JOURNAL OF VIROLOGY, June 1993, p. 3217–3225
0022-538X/93/063217-09$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 67, No. 6

Complex Nature of the Major Viral Polyadenylated Transcripts in Epstein-Barr Virus-Associated Tumors

PAUL R. SMITH,* YANNING GAO, LORAINE KARRAN, MICHAEL D. JONES, DEE SNUDDEN, AND BEVERLY E. GRIFFIN

Department of Virology, Royal Postgraduate Medical School, Du Cane Road, London W12 0NN, England

Received 28 December 1992/Accepted 7 March 1993

The most abundant polyadenylated viral transcripts in the Epstein-Barr virus (EBV)-associated tumor nasopharyngeal carcinoma are a family (apparent sizes, 4.8, 5.2, 6.2, and 7.0 kb) of highly spliced cytoplasmic RNAs expressed from the *Bam*HI-I and -A regions of the viral genome in an antisense direction with respect to several viral lytic functions encoded within the same region and concerned with the lytic cycle of the virus. We have called these complementary-strand transcripts. They are also expressed in B cells, including Burkitt's lymphoma and EBV-immortalized marmoset cell lines, and tumors generated in cottontop tamarins in response to EBV infection, but at a lower level. The complete structure of the major 4.8-kb RNAs (seven or eight exons) was determined in this study; the larger, but related, transcripts appear to be produced by differential splicing. The transcriptional promoter for the major complementary-strand transcripts, located in *Bam*HI-I, contains several well-characterized transcriptional control elements (E2A, SP1, and AP1) and is functionally active in both B lymphocytes and epithelial cells. It appears to be a bifunctional viral promoter, as it also contains the initiation codon for a gene (BILF2) that encodes a glycoprotein that is expressed off the other strand. Splicing events create a number of small AUG-initiated open reading frames, one of which has homology to functionally significant regions of the EBV-encoded nuclear antigen 2 and to E2 (in papillomavirus). The complex nature of these transcripts and their potential role in the virus association with malignancy are considered.

Epstein-Barr virus (EBV) is a causative agent of infectious mononucleosis and is implicated in the etiology of several distinct malignant tumors, primarily Burkitt's lymphoma (BL) (reviewed in reference 35) and anaplastic nasopharyngeal carcinoma (NPC) (44). There also appears to be a viral association with a subset of Hodgkin's lymphomas (27) and AIDS-related lymphomas (37). The role of EBV in the development of tumors is, however, based largely on the presence of viral DNA in tumor cells—and with NPC the correlation, originally made by Old et al. (38), is 100%—and is reinforced, in part, by the fact that EBV can immortalize B lymphocytes in vitro to produce continuously proliferating lymphoblastoid cell lines (LCLs). Analyses of the protein products in virus nonproducer BL lines and LCLs has allowed the identification of a number of EBV gene products, described as latent functions, that are associated with the establishment and/or maintenance of immortalized B cells in vitro (12). Viral genes expressed in latently infected LCLs were previously thought to be restricted to six EBNAs and three putative membrane proteins (latent membrane protein [LMP] and terminal proteins 1 and 2 [alternatively called LMP2a and LMP2b]), as well as two small, nonpolyadenylated RNA untranslated species, EBER1 and EBER2, related to the virus-associated RNAs of adenovirus (3). In contrast to this pattern, in BLs only one viral antigen, EBNA-1, has been detected and in NPC cells, only EBNA-1, and LMP (in approximately 60% of cases) have been reported to be consistently expressed (14, 41). Although the pattern of expression of the latent proteins has been well documented, little is known of the functions of most of them. EBNA-1 has been identified as necessary for maintenance of the EBV episome (43, 53) and is also thought to act as an enhancer for transcription under certain conditions (42, 48). EBNA-2 can act as a transcriptional transactivator and is strongly implicated in cellular immortalization of B cells (1, 11, 26). Expression of LMP has been shown to transform established rodent fibroblasts (50) and human keratinocytes (15) in culture, a function shared with another viral protein, the 33-kDa species expressed from the BARF1 open reading frame (52). We recently reported that the antisense transcript originally identified in NPCs, and discussed below, is also expressed in LCLs and BL-derived cell lines (9, 30).

Detailed analysis of EBV gene expression in epithelial cells has been hampered by the lack of representative cell lines. However, a number of human NPC tumors established recently in nude mice (6) retained characteristics typical of the original tumor and could be used to study viral gene expression. Analysis of a λgt10 cDNA recombinant library from one of these tumors (C15) produced a totally unexpected result (29). That is, the major EBV polyadenylated RNA species were found to be transcribed from a region of the viral genome (*Bam*HI-A) which had not previously been identified as transcriptionally active in virally associated B cells, nor was its existence predicted by analysis of putative open reading frames in the genomic sequence of EBV (4, 17). The reason for this oversight is that these transcripts were derived from the strand of EBV DNA complementary to genes already mapped to the same region. These results were confirmed in other studies (9, 21, 30). More recent evidence has shown that the major RNA species of approximately 4.8 kb, with associated bands at about 4.2, 5.5, 6.2, and 7.0 kb, not only is expressed in NPC tumors propagated in nude mice but can also be detected by Northern (RNA) blotting in primary biopsy material from Chinese NPCs (9, 30) and has been identified in B lymphomas generated in cottontop tamarins in response to EBV infection (54). Anal-

* Corresponding author.

TABLE 1. Sequences of oligonucleotides used in this study[a]

Oligonucleotide	Sequence	EBV genome coordinate	Exon-intron location[b]
1	CGAAGAGGCTAGTTGCCTACGT	150743	I
2	CGCGAATTCAGCTCAGTGACACGT	6524*	II
3	GTCATACGCCCGTATTCACA	9942*	II-IIIA
4	ATTCAGCTGACACGCTCCT	10086*	III
5	CGCGAATTCCTCAAGCCCTTCTTCGT	10253*	III-IIIB
6	TATTGCAGCTGGACGCGCAGT	10413*	Intron
7	GTTGAGGTCTACGATTC	155688	Intron
8	CATAGAATTCCGCTATAGGCGCATCCTGCT	155737	V
9	GGCTGGTACGCGGACTCC	158975	Intron
10	GTATGGCTGTTGTTGC	150623	5′ to start of exon I[c]
λF	AGCAAGTTCAGCCTGGTTAAG		
λA	CTTATGAGTATTTCTTCCAGGGTA		

[a] EBV coordinates refer to the positions of 3′ bases with respect to EBV strain B95-8 (4), except for oligonucleotides marked with asterisks, which refer to their positions in the Raji sequence, deleted in B95-8 (39). The sequences can be found in the GenBank-EMBL data base under accession numbers V01555 and M35547.
[b] See Table 2.
[c] See Fig. 2.

ysis of cDNA clones that encode these complementary RNAs has identified a number of alternative splicing patterns in the extreme 3′ region of the transcripts (9, 21, 54), indicative of a complicated mechanism for maturation from an initial large primary RNA transcript(s). A role has been suggested for these complementary-strand transcripts (CSTs) in suppressing expression of genes on the other strand of EBV, thereby possibly acting functionally as antisense controlling elements (30). There is no evidence that proteins are translated from these RNAs, although a polypeptide generated in vitro from a cDNA clone at the 3′ end of the transcripts can be expressed in vitro (31) and can be immunoprecipitated with NPC sera (22). No cDNA clones yet described have represented full-length RNA sequences.

In this report, we describe the isolation and analysis of overlapping cDNA clones which identify the structure of the major EBV-encoded 4.8-kb CST as a highly spliced, heterogeneous species containing a number of small open reading frames. Complex splicing patterns produce several exons with AUG initiator codons capable of being translated. One of these has homology with a functionally important region of EBNA-2 expressed from the BYRF1 open reading frame in another part of the genome. It is possibly of significance that most exons in the spliced RNA complement exon sequences from previously identified open reading frames on the other strand of the genome (17). We also identify the promoter region for the major RNA species and show that it contains a number of well-known transcriptional response elements, is active in both epithelial cells and B lymphocytes, and from its structure, appears capable of acting in a bidirectional manner.

MATERIALS AND METHODS

Cell culture. The C15 tumor was passaged in nude mice as previously described (6, 46). Hep-2 cells, derived from a human carcinoma of the larynx, were cultured in Dulbecco modified Eagle medium supplemented with 5% fetal calf serum and 5 mM glutamine at 37°C in a 10% carbon dioxide atmosphere. The B-lymphoblastoid cell lines B95-8, an EBV-positive marmoset line, and RAMOS, an EBV-negative Burkitt's lymphoma line, were cultured in RPMI 1640 supplemented with 10% fetal calf serum and 5 mM glutamine at 37°C in a 5% carbon dioxide atmosphere.

cDNA isolation and analysis. Clones described here were isolated from the C15 λgt10 cDNA library described by Hitt et al. (29) by using the polymerase chain reaction (PCR) to amplify sequences directly from the library. Oligonucleotides 1 and 2 (Table 1) were used in conjunction with lambda-specific oligonucleotides (Table 1) to amplify specific sequences as described by Friedman et al. (19). Briefly, 1 μl of the amplified cDNA library was diluted 1:100 with distilled water and heated at 70°C for 5 min. One-microliter aliquots of this material were then used in a standard PCR as described previously (46). Following amplification, positive samples were identified by Southern blotting and cloned into Bluescript vectors. For DNA analysis, EBV-positive clones were cloned into M13 vectors and sequenced by the dideoxy-chain termination method.

RACE analysis. To confirm the position of the 5′ end of the transcripts, the RACE protocol described by Frohman et al. (20) was used. One microgram of C15 mRNA was primed with oligonucleotide 3 or 8 (Table 1) and reverse transcribed with avian myeloblastosis virus reverse transcriptase. Following tailing with terminal transferase, the cDNA was subjected to amplification by PCR with oligonucleotides 1 and 2 (Table 1). Positive bands were identified by Southern blotting and subcloned and sequenced as described above.

Northern (RNA) blotting. Extraction and purification of polyadenylated RNA was performed as described previously (29). Probing of Northern blots containing 5 μg of poly(A)$^+$ RNA with ^{32}P-labelled oligonucleotides (Table 1) was performed as described by Karran et al. (30).

RNase protection assay. A subclone of the *Bam*HI I fragment of the B95-8 strain of EBV DNA was constructed by cloning the *Bgl*II (position 150461 in the B95-8 genome; 4)-*Ssp*I (position 151010) fragment into an *Sma*I-*Bam*HI-cut Bluescript vector. Single-stranded riboprobes were synthesized in both orientations, gel purified, and hybridized overnight at 45°C to 1 μg of polyadenylated or 10 μg of total C15 RNA or *Escherichia coli* tRNA. RNase digestion and subsequent steps were performed as described previously (46).

Analysis of DNA methylation. Total cellular DNA was digested with either *Hpa*II or its methylation-insensitive

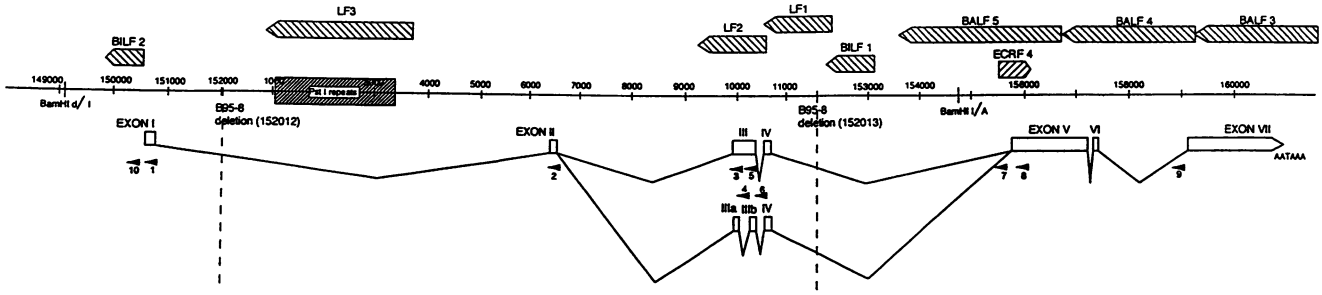


FIG. 1. Diagram showing the position on the genome and structure of the major EBV 4.8-kb mRNA in NPC cells. The rightward complementary transcripts are shown by open boxes. The coordinates refer to the numbering of the EBV genome (4, 39). The approximate positions of the oligonucleotides used for Northern blots or RACE protocols (Table 1) are shown by arrowheads. Previously described open reading frames are identified by cross-hatched boxes, and the directions of transcription are shown by arrow points. Functions for a number of these open reading frames have been proposed as follows: BALF3, possible glycoprotein transport protein; BALF4, homolog of herpes simplex virus glycoprotein B; BALF5, EBV DNA polymerase; BILF2, putative glycoprotein. ECRF4, identified as an open reading frame with an unknown function in the viral DNA (5), is truncated by splicing in the 4.8-kb RNA. Little is known of the functions of other open reading frames shown. An alternative oriLyt lies within the deletion found in B95-8 DNA (25).

isoschizomer *Msp*I. In each case, 10 μg of total cellular DNA was digested overnight at 37°C with 10 U of enzyme per μg of DNA. The digested DNA was separated on a 1.2% agarose gel, transferred to nitrocellulose membranes, and probed with subfragments from the B95-8 *Bam*HI I fragment. Probes were generated from the B95-8 *Bam*HI-I region by digestion with *Cla*I-*Sph*I, *Sph*I-*Bgl*II, and *Bgl*II-*Ssp*I (see Fig. 5).

Chloramphenicol acetyltransferase (CAT) assays. Vectors used to assess the efficiency of the promoter region were produced by PCR amplification of a fragment from positions 149930 to 150623 with engineered *Hin*dIII and *Sal*I sites. The resulting fragment was cloned into a Bluescript plasmid (Promega), and subfragments were generated with *Sph*I (EBV genome position 150200) and *Pst*I (position 150408) sites present in this region (4). These fragments were cloned into a pCAT BASIC plasmid (Promega) to produce the A422CAT and A216CAT plasmids, respectively.

A422CAT and A216CAT constructs were transfected into cells by electroporation. Briefly, approximately 5×10^6 LCLs or Hep-2 cells from a confluent 75-cm^2 flask were incubated on ice for 10 min with 20 μg of plasmid DNA. They were then electroporated with a Bio-Rad gene pulser at 960 μF with either 300 V (LCLs) or 250 V (Hep-2 cells), which typically resulted in a decay constant of 20 ms. Following electroporation, cells were incubated on ice for a further 10 min before medium was added and then incubated for either 24 or 48 h before being assayed for CAT activity. CAT activity was assessed by either thin-layer chromatography or liquid scintillation counting. The CAT activity of the extracts was semiquantitated by comparison to a reference curve constructed with the purified CAT enzyme and adjusted for the protein concentrations of the various extracts. The results were obtained from three separate transfection experiments, and each experiment was assayed in duplicate.

Protein homology. Potential protein homology was assessed with the FASTA program by screening against the EMBL protein sequence data bank. Homology with only a limited number of known functions was identified.

RESULTS

Structures of the major EBV polyadenylated transcripts in NPC. Two large, overlapping EBV cDNA clones from a λgt10 NPC-derived cDNA library were found previously to represent approximately 3.9 kb of the major 4.8-kb polyadenylated cytoplasmic RNA in C15 NPC (29, 30). The sequences of these cDNA clones, located in the EBV *Bam*HI A fragment, were used to identify oligonucleotides amplifying 5′ upstream sequences in the library. This strategy was employed to examine both splicing patterns and 5′ promoter sequences in the NPC transcripts. The data obtained by this procedure, and the oligonucleotides used, are shown schematically in Fig. 1 (also Tables 1 and 2). Analysis of clones produced following amplification by oligonucleotides 8 and λF identified two alternate splicing patterns for exons III and IV. When oligonucleotides 3 and λF (Table 1) were used to amplify sequences from the cDNA library by PCR, subsequent analysis of EBV-positive clones identified approximately 400 bp that corresponded to a number of exons located within the *Bam*HI I fragment. Similar amplification experiments with oligonucleotides 1 and 2 (Table 1) extended the sequences upstream, and the largest clones terminated at B95-8 EBV genome coordinate position 150640. This sequence, in *Bam*HI-I, is 69 bp downstream of a TAAATAT sequence, whose complement (ATATTTA) has previously been described as the initiation sequence for the BILF2 open reading frame (17), and 35 bp downstream of a TTTCATATT sequence (Fig. 2). Although not a perfect classical TATA promoter sequence, the latter is similar to the TACATAA sequence identified as the initiation sequence for transcription of LMP mRNA (17).

To identify the precise start site of the major NPC tran-

TABLE 2. Coordinates of exons of the 4.8-kb mRNA shown in Fig. 1

Exon	Size (bases)	EBV genome coordinates
I	134	150640–150774
II	101	6514–6615
III	497	9861–10358
IIIA	131	9861–9992
IIIB	154	10204–10358
IV	112	10517–10629
V	1,465	155730–157195
VI	77	157304–157381
VII	1,908	159083–160991

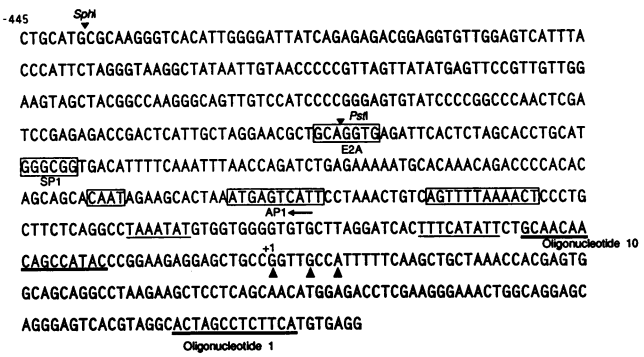


FIG. 2. Sequence of exon I and the upstream promoter region to position −445 (Fig. 1). Arrowheads show the positions of the 5′ ends of the RACE clones obtained with oligonucleotides 3 and 8 (Table 1); +1, the start site of the 4.8-kb RNA. Potential initiator sequences are underlined. Putative transcription factor-binding motifs for E2A, SP1, and AP1 (16) are indicated, as is a 12-base palindromic sequence (boxed). Positions of the restriction enzyme sites (*Sph*I and *Pst*I) used to generate CAT plasmids in conjunction with the *Sal*I site engineered into oligonucleotide 10 are indicated. Putative TATA initiation sequences are underlined. Oligonucleotides 1 and 10 are doubly underlined. The AUG initiator codon for BILF2 lies within the AP1 motif, implying that the promoter shown here could be functional in both polarities.

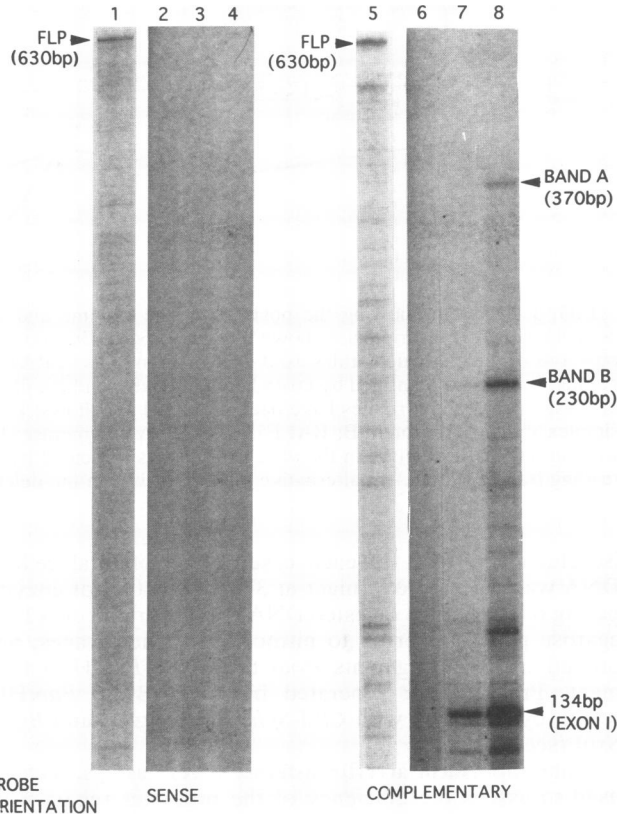


FIG. 3. RNase protection of C15 polyadenylated RNA. Lanes: 1 and 5, probe (630 bp) alone; 2 and 6, probe plus *E. coli* tRNA; 3 and 7, probe plus C15 poly(A)$^+$ (selected); 4 and 8, probe plus total C15 RNA. Lanes 5 to 8 show hybridization with an antisense probe complementary to the 4.8-kb RNA. Lanes 1 to 4 show hybridization with a sense probe colinear with the 4.8-kb RNA. The position of the largest protected fragment (370 bp) is shown, as are those of the products of incomplete protection of RNA by the probe. FLP, full-length probe.

scripts, the RACE protocol (20) was used. Sequencing of RACE clones from cDNAs initiated from both oligonucleotides 3 and 8 confirmed a transcription start site at position 150640 in both cases. The precise 5′ locations of the RACE clones are given in Fig. 2. A number of potential transcription factor-binding sites exist upstream of sequences identified in the putative promoter region, as indicated. RNase protection assays showed that a fragment of approximately 134 bp, covering exon I, was insensitive to digestion when hybridized to C15 polyadenylated RNA (Fig. 3, lane 7). This would position the start site of exon I, in *Bam*HI-I, at position 150640, in excellent agreement with data obtained from the cDNA library and from analysis of RACE clones. This was confirmed by hybridization to total C15 RNA (Fig. 3, lane 8), which besides protecting the 134-bp exon 1 band, also protected a band of approximately 370 bp, positioning the start site at 150640. The actual major transcript size determined by these methods is 4.3 kb, compared with the 4.8 kb estimated with Northern blots with marker RNAs.

To confirm that the exon sequences identified were actually present in the 4.8-kb RNA, a number of oligonucleotides (Fig. 1) were used to probe polyadenylated RNA on Northern blots. (Note that in our experience, short oligonucleotides are not labelled to the same specific activities, and thus data obtained by this technique can only be a qualitative, not a quantitative, measurement of messenger concentration; 30.) Our data (Fig. 4) confirmed that exons identified by analysis of the cDNA clones are present in the 4.8-kb species. In addition, a number of oligonucleotides (notably, 3, 4, and 8) recognized the larger, but less abundant, 5.5- and 6.2-kb species, suggesting that these RNAs contain sequences that overlap the 4.8-kb transcripts. Probing of Northern blots with oligonucleotide 4, which is present in only one of the two alternately spliced RNA species (Fig. 1), identified the 4.8-kb species, showing that the larger (as well as the smaller) form of the two different spliced clones (Fig. 1) is present in the major transcript. Probing of Northern blots with some oligonucleotides (for example, oligonucleotide 3 [Fig. 4]) emphasized the broadness of the band at 4.8 kb, suggesting that it consists of a number of differently spliced species of similar sizes. Two of these have been identified here (Fig. 1), and another, with a splice in exon VII, has been identified elsewhere (9). The small message seen with oligonucleotide 6 has not been further characterized. Oligonucleotide 2, in exon 2 (Table 1), reacts only weakly with the 4.8-kb species and is not shown in Fig. 4. In addition, oligonucleotide 9 (and 7 on a longer exposure of the autoradiograph) recognized only the 5.5- and 6.2-kb bands, suggesting that these bands are produced by alternate splicing of a large primary transcript that can be further processed to produce the 4.8-kb mRNA species. Oligonucleotide 10, lying 5′ to the ends of mapped transcripts (Fig. 1), gave a negative result. Detailed mapping of these larger transcripts is in progress.

Analysis of the promoter region for the 4.8-kb RNA family. Digestion of EBV DNA with methylation-sensitive and -insensitive restriction enzyme isoschizomers *Hpa*II and *Msp*I, respectively, has been shown previously to indicate actively transcribed EBV promoter regions (2, 13). The results of an analysis of this type suggest that most of the CpG sites in the EBV episomes in C15 DNA are methylated over the genomic region that includes the 5′ end of the 4.8-kb

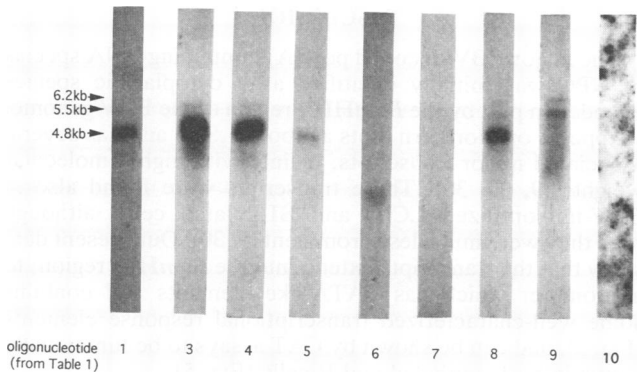
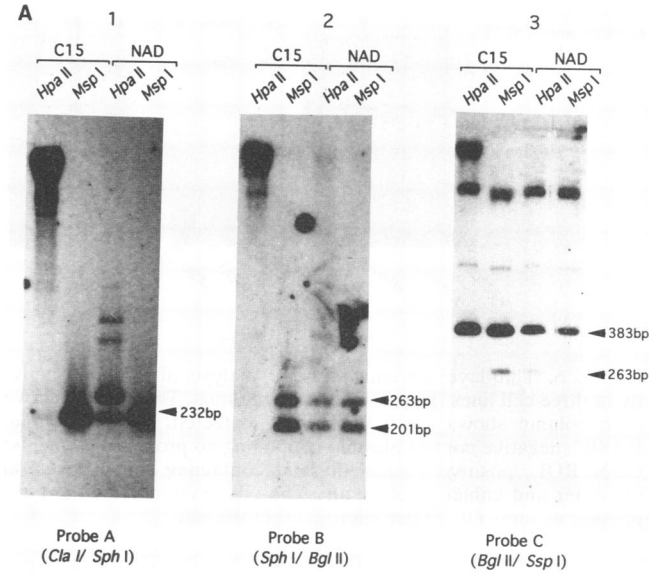


FIG. 4. Northern blots of C15 polyadenylated RNA probed with specific oligonucleotides (Table 1). The sizes of the main bands are indicated on the left. Note that oligonucleotide 9 identifies only the larger of the RNA species. The identity of the small band recognized by oligonucleotide 6 is not known. The blot shown with oligonucleotide 10 was deliberately exposed for a long time to search for any weakly hybridizing band. None was identified.



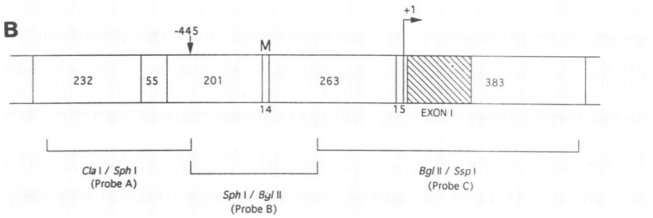


FIG. 5. (A) Methylation of the promoter region between EBV nucleotides 149924 and 151003 as assessed by *Hpa*II-*Msp*I restriction digestion and Southern blotting. The positions of the four bands (201, 232, 263, and 383 bp) obtained by complete *Msp*I digestion and recognized by the three probes used are indicated. (B) Diagram showing the expected restriction pattern for nonmethylated DNA. The *Hpa*II-*Msp*I sites are represented by open boxes. The position of the 4.8-kb mRNA start site is shown by the bent arrow above the diagram, and exon I is represented by the hatched box. The site(s) completely insensitive to *Hpa*II digestion and therefore methylated in all episomes is shown by the letter M above the diagram. The positions of the probes used for panel A are shown below the diagram.

RNA (probes A, B, and C, Fig. 5), as shown by the presence of a large (>10-kb) band in the *Hpa*II digest of C15 DNA. (A similar result, showing extensive methylation in the tumor, was obtained when C15 DNA was probed with fragments from the B95-8 *Bam*HI-A and Raji *Bam*HI-Ic regions [data not shown; 45]). However, although most of the CpG sites (within CCGG nucleotides) of EBV episomes appear to be methylated throughout the *Bam*HI-A and -I regions, some *Hpa*II sites in the sequence that contains the 5′ ends of the transcripts are available for restriction, as shown by the presence of bands at 232 and 383 bp obtained with probes A and C (Fig. 5A, panels 1 and 3, respectively). The absence in C15 DNA of bands at 201 and 263 bp when probe B was used (panel 2) and at 263 bp when probe C was used (panel 3) implies that *Hpa*II sites upstream of this (indicated by the letter M) are completely methylated. As a control for these experiments, the methylation patterns in NAD, an LCL produced with virus from C15, were also examined. NAD is not as restricted in expression as the tumor (29), and this is reflected in the digestion patterns obtained, although interestingly, the level of transcription of the 4.8-kb RNA in the B cell is considerably below that seen in the C15 tumor (30).

The presence of a high level of transcription initiated from the *Bam*HI-I region, from a limited number of EBV episomes (29), suggests that this region contains a strong promoter; alternatively, the transcripts may be remarkably stable. The ability of the 5′ upstream region of the transcripts identified to function as a promoter and its efficiency were assessed by CAT assays with constructions covering two regions of the promoter (Fig. 2). The results in Fig. 6 show that the promoter identified for the 4.8-kb RNA transcripts functions in both LCLs (B95-8 and Ramos) and epithelial cell lines (Hep-2). Quantitation of the levels of CAT activity in these different types of cells suggests that the promoter induces similar levels of CAT production in both types of cells in vitro and, as shown with data on the EBV-negative Ramos BL line, does not depend on the presence of EBV DNA. Thus, it is apparently independent of expression of other EBV genes as it functions in a similar manner in both EBV-positive and EBV-negative LCLs. However, since the level of transcription observed in vivo seems much higher than that observed with CAT assays, we must consider that elements acting as enhancers may exist either upstream or downstream of the mapped promoter that can attenuate activity. Alternatively, in NPCs, mutations that have not been recognized and affect promoter activity may exist.

Identification of open reading frames in the 4.8-kb polyadenylated RNA. The 5′ sequence of the 4.8-kb RNA, in conjunction with the clones previously characterized in our laboratory (29, 30), has allowed us to identify all of the open reading frames present in the C15 major RNA species. Figure 7 shows the positions of the potentially translatable regions in each of the reading frames. The largest open reading frame (exon VII in *Bam*HI-A) has two AUG initiation codons towards the middle of the sequence (BARF0 in Fig. 7), suggesting that if this open reading frame were translated, then a protein of approximately 16 kDa would be produced. The entire open reading frame, using an alternative initiation codon, has the potential to code for a protein of approximately 40 kDa. The position of the termination codon for BARF0 is, however, found downstream of the polyadenylation signal (AAUAAA) in the 4.8-kb RNA fam-

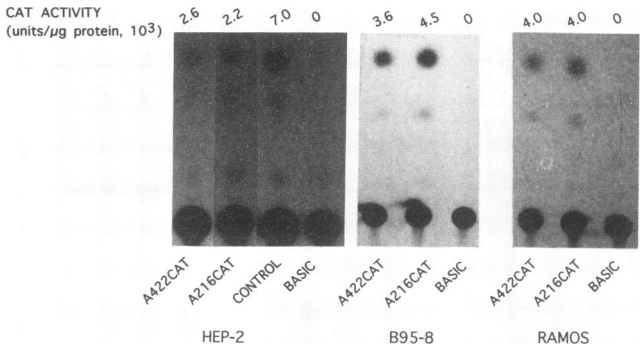


FIG. 6. Thin-layer chromatographic analysis of promoter activity in three cell lines, Hep-2, B95-8, and Ramos. The number above each column shows the CAT activity detected in each sample. BASIC, negative control plasmid containing no promoter elements; CONTROL, positive control plasmid containing simian virus 40 promoter and enhancer elements. The observed activities are expressed as units (10^{-3}) per microgram of protein.

ily of transcripts. This is an uncommon arrangement for an mRNA and suggests that this particular open reading frame may not be translated in vivo. A number of alternate open reading frames are, moreover, present, and the positions of those containing a potential initiator AUG are shown. From the sequence, it appears that a number of small proteins, none larger than about 15 kDa, could be translated from this family of transcripts.

Screening of the larger polypeptides potentially encoded by exon VII (Fig. 1 and 7) for homology with previously described proteins showed limited identity with the herpes simplex virus ICP4 gene product, as described elsewhere (9). Screening of all of the open reading frames (Fig. 7) identified two potential cyclic AMP kinase phosphorylation sites in RPMS1 (Fig. 8). In addition, this open reading frame was shown to have similarity to a region of EBNA-2 identified as functionally important for transcriptional transactivation (10) and immortalization of B cells and to papillomavirus E2.

DISCUSSION

The major EBV-encoded poly(A)-containing RNA species in NPC was initially identified as a cytoplasmic species encoded in part by the *Bam*HI-A region of the EBV genome. It appears on Northern blots at about 4.8 kb and has several associated minor transcripts, mainly with higher molecular weights (9, 29, 30). These transcripts were found also in EBV-immortalized LCLs and BL-related cells, although here they were much less prominent (9, 30). Our present data show that the transcripts extend into the *Bam*HI-I region, to a promoter which has TATA-like elements and contains some well-characterized transcriptional response elements (Fig. 2), and can be shown by CAT assays to be functionally active in both epithelial and B cells (Fig. 5).

The various RNAs that make up this family of highly expressed transcripts in NPC tumors appear to be related by differential splicing, and the broad 4.8-kb major band observed on Northern blots is composed of at least two species, of which one contains seven exons and the other contains eight (Fig. 1). Such findings emphasize the complexity of transcription in this region (*Bam*HI-I and -A) of the EBV genome. Several exons lie within a deletion found in the B95-8 virus strain, where a viral lytic origin of replication has also been mapped (25), and the promoter itself lies 5′ to the deletion; B95-8 EBV expresses a smaller major RNA, about 4.2 kb long (9, 30), consistent with the loss of exons II to IV (Fig. 1). The precise structure of the larger of these EBV complementary transcripts identified in NPCs remains to be elucidated, but two of them at least are identified by an oligonucleotide (oligonucleotide 9 in Table 1) that lies within an intron between exons VI and VII of the 4.8-kb message (Fig. 1), and further analysis of the λgt10 C15 cDNA library (29) identified recombinant clones mapping between exons I and II in the major 4.8-kb message (unpublished data), indicative of differences based on splicing patterns.

These abundant transcripts were originally overlooked in the EBV genomic sequence partly, one suspects, because they contain few long open reading frames and partly because viral genes had already been mapped onto the com-

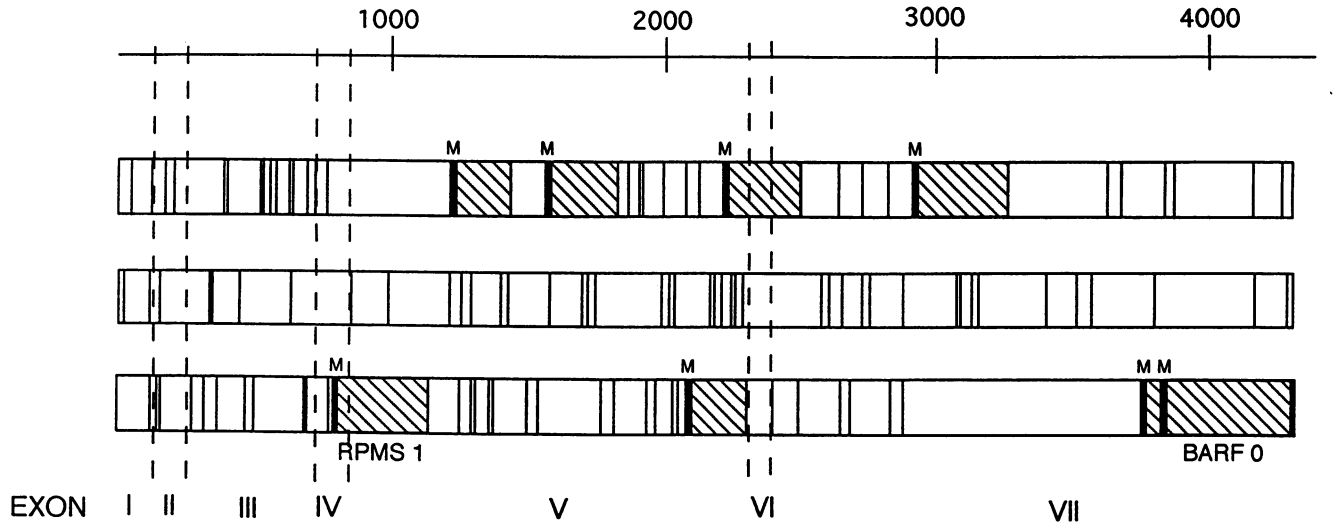


FIG. 7. Positions of termination codons (TGA, TAG, and TAA) in each of the three reading frames in the seven-exon 4.8-kb RNA species. Splice site positions are shown by broken lines. Potential open reading frames, with initiator AUG codons indicated by the letter M, are shown as hatched boxes. The coordinates at the top are derived from the sequences of clones containing all of large exon III (Fig. 1).

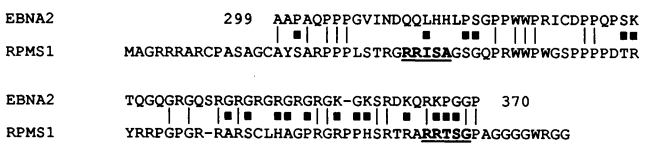


FIG. 8. Homology between the RPMS1 open reading frame (Fig. 7) and EBNA-2 residues 299 to 370. Identity is indicated by vertical lines, and conservative changes are indicated by squares. Potential cyclic AMP kinase motifs (consensus RRXS/TX; reference 18) are underlined. This homology, although not extensive, may be significant in that the triplet WWP is a very rare motif.

plementary strand (reviewed in reference 17). Further, sequence analysis, as a technique for predicting mRNAs and proteins, whether of EBV or of other viruses, has not generally identified potential biological functions on both strands of DNA in the same region of a viral genome, although experimentally they have been detected in herpes simplex virus (47) and more recently in papillomaviruses (49). (In the latter cases, the antisense RNAs have a predominant nuclear localization, however.) Such findings emphasize the importance of superimposing detailed analysis of transcripts onto sequence information. There are at least two obvious roles that might be attributed to the EBV CSTs, one taking cognizance of their antisense structures and the other based on their potential ability to encode functionally relevant polypeptides.

The importance of antisense transcripts in prokaryotic systems is well known (reviewed in reference 23), and in several eukaryotic organisms, partial antisense transcripts have recently been detected, a notable case in point being that of the EB4 gene locus in *Dictyostelium discoideum*, in which the prespore gene (EB4-PSV) appears to be stabilized, or protected, by an endogenous antisense transcript (28). A similar role was postulated for the complementary overlapping transcript identified for fibroblast growth factor bFGF in *Xenopus* oocytes (34). In both of these cases, the temporal relationship between the antisense transcripts resembles that identified in herpes simplex virus, in which the nuclear RNAs, called latency-associated transcripts, overlap a part of the viral ICP0 gene, which encodes a transactivator function (7). A simple model, as also with p53 (33), is that the antisense RNAs play a role in mRNA maturation and thus, as shown with herpes simplex virus, they themselves are not exported to the cytoplasm. However, the CSTs we describe from the *Bam*HI-I-A region of EBV are found in the cytoplasm (29), seem to bear little physical resemblence to the above-described examples, and are more complex. Moreover, in NPCs and in cottontop tamarin lymphomas (54), no messages transcribed from the other strand of the DNA could be identified (9, 30; unpublished data), suggesting that if double-stranded RNA is formed during transcription of the genome in the tumor setting, it results in complete down regulation of transcription from the alternative DNA strand. Thus, there are no other examples that describe genetic arrangements similar to those identified here (Fig. 1), that is, of a highly spliced EBV transcript in which most (at least five) of the exons share complementary sequences with genes on the other strand, including, here, those that specify the viral DNA polymerase (BALF5), a homolog of herpes simplex virus glycoprotein B (BALF4), a putative glycoprotein (BILF2), and a gene postulated, by analogy with herpes simplex virus (40), to be involved with glycoprotein transport (BALF3). Moreover, the promoter identified for the major CST (and possibly the other transcripts) contains the initiation codon for the glycoprotein gene product of BILF3 (36), which does not appear to be expressed in the tumors.

The major tumor RNAs described here, however, bear the hallmarks of mRNAs. Analysis of the sequences of the isolated cDNA clones from the C15 library identified a number of potential coding frames in the transcripts, one of which, the BARF0 open reading frame (Fig. 7), has been described previously (21, 29). In vitro translation of the BARF0 open reading frame has been reported to react with sera from NPC patients (22), although no protein has been identified in vivo and numerous polypeptide antibodies generated against the putative BARF0 product have, in our studies, failed to recognize a protein from C15 tumor extracts in a consistent manner. We noted the unconventional nature of BARF0 as a message in Results. Elsewhere (9), attention was drawn to homologies with herpes simplex virus transactivator protein ICP4. A computer search for homologies within the 4.8-kb polypeptide RNA that could be expressed from open reading frames created by splicing events and containing initiator AUG codons (Fig. 7) identified, in the open reading frame designated RPMS1, two cyclic AMP kinase motifs and about 30% identity with a region within EBNA-2 that includes a rare triplet motif, WWP (Fig. 8). Limited identity with the E2 gene of human papillomavirus was also identified (data not shown). In the case of EBNA-2, the identities lie within a region of the gene necessary for cellular transformation and LMP promoter transactivation (10). Functionally, this may be significant in that EBNA-2, in B cells at least, has been assigned a key role in malignancy, partly on the basis of the facts that naturally occurring EBNA-2-deficient viruses do not immortalize B cells in vitro, that an immortalizing function can be rescued with the gene (11, 26), and that transcription of LMP, an in vitro transforming function, appears to be regulated by EBNA-2 (1, 51). One interpretation of our findings is that certain functions of the EBNA-2 protein are also required for transformation in epithelial cells but that as the promoter region for EBNA-2 is presumably inactive—no EBNA-2 has been detected in NPCs—an alternate region of the viral genome has been recruited for this function. In this regard, it may not be a coincidence that the region covered by the major CSTs is included in the recombinant p31 fragment of the EBV genome, which can immortalize primary (human and primate) epithelial cells in culture (24, 32). Because of the complexities exhibited by these CSTs, future studies must consider the possibility that one or all of them may functionally act as antisense elements, controlling expression from the other viral strand (30), and also express polypeptides like those in adenovirus (8) with significance for tumor induction and/or maintenance.

The CpG methylation patterns of the EBV genome indicate the tight control that is exerted over transcriptional expression in the C15 NPC tumor, compared, for example, with that in B-lymphoblastoid line NAD, generated with virus from C15, as discussed elsewhere (2, 13). In the tumor, the methylation pattern over the *Bam*HI A and I fragments suggests that the EBV genome is probably transcriptionally active only in the region that contains the promoter in *Bam*HI-I (Fig. 7; unpublished data). The sequence of the promoter (Fig. 2) identifies several transcriptional response elements (16). In CAT assays (Fig. 6) carried out with this region, contrary to expectation, little difference in transcriptional expression was observed in the epithelial- and B-cell types analyzed. The fact that, in vivo, higher levels of CSTs were observed in the epithelial-cell environment than in

corresponding B cells may suggest an interaction with different cellular (or viral) transcription factors in the two cases. To solve this problem, further work on the elements that control the promoter activity are required. Alternatively, since only detailed sequence analysis is available for B cells, there may be mutations (or enhancers) in the NPCs relevant to promoter activities in the tumor setting that have not yet been recognized.

In summary, we show here, for the first time, the complete structure of the major species in a highly abundant family of EBV polyadenylated transcripts found in NPC cells. These transcripts are unique, even among known viral and eukaryotic antisense RNAs, in that most exons that they comprise complement an open reading frame on the other strand. This arrangement provides the potential for a novel method of transcriptional control over a large area of the viral DNA. The promoter for these transcripts exhibits a high degree of tissue specificity; although it contains a number of well-characterized transcriptional regulatory elements and can function in both B and epithelial cells, the high level of transcription observed in NPC tumors suggests specific regulation in this particular environment. We have also identified several potential open reading frames that are generated by splicing events, one of which, RPMS1, merits detailed attention by virtue of its homology to regions of EBV EBNA-2. The transcriptional complexity of the *Bam*HI-I-A region of the EBV genome and the diverse mechanistic opportunities it provides for control of gene expression could, we think, open a new area of research on this important human virus.

ACKNOWLEDGMENTS

We thank the Cancer Research Campaign for generous support of this work and M. H. Ng and Hong Lin Chen, Department of Microbiology, University of Hong Kong, and A. Morgan, Department of Pathology, University of Bristol, for helpful discussions and communication of results prior to publication.

REFERENCES

1. **Abbot, S. D., M. Rowe, K. Cadwallader, A. Ricksten, J. Gordon, F. Wang, L. Rymo, and A. B. Rickinson.** 1990. Epstein-Barr virus nuclear antigen 2 induces expression of the virus-encoded latent membrane protein. J. Virol. **64**:2126–2134.
2. **Allday, M. J., D. Kundu, and B. E. Griffin.** 1990. CpG methylation of viral DNA in EBV associated tumours. Int. J. Cancer **45**:1125–1130.
3. **Arrand, J. R., J. E. Walsh-Arrand, and L. Rymo.** 1983. Cytoplasmic RNA from normal and malignant human cells shows homology to DNAs of Epstein-Barr virus and human adenoviruses. EMBO J. **2**:1673–1683.
4. **Baer, R., R. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tufnell, and G. B. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
5. **Bankier, A. T., P. L. Deininger, S. C. Satchwell, R. Baer, P. J. Farrell, and B. G. Barrell.** 1983. DNA sequence analysis of the *Eco*RI-Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences. Mol. Biol. Med. **1**:425–446.
6. **Busson, P., G. Ganem, P. Flores, F. Mugneret, B. Clauss, B. Caillou, K. Braham, H. Wakasugi, M. Lipinski, and T. Tursz.** 1988. Establishment and characterisation of three transplantable EBV-containing nasopharyngeal carcinomas. Int. J. Cancer **42**:599–606.
7. **Cai, W., and P. A. Schaffer.** 1992. Herpes simplex virus type 1 ICP0 regulates expression of immediate-early, early, and late genes in productively infected cells. J. Virol. **66**:2904–2915.
8. **Carlin, C. R., A. E. Tolefson, H. A. Brady, B. L. Hoffman, and W. S. M. Wold.** 1989. Epidermal growth factor receptor is down-regulated by a 10,400 MW protein encoded by the E3 region of adenovirus. Cell **57**:135–144.
9. **Chen, H.-L., M. L. Lung, D. Choy, J. Sham, B. E. Griffin, and M. H. Ng.** 1992. Transcription of *Bam*HI-A region of the EBV genome in NPC tissues and B cells. Virology **191**:193–201.
10. **Cohen, J. I., F. Wang, and E. Kieff.** 1991. Epstein-Barr virus nuclear protein 2 mutations define essential domains for transformation and transactivation. J. Virol. **65**:2545–2554.
11. **Cohen, J. I., F. Wang, J. Mannick, and E. Kieff.** 1989. Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation. Proc. Natl. Acad. Sci. USA **86**:9558–9562.
12. **Dambaugh, T., K. Hennessy, S. Fennewald, and E. Kieff.** 1986. The EBV genome and its expression in latent infection, p. 13–45. *In* M. A. Epstein and B. G. Achong (ed.), The Epstein-Barr virus: recent advances. John Wiley & Sons, Inc., New York.
13. **Ernberg, I., K. Falk, J. Minarovits, P. Busson, T. Tursz, M. G. Masucci, and G. Klein.** 1989. The role of methylation in the phenotype dependent modulation of EB nuclear antigen 2 and LMP genes in cells latently infected with EBV. J. Gen. Virol. **70**:2989–3002.
14. **Fåhraeus, R., L.-F. Hu, I. Ernberg, J. Finke, M. Rowe, G. Klein, K. Falk, E. Nilsson, M. Yadav, P. Busson, T. Tursz, and B. Kallin.** 1988. Expression of EBV encoded proteins in NPC. Int. J. Cancer **42**:329–338.
15. **Fåhraeus, R., L. Rymo, J. S. Rhim, and G. Klein.** 1990. Morphological transformation of human keratinocytes expressing the LMP gene of Epstein-Barr virus. Nature (London) **345**:447–449.
16. **Faisst, S., and S. Meyer.** 1992. Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res. **20**:3–26.
17. **Farrell, P.** 1989. Epstein-Barr virus genome. Adv. Viral Oncol. **8**:103–132.
18. **Feramisco, J. R., D. B. Glass, and E. G. Krebs.** 1980. Optimal spatial requirements for the location of basic residues in peptide substrates for the cyclic AMP-dependent protein kinase. J. Biol. Chem. **255**:4240–4245.
19. **Friedman, K. D., N. L. Rosen, P. J. Newman, and R. R. Montgomery.** 1990. Screening of λgt 11 libraries, p. 253–260. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, London.
20. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full length cDNAs from rare transcripts. Amplification using a single gene specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**:8998–9002.
21. **Gilligan, K., H. Sato, P. Rajadurai, P. Busson, L. Young, A. Rickinson, T. Tursz, and N. Raab-Traub.** 1990. Novel transcription from the Epstein-Barr virus terminal *Eco*RI fragment, DIJhet, in a nasopharyngeal carcinoma. J. Virol. **64**:4948–4956.
22. **Gilligan, K. J., P. Rajadurai, J.-C. Lin, P. Busson, M. Abdel-Hamid, U. Prasad, T. Tursz, and N. Raab-Traub.** 1991. Expression of the Epstein-Barr virus *Bam*HI A fragment in nasopharyngeal carcinoma: evidence for a viral protein expressed in vivo. J. Virol. **65**:6252–6259.
23. **Green, P. J., O. Pine, and M. Inouye.** 1986. The role of antisense RNA in gene regulation. Annu. Rev. Biochem. **55**:569–597.
24. **Griffin, B. E., and L. Karran.** 1984. Immortalization of monkey epithelial cells by specific fragments of Epstein-Barr virus DNA. Nature (London) **309**:78–82.
25. **Hammerschmidt, W., and B. Sugden.** 1988. Identification and characterisation of oriLyt a lytic origin of DNA replication of Epstein-Barr virus. Cell **55**:427–433.
26. **Hammerschmidt, W., and B. Sugden.** 1989. Genetic analysis of immortalizing functions of Epstein-Barr virus in human B lymphocytes. Nature (London) **340**:393–397.
27. **Herbst, H., F. Dallenbach, M. Humel, G. Niedobitek, S. Pileri, N. Muller-Lantzsch, and H. Stein.** 1991. EBV expression in Hodgkin and Reed-Sternberg cells. Proc. Natl. Acad. Sci. USA **88**:4766–4770.
28. **Hildebrandt, M., and W. Nellen.** 1992. Differential antisense transcription from the Dictyostelium EB4 gene locus: implications on antisense-mediated regulation of mRNA stability. Cell

69:197–204.

29. **Hitt, M. M., M. A. Allday, T. Hara, L. Karran, M. D. Jones, P. Busson, T. Tursz, I. Ernberg, and B. E. Griffin.** 1989. EBV gene expression in an NPC related tumor. EMBO J. **8:**2639–2651.
30. **Karran, L., Y. Gao, P. R. Smith, and B. E. Griffin.** 1992. Expression of a family of EB virus complementary strand transcripts in latently infected cells. Proc. Natl. Acad. Sci. USA **89:**8058–8062.
31. **Karran, L., Y. Gao, P. R. Smith, M. Lung, M. H. Ng, J. Sham, D. R. Choy, M. Lui, and B. E. Griffin.** 1991. Characterization of the novel "18.8" family of transcripts from NPC tumors, p. 157–161. *In* D. V. Ablashi, A. T. Huang, J. S. Pagano, G. R. Pearson, and C. S. Yang (ed.), Epstein-Barr virus and human disease III. Humana Press, Clifton, N.J.
32. **Karran, L., C. G. Teo, D. King, M. M. Hitt, Y. Gao, N. Wedderburn, and B. E. Griffin.** 1990. Establishment of immortalized primate epithelial cells with sub-genomic EBV DNA. Int. J. Cancer **45:**763–772.
33. **Khochbin, S., and J.-J. Lawrence.** 1989. An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate. EMBO J. **8:**4107–4114.
34. **Kimelman, D., and N. W. Kirschner.** 1989. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. Cell **59:**687–696.
35. **Lenoir, G., G. O'Conor, and C. L. M. Olweny.** 1985. Burkitt's lymphoma. IARC Sci. Publ. **60:**1–484.
36. **Mackett, M., M. J. Conway, J. R. Arrand, R. S. Haddad, and L. M. Hutt-Fletcher.** 1990. Characterization and expression of a glycoprotein encoded by the Epstein-Barr virus *Bam*HI I fragment. J. Virol. **64:**2545–2552.
37. **MacMahon, E. M. E., J. D. Glass, S. D. Hayward, R. B. Mann, P. S. Becker, P. Charache, J. C. McArthur, and R. F. Ambinder.** 1991. Epstein-Barr virus in AIDS-related primary central nervous system lymphoma. Lancet **338:**969–973.
38. **Old, L. J., E. A. Boyse, H. F. Oettgen, E. deHarven, G. Geering, B. Williamson, and P. Clifford.** 1966. Precipitating antibody in human serum to an antigen present in cultured Burkitt's lymphoma cells. Proc. Natl. Acad. Sci. USA **56:**1699–1704.
39. **Parker, B. D., A. Bankier, S. Satchwell, B. Barrell, and P. J. Farrell.** 1990. Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region. Virology **179:** 339–346.
40. **Pellet, P. E., F. J. Jenkins, M. Ackermann, M. Sarmiento, and B. Roizman.** 1986. Transcription initiation sites and nucleotide sequence of a herpes simplex virus 1 gene conserved in the Epstein-Barr virus genome and reported to affect the transport of viral glycoproteins. J. Virol. **60:**1134–1140.
41. **Raab-Traub, N., R. Hood, C.-S. Yang, B. Henry II, and J. S. Pagano.** 1983. Epstein-Barr virus transcription in nasopharyngeal carcinoma. J. Virol. **48:**580–590.
42. **Reisman, D., and B. Sugden.** 1986. *trans* activation of an Epstein-Barr viral transcriptional enhancer by the Epstein-Barr viral nuclear antigen 1. Mol. Cell. Biol. **6:**3838–3846.
43. **Reisman, D., J. Yates, and B. Sugden.** 1985. A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two *cis*-acting components. Mol. Cell. Biol. **5:**1822–1832.
44. **Simons, M. J., and K. Shanmugaratnam.** 1982. UICC technical report series 71, report 16. UICC, Geneva.
45. **Smith, P. R., and B. E. Griffin.** 1991. Differential expression of Epstein-Barr viral transcripts for two proteins (TP1 and LMP) in lymphocyte and epithelial cells. Nucleic Acids Res. **19:**2435–2440.
46. **Smith, P. R., and B. E. Griffin.** 1992. Transcription of the Epstein-Barr virus gene EBNA-1 from different promoters in nasopharyngeal carcinoma and B-lymphoblastoid cells. J. Virol. **66:**706–714.
47. **Stevens, J. G., E. K. Wagner, G. B. Devi-Rao, M. L. Cork, and L. T. Feldman.** 1987. RNA complementary to a herpes virus α gene mRNA is prominent in latently infected neurons. Science **235:**1056–1059.
48. **Sugden, B., and N. Warren.** 1989. A promoter of Epstein-Barr virus that can function during latent infection can be transactivated by EBNA-1, a viral protein required for viral DNA replication during latent infection. J. Virol. **63:**2644–2649.
49. **Vormwald-Dogan, G., B. Fischer, H. Blundau, L. Gissmann, D. Glitz, E. Schwarz, and M. Dürst.** 1992. Sense and antisense transcripts of human papillomavirus type 16 in cervical cancers. J. Gen. Virol. **73:**1833–1838.
50. **Wang, D., D. Liebowitz, and E. Kieff.** 1985. An EBV membrane protein expressed in immunoblast lymphocytes transforms established rodent cells. Cell **43:**831–840.
51. **Wang, F., S.-F. Tsang, M. G. Kurilla, J. I. Cohen, and E. Kieff.** 1990. Epstein-Barr virus nuclear antigen 2 transactivates latent membrane protein LMP1. J. Virol. **64:**3407–3416.
52. **Wei, M. X., and T. Ooka.** 1989. A transforming function of the BARF1 gene encoded by Epstein-Barr virus. EMBO J. **8:**2897–2903.
53. **Yates, J., N. Warren, D. Reisman, and B. Sugden.** 1984. A cis-acting element from the EB viral genome that permits stable replication of recombinant plasmids in latently infected cells. Proc. Natl. Acad. Sci. USA **81:**3806–3810.
54. **Zhang, C. X., P. Lowrey, S. Finerty, and A. J. Morgan.** 1993. Analysis of Epstein-Barr virus gene transcription in lymphoma induced by the virus in the cottontop tamarin by construction of a cDNA library with RNA extracted from a tumour biopsy. J. Gen. Virol. **74:**509–514.