JOURNAL OF VIROLOGY, Feb. 1995, p. 1132–1141 Vol. 69, No. 2
0022-538X/95/$04.00+0

Structural Analyses of the Epstein-Barr Virus *Bam*HI A Transcripts

ROBERT H. SADLER AND NANCY RAAB-TRAUB*

Department of Microbiology and Immunology and the Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295

Received 6 September 1994/Accepted 8 November 1994

Epstein-Barr virus (EBV) gene expression in nasopharyngeal carcinoma (NPC) includes abundant rightward transcription of the *Bam*HI A fragment, consisting of mRNAs ranging in size from approximately 4.0 to 8.0 kb. These transcripts include several distinctly spliced forms which are 3′-end coterminal and contain the *Bam*HI A rightward frame 0 (BARF0) open reading frame (ORF) in the final exon. *Bam*HI A transcription is detected at a lower level of expression in EBV-infected lymphoid cells. In this study, cDNA cloning, reverse transcription-based PCR, and Northern (RNA) blotting were used to further define the structures of the *Bam*HI A transcripts and to characterize their expression in different EBV-infected tissues. Three *Bam*HI A cDNAs isolated from a passaged NPC represent previously unidentified mRNAs that contain BARF0 and additional ORFs encoded by multiple exons, including one which extends the size of the BARF0 ORF from 174 to 279 codons. The distinct exons were detected in multiple, differently sized mRNAs, indicating that these transcripts have complex patterns of alternate splicing. In support of this finding, 5′-end analysis confirmed the presence of a previously reported start site and also identified a subset of transcripts of 4.8 kb and larger that initiate further 5′ to this site. In addition, 3′-end analysis identified heterogeneous 3′-end processing in all of the *Bam*HI A mRNAs, resulting in transcripts that either contain the entire BARF0 ORF or are cleaved and polyadenylated 5′ of the stop codon. Finally, the expression of multiple, distinctly spliced *Bam*HI A transcripts was consistently detected in a wide range of EBV-infected samples, including NPC, Burkitt's lymphoma, and parotid carcinoma biopsy samples, and in type I and type III Burkitt's lymphoma lines and type III lymphoblastoid cell lines. This complex pattern of start site selection, alternate splicing, and heterogeneous 3′-end processing is likely to regulate the expression in vivo of the ORFs encoded by the EBV *Bam*HI A transcripts.

Epstein-Barr virus (EBV) is a ubiquitous human herpesvirus that likely plays a causative role in the development of several cancers. It is consistently detected in endemic Burkitt's lymphoma (BL) and nasopharyngeal carcinoma (NPC). Analysis of viral termini indicates that both BL and NPC are clonal proliferations of a single EBV-infected cell, suggesting that the virus has a role in the origin of these tumors (8, 26, 32). It is therefore important to identify the viral genes which are expressed in these transformed cells.

In vitro, EBV infection of B lymphocytes efficiently results in the outgrowth of immortalized lymphoblastoid cell lines (LCLs) (25, 31). This infection is predominantly nonproductive, with expression of six nuclear antigens (EBNAs), three membrane proteins (LMP1, -2A, and -2B), and two highly abundant small nuclear transcripts (EBER1 and -2) (17). This pattern of EBV gene expression has been defined as type III latency, whereas in BL, viral gene expression is more restricted and has been defined as type I latency (16). In BL in vivo, only EBNA-1 has been detected at the protein level. In NPC, defined as type II latency, EBNA-1 has been detected by immunoblotting in all samples, and LMP1 has been detected in 40 to 60% (9, 36). Studies of viral transcription in NPC by Northern (RNA) blotting, cDNA cloning, and reverse transcription-based PCR (RT-PCR) have also detected the EBER1 and EBER2 polymerase III RNAs, LMP1 and LMP2 mRNAs, and rightward transcription terminating in the *Bam*HI A region of the EBV genome (6, 11, 33).

Transcription from *Bam*HI A was not initially detected in latently infected LCLs and was first identified in NPC, in which these transcripts are relatively abundant. The initial studies identified several rightward messages, ranging in size from approximately 4.0 to 8.0 kb, on Northern blots (12, 13). These transcripts have subsequently been detected by RT-PCR at very low levels in LCLs (3, 7, 15). In situ hybridization has detected *Bam*HI A transcription in all cells of NPC tumors, in the nude-mouse-passaged NPC C15, and in a carcinoma of the parotid gland. The transcripts have also been detected at low levels in B cells maintained in vitro, including cell lines with type 1 latency in which only the EBNA-1 gene was thought to be expressed (3, 11). *Bam*HI A transcription is therefore a significant component of EBV latent gene expression, and it is essential to determine the structures of these mRNAs.

Several partial *Bam*HI A cDNAs have been isolated from different sources, including NPC, C15, and EBV-induced lymphomas of cottontop tamarins (7, 12, 13, 15, 49). These cDNAs are distinctly spliced but 3′-end coterminal, and all include a 174-codon open reading frame (ORF) designated *Bam*HI A rightward frame 0 (BARF0). This ORF has been translated in vitro and is immunoprecipitable with sera from patients with NPC, suggestive of its expression in vivo (11). Other unique ORFs have been identified in certain *Bam*HI A cDNAs (15). It has also been suggested that the *Bam*HI A transcripts have a role in the maintenance of latency by an antisense regulation of the replicative genes encoded by the opposite strand (13). In this regard, the *Bam*HI A transcripts are analogous to the

* Corresponding author. Mailing address: Campus Box 7295, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295. Phone: (919) 966-1701. Fax: (919) 966-3015. Electronic mail address: NRT@med.unc.edu.

latency-associated transcripts of herpes simplex virus that also contain sequences that are antisense to replicative genes. Although the structures of the larger latency-associated transcripts are still unknown, these transcripts potentially encode a number of small ORFs. A regulatory role for these RNAs has not been established; however, the expression in vivo of one of the ORFs has recently been described (21).

In this study, cDNA cloning, RT-PCR, and Northern blot hybridization were used to define the structures of the *Bam*HI A transcripts and to characterize their expression in different EBV-infected tissues. cDNA clones isolated from C15 contained novel exons and distinct patterns of alternate splicing. These cDNAs represent previously unidentified mRNAs that contain BARF0 and additional ORFs encoded by multiple exons, including one which extends the size of the BARF0 ORF from 174 to 279 codons. The expression of the distinct mRNAs represented by these cDNAs, and of previously reported *Bam*HI A cDNAs, was consistently detected in NPC, BL, and parotid carcinoma biopsy samples, in type I and type III BL lines, and in a type III LCL. The cDNAs presented here and elsewhere now identify at least five distinct patterns of splicing for the *Bam*HI A transcripts. Four of these splicing patterns were detected in multiple size-selected mRNA fractions, indicating additional levels of transcript heterogeneity and patterns of alternate splicing. In support of this finding, 5′-end analysis of the *Bam*HI A transcripts confirmed the presence of a previously reported start site and also identified a subset of transcripts of 4.8 kb and larger that initiate further 5′ to this site. In addition, 3′-end analysis identified heterogeneous 3′-end processing in all of the *Bam*HI A mRNAs, resulting in transcripts that either contain the entire BARF0 ORF or are truncated upstream of the stop codon. This complex pattern of start site selection, alternate splicing, and 3′-end processing is likely to regulate the expression in vivo of the ORFs encoded by the EBV *Bam*HI A transcripts.

MATERIALS AND METHODS

Cell lines and tumors. Cell lines were grown at 37°C in RPMI 1640 medium with 10% fetal calf serum and antibiotics. B95-8 is an EBV-positive marmoset LCL. Raji, Akata, and Mutu are EBV-positive BL cell lines. B95-8 and Raji have type III latency, while Akata and Mutu have type I latency (36). Louckes is an EBV-negative BL cell line. NPC tumors C15, C17, C18, and C19 were serially passaged in athymic nude mice and have been previously described (5, 6). C15 tumors were passaged by using Matrigel basement membrane matrix (Collaborative Biomedical Products). Biopsies were obtained as frozen samples and stored at −70°C.

RNA preparation and Northern blotting. Total RNA was prepared by suspending pelleted cells or pulverized frozen specimens in 4 M guanidine thiocyanate and then subjecting them to centrifugation through a 5.7 M cesium chloride step gradient (12). Poly(A)$^+$ RNA (mRNA) was prepared by oligo(dT)-Sepharose chromatography (Collaborative Biomedical Products) or by using Oligotex-dT (Qiagen) according to the manufacturer's instructions. For Northern blots, 30 μg of poly(A)$^+$ RNA was resolved by electrophoresis on 1% agarose formaldehyde gels with RNA molecular weight standards (Life Technologies) and blotted by capillary transfer onto BA-S85-supported nitrocellulose (Schleicher & Schuell) (11). ^{32}P-labeled single-stranded RNA probes were prepared from linearized templates by using bacteriophage RNA polymerases and purified by using Sephadex G-50 quick-spin RNA columns (Boehringer Mannheim). Hybridizations were performed at 50°C, using ≥10^6 cpm/ml in 50% formamide, with washing at 72°C in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS)–0.1% sodium pyrophosphate.

For preparation of size-selected mRNA fractions, 40 to 50 μg of C15 poly(A)$^+$ RNA was resolved by electrophoresis on 1% agarose denaturing gels. Fraction size was determined by comparison with RNA molecular weight standards (Life Technologies) and by hybridization of a parallel lane which was blotted and probed with a single-stranded RNA probe complementary to BARF0. RNA was extracted from gel slices by using an RNaid RNA purification kit (Bio 101). RT-PCR was performed by using Superscript II reverse transcriptase (Life Technologies) and 0.5 μg of oligo(dT)$_{20}$ primer for each fraction as described below.

RT-PCR and cDNA cloning. Two-microgram RNA samples were denatured by using 10 mM methylmercuric hydroxide (Alpha Chemical) for 10 min at room temperature, and then 0.14 M 2-mercaptoethanol was added. Oligo(dT)$_{20}$ primer (0.5 μg) was annealed by heating to 70°C for 10 min and then incubated on ice. Reverse transcription was performed in a 20-μl reaction volume at 45°C, using Superscript II reverse transcriptase (Life Technologies) according to manufacturer's instructions. Equivalent aliquots of RNA were processed in parallel without the addition of reverse transcriptase to serve as negative controls.

PCR was performed with 5 μl of a 1:20 cDNA dilution in a 25- or 50-μl reaction volume with *Taq* DNA polymerase (Promega), using standard reaction conditions including 1.5 mM $MgCl_2$, 0.5 pmol of each primer per μl, 200 μM dATP, dCTP, and dTTP, 50 μM dGTP, and 150 μM 7-deaza-dGTP (Boehringer Mannheim) for 40 cycles. Reaction conditions were as follows: 95°C for 5 min, 60°C for 1 min, and 72°C for 1 min for 1 cycle; 95°C for 1 min, 60°C for 1 min, and 72°C for 1 min for 39 cycles; and 72°C for 10 min. Nested PCRs were performed with a 1:50 template dilution for 20 to 40 cycles. PCR assays included negative control reaction mixtures containing no cDNA. Twenty microliters of each PCR reaction mixture was resolved by electrophoresis on 1% agarose gels with DNA molecular weight size standards. Gels were denatured, neutralized, and blotted by capillary transfer onto BA-S83-supported nitrocellulose (Schleicher & Schuell) (12). Hybridizations were performed at 45°C, using ^{32}P-end-labeled oligonucleotides at ≥10^6 cpm/pmol/ml and washed at 45°C in 1× SSC–0.1% SDS.

PCR primers and oligonucleotide probes (Table 1) are numbered at their 5′ ends according to the B95-8 sequence, except for primers within the 12-kb B95-8 deletion, which are numbered according to Raji sequence, and extend 5′ to 3′ in the direction indicated (R, rightward; L, leftward) (1, 28).

For cloning of PCR products, ethidium bromide-stained bands were extracted from agarose gels by using Qiaex (Qiagen), treated with T4 DNA polymerase and T4 polynucleotide kinase (Promega), and ligated into the pBS vector (Stratagene). Positive clones were identified by Southern blotting of plasmids digested with restriction enzymes, and an example of each distinctly sized insert was sequenced.

5′ RACE. Rapid amplification of cDNA 5′ ends (5′ RACE) was performed as previously described (10). Two micrograms of poly(A)$^+$ RNA was primed with 0.5 μg of oligo(dT)$_{20}$ and reverse transcribed as described above. cDNA 5′ ends were tailed with dATP and terminal deoxynucleotidyltransferase (Boehringer Mannheim) according to the manufacturer's instructions. PCR was performed with an oligo(dT)$_{17}$-adapter primer (5′ GACTCGAGTCGACATCGAT$_{17}$) and leftward gene-specific primer 9983-L (exon III) or 6572-L (exon II), using standard conditions as described above. PCR products were screened by Southern blot analysis using oligonucleotide probe 150708-L. For cloning, gel slices corresponding to hybridizing bands were extracted by using Qiaex (Qiagen) and cloned by treatment with T4 DNA polymerase and T4 polynucleotide kinase (Promega) prior to ligation into plasmid vector pGem3Z (Promega).

3′ RACE. Rapid amplification of cDNA 3′ ends (3′ RACE) was performed as previously described (10). Two micrograms of poly(A)$^+$ RNA was primed with oligo(dT)$_{17}$-adapter (5′ GACTCGAGTCGACATCGAT$_{17}$) and reverse transcribed as described above. PCR was performed with primer 160853-R and an adapter region primer (5′ GACTCGAGTCGACATCGAT), and products were screened by Southern blot analysis using oligonucleotide probe 160901-R. For cloning, gel slices corresponding to hybridizing bands were extracted by using Qiaex (Qiagen) and cloned by treatment with T4 DNA polymerase and T4 polynucleotide kinase (Promega) prior to ligation into pBS$^+$ vector (Stratagene).

3′ RACE analysis of size-selected mRNA fractions was performed by PCR of oligo(dT)$_{20}$-primed, size-selected C15 cDNA fractions (see above), using primers 160853-R and oligo(dT)$_{17}$-adapter (5′ GACTCGAGTCGACATCGAT$_{17}$). PCR products were extracted with phenol-chloroform and precipitated. Equivalent aliquots were subsequently cut with restriction enzyme *Mse*I or uncut and resolved by electrophoresis on 10% polyacrylamide gels. Gels were transferred by electroblotting onto Zeta-probe (Bio-Rad) and hybridized according to the manufacturer's instructions with ^{32}P-end-labeled oligonucleotide 160901-R.

RESULTS

Cloning of alternately spliced cDNAs. In C15 and NPC biopsy samples at least four rightward messages from *Bam*HI A, ranging in size from 4.0 to 8.0 kb with a prominent band at 4.8 kb, are detected on Northern blots. A similar pattern of transcription is detected at a lower level of expression in EBV-infected lymphoid cells. Several partial cDNAs representing distinctly spliced, rightward *Bam*HI A transcripts have been previously isolated from C15 and NPC cDNA libraries (7, 12, 13, 15, 49). Structures based on the assembly of partial cDNA sequences have been proposed to represent the major 4.8-kb *Bam*HI A transcript. These include a seven-exon form and an alternate form in which exon III is spliced into exons IIIa and IIIb (Fig. 1) (43). In this study, additional *Bam*HI A cDNAs containing novel exons and distinct patterns of alternate splicing were isolated.

TABLE 1. Oligonucleotides used for PCR and hybridization[a]

Oligonucleotide	Sequence	Location
150195-R	5′ CTGCATGCGCAAGGGTCACA	Exon I
150400-R	5′ AACTCTGCAGGTGAGATTCA	Exon I
150519-R	5′ GAGTCATTCCTAAACTGTCA	Exon I
150626-L	5′ CGGGTATGGCTGTTGTTGCA	Exon I
150708-L	5′ TCCATGTTGCTGAGGAGCT	Exon I
6572-L	5′ GCCCCCGTTTCCAGTTC	Exon II
9942-R	5′ TGTGAATACGGGCGTATGAC	Exons III and IIIa
9983-L	5′ AGCACACGCCCCCATTTCTA	Exons III and IIIa
10311-R	5′ CCGTAAAGTAGCGGCTGCCG	Exons III and IIIb
10350-L	5′ TCCTTCAACAGCACCACCTA	Exons III and IIIb
156051-R	5′ GTTGGCCTCAAAGATCCGAC	Exons V and Va′
156180-L	5′ CCGTTCCATTATGGGCTACG	Exons V and Va′
157313-R	5′ GTTTTGCGCCTGGAAGTTGT	Exon VI
157375-L	5′ CGTGCCTCCAACGTCTTTGA	Exon VI
159190-L	5′ AGCTTTCCTTTCCGAGTCTG	Exons VII, VIIa, and VIIa′
159847-R	5′ GTACTTGGAATTCTCAAAGG	Exons VII and VIIa′
160004-R	5′ TCTCGCTCTGTTGCGCCAGG	Exons VII and VIIa′
160070-R	5′ ACTGCTCCTTGCGCTTGAGC	Exon VII; intron VIIa′-VIIb
160387-L	5′ GGCTGACTCACCTGTTTG	Exons VII, VIIb, and VIIb′
160537-L	5′ AGCTCCTCTATTGGAACTCT	Exons VII, VIIb, and VIIb′
160853-R	5′ GCACAGAGGTGATGTAGA	Exons VII, VIIb, and VIIb′
160901-R	5′ GAAAGCGGCCCAGCAGATCC	Exons VII, VIIb, and VIIb′

[a] Oligonucleotides are numbered at the 5′ end according to the EBV strain B95-8 sequence or Raji sequence (exons II and III only) and extend in the direction indicated (R, rightward; L, leftward) relative to the standard EBV genome (1, 28). The locations of oligonucleotides within the structures of the *Bam*HI A cDNAs are indicated.

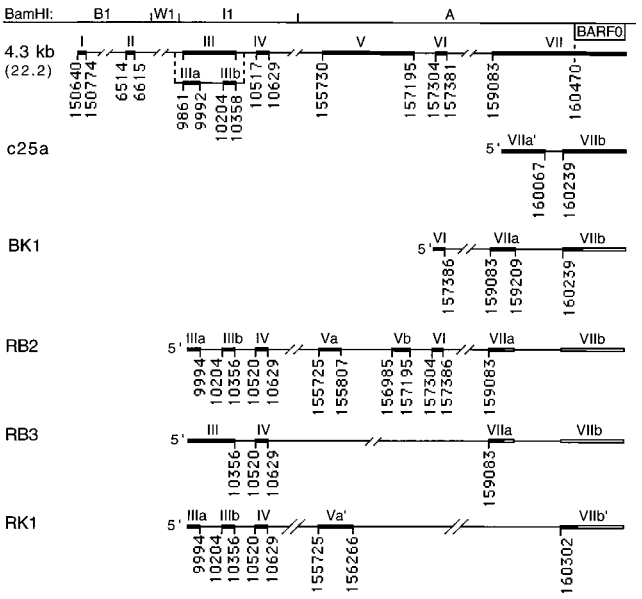

FIG. 1. Structures of the C15 *Bam*HI A cDNAs. cDNAs isolated from the C15 passaged NPC are positioned as exons (thick lines, roman numerals) and introns (thin lines) aligned with the *Bam*HI restriction map of Raji strain EBV sequence. Splice sites are numbered according to the B95-8 EBV sequence, except for exons II and III, which are numbered according to the Raji insertion relative to B95-8 in this region (1, 28). The position of BARF0 is also indicated. The seven-exon structure (4.3 kb [22.2]) was assembled from two overlapping cDNAs, and the alternately spliced exon IIIa-IIIb structure was derived from a third cDNA (43). c25a has also been previously reported (12). BK1, RB2, RB3, and RK1 were isolated by PCR amplification of C15 cDNA by using primers 157313-R and 160537-L for BK1, 9942-R and 159190-L for RB2 and RB3, and 9942-R and 160537-L for RK1. BK1 is spliced identically to the T13 cDNA isolated from an EBV-induced tamarin lymphoma (49). The unshaded exon regions are inferred by PCR analysis to be contained in the full-length transcripts (see Results).

cDNA BK1 was isolated by RT-PCR of C15 mRNA, using primers 157313-R (exon VI) and 160537-L (exon VIIb, downstream of the BARF0 start codon), and contained a splice pattern similar to that of the A73 cDNA, from an NPC biopsy, and identical to that of the T13 cDNA from an EBV-induced tamarin lymphoma (7, 49). cDNAs 22.2, BK1, and c25a have therefore identified three distinct forms of exon VII in NPC, all of which encode the BARF0 ORF (Fig. 1). In addition, BK1 encodes a 96-codon ORF that is not present in the A73 cDNA because of a 1-bp difference at the splice donor site of exon VI (see below).

Additional *Bam*HI A cDNAs were isolated by RT-PCR using primer 9942-R (exon III), within the *Bam*HI I1 fragment, and 159190-L (exon VIIa). cDNA RB2 contained the previously identified exons IIIa, IIIb, IV, and VI but had a distinctly spliced form of exon V (forming exons Va and Vb) (Fig. 1). To determine if the RB2 cDNA was derived from a transcript that contains the BARF0 ORF, RT-PCR was performed with primers 9942-R (exon III) and 160537-L (exon VIIb, downstream of the BARF0 start codon). This resulted in detection of a product of a predicted size (1,119 bp) which was identified by hybridization with an internal oligonucleotide probe (160387-L, exon VIIb), indicating that the RB2 cDNA was derived from a transcript that contains exons VIIa and VIIb (Fig. 1 and data not shown). This result, and the identical structures of exons VI and VIIa in RB2 and BK1, indicated that these cDNAs are likely derived from the same spliced transcript, which is referred to hereafter as RB2/BK1. This transcript would encode BARF0 as well as a 57-codon ORF spanning exons IIIa and IIIb, and a 126-codon ORF (spanning exons Vb to VIIa) that is an in-frame extension of the 96-codon ORF of BK1 (Fig. 2). As noted above, this ORF is not encoded by the A73 cDNA because of the different splice donor site of exon VI. To further analyze this distinction and to ensure that it was not the result of a misincorporation by *Taq* polymerase, three additional cDNAs from C15 and three cDNAs from an NPC biopsy were cloned by RT-PCR using primers in exon VI (157313-R) and

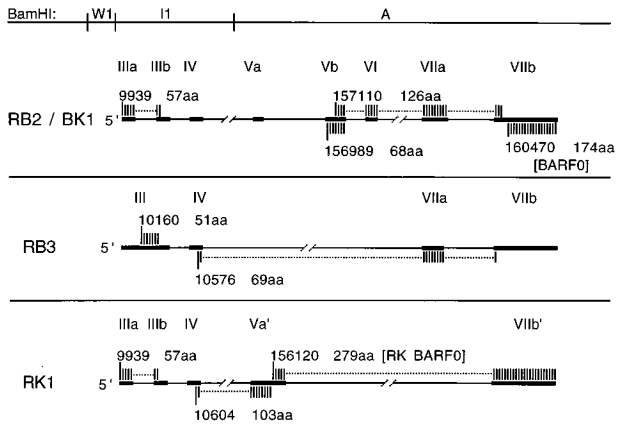

FIG. 2. ORFs of the C15 *Bam*HI A cDNAs. ORFs contained within each cDNA are shown as shaded regions, with positions of the start codons numbered as in Fig. 1. The number of amino acid codons in each ORF is indicated. The BARF0 ORF, shown here in the RB2/BK1 cDNA, was shown by PCR analysis to be contained in each cDNA (see Results). The 126-aa ORF of RB2/BK1 is an in-frame extension of a 96-aa ORF in the T13 cDNA, which has been previously described (49). In RK1, the 279-aa ORF, RK-BARF0, is an in-frame extension of the BARF0 ORF.

exon VIIb (160387-L) and sequenced. All six cDNAs were spliced identically to the RB2/BK1 cDNA, suggesting that this is the predominant form of the transcripts.

cDNA RB3, also isolated from C15 by using primers 9942-R (exon III) and 159190-L (exon VIIa), contained exons III and IV spliced directly to exon VIIa, excluding exons V and VI (Fig. 1). To determine if the RB3 cDNA was derived from a transcript that contains the BARF0 ORF, RT-PCR was performed with primers 10310-R (exon III) and 160537-L (exon VIIb). This resulted in detection of a product of a predicted size (583 bp) which was identified by hybridization with an internal oligonucleotide probe (160387-L, exon VIIb), indicating that the RB3 cDNA was also derived from a transcript that contains exons VIIa and VIIb. This transcript would encode BARF0, as well as a 51-codon ORF in exon III, and a 69-codon ORF spanning exons IV, VIIa, and VIIb (Fig. 2).

cDNA RK1 was isolated from C15 by using primers 9942-R (exon III) and 160537-L (exon VIIb, downstream of the BARF0 start codon) and contained exons IIIa, IIIb, and IV and distinct forms of exon V (Va′) and exon VII (VIIb′) (Fig. 1). Interestingly, exon Va′ contains an initiation codon and coding sequences that extend the size of the BARF0 ORF from 174 to 279 codons and would encode a protein of approximately 30 kDa (Fig. 2). Both RB2/BK1 and RK1 utilized the same exon V splice acceptor site at bp 155725, distinct from that for the 22.2 cDNA. Therefore, neither RB2/BK1 nor RK1 encoded the previously reported 71-codon ORF of c22.2, designated RPMS1, that spans exons IV and V in this cDNA (15). In RK1, however, this splice created a 103-codon ORF in an alternate reading frame (Fig. 2).

These additional cDNAs indicate that the multiple *Bam*HI A mRNAs are transcribed with additional exons and splicing patterns and with alternate splicing of exons III, IV, VI, VIIa, and VIIb. This complex pattern of transcription results in a number of previously unidentified ORFs that could be expressed during latent infection. The use of common splice donor and acceptor sites to generate distinct splicing patterns indicates that the variation in the cDNA structures accurately reflects the mRNA sequence and does not represent variation generated in the amplification and cloning process.

Splicing patterns of *Bam*HI A transcripts. The cDNAs identified in this study reveal that at least five distinctly spliced mRNAs comprise the *Bam*HI A family of transcripts (Fig. 1, cDNAs 22.2, 25a, RB2/BK1, RB3, and RK1). All of these cDNAs represent transcripts that include the BARF0 ORF in the final exon, suggesting that this ORF may be an important component of *Bam*HI A mRNA expression. Nevertheless, given that the cDNAs also encode unique ORFs that may also be expressed, it is important to determine which of these cDNAs represents the abundant 4.8-kb *Bam*HI A mRNA and which represent the larger transcripts that are detected at lower levels. Previous studies have used exon-specific and intron-specific oligonucleotide probes to determine the splicing patterns of the *Bam*HI A transcripts; however, these hybridizations cannot distinguish the various *Bam*HI A cDNAs which contain large regions of sequence overlap (7, 43).

To determine the splicing patterns contained within the differently sized transcripts detected by Northern blotting, RT-PCR was performed with gel-purified, size-selected mRNA fractions. C15 mRNA extracted from denaturing gels in fractions corresponding to the size of bands detected by Northern blotting (7.6, 6.2, and 4.8 kb) was used to synthesize cDNA. PCR with primer pairs that distinguish the unique splices within each of the *Bam*HI A cDNAs was followed by hybridization with internal oligonucleotide probes (Fig. 3A). The results demonstrated that each of the mRNA fractions contained multiple, distinctly spliced structures, indicating that complex patterns of alternate splicing are incorporated into transcripts of different size.

RT-PCR of the size-selected fractions performed using primers 159847-R (exon VIIa′) and 160387-L (exon VIIb) produced a 370-bp spliced product and an unspliced product of 541 bp, characteristic of the c25a and 22.2 cDNAs, respectively (Fig. 3A). By hybridization with an oligonucleotide probe (160004-R) in exon VII/VIIa′, the spliced product was predominantly detected in the 6.2-kb fraction and at a lower level in the 7.6- and 4.8-kb fractions, while the unspliced product was detected exclusively in the 7.6-kb fraction (Fig. 3B, probe 1). The identity of the unspliced product was confirmed by hybridization with an oligonucleotide probe (160070-R) within the intron of c25a (Fig. 3B, probe 2). This probe hybridized to the 541-bp unspliced product only in the 7.6-kb fraction but did not hybridize to the 370-bp product. A low level of hybridization to small, heterogeneous material was detected in the 6.2-kb fraction. This result indicated that the 22.2 cDNA is likely derived from a 7.6-kb *Bam*HI A mRNA. In addition, the detection of the unspliced form of exon VII only in the largest mRNA fraction demonstrated that the size-selected RNA fractions used in this study were discrete and that there was no cross-contamination of the fractions due to degradation of larger transcripts. Furthermore, the detection of both the unspliced and spliced forms of exon VII in the 7.6-kb fraction indicated that *Bam*HI A transcription of 7.6 kb consists of at least two differently spliced mRNAs.

RT-PCR using primers 157313-R (exon VI) and 160387-L (exon VIIb) produced a 350-bp spliced product characteristic of the RB2/BK1 cDNA (Fig. 3A). By hybridization with an oligonucleotide probe (157375-R) in exon VI, this product was detected at a very high level in the 4.8-kb mRNA fraction, at a trace level in the 7.6-kb fraction, and not at all in the 6.2-kb fraction (Fig. 3B, probe 3). This result demonstrated that the splicing pattern of the RB2/BK1 cDNA is contained within both the 4.8- and 7.6-kb mRNAs. To account for the difference in transcript size, these transcripts probably contain the RB2/BK1 exons alternately spliced to distinct 5′ exons. This result also demonstrated that *Bam*HI A transcription of 4.8 kb consists of at least two distinct splicing patterns (i.e., c25a and

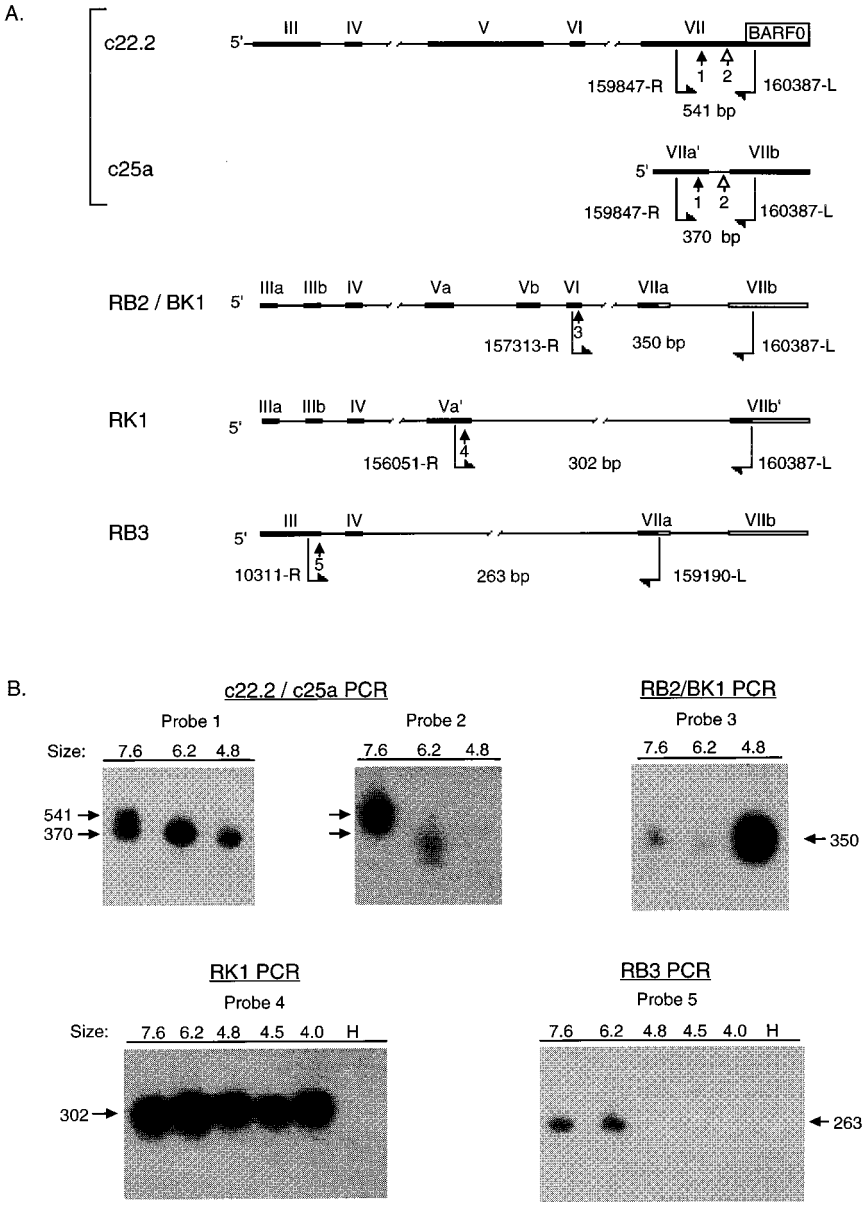

FIG. 3. RT-PCR analysis of size-selected C15 RNA fractions. (A) Detection of alternately spliced *Bam*HI A mRNAs. PCR primers (bent arrows, with genomic coordinates), oligonucleotide probes (straight arrows, numbered 1 to 5), and sizes of predicted products used for detection of alternately spliced *Bam*HI A mRNAs are indicated. The primers are numbered as in Fig. 1 at the 5′ ends and extend 5′ to 3′ in the direction indicated (R, rightward; L, leftward). (B) Southern blot hybridizations. Hybridizations to amplified products from size-selected mRNA fractions are shown for each set of PCR primers and oligonucleotide probes depicted in panel A. RNA fractions were excised from a denaturing gel; fraction sizes (in kilobases) were determined by comparison with RNA molecular weight standards and by hybridization of a parallel lane with a single-stranded RNA probe complementary to BARF0 sequences. Primers 159847-R and 160387-L amplify an unspliced product (541 bp) from c22.2 and a spliced product (370 bp) from c25a. Probe 1 hybridizes to both products, while probe 2, located within the intron of the c25a, hybridizes only to products having the unspliced structure. The identity of the unspliced, 541-bp product of the c22.2/c25a amplification was confirmed by hybridization of the same blot with probe 2, located within the c25a intron. Probes: 1, 160004-R; 2, 160070-R; 3, 157375-R; 4, 156180-L; 5, 10350-L. H, no cDNA (RT-PCR negative control).

RB2/BK1). In addition, the high level of detection of the RB2/BK1 cDNA in the 4.8-kb mRNA fraction suggests that the RB2/BK1 cDNA may represent a transcript that is expressed at high levels in vivo.

Further heterogeneity in the *Bam*HI A transcripts was detected with primers 156051-R (exon Va′) and 160387-L (exon VIIb), which resulted in amplification of a spliced product of 302 bp characteristic of the RK1 cDNA. By hybridization with an oligonucleotide probe (156180-L) in exon Va′, this product was detected at approximately equal levels in the 7.6-, 6.2-, and 4.8-kb fractions, as well as in a smaller 4.0 kb fraction, but not in the negative control (Fig. 3B, probe 4). This result suggested that the transcripts represented by RK1, and the 274-codon RK-BARF0 ORF that it encodes, may also be expressed at relatively high levels in vivo.

RT-PCR was also performed with primers 10311-R (exon III) and 159190-L (exon VIIa), which amplified a spliced product of 263 bp characteristic of the RB3 cDNA (Fig. 3A). This

product was detected by hybridization with an oligonucleotide probe (10350-L) in exon III in the 7.6- and 6.2-kb mRNA fractions but not in the smaller fractions of 4.8 and 4.0 kb (Fig. 3B, probe 5).

Finally, RT-PCR was performed with primers 157313-R (exon VI) and 159190-L (exon VII), which amplify a product of approximately 180 bp characteristic of the 22.2 and RB2/BK1 cDNAs. These cDNAs have predicted amplified products of 177 and 182 bp, respectively; however, this difference could not be distinguished in this assay. These primers span an intron of 1.7 kb in these cDNAs, and the sequence within this intron has been proposed to account for the larger size of the 6.2- and 7.6-kb transcripts (7, 43). Interestingly, however, the 180-bp spliced product was detected by hybridization with an oligonucleotide probe (157375-R) in all of the size-selected mRNA fractions but not in negative controls (data not shown). This result indicated that at least a subset of each of the differently sized *Bam*HI A mRNAs, including the larger mRNAs, do not contain the sequences of the intron between exons VI and VII.

In summary, this analysis demonstrated the presence of heterogeneous populations of *Bam*HI A transcripts within each of the discrete, size-selected mRNA fractions. The splicing patterns of certain cDNAs were detected in only one of the fractions (cDNA 22.2 in the 7.6-kb fraction), in a few of the fractions (RB3 in the 7.6- and 6.2-kb fractions and RB2/BK1 in the 4.8- and 7.6-kb fractions), or in all of the fractions (c25a and RK1). These results indicated the existence of complex patterns of alternate splicing that are incorporated into transcripts of different sizes. In addition, the identification of the RB2/BK1 and RK1 splicing patterns at relatively high levels in the major 4.8-kb mRNA fraction implicates these cDNAs, and the ORFs that they encode, as important components of *Bam*HI A transcription.

5′ analysis of the *Bam*HI A transcripts. A previous study has reported a transcriptional initiation site for the *Bam*HI A mRNAs and proposed that the major 4.8-kb mRNA initiates at this site and contains the splicing pattern of the 22.2 cDNA. It has also been proposed that the larger *Bam*HI A mRNAs initiate at this site but contain additional 3′ sequences from within the intron regions of the 22.2 cDNA structure. However, the existence of additional 5′ start sites was considered likely since previous studies have detected transcription only in the intron between exons VI and VII, and, as noted above, this intron is not present in at least a subset of the larger transcripts (7, 43). In addition, the length of the proposed structure for the 4.8-kb mRNA was only 4.3 kb. Furthermore, the unspliced exon VII characteristic of the 22.2 cDNA was detected only in the 7.6-kb mRNA fraction, not in the 4.8-kb fraction, while a subset of the transcripts of 7.6 and 6.2 kb include the relatively short, spliced structure of the RB3 cDNA. Therefore, to confirm the previously reported *Bam*HI A mRNA start site and to determine if additional start sites exist, 5′ RACE cloning, Northern blot hybridizations, and RT-PCR-based cDNA cloning were performed.

5′ RACE analysis was performed by PCR amplification of 5′-oligo(dA)-tailed C15 cDNA, using an oligo(dT) 5′ primer and a 3′ gene-specific primer within exon II (6572-L). Three identical cDNAs were isolated from independent amplifications and sequenced. These cDNAs were similar in structure to a previously reported cDNA having the same exon II splice acceptor site at bp 6514 but with an exon I splice donor site at bp 150769 differing by 5 nucleotides (nt) from that previously described at bp 150774 (43). 5′ RACE using a 3′ primer within exon III (9983-L) contained the identical structures of exon I, II, and III as identified in the shorter clones, differing from the previously reported exon I donor site. All of the 5′ RACE

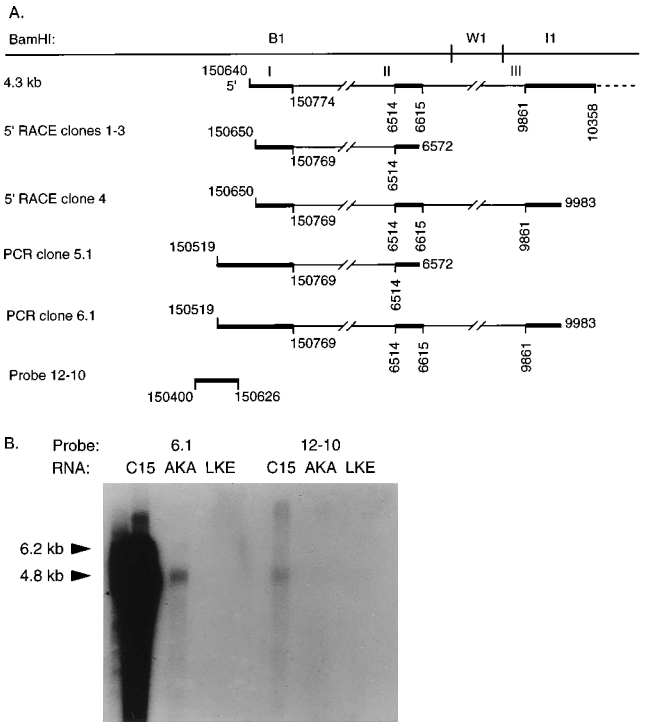

FIG. 4. 5′ analysis of *Bam*HI A transcripts. (A) Structures of C15 5′ RACE cDNAs. cDNAs isolated by amplifying with primer 6572-L in exon II (cDNAs 1-3) or primer 9983-L in exon III (cDNA 4) are shown. cDNAs 5.1 and 6.1 were isolated by RT-PCR of C15 mRNA, using primer 150519-R with 6572-L (c5.1) and 150519-R with 9983-L (c6.1). The 12-10 clone, which is 5′ of a previously reported transcription start site, was isolated from C15 DNA for use in Northern blot hybridizations (43). (B) Northern blot hybridization. Hybridization was performed with 30 μg of poly(A) mRNA per lane from C15, Akata (AKA), an EBV-positive BL line, and Loukes (LKE), an EBV-negative BL line, using leftward, single-stranded RNA probes derived from cDNA 6.1 and clone 12-10. The 12-10 probe hybridizes to messages of 4.8, 6.2, and 8.0 kb in C15.

cDNAs terminated at approximately bp 150650, in close agreement with the previously reported start site (Fig. 4A).

To determine if processed transcription initiated further 5′ to this site, C15 cDNA was amplified by using a 5′ primer (150519-R) located 120 nt 5′ to the identified start site and a 3′ primer within exon II (6572-L) or within exon III (9983-L). The amplified cDNAs, c5.1 and c6.1, respectively, contained the same splice patterns as the 5′ RACE cDNAs, but with exon I extended unspliced to the 5′ primer. This finding indicated the existence of transcription initiating further 5′ to the previously reported start site (Fig. 4A). The invariant structures of exons I and II in the 5′ RACE and RT-PCR cDNAs reported here suggest that this region of the *Bam*HI A transcripts is less heterogeneous than the 3′ exons.

To identify the *Bam*HI A mRNAs that contain sequences 5′ to the previously identified start site, a leftward, single-stranded RNA probe was prepared from a clone containing the sequences from bp 150400 to 150626 (probe 12-10; Fig. 4A) and hybridized to a Northern blot of C15, Akata, and Loukes (EBV-negative) mRNAs. Each lane contained 30 μg of poly(A) RNA, and equal loading was confirmed by ethidium bromide staining. In C15, probe 12-10 hybridized to the 4.8-kb mRNAs and to larger transcripts, which were visible upon longer exposure (Fig. 4B). Although hybridization to Akata RNA (EBV-positive BL line) was detected with the 6.1 probe, hybridization was not detected with probe 12-10 even upon long exposure. This result was consistent with the greatly reduced amount of

*Bam*HI A transcription in Akata detected by the 6.1 probe. Hybridization was not detected with either probe to RNA from the EBV-negative BL line Louckes (Fig. 4B). These results indicated that some of the larger *Bam*HI A mRNAs in C15 initiate further 5′ to the previously reported start site and also that a subset of the 4.8-kb mRNAs contain sequences 5′ to this start site. Thus, these data demonstrated that the *Bam*HI A mRNAs have multiple sites of transcription initiation in NPC, which is consistent with the identification of heterogeneous and alternately spliced structures in these transcripts.

3′ analysis of *Bam*HI A transcripts. The 3′ end of the BARF0 ORF is unusual in that there are no recognizable poly(A) signals located within several kilobases downstream of the ORF. A canonical poly(A) signal (AATAAA) is located within the ORF, 27 nt upstream of the stop codon, and a nonconsensus poly(A) signal (ATTAAA) overlaps the stop codon (Fig. 6A). Either of these signals could potentially direct cleavage 3′ of the BARF0 stop codon, giving rise to transcripts encoding the entire ORF. To determine if either or both of these signals are functional, or if a 3′ cryptic polyadenylation signal is used, cleavage sites of the *Bam*HI A transcripts were analyzed by 3′ RACE (Fig. 5). C15, B95-8, and Raji mRNAs were reverse transcribed by using a modified oligo(dT) primer containing a polylinker region. The 3′ ends of *Bam*HI A transcripts were amplified by PCR from these cDNAs by using 5′ primer 160853-R and a 3′ primer containing the polylinker sequence. Amplified products contained sequences from the 3′ end of BARF0 and from the poly(A) tail of the transcripts, allowing determination of the transcript cleavage site.

Twelve distinct cDNAs, that could be distinguished by the location of the cleavage site or by the length of the cloned poly(A) tail, were isolated from the three mRNA samples (five from C15, four from B95-8, and three from Raji). Three cDNAs from each sample were cleaved after the adenosine at bp 160990, 1 nt 5′ to the BARF0 stop codon, resulting in truncation of the ORF (Fig. 5B). Cleavage at this site is consistent with the use of the upstream poly(A) signal within the ORF (Fig. 5A, signal I). One cDNA from each mRNA sample was cleaved further 3′ within the four adenosine residues from 160993 to 160996. These cDNAs represent transcripts that contain the complete BARF0 ORF and include the stop codon. Cleavage in this region is also likely to be directed by signal I and demonstrates heterogeneous 3′-end processing of these transcripts. A single cDNA isolated from C15 terminated at bp 160986 and was unusual in that cleavage of the primary transcript occurred following a cytosine nucleotide (Fig. 5B).

To determine if heterogeneous 3′-end processing occurred consistently among the differently sized *Bam*HI A messages, the 3′ RACE protocol was modified to determine if the BARF0 stop codon was contained only in transcripts of a certain size. cDNAs containing the downstream cleavage site included an *Mse*I restriction site, TTAA, not present in those cDNAs that exclude the BARF0 stop codon. 3′ RACE was performed with cDNA from size-selected C15 mRNA fractions, and the amplified products were cut with *Mse*I. The presence of the *Mse*I site resulted in the release of a 140-bp fragment which was detected by electrophoresis on a 10% polyacrylamide gel followed by electroblotting and hybridization with an internal oligonucleotide probe. Transcripts containing the *Mse*I site, and therefore containing the complete BARF0 ORF, were detected with similar frequencies relative to the total amount of amplified product in each size-selected fraction (Fig. 5C). These data indicated that the *Bam*HI A transcripts, regardless of size, undergo heterogeneous 3′-end processing. Approximately 75% of the mRNAs are cleaved before the stop codon of the BARF0 ORF, while the remainder are cleaved further 3′, giving rise to processed transcripts that contain the entire ORF.

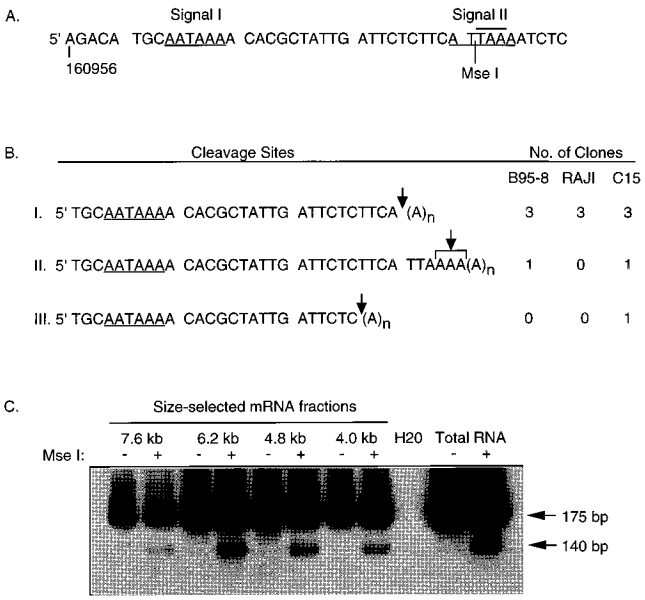

FIG. 5. 3′ analysis of *Bam*HI A transcripts. (A) Sequence of the 3′ end of the BARF0 ORF from position 160956. The two potential polyadenylation signals (underlined), the TAA stop codon (overlined), and the position of the *Mse*I restriction site are indicated. (B) Analysis of 3′ RACE cDNAs. cDNAs were isolated from B95-8, Raji, and C15 by RT-PCR using primer 160853-R and an oligo(dT)-adapter primer. Clones are divided into three groups according to the locations of the cleavage site (indicated by arrow), and the number of clones in each group from each sample is indicated. The exact cleavage site of type II clones, which are the only clones containing the BARF0 stop codon, could not be determined because of the run of dA in the sequence from bp 160993 to 160996. (C) 3′ analysis of size-selected C15 mRNA fractions. RNA fractions were excised from a denaturing gel, and fraction sizes were determined by comparison with RNA molecular weight standards and by hybridization of a parallel lane with a single-stranded RNA probe complementary to BARF0 sequences. RT-PCR was performed with primers 160853-R and the oligo(dT)-adapter, which amplified products of approximately 175 bp, depending on the location of the cleavage site and the length of amplified poly(A) tail. PCR products were cut with *Mse*I (+) or uncut (−), resolved by electrophoresis on a 10% polyacrylamide gel, blotted onto nitrocellulose, and hybridized with an internal oligonucleotide probe (160901-R). Transcripts containing the BARF0 stop codon give rise to products that are cleaved by *Mse*I (T/TAA), resulting in the release of a 140-bp fragment. C15 mRNA that was not size selected (Total RNA) was included as a positive control. A sample that contained no cDNA (H20) was also included as a negative control for RT-PCR.

Expression of alternately spliced *Bam*HI A mRNAs in vivo. The *Bam*HI A cDNAs presented in this and previous studies represent highly spliced transcripts that contain the BARF0 ORF and encode additional, unique ORFs by virtue of their distinct splicing patterns. Some of the splicing patterns present in the *Bam*HI A cDNAs have also been detected by RT-PCR in NPC biopsies, EBV-infected B cells, and EBV-positive LCLs (3). The primers used in the study by Brooks et al. (3), however, did not distinguish between the splicing patterns of the 22.2 and RK1 cDNAs or between the 22.2 and RB2/BK1 cDNAs. In this study, in order to characterize the expression in different tissue types of the *Bam*HI A transcripts represented by the reported cDNAs, RT-PCR was performed with primer pairs designed to detect the unique splicing patterns of each cDNA (Fig. 3A). RNA was isolated from additional passaged NPC tumors, primary-site NPC biopsies, metastatic-site NPC biopsies, a parotid carcinoma, a BL biopsy sample, type I (Akata and Mutu) and type III BL lines (Raji), and a type III LCL (B95-8). Equivalent aliquots of RNA that were not re-

TABLE 2. Expression of alternately spliced *Bam*HI A mRNAs in vivo[a]

Sample	Expression			
	c25a	RB2/BK1	RB3	RK1
Passaged NPC	4/4[b]	4/4	2/2	2/2
Biopsy				
Primary-site NPC	14/17	7/7	7/7	7/7
Secondary-site NPC	5/5	ND	ND	ND
BL	1/1	1/1	1/1	1/1
Parotid carcinoma	1/1	1/1	1/1	1/1
Cell line				
B95-8 (type III LCL)	+	+	+	+
Raji (type III BL)	+	+	+	+
Akata (type I BL)	+	+	+	+
Mutu (type I BL)	+	+	+	+

[a] RT-PCR and hybridizations to detect specific cDNAs were performed with oligo(dT)-primed total RNA, using primers and probes as shown in Fig. 3A and described in Materials and Methods. The passaged NPCs used for c25a and RB2/BK1 PCR were C15, C17, C18, and C19. C15 and C18 were used for RB3 and RK1 PCR. ND, not determined.

[b] Number expressing alternately spliced *Bam*HI A mRNA/number examined.

verse transcribed were processed and amplified in parallel for all samples as a control for false-positive signals caused by plasmid contamination or by PCR product carryovers. Transcripts represented by the c25a, RB2/BK1, RK1, and RB3 cDNAs were detected with high frequency in all tissues and cell types (Table 2). These data demonstrated that transcripts represented by each *Bam*HI A cDNA were consistently expressed in both epithelial and lymphoid samples. In addition, *Bam*HI A transcription was demonstrated for the first time in BL in vivo and was shown to include multiple, distinctly spliced forms.

DISCUSSION

This study has demonstrated a complex pattern of transcription for the *Bam*HI A mRNAs that includes multiple start sites, alternate splicing, and heterogeneous 3′-end processing. The analysis by RT-PCR of size-selected mRNA fractions has provided structural information that cannot be derived from Northern blot hybridizations used in previous studies. For example, in a previous study, hybridization with an oligonucleotide within the intron between exons IIIa and IIIb confirmed the presence of the unspliced exon III in a transcript of 4.8 kb, but this hybridization did not identify the exon IIIa-IIIb structure as part of these transcripts (43). In the present study, two cDNAs which contained the spliced exon IIIa-IIIb structure (RB2/BK1 and RK1) were detected by RT-PCR in mRNAs of 4.8 kb and in larger *Bam*HI A transcripts. Similarly, hybridization with an oligonucleotide probe within exon V did not distinguish transcripts containing the unspliced exon V of the 22.2 cDNA from those containing exon Va (RB2/BK1) or exon Va′ (RK1), and hybridization to exon VII did not distinguish transcripts containing the unspliced exon VII of the 22.2 cDNA from those containing exon VIIa (RB2/BK1 and RB3) or exon VIIa′ (c25a) (7, 43). In this study, the 4.8-kb mRNAs were shown to include the Va (RB2/BK1) and Va′ exon (RK1) as well as the VIIa′-VIIb exon structure of c25a. In contrast to a previous study, the unspliced exon VII structure of 22.2 was not detected in the 4.8-kb mRNAs but was present only in transcripts of 7.6 kb (43). In addition, a previous study using an oligonucleotide probe located within the intron between exons II and III demonstrated hybridization to a 4.8-kb mRNA, indicating that additional alternately spliced structures have yet to be identified (7).

The 4.8-kb *Bam*HI A mRNAs, which are the most abundant component of *Bam*HI A transcription, included transcripts containing the RK1 and the RB2/BK1 cDNA structures and, at a low level, the c25a structure. This finding suggests that the ORFs encoded by these cDNAs are perhaps an important component of *Bam*HI A expression. The largest of these ORFs are BARF0 (174 codons), present in all of the *Bam*HI A transcripts, RK-BARF0 (279 codons), which is an in-frame extension of BARF0 present in RK1, a 103-codon ORF also in RK1, and the 126-codon ORF spanning four exons of RB2/BK1. The initiation codons of these ORFs are in a favorable context for translation as defined by the consensus sequence for recognition by a scanning ribosomal subunit, A/GXXA UGG, with the exception of the RB2/BK1 126-codon ORF, which deviates from the consensus at the +1 position (AXXA UGA) (19). A purine at −3 is the more conserved element of this motif in eukaryotic mRNAs, however, and in fact several EBV genes expressed during latency, including EBNA-1, do not conform to the consensus sequence (18). The presence of extremely long 5′ untranslated regions that would be associated with these ORFs raises the possibility that internal ribosome binding rather than scanning might contribute to their efficient translation. Internal ribosome binding has been identified in picornavirus, cowpea mosaic virus, and certain cellular mRNAs (24, 29, 46).

Previously, sequence homology has been noted between BARF0 and the herpes simplex virus type 1 (HSV-1) ICP4 (IE175) protein consisting of 38% homology (20% identical and 18% similar residues) over a region of 178 codons (7). However, since similar levels of sequence homology are identifiable between this region of BARF0 and several other regions of ICP4, it is unlikely that this reflects functional homology between the two proteins. Likewise, the 103-codon ORF of RK1 is 66% similar (26% identical) to HSV-1 ICP4 over 94 codons yet 73% similar (34% identical) over 92 codons to the HSV-1 large tegument protein, UL36. No striking sequence homologies to known proteins were detected by examination of the *Bam*HI A ORFs by using the Genetics Computer Group sequence analysis software package (Genetics Computer, Inc.).

As has been previously noted, the BARF0 ORF is arginine rich (25 of 174 residues, 14 mol%), containing charged residues clustered in several regions of the ORF (12). Arginine-rich motifs (ARMs) mediate RNA binding in a number of viral and cellular proteins and consist of a highly basic stretch of 10 to 12 amino acid residues (4). An 11-amino-acid (aa) consensus sequence, which allows up to four mismatched residues, identifies this motif in the known ARM-containing, RNA-binding proteins. For example, adenovirus type 2 E4 protein differs from the consensus at two residues, while human immunodeficiency virus type 1 Tat differs at four residues (22). The hepatitis delta virus antigen contains two ARMs which differ from the consensus at four and six residues, respectively (23). Interestingly, within a 75-aa segment of the BARF0 ORF are three 11-aa regions (two of which overlap) that differ from the consensus ARM sequence at four residues and a fourth region that differs at five residues. This raises the possibility that the function of the putative BARF0 protein may involve an interaction with viral or cellular RNAs. Other ORFs of the *Bam*HI A cDNAs are similarly arginine rich, most notably the 103-codon ORF of RK1 (22 mol% arginine), but none contains a region having less than five mismatches to the ARM consensus sequence.

The RK-BARF0 ORF contained within the RK1 cDNA includes the entire BARF0 ORF with an additional 105 N-

terminal residues and would encode a protein of 30 kDa. This ORF contains a strongly hydrophobic region at the amino terminus (Leu-Val-Val-Val-Ala-Gly-Met-Val-Ala-Val-Ala), which closely resembles an endoplasmic reticulum-targeting signal peptide sequence. Signal peptides have been identified in numerous secretory and transmembrane proteins, and although no stringent consensus sequence exists, they contain characteristic features to which the putative signal peptide of RK-BARF0 conforms (30). The hydrophobic core of 10 to 12 aa, located at the amino terminus, most frequently includes a high proportion of alanine, leucine, and valine residues. Further, alanine residues in the middle of the core frequently occur in two- or three-residue symmetrical repeat elements (in RK-BARF0, Val-Ala-X-X-X-Ala-Val). The hydrophobic core of RK-BARF0 is immediately followed by a stretch of positively charged residues. This configuration may be sufficient to cause stop-translocation of the nascent peptide, resulting in a type III (N terminus extracytoplasmic, C terminus cytoplasmic) integral membrane protein (20, 48).

The BARF0 and RK-BARF0 ORFs are unusual in that the polyadenylation signal in each is internal to the ORF (5′ of the stop codon), and it has been suggested that this arrangement is inconsistent with translation of these ORFs (15). Functional internal polyadenylation signals have been reported for other genes, however, including the chicken growth hormone receptor gene and polyomavirus middle T antigen (PyMT) gene (27, 37). In this study, the internal polyadenylation signal of BARF0 has been shown to direct heterogeneous 3′-end processing. Approximately 25% of the transcripts contain the complete BARF0 ORF, while the remainder are truncated and do not include the stop codon. The 3′ RACE analysis has also indicated that transcripts containing the RK-BARF0 ORF are similarly processed. Heterogeneous 3′-end processing has been reported for several genes, including the mouse ribosomal protein L30, luteinizing hormone receptor, bovine prolactin, and hepatitis B virus surface antigen gene (14, 40, 42, 47). Interestingly, the internal polyadenylation signal of PyMT also directs heterogeneous 3′-end processing, and about 50% of the transcripts truncate the ORF. It is not known if the truncated PyMT transcripts are translated (37).

This study has demonstrated that a subset of *Bam*HI A mRNAs of 4.8 kb and larger initiate 5′ to a previously reported start site. The cDNAs containing these sequences were unspliced in exon I and were detected by Northern blotting as a small proportion of the total *Bam*HI A transcripts. Multiple start sites have also been identified for the EBV EBNA mRNAs, which initiate from two distinct promoters (Cp and Wp) for transcripts encoding EBNA-2, -3A, -3B, -3C, and -LP and from three promoters (Cp, Wp, and Fp) for transcripts encoding EBNA-1 (2, 38, 39, 41, 44, 45).

The *Bam*HI A transcripts encode a number of ORFs that are formed by alternate splicing, and this feature is also shared by other EBV latent-gene mRNAs. Alternate splicing creates a subset of EBNA transcripts that contain the initiation codon for the EBNA-LP ORF (35, 39, 45). In addition, the EBNA transcripts utilize common 5′ exons but are distinctly spliced to incorporate the various EBNA ORFs in the 3′ exons (17). The *Bam*HI A mRNAs, in contrast, are 3′-end coterminal but are distinctly cleaved to incorporate the BARF0 ORF stop codon in a subset of the transcripts.

Previous studies have demonstrated expression of the *Bam*HI A mRNAs in NPC and lymphoid samples, but the structures of these transcripts in various tissues had not been well characterized. In NPC, the relative level of *Bam*HI A transcription is significantly higher than in BL lines or in LCLs, and the detection by RT-PCR of several, distinct cDNAs in numerous NPC samples indicates that this elevated expression consists of many mRNA species. In lymphoid cells displaying both type I and type III gene expression, the pattern of transcription was shown to be similarly complex. The distinct cDNA structures were also demonstrated in parotid carcinoma and BL in vivo, for which *Bam*HI A transcription has not been previously reported.

The data presented in this study demonstrate additional levels of transcriptional complexity in the *Bam*HI A mRNAs. Multiple start sites and heterogeneous 3′-end processing suggest potential levels of regulation for the ORFs encoded by these alternately spliced transcripts. The BARF0 ORF is contained in all of these transcripts, and ORFs spanning multiple exons are created by distinct splicing patterns. These splicing patterns were detected in a wide range of EBV-infected cell types, suggesting that these ORFs may be a significant component of *Bam*HI A expression in vivo. The heterogeneous splicing of these transcripts may also serve as a regulatory mechanism by conferring upon a particular ORF a favorable context for translation by internal ribosomal loading. Although genetic studies have shown that the complete sequences spanned by the *Bam*HI A mRNAs can be deleted from the virus without affecting lymphocyte transformation, these transcripts may have an epithelial cell-specific function that is not assayable by the existing in vitro models of infection (34). The highly abundant expression of these transcripts in the latently infected epithelial cells of NPC suggests that the additional viral proteins potentially encoded by these RNAs are important for epithelial infection in vivo.

ACKNOWLEDGMENTS

This study was supported by grant CA32979 from the National Institutes of Health.

Tumor samples were provided by C.-L. Chen, M. Abdel-Hamid, and R. Pathmanathan, and Mutu RNA was provided by J. Sample.

REFERENCES

1. **Baer, R., R. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Sequin, P. S. Tufnell, and G. B. Barell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310:**207–211.
2. **Bodescot, M., M. Perricaudet, and P. J. Farrell.** 1987. A promoter for the highly spliced EBNA family of RNAs of Epstein-Barr virus. J. Virol. **61:** 3424–3430.
3. **Brooks, L. A., A. L. Lear, L. S. Young, and A. B. Rickinson.** 1993. Transcripts from the Epstein-Barr virus *Bam*HI A fragment are detectable in all three forms of latency. J. Virol. **67:**3182–3190.
4. **Burd, C. G., and G. Dreyfuss.** 1994. Conserved structures and diversity of functions of RNA-binding proteins. Science **265:**615–621.
5. **Busson, P., G. Ganem, P. Flores, F. Mugneret, B. Clauss, B. Caillou, K. Braham, H. Wakasugi, M. Lipinski, and T. Tursz.** 1988. Establishment and characterization of three transplantable EBV-containing nasopharyngeal carcinomas. Int. J. Cancer **42:**599–606.
6. **Busson, P., R. McCoy, R. Sadler, K. Gilligan, T. Tursz, and N. Raab-Traub.** 1992. Consistent transcription of the Epstein-Barr virus LMP2 gene in nasopharyngeal carcinoma. J. Virol. **66:**3257–3262.
7. **Chen, H. L., M. L. Lung, J. S. T. Sham, D. T. K. Choy, B. E. Griffin, and M. E. Ng.** 1992. Transcription of the BamHI A region of the EBV genome in NPC tissues and B cells. Virology **191:**193–201.
8. **Desgranges, C., H. Wolf, G. de Thé, K. Shanmugaratnam, R. Ellouz, N. Cammoun, G. Klein, and H. zur Hausen.** 1975. Nasopharyngeal carcinoma. X. Presence of Epstein-Barr virus genomes in epithelial cells of tumors from high and medium risk areas. Int. J. Cancer **16:**7–15.
9. **Fahraeus, R., H. L. Fu, I. Ernberg, J. Finke, M. Rowe, G. Klein, K. Falk, E. Nilsson, M. Yadav, P. Busson, T. Tursz, and B. Kallin.** 1988. Expression of Epstein-Barr virus-encoded proteins in nasopharyngeal carcinoma. Int. J. Cancer **42:**329–338.
10. **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.
11. **Gilligan, K., P. Rajadurai, J. Lin, P. Busson, M. Abdel-Hamid, U. Prasa, T. Tursz, and N. Raab-Traub.** 1991. Expression of the Epstein-Barr virus *Bam*HI A fragment in nasopharyngeal carcinoma: evidence for a viral pro-

tein expressed in vivo. J. Virol. **65:**6252–6259.
12. **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.
13. **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.
14. **Hu, Z. Z., E. Buczko, L. Zhuang, and M. L. Dufau.** 1994. Sequence of the 3′ non-coding region of the luteinizing hormone receptor gene and identification of two polyadenylated domains that generate the major mRNA forms. Biochim. Biophys. Acta **1220:**333–337.
15. **Karran, L., Y. Gao, P. Smith, and B. Griffin.** 1992. Expression of a family of EB virus complementary strand transcripts in latently infected cells. Proc. Natl. Acad. Sci. USA **89:**8058–8062.
16. **Kerr, B. M., A. L. Lear, M. Rowe, D. Croom-Carter, L. S. Young, S. M. Rookes, P. H. Gallimore, and A. B. Rickinson.** 1992. Three transcriptionally distinct forms of Epstein-Barr virus latency in somatic cell hybrids: cell phenotype dependence of virus promoter usage. Virology **187:**189–201.
17. **Kieff, E., and D. Liebowitz.** 1990. Epstein-Barr virus and its replication, p. 1889–1920. *In* B. N. Fields et al. (ed.), Virology, 2nd ed. Raven Press, New York.
18. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulate translation by eukaryotic ribosomes. Cell **44:** 283–292.
19. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108:**229–241.
20. **Kuroiwa, T., M. Sakaguchi, K. Mihara, and T. Omura.** 1991. Systematic analysis of stop-transfer sequence for microsomal membrane. J. Bio. Chem. **266:**9251–9255.
21. **Lagunoff, M., and B. Roizman.** 1994. Expression of a herpes simplex virus 1 open reading frame antisense to the $\gamma_1$34.5 gene and transcribed by an RNA 3′ coterminal with the unspliced latency-associated transcript. J. Virol. **68:** 6021–6028.
22. **Lazinski, D., E. Grzadzielska, and A. Das.** 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell **59:**207–218.
23. **Lee, C.-Z., J.-H. Lin, M. Chao, K. McKnight, and M. C. Lai.** 1993. RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus replication. J. Virol. **67:**2221–2227.
24. **Macejak, D. G., and P. Sarnow.** 1991. Internal initiation of translation mediated by the 5′ leader of a cellular mRNA. Nature (London) **353:**90–94.
25. **Miller, G., and M. Lipman.** 1973. Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. Proc. Natl. Acad. Sci. USA **70:**190–194.
26. **Nonoyama, M., C. H. Huang, J. S. Pagano, G. Klein, and S. Singh.** 1973. DNA of Epstein-Barr virus detected in tissue of Burkitt's lymphoma and nasopharyngeal carcinoma. Proc. Natl. Acad. Sci. USA **70:**3265–3268.
27. **Oldham, E. R., B. Bingham, and W. R. Baumbach.** 1993. A functional polyadenylation signal is embedded in the coding region of chicken growth hormone receptor RNA. Mol. Endocrinol. **7:**1379–1390.
28. **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.
29. **Pelletier, J., and N. Sonenberg.** 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature (London) **334:**320–325.
30. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167:**391–409.
31. **Pope, J., M. Horne, and W. Scott.** 1968. Transformation of fetal human leukocytes *in vitro* by filtrates of a human leukemic cell line containing herpes-like virus. Int. J. Cancer **3:**857–866.
32. **Raab-Traub, N., K. Flynn, G. Pearson, A. Huang, P. Levine, A. Lanier, and J. Pagano.** 1987. The differentiated form of nasopharyngeal carcinoma contains Epstein-Barr virus DNA. Int. J. Cancer **39:**25–29.
33. **Raab-Traub, N., and K. Gilligan.** 1991. Comparison of EBV transcription in lymphoid and epithelial cells, p. 339–362. *In* E. K. Wagner (ed.), Herpesvirus transcription and its regulation. CRC Press, Boston.
34. **Robertson, E. S., B. Tomkinson, and E. Kieff.** 1994. An Epstein-Barr virus with a 58-kilobase deletion that includes BARF0 transforms B lymphocytes in vitro. J. Virol. **68:**1449–1458.
35. **Rogers, R. P., M. Woisetschlaeger, and S. H. Speck.** 1990. Alternative splicing dictates translational start in Epstein-Barr virus transcripts. EMBO J. **9:**2273–2277.
36. **Rowe, M., D. T. Rowe, C. D. Gregory, L. S. Young, P. S. Farell, H. Rupani, and A. B. Rickinson.** 1987. Differences in B cell growth phenotype reflect novel patterns of Epstein-Barr virus latent gene expression in Burkitt's lymphoma cells. EMBO J. **6:**2743–2751.
37. **Salzman, N. P., V. Natarajan, and G. B. Selzer.** 1986. Transcription of SV40 and polyoma virus and its regulation. *In* N. P. Salzman (ed.), The Papovaviridae, vol. 1. The polyomaviruses. Plenum Publishing Corp., New York.
38. **Sample, J., L. Brooks, C. Sample, L. Young, M. Rowe, C. Gregory, A. Rickinson, and E. Kieff.** 1991. Restricted Epstein-Barr virus protein expression in Burkitt lymphoma is due to a different Epstein-Barr nuclear antigen 1 transcriptional initiation site. Proc. Natl. Acad. Sci. USA **88:**6343–6347.
39. **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.
40. **Sasavage, N., L. M. Smith, S. S. Gillam, R. P. R. P. Woychik, and F. M. Rottman.** 1982. Variation in the polyadenylation site of bovine prolactin mRNA. Proc. Natl. Acad. Sci. USA **70:**223–227.
41. **Schaeffer, B. C., M. Woisetschlaeger, J. L. Strominger, and S. H. Speck.** 1991. Exclusive expression of Epstein-Barr virus nuclear antigen 1 in Burkitt lymphoma arises from a third promoter, distinct from the promoters used in latently infected lymphocytes. Proc. Natl. Acad. Sci. USA **88:**6550–6554.
42. **Simonsen, C. C., and A. D. Levinson.** 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis virus chimeric plasmids. Mol. Cell. Biol. **3:**2250–2258.
43. **Smith, P. R., Y. Gao, L. Karran, M. D. Jones, D. Snudden, and B. E. Griffin.** 1993. Complex nature of the major viral polyadenylated transcripts in Epstein-Barr virus-associated tumors. J. Virol. **67:**3217–3225.
44. **Smith, P. R., and B. E. Griffin.** 1992. Transcription of the Epstein-Barr virus gene EBNA1 from different promoters in nasopharyngeal carcinoma and B-lymphoblastoid cells. J. Virol. **66:**706–714.
45. **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.
46. **Thomas, A. A. M., E. ter Haar, J. Wellnik, and H. O. Voorma.** 1991. Cowpea mosaic virus middle component RNA contains a sequence that allows internal binding of ribosomes and that requires eukaryotic initiation factor 4F for optimal translation. J. Virol. **65:**2953–2959.
47. **Weidemann, L. M., and R. M. Perry.** 1984. Characterization of the expressed gene and several processed pseudogenes of the mouse ribosomal protein L30 gene family. Mol. Cell. Biol. **4:**2518–2528.
48. **Wilson-Rawls, J., S. Deutscher, and W. S. Wold.** 1994. The signal-anchor domain of adenovirus E3-6.7K, a type III integral membrane protein, can direct adenovirus E3-gp19K, a type I integral membrane protein, into the membrane of the endoplasmic reticulum. Virology **201:**66–76.
49. **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.