Journal of Virology, May 1986, p. 339–347
0022-538X/86/050339-09$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 58, No. 2

Characterization of a Pseudorabies Virus Glycoprotein Gene with Homology to Herpes Simplex Virus Type 1 and Type 2 Glycoprotein C

A. K. ROBBINS,† R. J. WATSON,‡ M. E. WHEALY,† W. W. HAYS, and L. W. ENQUIST†*
Molecular Genetics, Inc., Minnetonka, Minnesota 55343

Received 30 September 1985/Accepted 16 January 1986

A pseudorabies virus (Becker strain) glycoprotein gene was located in the U_L region at map position 0.40. The gene was identified by using open reading frame *Escherichia coli* plasmid expression vectors and specific antibody reagents. A 1.55-kilobase unspliced transcript from the gene was detected in pseudorabies virus-infected tissue culture cells. The DNA sequence revealed a single open reading frame of 1,437 base pairs encoding 479 amino acids. The predicted primary translation product has a molecular weight of 50,860 and contains features of a typical herpesvirus glycoprotein. An *E. coli* expression plasmid was constructed that contained essentially all of the open reading frame for this gene. Antibodies raised in rabbits against the protein expressed in bacteria by this plasmid immunoprecipitated pseudorabies virus-specific glycoproteins of 92,000 and 74,000 daltons from infected cell extracts. It is likely that these two forms represent different glycosylation states of the protein.

Pseudorabies virus (PRV; Suid herpesvirus 1) is a naturally occurring herpesvirus of swine. The PRV genome consists of a linear, double-stranded DNA duplex (approximately 90×10^6 daltons) with an estimated coding capacity of 100 to 200 genes. The genome contains a short unique sequence (U_S) of 6×10^6 daltons bracketed by an inverted repeat sequence of 10×10^6 daltons. The remainder of the molecule is the long unique region (U_L) that is not bracketed by inverted repeat regions. Of interest for this report are the genes encoding the envelope or structural glycoproteins. After PRV infection of tissue culture cells, at least five groups of structural glycoproteins (1, 14) and a secreted glycoprotein (1, 25) are synthesized. The structural glycoproteins are found embedded in the nuclear and cellular membranes of infected cells as well as on the surface of the mature, enveloped virions. These PRV glycoproteins play a pivotal role in the life cycle of the virus and are likely to be the major antigens that interact with the host immune system eliciting both humoral and cell-mediated immune responses (reviewed in references 9, 23, 28).

In general, the role of individual PRV glycoproteins in the virus life cycle or in viral pathogenicity is poorly understood. Only recently have genome locations of PRV-encoded glycoproteins begun to be described. A 90,000-dalton nonstructural PRV glycoprotein secreted into the medium of infected cells was reported to be encoded by the U_S region of the PRV genome (25). Similarly, the tentative locations in the U_S region of genes encoding PRV glycoproteins of 130,000 and 50,000 daltons have been described (22, 31). A. K. Robbins et al. (Herpesvirus Meeting, 1985, Ann Arbor, Mich., p. 130) reported the localization of the gII glycoprotein gene in the U_L region at map position 0.1. We have previously localized a PRV glycoprotein gene in the U_L region through the use of *Escherichia coli* expression plasmid libraries containing random fragments of PRV DNA (27).

We now report the further characterization of this PRV glycoprotein gene found in the U_L region. Previously, we described two different expression plasmids, designated pDPR7 and pDPR123, carrying adjacent but nonoverlapping PRV DNA segments that specified PRV–β-galactosidase fusion proteins. Antisera produced in rabbits against either fusion protein immunoprecipitated the same two, apparently related, PRV-encoded glycoproteins of 92,000 and 74,000 apparent molecular weight. Our hypothesis was that the PRV DNA segments cloned in either expression plasmid represented different regions of the same gene. In this report, we confirm this hypothesis and describe the detailed mapping, mRNA characterization, and complete nucleotide sequence of this PRV glycoprotein gene.

MATERIALS AND METHODS

Animal cells and virus. The swine kidney cell line PK15 and the Becker strain of PRV have been described previously (27).

***E. coli* strains.** Strain MC1000 has the genotype *araD139* Δ(*araABC-leu*)*7679 galU galK* Δ(*lac*)X74 *rpsL thi*. Strain NF1829 is MC1000 carrying an F′ plasmid with the *lacI*^{q} mutation which results in overproduction of the Lac repressor. The F′ also carries the *lac* operon with a Tn*5* insertion in *lacZ*. Both strains were obtained from T. Silhavy.

Bacterial plasmids. All plasmids were constructed by standard recombinant DNA techniques. Plasmid polink26 is a derivative of pBR328 (10). It was constructed by replacing the 1,850-base-pair (bp) *Sal*I-*Eco*RI fragment with a synthetic 65-bp linker containing multiple restriction endonuclease cleavage sites. Plasmids pDPR7 and pDPR123 express open reading frame segments of PRV DNA fused to the *lacZ* gene under control of the *E. coli lac* promoter. These β-galactosidase fusion vectors have been described elsewhere (27). Plasmid pPRV49 contains the PRV *Bam*HI 2 fragment inserted in the *Bam*HI site of pBR322. Plasmid p7-123 is a subclone of pPRV49, containing a 2.5-kilobase

* Corresponding author.
† Present address: DuPont Experimental Station, Wilmington, DE 19898.
‡ Present address: ICRF Laboratories, St. Bartholomew's Hospital, London EC1A 7BE, England.

(kb) PRV *SphI-BamHI* fragment replacing the *SphI-BamHI* fragment of pBR322. Plasmid pK64 contains the PRV *KpnI* J fragment cloned into the *KpnI* site of polink26. Plasmid p7-123AM is described in detail below.

Construction of expression plasmid p7-123AM. The basic expression vector, ptc412, is derived from pJS413 (27). In this vector, the first 23 amino acids of the bacteriophage lambda *cro* gene are used as a "leader" for efficient translation. The DNA encoding this leader is joined in phase by a synthetic polylinker sequence containing several restriction enzyme cleavage sites to a sequence encoding an *lacI-lacZ* hybrid gene. If a sequence containing an open reading frame is inserted at the polylinker site in frame with the *cro* leader and in frame with the *lacI-lacZ* sequences, a hybrid gene will result encoding a *cro*–open reading frame–β-galactosidase fusion protein. This protein is expected to have β-galactosidase activity. To make ptc412, the *lac* promoter was replaced with the stronger *tac* promoter obtained from plasmid pDR540 (P-L Biochemicals, Inc., Milwaukee, Wis.). In addition, the polylinker between the DNA encoding Cro and β-galactosidase of pJS413 was replaced with a synthetic DNA linker containing *BglII*, *HindIII*, *SmaI*, and *BamHI* restriction enzyme cleavage sites.

p7-123AM is designed to express either the Cro-PRV fusion protein alone or both Cro-PRV and Cro–PRV–β-galactosidase fusion proteins concomitantly. This is accomplished by the insertion of a synthetic DNA sequence containing an amber nonsense codon at the PRV-*lacZ* DNA junction (see Fig. 5). When such a plasmid is carried by a wild-type cell, translation terminates at the nonsense codon, resulting in the production of only the Cro-PRV fusion protein. However, when the plasmid is carried by a cell with a nonsense suppressor mutation, the nonsense codon is translated, and the entire Cro–PRV–β-galactosidase fusion protein is made. In addition, however, the Cro-PRV fusion protein is also produced because suppression of the nonsense codon is less than 100% efficient. This concomitant production of both fusion proteins has proven to be of value for isolation of the Cro-PRV fusion protein since it will coaggregate with the Cro–PRV–β-galactosidase protein in large intracellular inclusions, simplifying purification (36). In the case of p7-123AM, concomitant expression of both fusion proteins was not necessary to stabilize the Cro-PRV fusion protein, presumably because the Cro-PRV fusion protein itself forms insoluble aggregates.

The construction of p7-123AM was accomplished in two steps. First a Cro–PRV–β-galactosidase protein fusion was made; second, a DNA linker containing an in-frame nonsense codon was inserted at the PRV-*lacZ* junction.

A 2.5-kb *SphI-BamHI* fragment of PRV genomic DNA (located from 0.0 to 2.5 kb in Fig. 1C) was cloned in pBR322. This plasmid, designated p7-123, carries the PRV gene of interest. p7-123 DNA was digested with *NcoI*, and the cohesive ends were filled in by using the Klenow fragment of DNA polymerase I and all four deoxyribonucleotides. The DNA was then digested with *BamHI*, and a PRV DNA fragment containing 1,380 bp of the putative PRV glycoprotein-coding sequence was isolated. By DNA sequence analysis (see Fig. 4), this fragment contains the coding sequence from amino acid 1 through 460 or about 96% of the gene. This 1.38-kb PRV DNA fragment was then inserted into the expression vector ptc412 at the *SmaI-BamHI* sites between the *cro* and *lacZ* gene sequences. Plasmids resulting from this ligation were introduced into NF1829 by $CaCl_2$ transformation followed by selection for ampicillin resistance. A plasmid designated ptacNB expressed a Cro–PRV–β-galactosidase fusion protein and was subsequently shown to have the PRV DNA in the correct orientation.

Plasmid ptacNB DNA was then cleaved at the unique *BamHI* restriction endonuclease site located at the junction between the PRV gene and the *lacZ* gene. The cleaved plasmid was religated by using T4 DNA ligase in the presence of the DNA linker:

This linker contains an amber nonsense codon, TAG. The reading frame is indicated by spacing. The linker was synthesized as described by Chow et al. (6). Introduction of the linker sequence destroyed the *BamHI* site but introduced a *BglII* and *XbaI* site as well as the in-frame nonsense codon. The structure of the resulting plasmid, designated p7-123AM, was verified by restriction enzyme mapping and nucleotide sequence of the fusion joints (see Fig. 5). Immunological analysis of the fusion protein produced by p7-123AM is described in Results.

Production and isolation of Cro-PRV fusion proteins from *E. coli*. The production and isolation methods were essentially those described by Watson et al. (36). The Cro-PRV fusion protein specified by p7-123AM was produced at approximately 3% of total cellular *E. coli* protein as determined by Coomassie brilliant blue staining of sodium dodecyl sulfate-polyacrylamide gels (see Fig. 6).

Northern blot analysis of PRV RNA. Total cytoplasmic RNA was extracted from uninfected or PRV-infected cells at various time points as described previously (33) and fractionated by electrophoresis on a 1% agarose gel containing 2 M formaldehyde–0.02 M morpholinepropanesulfonic acid–5 mM sodium acetate–1 mM EDTA, pH 7 (20). The RNA was subsequently transferred to a nitrocellulose membrane (30) and hybridized at 42°C for 16 h with DNA probes labeled with ^{32}P by nick translation (26). The hybridization solution contained 50% formamide, 1 M sodium chloride, 50 mM sodium phosphate, 5 mM EDTA, 0.1% sodium dodecyl sulfate, 2× Denhardt solution (0.04% each of Ficoll 400 [Pharmacia Fine Chemicals], bovine serum albumin, and polyvinylpyrollidone [Sigma Chemical Co.]), 400 μg of yeast tRNA per ml, and 250 μg of sonicated calf thymus DNA per ml. After hybridization, the membrane was washed four times for 30 min each at 65°C with 0.25× SSC (1× SSC contains 0.15 M sodium chloride and 0.015 M sodium citrate, pH 7.0)–0.1% sodium dodecyl sulfate before the membrane was analyzed by autoradiography.

S1 nuclease and *ExoVII* analysis of RNA. The protocols of Berk and Sharp (2) were followed by using 5′ or 3′ end-labeled fragments as probes (Fig. 1D, E, and F). Probes were labeled at their 5′ or 3′ ends as described previously (3).

DNA sequencing. The method of Maxam and Gilbert (21) was used for DNA sequence analysis.

Preparation of cell extracts and immunoprecipitations. The techniques for making cell extracts for immunological analysis were described previously (27).

Preparation of Cro-PRV fusion proteins expressed in *E. coli*. *E. coli* cells containing the plasmid p7-123AM were grown overnight at 37°C with shaking in L-broth containing 100 μg of ampicillin per ml. The cells were subsequently diluted 1:50 in similar fresh medium and shaken for 4 h at 37°C. The expression of the Cro-PRV fusion protein was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to 1 mM final concentration. At 4 h after IPTG addition, the induced cells were collected by centrifugation, the supernatant was decanted, and the cell pellets were

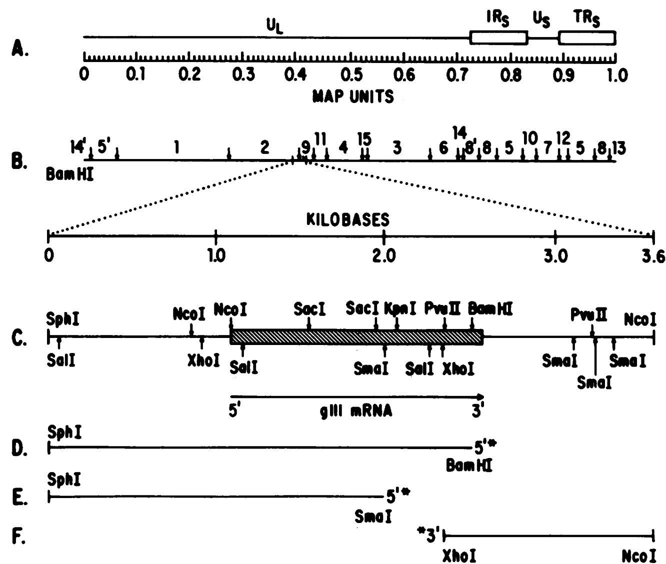

FIG. 1. Map of the PRV genome indicating relevant sites and DNA fragments used to study the U_L glycoprotein gene. (A) Genome of PRV depicting the U_L and U_S regions as well as the inverted and terminal repeat sequences bracketing U_S. (IR_S and TR_S, respectively). The genome is divided in map units as indicated. (B) *Bam*HI restriction enzyme map for the PRV genome. The region encompassing *Bam*HI fragments 2 and 9 is expanded and detailed in (C) as a 3.6-kb *Sph*I-*Nco*I fragment with relevant restriction enzyme sites noted. The shaded box corresponds to the gIII glycoprotein gene whose sequence is given in Fig. 4. The mRNA transcript for this gene and its orientation is given below the line. (D, E, and F) Specific end-labeled probes used to establish the 5′ and 3′ mRNA termini. The unique end label for each probe is indicated by an asterisk.

frozen at −70°C. The Cro-PRV fusion protein is produced as an insoluble aggregate or inclusion body. Purification of the aggregated protein was done essentially as described by Watson et al. (36).

Production of rabbit antisera. A sample of Cro-PRV fusion protein suspension was adjusted to 50 mM NaOH and incubated at 60°C for 10 min. The suspension was then neutralized with 1 M hydrochloric acid and then emulsified with complete Freund adjuvant. Approximately 250 μg of emulsified fusion protein was injected intramuscularly into New Zealand White rabbits. Three weeks after the first injection, the rabbits were given booster injections with 250 μg of protein in incomplete Freund adjuvant. One week after the second injection, the rabbits were bled from the marginal ear vein, and the antisera were used for immunological studies.

RESULTS

Previously, we had suggested that the PRV DNA sequences represented in pDPR7 and pDPR123 were likely to come from a single PRV glycoprotein gene located in the U_L region (27). These plasmids carried DNA segments of 495 and 363 bp, respectively, that hybridized to the *Bam*HI 2 fragment of PRV cloned in pPRV49. Further analyses indicated that the cloned open reading frame segments were adjacent, but not overlapping. To facilitate a more detailed study, subclones of this region were constructed with the vectors pBR322 and polink26 (see Materials and Methods). The results of restriction enzyme analysis of these plasmids with the enzymes *Nco*I, *Xho*I, *Sal*I, *Sac*I, *Sma*I, *Kpn*I, *Sph*I, *Pvu*II, and *Bam*HI are summarized in Fig. 1. The region of specific hybridization of both pDPR7 and pDPR123 DNA is contained within a 3.6-kb *Sph*I-to-*Nco*I fragment of PRV DNA.

Analysis of RNA from PRV-infected cells. If the PRV DNA carried by pDPR7 and pDPR123 indeed is derived from the same gene, then hybridization analysis using either plasmid as a probe should identify the same characteristic RNA species. RNA extracted from uninfected cells and cells infected with PRV for 4, 8, and 16 h was fractionated by electrophoresis on an agarose-formaldehyde gel. The separated RNA was subsequently transferred to nitrocellulose paper for hybridization. Separate nitrocellulose filters on which the RNA had been immobilized were then incubated under appropriate hybridization conditions (see Materials and Methods) with plasmid DNA from either pDPR7 or pDPR123 that had been labeled with ^{32}P by nick translation. The results are presented in Fig. 2 for pDPR7. Identical results were found with pDPR123 (data not shown). Both plasmids hybridized to a single RNA species of approximately 1.55 kb. No hybridization of plasmid DNA was observed to RNA extracted from uninfected cells. These data are consistent with the hypothesis that both pDPR7 and pDPR123 carry segments of the same gene. The detailed analysis of the RNA is presented below.

S1 nuclease protection of labeled DNA by hybridization to RNA. The S1 endonuclease and *Exo*VII exonuclease procedures of Berk and Sharp (2, 3) were used to determine the direction of transcription, the approximate location of the 5′ and 3′ ends of the RNA transcript defined by the previous hybridization experiments, and the presence or absence of RNA splicing.

Plasmid p7-123 carries the *Sph*I-*Bam*HI fragment as previously described. The 1.6-kb *Xho*I fragment that contains the DNA carried by both pDPR7 and pDPR123 was labeled at the 5′ or 3′ end as described previously (34). The 5′ or 3′

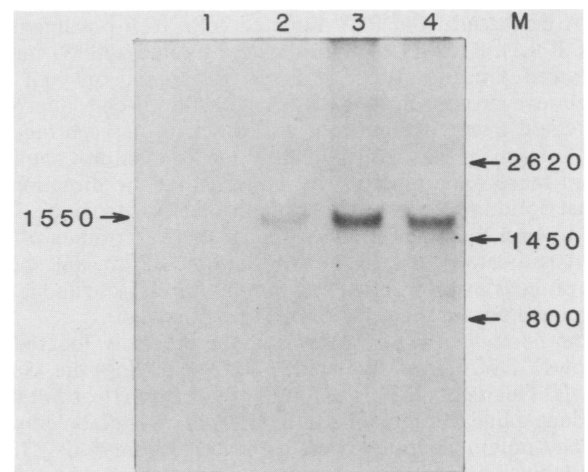

FIG. 2. Northern blot analysis of RNA from PRV-infected PK15 cells. Total cytoplasmic RNA was extracted from uninfected and PRV-infected cells at various time points postinfection and fractionated on an agarose-formaldehyde gel. The RNA was transferred to a nitrocellulose membrane and hybridized with nick-translated DNA from plasmid pDPR7. Lanes: 1, RNA extracted from uninfected cells; 2, 3, and 4, RNA extracted from cells at 4, 8, and 16 h, respectively, after infection with virus. The hybridized blots were washed and autoradiographed at −70°C. The size of the fragments hybridizing to the radioactive probe was estimated by using ^{32}P-labeled DNA markers as indicated.

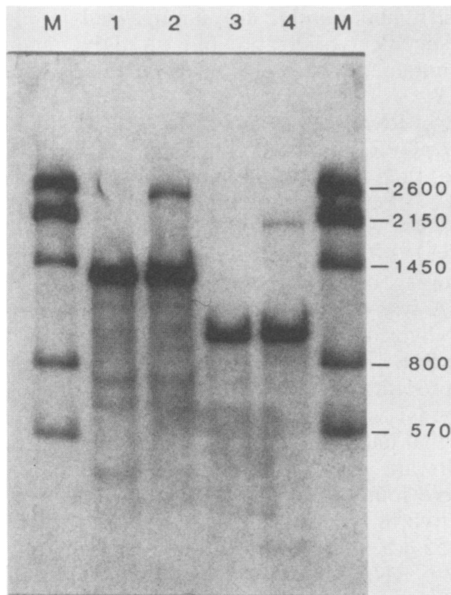

FIG. 3. Nuclease S1 mapping of glycoprotein gIII-specific RNA. DNA probes of 2.1 and 2.5 kb uniquely labeled at a 5′ end as indicated in Fig. 1D and E were hybridized with total cytoplasmic RNA extracted from either uninfected or virus-infected cells. After hybridization, RNA-DNA hybrids were digested with either S1 endonuclease or *Exo*VII nuclease and fractionated by electrophoresis on an alkaline agarose gel as described in Materials and Methods. Lanes: 1, hybrids formed with the 2.5-kb probe digested with S1 nuclease; 2, hybrids formed with the 2.5-kb probe digested with *Exo*VII; 3, hybrids formed with the 2.1-kb probe digested with S1 nuclease; 4, hybrids formed with the 2.1-kb probe digested with *Exo*VII. The sizes of ^{32}P-labeled DNA markers are indicated. RNA extracted from uninfected cells did not protect either DNA probe from S1 or *Exo*VII digestion (data not shown).

end-labeled DNA was then hybridized to total cytoplasmic RNA extracted from PRV-infected cells 16 h postinfection (late RNA). RNA-DNA hybrids were treated with S1 nuclease and fractionated by electrophoresis on a 1.4% alkaline–agarose gel. Only DNA with the 5′-end label was protected from S1 digestion, and the size of the protected DNA fragment was approximately 1.2 kb (data not shown). From these experiments, we inferred that the direction of transcription of the gene encoding this RNA is from the *Sph*I site toward the *Bam*HI site (Fig. 1), the 5′ terminus of the RNA is located close to the *Nco*I site at 1.2 kb, and the 3′ end of the transcript is just beyond the *Bam*HI site and is not contained within the *Sph*I-*Bam*HI fragment.

The 5′ end of the RNA was more precisely located as follows. DNA from plasmid p7-123 carrying the *Sph*I-*Bam*HI fragment (Fig. 1) cut with either *Bam*HI or *Sma*I to produce a linear fragment. Both 5′ termini were labeled with ^{32}P and polynucleotide kinase as described previously (34). The labeled DNA was subsequently cut with *Sph*I to yield fragments of 2.5 kb (*Sph*I to *Bam*HI) and 2.1 kb (*Sph*I to *Sma*I). Each labeled fragment was then hybridized to total cytoplasmic RNA from PRV-infected cells as described in Materials and Methods. The RNA-DNA hybrids formed were digested with S1 nuclease or *Exo*VII exonuclease and then fractionated by electrophoresis on a 1.4% alkaline–agarose gel. The results (Fig. 3) indicated that a 1.4-kb fragment was protected with the 2.5-kb *Sph*I-*Bam*HI probe and a 0.95-kb fragment was protected with the 2.1-kb *Sph*I-*Sma*I probe. From these results, we conclude that the 5′ end of the RNA is situated 1.4 kb from the *Bam*HI site and 0.95 kb from the *Sma*I site close to the *Nco*I site at 1.1 kb (Fig. 1). The sizes of the protected fragments when *Exo*VII was used to digest the RNA-DNA hybrids were identical to those seen for S1 digestion (Fig. 3). Such results are consistent with the absence of intervening sequences in this RNA.

A more precise location of the 3′ end was done as follows. Plasmid pK64 carries the 5.7-kb PRV *Kpn*I J fragment that overlaps the PRV *Bam*HI 2 fragment, which, based on the previous data (27), should encode the 3′ terminus of the transcript. Plasmid pK64 was cleaved with *Xho*I, and the 3′ termini were labeled with ^{32}P as described previously (34). The labeled DNA was subsequently digested with *Nco*I to yield an *Xho*I-*Nco*I fragment of 1.1 kb (Fig. 1F). This fragment was then hybridized to total cytoplasmic RNA from PRV-infected cells as described previously (34). RNA-DNA hybrids were digested with S1 or *Exo*VII and analyzed as previously described. A 0.2-kb fragment was protected from either S1 or *Exo*VII nuclease digestion (data not shown). This establishes the location of the transcript's 3′ end at 0.2 kb from the *Xho*I site (Fig. 1C). In addition, these experiments further indicate that the transcript is not spliced. Preliminary S1 nuclease protection experiments indicate that the 5′ mRNA terminus is somewhat heterogenous and located between positions 180 and 189 (Fig. 4) (data not shown).

DNA sequence analysis. The entire nucleotide sequence (both strands) of the region was determined (Fig. 4). A single open reading frame of 1,437 bp was found that contained both the sequences cloned in pDPR7 (bases 413 through 907) and pDPR123 (bases 1127 through 1489) as well as those sequences defined by RNA-DNA hybridization (Fig. 3). The overall base composition of the open reading frame was 11.2% A, 13.6% T, 34.3% G, and 40.9% C. The codon usage for the 479 amino acids is given in Table 1. Given the high G+C content, it perhaps is not surprising that codon bias exists. For example, 23 codons are not used at all. It is remarkable that more than one-third of the amino acid content is composed of only three amino acids: valine (11.3%), alanine (14.4%), and proline (9.2%).

The DNA sequence for 196 bp upstream and 90 bp downstream of the coding sequence was determined (Fig. 4). The base composition of these regions was not markedly different from that of the open reading frame. The TATA box with the consensus sequence of 5′-TATAA/TAA/T-3′ is usually found 26 to 34 bp upstream of the site of initiation of transcription in higher eucaryotes (8). The sequence TTTT TAAAA (residues 151 through 159) may represent the TATA sequence for this gene. In addition, the sequence GCATTAAA (residues 111 through 118) may be analogous to the CAT box motif characteristic of many eucaryotic promoters (4, 8). The initiation codon of the gene is likely the ATG at residue 196. It is the first ATG in the transcript and is followed by an open reading frame of 478 codons. We note that there is a subsequent ATG codon in the same reading frame (residue 217). It has been predicted that efficient translation initiation in eucaryotes depends on a purine residue, usually an A, at position −3 relative to the initiating ATG codon (15). The −3 position of either ATG (position 196 or 217) is a G. Downstream of the ATG, a purine at position +4 (16) has been implicated in playing a role in the efficiency of translation. By this argument, the first ATG would be more efficient (+4 = G) than the second (+4 = C).

The proposed stop codon, TGA, is found at nucleotide 1,633. A potential polyadenylation signal, AATAAA, was present downstream from the coding sequence beginning at

nucleotide 1,666. This site would be consistent with the observed size of the transcript localized to this region.

The predicted amino acid sequence for the polypeptide encoded by this gene is also indicated in Fig. 4. The molecular weight of the primary translation product would be 50,860. The protein predicted from the DNA sequence has features in common with envelope glycoproteins from other herpesviruses (5, 11, 12, 24, 29, 32, 35). This can best be seen by using the hydropathic analysis of Kyte and Doolittle (17) and the empirically based secondary structure analysis of Garnier et al. (13). The first 22 amino acids are hydrophobic, with the exceptions of serine at position 3, arginine at position 6, and tyrosine at position 15. The hydrophobic core (amino acids 7 through 22) of the first 22 amino acids is predicted to be in an alpha-helical arrangement. This sequence could correspond to a signal peptide for membrane insertion and may well be removed during translation and transport. Amino acids present at positions 451 through 470 are also strongly hydrophobic and have the characteristics of a membrane-spanning alpha-helical region. The carboxy-terminal nine amino acids have a net basic charge and may function as a cytoplasmic anchor sequence. Eight potential glycosylation sites (Asn-X-Ser/Thr) are present in the region between the putative signal sequence and transmembrane sequence and are indicated in Fig. 4.

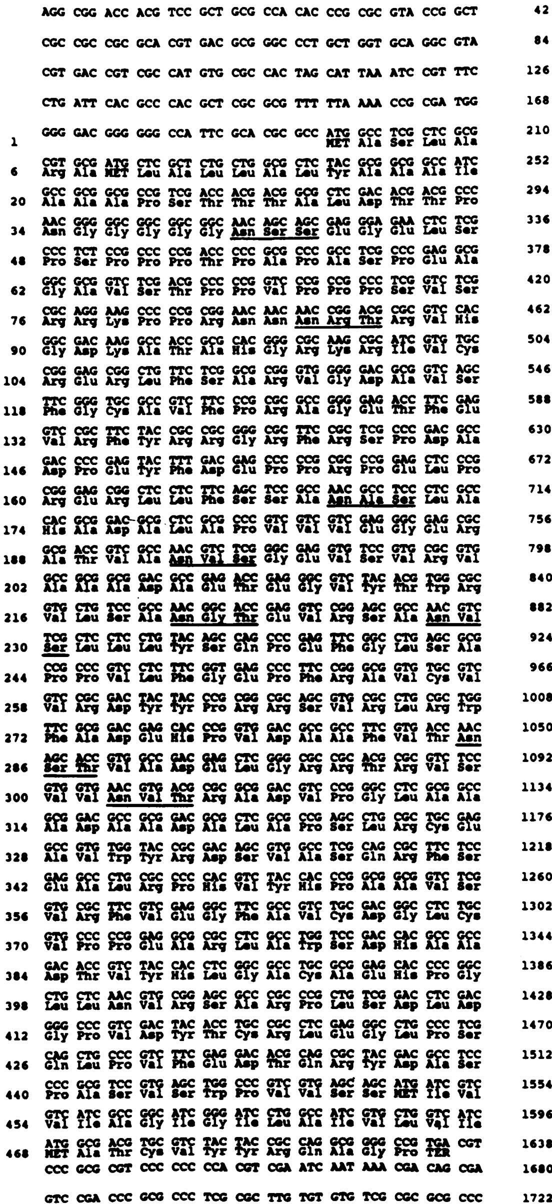

FIG. 4. DNA sequence of the PRV glycoprotein gIII gene. The nucleotide sequence was determined for both strands as described in Materials and Methods. The sequence represented here is that of the noncoding strand. Nucleotides are numbered at the right of the sequence, which is transcribed from left to right. The deduced amino acid sequence is indicated in the conventional three-letter code, and the amino acids are numbered at the left of the sequence. The underscored sequences are the Asn-X-Thr/Ser motif proposed to be N-linked glycosylation sites.

TABLE 1. Codon usage in the gIII gene of pseudorabies virus

Codon	Frequency of occurrence	% of total[a]	Codon	Frequency of occurrence	% of total[a]
TTT-Phe	1	0.2	TAT-Tyr	0	0.0
TTC-Phe	16	3.3	TAC-Tyr	14	2.9
TTA-Leu	0	0.0	TAA-TER	0	0.0
TTG-Leu	0	0.0	TAG-TER	0	0.0
CTT-Leu	0	0.0	CAT-His	0	0.0
CTC-Leu	22	4.6	CAC-His	9	1.9
CTA-Leu	0	0.0	CAA-Gln	0	0.0
CTG-Leu	15	3.1	CAG-Gln	5	1.0
ATT-Ile	0	0.0	AAT-Asn	0	0.0
ATC-Ile	8	1.7	AAC-Asn	12	2.5
ATA-Ile	0	0.0	AAA-Lys	0	0.0
ATG-MET	4	0.8	AAG-Lys	3	0.6
GTT-Val	0	0.0	GAT-Asp	0	0.0
GTC-Val	32	6.7	GAC-Asp	24	5.0
GTA-Val	0	0.0	GAA-Glu	1	0.2
GTG-Val	22	4.6	GAG-Glu	26	5.4
TCT-Ser	1	0.2	TGT-Cys	0	0.0
TCC-Ser	9	1.9	TGC-Cys	9	1.9
TCA-Ser	0	0.0	TGA-TER	0	0.0
TCG-Ser	15	3.1	TGG-Trp	5	1.0
CCT-Pro	0	0.0	CGT-Arg	1	0.2
CCC-Pro	24	5.0	CGC-Arg	32	6.7
CCA-Pro	0	0.0	CGA-Arg	0	0.0
CCG-Pro	20	4.2	CGG-Arg	11	2.3
ACT-Thr	0	0.0	AGT-Ser	0	0.0
ACC-Thr	11	2.3	AGC-Ser	15	3.1
ACA-Thr	0	0.0	AGA-Arg	0	0.0
ACG-Thr	11	2.3	AGG-Arg	1	0.2
GCT-Ala	1	0.2	GGT-Gly	1	0.2
GCC-Ala	33	6.9	GGC-Gly	18	3.8
GCA-Ala	0	0.0	GGA-Gly	1	0.2
GCG-Ala	36	7.5	GGG-Gly	10	2.1

[a] Total = 479 amino acids.

Expression of the open reading frame in *E. coli*. The identity of the protein encoded by this open reading frame was deduced by expression of the protein in *E. coli*, production of antisera against the bacterially produced protein, and by use of sera so obtained for immunoprecipitation of specific PRV polypeptides. We had previously expressed portions of this open reading frame in *E. coli* as β-

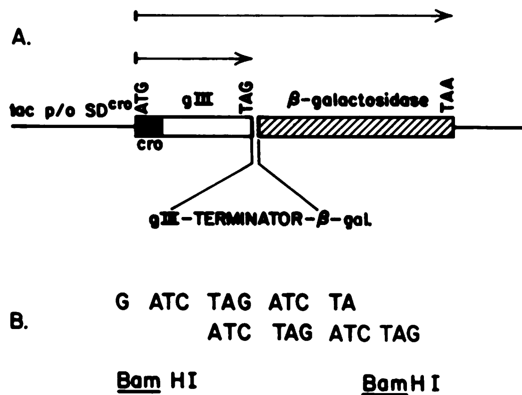

FIG. 5. Diagram of the fusion protein-coding region of expression plasmid p7-123AM. Details of construction of this plasmid are described in Materials and Methods. (A) tac p/o corresponds to the *tac* promoter-operator region used for transcription of the fusion protein gene. SD_{cro} indicates the Shine-Dalgarno region or ribosome-binding site for translation initiation taken from the bacteriophage lambda *cro* gene. The boxed-in region indicates the three parts of the fusion protein gene. The solid shading corresponds to the first 23 amino acids of the bacteriophage lambda *cro* gene (the ATG initiator codon is shown). Next, in the unshaded box, is the open reading frame from the glycoprotein gIII gene (the *Nco*I-to-*Bam*HI fragment from the hatched region in Fig. 1C) fused in frame with the upstream *cro* sequences. Next, a synthetic linker with *Bam*HI cohesive ends and an in-frame nonsense (amber) codon (sequence is given in panel B) is indicated in the space between the open box (viral sequences) and the hatched box (*lacZ* gene). This linker is in frame with both the viral DNA sequence and the *lacZ* sequences. The TAG and TAA translational terminators of the linker and *lacZ* are indicated above the boxes. The two arrows above the gene indicate the two types of protein expected to be translated from the mRNA. One protein, Cro-PRV, results from translation terminating at the synthetic TAG codon at the junction of viral and *lacZ* sequences. The other, Cro–PRV–β-galactosidase, results from translation terminating at the natural TAA codon at the end of *lacZ*. Only one of the three TAA *lacZ* terminator codons is indicated.

galactosidase fusion proteins from the expression plasmids pDPR7 and pDPR123. We had demonstrated that antisera made against either of these fusion proteins immunoprecipitated similar glucosamine-labeled PRV proteins of 92,000 and 74,000 apparent molecular weight. The problem with these antisera was that they only reacted with denatured protein; they did not react with native material. In addition, major portions of the protein, particularly the amino-terminal segments, were not represented in either fusion protein. A further complication was that the majority of the mass of each fusion protein was β-galactosidase. These shortcomings were eliminated as follows. By knowing the complete DNA sequence, we were able to express better than 96% of the open reading frame in *E. coli* as a Cro-PRV protein unfused to β-galactosidase as described by Watson et al. (36) and in Materials and Methods.

This plasmid, p7-123AM, is diagrammed in Fig. 5. The construction of this plasmid is described in Materials and Methods. The hybrid *trp-lac* or *tac* promoter used was regulated by the *lac* repressor, and IPTG was used as the inducer. Upon induction in the suppressor-negative host, NF1829, p7-123AM directs the production of only the Cro-PRV fusion protein (Fig. 6). The Cro-PRV fusion protein is produced in significant quantities and is found in large, insoluble aggregates. Purification of the Cro-PRV fusion protein is easily done by collecting the aggregates by centrifugation as described in Materials and Methods.

The physiology of *E. coli* carrying p7-123AM was markedly affected upon IPTG induction. Cell division rapidly stopped, although the Cro-PRV fusion protein continued to accumulate. The phenomenon was observed in both suppressor-positive and -negative strains. We have not investigated this phenomenon further to date, although similar phenomena have been observed previously (36).

Analysis of antisera directed toward the Cro-PRV fusion protein. The Cro-PRV fusion protein was partially purified from a culture of *E. coli* carrying the plasmid p7-123AM as described in Materials and Methods. The fusion protein was then injected into rabbits to raise antibodies as detailed in Materials and Methods. Extracts prepared from [^{3}H]glucosamine-labeled PRV-infected cells were reacted with anti-PRV serum as a control or antisera raised in rabbits against the Cro-PRV fusion protein. The anti-PRV serum specifically immunoprecipitated PRV glycoproteins of 110,000, 92,000, 74,000, and 55,000 apparent molecular weight as described previously (27). Rabbit antisera raised against the Cro-PRV fusion protein containing 96% of the open reading frame specifically immunoprecipitated two PRV glycoproteins of 92,000 and 74,000 apparent molecular weight (Fig. 7). These results are similar if not identical to those obtained previously with Cro–PRV–β-galactosidase fusion proteins representing small segments of the open reading frame, except that immunoprecipitations could be done without prior denaturation. Although it is possible that these two proteins result from differential processing or proteolytic degradation products, we favor the hypothesis that the

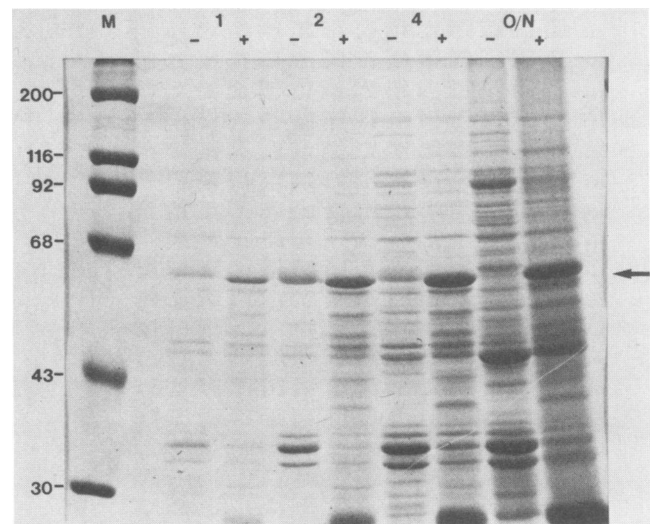

FIG. 6. Induction of fusion protein expressed by *E. coli* NF1829 carrying plasmid p7-123AM. Bacteria were grown at 37°C in L-broth containing 100 μg of ampicillin per ml. Synthesis of fusion protein was induced by the addition of IPTG to a final concentration of 1 mM. Samples (3 ml) of culture from uninduced (−) or induced (+) were harvested at 1 h, 2h, and 4 h and after overnight (O/N) incubation with inducer. The Cro-PRV fusion proteins were found in insoluble aggregates and were prepared as previously described (36). The aggregated proteins were fractionated by electrophoresis on a 10% polyacrylamide–sodium dodecyl sulfate gel followed by staining with Coomassie blue. The position of the Cro-PRV fusion protein is indicated by the arrow. The locations of molecular weight standards are indicated (×1,000).

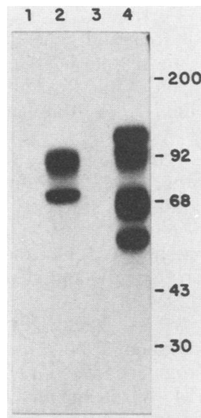

FIG. 7. Immunoprecipitation of PRV-specific proteins with antisera raised against the Cro-PRV fusion protein produced in *E. coli*. PRV-infected cells were labeled with [^{3}H]glucosamine, and cell extracts were reacted with either normal rabbit serum (lane 1), rabbit anti Cro-PRV serum (lane 2), normal goat serum (lane 3), or goat anti-PRV serum (lane 4). Immune complexes were collected by adsorption to *Staphylococcus aureus* and fractionated by electrophoresis on a 10% polyacrylamide–sodium dodecyl sulfate gel. ^{3}H-labeled polypeptides were detected by fluorography. The positions of molecular weight standards (×1,000) are indicated.

diffuse family of proteins migrating at about 92,000 apparent molecular weight represents highly glycosylated forms of the primary translation product. Similarly, the sharper band at 74,000 apparent molecular weight may be a less-glycosylated form of the same protein. A similar precursor-product relationship has been observed with herpes simplex virus type 1 glycoprotein C (7). We have observed that the 74,000-molecular-weight protein does not appear to be present in purified virions (data not shown). A test of this hypothesis will come from experiments that inhibit glycosylation in vivo, from in vitro translation of specific mRNA, or from removal of the sugar modifications by enzymatic treatment of purified proteins.

DISCUSSION

In this paper we expand our initial observations in which we located a PRV glycoprotein gene within the PRV genome (27). Since the two expression plasmids used to localize the new PRV gene carried different DNA segments that did not overlap, it was important to prove directly that they carried DNA from the same PRV gene. We have proven this unequivocally by DNA sequence analysis, mRNA analysis, and immunological studies of bacterially produced fusion proteins.

This PRV glycoprotein gene is defined by an open reading frame of 1,437 bases or 479 codons. The 1.55-kb mRNA transcript of this sequence is not spliced. The transcript was observed as early as 4 h postinfection and seemed to be present at increased levels at 8 h postinfection. Since the 4-h time point was the earliest examined, it may well be that the transcript is synthesized earlier in the cycle.

Analysis of the primary sequence and the predicted amino acid sequence indicates that the protein has a structure with some features in common with most herpesvirus glycoproteins. There is a hydrophobic sequence of about 22 amino

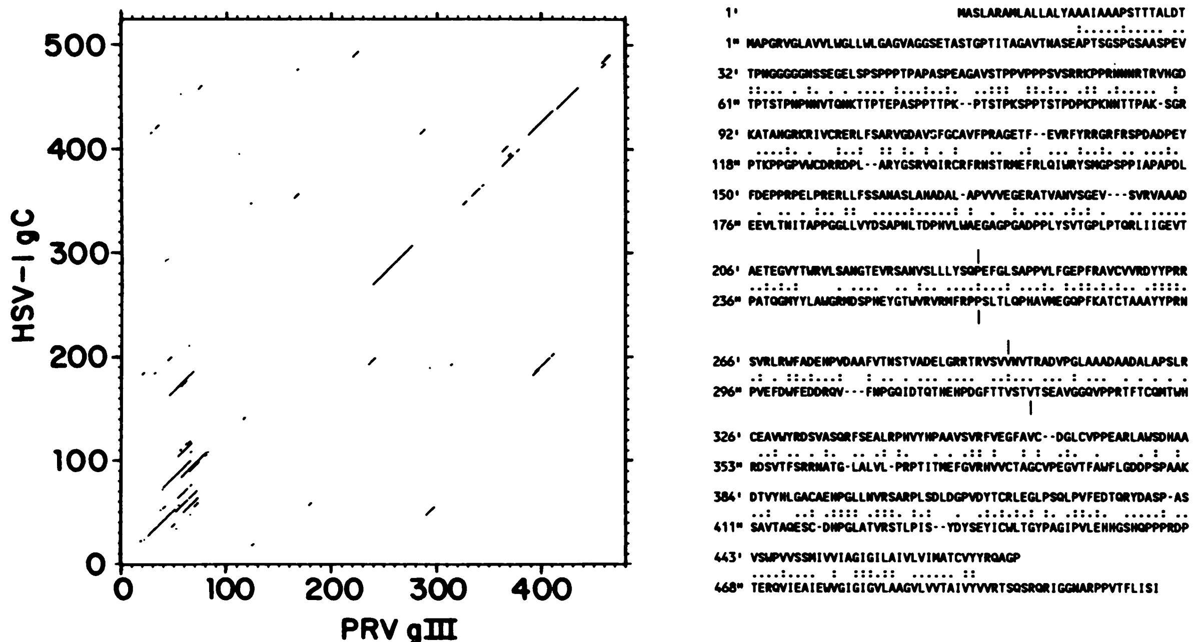

FIG. 8. Protein similarity between herpes simplex virus type 1 glycoprotein C and the PRV glycoprotein gIII. Shown to the left is a dot matrix study with the deduced amino acid sequence of herpes simplex virus type 1 glycoprotein C (13, 29) on the *Y*-axis and the deduced amino acid sequence of PRV gIII on the *X*-axis. The method was described previously (19). The window was 30 and the stringency was set at 8. Shown to the right is the printout from an FASTP homology search of the Dayhoff protein sequence library by the method of Lipman and Pearson (18). The ktup value was set at 1 for this search. The PRV gIII amino acid sequence in single-letter code begins on the first line and is numbered on the left as 1′. In a similar fashion, the herpes simplex virus type 1 glycoprotein C amino acid sequence begins on the second line and is numbered on the left as 1″. Identical amino acids (:), similar amino acids (.), and gaps introduced by the alignment algorithm (--) are indicated. In this alignment, there was 23.3% identity in an overlap of 466 amino acids.

acids at the amino terminus that may correspond to a signal peptide that should be removed during protein localization in the membranes of infected cells. Similarly, there is a hydrophobic sequence near the carboxy terminus that is predicted to span the membrane of the virus envelope or infected cells.

Computer analysis revealed no significant DNA homology to the known DNA sequences of other herpesvirus glycoproteins (data not shown). Homology at the amino acid level was also very minimal with the exception of glycoprotein C from herpes simplex virus type 1 and type 2. The homology presented in Fig. 8 was detected by using the FASTP homology search program of Lipman and Pearson (18), scanning the Dayhoff protein data base. For herpes simplex virus type 1 glycoprotein C, the 22.3% homology extends over 466 amino acids. The majority of matches are for similar, but not identical, amino acids. The similarity was significant as judged by a Monte Carlo shuffle test with the RDF program of Lipman and Pearson (18). The dot matrix homology analysis of Maizel and Lenk (19) also provided evidence of similarity (Fig. 8). The herpes simplex virus type 2 glycoprotein C gene (29) exhibited similar homology patterns (19.0% homology over 458 amino acids; data not shown). Although it may be that the PRV and herpes simplex virus proteins have evolved portions of a common primary protein structure, there is no evidence for immunological cross-reaction (unpublished observations). Further work is necessary to determine whether these proteins share any common functions and indeed are analogs.

Hampl et al. (14) used monoclonal antibodies to define specific PRV glycoproteins. The nomenclature proposed would number them with roman numerals gI, gII, gIII, gIV, and gV. We have attempted to use these reagents to define more precisely this new glycoprotein. Unfortunately these monoclonal antibodies did not react with the protein expressed in *E. coli*, presumably because they recognize conformational determinants not present in the bacterial product. However, using recombinant DNA techniques we have introduced defined mutations into the gene identified in this paper and have demonstrated appropriately altered glycoproteins that react with gIII-specific monoclonal antibodies (A. K. Robbins, M. E. Whealy, R. J. Watson, and L. W. Enquist, manuscript in preparation). Although our estimated molecular weights differ slightly from those of Hampl et al. (14), we suggest that the open reading frame at 0.4 map unit encodes the gIII glycoprotein gene.

ACKNOWLEDGMENTS

T. Ben-Porat provided us with valuable advice and reagents. T. Silhavy provided bacterial strains. Dennis Anderson provided excellent computer support. David Mount kindly provided computer software for DNA and protein analysis as well as advice for its use. We thank Thomas Kempe for synthesizing oligonucleotides. We also thank Walt Hagelstein and Dorothy Tinker for excellent assistance with art work. The Will Hop Inn and Lake Waconia supplied inspiration.

LITERATURE CITED

1. **Ben-Porat, T., and A. S. Kaplan.** 1970. Synthesis of proteins in cells infected with herpesviruses. V. Viral glycoproteins. Virology **41**:265–273.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNA by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
3. **Berk, A. J., and P. A. Sharp.** 1978. Structure of the adenovirus 2 early mRNAs. Cell **14**:695–711.
4. **Busslinger, M., R. Portmann, J. C. Irminger, and M. D. Birnsteil.** 1980. Ubiquitous and gene-specific regulatory 5′ sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res. **8**:957–965.
5. **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.
6. **Chow, F., T. Kempe, and G. Palm.** 1981. Synthesis of oligodeoxyribonucleotides on silica gel support. Nucleic Acids Res. **12**:2807–2817.
7. **Cohen, G. H., D. Long, and R. J. Eisenberg.** 1980. Synthesis and processing of glycoproteins gD and gC of herpes simplex virus type 1. J. Virol. **36**:429–439.
8. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
9. **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.
10. **Covarrubias, L., L. Cervantes, A. Covarrubias, X. Soberon, I. Vichido, A. Blanco, Y. M. Kupersztoch-Portnoy, and F. Bolivar.** 1981. Construction and characterization of new cloning vehicles. V. Mobilization and coding properties of pBR322 and several deletion derivatives including pBR327 and pBR328. Gene **13**:25–35.
11. **Dowbenko, D. J., and L. A. Lasky.** 1984. Extensive homology between the herpes simplex virus type 2 glycoprotein F gene and the herpes simplex virus type 1 glycoprotein C gene. J. Virol. **52**:154–163.
12. **Frink, R. J., R. Eisenberg, G. Cohen, and E. K. Wagner.** 1983. Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. J. Virol. **45**:634–647.
13. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implication of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
14. **Hampl, H., T. Ben-Porat, L. Ehrlicher, K. O. Habermehl, and A. S. Kaplan.** 1984. Characterization of the envelope proteins of pseudorabies virus. J. Virol. **52**:583–590.
15. **Kozak, M.** 1984. Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microbiol. Rev. **47**: 1–45.
16. **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.
17. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
18. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
19. **Maizel, J. V., and R. P. Lenk.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78**:7665–7669.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Maxam, A. M., and W. Gilbert.** 1977. Chemical sequencing of DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
22. **Meltenleiter, T. C., N. Lukacs, and H.-J. Rziha.** 1985. Mapping of the structural gene of pseudorabies virus glycoprotein A and identification of the two non-glycosylated precursor polypeptides. J. Virol. **53**:52–57.
23. **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.
24. **Pellett, P. E., K. G. Kousoulas, L. Pereira, and B. Roizman.** 1985. Anatomy of the herpes simplex virus type 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.

25. **Rea, T. J., J. G. Timmins, G. W. Long, and L. E. Post.** 1985. Mapping and sequence of the gene for the pseudorabies virus glycoprotein which accumulates in the medium of infected cells. J. Virol. **54**:21–29.
26. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
27. **Robbins, A. K., J. H. Weis, L. W. Enquist, and R. J. Watson.** 1984. Construction of *E. coli* expression plasmid libraries: localization of a pseudorabies virus glycoprotein gene. J. Mol. Appl. Gen. **2**:485–496.
28. **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 Raton, Fla.
29. **Swain, M. A., R. W. Peet, and D. A. Galloway.** 1985. Characterization of the gene encoding herpes simplex virus type 2 glycoprotein C and comparison with the type 1 counterpart. J. Virol. **53**:561–569.
30. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
31. **Wathan, M. W., and L. M. K. Wathan.** 1984. Isolation, characterization and physical mapping of a pseudorabies virus mutant containing antigenically altered gp50. J. Virol. **51**:57–62.
32. **Watson, R. J.** 1983. DNA sequence of the herpes simplex virus type 2 glycoprotein D gene. Gene **26**:307–312.
33. **Watson, R. J., C. M. Preston, and J. B. Clements.** 1979. Separation and characterization of herpes simplex virus type 1 immediate-early mRNA's. J. Virol. **31**:42–52.
34. **Watson, R. J., K. Umene, and L. W. Enquist.** 1981. Reiterated sequences within the intron of an immediate-early gene of herpes simplex virus type 1. Nucleic Acids Res. **9**:4189–4199.
35. **Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist.** 1982. Herpes simplex virus type 1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. Science **218**:381–384.
36. **Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist.** 1984. Bacterial synthesis of herpes simplex virus types 1 and 2 glycoprotein D antigens. J. Invest. Dermatol. **83**:102s–111s.