Nucleotide Sequence of an mRNA Transcribed in Latent Growth-Transforming Virus Infection Indicates That It May Encode a Membrane Protein

SUSAN FENNEWALD, VICKY VAN SANTEN, AND ELLIOTT KIEFF*

Kovler Viral Oncology Laboratories, University of Chicago, Chicago, Illinois 60637

Received 23 February 1984/Accepted 23 April 1984

The most abundant Epstein-Barr virus mRNA in a latently infected cell line, IB4, established by *in vitro* growth transformation with virus, was a 2.8-kilobase RNA encoded by largely unique DNA near the right end of the genome. The RNA was transcribed from right to left, and two introns were spliced out. This region of the genome was sequenced, and the exons of the RNA were identified by S1 analysis of DNA-RNA hybrids and primer extension. The first start codon in the RNA was 40 nucleotides from its 5' end. Beginning with the start codon, there was a 1,158-nucleotide open reading frame which crossed both introns. The important characteristics of the translated protein were as follows. (i) The amino terminus was highly charged and not suggestive of a leader sequence. (ii) There were six markedly hydrophobic alpha-helical domains, each having 21 amino acids and connected by 5 to 7 amino acid segments predicted to be reverse turns. (iii) The carboxy-terminal 200 amino acids were markedly acidic, containing 6 glutamic and 37 aspartic acids. The hydrophobic region is predicted to form six membrane-spanning regions, leaving the short charged amino terminus and long acidic carboxy terminus on the inside of the membrane. This protein could be responsible for the new antigen detected in the plasma membrane of Epstein-Barr virus-transformed cells, lymphocyte-determined membrane antigen. There were two other open reading frames in the RNA.

Epstein-Barr virus (EBV) infection of human lymphocytes results in continuous cell proliferation and little or no virus replication (21, 29, 46). The entire viral genome persists in the proliferating lymphocytes as linear-integrated (28; M. Heller, T. Matsuo, E. Oshiro, and E. Kieff, manuscript in preparation) and covalently closed circular episomal DNA (38, 45). The viral genome is 170,000 base pairs and consists of largely unique (U), internal direct repeat (IR), and terminal direct repeat (TR) elements arranged in the format TR-U1-IR1-U2-IR2-U3-IR3-U4-IR4-U5-TR (11, 13, 23, 27, 48).

Three virus mRNAs have been identified in latently infected growth-transformed cells. These RNAs are a 3.0-kilobase (kb) RNA encoded by IR1-U2, a 3.7-kb RNA encoded by U3-IR3-U4, and a 2.8-kb RNA encoded by U5 (34, 61, 62). These EBV DNA domains are also transcribed in latently infected Burkitt tumor cells (2, 12, 35, 47). Burkitt tumor is a B lymphocyte malignancy associated with EBV infection.

The RNAs encoded by the IR1-U2 and U3-IR3-U4 domains are transcribed from left to right from the EBV genome (Fig. 1) and are spliced (27, 30, 61). There are approximately three copies of the mRNAs encoded by these domains in latently infected, *in vitro*-transformed, IB4 cells. A 2-kb exon of the 3.7-kb U3-IR3-U4 RNA begins in U3 and extends through IR3 into U4. An EBV-specified nuclear antigen (EBNA; 52) is encoded by this RNA (30, 31). Much of the nucleotide sequence and the reading frame for translation of this protein from the U3-IR3-U4 domain have been previously determined (31). Some human sera identify a second EBNA protein in latently infected, growth-transformed B lymphocytes. This protein is not encoded by the U3-IR3-U4 domain (31). Preliminary data now suggest that this protein is encoded by the IR1-U2 domain (T. Dambaugh, K. Hennessey, and E. Kieff, manuscript in preparation).

No detailed analyses have been previously published of

the U5 mRNA. We now report the complete nucleotide sequence of the exons and introns and of the 3' and 5' proximal regions of the EBV U5 domain which encodes this RNA. The data indicate that the RNA has a single large open reading frame which may encode a membrane protein. This could be related to the new lymphocyte-determined membrane antigen (LYDMA), which is on the surface of EBV growth-transformed lymphocytes (40, 53, 60).

MATERIALS AND METHODS

Cell culture, RNA preparation, and RNA blots. The IB4 cell line was established by *in vitro* transformation of human cord blood lymphocytes by the B95-8 isolate of EBV and cloned in soft agarose (34). The B95-8 cell line is negative for expression of antigens associated with early and late EBV replication. The conditions for cell growth, RNA extraction, polyadenylated RNA fractionation, RNA gel electrophoresis, and RNA blot analyses have been previously described (61).

Recombinant DNAs, DNA sequencing, and Southern blots. For initial S1 mapping studies, the entire *EcoRI* D_{het} fragment of the B95-8 EBV isolate was subcloned from pCB281, a plasmid derivative of lambda CB28 (11), into pKH47 (26) to yield the recombinant clone pVRF1. For more precise S1 mapping and for DNA sequencing, the *BamHI*-*BglII* fragment of EBV DNA (Fig. 1) was cloned from viral DNA into pUC9 (63) to create pSF1. This clone was then used to prepare a library of random subclones in M13 by sonication to 400 to 600 base pairs, repair with T4 DNA polymerase, and ligation into the *SmaI* site of M13mp10 (14). Defined *XhoI*, *Sau3A*, and *Avall* fragments of pCB281 from the *BamHI* site to the rightward *EcoRI* site (Fig. 1D) were subcloned into the *Sall*, *BamHI*, and *SmaI* sites, respectively, of M13mp10.

Nucleotide sequence was determined by M13 dideoxy-chain termination (39, 56, 57) with labeled dCTP (Amersham Corp.) and the Klenow fragment of DNA polymerase (Be-

* Corresponding author.

FIG. 1. S1 mapping. Single-stranded DNAs derived from pVRF1 were used in S1 protection experiments with IB4 cytoplasmic polyadenylated RNA (I) or yeast RNA (Y) as a control. "Uncut" DNAs in all panels are the *EcoRI* D_{het} to *EcoRI* I_J fragment (D). *Bam*HI and *Xho*I DNAs are the fragments extending from the *Bam*HI site or the leftward *Xho*I site through the TR to the rightward *EcoRI* site (D). Uncut and *Bam*HI-cut probes were hybridized to IB4 RNA, and the products of S1 digestion were run on a neutral agarose gel (A) or on a denaturing formaldehyde agarose gel (B). Uncut and partial *Xho*I-cut probes were hybridized to IB4 RNA, and the products of S1 digestion were run on a denaturing alkaline gel (C). D is a map of the EBV genome, with the IRs and U regions (U1 to U5) shown on top, and the *EcoRI* restriction map shown below. The DNA fragments used for hybridization to RNA and the protected regions (in black) are shown in E. All numbers represent sizes in kb.

thesda Research Laboratories, Inc.). Sequences were aligned and analyzed with the DATBAS programs (59) modified to run on a Dec-20 computer.

S1 and primer extension analyses. Radioactively labeled single-stranded DNAs for use as primers or S1 probes were synthesized from recombinant M13 clone templates by the "prime-cut" method (17). Single-strand recombinant M13 clone DNA was annealed to sequencing primer (17-mer; Collaborative Research, Inc.), and the complementary strand was synthesized with [³²P]dCTP and the Klenow fragment of DNA polymerase. The resultant double-stranded DNA was cut with a restriction enzyme which cut downstream from the desired sequences. The labeled single strand, having the 17-mer at its 5' end, was then purified on a denaturing polyacrylamide gel (55).

S1 nuclease analysis of IB4 polyadenylated RNA (3, 61) was done with single-strand probes from pVRF1 or from recombinant EBV M13 clones. RNA-protected DNA seg-

ments of pVRF1 were separated by electrophoresis in agarose gels, transferred to nitrocellulose, and hybridized to DNA probes as previously described (61). Labeled M13-derived probes were analyzed on denaturing polyacrylamide gels in parallel with a known DNA sequence. For primer extension, hybridization of primer with RNA followed a procedure described previously for S1 reactions (49). Cytoplasmic polyadenylated RNA (10 to 50 µg) in 20 µl of hybridization buffer (80% recrystallized formamide, 0.1% sodium dodecyl sulfate, 0.4 M NaCl, 0.01 M PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid)] [pH 6.4]) was mixed with labeled single-stranded DNA in 2 to 3 µl of 5 mM NaPO₄ (pH 7.5)–5 mM EDTA. The hybridization mixture was sealed in a capillary tube, heated at 85°C for 5 min, and then incubated at 50°C for 12 to 18 h. Five volumes of 0.3 M sodium acetate were added, and the nucleic acid was ethanol precipitated. The nucleic acid was suspended in 200 µl of 25 mM Tris-hydrochloride (pH 8.3)–50 mM NaCl–7.5 mM MgCl₂–2 mM dithiothreitol–500 µM each of dATP, dGTP, dCTP, and dTTP–750 U of RNasin (Biotech) per ml. Reverse transcriptase was added to 400 U/ml, and the reaction was incubated at 42°C for 1 h. The reaction was brought to 0.2 N NaOH–15 mM EDTA, incubated at 42°C for 1 h to destroy the RNA, phenol extracted, and ethanol precipitated. The labeled DNA was analyzed on 4 or 6% denaturing polyacrylamide gels (55) alongside sequencing ladders as size markers.

RESULTS

S1 analyses with pVRF1. The EBV *EcoRI* D_{het} fragment identifies a 2.8-kb RNA in blots of denatured IB4 cytoplasmic polyadenylated RNA (62). The recombinant plasmid pVRF1 (*EcoRI* D_{het} to *EcoRI* I_Jhet cloned into the strand separation vector pKH46) was used to determine the direction of transcription of this RNA. IB4 RNA hybridized to the EBV DNA strand with its 5' end to the left (Fig. 1) and protected no DNA of the opposite strand, indicating that the RNA was transcribed from right to left. The size of S1-treated pVRF1 DNA-IB4 RNA hybrids in neutral agarose gels was 2.5 kb, in reasonable agreement with the size expected for the 2.8-kb RNA without its polyadenylated tail. In alkaline gels, the S1 RNA-protected DNA segment was 2.0 kb. The data (Fig. 1) therefore indicate that the RNA is spliced and that one exon is 2 kb.

Comparison of the intensity of hybridization of the labeled probe to the RNA-protected DNA with that to known amounts of EBV DNA in parallel indicated that there was 9 to 13 ng of RNA per 50 µg of cytoplasmic polyadenylated RNA. Since 3.8×10^{-7} µg of cytoplasmic polyadenylated RNA was recovered per cell, there were 60 to 90 RNA molecules per cell.

pVRF1 DNA was cut with *Bam*HI to purify single-stranded DNA to the right of the *Bam*HI site (Fig. 1). *Bam*HI cutting did not affect the size of the RNA-protected DNA in alkaline, formaldehyde, or neutral gels (Fig. 1A and B), indicating that the RNA mapped entirely to the right of the *Bam*HI site.

The 2.0-kb exon was further mapped by S1 nuclease analysis, using DNA cut with restriction enzymes to the right of the *Bam*HI site. Incomplete digestion with *Xho*I yielded two fragments, each extending right from an *Xho*I site (Fig. 1E). When the *Xho*I fragments were used for S1 nuclease analysis of IB4 cytoplasmic polyadenylated RNA, a 1.5-kb RNA-complementary DNA fragment was detected (Fig. 1C). The simplest interpretation is that the 5' end of the

dideoxy-chain termination method. The *Bam*HI-*Bgl*II fragment containing the 2.0-kb exon (Fig. 1) was sequenced first. The sequence was then extended 1 kb from the *Bgl*II site toward TR (Fig. 1). The *Bam*HI-*Bgl*II fragment was sequenced from random clones. Sequencing beyond the *Bgl*II site was from *Ava*II, *Xho*I, and *Sau*3A fragment clones. All sequencing was confirmed with both strands. Within the

FIG. 3. S1. IB4 cytoplasmic polyadenylated RNA (I) was used in S1 experiments with labeled single-stranded DNA probes derived from M13 EBV DNA clones. The probes were from regions as depicted in the schematic, with the protected regions in black. The derived RNA map is shown above the DNA map. Lanes D and Y in each panel are the probes hybridized with the opposite strand of DNA (D) or with yeast RNA (Y) and treated the same as the IB4 RNA reactions (I). A and B are probes for the 3' end of the RNA. Fragments C through H are probes for the 5' and the 2-kb exon and for the S1 exons. All numbers are in b.

region of the 2-kb exon there were two unexpected results. First, there was a repeat element. The repeat is described by XYXYXZYXZ, where X = TGACAATGGCCCACA, Y = GGACCCTGACAACACTGA, and Z = TGACCCGCTGCCTCA. Second, the longest open reading frame was 0.8 kb, which started at nucleotide 491 shortly before a possible splice acceptor (41) and ended at nucleotide 1312, where there was a stop codon. The stop codon was 1.2-kb before the end of the message indicated by the S1 results in Fig. 1. To rule out the possibility that the unusual repeat array and long 3' untranslated region may have resulted from alteration during the cloning process, two other pUC9 clones of the *Bam*HI-*Bgl*II fragment of virus DNA were subcloned into M13mp9 and sequenced. The results through the repeats and stop codons were identical. Another important feature of the sequence is a typical polyadenylation signal, AAUAAA (49), near the 3' terminus of the 2-kb exon.

Mapping of the exons of the 2.8-kb RNA. The pVRF1 analysis indicated that the 2.8-kb RNA was transcribed right to left and was spliced, a 2-kb exon mapped as shown in Fig. 1E, and the rest of the RNA, ca. 0.5 kb, was encoded by DNA to the right of the *Bam*HI site. The identification of a polyadenylation consensus sequence in a region which should contain the 3' end of the 2-kb exon and a potential splice acceptor site near its 5' end further suggested that the 2-kb exon was the 3' end of the RNA and that the other

exons were 5' to the 2-kb exon. Labeled single-strand probes made from M13 clones used in sequencing were hybridized to RNA to confirm the 3' and 5' ends of the 2-kb exon in S1 protection experiments and to identify and map the putative 5' exons. The 3' end of the 2-kb exon mapped (see Fig. 3A and B) to nucleotide 2556 ± 5 of the sequence shown in Fig. 2, 20 nucleotides 3' to the polyadenylation signal identified in the nucleotide sequence. The 5' end of this exon mapped (Fig. 3C and D) to nucleotide 510 ± 5 . Nucleotide 510 was a potential splice acceptor site (Fig. 2). Two other exons were then identified 5' to the 2-kb exon (Fig. 3E to H); a 308-base (b) exon and an 86-b exon. The 308-b exon began at nucleotide -40 ± 3 and ended at nucleotide 268 ± 3 at a potential splice donor site (Fig. 3E to H). The 86-b exon was in a region of 175 b between the 307-b exon and the *Bgl*II site (Fig. 3E and F). Within the 175-b region there was only one splice acceptor site. It was followed 86 b later by a splice donor site and maintained an open reading frame across both splice junctions. The 86-b exon was therefore assigned to this position.

Primer extensions. To confirm that the 86- and 308-b exons are the entire set of exons at the 5' end of the RNA, primer extension experiments were carried out. The map position of the primers and the extension results are shown in Fig. 4. Primer A was derived from a random clone in the 2-kb exon. Reverse transcriptase extended this primer on an IB4 RNA

template to a final length of $675 \text{ b} \pm 10$. The length expected from the S1 data was 675 b. Primer B was a *Sau3A-XhoI* fragment, which was in the 307-b exon. This primer extended to give five bands. The largest, although not the darkest, gave the expected size of 271 b. The smaller bands are presumably due to premature stopping of the reverse transcriptase and do not agree with the S1 results or the other primer extension reaction. The primer extensions mapped the 5' end of the RNA to nucleotide -40, 40 b 5' to an AUG (Fig. 2) and confirmed the S1 results. A TATA box, TACATAA, lies 25 b and a CAAT box lies 85 b 5' to the RNA cap site (7).

Protein structure. The DNA sequence and the S1 mapping yielded an RNA sequence which started 40 b 5' to an ATG (at position 1 in Fig. 2). There was a long open reading frame which began with this codon for methionine, extended across two introns, and ended with a UAA at nucleotide 1312, for a total of 386 codons. UAA has not been reported to be suppressible to eucaryotic cells (20). The RNA extended for 1,200 b beyond this but contained multiple stops in all three frames. The protein encoded by the long open reading frame was 386 amino acids and would have a molecular weight of 42,000. The most notable feature of the protein was a long hydrophobic region extending through most of the amino half of the protein. The hydrophobicity profile (37) of the predicted sequence (Fig. 5A) indicated six very hydrophobic regions in the amino half of the protein, each long enough to span a membrane. The six hydrophobic regions are also predicted to be alpha-helical (8, 19). Five short domains which meet criteria for reverse turns bridge between each of the six hydrophobic, alpha-helical segments (8). A schematic of the possible organization with respect to the membrane is shown in Fig. 5B.

Other notable features of the protein are as follows. (i) The amino acids preceding the first hydrophobic region were highly charged, having four positive and three negative charges. (ii) This region was also proline rich, containing three Pro-Pro pairs. (iii) The carboxy half of the protein, after the six hydrophobic domains, was highly negatively charged, containing 6 glutamic and 37 aspartic acids and only 1 lysine and 2 arginines. (iv) The carboxy-terminal domain also contained 30 prolines and 35 glycines. Therefore, little helical or sheet character is predicted by Chou and Fasman (8) and Garnier et al. rules (19). (v) The repeat DNA sequence was part of the putative coding domain. Triplet symmetry was preserved through the repeat domain, creating a repeat peptide (Asp-Asn-Gly-Pro-Gln-Asp-Pro-Asp-Asn-Thr-Asp).

DISCUSSION

These experiments delineate the nucleotide sequence of a region of EBV U5 DNA and of the mRNA encoded by that region in latently infected growth-transformed cells. Two other mRNAs, an IR1-U2 RNA and a U3-IR3-U4 RNA, have been partially delineated in these cells (30, 31, 61). The transcriptional start for the EBV IR1-U2 RNA has been previously identified (61). Since the IR1-U2 and U5 RNAs are both transcribed in latently infected B lymphocytes, the 500 bp 5' to the RNA cap sites were compared to identify similar sequences which could be regulatory elements. There was no decanucleotide or longer homology at similar distances in the 500 bp 5' to the U5 and IR1-U2 cap sites, even allowing for 25% mismatch (50). In a broader search for homology between the two sequences, the sequence CCCGCCTCCCTACGGTTACCCACAGCCTTGCT-

CACC, which was 500 nucleotides 5' to the U5 cap site, was noted to be similar to the sequence CCCGCCTCCACCC-GAGCCCCCTCAGC-TTGCTCCCC, which was 1,000 nucleotides 5' to the IR1-U2 start site. Because of the large and different distances between these sequences and the cap sites, the possibility that these were enhancer elements was considered. However, no homology was noted to viral or immunoglobulin enhancers (22, 33). Comparison of the two EBV 5' sequences with eucaryotic viral and cellular 5' sequences in the November 1983 Genbank databank (66) also revealed no significant matches. An alternating purine-pyrimidine sequence, ACATGCACACACAC, which was -177 to -191 from the U5 RNA cap site, was near the mouse H-chain V region (9) and could form Z DNA (64).

Two features of the putative U5 messenger RNA were noted. One was the direct repeat near the 3' end of the open reading frame. Since the repeats were multiples of three nucleotides, they encoded similar repeating oligo peptides. A simpler, more highly reiterated triplet nucleotide repeat array, IR3, was part of the 2-kb exon of the U3 IR3-U4 RNA and was translated into a glycine-alanine copolymer domain of EBNA1 (27, 30, 31). Thus, nucleotide sequence amplifica-

FIG. 4. Primer extensions. Cytoplasmic polyadenylated IB4 RNA was used in primer extension reactions with primers A and B depicted in the schematic. Each primer consisted of M13 sequences, whose length is indicated on the upright, and an EBV-specific portion, whose length is shown on the horizontal. All numbers are in b.

FIG. 5. Protein encoded by the U5 RNA. (A) Hydrophobicity profile by the method of Kyte and Doolittle (37). Each amino acid has been assigned a value reflecting its relative hydrophobicity. The computer program continuously determines the average of a segment of seven amino acids as it advances through the sequence from amino to carboxyl terminus. Isoleucine has the highest hydrophobicity index, 4.5; arginine has the lowest index, -4.5. Relatively hydrophobic regions appear above the zero line. (B) Schematic of the proposed arrangement of the protein in the membrane based on its hydrophobicity profile. Plus and minus signs indicate charged residues.

tion may be a common strategy for amplification of EBV protein domains. Since repeat elements are expected to diverge in reiteration frequency, polymorphism of the protein encoded by the repeat can be expected and may be used to map the protein to a repeat DNA segment (30). A second remarkable aspect of the U5 messenger RNA was that the principal open reading frame terminated 1.2 kb before the end of the exon. No dyad symmetry was evident between the 3' 1.2-kb sequence and the rest of the RNA, which could suggest a role for the 3' sequence in forming secondary structures which might be important in processing or translation (50). There are two open reading frames (of 200 to 300 b each) within the 3' 1.2-kb region. ATGs near the beginning of both open reading frames could be starts for translation of small proteins from this RNA. One protein would begin with a methionine at nucleotide 1492 and end with a glutamine at nucleotide 1716 in the sequence shown in Fig. 2. The second

protein would begin with a methionine at nucleotide 2180 and end with a glycine at nucleotide 2474 in the sequence shown in Fig. 2. Of the 97 amino acids of the latter open reading frame, 20 are lysines or arginines. A start site for transcription of an RNA expressed in productive infection has been identified at nucleotide 1972, 200 nucleotides 5' to this open reading frame (16). The arginine-lysine-rich protein would therefore likely be expressed in productive infection. Its presence at the end of the U5 latent RNA could mean that it is rarely translated in latent infection as well. Differential expression of this region in latent versus productive infection is consistent with a regulatory role. An analogous situation pertains in adenovirus transformation in which the 3' end of the early region 1b message has no known function but is a sequence which encodes a protein expressed late in adenovirus replication (1).

The principal translation product of the EBV U5 RNA is

almost certainly the protein encoded by the 5' long open reading frame. This protein is expected to be a membrane protein because of its markedly hydrophobic domains. Although the amino terminal region does not resemble a leader sequence thought to be responsible for initiating vectorial transfer across membranes, the six helical, hydrophobic domains linked by reverse turns are consistent with an internal insertion sequence and spontaneous association of the nascent, or complete, protein with a membrane (5, 54, 58). Hydrophobic attractions between internal hydrophobic sequences such as these and the membrane lipids are thought to be strong enough to pull the short regions of turn through the membranes. Only a few proteins have been studied which seem to span the membrane multiple times: the calcium pump protein (36), bacteriorhodopsin (15), bovine rhodopsin (43), lactose permease (18), acetylcholine receptor (10, 44), and erythrocyte band 3 (5). These proteins have all been identified as transport proteins, suggesting that the similar structure in the EBV protein may indicate that it too has a transport function. Transformed cells are known to have altered ion permeability, and a protein such as this could conceivably be responsible for such a change.

Because there are an even number of transmembrane domains, the amino and carboxy termini of the putative EBV membrane protein are likely to be on the same side of the membrane. The cytoplasmic amino-terminal part of one of the multiple membrane-spanning proteins, erythrocyte band 3, is highly acidic and therefore similar to the carboxy-terminal part of the putative U5 protein. The acidic amino terminus of erythrocyte band 3 binds glycolytic enzymes (32). Another cytoplasmic domain binds ankyrin, which in turn binds to the cytoskeletal protein spectrin (6). The highly negatively charged part of the carboxy end of the EBV proteins could be involved in a similar association. *v-fgr* (42), a feline sarcoma virus oncogene, has recently been shown to be a fusion between actin and a protein kinase, suggesting that the interphase between membrane and cytoskeletal proteins may be important in regulation of cell growth.

Proteins encoded by growth-transforming viruses, including EBV, appear to fall into several categories. Some of the proteins, such as simian virus 40 T and t, polyomavirus large T, adenovirus early region 1a, avian myeloblastosis virus *myc* and *myb*, and EBV EBNA1 and EBNA2, are nuclear proteins which may regulate transcription or DNA synthesis. Other proteins, such as *src*, *abl*, *fes*, *fps*, *yes*, *frms*, *ras*, *sis*, *erb*, and *fms*, are membrane proteins (4, 65) which may be similar to growth factors or growth factor receptors. Some of the latter class of proteins, such as *src*, *abl*, *fes*, *fps*, *yes*, *frms*, and *ras*, are phosphoprotein kinases, whereas others, such as *fms*, appears to lack such activity. None of the previously described membrane oncogene products resembles the putative EBV U5 proteins, and a computer search of the protein sequence data bank with the algorithms of Wilbur and Lipman does not reveal any homology to oncogenes. Specific comparison with the conserved regions of *src*, *abl*, *fes*, *fps*, *yes*, and *fms* (24, 25) also reveals no similar regions in the putative EBV U5 protein.

The only new antigen detected in EBV-transformed cells other than the EBNA is LYDMA (60). LYDMA can only be detected on the surface of EBV-transformed cells by EBV-immune T lymphocytes (40, 53, 60). EBV-immune T lymphocytes specifically kill EBV-transformed cells. No human antisera have been identified which recognize LYDMA in immunoblots or in complement-mediated killing of EBV-transformed cells. The putative protein described in this

paper may be LYDMA, since it is expressed in latent infection, it is a membrane protein, and there is little extension of the protein to the outside of the cell which might otherwise be expected to induce and bind antibodies. Furthermore, this is the only latent EBV gene product found to date that is a likely membrane protein. To identify the protein in IB4 cells, the DNA encoding the carboxy-terminal part of the protein has been inserted into the beta-galactosidase gene. Antisera, raised in rabbits, to the fusion protein will be used to attempt to identify LYDMA in immunoblots of IB4 cell membrane proteins. The immune serum may also permit us to confirm the orientation of the protein in the cell plasma membrane.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants CA 19264 and CA 17281 from the National Institutes of Health and by American Cancer Society grant ACS MV321. S.F. and V.v.S. were supported by National Research Service awards AI 07182 and CA 09267, respectively, from the Public Health Service. E.K. is the recipient of a Faculty Research award from the American Cancer Society.

The sequence analysis programs were modified by K. Isono. Larven Chanmankit and Lisa Petti provided expert assistance with DNA sequencing. Terri Cole with tissue culture, and Patricia Morrison with manuscript preparation. We thank Ted Steck, Donald Steiner, and Kurt Drickhamer for helpful discussion of membrane proteins, and David Lipman and W. Wilbur for comparison of the nucleotide and amino acid sequences with the appropriate databanks.

LITERATURE CITED

1. Alestrom, P., G. Akusjarvi, M. Perricaudet, M. B. Mathews, D. F. Klessig, and U. Pettersson. 1980. The gene for polypeptide IX from adenovirus type 2 and its unspliced messenger RNA. *Cell* **19**:671-681.
2. Arrand, J. R., J. E. Walsh-Arrand, and L. Rymo. 1983. Cytoplasmic RNA from normal and malignant human cells shows homology to the DNAs of Epstein-Barr virus and human adenoviruses. *EMBO J.* **2**:1673-1683.
3. Berk, A., and P. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**:721-732.
4. Bishop, M. J. 1983. Cellular oncogenes and retroviruses. *Annu. Rev. Biochem.* **52**:301-354.
5. Braell, W. A., and H. F. Lodish. 1982. The erythrocyte anion transport protein is cotranslationally inserted into microsomes. *Cell* **28**:23-31.
6. Branton, D., C. M. Cohen, and J. Tyler. 1981. Interaction of cytoskeletal proteins on the human erythrocyte membrane. *Cell* **24**:24-32.
7. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**:349-383.
8. Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**:251-276.
9. Cohen, J. B., K. Effron, G. Rechavi, Y. Ben Neriah, R. Zakut, and D. Givol. 1982. Simple DNA sequences in homologous flanking regions near immunoglobulin Vh genes: a role in gene interaction? *Nucleic Acids Res.* **10**:3353-3370.
10. Conti-Tronconi, B. M. 1982. The nicotinic cholinergic receptor: correlation of molecular structure with functional properties. *Annu. Rev. Biochem.* **51**:491-530.
11. Dambaugh, T., C. Beisel, M. Hummel, W. King, S. Fennwald, A. Cheung, M. Heller, N. Raab-Traub, and E. Kieff. 1980. Epstein-Barr virus (B95-8) DNA. VII. Molecular cloning and detailed mapping. *Proc. Natl. Acad. Sci. U.S.A.* **77**:2999-3003.
12. Dambaugh, T., E. Kieff, F. K. Nkrumah, and R. J. Biggar. 1979. Epstein-Barr virus RNA in Burkitt tumor tissue. *Cell* **16**:313-322.
13. Dambaugh, T. R., and E. Kieff. 1982. Identification and nucleotide sequences of two similar tandem direct repeats in Epstein-

- Barr virus DNA. *J. Virol.* **44**:823-833.
14. Deininger, P. L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* **129**:216-223.
15. Engelman, D. M., R. Henderson, A. D. McLachlan, and B. A. Wallace. 1980. Path of the polypeptide in bacteriorhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **77**:2023-2027.
16. Farrell, P., A. Bankier, C. Sequin, P. Deininger, and B. Barrell. 1983. Latent and lytic cycle promoters of Epstein-Barr virus. *EMBO J.* **2**:1331-1338.
17. Farrell, P. J., P. L. Deininger, A. Bankier, and B. Barrell. 1983. Homologous upstream sequences near Epstein-Barr virus promoters. *Proc. Natl. Acad. Sci. U.S.A.* **80**:1565-1569.
18. Foster, D. L., M. Boublik, and H. R. Kaback. 1983. Structure of the *lac* carrier protein of *Escherichia coli*. *J. Biol. Chem.* **258**:31-34.
19. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**:97-120.
20. Geller, A., and A. Rich. 1980. A UGA termination suppression tRNA Trp active in rabbit reticulocytes. *Nature (London)* **283**:41-46.
21. Gerber, P., J. Whang-Peng, and J. H. Monroe. 1969. Transformation and chromosome changes induced by Epstein-Barr virus in normal human leukocyte cultures. *Proc. Natl. Acad. Sci. U.S.A.* **63**:740-747.
22. Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**:717-728.
23. Given, D., and E. Kieff. 1978. DNA of Epstein-Barr virus. IV. Linkage map of restriction enzyme fragments of the B95-8 and W91 strains of Epstein-Barr virus. *J. Virol.* **28**:524-542.
24. Groften, J., N. Heisterkamp, F. H. Reynolds, Jr., and J. R. Stephenson. 1983. Homology between phosphotyrosine acceptor site of human c-abl and viral oncogene products. *Nature (London)* **304**:167-169.
25. Hampe, A., M. Gobet, C. J. Sherr, and F. Galibert. 1984. Nucleotide sequence of the feline retroviral oncogene v-fms shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **81**:85-89.
26. Hayashi, S. 1980. A cloning vehicle suitable for strand separation. *Gene* **11**:109-115.
27. Heller, M., V. van Santen, and E. Kieff. 1982. Simple repeat sequence in Epstein-Barr virus DNA is transcribed in latent and productive infections. *J. Virol.* **44**:311-320.
28. Henderson, A., S. Ripley, M. Heller, and E. Kieff. 1983. Chromosome site for Epstein-Barr virus DNA in a Burkitt tumor cell line and in lymphocytes growth transformed in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **80**:1987-1991.
29. Henle, W., V. Diehl, G. Kohn, H. zur Hausen, and G. Henle. 1967. Herpes-type virus and chromosome marker in normal lymphocytes after growth with irradiated Burkitt cells. *Science* **157**:1064-1065.
30. Hennessy, K., M. Heller, V. van Santen, and E. Kieff. 1983. A simple repeat array in Epstein-Barr virus DNA encodes part of EBNA. *Science* **220**:1396-1398.
31. Hennessy, K., and E. Kieff. 1983. One of two Epstein-Barr virus nuclear antigens contains a glycine-alanine copolymer domain. *Proc. Natl. Acad. Sci. U.S.A.* **80**:5665-5669.
32. Kaul, R. K., S. N. P. Murthy, A. G. Reddy, T. L. Steck, and H. Kohler. 1983. Amino acid sequence of the N-terminal 201 residues of human erythrocyte membrane band 3. *J. Biol. Chem.* **258**:7981-7990.
33. Khoury, G., and P. Gruss. 1983. Enhancer elements. *Cell* **33**:313-314.
34. King, W., A. L. T. Powell, N. Raab-Traub, M. Hawke, and E. Kieff. 1980. Epstein-Barr virus RNA. V. Viral RNA in a restrictingly infected, growth-transformed cell line. *J. Virol.* **36**:506-518.
35. King, W., V. van Santen, and E. Kieff. 1981. Epstein-Barr virus RNA. VI. Viral RNA in restrictingly and abortively infected Raji cells. *J. Virol.* **38**:649-660.
36. Klip, A., R. A. F. Reithmeier, and D. H. MacLennan. 1980. Alignment of the major tryptic fragments of the adenosine triphosphatase from sarcoplasmic reticulum. *J. Biol. Chem.* **255**:6562-6568.
37. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105-132.
38. Lindahl, T., A. Adams, G. Bjursell, G. W. Bornkamm, C. Kaschka-Dierich, and U. Jehn. 1976. Covalently closed circular duplex DNA of Epstein-Barr virus in human lymphoid cell line. *J. Mol. Biol.* **102**:511-530.
39. Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**:309-321.
40. Misko, J. S., D. J. Moss, and J. H. Pope. 1980. HLA antigen-related restriction of T lymphocyte cytotoxicity to Epstein-Barr virus. *Proc. Natl. Acad. Sci. U.S.A.* **77**:4247-4250.
41. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
42. Naharro, G., K. C. Robbins, and E. P. Reddy. 1984. Gene product of v-fgr onc: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. *Science* **223**:63-66.
43. Nathans, J., and D. S. Hogness. 1983. Isolation, sequence analysis and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* **34**:807-814.
44. Noda, M., H. Takahashi, T. Tanabe, M. Toyosato, S. Kikuyotani, Y. Furutani, T. Hirose, H. Takashima, S. Inayama, T. Miyata, and S. Numa. 1983. Structural homology of *Torpedo californica* acetylcholine receptor subunits. *Nature (London)* **302**:528-532.
45. Nonoyama, M., and J. S. Pagano. 1972. Separation of Epstein-Barr virus DNA from large chromosomal DNA in non-virus producing cells. *Nature (London) New Biol.* **238**:169-711.
46. Pope, J. H., M. K. Horne, and W. Scott. 1968. Transformation of fetal human leukocytes in vitro in filtrates of a human leukemic cell line containing herpes-like virus. *Int. J. Cancer* **3**:857-866.
47. Powell, A. L. T., W. King, and E. Kieff. 1979. Epstein-Barr virus-specific RNA. III. Mapping of DNA encoding viral RNA in restrictive infection. *J. Virol.* **29**:261-274.
48. Pritchett, R. F., S. D. Hayward, and E. D. Kieff. 1975. DNA of Epstein-Barr virus. I. Comparative studies of the DNA of Epstein-Barr virus from HR-1 and B95-8 cells: size, structure, and relatedness. *J. Virol.* **15**:556-569.
49. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature (London)* **263**:211-214.
50. Queen, C., and L. Korn. 1980. Computer analysis of nucleic acids and proteins. *Methods Enzymol.* **65**:595-609.
51. Raab-Traub, N., T. Dambaugh, and E. Kieff. 1980. DNA of Epstein-Barr virus. VIII. B95-8, the previous prototype, is an unusual deletion derivative. *Cell* **22**:257-267.
52. Reedman, B. M., and G. Klein. 1973. Cellular localization of an Epstein-Barr virus (EBV)-associated complement-fixing antigen in producer and non-producer lymphoblastoid cell lines. *Int. J. Cancer* **11**:499-520.
53. Rickinson, A., L. Wallace, and M. A. Epstein. 1980. HLA restricted T-cell recognition of Epstein-Barr virus infected B cells. *Nature (London)* **283**:865-867.
54. Sabatini, D. D., G. Kreibach, T. Morimoto, and M. Adesnik. 1982. Mechanism for the incorporation of proteins in membranes and organelles. *J. Cell Biol.* **92**:1-22.
55. Sanger, F., and A. R. Coulson. 1978. The use of thin acrylamide gels for DNA sequencing. *FEBS Lett.* **87**:107-110.
56. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.
57. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5463-5467.
58. Silhavy, T. J., S. A. Benson, and S. D. Emr. 1983. Mechanisms of protein localization. *Microbiol. Rev.* **47**:313-344.
59. Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing.

- ing. *Nucleic Acids Res.* **10**:4731–4751.
60. **Svedmyr, E., and M. Jondal.** 1975. Cytotoxic effector cells specific for B cell lines transformed by Epstein-Barr virus are present in patients with infectious mononucleosis. *Proc. Natl. Acad. Sci. U.S.A.* **72**:1622–1626.
61. **van Santen, V., A. Cheung, M. Hummel, and E. Kieff.** 1983. RNA encoded by the IR1-U2 region of Epstein-Barr virus DNA in latently infected, growth-transformed cells. *J. Virol.* **46**:424–433.
62. **van Santen, V., A. Cheung, and E. Kieff.** 1981. Epstein-Barr virus RNA. VII. Size and direction of transcription of virus specified cytoplasmic RNA in a cell line transformed by EBV. *Proc. Natl. Acad. Sci. U.S.A.* **78**:1930–1934.
63. **Vieira, J., and J. Messing.** 1982. The pUC9 plasmids, an M13mp7-derived system for insertion, mutagenesis and sequencing with synthetic universal primers. *Gene* **19**:259–268.
64. **Wang, A. H.-J., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich.** 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature (London)* **282**:680–686.
65. **Waterfield, M. D., G. T. Scrace, N. Whittle, P. Stroobant, A. Johnson, A. Wasteson, B. Westermark, C.-H. Heldin, J. S. Huang, and T. Deuel.** 1983. Platelet-derived growth factor is structurally related to the putative transforming protein p28 sis of simian sarcoma virus. *Nature (London)* **304**:35–39.
66. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. U.S.A.* **80**:726–730.