JOURNAL OF VIROLOGY, Oct. 1993, p. 5955–5961
0022-538X/93/105955-07$02.00/0

Vol. 67, No. 10

A 3′ Coterminal Gene Cluster in Pseudorabies Virus Contains Herpes Simplex Virus UL1, UL2, and UL3 Gene Homologs and a Unique UL3.5 Open Reading Frame

HANSI J. DEAN AND ANDREW K. CHEUNG*

Virology Swine Research Unit, National Animal Disease Center, U.S. Department of Agriculture Agricultural Research Service, P.O. Box 70, 2300 Dayton Avenue, Ames, Iowa 50010

Received 14 April 1993/Accepted 1 July 1993

We have determined the nucleotide sequence and transcription pattern of a group of open reading frames of pseudorabies virus (PRV), which are located at the right end of the *Bam*HI-G fragment from 0.664 to 0.695 map units in the unique long region of the genome. Nucleotide sequence analysis revealed four open reading frames. The first three correspond in genome location to the herpes simplex virus type 1 (HSV-1) open reading frames UL1, which codes for glycoprotein L (gL); UL2, which codes for a uracil-DNA glycosylase; and UL3, which codes for a polypeptide of unknown function. The fourth open reading frame, UL3.5, is not present in the HSV-1 genome. Northern (RNA) blot analysis with oligonucleotide and cDNA probes revealed four abundant mRNA species of 3.3, 2.7, 1.8, and 0.9 kb, which are likely to yield polypeptides encoded by the UL1, -2, -3, and -3.5 open reading frames, respectively. All four transcripts were of the early-late kinetic class, transcribed in the same direction, and 3′ coterminal. The UL2 and UL3 genes of PRV and HSV-1 have significant amino acid sequence homology, while the UL1 genes are positional homologs and the UL3.5 gene is unique to PRV.

Pseudorabies virus (PRV) is a member of the *Alphaherpesvirus* subfamily, which includes the human viruses herpes simplex virus (HSV) types 1 (HSV-1) and 2 (HSV-2) and varicella-zoster virus (VZV). The natural host for PRV is swine, although some other animal species are susceptible to infection as well (26). PRV is similar to HSV-1 in genome size, structure, and arrangement (2). Both genomes are divided into unique long (U_L) and unique short (U_S) regions. However, only the U_S region is invertible in PRV, while both the U_L and U_S regions are invertible in HSV-1. Genomic differences reported for PRV relative to the HSV-1 genome include contraction of the PRV U_S region (24) and a large inversion from approximately 0.1 to 0.4 map units in the PRV genome (2, 8).

The entire 152-kb HSV-1 genome has been sequenced and reported to contain at least 72 genes (14). The complete nucleotide sequence of the estimated 140-kb PRV genome has not yet been determined. Approximately 30 PRV genes have been sequenced and mapped to date. While many PRV genes are homologous in position and sequence to their HSV-1 counterparts, some genes present in HSV-1 are absent in PRV. The kinetics of expression of some PRV and HSV-1 homologous genes also differs (2, 13).

Previous studies identified a complex pattern of transcription in the PRV *Bam*HI-G fragment (2). In this study, we present the DNA sequence, arrangement, and predicted amino acid sequence of a group of four PRV open reading frames located from 0.664 to 0.695 map units in the *Bam*HI-G fragment of the U_L region. We designated these PRV open reading frames UL1, UL2, UL3, and UL3.5, to reflect organizational and sequence similarities between PRV and HSV-1. Northern (RNA) blot analysis was performed to map and identify the kinetic class of the UL1, UL2, UL3, and UL3.5 transcripts. The results revealed a complex pattern of overlapping and possibly spliced transcripts which have similar transcription termination sequences and may also have common promoter sequences.

MATERIALS AND METHODS

Cells and virus. The Indiana-Funkhauser (InFh) strain of PRV was grown on Madin-Darby bovine kidney (MDBK) cells cultivated in Eagle's minimal essential medium supplemented with 10% fetal bovine serum.

Nucleotide sequencing and analysis. The *Bam*HI-G fragment was cloned into the Bluescript SK+ plasmid, and the nucleotide sequence of both strands was determined by the dideoxy chain-termination method with a series of oligonucleotide primers (23). Searches of the GenBank version 72 and Swiss-Protein version 22 data bases were performed with the program FASTA from the University of Wisconsin Genetics Computing Group (19). Multiple protein sequence alignments were computed with the Alignment program from GeneWorks (Intelligenetics).

RNA isolation and Northern blot analysis. Total cellular RNA was isolated as described previously (5). Glyoxal-treated total-RNA samples were subjected to electrophoresis and transferred to nylon membranes by using positive pressure in 10× SSC (1× SSC is 0.15 M NaCl with sodium citrate at 0.015 M). Hybridizations with nick-translated probes were done as described previously (6). Hybridizations with ^{32}P-end-labeled 18-mer oligonucleotide probes were performed at 45°C in the presence of 5× SSC–5× Denhardt's solution (1× Denhardt's solution is 0.2% each bovine serum albumin, Ficoll, and polyvinylpyrrolidone). The blots were washed at 50°C with 5× SSC–5× Denhardt's solution for 30 min, then at 51°C with 3.2 M tetramethylammonium chloride containing 1% sodium dodecyl sulfate for 30 min, and then twice for 15 min each in 2× SSC at room temperature.

Nucleotide sequence accession number. The DNA sequence

* Corresponding author.

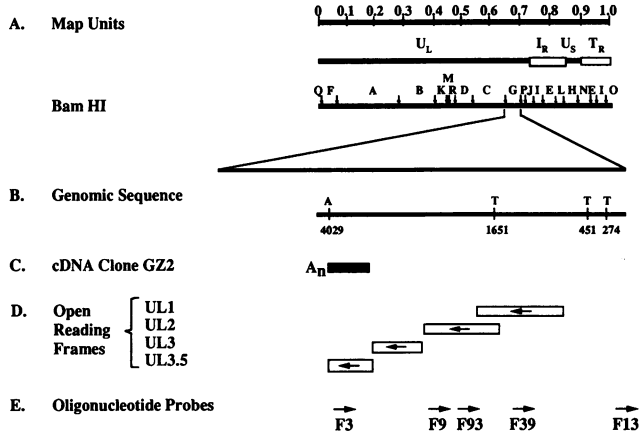

FIG. 1. (A) Schematic diagram of the PRV genome and *Bam*HI restriction enzyme map. The genome is organized into the unique long (U_L), internal repeat (I_R), unique short (U_S), and terminal repeat (T_R) sequences. (B) Expanded diagram of the *Bam*HI-G restriction fragment. T, TATA box; A, polyadenylation signal. (C) Location of the polyadenylated (A_n) cDNA clone GZ2. The λ ZAP II cDNA library was constructed with polyadenylated RNAs isolated from MDBK cells infected with PRV InFh, as described previously (6). The *Bam*HI-G fragment, ^{32}P labeled by nick translation, was used to screen for *Bam*HI-G-specific clones. DNA inserts were excised and rescued, and the DNA sequences were determined by the dideoxy chain-termination method (23). (D) Locations of open reading frames UL1, UL2, UL3, and UL3.5. The direction of translation is indicated by an arrow. (E) Location and orientation of oligonucleotide probes.

data in this report have been submitted to the GenBank data base under accession number L13855.

RESULTS

Nucleotide sequence of the *Bam*HI-G region. We have determined the nucleotide sequence of the right-hand portion of the *Bam*HI-G fragment of PRV strain InFh, from 0.664 to 0.695 map units in the conventional PRV genome representation. Consistent with the entire PRV genome (2), the region sequenced is 73% G+C. Figure 2 shows the nucleotide sequence numbered from right to left with respect to the conventional PRV genome map, beginning with the *Bam*HI site at the junction of the *Bam*HI-G and -P fragments. Computer-aided sequence analysis revealed four potential open reading frames, named UL1, UL2, UL3, and UL3.5 to be consistent with the corresponding HSV open reading frames. The four open reading frames were all oriented in the same direction. The locations and the predicted amino acid sequences of these open reading frames are shown in Fig. 1 and 2.

There are three potential TATA boxes in the UL1, -2, -3, and -3.5 gene cluster, located at nucleotides 274 (TATATT), 451 (TATATA), and 1651 (TATAAT) (Fig. 2). The first two TATA boxes are upstream of UL1, and the third is within the UL2 coding sequence, upstream of UL3. The only potential polyadenylation signal for this gene cluster (AATAAA) is located 3′ to the UL3.5 open reading frame at nucleotide 4029. The polyadenylation signal 5′ of the gene cluster (nucleotide 262) is utilized by the EP0 gene (6).

UL1. The UL1 open reading frame (nucleotides 1024 to 1675) is the first open reading frame downstream of the two TATA boxes at nucleotides 274 and 451. Immediately upstream of UL1, from nucleotides 912 to 999, is a region of eight direct repeats, consisting of the sequence CCCCCC CGACT. UL1 has the capacity to code for a 217-amino-acid protein with a predicted molecular mass of 23 kDa and an isoelectric point of 6.5. Examination of the arrangement of PRV genes in the vicinity of UL1 suggested that PRV UL1 is equivalent in genome location to the VZV gene 60 (7) and the HSV-1 UL1 gene (14), which codes for the glycoprotein gL (11). Both the HSV-1 and PRV UL1 open reading frames are located upstream of and partially overlap their respective UL2 open reading frames (see below). Immediately upstream of the UL1 open reading frame is the early protein 0 (EP0) gene in PRV and the homologous ICP0 gene in HSV-1 (6, 14). Thus, the UL1 open reading frames in PRV and HSV-1 are flanked by homologous genes. The predicted PRV UL1 translation product has limited sequence homology to the translated VZV gene 60 (Fig. 3) and to HSV-1 UL1. As computed by the Alignment program from GeneWorks, PRV UL1 has 21% identity with VZV gene 60 and 13% identity with HSV-1 UL1. The identity between HSV-1 UL1 and the VZV gene 60 is 12%.

UL2. The PRV UL2 open reading frame is located from nucleotides 1586 to 2602. Examination of the sequence revealed no TATA box elements immediately upstream of UL2. The closest upstream TATA box is located 5′ to UL1. UL2 has the capacity to code for a 339-amino-acid basic protein with a predicted molecular mass of 36 kDa and an isoelectric point of 10.2. The predicted translation product contains a consensus uracil-DNA glycosylase signature sequence (WARRGVLLLN) (17, 18, 22) from amino acids 232 to 241. A comparison of the PRV UL2 translated sequence with known and predicted viral, bacterial, and eukaryotic uracil-DNA glycosylase sequences is shown in Fig. 4 (1, 4, 7, 14, 17, 25, 27). PRV UL2 is 48% homologous with the 335-amino acid HSV-1 UL2 and 55.8% homologous with the 295-amino-acid HSV-2 UL2, both of which have been shown to have uracil-DNA glycosylase activity (16, 27). The greatest divergence among these sequences occurs at the 5′ end. The overall homology among the eight uracil-DNA glycosylase sequences shown in Fig. 4 is 12.8%.

UL3. The PRV UL3 open reading frame is located from nucleotides 2661 to 3371. A potential TATA box is located upstream of UL3 at nucleotide 1651 (TATAAT), within the UL2 coding sequence. With the first potential methionine codon at nucleotide 2661, UL3 has the capacity to code for a 237-amino-acid protein with a predicted molecular mass of 26 kDa and an isoelectric point of 5.8. UL3 contains two additional in-frame ATG codons at nucleotides 2679 and 2682, which, if used, would yield protein products of 231 and 230 amino acids, respectively. Figure 5 shows that the deduced PRV UL3 product has homology with the deduced products of HSV-1 UL3 (14) and VZV gene 58 (7). All three predicted proteins are similar in size. The greatest sequence divergence occurs at the 5′ end of the proteins. No protein product or function has been described for HSV-1 UL3 or VZV gene 58 to date.

UL3.5. The PRV UL3.5 open reading frame is located from nucleotides 3371 to 4030. The only potential polyadenylation signal for the entire UL1 to UL3.5 gene cluster is located immediately downstream of the UL3.5 open reading frame. The closest TATA box for UL3.5 is located upstream of UL3. UL3.5 has the capacity to code for a 220-amino-acid basic protein with a predicted molecular mass of 24 kDa and an isoelectric point of 12. The amino acid composition of UL3.5 includes 17% alanine and 16% arginine residues. In addition, the amino acid sequence contains a histidine- and

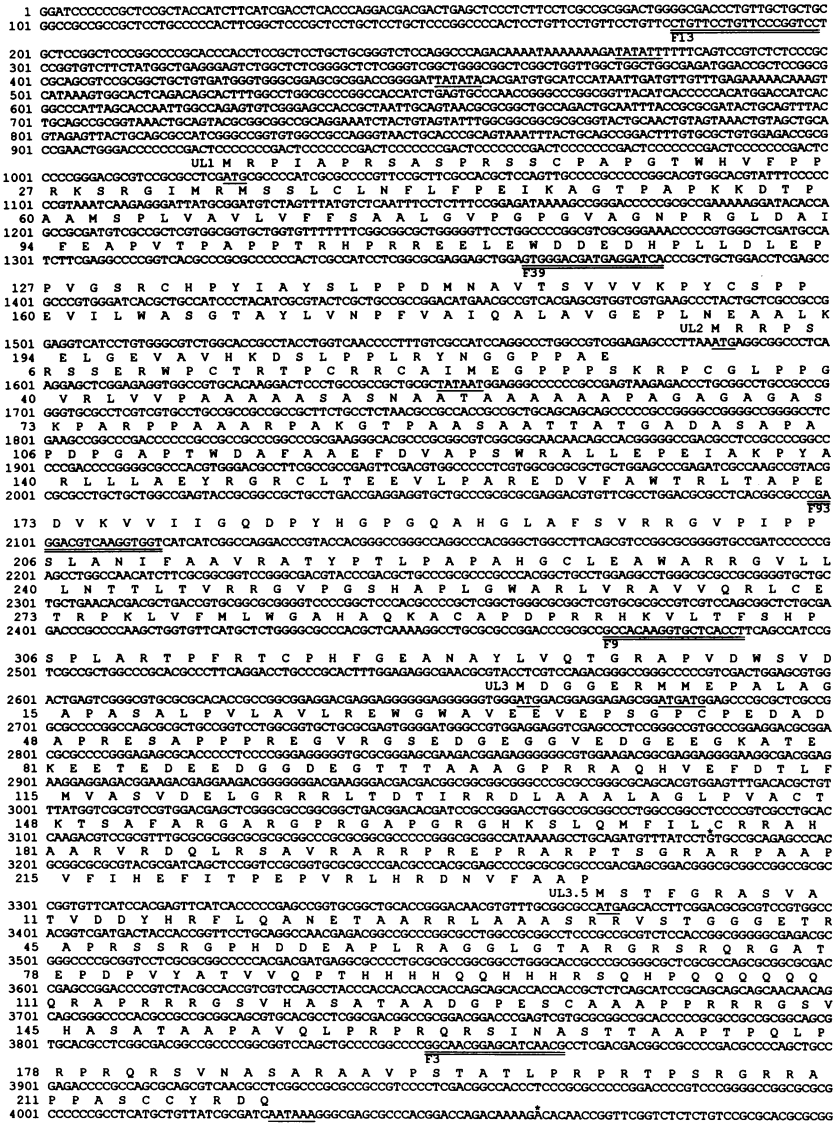

FIG. 2. Nucleotide sequence of the right end of the *Bam*HI-G fragment of PRV InFh. Numbering is from right to left with respect to the conventional PRV genome arrangement. Numbering starts with the *Bam*HI site at the junction of the *Bam*HI-G and -P fragments. The predicted amino acid sequences of UL1, UL2, UL3, and UL3.5 are indicated. Putative TATA boxes, polyadenylation site, and translation initiation ATG codons are indicated by a single underline. The 18-mer oligonucleotide probes are complementary to the nucleotide sequences indicated by a double underline. The starting and ending coordinates of cDNA clone GZ2 are indicated by asterisks.

glutamine-rich domain from amino acids 91 to 111. The beginning of this domain consists mainly of histidine, and it ends with seven glutamines. The PRV UL3.5 open reading frame does not have a counterpart in HSV-1. A search of the Swiss Protein data base with the UL3.5 translated sequence revealed no detectable sequence homology to any herpesvirus or other protein sequences.

Kinetic class of UL1 to UL3.5 transcripts. To determine the transcriptional activity of the *Bam*HI-G region, Northern blot analysis was conducted on mRNA isolated from PRV-infected MDBK cells at various times postinfection (p.i.). As shown in Fig. 6, the ^{32}P-labeled *Bam*HI-G probe hybridized to four abundant transcripts of 3.3, 2.7, 1.8, and 0.9 kb. A minor RNA species of 3.6 kb, derived from the left end of *Bam*HI-G (unpublished data), was also detected. All four major transcripts were detected at 4 h p.i. and did not

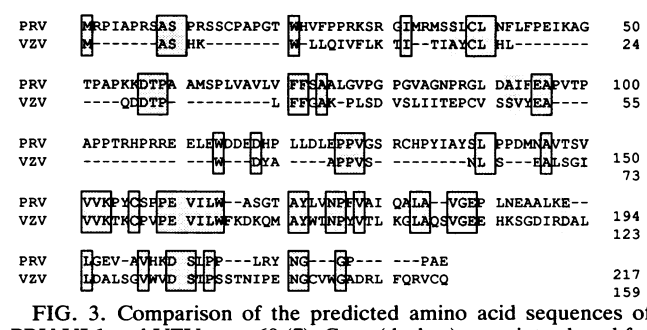

FIG. 3. Comparison of the predicted amino acid sequences of PRV UL1 and VZV gene 60 (7). Gaps (dashes) were introduced for best alignment. Identical residues are boxed; similar amino acids are shaded. The 217-amino-acid PRV UL1 and the 159-amino-acid VZV gene 60 products show 21% sequence identity.

```
PRV      MRRPSRSSER WPCTRTPCRR CAIMEGPPPS KRPCGLPPGV RLVVPAAAAA   50
HSV1     MKRACSRSPS -PRRRPSSPR RTPPRDGTPP QKADADDPTP GASNDASTET   49
HSV2     M--------A -MKRNPSRVF CAYSKNGT-- HRSAA--PT- ---------T   27
VZV      M--------- --DVSGEPTV CSNAYANE-- MKLSDSKDIY VLAHPVTKKT   37
ECOLI    A--------- -------NEL ---------- TW-------- ----------    6
HUMAN    M--------- -------GVF C-------LG PW-------- GLGRKLRTPG   19
EBV      M--------- -------ASR GLD------- LW-------- ----------    9
HCMV     M--------- -------ALK ---------- QW-------- ----------    6

PRV      SASNAATAAA AAAPAGAGAG ASKPARPPAA ARPAKGTPAA SAATTATGAD  100
HSV1     RPGSGGEPAA CRSSGPAALL AALEAGPAGV TFSSSAPPDP PMDLTNGGVS   99
HSV2     HR-------- CIAGGGR--- GALDAGAE-- --NTQGHPE- SRCFPGGRPP   61
VZV      RK-------- -RPRGLP--- LGVKLDPPTF KLNNMSHHYD TETFTPVSSQ   75
ECOLI    ---------- ---------- -HDV------ ---------- ---LAEE---   13
HUMAN    KGPLQLLSRL C------GDH LQAIPAKKAP AGQEEPGTPP SSPLSAEQLD   63
EBV      ---------- ---------- LDEHVWKR-- ---------- KQEIGVKGEN   27
HCMV     ---------- ---------M LANIADNK-- ---------- GSLLTPDEQA   25

PRV      ASAPAPDPGA PTWDAFAAEF DVAPSWRALL EPEIAKPYAR LLLAEYRG-R  149
HSV1     PAATSA---P LDWTTFRRVF LIDDAWRPLM EPELANPLTA HLLAEYNR-R  145
HSV2     QTGPSW---C LG-AAFRRAF LIDDAWRPLL EPELANPLTA RLLAEYDR-R  106
VZV      LDSV------ ---EVFSK-F NISPEWYDLL SDELKEPYAK GIFLEYNRLL  115
ECOLI    ---------- ---------- ------KQQ- --PY---FLN TLQTVASE-R   30
HUMAN    RIQRNKAAAL LRLAARNVPV GFGESWKKHL SGEFGKPYFI KLMGFVAE-E  112
EBV      LL-------- --LPDLWLDF ------LQL- -SPIFQRKLA AVIACVRR-L   58
HCMV     RV-------- FCLSADWIR- -F-LSLPDH- -DTV---LLR DTVAAVEG-A   58

PRV      CLTEEVLPAR EDVFAWTRLT APEDVKVVII GQDPYHGPGQ AHGLAFSVRR  199
HSV1     CQTEEVLPPR EDVFSWTRYC TPDEVRVVII GQDPYHHPGQ AHGLAFSVRA  195
HSV2     CQTEEVLPPR EDVFSWTRYC TPDDVRVVII GQDPYHHPGQ AHGLAFSVRA  156
VZV      NSGEEIIPST GDIFAWTRFC GPQSIRVVII GQDPYPTAGH AHGLAFSVKR  165
ECOLI    QSGVTIVPPQ KDVFNAFRFT ELGDVKVVIL GQDPYHGPGQ AHGLAFSVRP   80
HUMAN    RKHYTVVPPP HQVFTWTQMC DIKDVKVVIL GQDPYHGPNQ AHGLCFSVQR  162
EBV      RTQATVVPEE DMCMAWARFC DPSDIKVVIL GQDPYHG-GQ ANGLAFSVAY  107
HCMV     RQLEMVVPAP EHVHRWSYLC PPEQVRVVIV GQDPYCD-GS ASGLAFSTLA  107

PRV      GVPIPPSLAN IFAAVRATYP TLPAPAHGCL EAWARRGVLL LNTTILTVRR  249
HSV1     NVPPPHSLAN VLAAVKNCYP EARMSGHGCL EKWARDGVLL LNTTILTVKR  245
HSV2     DVPVPHSLAN VLAAVKNCYP DARMSGHGCL EKWARDGVLL LNTTILTVKR  206
VZV      GITPPSSLAN IFAALMESYP NMTPPTHGCL ESWARQGVLL LNTTILTVRR  215
ECOLI    GIAIPHSLLN MYKELENTIP GFTRPNHGVL ESWARQGVLL LNTVLTVRAG  130
HUMAN    PVPPPHSLEN IYKELSTDIE DFVHPGHGCL SGWAKQGVLL LNAVLTVRAH  212
EBV      GFPVPHSLHN IYAELHRSLP EFSPPDHGCL DAWASQGVLL LNTILTVQKG  157
HCMV     GRPPPHSLAN VFRELARTVD GFQRPASGCL DAWARRGVLL LNTVFTVVHG  157

PRV      VPGSHAPLGW ARLVRAVVQR LCETRPKLVF MLWGAHAQK- ACAPDPRHK  298
HSV1     AAASHSRIGW DRFVGGVIRR LAARRPGLVF MLWGTHAQN- AIRPDPRVHC  294
HSV2     AAASTSKIGW DRFVGGVVRR LAARRPGLVF MLWGAHAQN- AIRPDPRCHY  255
VZV      TPGSHVYIGW GRLVQRVLQR LCENRTGLVF MLWGAHAQK- TTQPNSRCHL  264
ECOLI    QAHSHASLGW ETFTDKVISL INQHREGVVF LLWGSHAQKK GAIIDKQRHH  180
HUMAN    QANSHKERGW EQFTDAVWSW LNQNSNGLVF LLWGSYAQKK GSAIDRKFHH  262
EBV      KPGSHADIGW AWFTDHVISL LSERLKACVF MLWGAKAGDK ASLINSKKHL  207
HCMV     QPGSHRHIGW QTLSNHVIRR LSERREHLVF MLWGADAHTC EYLIDRRFHL  207

PRV      VLTFSHPSPL ---A-RTP-- --FRTCPHFG EANAYLVQTG RAPVDWS-VD  339
HSV1     VLKFSHPSPL ---S-KVP-- --FGTCCHFL VNNRYLETRS ISPIDWSV--  334
HSV2     VLKFSHPSPL ---S-KVP-- --FGTCCHFL AANRYLETRD IMPIT-VV--  294
VZV      VLTHAHPSPL ---S-RVP-- --FRNCRHFV QNNEYFTRKG EPEIDWS-VI  305
ECOLI    VLKAHHPSPL ---SAHRG-- --FFGCNHFV LANQWLEQRG ETPIDWMPVL  223
HUMAN    VLQTAHPSPL ---SVYRG-- --FFGCRHFS KTNELLQKSG KKPIDWKEL-  304
EBV      VLTSQHPSPL AQNSTRKSAQ QKFLGNNHFV LANNFLREKG LGEIDWRL--  255
HCMV     VLKSCHPSPR ---NTTRA-- --FVGNDHFI LANAYLDTHY RETMDWR-L-  248

PRV      -----                                                   339
HSV1     -----                                                   334
HSV2     -----                                                   294
VZV      -----                                                   305
ECOLI    PAESE                                                   228
HUMAN    -----                                                   304
EBV      -----                                                   255
HCMV     ---CG                                                   250
```

FIG. 4. Comparison of the predicted amino acid sequences of PRV UL2 with those of known and predicted HSV-1 (15), HSV-2 (27), VZV (7), *Escherichia coli* (25), human (17), EBV (1), and HCMV (4) uracil-DNA glycosylases. The uracil-DNA glycosylase signature sequence (heavy overline) was identified by the Prosite program in PC Gene (Intelligenetics), based on data in references 17, 18, and 22. Gaps (dashes) were introduced for best alignment. Identical residues are boxed; similar amino acids are shaded.

decrease in abundance for the remaining course of the infection (Fig. 6A). The increase in RNA levels at 12 h p.i. is likely due to second-round virus replication, since the replicative cycle of InFh in MDBK cells is approximately 8 h (5). The high-molecular-weight RNA species which appeared in the 12-h sample has not been identified.

To determine whether viral DNA replication was required for transcription, parallel cultures were treated before and

```
PRV      MDGGERMMEP ALAGAPASAL P-VLAVLREW GWAVEEVEPS GPCPEDADAP   49
HSV      MVKPLVSYGS VMSGVGGEGV PSALAILASW GWTFD--TPN ----HESGIS   44
VZV      M--------- -FSELP-PSV P---TALLQW GWGLH----- --------RG   23

PRV      RESAPPPREG VRGSEDGEGG VEDGEEGKAT EKEETEDEED GGDEGTTTAA   99
HSV      PDTTPA--DS IRGA--AVAS PDQPLHG-GP EREATAPSFS PTRADDGPPC   89
VZV      PCSIP----N FK----QVAS QHSVQNDFTE NSVDANEKFP IGHAGCIEKT   65

PRV      AGPRRAQHVE FDTLFMVASV DELGRRELTD TIRRDLAAAL AGLPVACTKT  149
HSV      TD---GPYVT FDTLFMVSSI DELGRRQLTD TIRKDLRLSL AKFSIACTKT  136
VZV      KD----DYVP FDTLFMVSSI DELGRRQLTD TIRRSLVMNA CEITVACTKT  111

PRV      SAFARGA--- --RGP--RGA PG-RGHKSLQ MFTLCRRAHA PRVRDCLRSA  191
HSV      SSFSGNAPRH HRRGAFQRGT RAPRSNKSLQ MFVLCKRAHA ARVREGLRVV  186
VZV      AAFSGRGVSR QKHVTLSKNK FNPSSHKSLQ MFVLCQKTHA FRVRNILYES  161

PRV      VRARRPREPR ARPTSGRARP AAPVFTHEFI TPEPVRLHRD NVFA------  235
HSV      IQSRKPRKYY TRSSDGRLCP AVPVFVHEFV SSEPMRLHRD NVML------  230
VZV      IRARRPRRYY TRSTDGKSPP LVPVFVYEFT ALDRVILHKE NTLTDQPINT  211

PRV      -----AP---                                              237
HSV      -----ASGAE                                              235
VZV      ENSGHGRTRT                                              221
```

FIG. 5. Comparison of the predicted amino acid sequence of PRV UL3 with the predicted amino acid sequences of the HSV-1 UL3 (14) and VZV gene 58 (7) products. Identical residues are boxed; similar amino acids are shaded. The three amino acid sequences shown 24.2% identity and 27.7% similarity.

during infection with a viral DNA synthesis inhibitor. The abundance of all four transcripts was reduced but not eliminated in the presence of phosphonoacetic acid (Fig. 6B). Therefore, these transcripts are all of the early-late or γ-1 kinetic class. In comparison, the EP0 transcript, which is of the early kinetic class, was not inhibited by phosphonoacetic acid in a previous experiment with the same blot (6).

Mapping of UL1 to UL3.5 transcripts. The *Bam*HI-G fragment was used to screen a cDNA library from InFh-infected MDBK cells (6). All positive clones were sequenced, and one cDNA clone, GZ2, was selected for further analysis because it was found to contain a poly(A) tract. Comparison of the GZ2 sequence with the genomic sequence revealed that GZ2 starts at nucleotide 3186 and contains a poly(A) tract after nucleotide 4063, 34 nucleotides

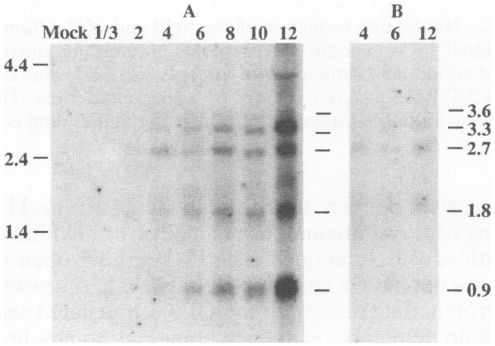

FIG. 6. Kinetics of UL1, UL2, UL3, and UL3.5 mRNA synthesis. Northern blots of RNA from PRV-infected MDBK cells were hybridized to the *Bam*HI-G fragment of PRV InFh, ^{32}P labeled by nick translation as described in Materials and Methods. (A) Total mRNA was isolated at 0.3, 2, 4, 6, 8, 10, or 12 h p.i. and from mock-infected cells. (B) Cells were treated with phosphonoacetic acid at 0.1 mg/ml for 1 h prior to and during infection. Ten micrograms of total mRNA isolated at 4, 6, or 12 h p.i. was electrophoresed. The transcripts of 3.6, 3.3, 2.7, 1.8, and 0.9 kb are indicated at right. The positions RNA molecular size markers (GIBCO BRL) (in kilobases) are indicated at the left.

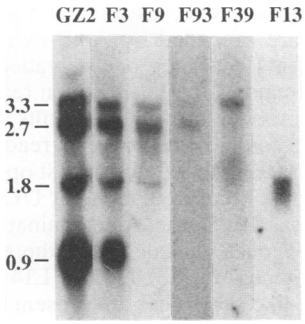

FIG. 7. Mapping of UL1 to UL3.5 transcripts by Northern blot analysis. Northern blots of total RNA from PRV-infected MDBK cells were hybridized to cDNA clone GZ2 or to 18-mer oligonucleotide probe F3, F9, F93, or F13 as described in Materials and Methods. The complementary sequences and locations of the probes are shown in Fig. 1 and 2. The positions of the transcripts of 3.3, 2.7, 1.8, and 0.9 kb from UL1 to UL3.5 are indicated.

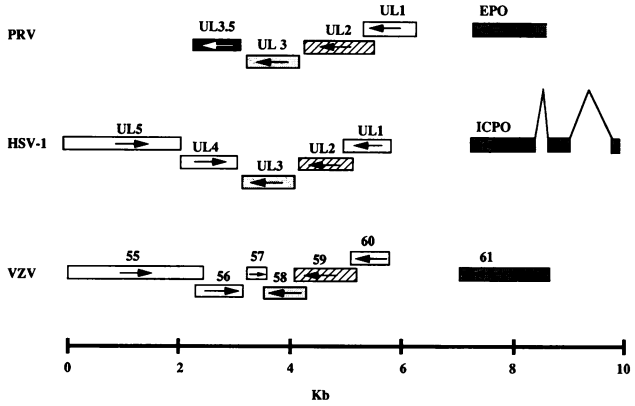

FIG. 8. Comparison of gene arrangement in PRV, HSV-1 (14), and VZV (7) in the region of the UL1, UL2, and UL3 gene cluster. For comparison, the HSV-1 sequence is inverted from the conventional representation. Genes with significant amino acid sequence homology are indicated by the same fill pattern. The direction of transcription is indicated by an arrow. The 5′ end of the PRV EP0 gene and the sequences 3′ to the UL3.5 gene have not been determined.

downstream of the polyadenylation signal at nucleotide 4029 (Fig. 1 and 2). In Northern blot analysis, GZ2 hybridized to four transcripts of 3.3, 2.7, 1.8, and 0.9 kb (Fig. 7). To determine the orientations of the four transcripts, Northern blots were probed with oligonucleotides derived from the GZ2 sequence. Oligonucleotide F3, complementary to the sequence shown in Fig. 2, hybridized to all four transcripts (Fig. 7), while an oligonucleotide complementary to the opposite strand did not. Therefore, all four mRNA species are transcribed in the same direction and have some common sequences. Oligonucleotide and nick-translated probes downstream of the polyadenylation site did not hybridize to any of the four transcripts, indicating that the four mRNA species all terminate at the polyadenylation site at nucleotide 4029.

To determine the coding potential of each transcript, Northern blots were probed with oligonucleotides predicted to hybridize to transcripts containing open reading frames UL1, UL2, UL3, and UL3.5. The complementary sequences and locations of selected oligonucleotides are indicated in Fig. 1 and 2, and the hybridization results are shown in Fig. 7. Probe F39, which is located within the UL1 open reading frame, hybridized to a single transcript of 3.3 kb. Probe F93, which is located within the UL2 open reading frame, hybridized to the 3.3-kb transcript and an additional transcript of 2.7 kb. Probe F9, located within the UL2 ORF and 150 nucleotides upstream of the first UL3 ATG, hybridized to three transcripts of 3.3, 2.7, and 1.8 kb. As mentioned above, probe F3, which is located within the UL3.5 open reading frame, hybridized to all four transcripts of 3.3, 2.7, 1.8, and 0.9 kb. Oligonucleotide probe F13, located immediately upstream of the two TATA boxes at nucleotides 279 and 471, hybridized only to the 1.7-kb EP0 transcript and not to any of the UL1 to UL3.5 gene cluster transcripts. Therefore, the 3.3-kb transcript contains sequences for UL1, UL2, UL3, and UL3.5, the 2.7-kb transcript contains sequences for UL2, UL3, and UL3.5, the 1.8-kb transcript contains sequences for UL3 and UL3.5, and the 0.9-kb transcript contains only sequences for UL3.5. The 1.8-kb mRNA hybridized with probe F9 and not with probe F93, indicating that this transcript contains UL3 and UL3.5 sequences and some 5′ sequence that extends into the UL2 open reading frame.

DISCUSSION

Previous studies showed that transcription in the *Bam*HI-G region of PRV, where the UL1 to UL3.5 genes are located, includes four early transcripts of 1.9, 3.0, 3.5, and 3.7 kb and one 1.2-kb early-late transcript (2). Early transcripts are generally defined as those which appear at and accumulate for up to 3 to 4 h p.i. and are not sensitive to phosphonoacetic acid, while early-late transcripts accumulate between 3 and 9 h p.i. and are less abundant in the presence of phosphonoacetic acid (2). The 3.5-, 3.0-, and 1.9-kb early transcripts and the 1.2-kb early-late transcript reported previously likely correspond to the 3.3-kb UL1, 2.7-kb UL2, 1.8-kb UL3, and 0.9-kb UL3.5 transcripts shown in this work. Our data indicated that all four transcripts are early-late mRNAs, as they continue to accumulate after 4 h p.i. and their levels are sharply reduced by treatment with phosphonoacetic acid (Fig. 6).

The arrangement of homologous genes in the UL1 to UL3 cluster of selected alphaherpesviruses is shown in Fig. 8. Previous studies showed that the *Bam*HI-G fragment of PRV cross-hybridizes with restriction fragments from the same relative genome location in HSV-1 (2, 3, 8). In HSV-1, the region contains three genes, while in PRV and VZV there are four open reading frames. Of the four PRV open reading frames, UL2 and UL3 are conserved in both genome location and sequence relative to the genes in HSV and VZV. The UL1 open reading frame is conserved in genome location but not sequence, and the UL3.5 open reading frame is unique to PRV. The 71-amino-acid-coding VZV gene 57 is located comparably to PRV UL3.5 but is in the opposite orientation and does not have sequence homology with UL3.5. The genes immediately upstream of UL1, EP0 in PRV, ICP0 in HSV-1, and gene 61 in VZV, are homologs (6, 7, 14, 15, 20). However, HSV ICP0 is located in the inverted repeat, while EP0 and gene 61 are located in the U_L region.

PRV and HSV-1 UL1 are both preceded by a stretch of direct repeat sequences. In the case of HSV-1 UL1, the repeats (GGGGGGGAGAGGGGAGA) are part of the terminal repeat of the U_L (14). It is possible that the repeat sequences upstream of PRV UL1 (CCCCCCCGACT) repre-

sent part of a vestigial U_L terminal repeat sequence, as has been suggested previously for U_L sequences adjacent to the IR region (9).

The PRV UL1 open reading frame does not show significant sequence homology with the HSV-1 UL1 open reading frame. However, the genome location and orientation of UL1 are equivalent for both viruses (14). In addition, the UL1 open reading frame is flanked by and partially overlaps genes which are equivalent between HSV-1 and PRV. Human cytomegalovirus (HCMV) and Epstein-Barr virus (EBV) also contain positional homologs of the HSV-1 UL1 open reading frame, HCMV UL115 (4) and EBV BKRF2 (1). The product of the HSV-1 UL1 gene is the essential glycoprotein gL (11, 21). HSV-1 gL has been shown to complex with the viral glycoprotein gH and is required for processing and cell surface expression of gH (11). Despite the lack of significant sequence homology, UL115 and BKRF2 have been shown to be functionally equivalent to gL, as each is required for cell surface expression of their respective gH homologs (12, 28). Therefore, variation in UL1 sequence could be one determinant of cell tropism between these herpesviruses. Additional experiments will be required to determine whether the product of PRV UL1 interacts with PRV gH.

PRV UL2 is a homolog of HSV UL2, which codes for a nonessential enzyme with uracil-DNA glycosylase activity (16, 27). The translated PRV UL2 open reading frame is 48% homologous to HSV-1 UL2 and 55.8% homologous to HSV-2 UL2. PRV UL2 also has homology with the putative EBV and HCMV uracil-DNA glycosylase genes, which are equivalent in genome location to UL2 (1, 4). The proteins are similar in size and have the uracil-DNA glycosylase signature sequence (WARRGVLLLN), as well as an overall basic character. In vitro transcription and translation of HSV-2 UL2 yields uracil-DNA glycosylase activity (27). In HSV-1, interrupting the UL2 open reading frame by *lacZ* insertion mutagenesis results in a virus which grows normally in tissue culture yet is negative for uracil-DNA glycosylase activity (16).

PRV UL3 has significant homology with HSV-1 UL3 and VZV gene 58 (7, 14). Both the HSV-1 and PRV UL3 open reading frames contain multiple in-frame methionine residues at the amino terminus, followed by a pronounced hydrophobic region. The function of the UL3 family members is unknown at this time. It has been suggested that HSV-1 UL3 might be a membrane-associated protein (14).

The UL3.5 open reading frame appears to be unique to PRV. The predicted translated sequence has no detectable homology to herpesvirus or other known proteins in the Swiss Protein data base. The sequence contains a basic region rich in histidine, glutamine, and arginine residues, which could possibly participate in protein-DNA interactions.

In general, most HSV genes studied to date are simple in structure. Most open reading frames are unspliced and are preceded by separate promoter elements, including a TATA box (13). Many HSV open reading frames are followed by a polyadenylation signal; however, families of HSV genes which share a polyadenylation site are common (13). For both HSV-1 and HSV-2, the UL1, UL2, and UL3 open reading frames are each preceded by a TATA box, and there are polyadenylation signals downstream of UL2 and UL3 (14, 27). The transcription of HSV-2 UL1, -2, and -3 is complex and has not been fully elucidated. Five overlapping transcripts encompassing all or part of the UL1, -2, and -3 gene cluster have been detected (27).

In contrast, in the PRV UL1 to UL3.5 gene cluster, only UL1 and UL3 are preceded by A+T-rich eukaryotic promoter-like elements, and only one polyadenylation signal is present, downstream of the UL3.5 open reading frame. The oligonucleotide probe hybridization results indicated that the larger transcripts contain multiple open reading frames of the gene cluster. It is likely that only the first open reading frame in each transcript is translated, as UL1, UL2, and UL3 each lie in a different reading frame. An examination of previously sequenced PRV genes revealed that there is at least one other group of genes, the UL12 to UL14 gene cluster, in which no TATA-like sequences are present between sequential open reading frames (10). The PRV genome has a marked bias for guanine and cytosine, with a G+C content of 73%, compared with 68% for the HSV-1 genome (2, 14). Consequently, in PRV there is a low frequency of A+T-rich regions, including potential TATA and polyadenylation signals. Thus, some PRV genes may have a transcription strategy different from that commonly observed for HSV genes.

The lack of a TATA box before each gene and the overlapping nature of the open reading frames suggest that either the genes in the UL1 to UL3.5 cluster share a single promoter or they do not use a TATA box. Alternatively, UL1 and UL2 may use the first promoter sequence upstream of UL1, while UL3 and UL3.5 originate from a second promoter sequence upstream of UL3. A common promoter for all four PRV genes is consistent with the observation that all of the transcripts are of the same kinetic class. Comparison of the size of each transcript with the nucleotide sequence shows that each transcript is roughly 300 to 400 bases longer than the size calculated from the first ATG to the polyadenylation site. Assuming similar degrees of polyadenylation, each transcript would contain approximately 100 to 200 bases of additional 5′ untranslated sequence. The closest TATA sequence upstream of any individual open reading frame varies from 600 to 1,700 nucleotides away, suggesting that the transcripts may be spliced, perhaps to a 100- to 200-base leader sequence. We are currently investigating the promoter sequences of this gene cluster.

ACKNOWLEDGMENTS

We thank Richard Zuerner for assistance with computer-aided sequence analysis. We credit Lory Patrilla for nucleotide sequencing of the cDNA clone; Judy Lundy of the Iowa State University Nucleic Acids Facility for oligonucleotide synthesis; Sue Ohlendorf for administrative assistance and typing of the manuscript; and Gene Hedberg, Tom Glasson, and Wayne Romp for graphics and photography.

REFERENCES

1. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310:**207–211.
2. **Ben-Porat, T., and A. S. Kaplan.** 1985. Molecular biology of pseudorabies virus, p. 105–173. *In* B. Roizman (ed.), The herpesviruses, vol. III. Plenum Press, New York.
3. **Ben-Porat, T., R. A. Veach, and S. Ihara.** 1983. Localization of the regions of homology between the genomes of herpes simplex virus, type 1, and pseudorabies virus. Virology **127:**194–204.
4. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerney, T. Horsnell, C. A. Hutchinson III, T. Kouzarides, J. A. Martignetty, and B. G. Barrell.** 1990. Analysis of the protein coding content of the sequence of the human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**127–169.
5. **Cheung, A. K.** 1989. Detection of pseudorabies virus transcripts

in trigeminal ganglia of latently infected swine. J. Virol. **63:** 2908–2913.
6. **Cheung, A. K.** 1991. Cloning of the latency gene and the early protein 0 gene of pseudorabies virus. J. Virol. **65:**5260–5271.
7. **Davison, A. J., and J. E. Scott.** 1986. The complete DNA sequence of varicella-zoster virus. J. Gen. Virol. **67:**1759–1816.
8. **Davison, A. J., and N. M. Wilkie.** 1983. Location and orientation of homologous sequences in the genomes of five herpesviruses. J. Gen. Virol. **64:**1927–1942.
9. **DeMarchi, J. M., Z. Lu, G. Rall, S. Kupershmidt, and T. Ben-Porat.** 1990. Structural organization of the termini of the L and S components of the genome of pseudorabies virus. J. Virol. **64:**4968–4977.
10. **de Wind, N., J. Domen, and A. Berns.** 1992. Herpesviruses encode an unusual protein-serine/threonine kinase which is nonessential for growth in cultured cells. J. Virol. **66:**5200–5209.
11. **Hutchinson, L., H. Browne, V. Wargent, N. Davis-Poynter, S. Primorac, K. Goldsmith, A. C. Minson, and D. C. Johnson.** 1992. A novel herpes simplex virus glycoprotein, gL, forms a complex with glycoprotein H (gH) and affects normal folding and surface expression of gH. J. Virol. **66:**2240–2250.
12. **Kaye, J. F., U. A. Gompels, and A. C. Minson.** 1992. Glycoprotein H of human cytomegalovirus (HCMV) forms a stable complex with the HCMV UL115 gene product. J. Gen. Virol. **73:**2693–2698.
13. **McGeogh, D. J.** 1991. Correlation between HSV-1 DNA sequence and viral transcription maps, p. 29–48. *In* E. K. Wagner (ed.), Herpesvirus transcription and its regulation. CRC Press, Boca Raton, Fla.
14. **McGeogh, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, J. E. Scott, and P. Taylor.** 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69:**1531–1574.
15. **Moriuchi, H., M. Moriuchi, H. A. Smith, S. E. Straus, and J. I. Cohen.** 1992. Varicella-zoster virus open reading frame 61 protein is functionally homologous to herpes simplex virus type 1 ICP0. J. Virol. **66:**7303–7308.
16. **Mullaney, J., H. W. Moss, and D. J. McGeogh.** 1989. Gene UL2 of herpes simplex virus type 1 encodes a uracil-DNA glycosylase. J. Gen. Virol. **70:**449–454.
17. **Muller, S. J., and S. Caradonna.** 1991. Isolation and characterization of a human cDNA encoding uracil-DNA glycosylase. Biochim. Biophys. Acta **1088:**197–207.
18. **Olsen, L. C., R. Aasland, C. U. Wittwer, H. E. Krokan, and D. E. Helland.** 1989. Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme. EMBO J. **8:**3121–3125.
19. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
20. **Perry, L. J., F. J. Rixon, R. D. Everett, M. C. Frame, and D. J. McGeogh.** 1986. Characterization of the IE110 gene of herpes simplex virus type 1. J. Gen. Virol. **67:**2365–2380.
21. **Roop, C., L. Hutchinson, and D. C. Johnson.** 1993. A mutant herpes simplex virus type 1 unable to express glycoprotein L cannot enter cells, and its particles lack gH. J. Virol. **67:**2285–2297.
22. **Sancar, A., and G. B. Sancar.** 1988. DNA repair enzymes. Annu. Rev. Biochem. **57:**29–67.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
24. **van Zilj, M., H. van der Gulden, N. de Wind, A. Gielkens, and A. Berns.** 1990. Identification of two genes in the unique short region of pseudorabies virus; comparison with herpes simplex virus and varicella-zoster virus. J. Gen. Virol. **71:**1747–1755.
25. **Varshney, U., T. Hutcheon, and J. H. Van de Sande.** 1988. Sequence analysis, expression, and conservation of *Escherichia coli* uracil-DNA glycosylase and its gene (ung). J. Biol. Chem. **263:**7776–7784.
26. **Wittman, G., and H.-J. Rziha.** 1989. Aujeszky's disease (pseudorabies) in pigs, p. 230–325. *In* G. Wittmann (ed.), Herpesvirus diseases of cattle, horses, and pigs. Kluwer, Boston.
27. **Worrad, D. M., and S. Caradonna.** 1988. Identification of the coding sequence for herpes simplex virus uracil-DNA glycosylase. J. Virol. **62:**4774–4777.
28. **Yaswen, L. R., C. Campbell, L. C. Davenport, E. B. Stephens, and L. M. Hutt-Fletcher.** 1992. Expression and analysis of the Epstein-Barr virus gH. Program Abstr. 17th Int. Herpesvirus Workshop, abstr. 273.