JOURNAL OF VIROLOGY, Oct. 1990, p. 4968–4977
0022-538X/90/104968-10$02.00/0

Vol. 64, No. 10

Structural Organization of the Termini of the L and S Components of the Genome of Pseudorabies Virus

JEAN M. DEMARCHI, ZHIQIANG LU, GLENN RALL, SABINA KUPERSHMIDT, AND TAMAR BEN-PORAT*

Department of Microbiology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Received 17 May 1990/Accepted 12 July 1990

The sequences of several hundred nucleotides around the junctions between the L and S components in concatemeric DNA and in mature virion DNA were ascertained. The two ends of the mature genome (which are joined in concatemeric DNA) show no sequence homology. Several directly repeated elements are present near both ends of the genome. Furthermore, the last 82 nucleotides at the left end of the L component (and of the genome) are repeated in inverted form (inverted repeat within the L component [IRL]) approximately 350 to 600 nucleotides downstream (depending on the virus isolate) bracketing the U_{L2} component. A comparison between the sequences at the right and left ends of the L component of the genome showed patchy homology, probably representing a vestigial inverted repeat bracketing the L component (IRL). Furthermore, less than 5% of the genomes have an L component that is in the orientation opposite to that of most of the viral genomes, indicating that the vestigial IRL that brackets the U_L sequence may be sufficient to mediate inversion of the L component in some of the genomes. On the other hand, the U_{L2} component, which is bracketed by a perfect IRL, does not invert to a greater extent than does the L component (if it inverts at all). Analysis of the nucleotide sequence at the concatemeric junction of three different pseudorabies virus isolates showed almost complete sequence conservation. The sequence and organization of the repeated elements in the different isolates were almost identical, despite their different histories and origins. The high degree of conservation of these repeated elements implies that they may fulfill an essential function in the life cycle of the virus.

The genome of pseudorabies (PRV) is a linear, double-stranded class 2 (4, 12, 25) herpesvirus DNA molecule approximately 90×10^6 daltons in size. It is composed of two components, L and S. The S component consists of a short unique (U_S) sequence bracketed by large (approximately 10^7 daltons) inverted repeats (IR). High-frequency inversion of the U_S component, but not of the L component, is observed during virus growth. The S genome is found in two isomeric forms, and isomerization is complete by the time a small plaque has formed from a single virion (1). The ends of the genome of PRV are unique; no terminal redundancy is present.

Varicella-zoster virus (VZV), like PRV, has a class 2 genome. The genome is not terminally redundant, but both its L and S components are bracketed by inverted repeats. The repeat bracketing the L component is only 88 nucleotides in size. The S component inverts and is found in two equimolar orientations relative to the L component, but only a small percentage of the viral genomes have an L component that is inverted relative to that found in most of the genomes (7, 11). The genome of bovine herpesvirus (BHV) also has a structure that is very similar to that of PRV; the U_S sequence is bracketed by large IR and is found in two orientations relative to the unique long (U_L) sequence. The U_L sequence is bracketed by a small inverted repeat only (25 base pairs [bp]), and the L component does not appear to invert itself relative to the S component. The genome of BHV also has a small terminal redundancy (9).

The structures of class 3 herpesvirus genomes, such as herpes simplex virus (HSV) and human cytomegalovirus, and class 2 genomes, such as VZV, BHV, and PRV, are related. The genome of HSV is also composed of L and S components. Both of these components are flanked by inverted repeats which include an *a* sequence. The *a* sequence is present as a direct terminal redundancy and in inverted form at the junction between the L and S components. In HSV, both the L and S components invert and the genome is found in four isomeric forms (12, 21).

A comparison of the sequences near the termini of the S and L components of different herpesviruses is of interest, because sequences near the ends of their genomes appear to fulfill several functions. In PRV, signals that play a role in cleavage-encapsidation of genomic DNA are present at both ends of the mature linear genome (28). In class 3 genomes, such as HSV, the *a* sequence (which is present as a direct repeat at both ends of the genome) appears to include *cis* functions that have been separated by the cleavage site in class 2 genomes. *cis* functions near the ends of the S component of PRV may, in principle, also be instrumental in promoting inversion of this component, as do *cis* functions that have been reported to be present in the *a* sequence of HSV (6, 17, 18, 23). Furthermore, sequences that can function as an origin of replication are present near the terminus of the L component of the genome of PRV (28).

We have previously compared the nucleotide sequence of 140 bp at each end of the mature genome of PRV with that of the concatemeric junction of these ends (10). We have now further investigated the structure of the regions adjacent to the termini, as well as that adjacent to the junction of the L and S components in the mature PRV genome. Several thousand nucleotides around each of these regions have been sequenced. This project was undertaken to gain insight into the *cis* functions that might be involved in cleavage-encapsidation of concatemeric DNA, as well as those that might be instrumental in mediating the inversion that results in isomerization of the genome.

* Corresponding author.

MATERIALS AND METHODS

Virus strains and cell culture. PRV Ka is a strain that has been carried in our laboratory for more than 25 years. PRV Norden is an attenuated vaccine strain; its origins have been described previously (19). PRV USA7 is a primary isolate obtained from an infected pig in the United States. The viruses were grown in PK or RK cells which were cultivated in Eagle synthetic medium containing 5% bovine serum.

Enzymes and chemicals. Restriction enzymes, DNA polymerase I, and T4 DNA ligase were purchased from Bethesda Research Laboratories, Inc. T4 polynucleotide kinase was purchased from U.S. Biochemical Corp. [α-^{32}P]ATP and [α-^{32}P]dCTP were purchased from Dupont, NEN Research Products.

Purification of virions and viral DNA. A continuous line of PK cells was infected (multiplicity of infection, 0.5 PFU per cell) and incubated for 24 h in Eagle medium at 37°C. Virions were purified as described previously (2) and were lysed by addition of Sarkosyl (final concentration, 2%; CIBA-GEIGY Corp.). The samples were incubated for 15 min at 60°C and digested with nuclease-free pronase (1 mg/ml) for 2 h. The DNA was extracted with phenol-chloroform-isoamyl alcohol (50:48:2, vol/vol) and dialyzed against TE buffer (0.01 M Tris, 0.001 M EDTA, pH 7.6).

Cloning of restriction fragments. The procedure used to clone the terminal and junction *Bam*HI restriction fragments have been described previously (10, 15). Appropriate subfragments were isolated, blunt ended with the Klenow fragment, and cloned into the *Sma*I site of pUC13. In some cases, the uncloned end fragments of the genome were isolated and sequenced.

DNA sequencing. DNA was sequenced by the method of either Sanger et al. (22) or Maxam and Gilbert (16).

Analysis of intracellular concatemeric DNA. Cells were infected with PRV Ka and incubated at 37°C until 5 h postinfection. The cells were harvested, and DNA was extracted, digested with the appropriate restriction enzyme, and analyzed by Southern blotting. In some cases, cells were infected at 41°C with a temperature-sensitive mutant of PRV, TsN, which is DNA$^+$ but defective in cleavage-encapsidation (i.e., concatemeric DNA accumulates in these infected cells) at the nonpermissive temperature (13). The cells were harvested at 10 h postinfection, and the DNA was analyzed as described above.

Southern analysis. Restriction fragments were separated on agarose gels and transferred to nitrocellulose filter paper (24). The immobilized DNA was hybridized to [α-^{32}P]dCTP nick-translated DNA probes (20).

Isolation and sequencing of uncloned end fragments of the PRV genome. Mature virion DNA was cleaved with *Bam*HI or *Kpn*I. The end fragments, *Bam*HI fragments 13 and 14 or *Kpn*I fragment D, were excised from the gel, phosphatase treated, end labeled, and digested with either *Xho*I or *Nco*I, and the 5'-end-labeled ends were sequenced by the method of Maxam and Gilbert (16).

Computer-generated sequence analysis. Computer-generated sequence analysis was done by using the DNA Inspector IIe and Microgenie programs.

Nucleotide sequence accession numbers. The sequences reported in this article are registered with GenBank under accession numbers M37112 and M37113.

RESULTS

Sequence analysis of the concatemeric junction of PRV Ka. The nucleotide sequence of the terminal 140 bp at both ends of the mature DNA of PRV Ka, as well as the nucleotide sequence at the junction of the genome termini in concatemeric DNA, has been reported previously (10). From these results, the structure of the termini and the mechanism of concatemerization of the PRV genome have been deduced. We have now sequenced several thousand nucleotides near the termini of the mature genome, as well as near the junctions between the L and S components, i.e., the region adjacent to the internal IR (Fig. 1). The position of the *Bam*HI restriction fragments spanning the junctions between the L and S components present in concatemeric (region C) and mature (region I) DNAs that were cloned into the *Bam*HI site of pBR325 and that have been sequenced are depicted in Fig. 1. Clone I4 contains *Bam*HI fragment 8'; i.e., the standard junction between the L and S components present in mature DNA (region I) and clone J31 contains the concatemeric junction, i.e., the fused genomic termini (region C). A detailed restriction enzyme map of the regions of interest in clones J31 and I4 of PRV Ka and the strategy used to sequence these regions of the genome of PRV are shown in Fig. 2. Although several thousand nucleotides have been sequenced, the sequence of only approximately the last 600 to 700 nucleotides at the genomic termini, as well as near the junctions between the L and S components in the mature genome are relevant to the problems discussed in this report and are presented in Fig. 3 and 4.

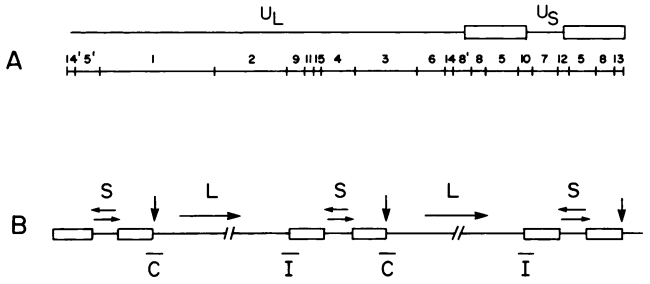

FIG. 1. Regions of the PRV genome that have been sequenced. (A) *Bam*HI restriction map of mature PRV DNA. (B) Origin of the junction fragments that have been sequenced. The horizontal arrows indicate the orientations of the L and S components. The vertical arrows indicate the junctions. C, Concatemeric junction consisting of the two genomic ends present in concatemeric DNA (clone J31); I, internal standard junction present in mature DNA (clone I4).

Analysis of the nucleotide sequence near the termini of the genome. Computer-generated dot matrix analyses of the sequences spanning the junctions of the two ends of the genome present in clone J31 are shown in Fig. 5A and B. No significant direct (Fig. 5A) or inverted (Fig. 5B) homology between the two ends of the genome was detected.

Several types of inverted or direct repeats are evident within the sequences at each terminus. Small direct repeats are prominent within the 700 nucleotides at the right (S) terminus of the genome (R1 to R6; Fig. 3 and 5A). An inverted repeat (IRL) is present at the left terminus in the L component of the genome (Fig. 3 and 5B); the terminal 82 nucleotides are repeated in inverted form 353 nucleotides upstream. The IRL brackets a unique sequence (U_{L2}) that includes two direct repeats of 34 nucleotides (R7), as well as a partial repeat of this sequence (R7a).

Analysis of the sequence at the junction between the L and S components in the mature genome. Dot matrix analyses of the nucleotide sequence at the standard junction between the L and the S components present in the mature genome (clone I4) and comparison of these sequences with those

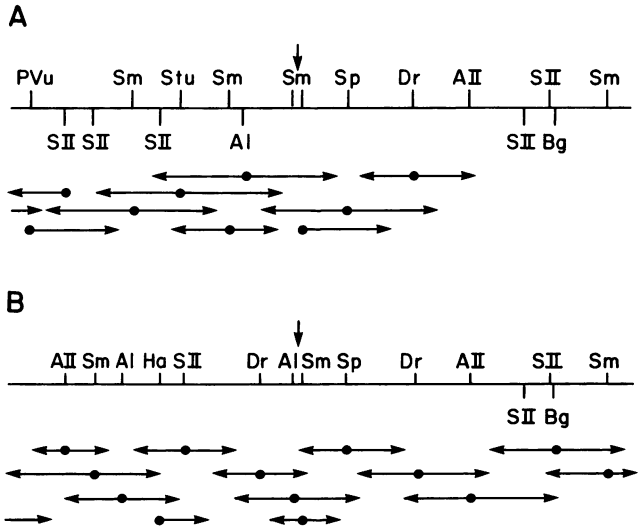

FIG. 2. Restriction maps of clones J31 (A) and I4 (B) and sequencing strategy used. The nucleotide sequence was obtained by the method of Maxam and Gilbert. AI, *Alu*I; AII, *Ava*II; Bg, *Bgl*I; Dr, *Dra*I; Ha, *Hae*II; Pvu, *Pvu*I; Sm, *Sma*I; Sp, *Sph*I; SII, *Sst*II; Stu, *Stu*I. The vertical arrows indicate the junction between the L and S components; the horizontal arrows indicate the directions of sequencing.

present in the concatemeric junction (clone J31) are shown in Fig. 6 and 7. The sequences at the ends of the IR which bracket the U_S sequence are identical in both junction fragments; the terminal IR (which is derived from the right end of the genome) is thus identical to the internal IR. The nucleotide sequences of the two junction fragments diverge thereafter, and no homology between the two ends of the L component was evident in the computer-generated analysis. However, the sequences at the two ends of the L component could be aligned in a manner that revealed significant patchy inverted homology between the two (Fig. 8). Thus, while the L component of PRV is not bracketed by a perfect inverted repeat, as are those of VZV and BHV (7, 9), a sequence homology that appears to represent a vestigial inverted repeat is evident.

Inversion of the unique sequences bracketed by the patchy IRL. The sequence analysis described above showed that 82 bp at the left terminus of the molecule are present in inverted form approximately 350 nucleotides downstream, forming a perfect inverted repeat (IRL) that brackets a unique sequence (U_{L2}). Furthermore, what appeared to be a vestigial form of the IRL sequence was also found at the right end of the L component. The experiments described in this section were performed to ascertain whether the unique sequences bracketed by either the IRL (i.e., the U_{L2} sequence) or the

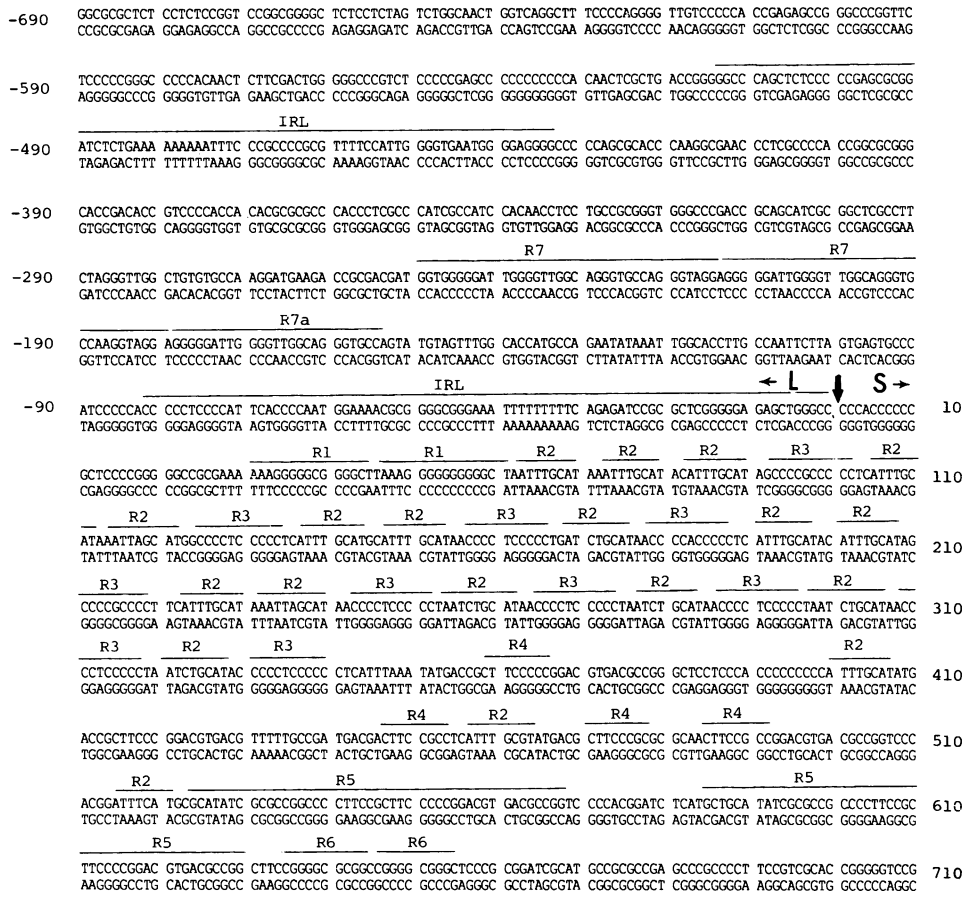

FIG. 3. Nucleotide sequence at the concatemeric junction. The vertical arrow indicates the position of the junction between the left and the right ends of the genome. S and L indicate the S and L components of the genome, respectively. The different types of repeated elements (R1 to R7) are indicated. Each set of repeated elements may vary by one nucleotide. IRL indicates the 82-bp inverted repeat. Nucleotide 1 is the first nucleotide of the S component (IR). Nucleotide −1 is the first nucleotide of the L component (left end of the genome and of the L component).

FIG. 4. Nucleotide sequence at the standard junction. The vertical arrow indicates the position of the junction between the S and L components. The nucleotide sequence of the inverted repeat is identical to that of the inverted repeat illustrated in Fig. 3. The numbering of the nucleotides of the S component is also the same as in Fig. 3. Nucleotide −1 is the first nucleotide of the L component (right end of the L component).

vestigial IRL (i.e., the U_L component) are found in two orientations in either mature or concatemeric DNA.

(i) Inversion of the U_L component of the genome. Populations of genomes obtained form mature virions of PRV include a small minority of genomes that consists of molecules in which the L component is in the orientation opposite to that found in most of the viral genomes (Fig. 9). In this experiment, DNA obtained from virions was digested with *Bam*HI, transferred to nitrocellulose filters, and hybridized to cloned *Bam*HI fragment 8′. The sequences of *Bam*HI fragment 8′ should hybridize to *Bam*HI fragment 8′, as well as to *Bam*HI fragment 13 (Fig. 1). Indeed, two major bands of the appropriate size did hybridize to the *Bam*HI fragment 8′ probe. However, the digest also contained two minor bands that hybridized to *Bam*HI fragment 8′. One of the bands (fragment 11a) had the expected hybridization pattern (i.e., it hybridized to *Bam*HI fragment 13, *Bam*HI fragment 8′, and *Bam*HI fragment 14′) and the expected size (2.8 kilobases) of the internal junction that would be generated from mature genomes in which inversion of the L component has occurred (Fig. 10). This fragment is equivalent to a fragment composed of the joined standard genome ends (clone J31). The other fragment, 10a, is 3.2 kilobases long and hybridized only to *Bam*HI fragment 8′. The hybridization pattern, as well as the size of that fragment, is that expected of the end fragment generated from genomes in which inversion of the L component has occurred (Fig. 10). Analysis of mature viral DNA with other restriction fragments (*Kpn*I and *Sal*I) also revealed the presence of fragments that showed that inversion of the L component had occurred in a minority of the genomes (data not shown). On the basis of the relative intensities of the signals, we estimate that less than 5% of the viral genomes have an inverted L component. However, the presence of even this small minority of genomes with an inverted L component indicates that the patchy inverted homology between the two ends of the L component may be sufficient to mediate inversion. Inversion of the L component of PRV had not been detected previously, because the Southern blots had not been sufficiently overexposed to detect the minority of fragments generated from molecules in which the L component was inverted.

(ii) Inversion of the unique sequences (U_{L2}) bracketed by the perfect IRL. The results described above indicate that the vestigial IRL sequence present at the right end of the L component appears to be able to mediate inversion of the L component of a minority of the viral genomes. We were therefore interested in ascertaining whether the perfect IRL bracketing the 353-bp U_{L2} sequence (which is present near the left end of the genome) would mediate high-frequency inversion of U_{L2}.

To ascertain whether U_{L2} is found mainly in one orientation in the viral genomes, we sequenced the uncloned genomic L terminus isolated from mature virions. If a large

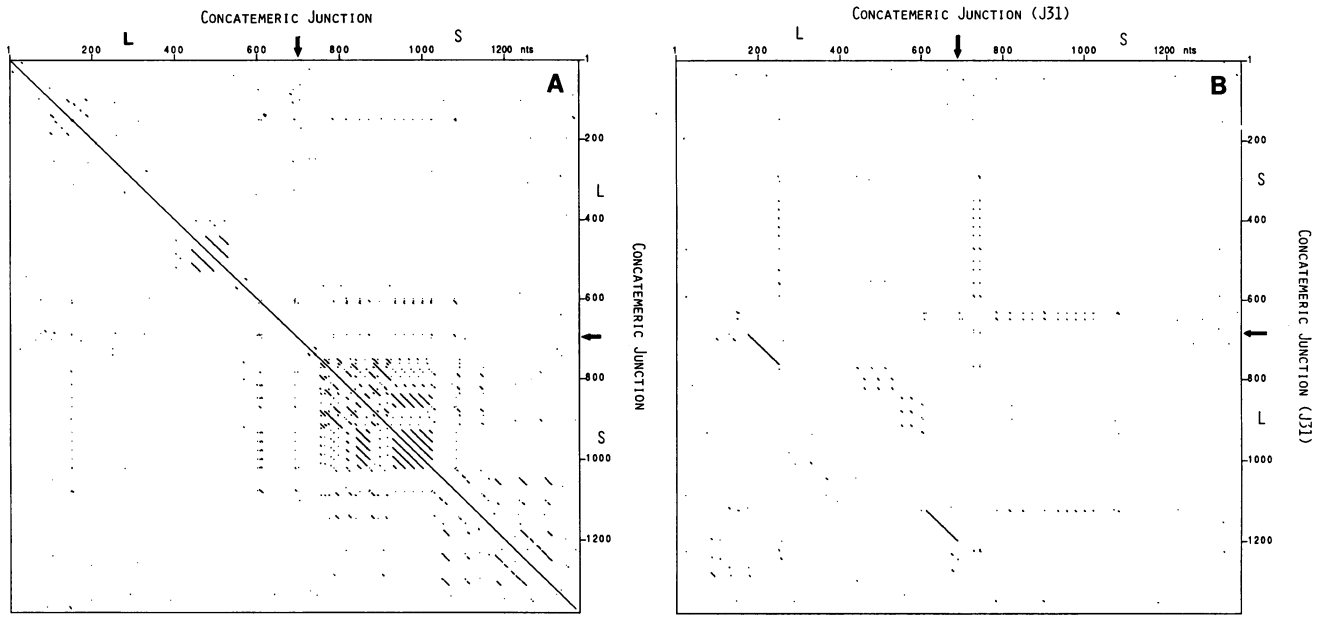

FIG. 5. Dot matrix analysis of the sequence of the concatemeric junction. The nucleotide sequence in Fig. 3 was analyzed by using a DNA Inspector IIe program. Panels: A, analysis for the presence of direct repeats; B, analysis for the presence of inverted repeats. nts, Nucleotides. The arrows indicate the junctions between L and S components.

proportion of the U_{L2} inverted, direct sequencing of the uncloned L terminus beyond 82 nucleotides (i.e., beyond the IRL) would not generate a readable sequence ladder. However, when the uncloned end fragment was sequenced, we obtained clear sequence data that were identical to those obtained by sequencing the cloned end fragment. We conclude, therefore, that only a very small proportion (if any) of U_{L2} is found in inverted form in the mature genomes.

The failure to detect inversion of the U_{L2} sequence in the genomes of mature virions could be due to the fact that U_{L2} does not invert. Alternatively, it could also reflect the lack of cleavage of concatemeric DNA or the lack of encapsidation of DNA in which U_{L2} is inverted. To ascertain whether inversion of the U_{L2} sequence would be detectable in intracellular uncleaved viral DNA, we cleaved intracellular DNA obtained from cells infected with a temperature-sensitive mutant (TsN) in which cleavage of concatemeric DNA does not occur at the nonpermissive temperature (13) with enzymes that do not cleave in the IRL and cleave asymmetrically in the U_{L2} sequence (*Dde*I or *Sst*II and *Sma*I). The

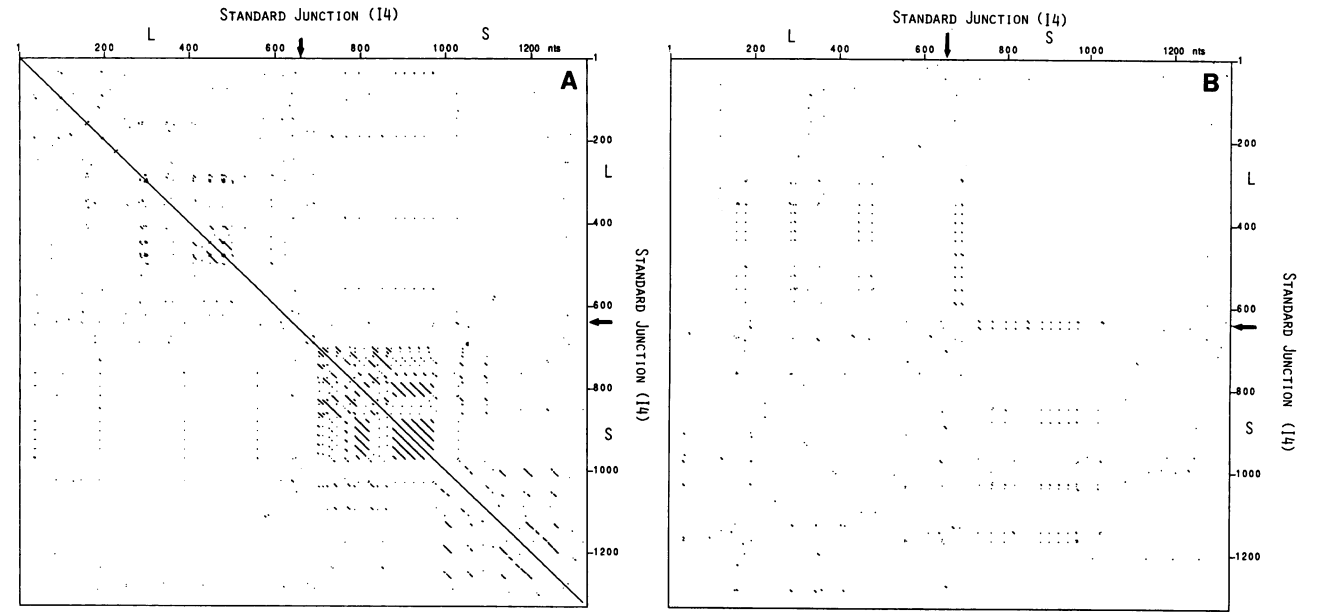

FIG. 6. Dot matrix analysis of the sequence of the internal standard junction. The nucleotide sequence in Fig. 4 was analyzed by using a DNA Inspector IIe program. Panels: A, analysis for the presence of direct repeats; B, analysis for the presence of inverted repeats. nts, Nucleotides. The arrows indicate the junctions between L and S components.

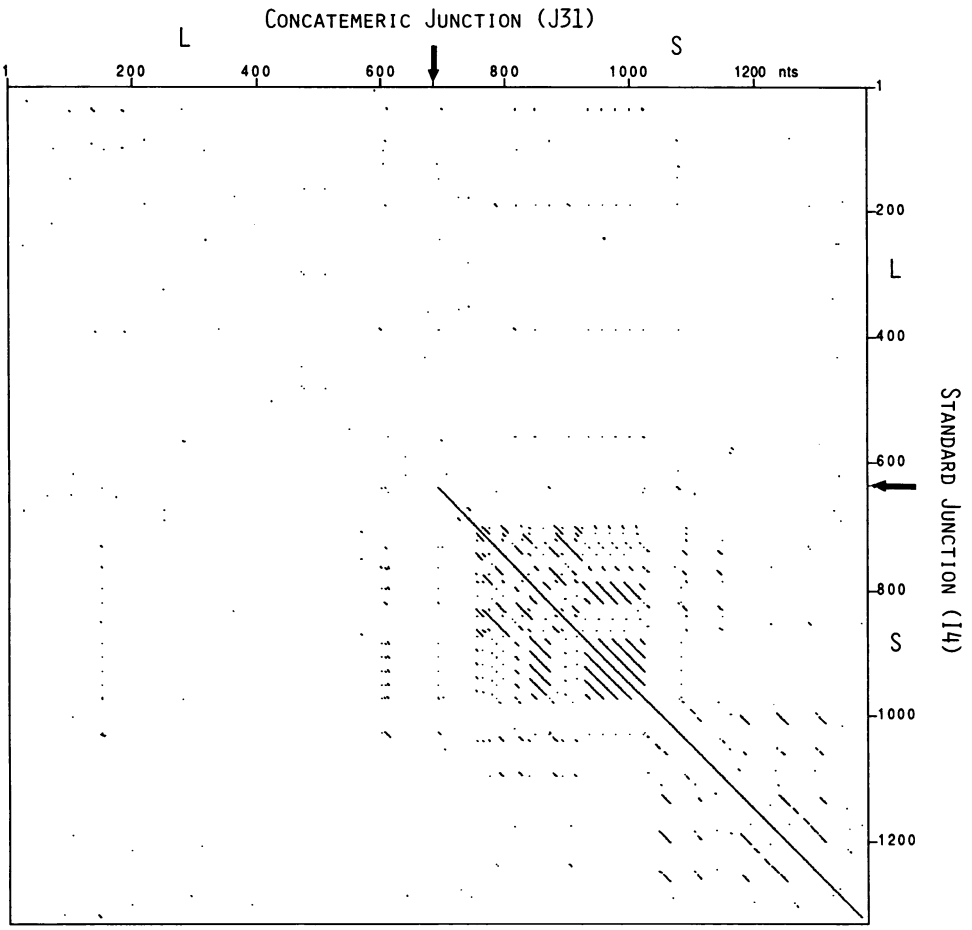

FIG. 7. Comparison of the sequences of the concatemeric and internal standard junctions. The nucleotide sequences in Fig. 3 and 4 were compared. The S components in both were identical. No homology between the sequences in their L components (the right and left end sequences of the L component) was detected. nts, Nucleotides. The arrows indicate the junctions between L and S components.

DNA was transferred to nitrocellulose paper and probed with a nick-translated DNA fragment consisting of the leftmost 600 nucleotides of the genome. By using this method, we should have been able to detect inversion of U_{L2} had it occurred in a significant proportion of the genomes. The results of these experiments (data not shown) revealed no evidence that the U_{L2} sequence inverts; only bands generated from the prototype orientation of U_{L2} were detected. On the basis of our results, we cannot, however, exclude the possibility that a small proportion (less than 5%) of the viral genomes had an inverted U_{L2} sequence; a small percentage of restriction fragments generated from the inverted form of U_{L2} would probably not have been detected by Southern analysis because the restriction fragments generated from the minority isomeric form of U_{L2} would have migrated sufficiently close to those generated from the majority isomeric form of U_{L2} to be obscured upon overexposure.

We conclude that if the U_{L2} sequence inverts, it does so only in a small proportion of the viral genomes, a proportion too small to be detected by the methods we used (Southern and sequence analyses of the uncloned DNA fragment). We estimate that we would have been able to detect isomerization of U_{L2} had it occurred in more than 5% of the genomes.

Analysis of ORFs. Several open reading frames (ORFs) were identified in the sequences of both J31 and I4. In particular, a large ORF terminating 445 nucleotides from the left genomic terminus (in the IRL) was identified in J31. The 5′ region of this ORF is not included in the sequence in Fig. 3. Northern (RNA) blot and S1 nuclease analyses showed that this ORF is transcribed (data not shown). Several ORFs were also observed in the sequences within I4 (i.e., the sequences near the standard junction of the L and S components present in mature virion DNA). The significance of these ORFs is unknown. Since these sequences can be deleted from the genome of viable virions (14, 15), they may be nonfunctional or may be genes that are not essential for growth in vitro. A detailed transcriptional analysis of the regions of the genome contained within clones I4 and J31 will be published elsewhere.

An ORF spanning the junction of the two genomic termini (in clone J31), which starts in the S component at nucleotide

LEFT L TERMINUS	5′	GGCCC AGCT[10]	CTCCCCC G[20]	AGCG CGGAT[30]
RIGHT L TERMINUS	5′	ATGCCGAGC	CTCCGCCCCG	GGGGTCGGAT
		CTCTGAAAAA[40]	AAAATTTCCC[50]	GCCC CGCGT[60]
		C GGATTA	TTTTCCT	CCCCGCGCGA

FIG. 8. Alignment of the sequences adjacent to the S components of the concatemeric and standard junctions. The sequences at the two ends of the L component can be aligned to reveal patchy homology.

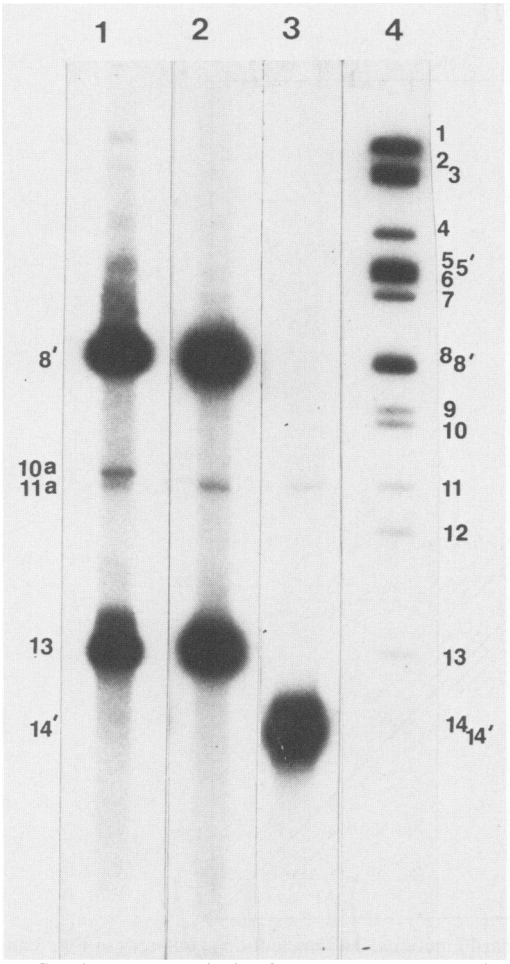

FIG. 9. Southern blot analysis of mature PRV DNA. Viral DNA isolated from purified virions was digested with *Bam*HI, electrophoresed, and transferred to nitrocellulose filters. Filter strips were probed with nick-translated cloned *Bam*HI fragment 8' (lane 1), cloned *Bam*HI fragment 13 (lane 2), cloned *Bam*HI fragment 14' (lane 3), or PRV DNA (lane 4).

81 and ends in the L component at nucleotide −654, was identified. Possible regulatory sequences (in the appropriate positions) were observed upstream. This ORF was of particular interest, because PRV mutants with genomes in which a segment of DNA derived from the left end of the L component has been translocated next to the L-S junction can be derived by serial passage of PRV in cells of avian origin in which they have a growth advantage over standard

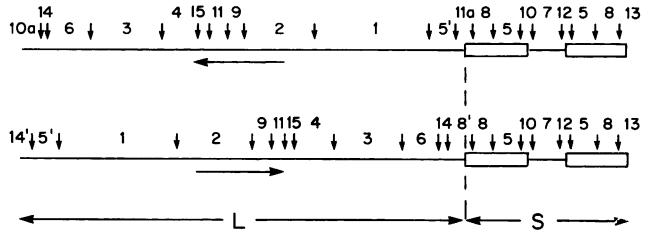

FIG. 10. *Bam*HI restriction map of the PRV genomes with L components in the inverted (top) and prototype (bottom) orientations.

PRV (14). Because the sequences derived from both ends of the genome are permanently juxtaposed in the mature genome of these mutants (15), it seemed possible that the growth advantage of these mutants is related to the presence of the ORF that spans the junction between the genome termini. Despite extensive efforts to detect a transcript originating from these sequences, we were, however, unable to do so. Therefore, the possibility that this ORF was not functional had to be considered (see below).

Nucleotide sequences at the concatemeric junctions of different PRV isolates. Concatemeric junction fragments comparable to J31 of PRV Ka from two other PRV isolates (Norden and USA7) were also cloned, and the nucleotide sequences spanning the junction were ascertained and compared with each other, as well as with those of J31 of PRV Ka. This was done to ascertain whether (i) insertions or deletions of nucleotides that would disrupt the ORF spanning the concatemeric junction of PRV Ka would be observed, thereby indicating that it is nonfunctional; (ii) differences between isolates exist in the nucleotide sequence at the junctions between the termini of the virions, i.e., at the cleavage site; (iii) differences between isolates exist in the sequence and in the number of different direct repeats present near the genomic termini.

A comparison of the nucleotide sequences at the concatemeric junction between the L and S termini of different isolates (Fig. 11) showed almost complete sequence conservation. The sequence and organization of the repeated sequences were strikingly similar, despite the different origins of the isolates. Only occasional differences in the nucleotide sequences were observed within the first 180 nucleotides of the S component. This sequence is composed of different sets of repeat elements interspersed among each other. Between the repeated elements, one or two spacer nucleotides are present. Interestingly, the occasional difference in nucleotides between the isolates involved mainly spacer nucleotides. Thus, in PRV Norden, the spacer GG doublet between R2 and R3 at nucleotides 123 and 124 were replaced by AA. Similarly, the spacer nucleotides GA between R3 and R2 at nucleotides 168 and 169 were replaced by AA.

The ORF that spans the concatemeric junction of the termini of the genome of PRV Ka (see above) is disrupted in the two other isolates; in these isolates, a single G was inserted at position −52, thereby disrupting this ORF.

Differences in the numbers of direct repeats present near the termini of the genome were observed. Thus, the direct repeat near the L terminus (R7) was repeated two and one-half times in the clone obtained from PRV Ka, eight times in that obtained from PRV Norden, and five times in that obtained from PRV USA7. Interestingly, the number of times R7 was repeated in the cloned fragment isolated from the genomes of the different isolates was similar to the number of times these repeats were reiterated in most of the genomes in the virion population; sequence analysis of the uncloned ends of the genome of these virus isolates revealed that most of the genomes in each population had a given number of R7 repeats, eight in PRV Norden, three in PRV Ka, and five in PRV USA7.

The repeated sequences R2 and R3 near the ends of the S component were also reiterated different numbers of times in the different isolates because of deletions between nucleotide 189 and 297 in PRV Norden and between nucleotides 197 and 262 in PRV USA7.

A difference in the nucleotides directly adjacent to the cleavage site between the genome of PRV Norden and the

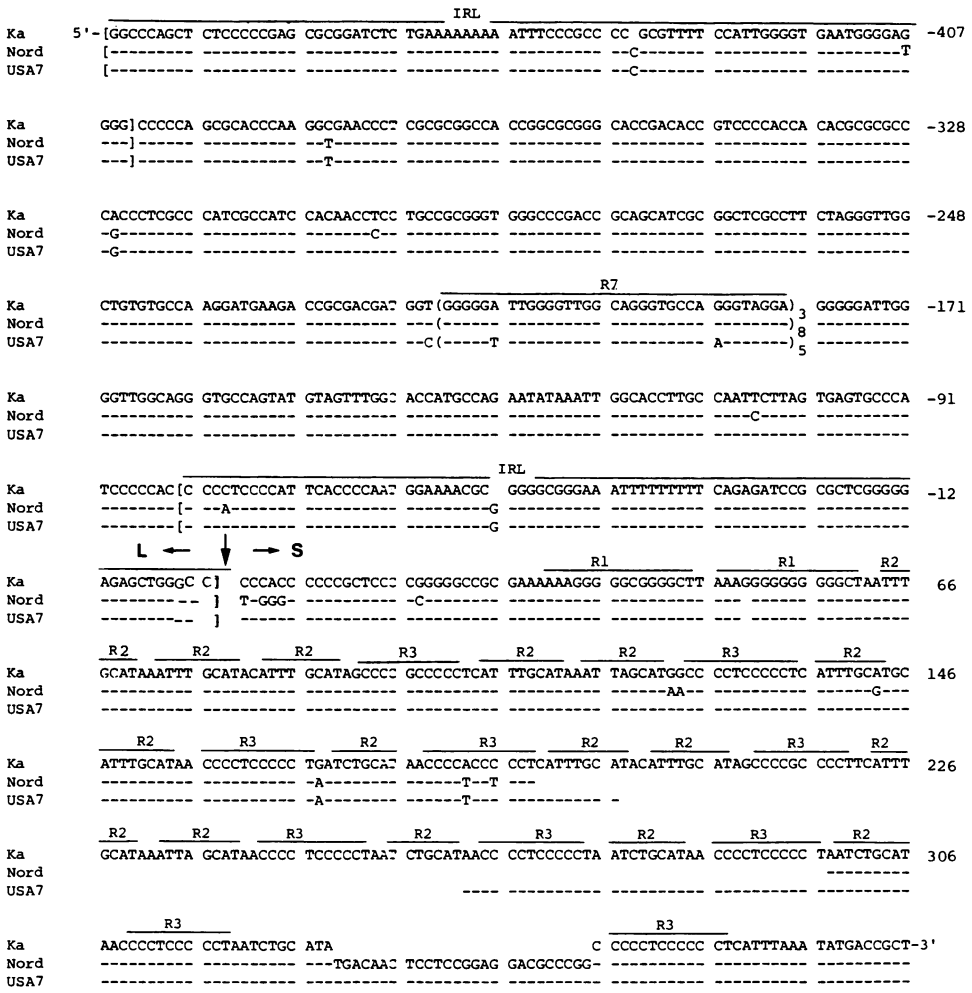

FIG. 11. Nucleotide sequences of the concatemeric junctions of three different isolates of PRV. The sequencing strategies were similar to that illustrated in Fig. 2. The vertical arrow indicates the junction between the S and L components. The repeated elements R1 to R7 are indicated. Blanks indicate deletions of nucleotides.

other two isolates was also observed. This difference was verified by sequencing the uncloned end of PRV Norden and PRV Ka (Fig. 12). The sequence illustrated in Fig. 12 was obtained by sequencing the 5′-end-labeled S terminus. This end carries a 3′ 2-base overhang (10). Thus, in addition to the difference in the three 5′-terminal nucleotides seen in Fig. 12, the 2-base overhang would, according to the sequence of the junction fragments (Fig. 11), also differ between PRV Ka and PRV Norden.

DISCUSSION

Sequence analysis of the several thousand nucleotides spanning the concatemeric and standard junctions between

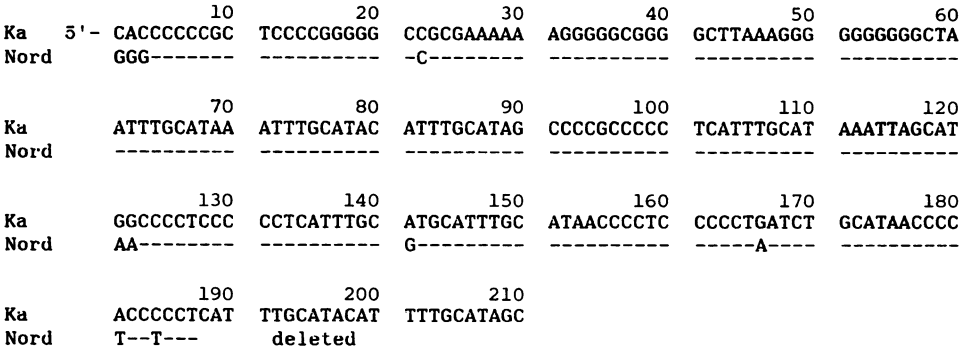

FIG. 12. Nucleotide sequences of the uncloned S termini of PRV Ka and PRV Norden. Virion DNA was digested with *Bam*HI, and fragment 13 was purified and sequenced as described previously (10).

the L and S components of PRV has revealed the following. (i) No significant homology between the two genome termini exists. This finding extends our previously published conclusions based on the analysis of 140 nucleotides at each end of the genome (10). (ii) There is patchy inverted homology between the left and right ends of the L component. This patchy homology led us to reexamine the question of the invertibility of the L component of PRV. We found that in PRV, as in VZV (7), few (less than 5%) of the mature genomes have an L component that is found in an inverted orientation compared with that found in most of the genomes. (iii) Eighty-two nucleotides present at the left terminus of the genome are found in inverted form (IRL) approximately 350 to 500 nucleotides (depending on the virus isolate) downstream. The distance between the IRLs depends on the number of times a direct repeat (R7) present in the intervening sequence between the repeats is reiterated. The unique sequence (U_{L2}) bracketed by the IRL is found mainly in one orientation in the mature genome and does also not appear to invert significantly during DNA replication. (iv) Several distinct direct repeats, some of which may vary in number in different isolates, are present near both termini of the genome. The sequences of these repeats and their organization are remarkably conserved in different isolates. (v) The sequences at the termini of the genomes of different isolates of PRV show some variability; the actual nucleotide sequence at which cleavage can occur can differ.

Direct as well as inverted repeats have been observed near the ends of the genomes of several herpesviruses and appear to be a hallmark of this family of viruses. Their function, however, remains largely unknown. Not surprisingly, several directly repeated sequences were found near both the L and S termini of PRV. The repeated sequences in the L and S components share no sequence homology.

Comparison of the nucleotide sequences at the termini of different isolates of PRV showed that the nucleotide sequences were almost identical in all of the isolates examined, except for an occasional nucleotide change, insertion, or deletion. Although some segments of DNA which included some of the repeats present near the ends of the genomes were deleted in some of the isolates, their sequence and organization were remarkably similar, despite the different geographic origins and histories of the isolates. (PRV Ka is a laboratory strain, PRV USA7 was recently isolated in the United States, and PRV Norden is a vaccine strain derived from a Romanian isolate.) For example, the organization of the repeat elements within the first 200 nucleotides of the S component is $(R1)_2\ (R2)_3\ (R3)_1\ (R2)_2\ (R3)_1\ (R2)_2\ (R3)_1\ (R2)_1$ in all three isolates. These repeated elements are separated by one or two spacer nucleotides that may differ. This sequence organization is the same in all three strains. Furthermore, many of the nucleotide variations observed between the isolates occurred in the spacer nucleotides. The similarity of the organization and sequences of the repeated elements is further illustrated by the finding that the repeated elements sometimes differed by one nucleotide. For example, R2 is usually ATTTGCAT, but in some cases it degenerates into ATCTGCAT or ATTAGCAT and the same sequence changes were observed in a given R2 in all three isolates analyzed. While these findings point to the possibility that the sequences and organization of the repeated elements may play an important role in the biology of the virus, the presence of all of the repeated elements does not, however, appear to be of pivotal importance to survival of the virus. Thus, stretches of DNA that include some of the R2 and R3 repeats are deleted from the genomes of different isolates. These stretches of DNA were, however, found beyond the first 190 nucleotides of the S component.

One of the repeated elements near the L terminus (R7) which consists of an almost perfect repetition of 34 bp differs greatly in number between different isolates. It is present in only three copies in PRV Ka but in eight copies in PRV Norden. Direct sequencing of the uncloned ends of the mature genomes of the different isolates revealed the same differences in the numbers of these repeats, indicating that each population of virions contained a majority of genomes with a given number of these repeats.

While an overall high degree of nucleotide conservation near the ends of the genomes was observed, the actual sequences at which cleavage of concatemeric DNA (to yield mature DNA) occurred differed significantly between PRV Ka and PRV Norden. The lack of consensus sequences at the actual cleavage site (despite almost total conservation of sequences surrounding the cleavage site) indicates that sequence specificity at the actual site of cleavage is not necessary for efficient cleavage of the genome of PRV and that cleavage is probably mediated mainly by a mechanism that recognizes distal signals. Indeed, when the last 15 nucleotides of the S component of PRV are replaced by unrelated sequences, the concatemeric DNA is cleaved and the virus is viable (S. Kuperschmidt and T. Ben-Porat, unpublished data), supporting a measuring mechanism. These findings confirm the report by Varmuza and Smiley (27), who showed that cleavage of concatemeric DNA of HSV is determined by the distance from specific signals and is not dependent on the actual sequence at the cleavage site.

The sequences at the right and left ends of the L component show patchy inverted homology; a vestigial inverted repeat appears to bracket the U_L sequence. In VZV, a perfect 88-bp repeat bracketing the U_L sequence has been observed, and approximately 5% of VZV genomes have a U_L sequence that is in an orientation different from that of the U_L sequences in most VZV genomes (7). We show here that, despite only patchy inverted homology between the ends of the L component of PRV, approximately 5% of the genomes have an L component that is inverted relative to that present in most of the genomes. The proportion of PRV genomes in which the L component is inverted remained similar after repeated passage at low or high multiplicity in RK cells. It was also similar in the genomes of the different isolates that we analyzed. Indeed, one would not expect enrichment of this form of genome upon replication, because the genome of PRV circularizes upon entering the cells and replicates as concatemers (3, 4); after circularization, genomes with either one of the two isomeric forms of the L component are indistinguishable from each other.

Isomerization of the L component probably results from the presence at both of its ends of sequences that share an inverted patchy homology. Isomerization could result from active inversion of the L component; alternatively, it could be generated by alternative cleavage at the junction between either end of the L component and the IR bracketing the U_S sequence. Indeed, Davison (7) has previously mentioned the possibility that isomerization of the L component of VZV arises by cleavage at these two alternative cleavage sites. We also favor this possibility on the basis of the following observation. While a small proportion of the L component (which is bracketed by a vestigial IRL only) inverts, the U_{L2} sequence (which is bracketed by an exact IRL) does not appear to invert to a significantly greater extent and our results exclude the possibility that more than 5% of the genomes contain an inverted U_{L2} sequence. If isomerization

of the L component were mediated by active inversion via the inverted repeats, one would expect the perfect repeat bracketing the U_{L2} sequence to be considerably more efficient in inversion than the patchy homology bracketing the U_L sequence.

Recognition signals necessary for cleavage have been shown to be present both upstream and downstream of the cleavage site of PRV (28), and cleavage occurs between the sequences of the IRL and the IR in the concatemeric junction. As mentioned above, the L component could therefore become inverted if cleavage between the partially homologous IRL and the internal IR occurred at the standard junction. However, because the IRL sequence at the standard junction is only partially homologous to the IRL present at the end of the L component, and possibly also because only part of the signals necessary for cleavage may be duplicated at the internal end of the L component, this cleavage site would be used at low frequency only.

Deiss et al. (8) and Hammerschmidt et al. (9) have previously aligned the sequences at the ends of the genomes of several herpesviruses. According to these alignments, the two unique ends of the class 2 genomes (VZV, BHV, and PRV) combine upon circularization of the genome to produce a sequence that has homology to the *a* sequence of HSV. In HSV, cleavage occurs in a DR1 repeat within adjacent *a* sequences and signals from both terminal *a* sequences are necessary (18, 26). Several conserved motifs can be discerned near the ends of the genomes of various herpesviruses, indicating that these sequences may be recognition signals for cleavage encapsidation. It is interesting that the elements that have been conserved to the greatest extent in the patchy inverted repeats bracketing the L component of PRV are not those that have been implicated on the basis of their conservation in different herpesvirus strains as putative signals for cleavage or inversion (8, 9). Thus, the An (or Tn) motif that has been noted 31 to 34 nucleotides upstream of the left terminus of class 2 genomes (and at an equivalent position in the *a* sequence of class 3 genomes) (7–9) is not conserved in the vestigial repeat at the right end of the L component of PRV. Thus, if inversion of the L component of PRV is indeed the result of alternative cleavage due to the presence of cleavage signals in the internal vestigial IRL (as we believe it to be), the long An (or Tn) element that is present near the cleavage site of many herpesviruses does not appear to be an essential element for cleavage recognition.

ACKNOWLEDGMENT

This investigation was supported by Public Health Service grant AI-10947 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. **Ben-Porat, T.** 1981. Replication of herpesvirus DNA. Curr. Top. Microbiol. Immunol. **91**:81–107.
2. **Ben-Porat, T., J. M. DeMarchi, and A. S. Kaplan.** 1974. Characterization of defective interfering viral practices present in a population of pseudorabies virions. Virology **61**:29–37.
3. **Ben-Porat, T., and F. J. Rixon.** 1979. Replication of herpesvirus DNA. IV. Analysis of concatemers. Virology **94**:61–70.
4. **Ben-Porat, T., F. J. Rixon, and M. Blankenship.** 1979. Analysis of the structure of the genome of pseudorabies virus. Virology **95**:285–294.
5. **Ben-Porat, T., R. A. Veach, and B. F. Ladin.** 1980. Replication of herpesvirus DNA. VI. Virions containing either isomer of pseudorabies virus DNA are infectious. Virology **102**:370–380.
6. **Chou, J., and B. Roizman.** 1985. Isomerization of herpes simplex type 1 genome. Identification of the cis-action and recombinational sites within the domain of the "a" sequence. Cell **41**:803–811.
7. **Davison, A.** 1984. Structure of the genome termini of varicella-zoster virus. J. Gen. Virol. **65**:1969–1977.
8. **Deiss, L. P., J. Chou, and N. Frenkel.** 1986. Functional domains within the *a* sequence involved in the cleavage-packaging of herpes simplex virus DNA. J. Virol. **59**:605–618.
9. **Hammerschmidt, W., H. Ludwig, and H.-J. Buhk.** 1988. Specificity of cleavage in replicative-form DNA of bovine herpesvirus 1. J. Virol. **62**:1355–1363.
10. **Harper, L., J. DeMarchi, and T. Ben-Porat.** 1986. Sequence of the genome ends and of the junction between the ends in concatemeric DNA of pseudorabies virus. J. Virol. **60**:1183–1185.
11. **Hayakawa, Y., and R. W. Hyman.** 1987. Isomerization of the U_L region of varicella-zoster virus DNA. Virus Res. **8**:25–31.
12. **Honess, R. W., and D. H. Watson.** 1977. Unity and diversity in the herpesvirus. J. Gen. Virol. **37**:15–37.
13. **Ladin, B. F., S. Ihara, H. Hampl, and T. Ben-Porat.** 1982. Pathway of assembly of herpesvirus capsids. An analysis using DNA^+ temperature-sensitive mutants of pseudorabies virus. Virology **116**:544–561.
14. **Lomniczi, B., A. Gielkens, I. Csobai, and T. Ben-Porat.** 1987. Evolution of pseudorabies virions containing genomes with an invertible long component after repeated passage in chicken embryo fibroblasts. J. Virol. **61**:1772–1780.
15. **Lu, Z., J. M. DeMarchi, L. Harper, G. F. Rall, and T. Ben-Porat.** 1989. Nucleotide sequences at recombinational junctions present in pseudorabies virus variants with an invertible L component. J. Virol. **63**:2690–2698.
16. **Maxam, A. M., and T. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
17. **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.
18. **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.
19. **Platt, K. B., C. J. Mare, and P. N. Hinz.** 1979. Differentiation of vaccine strains and field isolates of pseudorabies virus: thermal sensitivity and rabbit virulence markers. Arch. Virol. **60**:13–23.
20. **Rigby, P. W., J. M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
21. **Roizman, B.** 1982. The family Herpesviridae: general description, taxonomy, and classification, p. 1–23. *In* B. Roizman (ed.), The herpesviruses, vol. 1. Plenum Publishing Corp., New York.
22. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
23. **Smiley, J. R., B. S. Fong, and W.-C. Leung.** 1881. Construction of a double-jointed herpes simplex viral DNA molecule: inverted repeats are required for segment inversion and direct repeats promote deletions. Virology **113**:345–362.
24. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
25. **Stevely, W. S.** 1977. Inverted repetition in the chromosome of pseudorabies virus. J. Virol. **22**:232–234.
26. **Stow, N. D., E. C. McMonagle, and A. J. Davison.** 1983. Fragments from both termini of HSV-1 genome contain signals required for the encapsidation of viral DNA. Nucleic Acids Res. **11**:8205–8220.
27. **Varmuza, S. L., and R. J. Smiley.** 1985. Signals for site-specific cleavage of HSV DNA: maturation involves two separate cleavage events at sites distal to the recognition sequences. Cell **41**:793–802.
28. **Wu, C. A., T. Harper, and T. Ben-Porat.** 1986. *cis* functions involved in replication and cleavage-encapsidation of pseudorabies virus. J. Virol. **59**:318–327.