# Analysis of the transcript encoding the latent Epstein–Barr virus nuclear antigen I: A potentially polycistronic message generated by long-range splicing of several exons

(primer extension/cDNA cloning)

SAMUEL H. SPECK AND JACK L. STROMINGER

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

ABSTRACT     The Epstein–Barr virus nuclear antigen (EBNA I) present in latently infected cells is encoded in a 2-kilobase exon contained in the BamHI K viral genomic fragment. This exon is, however, found within a 3.7-kilobase mRNA transcript. The origin of the remaining 1.7 kilobases is unknown, although it is not derived from adjacent Epstein–Barr virus DNA sequences. A 1.1-kilobase cDNA clone generated by primer extension using an oligonucleotide complimentary to a sequence 245 base pairs 3' to the putative initiation codon for EBNA I in the BamHI K fragment has been isolated. This clone contains seven exons (from the BamHI W, Y, U, E, and K viral genomic fragments), which are spread over approximately 70 kilobases of the viral genome. However, this clone does not appear to contain the complete 5' end of the transcript. In addition to the open reading frame in the exon encoding EBNA I, three other open reading frames are found in this transcript that potentially encode other viral antigens present in latently infected cells.

Epstein–Barr virus (EBV) is a human lymphotropic herpesvirus that is the etiologic agent of infectious mononucleosis and is also associated with two forms of malignancy, Burkitt lymphoma and nasopharyngeal carcinoma. Viral infection leads to the establishment of a latent "carrier state" in which the virus persists for life in B lymphocytes. Infection of cord blood lymphocytes in vitro with EBV results in a latent infection with little or no virus production and a concomitant growth transformation of the infected B lymphocytes (reviewed in refs. 1–3).

To date, several transcriptionally active regions of the viral genome have been identified in latently infected, immortalized B-lymphoblastoid lines (4). The most abundant transcript is a 2.8-kilobase (kb) message encoded in the BamHI Nhet fragment at the right-hand end of the genome (4, 5). Other transcripts present in latently infected cells have been mapped to the BamHI K fragment (a 3.7-kb transcript); to the BamHI W, Y, and H fragments at the left-hand end of the genome (a 3.0-kb transcript); and to the BamHI M fragment near the middle of the genome (6) (Fig. 1).

Of particular interest is the large transcript (3.7 kb) mapping to the BamHI K viral genomic fragment (4). Transfection of genomic fragments has identified this region as encoding one of the viral nuclear antigens (EBNA I) (8). Nuclease S1 protection experiments have identified only a single 2-kb exon from this region of the genome (9). Subsequent sequence analyses have clearly identified a 1912-base-pair (bp) open reading frame mapping to the left-hand end of the BamHI K fragment (10), consistent with the transfection data. However, there does not appear to be a consensus eukaryotic promoter in the 5' flanking sequences (10), al-

though there is a polyadenylylation signal 62 bp from the termination codon (10). It is likely, therefore, that the 2-kb exon encoding EBNA I is spliced to a region upstream from the BamHI K fragment. The possibility of long-range splicing in EBV transcripts in the productive cycle has already been discussed (11).

In this paper, we report the characterization of a cDNA clone containing a portion of the 5' sequences flanking the exon encoding EBNA I. Sequence analysis shows that this cDNA clone contains seven exons from the BamHI W, Y, U, and E fragments as well as the BamHI K fragment. In addition to the open reading frame in the BamHI K exon, there are three other open reading frames, which may encode other viral antigens present in latently infected cells.

## MATERIALS AND METHODS

Source of RNA. RNA was prepared from the EBV latently infected lymphoblastoid cell line, JY, established by infection of the peripheral blood lymphocytes of an individual from the Indiana Amish population (12). JY was grown at the Cell Culture Facility of the Massachusetts Institute of Technology (Cambridge, MA). Cells were harvested and stored frozen at −70°C.

RNA Preparation and Blotting. Total cellular RNA was prepared by the method of Auffray and Rougeon (13). Briefly, 5 g of frozen cells was suspended in 50 ml of lysis buffer (6 M urea/3 M lithium chloride containing heparin at 100 μg/ml) and disrupted in a Waring blender for 2 min. The lysate was stored on ice at 4°C for 4–16 hr to allow precipitation of the RNA. The RNA was recovered by centrifugation for 30 min at 10,000 × g, suspended in 20 ml of 10 mM Tris·HCl, pH 7.6/0.5% NaDodSO₄, and then extracted several times with phenol/chloroform, 1:1 (vol/vol), and with chloroform/isoamyl alcohol, 1:24 (vol/vol) and precipitated. Poly(A)⁺ RNA was isolated by fractionation on oligo(dT)-cellulose (14).

JY poly(A)⁺ RNA (10 μg) was fractionated by electrophoresis on a formaldehyde/agarose gel (35) and subsequently transferred to a nylon membrane (Pall Biodyne A). RNA blots were hybridized with a ³²P-labeled nick-translated BamHI K viral genomic fragment at 42°C for 15–20 hr in 50% (vol/vol) formamide/50 mM Hepes, pH 7.0/1x Denhardt's solution (17)/6x NaCl/Cit (1x NaCl/Cit 0.15 M NaCl/0.015 M sodium citrate, pH 7.0) containing denatured salmon sperm DNA at 100 μg/ml. The filters were washed twice with 2x NaCl/Cit/0.1% NaDodSO₄ for 15 min at room temperature and four times with 0.1x NaCl/Cit/0.1% NaDodSO₄ for 15 min at 68°C. Autoradiography was carried out for 12–48 hr at −70°C using a Dupont intensifying screen.
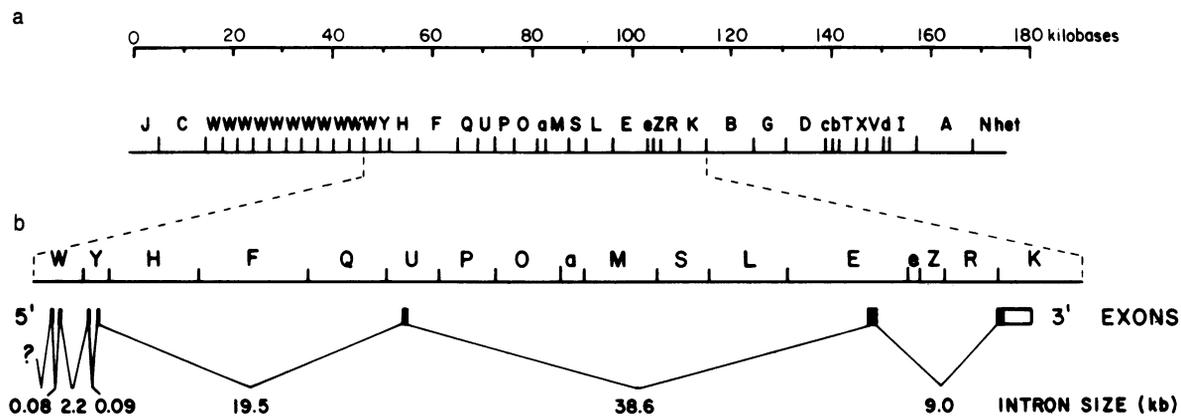
FIG. 1. (a) *Bam*HI restriction endonuclease map (8) of the B95-8 strain of EBV. (b) Proposed structure for the EBNA I transcript from the JY cell line. Filled-in boxes represent sequences present in the JYK2 cDNA clone. The indicated splicing of the 5' exon from the *Bam*HI W fragment is inferred from the results of Bodescot *et al.* (7).

**cDNA Synthesis.** A cDNA library was prepared from poly(A)$^+$ JY RNA by a modification of the method of Gubler and Hoffman (15). Briefly, 10 $\mu$g of poly(A)$^+$ RNA was denatured at 68°C for 3 min, quick cooled on ice, and incubated with oligo(dT)$_{12-14}$ (100 $\mu$g/ml) and a *Bam*HI K fragment-specific synthetic oligonucleotide (17-mer; 5' G-C-A-G-C-C-A-A-T-G-C-A-A-C-T-T-G 3', 300 $\mu$g/ml) for 30 min at 42°C in 70 mM Tris·HCl, pH 8.3/14 mM MgCl$_2$/6 mM sodium pyrophosphate/14 mM dithiothreitol/50 mM KCl containing RNasin (Promega Biotech, Madison, WI) at 1 unit/$\mu$l and nuclease-free bovine serum albumin (Enzo Biochemicals, New York) at 50 $\mu$g/ml; then, 1.25 mM dNTPs and avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL) at 3 units/$\mu$l were added and incubation was continued for an additional 30 min. Second-strand synthesis was carried out using RNase H (Bethesda Research Laboratories) and *Escherichia coli* DNA polymerase I (New England Biolabs) as described (15), except that *E. coli* DNA ligase and $\beta$-NAD$^+$ were omitted. The double-stranded cDNA was blunted by using T4 DNA polymerase (New England Biolabs), although we have since found this step to be unnecessary. *Eco*RI linkers (New England Biolabs) were ligated to the blunted cDNA with T4 DNA ligase (New England Biolabs), and this was followed by digestion with excess *Eco*RI and chromatography over a Sepharose CL-4B column to separate free linkers. The cDNA was cloned into *Eco*RI-digested calf intestinal alkaline phosphatase (Boehringer Mannheim)-treated $\lambda$gt10 (16), *in vitro* packaged, and plated on the C600 *hfl* strain of *E. coli*.

**Screening of the cDNA Library.** Approximately 5 × 10$^4$ unamplified phage was plated on a lawn of C600 *hfl* on each of twenty 150-mm plates. The phage was grown at 37°C until subconfluent (≈6 hr). Plaques were transferred to nitrocellulose filters, and the phage DNA was denatured by a 60-sec treatment with 0.5 M NaOH/1.5 M NaCl, which was followed by neutralization for 5 min with 0.5 M Tris·HCl, pH 8.0/1.5 M NaCl and washing with 0.36 M NaCl/20 mM NaH$_2$PO$_4$/2 mM EDTA for 5 min. The filters were dried at room temperature for 30 min and baked *in vacuo* at 80°C for 2 hr. They were then prehybridized for 4 hr at 68°C in 6× NaCl/Cit (1× NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0)/5× Denhardt's solution (17)/0.1% NaDodSO$_4$ containing denatured salmon sperm DNA at 100 $\mu$g/ml. $^{32}$P-labeled nick-translated probe (0.2 $\mu$g, 2–4 × 10$^8$ cpm/$\mu$g) was added and the mixture was hybridized for 16–20 hr at 68°C. Filters were washed twice in 2× NaCl/Cit/0.1% NaDodSO$_4$ for 15 min at room temperature and twice in 0.3× NaCl/Cit/0.1% NaDodSO$_4$ for 1 hr at 68°C. Autoradiography was carried out for 4–48 hr at −70°C using Dupont intensifying screens. Positive plaques were picked and plaque

purified. Inserts were excised from the phage DNA with *Eco*RI and subcloned into *Eco*RI-digested phosphatase-treated pUC18 vector.

**DNA Sequencing.** Appropriate restriction endonuclease-digested DNA fragments were subcloned into phage M13 mp18 and M13 mp19 directly from low-melting-point agarose as described by Crouse *et al.* (18). DNA sequencing was done by the chain termination method of Sanger *et al.* (19) using [$^{35}$S]dATP.

**Southern Blotting.** One microgram of high molecular weight DNA prepared from the EBV-transformed marmoset cell line B95-8, which carries >100 viral genomes per cell, was digested for 5 hr at 37°C with a 5-fold excess of *Bam*HI and was separated by electrophoresis on a 0.7% agarose gel. DNA was transferred to a nitrocellulose filter by the method of Southern (20). Hybridizations were carried out as described above for the screening of cDNA libraries except the probe concentration was ≈5 × 10$^6$ cpm/ml.

## RESULTS

Hybridization of a $^{32}$P-labeled nick-translated *Bam*HI K viral genomic fragment to total cellular poly(A)$^+$ RNA from the latently infected lymphoblastoid cell line, JY, revealed a single large transcript present in these cells (Fig. 2a). The size of the detected message, 3.9 kb, is in reasonable accord with that reported for another latently infected cell line, IB4 (4). RNA from the JY lymphoblastoid cell was subsequently used to generate cDNA libraries.

Repeated attempts to clone the 5' region flanking the exon encoding EBNA I, using oligo(dT) as a primer for cDNA synthesis, have failed to generate clones of sufficient length. To circumvent this problem, a $\lambda$ gt10 cDNA library was constructed employing as a primer a synthetic oligonucleotide homologous to a region near the 5' end of the EBNA I exon. The synthetic oligonucleotide used (17-mer) was homologous to a region of the transcript 228–245 nucleotides 3' to the putative initiation codon for EBNA I within the *Bam*HI K viral genomic fragment (10) and, in addition, oligo(dT) priming was used to ensure sufficient cDNA synthesis. Approximately 1 × 10$^6$ recombinants in this library (average insert size, 1–1.2 kb) were screened with the *Bam*HI K fragment and five positive clones were isolated. One of these clones (JYK2) contained sequences from the 5' end of the *Bam*HI K fragment open reading frame encoding EBNA I, as determined by screening with a specific 5' fragment of the *Bam*HI K fragment (500 bp, the *Bam*HI–*Nco* I fragment) and probably originated from extension of the oligonucleotide primer. The other clones resulted from priming of the transcript by oligo(dT), as determined by DNA sequence
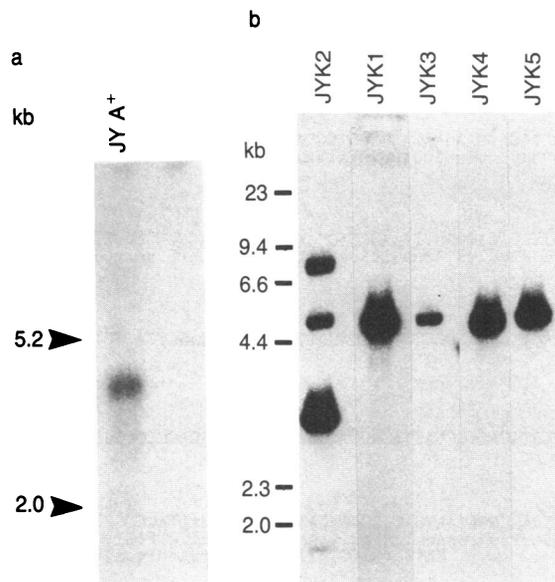
FIG. 2. (a) RNA blot of total cellular poly(A)$^+$ RNA from the B-lymphoblastoid cell line, JY, probed with the $^{32}$P-labeled nick-translated BamHI K viral genomic fragment. (b) Southern blot of BamHI restriction endonuclease-digested genomic DNA from the B95-8 cell line probed with $^{32}$P-labeled nick-translated EBNA I cDNA clones obtained from a cDNA library prepared from the JY cell line. The JYK2 cDNA clone hybridizes to the 8.0-kb (BamHI E), 5.1-kb (BamHI K), 3.2- to 3.4-kb doublet (BamHI U and W), and 1.8-kb (BamHI Y) viral genomic fragments. The other cDNA clones hybridize to a single fragment (BamHI K).

analysis (data not shown). A Southern blot of BamHI-digested genomic DNA from the EBV-infected marmoset cell line B95-8 (Fig. 2b), probed with the JYK2 cDNA clone, revealed hybridization of this clone to the BamHI K viral fragment and, in addition, to several other EBV DNA fragments. Subsequent sequence analysis (Fig. 3) confirmed the assignment of these bands to the BamHI W, Y, U, E, and K EBV genomic DNA fragments (10).

Comparison of the DNA sequence determined for the JYK2 cDNA clone to the EBV genomic DNA sequence (10) revealed that this cDNA clone has seven exons (Figs. 1b and 3). There are two exons (W1 and W2, see below) from within the BamHI W fragment (IRl repeats) of 63 bp and 131 bp, respectively, separated by an 82 bp intron. There are also two exons (Y1 and Y2) from the unique region (U2) of the BamHI Y fragment of 31 and 121 bp, respectively, separated by an 85-bp intron. The W2 and Y1 exons are separated by an 2208-bp intron. The Y2 exon is spliced approximately 19.5 kb downstream to a 172-bp exon from the BamHI U fragment. This exon in turn is spliced approximately 38.6 kb downstream to a 367-bp exon from the BamHI E fragment, which is spliced to the BamHI K exon 9.0 kb away. Thus, the latent transcript encoding EBNA I is composed of exons spread over at least 70 kb of the viral genome.

The appropriate splice donor and acceptor are present in the flanking regions of each exon (Table 1), making it very unlikely that this cDNA is an artifact of cloning. Furthermore, since the latent EBV transcripts are very low abundance messages (4), it is extremely improbable that two cDNA clones encoding latent viral transcripts would be ligated together. Indeed, the cDNA library employed in this study has been screened for a number of other higher abundance cellular transcripts, and there is no evidence that blunt end cDNA dimers are present in the library.

Interestingly, the four exons from the BamHI W and Y fragments have also been found in another viral transcript from the EBV latently infected cell line Raji (22). The W1 and W2 exons (encoded in the 3.2-kb IRl direct repeat sequences)

are repeated at least once, in that transcript, and probably several times, since the cDNA clone isolated by Bodescot et al. (22) did not extend to the 5′ end of the message. In addition, this cDNA clone contained a third exon (Y3) from the BamHI Y fragment and an exon from the BamHI H fragment (22), clearly not present in JYK2.

An examination of the 5′ region of the EBNA I transcript in JYK2 reveals three other open reading frames in addition to the open reading frame encoding EBNA I (Fig. 3). The first open reading frame extends through exons W1, W2, and Y1 and nearly to the end of Y2. The JYK2 cDNA has a 5′ terminus 3 bp from the 5′ boundary of the W1 exon found by Bodescot et al. (22). It is likely, as in the transcript from RAJI cells, that the W1 and W2 exons are repeated several times in the EBNA I transcript since JYK2 accounts for 900 bp of the estimated 1.7 kb of 5′ sequences flanking the exon encoding EBNA I (9). In addition, there may be an exon(s) 5′ to the W1 and W2 exons in the transcript.

No open reading frame is apparent in the BamHI U exon. However, two short open reading frames are present in the BamHI E exon that use alternative reading frames. The larger open reading frame begins 9 bp from the 5′ end of the BamHI E exon and terminates 48 bp within the BamHI K exon. It would encode a protein of 135 residues (15 kDa). The shorter open reading frame starts in the BamHI E exon 57 nucleotides after the start of the longer open reading frame and terminates within the BamHI E exon 216 bp downstream. It would encode a protein 72 residues long (8.4 kDa). Interestingly, there are three potential initiation codons within 37 bp clustered at the start of the shorter open reading frame. It is also noteworthy that the 5′ exon border for the exon encoded in the BamHI E genomic fragment is only 9 nucleotides from the initiation codon for the putative 15-kDa protein, nearly the same distance as the splice acceptor site for the EBNA I exon from its initiation codon (Fig. 3).

## DISCUSSION

In an effort to characterize the 5′ region flanking the exon encoding EBNA I in the transcript from an EBV-immortalized B-lymphoblastoid cell line, a cDNA library was constructed using a specific oligonucleotide primer. From this library, a single clone was recovered that contained 5′ flanking sequences. This cDNA clone, JYK2, is composed of seven exons that are spliced together over ≈70 kb of the viral genome. The data presented suggest that a very long primary transcript is subsequently spliced to generate the EBNA I transcript. Such long primary transcripts have been postulated to account for the long-range splicing exhibited in transcripts from the antennapedia and bithorax homeotic loci of Drosophila (7, 23). Alternatively, discontinuous transcription, recently postulated to account for the common spliced leader sequence present in trypanosome transcripts (24, 25), or trans-splicing (26) are possible mechanisms by which the EBNA I transcript might be generated. In addition, alternative splicing patterns may exist, giving rise to a heterogenous population of transcripts, as in the case of the developmentally regulated splicing of the troponin T transcript in rat cells (27). The nature of the EBNA I transcript in P3HR-1 cells will be of interest in this connection, since the entire BamHI Y fragment is deleted in the P3HR-1 virus and the EBNA I transcript has been reported to be 1.2 kb long (28).

Clearly, another important question raised by these data is whether this cDNA clone corresponds to the structure of the mature message or whether it represents a spliced intermediate. This is especially relevant since the RNA preparation used in the cDNA synthesis was the polyadenylylated fraction from total cellular RNA not just the cytoplasmic polyadenylylated RNA fraction. However, since RNA blots of total cellular polyadenylylated RNA from JY cells identify only a single transcript of 3.9 kb (Fig. 2a), similar to that

**W1**

```
(EcoRI Linker) AGGGGAGACCGAAGTGAAGGCCCTGGACCAACCCGGCCCGGGCCCCCCGGTATCGGGCCAGAGGGTCCCCTCGGACAGCTCCTAAGAAGGCACCGGTCGC 100
               ArgGlyAspArgSerGluGlyProGlyProThrArgProGlyProProGlyIleGlyProGluGlyProLeuGlyGlnLeuLeuArgArgHisArgSerP
```

**W2**

```
CCAGTCCTACCAGAGGGGGCCAAGAACCCAGACGAGTCCGTAGAAGGGTCCTCGTCCAGCAAGAAGAGGAGGTGGTAAGCGGTTCACCTTCAGGGCCACT 200
roSerProThrArgGlyGlyGlnGluProArgArgValArgArgArgValLeuValGlnGlnGluGluGluValValSerGlySerProSerGlyProLe
```

**Y1**                                                      **Y2**

```
ACGGCCACGTCCCCGGCCTCCAGCTCGGTCTCTTAGAGAGTGGCTGCTACGCATTAGAGACCACTTTGAGCCACCCACAGTAACCACCCAGCGCCAATCT 300
uArgProArgProArgProProAlaArgSerLeuArgGluTrpLeuLeuArgIleArgGlyHisPheGluProProThrValThrThrGlnArgGlnSer
```

```
GTCTACATAGAAGAAGAAGAGGATGAAGACTAAGTCACAGGCTTAGCCAGGTAACTTAGGAAGCGTTTCTTGAGCTTCCCTGGGAGAGCTGATTCTGCAG 400
ValTyrIleGluGluGluGluAspGluAsp
```

```
CCCAGAGAGTAGTCTCAGGGCATCCTCTGGAGCCTGACCTGTGATCGTCGCATCATAGACCGCCAGTAGACCTGGGAGCAGATTCACCGCCGCGGCCGTC 500
```

**U**

```
TCCTTTAAGACCCCACCATGGAATCATTTGAAGGACAGGGGGACTCTAGACAGTCACCCGACAATGAGCGGGGAGATAATGTACAGACTACCGGCGAGCA 600
               METGluSerLeuGluGlyGlnGlyAspSerArgGlnSerProAspAsnGluArgGlyAspAsnValGlnIleThrGlyGluHi
                                                           METSerGlyGluIleMETTyrArgLeuProAlaSerM
```

```
TGATCAGGACCCTGGGCCGGGGCCTCCATCCAGTGGGGCTTCTGAGAGATTGGTACCAGAAGAGTCATACTCAAGAGATCAGCAACCTTGGGGGCAAAGC 700
sAspGlnAspProGlyProGlyProProSerSerGlyAlaSerGluArgLeuValProGluGluSerTyrSerArgAspGlnGlnProTrpGlyGlnSer
ETIleArgThrLeuGlyArgGlyLeuHisProValGlyLeuLeuArgAspTrpTyrGlnLysSerHisThrGlnGluIleSerAsnLeuGlyGlyLysAl
```

**E**

```
AGGGGTGATGAAAACAGAGGCTGGATGCAGCGCATCAGGCGAAGGCGGAGAAGACGGGCTGCCTTGTCCGGCCATCTTTTAGACACGGAAGACAATGTGC 800
ArgGlyAspGluAsnArgGlyTrpMETGlnArgIleArgArgArgArgArgArgAlaAlaLeuSerGlyHisLeuLeuAspThrGluAspAsnValP
aGlyValMETLysThrGluAlaGlyCysSerAlaSerGlyGluGlyGlyGluAspGlyLeuProCysProAlaIlePhe
```

```
CGCCATGGTTGCCTCCACACGACATCACACCATATACCGCAAGGAATATCAGGGATGCTGCCTGCCGGGCTGTCAAGTGTGAATCATGTCTGACGAGGGG 900
roProTrpLeuProProHisAspIleThrProTyrThrAlaArgAsnIleArgAspAlaAlaCysArgAlaValLysCysGluSerCysLeuThrArgGl
                                                                     METSerAspGluGly
```

```
CCAGGTACAGGACCTGGAAATGGCCTAGGAGAGAAGGGAGACACATCTGGACCAGAAGGCTCCGGCCGCAGTGGACCTCAAAGAAGAGGGGGTGATAACC 1000
yGlnValGlnAspLeuGluMETAla
ProGlyThrGlyProGlyAsnGlyLeuGlyGluLysGlyAspThrSerGlyProGluGlySerGlyArgSerGlyProGlnArgArgGlyGlyAspAsnH
```

**K**

```
ATGGACGAGGACGGGGAAGAGGACGAGGACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGATGGTGTCCGGAG 1100
isGlyArgGlyArgGlyArgGlyArgGlyGlyGlyArgProGlyAlaProGlyGlySerGlySerGlyProArgHisArgAspGlyValArgAr
```

```
GCCCCAAAAACGTCCAAGTTGCATTGGCTGC (EcoRI Linker)
gProGlnLysArgProSerCysIleGlyCys
```

FIG. 3. Nucleotide sequence of the JYK2 cDNA clone and translation of the open reading frames. The exon borders are indicated by arrowheads, and the exons are labeled corresponding to the appropriate *Bam*HI genomic restriction fragments. The bar over the 3' 17 bp of the JYK2 cDNA sequence indicates the region homologous to the primer used to generate this clone.

observed for cytoplasmic polyadenylylated RNA from the latently infected IB4 cell line (4), it is unlikely that a nuclear precursor has been cloned. Furthermore, S1 nuclease protection experiments showed that the 5' sequences flanking the *Bam*HI K exon in the EBNA I transcript must be spliced from relatively far upstream (9), consistent with the structure of the cDNA clone characterized in this paper.

The location of the latent promoter of the EBNA I transcript has not yet been mapped. The promoter for the latent transcript of RAJI cells, which shares the W1, W2, Y1, and Y2 exons in common with the EBNA I transcript, has been suggested to map within the *Bam*HI C fragment (22). These two transcripts, and possibly other latent messages, may use a common promoter. Furthermore, since a putative EBV origin of replication (*oriP*), which requires the presence of EBNA I to function, is located in the *Bam*HI C fragment (29), it is intriguing to speculate that the region containing *oriP* may also contain the latent promoter for the transcription of the EBNA I message. Such an organization would be

similar to the tight linkage of the simian virus 40 origin and early promoter (30). Indeed, based on the genomic sequence, Baer *et al.* (10) have identified two putative RNA polymerase II promoters in the *Bam*HI C fragment, one of which is near the region containing *oriP*. A polymerase II promoter was also found in the *Bam*HI W (IR*l* repeat) fragment and might also initiate transcription of the EBNA I transcript (10).

Eukaryotic messages are generally monocistronic. However, in the case of the EBNA I transcript, there are three other open reading frames in addition to the EBNA I open reading frame. The open reading frame, which extends through the W1, W2, and Y1 exons and nearly to the end of the Y2 exon, may be translated in addition to the EBNA I exon, since this same open reading frame is found in another latent viral transcript in which it is the only open reading frame (22). The region of the EBV genome containing the *Bam*HI Y fragment is of particular interest since deletion of this region results in nontransforming viruses (31), and it is likely that this deletion is functionally important in establishing

Biochemistry: Speck and Strominger

*Proc. Natl. Acad. Sci. USA 82 (1985)* 8309

Table 1. Analysis of splice junctions in the JYK2 cDNA clone

| Consensus sequence | Donor | Acceptor |
|---|---|---|
| Exon border | $^C_A$–A–G / G–T–$^A_G$–A–G–T | $\left(^T_C\right)_n$ N–$^C_T$–A–G / G |
| W2/W1* | G–G–G / $\boxed{G-T}$ A–A–G–T | (C–G–C–C–A–T–C) C–A$\boxed{A-G}$ / C |
| W1/W2 | G–A–G / $\boxed{G-T}$ A–A–G–T | (A–C–C–C–G–T–C) T–C$\boxed{A-G}$ / G |
| W2/Y1 | G–G–G / $\boxed{G-T}$ A–A–G–T | (T–T–A–C–A–A–C) C–A$\boxed{A-G}$ / C |
| Y1/Y2 | C–G–G / $\boxed{G-T}$ A–A–G–T | (T–T–C–C–A–A–T) G–T$\boxed{A-G}$ / T |
| Y2/U | C–A–G / $\boxed{G-T}$ G–A–T–T | (A–A–T–T–T–C–T) G–C$\boxed{A-G}$ / G |
| U/E | A–A–G / $\boxed{G-T}$ G–C–T–G | (C–A–T–A–T–T–T) T–C$\boxed{A-G}$ / A |
| E/K | A–A–G / $\boxed{G-T}$ G–A–G–T | (A–T–C–T–C–T–T) T–T$\boxed{A-G}$ / T |

Consensus sequences are from Mount (21). The intron borders that are not present in the cDNA clone were obtained from the EBV genomic sequence (10).
*It is possible that the 5' end of the transcript, which has not yet been obtained, contains repeats of the W1 and W2 exons as found in another latent cDNA clone (7). In this case, a W2/W1 exon border would occur and it is shown for comparison.

the nontransforming phenotype (32). Unfortunately, since the JYK2 cDNA clone does not contain the complete 5' region, the complete sequence of the protein putatively encoded in this region is not available. However, its carboxyl-terminal portion is obviously both arginine and proline rich (ref. 7 and Fig. 3).

The splice junction between the *Bam*HI U and *Bam*HI E exons is 9 bp in front of the longer open reading frame in the *Bam*HI E exon. Examination of the putative initiation codons for the two open reading frames in the *Bam*HI E exon, as well as that for the EBNA I open reading frame, reveals that they all conform to the generalized favorable start sequence of R-N-N-A-T-G(G) (33) except at the last position (where N = any nucleotide, R = purine). Moreover, the initiation sequence for the longer *Bam*HI E exon open reading frame is identical to the consensus eukaryotic initiation sequence C-C-R-C-C-A-T-G(G) (33). Thus, based on the presence of a consensus initiation sequence it seems likely that this open reading frame is translated in latently infected cells. If so, it would encode a 15 kDa-protein with an overall charge of about −2. The distribution of charged residues appears to preclude its being a membrane antigen (Fig. 3). One striking feature is that, like EBNA I and the putative antigen encoded by the W1, W2, Y1, and Y2 exons, this putative viral protein would be relatively proline rich. Furthermore, both of these proteins contain a stretch of basic amino acid residues similar, although not identical, to a proposed nuclear targeting signal identified for Simian virus 40 large tumor antigen (34).

The data presented here underscore the importance of precise characterization of the latent viral transcripts in EBV-transformed lymphocytes. It may be possible to use antisera to β-galactosidase fusion proteins, constructed with the various open reading frames contained in the EBNA I transcript, to address the question of expression of these proteins in latently infected cells and obtain information pertinent to the understanding of the mechanisms by which EBV regulates its own gene expression, as well as the growth of infected B lymphocytes.

1. de-Thé, G. (1980) in *Viral Oncology*, ed. Klein, G. (Raven, New York), pp. 769–798.
2. Miller, G. (1980) in *Viral Oncology*, ed. Klein, G. (Raven, New York), pp. 713–738.
3. Kieff, E., Dambaugh, T., Hummel, M. & Heller, M. (1983) in *Advances in Viral Oncology*, ed. Klein, G. (Raven, New York), Vol. 3, pp. 133–182.
4. van Santen, V., Cheung, A. & Kieff, E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1930–1934.
5. Fennewald, S., van Santen, V. & Kieff, E. (1984) *J. Virol.* **51**, 411–419.
6. Grogan, E. A., Summers, W. P., Dowling, S., Shedd, D., Gradoville, L. & Miller, G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7650–7653.
7. Garber, R. L., Kuraiwa, A. & Gehring, W. J. (1983) *EMBO J.* **2**, 2027–2036.
8. Summers, W., Grogan, E., Shedd, D., Robert, M., Lui, C. & Miller, G. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5688–5692.
9. Hennessy, K., Heller, M., van Santen, V. & Kieff, E. (1983) *Science* **220**, 1396–1398.
10. Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatful, G., Hudson, G. S., Satchwell, S. C., Séguin, C., Tuffnell, P. S. & Barrell, B. G. (1984) *Nature (London)* **311**, 207–211.
11. Cohen, L. K., Speck, S. H., Roberts, B. E. & Strominger, J. L. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4183–4187.
12. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 910–914.
13. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
14. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
15. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
16. Huynh, T., Young, R. & Davis, R. (1985) in *Practical Approaches in Biochemistry*, ed. Grover, D. (IRL, Oxford), pp. 49–78.
17. Denhardt, D. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–652.
18. Crouse, G. F., Forschauf, A. & Lehrach, T. (1983) *Methods Enzymol.* **101**, 78–79.
19. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
21. Mount, S. (1982) *Nucleic Acids Res.* **10**, 459–471.
22. Bodescot, M., Chambraud, B., Farrell, P. & Perricaudet, M. (1984) *EMBO J.* **3**, 1913–1917.
23. Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. & Hogness, D. S. (1983) *Science* **221**, 23–29.
24. Parsons, M., Nelson, R. G., Watkins, K. P. & Agabian, N. (1984) *Cell* **38**, 309–316.
25. Guyaux, M., Cornelissen, A. W. C. A., Pays, E., Steinert, M. & Borst, P. (1985) *EMBO J.* **4**, 995–998.
26. Konarska, N. N., Padgett, R. A. & Sharp, P. A. (1985) *Cell*, in press.
27. Medford, R. M., Nguyen, H. T., Destree, A. T., Summers, E. & Nadal-Ginard, B. (1984) *Cell* **38**, 409–421.
28. Weigel, R. & Miller, G. (1983) *Virology* **125**, 287–298.
29. Yates, J. L., Warren, N. & Sugden, B. (1985) *Nature (London)* **313**, 812–815.
30. Griffin, B. E. (1980) in *DNA Tumor Viruses*, ed. Tooze, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 61–123.
31. Miller, G., Robinson, J., Heston, L. & Lipman, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4006–4010.
32. Skare, J., Farley, J., Strominger, J. L., Fresen, K. O., Cho, M. S. & zur Hausen, H. (1985) *J. Virol.*, in press.
33. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
34. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. (1984) *Cell* **39**, 499–509.
35. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* **16**, 4743–4751.