JOURNAL OF VIROLOGY, Apr. 1990, p. 1812–1816
0022-538X/90/041812-05$02.00/0

Vol. 64, No. 4

cDNA Cloning and Transient Expression of the Epstein-Barr Virus-Determined Nuclear Antigen EBNA3B in Human Cells and Identification of Novel Transcripts from Its Coding Region

BRIGITTE KERDILES,[1] DERMOT WALLS,[1] HINDA TRIKI,[2] MICHEL PERRICAUDET,[1]* AND IRÈNE JOAB[1]

Laboratoire de Génétique des Virus Oncogènes, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94800 Villejuif,[1] and Laboratoire de Virologie, Hôpital Trousseau, 75012 Paris,[2] France

Received 24 August 1989/Accepted 18 December 1989

Recombinant plasmids containing sequences from the *Bam*HI-E rightward reading frames 2a and 2b (BERF2a and 2b) of the Epstein-Barr virus (EBV) genome were isolated from a library of cDNA clones which had been previously made from the EBV B95-8 lymphoblastoid cell line (M. Bodescot, O. Brison, and M. Perricaudet, Nucleic Acids Res. 14:7103–7114, 1986). The characterization of these clones in combination with RNase mapping experiments led to the identification of one leftward and several rightward transcripts traversing the EBV-determined nuclear antigen EBNA3B coding region. One cDNA (T7) contains a continuous open reading frame generated by the splicing together of BERF2a and BERF2b. The T7 clone was used to reconstruct a complete fused BERF2a/2b open reading frame in an adenovirus-based expression vector. Western immunoblotting and immunofluorescence experiments using human 293 cells showed that the recombinant plasmid is capable of expressing a protein with a size, immunological characteristics, and a subcellular localization indistinguishable from those of native B95-8 EBNA3B.

The Epstein-Barr virus (EBV) has been identified as an etiological agent of infectious mononucleosis and is now clearly established to be associated with two human cancers, Burkitt's lymphoma and nasopharyngeal carcinoma. Its host range is limited to human and closely related primate B lymphocytes and certain epithelial cells. Resting B lymphocytes are immortalized to perpetual growth in vitro upon infection with EBV, thus generating lymphoblastoid cell lines. Most information on EBV gene expression which has been gathered to date has come from the study of lymphoblastoid cell lines or B-cell lines derived from Burkitt's lymphoma biopsies. In the majority of cases, a state of viral latency is established, with a restricted set of viral genes being expressed (2). The identification of these genes is a key step towards deciphering the mechanisms responsible for the establishment and maintenance of the immortalized state and for the control of virus gene expression itself.

In lymphoblastoid cell lines, a family of at least six EBV-determined nuclear antigens (EBNAs) is expressed. Three of these proteins, EBNA3A, -3B, and -3C (EBNA3, -4, and -6, respectively, by alternative nomenclature) are encoded by genes which have been mapped to the *Bam*HI-E region of the EBV genome. The EBNA3A and -3C transcripts have been shown to be rightward in direction and structurally similar, with each containing short and long open reading frames joined together by splicing to generate long open reading frames encoding each protein (BLRF3 plus BERF1 [3, 4, 7, 9] and BERF3 plus BERF4 [1, 4, 12, 15], respectively). A description of this region of the EBV genome is presented in Fig. 1A and B. The coding sequence for EBNA3B has only been partially mapped to BERF2b (11, 13, 16). In this report, we describe the structure of the EBNA3B transcript in the region of its coding exons and demonstrate the transient expression of the complete EBNA3B protein in a human cell line by using an adenovirus expression vector.

A cDNA bank made from the cytoplasmic polyadenylated mRNAs of the EBV B95-8 cell line (4) was screened by in situ hybridization for clones from the EBNA3B region. The DNA probe used was a 727-base-pair (bp) *Nco*I subfragment of BERF2b in Fig. 1B, hatched box. cDNAs thus isolated were characterized by restriction endonuclease mapping and partial DNA sequencing after subcloning into pUC13 or M13 vectors.

The structures of five cDNA clones (T7 to T11) are presented in Fig. 1C. The T7 clone, which is 1.9 kbp long, begins in BERF2a and ends in BERF2b (EBV genome positions 95524 to 97524) (2). It contains a continuous open reading frame generated by the splicing together of the short and long open reading frames. The splice involves the excision of a 78-base intron (95710 to 95787) from a rightward transcript and has not been previously demonstrated. Since the flanking EBNA3A and EBNA3C are each encoded by a short and long exon pair (processed from BLRF3 and BERF1 and from BERF3 and BERF4, respectively), it is likely by analogy and by the splicing motif used that the T7 clone represents a part of the EBNA3B transcript. None of the other clones were found to contain this splice, making it impossible to further describe this particular transcript by this means. The 4.4-kbp T8 clone (which begins 3′ to the splice acceptor site identified in T7) represents part of a novel transcript extending into the EBNA3C coding sequence. This clone contains the corresponding EBNA3C splicing motif between BERF3 and BERF4 and as such is derived from a transcript potentially capable of expressing the EBNA3C protein. In EBV-infected cells, transcripts containing the same splice as that in the T8 clone have previously been described by different groups (4, 12, 15), and one of these has been shown to encode full-size EBNA3C (12). These latter transcripts, however, do not contain BERF2b sequences, as they are spliced from upstream noncoding sequences to an acceptor site just 5′ to BERF3 and thus bear no relationship to the T8 clone. The T8 cDNA, if derived from the same transcript as T7, would suggest that

* Corresponding author.

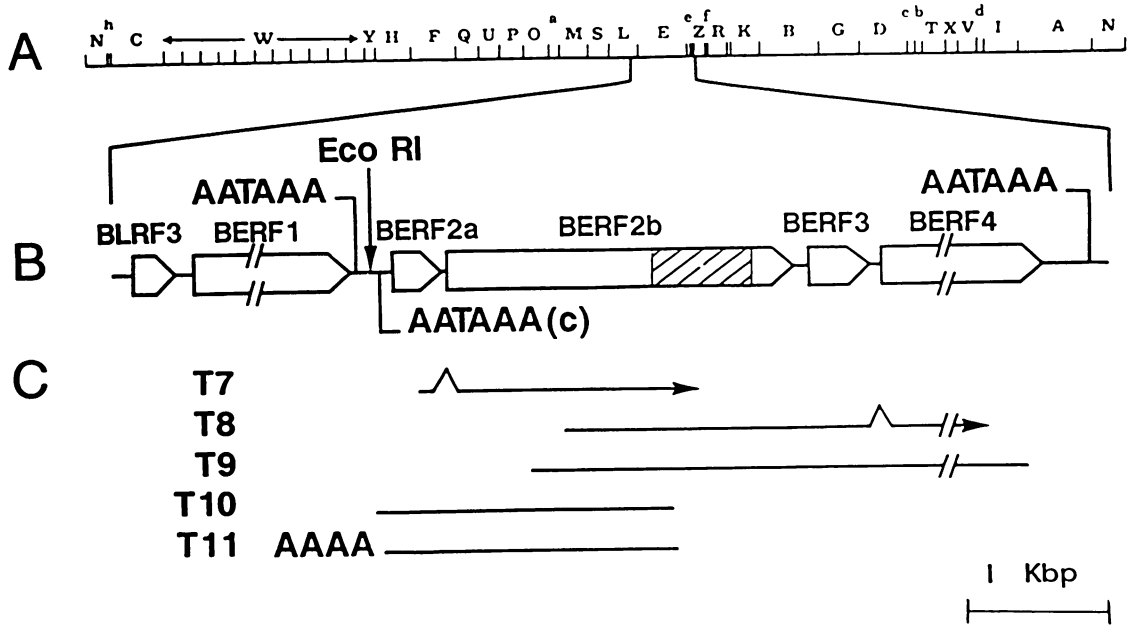

FIG. 1. (A) EBV B95-8 *Bam*HI restriction map. (B) Enlarged *Bam*HI-L to -E region showing open reading frames encoding EBNA3A, -3B, and -3C (described in text). The position of the probe region in BERF2b used to isolate the cDNAs is indicated (▨). Polyadenylation signals (AATAAA) are indicated. C, Complementary strand. (C) Structures and locations of cDNAs T7 to T11; arrows indicate the direction of transcription where known.

the EBNA3B and -3C mRNAs share the same 3′ end. Whether or not EBNA3B and -3C can be translated from a bicistronic message is impossible to prove, however, in the absence of the corresponding full-length cDNA.

The remaining three clones indicated in Fig. 1C are not spliced. The T11 clone represents part of a novel mRNA transcribed from the strand complementary to that encoding EBNA3B and is polyadenylated 12 bp after an AATAAA sequence (position 95272) located just 5′ to BERF2a. The discovery of this leftward transcript implies that any of the unspliced cDNAs (except T10; see below) may also represent sequences from it. No major open reading frame could be identified within its sequence, and as it does not begin near any consensus polymerase II promoter sequences, it is likely to be incomplete. It may be the case that this left-to-right mRNA simply has a long untranslated 3′ end and/or serves to down regulate the expression of EBNA3B by in vivo hybridization to its transcript. We have not yet determined whether this transcript is produced during latent, early, or late viral transcription. Nonetheless, it is remarkable that a virus with such a large genome should generate different mRNAs with long complementary tracts. The unspliced T10 cDNA which begins at the endogenous *Eco*RI site 5′ to BERF2a (and the polyadenylation signal used in the T11 clone) may also represent another species of transcript from this region. Because of the fact that the T9 and T10 cDNA clones are neither spliced nor polyadenylated, it cannot be ruled out that they were generated by the reverse transcription of nuclear RNAs during cDNA synthesis and therefore may not represent parts of processed cytoplasmic mRNAs.

RNase mapping was used to further investigate the nature of the rightward transcripts containing the BERF2a/2b junction (Fig. 2). An *Eco*RI-*Sph*I subfragment of *Bam*HI-E was cloned into the plasmid pGEM 3Zf(+), thus allowing the generation of a leftward RNA probe from its Sp6 promoter. Total cellular RNA was prepared by the guanidium thiocyanate method (5). Labeled transcript was synthesized by using Sp6 polymerase and [α-^{32}P]UTP as recommended by Promega Biotec. Forty micrograms of total RNA was hybridized with 10^6 cpm of probe and then digested with RNase A and T1 as described previously (10). Samples were run on a 4.5% denaturing polyacrylamide gel which was then dried and autoradiographed at −70°C with an intensifying screen. This experiment established several facts. (i) The splice acceptor site identified in the T7 cDNA clone is used in the EBV-positive cell lines B95-8, IB4, BJAB/B95-8, and Namalwa (Fig. 2A, band C). (ii) A splice acceptor site was mapped in three of these lines to a position about 40 bp 5′ to BERF2a (Fig. 2A, band B). This corresponds well with an excellent consensus splice acceptor site present on the DNA sequence (position 95788) and when used would thus give splicing motifs similar to those of EBNA3A and -3C. (iii) The existence of a novel transcript containing the sequence from *Eco*RI to the splice donor site at the end of BERF2a can be implied from band A.

The identification by RNase mapping of a functional splice acceptor site just 5′ to BERF2a implied that the ATG located at the left extremity of BERF2a (position 95353) is the first of the open reading frames created by the fusion of BERF2a and BERF2b. Two in-phase stop codons between the splice acceptor site and this ATG eliminated the possibility that upstream candidate initiation codons exist. Three downstream ATGs located close to the BERF2a/2b junction (positions 95704, 95857, and 95865) could theoretically also be used to translate the BERF2b sequence alone. We have not succeeded in identifying the polyadenylation site used to terminate EBNA3B transcripts. However, on the basis of some of the cDNAs obtained and the fact that the closest poly(A) addition site is situated at the end of EBNA3C (Fig. 1B), it is likely that transcripts for both of these proteins share the use of this signal. The data lead one to conclude that the EBNA3B coding sequence is assembled from exons structurally similar to those of EBNA3A and -3C.

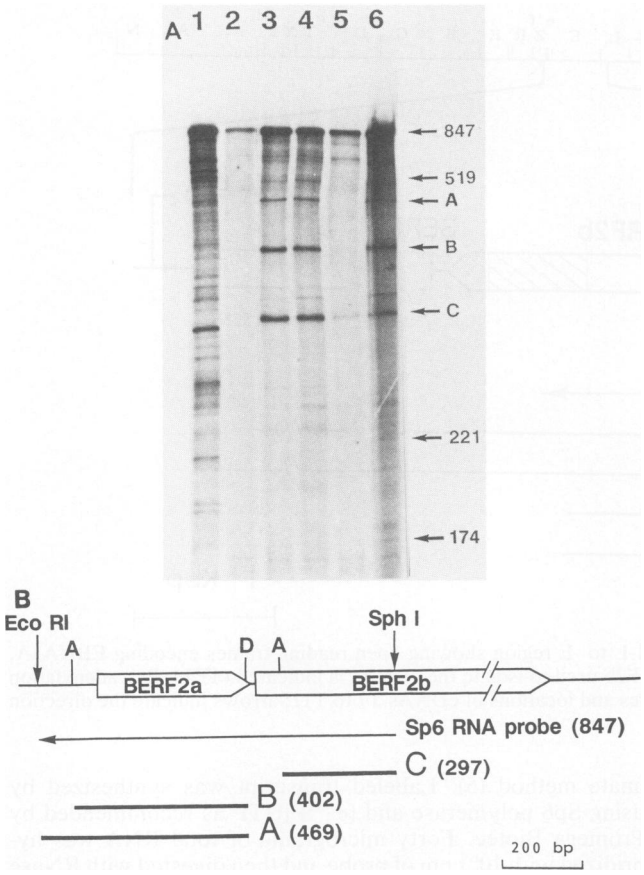

FIG. 2. RNase protection analysis of rightward transcripts containing the BERF2a/2b junction. (A) Lanes: 1, probe only; 2, Jurkat EBV-negative T-cell line; 3 to 6, EBV-positive cell lines JC5, IB4, Namalwa, and B95-8, respectively. Size markers (in base pairs) are indicated on the right. Bands A, B, and C are discussed in the text. (B) Map of the EBNA3B coding region showing splice donor (D) and acceptor (A) sites, the location of the RNA probe used, and the identities of bands A, B, and C indicated on the autoradiograph in panel A.

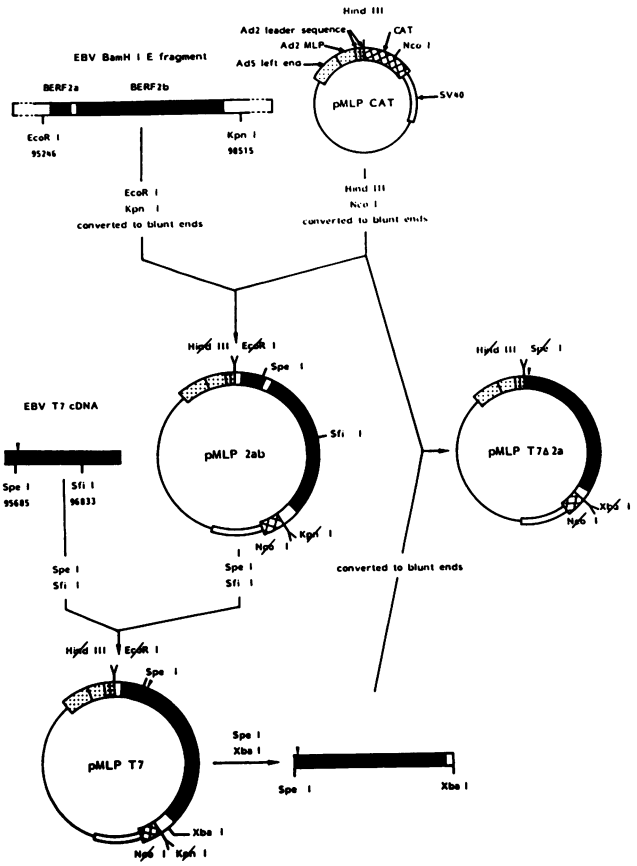

FIG. 3. Construction of pMLP T7 and pMLP T7Δ2a plasmids. The *Eco*RI-*Kpn*I subfragment of EBV *Bam*HI-E was first substituted for the *Hin*dIII-*Nco*I fragment of the pMLP CAT plasmid (8), thus generating pMLP 2ab. To construct pMLP T7, the *Spe*I-*Sfi*I fragment of pMLP 2ab was substituted by the *Spe*I-*Sfi*I fragment of the T7 cDNA in which BERF2a and BERF2b were fused through splicing. To construct pMLP T7Δ2a, the *Hin*dIII-*Nco*I fragment of pMLP CAT was substituted for the *Spe*I-*Xba*I fragment of pMLP T7. pMLP T7Δ2a thus contains the right end of BERF2a fused to BERF2b. EBV open reading frames (■) are indicated.

The T7 cDNA clone was used to reconstruct a complete fused BERF2a/2b open reading frame under the transcriptional control of the adenovirus major late promoter and using the simian virus 40 T-antigen gene polyadenylation signal (construction pMLP T7) (Fig. 3). In a second construction (pMLP T7Δ2a), most of the BERF2a component of the fused open reading frame (including the first ATG) was deleted in order to determine whether downstream initiation codons were also functional. These two plasmids were tested for the ability to direct the synthesis of EBNA3B in a transient expression assay after transfection of human 293 cells. Proteins corresponding to 2×10^6 cells were run on a 9% sodium dodecyl sulfate–polyacrylamide gel. After being blotted onto a nitrocellulose filter, specific proteins were detected with a 1/25 dilution of a human serum, RS22, taken from a rheumatoid arthritis patient. This serum has been previously shown to react specifically with the EBNA3 family of proteins (14). To eliminate antibody reactivity with cellular proteins migrating close to EBNA3B, the serum was preincubated with mock-transfected 293 cellular protein extract. Western (immuno-) blotting was performed as described previously (9). Only the pMLP T7 construction was seen to express an immunologically reactive protein which comigrated with the 165-kilodalton EBNA3B band of EBV B95-8 (Fig. 4A). This strongly implies that the EBV sequence contained in pMLP T7 is sufficient for the expression of a complete EBNA3B protein. Nevertheless, it cannot be ruled out that the failure to detect EBNA3B expression from pMLP T7Δ2a is due to the lack of antibodies directed against the BERF2b-encoded portion of the protein or to the deletion of sequences critical for proper translation initiation from the methionine codons located downstream. An immunofluorescence experiment (6) (Fig. 4B and C) demonstrated that the protein which is transiently synthesized in 293 cells transfected with pMLP T7 is located in the nucleus. Cells were harvested 24 h after transfection and treated as described previously (9), and RS22 serum (1/10 dilution) was again used to detect EBNA3B protein. In summary, these experiments showed that the pMLP T7 plasmid is capable of expressing a protein with a size, immunological characteristics, and a subcellular localization similar to those of native B95-8 EBNA3B.

Identification of the intron sequence excised in the T7 clone enabled a prediction of the primary structure of EBNA3B to be made. The 938-amino-acid sequence is

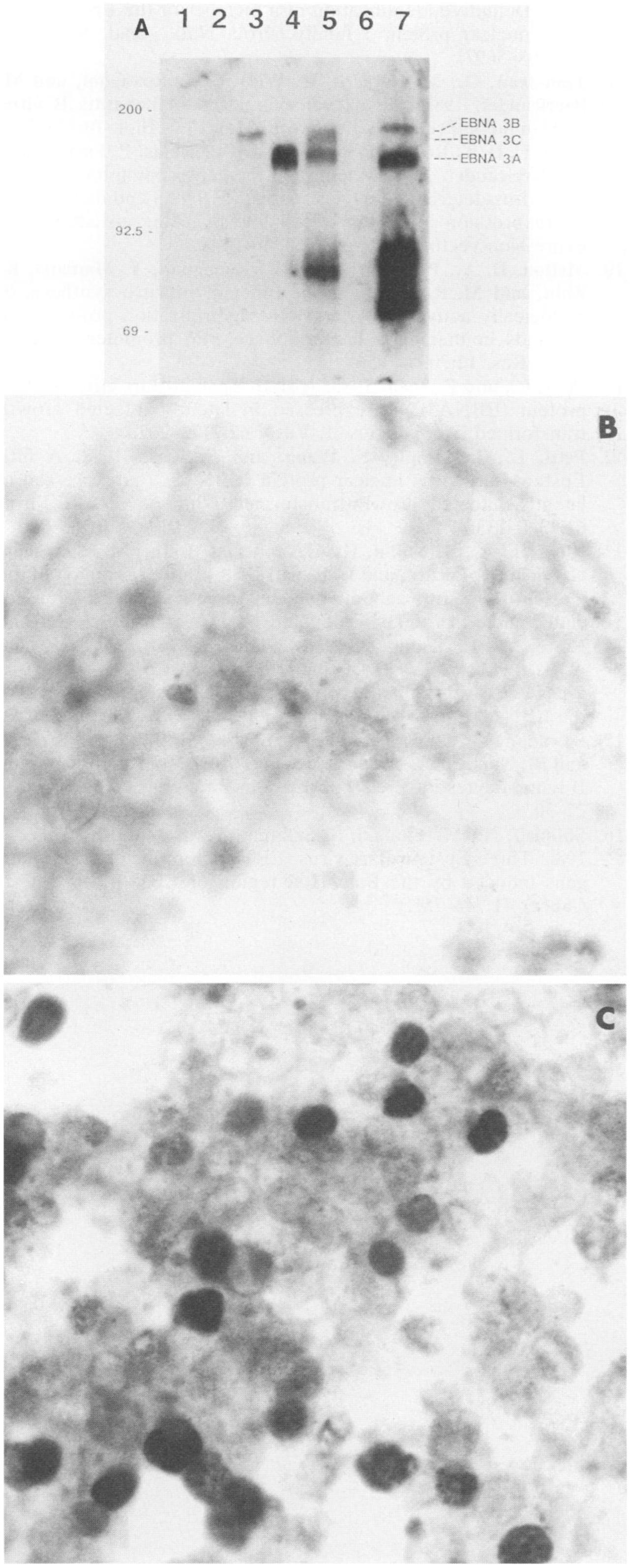

FIG. 4. (A) Western blot analysis of proteins extracted from 293 cells transfected with carrier DNA (lane 1), pMLP T7Δ2a (lane 2), pMLP T7 (lane 3), and pMLP 10T2, which expresses the EBNA3A protein (9) (lane 4), and from BJAB cells superinfected with EBV B95-8 (lane 5), the EBV-negative cell line BJAB (lane 6), and Raji cells, which are EBNA3C negative (lane 7). Molecular masses in kilodaltons are shown on the left and were determined with prestained standard proteins (Amersham Corp.). The positions of EBNA3A, -3B, and -3C are indicated on the right. (B and C) Indirect immunofluorescence staining of 293 cells transfected with carrier DNA (B) or pMLP T7 (C).

```
        10         20         30         40         50
MKKAWLSRAQ QADAGGASGS EDPPDYGDQG NVTQVGSEPI SPEIGPFELS

        60         70         80         90        100
AASEDDPQSG PVEENLDAAA REEEEPHEQE HNGGDDPLDV HTRQPRFVDV

       110        120        130        140        150
NPTQAPVIQL VHAVYDSMLQ SDLRPLGSLF LEQNLNIEEF IWMCMTVRHR
                    ★
       160        170        180        190        200
CQAIRKKPLP IVKQRRWKLL SSCRSWRMGY RTHNLKVNSF ESGGDNVHPV

       210        220        230        240        250
LVTATLGCDE GTRHATTYSA GIVQIPRISD QNQKIETAFL MARRARSLSA

       260        270        280        290        300
ERYTLFFDLV SSGNTLYAIW IGLGTKNRVS FIEFVGWLCK KDHTHIREWF

       310        320        330        340        350
RQCTGRPKAA KPWLRAHPVA IPYDDPLTNE EIDLAYARGQ AMNIEAPRLP

       360        370        380        390        400
DDPIIVEDDD ESEEIEAESD EEEDKSGMES LKNIPQTLPY NPTVYGRPAV

       410        420        430        440        450
FDRKSDAKST KKCRAIVTDF SVIKAIEEEH RKKKAARTEQ PRATPESQAP

       460        470        480        490        500
TVVLQRPPTQ QEPGPVGPLS VQARLEPWQP LPGPQVTAVL LHEESMQGVQ

       510        520        530        540        550
VHGSMLDLLE KDDEVMEQRV MATLLPPVPQ QPRAGRRGPC VFTGDLGIES

       560        570        580        590        600
DEPASTEPVH DQLLPAPGPD PLEIQPLTSP TTSQLSSSAP SCAQTPWPVV

       610        620        630        640        650
QPSQTPDDPT KQSRPPETAA PRQWPMPLRP IPMRPLRMQP IPFNHPVGPT

       660        670        680        690        700
PHQTPQVEIT PYKPTWAQIG HIPYQPTPTG PATMLLRQWA PATMQTPPRA

       710        720        730        740        750
PTPMSPPEVP PVPRQRPRGA PTPTPPPQVP PVPRQRPRGA PTPTPPPQVL

       760        770        780        790        800
PTPMQLALRA PAGQQGPTKQ ILRQLLTGGV KKGRPSLKLQ AALERQAAAG

       810        820        830        840        850
WQPSPGSGTS DKIVQAPIFY PPVLQPIQVM GQGGSPTAMA ASAVTQAPTE

       860        870        880        890        900
YTRERRGVGP MPPTDIPPSK RAKIEAYTEP EMPHGGASHS PVVILENVGQ

       910        920        930
GQQQTLECGG TAKQERDMLG LGDIAVSSPS SSETSNDE
```

FIG. 5. Primary structure of the EBNA3B protein as predicted from the spliced T7 cDNA clone and genomic BERF2a and -2b DNA sequences. The amino acid sequence is presented in the one-letter code, and proline residues are highlighted by shading. The spliced junction between the BERF2a and BERF2b sequences is indicated (★). Amino acid residues found at the same positions in each of EBNA3A, -3B, and -3C are underlined. Arrows show the location of a proline-rich repetition.

shown in Fig. 5. The sequence displays a high proline content (13.5% overall), in particular between amino acids 525 and 755, indicating a structural similarity with the EBNA3A (3, 9) and -3C (12) proteins, which share this feature at approximately the same locations. No major hydrophobic regions were apparent, as predicted from the primary structure, indicating that the protein is not directly associated with the nuclear membrane. Alignment of the three EBNA3 primary sequences revealed a homologous region located between amino acids 140 and 320 (approximately) where 32 residues are conserved at the same position in all three sequences (Fig. 5). The homologous region in EBNA3A constitutes the major part of a sequence already

shown to be of importance in the process leading to nuclear localization (9). We have not, however, been able to locate any known consensus nuclear localization signals on the EBNA3B protein sequence (1). The ability to express the complete EBNA3B protein from plasmid constructs will enable the performance of experiments directed toward the elucidation of its function.

We thank Alan Rickinson for the gift of RS22 antibodies and Jean-Claude Nicolas for help with the immunofluorescence experiments.

D. Walls was the recipient of a long-term European Molecular Biology Organization fellowship.

LITERATURE CITED

1. **Allday, M. J., D. H. Crawford, and B. E. Griffin.** 1988. Prediction and demonstration of a novel Epstein-Barr virus nuclear antigen. Nucleic Acids Res. **16**:4353–4367.
2. **Baer, R., A. Bankier, M. Biggin, P. Deininger, P. Farrel, G. Gibson, G. Hatfull, G. Hudson, C. Satchwell, C. Sequin, P. Fuffnell, and B. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**: 207–211.
3. **Bodescot, M., O. Brison, and M. Perricaudet.** 1986. An Epstein-Barr virus transcription unit is at least 84 Kb long. Nucleic Acids Res. **14**:2611–2620.
4. **Bodescot, M., and M. Perricaudet.** 1986. Epstein-Barr virus mRNAs produced by alternative splicing. Nucleic Acids Res. **14**:7103–7114.
5. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
6. **Hearing, C., J. C. Nicolas, and A. J. Levine.** 1984. Identification of Epstein-Barr virus sequence that encodes nuclear antigen expressed in latently infected lymphocytes. Proc. Natl. Acad. Sci. USA **84**:4373–4377.
7. **Hennessy, K., F. Wang, E. Woodland-Bushman, and E. Kieff.** 1986. Definitive identification of a member of the Epstein-Barr virus nuclear protein 3 family. Proc. Natl. Acad. Sci. USA **83**:5693–5697.
8. **Jean-Jean, O., M. Levrero, H. Will, S. M. Rossignol, and M. Perricaudet.** 1987. High level expression of hepatitis B virus c-orf in human cells. UCLA Symp. Mol. Cell. Biol. **70**:223–235.
9. **Joab, I., D. T. Rowe, M. Bodescot, J. C. Nicolas, P. Farrell, and M. Perricaudet.** 1987. Mapping of the gene coding for Epstein-Barr virus-determined nuclear antigen EBNA3 and its transient overexpression in a human cell line by using an adenovirus expression vector. J. Virol. **61**:3340–3344.
10. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
11. **Petti, L., and E. Kieff.** 1988. A sixth Epstein-Barr virus nuclear protein (EBNA3B) is expressed in latently infected growth transformed lymphocytes. J. Virol. **62**:2173–2178.
12. **Petti, L., J. Sample, F. Wang, and E. Kieff.** 1988. A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in latently infected growth-transformed lymphocytes. J. Virol. **62**:1330–1338.
13. **Ricksten, A., B. Kallin, H. Alexander, J. Dillner, R. Fahraeus, G. Klein, R. Lerner, and L. Rymo.** 1988. BamHI E region of the Epstein-Barr virus genome encodes three transformation associated nuclear proteins. Proc. Natl. Acad. Sci. USA **85**:995–999.
14. **Rowe, M., L. Young, K. Cadwallader, L. Petti, E. Kieff, and A. B. Rickinson.** 1989. Distinction between Epstein-Barr virus type A (EBNA 2A) and type B (EBNA 2B) isolates extends to the EBNA3 family of nuclear proteins. J. Virol. **63**:1031–1039.
15. **Sawada, K., M. Yamamoto, T. Tabata, M. Smith, A. Tanaka, and M. Nonoyama.** 1989. Expression of EBNA 3 family in fresh B lymphocytes infected with Epstein-Barr virus. Virology **168**: 22–30.
16. **Shimizu, N., M. Yamaki, S. Sakuma, Y. Ono, and K. Takada.** 1988. Three Epstein-Barr virus (EBV)-determined nuclear antigens induced by the BamHI E region of EBV DNA. Int. J. Cancer **41**:744–751.