Identification and Nucleotide Sequences of Two Similar Tandem Direct Repeats in Epstein-Barr Virus DNA

TIMOTHY R. DAMBAUGH AND ELLIOTT KIEFF*

Kovler Viral Oncology Laboratory, Departments of Medicine and Microbiology, The University of Chicago, Chicago, Illinois 60637

Received 14 June 1982/Accepted 16 August 1982

Epstein-Barr virus DNA is known to have partially homologous segments, designated DL and DR, near the left and right ends of the long unique region (Raab-Traub et al., *Cell* 22:257-267, 1980). DL and DR are each partially composed of tandem direct repeat sequences. DL contains 11 to 14 repeats of a 124-base-pair sequence designated IR2. DR contains approximately 30 direct repeats of a 103-base-pair sequence designated IR4. The DL and DR sequences have colinear partial homology for approximately 2.4 and 1.5 kilobase pairs to the right of IR2 and IR4, respectively. IR2 and IR4 are similar sequences and evolved in part from a common ancestor. Both sequences are 84% guanine and cytosine and have limited homology to Epstein-Barr virus IR1 and to the herpes simplex virus type 1 inverted terminal repeat "a" sequence. IR2 encodes part of an abundant 2.5-kilobase persistent early EBV RNA expressed in productively infected cells, but does not encode part of the 3-kilobase Epstein-Barr virus RNA which is transcribed from the adjacent IR1-U2 region of the Epstein-Barr virus genome in latently infected cells.

The Epstein-Barr virus (EBV) genome is a non-isomerizing linear DNA molecule of approximately 180 kilobase pairs (kbp) which consists of several distinct direct tandem repeat sequences which separate discrete regions of unique sequence complexity DNA (for reviews, see references 5, 6, and 17b). The structural features of EBV DNA are shown in Fig. 1A. Both ends of virion DNA contain a variable number of tandem direct repeats of a 500-base-pair (bp) sequence, designated TR, which may facilitate circularization of viral DNA in infected cells (10, 21). A variable number of direct tandem repeats of a 3,071-bp sequence (2-4, 9, 11), designated IR1, divides the genome into two largely unique sequence domains: a 15-kbp region, US1, and a 150-kbp region, UL. Located near the center of UL is an unusual direct repeat array of three nucleotide triplets (14a, 16).

Two sequences of limited homology, DL and DR, are located near the left and right ends of UL, respectively (27). The homologous DL and DR sequences are at least partially conserved in the genomes of EBV-like viruses that infect baboons, herpesvirus Papio, and chimpanzees, herpesvirus Pan (13-15). Restriction endonuclease fragments containing DL and DR from different EBV isolates vary in size by increments of approximately 120 bp, suggesting that DL and DR each contain a variable number of copies of

a repetitious sequence (4, 19, 27). Digestion of cloned DNA fragments containing DL with *Hae*II yields a supermolar fragment of approximately 120 bp (19).

The *Bam*HI or *Eco*RI restriction endonuclease fragments containing DL or DR from several different EBV isolates have previously been cloned into bacterial plasmids (4, 27). This paper presents detailed restriction endonuclease maps for the DL and DR regions and the nucleotide sequences of the repeat elements within DL and DR. These repeats are designated IR2 and IR4, respectively. IR2 and IR4 differ in length, but are similar in nucleotide sequence, indicating that both have evolved in part from a common progenitor sequence.

MATERIALS AND METHODS

Recombinant plasmids and isolation of EBV DNA fragment. Recombinant plasmids were grown in *Escherichia coli* and purified by isopycnic banding in cesium chloride with ethidium bromide (4). Plasmids pDK286, pNT318, and pDA113 contain EBV *Bam*HI-H DNA from the B95-8, W91, and AG876 EBV isolates, respectively (4, 27). Plasmids pDA70 and pNT27 contain EBV *Bam*HI-B1 DNA from the AG876 and W91 EBV isolates, respectively (4, 27). Plasmid DNAs were digested with restriction endonucleases under the conditions suggested by the manufacturer (New England Biolabs and Bethesda Research Labo-

Fragment linkages were confirmed by hybridization of purified ^{32}P -labeled subfragments to Southern blots of fragments produced by other enzymes or by partial digestion of DNA fragments labeled at one 5' or 3' end (31). Alkaline phosphatase, polynucleotide kinase, and terminal deoxynucleotidyl transferase were obtained from P-L Biochemicals and were used for end labeling DNA. Radiochemicals were obtained from Amersham Corp.

Nucleotide sequencing analyses. Nucleotide sequences were determined by chemical degradation of 5' or 3' end-labeled fragments (22). The IR2 sequence of pDK286, a clone of B95-8 *Bam*HI-H, was determined by degradation of the separated strands of the 3' end-labeled *Hae*II-c fragment of *Bam*HI-H and by sequencing leftward from both the 5' and 3' end-labeled *Sau*3A site in *Hin*fI-a (Fig. 1B and C). The IR4 sequence of pDA70, a clone of AG876 *Bam*HI-B1, was determined by degrading the separated 3' end-labeled strands of the *Sst*II-h fragment and by sequencing 150 bp to the left and right of the 3' end-labeled *Sst*II site between *Sst*II-b and -a (Fig. 1B and C). Nucleotide sequences were analyzed by using previously derived programs (26).

RNA blots and hybridization. Cytoplasmic polyadenylated RNAs were prepared from the latently infected IB4 cell line which was growth transformed by the B95-8 EBV isolate (20, 23, 36) or from productively infected, induced cultures of B95-8 cells (17). The B95-8 *Bam*HI-H-*Hae*II-c fragment was purified by acrylamide gel electrophoresis and 3' end labeled with [^{32}P]dCTP (7). RNAs were denatured in aqueous formamide and formaldehyde, separated in agarose gels containing formaldehyde, and transferred to nitrocellulose (17, 35, 36). RNA blots were hybridized at 50°C for 18 h in 6 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and 50% formamide and washed, and fluorograms were prepared as previously described (17, 36).

RESULTS

Restriction endonuclease maps (Fig. 1B) of EBV B95-8 *Bam*HI-H (pDK286) and of AG876 *Bam*HI-B1 (pDA70) were derived as described above. Cleavage of recombinant plasmids containing the B95-8 or W91 *Bam*HI-H or AG876 *Bam*HI-D1 fragments with *Hae*II yielded a supermolar fragment of approximately 120 bp (data not shown). Similarly, cleavage of recombinant plasmids containing the W91 or AG876 *Bam*HI-B1 fragment with *Sst*II yielded a supermolar fragment of approximately 110 bp (see Fig. 5B, lane 1). *Hae*II partial digestion of the *Sst*II end-labeled *Sst*II-a fragment from B95-8 *Bam*HI-H (Fig. 2A) and *Sst*II partial digestion of the *Sst*II end-labeled *Sst*II-a-*Pvu*II-a and *Sst*II-b-*Bgl*II-a fragments from AG876 *Bam*HI-B1 (Fig. 2B) indicated that B95-8 *Bam*HI-H contained 11 tandem *Hae*II-c fragments of approximately 120 bp and that AG876 *Bam*HI-B1 contained 30 tandem *Sst*II-h fragments of approximately 110 bp. Similar analyses of the

FIG. 2. Identification of tandem repeat sequences in *Bam*HI-H, -D1, and -B1. Panels A and B are autoradiograms of 1.0% agarose gels containing partial digests of 3' end-labeled fragments from B95-8 *Bam*HI-H and AG876 *Bam*HI-D1 and -B1. (A) *Hae*II partial digestion products of the 3' *Sst*II-a fragments, 3' end-labeled at the *Sst*II site, from cloned B95-8 *Bam*HI-H (lane 1) and cloned AG876 *Bam*HI-D1 (lane 2). (B) *Sst*II partial digestion products of the 3' *Sst*II end-labeled *Sst*II-a-*Pvu*II-a fragment (lane 1) and 3' *Sst*II end-labeled *Sst*II-b-*Bgl*II-a fragment (lane 2) from cloned AG876 *Bam*HI-B1. Lanes 3 in both panels are end-labeled *Eco*RI fragments of λ (34) and *Hpa*II fragments of ϕ X174 RFI (29) DNAs. The sizes are indicated in base pairs. The arrows in both panels indicate the positions of the end-labeled fragment of *Bam*HI-H, -D1, or -B1 generated by complete digestion with *Hae*II or *Sst*II, respectively. The difference in molecular weight between successive bands is the size of adjacent *Hae*II fragments in the *Sst*II-a fragment from *Bam*HI-H or -D1 (panel A, lanes 1 and 2) or adjacent *Sst*II fragments in the *Sst*II-a and -b fragments from *Bam*HI-B1 (panel B, lanes 1 and 2). The largest fragment in lanes 1 and 2 corresponds to the undigested 3' end-labeled fragment.

W91 EBV *Bam*HI-H and the *Bam*HI-H containing AG876 *Bam*HI-D1 fragment (12, 19) indicated that W91 *Bam*HI-H had 14 or 15 tandem *Hae*II-c fragments and AG876 *Bam*HI-D1 had

13 tandem *HaeII*-c fragments (Fig. 2A). The *HinfI*-a (Fig. 1B) fragment of W91 *BamHI*-B1 is approximately 500 bp smaller than the *HinfI*-a fragment from AG876 *BamHI*-B1 (27), suggesting that there may be five fewer tandem copies of the *SstI*-h fragment in W91 *BamHI*-B1. The tandemly repeated *HaeII* and *SstI* sites in *BamHI*-H and -B1, respectively, were each likely to result from tandem direct repeat elements which were designated IR2 and IR4, respectively. Variation in the average number of copies of IR2 and IR4 in the *BamHI*-H and -B1 fragments, respectively, could account for the observed differences in size of these fragments from different EBV isolates (12).

Nucleotide sequences of IR2 and IR4. The nucleotide sequences of IR2 and IR4 (Fig. 3) were determined as described above. IR2 was 124 bp and was 84% guanine and cytosine. IR4 was 103 bp and was also 84% guanine and cytosine. The 5'-to-3' nucleotide sequence of IR2 is shown in Fig. 3 as it occurs in the genome orientation indicated in Fig. 1. The *HaeII* site in the rightmost IR2 repeat is located two nucleotides before the transition into the unique sequence region U3. IR4 is also shown 5' to 3' in the genome orientation indicated in Fig. 1. Since neither of the junctions with unique sequences in the adjacent *SstI*-b and -c fragments has been determined for IR4, the sequence is shown with the single *SstI* site in IR4 at the beginning of the sequence. *SstI*-h (IR4) hybridizes to both adjacent *SstI* fragments, b and c, of *BamHI*-B1 (data not shown), indicating that part of the 103-bp IR4 repeat sequence is present in both adjacent fragments. The *SstI* site is therefore likely to be near the center of the IR4 repeat. Both strands

of IR2 and IR4 lack translational termination codons in all three reading frames.

Transcription of IR2. The *BamHI*-H fragment, which contains IR2, encodes part of a 3-kilobase (kb) cytoplasmic polyadenylated RNA in latently infected cells growth transformed by B95-8 EBV (20, 36) and also encodes part of an abundant 2.5-kb persistent early cytoplasmic polyadenylated RNA expressed in productively infected B95-8 cells (17). To determine whether these RNAs are encoded in part by IR2, the ³²P-labeled *HaeII*-c fragment was hybridized to RNA blots (Fig. 4). Labeled IR2 hybridized to the abundant 2.5-kb persistent early RNA in productively infected cells, but not to the 3-kb RNA from latently infected cells. Longer exposure of the fluorograms indicated that IR2 also hybridized to low-abundance 4.5- and 8-kb RNAs from productively infected cells. A trace amount of the 4.5-kb RNA could also be detected in latently infected cells. No RNAs are transcribed from IR4 or DR in B95-8 cells since all of DR is deleted from the B95-8 genome (27, 28).

Homology between *BamHI*-H and -B1. A previous report (27) demonstrated that the homology between *BamHI*-H (DL) and *BamHI*-B1 (DR) is confined to the *HinfI*-a and -b fragments of *BamHI*-H and to the *HinfI*-a and -d fragments of *BamHI*-B1 (Fig. 1). To establish the extent of homology between *BamHI*-H and -B1 within these regions, ³²P-labeled *BamHI*-B1 was hybridized to Southern blots of *BamHI*-H digested with enzymes that cut within *HinfI*-a and -b (Fig. 5A and data not shown). Hybridization was observed to all *BamHI*-H fragments which lie between the *HincII*-b and *BglII*-b fragments (Fig. 5A). *BamHI*-B1 did not hybridize to the

FIG. 3. Nucleotide sequence of IR2 and IR4. The 5'-to-3' direction indicated for IR2 and IR4 corresponds to the polarity of these sequences within the viral genome as shown in Fig. 1. The recognition sites for *HaeII* and *SstI* enzymes, which cut once within IR2 and IR4, respectively, are indicated in boxes. IR2 and IR4 each contain different 5-bp dyad symmetries indicated by a vertical arrow at the axis of the symmetry. The single solid line indicates the positions of a perfect direct repeat of 12 bp in IR2. The double line indicates the position of a 9-bp perfect tandem direct repeat in IR4 and the same 9-bp tandem direct repeat in IR2. The $\blacktriangle$ marks the position of a missing T residue in the second 9-bp repeat in IR2. The dashed lines indicate imperfect direct repeats of 22 bp in IR2 and 16 bp in IR4, and the dotted line marks the position of an imperfect direct repeat of 15 bp in IR4. The imperfect direct repeats shown have greater than 85% base pair matching.

FIG. 4. Homology between IR2 and EBV RNA. Panels A and B are short and long exposures of the same Northern blot of EBV RNA hybridized to a ^{32}P -labeled *Bam*HI-H *Hae*II-c probe. Lane 1 of the blots contains polyadenylated cytoplasmic RNA from latently infected IB4 cells (36). Lane 2 contains polyadenylated cytoplasmic RNA from induced cultures of productively infected B95-8 cells (17). The positions of the 28S and 18S rRNA markers are shown on the gel. The sizes of EBV RNAs which hybridized to the IR2 probe are indicated.

*Hinc*II-b and *Bgl*II-b fragments of *Bam*HI-H (data not shown). The lack of homology to *Hinc*II-b indicates that the left boundary of DL is adjacent to the left boundary of IR2. The limited hybridization of *Bam*HI-B1 to the *Taq*I-c fragment (Fig. 5A) indicates that the right boundary of DL is a short distance to the right of the *Taq*I site. These results indicate that DL sequences span a region of approximately 3.3 kbp including approximately 2.4 kbp to the right of IR2.

To define the DR sequences in *Bam*HI-B1 and determine the polarity of these sequences relative to DL, the labeled *Bam*HI-H *Hae*II-c (IR2), *Hinf*I-a-*Sst*I-a, *Hinf*I-a-*Sst*I-b, and *Hinf*I-b fragments which comprise DL were hybridized to blots of *Bam*HI-B1 digested with restriction endonucleases which cleave within the *Hinf*I-a and -d fragments (Fig. 5B and data not shown). The data which follow indicate that DR is a partially divergent direct repeat of DL. The *Bam*HI-H *Hinf*I-a-*Sst*I-a probe, which contained IR2, hybridized only to the *Hinf*I-a fragment of *Bam*HI-B1 which contained IR4 (Fig. 5B, lane 2). When *Bam*HI-B1 was cleaved with *Sst*I, the 103-bp *Sst*I-h (IR4) fragment bound

inefficiently to nitrocellulose and was barely detected with *Hinf*I-a-*Sst*I-a or *Hae*II-c (IR2) probes. Since *Bam*HI-H did not hybridize to the *Hinf*I-b fragment of *Bam*HI-B1, the left margin of DR sequences probably coincides with the leftmost copy of IR4. The *Bam*HI-H *Hinf*I-a-*Sst*I-b fragment, located 300 bp to the right of IR2, hybridized to the *Bam*HI-B1 *Hinf*I-a-*Sst*I-b, *Hinf*I-i, and *Pvu*II-a and -c fragments which lie to the right of IR4 (Fig. 5B, lane 3, and data not shown). (The weak hybridization observed to the *Hinf*I-d and other fragments on the blot of the *Hinf*I-*Sst*I digest of *Bam*HI-B1 results from hybridization of trace amounts of labeled *Bam*HI-H *Hinf*I-b and pBR322 sequences contaminating the *Hinf*I-a-*Sst*I-b probe.) The *Bam*HI-H *Hinf*I-b fragment which is to the right of the *Bam*HI-H *Hinf*I-a-*Sst*I-b fragment hybridized only to the *Bam*HI-B1 *Hinf*I-d fragment (Fig. 5B, lane 4) and the *Pvu*II-c, -d, and -b fragments (data not shown). Thus, there is direct colinearity between the DL and DR regions. These results indicate that the DR sequences span a region of approximately 4.5 kbp including approximately 1.5 kbp of unique sequence complexity to the right of IR4. The extent of homology of the *Bam*HI-H fragments to *Bam*HI-B1 was assessed qualitatively by including blots containing comparable amounts of *Bam*HI-H fragments in each of the hybridizations with *Bam*HI-H probes. The *Bam*HI-H probes hybridized approximately twofold more to *Bam*HI-H fragments than to *Bam*HI-B1 fragments, suggesting that each probe has only partial homology to *Bam*HI-B1 (data not shown). The *Pvu*II-c fragment of *Bam*HI-H is identical in size to the *Pvu*II-d fragment of *Bam*HI-B1. The *Bam*HI-H *Pvu*II-c and *Bam*HI-B1 *Pvu*II-d fragments are approximately 1.1 and 0.4 kbp to the right of IR2 and IR4, respectively, and are within homologous sequence domains of DL and DR (Fig. 5). The regions of DL and DR which includes the identical *Pvu*II fragments are probably more highly conserved than other regions of DL and DR where restriction endonuclease sites are not conserved.

DISCUSSION

The extent of homology between IR2 and IR4 and the relative orientation of two different direct repeat elements within each sequence (Fig. 3 and 6) indicates that both probably evolved from a single core sequence of approximately 50 bp and that the divergent regions of IR2 and IR4 arose from duplications of different parts of the original core sequence. This putative ancestral core sequence presumably resembles the first 50 bp of the IR2:IR4 consensus sequence in Fig. 6A. The 16- and 15-bp repeat

FIG. 5. Identification of homologous sequences in *BamHI-H* (DL) and *BamHI-B1* (DR). Ethidium bromide-stained agarose gels are shown in lanes 1. Specific DNA fragments generated by single or double digestion are identified by letter. Fragments from pBR322 are marked by a dot. Terminal *HinfI* or *SstI* fragments attached to pBR322 or to *HinfI* fragments of pBR322 are identified as e+pBR, for example. Panel A, lane 2, shows the results of hybridization of a *BamHI-B1* probe to blots of digests of *BamHI-H* plasmid; Panel B, lanes 2 through 4, shows the results of hybridization of blots of digests of *BamHI-B1* with the indicated *BamHI-H* fragment probes.

FIG. 5B.

The location of the homologous DL and DR sequences in *Bam*HI-H and -B1 are indicated below the restriction endonuclease maps of these fragments. The broken lines identify regions of limited or uncertain cross-homology. The DL and DR sequences are depicted in three domains (1, 2, and 3) which correspond to the homologous sequences identified by the three specific labeled fragments of *Bam*HI-H listed in panel B.

elements within IR4 are duplicated in the same order as the putative core sequence, suggesting that the duplication could have been a single event. The 12- and 22-bp repeat elements within IR2 are in inverse order relative to the putative core sequence and therefore are likely to have evolved in two stages. IR2 and IR4 both contain the same 9-bp tandem direct repeat, except for a missing T residue in the second 9-bp repeat in IR2 (Fig. 3). Duplications of parts of the putative core sequence would be expected to either be conserved or eliminated in all copies of IR2 or IR4 (30). The extensive homology between the left part of IR2 and IR4 (Fig. 6A) and the conservation of the 9-bp tandem direct repeats within this region could indicate the importance of this region as some type of regulatory or recognition sequence.

A further observation supportive of the functional importance of IR2 and IR4 is the conservation of sequences homologous to IR2 and IR4 in analogous regions of the genomes of the EBV-related primate viruses, herpesvirus Pan and herpesvirus Papio (13, 14, 19). The presence of both IR2 and IR4 is not required for EBV replication, however, since all copies of IR2 are deleted from P3HR-1 EBV DNA (1a, 19) and all copies of IR4 are deleted from B95-8 EBV DNA (27), yet both of these EBV isolates can replicate (23). IR4, DR, and the 14-kbp DNA segment which contains IR4 and DR are also not required for *in vitro* growth transformation of lymphocytes since the B95-8 virus is deleted for this region but remains competent for growth transformation.

IR2 and IR4 are relatively simple direct repeats and are unlikely to encode proteins. However, both sequences have three open reading frames on both DNA strands. Comparison with published sequences reveals that there are no likely splice donor or acceptor sites within IR2 and IR4 (25). It is surprising therefore that IR2 encodes part of the abundant 2.5-kb persistent early mRNA in productively infected B95-8 cells (18). This RNA appears to encode an 18K polypeptide (17a) which does not appear to be heterogeneous in size as might be expected if it were encoded in part by repeated sequences (32). The 2.5-kb mRNA is large enough to be

transcribed, in part, from IR2 (1.4 kb overall) and still have sufficient coding capacity exclusive of IR2 (1.1 kb) to encode the 18K polypeptide which has been identified as the *in vitro* translation product of this RNA (17, 17a). The ability of P3HR-1 EBV to replicate although it is deleted for IR2 (1a, 19) suggests (i) that either the IR4-DR region of P3HR-1 encodes an equivalent polypeptide or provides an equivalent regulatory or recognition site or (ii) that the IR2-encoded function is not required for replication of P3HR-1 virus. P3HR-1 virus could be unique in its relaxed requirement for IR2 as a consequence of the deletion of the adjacent U2 sequences (1a, 10a, 19) which are part of the template for the 3-kb RNA transcribed from IR1-U2 in latently infected, growth-transformed cells (36).

IR2 and, to a smaller extent, IR4 have extensive homology to oligonucleotides within the 3,071-bp IR1 repeat (Fig. 6B) (2, 3). IR1 also has internal repeat elements and may have evolved by expansion of a more simple sequence (2, 3). Further, the homology of IR2 and IR4 to parts of IR1 suggests the possibility that these parts of IR1 may have evolved by recombination between IR1 and IR2 or IR4 sequences with similar high guanine and cytosine content. The P3HR-1 isolate which is missing the entire U2 and IR2 regions (1a, 19) may be a recent example of a recombinational event between IR1 and IR2. That two deletion derivatives of EBV, P3HR-1 and B95-8, are deleted for U2-IR2 and part of U5-IR4- part of U6, respectively, suggests that IR2 and IR4 may facilitate deletion mutation.

The entire DL and DR regions are imperfect direct repeats of each other. Although there is colinear homology between DL and DR, the homology is limited. Except for the 190-bp *PvuII* fragment, there is little conservation of restriction endonuclease sites between DL and DR. These observations indicate that there is nucleotide sequence divergence throughout the DL and DR homology. These data are compatible with the following alternative hypotheses: (i) that duplication of an entire ancestral DL or DR sequence containing a short tandem direct repeat sequence occurred in a progenitor EBV genome or (ii) that IR2 and IR4 evolved from a

FIG. 6. Consensus sequence of IR2 and IR4. (A) The core nucleotide sequence common to both IR2 and IR4 is indicated by the nucleotides on the horizontal line. Nucleotide residues which are mismatched or deleted in IR2 and IR4 are indicated above or below the consensus sequence, respectively. The numbering scheme indicated above and below the consensus sequence refers to the IR2 and IR4 sequences in Fig. 3. The solid, dashed, and dotted lines indicate the positions of perfect and imperfect direct repeats in IR2 and IR4 described in Fig. 3 (B and C). Regions of homology between IR2 or IR4 (2, 3) and EBV IR1 or the HSV-1 "a" terminal repeat sequence (24) are indicated by horizontal bars below the consensus diagram. The residue numbers at the ends of the bars indicate the homologous residues in IR1 (2, 3) or HSV-1 "a" (24). Symbols: ▼, mismatched or missing residue in IR2 or IR4; ▲, extra residue in IR2 or IR4. The broken horizontal bars indicate the regions of the consensus sequence where several bases are missing from IR2 or IR4.

common sequence and were amplified before recombination between them resulted in the duplication of the sequences to the right of IR2 or IR4 in a progenitor EBV genome.

The genomes of the EBV group differ from those of herpes simplex virus, cytomegalovirus, and pseudorabies virus in the presence of tandem direct repeats in EBV DNAs and the absence of inverted repeats which bracket the unique sequence domains of these other herpesviruses (1, 18, 24). Although the inverted repeat elements in HSV DNA permit isomerization of the DNA, the duplication of inverted repeats in HSV DNA is not necessary for virus replication (K. Poffenberger, E. Paberas, and B. Roizman; personal communication). The HSV inverted repeat element has been sequenced previously (24). Like IR2 and IR4, the sequence is guanine and cytosine rich (77%) and contains numerous internal direct repeat elements, suggesting that it evolved from simpler sequences. Several regions of nucleotide sequence identity to IR2 and IR4 are noted (Fig. 6C). This suggests that the EBV and herpes simplex virus families may have evolved from a common ancestor.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Larvin Chankit in the DNA sequencing, the advice on sequencing from Mark Heller and Andrew Cheung, and the assistance of Christopher Beisel and Susan Fennwald in the computer analysis.

This research is supported by Public Health Service research grants CA 19264 and CA 17281 from the National Institutes of Health and by American Cancer Society grant MV 32G. Elliott Kieff is a Faculty Research Awardee of the American Cancer Society.

LITERATURE CITED

- Ben-Porat, T., R. Veach, and B. Ladin. 1980. Replication of herpesvirus DNA. VI. Virions containing either isomer of pseudorabies virus DNA are infectious. *J. Virol.* 102:370-380.
- Bornkamm, G. W., J. Hudewentz, U. K. Freese, and U. Zimmer. 1982. Deletion of the nontransforming Epstein-Barr virus strain P3HR-1 causes fusion of the large internal repeat to the DSL region. *J. Virol.* 43:952-968.
- Cheung, A., and E. Kieff. 1981. Epstein-Barr virus DNA. X. Direct repeat within the internal direct repeat of Epstein-Barr virus DNA. *J. Virol.* 40:501-507.
- Cheung, A., and E. Kieff. 1982. Long internal direct repeat in Epstein-Barr virus DNA. *J. Virol.* 44:286-294.
- Dambaugh, T., C. Beisel, M. Hummel, W. King, S. Fennwald, A. Cheung, M. Heller, N. Raab-Traub, and E. Kieff. 1980. Epstein-Barr virus (B95-8) DNA. VII. Molecular cloning and detailed mapping. *Proc. Natl. Acad. Sci. U.S.A.* 77:2999-3003.
- Dambaugh, T., M. Heller, N. Raab-Traub, W. King, A. Cheung, C. Beisel, M. Hummel, V. van Santen, S. Fennwald, and E. Kieff. 1980. DNAs of Epstein-Barr virus and Herpes Virus Papio, p. 85-90. In A. Nahmias, W. Dondle, and R. Schinazi (ed.), *Human herpes viruses*. Elsevier-North Holland Publishing Co., New York.
- Dambaugh, T., N. Raab-Traub, M. Heller, C. Beisel, M. Hummel, A. Cheung, S. Fennwald, W. King, and E. Kieff. 1980. Variations among isolates of Epstein-Barr Virus. *Ann. N.Y. Acad. Sci.* 602:711-719.
- Deng, G., and R. Wu. 1981. An improved procedure for utilizing terminal transferase to add homopolymers to the 3' termini of DNA. *Nucleic Acids Res.* 9:4173-4188.
- Given, D., and E. Kieff. 1978. DNA of Epstein-Barr virus. IV. Linkage map for restriction enzyme fragments of the B95-8 and W91 strains of EBV. *J. Virol.* 28:524-542.
- Given, D., and E. Kieff. 1979. DNA of Epstein-Barr virus. VI. Mapping of the internal tandem reiteration. *J. Virol.* 31:315-324.
- Given, D., D. Yee, K. Griem, and E. Kieff. 1979. DNA of Epstein-Barr virus. V. Direct repeats at the ends of EBV DNA. *J. Virol.* 30:852-862.
- Hayward, S. D., S. G. Lazarowitz, and G. S. Hayward. 1982. Organization of the Epstein-Barr virus DNA molecule. II. Fine mapping of the boundaries of the internal repeat cluster of B95-8 and identification of additional small tandem repeats adjacent to the HR-1 deletion. *J. Virol.* 43:201-212.
- Hayward, S., L. Nogee, and G. Hayward. 1980. Organization of repeated regions within the Epstein-Barr virus DNA molecule. *J. Virol.* 33:507-521.
- Heller, M., T. Dambaugh, and E. Kieff. 1981. Epstein-Barr Virus DNA. IX. Variation among viral DNAs from producer and nonproducer infected cells. *J. Virol.* 38:632-648.
- Heller, M., P. Gerber, and E. Kieff. 1981. Herpesvirus papio DNA is similar in organization to Epstein-Barr virus DNA. *J. Virol.* 37:698-709.
- Heller, M., P. Gerber, and E. Kieff. 1981. DNA of herpesvirus Pan, a third member of the Epstein-Barr virus-herpesvirus papio group. *J. Virol.* 41:931-939.
- Heller, M., A. Henderson, and E. Kieff. 1982. Repeat array in Epstein-Barr virus DNA is related to cell DNA sequences interspersed on human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 79:5916-5920.
- Heller, M., and E. Kieff. 1981. Colinearity between the DNAs of Epstein-Barr virus and herpesvirus papio. *J. Virol.* 37:821-826.
- Heller, M., V. van Santen, and E. Kieff. 1982. Simple repeat sequence in Epstein-Barr virus DNA is transcribed in latent and productive infections. *J. Virol.* 44:311-320.
- Hummel, M., and E. Kieff. 1982. Epstein-Barr virus RNA. VIII. Viral RNA in permissively infected B95-8 cells. *J. Virol.* 43:262-272.
- Hummel, M., and E. Kieff. 1982. Mapping of polypeptides encoded by the Epstein-Barr virus genome in productive infection. *Proc. Natl. Acad. Sci. U.S.A.* 79:5698-5702.
- Kieff, E., T. Dambaugh, M. Heller, W. King, A. Cheung, V. van Santen, M. Hummel, C. Beisel, S. Fennwald, K. Hennessy, and T. Heineman. 1982. Biology and chemistry of Epstein-Barr virus. *J. Infect. Dis.* 146:506-517.
- Kilpatrick, B. A., and E.-S. Huang. 1977. Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. *J. Virol.* 24:261-276.
- King, W., T. Dambaugh, M. Heller, J. Dowling, and E. Kieff. 1982. Epstein-Barr virus DNA. XII. A variable region of the Epstein-Barr virus genome is included in the P3HR-1 deletion. *J. Virol.* 43:979-986.
- King, W., A. Powell, N. Raab-Traub, M. Hawke, and E. Kieff. 1980. Epstein-Barr virus RNA. V. Viral RNA in a restringently infected, growth-transformed cell line. *J. Virol.* 36:506-518.
- Kinter, C., and B. Sugden. 1979. The structure of the termini of the DNA of Epstein-Barr virus. *Cell* 17:661-671.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
- Miller, G., D. Coope, J. Niederman, and J. Pagano. 1976. Biological properties and viral surface antigens of Burkitt lymphoma and mononucleosis derived strains of Epstein-Barr virus released from transformed marmoset cells. *J. Virol.* 18:1071-1080.
- Mocarski, E., and B. Roizman. 1981. Site-specific inversion sequence of the herpes simplex virus genome: do-

- main and structural features. *Proc. Natl. Acad. Sci. U.S.A.* **78**:7047-7051.
25. Mount, S. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
26. Queen, C., and Korn, L. 1980. Computer analysis of nucleic acids and proteins. *Methods Enzymol.* **65**:595-609.
27. Raab-Traub, N., T. Dambaugh, and E. Kieff. 1980. DNA of Epstein-Barr virus. VIII. B95-8, the previous prototype, is an unusual deletion derivative. *Cell* **22**:257-267.
28. Raab-Traub, N., R. Pritchett, and E. Kieff. 1978. DNA of Epstein-Barr virus. III. Identification of restriction enzyme fragments which contain DNA sequences which differ among strains of Epstein-Barr virus. *J. Virol.* **27**:388-398.
29. Sangar, F., A. R. Coulson, T. Friedmann, G. M. Air, B. G. Barrell, N. L. Brown, J. C. Fiddes, C. A. Hutchinson III, P. M. Slocombe, and M. Smith. 1978. The nucleotide sequence of bacteriophage ϕ X174. *J. Mol. Biol.* **125**:225-246.
30. Smith, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* **191**:528-535.
31. Smith, H. O., and M. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucleic Acids Res.* **3**:2387-2398.
32. Sprague, K., M. Roth, R. Manning, and P. Gage. 1979. Alleles of the fibroin gene coding for proteins of different lengths. *Cell* **17**:407-413.
33. Sutcliffe, J. G. 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. *Nucleic Acids Res.* **5**:2721-2728.
34. Szybalski, E., and W. Szybalski. 1979. A comprehensive molecular map of bacteriophage lambda. *Gene* **7**:217-270.
35. Thomas, P. A. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. U.S.A.* **77**:5201-5205.
36. van Santen, V., A. Cheung, and E. Kieff. 1981. Epstein-Barr virus RNA. VII. Size and direction of transcription of virus specified cytoplasmic RNA in a cell line transformed by EBV. *Proc. Natl. Acad. Sci. U.S.A.* **78**:1930-1934.