JOURNAL OF VIROLOGY, Sept. 1986, p. 605–618
0022-538X/86/090605-14$02.00/0

Vol. 59, No. 3

Functional Domains within the *a* Sequence Involved in the Cleavage-Packaging of Herpes Simplex Virus DNA

LOUIS P. DEISS, JOANY CHOU, AND NIZA FRENKEL*

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637

Received 24 February 1986/Accepted 16 May 1986

Newly replicated herpes simplex virus (HSV) DNA consists of head-to-tail concatemers which are cleaved to generate unit-length genomes bounded by the terminally reiterated *a* sequence. Constructed defective HSV vectors (amplicons) containing a viral DNA replication origin and the *a* sequence are similarly replicated into large concatemers which are cleaved at *a* sequences punctuating the junctions between adjacent repeat units, concurrent with the packaging of viral DNA into nucleocapsids. In the present study we tested the ability of seed amplicons containing specific deletions in the *a* sequence to become cleaved and packaged and hence be propagated in virus stocks. These studies revealed that two separate signals, located within the Ub and Uc elements of the *a* sequence, were essential for amplicon propagation. No derivative defective genomes were recovered from seed constructs which lacked the Uc signal. In contrast, propagation of seed constructs lacking the Ub signal resulted in the selection of defective genomes with novel junctions, containing specific insertions of *a* sequences derived from the helper virus DNA. Comparison of published sequences of concatemeric junctions of several herpesviruses supported a uniform mechanism for the cleavage-packaging process, involving the measurement from two highly conserved blocks of sequences (*pac-1* and *pac-2*) which were homologous to the required Uc and Ub sequences. These results form the basis for general models for the mechanism of cleavage-packaging of herpesvirus DNA.

The 150-kilobase (kb) DNA genomes of standard herpes simplex viruses 1 and 2 (HSV-1 and HSV-2) are composed of two components, L and S, each consisting of unique sequences bracketed by inverted repeats, *ab* and *b'a'* flanking U_L, and *ac* and *c'a'* flanking U_S. The L and S components invert relative to each other, yielding four genomic isomers (reviewed in reference 25). The L-S junctions and the L termini of all four isomers contain a variable number of tandemly reiterated *a* sequences. In contrast, each of the S termini contains only a single copy of the *a* sequence (15, 38). The standard virus genome can thus be represented by $a_n b$-U_L-$b'a'_m c'$-U_S-*ca*, with *n* and *m* varying from 1 to greater than 10.

The linear viral DNA molecules circularize soon after infection (22), and newly replicated viral DNA molecules consist of large head-to-tail concatemers, arising most probably by rolling-circle replication (2, 3, 12). The large concatemers are subsequently cleaved within *a* sequences situated at the junctions between adjacent viral genomes (6, 21, 25). Studies with temperature-sensitive mutants suggested that the cleavage process may be linked to packaging (13, 14, 23; S. Bachenheimer, personal communication).

Defective virus genomes present in serially passaged virus stocks are typically composed of head-to-tail repeat units, each containing a DNA replication origin and the *a* sequence (reviewed in reference 8). In the presence of helper virus, multimeric defective genomes can also be generated from seed plasmids (amplicons) containing a DNA replication origin and the *a* sequence, which is required for concatemeric cleavage and packaging (7, 28, 29, 31, 32, 36). Any pair of *a* sequences along the concatemeric molecule could become cleaved, resulting in the packaging of monomers, dimers, and higher multimeric forms of defective genomes into nuclear capsids. However, only full-length (approximately 150 kb) defective genomes were found to be present in cytoplasmic virions (37). Like the L-S junctions in standard viral DNA, the junctions separating the repeat units of defective genomes contain a variable number of tandemly reiterated *a* sequences, which are cleaved to generate a variable number of *a* sequences at one terminus and predominantly a single *a* sequence at the other terminus (7).

The *a* sequences of HSV-1 and HSV-2 strains range in size from 250 to 500 base pairs (bp) and contain both unique (U) and directly repeated (DR) elements (6, 18, 19, 35). The size differences predominantly reflect variability in the copy number of the DR elements. For example, the 501-bp *a* sequence of HSV-1 (F) is represented by DR1-Ub-$(DR2)_{19}$-$(DR4)_3$-Uc-DR1, whereas the 264-bp *a* sequence of the Justin strain contains only two copies of DR2, a single copy of DR4, and two copies of another reiteration designated DR3.5 (18, 19). In a double *a* (*baac*) junction the tandemly reiterated *a* sequences share the intervening DR1 (20 bp) element (6, 19, 21). Furthermore, in the F strain, the L terminus (*ba*) contained 18 bp plus one 3' overhang base of the DR1 sequence, whereas the S terminus (*ca*) carried the remaining 1 1/2 bp. These findings led Mocarski and Roizman (21) to propose that the double *a* junctions formed the substrate for cleavage. In this regard, the cleavage of viral DNA concatemers represented a puzzle inasmuch as, although the number of *a* sequences at the junctions is variable, the predominant junctions in both standard and defective viral genomes contain a single *a* sequence (7, 15, 38). We have recently shown that the junctions containing a single *a* sequence can indeed serve as the target for cleavage-packaging. Furthermore, the cleavage process itself appears to be accompanied by the amplification of an *a* sequence, inasmuch as molecules carrying junctions of the structure *xay* are cleaved to generate *xa* and *ya* termini (7).

The present investigation was designed to further map the signals within the *a* sequence which are involved in the

* Corresponding author.

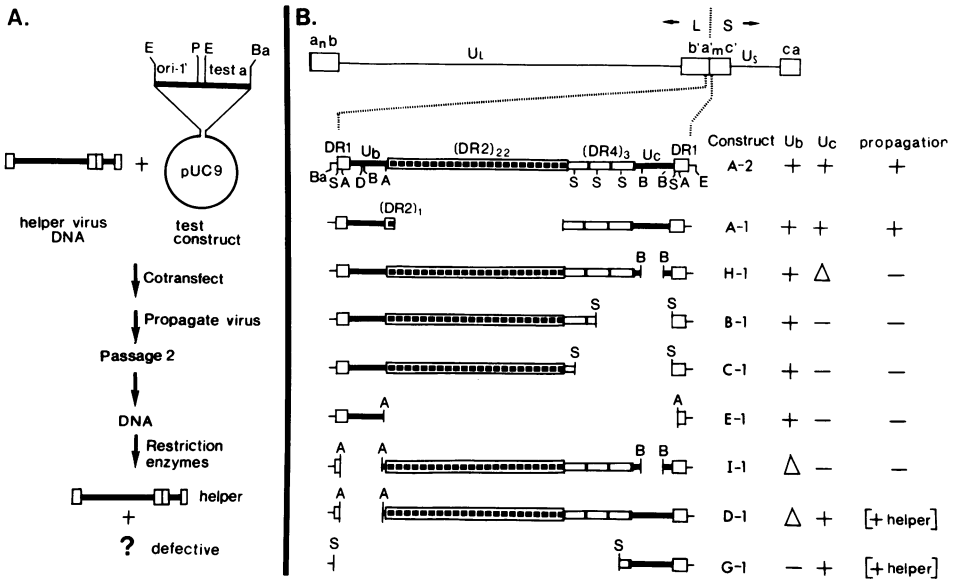


FIG. 1. Propagation test and structure of test constructs. (A) Scheme of the propagation test (28). (B) Structure of the test *a* sequences. Top line, schematic representation of the standard viral genome and the location of the *a* sequence in the L-S junction. Second line, arrangement of the directly repeated (DR) and unique (U) elements in the *a* sequence of HSV-1 F DNA as described by Mocarski and Roizman (19). Shown are the sites for *Bam*HI (Ba), *Sst*II (S), *Apa*I (A), *Dra*I (D), *Bss*HII (B), and *Eco*RI (E). The test *a* sequences (4) are contained between the *Bam*HI and the *Eco*RI sites in the general construct shown in panel A. Symbols: + or −, presence or absence of intact Ub or Uc elements; Δ, deletions within the Ub or Uc elements. The results of the propagation tests are described as efficient propagation (+), lack of propagation (−), and propagation at low efficiency coupled to selection of genomes with inserted helper *a* sequence (+ helper).

cleavage-packaging process. Analyses of the products resulting from the propagation of seed amplicons containing various deleted versions of the *a* sequence revealed the presence of two regions within the *a* sequence that are essential for efficient propagation. The results also revealed aspects of the mechanism by which single *a*-containing junctions become cleaved.

MATERIALS AND METHODS

Cells and viruses. HEp-2 cells and Vero cells were obtained from American Type Culture Collection, Rockville, Md., rabbit skin cells and HSV-1 strain F were obtained from B. Roizman, University of Chicago, and HSV-1 strain Justin was obtained from A. Sabin, who was then at the University of Cincinnati, Cincinnati, Ohio.

Derivation of plasmids. Cloning was done as previously described (28) with *Escherichia coli* DH1 (gift of D. Hanahan, Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.). Restriction enzymes and T4 DNA ligase were purchased from New England BioLabs and Bethesda Research Laboratories. Calf alkaline phosphatase was purchased from Boehringer Mannheim Biochemicals, and polynucleotide kinase was purchased from P-L Biochemicals, Inc. All enzymes were used as recommended by the suppliers. The test constructs contained the sets of deleted *a* sequences which were constructed by Chou and Roizman (4) as described in detail by them, starting from a 537-bp *a* sequence from HSV-1 strain F. Briefly, the constructs were made in pUC9 and contained 3 bp of *b* sequences and 12 bp of *c* sequences in addition to the *a* sequences. A series of deleted *a* subclones were made (Fig. 1B). In the present study, a 1.90-kb *Eco*RI fragment, containing the HSV DNA replication origin ori-1′ (ori_S; 20,31,36) was inserted between the *Eco*RI sites of the deleted *a* clones derived by Chou and Roizman (4), yielding the general structure in Fig. 1A. The inserted *Eco*RI fragment contained a 0.26-kb *Eco*RI-to-*Pvu*II segment of pKC7 (24) and the 1.64-kb ori-1′ *Pvu*II-to-*Eco*RI segment, derived from map coordinates 0.856 to 0.867 of HSV-1 F DNA. This ori-1′ segment was derived originally from pRB373 (20) and was transferred to pKC7 as described in our previous study (7). The construct A-2 contained the intact *a* sequence (537 bp) and was derived by the insertion of ori-1′ into pRB3119; clone A-1 was picked as a colony during the preparation of the A-2 construct. H-1 was derived from pRB3613 which contained a *Bss*HII deletion in Uc; B-1 was derived from pRB3384 which contained a *Sst*II deletion fusing the second DR4 element to the DR1 element; C-1 was derived from pRB3385 which contained a *Sst*II deletion fusing the first DR4 element to DR1; E-1 was derived from pRB3387 which contained an *Apa*I deletion fusing the Ub element to DR1; I-1 was derived from pRB3614 which contained a *Bss*HII deletion in Uc as well as an *Apa*I deletion in Ub; D-1 was derived from pRB3386 which contained an *Apa*I deletion in Ub; G-1 was derived from pRB3389 which contained a *Sst*II deletion fusing the DR1 element to the third DR4 element. Plasmids pRB3386, pRB3613, and pRB3614 underwent intermediate recloning into pUC9 before the ori-1′ segment was inserted, to obtain a uniform orientation of the deleted *a* segments in the plasmids. A sister set of plasmids was constructed with the *Eco*RI origin fragment in the opposite orientation. The constructs G-2, G-3, and G-4 were rescued from G-1-propagated defective genomes, as described in the text. Constructs G-5 and G-6 were constructed (to eliminate *Dra*I restriction sites) by insertion of a *Pvu*II fragment which contained ori-1′ and the G-deleted *a* sequence into a *Dra*I-cleaved pBR322 plasmid. The *Pvu*II fragment was taken from the sister clone of G-1, which contained the *Eco*RI ori-1′ segment in reverse orientation to that in G-1. The

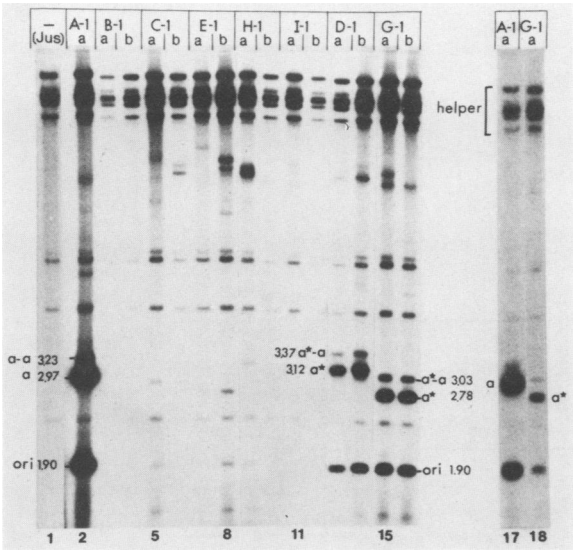


FIG. 2. Analyses of DNA from transfection-derived virus stocks. *Eco*RI digests of ^{32}P-labeled DNA from cells infected with passage 2 of a transfection receiving helper HSV-1 Justin alone (lane 1) or with passage 2 of duplicate cotransfections (a and b) receiving helper virus along with the test constructs shown (lanes 2 through 16). The group of fragments visible in lane 1 correspond to helper virus DNA fragments. Shown are bands containing the replication origin (ori), the intact single and double *a* junctions (*a* and *a-a*, respectively), the test *a* sequences (*a**), and the chimeric junctions (*a**-*a*). Fragments arising from digestion of the generated defective genome are marked by their sizes in kilobases. Lanes 17 and 18 show a shorter autoradiographic exposure of lanes 2 and 15, respectively.

*Pvu*II-*Dra*I fusion resulted in the loss of both restriction sites. G-5 and G-6 differed in the orientation of the *Pvu*II insert.

Cotransfection and propagation of virus stocks. Cotransfection of rabbit skin cells with helper virus DNA and test constructs was done as previously described (28) and as outlined in Fig. 1A.

Analysis of infected cell DNA. Vero cell cultures (25 cm^2) were infected with passage 2 virus and labeled with $^{32}P_i$ as previously described (28). At 20 h postinfection, DNA was extracted from the infected cells by lysis with a mixture of sodium dodecyl sulfate and proteinase K which was followed by phenol-chloroform extraction and ethanol precipitation as previously described (28). Following digestion with the appropriate restriction enzyme, the DNA was electrophoresed in 0.5% agarose gels; 1/50 of the DNA produced from an infected 25-cm^2 culture was used per lane.

Nucleotide sequence analysis. For the nucleotide sequence analyses the constructs G-2, G-3, and G-4 were each cleaved with either *Eco*RI-*Pst*I or *Bam*HI-*Sst*I. The appropriate fragments containing the *a* sequence were electroeluted, and the *Eco*RI or *Bam*HI sites were end labeled with polynucleotide kinase. The nucleotide sequence was determined by the Maxam-Gilbert technique (17).

RESULTS

Propagation of amplicons containing deletions in the *a* sequence. As reviewed above, the *a* sequence of HSV DNA was found to be essential for the cleavage and packaging of viral DNA and therefore for the propagation of defective genomes in serially passaged virus stocks. To map the cleavage-packaging signal(s) within the *a* sequence we tested the ability of test constructs containing specific deletions in the *a* sequence to become propagated in virus stocks. Figure 1A summarizes the general structure of these test constructs and the features of the propagation assay. All of the test constructs contained the replication origin ori-1′. In addition, they contained either an intact *a* sequence derived from HSV-1 F or various deleted *a* segments derived from it (Fig. 1B). The intact *a* as well as the deleted *a* segments were derived from plasmids constructed previously by Chou and Roizman (4), starting from the cloned *a* sequence of HSV-1 F DNA. All constructs used in the present study were designated according to the *a* sequence which they contained; the lettering system of the original study of Chou and Roizman (4) was used. The construct A-1, which was used as a control for this set of propagation tests, contained an *a* sequence of 285 bp, i.e., close to the average size of the deleted test *a* constructs (range 86 to 497 bp). The 285-bp variant *a* sequence contained only a single DR2 element and in that respect resembled the shorter type *a* sequences found in some HSV-1 strains (6, 18, 35).

In the propagation tests, duplicate cell cultures (a and b) were cotransfected with helper HSV-1 Justin DNA and the control or test constructs. Virus stocks derived from these transfections were serially passaged, and ^{32}P-labeled DNAs from cells infected with passage 2 of the resultant series were analyzed by digestion with *Eco*RI (Fig. 2) and additional enzymes (not shown) to test for the presence of defective genomes generated from the seed constructs. As seen in the maps shown in Fig. 3 (e.g., for the D-1 and G-1 constructs), two fragments were expected to arise by *Eco*RI digestion of concatemers generated by replication of the seed constructs: (i) a fragment of 1.90 kb, which contains the replication origin and was common to all constructs, and (ii) a variable-sized fragment containing the respective test *a* sequences.

The analyses revealed the following. (i) As expected, the control construct A-1 gave rise to abundant defective genomes composed of head-to-tail repeats of identical size to that of the input seed construct (represented by the 1.90- and 2.97-kb fragments; Fig. 2, lane 2). A minor proportion of the repeat units in the generated defective genomes contained tandemly reiterated *a* sequences, apparent as the "ladder bands" with increments of ca. 260 bp, extending from the 2.97-kb *Eco*RI band. (ii) Virus stocks derived from transfections receiving the constructs B-1, C-1, E-1, H-1, and I-1 did not contain authentically propagated defective genomes (Fig. 2, lanes 3 to 12). All of these constructs contained deletions affecting the Uc element, with the smallest deletion present in the construct H-1 (40-bp deletion bounded by the *Bss*HII sites in Uc). (iii) Virus stocks derived from cotransfections receiving D-1 and G-1 contained defective genomes with repeat units identical to the input seed constructs (yielding the 1.90-kb origin fragment and the fragments designated *a** in Fig. 2, lanes 13 through 16). However, the abundance of generated defective genomes in the population was relatively low compared with their counterparts generated from the control construct A-1 (compare the relative intensities of defective and helper virus fragments in the short autoradiographic exposures shown in Fig. 2, lanes 17 and 18). In addition, the defective genomes contained significant proportions of repeats that were ca. 240 bp larger than the input seed plasmids (the 3.37-kb fragment in lanes 13 and 14 and the 3.03-kb fragment in lanes 15 and 16). As shown below, the increase in size resulted from the insertion of an intact *a* sequence, derived from the helper virus DNA, within the repeat unit junctions. Consequently, these bands have been designated *a**-*a* in Fig. 2.

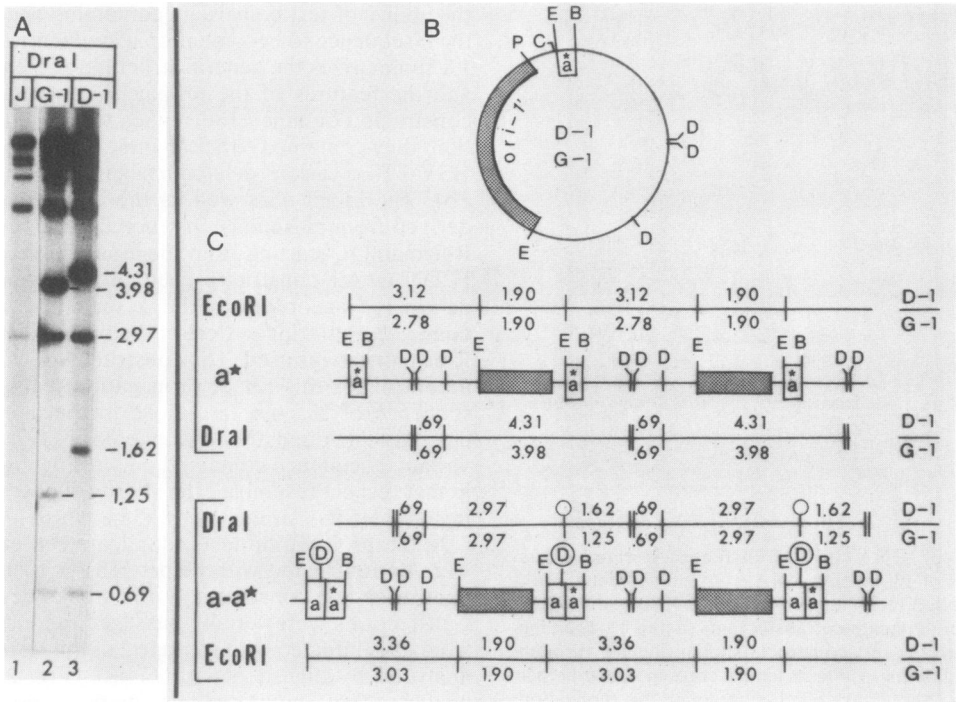


FIG. 3. Structure of chimeric junctions. (A) *Dra*I-digested ^{32}P-labeled DNAs from cells infected with passages 2 of transfections receiving HSV-1 (Justin) helper DNA alone (J) or the constructs G-1 and D-1. The DNAs correspond to those shown in Fig. 2, lanes 1, 16, and 14, respectively. (B) structure of the two test constructs. Shown are locations of the ori-1′ fragment (stippled area), the test *a* segment (a*) and the sites for *Eco*RI (E), *Bam*HI (B), *Dra*I (D), *Cla*I (C), and *Pvu*II (P). (C) predicted structures of generated defective genomes with repeat units containing the original deleted *a* sequence (a*) or the chimeric repeats (a-a*) with inserted helper virus *a* sequence. Sizes of generated *Eco*RI and *Dra*I fragments are given in kilobases for the D-1-generated (above the line) and G-1-generated (below the line) concatemers. The *Dra*I site present in the Ub element of the inserted helper virus *a* sequence is circled.

Two conclusions can be reached on the basis of these observations. First, the failure of the constructs lacking the Uc element (Uc^-) to be propagated in virus stocks revealed that the sequences bounded by the *Bss*HII sites within the Uc element were essential for cleavage-packaging. These boundaries are based on the 40-bp deletion in the H-1 construct. Second, the low-efficiency propagation of D-1 (containing a 64-bp deletion in Ub) suggested that the sequences bounded by the *Apa*I sites in Ub affected the efficiency of amplicon seed propagation. In addition to testing the constructs above, we tested (in duplicate sets) the corresponding *a* deletions in plasmids which contained the *Eco*RI 1.90-kb fragment (carrying the replication origin) in the opposite orientation to that in Fig. 1A. The results of these studies (data not shown) yielded the same conclusions as above.

Structure of the larger repeat units derived from the Ub^--Uc^+ constructs. The experiments described in this and the next sections revealed that the larger repeats found in the D-1- and G-1-derived defective genomes (Fig. 2, lanes 13 through 16; band a-a*) contained an insertion of an *a* sequence derived from the helper virus DNA. Specifically, because the increased size of the repeat units reflected changes in the fragment carrying the *a* sequence (Fig. 2; data not shown), the larger repeats could be derived by tandem reiterations of the deleted *a* sequence or, alternatively, by insertion of an *a* sequence derived from the helper virus DNA. To test these alternatives we took advantage of the fact that the input test constructs did not contain the Ub sequence and therefore lacked the *Dra*I site (TTTAAA) at position 31 of the *a* sequence. The *Dra*I enzyme cleaves in three locations within the bacterial plasmid portion of G-1 and D-1, generating common fragments of 19 and 692 bp, and the variable (*a*-containing) fragments of 3.98 and 4.31 kb (Fig. 3). Self-amplification of the test *a* sequences in G-1 and D-1 was predicted to result in increased size of the variable bands. In contrast, insertion of the *a* sequence from the Justin helper DNA should result in cleavage of these fragments, owing to the presence of a *Dra*I site in the inserted Ub sequence. In this case, two new fragments would be generated, with sizes dependent on the extract site of insertion. The *Dra*I cleavage patterns (Fig. 3) revealed that the novel repeats indeed contained an additional *Dra*I site, consistent with an insertion of an intact (Ub^+) *a* sequence derived from the helper virus DNA. In addition, the results showed that the helper virus *a* sequence was adjacent to the test (Ub-deleted) *a* sequence on the side proximal to Uc, creating a chimeric double *a* junction. Specifically, defective genomes generated from the constructs D-1 and G-1 yielded a common 0.69-kb *Dra*I fragment generated from the plasmid region, the *a*-containing *Dra*I bands of sizes 4.31 and 3.98 kb representing the input DNA (see map labeled a* in Fig. 3), and two novel fragments not present in the input seed, of sizes 2.97 and 1.62 kb for D-1, and 2.97 and 1.25 kb for G-1. The sizes of these novel fragments were consistent with the insertions of helper *a* (Ub^+ and hence *Dra*I containing) occurring on the side proximal to Uc (see map labeled a-a* in Fig. 3).

Nucleotide sequence of the novel chimeric junction. To further analyze the events which led to the formation of the chimeric junctions we shuttled the repeat units of the G-1-derived defective genomes back into bacteria and obtained the nucleotide sequence for three of the resultant clones. Specifically, unlabeled passage 2 DNA from a G-1-

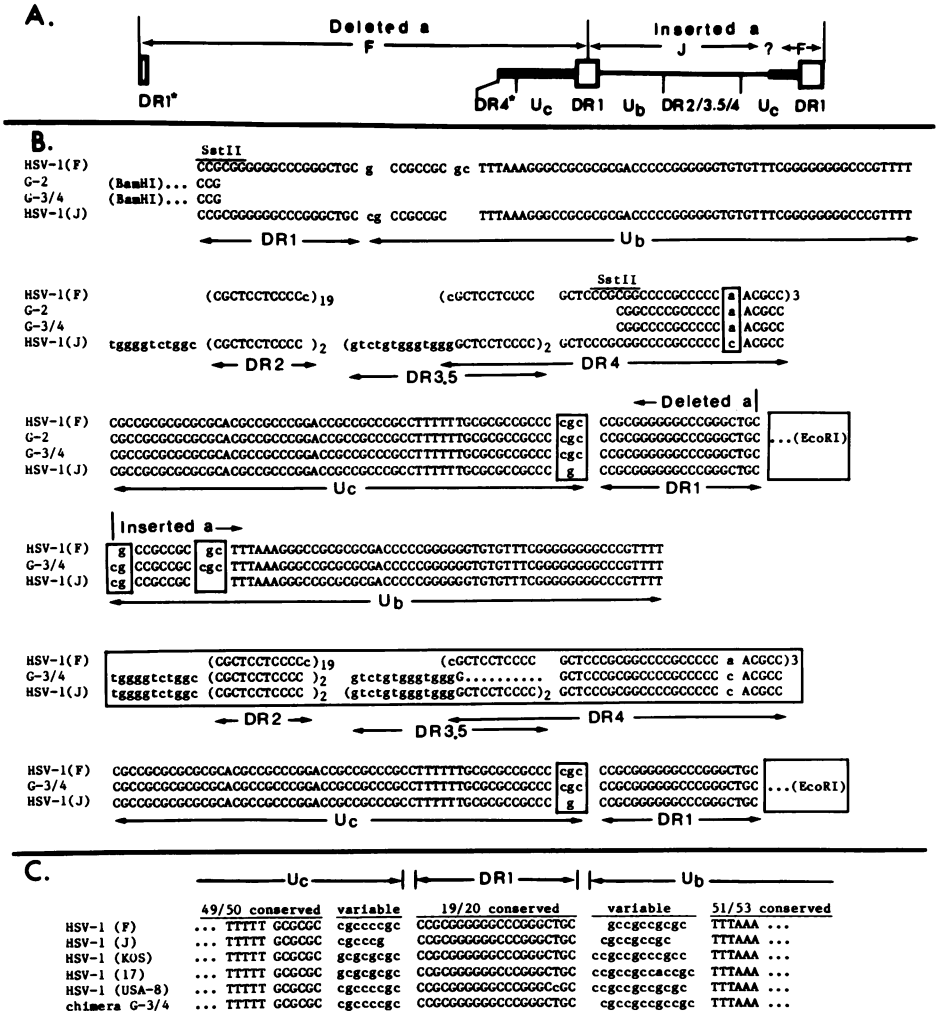


FIG. 4. Nucleotide sequences of the chimeric junctions and sequence conservation in the Ub and Uc elements of different HSV-1 strains. (A) General structure of the chimeric junctions in clones G-3 and G-4. The Ub^--Uc^+ *a* sequence from the construct (denoted as deleted *a*) and the inserted *a* sequence are arranged in tandem. The heavy line denotes sequences originating from F; thinner line denotes sequences originating from strain Justin (J). DR1* and DR4* indicate partial elements that are fused at the *Sst*II site as in the seed G-1 construct. ?, A region conserved between F and Justin; therefore, the site of transition from Justin to F sequences in Uc is unknown. (B) Nucleotide sequences of the *a* segments in the authentically sized (G-2), and larger (G-3 and G-4) repeats and the sequences of double *a* junctions of HSV-1 strains F (19, 21) and Justin (18). Sequences which differ in the Justin and F strains are boxed. The locations of the *Eco*RI and *Bam*HI sites in the polylinker are shown. The dotted line in the DR3.5 repeat indicates area of sequence uncertainty. (C) Sequence conservation in the Ub and Uc elements in different HSV-1 strains and in the G-3 and G-4 chimeric repeats. The regions proximal to the DR1 are variable, whereas those farther away are conserved. Nucleotide sequences are from Mocarski and Roizman (19, 21) for strain F, Mocarski et al. (18) for Justin (J), Varmuza and Smiley (35) for KOS, and Davison and Wilkie (6) for strains 17 and USA-8.

propagated virus series was cleaved with *Cla*I which cleaves once within each repeat unit (site denoted by C in map [Fig. 3, top]). The cleaved fragments were resolved by gel electrophoresis, and monomeric fragments corresponding to the authentically sized repeats and the larger (chimeric) repeats were eluted from the gel, circularized, and used to transform bacteria. We then determined the nucleotide sequences of the *a* sequences in one of the authentically sized clones (designated G-2) and in two of the larger chimeric clones (designated G-3 and G-4). The results are shown in Fig. 4 along with the corresponding DNA sequences of HSV-1 strain F (21), from which the deleted *a* construct was derived, and HSV-1 strain Justin (18), which served as helper virus. The *a* sequence in G-2 was found to be identical to the deleted *a* sequence in the seed HSV-1 (F)-derived G-1 clone (4). Thus, it contained the fused (at the *Sst*II site) DR1 and DR4 elements and the entire Uc element followed by the DR1 and linker sequences. The nucleotide sequences of the chimeric double *a* junctions in the two larger cloned repeat units G-3 and G-4 were identical and contained several interesting features. (i) Starting from the DR1-DR4 fusion site (*Sst*II), the sequences were identical to those in the rescued, authentically sized clone (G-2) up to the DR1 junction. This included the DR4 and Uc elements in which the Justin strain differs from the F strain by 1 and 2 nucleotides, respectively (Fig. 4B, the boxed A and C in DR4 and the boxed CGC and G in Uc). Thus, the insertion of the helper virus *a* sequence evidently occurred within the DR1 sequence. (ii) The inserted *a* sequence was generally similar to the *a* sequence of the Justin helper virus DNA. Thus, the chimeric junctions contained the Justin-specific sequence TGGGGTCTGGC between the Ub and DR2 ele-

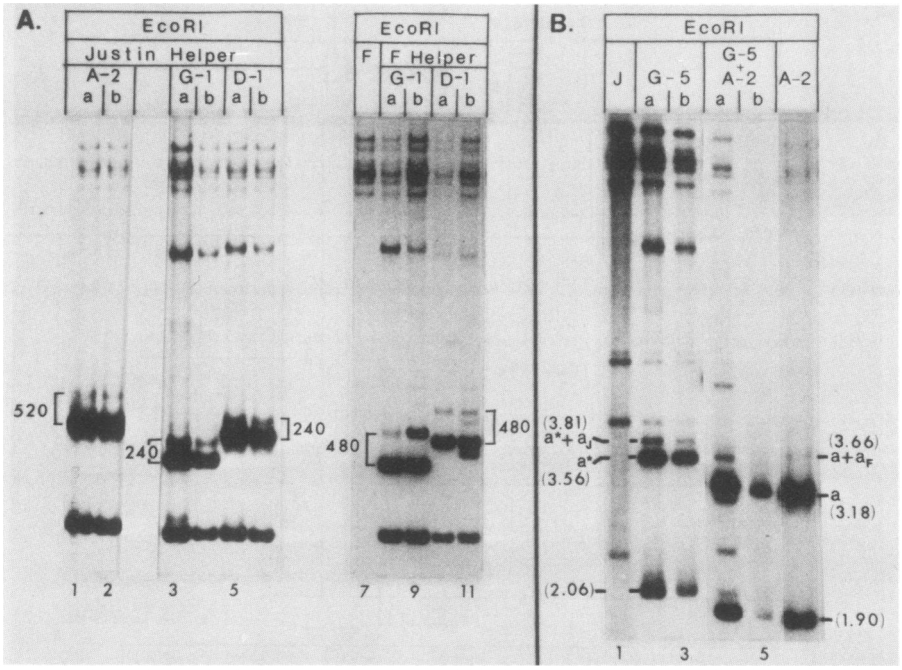


FIG. 5. Propagation of test constructs in the presence of Justin and F helper viruses and in transfection competition tests. (A) *Eco*RI digests of ^{32}P-labeled DNAs from duplicate transfections (lanes a and b) receiving the HSV-1 strains Justin (lanes 1 through 6) or F (lanes 8 through 11) helper virus DNA along with the indicated test constructs. Lane 7 contains the *Eco*RI-digested F helper DNA. The increments (in base pairs) between fragments containing single *a* and double *a* junctions are indicated. (B) Propagation competition tests. *Eco*RI digests of ^{32}P-labeled passage 2 DNAs from transfections receiving helper virus DNA alone (lane 1) or with the G-5 construct (lanes 2 and 3), the A-2 construct (lane 6), or an equal mixture of both constructs (lane 4 and 5). Fragments derived from the defective genomes are indicated by size. *a* and *a** denote fragments containing the intact and Ub$^-$-Uc$^+$ *a* sequences present in the seed A-2 and G-5 constructs, respectively. $a^* + a_J$ denotes the fragment containing the chimeric junction with helper Justin *a* sequence; $a + a_F$ denotes the double *a* junction arising by self-amplification of the intact *a* sequence in A-2.

ments, two copies of the DR2 repeat, the Justin-specific DR3.5 repeat, and a single copy of the DR4 repeat (these elements are boxed in Fig. 4B). (iii) two sequence alterations appeared in both chimeric repeat clones. First, 10 bp into the Ub element, the chimeric junctions in G-3 and G-4 contained the triplet CGC (boxed in Fig. 4B) which was absent from the Justin helper virus *a* sequence. Second, the distal portion of the Uc element appeared to be derived from F (construct) rather than Justin (helper) DNA (Fig. 4A), inasmuch as at the Uc-DR1 transition, the inserted *a* sequences resembled the sequance of the F rather than the Justin strain (Fig. 4B, boxed CGC and G).

Two points are noteworthy concerning this result. First, it is likely that the sequence variations reflect features of the recombination events leading to the formation of the chimeric junctions, rather than sequence variation related to propagation in virus stocks. Thus, the nucleotide sequence of the *a* sequence in the G-2 repeat (cloned following propagation) was identical to that in the input *a* construct (G-1), whereas both chimeric *a* clones exhibited the same sequence alterations. Second, the sequence alterations were of interest in that, in the HSV-1 strains sequenced thus far, the sequences immediately adjacent to the DR1 (at the Uc-DR1-Ub transitions) are quite variable, whereas the bulk of the sequences making up the Uc and Ub elements are highly conserved (Fig. 4C). It is possible that the variability in the Uc-DR1 and DR1-Ub transitions within the chimeric double *a* junction reflected inherent variability in these regions. A possible explanation for the source of sequence variations is proposed in the Discussion.

Essential role of Ub in cleavage-packaging. Insertions of helper virus DNA sequences into defective genomes had not been detected in our previous studies that involved the use of amplicons with heterologous helper viruses (F strain-derived amplicons in combination with Justin helper, and vice versa; see, e.g., reference 7). A possible explanation for the observed insertions into the D-1- and G-1-derived defective genomes was that the Ub deletion in these constructs affected the cleavage-packaging signal and therefore chimeric junctions which restored the deleted Ub sequences in an appropriate arrangement were selected for propagation. Three predictions of this hypothesis were tested below and found to be correct. The first was that helper virus DNA insertions would not be detected in defective genomes propagated from constructs containing an intact *a* sequence. The second was that in competition tests, the Ub$^-$-Uc$^+$ test plasmids will be propagated with lower efficiency than the intact *a* constructs. The third was that the Ub$^-$-Uc$^+$ repeat units will be propagated from passage to passage within heteroconcatemers, i.e., mixed concatemers containing (i) repeats with the Ub$^-$ *a* sequence and (ii) repeats with the chimeric (helper *a*-containing) junctions. The cleavage-packaging functions of these heteroconcatemers could be provided, at least in part, by the chimeric junctions.

Transfer of the helper *a* sequence into different seed constructs. To test the first prediction we analyzed the pattern of *a* amplification in defective genomes derived by propagation of the intact *a* construct A-2 and the Ub$^-$ constructs D-1 and G-1 (Fig. 1B). Each of these constructs was used to cotransfect duplicate cultures in the presence of HSV-1 (F) or HSV-1 (Justin) helper DNAs (possessing *a* sequences of 501 and 264 bp, respectively) to determine whether the

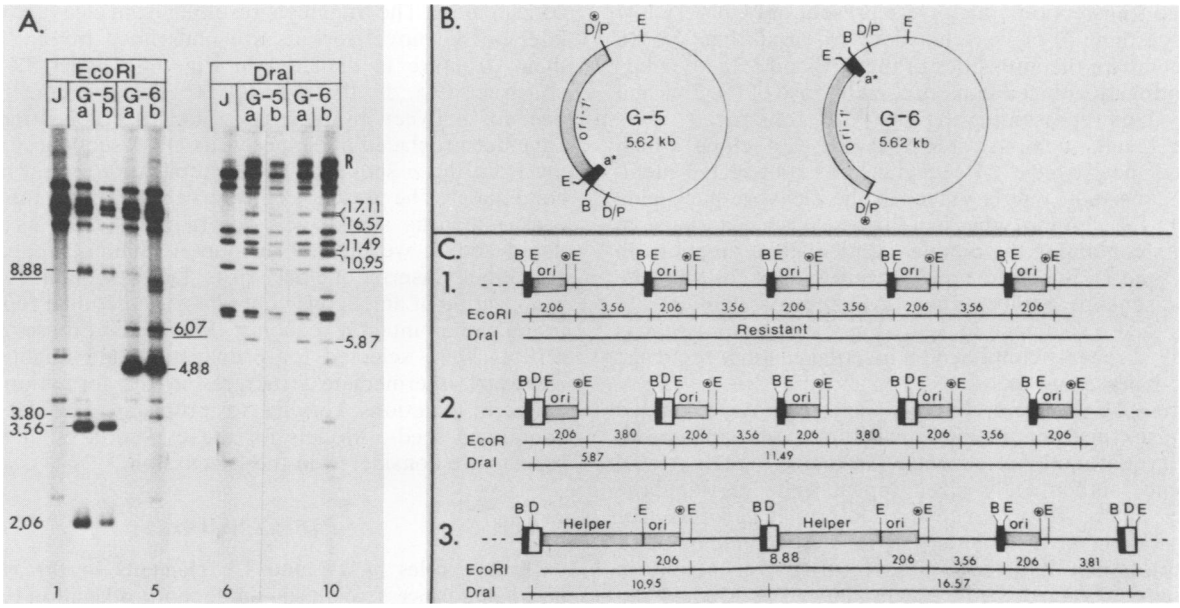


FIG. 6. Structure of concatemers arising by propagation of the G-5 and G-6 test constructs. (A) *Eco*RI and *Dra*I digests of ^{32}P-labeled passage 2 DNAs from transfections receiving helper Justin DNA alone or with the G-5 or G-6 test constructs. Sizes are indicated in kilobases. The underlined sizes correspond to fragments which arise from the large inserts of helper virus DNA. The corresponding maps for the G-5-derived genomes are shown in panel C3 and account for the novel 8.88-kb *Eco*RI band and the 10.95 and 16.57-kb *Dra*I bands. The construct G-6 also produces large recombinant repeat units with expected 6.07-kb novel *Eco*RI fragment and 10.95- and 16.57-kb novel *Dra*I fragments (maps not shown). (B) Structure of the constructs G-5 and G-6. The *Eco*RI-to-*Pvu*II segment contains ori-1′; *a** denotes the Ub$^-$-Uc$^+$ *a* segment. The locations of the *Eco*RI (E), *Bam*HI (B), and the fused *Dra*I-*Pvu*II (D/P) sites are shown. (C) Concatemeric structures for the G-5 construct. C1, Homoconcatemers containing authentically sized (input) repeat units. ⊛, *Dra*I-*Pvu*II fused site at the border of the ori-1′ fragment. C2, Structure of heteroconcatemers containing interspersed chimeric and authentic junctions. Black rectangles denote the Ub$^-$-Uc$^+$ *a* sequence (*a**), and open rectangles denote the inserted helper virus *a* sequence. Cleavage of the heteroconcatemers with *Dra*I is expected to produce bands with increments of authentically sized repeats. The first two bands (5.87 and 11.49 kb) are shown both here and in panel A, lanes 7 through 10. C3, Heteroconcatemers containing repeat units with an inserted helper DNA segment spanning the region from the *a* sequence to the replication origin ori-1′. The predicted sizes of fragments generated from *Eco*RI and *Dra*I cleavages of the heteroconcatemers are shown for the construct G-5. The corresponding first two bands in the predicted ladder generated by *Dra*I cleavage (10.95 and 16.57 kb) are underlined in panel A, lanes 7 through 10, as is the novel 8.88-kb *Eco*RI fragment.

amplified *a* sequences were derived from the helper virus DNA or from the seed test plasmid. Restriction enzyme analyses of passage 2 DNAs from these transfections (Fig. 5A for *Eco*RI; additional data not shown) revealed the following. (i) The increment between the bands representing the double *a* and single *a* junctions in defective genomes propagated from A-2 (intact *a*) in the presence of the Justin helper virus was ca. 520 bp (fig. 5A, lanes 1, 2). Thus, the double *a* junction was derived by self-amplification of the 537-bp *a* sequence of the seed construct (minus the shared DR1 element) rather than by insertion of the 264-bp helper virus *a* sequence. (ii) In contrast, the increments between the double *a* and single *a* junctions in defective genomes propagated from the D-1 and G-1 constructs varied with the helper virus. Specifically, in the presence of the Justin helper (possessing a 264-bp *a* sequence) the increment corresponded to ca. 240 bp (Fig. 5, lanes 3 through 6), whereas in the presence of the F helper (501-bp *a* sequence) it was ca. 480 bp (lanes 8 through 11). These values were consistent with insertions of the helper *a* sequences into the chimeric junctions. Furthermore, bands arising from self-amplification of the Ub$^-$ *a* sequences (100 bp in G-1 and 470 bp in D-1) were not detected.

The data support three conclusions. (i) The intact *a* sequence contains a complete set of *cis*-acting signals needed for the cleavage-packaging of viral DNA, as well as for the self-amplification of *a* sequences. (ii) For defective genomes containing intact *a* sequence, the *trans* insertion of *a* sequences from helper virus DNA is less efficient than self-amplification of the *a* sequences. (iii) the Ub deletion affects function(s) required for *a* amplification, as well as for defective genome propagation. Hence, defective genomes derived from these constructs require the presence of at least some chimeric junctions (in the same concatemer, see below) for their propagation.

Competition propagation tests. Studies described below revealed that constructs containing the Ub$^-$-Uc$^+$ *a* sequences were propagated in virus stocks less efficiently than their counterparts containing the intact *a* sequence. Specifically, duplicate cultures were cotransfected with helper Justin virus DNA and (i) the A-2 intact *a* construct, (ii) the construct G-5, which contained a Ub$^-$-Uc$^+$ *a* sequence identical to that in G-1 (map in Fig. 6), or (iii) an equal mixture of both plasmids. The relative propagation efficiency of A-2 and G-5 could be determined from the ratios of the distinct restriction enzyme fragments which they were expected to yield.

Figure 5B shows the *Eco*RI digests of passage 2 DNAs from these transfections. Defective genomes with repeat units derived from the intact *a* construct were abundant in stocks derived from the single (A-2; lane 6) and mixed (A-2 + G-5, lanes 4 and 5) transfections (see bands 1.9 and 3.18 kb in these lanes). Defective genomes derived from the Ub$^-$ *a* construct were propagated in stocks derived from the transfections receiving the G-5 plasmid singly (lanes 2 and 3; bands 2.06 and 3.56 kb). However, in stocks derived from

the mixed transfections, they were present only in very low abundance (lane 4) or were not detected at all (lane 5). (In lane 4, compare the intensities of the 1.90 and 3.18 kb bands representing the intact *a* construct with those of the 2.06 and 3.56 kb bands representing the Ub^--Uc^+ construct.)

These results demonstrated that the G-5 clone which contained the Ub^--Uc^+ *a* sequence was propagated inefficiently, consistent with a defect in the cleavage-packaging signal(s). In addition, the increments observed between fragments containing the double *a* and single *a* junctions in the A-2- and G-5-derived repeats were consistent with the results discussed earlier. Thus, A-2 repeats contained a self-amplified *a* sequence ($a + a_F$; Fig. 5B, lane 6), whereas G-5-derived repeats contained an inserted Justin *a* sequence ($a^* + a_J$; lanes 2 and 3).

Structure of concatemeric DNA molecules. As discussed above, test constructs lacking the Ub element appeared to contain a nonfunctional cleavage-packaging signal. At the same time, authentically sized repeat units carrying the Ub^--Uc^+ junctions were recovered in propagated defective genomes (along with chimeric junctions containing helper *a* sequence). In fact, they represented the major repeat units in the serially passaged virus populations. The studies described below provide an explanation for this puzzling observation in that they revealed that repeat units containing the Ub^--Uc^+ junctions were propagated as constituents of heteroconcatemers containing interspersed chimeric junctions. Specifically, in these studies the Ub^--Uc^+ *a* sequence was first introduced into plasmids lacking *Dra*I sites. Since the *Dra*I site in the Ub element could serve as the hallmark for the inserted helper virus *a* sequences, the degree of interspersion of chimeric and authentic repeat units within the same defective genome molecules could be determined by analyzing the distribution of *Dra*I sites along defective genome concatemers.

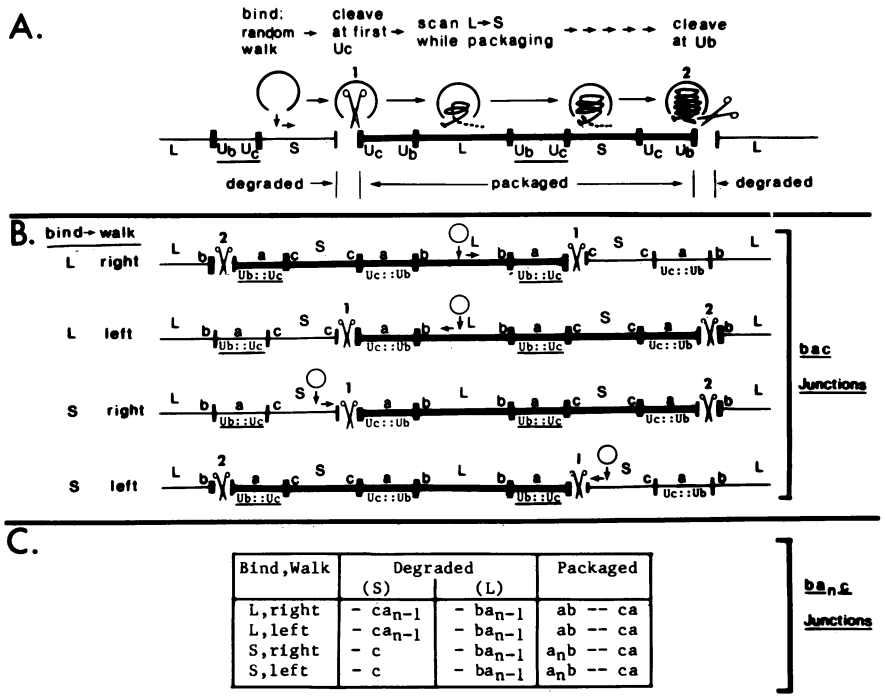
For the construction of the test clones lacking *Dra*I, a *Pvu*II fragment containing the replication origin and the Ub^--Uc^+ *a* sequence was introduced into a *Dra*I-cleaved pBR322, resulting in the inactivation of both sites and yielding the plasmids G-5 and G-6, differing in the orientation of the *Pvu*II insert (Fig. 6B, top). The predicted *Eco*RI and *Dra*I cleavages of homoconcatemers and (the simplest) heteroconcatemers are shown for the construct G-5 in Fig. 6C, panels 1 and 2. Using the G-5 and G-6 test constructs as input seed clones, we expected that homoconcatemers containing solely the Ub^- *a* sequences would be resistant to *Dra*I (Fig. 6C, panel 1). In contrast, random interspersion of the chimeric junctions (each containing a *Dra*I site in Ub) within heteroconcatemers containing the authentic repeat units (devoid of the *Dra*I site) would create a ladder of *Dra*I fragments with increments of single, authentically sized repeats. Because the authentically sized repeats were 5.62 kb and because the repeats with chimeric double *a* were 5.87 kb, we predict that *Dra*I cleavage of the heteroconcatemers will generate a ladder described by 5.87 kb + *n* (5.62 kb), where *n* is an integer, i.e., 5.87, 11.49, 17.11 kb, etc. (Fig. 6C; panel 2).

The *Eco*RI and *Dra*I digestion patterns of passage 2 DNAs from duplicate transfections receiving HSV-1 (Justin) helper along with the G-5 and G-6 constructs are shown in Fig. 6A. The *Dra*I cleavage pattern of the G-5 and G-6 defective genomes contained bands of sizes predicted from random interspersion of chimeric junctions within the concatemeric defective genomes (Fig. 6A). Interestingly, the *Dra*I and *Eco*RI patterns revealed the presence of an additional type of chimeric repeat unit within the same defective genome concatemers. The fragments resulting from cleavage of these interspersed novel repeats are underlined in Fig. 6A and their structure is depicted in Fig. 6C, panel 3, for the construct G-5. In these novel repeats, the construct sequences between the deleted *a* and the replication origin have been replaced by helper virus DNA sequences extending from the *a* sequence to the replication origin in the S component. The presence of these repeat units in the serially passaged virus stocks was surprising in that, as already noted above, we did not encounter chimeric repeat units containing inserted helper virus DNA sequences during propagation of amplicons containing a functional replication origin and an intact *a* sequence. Most likely, these chimeric repeats were selected for propagation and might in fact represent intermediate structures in the formation of the chimeric junctions serving to propagate the Ub^--Uc^+-containing seeds. Models for the evolution of these large repeats are considered in the Discussion.

DISCUSSION

Distinct roles of Uc and Ub elements in the cleavage-packaging process. We used amplicon propagation tests (28) to map the signals required for the cleavage and packaging of HSV DNA. The constructs tested in these studies fell into three distinct categories. Constructs in the first category (A-1 and A-2) contained an intact *a* sequence and yielded defective genomes which did not contain additional sequences derived from the helper virus genome. Constructs in the second category (D-1, G-1, G-5, and G-6) contained deletions in the Ub element but retained the Uc element. Defective genomes with authentically sized repeats were recovered from these constructs. However, propagation resulted in the selection of chimeric repeat units containing intact *a* sequences derived from helper virus DNA. Moreover, in transfection competition tests the Ub^--Uc^+ defective genomes did not compete efficiently with their counterparts containing an intact *a* sequence. Finally, test constructs in the third category (B-1, C-1, E-1, and H-1) contained deletions in the Uc element and could not be propagated into defective genomes. Taken together, these data revealed that the Ub and Uc elements played critical yet distinct roles in the cleavage-packaging process. The involvement of the DR2 and DR4 elements in the cleavage-packaging process is at present unknown.

Directional packaging model for cleavage-packaging of viral DNA. In this and the next section we propose two alternative models for the cleavage-packaging of viral DNA. Both models are designed to explain the data summarized in this paper and to account for previous observations related to the cleavage-packaging process. The first model (the directional cleavage model) modifies the "theft" model proposed recently by Varmuza and Smiley (35) to include polarity in the cleavage-packaging process. The model is schematically diagrammed in Fig. 7A for standard virus DNA concatemers and includes the following steps. (i) A packaging complex encounters the HSV DNA anywhere in the L or S component and traverses along the concatemeric DNA in either direction (random walk; Fig. 7A) until a junction containing a Uc signal is found. (ii) Cleavage (producing 3′ single-base overhangs) occurs at the DR1 element proximal to the first Uc signal encountered (cleavage 1; Fig. 7). If the complex traverses into a ba_nc type junction starting from the L component (ca. 80% of the cases, based on the relative sizes of L and S, and assuming random initial binding) this first cleavage will produce *ba* and ca_{n-1} termini (Fig. 7B and C).



C.

Bind,Walk	Degraded (S)	Degraded (L)	Packaged
L,right	- ca_{n-1}	- ba_{n-1}	ab -- ca
L,left	- ca_{n-1}	- ba_{n-1}	ab -- ca
S,right	- c	- ba_{n-1}	a_nb -- ca
S,left	- c	- ba_{n-1}	a_nb -- ca

ba_nc Junctions

FIG. 7. Directional cleavage model. (A) Basic features of the model. The packaging complex (open circle) binds to the standard viral DNA concatemer at a random location and traverses the concatemer in a random walk (right or left in the figure). The first cleavage [1] occurs at a constant distance from the first Uc signal encountered, leaving a 3′ single-base overhang. In the standard viral genome this cleavage occurs in the DR1 element. Packaging begins at the generated L terminus (which contains at least one *a* sequence), whereas the generated S terminus (which is devoid of *a* sequences) is rapidly degraded. While packaging in the L to S direction, the DNA is scanned for the next directly repeated junction, skipping the inverted junction (underlined). A second cleavage [2] occurs at a constant distance from the first Ub sequence encountered in the second junction, leaving 3′ single-base overhangs and producing a packaged molecule (heavy line) which carries a *ca* terminus, and an unpackaged terminus which is degraded. (B) Diagram of the cleavage within single *a* junctions in a standard virus DNA concatemer, featuring the four possible permutations of initial binding and random walk. Assuming random binding, ca. 80% of the initial binding events should occur in L (lines 1 and 2) and ca. 20% should occur in S (lines 3 and 4). (C) Termini produced in packaged and free molecules as a consequence of the cleavage-packaging process starting from each of the four initial binding permutations. n represents the number of *a* sequences in the junctions which are being cleaved.

If the process began within the S component, the first cleavage will generate ba_n and *c* termini. In both cases directional packaging will begin from the generated L termini (*ba* or ba_n), whereas the generated S termini (ca_{n-1} or *c*) will be rapidly degraded. (iii) Packaging will continue by scanning in the L-to-S direction, until a directly repeated junction is encountered. (iv) A second cleavage (also yielding 3′ single-base overhangs) will then occur proximal to the first Ub signal encountered within that junction. This cleavage is expected to produce a *ca* terminus on the packaged molecule and a free (ba_{n-1}) terminus which might be rapidly degraded or form the substrate for a new round of packaging (into a separate capsid).

The directional cleavage model can account for several previously unexplained observations. First, the proposed scanning for a second directly repeated junction (step 3) accounts for the observed cleavage at alternate (i.e., directly repeated) junctions within standard virus concatemers. Furthermore, the polarity requirement for the two prospective termini has also been demonstrated in studies with constructed defective genomes, supporting a scanning mechanism (9; L. P. Deiss and N. Frenkel, manuscript in preparation). Second, the proposed specificities of junction cleavages and termini degradation account for the production of two *a*-containing termini (with 3′ overhangs [20]) from single *a*-containing junctions (7). Furthermore, the model accounts for the presence of a variable number of *a* sequences at the L terminus, and a single *a* sequence at the S terminus. Third, the predicted generation of *ba* termini from ba_nc junctions in the majority (those resulting with first encounter in L [ca. 80%]) of first cleavages might account for the observed predominance of L termini carrying a single *a* sequence (15). Moreover, it might constitute the mechanism for the diminution of *a* sequences from multiple *a*-containing junctions arising as consequence of genome circularization, following virus entry into the cells (21, 22). Finally, the proposed roles for the Ub and Uc elements in cleavage-packaging might account for the distinct propagation patterns of amplicon constructs tested in the present study. Specifically (Fig. 8 a through e), the model predicts that defective genome concatemers arising by replication of Ub^--Uc^+ constructs (Fig. 8a) could nevertheless undergo the first cleavage at the Uc signal, unlike their Uc^- counterparts. The resultant terminus, which carries the Ub^- *a* sequence, might undergo an aberrant fusion with the S terminus of the helper virus DNA (Fig. 8b), yielding chimeric junctions, after rounds of homologous recombinations and replication (Fig. 8 c through e). Finally, the resultant concatemers might serve as substrates for the normal cleavage packaging process, with the first (Uc proximal) cleavage occurring within a Ub^--Uc^+ junction, followed by the second (Ub proximal) cleavage within a chimeric junction present in the same heteroconcatemer.

While the directional packaging model may account for

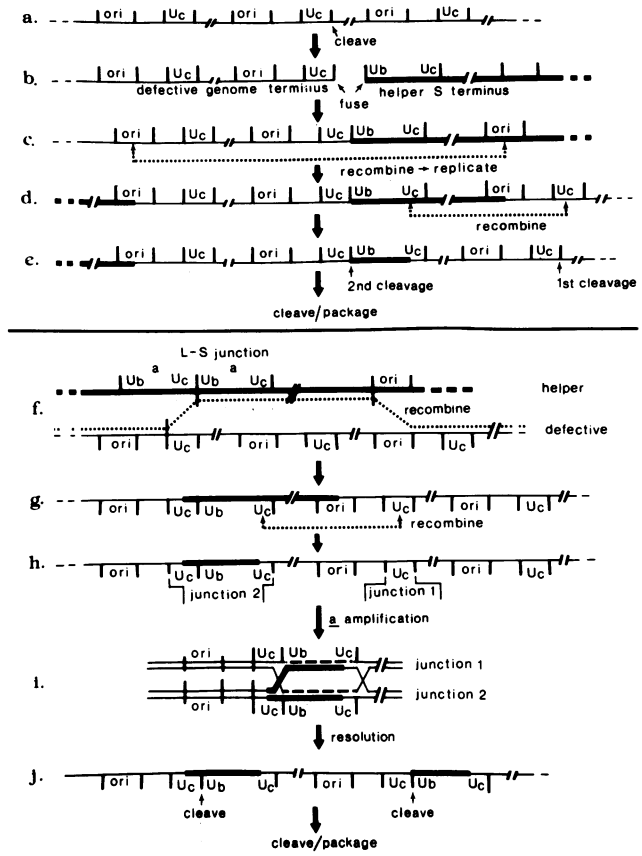
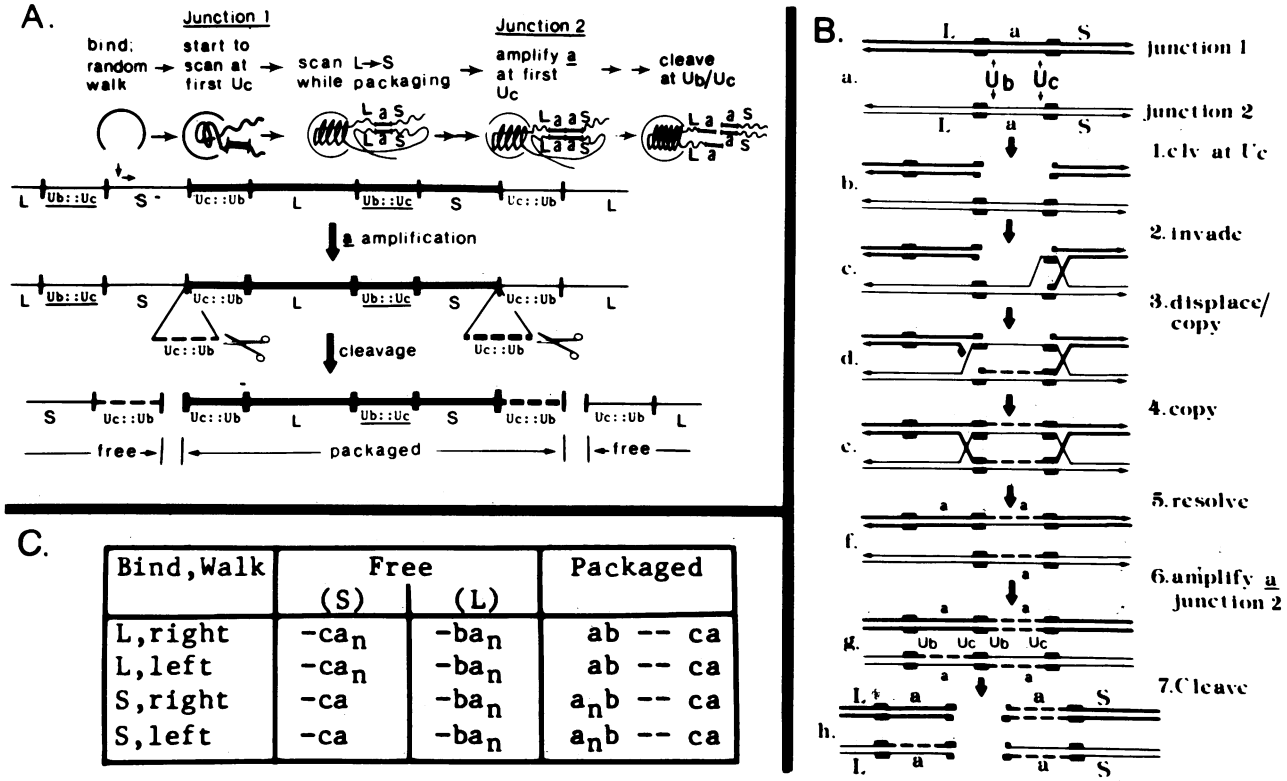


FIG. 8. Generation and cleavage of the heteroconcatemers from Ub$^-$-Uc$^+$ constructs. (a) through (e) Chimeric junctions are generated and cleaved by the directional cleavage mechanism. Concatemers arising by replication of the Ub$^-$-Uc$^+$ seed constructs (thin line in panel a) are cleaved at the DR1 proximal to the Uc signal by the first cleavage step. The generated terminus, carrying Uc, fuses with the S terminus of standard virus DNA (thick line in panel b). Homologous recombination at the ori-1 region followed by rolling-circle replication of the resulting circular repeat generates heteroconcatemers (panel d). Homologous recombination in the Uc elements within a large chimeric repeat and an authentic repeat along the same concatemer results in the formation of chimeric junctions (panel e) containing the inserted helper *a* sequence, with the last portions of the Uc sequence derived from the seed construct. Cleavage of the heteroconcatemers occurs by the directional cleavage mechanism, the first cleavage occurring within the deleted *a* junction (at the Uc signal) and the second cleavage occurring in the chimeric junction (at the Ub signal). The model predicts that concatemers derived by replication of the Uc$^-$ constructs will not be cleaved in step (a). (f) through (j) Generation and cleavage of the chimeric repeats by the double-stranded break-gap repair model. Homologous double crossover recombination occurs between helper virus DNA with a double *a* L-S junction (heavy line) and a defective viral genome (thin line) which was generated by replication of the input Ub$^-$-Uc$^+$ construct (panel f). The crossover points are in the homologous Uc and ori-1′ regions, yielding the heteroconcatemer shown (panel g). A chimeric junction arises by homologous recombination between the Uc sequences in an authentic repeat and the Uc segment in the inserted helper segment. The chimeric junction contains the deleted *a* and the intact inserted *a* carrying a Uc element which is derived in part from the construct (panel h). The heteroconcatemer is a substrate for the cleavage-packaging process, occurring by the double-stranded break-gap repair mechanism depicted in Fig. 9. In the *a* amplification process the Ub$^-$-Uc$^+$ junctions is used as junction 1 and the chimeric junction is used as junction 2. The sequence variability in Ub arises during the resolution of the Holliday (11) structures. This model predicts that constructs containing the Uc$^+$ repeats will form heteroconcatemers which will not be cleaved and packaged owing to the lack of Uc element in junction 1.

many of the previous observations, it suffers from two drawbacks. First, as recognized by Varmuza and Smiley (35), the model predicts the formation of termini devoid of *a* sequences. Such termini have not been observed in total nuclear DNA (7, 20). Therefore, it must be assumed that they are rapidly degraded. Second, although the model explains some of the structural features of the chimeric junctions, it fails to explain the observed variations in the Ub-DR1 portion of the inserted *a* sequence. It appears unlikely that such variations could result from the fusion of the defective genome Uc terminus with the S terminus of standard virus DNA.

Double-strand break and gap repair model for a amplification. The second model proposed for the cleavage-packaging process is based on the double-strand break and gap repair mechanism proposed by Szostak et al. (33) to explain recombination events resulting in gene conversion. Specifically, the model involves the interaction of two directly repeated junctions (Fig. 9; junctions 1 and 2), resulting in *a* amplification by a gene conversion-like mechanism. The resultant junctions are then cleaved to yield the genomic termini. Cleavage-packaging thus includes the following steps. (i) As in the directional cleavage model, the packaging complex traverses along concatemeric viral DNA, starting from any point in the S or L components, until the first Uc signal is encountered in junction 1 (Fig. 9A). (ii) Scanning-packaging begins, proceeding in the L-to-S direction, until the directly repeated junction 2, along the same concatemer, is encountered. The two junctions are juxtaposed for the next series of events (depicted schematically in Fig. 9B). (iii) A Uc signal-directed double-stranded cleavage occurs within the DR1 element of either junction, e.g., in junction 1, (Fig. 9B, step b). (iv) The resultant 3′-terminated strand invades the homologous sequence in junction 2 and is extended by copying the *a* sequence in that junction, while displacing the equivalent strand (steps c and d). (v) The displaced strand from junction 2 serves as the template for repair synthesis of the second strand of junction 1 (step e). (vi) The process terminates by the resolution of the two Holliday structures (11); this is predicted to occur at or near the DR1 sequences (step f). (vii) If junction 2 contained a single *a* sequence (with Uc$^+$ element), it too will become a recipient for an *a* sequence from junction 1 in a similar sequence of events as above. The entire process thus results in the amplification of the *a* sequence within each of the two junctions (step g). Each of the double *a* junctions now contains a DR1 sequence flanked by the Uc and Ub signals. (viii) Following the *a* amplification, both double *a* junctions are cleaved (generating a 3′ single-base overhang) between the newly inserted and the original *a* sequence (step h). This cleavage may require a structure in which the DR1 element is flanked by Ub and Uc signals, an arrangement which is present only in the double *a* junctions. The model predicts that ca. 80% of the packaged genomes (initial encounter in L) will carry *ba* and *ca* termini, regardless of the number of *a* sequences in the cleaved junctions, whereas the remaining packaged genomes (initial binding in S) will be terminated with *ca* and ba_n sequences (Fig. 9C). The termini of the generated free molecules (predicted to be of the type ba_n and ca_n or *ca*; Fig. 9C) might either become degraded or serve as target for new packaging cycles.



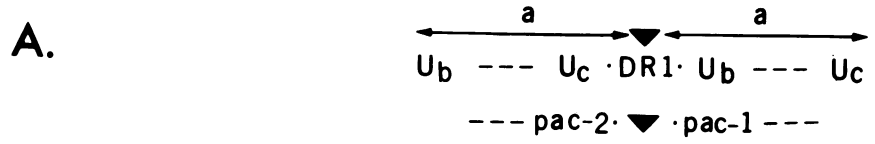
C.

Bind, Walk	Free (S)	Free (L)	Packaged
L, right	$-ca_n$	$-ba_n$	ab -- ca
L, left	$-ca_n$	$-ba_n$	ab -- ca
S, right	-ca	$-ba_n$	a_nb -- ca
S, left	-ca	$-ba_n$	a_nb -- ca

FIG. 9. The double-stranded break-gap repair model for *a* amplification. (A) Basic features of the model are shown for two single *a* junctions. Following random binding and random walk, packaging begins at the first Uc sequence encountered, proceeding in an L to S direction, until a directly repeated junction is encountered. The two junctions interact by the double-stranded break-gap repair mechanism (detailed in panel B), resulting in reciprocal copying of an *a* sequence from one junction to the other, the *a* insertions occurring at a the first Uc signal encountered in both junctions. The generated double *a* junctions are cleaved within the DR1 elements flanked by the Uc and Ub signals. Unpackaged molecules are either degraded or utilized in a new round of packaging. (B) Proposed steps in the double-strand break-gap repair model for *a* amplification, featured after the gene conversion model of Szostak et al. (33). (a) Two directly repeated single *a* junctions (junction 1, heavy line; junction 2, thin line) are juxtaposed. (b) A double-strand break occurs at the DR1 element proximal to the first Uc signal. (c) The resultant 3′ terminus of junction 1 invades junction 2, displacing the homologous strand. (d) The invading 3′ terminus has been extended (dotted line) by copying the complementary junction 2 strand; the displaced strand pairs in the DR1 region with the second cleaved strand of junction 1, forming a template for extension of this strand (dotted line). (e) The newly synthesized portions join the original corresponding strands, forming the Holliday structures (11). (f) Resolution has occurred within the DR1 elements or adjacent sequences. (g) A similar series of steps results in the amplification of the *a* sequence in junction 2. (h) Cleavage producing 3′ single-base overhangs has occurred within the DR1 elements, separating the new and original *a* sequences in each junction. This cleavage could be directed by the Ub and Uc signals which flank the cleavage site. Small arrows indicate the 3′ end of each strand. (C) Termini produced in packaged and free molecule as consequence of cleavage-packaging starting from each of the four initial binding permutations. n represents the number of *a* sequences in the junctions which are being cleaved.

Unlike the directional packaging model, the double-stranded gap repair model does not predict the generation of termini devoid of *a* sequences. Furthermore, multiple-*a*-containing S termini have been observed in stocks containing very high ratios of defective to helper virus genomes (7), supporting the hypothesis that they represent intermediary products of the packaging process. In addition, the double-stranded break-gap repair model accounts for the same set of observations explained by the directional cleavage model, including the cleavage of alternate junctions in standard virus DNA and the generation of packaged DNA molecules terminated with a single *a* sequence at the S end and with predominantly single *a* (but also multiple *a*) sequences at the L ends. Finally it also provides a scheme for the propagation of the Ub^--Uc^+ construct and explains the structural features of the resultant chimeric repeats (Fig. 8f through j). Briefly, this scheme involves homologous recombination between helper virus DNA and the replicated Ub^--Uc^+-defective genomes (Fig. 8f), followed by intramolecular recombination between the created recombinant junction and an adjacent deleted junction in the same concatemer (Fig. 8g). This last recombination creates chimeric junctions with the inserted Justin *a* sequence carrying a Uc element derived in part from the F construct (Fig. 8h). The heteroconcatemer is then cleaved-packaged by the double-stranded break-gap repair mechanism (Fig. 8i), with the sequence variation in Ub arising as a result of slippage during the resolution of the Holliday structures (Fig. 8i and j). In contrast, heteroconcatemers, which are predicted to arise by similar recombinational events with Ub^+-Uc^- constructs, would not become cleaved and packaged owing to the lack of Uc element in junction 1. Thus, unlike the directional cleavage model, the present model requires that both junctions (which are cleaved to generate the termini) contain the Uc signal.

***pac-1* and *pac-2* homologies.** The proposed role of the



B. pac-2

		CG motif	BssHII		T_n	BssHII	L Terminal region	SmaI	end
F	....	CGCCGCG	cgcgcgcacgccgcccggaccgccgcccgcc		TTTTTT	gcgcgccgccccgcccgcgggggccccgggctgc			L
Justin	....	CGCCGCG	cgcgcgcacgccgcccggaccgccgcccgcc		TTTTTT	gcgcgccgccccgccgcgggggccccgggctgc			L
KOS	....	CGCCGCG	cgcgcgcgcgccgcccggaccgccgcccgcc		TTTTTT	gcgcgcgcgcgcgccccgcgggggccccgggctgc			L
17	....	CGCCGCG	cgcgcgcacgccgcccggaccgccgcccgcc		TTTTTT	gcgcgcgcgcgcgccccgcgggggccccgggctgcc			L
USA-8	....	CGCCGCG	cgcgcgcgcgccgcccggaccgccgcccgcc		TTTTTT	gcgcgccgccccgcccgcgggggccccgggccgcc			L
HG52	....	CGCCGCG	ggctgccttccgcggcgcccccgcgcggc		TTTTTT	cccgcgcccgcccccgcgcggcaggacgggggactcccgccc			L
VZV	....	CGCCcCG	caaacgcgcggggaggtggggtcgc		TTTTTTTTTT	ctctctcgaggggggccgcgagagggctggcctcc			L
EBV	....	gGCCGCG	ggggccagccacgccccctccac		TTTTT	ccaggaatgcgcggcggcccatgctccccccatgc			-
HVS 11	....	CcCCGaG	ctcgctctagccacgcccaggaca		TTTTT	ccagctgcccagcgcccactgcttggggcccc			-
BHV-1	....	CGCCGCG	cagagaaaa		TTTTTTTTTT	ctccgcgcccccccgcgcgggggctgggcc			L

C. pac-1

	end	S terminal region	C_n	G_n	T motif	G_n	
		DraI	SmaI			ApaI	
HSV-1 (F)	S	gcgccgccgcgctttaaagggccgcgcgcga	CCCCC	GGGGGG	TGTGTTTC	GGGGGGGG	cccgt ....
HSV-1 (Justin)	S	gccgccgccgctttaaagggccgcgcgcga	CCCCC	GGGGGG	TGTGTTTC	GGGGGGGG	cccgt ....
HSV-1 (KOS)	S	gcccgccgcccgctttaaagggccgcgcgcga	CCCCC	GGGGGG	TGTGTTTT	GGGGGGGG	cccgt ....
HSV-1 (17)	S	tgcccgccgccaccgctttaaagggccgcgcgcga	CCCCC	GGGGGG	TGTGTTTT	GGGGGGGG	cccgt ....
HSV-1 (USA-8)	S	gcccgccgccgcgctttaaagggccgcccgcga	CCCCC	GGGGGG	TGTGTTTC	GGGGGGGG	cccgt ....
HSV-2 (HG52)	S	ctcccgccccgcctcttttccccgggagtcaa	CCCCC	GGGGGGGG	TGTTTTTT	GGGGGGGGG	cgcga ....
VZV	S	cctctcccggggtccgccgggcgcccagaaa	CC	GGGGGGGGG	TTATTTTC	GGGGGGGGG	tccga ....
EBV/Namalwa	-	ccccccatgccctctccaacacccccgaag	CCCCCC	GGGGG	TCTTTCCT	GGGGG	cctttt ....
HVS (11)	-	cccctttcccctctttgcctaccaagttat	CCCCC	GGGGGG	AAAATCAGT	GGGGG	ctgca ....

D.

Virus	Reference	S terminus/pac-1 Distance end → T	S terminus/pac-1 %GC Ster region	L Terminus/pac-2 Distance T_n → end	L Terminus/pac-2 %GC Lter region	Ave %GC genome
F	Mocarski and Roizman, 1982	41-42	83	33-34	97	67
Justin	Mocarski et al., 1985	40-41	83	31-32	97	67
KOS	Varmuza and Smiley, 1985	43-44	84	33-34	97	67
17	Davison and Wilkie, 1981	44-46	80	33-35	97	67
USA-8	Davison and Wilkie, 1981	42-44	84	33-35	100	67
HG52	Davison and Wilkie, 1981	44-47	79	30-41	90	69
VZV	Davison, 1984	41-42	83	33-34	76	46
EBV	Matsuo et al., 1984	?	79	?	74	59
HVS 11	Bankier et al., 1985	40-41	66	31-32	78	46
BHV-1	Hammerschmidt et al., 1986	-	-	30	93	72

E. summary:

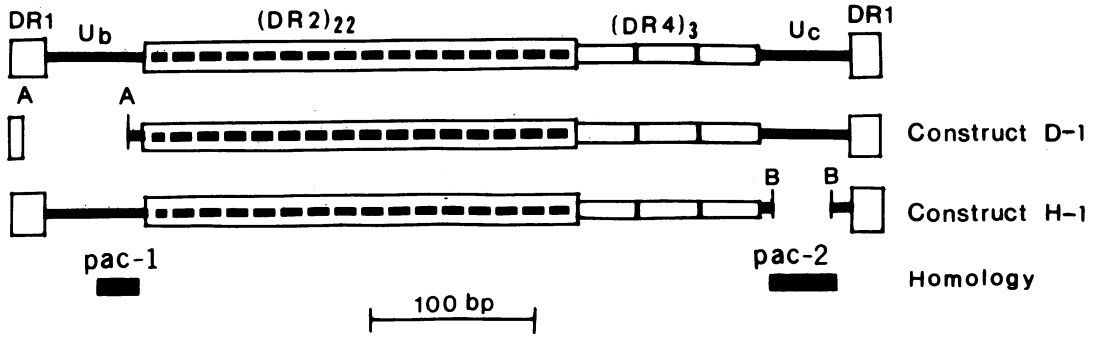


FIG. 10. *pac* homologies in different herpesviruses. (A) Alignment of the *pac-1* and *pac-2* sequences flanking the cleavage site. (B) Conserved *pac*-2 (Uc homologous) sequences (capital letters) found at concatemeric junctions of various herpesviruses. References for the nucleotide sequences are listed in panel D. Cleavage sites reported in the literature are underlined at the end of the sequences shown. The Epstein-Barr virus sequence shown is that of the integrated genome in the Namalwa cell line (16). Hence no cleavage (underlined) site is indicated. (C) Conserved *pac-1* (Ub homologous) sequences found at concatemeric junctions of various herpesviruses. (D) Summary of distances from the T motif in *pac-1* (arrow in panel A) and T_n element in *pac-2* (arrow in panel B) to the respective genomic termini, and the percent G+C residues in the S and L terminal regions. The average G+C content of the respective viral genomes are taken from reference 26. (E) Deletions defined by the D-1 (Ub) and H-1 (Uc) constructs overlap the conserved *pac-1* and *pac-2* sequences. The *Apa*I (A) and *Bss*HII (B) sites which define the deletions are indicated.

HSV-1 Ub and Uc elements in the cleavage-packaging process raises the question of whether similar sequences are present in concatemeric junctions of additional herpesviruses. Figure 10 summarizes the nucleotide sequences in the corresponding DNA regions of a number of herpesviruses, including five HSV-1 strains (6, 18–21, 35), HSV-2 strain HG52 (6), varicella-zoster virus (5), the integrated Epstein-Barr virus genome in the Namalwa cell line (16), herpesvirus saimiri (1), and bovine herpes virus 1 (10). It is convenient to consider the concatemeric junctions of these viruses as consisting of two highly conserved regions flanking a more variable cleavage site (in HSV-1 this will correspond to the arrangement Uc-DR1-Ub; Fig. 10A).

On one side of the cleavage site (e.g., proceeding from Uc into the DR1 in a HSV double *a* junction, the region which will become the L terminus; Fig. 10B) a consensus sequence CGCCGCG (the CG motif) precedes a short stretch of unconserved nucleotides, followed by a highly conserved run of T residues (T_n) and terminating with a G+C-rich region (the L-terminal region; the G+C content is given in Fig. 10D).

Two lines of evidence support the involvement of this region in the cleavage-packaging process. First, the T_n sequence is located some 30 to 35 bp away from the cleavage site in all the herpesviruses above, supporting a measurement mechanism for the cleavage reaction. Second, the *Bss*HII deletion in the construct H-1 described in this paper deletes the conserved T_n sequence and abolishes the ability of defective genomes to propagate in virus stocks. Consequently, we have designated the CG motif followed by the runs of T residues as the *pac-2* homology. It is noteworthy that Hammerschmidt et al. (10) recently reported the conservation of the stretch of A residues (reading on the opposite strand) between bovine herpes virus 1 and varicella-zoster virus.

Proceeding rightward from the double *a* junction (e.g., from DR1 into Ub in the HSV double *a* junction; Fig. 10C) a 40- to 50-bp stretch of G+C-rich sequences can be found (S-terminal region; the G+C content is given in Fig. 10D) at the edge of which are uninterrupted runs of C (C_n) and G (G_n) residues. This is then followed by a T-rich element (T motif) and another run of G residues (G_n). The conservation of sequences in this region was recognized previously (5, 6, 16, 18, 30, 34). Similarly to the *pac-2* region, the start of the T motif is located at a constant distance (41 to 47 bp) from the cleavage site (Fig. 10D), suggesting a measurementlike mechanism. Furthermore, the deletion present in the D-1 (Ub^-) construct spans this region in the HSV-1 Ub sequence, starting from the *Apa*I site in the junction DR1 to the *Apa*I site in the second G_n stretch (Fig. 10C and E). Since this deletion affected the ability of defective genomes to propagate autonomously, we have designated the highly conserved C_n-G_n-T motif-G_n as the *pac-1* homology.

The function of these various conserved elements is at present unknown. However, from the results presented in this paper it is clear that one or more of the *pac-2* and the *pac-1* elements must play a role in the cleavage-packaging process. It is tempting to speculate that the T elements within the *pac-2* and *pac-1* homologies serve as DNA bending sites, as recently shown for T_n elements present within the T-antigen binding site I of simian virus 40 (27) and in trypanosome kinetoplast DNA (39).

An exception to the conserved organization of the cleavage target is the *a* sequence of human cytomegalovirus (30, 34), in which the *pac-1* and *pac-2* sequences appear in several locations. It remains to be seen whether some of the complex patterns of *a* amplifications in human cytomegalovirus strain AD169 (34) are directed by the *pac-2* signals.

Role of *a* amplification in different herpesviruses. The herpesviruses listed in Fig. 10 fall into two groups in terms of the arrangement of the *pac* signals in their genomic termini. The first group includes thus far varicella-zoster virus and bovine herpes virus 1 (5, 10), in which the *pac-1* and *pac-2* homologies are present at opposite termini of the genome. Consequently, fusion of the ends to generate the putative rolling-circle template creates junctions in which the *pac-1* and *pac-2* are flanking the cleavage site. No *a* amplifications have been reported in these virus genomes. In the second group of herpesviruses (e.g., HSV, cytomegalovirus, and herpesvirus saimiri), the *pac-1* and *pac-2* signals are separated by internal *a* sequences. DNA molecules of these viruses contain variable numbers of *a* sequence reiterations at the termini and concatemeric junctions. It is possible that *a* amplification must precede the final cleavage event in these viruses to bring the *pac-1* and *pac-2* signals into the appropriate arrangement flanking the cleavage site. It remains to be seen whether junctions and terminal sequences of additional herpesviruses will be found to contain the *pac-1* and *pac-2* homologies in a similar organization relative to the termini of the viral genome.

ACKNOWLEDGMENTS

We thank Bernard Roizman for making available the constructs containing the *a* sequences before publication. We thank Malera Traylor and Susan Reiser for technical help.

These studies were supported by Public Health Service research grants AI-15488 and CA-19264 from the National Cancer Institute and by National Science Foundation grant PCM 8118303. L.P.D. is a Public Health Service predoctoral trainee (CA-09241). J.C. was a Public Health Service predoctoral trainee (GM 07197).

LITERATURE CITED

1. **Bankier, A. K., W. Dietrich, R. Baer, B. G. Barrell, F. Colbere-Garapin, B. Fleckenstein, and W. Bodemer.** 1985. Terminal repetitive sequences in herpesvirus saimiri virion DNA. J. Virol. **55**:133–139.
2. **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.
3. **Ben-Porat, T., and S. A. Tokazewski.** 1977. Replication of herpesvirus DNA. II. Sedimentation characteristics of newly-synthesized DNA. Virology **79**:292–301.
4. **Chou, J., and B. R. Roizman.** 1985. Isomerization of herpes simplex type 1 genome: identification of the cis-acting and recombinational sites within the domain of the *a* sequence. Cell **41**:803–811.
5. **Davison, A. J.** 1984. Structure of the genome termini of varicella-zoster virus. J. Gen. Virol. **65**:1969–1977.
6. **Davison, A. J., and N. M. Wilkie.** 1981. Nucleotide sequences of the joint between the L and S segments of herpes simplex virus type 1 and 2. J. Gen. Virol. **55**:315–331.
7. **Deiss, L. P., and N. Frenkel.** 1986. The herpes simplex amplicon: cleavage of concatemeric DNA is linked to packaging and involves the amplification of the terminally reiterated *a* sequence. J. Virol. **57**:933–941.
8. **Frenkel, N.** 1981. Defective interfering herpesviruses, p. 91–120. *In* A. J. Nahmias, W. R. Dowdle, and R. S. Schinazi. (ed.), The human herpesviruses—an interdisciplinary perspective. Elsevier Science Publishing, Inc., New York.
9. **Frenkel, N., L. P. Deiss, and R. R. Spaete.** 1984. Studies of HSV propagation using HSV amplicons, p. 289–299. *In* F. Rapp (ed.), Herpesvirus. Proceedings of a Burroughs-Wellcome-UCLA Symposium. Alan R. Liss, Inc., New York.

10. **Hammerschmidt W., H. Ludwig, and H.-J. Buhk.** 1986. Short repeats cause heterogeneity at genomic terminus of bovine herpesvirus 1. J. Virol. **58**:43–49.
11. **Holliday, R.** 1964. A mechanism for gene conversion in fungi. Genet. Res. **5**:282–304.
12. **Jacob, R. J., L. S. Morse, and B. Roizman.** 1979. Anatomy of herpes simplex virus DNA. XII. Accumulation of head to tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. J. Virol. **29**:448–457.
13. **Ladin, B. F., M. L. Blankenship, and T. Ben-Porat.** 1980. Replication of herpesvirus DNA. V. Maturation of concatemeric DNA of pseudorabies virus to genome length is related to capsid formation. J. Virol. **33**:1151–1164.
14. **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.
15. **Locker, H., and N. Frenkel.** 1979. The *Bam*I, *Kpn*I, and *Sal*I restriction enzyme maps of the DNAs of herpes simplex virus strains Justin and F: occurrence of heterogeneities in defined regions of the viral DNA. J. Virol. **32**:429–441.
16. **Matsuo, T., M. Heller, L. Petti, E. O'Shiro, and E. Kieff.** 1984. Persistence of the entire Epstein-Barr virus genome integrated into human lymphocyte DNA. Science **226**:1322–1325.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **Mocarski, E. S., L. P. Deiss, and N. Frenkel.** 1985. The nucleotide sequence and structural features of a novel U_s-*a* junction present in a defective herpes simplex virus genome. J. Virol. **55**:140–146.
19. **Mocarski, E. S., and B. Roizman.** 1981. Site-specific inversion sequence of the herpes simplex virus genome: domain and structural features. Proc. Natl. Acad. Sci. USA **78**:7047–7051.
20. **Mocarski, E. S., and B. Roizman.** 1982. Herpesvirus-dependent amplification and inversion of cell-associated viral thymidine kinase gene flanked by viral *a* sequences and linked to an origin of viral DNA replication. Proc. Natl. Acad. Sci. USA **79**: 5626–5630.
21. **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.
22. **Poffenberger, K. L., and B. Roizman.** 1985. A noninverting genome of a viable herpes simplex virus 1: presence of head-to-tail linkages in packaged genomes and requirements for circularization after infection. J. Virol. **53**:587–595.
23. **Preston, V. G., J. A. V. Coates, and F. Rixon.** 1983. Identification and characterization of a herpes simplex virus gene product required for encapsidation of viral DNA. J. Virol. **45**:1056–1064.
24. **Rao, R. N., and S. G. Rogers.** 1979. Plasmid pKC7: a vector containing ten restriction endonuclease sites suitable for cloning DNA segments. Gene **7**:79–82.
25. **Roizman, B.** 1979. The structure and isomerization of herpes simplex virus genomes. Cell **16**:481–494.
26. **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.
27. **Ryder, K., S. Silver, A. L. DeLucia, E. Fanning, and P. Tegtmeyer.** 1986. An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. Cell **44**:719–725.
28. **Spaete, R. R., and N. Frenkel.** 1982. The herpes simplex virus amplicon: a new eucaryotic defective-virus cloning-amplifying vector. Cell **30**:295–304.
29. **Spaete, R. R., and N. Frenkel.** 1985. The herpes simplex amplicon. III. Analysis of cis replication functions. Proc. Natl. Acad. Sci. USA **82**:694–698.
30. **Spaete, R. R., and E. S. Mocarski.** 1985. The *a* sequence of the cytomegalovirus genome functions as a cleavage/packaging signal for herpes simplex virus defective genomes. J. Virol. **54**:817–824.
31. **Stow, N. D.** 1982. Localization of an origin of DNA replication within the TR_s/IR_s repeated region of the herpes simplex virus type 1 genome. EMBO J. **1**:863–867.
32. **Stow, N. D., E. C. McMonagle, and A. J. Davison.** 1983. Fragments from both termini of the herpes simplex virus type 1 genome contain signals required for the encapsulation of viral DNA. Nucleic Acids Res. **11**:8205–8220.
33. **Szostak, J. K., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33**:25–35.
34. **Tamashiro, J. C., D. Filpula, T. Friedman, and D. H. Spector.** 1984. Structure of the heterogeneous L-S junction region of human cytomegalovirus strain AD169 DNA. J. Virol. **52**: 541–548.
35. **Varmuza, S. L., and J. R. Smiley.** 1985. Signals for site-specific cleavage of herpes simplex virus DNA: maturation involves two separate cleavage events at sites distal to the recognition site. Cell **41**:792–802.
36. **Vlazny, D. A., and N. Frenkel.** 1981. Replication of herpes simplex virus DNA: localization of replication recognition signals within defective virus genomes. Proc. Natl. Acad. Sci. USA **78**:742–746.
37. **Vlazny, D. A., A. D. Kwong, and N. Frenkel.** 1982. Site specific cleavage/packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full length viral DNA. Proc. Natl. Acad. Sci. USA **79**:1423–1427.
38. **Wagner, M. J., and W. C. Summers.** 1978. Structure of the joint region and the termini of the DNA of herpes simplex virus type 1. J. Virol. **27**:374–387.
39. **Wu, H., and D. M. Crothers.** 1984. The locus of sequence directed and protein-induced DNA bending. Nature (London) **308**:509–513.