The EMBO Journal vol.9 no.7 pp.2273–2277, 1990

Alternative splicing dictates translational start in Epstein–Barr virus transcripts

Robert P.Rogers, Maximilian Woisetschlaeger and Samuel H.Speck

Division of Tumor Virology, Dana–Farber Cancer Institute and Department of Pathology, Harvard Medical School, 44 Binney Street, Boston, MA 02115, USA

Communicated by A.Rich

Of the 10 viral genes known to be expressed during Epstein–Barr virus (EBV) latency, six encode nuclear antigens (EBNAs), all of which are expressed from a long primary transcript by means of alternative splicing and alternative polyadenylation sites. The primary transcript is generated by either of two promoters which operate in a mutually exclusive fashion in different clonal cell lines. All mRNAs from either promoter have exons in common from the *Bam*HI W viral genomic fragment (the major internal repeat, IR1) which encode the N-terminal portion of one of the nuclear antigens (EBNA 4). In addition to the coding regions for EBNA 4, EBNA mRNAs encode another EBNA (i.e. EBNA 1, 2, 3A, 3B or 3C) downstream. We show that alternative splicing determines whether the translation initiation codon for EBNA 4 is present or absent, thus permitting the generation of mRNAs in which the first translation initiation codon is either that for the EBNA 4 gene or for the other EBNA gene encoded downstream. This mechanism presumably ensures efficient translation of all the EBNA genes.

Key words: alternative splicing/Epstein–Barr virus/cDNA cloning translation initiation codon/S1 nuclease protection

Introduction

Infection of human B lymphocytes with Epstein–Barr virus (EBV) predominantly results in a latent infection with little production of virus, and a concomitant growth tranformation of the infected cells (immortalization). There are at least eight viral antigens expressed during latent infection, and the genes encoding six of these (the EBNAs) are transcribed from a complex transcriptional unit which spans the left-hand ~100 kb of the viral genome (for review, see Speck and Strominger, 1989). This transcriptional unit is driven by two viral promoters, Cp (Bodescot *et al.*, 1987) and Wp (Sample *et al.*, 1986; Speck *et al.*, 1986), which we have shown to operate in a mutually exclusive fashion in all clonal cell lines examined (Figure 1A) (Woisetschlaeger *et al.*, 1989).

Transcripts initiating from Cp contain two exons, C1 and C2 (Bodescot *et al.*, 1986), at their 5′ ends, while transcripts initiating from Wp have a short exon, W0, at their 5′ ends (Sample *et al.*, 1986; Speck *et al.*, 1986). C2 and W0 are spliced to a variable number of two repeated exons (W1 and W2) encoded within the viral *Bam*HI W fragments, a series of 3.1 kb direct repeats which vary in number among viral strains. cDNAs characterized to date have shown that C2 splices to the W1 exon (Bodescot and Perricaudet, 1986; Bodescot *et al.*, 1986; Sawada *et al.*, 1989), while W0 splices to the W1′ exon, a slightly shorter form of the W1 exon generated by the utilization of a splice acceptor site located 5 bp downstream of the acceptor for W1 (Sample *et al.*, 1986; Speck *et al.*, 1986). Both the W1 and W1′ exons utilize the same splice donor site to splice to the W2 exon. The W2 exons splice either to another W1 exon, or in the case of the ultimate W2 exon, to two or three exons (Y1, Y2 and Y3) encoded within the adjacent *Bam*HI Y fragment. The Y2 or Y3 exon is in turn spliced to (or contiguous with) unique 3′ coding sequences. In some transcripts, splicing from the W1 exon bypasses the *Bam*HI Y exons (Bodescot and Perricaudet, 1986).

Interestingly, the *Bam*HI W exons encode the N-terminal portion of a variably sized, highly repetitive nuclear antigen, EBNA 4 (Bodescot *et al.*, 1984; Dillner *et al.*, 1986; Sample *et al.*, 1986; Speck *et al.*, 1986). The C-terminal portion of EBNA 4 is encoded in the Y1 and Y2 exons. In all EBNA mRNAs characterized to date, some or all of the EBNA 4 coding exons are situated 5′ to the coding sequences for another EBNA on the same message. However, the cDNA evidence indicates that the only translation initiation codon present in the EBNA 4 gene is generated by a splice junction associated with Wp usage (the W0/W1′ splice). The AUG-creating C2/W1′ splice has not been observed, and the C2/W1 splice, generated from either of two splice donor sites (Sawada *et al.*, 1989), does not generate a translation initiation. This raises two interesting questions about EBV gene expression. First, can there be efficient translation of EBNAs encoded downstream of EBNA 4 on messages initiated from Wp? Second, can EBNA 4 be expressed from Cp?

In this paper we show that a mechanism exists for generating transcripts from Wp that lack a translation initiation codon for the EBNA 4 gene, thereby allowing the synthesis of monocistronic mRNAs for expression of the downstream EBNA gene. With regard to the second question, it is known that cell lines exhibiting exclusive utilization of Cp do express EBNA 4 (Finke *et al.*, 1987; Woisetschlaeger *et al.*, 1989). We show that the C2 exon associated with Cp usage can splice to the W1′ exon to generate a translation initiation codon for the EBNA 4 gene.

Results

Characterization of a transcript initiating from Wp which does not contain a translation initiation codon for EBNA 4

We have recovered a cDNA clone (IB4-W2.16) initiating from Wp which does not contain the translation initiation codon for EBNA 4 (Figure 2). Unlike previously characterized cDNAs (Sample *et al.*, 1986; Speck *et al.*,

1986), the W0 exon of IB4-W2.16 splices to the W1 exon instead of the W1′ exon. In Figure 3, the IB4-WY1 cDNA clone (Speck *et al.*, 1986) illustrates the previously known W0/W1′ splice junction which joins A and U residues from W0 with a G residue from W1′ to generate the EBNA 4 translation initiation codon. The W0/W1 splice of IB4-W2.16 joins A and U residues from W0 with a C residue from W1, failing to generate a translation initiation codon.

The IB4-W2.16 cDNA contains three pairs of the W1, W2 repeat exons, but is incomplete at its 3′ end. Interestingly, the last W2 exon fails to splice at its normal splice donor site, and extends into the downstream IR1 sequences which encode a short open reading frame (ORF). The initiation codon of this ORF lies 51 nucleotides downstream of the usual W2 splice donor site.

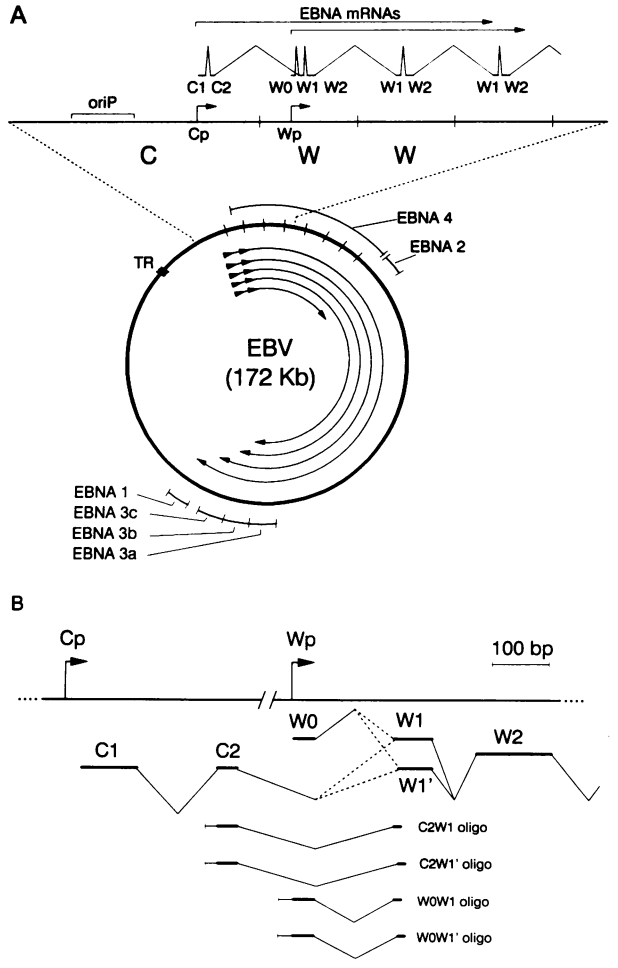

Fig. 1. (**A**) Transcriptional organization of the EBNA genes.The concentric arcs illustrated within the viral episome represent the primary transcripts which give rise to the EBNA mRNAs. The two arrowheads at the 5′ ends represent transcription initiation at either Cp or Wp. The exploded portion of the diagram illustrates the organization of the exons present at the 5′ end of all EBNA mRNAs. The proximity to Cp and Wp of the latent origin of replication (*ori*P), which also functions as an EBNA 1-dependent enhancer (Sugden and Warren, 1989), is shown. Only the left-hand most Wp is illustrated. (**B**) Schematic illustration of alternative splicing from the C2 and W0 exons to the W1 and W1′ exons. The structures of the antisense splice junction oligonucleotides are depicted, the heavy lines indicating regions complementary to exon sequences. The size difference between the W1 and W1′ exons is exaggerated for clarity (see text and Figure 3).

Both the W1 and W1′ splice acceptor sites are utilized in splicing from the W0 exon

To determine whether both the W0/W1′ and W0/W1 splice junctions are present in all known Wp-utilizing cell lines, oligonucleotides were generated to assay for the presence of these splice junctions by S1 nuclease protection analysis (Figures 1B and 4A). The 'anti-sense' oligonucleotides employed were complementary to 12 nucleotides of the W1 or W1′ exon, all of the W0 exon, and ~20 nucleotides upstream of the transcription cap site (Figure 1B). They were labeled and hybridized to polyadenylated total cellular RNA from a number of cell lines, followed by digestion with S1 nuclease (Figure 4A). Signals generated from protected oligonucleotide fragments of appropriate size demonstrate exclusive utilization of either Cp or Wp in all cell lines examined. As previously observed (Woisetschlaeger *et al.*, 1989), the Daudi, Clone-13, IB4 and X50-7 cell lines utilize Wp. RNA from these cell lines protected fragments of appropriate size for both the W0/W1′ and W0/W1 splice junction oligonucleotides. This confirms the cDNA data and demonstrates that both splice junctions occur in Wp-utilizing cell lines. Total protection of the W0/W1 and W0/W1′ probes appears to be due to partially spliced heteronuclear RNA, since this signal was greatly diminished when cytoplasmic polyadenylated RNA was examined (data not shown).

The C2 exon splices to both W1 and W1′ splice acceptor sites

The results presented above raised the question of whether alternative splicing from the C2 exon might also occur. This is of particular interest since splicing from C2 to W1′, an event not yet observed, would generate a translation initiation codon for EBNA 4 in Cp-utilizing cell lines. We therefore generated an 'anti-sense' oligonucleotide to assay for the C2/W1′ splice junction. As shown in Figure 4A, Cp-utilizing cell lines (BJAB/B95-8, Jijoye, JY, JC5 and B95-8) protected the C2/W1′ oligonucleotide, demonstrating that this splice junction occurs in these cell lines. Furthermore, the lower level of protection of this oligonucleotide with RNA prepared from the BJAB/B95-8, Jijoye and B95-8 cell lines, compared to the level of protection observed with RNA from the JY and JC5 cell lines, is consistent with a greater level of EBNA 4 protein detected in the latter cell lines by an EBNA 4-specific antiserum on immunoblots (data not shown).

Splicing from both the C2 and W0 exons favors the W1 splice acceptor over the W1′ splice acceptor

The experiment presented in Figure 4B was designed to determine the relative abundance of the W0/W1 and W0/W1′ splice junctions in a Wp-utilizing cell line (X50-7), and the relative abundance of C2/W1 and C2/W1′ splice junctions in a Cp-utilizing cell line (JY). The splice junction oligonucleotide pairs were labeled to approximately the same specific activity and assayed with total cellular RNA. For the Wp-utilizing X50-7 cell line, the non-AUG-creating W0/W1 splice junction accounted for 78% of the sum of the signal intensities observed with these two partially protected oligonucleotides (W0/W1 and W0/W1′). For the Cp-utilizing JY cell line, the non-AUG-creating C2/W1 splice junction accounted for 59% of the sum of the signal intensities observed with these two partially protected oligonucleotides (C2/W1 and C2/W1′).

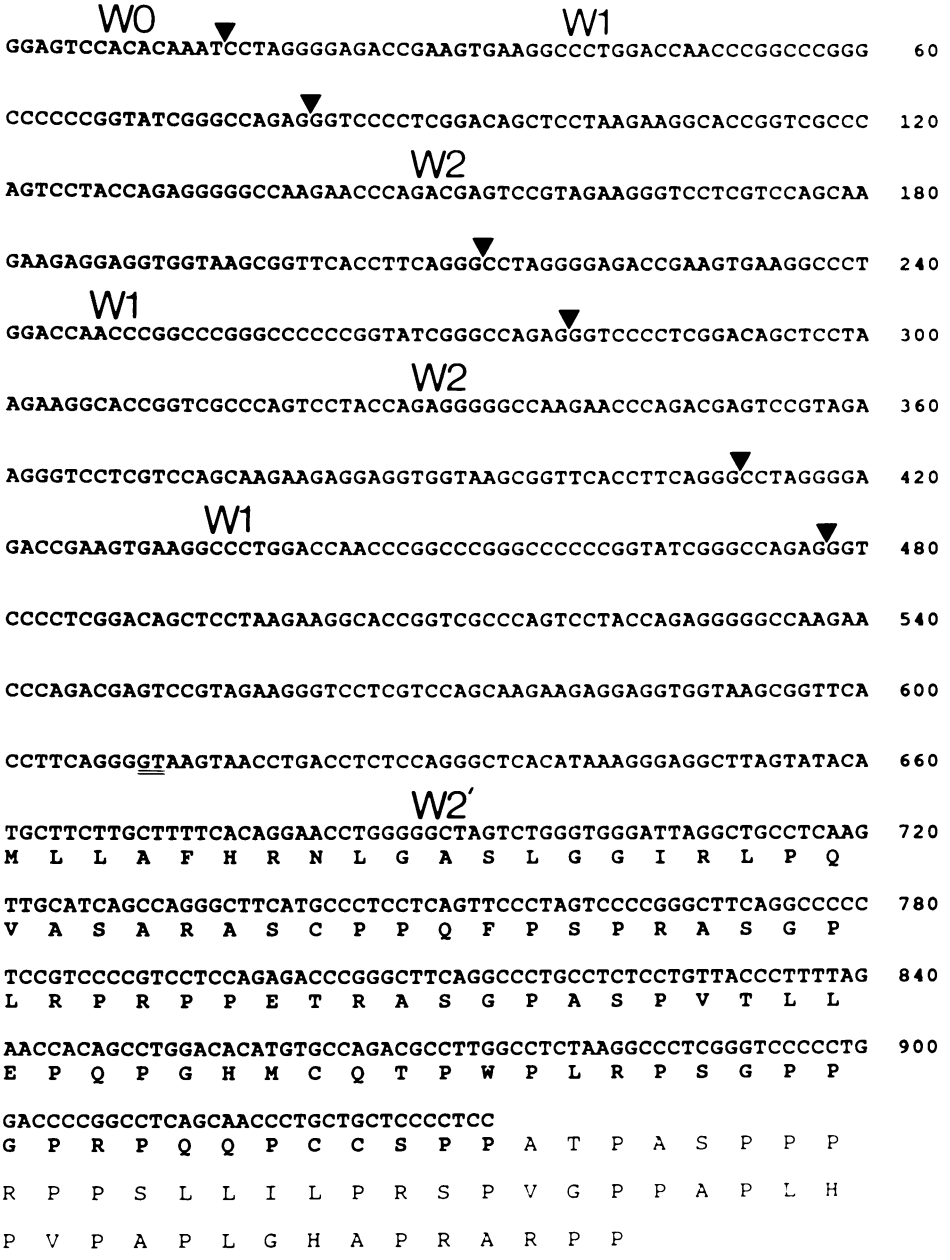

Fig. 2. Nucleotide sequence of the IB4-W2.16 cDNA clone and the predicted amino acid sequence of a short ORF. The exon boundaries are indicated by arrowheads.

Discussion

Some or all of the exons encoding EBNA 4 are present in the 5′ exons of all the EBNA mRNAs in latently infected B lymphocytes (Speck and Strominger, 1989). Two cDNA clones have previously been characterized which contain the complete EBNA 4 ORF (Sample *et al.*, 1986; Speck *et al.*, 1986). These clones, both of which were isolated from cDNA libraries prepared from a Wp-utilizing cell line, IB4, revealed that the EBNA 4 translation initiation codon is generated by the W0/W1′ exon splice junction. (The EBNA 4 ORF contains no internal methionines.) The W0/W1′ splice junction utilizes a splice acceptor which is 5 bp downstream of the splice acceptor employed in the W2/W1 splice junctions (Bodescot *et al.*, 1984).

In this paper we have characterized another cDNA clone from the Wp-utilizing cell line, IB4, which, in contrast to EBNA 4-expressing mRNAs, contains a W0/W1 exon splice junction, thereby failing to generate the EBNA 4 translation initiation codon. This cDNA clone has another interesting feature. While being highly spliced in its 5′ half, the last W2 exon is not spliced at the normal splice donor site, but retains the IR1 sequences immediately downstream. Thus, the 3′ sequences present in this clone have previously only been characterized as an intron. This sequence contains the 5′ portion of a short ORF. The predicted protein would be basic, 135 amino acids in length and proline rich (~30%). A search of the protein sequence data base revealed sequence similarities of proline-rich proteins such as collagen, as well as the herpes simplex type on US11 DNA-binding protein

```
Genomic DNA  CCAATTGTCAGTTCTAGGGAGGGGGACCACTGCCCCTGGTATAAAGT   14,360
IB4-WY1
IB4-W2.16                                                W0

Genomic DNA  GGTCCTGCAGCTATTTCTGGTCGCATCAGAGGCCCAGGAGTCCACAC   14,407
IB4-WY1                                  TCAGAGGCCCAGGAGTCCACAC
IB4-W2.16                                    GGAGTCCACAC

Genomic DNA  AAATGTAAGAGGGGGTCTTCTACCTCTCCCTAGCCCTCCGCCCCCTC   14,454
IB4-WY1      AAAT
IB4-W2.16    AAAT

Genomic DNA  CAAGGACTCGGGCCCAGTTTCTAACTTTTCCCCTTCCCTCCCTCGTCT   14,502
IB4-WY1
IB4-W2.16

Genomic DNA  TGCCCTGCGCCCGGGGCCACCTTCATCACCGTCGCTGACTCCGCCA   14,548
IB4-WY1
IB4-W2.16                                  W1
Genomic DNA  TCCAAGCCTAGGGGAGACCGAAGTGAAGGCCCTGGACC...
IB4-WY1                 GGGAGACCGAAGTGAAGGCCCTGGACC...
IB4-W2.16          CCTAGGGGAGACCGAAGTGAAGGCCCTGGACC...
```

Fig. 3. Comparison of the EBV genomic sequence with the 5′ sequences present in the IB4-WY1 and the IB4-W2.16 cDNA clones. The Wp CCAAT and TATA boxes are denoted with overbrackets. The splice junction donor and acceptor sequences are underlined. The closed box indicates the splice acceptor for the W1 exon. The open box indicates the splice acceptor for the W1′ exon.

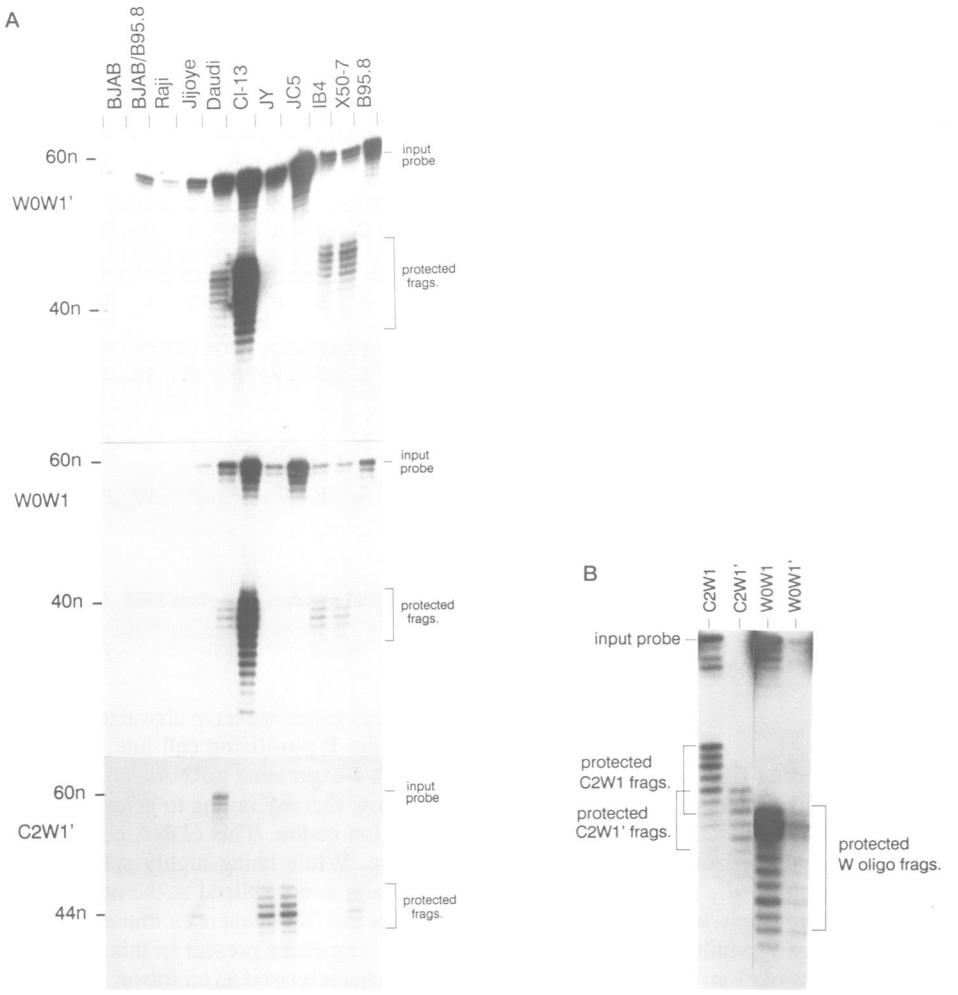

Fig. 4. (**A**) S1 nuclease protection employing splice junction-specific oligonucleotides. S1 analyses were carried out with synthetic oligonucleotides that were complimentary to a region spanning the W0/W1, W0/W1′ or C2/W1′ splice junctions. (**B**) Relative abundance of AUG- versus non-AUG-producing splice junctions. S1 nuclease assays were performed as described, except that 20 μg of total cellular RNA was employed and the S1 nuclease concentration was 500 U/ml. RNA prepared from the JY cell line was employed to assay for the relative abundance of C2W1 and C2W1′ splice junctions, and RNA prepared from the X50-7 cell lines was used to determined the relative abundance of W0W1 and W0W1′ splice junctions.

(McGeoch *et al.*, 1985) and a number of retroviral gag proteins.

The IB4-W2.16 cDNA clone was recovered from a library prepared from polyadenylated total cellular RNA, and transcripts bearing the IR1 sequence immediately downstream of W2 are largely confined to the nucleus (Rogers and Speck, 1990). Thus, it is probable that IB4-W2.16 was derived from incompletely spliced heteronuclear polyadenylated RNA. However, it cannot be ruled out that a minor fraction of similar transcripts is processed into mature mRNA. We are therefore investigating the possibility that the ORF encoded in the W2/W1 'intron' is expressed during latency.

The observation that the first splice junction of the IB4-W2.16 cDNA fails to generate the EBNA 4 translation initiation codon led us to seek additional evidence for alternative splicing from the W0 exon as well as the C2 exon. Although cDNA evidence prior to this work only provided examples of C2/W1 and W0/W1′ splice junctions, we demonstrated the occurrence of C2/W1′ and W0/W1 splice junctions. The former allows expression of EBNA 4 in Cp-utilizing cell lines, while the latter presumably allows the generation (in Wp-utilizing cell lines) of monocistronic mRNAs which can be efficiently translated to produce the other EBNAs.

Several recent examples illustrate the important role of alternative splicing for controlling the expression of cellular genes. In *Drosophila melanogaster* the sex-specific expression of the *transformer* and *Sex-lethal* gene products is determined in both cases by the usage of an alternative splice acceptor site which only occurs in females (Sosnowski *et al.*, 1989). Tissue-specific expression of the different isoforms of α-tropomyosin (Wieczorek *et al.*, 1988), and regulation of the H-*ras* proto-oncogene (Cohen *et al.*, 1989) are also mediated by alternative splicing.

The complex transcriptional pattern for expressing the EBNAs is probably involved in the tight regulation of these gene products. Whether the alternative splicing pattern described here is actively regulated by the abundance of the various EBNAs, or is preset to a fixed ratio, will require further investigation.

Materials and methods

Cell culture and RNA preparation

Cell lines were grown in RPMI 1640 medium supplemented with 10% (v/v) fetal bovine serum. Total cellular RNA was prepared (Auffray and Rougeon, 1980). Poly(A)$^+$ RNA was isolated by fractionation on oligo(dT)–cellulose (Aviv and Leder, 1972).

cDNA synthesis and DNA sequencing

A cDNA library prepared with poly(A)$^+$ RNA isolated from the latently infected lymphoblastoid cell line, IB4, and was cloned into the λgt10 bacteriophage vector (Speck *et al.*, 1986).The library was screened with ^{32}P-labeled viral *Bam*HI W fragment. The IB4-W2.16 cDNA clone was sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977), and the number of W1, W2 repeat exons was determined by partial digestion with *Sma*I.

S1 nuclease protection

Between 50 and 100 fmol of ^{32}P-labeled oligonucleotide was hybridized to 2 μg of poly(A)$^+$ RNA in 50 μl of hybridization buffer [40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 0.1% SDS, 50% (v/v) formamide] at 42°C for 12 h. The hybridization reactions were subsequently treated with S1 nuclease (Bethesda Research Laboratories) at a concentration of 100 U/ml in a final reaction volume of 0.35 ml (0.28 M NaCl, 50 mM NaOAc, pH 4.6, 4.5 mM $ZnSO_4$) for 30 min at 37°C. The protected fragments were resolved on a 10% denaturing polyacrylamide gel. Quantitation of the relative abundance of C2W1 versus C2W1′ and W0W1 versus W0W1′ splice junctions was determined using a Betascope 603 blot analyzer (Betagen).

The following single stranded oligonucleotides were synthesized to assay for specific splice junctions: C2/W1, 5′-TTCGGTCTCCCCTAGGATGC-TCACGTGCAGGAGGCTGTTTCTTCAGTCCTAGAGGGAAGG-3′; C2/W1′, 5′-TTCGGTCTCCCATGCTCACGTGCAGGAGGCTGTT-TCTTCAGTCCTAGAGGGAAGG-3′; W0/W1, 5′-GTCTCCCCTAG-GATTTGTGTGGACTCCTGGTCGCTCTGATGCGACCAGAAATAG-TCGCAG-3′; W0/W1′, CTTCGGTCTCCCATTTGTGTGGACTCC-TGGTCGCTCTGATGCGACCAGAAATAGTCGCAG-3′.

Acknowledgements

This research was supported by National Institutes of Health Grant 5 R01 CA 43143 and a Leukemia Society Special Fellowship to S.H.S., NIH Grant DE 00194-03 to R.P.R., postdoctoral fellowship by the Fonds zur Foerderung der Wissenschaftlichen Forschung, Vienna to M.W. and NIH Grant CA47554 to J.L.S.

References

Auffray,C. and Rougeon,T. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408–1412.
Bodescot,M. and Perricaudet,M. (1986) *Nucleic Acids Res.*, **14**, 7103–7114.
Bodescot,M., Chambraud,B., Farrell,P. and Perricaudet,M. (1984) *EMBO J.*, **3**, 1913–1917.
Bodescot,M., Brison,O. and Perricaudet,M. (1986) *Nucleic Acids Res.*, **14**, 2611–2620.
Bodescot,M. Perricaudet,M. and Farrell,P.J. (1987) *J. Virol.*, **61**, 3424–3430.
Cohen,J.B., Broz,S.D. and Levinson,A.D. (1989) *Cell*, **58**, 461–472.
Dillner,J., Kallin,B., Alexander,H., Ernberg,I., Uno,M., Ono,Y., Klein,G. and Lerner,R.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6641–6645.
Finke,J., Rowe,M., Kallin,B., Ernberg,I., Rosen,A., Dillner,J. and Klein,G. (1987) *J. Virol.*, **61**, 3870–3878.
McGeoch,D.J., Dolan,A., Donald,S. and Rixon,F.J. (1985) *J. Mol. Biol.*, **181**, 1–13.
Rogers,R.P. and Speck,S.H. (1990) *J. Virol.*, **64**, 2426–2429.
Sample,J., Hummel,M., Braun,D., Birkenbach,M. and Kieff,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5096–5100.
Sanger,F., Nicklen,S. and Coulson,A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sawada,K., Yamamoto,M., Tabata,T., Smith,M., Tanaka,A. and Nonoyama,M. (1989) *Virology*, **168**, 22–30.
Sosnowski,B.A., Belote,J.M. and McKeown,M. (1989) *Cell*, **58**, 449–459.
Speck,S.H. and Strominger,J.L. (1989) In Klein,G. (ed.), *Advances in Viral Oncology, Vol. 8, Transcription of Epstein–Barr Virus in Latently Infected Growth-Transformed Lymphocytes*. Raven Press, New York, pp. 133–150.
Speck,S.H., Pfitzner,A.J. and Strominger,J.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9298–9302.
Sugden,B. and Warren,N. (1989) *J. Virol.*, **63**, 2644–2649.
Wieczorek,D.F., Smith,C.W.J. and Nadal-Ginard,B. (1988) *Mol. Cell. Biol.*, **8**, 679–694.
Woisetschlaeger,M., Strominger,J.L. and Speck,S.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6498–6504.

Received on February 16, 1990