J. gen. Virol. (1989), **70**, 3079–3084. *Printed in Great Britain*

Key words: *EBV/transcription/latent cycle*

The Terminal Protein Gene 2 of Epstein–Barr Virus Is Transcribed from a Bidirectional Latent Promoter Region

By GERHARD LAUX, ANDROULLA ECONOMOU AND PAUL J. FARRELL*

Ludwig Institute for Cancer Research, St Mary's Branch, St Mary's Hospital Medical School, Norfolk Place, London W2 1PG, U.K.

(*Accepted 10 July 1989*)

SUMMARY

The intact terminal protein genes (TP1 and TP2) of Epstein–Barr virus (EBV) are created upon infection by circularization of the linear viral genome at its terminal repeats. The structure of the 1·7 kb TP2 latent mRNA has been determined by cDNA analysis and Northern blotting, revealing its close relation to TP1 mRNA. The 1·7 kb transcript is expressed from a different promoter and has a different 5′ exon from TP1 but is also spliced across the terminal repeats. The last eight exons are common to the TP1 and TP2 RNAs. The TP2 promoter is 3·3 kb downstream of the TP1 promoter and is part of a bidirectional latent EBV promoter region transcribing the TP2 and the latent membrane protein RNAs in opposite directions.

Epstein–Barr virus (EBV), a ubiquitous human herpesvirus, causes infectious mononucleosis and is linked to African Burkitt's lymphoma (BL) and undifferentiated nasopharyngeal carcinoma (reviewed in Epstein & Achong, 1986). The virus efficiently immortalizes human B lymphocytes in tissue culture (Henle *et al.*, 1967) and a latent infection is usually established in human lymphocytes. After infection, the linear viral genome of about 170 kb dsDNA circularizes via its terminal repeats and is amplified in copy number (Sugden *et al.*, 1979; Hurley & Thorley-Lawson, 1988). It is maintained as episomes in the immortalized cells (Lindahl *et al.*, 1976) but there are occasionally integrated copies as well (Matsuo *et al.*, 1984). About 11 of the approximately 90 EBV genes are constitutively expressed in latently infected lymphocytes. These latent cycle EBV genes (reviewed in Farrell, 1989), which include six different nuclear proteins, EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C and the leader protein (EBNA4, EBNA5), the latent membrane protein (LMP), the EBER RNAs and two terminal proteins (TP1 and TP2), appear to achieve and maintain the virus-induced cell proliferation.

The two TP RNAs were identified on Northern blots by Hudson *et al.* (1985) and this report concerns the characterization of the 1·7 kb RNA encoding TP2. The organization and expression of the TP1 mRNA, which is 2·0 kb in length and consists of nine exons (Fig. 1), have been described previously (Laux *et al.*, 1988). The first exon is located in the long unique region of the EBV genome (Baer *et al.*, 1984) at the right-hand end (nucleotides 166498 to 166916) whereas all the other exons are derived from the short unique region so the transcription unit crosses the terminal repeats. While the studies described in this manuscript were in progress, the sequence of a cDNA clone (LMP-2B) for the TP2 mRNA from IB-4 cells was reported (Sample *et al.*, 1989). That cDNA clone sequence started at nucleotide 169752. We now report the characterization of several cDNA clones from a B95-8 library (Bodescot *et al.*, 1986) which are derived from the 1·7 kb TP2 transcript. TP2 and TP1 are closely related genes and both require the circularization of the linear viral DNA molecule at the terminal repeats for their expression. We deduce that TP2 is transcribed from a bidirectional latent promoter region which expresses the LMP in the opposite direction.

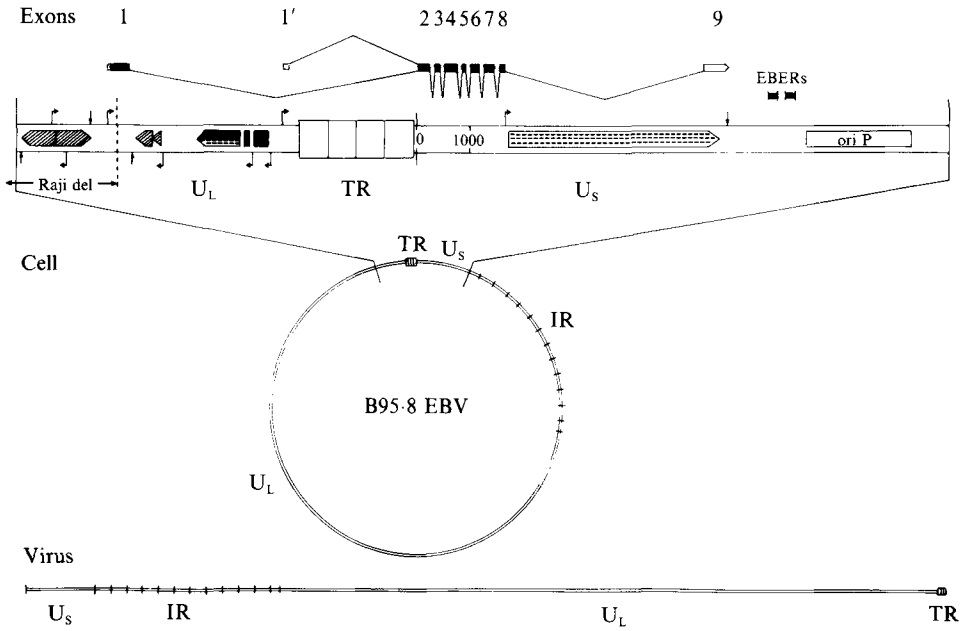

Fig. 1. Schematic representation of the EBV genome. The internal repeats (IR) divide the genome into the unique short (U_S) and unique long (U_L) regions. The linear viral dsDNA molecule (bottom line) circularizes via its terminal repeats (TR) in the infected cell (middle part). Part of the episomal EBV genome is enlarged (top part) showing major open reading frames inside the double line as broad boxes, pointed to show direction and shaded according to their expression class (solid black for latent, diagonally hatched for early productive and horizontally dashed for late productive cycle genes). Above and below the double line, small vertical and small bent arrows represent polyadenylation signals and promoters, respectively. Locations of the origin of episomal DNA replication (ori P), two small polymerase III transcripts (EBERs) and part of a deletion in the Raji EBV genome (Raji del) are also indicated. The structure of the cDNAs relative to the viral genome is shown above the double line. The TP1 cDNA derived from the 2·0 kb mRNA consists of exons 1 to 9 (excluding exon 1′) whereas the TP2 cDNA derived from the 1·7 kb mRNA consists of exons 1′ to 9. Coding and non-coding regions are indicated by solid black and open rectangles, respectively. The most 3′ exon is pointed to show the direction of transcription. Splicing is represented by thin lines between the exons.

We already knew that the TP RNA structures overlapped (Hudson *et al.*, 1985; Laux *et al.*, 1988) and that exon 1 of TP1 is not present in TP2 mRNA but exon 2 of TP1 hybridizes to both TP1 and TP2 mRNA. It therefore seemed likely that clones hybridizing to the second but not to the first exon of TP1 cDNA could be derived from TP2 transcripts. Twenty positive recombinant λgt10 bacteriophages had been isolated by screening the B95-8 cDNA library with the M13 clone 349.EI (B95-8 EBV nucleotides 736 to 482) as described (Laux *et al.*, 1988). DNAs of these bacteriophages were re-screened with the M13 clones B7.HET (B95-8 EBV nucleotides 166677 to 166465) and 669.EI (B95-8 EBV nucleotides 332 to 6) located in the first and second exons of the TP1 cDNA, respectively. Nine λgt10 *Eco*RI inserts hybridizing to 669.EI but not to B7.HET were subcloned into M13mp8 (Messing & Vieira, 1982) and about 200 bases at each end of each clone were sequenced by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). The 5′ ends of five of the inserts were located at EBV genome positions 169812, 169813, 169817, 169818 and 169831. All the most 5′ exons of these clones used a splice donor site at nucleotide 169906 and the acceptor site at nucleotide 58. The splice junctions conform to the published consensus sequences (Mount, 1982) and the acceptor site at nucleotide 58 is the same as the site used for exon 2 of the TP1 RNA. The 3′ ends showed polyadenylation at nucleotide 5856 (also used by TP1 mRNA) and at an alternative site at nucleotide 5836. It appeared therefore that these five cDNA clones were derived from an mRNA which is collinear at its 3′ end with exons 2 to 9 of TP1 mRNA but has a different 5′ exon (which we call 1′) and a transcription start apparently different from that of TP1 mRNA.

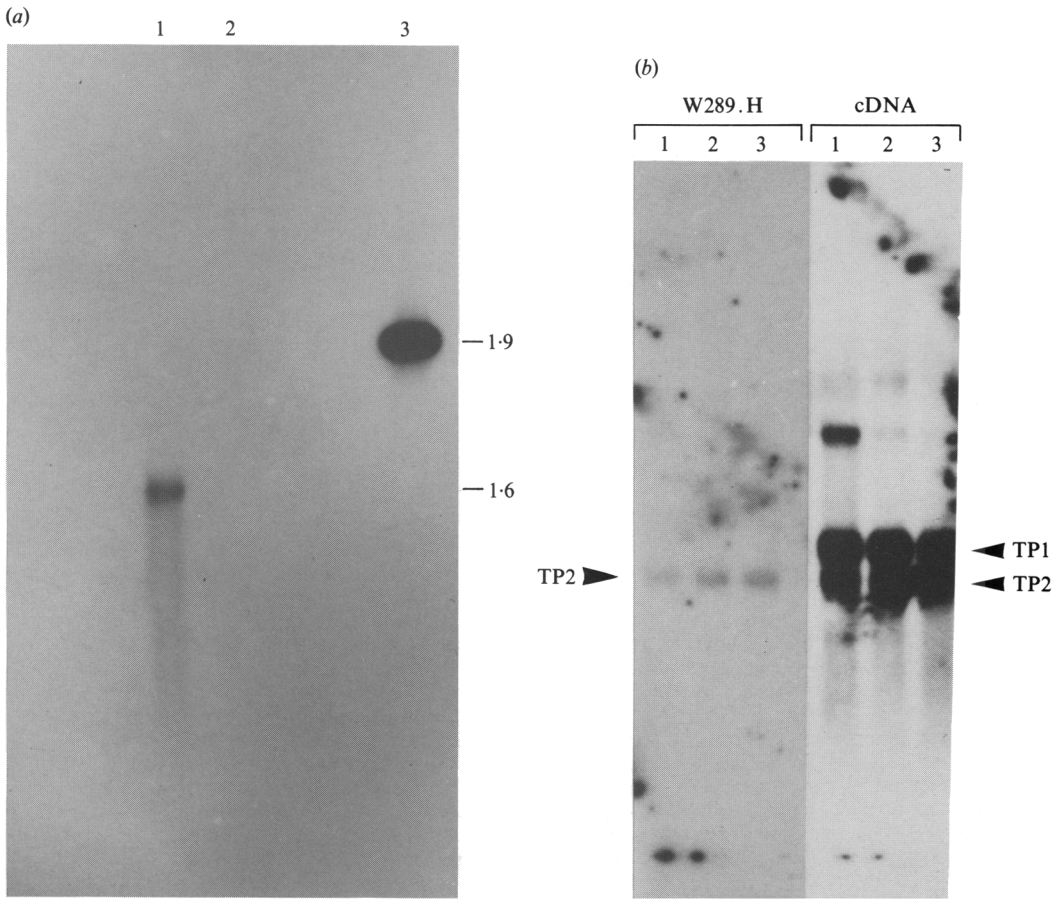

Fig. 2. (*a*) S1 nuclease analysis to determine the collinearity between TP1 and a TP2 cDNA clone. TP2 cDNA clone of sense (lane 1) or antisense (lane 2) orientation was hybridized to an antisense TP1 cDNA clone. The hybrids were treated with S1 nuclease, electrophoresed on an alkaline agarose gel, blotted to nitrocellulose and probed with radioactive TP1 cDNA. Lane 3 shows the TP1cDNA insert used for hybridization of the nitrocellulose filters. The sizes on the right are indicated in kb. (*b*) Autoradiogram of a Northern blot to cytoplasmic poly(A)$^+$ RNA (3 μg per lane) from TPA-induced (lanes 1). TPA + PAA-treated (lanes 2) and untreated (lanes 3) B95-8 cells. The blot was first probed with ^{32}P-labelled DNA of the M13 clone W289.H covering the entire exon 1′ of TP2 cDNA. The probe was removed and the filter rehybridized with ^{32}P-labelled DNA of the TP1 cDNA clone.

S1 nuclease analysis (Berk & Sharp, 1978) and restriction enzyme digestions were performed in order to verify the collinearity of these cDNAs with the TP1 cDNA through exons 2 to 9. Single-stranded DNA from M13 subclones containing inserts in either sense or antisense orientation was hybridized to an M13 clone of TP1 cDNA in antisense orientation and subsequently subjected to S1 nuclease digestion. The products were electrophoresed through an alkaline agarose gel (Maniatis *et al.*, 1982). The gel was blotted to nitrocellulose and the filter probed with a ^{32}P-labelled nick-translated cDNA clone of TP1. The probe detected a 1·6 kb band using the sense orientation of TP2 cDNAs (Fig. 2*a*, lane 1). No band was seen with the antisense orientation (Fig. 2*a*, lane 2). This demonstrates that the cDNAs derived from the 2·0 (TP1) and 1·7 kb (TP2) mRNAs (Hudson *et al.*, 1985) are collinear over 1·6 kb and the splicing pattern in exons 2 to 9 is identical for the two transcripts, since the sequences of the cDNA clones differed only at their 5′ but not at their 3′ ends. Three independent TP2 cDNAs were tested in this manner and they all gave the same result as in Fig. 2(*a*) (not shown). Restriction

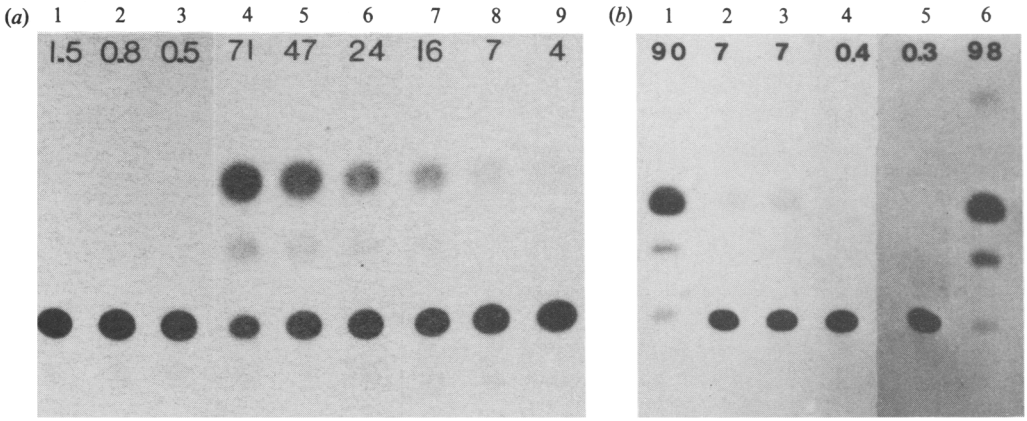

Fig. 3. Transcriptional activity of the bidirectional latent promoter region. The LMP promoter (pTP2RCAT), the TP2 promoter region (pTP2CAT) or no promoter (pUCCAT) were linked to the CAT gene and transfected into BL2/B95-8 (*a*) or BL41/C16 (*b*) cells. (*a*) Lanes 1 to 3, no promoter; 4 to 6, LMP promoter; 7 to 9, TP2 promoter; (*b*) lane 1, LMP promoter; lanes 2 and 3, TP2 promoter; lane 4, no promoter. Aliquots of the cell extracts were assayed at various dilutions for CAT activity and the percentage acetylation of [^{14}C]chloramphenicol is shown below the numbers of the lanes. Twenty μl of extract was used in (*a*) lanes 1, 4 and 7, and in (*b*) lanes 1 to 6; 10 μl was used in (*a*) lanes 2, 5 and 8 and 5 μl in (*a*) lanes 3, 6 and 9. Further controls were CAT assays without extract (*b* lane 5) and with purified CAT (*b* lane 6). To prepare the constructs for transfection, the CAT gene was isolated as a *Hin*dIII/*Bam*HI fragment from pSV2 CAT and cloned into pAT153 cut with *Hin*dIII and *Bam*HI giving plasmid p153CAT. The TP2 promoter region was isolated as an *Mst*II/*Bss*HII fragment (nucleotides 169477 to 169848) and cloned into *Hin*dIII-cut p153CAT after conversion of the *Mst*II and *Bss*HII ends to *Hin*dIII sites with linkers. The resulting plasmids were designated p2CAT and p2RCAT according to the orientation of their inserts with respect to TP2. The promoter/CAT parts of p153CAT, p2CAT and p2RCAT were then recloned as *Cla*I/*Bam*HI fragments into pUC18 cut with *Acc*I and *Bam*HI, giving pUCCAT, pTP2CAT and pTP2RCAT respectively. The electroporation used the BioRad Gene Pulser set to 200 V and 960 μF with 20 μg of plasmid DNA and 5×10^6 cells in 0·3 ml RPMI medium plus 10% foetal calf serum (Gibco) resulting in a decay constant between 50 and 60 ms (Cann *et al.*, 1988). After 24 h, cell extracts were prepared and assayed for acetylation of [^{14}C]chloramphenicol essentially as described by Gorman *et al.* (1982).

enzyme analysis (not shown) performed with *Ava*II, *Bst*XI, *Hin*cII, *Nhe*I, *Pvu*II and *Sty*I on cDNA clones of TP1 and TP2 revealed specific EBV fragments of the same length in the cDNAs. This confirmed (with the S1 analysis) the close relationship of the two TP RNAs.

To establish that the 1·7 kb mRNA was equivalent to the TP2 cDNAs, a Northern blot with RNA from uninduced, 12-*O*-tetradecanoylphorbol 13-acetate (TPA)-induced and TPA + phosphonoacetic acid (PAA)-induced B95-8 cells was hybridized with a single-stranded probe (Farrell *et al.*, 1983) from the M13 clone W289.H/*Ava*I (nucleotides about 169975 to 169719), which covers the exon 1′ of TP2 cDNA (Fig. 2*b*) and would hybridize to rightward transcripts. The W289.H probe hybridized to only one poly(A)$^+$ RNA of about 1·7 kb length. After stripping of the probe off the filter and re-exposure to check the absence of any signals, the filter was rehybridized with ^{32}P-labelled DNA of the TP1 cDNA clone. The cDNA clone detected both TP mRNAs of 2·0 and 1·7 kb and additionally a 4·5 kb transcript in the TPA-treated B95-8 cells (Fig. 2*b*). This mRNA is from a late viral gene in the short unique region (Hudson *et al.*, 1985) which partly overlaps the 3′ ends of the TP1 and TP2 RNAs. The Northern analysis confirmed that the cDNA was derived from the 1·7 kb RNA and that the first exon of TP2, exon 1′, is not present in the longer 2·0 kb transcript. The 1·7 and 2·0 kb transcripts did not change in abundance after TPA or TPA + PAA treatment, consistent with their classification as latent cycle genes.

Since the TP2 cDNAs were partial, primer extension and S1 analysis were performed to try to identify the 5′ end of the RNA. No clear results were obtained with either method (not shown) but comparison of the length of the cDNAs and the mRNA indicated that the cDNAs were

almost full-length. The 5′ end of the part of exon 1′ in the longest TP2 cDNA was 103 nucleotides downstream of the third of three closely spaced TAGAAA sequences possibly representing Goldberg–Hogness boxes for the initiation of eukaryotic transcription (Corden *et al.*, 1980). The negative results in the 5′ end mapping are probably explained by the low abundance of the TP2 mRNA and the possibility of several closely spaced cap sites caused by the close spacing of the three TATA-like elements (TAGAAA).

In order to show that this region can serve as a promoter, plasmids were constructed containing the putative promoter region in front of the chloramphenicol acetyltransferase (CAT) reporter gene. The plasmids were transfected by electroporation into two BL cell lines, BL2/B95-8 and BL41/C16, converted by EBV strains B95-8 and P3HR-1 clone 16 respectively (Rooney *et al.*, 1988). The transfected cells were extracted and assayed for CAT activity and the results are shown in Fig. 3; (*a*) represents the BL2/B95-8 cells and (*b*) the BL41/C16 cells. In both cell lines, the reverse orientation of the TP2 promoter (plasmid pTP2RCAT) showed very strong activity (Fig. 3*a*, lanes 4 to 6; *b*, lane 1). This was expected, since the upstream region of TP2 contains in reverse orientation the upstream regulatory region of LMP, the most abundant of the latent cycle mRNAs. The TP2 promoter (plasmid pTP2CAT) showed a substantially lower promoter activity than the LMP promoter (Fig. 3*a*, lanes 7 to 9; *b*, lanes 2 and 3; the percentage conversion is shown in each lane) whereas the control plasmid (pUCCAT) without any promoter did not show activity above background levels (Fig. 3*a*, lanes 1 to 3; *b*, lanes 4 and 5). Transfection of the plasmids into Raji cells gave similar results but the efficiency of the transfection was lower (data not shown). The transfection experiments demonstrated the presence of a bidirectional latent promoter region at the right-hand end of the EBV genome expressing the LMP gene leftward and the TP2 gene rightward. Compared to the LMP promoter, the TP2 promoter seemed generally to function about two- to threefold better in BL2/B95-8 cells than in BL41/C16 cells. If a single constitutively active enhancer region were driving the promoters for both LMP and TP2, the activity of the promoters might be coordinately regulated when, for example, LMP expression is restricted in BL cells *in vivo*.

Our results and those of Sample *et al.* (1989), together with the fact that there are no consensus splice sites between the TATA-like elements upstream of TP2 and the start of the cDNA, make it likely that the cap site of the TP2 mRNA is around nucleotide 169740. The first AUG in exon 1′ at position 169764 is not in a favourable context (Kozak, 1986) and is followed by a stop codon after 19 amino acids. A following AUG codon is located in the first common exon (exon 2) and in frame with the open reading frame of the TP1 mRNA giving rise to a predicted protein of about 40K. This protein would be 119 amino acids shorter than TP1 at its amino terminus. It would lack the hydrophilic part encoded by exon 1 of TP1 and would consist only of a very hydrophobic part possibly structured into 12 membrane-spanning domains. We are now attempting to identify the TP2 protein in EBV-infected cells and determine its function.

We thank Michel Perricaudet and Myriam Bodescot for access to the B95-8 cDNA library and David Rowe for helpful comments.

REFERENCES

BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SÉGUIN, C., TUFFNELL, P. S. & BARRELL, B. G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature, London* **310**, 207–211.

BERK, A. J. & SHARP, P. A. (1978). Spliced early RNAs of SV40. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 1274–1278.

BODESCOT, M., BRISON, O. & PERRICAUDET, M. (1986). An Epstein-Barr virus transcription unit is at least 84 kilobases long. *Nucleic Acids Research* **14**, 2611–2620.

CANN, A. J., KOYANAGI, Y. & CHEN, I. S. Y. (1988). High efficiency transfection of primary human lymphocytes and studies of gene expression. *Oncogene* **3**, 123–128.

CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C. & CHAMBON, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* **209**, 1406–1414.

EPSTEIN, M. A. & ACHONG, B. G. (1986). *The Epstein-Barr Virus: Recent Advances.* London: Heinemann.

FARRELL, P. J. (1989). The Epstein-Barr virus genome. In *Advances in Viral Oncology*, vol. 8, pp. 103–132. Edited by G. Klein. New York: Raven Press.

FARRELL, P. J., BANKIER, A., SÉGUIN, C., DEININGER, P. & BARRELL, B. G. (1983). Latent and lytic cycle promoters of the Epstein-Barr virus. *EMBO Journal* **2**, 1331–1338.

GORMAN, C. M., MOFFAT, L. F. & HOWARD, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Molecular and Cellular Biology* **2**, 1044–1051.

HENLE, W., DIEHL, V., KOHN, G., ZUR HAUSEN, H. & HENLE, G. (1967). Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. *Science* **157**, 1064–1065.

HUDSON, G. S., BANKIER, A. T., SATCHWELL, S. C. & BARRELL, B. G. (1985).The short unique region of the B95-8 Epstein-Barr virus genome. *Virology* **147**, 81–98.

HURLEY, E. A. & THORLEY-LAWSON, D. A. (1988). B cell activation and the establishment of Epstein-Barr virus latency. *Journal of Experimental Medicine* **168**, 2059–2077.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

LAUX, G., PERRICAUDET, M. & FARRELL, P. J. (1988). A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome. *EMBO Journal* **7**, 769–775.

LINDAHL, T., ADAMS, A., BJURSELL, G., BORNKAMM, G., KASCHKA-DIERICH, C. & JEHN, U. (1976). Covalently closed circular duplex DNA of Epstein-Barr virus in a human lymphoid cell line. *Journal of Molecular Biology* **102**, 511–530.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

MATSUO, T., HELLER, M., PETTI, L., O'SHIRO, E. & KIEFF, E. (1984). Persistence of the entire Epstein-Barr virus genome integrated into human lymphocyte DNA. *Science* **226**, 1322–1325.

MESSING, J. & VIEIRA, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269–276.

MOUNT, S. (1982). A catalogue of splice junction sequences. *Nucleic Acids Research* **10**, 459–472.

ROONEY, C., TAYLOR, N., COUNTRYMAN, J., JENSON, H., KOLMAN, J. & MILLER, G. (1988). Genome rearrangements activate the Epstein-Barr virus gene whose product disrupts latency. *Proceedings of the National Academy of Sciences, U.S.A.* **85**, 9801–9806.

SAMPLE, J., LIEBOWITZ, D. & KIEFF, E. (1989). Two related Epstein-Barr virus membrane proteins are encoded by separate genes. *Journal of Virology* **63**, 933–937.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SUGDEN, B., PHELPS, M. & DOMORADZKI, J. (1979). Epstein-Barr virus DNA is amplified in transformed lymphocytes. *Journal of Virology* **31**, 590–595.

(*Received 25 February 1989*)