JOURNAL OF VIROLOGY, Dec. 1986, p. 1134–1140
0022-538X/86/121134-06$02.00/0

Vol. 60, No. 3

Transcription Initiation Sites and Nucleotide Sequence of a Herpes Simplex Virus 1 Gene Conserved in the Epstein-Barr Virus Genome and Reported To Affect the Transport of Viral Glycoproteins

PHILIP E. PELLETT, FRANK J. JENKINS, MATHIAS ACKERMANN, MARION SARMIENTO, AND BERNARD ROIZMAN*

The Marjorie B. Kovler Viral Oncology Laboratories, The University of Chicago, Chicago, Illinois 60637

Received 28 April 1986/Accepted 12 August 1986

Earlier reports have localized mutations which affect the processing and transport of herpes simplex virus 1 glycoproteins to a region located between the genes specifying glycoprotein B and the major viral DNA-binding protein (β8). The nucleotide sequence of this region contains a single long open reading frame encoding a 780-amino-acid protein with a predicted molecular weight of 83,845. To confirm the existence of this protein, rabbit polyclonal antibody was made against a synthetic peptide made according to the predicted sequence of a hydrophilic domain near the carboxy terminal of the protein. This antibody reacted with an infected cell protein of an apparent molecular weight of 95,500. We designated this protein infected cell protein 18.5 (ICP18.5). S1 nuclease analysis suggested that the 5.6-kilobase mRNA encoding ICP18.5 is initiated predominantly from one site, but three weaker initiation sites also seemed to occur within a 74-base-pair stretch of DNA. This gene appears to be conserved in the Epstein-Barr virus (EBV) genome, inasmuch as 174 of the 780 amino acids of ICP18.5 align with corresponding amino acids predicted by the EBV open reading frame BALF3. The EBV gene is located adjacent to the gene specifying a homolog of the herpes simplex virus 1 glycoprotein B.

Analyses of the gene directing the synthesis of the glycoprotein B (gB) of herpes simplex virus 1 (HSV-1) revealed that the 5′ terminus of the gene overlaps the 3′ terminus of an open reading frame. A portion of the open reading frame was sequenced at that time (4, 27). In this paper we report on the complete nucleotide sequence of this gene. Pending analyses of the regulation of this gene, we designated its product infected cell protein 18.5 (ICP18.5).

Interest in the gene specifying ICP18.5 stems from unresolved questions regarding the gene organization in the HSV-1 genome, specifically the significance and regulatory interdependence of clusters of genes found throughout the genome that (i) may be functionally related but apparently do not share *cis*-acting sites such as transcription termination and polyadenylation signals (e.g., the genes specifying gD and gE [30] and the genes specifying the major DNA-binding protein and DNA polymerase [29], (ii) genes that share *cis*-acting sites but do not appear to be functionally related (e.g., genes specifying a 42,000-molecular-weight protein and the α *trans*-inducing factor (5, 14), and (iii) genes that both share *cis*-acting sites and have clear functional relatedness. One cluster that fullfils the last characteristic is that specifying the subunits of the ribonucleotide reductase (8, 11, 24). The genes specifying gB and ICP18.5 may also fulfill this characteristic. Specifically, these genes specify mRNAs that are 3′ coterminal (15). gB is an essential membrane protein; virions produced in cells infected with temperature-sensitive (*ts*) mutants in this gene lack the capacity to be transported to the cell surface or to infect cells (23, 31). Analyses of HSV-1 and HSV-2 *ts* mutants (7, 13, 15, 25) carrying lesions mapping in the domain of the ICP18.5 gene, i.e., between the transcribed sequences of the gB gene and the coding sequences of the major DNA-binding protein designated β8 or ICP8, suggest that the product of this gene affects translocation of HSV glycoproteins to membranes. One mutant, HSV-1(KOS1.1) *ts*8 (13), shows wild-type patterns and levels of protein synthesis, glycoprotein synthesis, cell surface localization of glycoproteins, and sensitivity to immune cytolysis. Another mutant, HSV-1(KOS)icr *ts*78 (25), also exhibits wild-type patterns and levels of protein and glycoprotein synthesis and intracellular accumulation, but the amounts of glycoprotein transported to the cell surface are at greatly reduced levels. Moreover, cells infected with HSV-1(KOS)icr *ts*78 and maintained at the nonpermissive temperature are resistant to immune cytolysis by polyclonal anti-gB antiserum. The molecular basis of the *ts* phenotype is not known. All the mutants mapping in this region and characterized to date appear to produce viral DNA in quantities equivalent to those produced by the wild-type parent.

Holland et al. (15) mapped a 5.6-kilobase mRNA to a position overlapping the gB gene but 3′ to the HSV-1 β8 gene. Earlier reports from this (27) and another laboratory (4) identified an open reading frame extending 5′ from the gB gene. The coding sequence for the putative protein product of this mRNA spanned the upstream promoter and regulatory region of gB and terminated 10 nucleotides before the translation initiation site of gB. The 3′ terminus of this mRNA appears to be coterminal with the mRNA directing the synthesis of gB (15). To precisely define the 5′ end of the ICP18.5 mRNA transcript, S1 analysis was done with the recombinant plasmid pRB2041, which contains a 550-base-pair (bp) DNA fragment from HSV-1 *Bam*HI-G cloned into pUC9 (Fig. 1E). pRB2041 was digested at a *Nar*I restriction endonuclease site and was 5′ end labeled. Infected cell RNA was isolated as previously described (28) from confluent Vero cell monolayers infected at a multiplicity of infection of 10 to 20 with HSV-1(F) (10). S1 analyses of cellular RNA harvested at 14 h postinfection protected one major and three minor DNA fragments whose sizes were 192, 157, 151,

* Corresponding author.

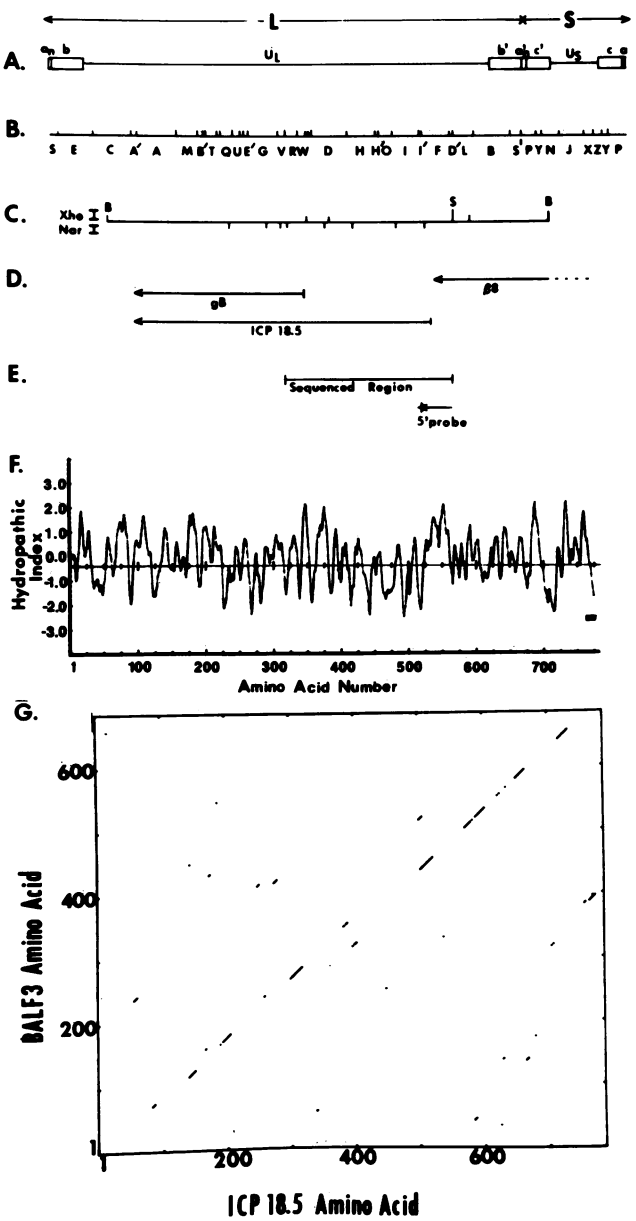

FIG. 1. (A) Sequence arrangement in HSV-1 DNA showing the location of the unique sequences of the L and S components (U_L and U_S) and of the inverted repeats (a_nb, b'a'_nc', and ca) flanking the unique sequences. The subscript n indicates the number of repeats (>1) of the a sequence. (B) *Bam*HI restriction endonuclease map of the prototype arrangement of HSV-1(F) DNA. (C) Relevant restriction endonuclease maps of the *Bam*HI G fragment. S, *Sal*I; B, *Bam*HI. (D) The domains of gB, ICP18.5, and β8. (E) The position of the sequenced region and the DNA probe used for mapping the 5′ end of the ICP18.5 mRNA. Panels D and E are shown in relation to the restriction endonuclease maps in panel C. (F) Hydropathic analysis of HSV-1 ICP18.5. The hydropathic profile was obtained by using the algorithim of Kyte and Doolittle (21) with a moving window of seven residues. The subsequent profile was smoothed for plotting by taking its average in a moving three-residue-wide window. The x axis of the plot is drawn at the average hydropathicity value deducted by Kyte and Doolittle so that points above the line are of above-average hydrophobicity and those below the line are of above-average hydrophilicity. ▬, location of the sequence used for preparation of the synthetic polypeptide. (G) Dot homology matrix plot at the protein level between HSV-1 ICP18.5 and EBV BALF3. ICP18.5 is represented on the horizontal axis and BALF3 is represented on the vertical axis. A dot is plotted if a score of evolutionary relatedness (32) summed over a 19-amino-acid window is equal to or greater than 222. Details of the procedure have been described previously (26).

and 118 bp, respectively (Fig. 2A), indicating that transcription of the ICP18.5 gene may initiate at several sites. The intensity of the protected fragments suggests that most of the mRNA protected the 192-bp fragment, i.e., transcription initiated 3′ to this site occurred less frequently.

The nucleotide sequence of the ICP18.5 gene was determined by previously described procedures (27). A 6-kilobase-pair *Sal*I-*Bam*HI subclone of the *Bam*HI G fragment (pRB2017) was subjected to a DNase I deletion subcloning strategy as previously described (12, 17). One set of derived clones contained staggered deletions from the *Sal*I site toward the *Bam*HI terminus, whereas the second set contained staggered deletions from the *Bam*HI terminus toward the *Sal*I site. The DNA sequence of the terminal 200 to 300 nucleotides from each DNase I deletion subclone as well as other clones obtained by conventional cloning techniques was determined by the dideoxy sequencing method. The DNA sequence of a region of 2,889 nucleotides was obtained, beginning with the *Sal*I site located 80 amino acids from the carboxy terminus of the β8 gene and ending 7 nucleotides from the site of translation initiation of gB (Fig. 3). The sequence was determined in its entirety on both strands. There is an overlap of 787 nucleotides with the previously reported sequence of the genome domain 5′ to the coding sequences of the gB gene (27). Comparison of sequences of new clones as well as resequencing of old clones indicated that the nucleotides reported to exist at positions 81, 87, 89, and 101 of the previously reported sequence (i.e., upstream from the transcribed domains of gB) (27) do not exist. The sequence reported here overlaps with that of the HSV-1(17) DNA polymerase and β8 genes (26) by 444 nucleotides. Our sequence of the carboxy terminal 80 amino acids of the predicted structure of the β8 protein of HSV-1(F) is in complete agreement with that of the reported sequence of the HSV-1(17) protein (29).

The major ICP18.5 mRNA start site determined by S1 analysis (Fig. 2A) and located at nucleotide 347 (Fig. 3) is 21 nucleotides 3′ to the sequence TTGTTT, which could serve as a TATA box (3). The next most abundant mRNA species has its 5′ end located at nucleotide 421, 34 nucleotides 3′ of the sequence ATAGAA, which could also serve as a TATA box. It is not immediately obvious which sequence serves the TATA function for the two least abundant mRNA termini which are located at nucleotides 382 and 388. Sequences with single-nucleotide deviations from the consensus SP1 transcription-factor-binding sequence, $\frac{G}{T}GGGCGG\frac{G}{A}\frac{G}{A}\frac{C}{T}$ (16), were found beginning at nucleotides 268, 273, and 346. The first translation initiation codon beyond the transcriptional start sites was found beginning at nucleotide 547. The environment of the start codon qualifies as one of moderate efficiency (18–20) and should yield the predicted 5′-transcribed noncoding regions of 126, 159, 165, and 200 bp for the four predicted mRNA start sites at nucleotides 347, 382, 388, and 421, respectively. The first in-frame stop codon was found at position 2887, and its first nucleotide was located 10 nucleotides 5′ to the translation initiation codon of gB. The predicted translated molecular weight of the 780-amino-acid polypeptide encoded by the open reading frame is 83,845. Codon usage patterns in this reading frame were consistent with those seen in other HSV genes (data not

shown). For the 2,889 nucleotides in the strand of DNA shown in Fig. 3, 15.0% are adenine, 33.4% are cytosine, 35.6% are guanine, and 16.0% are thymine. Of the 780 amino acids predicted in ICP18.5, the most abundant is alanine, with 104 residues, and the least abundant are lysine and tryptophan, each with 7 residues. The proline residues account for 6.7% of the total amino acids. The acidic residues, glutamic acid and aspartic acid, account for 11.8% of the amino acid residues, whereas the positively charged residues, arginine and lysine, account for 10% of the amino acids, giving the ICP18.5 protein a net negative charge. The average hydropathicity per residue determined by using the hydropathic values of Kyte and Doolittle (21), is −0.13, i.e., a more hydrophobic average than the value of −0.4 determined to be the grand average hydropathicity of a large group of proteins analyzed by Kyte and Doolittle.

The hydropathic profile of ICP18.5 determined by the relative hydropathy as a function of position along the amino acid sequence (21) revealed no regions remarkable for their hydrophilicity (Fig. 1F). Several relatively hydrophobic stretches, e.g., residues 360 to 380 and 525 to 560, were suggestive of membrane-spanning segments. Given the possible role of ICP18.5 in processing and transport of viral glycoproteins to the cell surface, it seemed plausible that ICP18.5 might be a membrane-spanning protein itself, or even a glycoprotein, since three potential N-linked glycosylation sites (Asn-X-Ser or Thr, X, is not Pro) were found in the sequence. Eisenberg (9) as well as Kyte and Doolittle (21) have established guidelines for predicting whether proteins may be membrane associated by virtue of possessing one or more membrane-spanning segments. Analysis of the ICP18.5 amino acid sequence by either of these two methods failed to detect any membrane-spanning segments. In addition, when the Epstein-Barr virus (EBV) homolog of ICP18.5 (see below) was analyzed in a similar fashion, no potential N-linked glycosylation sites were found and no membrane-spanning segments could be predicted.

The hydrophilicity analysis (Fig. 1F) suggests that a hydrophilic domain with the predicted amino acid sequence of Leu-Ala-Pro-Arg-Leu-Pro-Asp-Gly-Gly-His-Asp-Gly-Pro-Pro which corresponds to amino acids 767 to 780 may be suitable for raising polyclonal antibody reactive with this protein. The rabbit antiserum made against this synthetic peptide reacted with an infected cell polypeptide which was electrophoretically separated in denaturing polyacrylamide gels and electrically transferred to nitrocellulose sheets (Fig. 2B). The observation that the apparent molecular weight of this protein (95,500) was higher than the molecular weight predicted from the nucleotide sequence of the gene (83,845) was not unexpected. All HSV proteins analyzed to date appear to be relatively rich in proline and exhibit an apparent molecular weight higher than that predicted for the coding sequences of their genes (6, 27, 28). Since this polypeptide migrated between HSV-1 ICP18 and ICP19, we designated it ICP18.5.

The published DNA sequences of HSV-1 gB (4, 27) contain the region encoding the carboxyl terminus of ICP18.5. As observed independently in several laboratories (26; B. Barrell, personal communication; this report), the carboxyl terminus of ICP18.5 displays significant sequence homology with the BALF3 reading frame (2) of EBV located adjacent to the BALF4 reading frame encoding a homolog of gB. In the dot homology matrix shown in Fig. 1G, where perfect identity between the sequences being compared would appear as a solid diagonal line extending from the lower left corner and extending to the upper right corner, a

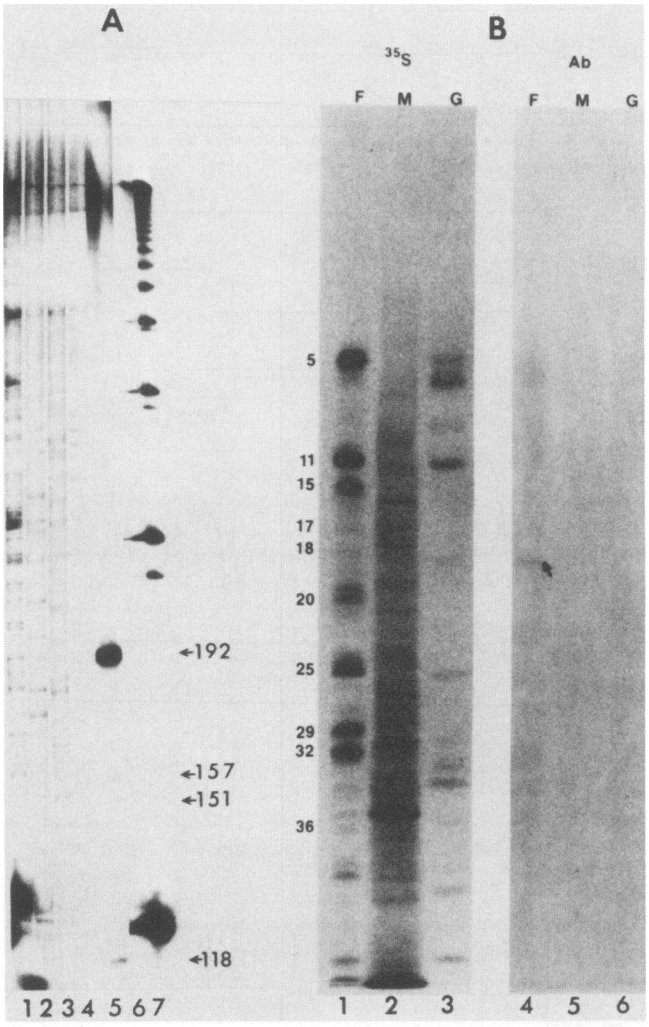

FIG. 2. (A) Autoradiographic images of the labeled DNA probe fragments protected from digestion by S1 nuclease. The 5′-end-labeled probe DNA molecules protected from digestion with S1 nuclease by hybridization with cytoplasmic mRNA from infected cells were electrophoretically separated in denaturing 8% polyacrylamide gels with molecular size markers. The sizes of the protected fragments are indicated in nucleotide numbers to the right of lane 7. Lanes: 1 to 4, DNA sequencing ladder generated as described in the text; 5, 5′-probe fragments protected by 12 to 14 h postinfection total cytoplasmic RNA; 6 and 7, end-labeled 123-nucleotide ladder. (B) Autoradiographic images and immune reactivity of polypeptides from infected and mock-infected HEp-2 cells electrophoretically separated in denaturing gels and then electrically transferred to nitrocellulose sheets. Techniques for the synthesis of the peptide, the raising of antipeptide antiserum in rabbits, and the subsequent use of the antipeptide antiserum to detect an infected cell polypeptide were as previously described (1). Replicate cultures of HEp-2 cells were infected with HSV-1(F) (F) or HSV-2(G) (G) or were mock infected (M) and then labeled with [^{35}S]methionine (^{35}S) from 8 to 24 h postinfection. Cell lysates were subjected to electrophoresis in denaturing polyacrylamide gels and transferred electrically to nitrocellulose sheets. Numbers to the left indicate infected cell protein designations. Lanes: 1 to 3, autoradiographic images of electrophoretically separated polypeptides in lysates of infected and mock-infected cells; 4 to 6, reaction of rabbit polyclonal antiserum (Ab) made against a synthetic polypeptide made according to the predicted sequence of the last 14 amino acids at the carboxy terminus of ICP18.5 with a polypeptide (→) which migrates in denaturing polyacrylamide gel between ICP18 and ICP19 of HSV-1(F).

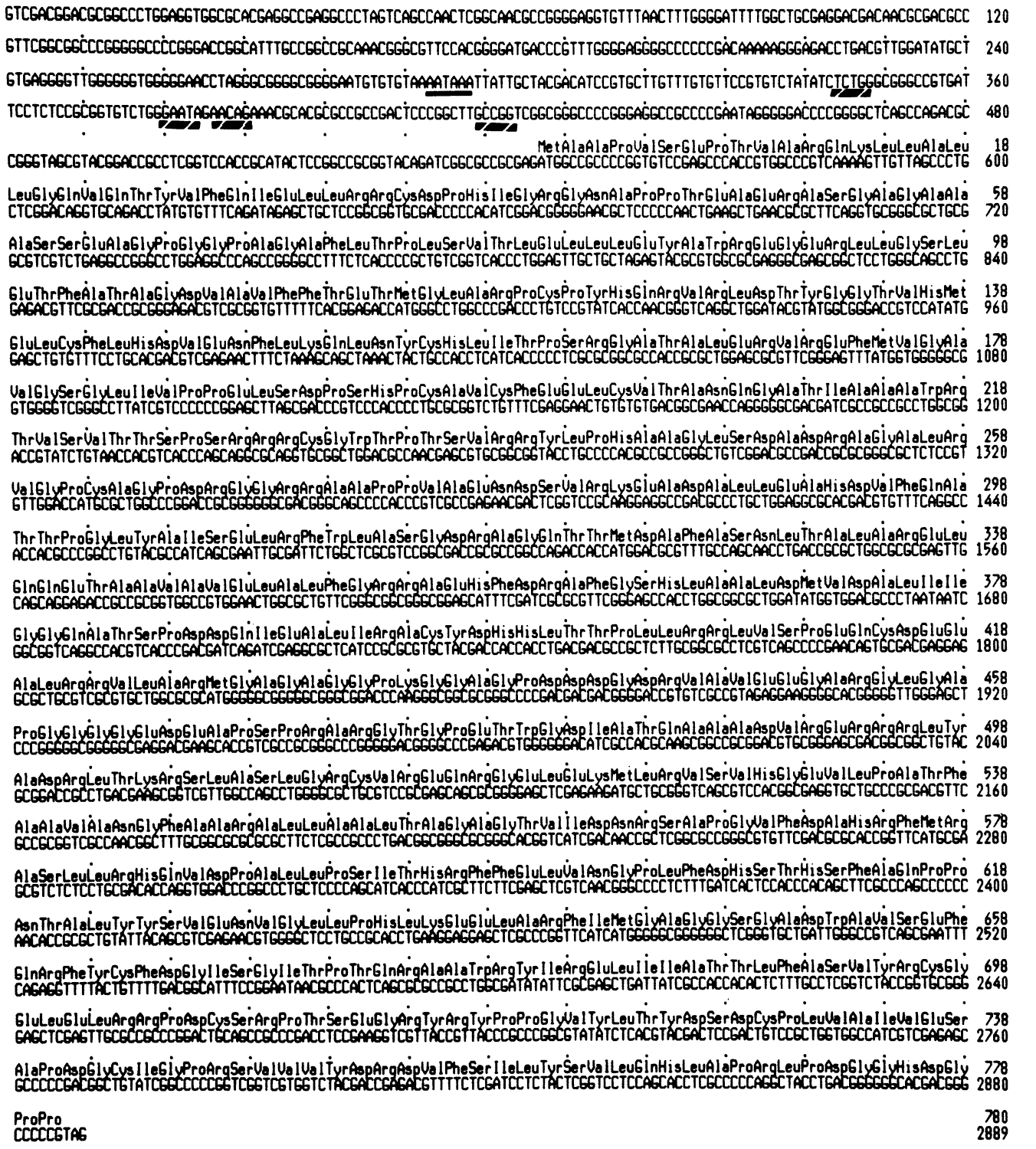

FIG. 3. Nucleotide and predicted amino acid sequences of the domain of the ICP18.5 gene. ▬, polyadenylation site for ICP8 gene; ◤▬◢, transcription initiation sites of ICP18.5 mRNA indicated by the S1 analyses.

less than perfect but nonetheless well above background diagonal of homology extends in patches for almost the entire length of the molecules. The level of relatedness seen between BALF3 and ICP18.5 is less than that seen between the adjacent EBV reading frame BALF4 and HSV-1 gB (26), where a stronger diagonal with less background detail was seen by using a minimum relatedness value of 235 for plotting, compared with the value of 222 used in this study. The amino acid sequences of ICP18.5 and BALF3 were aligned to demonstrate the areas of identity. In the alignment shown in Fig. 4, ICP18.5 residues 76 to 95, 132 to 221, and 487 to 732 were aligned by the alignment optimization

```
  1 MetAlaAlaProValSerGluProThrValAlaArgGlnLysLeuLeuAlaLeuLeuGlyGlnValGlnThrTyrValPheGlnIleGluLeuLeuArgArgCysAspProHisIleGly
  1 MetSerGlyLeuLeuAlaAlaAlaTyrSerGlnValTyrAlaLeuAlaValGluLeu---SerValCysThrArgLeuAspProArgSerLeu---------AspValAlaAlaValVal

 41 ArgGlyAsnAlaProProThrGluAlaGluArgAlaSerGlyAlaGlyAlaAlaAlaSerSerGluAlaGlyProGlyGlyProAlaGlyAlaPheLeuThrProLeuSerValThrLeu
 37 ArgAsnAlaGlyLeuLeuAlaGluLeuGlu----------------------AlaIleLeuLeuProArgLeuArgArgGlnAsnAspArgAlaCysSerAlaLeuSerLeuGluLeu

 81 GluLeuLeuLeuGluTyrAlaTrpArgGluGlyGluArgLeuLeuGlySerLeuGluThrPheAlaThrAlaGlyAspValAlaValPhePheThrGluThrMetGlyLeuAlaArgPro
 70 ValHisLeuLeuGluAsnSerArgGluAlaSerAlaAlaLeuLeu----------------------------------AlaProGlyArgLysGlyThrArgValProProLeuArg

121 CysProTyrHisGlnArgValArgLeuAspThrTyrGlyGly---ThrValHisMetGluLeuCysPheLeuHisAspValGluAsnPheLeuLysGlnLeuAsn---------TyrCys
 97 ThrPro---SerValAlaTyrSerValGluPheTyrGlyGlyHisLysValAspValSerLeuCysLeuIleAsnAspIleGluIleLeuMetLysArgIleAsnSerValPheTyrCys

157 HisLeuIleThrProSerArgGlyAlaThrAlaLeuGluArgValArgGluPheMetValGlyAlaValGlySerGlyLeuIleValProProGluLeuSerAspProSerHisProCys
136 MetSerHisThr------MetGlyLeuGluSerLeuGluArgAlaLeuAspLeuLeuGlyArgPheArgGlyValSerProIleProAspProArgLeuTyrIleThrSerValProCys

197 AlaValCysPheGluGluLeuCysValThrAlaAsnGlnGlyAlaThrIleAlaAlaAlaTrpArgThrValSerValThrThrSerProSerArgArgArgCysGlyTrpThrProThr
174 TrpArgCysValGlyGluLeuMetValLeuProAsnHisGlyAsnProSerThrAlaGluGlyThrHisValSer--------------------------CysAsnHisLeuAlaVal

237 SerValArgArgTyrLeuProHisAlaAlaGlyLeuSerAspAlaAspArgAlaGlyAlaLeuArgValGlyProCysAlaGlyProAspArg---GlyGlyArgArgAlaAlaProPro
205 ProValAsnProGluProValSerGlyLeuPheGluAsnGluValArgGlnAlaGlyLeuGlyHisLeuLeuGluAlaGluGluLysAlaArgProGlyGlyProGluGluGlyAlaVal

276 ValAlaGluAsnAspSerValArgLysGluAlaAspAlaLeuLeuGluAlaHisAspValPheGlnAlaThrThrProGlyLeuTyrAlaIleSerGluLeuArgPheTrpLeuAlaSer
245 ProGlyProGlyArgProGluAlaGluGlyAlaThrArgAlaLeuAspThrTyrAsnValPheSerThrValProProGluValAlaGluLeuSerGluLeuLeuTyrTrpAsnSerGly

316 GlyAspArgAlaGlyGlnThrThrMetAspAlaPheAlaSerAsnLeuThrAlaLeuAlaArgGluLeuGlnGlnGluThrAlaAlaValAlaValGluLeuAlaLeuPheGlyArgArg
285 GlyHisAlaIleGlyAlaThr----------------------------------------------------------------------------------------------------

356 AlaGluHisPheAspArgAlaPheGlySerHisLeuAlaAlaLeuAspMetValAspAlaLeuIleIleGlyGlyGlnAlaThrSerProAspAspGlnIleGluAlaLeuIleArgAla
292 ---GlyGlnGlyGluGlyGlyGlyHisSerArgLeuSerAlaLeu-----------PheAlaArgGluArgArgLeuAlaLeuValArgGlyAlaCysGluGluAlaLeuAlaGlyAla

396 CysTyrAspHis--------------------HisLeuThrThrProLeuLeuArgArgLeuVal---SerProGluGlnCysAspGluGluAlaLeuArgArgValLeuAlaArgMet
327 ArgLeuThrHisLeuPheAspAlaValAlaProGlyAlaThrGluArgLeuPheCysGlyGlyValTyrSerSerSerGlyAspAlaValGluAlaLeuLysAlaAspCysAla------

428 GlyAlaGlyAlaGlyGlyProLysGlyGlyAlaGlyProAspAspAspGlyAspArgValAlaValGluGluGlyAlaArgGlyLeuGlyAlaProGlyGlyGlyGlyGluAspGluAla
365 ---------AlaAlaPheThrAlaHisProGlnTyrArgAlaIleLeuGlnLysArgAsnGluLeuTyrThrArgLeuAsnArgAlaMetGlnArgLeuGlyArgGlyGluGluAla

468 ProSerProArgAlaArgGlyThrGlyPro---GluThrTrpGlyAspIleAlaThrGlnAlaAlaAlaAspValArgGluArgArgArgLeuTyrAlaAspArgLeuThrLysArgSer
402 SerArgGluSerProGluValProArgProAlaGlyAlaArgGluProGlyProSerGlyAlaLeuSerAspAlaLeuLysArgLysGluGlnTyrLeuArgGlnValAlaThrGluGly

507 LeuAlaSerLeuGlyArgCysValArgGluGlnArgGlyGluLeuGluLysMetLeuArgValSerValHisGlyGluValLeuProAlaThrPheAlaAlaValAlaAsnGlyPheAla
442 LeuAlaLysLeuGlnSerCysLeuAlaGlnGlnSerGluThrLeuThrGluThrLeuCysLeuArgValTrpGlyAspValValTyrTrpGluLeuAlaArgMetArgAsnHisPheLeu

547 AlaArgAlaLeuLeuAlaAlaLeuThrAlaGlyAlaGlyThrValIleAspAsnArgSerAlaProGly---ValPheAspAlaHisArgPheMetArgAlaSerLeuLeuArgHisGln
482 TyrArg---------ArgAlaPheValSerGlyPro------TrpGluAspArgArgAlaGlyGluGlyAlaAlaPheGluAsnSerLysTyrIleLysThrHisLeuPheThrGlnThr

586 ValAspProAlaLeuLeuProSerIleThrHisArgPhePheGluLeuValAsnGlyProLeuPheAspHisSerThrHisSerPheAlaGlnProProAsnThrAlaLeuTyrTyrSer
517 LeuSerSerGluHisLeuHisAlaLeuThrHisSerLeuTyrThrPheIleThrGlyProLeuAlaGluGluSerGlyLeu---PheProProProSerAsnValAlaLeuAlaArgCys

626 ValGluAsnValGlyLeuLeuProHisLeuLysGluGluLeuAlaArgPheIleMetGlyAlaGlyGlySerGlyAlaAspTrpAlaValSerGluPheGlnArgPheTyrCysPheAsp
556 CysAspAlaAlaGlyThrLeuProHisGlnLysAlaPheLeuThrSerLeuIleTrp---ProGlyIleGluProSerAspTrpIleGluThrSerPheAsnSerPheTyrSerValPro

666 GlyIleSerGlyIleThrProThrGlnArgAlaAlaTrpArgTyrIleArgGluLeuIleIleAlaThrThrLeuPheAlaSerValTyrArgCysGlyGluLeuGluLeuArgArgPro
595 GlyGlySerLeuAlaSerSerGlnGln---IleLeuCysArgAlaLeuArgGluAlaValLeuThrValSerLeuTyrAsnLysThr---TrpGlyArgSerLeuIleLeuArgArgAla

706 AspCysSerArgProThrSerGluGlyArgTyrArgTyrProProGlyValTyrLeuThrTyrAspSerAspCysProLeuValAlaIleValGluSerAlaProAspGlyCysIleGly
633 AspAlaValSerPro---------GlyGlnAlaLeuProProAspGlyLeuTyrLeuThrTyrAspSerAspArgProLeuIleLeuLeuTyrLysGlyArgGlyTrpValPheLysAsp

746 ProArgSerValValValTyrAspArgAspValPheSerIleLeuTyrSerValLeuGlnHisLeuAlaProArgLeuProAspGlyGlyHisAspGlyProPro
670 LeuTyrAlaLeuLeuTyrLeuHisLeuGlnMetArgAspAspSerAla
```

FIG. 4. Alignment of the HSV-1 ICP18.5 amino acid sequence and the amino acid sequence encoded in the BALF3 reading frame of EBV. The alignment was manually constructed as described in the text. HSV-1 ICP18.5 is the upper sequence and EBV BALF3 is the lower sequence. :, Identical residues between the sequences; ----, gaps inserted to allow alignment of homologous seqments.

program of Lipman and Pearson (22). The remainder of the alignment was constructed manually by using the criterion that a gap would be inserted into a sequence only if it allowed a minimum of three residues to be brought into alignment. As seen in the homology dot matrix plot (Fig. 1G), the most homologous portions between the two proteins occurred near the carboxy terminal of the molecules. The longest stretch of uninterrupted identity ranges from ICP18.5 residues 723 to 729. Of the 780 predicted amino acids of ICP18.5, 174 or 22.3% aligned with predicted sequences of BALF3.

The arrangement of the coding sequences of the ICP18.5 and gB genes indicates that a total of 10 bp separated the third nucleotide of the last codon of ICP18.5 from the first

nucleotide of the initiation codon of gB. The predicted transcription initiation site of gB (27) is within the coding sequence of ICP18.5 and, since its mRNA is not spliced, the promoter-regulatory domain of gB, although at this time poorly defined, must also be located within the domain of the ICP18.5 gene. An interesting issue which remains to be resolved is the effect of transcription of the ICP18.5 gene on the transcription of the gB gene and, conversely, the effect of transcription of the gB gene on the expression of the 5′ gene specifying ICP18.5. Studies designed to define the relationship between the expression of these two genes are in progress.

Nothing is known of the function of the ICP18.5 protein other than the phenotype of viruses carrying mutations in this gene. The predicted amino acid sequence of the ICP18.5 does not indicate that its domain spans across cell or viral membranes. We cannot exclude the possibility, however, that it interacts with the inner surface of cell or viral membranes or with membrane proteins.

The relationship between ICP18.5 and gB is of interest primarily because the phenotype of ICP18.5 mutants suggests that they affect membrane proteins and because the gene specifying these proteins are not only juxtaposed but actually overlap in domain. This overlap in domain observed in the HSV-1 genome is conserved in the genome of the EBV. Thus, the BALF3 open reading frame encoding a putative protein related by sequence to ICP18.5 is juxtaposed 5′ to the BALF4 open reading frame expressing a glycoprotein related by sequence to gB (2, 26). Of special interest is the observation that the most closely related domains of the ICP18.5 and the protein encoded by the EBV BALF3 reading frame were those encoding the carboxyl terminus of the protein and the promoter-regulatory domain of the gB gene. If the conservation of homology is owing to the function of the ICP18.5 and the corresponding BALF3 protein, this would suggest that the important functional domain of ICP18.5 is located at the carboxyl end of the protein. Alternatively, the sequence conservation would suggest that the regulation of gB and its counterpart in the EBV genome has special, well-conserved requirements. This issue also remains to be resolved.

We thank Susanna Rudofsky for excellent technical assistance and Bart Barrell for useful discussions.

These studies were aided by Public Health Service grants CA-08494 and CA-19264 from the National Cancer Institute and by American Cancer Society grant MV-2R to the University of Chicago. P.E.P. was a predoctoral trainee, Public Health Service training grant T32-PHS-AI-07182; F.J.J. is a Public Health Service postdoctoral trainee, grant PHS-5-T32-CA-09241; M.A. was a fellow of the Swiss Nationalfond.

LITERATURE CITED

1. **Ackermann, M., M. Sarmiento, and B. Roizman.** 1985. Application of antibody to synthetic peptides for characterization of the intact and truncated α22 protein specified by herpes simplex virus 1 and the R325 α22⁻ deletion mutation. J. Virol. **56**:207–215.
2. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. G. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
3. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
4. **Bzik, D. J., B. A. Fox, N. A. DeLuca, and S. Person.** 1984. Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. Virology **133**:301–314.
5. **Campbell, M. E. M., J. W. Palfreyman, and C. M. Preston.** 1984. Identification of herpes simplex virus DNA sequences which encode a *trans*-acting polypeptide responsible for stimulation of immediate early transcription. J. Mol. Biol. **180**:1–19.
6. **Chou, J., and B. Roizman.** 1986. The terminal *a* sequence of the herpes simplex virus genome contains the promoter of a gene located in the repeat sequences of the L component. J. Virol. **57**:629–637.
7. **Dixon, R. A. F., D. J. Sabourin, and P. A. Schaffer.** 1983. Genetic analysis of temperature-sensitive mutants which define the genes for the major herpes simplex virus type 2 DNA-binding protein and a new late function. J. Virol. **45**:343–353.
8. **Draper, K. G., R. J. Frink, and E. K. Wagner.** 1982. Detailed characterization of an apparently unspliced β gene mapping in the interior of another. J. Virol. **43**:1123–1128.
9. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
10. **Ejercito, P. M., E. D. Kieff, and B. Roizman.** 1968. Characterization of herpes simplex virus strains differing in their effect on social behavior of infected cells. J. Gen. Virol. **3**:357–364.
11. **Frame, M. C., H. S. Marsden, and B. M. Dutia.** 1985. The ribonucleotide reductase induced by herpes simplex virus type 1 involves minimally a complex of two polypeptides (136K and 38K). J. Gen. Virol. **66**:1581–1587.
12. **Frischauf, A. M., H. Garrof, and H. Lehrach.** 1980. A subcloning strategy for DNA sequence analysis. Nucleic Acids Res. **8**:5541–5549.
13. **Glorioso, J. C., M. Levine, T. C. Holland, and M. S. Szczesiul.** 1980. Mutant analysis of herpes simplex virus-induced cell surface antigens: resistance to complement-mediated immune cytolysis. J. Virol. **35**:672–681.
14. **Hall, L. M., K. G. Draper, R. J. Frink, R. H. Costa, and E. K. Wagner.** 1982. Herpes simplex virus mRNA species mapping in *Eco*RI fragment I. J. Virol. **43**:594–607.
15. **Holland, L. E., R. M. Sandri-Goldin, A. L. Goldin, J. C. Glorioso, and M. Levine.** 1984. Transcriptional and genetic analyses of the herpes simplex virus type 1 genome: coordinates 0.29 to 0.45. J. Virol. **49**:947–959.
16. **Kadonaga, J. T., K. A. Jones, and R. Tijan.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
17. **Kousoulas, K. G., P. E. Pellett, L. Pereira, and B. Roizman.** 1984. Mutations affecting conformation or sequence of neutralizing epitopes identified by reactivity of viable plaques segregated from *syn* and *ts* domains of HSV-1(F) gB gene. Virology **135**:379–394.
18. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**: 1–45.
19. **Kozak, M.** 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature (London) **308**:241–246.
20. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
22. **Lipman, D., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
23. **Little, S. P., J. T. Jofre, R. J. Courtney, and P. A. Schaffer.** 1981. A virion associated glycoprotein necessary for infectivity of herpes simplex virus type 1. Virology **115**:149–160.
24. **McLaughlan, J., and J. B. Clements.** 1983. Organization of the herpes simplex virus type 1 transcription unit encoding two early proteins with molecular weights of 140,000 and 40,000. J. Gen. Virol. **64**:997–1006.
25. **Pancake, B. A., D. P. Aschman, and P. A. Schaffer.** 1983. Genetic and phenotypic analysis of herpes simplex type 1 mutants conditionally resistant to immune cytolysis. J. Virol. **47**:568–585.
26. **Pellett, P. E., M. D. Biggin, B. Barrell, and B. Roizman.** 1985.

Epstein-Barr virus genome may encode a protein showing significant amino acid and predicted secondary structure homology with glycoprotein B of herpes simplex virus 1. J. Virol. **56**:807–813.

27. **Pellett, P. E., K. G. Kousoulas, L. Pereira, and B. Roizman.** 1985. The anatomy of the herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody-resistant mutants. J. Virol. **53**:243–253.
28. **Pellett, P. E., J. L. C. McKnight, F. J. Jenkins, and B. Roizman.** 1985. Nucleotide sequence and predicted amino acid sequence of a protein encoded in a small herpes simplex virus DNA fragment capable of *trans* inducing α genes. Proc. Nat. Acad. Sci. USA **82**:5870–5874.
29. **Quinn, J. P., and D. J. McGeoch.** 1985. DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. Nucleic Acids Res. **13**:8143–8150.
30. **Rixon, F. J., and D. J. McGeoch.** 1985. Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1. Nucleic Acids Res. **13**:953–973.
31. **Sarmiento, M., M. Haffey, and P. G. Spear.** 1979. Membrane proteins specified by herpes simplex viruses. III. Role of glycoprotein VP7(B_2) in virion infectivity. J. Virol. **29**: 1149–1158.
32. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. **10**:2951–2961.