JOURNAL OF VIROLOGY, Apr. 1985, p. 21–29
0022-538X/85/040021-09$02.00/0

Vol. 54, No. 1

Mapping and Sequence of the Gene for the Pseudorabies Virus Glycoprotein Which Accumulates in the Medium of Infected Cells

THOMAS J. REA, JAMES G. TIMMINS, GRAHAM W. LONG,† AND LEONARD E. POST*
The Upjohn Company, Kalamazoo, Michigan 49001

Received 3 October 1984/Accepted 28 December 1984

RNA from pseudorabies virus (PRV)-infected cells was translated in a reticulocyte lysate with and without the addition of dog pancreas microsomes. Upon addition of the microsomes to the translation reaction, an additional prominent protein product was observed that was not present when microsomes were omitted. The gene coding for this processed protein and its lower-molecular-weight precursor was mapped within the small unique region of the genome by hybridization of mRNA to cloned fragments of PRV DNA and translation of the selected mRNAs. A fragment of the coding region of this gene was inserted into an open reading frame cloning vector to express part of this gene as a hybrid protein in *Escherichia coli*. This hybrid protein was injected into mice to raise an antiserum which was found to precipitate the glycoprotein which accumulates in the medium of PRV-infected cells. This allows us to conclude that the gene for the "excreted" glycoprotein (gX) maps to the small unique region of the genome, and that the precursor of this glycoprotein is readily processed by dog pancreas microsomes. The region of the PRV genome which codes for this glycoprotein was sequenced and found to include an open reading frame coding for 498 amino acids, flanked by sequences which contain features common to eucaryotic promoters and polyadenylation signals. The predicted protein sequence includes a hydrophobic sequence at the N-terminus which could be a signal sequence, and a hydrophobic sequence followed by a hydrophilic sequence at the C-terminus.

Some of the most important interactions between a herpesvirus and the immune system of its host are mediated by the viral glycoproteins. The glycoproteins are found on the surfaces of both virions and infected cells. In the best-studied herpesvirus, herpes simplex virus, both humoral and cellular immune responses are evoked against the viral glycoproteins, and these responses are important for recovery of the host from viral infection and prevention of viral reinfection (reviewed in references 6, 26, and 36). The glycoproteins on the surfaces of virions and cells serve several functions: facilitation of entry into the host of the infecting virus, receptor activity for the Fc segment of antibody molecules, and receptor activity for the C3b component of complement (12, 28, 37). Several herpesviruses have been observed to release glycoproteins into the medium of infected cells. Herpes simplex virus releases glycoprotein C and several truncated forms of glycoprotein D into the medium (27, 35). Marek's disease virus releases a considerable amount of the virion glycoprotein A into the medium (38), and herpes saimiri also releases a virion glycoprotein into the medium (34). Pseudorabies virus (PRV) releases a glycoprotein into the medium which reportedly is not present in virions (2). The role in viral infection of glycoproteins released into the medium is not understood. In some cases, release in the medium could be a fortuitous consequence of proteolysis of a membrane glycoprotein. The PRV glycoprotein in the medium is the only one of these examples reported to be distinct from virion glycoproteins, but the evidence for this is based on a difference in electrophoretic mobilities and extent of sulfation. This still does not preclude the possibility that the medium form could be a proteolysis product of a viral membrane glycoprotein.

In this report, we describe the mapping and sequence of the gene for the PRV glycoprotein released into the medium of infected cells. The method for identification of the gene relies on production of a monospecific polyclonal antiserum raised against a protein produced in *Escherichia coli*. This antiserum suggests that the PRV glycoprotein released into the medium is antigenically distinct from the major viral membrane glycoproteins, and that this glycoprotein is not simply a proteolytic product of one of the major membrane glycoproteins. However, the sequence of the protein predicted from the DNA sequence has features in common with many membrane glycoproteins.

MATERIALS AND METHODS

Cells and virus. For propagation of PRV, preparation of PRV DNA and RNA, and preparation of radioactive PRV proteins, Vero cells (ATCC CCL 81) were used. Cells were grown in Dulbecco modified Eagle medium (GIBCO Laboratories) supplemented with 100 U of penicillin per ml, 100 μg of streptomycin per ml, 10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid), and 10% fetal bovine serum. PRV Rice strain was originally provided by D. P. Gustafson, Purdue University, West Lafayette, Ind. Virus stocks were prepared by infecting Vero cells at a multiplicity of infection of less than 0.01, in medium 199 (GIBCO) supplemented with 100 U of penicillin per ml, 100 μg of streptomycin per ml, and 1% fetal bovine serum.

DNA. PRV DNA was prepared from the cytoplasm of infected Vero cells ca. 16 h after infection at a multiplicity of infection of 5. Infected cells were harvested by centrifugation, washed in phosphate-buffered saline, and suspended in ice-cold TE buffer (10 mM Tris-hydrochloride [pH 7.4], 1 mM EDTA) at 2 ml per 850-cm^2 roller bottle. After the cells were allowed to swell to 15 min on ice, Nonidet P-40 (NP-40) was added to 0.2% and the cells were incubated for another 15 min on ice. The nuclei were removed by low-speed

* Corresponding author.
† Present address: Wayne State University Medical School, Detroit, MI 48202.

centrifugation, and DNA was purified from the cytoplasm by sodium iodide density gradient centrifugation as described previously (40). This viral DNA was used to clone the PRV *Bam*HI fragments into plasmid vector pSV2gpt (24), by methods previously described for cloning herpes simplex DNA (32).

RNA. Cytoplasmic RNA was prepared from Vero cells 12 to 16 h after infection at a multiplicity of infection of 5 to 10. Cells were harvested by centrifugation, washed once in cold phosphate-buffered saline, and lysed by suspension in 5 mM Tris-hydrochloride (pH 7.5)–10 mM NaCl–1% NP-40–10 mM vanadyl ribonucleoside complex (Bethesda Research Laboratories). After a 10 min incubation on ice, the nuclei were removed by low-speed centrifugation. The supernatant was extracted three times with phenol-chloroform and ethanol precipitated. Polyadenylated RNA was selected on oligodeoxythymidylate-cellulose (Collaborative Research, Inc.).

Hybrid selection of RNA. Plasmids carrying cloned PRV fragments were immobilized on nitrocellulose filters as described by Paterson and Roberts (29). Hybridizations were carried out under conditions described by Mackem and Roizman (16).

In vitro translation. Total PRV polyadenylated RNA or hybrid-selected RNA in water was translated by the use of a reticulocyte lysate system from New England Nuclear Corp. The RNA was translated as recommended by the supplier, except that the reaction was performed in 0.4 mM magnesium acetate–80 mM potassium acetate–32 μM spermine (16). Translation reactions with the cotranslational processing preparation contained 3% (vol/vol) extract from canine pancreatic microsomal membranes (New England Nuclear). Products of the in vitro translational reactions were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis, using the system described by Morse et al. (23).

Preparation of antiserum against PRV fusion protein 60-11. *E. coli* MC1061 (5) carrying plasmid p60-11 (which directs the expression of a hybrid protein including gX sequences) was grown to stationary phase in L broth and diluted 100-fold into M9 medium plus 1% Casamino Acids (18). When the optical density of the culture was between 1 and 2, the bacteria were harvested by centrifugation. The cells were suspended in 50 mM Tris-hydrochloride (pH 7.6)–0.3 M NaCl–1 mM EDTA (5 to 10 ml per liter of culture), and egg white lysozyme (Sigma Chemical Co.) was added to 1 mg/ml. After 15 min on ice, NP-40 was added to 0.2% and incubation was continued for an additional 15 min on ice. The extract was sonicated until no longer viscous and then centrifuged at 15,000 × *g* for 15 min. The recombinant protein was found almost entirely in the resulting pellet. The pellet was dissolved by boiling in SDS gel sample buffer and was loaded onto a preparative (3-mm) SDS slab gel (23). The location on the gel of the recombinant protein was determined by cutting off a small strip for staining with Coomassie blue, and the protein band was cut from the unstained portion of the gel. The recombinant protein was recovered from the gel by mincing the gel, soaking in 0.05% SDS to 0.05 M ammonium bicarbonate, and acetone precipitation from the resulting supernatant (42). The protein was suspended in Tris-saline buffer. The protein solution was mixed 1:1 with complete Freund adjuvant, and ca. 50 μg was injected subcutaneously into female ICR mice. The animals were reimmunized intraperitoneally with ca. 50 μg of protein in incomplete Freund adjuvant (1:1) every 10 days over a 6-week period. Animals were bled from the tail, and serum was stored at −20°C.

Analysis of PRV-infected cell polypeptides. [^{14}C]glucosamine-labeled viral proteins were prepared by infection of Vero cells (MOI, 5 to 10) in medium 199 plus 1% fetal bovine serum, with the addition of 3 μCi of [^{14}C]glucosamine per ml at 4 h after infection. At 12 to 16 h after infection, cells were shaken into the medium and harvested by centrifugation. The medium was saved as a source of radioactive gX. The cells were washed once in phosphate-buffered saline and suspended in phosphate-buffered saline plus 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and 200 μg of ovalbumin per ml. Debris was removed from the cellular extract preparation by centrifugation in an Eppendorf microcentrifuge. For analysis by immunoprecipitation, the medium or cellular extract samples were preabsorbed for 30 min with *Staphylococcus aureus* cells (Calbiochem) (*S. aureus* cells were washed before use with 50 mM Tris-hydrochloride, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 0.5% NP-40; they were then washed with the same buffer with only 0.05% NP-40 and finally suspended in the low NP-40 buffer with 200 μg of ovalbumin per ml). The mixture was centrifuged for 45 min in a Beckman SW50.1 rotor at 35,000 rpm, and the supernatant was used for immunoprecipitations. Antiserum was incubated with the samples for 1 h on ice, then *S. aureus* cells washed as described above were added and the incubation continued for an additional 30 min on ice. The precipitates were collected by centrifugation, washed three times in RIPA/SDS buffer (10 mM Tris-hydrochloride [pH 7.2], 150 mM NaCl,

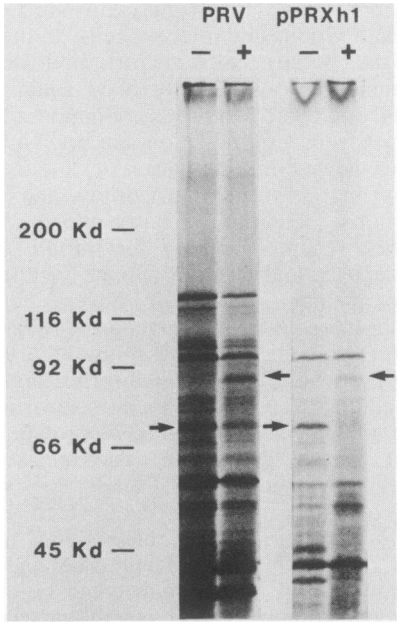


FIG. 1. In vitro cotranslational processing of PRV proteins. The lanes labeled PRV represent translations of total polyadenylated RNA from PRV-infected Vero cells. The lanes labeled pPRXh1 represent translations of RNA that was hybrid selected by plasmid pPRXh1, which contains an *Xho*I fragment that includes all of *Bam*HI fragments 7 and 12 and parts of *Bam*HI fragments 10 and 6. Lanes labeled with a minus sign are the products of a reticulocyte lysate translation. Lanes labeled with a plus sign are the products of a reticulocyte lysate to which dog pancreas microsomes were added as a source of cotranslational processing activity. The products of the translation were labeled with [^{35}S]methionine, and an autoradiogram is shown. The arrows point out the unprocessed and cotranslationally processed forms of gX.

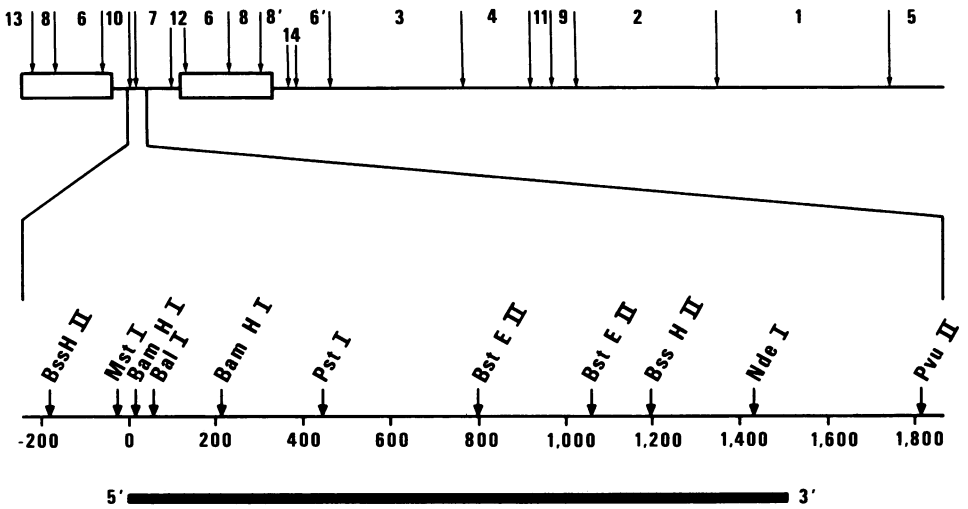


FIG. 2. Structure of the PRV genome and the gX gene. The PRV genome is represented at the top, with the inverted repeats of the small component of the genome indicated by boxes. The arrows indicate the positions of *Bam*HI cleavage sites on the genome of PRV Rice, which have been shown to be the same as in the strain mapped by Feldman and co-workers (11), although the *Bam*HI fragments within the repeats appear to comigrate with a smaller *Bam*HI fragment, so on Rice we identify them as *Bam*HI 6 rather than *Bam*HI 5 (data not shown). The *Bam*HI fragments are numbered in decreasing order of size. The region of the genome coding for gX is expanded with several restriction enzyme cleavage sites shown, and the coding sequences are indicated by the bar beneath the line. Numbers on the lower line indicate DNA size in base pairs, with 1 being the first base of the initiation codon of the gX open reading frame, as in Fig. 6.

1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1 mg of ovalbumin per ml). The precipitates were boiled in SDS gel sample buffer and analyzed by SDS-polyacrylamide gel electrophoresis (23).

DNA sequencing. DNA sequencing was done by the dideoxy method after fragments were cloned into M13 vectors. Two series of phage with one random end were constructed by the method of Poncz et al. (31). In one series, pPR28-4 (a clone of the 1.5-kilobase (kb) *Bam*HI-*Pvu*II fragment from *Bam*HI 7 in pBR322) was digested with *Pvu*II, and random ends were generated with BAL 31 nuclease by digestion with BAL 31 nuclease for various times (5 to 30 min) under the conditions of Poncz et al. (31). The resulting mixture was digested with *Bam*HI, and the fragments were cloned between the *Bam*HI and *Sma*I sites of M13 mp8. In the other series, pPR28 (a clone of *Bam*HI 7 in pSV2gpt) was digested with *Bam*HI, random ends were generated with BAL 31 nuclease and digested with *Sal*I (which cleaves pPR28 400 nucleotides to the right of the *Pvu*II site [see Fig. 2]), and the fragments were cloned between the *Sal*I and *Sma*I sites of M13 mp9. Other phage were made by insertion of specific restriction fragments into the mp8 and mp9 vectors. All methods for working with the M13 vectors were described by Messing (22). Presumably because of the high GC content of the DNA, we found the dideoxy sequencing methods involving Klenow fragment of DNA polymerase I to give poor results. Reactions were done with avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) at 50°C in 34 mM Tris-hydrochloride (pH 8.3)–50 mM NaCl–6 mM $MgCl_2$–5 mM dithiothreitol, plus deoxynucleoside triphosphates and dideoxynucleoside triphosphates in concentrations empirically determined to give desirable sequencing patterns. The labeled triphosphate was deoxycytosine 5'-[^{35}S thio]triphosphate (New England Nuclear) (3). Other details of dideoxy sequencing methods were as described by Messing (22). Parts of the sequence were confirmed by the method of Maxam and Gilbert (20). Sequences of both strands of the DNA were determined at least twice.

RESULTS

In vitro translation of a processed PRV protein. For examination of the effects of cotranslational processing on the in vitro synthesis of PRV proteins, PRV mRNA was translated in a reticulocyte lysate in the presence and absence of dog pancreas microsomes. When the products of the translation reactions were examined by gel electrophoresis (Fig. 1), a number of protein products were eliminated or reduced in yield by addition of the microsomes. More importantly, a protein product of apparent molecular weight 95,000 appeared when microsomes were added to the system, but not in the absence of the microsomes. Dog pancreas microsomes have been shown by others to provide cotranslational processing of proteins by removal of signal peptides and core glycosylation of glycoprotein products (15). It seemed possible that the 95-kilodalton (kd) product represented a glycosylated PRV protein.

Because several protein products disappeared upon addition of the microsomes, it was not obvious from translation of total RNA which product was the precursor of the prominent processed product. This required fractionation of the mRNAs before translation. A series of hybrid selection experiments was carried out with cloned PRV DNA fragments from across the viral genome. All fragments from within the large component of the genome failed to select RNA coding for the 95-kd processed protein. Figure 1 shows that the large *Xho*I fragment extending from within *Bam*HI 10 to *Bam*HI 6, including *Bam*HI 7 and *Bam*HI 12 (Fig. 2), did select an mRNA which was translated in the presence of microsomes to produce the prominent processed 95-kd protein (Fig. 1). When this same mRNA was translated in the absence of microsomes, a prominent product of molecular weight 70,000 was produced that was not observed in the translation with microsomes.

Further hybrid selection experiments were done to locate more precisely the coding region of DNA for the processed protein and to confirm that the putative precursor did map to the same region of DNA. Fragments *Bam*HI 7 and, to a

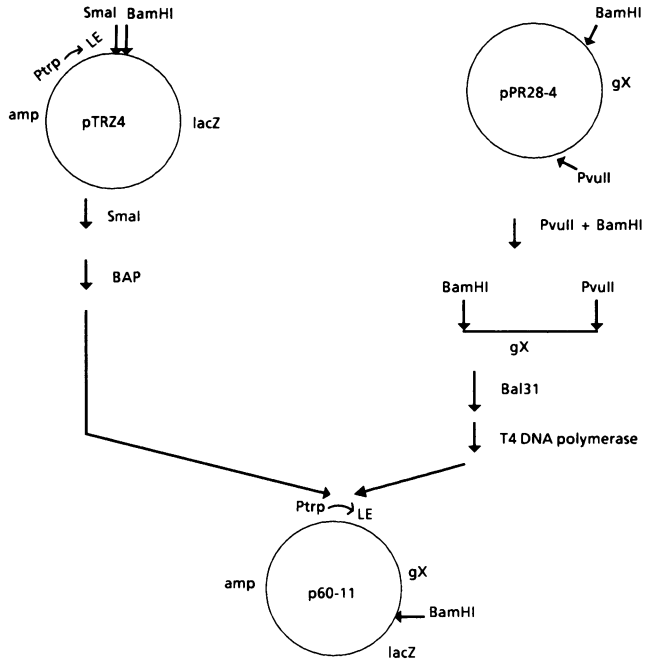


FIG. 3. Construction of a plasmid to express the gX open reading frame in *E. coli*. The pTRZ4 vector contains an out-of-phase *trp-lac* fusion, with the components derived as follows: the fragment carrying the *trp* promoter, the translation initiation region of *trpE*, and the first 13 amino acids of the *trpLE* fusion Δ1413 was derived from pVV1 (25); the fragment carrying the *lacZ* gene minus its translation initiation region was from pMC1403 (5). The fusion was made by ligating a DNA polymerase-treated *Bgl*II end from pVV1 to a DNA polymerase-treated *Eco*RI end from pMC1403. pTRZ4 was cleaved with *Sma*I to insert random pieces of PRV DNA from the gX coding region between the *trp* and *lac* sequences. The plasmid pPR28-4 contains the PRV DNA between the *Pvu*II and *Bam*HI cleavage sites shown in the expanded region of Fig. 2. The PRV DNA was made to have random ends by treatment for various times (5 to 30 min) with BAL 31 nuclease under the conditions of Poncz et al. (31) and then was ligated into pTRZ4. The plasmid p60-11 was derived from a transformed colony which was particularly dark blue on X-gal plates.

lesser extend *Bam*HI 10 were observed to select mRNA coding for both the processed protein and the identified precursor. A 1.5-kb *Bam*HI-*Pvu*II fragment (Fig. 2) from within *Bam*HI 7 was found to select the mRNA for both processed protein and precursor, and the remainder of *Bam*HI 7 did not select any detectable mRNA for the processed protein. That both the processed protein and the putative precursor mapped to the same small region of the genome strongly indicated a precursor-product relationship between the 70-kd product from the reticulocyte lysate reaction and the 95-kd product produced by the lysate in the presence of microsomes. Since the presence of the microsomes during translation caused an increase in molecular weight of the product, the processed product must be glycosylated and presumably was from translation of a PRV glycoprotein mRNA.

Identification of the processed glycoprotein. The hybrid selection experiments mapped the glycoprotein gene to a 1.5-kb region of *Bam*HI 7; in addition, a small amount of mRNA coding for the glycoprotein was selected by *Bam*HI 10 in some experiments. Even if the inefficient selection by *Bam*HI 10 might be due to a very small coding region for the glycoprotein on the fragment, and if so, the 300-base-pair fragment between *Bam*HI 7 and *Bam*HI 10 also might code for the glycoprotein mRNA, coding capacity requirements for a protein the size of the glycoprotein precursor lead to the conclusion that most of the 1.5-kb *Pvu*II-*Bam*HI fragment from *Bam*HI 7 must code for the glycoprotein. A random collection of DNA fragments from within the *Pvu*II-*Bam*HI fragment was generated by BAL 31 nuclease treatment of the fragment (see Fig. 3). Some of the resulting fragments were presumed to be from within the coding region of the glycoprotein gene.

The collection of fragments produced by BAL 31 was ligated into the open reading frame cloning vector pTRZ4. The plasmid pTRZ4 is shown in Fig. 3. The pTRZ4 plasmid consists of the *trp* promoter, the translation initiation region of the *trpL* gene, the coding region for the first 14 amino acids of a *trpLE* fusion, followed by a *lacZ* gene minus its translation initiation region. The crucial feature of the vector is that the *trp* and *lac* coding sequences are not in phase, but restriction sites between the two coding regions offer the possibility that insertion of appropriate DNA fragments could bring them in phase and thus allow synthesis of β-galactosidase activity. The ligation of the BAL 31 fragments of the glycoprotein coding region into pTRZ4 was transformed into a *lac* minus *E. coli* strain, and transformants were plated onto X-gal plates where *lac* plus colonies could be identified by blue color. A dark blue colony was picked and found to contain an insertion of ca. 600 base pairs of PRV DNA. This plasmid expressing β-galactosidase activity was called p60-11.

Figure 4 shows the proteins produced by *E. coli* carrying p60-11 when grown in minimal medium in which the *trp* promoter is fully induced. The most abundant protein in these cells was a large protein not present in cells without a plasmid or cells containing a plasmid without an insertion of foreign DNA. As has been found for many proteins overproduced in *E. coli* by recombinant DNA methods, the p60-11 protein accumulates in the cell as an aggregate that is readily harvested by centrifugation of the lysed *E. coli*. Using this property of the protein for a simple preparative step as described above, the protein was extracted and purified by preparative SDS-polyacrylamide gel electrophoresis. Since this recombinant protein is the largest protein in the host cells, the electrophoretic purification gave a protein preparation of greater than 95% purity, estimated by silver staining of an SDS gel of the preparation.

The recombinant protein was injected into mice and was found to raise antibodies which reacted in an enzyme-linked immunosorbent assay against PRV-infected Vero cells (data not shown). The serum from mice injected with the recombinant protein was used in immunoprecipitations with [^{14}C]glucosamine-labeled proteins from PRV-infected cells (Fig. 5). When the [^{14}C]glucosamine-labeled proteins from infected cells were used as the antigen, a precipitated labeled protein was visible only on overexposure of an autoradiogram, and this protein did not comigrate with any of the major PRV glycoproteins associated with infected cells. However, in addition to the glycoproteins that accumulate in the cells, a glycoprotein accumulates in the medium of PRV-infected cells. This glycoprotein from the medium was efficiently precipitated by antiserum to the p60-11 recombinant protein. Since p60-11 is coded by the same PRV sequences that were found to select a glycoprotein mRNA, this is strong evidence that the glycoprotein gene mapped in the hybrid selection experiments codes for the glycoprotein that accumulates in the medium of PRV-infected cells.

Because there is as yet no systematic nomenclature for the PRV glycoproteins and the relationship between the glycoprotein in the medium of infected cells and the viral membrane glycoproteins is not well established, we provisionally refer to the glycoprotein in the medium as gX.

DNA sequence of the gX coding region. The hybrid selection experiments described above mapped the DNA coding for gX to within ca. 2 kb of DNA, assuming that a very limited homology was responsible for the inefficient selection of the gX mRNA by *Bam*HI 10. The 1.5-kb *Pvu*II-*Bam*HI fragment from *Bam*HI 7, the small *Bam*HI fragment from between *Bam*HI 7 and *Bam*HI 10, and the end of *Bam*HI 10 adjacent to *Bam*HI 7 was sequenced. The DNA sequencing was mostly by the dideoxy method, using M13 subclones of the PRV DNA, with some by the Maxam-Gilbert method. The high GC content (1) of PRV DNA made the usual procedure of dideoxy sequencing with the Klenow fragment of DNA polymerase I highly unsatisfactory. Therefore, a procedure with reverse transcriptase at 50°C was used as described above. The sequence determined from the gX coding region is shown in Fig. 6. The sequence contains an open reading frame that could code for a protein of 498 amino acids, which we presume to be gX, as discussed below.

The DNA sequence of the junctions of PRV DNA with vector DNA in p60-11 was determined. The PRV insertion was found to include nucleotides 777 to 1341, inserted such that the phase translated in the hybrid protein would be included in the reading frame shown in Fig. 6 (Fig. 7).

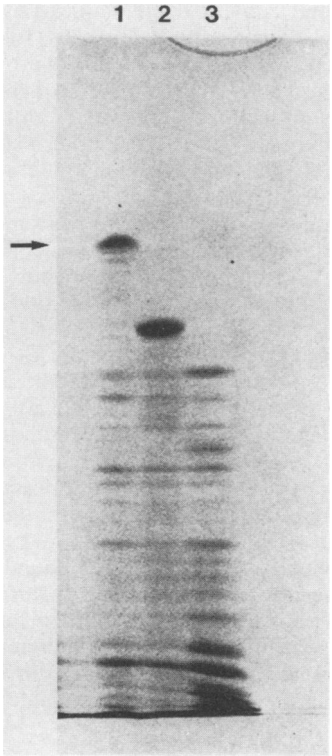


FIG. 4. Expression of the gX open reading frame in *E. coli*. *E. coli* strains were grown in M9 as described in the text, harvested by centrifugation, and boiled in SDS gel sample buffer without fractionation for an examination of the total protein being expressed in the cells. Lane 1, protein being made in a strain carrying p60-11; the arrow points out the fusion protein which includes the gX sequences. Lane 2, strain carrying plasmid pTRZ5, which is identical to pTRZ4 except that the *trp* and *lac* sequences are in phase without any insertion of foreign DNA, and therefore the pTRZ5 fusion protein shows the size for the *trp-lac* fusion protein without any insertion. Lane 3, *E. coli* containing no plasmid.

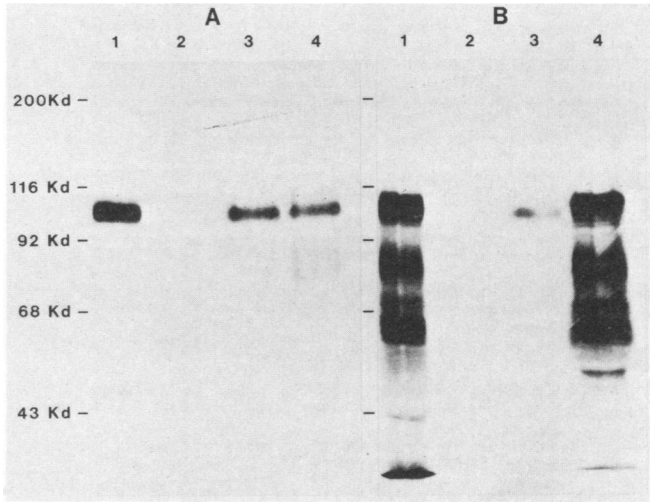
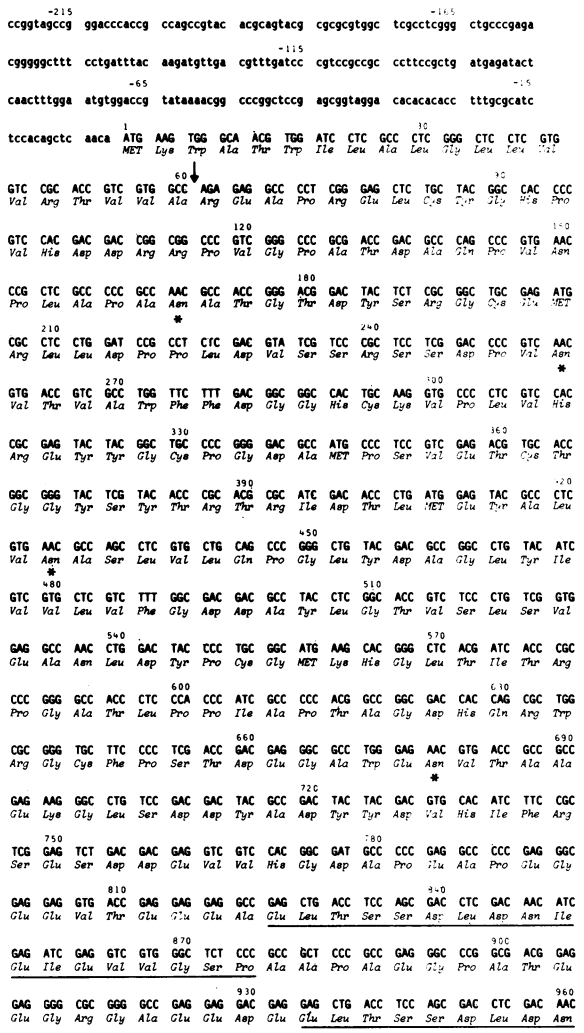


FIG. 5. Immunoprecipitation of PRV proteins with antiserum raised against the p60-11 fusion protein. (A) Immunoprecipitation of medium from PRV-infected cells grown in the presence of [^{14}C]glucosamine. (B) Immunoprecipitation of the proteins from the infected cells. In both cases, lanes 1 are the total labeled proteins before immunoprecipitation, lanes 2 are the result of precipitation with preimmune mouse serum, lanes 3 are precipitated with mouse serum raised against the p60-11 fusion protein, and lanes 4 are precipitated with serum from pigs recovered from PRV infection. B is overexposed to show the small amount of precipitating glycoprotein in lane 3.

DISCUSSION

This report describes the identification of a PRV glycoprotein gene from the initial observation that a prominent new product appeared in the products of a reticulocyte lysate translation of PRV RNA when dog pancreas microsomes were added to the translation. The coding sequence for this product was mapped by hybrid selection experiments and was found to be the same as for a lower-molecular-weight product of the same translation reaction in the absence of the microsomes. Since the only known activity of the microsomes which could increase the molecular weight of a product is the cotranslational glycosylating activity, it was inferred that the mapped gene codes for a PRV glycoprotein. A detailed study of the processing of herpes simplex glycoprotein D by the reticulocyte lysate plus microsomes concluded that the in vitro system was adding the core glycosylation to authentic in vivo glycosylation sites (19). The proof that the mapped gene is indeed a glycoprotein gene, and the identification of the coded glycoprotein, were by expression of the sequences in *E. coli* and use of the bacterial protein for production of a monospecific antiserum. Immunoprecipitation of the glycoprotein accumulated in the medium of PRV-infected cells by this serum is strong evidence that the sequences inserted into the expression vector code for the glyprotein. Others have used similar open reading frame expression vectors to produce proteins in *E. coli* for the purpose of raising a monospecific antiserum (13, 43). We note that in the case of the glycoprotein gene, the expression in *E. coli* was done before the precise location of the gene



was known. It is not necessary to have a defined DNA sequence to use the open reading frame expression vectors to raise an antiserum to a coded gene product. The antiserum was used to identify the mapped gene as that for the PRV glycoprotein which accumulates in the medium of infected cells. This glycoprotein was first identified by Ben-Porat and Kaplan (2), who named the protein 3a on the basis of its electrophoretic mobility relative to other proteins in their gel system. Since there is no systematic nomenclature yet for PRV glycoproteins, we refer to this medium glycoprotein as gX. Our finding that the anti-gX serum does not immunoprecipitate the major PRV proteins that accumulate in infected cells is consistent with the conclusion of Ben-Porat and Kaplan (2) that the glycoprotein in the medium is not one of the viral glycoproteins that accumulate in infected cells. However, we did find protein reacting with the anti-gX serum in the cellular protein preparations. This may have been newly synthesized gX before release into the medium, but it could also indicate a role for gX as a cellular glycoprotein.

The sequence of the DNA which selected the gX mRNA contains an open reading frame coding for 498 amino acids (Fig. 6). The predicted molecular weight of the coded amino acid sequence is 53,700, considerably smaller than the apparent mass of the 70-kd precursor seen with in vitro translation. Perhaps this is because of the high percentage (8.8%) of proline residues in the sequence. The protein sequence predicted from the open reading frame has features of a membrane protein. At the N-terminus, following a basic amino acid lysine, is a sequence of hydrophobic amino acids. At the end of this hydrophobic sequence (nucleotide 46 in Fig. 6) is a sequence consistent with the rules of Perlman and Halvorson (30) for predicting a signal peptidase cleavage site after the alanine at nucleotide 60 in Fig. 6. At the C-terminus of the predicted protein sequence is a very hydrophobic sequence of 25 amino acids (nucleotides 1374 to 1449) followed by a hydrophilic sequence. This C-terminal sequence looks very much like the "anchor" sequences of extrinsic membrane proteins, raising the possibility that gX accumulating in the medium is a cleavage product rather than a true secreted protein. Because the gX in the medium had the same apparent molecular weight as the protein detected in the cells, proteolysis could not be extensive. However, cleavage of putative anchor sequence could involve removing less than 5% of the amino acids of the precursor and none of the carbohydrate, producing a change in molecular weight difficult to detect in a diffuse glycoprotein band on an SDS gel. If the gX in the medium is a proteolysis product of a membrane bound form of the protein, the question remains which form is biologically relevant for virus infection or whether both forms are important. Herpes simplex has been shown to accumulate forms of glycoproteins C (27) and D (34) in the medium of infected cells, but no information is available on the biolog-

FIG. 6. DNA sequence of the gX open reading frame and flanking area. The sequence is numbered, with +1 as the first base of the initiation codon of the open reading frame. The sequence spans the expanded region of Fig. 2, and the last three bases represent half of the *Pvu*II cleavage site. The arrow after nucleotide 60 represents the position in the protein sequence at which the signal peptide would be cleaved according to the rules of Perlman and Halvorson (30). An asterisk has been placed under the asparagine residues which are part of consensus N-linked glycosylation sequences. The direct repeat within the sequences is underlined.

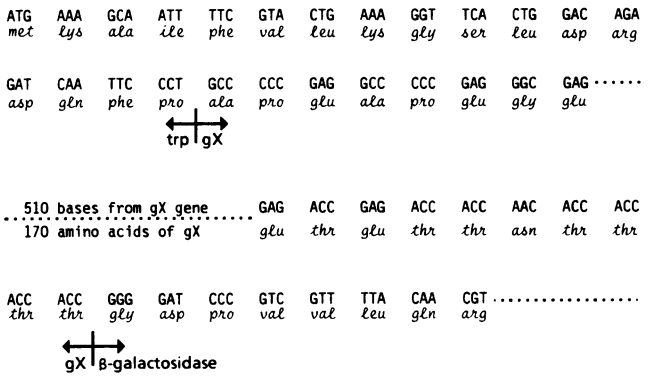


FIG. 7. DNA sequence and predicted amino acid fusion protein sequence from p60-11. The first codon is the initiation codon of the *trpL* gene. The sequence before the trp/gX dividing point is that coded by the *trpLE* fusion Δ1413 (25) and the linker created by fusion of the *Bgl*II site within the *Eco*RI site from pMC1403 (5). The sequence after the trp/gX dividing point begins at position 778 of Fig. 6. The sequence before the gX/β-galactosidase fusion is the linker from pMC1403 and the remainder of the *E. coli* β-galactosidase sequence.

ical significance, if any, of the released glycoproteins. Also, Marek's disease virus (38) and herpesvirus saimiri (34) release membrane glycoproteins into the medium. Relase of immunologically relevant molecules such as glycoproteins by virus-infected cells could be an important mechanism for interaction with the immune system of the host. The PRV system in which a substantial fraction of the gX protein is released (2) may provide an opportunity to study such a possibility. Although the ends of the predicted gX sequence resemble a typical membrane protein, the sequence has some unusual features. The most remarkable feature is an internal duplication of an 18-amino acid sequence (nucleotides 823 to 876 repeated in nucleotides 934 to 987). The DNA sequence is also duplicated, although this may not be surprising given the highly nonrandom codon usage throughout the gene, discussed below. Between the repeats is the hydrophilic maximum of the gX protein sequence as determined by the Hopp and Woods algorithm (14). The regions surrounding the repeats also have unusual amino acid sequences: on the N-terminal side of the first repeat is a sequence in which 8 of 14 amino acids are glutamic acid (nucleotides 784 to 825); and on the C-terminal side of the second repeat is a sequence in which 11 of 17 amino acids are proline (nucleotides 985 to 1035). Another unusual sequence occurs between nucleotides 1315 and 1347 in which 8 of 11 amino acids are threonine. The structural and functional significance of these unusual amino acid sequence features is difficult to predict. As expected from the GC-rich PRV genome (1), the codon usage in the gX gene is such that codons with G or C in the third position are used preferentially. The extent of this preference is such that 94% of all codons in the gene have G or C in the third position. In fact, 30% of the codons in the gene are entirely G and C, which raises the possibility that whatever selective pressure led to the GC-rich genome may even have influenced the amino acid composition of the protein.

The open reading frame identified as the gX gene is flanked by sequences that have common features in eucaryotic transcription initiation and termination signals. The sequence AATAAA is common to polyadenylation signals of eucaryotic genes (33) and is found 58 nucleotides downstream from the termination codon of the gX gene. At the 5' end of the gene, the TATAA sequence at position −64 from the initiation codon is the canonical TATA box found in eucaryotic promoters (4). Since most eucaryotic mRNAs are initiated 20 to 30 bases downstream from a TATA box, this interpretation would predict that the mRNA start position of the gX gene would be in the CACACACA sequence at position −24 from the initiation codon. The region upstream from the putative TATA box has a peculiar distribution of base composition. The TATA box and 30 base pairs upstream are 65% AT, in marked contrast to the average 71% GC of the sequenced region of DNA and the PRV genome as a whole. Upstream from the AT-rich region is a sequence with 82% GC (positions −14 through −95), even more GC rich than the average of the sequence. Upstream from that is a sequence with 82% AT (−147 through −113), and further upstream is another GC-rich sequence (positions −159 through −148) of 83% GC. Blocks of GC-rich sequence interspersed with less GC-rich sequences have been implicated as important features of herpes simplex promoters (9, 17, 21, 44), the simian virus 40 early promoter (10), and the rabbit β-globin promoter (8). We conclude that the sequence upstream from the gX open reading frame looks very much like a promoter region and is probably the sequence responsible for transcription of the gX gene.

It should be noted that all of the above discussion rests on the assumption that the open reading frame identified in the DNA sequence is indeed the coding sequence for gX. The gX mRNA has not been characterized in detail, but Northern blot analysis indicates an mRNA of 1.6 kb (K. R. Marotti, unpublished data) as predicted from the location of the consensus transcription initiation and termination sequences. The reading frame expressed in p60-11 was determined by DNA sequencing to insert the amino acids coded by nucleotides 778 to 1341 into the hybrid protein. Since this amino acid sequence raised antibodies that reacted with gX, the identified reading frame certainly codes for amino acids in the gX sequence. Our interpretation of the DNA sequence predicts that the gX mRNA is not spliced. Although this would be rare for mammalian mRNAs in general, unspliced mRNAs seem to be common in the α-herpesviruses. Most herpes simplex virus RNAs are not spliced (39), and apparently many varicella-zoster virus genes also code for unspliced RNAs (7).

While this manuscript was in preparation, Wathen and Wathen (41) reported that the gene for a 50-kd glycoprotein which they call gp50 maps at least partially within a *Bam*HI-*Sal*I fragment within *Bam*HI 7. That *Bam*HI-*Sal*I fragment contains only an additional 400 bases beyond the *Bam*HI-*Pvu*II fragment sequenced in Fig. 6. The location of the genes for these two glycoprotein genes within the same small region raises the interesting possibilities that two glycoprotein genes map very close to each other, or that the glycoproteins are somehow related.

ACKNOWLEDGMENTS

We thank J. A. Wolcott and C. C. Marchioli for helpful comments on the manuscript, and K. J. Hiestand for typing the manuscript and figures. G.W.L. was involved in this work while a student at Kalamazoo College.

LITERATURE CITED

1. **Ben-Porat, T., and A. S. Kaplan.** 1982. The chemical composition of herpes simplex and pseudorabies viruses. Virology

16:261–266.
2. **Ben-Porat, T., and A. S. Kaplan.** 1970. Synthesis of proteins in cells infected with herpesvirus. V. Viral glycoproteins. Virology **41**:265–273.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an acid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. U.S.A. **80**:3963–3965.
4. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
5. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
6. **Courtney, R. J.** 1984. Virus-specific components of herpes simplex virus involved in the immune system p. 33–44. *In* B. T. Rouse and C. Lopez (ed.), Immunobiology of herpes simplex virus infection. CRC Press, Boca Raton, Fla.
7. **Davison, A. J.** 1983. DNA sequence of the Us component of the varicella-zoster virus genome. EMBO J. **2**:2203–2209.
8. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell **32**:695–706.
9. **Everett, R. D.** 1984. A detailed analysis of an HSV-1 early promoter: sequences involved in trans-activation by viral immediate early gene products are not early gene specific. Nucleic Acids Res. **12**:3037–3056.
10. **Everett, R. D., D. Baty, and P. Chambon.** 1983. The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. **11**:2447–2464.
11. **Feldman, L., F. J. Rixon, J. Jean, T. Ben-Porat, and A. S. Kaplan.** 1979. Transcription of the genome of pseudorabies virus (a herpesvirus) is strictly controlled. Virology **97**:316–327.
12. **Friedman, H. M., G. H. Cohen, R. J. Eisenberg, C. A. Seidel, and D. B. Cines.** 1984. Glycoprotein C of herpes simplex virus I acts as a receptor for the C3b complement component on infected cells. Nature (London) **309**:633–635.
13. **Gray, M. R., H. V. Colot, L. Guarerte, and M. Rosbash.** 1982. Open reading frame cloning: identification, cloning, and expression of open reading frame DNA. 1982. Proc. Natl. Acad. Sci. U.S.A. **79**:6598–6602.
14. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
15. **Katz, F. N., J. E. Rothman, V. R. Lingappa, G. Blobel, and H. F. Lodish.** 1977. Membrane assembly *in vitro*: synthesis, glycosylation, and asymmetric insertion of a transmembrane protein. Proc. Natl. Acad. Sci. U.S.A. **74**:3278–3282.
16. **Mackem, S., and B. Roizman.** 1980. Regulation of herpesvirus macromolecular synthesis: transcription-initiation sites and domains of α genes. Proc. Natl. Acad. Sci. U.S.A. **77**:7122–7126.
17. **Mackem, S., and B. Roizman.** 1982. Structural features of the herpes simplex virus α gene 4, 0, and 27 promoter-regulatory sequences which confer α regulation on chimeric thymidine kinase genes. J. Virol. **44**:939–949.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning—a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Matthews, J. T., G. H. Cohen, and R. J. Eisenberg.** 1983. Synthesis and processing of glycoprotein D of herpes simplex virus types 1 and 2 in an in vitro system. J. Virol. **48**:521–533.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **McKnight, S. L.** 1982. Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell **31**:355–365.
22. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
23. **Morse, L., L. Pereira, B. Roizman, and P. A. Schaffer.** 1978. Anatomy of herpes simplex virus (HSV) DNA. X. Mapping of viral genes by analysis of polypeptides and functions specified by HSV-1 × HSV-2 recombinants. J. Virol. **26**:389–410.
24. **Mulligan, R. C., and P. Berg.** 1981. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoriboxyltransferase. Proc. Natl. Acad. Sci. U.S.A. **78**:2072–2076.
25. **Nichols, B. P., and C. Yanofsky.** 1983. Plasmids containing the *trp* promoters of *Escherichia coli* and *Serratia marcescens* and their use in expressing cloned genes. Methods Enzymol. **101**:155–164.
26. **Norrild, B., H. Emmertsen, H. J. Krebs, and B. Pedersen.** 1984. Antibody-dependent immune mechanisms and herpes simplex virus infections, p. 91–105. *In* B. T. Rouse and C. Lopez (ed.), Immunobiology of herpes simplex virus infection. CRC Press, Boca Raton, Fla.
27. **Norrild, B., and B. F. Vestergaard.** 1979. Immunoelectrophoretic identification and purification of herpes simplex virus antigens released from infected cells in tissue culture. Intervirology **11**:104–110.
28. **Para, M. F., R. B. Baucke, and P. G. Spear.** 1980. Immunoglobulin G(Fc)-binding receptors on virions of herpes simplex virus type 1 and transfer of these receptors to the cell surface by infection. J. Virol. **34**:512–520.
29. **Paterson, B. M., and B. E. Roberts.** 1981. Structural gene identification utilizing eukaryotic cell-free translational systems, p. 417–437. *In* J. G. Chirikjian and T. S. Papas (ed.), Gene amplification and analysis, vol. 2 Elsevier/North-Holland Publishing Co., New York.
30. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
31. **Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey.** 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. U.S.A. **79**:4298–4302.
32. **Post, L. E., A. J. Conley, E. S. Mocarski, and B. Roizman.** 1980. Cloning of reiterated and nonreiterated herpes simplex I sequences as BamHI fragments. Proc. Natl. Acad. Sci. U.S.A. **77**:4201–4205.
33. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′-Non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
34. **Randall, R. E., and R. W. Honess.** 1980. Proteins of herpesvirus saimiri: identification of two virus polypeptides released into the culture medium of productively infected cells. J. Gen. Virol. **51**:445–449.
35. **Randall, R. E., R. A. Killington, and D. H. Watson.** 1980. Glycoproteins with type common and type specific antigenic sites excreted from cells infected with herpes simplex virus types 1 and 2. J. Gen. Virol. **48**:297–310.
36. **Rouse, B. T.** 1984. Cell-mediated immune mechanisms, p. 107–120. *In* B. T. Rouse and C. Lopez (ed.), Immunobiology of herpes simplex virus infection. CRC Press, Boca Ratn, Fla.
37. **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.
38. **Van Zaane, D., J. M. Brinkhof, F. Westenbrink, and A. L. Gielkens.** 1982. Molecular-biological characterization of Marek's disease virus. 1. Identification of virus-specific polypeptides in infected cells. Virology **121**:116–132.
39. **Wagner, E. K.** 1983. Transcription patterns in HSV Infections. Adv. Viral Oncol. **3**:239–270.
40. **Walboomers, J. M., and J. ter Schegget.** 1976. A new method for the isolation of herpes simplex virus type 2 DNA. Virology **74**:256–258.
41. **Wathen, M. W., and L. M. K. Wathen.** 1984. Isolation, characterization, and physical mapping of a pseudorabies virus mutant containing antigenically altered gp50. J. Virol. **51**:57–62.
42. **Weber, K., and M. Osborn.** 1975. Proteins and sodium dodecyl sulfate: molecular weight determinations on polyacrylamide

gels and related procedures, p. 179–223. *In* H. Neurath and R. L. Hall (ed.), The proteins, vol. I. Academic Press, Inc., New York.

43. **Weinstock, G. M., C. apRhys, M. L. Berman, B. Hampor, D. Jackson, T. J. Silhavy, J. Weisemann, and M. Zweig.** 1983. Open reading frame expression vectors: a general method for antigen production in *E. coli* using protein fusions to β-galactosidase. Proc. Natl. Acad. Sci. U.S.A. **80**:4432–4436.
44. **Whitton, J. L., and J. B. Clements.** 1984. Replication origins and a sequence involved in coordinate induction of the immediate-early gene family are conserved in an intergenic region of herpes simplex virus. Nucleic Acids Res. **12**:2061–2079.