proportion of minor arrangements is determined by the relative preference of the cleavage system for the novel L–S joint. If this is correct, part of the signal recognized by the cleavage system in producing the major arrangements must be located in the portion of Uₗ adjacent to the novel joint, at least 90 bp from the cleavage site. A second part of the signal may reside in IRₗ/IRₗ or IRₛ/TRₛ, and occasional recognition of this at the normal L–S joint would result in the minor arrangements.

Two less likely hypotheses may be envisaged which would account for the observed proportions of genome arrangements. In the first, cleavage occurs only at the novel joint, generating the major arrangements, and then the minor arrangements are produced by intra- or intermolecular recombination between IRₗ/TRₗ in unit length genomes. However, mature genomes may be unable to recombine thus since cleavage and encapsidation seem to be intimately associated in PRV (Ladin et al., 1980). In the second hypothesis, the VZV strain used comprises a mixed population in which the major arrangements generate concatemers which are cleaved at the novel L–S joint and the minor arrangements give rise to identical concatemers which are cleaved at the normal L–S joint. However, this hypothesis does not readily explain the exclusive operation of the cleavage system on the novel L–S joint in some concatemers and on the normal L–S joint in others. Although all clinical isolates analysed to date contain predominantly P and Iₛ, an insufficient number has been examined to determine whether isolates containing predominantly Iₙ and Iₛₘ exist in nature. The hypothesis could be tested in vitro by attempting to isolate virus containing predominantly Iₙ and Iₛₘ by plaque purification.

The author is grateful to Jose Cebrian and Peter Sheldrick for discussion of their unpublished results, which confirm the sequences at the termini, and to John Subak-Sharpe and Ron Hay for critical reading of the manuscript.
REFERENCES


(Received 5 June 1984)