JOURNAL OF VIROLOGY, Dec. 1986, p. 1183–1185
0022-538X/86/121183-03$02.00/0

Vol. 60, No. 3

Sequence of the Genome Ends and of the Junction between the Ends in Concatemeric DNA of Pseudorabies Virus

LINDA HARPER, JEAN DEMARCHI, AND TAMAR BEN-PORAT*

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

Received 14 July 1986/Accepted 1 September 1986

The nucleotide sequences of the termini of the mature pseudorabies virus genome and of the junction between these termini in concatemeric DNA were compared. To ensure conservation of unmodified 5′ and 3′ termini, the end fragments obtained directly (uncloned) from mature viral DNA were sequenced. The sequence obtained from 5′ and 3′ end labeling revealed that whereas the L terminus was blunt ended, the S terminus had a 2-base (GG) 3′ overhang. The sequences spanning the junction between the termini present in concatemeric DNA was also determined and compared with that expected when the two ends of the mature DNA were juxtaposed. This comparison showed that in concatemeric DNA the ends of the mature genome had become joined by blunt-end ligation of one of the strands and that the 2-nucleotide gap on the other strand had been repaired. A significant degree of homology between the sequences spanning the junction between the ends of the varicella-zoster virus and pseudorabies virus genomes was found.

The genome of pseudorabies virus (PrV), a herpesvirus, is divided into the unique short (U_S) and unique long (U_L) components; the U_S component is bracketed by inverted repeats and is found in two orientations relative to the U_L component (1). Upon entering the cell nucleus, the PrV genome forms circles (6, 10). The first round of replication occurs mainly as theta-type structures (10), but thereafter the DNA is found in the form of head-to-tail concatemers (3). Circle formation of the parental genome does not require the expression of viral functions (10).

In contrast to herpes simplex virus (HSV), in which sequence homology between the two ends of the viral genome exists (the "a" sequence) (17), no cross hybridization between the two end fragments of the PrV genome can be detected (see, for example, reference 12). Furthermore, in contrast to the HSV genome, which forms circles after in vitro exposure of its complementary single-stranded ends by digestion with exonucleases, no such circle formation of PrV DNA molecules is observed after similar treatment (4). It is unlikely, therefore, that PrV DNA can circularize by ligation of cohesive ends that may become exposed within the cells. Indeed, the structures of the joints between the end fragments found in the concatemeric genomes of both varicella-zoster virus (VZV) and HSV are consistent with blunt ligation of their ends (7, 16), and this may also be the case for PrV.

The experiments described here were designed to determine the structure of the ends of the PrV genome and to clarify the mechanism by which circularization of the PrV genome occurs. To this end, we compared the nucleotide sequence of the termini of the mature genome with that of the junction between the termini found in concatemeric intracellular viral DNA. Because of possible alterations in the end sequences of the genome as a result of cloning procedures, i.e., to ensure conservation of unmodified 5′ and 3′ termini, the end fragments from mature DNA were sequenced directly.

The *Bam*HI restriction fragments originating from the ends of the mature genome (Fig. 1) were extruded from agarose gels. They were purified and labeled either at the 5′ end with [^{32}P]ATP or at the 3′ end with [α-^{32}P]-cordycepin-5′-triphosphate essentially as described by Maniatis et al. (14) and Maxam and Gilbert (15). The end-labeled fragments were digested with a second restriction enzyme (Fig. 1) and were sequenced by the chemical degradation technique as described by Maxam and Gilbert (15). An outline of the experimental approach used in this experiment is shown in Fig. 1A.

The 140-nucleotide sequence at either end of the genome is shown in Fig. 2. The L component of the viral genome was found to be blunt ended; the S terminus, however, had a 2-base (GG) 3′ overhang. A computer-aided analysis of the end sequences did not reveal any extensive direct or inverted sequence homologies between them.

To ascertain how the two ends of the genome are joined within concatemeric DNA, the *Bam*HI junction fragment obtained from intracellular concatemeric DNA was cloned in pBR325 as described previously (11). The region spanning the joint between the ends present in this clone (J31) was sequenced by using as a guide the previously determined sequence of the genome ends. The sequence of the end fragments that was available (Fig. 2) showed that a *Sma*I restriction site was present 14 nucldotides from the S terminus and that a *Fok*I site was present 88 nucleotides from the L terminus. Therefore, clone J31 was digested with *Sma*I, the fragments were mapped, *Sma*I fragment 4 (which spans the junction) was 5′ end labeled and cleaved with *Fok*I, and the resulting end-labeled fragment of 102 nucleotides was sequenced. The sequence spanning the joint that was thus determined, as well as that expected when the ends of the mature DNA are juxtaposed, is shown in Fig. 3. A comparison of the two shows that in concatemeric DNA the mature genome ends had become joined by blunt-end ligation of one of the strands and that repair of the 2-nucleotide gap on the other strand had taken place.

On the basis of the analysis of the nucleotide sequence of the mature genome ends, as well as that of the junction between the ends in concatemeric DNA, we conclude the following. (i) Concatemeric PrV DNA is cleaved by a mechanism giving rise to a 2-nucleotide 3′ overhang. The

* Corresponding author.

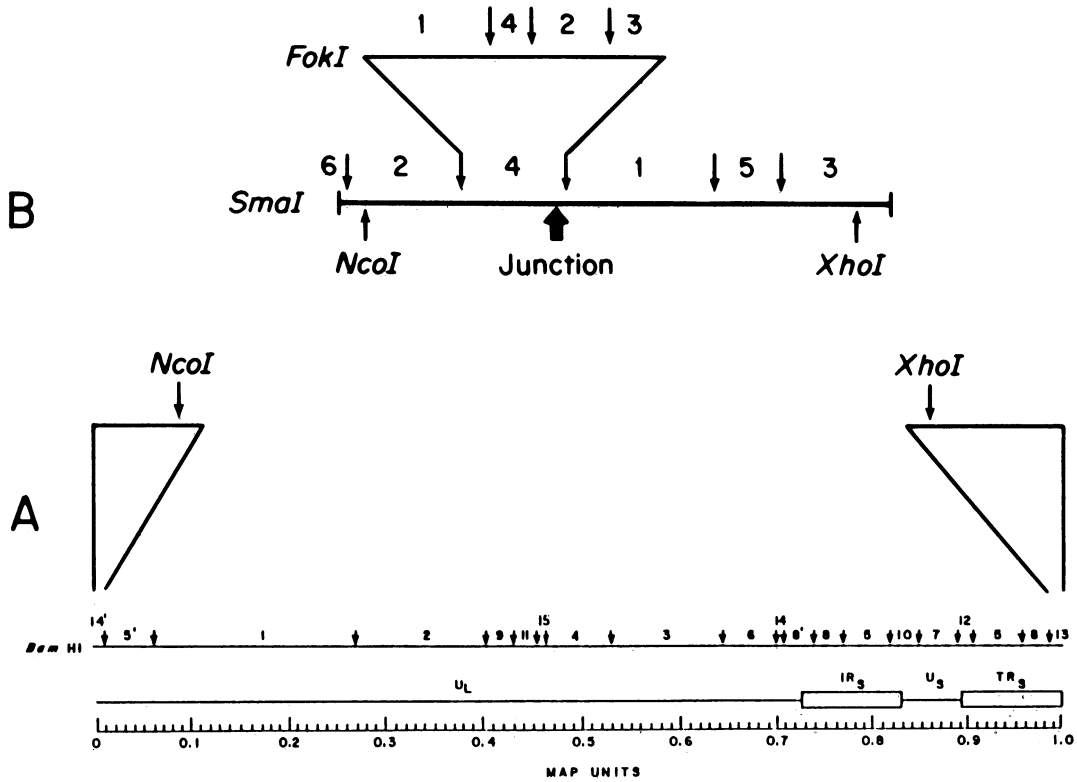

FIG. 1. Strategy used to sequence termini of mature genome and junction between these termini in concatemeric DNA of PrV. (A) The band containing the *Bam*HI end fragments was excised from agarose gels, and the DNA was electroeluted and purified. The fragments were labeled at their 3′ or 5′ ends, as described in the text, and cleaved with either *Xho*I (fragment *Bam*HI 13; S terminus) or *Nco*I (fragment *Bam*HI 14′; L terminus). U_L, Unique long component; U_S, unique short component; IR_S, internal repeat (short); TR_S, terminal repeat (short). (B) The cloned junction fragment (J31) obtained from concatemeric DNA was digested with *Sma*I, and SmaI fragment 4, which spans the junction, was labeled at its 5′ end and cleaved with *Fok*I. The resulting end-labeled fragment that spans the junction was sequenced.

overhang is retained on the S terminus of the genome only; the L component becomes blunt ended. (ii) The DNA forms circles (and concatemers) by blunt ligation of the molecule ends, the gap resulting from the 2 missing nucleotides on the L terminus being repaired. (iii) There are no extensive stretches of sequence homology between the 140 nucleotides at the L and S termini of the PrV genome. (iv) An octamer

```
S TERMINUS  5′  CACCCCCCGC TCCCCGGGGG CCGCGAAAAA AGGGGGCGGG
            3′  GGGTGGGGGGCG AGGGGCCCCC GGCGCTTTTT TCCCCCGCCC

                GCTTAAAGGG GGGGGGGCTA ATTTGCATAA ATTTGCATAC
                CGAATTTCCC CCCCCCCGAT TAAACGTATT TAAACGTATG

                ATTTGCATAG CCCCGCCCCC TCATTTGCAT AAATTAGCAT
                TAAACGTATC GGGGCGGGGG AGTAAACGTA TTTAATCGTA

                GGCCCCTCCC CCTCATTTGC
                CCGGGGAGGG GGAGTAAACG ---------- ----------

L TERMINUS                          CACCATGCCG GAATATAAAT
                ---------- ---------- GTGGTACGGC CTTATATTTA

                TGGCACCTTG CCAATTCTTA GTGAGTGCCC ATCCCCCACC
                ACCGTGGAAC GGTTAAGAAT CACTCACGGG TAGGGGGTGG

                CCCTCCCCAT TCACCCCAAT GGAAAACGCG GGGCGGGAAA
                GGGAGGGGTA AGTGGGGTTA CCTTTTGCGC CCCGCCCTTT

                TTTTTTTTTC AGAGATCCGC GCTCGGGGGA GAGCTGGGCC 3′
                AAAAAAAAAG TCTCTAGGCG CGAGCCCCCT CTCGACCCGG 5′
```

FIG. 2. Nucleotide sequence of termini of PrV genome. The termini of the genome were labeled at either the 5′ end with [γ-^{32}P]ATP or the 3′ end with [α-^{32}P]-cordycepin-5′-triphosphate, and the sequence of the terminal 140 nucleotides of both ends of the genome was determined.

TAAACGTA (which is underlined in Fig. 2) is repeated several times within the sequences near the S terminus.

The free termini of the L and S segments of HSV DNA share a DR-I sequence, and both the S and the L compo-

```
L TERMINUS                    S TERMINUS
                                  SMA
5′ GGCCCAGCTC TCCCCCGAGC      CCCCCGGGGA GCGGGGGGTGGG 3′
3′ CCGGGTCGAG AGGGGGCTCG      GGGGGCCCCT CGCCCCCCAC 5′

S TERMINUS                    L TERMINUS

JUNCTION (DEDUCED)
GGGGGGTGGGGGCCCAGCTC
CCCCCCAC CCGGGTCGAG

JUNCTION (ASCERTAINED)
SMAI
5′ GGGGAGCGGGGGGTGGGGGCCCAGCTCT
CCCCTCGCCCCCCACCCCCGGGTCGAGA
```

FIG. 3. Comparison of sequences spanning the joint in concatemeric DNA and of the juxtaposed termini of mature viral DNA.

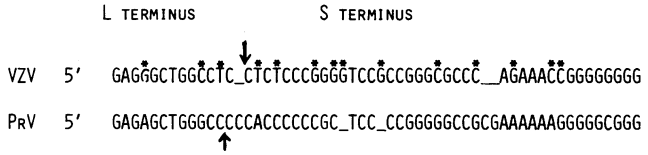

FIG. 4. Comparison of sequences of VZV and PrV junctions. The sequence of the VZV junction was reported by Davison (7). Asterisks indicate differences in nucleotides; dashes indicate missing nucleotides. The arrows indicate the points of cleavage of the concatemeric DNA.

nents have a 3′ complementary single-base extension at their ends (16). Similarly, the ends of the VZV genome, although otherwise nonhomologous, have a complementary base 3′ overhang (7). The PrV genome is somewhat different; a 2-base overhang is present on the 3′ end of the S component, whereas the L component is blunt ended. It appears that 2 bases from the 3′ end of the L component must somehow be lost from the mature genomes but are regenerated as the genomes form circles or concatemers, since they are present in the concatemeric DNA. The events leading to the loss of the 2-base 3′ overhang on the L terminus are unknown.

A computer-aided analysis of sequence homology between the ends of the PrV and HSV genomes revealed no striking similarities between the two, although short stretches of homology were observed.

A much more striking homology was observed between the junctions of the PrV and VZV genomes. Interestingly, the overall sequences of the VZV and PrV genomes differ greatly; the G+C content of VZV is 47 mol% (13), whereas that of PrV is 73 mol% (2), and the genomes of the two viruses share a maximum of only 1% homology (8). The homology between the junction sequences of the VZV and PrV genomes may, therefore, be significant. The sequences of the joints between the L and S components found in concatemeric DNA of the two viruses are compared in Fig. 4.

A striking homology between the sequences of the L terminus of PrV and bovine herpesvirus I (9) also exists. With some adjustment for sequence insertion or deletion, a homology of 75% was observed (data not shown).

A comparison of the sequence homology between the termini of PrV, VZV, and bovine herpesvirus I genomes thus indicates extensive homology. All of these viruses have similar genome structures. It remains to be established whether the observed sequence homology between the termini of the genomes of these viruses has particular significance with respect to the functions of these sequences in their life cycles.

We thank M. Haralson for performing the computer-aided analysis of the nucleotide sequences.

This investigation was supported by Public Health Service grant AI-10947 from the National Institutes of Health.

LITERATURE CITED

1. **Ben-Porat, T.** 1982. Organization and replication of herpesvirus DNA, p. 147–172. *In* A. S. Kaplan (ed.), Organization and replication of viral DNA. CRC Press, Inc., Boca Raton, Fla.
2. **Ben-Porat, T., and A. S. Kaplan.** 1962. The chemical composition of herpes simplex and pseudorabies viruses. Virology **16**:261–266.
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., and R. A. Veach.** 1980. Origin of replication of the DNA of a herpesvirus (pseudorabies). Proc. Natl. Acad. Sci. USA **77**:172–175.
6. **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.
7. **Davison, A. J.** 1984. Structure of the genome termini of varicella zoster virus. J. Gen. Virol. **65**:1969–1977.
8. **Davison, A. J., and N. M. Wilkie.** 1983. Location and orientation of homologous sequences in the genomes of five herpesviruses. J. Gen. Virol. **64**:1927–1942.
9. **Hammerschmidt, W., H. Ludwig, and H.-J. Buhk.** 1986. Short repeats cause heterogeneity at the genome terminus of bovine herpesvirus 1. J. Virol. **58**:43–49.
10. **Jean, J. H., M. L. Blankenship, and T. Ben-Porat.** 1977. Replication of herpesvirus DNA. I. Electron microscopic analysis of replicative structures. Virology **79**:281–291.
11. **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.
12. **Lomniczi, B., M. L. Blankenship, and T. Ben-Porat.** 1984. Deletions in the genomes of pseudorabies virus vaccine strains and existence of four isomers of the genomes. J. Virol. **49**:970–979.
13. **Ludwig, H., H. G. Haines, N. Biswal, and M. Benyesh-Melnick.** 1972. The characterization of varicella zoster DNA. J. Gen. Virol. **14**:111–114.
14. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **Mocarski, E. S., and B. Roizman.** 1982. Structure and role of herpes simplex virus DNA termini in inversion, circularization and generation of virion DNA. Cell **31**:87–97.
17. **Roizman, B.** 1979. The structure and isomerization of herpes simplex virus genomes. Cell **16**:481–494.