JOURNAL OF VIROLOGY, Sept. 1987, p. 2902–2909
0022-538X/87/092902-08$02.00/0

Vol. 61, No. 9

Isolation and Characterization of cDNA Clones Corresponding to Transcripts from the *Bam*HI H and F Regions of the Epstein-Barr Virus Genome

ARTUR J. PFITZNER,† EUGENE C. TSAI, JACK L. STROMINGER, AND SAMUEL H. SPECK*

Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115

Received 28 January 1987/Accepted 2 June 1987

The Epstein-Barr virus (EBV) mutant P3HR1 is incapable of immortalizing B lymphocytes because of a 6.8-kilobase deletion in the *Bam*HI W, Y, and H regions of the viral genome (M. Rabson, L. Gradoville, L. Heston, and G. Miller, J. Virol. 44:834–844, 1982). To characterize transcripts that are encoded in this region, poly(A)$^+$ RNA from the EBV-transformed lymphoblastoid cell line JY was isolated, and this RNA was used to generate a cDNA library in λ gt10. By screening 500,000 recombinant bacteriophages with the *Bam*HI H fragment, we isolated 10 cDNA clones and characterized them in detail. One group of six cDNA clones was derived from a 2.9-kilobase early transcript encoded by the IR2 repeat element and showed restriction site polymorphism for the enzyme *Sma*I. The second group consisted of four cDNA clones, all of which contained the *Bam*HI-H right reading frame (BHRF1), and used the polyadenylation signal at base pair 662 in the *Bam*HI F fragment. Computer analysis of the hydrophobicity of the BHRF1 protein revealed that it is likely to be a membrane protein. Northern blotting experiments with RNA from an EBV producer line, B95-8, and a tightly latent lymphoblastoid B-cell line, IB4, revealed that BHRF1 is contained in at least two different mRNA species which can be detected during the latent cycle of EBV. These data and the recent characterization of a spliced transcript (containing five exons in common with other known latent messages [M. Bodescot and M. Perricaudet, Nucleic Acids Res. 14:7103–7113, 1986]) suggest that alternative splicing is used to generate transcripts containing BHRF1, as for the EBV nuclear antigen 1 transcripts. Furthermore, the observation that a potential oncogene activated in human follicular lymphomas is homologous to the BHRF1-encoded polypeptide (M. L. Cleary, S. D. Smith, and J. Sklar, Cell 47:19–28, 1986) suggests a possible role for this putative viral protein in EBV-induced growth transformation of B lymphocytes.

Epstein-Barr virus (EBV) is a human lymphotropic herpesvirus which is the etiologic agent of infectious mononucleosis and is associated with two human cancers, Burkitt's lymphoma and nasopharyngeal carcinoma. A characteristic feature of EBV is its ability to transform resting human B lymphocytes into continuously growing lymphoblastoid cell lines (LCLs; 13). One important clue to which region of the EBV genome is essential for growth transformation is provided by the nontransforming EBV strain, P3HR1. This virus has a 6.8-kilobase (kb) deletion when compared with several other strains of EBV that are immortalization competent (5, 19, 29). The deletion starts at the right end of the major internal repeat (IR1) and extends through the *Bam*HI Y fragment and the internal repeats (IR2) 3.5 kb into *Bam*HI-H (Fig. 1). The importance of this deletion is emphasized by the analysis of a series of recombinant viruses generated by superinfection of a Burkitt's lymphoma cell line, Raji, with P3HR1. Immortalizing viruses recovered from the culture supernatant were recombinants, all of which retained the *Bam*HI W, Y, and H regions and parts of the F region from the Raji genome (34). One open reading frame (ORF) in this region, the *Bam*HI-Y right reading frame (BYRF1), has been shown to code for an EBV nuclear antigen, EBNA 2 (30). The sequence analysis of cDNA clones revealed a complicated splicing pattern, with repetitive use of exons from the *Bam*HI W fragments (31, 36). The other transcripts have only been characterized by Northern (RNA) blotting, and even their number is controversial (39, 42).

Here we report the results of cDNA cloning of mRNAs encoded by the *Bam*HI H fragment. We present DNA sequencing data of 10 cDNA clones which indicate that they represent transcripts from the IR2 repeat region and BHRF1. These transcripts are further characterized by Northern blotting experiments with RNA from the EBV-transformed B95-8 marmoset cell line and from JY and IB4 LCLs.

MATERIALS AND METHODS

Cell culture. The B-cell lines IB4 (20), B95-8 (26), and JY (38) were used. Cultures were grown at 37°C in RPMI 1640 medium containing 10% fetal calf serum with antibiotics as previously described (15).

RNA preparation and Northern blotting. Total cellular RNA was prepared by the method of Auffray and Rougeon (1). Frozen cells (5 g) were suspended in 50 ml of lysis buffer (6 M urea and 3 M lithium chloride containing heparin at 100 μg/ml) and disrupted in a Waring blender for 2 min. The lysate was stored on ice at 4°C for 4 to 16 h to allow precipitation of the RNA. The RNA was recovered by centrifugation for 30 min at 10,000 × *g*, suspended in 20 ml of 10 mM Tris hydrochloride (pH 7.6)–0.5% sodium dodecyl sulfate (SDS), and then extracted several times with chloroform-isoamyl alcohol (24:1 [vol/vol]) and precipitated. Poly(A)$^+$ RNA was isolated by fractionation on oligo(dT)-cellulose (2). Total or poly(A)$^+$ RNA (10 μg) was fractionated on a formaldehyde-agarose gel (23) and subsequently transferred to a nylon membrane (Pal Biodyne A). RNA

* Corresponding author
† Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

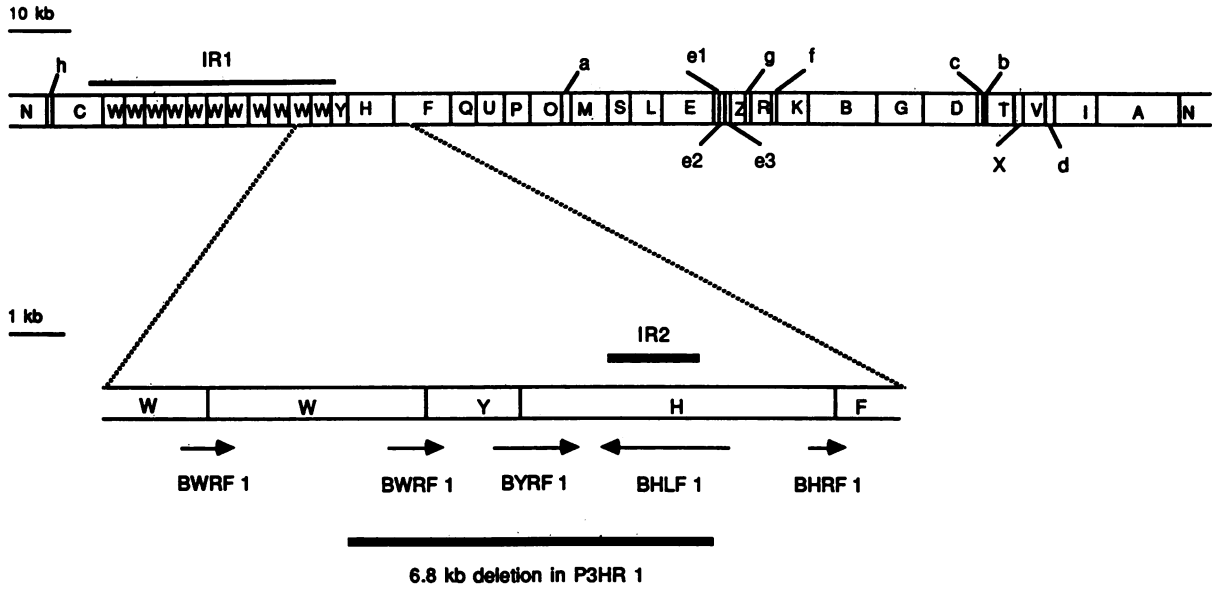

FIG. 1. *Bam*HI restriction endonuclease map of the B95-8 strain of EBV. The *Bam*HI W, Y, H, and F regions are enlarged to show the sizes of the deletion in the genome of the nontransforming virus strain P3HR1. The sizes and orientations of the ORFs BWRF1, BYRF1, BHLF1, and BHRF1 are indicated by arrows. The position of the IR2 repeat elements in *Bam*HI-H is represented by a solid bar.

blots were hybridized with a ^{32}P-labeled nick-translated DNA fragment in 50% (vol/vol) formamide–50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid, pH 7.0)–1× Denhardt solution (9)–6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.0) containing denatured salmon sperm DNA at 100 μg/ml. The filters were washed twice with 1× SSC–0.05% SDS for 15 min at room temperature and four times with 0.1× SSC–0.05% SDS for 15 min at 42°C. Autoradiography was carried out for 12 h at −70°C by using an intensifying screen (Du Pont Co., Wilmington, Del.).

cDNA synthesis. A cDNA library was prepared from poly(A)$^+$ JY RNA by a modification of the method of Gubler and Hoffmann (12). Poly(A)$^+$ (5 μg) RNA was denatured at 68°C for 3 min, quickly cooled on ice, and incubated with oligo(dT)$_{12-14}$ (100 μg/ml) for 30 min at 42°C in 70 mM Tris hydrochloride (pH 8.3)–14 mM dithiothreitol–6 mM sodium PP_i–50 mM KCl containing RNasin (Promega Biotech, Madison, Wis.) at 1 U/μl and nuclease-free bovine serum albumin (Enzo Biochemicals, New York, N.Y.) at 50 μg/ml. Deoxynucleoside triphosphates (1.25 mM) and avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc., St. Petersburg, Fla.) at 3 U/μl were added, and incubation was continued for an additional 30 min. Second-strand synthesis was carried out by using RNase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) and *Escherichia coli* DNA polymerase I (New England BioLabs, Inc., Beverly, Mass.) as previously described (12), except that *E. coli* DNA ligase and β-NAD$^+$ were omitted. The double-stranded cDNA was ligated to *Eco*RI linkers with T4 DNA ligase (both from New England BioLabs), digested with *Eco*RI, and fractionated over a Sepharose CL-4B column. The cDNA was cloned into *Eco*RI-digested calf intestinal alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.)-treated λ gt10 (17). Recombinant bacteriophage was packaged in vitro and plated on the C600 *hfl* strain of *E. coli*.

Screening of the cDNA library. Approximately 5 × 10^4 bacteriophage were plated on a lawn of C600 *hfl E. coli* on each of 10 150-mm plates, and the phage was grown at 37°C for about 6 h. Plaques were transferred to nitrocellulose filters, and the phage DNA was denatured by a 30-s treatment with 0.5 M NaOH–1.5 M NaCl, which was followed by neutralization for 15 min with 0.5 M Tris hydrochloride (pH 8.0)–1.5 M NaCl and washing with 2× SSC. The filters were dried at room temperature for 30 min and then baked at 80°C under vacuum for 2 h. They were then prehybridized for 2 h at 68°C in 6× SSC–5× Denhardt solution (9)–0.1% SDS containing denatured salmon sperm DNA at 100 μg/ml. ^{32}P-labeled nick-translated probe (200 ng; 2 × 10^8 to 6 × 10^8 cpm/μg) was added, and the mixture was hybridized for 16 to 20 h at 68°C. Filters were washed twice in 2× SSC–0.1% SDS for 15 min at room temperature and twice in 0.3× SSC–0.1% SDS for 1 h at 68°C. Autoradiography was carried out overnight at −70°C with intensifying screens. Positive plaques were picked and purified (25). Inserts were excised from the phage DNA with *Eco*RI and subcloned into *Eco*RI-digested phosphatase-treated pUC18 vector.

DNA sequencing. Appropriate restriction endonuclease fragments were subcloned into phage M13 mp18 and M13 mp19 directly from low-melting-point agarose as described by Crouse et al. (8). Random deletion clones were generated by using *E. coli* DNase I in the presence of manganese ions, as described by Lin et al. (24) and cloned into M13 (27). DNA was sequenced by the chain termination method of Sanger et al. (33) by using [^{35}S]dATP.

RESULTS

Isolation and mapping of cDNA clones. Approximately 500,000 recombinant λ phages of a λ gt10 cDNA library generated from poly(A)$^+$ RNA of the EBV-transformed B-cell line JY were screened with the *Bam*HI H fragment. Positive clones were subcloned into plasmids and used to probe Southern blots (35) of *Bam*HI restriction endonuclease-digested EBV genomic DNA (Fig. 2). Two groups of cDNA clones were evident. The first class consists of six clones (H1, 1.3 kb; H2, 1.7 kb; H3, 1.7 kb; H4, 1.8 kb; H5,

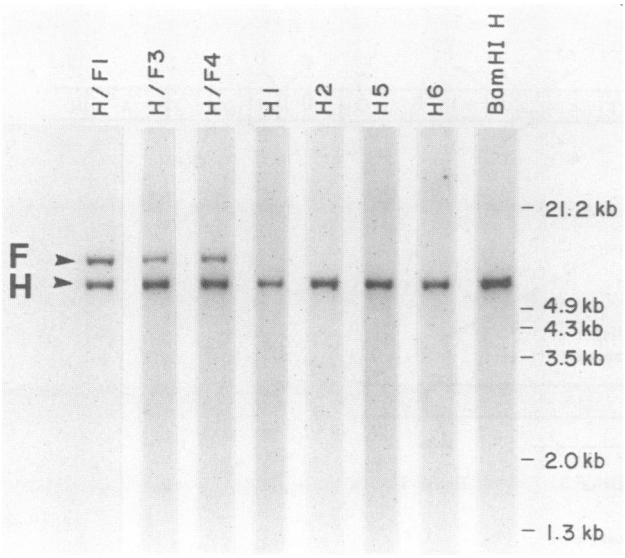

FIG. 2. Southern blot of *Bam*HI-digested B95-8 DNA probed with different cDNA clones and the *Bam*HI H fragments. DNA (10 μg per lane) was loaded, separated on a 0.7% agarose gel, and transferred to nitrocellulose (25, 35). The positions of the *Bam*HI H and the F fragments are indicated by arrowheads.

2.1 kb; and H6, 2.8 kb) which hybridized only to the *Bam*HI H fragment. Clones of the second group (H/F1, 1.5 kb; H/F2, 2.2 kb; H/F3, 2.4 kb; and H/F4, 2.8 kb) hybridized to two different fragments, *Bam*HI H and *Bam*HI F. Clone H/F2 was recovered from screening a cDNA library prepared from the RNA of the IB4 LCL.

A common feature of these isolates is that they are not spliced. Detailed restriction mapping and Southern blot hybridizations more precisely mapped these transcripts within the *Bam*HI H and F restriction fragments, and the results are summarized in Fig. 3.

Sequence analysis of the BHRF1 containing cDNA clones. Clones H/F1, H/F2, H/F3, and H/F4 were digested with appropriate restriction enzymes. The fragments were purified on low-melting-point agarose gels and subcloned into mp18 and mp19. The sequence was determined by the method of Sanger et al. (33). All clones have poly(A)$^+$ tracts which start, with one exception, 20 base pairs (bp) downstream from the AATAAA sequence at bp 665 in the *Bam*HI F fragment. In clone H/F4, the poly(A)$^+$ is added 3 bp downstream, which may reflect some flexibility in the site specificity of RNA processing. The 5′ ends of the four cDNAs are overlapping but of different lengths. The largest clone is H/F4, which includes 2,051 bp of the *Bam*HI H fragment and has a total size of 2,747 bp. The nucleotide sequence of this clone and the translation of BHRF1 are shown in Fig. 4. These clones exhibit several interesting features which are worth noting. (i) All of these clones are unspliced. (ii) The corresponding transcripts appear to be much larger than the encoded BHRF1 (i.e., clone H/F4 has a 593-bp 3′ untranslated region, a 473-bp coding region, and 1,571-bp 5′ untranslated region, indicating that only 17% of the transcript is actually used to generate a translation product). (iii) Clone H/F4 overlaps the promoter region of the leftward transcript containing BHLF1 (Fig. 3), which was mapped by Laux et al. (22).

BHRF1-encoded polypeptide may be a membrane protein. Analysis of the primary structure of the putative BHRF1-encoded viral protein with the computer program of Kyte and Doolittle (21) suggests that this protein has the characteristics of a typical cell surface protein. The hydropathy plot is shown in Fig. 5. Near the carboxy terminus is a group of 21 hydrophobic amino acids, which could function as the transmembrane region, while the 5 hydrophilic residues at the carboxy terminus contain two positive charges (arginine residues), typical of a cytoplasmic region for many membrane proteins. In addition, there are two N-linked glycosylation sites at residues 22 and 118, also a typical feature of surface antigens, such as human lymphocyte antigen (28). Furthermore, the amino terminus of this protein fulfills the requirements of a signal peptide. After a relatively hydro-

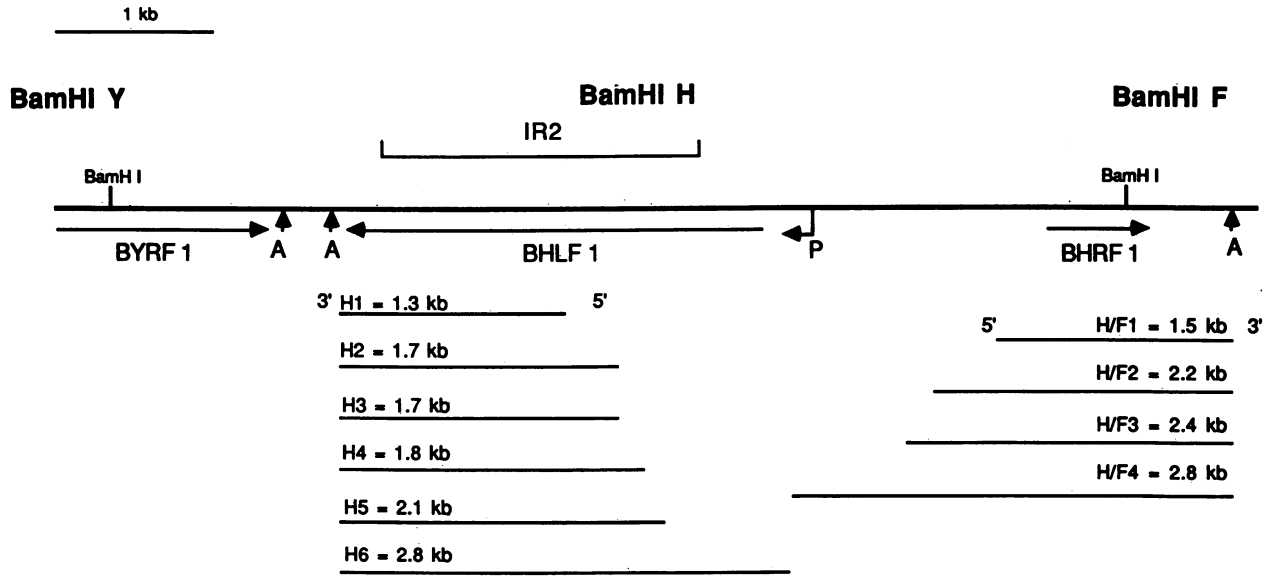

FIG. 3. Scheme of the *Bam*HI H and F regions indicating the two groups of cDNA clones isolated. The sizes and orientations of the ORFs are shown by arrows. The position of the IR2 repeat elements in the *Bam*HI H fragment is depicted. A, Polyadenylation signals used for the adjacent ORFs; P, promoter region for the BHLF1 transcript.

```
   1 TCGCCTTCTTTTATCCTCTTTTTGGGGTCTCTGTGTAATACTTTAAGGTTTGCTCAGGAGTGGGGGCTTCTTATTGGTTAATTCAGGTGTGTCATTTTAG  100
 101 CCCGTTGGGTTTCATTAAGGTGTGTCACCAGGTGGGTGGTACCTGGAGGTTATTCTATTGGGATAACGAGAGGAGGAGGGGCTAGAGGTCCGCGAGATTT  200
 201 GGGGTAGGCGGAGCCTCAGGAGGGTCCCCTCCATAGGGTTGAACCAGGAGGGGGAGGATTGGGCTCCGCCCCGATATACCTAGTGGGTGGAGCCTAGAGG  300
 301 TAGGTATCCATAGGGTTCCATTATCCTGGAGGTATCCTAAGCTCCGCCCCTATATACCAGGTGGGTGGAGCTAGGTAGGATTCAGCTAGGTTCCTACTGG  400
 401 GGTACCCCCCTACCCTACCTTAAGGTGCGCCACCCTTCCTCCTTCCGTTTTAATGGTAGAATAACCTATAGGTTATTAACCTAGTGGTGGAATAGGGTAT  500
 501 TGCAGCTGGGTATATACCTATAGGTATATAGAACCTAGAGGAAGGGAACCCTATAGTGTAATCCCTCCCCCCCCTACCCCCCCCTCCCTTACGGTTGCCT  600
 601 GAGCCCATCCCCCACCCCAGCACCCCGGGGTGACGTGGCACCCCGCGTGCCTTACTGACTTGTCACCTTTGCACATTTGGTCAGCTGACCGATGCTCGCC  700
 701 ACTTCCTGGGTCATGACCTGGCCTGTGCCTTGTCCCGTGGACAATGTCCCTCCAGCGTGGTGGCTGCCTTTGGGATGCATCACTTTGAGCCACTAAGCCC  800
 801 CCGTTGCTCGCCTTGCCTGCCTCACCATGACACACTAAGCCCCTGCTAATCCATGAGCCCCGCCTTTAGGAAGCACCACGTCCCGGGGACGGAAGGGGAC  900
 901 TTGGGGTGATTTTCTATGTGGGGGTGGAAATATGAGCAAGAATAAGGACGGCTCCTTATTAACCTGATCAGCCCCGGAGTTGCCTGTTTCATCACTAACC 1000
1001 CCGGGCCTGAAGAGGTTGACAAGAAGGGTCAAGGTTTCGTCTGTGTGTTGAAGGGCAGGGGCTGTTGGGTGCATCTGGAACGGCTTACCTCGGGTAACTG 1100
1101 TTTGCCATTAAAAGGTTGGGGATTAGGTTTAGCCCCTTTAGCTGCCATTTCGAACCGGGGTGTGCAGATGCAGGTCTCCGGGTGGGCAGGCAGTACGAGA 1200
1201 TGTCACGTTGTGTTGTCTTTCCTCCCACCCCTGTCCTGGCTGTGGCAAATGCGACCCTCATAGAGTTGTGTTTCAGGTCTGTGTCCTGTTTTGCGGTGGG 1300
1301 TTATTTCTTCCCTCAGTGTTTGCCAGCTTATTTCCCCAGTTTTCACGTACTGGGGCCTGTGGACACCTGAGGGAGCGGCCGTTGGTGGGTATGTGTTGGA 1400
1401 ATTGCTCCCACCCTCAATTTTCGGCTTGCCTTCTTCCCTTGTTAACCTGATAGCATAGCCTCTAGGTTTCCTTGTAGGTCTGTTTGGGTTTGTTGGTTCAC 1500
1501 GTGGTGCTAACTTGAATTTTTTGGTTTTCTAGTTCCCTCTTAATTACATTTGTGCCAGATCTTGTAGAGCAAGATGGCCTATTCAACAAGGGAGATACTG 1600
                                                                                  M  A  Y  S  T  R  E  I  L
1601 TTAGCCCTGTGTATACGGGACAGTCGTGTGCATGGAAATGGTACCCTGCATCCTGTGTTGGAGCTAGCAGCAAGAGAAACACCTCTCCGCCTTTCGCCAG 1700
     L  A  L  C  I  R  D  S  R  V  H  G  N  G  T  L  H  P  V  L  E  L  A  A  R  E  T  P  L  R  L  S  P  E
1701 AGGACACTGTAGTTCTGCGTTATCATGTGTTGCTTGAGGAGATAATTGAACGAAATTCAGAGACATTTACAGAAACTTGGAACAGATTTATAACACACAC 1800
       D  T  V  V  L  R  Y  H  V  L  L  E  E  I  I  E  R  N  S  E  T  F  T  E  T  W  N  R  F  I  T  H  T
1801 CGAACATGTGGATCTGGATTTTAACTCAGTATTTTTAGAGATATTTCACCGTGGAGACCCAAGCCTTGGGCGCGCGTTGGCCTGGATGGCCTGGTGCATG 1900
       E  H  V  D  L  D  F  N  S  V  F  L  E  I  F  H  R  G  D  P  S  L  G  R  A  L  A  W  M  A  W  C  M
1901 CATGCCTGCAGGACATTGTGTTGTAACCAGTCTACTCCTTACTATGTTGTGGACCTGTCAGTTCGTGGGATGTTAGAAGCCAGCGAAGGCCTGGATGGTT 2000
       H  A  C  R  T  L  C  C  N  Q  S  T  P  Y  Y  V  V  D  L  S  V  R  G  M  L  E  A  S  E  G  L  D  G  W
2001 GGATTCATCAACAGGGCGGCTGGTCTACATTAATTGAAGACAACATTCCTGGATCCAGAAGGTTTAGCTGGACTTTGTTTCTTGCTGGACTGACTTTGAG 2100
       I  H  Q  Q  G  G  W  S  T  L  I  E  D  N  I  P  G  S  R  R  F  S  W  T  L  F  L  A  G  L  T  L  S
2101 TCTGTTAGTTATATGTAGTTATTTATTTATCTCCAGAGGAAGACACTAATCTATACATTTTCTCAGCACTTTATATGAATCAGGGTCATTGGGCCTGCGG 2200
       L  L  V  I  C  S  Y  L  F  I  S  R  G  R  H
2201 GGAACTGAGCCAGTAGGATATTAGGCAAGGGTGACACAGTGCCCATGCATTATAATTTAACCAAACAGTGGTCGTGAGTTTTAGGCCGGCCATGGGGGCT 2300
2301 TACAAGAATAACATGCCAATGACCCGGCCCCCACTTTTAAATTCTGTTGCAGCAGATAGCTGATACCCAATGTTATCTTTTGCGGCAGAAATTGAAAGTG 2400
2401 CTGGCCATATCTACAATTGGGTGTCCTAGGTGGGATATACGCCTGTGGTGTTCTAACGGGAAGTGTGTAAGCACACACGTAATTTGCAAGCGGTGCTTCA 2500
2501 CGCTCTTCGTTAAAATAACACAAGGACAAGATACTAAAGAAATAACTGAGGTGAGTGTGGGAAGATGGGAATACTATGTGTTATGTTAACGGGTGAGAGC 2600
2601 CTATACTGCAGCCCAGACTCGGGGGGAGGAGGAAATGGTAAGAGTTATACTCTACTTATCTTTTTTGACACTACATTTAACTGTTATGTAACAATGTTTG 2700
2701 CTTATTTTCATGTTCAATAAACGCTATGTTAATGATGAAAAAAAAAA 2747
```

FIG. 4. DNA sequence of the cDNA clone H/F4 and translation of BHRF1. The splice site to an exon of the *Bam*HI Y region described by Bodescot and Perricaudet (4) for a BHRF1 cDNA clone is underlined, and the polyadenylation signal is overlined. The start of the shortest cDNA clone containing BHRF1 is indicated by an arrowhead. The potential N-linked glycosylation sites in BHRF1 are labeled with squares. The standard single-letter abbreviations are used for amino acids.

philic amino terminus is a stretch of 8 hydrophobic amino acids which represent the apolar core sequence found in all signal peptides; the processing site could be between residues 19 and 20. This would be in agreement with the rules of von Heijne (40), which require the signal peptide to be about 20 to 30 amino acids long and the residues −3 and −1 to be small (serine and valine in the case of BHRF1). Residues −2 and +1 for many eucaryotic signal peptides are often occupied by amino acids with a positive charge or with a long side chain or both. Arginine and histidine at these positions in BHRF1 are a good match for these requirements.

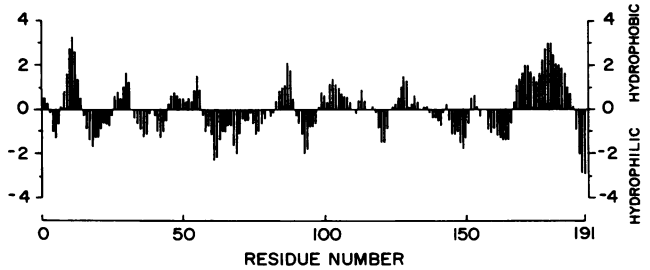

FIG. 5. Hydropathicity plot of the BHRF1 protein according to Kyte and Doolittle (21). The windowing average at residue *i* is calculated across 6 residues from *i* − 3 through and including *i* + 3.

Sequence analysis of cDNA clones containing BHRF1. The six clones H1, H2, H3, H4, H5, and H6 were sequenced as described in Materials and Methods. All the cDNA clones appear to be generated from the same transcript containing the highly repetitive BHLF1 ORF. This RNA is transcribed in the opposite direction to that of the message containing BHRF1. All of the clones have poly(A)$^+$ sequences which are attached at bp 1453 of the *Bam*HI H sequence, 17 bp downstream from the consensus polyadenylation signal. From here, the clones extend into the region of the IR2 repeats. Only H6 spans the entire IR2 region; its 5′ end is at bp 3482 of *Bam*HI-H (Fig. 6). Analysis of the size of the repeat region in these cDNA clones by digestion with the

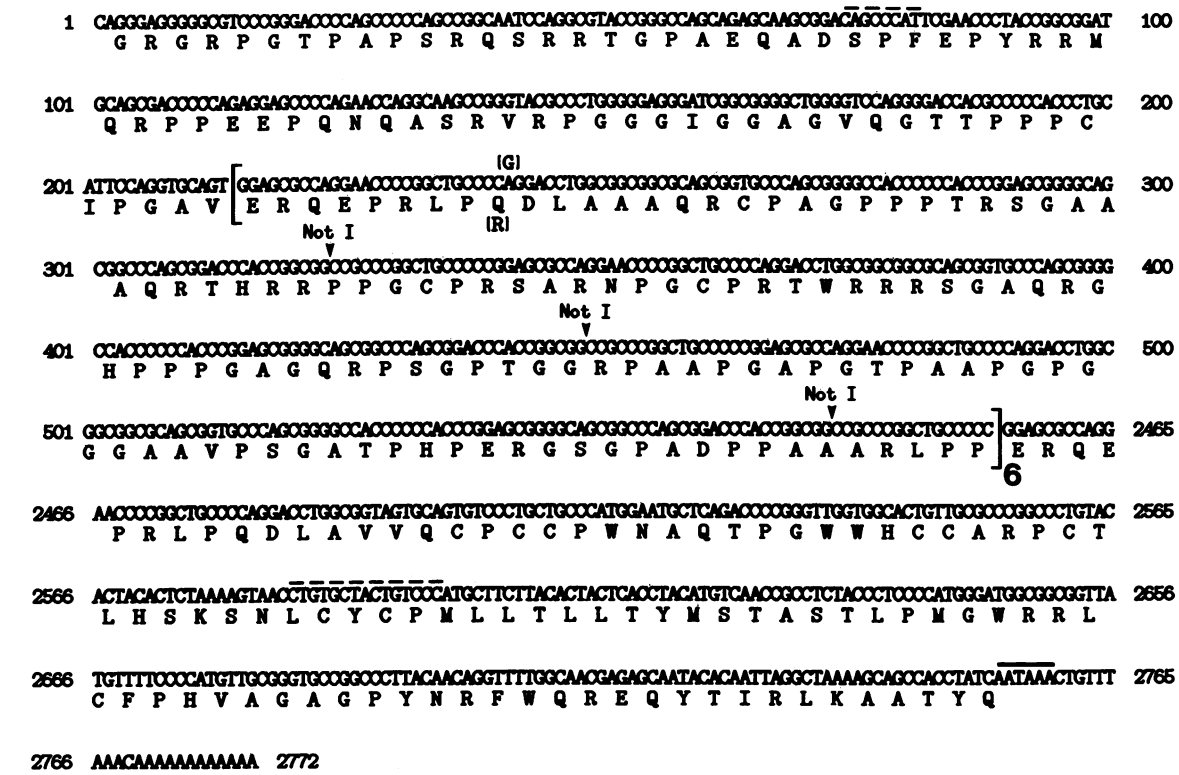

FIG. 6. DNA sequence of the longest BHLF1 cDNA clone H6. Only one ORF is translated, although all three reading frames are open. The structure of three repeat elements code for a repeated peptide is shown. The position of the one base pair change which was found in two repeat elements, resulting in the creation of new *Sma*I sites, is indicated. The polyadenylation signal is overlined. Other regions of the nucleotide sequence different than the B95-8 sequence are overlined with a broken line. The standard single-letter abbreviations are used for amino acids.

restriction endonucleases *Dde*I and *Eco*RI or partial digests with *Eag*I reveal that this area is about 750 bp (six repeat elements) longer in this JY clone than in the standard B95-8 genome. This result is not surprising, since it is well known (14) that different EBV strains are variable in the number of these repeat elements. H6 and the other BHLF1 cDNA clones are not completely identical. There is a single base pair change in two of the repeat elements present in clone 6.4, which creates two new *Sma*I sites (Fig. 6). This kind of base pair exchange in repeat elements even within the same EBV strain has been documented for other EBV isolates (14, 22). In general, however, there are few changes in the nucleotide sequence of these cDNA clones compared with the standard B95-8 genome. The two other examples of base changes found in the BHLF1 cDNA sequence are between *Bam*HI-H bp 1622 and 1640, where GGCGAAGTAGAC AGG is changed to GGGACAGTAGCACAGG, and between bp 3400 and 3410, where ATGGGCGTGGT is changed to ATGGGCTGT. Both mutations cause a frameshift but do not interrupt the protein sequence by creating a stop codon.

Identification of transcripts containing BHRF1 and BHLF1 by Northern blotting. The transcript containing BHLF1 has been well characterized by Northern blotting and S1 nuclease mapping (11, 14, 22). In most EBV-infected cell lines, this message is about 2,600 bp (depending on the number of IR2 repeat elements) and transcription is initiated downstream of a promoter located at bp 3400 of the *Bam*HI H fragment (18, 22). This transcript is inducible by the phorbol ester TPA and appears to be the most abundant early message in the lytic cycle of EBV (15). However, there is no published data on transcripts containing BHRF1. To identify transcripts which contain BHRF1, poly(A)$^+$ RNA from the JY cell line and total RNA from the productively infected

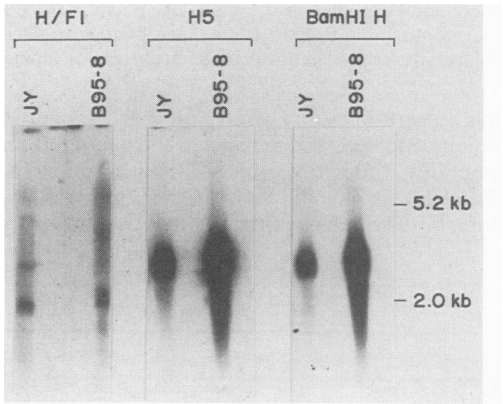

FIG. 7. Northern blot with RNA from the cell lines JY and B95-8. Poly(A)$^+$ RNA from JY or total RNA from B95-8 (10 μg per lane) was separated on a 1.5% formaldehyde agarose gel and transferred to nitrocellulose. The filters were probed with nick-translated cDNA clones H/F1 (containing BHRF1) or H5 (containing BHLF1) or with the *Bam*HI fragment. The positions and sizes of the two rRNA bands are indicated. Nitrocellulose strips with JY RNA were exposed four times longer than the strips with B95-8 RNA.

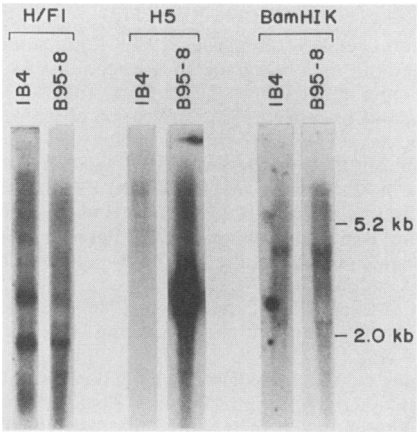

FIG. 8. Northern blot with RNA from the cell lines IB4 and B95-8. Poly(A)$^+$ RNA from IB4 or total RNA from B95-8 (10 μg per lane) was separated on a 1.5% formaldehyde agarose gel and transferred to nitrocellulose. The filters were probed with nick-translated cDNA clones H/F1 (containing BHRF1) or H5 (containing BHLF1) or with the *Bam*HI K fragment (containing the EBNA 1-coding region). The positions and sizes of the two ribosomal RNA bands are indicated. Nitrocellulose strips with JY RNA were exposed four times longer than the strips with B95-8 RNA.

B95-8 cell line were fractionated on formaldehyde-agarose gels and blotted and probed with appropriate cDNA clones containing either BHRF1 or BHLF1 (Fig. 7). Clone H/F1, which contains only the 3′ end of the BHRF1 transcript, including the coding region, hybridized to at least two different messages of about 1.95 and 2.7 kb and perhaps a third transcript of 4.5 kb. The smallest transcript is the most abundant message containing BHRF1 and represents about 75% of these transcripts in the JY and B95-8 cell lines.

The H5 cDNA clone (containing the 3′ end of the BHLF1 mRNA and most of the IR2 region) detected a reasonably high level of expression of this gene region in the JY cell line. The signal, however, was at least 200 to 300 times more intense for the RNA from B95-8 (there was 50-fold-less B95-8 mRNA than JY mRNA on these blots). These results indicate, however, that some transcription of early genes occurred in the JY cell line. Since the number of transcripts containing BHRF1 was much lower than the level of expression of the early transcript containing BHLF1, we decided to look for BHRF1 transcription during the latent cycle. Poly(A)$^+$ RNA from the tightly latent LCL IB4 was prepared. Northern blots of poly(A)$^+$ IB4 RNA were probed with BHRF1 and also with BHLF1 and the *Bam*HI K fragment (which encodes EBNA 1) as a control for early and latent transcription, respectively (Fig. 8). In contrast to the B95-8 and the JY cell lines, no BHLF1 RNA could be detected in IB4. The BHRF1 transcripts, however, were easily detectable, as was the EBNA 1 transcript. Of interest is that in the IB4 cell line, the 2.7-kb mRNA was relatively abundant compared with the 1.95-kb BHRF1 mRNA. However, it should be emphasized that the abundance of the transcripts encoding BHRF1 and BHLF1 was much greater in the productively infected marmoset cell line B95-8 than in either of the LCLs JY and IB4.

DISCUSSION

The IR2 region located within the *Bam*HI H fragment has been of great interest for two reasons. (i) The complete IR2 region is deleted in P3HR-1, and Northern blotting revealed that the BHLF1 transcript is missing (18). (ii) The structure of the predicted BHLF1 protein is very unusual, because the number of nucleotides (125 nucleotides) in each repeat element results in a frameshift after every *Not*I site. Thus, after three repeats, the same amino acid sequence begins again, and since all three reading frames in the IR2 region are open, a translation product consisting of the same 125-amino-acid repeat is generated regardless of which AUG is used for the initiation of translation. It is apparent from the cDNA clones in this study that the polyadenylation signal at bp 1470 in *Bam*HI-H is used for the transcripts containing BHLF1 and that there are minor heterogeneities in the sequence of the IR2 repeats, which leave the ORFs intact. Therefore, 90% of the putative BHLF1 polypeptide structure is unchanged in the JY EBV strain compared with that in B95-8, although two frameshift mutations upstream and downstream of the IR2 repeats were identified.

Several unsuccessful attempts have been made to identify this protein by hybrid selection and in vitro translation (16, 18, 22). To determine whether the hypothetical BHLF1 protein is expressed during the EBV life cycle, clones H1, H4, and H6 were cloned in frame to the *lac* promoter in pUC18 and induced with isopropyl-β-D-thiogalactopyranoside. With these constructs, we could not detect the expression of any EBV proteins by immunoblotting with the polyclonal antiserum WC (10). This approach was successful for the BMRF1-encoded protein (data not shown).

The other class of *Bam*HI-H-hybridizing cDNA clones isolated correspond to transcripts containing BHRF1. Northern blotting analysis with B95-8, JY, and IB4 RNA revealed that at least two and possibly three different mRNAs contain this ORF. Because of its size, 2.8 kb, clone H/F4 either is nearly full length or corresponds to the low-abundance 4.5-kb message. One obvious question concerns the structure of the 5′ end of these transcripts. Since there is no evidence for a promoter with the appropriate orientation in the *Bam*HI H or Y fragment, one hypothesis is that these transcripts are initiated at one of the promoters identified in the *Bam*HI C and W fragments (3). This assumption would implicate at least one splicing event, because the closest *Bam*HI W restriction fragment is about 7 kb upstream of BHRF1. Indeed, the recent characterization (4) of a cDNA clone isolated from the marmoset cell line B95-8 supports this hypothesis. In this study (4), clones corresponding to five different transcripts were isolated, all of which initiate downstream of a consensus eucaryotic promoter in the *Bam*HI C fragment. In addition, they all contain another exon from *Bam*HI-C and a variable number of exons from the *Bam*HI W and Y fragments. They are then spliced to coding regions in the *Bam*HI K (EBNA 1), E (thought to encode either EBNA 3 or EBNA 4), or the H (BHRF1) fragment. This BHRF1 transcript from B95-8 cells, however, is quite different from the mRNAs we have characterized here, because its 5′ end is spliced from an exon in *Bam*HI-Y to a splice junction only 41 bp upstream of the BHRF1 AUG (Fig. 4). However, all of the cDNA clones reported here contain sequences further in the *Bam*HI H fragment (the shortest contains 267 bp of additional 5′ sequences, and the longest contains 1,532 bp). Different splicing patterns are obviously used for BHRF1 mRNAs. Similarly, three structurally distinct cDNAs have been characterized for the EBNA 1 transcripts (4, 37). Since the BHRF1 cDNA clone from B95-8 cells (4) does not contain the complete 3′ end of the transcript, about 500 bp must be added to the cDNA clone in the previous study if use of the

same polyadenylation signal is assumed. Together with an average-sized poly(A)$^+$ tail, the size of this transcript is in good agreement with the 2.7-kb mRNA detected with BHRF1 on Northern blots.

Because several transcripts which utilize common exons from the IR1 region have been shown to code for EBV antigens associated with the latent viral life cycle (31, 36, 37), it has been proposed that long primary transcripts are generated that are spliced to give rise to mRNAs coding for these viral antigens. If this pattern of viral transcription is characteristic of latent transcripts, this would imply that BHRF1 is transcribed during the latent cycle. Northern blots of poly(A)$^+$ RNA from the tightly latent LCL IB4 indeed showed that all BHRF1 messages were present in this cell line, while the abundant 2.6-kb early message containing BHLF1 could not be detected. However, the BHRF1 transcripts were expressed at even higher levels in B95-8, a cell line with a relatively high level of virus production (in contrast to the transcripts encoding EBNAs, which are not transcribed at a higher level during the lytic cycle). In this respect, the BHRF1 product may resemble the EBV latent antigen LMP (latent membrane protein), which is expressed at relatively low levels in transformed cells (41) and can be induced by TPA or dilution of the cells (N. Müller-Lantzch, H. Boos, R. Berger, and C. Kuklik-Roos, 11th Int. Herpesvirus Meet., abstr. 97, 1986). These data indicate that during the lytic phase, transcriptional up-regulation of some transcripts present in the latent phase occurs.

Two independent pieces of information suggest that expression of BHRF1 may be important in the latent cycle of EBV. (i) Studies of viral transcription after infection of peripheral blood lymphocytes (32) showed that the first transcriptional activity observed is in the *Bam*HI W, Y, H, and F regions and in the *Bam*HI K fragment. Moreover, the signal for the *Bam*HI F fragment (which is the only unique index for BHRF1 expression, since W and Y exons are contained in all of the EBNA transcripts characterized to date and *Bam*HI-H encodes a portion of EBNA 2) is much stronger than the signal for *Bam*HI-K (EBNA 1). This is in good agreement with our Northern blotting data for poly(A)$^+$ RNA from the latently infected cell line IB4 (Fig. 8). (ii) Analysis of chromosomal rearrangements in human follicular lymphomas (7) revealed that the t(14;18) translocation, observed in greater than 90% of these lymphomas, joins the *bcl-2* gene to the immunoglobulin H gene. This rearrangement has been proposed to lead to an activation of the transcription of *bcl-2* analogous to that of the c-*myc* gene in Burkitt's lymphoma. A homology search revealed that the putative *bcl-2* gene product, a 26-kilodalton protein, shows homology to the BHRF-1 ORF. It is therefore conceivable that this EBV-encoded protein may serve a similar function in the growth transformation of B lymphocytes. In addition there is also a structural similarity between the *bcl-2* transcript and the BHRF1 mRNAs, in that both have very long 3′ untranslated regions, which could be involved in translational regulation of protein expression (6).

ACKNOWLEDGMENTS

We thank E. Kieff for the IB4 cell line.

This work was supported by Public Health Service grants 5P01CA21082, 5F32CA07147, and 1R01CA43143 (to S.H.S.) from the National Institutes of Health and a fellowship from the Deutsche Forschungsgemeinschaft (to A.J.P.). S.H.S. is a Special Fellow of the Leukemia Society of America.

LITERATURE CITED

1. **Auffray, C., and T. Rougeon.** 1980. Purification of mouse immunoglobin heavy chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
3. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. L. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
4. **Bodescot, M., and M. Perricaudet.** 1986. Epstein-Barr virus mRNA produced by alternative splicing. Nucleic Acids Res. **14**:7103–7113.
5. **Bornkamm, G. W., J. Hudewentz, U. K. Freese, and U. Zimber.** 1982. Deletion of the nontransforming Epstein-Barr virus strain P3HR-1 causes fusion of the large internal repeat to the DS_L region. J. Virol. **43**:952–968.
6. **Caput, D., B. Beutler, K. Hartog, R. Thayer, S. Brown-Shimer, and A. Cerami.** 1986. Identification of a common nucleotide sequence in the 3′-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA **83**:1670–1674.
7. **Cleary, M. L., S. D. Smith, and J. Sklar.** 1986. Cloning and structural analysis of cDNA for *bcl-2* and a hybrid *bcl-2*/immunoglobulin transcript resulting from the t(14;18) translocation. Cell **47**:19–28.
8. **Crouse, G. F., A. Froschauf, and T. Lehrach.** 1983. An integrated and simplified approach to cloning into plasmids and single-stranded phages. Methods Enzymol. **101**:78–79.
9. **Denhardt, D. T.** 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. **23**:641–646.
10. **Edson, C. M., L. K. Cohen, W. Henle, and J. L. Strominger.** 1983. An unusually high-titer human anti-Epstein Barr virus (EBV) serum and its use in the study of EBV-specific proteins synthesized in vitro and in vivo. J. Immunol. **130**:919–924.
11. **Freese, U. K., G. Laux, J. Hudewentz, E. Schwarz, and G. W. Bornkamm.** 1983. Two distant clusters of partially homologous small repeats of Epstein-Barr virus are transcribed upon induction of an abortive or lytic cycle of the virus. J. Virol. **48**: 731–743.
12. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
13. **Henderson, E., G. Miller, J. Robinson, and L. Heston.** 1977. Efficiency of transformation of lymphocytes by Epstein-Barr virus. Virology **76**:152–163.
14. **Hudewentz, J. H., H. Delius, U. K. Freese, U. Zimber, and G. W. Bornkamm.** 1982. Two distinct regions of the Epstein-Barr virus genome with sequence homologies have the same orientation and involve tandem repeats. EMBO J. **1**:21–26.
15. **Hummel, M., and E. Kieff.** 1982. Epstein-Barr virus RNA. VIII. Viral RNA in permissively infected B95-8 cells. J. Virol. **43**:262–272.
16. **Hummel, M., and E. Kieff.** 1982. Mapping of the polypeptides encoded by the Epstein-Barr virus genome in productive infection. Proc. Natl. Acad. Sci. USA **79**:5698–5702.
17. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in λ gt10 and λ gt11, p. 49–78. *In* D. Grover (ed.), Practical approaches in biochemistry. IRL, Oxford.
18. **Jeang, K.-T., and S. D. Hayward.** 1983. Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the *Not*I repeat units that form part of the template for an abundant 12-*O*-tetradecanoylphorbol-13-acetate-induced mRNA transcript. J. Virol. **48**:135–148.
19. **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.
20. **King, W., A. Thomas-Powell, N. Raab-Traub, M. Hawke, and E.**

Kieff. 1980. Epstein-Barr virus RNA. VI. Viral RNA in a restringently infected, growth-transformed cell line. J. Virol. **36**:506–518.

21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**: 105–132.
22. **Laux, G., U. K. Freese, and G. W. Bornkamm.** 1985. Structure and evolution of two related transcription units of Epstein-Barr virus carrying small tandem repeats. J. Virol. **56**:987–995.
23. **Lehrach, H., D. Daimond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
24. **Lin, H.-C., S.-P. Lei, and G. Wilcox.** 1985. An improved DNA sequencing strategy. Anal. Biochem. **147**:114–119.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Miller, G., and M. Lipman.** 1973. Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. Proc. Natl. Acad. Sci. USA **70**:190–194.
27. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
28. **Ploegh, H. L., H. T. Orr, and J. L. Strominger.** 1981. Major histocompitibility antigens: the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. Cell **24**:287–299.
29. **Rabson, M., L. Gradoville, L. Heston, and G. Miller.** 1982. Non-immortalizing P3J-HR-1 Epstein-Barr virus: a deletion mutant of its transforming parent, Jijoye. J. Virol. **44**:834–844.
30. **Rowe, D., L. Heston, J. Metlay, and G. Miller.** 1985. Identification and expression of a nuclear antigen from the genomic region of the Jijoye strain of Epstein-Barr virus that is missing in its nonimmortalizing deletion mutant, P3HR-1. Proc. Natl. Acad. Sci. USA **82**:7429–7433.
31. **Sample, J., M. Hummel, D. Braun, M. Birkenbach, and E. Kieff.** 1986. Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a probable transcriptional initiation site. Proc. Natl. Acad. Sci. USA **83**:5096–5100.
32. **Sample, J., A. Tanaka, G. Lancz, and M. Nonoyama.** 1984. Identification of Epstein-Barr virus genes expressed during the early phase of virus replication and during lymphocyte immortalization. Virology **139**:1–10.
33. **Sanger, F., S. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
34. **Skare, J., J. Farley, J. L. Strominger, K. O. Fresen, M. S. Cho, and H. zur Hausen.** 1985. Transformation by Epstein-Barr virus requires DNA sequences in the region of *Bam*HI fragments Y and H. J. Virol. **55**:286–297.
35. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
36. **Speck, S. H., A. Pfitzner, and J. L. Strominger.** 1986. An Epstein-Barr virus transcript from a latently infected/growth transformed B-cell line encodes a highly repetitive polypeptide. Proc. Natl. Acad. Sci. USA **83**:9298–9302.
37. **Speck, S. H., and J. L. Strominger.** 1985. Analysis of the transcript encoding the latent Epstein-Barr virus nuclear antigen I. A potentially polycistronic message generated by long-range splicing of several exons. Proc. Natl. Acad. Sci. USA **82**:8305–8309.
38. **Terhorst, C., P. Parkam, D. L. Mann, and J. L. Strominger.** 1976. Structure of HLA antigens: amino-acid and carbohydrate composition and NH_2-terminal sequences of four antigen preparations. Proc. Natl. Acad. Sci. USA **73**:910–914.
39. **Van Santen, V., A. Cheung, and E. Kieff.** 1981. Epstein-Barr virus RNA. VII. Size and direction of transcription of virus-specified cytoplasmic RNAs in a transformed cell line. Proc. Natl. Acad. Sci. USA **78**:1930–1934.
40. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**:4683–4690.
41. **Wang, D., D. Liebowitz, and E. Kieff.** 1958. An EBV membrane protein expressed in immortalized lymphocytes transforms established rodent cells. Cell **43**:831–840.
42. **Weigel, R., and G. Miller.** 1983. EB virus-specific cytoplasmic transcripts in a cellular clone of the HR-1 Burkitt lymphoma line during latency and after induction of viral replicative cycle by phorbol esters. Virology **125**:287–298.