Homologous upstream sequences near Epstein–Barr virus promoters

(transcription/RNA polymerase II/DNA sequence determination)

PAUL J. FARRELL, PRESCOTT L. DEININGER*, ALAN BANKIER, AND BART BARRELL

Laboratory of Molecular Biology, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, England

ABSTRACT   The sequence of the 17,166-base-pair EcoRI C fragment of Epstein–Barr virus DNA, cell line B95-8, was determined. In vitro transcription was used to identify three RNA polymerase II promoters within this fragment of the virus. Cytoplasmic poly(A)$^+$RNAs starting at these points were demonstrated in B95-8 cells induced into virus production with 12-O-tetradecanoylphorbol 13-acetate. Uninduced B95-8 cells contained much less of these RNAs. The upstream sequences near the three promoters show striking homologies that may be involved in transcriptional control.

Epstein–Barr virus (EBV) is a human herpesvirus that is the infectious agent in infectious mononucleosis and is involved with Burkitt lymphoma and nasopharyngeal carcinoma (for review, see ref. 1). Cell lines that produce EBV have been obtained from tumor explants and by transforming tissue culture cells. The B95-8 cell line was generated by transforming marmoset blood leukocytes with EBV (2). Virus production can be stimulated up to 50-fold (3) by treatment of the cells with 12-O-tetradecanoylphorbol 13-acetate (TPA). The virus from the B95-8 line has been widely studied: the viral genome in the capsid is linear double-stranded (ds) DNA of ≈170-kilobase (kb) length. The restriction map of EcoRI sites and BamHI sites is known, and these fragments of the virus have been cloned (4–6).

During infection, the viral DNA is transcribed by the host cell RNA polymerases II and III. Viral RNAs have been mapped onto the viral genome by RNA blot analysis (7). This method gives information on the size of the RNAs but does not closely map the ends of the RNAs or their orientation on the viral genome. We analyzed EBV gene expression by determining the viral DNA sequence and then mapping polymerase II transcription promoters on that sequence. We determined the sequence of the EcoRI C fragment and identified three RNA polymerase II promoters within that sequence. The promoters function in vivo because RNAs with the predicted 5' termini accumulate in B95-8 cells in response to TPA induction. The promoters we identified contain conserved blocks of sequence homology upstream of the start points of transcription.

MATERIALS AND METHODS

The B95-8 and Raji EcoRI C fragment clones were isolated from cosmid libraries (6). The M13 clones 3BV20, SON28, and LEG34 were isolated during the DNA sequence determination. The pL2 clone contains the EcoRI–HindIII (bases 0–4,230) portion of the EcoRI C fragment in the large EcoRI–HindIII fragment of pBR322. The pR clone contains the HindIII–Bgl II (bases 4,230–7,775) portion of the EcoRI C fragment in the large HindIII–BamHI fragment of pBR322. The pL1 clone contains the Bgl II–

EcoRI (bases 14,902–17,167) portion of the EcoRI C fragment in the large EcoRI–BamHI fragment of pBR322. The M13 clones (LEG 156, LEG 28, and EBDS 27) used for preparation of nuclease S1 mapping probes contain, respectively, bases 5,540–6,033, 16,612–17,050, and 419–1,002 of the EcoRI C fragment of B95-8 EBV DNA.

The sequence of the EcoRI C fragment was determined by the M13/dideoxy method (8–10) with random clones generated by sonication. The DNA sequence and a detailed analysis of its coding properties will be published elsewhere.

In vitro transcription was in HeLa (S3) cell extracts (11) as described (12). Transcripts were analyzed by electrophoresis through denaturing polyacrylamide gels (13) or agarose gels after glyoxylation (14).

B95-8 cells were grown in RPMI 1640 medium containing 10% fetal calf serum. Permissive infection was induced by adding TPA (20 ng/ml) and harvesting the cells 3 days later. Cytoplasmic RNA was extracted and chromatographed on oligo(dT)-cellulose as described (15).

RNAs transcribed in vitro were mapped by the nuclease S1 technique (16) with end-labeled ds DNA probes (see legends for details).

Single-stranded probes for mapping B95-8 cell RNA from the R and L1 promoters were prepared by the "prime, cut" method (D. Bentley, personal communication). Single-stranded M13 clone DNA was hybridized to the sequence assay primer. A radioactive complementary strand was synthesized by using Klenow DNA polymerase and [α-^{32}P]dATP as label. After digestion with the appropriate restriction enzyme, the probe was separated from its complementary cold strand on a denaturing polyacrylamide gel (13).

The ds DNA probe used for the L2 promoter was prepared by complementary strand synthesis from EBDS 27 as above but was digested with BstNI, and the probe was recovered from a native gel.

RESULTS

Sequence Analysis of the EcoRI C Fragment of EBV DNA. The sequence of the 17,166-base EcoRI C fragment was determined; major open reading frames, located by computer, are indicated on the diagram of the sequence (Fig. 1). We searched the sequence for the A-A-U-A-A-A sequence, which is part of the signal for generation of the 3' end of mRNA (17). Positions of several of these also are shown in Fig. 1. Although RNA splicing is likely to be important in EBV gene expression, we cannot reliably predict from the sequence where splices actually occur.

The EcoRI C fragment of the B95-8 virus contains a deletion of 13.6 kb relative to most strains of EBV (18). By comparative

Abbreviations: EBV, Epstein–Barr virus; TPA, 12-O-tetradecanoylphorbol 13-acetate; kb, kilobase(s); ds, double-stranded.
* Present address: Dept. of Biochemistry, L.S.U. Medical Center, 1901 Perdido St., New Orleans, LA 70112.
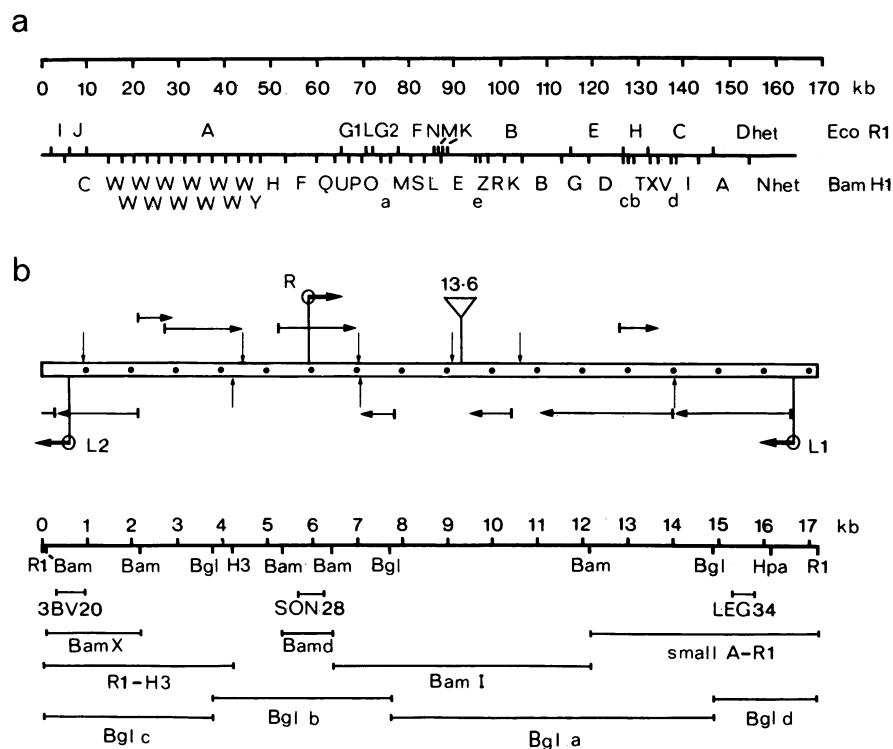
FIG. 1. (*a*) Restriction map of *Eco*RI and *Bam*HI sites in B95-8 EBV DNA from ref. 6. (*b*) Map of the *Eco*RI C fragment of B95-8 EBV DNA showing open reading frames (horizontal thin arrows), A-A-U-A-A-A sequences (vertical thin arrows), and the deletion relative to EBV from Raji cells (vertical line and triangle 13.6). We selected reading frames on grounds of length, AUG codon content, and patterns of codon usage. The positions of the three promoters (L1, L2, and R) are marked. On the restriction map of the *Eco*RI C fragment, all of the *Bam*HI (Bam), *Bgl* II (Bgl), and *Hin*dIII (H3) sites are shown, but there are other *Hpa* I (Hpa) sites. Restriction fragments used for *in vitro* transcription experiments are shown.

sequence assay of the B95-8 and Raji *Eco*RI C fragments, the deletion was found to be located after base 9,326 of the B95-8 sequence.

Transcription from Promoters Within the *Eco*RI C Fragment. We used *in vitro* transcription (11) to screen the *Eco*RI C fragment for RNA polymerase II promoter activity. Although

transcription in this system does not depend on upstream promoter elements, as does *in vivo* transcription, the system initiates accurately (11, 19–21). We used run-off transcription assays (22) to test various restriction fragments (designated as in Fig. 1) from the *Eco*RI C fragment. With this assay, the start of transcription is located at a distance equal to the length of the
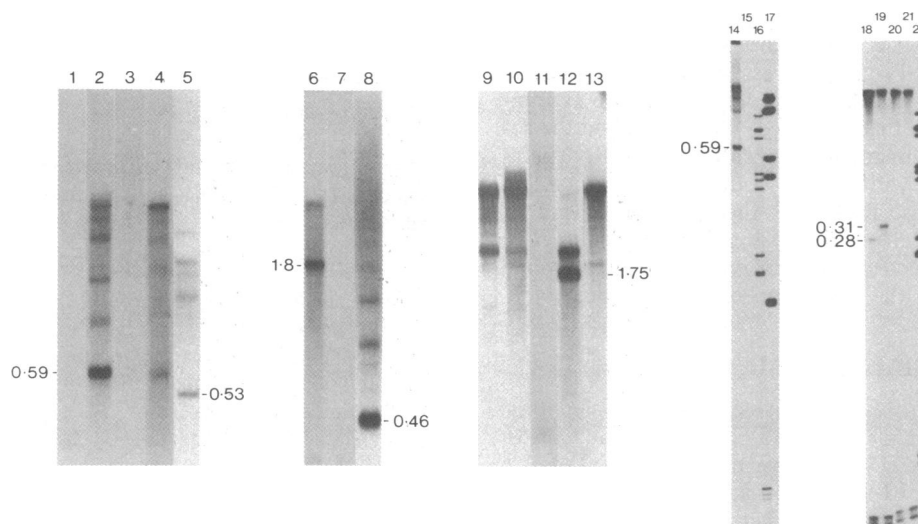

FIG. 2. Transcription of various restriction fragments of the *Eco*RI C segment in the Hela cell extract; see Fig. 1 to identify templates. (*Left*) Autoradiographs of 1.4% agarose gels. Tracks 1–13: 1, no DNA; 2, *Eco*RI–*Hin*dIII (R1-H3) DNA; 3, R1-H3 DNA with α-amanitin at 0.5 μg/ml; 4, *Bgl* c DNA; 5, *Bam* X DNA; 6, *Bgl* b DNA; 7, *Bgl* b DNA with α-amanitin at 0.5 μg/ml; 8, *Bam* d DNA; 9, pBR322 DNA cut with *Bam*HI; 10, *Bam* I DNA cut with *Bam*HI (pBR322 vector present); 11, *Bgl* a DNA; 12, *Bgl* d DNA; 13, small A-R1 DNA. Sizes of transcripts were determined from coelectrophoresed rRNAs and a *Hin*dIII digest of simian virus 40 DNA. (*Right*) Autoradiographs of 5% polyacrylamide gels. Tracks 14–22: 14, R1-H3 DNA; 15, no DNA; 16 and 22 (size markers), *Hae* III digests of bacteriophage G4 DNA; 17, *Hin*dIII digest of simian virus 40 DNA; 18, 3BV20 DNA cut with *Eco*RI; 19, SON28 DNA cut with *Eco*RI; 21, LEG 34 DNA cut with *Eco*RI. In tracks 18, 19, and 21, the M13 vector DNA was present, so these transcriptions should be compared with track 20 (M13 vector cut with *Eco*RI).

Biochemistry: Farrell *et al.*

Proc. Natl. Acad. Sci. USA 80 (1983) 1567

observed transcript back from the restriction site marking the end of the template fragment. Testing various overlapping DNAs allows one to deduce the direction of transcription.

Transcription of the *Eco*RI–*Hin*dIII fragment (bases 0–4,230; designated R1–H3) gave a strong transcript of ≈0.59 kb (Fig. 2, tracks 2 and 14). A transcript of identical size was obtained from the restriction fragment designated *Bgl* c (Fig. 2, track 4), which shares its left end but not its right end with the R1–H3 fragment. This suggests that there is a promoter reading leftwards from base ≈590 (designated promoter L2). Larger, weaker bands present in these transcriptions were not found reproducibly and were not studied further. Transcription of the fragment designated *Bam* X (bases 54–2,176) (Fig. 2, track 5) gave rise to the expected 0.53-kb run-off RNA. Transcription of the 0.59-kb run-off product from the R1–H3 fragment was inhibited by α-amanitin at 0.5 μg/ml (Fig. 2, track 3), indicating that RNA polymerase II was responsible for the transcription (23).

Transcription of the fragment designated *Bgl* b (bases 3,787–7,775) gave a 1.8-kb run-off RNA (Fig. 2, track 6). This comes from a promoter reading rightwards because the fragment designated *Bam* d (bases 5,321–6,429) gave a strong transcript, which was only 0.46 kb in length (Fig. 2, track 8). Thus, a transcription start is located at about base 5,970 reading rightwards (the R promoter). The transcription of the *Bgl* b fragment was inhibited by α-amanitin at 0.5 μg/ml (Fig. 2, track 7), so transcription was by RNA polymerase II.

The fragment designated *Bam* I (bases 6,429–12,061) was not separated from the pBR322 vector, so the transcription pattern (Fig. 2, track 10) should be compared with that of pBR322 cut with *Bam*HI (Fig. 2, track 9). No specific transcripts were detected from the *Bam* I fragment, though the pBR322 gave a substantial amount of transcription as has been reported (24). The fragment designated *Bgl* a (bases 7,775–14,902) did not yield any specific transcripts (Fig. 2, track 11).

The *Bgl* d fragment yielded a 1.75-kb run-off RNA (Fig. 2, track 12). The band at 2.1 kb in this track resulted from end-to-end transcription of the DNA template, which is often found if the template concentration is greater than optimal. The transcription of the 1.75-kb RNA from *Bgl* d was inhibited by α-amanitin at 0.5 μg/ml (data not shown), showing that it is a RNA polymerase II transcript. The *Bgl* d fragment and the *Eco*RI small fragment A (designated small A-R1) share their right-hand but not their left-hand ends. The 1.75-kb run-off transcript from the *Bgl* d fragment was not observed when the small A-R1 fragment was transcribed (Fig. 2, track 13), so the direction of transcription from *Bgl* d is leftwards from a start at about base 16,650. We call this the L1 promoter. This promoter should yield a 4.6-kb run-off transcript on the small A-R1 fragment. There were transcripts of this size (Fig. 2, track 13), but we are not certain about distinguishing these from end-to-end transcripts of 5 kb. We also transcribed (not shown) a subclone cut with *Hpa* I, which gave a run-off of about 500 bases, confirming that the L1 transcription start is leftwards from about base 16,650.

To provide further support for the L2 and R promoters, M13 ds DNA clones containing the promoters (3BV20 and SON28) were transcribed. The predicted run-off transcripts of 280 and 310 bases (Fig. 2, tracks 18 and 19) were obtained.

Nuclease S1 Mapping of *in Vitro* Transcribed RNA. The 5′ termini of RNAs transcribed from the L2, R, and L1 promoters were determined more precisely by nuclease S1 digestion of hybrids formed between 5′ end-labeled DNA probes and *in vitro* transcripts (details in Fig. 3). For promoter L2, 535 bases of the *Bam* X fragment were protected by *in vitro* transcripts (Fig. 3, track 3), indicating a 5′ terminus at base 590. For promoter R, the 465 protected bases of the *Bam* d probe (Fig. 3, track 7) indicate a start at about base 5,970. For promoter L2, the 65 protected bases of the *Bst*NI probe (Fig. 3, track 13) map the

start to about base 16,650. Background bands in the L1 assay could not be avoided even at higher hybridization temperatures and may result from secondary structure in the DNA probe. For each promoter, run-off and nuclease S1 mapping results were in close agreement.

The Promoters Work in B95-8 Cells. We tested whether the promoters work in B95-8 cells by nuclease S1 mapping cytoplasmic RNAs from cells induced to a productive infection with TPA. We also tested RNAs from control B95-8 cells, which have a low level of virus production. To give greater sensitivity, uniformly labeled probes were used. The R promoter was mapped with a probe from M13 clone LEG 156, containing EBV DNA from base 6,033 to a *Bal* I site at base 5,914. The probe is 155 bases long of which 120 bases are complementary to EBV DNA. A transcription start at base 5,970 predicts a protected fragment of 63 bases or slightly longer if the cap hinders nuclease S1 digestion. Major protected fragments of 65–66 bases were ob-
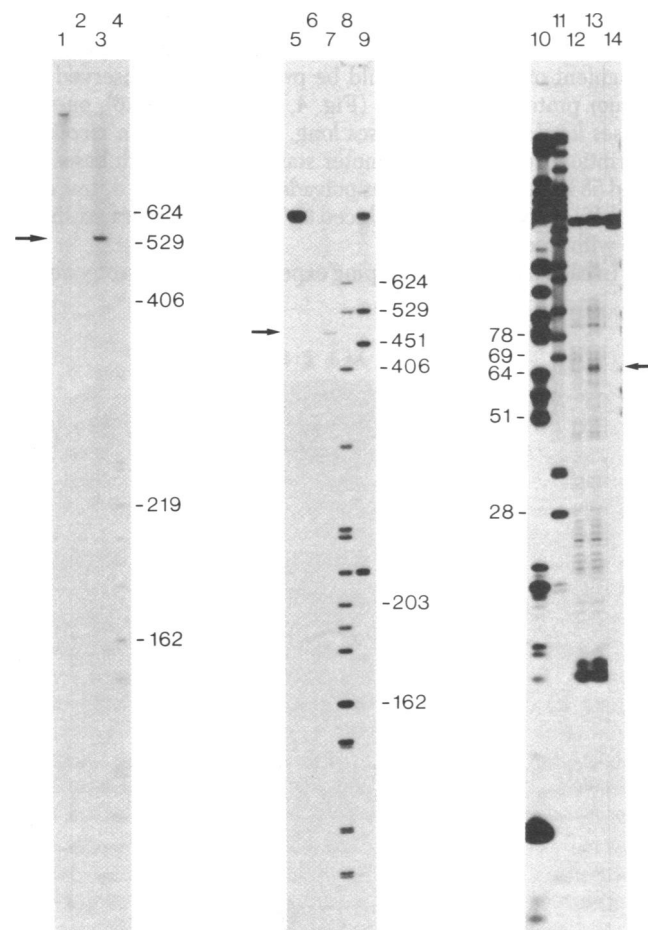

FIG. 3. Nuclease S1 mapping 5′ ends of *in vitro* transcripts from *Eco*RI C fragment promoters. The probe for mapping the L2 promoter was the *Bam* X fragment and the probe for the R promoter was the *Bam* d fragment. These were 5′ end-labeled by using T4 polynucleotide kinase. The L1 promoter was mapped with a *Bst*NI–*Hae* III fragment (bases 16,588–16,761). RNA (from transcriptions containing no radioactive label) was hybridized to the relevant probe in 80% formamide buffer (16) for 4 hr at 50°C for L2, 59°C for R, and 47°C for L1. After nuclease S1 digestion (16), the protected fragments were run on polyacrylamide gels; autoradiographs are shown. Tracks: 1–3, probe L2; 5–7, probe R; 12–14, probe L1; 1, 5, and 14, untreated probes; 2, 3, 6, 7, 12, and 13, nuclease S1 mapping reactions. Transcription reactions contained the following template DNAs: no template (tracks 2, 6, and 12), pL2 DNA cut with *Eco*RI (track 3), pR DNA cut with *Sal* I (track 7), pL1 DNA (track 13). Tracks 4, 8, and 11 (size markers) contain an *Msp* I digest of pBR322 DNA, track 9 contains a *Hin*dIII digest of simian virus 40 DNA, and track 10 contains a *Hae* III digest of pBR322 DNA.

served (Fig. 4, tracks 3 and 4). The signal was stronger with the poly(A)$^+$RNA (Fig. 4, track 3) than with the poly(A)$^-$RNA (track 4), even though 50 times as much poly(A)$^-$RNA was used, so the RNA transcribed from the R promoter is a poly(A)$^+$RNA. The RNA also is induced strongly by TPA; tracks 5 and 6 contain the parallel S1 mapping of poly(A)$^+$RNA and poly(A)$^-$RNA from uninduced B95-8 cells.

The L1 promoter was mapped with a probe from clone LEG 28 containing EBV DNA from base 16,612 to a *Bst*NI site at base 16,764. The probe was 187 bases long, containing 152 bases complementary to EBV DNA. We predicted a protected fragment of ≈38 bases or a little longer, with a transcription start around base 16,650. The protected fragment was ≈40 bases long (Fig. 4, tracks 9 and 10), confirming the *in vitro* start. Again the RNA was polyadenylylated (compare tracks 9 and 10 in Fig. 4), and the RNA was strongly induced by TPA (compare tracks 9 and 10 with 11 and 12 in Fig. 4).

The probe for the L2 promoter was made from clone EBDS 27 and contained EBV DNA from base 472 to base 749. If the L2 promoter starts transcription at about base 590, a protected fragment of 118 bases would be predicted. We observed two major protected fragments (Fig. 4, tracks 15 and 16), one 118 bases long and one 112 bases long, implying that *in vivo* transcription from the L2 promoter starts at about both bases 590 and 583. Again the RNA was polyadenylylated (Fig. 4, tracks 15 and 16) and was strongly induced by TPA (compare tracks 15 and 16 with tracks 17 and 18).

In each nuclease S1 mapping experiment, some protection of

the full length of the EBV sequence in the probe was seen. This probably represents EBV RNA transcribed right through all of the probe regions. The other possibility, protection by EBV DNA, is unlikely because the protected material is apparently polyadenylylated. In summary, all three promoters produce polyadenylylated RNAs, which accumulate during TPA induction of the productive virus cycle.

DISCUSSION

We tried to correlate the transcripts detected by RNA blotting analysis (7) with our results. There is (7) a 1.4-kb RNA that hybridizes to only the Bam d (Bam b in ref. 7) and Bam I segments. The R promoter transcribes from position 5,970 and the first downstream A-A-U-A-A-A sequence is at position 7,041. With, say, poly(A)$_{150}$, this would give an RNA of about 1.3 kb, correlating with the 1.4-kb one detected in ref. 7. The small A-R1 segment (Fig. 1) gives rise to a 3.0-kb late RNA in B95-8 cells. This may be a transcript from the L1 promoter. This promoter leads into an open reading frame which stops just before a poly(A) addition site at base 14,018. Transcription from the L2 promoter may give rise to the 2.7-kb late RNA which comes from the segments designated Bam T and X (T and W in ref. 7). There is an open reading frame running from the L2 promoter through the *Eco*RI–*Hind*III segment compatible with this.

Sequence Comparison of Promoters. The DNA sequences around the L2, R, and L1 promoters are shown in Fig. 5. Each transcription start is about 30 bases downstream from a se-
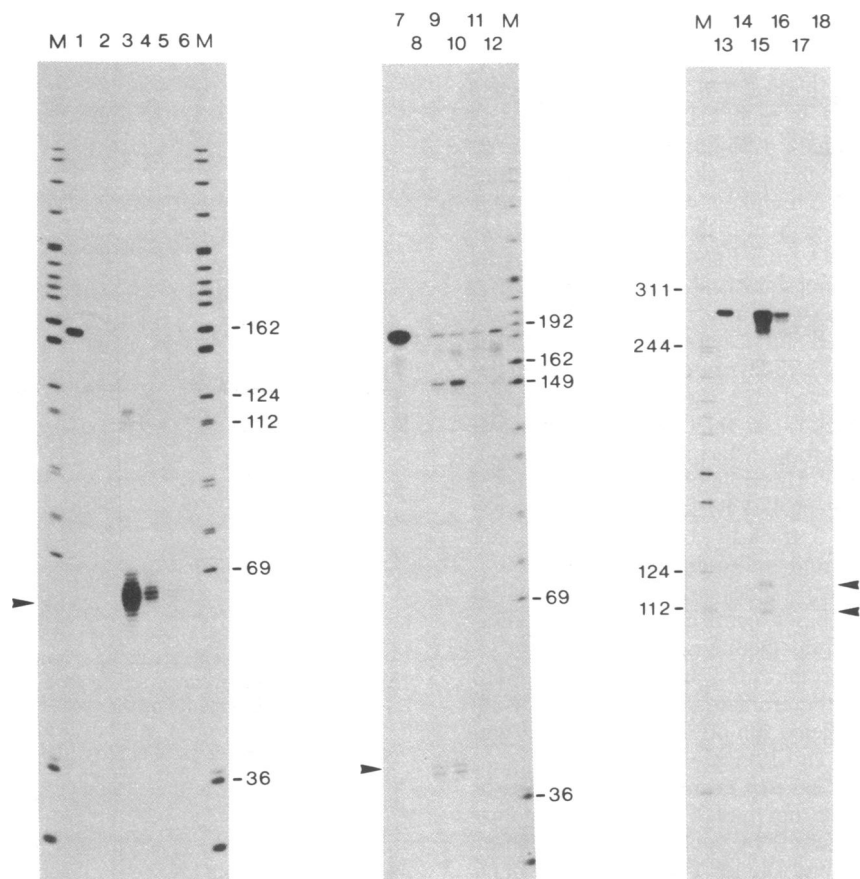

FIG. 4. Nuclease S1 mapping 5′ ends of transcripts from the R (tracks 1–6), L1 (tracks 7–12), and L2 (tracks 13–18) promoters in B95-8 cells. Hybridizations were in 50% formamide/0.5 M NaCl/40 mM Pipes, pH 6.4/1 mM EDTA overnight at 42°C (tracks 1–12) or in 80% formamide/0.5 M NaCl/50 mM Pipes, pH 6.4/1 mM EDTA overnight at 49°C (tracks 13–18). Samples were then treated with nuclease S1 (16) and analyzed on gels (13). Tracks: 1, 7, and 13, untreated probes; 2, 8, and 14, hybridizations with no RNA; 3, 9, and 15, 1 μg of poly(A)$^+$RNA from B95-8 cells induced with TPA; 4, 10, and 16, 50 μg of poly(A)$^-$RNA from B95-8 cells induced with TPA; 5, 11, and 17, 1 μg of poly(A)$^+$RNA from B95-8 cells not treated with TPA; 6, 12, and 18, 50 μg of poly(A)$^-$RNA from B95-8 cells not treated with TPA. DNA size markers (M) are an *Msp* I digest of pBR322.

```
        -90        -80        -70        -60        -50        -40        -30        -20        -10         0         10
L2   ACAGTATGTCATCAAAGCCACACCCCCATAATTCGAAGTGGTGTGGCCGAGAAGCTGTATAAGAACAGCATCTTTAGTGTACTTAAGGAAGTCATACAGCC

R    AACCGCGTGGAGACTCTGAGGCAGGACAGGGGTGTGGCCAATATCCCAGCCGAGTCGTATTTAAAGGCCAGCGACGCCCCGGACCTACAAAAGCCGGACAA

L1   ACCTCACGTACGACTCTGACCGCCCCCTAATTCTTCTGTATAAGGGCAGGGGTGGGTATTTAAGGATCTATATGCCCTTCTCTACCTGCACCTCCAAATG
```
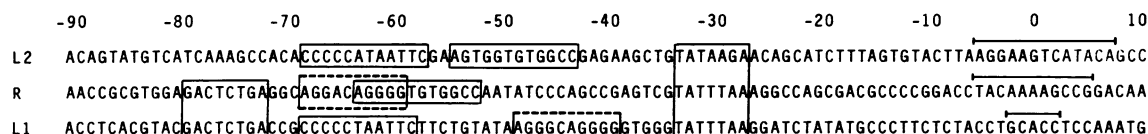
FIG. 5. Comparison of the sequences around the L1, R, and L2 promoters in the *Eco*RI C fragment of EBV DNA. The bars around base 0 indicate the error limit of the mapping of the start points of transcription. The boxes indicate the TATA boxes and the regions of sequence homology.

quence homologous to the "TATA" box (25), which is found up-stream of most polymerase II transcription starts. A TATA box alone is insufficient to promote *in vitro* transcription because the sequence T-A-T-A-A-A-A (identical with the adenovirus major late promoter TATA box) at position 15,520 in clone LEG 34 DNA is inactive (Fig. 2, track 21). Perhaps the sequence near the TATA box contributes to the activity of the promoter *in vitro*; the adenovirus major late promoter sequence is extremely G+C rich around the TATA box, whereas the LEG 34 sequence is not.

In vivo, promoter elements upstream of the TATA box, usually in the region −40 to −100, are crucial for promoter function (26–28). There are remarkable sequence homologies in the DNA sequences upstream of the TATA boxes in the three EBV promoters (Fig. 5). The sequence G-A-C-T-C-T-G-A is shared in an exactly equivalent position between promoters R and L1. The sequence C-C-C-C-C-T-A-A-T-T-C in promoter L1 is present in an identical position in L2 with a single A residue inserted in the middle, and the sequence A-G-T-G-G-T-G-T-G-G-C-C in L2 at around position −45 is also present (one mismatch) in R at around −55. The sequence A-G-G-G-C-A-G-G-G-G present at −40 in L1 is also present (one mismatch) around −62 in R. There is the capacity for formation of a stem loop structure in L2 between the sequences G-G-T-G-T-G-G-C (−45) and G-C-C-A-C-A-C-C (−67), and there is also an inverted homology between the sequences around −60 in R (G-G-G-G-T-G-T-G-G-C) and −70 in L2 (G-C-C-A-C-A-C-C-C-C). The sequences of promoters L1, R, and L2 downstream of the TATA boxes or upstream of position −100 are not so closely related.

We asked whether other eukaryotic promoters also contain these sequences. The G-A-C-T-C-T-G-A and C-C-C-C-C-T-A-A-T-T-C sequences are restricted to EBV DNA, but the A-G-G-G-G-T-G-T-G-G-C-C sequence is distributed more generally. With a consensus sequence of G-G-G-G-T-G-T-G-G-C-C and looking at 35 promoters, we found a homologous sequence in the EBV L2 and R promoters, the herpes simplex virus thymidine kinase gene (9/11 match), a herpesvirus early gene (9/11), a mouse β-globin (minor) gene (9/11), a mouse β-globin (major) gene (9/11), a rabbit β-globin gene (9/11), a human α-globin gene (9/11), a human β-globin gene (9/11), and the adenovirus EIIa gene (10/11). The homologous sequences are in the same orientation as in the EBV promoters in the thymidine kinase gene and in the adenovirus EIIa gene but in the opposite orientation in the others. In the globin genes, the herpesvirus early gene, and the adenovirus EIIa gene, the homologous sequence is around −90 from the transcription start, but, in the thymidine kinase gene, the homologous sequence is around −10. There is only a 1% chance of finding a 9/11 match in 50 bases of ds DNA sequence, so the homologies are highly statistically significant. A recent report (29) on the rabbit β-globin gene promoter identifies the two regions required for efficient transcription of this gene as the "CAAT" box and the −81 to −95 region, which contains the 9/11 match to the G-G-G-G-T-G-T-G-G-C-C sequence element.

Some other promoters we have located in other parts of the virus have some of the sequence homology blocks at similar locations upstream of their TATA boxes, but others do not (unpublished results). This suggests the interesting possibility that different combinations of these homologous sequences might comprise signals for the control of the promoters, perhaps allowing them to be active at particular stages in the virus life cycle or to be independently regulated.

1. Tooze, J., ed. (1980) *DNA Tumor Viruses* (Cold Spring Harbor Publications, Cold Spring Harbor, NY), 2nd Ed.
2. Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 383–387.
3. zur Hausen, H., Hecker, E., O'Neill, F. J. & Freese, U. K. (1978) *Nature (London)* **272**, 373–375.
4. Dambaugh, T., Beisel, C., Hummel, M., King, W., Fennewald, S., Cheung, A., Heller, M., Raab-Traub, N. & Kieff, E. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2999–3003.
5. Skare, J. & Strominger, J. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3860–3864, and correction (1980) **77**, 7510.
6. Arrand, J. R., Rymo, L., Walsh, J. E., Bjorck, E., Lindahl, T. & Griffin, B. E. (1981) *Nucleic Acids Res.* **9**, 2999–3014.
7. Hummel, M. & Kieff, E. (1982) *J. Virol.* **43**, 262–272.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5468.
9. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. (1980) *J. Mol. Biol.* **143**, 161–178.
10. Heidecker, G., Messing, J. & Gronenbourn, B. (1980) *Gene* **10**, 69–73.
11. Manley, J. L., Fire, A., Cano, A., Sharp, P. & Gefter, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3855–3859.
12. Bentley, D. L., Farrell, P. J. & Rabbitts, T. H. (1982) *Nucleic Acids Res.* **10**, 1841–1856.
13. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
14. McMaster, G. K. & Carmichael, G. C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4835–4838.
15. Farrell, P. J., Broeze, R. J. & Lengyel, P. (1978) *Nature (London)* **279**, 523–525.
16. Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1274–1278.
17. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
18. Raab-Traub, N., Dambaugh, T. & Kieff, E. (1980) *Cell* **22**, 257–267.
19. Tsai, S. Y., Tsai, M. & O'Malley, B. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 879–883.
20. Rio, D., Robbins, A., Myers, R. & Tjian, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5706–5710.
21. Proudfoot, N. J., Shander, M. H. M., Manley, J. L., Gefter, M. L. & Maniatis, T. (1980) *Science* **209**, 1329–1336.
22. Weil, P. A., Luse, D. S., Segall, J. & Roeder, R. G. (1979) *Cell* **18**, 469–484.
23. Blatti, S. P., Ingles, C., Lindell, T., Morris, P., Weaver, R., Weinberg, F. & Rutter, W. J. (1970) *Cold Spring Harbor Symp. Quant. Biol.* **35**, 649–657.
24. Sassone-Corsi, P., Corden, J., Kedinger, C. & Chambon, P. (1981) *Nucleic Acids Res.* **9**, 3941–3958.
25. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1414.
26. McKnight, S. L., Gavis, E. R., Kingsbury, R. & Axel, R. (1981) *Cell* **25**, 385–398.
27. Pelham, H. R. B. (1982) *Cell* **30**, 517–528.
28. Dierks, P., van Ooyen, A., Mantei, N. & Weissmann, C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1411–1415.
29. Grosveld, G. C., Rosenthal, A. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 4951–4971.