The Marek's Disease Virus (MDV) Unique Short Region: Alphaherpesvirus-Homologous, Fowlpox Virus-Homologous, and MDV-Specific Genes

PETER BRUNOVSKIS[1] AND LELAND F. VELICER[2]

Department of Microbiology, Michigan State University, East Lansing, Michigan 48824-1101

Despite its previous classification as a gammaherpesvirus, primarily due to its lymphotropism, Marek's disease virus (MDV), an oncogenic avian herpesvirus, is phylogenetically more related to the "neurotropic" alphaherpesviruses, characterized by its prototype, herpes simplex virus (HSV) (Buckmaster et al., 1988, J. Gen. Virol. 69, 2033–2042). In this report we present the DNA sequence of an 11,286-bp DNA segment encompassing the entire 11,160-bp-long U_S region of the oncogenic avian herpesvirus, Marek's disease virus. Eleven open reading frames (ORFs) likely to code for proteins were identified; of these, 7 represent homologs exclusive to alphaherpesvirus S component genes. These include MDV counterparts of HSV US1 (ICP22), US2, US3 (a serine-threonine protein kinase), US6, US7, and US8 (HSV glycoproteins gD, gI, and gE, respectively), and US10. Three additional ORFs were identified with no apparent relation to any sequences currently present in the SwissProt or GenBank/EMBL databases, while a fourth was found to exhibit significant homology to an uncharacterized fowlpox virus (FPV) ORF. Having precisely identified the IR_S–U_S and U_S–TR_S junctions, we have corrected and clarified their previously reported locations. By characterizing genes encoding three new alphaherpesvirus-related homologs (US1, US8, and US10), completing the sequence for a fourth (US7), and identifying 2 new MDV-specific ORFs (SORF1 and SORF3) and a fowlpox homolog (SORF2), our sequence analysis of the "virulent" GA strain of MDV (vMDV) extends upon that of a 5255-bp segment located in the U_S region of the "very virulent" RB1B strain of MDV (vvMDV) (Ross et al., 1991, J. Gen. Virol. 72, 939–947; 949–954). These two sequences were found to exhibit 99% identity at both nucleotide and predicted amino acid levels. Combined with the fact that MDV U_S sequences failed to show statistically significant CpG deficiencies, our analysis is consistent with MDV bearing a closer phylogenetic relation to alphaherpesviruses than to gammaherpesviruses. Because alphaherpesvirus-specific U_S region genes are primarily nonessential for virus replication, they are thought to be important biological property determinants. Thus, our sequence provides a foundation for further MDV studies aimed at resolving the apparent discrepancy between MDV's genetic and biologic properties. © 1995 Academic Press, Inc.

INTRODUCTION

Marek's disease virus (MDV) is a highly pathogenic herpesvirus of chickens, which can cause: (i) T-cell lymphomas as early as 3 weeks postinfection; (ii) peripheral neural lesions, characterized by lymphoproliferative infiltration and demyelination, occasionally leading to paralysis and/or blindness; (iii) various phenomena of acquired immunodeficiency; and (iv) atherosclerosis in normocholesterolemic chickens, bearing a remarkable resemblance to the human disease, in both character and distribution of arterial lesions (reviewed in Calnek and Witter, 1991). Marek's disease (MD) most commonly refers to the lymphoproliferative conditions above and is noteworthy for being the first naturally occurring lymphomatous neoplasm to be effectively controlled by vaccination (Churchill et al., 1969).

[1] Present Address: Department of Molecular and Microbiology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4960.

[2] To whom reprint requests should be addressed.

Because of similar biological properties, especially its lymphotropism, MDV and its antigenically related, apathogenic vaccine virus, herpesvirus of turkeys (HVT), have until recently been classified as gammaherpesviruses (Roizman et al., 1981; Roizman, 1992). In contrast to gammaherpesviruses, MDV and HVT have genome structures more closely resembling those of alphaherpesviruses (Cebrian et al., 1982; Fukuchi et al., 1985; Igarashi et al., 1987). Consistent with their structural relatedness to alphaherpesviruses, recent data indicate that MDV and HVT are phylogenetically more related to alphaherpesviruses than gammaherpesviruses (Buckmaster et al., 1988). This raises interesting questions regarding the seeming incongruence between MDV's genetic and biologic properties. To understand the nature of these differences and to identify new glycoproteins potentially important in virus–host cell interactions and mechanisms of protective immunity against MD, we have become particularly interested in the MDV U_S region. This stems from the observation that alphaherpesvirus U_S regions are known to contain a cluster of glycoprotein genes and appear to specify determinants for pathogenesis and viral dissemination, rather than those essential for virus production (Roizman, 1990a). These determi-

nants are encoded by a cluster of "nonessential" or *supplementary essential* (Roizman, 1990a) genes which are likely to account for many of the unique *in vivo* properties characteristic for a given alphaherpesvirus. The natural host–MDV system affords a unique opportunity to examine the *in vivo* role and function of this putative class of supplementary essential genes.

The alphaherpesvirus U$_S$ region is flanked by a pair of inverted repeat sequences (inverted and terminal repeat short, IR$_S$ and TR$_S$, respectively, or simply, repeat short, R$_S$). Together, these components make up the S region. Alphaherpesvirus U$_S$ and other S region genes originate from an area specific to members of this group, arguably their most divergent coding region. Second they specify a cluster of glycoprotein genes important for virus–cell interactions and mechanisms of protective immunity. A similar cluster in MDV would be particularly significant given the paucity of details regarding protective immunity against naturally occurring MD tumors.

The DNA sequence of a 5255-bp segment from U$_S$ region of the "very virulent" RB1B strain of MDV (vvMDV) was recently reported (Ross *et al.,* 1991). This region was found to contain open reading frames (ORFs) homologous to proteins encoded by HSV US2, US3 (protein kinase), US6 (glycoprotein D), part of US7 (glycoprotein I), and an additional MDV-specific ORF. In this report, we extend these results and present a sequence analysis of the entire 11.2-kbp MDV U$_S$ region ("virulent" GA strain, vMDV).

MATERIALS AND METHODS

Recombinant plasmids, M13 subcloning, and DNA sequencing

Pathogenic MDV GA strain subclones included *Eco*RI-O, -I, and -V cloned into pBR328 (Gibbs *et al.,* 1984) (pE328-O, pE328-I, and pE328-V, Fig. 1B); *Bam*HI-A and *Bam*HI-P1, cloned into pACYC184 and pBR322, respectively (Fukuchi *et al.,* 1985) (pBACYC-A and pB322-P1 (Fig. 1B), kindly provided by Dr. Meihan Nonoyama of the Tampa Bay Research Institute, St. Petersburg, FL); and GA-02, a phage clone containing a partially digested MDV *Sau*3A insert cloned into the *Sal*I site of EMBL3, kindly provided by Dr. Paul J. A. Sondermeijer, Intervet International, Boxmeer, The Netherlands. The latter clone contains most of *Bam*HI-A, all of *Bam*HI-P1, and additional 3' flanking sequences, including some of those present in pE328-V. This phage clone was used to generate the pUC18 subclone, pSP18-A (Fig. 1B). This clone contains a 2.5-kb *Sal*I insert with approximately 20 bp of EMBL-3's multiple cloning site at its 3' end. Together, the above clones (Fig. 1B) were used to generate M13mp18 and -19 subclones for use as templates for nucleotide sequencing.

DNA sequencing of both strands was performed by the dideoxy chain-termination method (Sanger *et al.,*

1977) using single-stranded M13 templates. Reaction products were synthesized and labeled using a 17-mer M13 primer, a modified T7 DNA polymerase (Sequenase), [^{35}S]thio-dATP (NEN), and appropriate deoxy- and dideoxynucleotides according to instructions by the manufacturer (Sequenase sequencing kit; United States Biochemical Corp., Cleveland, OH) and electrophoresed through 7% polyacrylamide/50% urea/Tris–Borate–EDTA gels. Remaining sequence gaps were determined by substituting M13 primers with synthetic 17-mer oligonucleotides (under similar reaction conditions, 0.5 pmol/reaction).

Analysis of sequence data

Sequences were assembled and analyzed with an IBM Personal System 2/Model 50 microcomputer utilizing Genepro (Version 4.10; Riverside Scientific Enterprises, Seattle, WA) sequence analysis software packages or programs obtained from the University of Wisconsin Genetics Computer Group (UWGCG, Versions 6.2 and 7.0; Devereaux *et al.,* 1984) and run through a VAX 8650 minicomputer. Homology searches of the SwissProt (Release 18.0), GenBank (Release 71.0), and EMBL (Release 30.0) databases were performed using the UWGCG programs FASTA and TFASTA (Pearson and Lipman, 1988).

RESULTS

Defining the MDV U$_S$ region and the location of the unique-repeat region junctions

Figure 1 contains a map of the area that was sequenced. This segment is bounded by a pair of *Pvu*II sites and spans the 3' half of *Bam*HI-A, extending an additional 1.5 kbp to the right of the 3' end of *Bam*HI-P$_1$ (Figs. 1B and 2). This 11,286-bp segment spans the entire U$_S$ region, which is 11,160 bp in length and flanked at the 5' and 3' ends by a 63-bp stretch of IR$_S$ and TR$_S$ DNA, respectively, each inversely complementary to the other (Figs. 1B and 2). Based on Southern blot analysis, the IR$_S$–U$_S$ junction was previously localized to a 1.4-kb *Bgl*I fragment (Fukuchi *et al.,* 1985) located in the second of five *Eco*RI subfragments of *Bam*HI-A (for *Bam*HI-A/*Eco*RI map, see Wen *et al.,* 1988). Our sequence analysis demonstrates that the IR$_S$–TR$_S$ junction is instead located in the middle of the third *Eco*RI subfragment, approximately 2–3 kbp downstream from the position reported above. This is further supported by Southern blot analysis of genomic DNA from the same strain (GA) as above (data not shown). Consequently, the U$_S$–TR$_S$ junction was localized 263 bp downstream of the US8 termination codon (following position 11,223, Fig. 2).

Nucleotide sequence and identification of open reading frames

The overall guanine plus cytosine ratio of the region sequenced was found to be 41%, somewhat below the
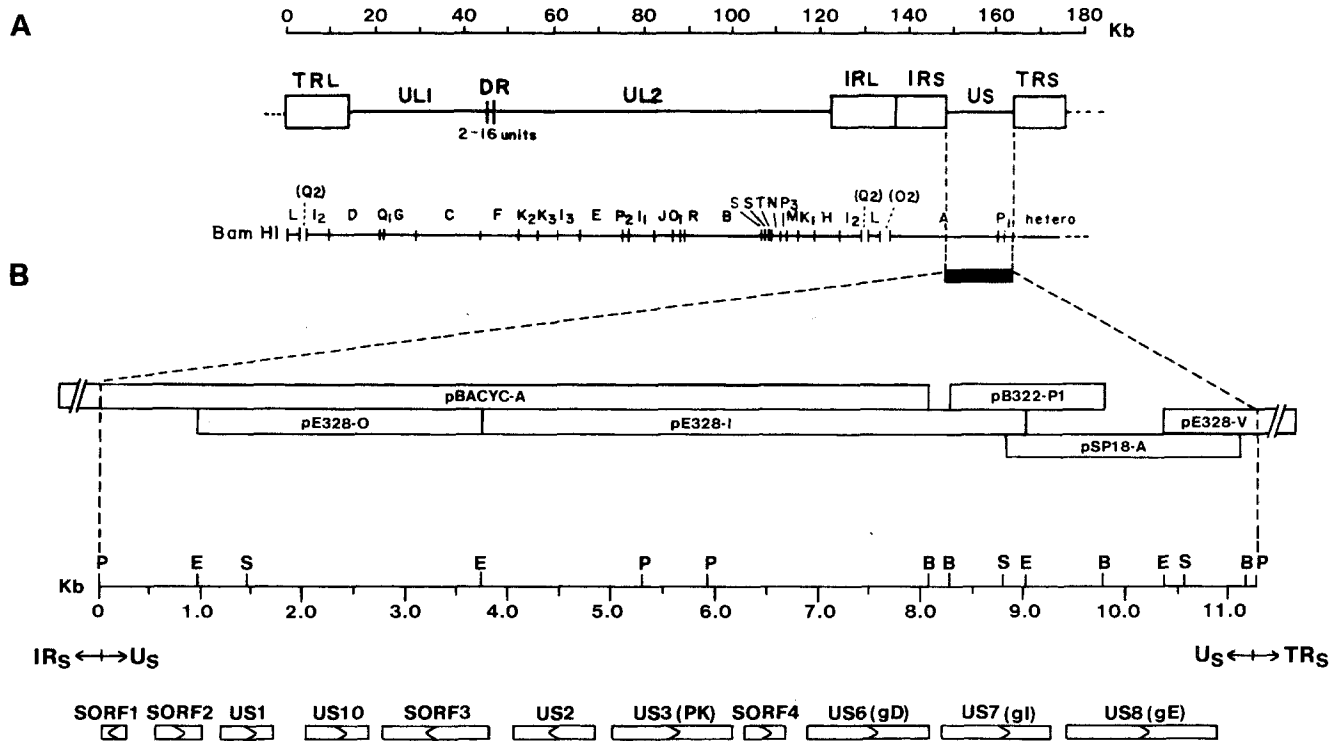
FIG. 1. Map location of area sequenced and organization of MDV U$_S$ ORFs. (A) MDV genome structure and BamHI restriction map outlining area sequenced. Restriction enzymes: B, BamHI; E, EcoRI; P, PstI; and S, SaII. Upper boxes define plasmid clones with BamHI, EcoRI, or SaII-bound inserts that were used to generate M13mp18 and -19 templates for sequencing. (B) Organization of MDV U$_S$ ORFs. Lower boxes (with arrows) represent location of MDV ORFs. Arrows define direction of transcription/translation. Names of ORFs are displayed above boxes. Basis for nomenclature is outlined under Results.

reported genomic value of 46% (Calnek and Witter, 1991). Observed frequencies of CpG dinucleotides in the whole sequence, or in the coding regions only, did not differ significantly from those expected from their mononucleotide compositions (data not shown). This result agrees with those obtained from alphaherpesviruses, while sharply contrasting with the CpG deficiencies associated with all gammaherpesviruses thus far studied (Efstathiou et al., 1990; Honess et al., 1989).

The region sequenced contains at least 11 ORFs likely to code for proteins (Fig. 1B; basis for names is defined below). This prediction was primarily based on homology and positional organization comparisons to other alphaherpesvirus genes, as well as the observation that alphaherpesviruses generally tend to contain relatively tightly packed, unspliced coding regions (Davison and Scott, 1986; McGeoch et al., 1985, 1987, 1988; Telford et al.,

1992). Methods for detecting protein coding regions based on the use of MDV-derived codon frequency tables (using these and previously published MDV sequences; Binns and Ross, 1989; Chen et al., 1992; Jones et al., 1992; Ross et al., 1989; Scott et al., 1989) or analysis of compositional bias (using the UWGCG programs CODONPREFERENCE and TESTCODE) were inconclusive. However, as pointed out previously (Ross et al., 1991), MDV-encoded ORFs do exhibit a detectable bias for A–T residues in the wobble position. Furthermore, using the UWGCG program FRAMES, together with the MDV-derived codon frequency table above, the 11 identified ORFs clearly show a significantly low pattern of rare codon usage not observed following computer-based translation of the remaining reading frames (data not shown).

The predicted amino acid sequences of the above

FIG. 2. Nucleotide and predicted amino acid sequences. The nucleotide sequence is given as the rightward 5' to 3' strand only (numbered 1 to 11,286). IR$_S$ and TR$_S$ sequences are located at the 5' and 3' ends, respectively, and are depicted using lowercase symbols; U$_S$ sequences are in uppercase. Rightward- and leftward-directed predicted amino acid sequences are shown above and below the corresponding nucleotide sequences, respectively, in single-letter code. The name of each ORF is given to the left of the first line of its respective sequence. Amino acid sequences are numbered from the N terminus, beginning with the first in-frame methionine codon and ending with the amino acid at the C-terminus, which precedes the termination codon. Dotted lines identify potential polyadenylation signals. Putative signal peptide and transmembrane domain regions of MDV US6 (gD), -US7 (gI), and -US8 (gE) are overlined at the amino and carboxy ends, respectively. Signal peptide overlining continues through to the last amino acid to the left of the predicted cleavage site (von Heijne, 1985). Potential N-glycosylation sites (N-X-S/T) are indicated by dashed lines.

```
                                              <----IRs
                                              .......
     1  cagctgcgtattttcccccgtgcatctcataccgcccatttttgggtagaggtatattttattAGCCAAATCGTATTCCTGGAAGTTTGACAAAACCTGT  100
                                                                   89  *  G  F  R  I  G  P  L  K  V  F  G  T   78

    101 CAAACTGACACGGTCCAAGCGAAACTCGAAAAAAAAAGGGGGGGGGGGGAGAATATTCTGTAGGACCGGCAGAACTTCTCAAGGCAGAGGAAAGATACACA  200
    78    L  S  V  R  D  L  R  F  E  F  F  F  P  P  P  P  S  Y  E  T  P  G  A  S  S  R  L  A  S  S  L  Y  V    45

    201 TTATTTTTTGTTAGATTTAGGCAAGTTTTGCAGAACCTGCAGGGAATGTATACACCATCAAATCTACTCGACTTATTGCTTGAGTCCAATTTAACAGAAA  300
    44    N  N  K  T  L  N  L  C  T  K  C  F  R  C  P  I  Y  V  G  D  F  R  S  S  K  N  S  S  D  L  K  V  S    12

    301 TTAAAATATATTGATGTTGCGACATATGCATCCTCGCATATGGGGGTGGGACACAGGACGATTATATCCCCAGACATGAACCTCAAACTGCCATTTTGAT  400
SORF1 11  I  L  I  Y  Q  H  Q  S  M  H  M                                                                      1

    401 CCCATCATTGGAGAGACAAATTCGCATACATCCTACTTATCGCACACATTGGATGTCGGTCTTTATTCAGGCCATATCAGCTTTCACGGGGGCAAATTCG  500

SORF2                         M  Q  R  Q  T  G  H  M  E  D  K  K  R  T  G  L  E  S  Q  G  T  E  N  A  F  S  D    27
    501 TATTCATAGATCCGTCATCGATGCAGCGCCAAACCGGACATATGGAAGACAAAAAGAGAACCGGTTTGGAATCGCAGGGGACCGAGAATGCTTTTTCAGA  600

    28    G  R  D  G  K  D  G  L  L  H  E  G  I  N  E  P  I  L  I  P  S  T  I  A  D  L  E  G  I  R  E  L  V    60
    601 TGGCAGAGATGGCAAAGATGGATTGTTACATGAAGGAATTAATGAGCCCATTTTGATTCCGTCTACCATCGCAGATCTCGAGGGGATTCGTGAATTGGTC  700

    61    R  K  F  R  G  R  L  L  P  F  E  K  C  P  D  F  C  L  R  I  G  G  L  E  A  S  F  H  K  G  Q  E  E  L  94
    701 CGAAAATTCCGTGGTCGTCTACTGCCCTTTGAAAAGTGTCCCGATTTTTGTCTGAGAATTGGGGGTTTGGAGGCCAGCTTTCATAAAGGGCAGGAGGAGC  800

    95    L  E  Y  C  E  A  L  Y  L  P  Q  P  V  K  M  E  I  V  G  I  V  D  D  V  P  C  L  A  T  G  M  Q  L   127
    801 TGTTAGAGTATTGTGAAGCACTTTATTTACCACAACCTGTTAAGATGGAAATAGTAGGCATTGTAGACGATGTGCCATGTCTGGCAACGGGGATGCAATT  900

    128   L  I  L  V  A  E  G  G  E  V  Y  A  Y  E  E  D  T  L  H  K  L  A  T  S  F  S  E  F  L  E  I  G  V   160
    901 ACTCATTCTTGTTGCCGAGGGGGGAGAGGTATATGCCTATGAAGAAGATACTCTGCATAAGTTAGCCACGAGTTTTTCCGAATTCCTTGAAATTGGAGTG 1000

    161   K  S  L  G  R  E  V  Y  H  C  G  E  Y  I  E  Q  V  V  H  *  179
   1001 AAATCTTTAGGGAGGGAGGTTTACCATTGTGGAGAATATATAGAGCAAGTAGTACATTAGGGGCTGGGTTAAAGACCAAGTAATTTTTGACCGGATATCA 1100
   1101 CGTGATGTAAATTCTAGCAATTATTGTTCCTAGCAGAAGATAAAAGCTGGTAGCTATATAATACAGGCCAAAGTCTCCAAATTACACTTGAGCAGAAAC 1200

US1                           M  S  R  D  R  D  R  A  R  P  D  T  R  L  S  S  S  D  N  E  S  D  D  E  D    25
   1201 CTGCTTTCGGCTCCATCGGAGGCAACATGAGTCGTGATCGAGATCGAGCCAGACCCGATACACGATTATCATCGTCAGATAATGAGAGCGACGACGAAGA 1300

    26    Y  Q  L  P  H  S  H  P  E  Y  G  S  D  S  S  D  Q  D  F  E  L  N  N  V  G  K  F  C  P  L  P  W  K    58
   1301 TTATCAACTGCCACATTCACATCCGGAATATGGCAGTGACTCGTCCGATCAAGACTTTGAACTTAATAATGTGGGCAAATTTTGTCCTCTACCATGGAAA 1400

    59    P  D  V  A  R  L  C  A  D  T  N  K  L  F  R  C  F  I  R  C  R  L  N  S  G  P  F  H  D  A  L  R  R  A  92
   1401 CCCGATGTCGCTCGGTTATGTGCGGATACAAACAAACTATTTCGATGTTTTATTCGATGTCGACTAAATAGCGGTCCGTTCCACGATGCTCTTCGGAGAG 1500

    93    L  F  D  I  H  M  I  G  R  M  G  Y  R  L  K  Q  A  E  W  E  T  I  M  N  L  T  P  R  Q  S  L  H  L   125
   1501 CACTATTCGATATTCATATGATTGGTCGAATGGGATATCGACTAAAACAAGCCGAATGGGAAACTATCATGAATTTGACCCCACGCCAAAGTCTACATCT 1600

    126   R  R  T  L  R  D  A  D  S  R  S  A  H  P  I  S  D  I  Y  A  S  D  S  I  F  H  P  I  A  A  S  S  G   158
   1601 GCGCAGGACTCTGAGGGATGCTGATAGTCGAAGCGCCCATCCTATATCCGATATATATGCCTCCGATAGCATTTTTCACCCAATCGCTGCGTCCTCGGGA 1700

    159   T  I  S  S  D  C  D  V  K  G  M  N  D  L  S  V  D  S  K  L  H  *  179
   1701 ACTATTTCTTCAGACTGCGATGTAAAAGGAATGAACGATTTGTCGGTAGACAGTAAATTGCATTAACTATCCAGACTTGAAGAGAAAGCTCTTATTATAT 1800
   1801 AATTTTAATTGTTAGACATAGAGCCGACATTCTTTGATCTATCTAATGAGATAAAATAATAGATTTTGGATTTATTTGTCATGATCTGTTGCAACAAACG 1900
   1901 CTGACCCCCCCCATCCATGAAGGGGCGTGTCAAATAACGTGTTGCCTTTTTGTTGTATATGAAGATATTTAATGTGGCGTTGAGCCTAATGAGAGGAGAA 2000

US10     1                                       M  A  M  W  S  L  R  R  K  S  S  R  S  V  Q    15
   2001 CGTGTTTGAATACTGGAGACGAGCGCCGTGTAAGATTAAAACATATTGGAGAGGTATGGCCATGTGGTCTCTACGGCGCAAATCTAGCAGGAGTGTGCAA 2100

    16    L  R  V  D  S  P  K  E  Q  S  Y  D  I  L  S  A  G  G  E  H  V  A  L  L  P  K  S  V  R  S  L  A  R  T  49
   2101 CTCCGGGTAGATTCTCCAAAAGAACAGAGTTATGATATACTTTCTGCCGGCGGGGAACATGTTGCGCTATTGCCTAAATCTGTACGCAGTCTAGCCAGGA 2200

    50     I  L  T  A  A  T  I  S  Q  A  A  M  K  A  G  K  P  P  S  S  R  L  W  G  E  I  F  D  R  M  T  V  T   82
   2201 CCATATTAACCGCCGCTACGATCTCCCAGGCTGCTATGAAAGCTGGAAAACCACCATCGTCTCGTTTGTGGGGTGAGATATTCGACAGAATGACTGTCAC 2300

    83    L  N  E  Y  D  I  S  A  S  P  F  H  P  T  D  P  T  R  K  I  V  G  R  A  L  R  C  I  E  R  A  P  L   115
   2301 GCTTAACGAATATGATATTTCTGCTTCGCCATTCCACCCGACAGACCCGACGAGAAAAATTGTAGGCCGGGCTTTACGGTGTATTGAACGTGCTCCTCTT 2400

    116   T  H  E  E  M  D  T  R  F  T  I  M  M  Y  W  C  C  L  G  H  A  G  Y  C  T  V  S  R  L  Y  E  K  N  V  149
   2401 ACACACGAAGAAATGGACACTCGGTTTACTATCATGATGTATTGGTGTTGTCTTGGACATGCTGGATACTGTACTGTTTCGCGCTTATATGAGAAGAATG 2500

    150   R  L  M  D  I  V  G  S  A  T  G  C  G  I  S  P  L  P  E  I  E  S  Y  W  K  P  L  C  R  A  V  A  T   182
   2501 TCCGTCTTATGGACATAGTAGGTTCGGCAACGGGCTGTGGAATAAGTCCACTCCCCGAAATAGAGTCTTATTGGAAACCTTTATGTCGTGCCGTCGCTAC 2600

    183   K  G  N  A  A  I  G  D  D  A  E  L  A  H  Y  L  T  N  L  R  E  S  P  T  G  D  G  E  S  Y  L  *  213
   2601 TAAGGGGAATGCAGCAATCGGTGATGATGCTGAATTGGCACATTATCTGACAAATCTTCGGGAATCGCCAACAGGAGACGGGGAATCCTACTTATAACTA 2700

                                                                     .......  .......
   2701 ATCGCACAATTATTAATAGGATTTTAGGAAAAACTGCTACTAACGTTGTTTAAATAATAAAATTTTATTTTCAATAAGGCATTACAGTGTTGTCATGATT 2800
   2801 GTATGTATTATATGGGGTATGCATGAGGATTACTTCGATTGAAACTTTGTCTAAATGTCTGTAGGATTTTACTATTCATTAGTCTGGATCGAGGCGGACG 2900
          351  *  I  P  Y  A  H  P  N  S  R  N  F  S  Q  R  F  T  Q  L  I  K  S  N  M  L  R  S  R  P  P  R   322

   2901 TAAATGGAGATTGCGGCAAATGTAGGGGTGCTGGTACATAAGACCTCCAACATCCATTCGACTCATCGGCCTGCGTCCAAATGGATATGTTGATGTACCT 3000
    321   L  H  L  N  R  C  I  Y  P  H  Q  Y  M  L  G  G  V  D  M  R  S  M  P  R  R  G  F  P  Y  T  S  T  G   289

   3001 TGTAAAGTTATGACATTAGAAGATCGATGGTGAATAGTGGGATCTATATCCATGCTATTCTCAATATTGCATGATATGCAATGTTCCCGGTTAGGTTTGA 3100
    288   Q  L  T  I  V  N  S  S  R  H  H  I  T  P  D  I  D  M  S  N  E  I  N  C  S  I  C  H  E  R  N  P  K   256

                                                        .......
   3101 TAAGATCATGTATGGTTCTATAATACAACTCCTCTTCAGAAGAATCATTTATTTTTATGTCCACTGTCCTTGGATATTCCAGTTTCTGTCAATCGATTCGC 3200
    255   I  L  D  H  I  T  R  Y  Y  L  E  E  E  S  S  D  N  I  K  H  G  S  D  K  S  I  G  T  E  T  L  R  N  A  222
```

```
3201 TTGCATTTGCGTGCAGCATGTCTTGATGGCATTTCCTATGCTATCATCCGGCAGGCCTAAGGGTGTTCTATACTCGCACACAGGTAGAGCAAGAACCACG 3300
 221  Q  M  Q  T  C  C  T  K  I  A  N  G  I  S  D  D  P  L  G  L  P  T  R  Y  E  C  V  P  L  A  L  V  V   189

3301 GCATATCGAGCTACCTCTATTGCCCCGCTAAGGACATTTCTTGCAGACTGTATTGTCATGAACATATTTCGTGTATTGTGTCGATCATAACCCTTGTTGA 3400
 188  A  Y  R  A  V  E  I  A  G  S  L  V  N  R  A  S  Q  I  T  M  F  M  N  R  T  N  H  R  D  Y  G  K  N   156

3401 TTCCTATGGAAAGCATTGTGGTCCAGTTTTCCAGATGAAATGAAAACAATGCGGGCAAAAATGGTCCCACCTGTTTCATCTTCAATGCATCTCTCACATC 3500
 155  I  G  I  S  L  M  T  T  W  N  E  L  H  F  S  F  L  A  P  L  F  P  G  V  Q  K  M  K  L  A  D  R  V  D  122

3501 CCAAGTTCTATAGAATATTCTCCACTGACCAGTTTCGGTAAGATCAGTTTCTGTAAAATTTGTGATAGTTTCAATCGAAAACATTTTGTCCATCATGGCA 3600
 121  W  T  R  Y  F  I  R  W  Q  G  T  E  T  L  D  T  E  T  F  N  T  I  T  E  I  S  F  M  K  D  M  M  A   89

3601 AAAAATCTATAGGCAGACCAGATAACCATTTGACACCACATATCCTTGTGTATATCAAACGATGTAATAGATCCCTCGTTAGTAGATATGGTACATAAAA 3700
  88  F  F  R  Y  A  S  W  I  V  M  Q  C  W  M  D  K  H  I  D  F  S  T  I  S  G  E  N  T  S  I  T  C  L   56

3701 GGCCTAATCTCTCTCGGGCTTCCATACATTGAACGATTCCTTCTGTGAATTCATCAACAACCACATGCCAAAAATTTACATTAGTAATCTTTCTCGGTGG 3800
  55  L  G  L  R  E  R  A  E  M  C  Q  V  I  G  E  T  F  E  D  V  V  V  H  W  F  N  V  N  T  I  K  R  P  P   22

3801 CTTACCAAATCGTCCTCTTGGTATATCCATATCATCGAACATTGTAGCATTGACTCTGCTCATCGTTGTCTTTCAAATGCGCTCGATTGTTGAATCTCTC 3900
SORF3 21  K  G  F  R  G  R  P  I  D  M  D  D  F  M  T  A  N  V  R  S  M                                       1

3901 CTGATGTTAGAAGTATATGGAAGATAGCCTGGATACATAAGTGATCTAGAAGGGTTTGTTATTGCACTAATATACAAATTATACGTGACACTATAGCGAC 4000
4001 GGTTGTAGCGATGCACCTAATCGTAATGTGTATACGCCCCATCATGTAATTATATCTAATTGGTAGCAAGTAGGTCTGTCGAATAACAGCTAATGACTAC 4100
                                                                                        270  *  H  S      269

4101 CGGCTCTACATTTTTTCTGTATTCGTGACTTTCCTGTCGCAGTGTAACGAACCGGAATTGCAATCGCATCTCTATCTTCTTTCTTGCAACATTTTCCACA 4200
 268  G  A  R  C  K  K  Q  I  R  S  K  G  T  A  T  Y  R  V  P  I  A  I  A  D  R  D  E  K  K  C  C  K  G  C  235

4201 ACAGAATAATCTGCCGGGTGTACTACTCATTTGAGGTGGTTCGATTTCCGGAGGTTTTAGAGGATTGGGTGGGGGACCCGAGGATTTTGTATACACATACC 4300
 234   C  F  L  R  G  P  T  S  S  M  Q  P  P  E  I  E  P  P  K  L  P  N  P  P  S  G  L  I  K  Y  V  C  V   202

4301 ATATCACTGTCGCAAAAATGCGCTCTATCTTCTGGGGTGTCGAACTTCGGTTCCCATGTAGATGTCAAGAGAGTTTGAATATTGTCGGGAATGGCCCACG 4400
 201  M  D  S  D  C  F  H  A  R  D  E  P  T  D  F  K  P  E  W  T  S  T  L  L  T  Q  I  N  D  P  I  A  W   169

4401 GCATACCGGACCAGGTCCCAGACACTTTGATTGCAAGTAACCTTTTTGGCAAAGGAATACATTCGAGCGCAATGGCACATATATCTGCCGCCCCAACTAT 4500
 168  P  M  G  S  W  T  G  S  V  K  I  A  L  L  R  K  P  L  P  I  C  E  L  A  I  A  C  I  D  A  A  G  V  I  135

4501 CCACAAGCTATGTGGAGCATTACCAGAAACTTCAGATTCCAACATCAAATATCCAGATAGAACATCCTGCCATTCTGTGGAACATCCTGCAACATCTTCA 4600
 134  W  L  S  H  P  A  N  G  S  V  E  S  E  L  M  L  Y  G  S  L  V  D  Q  W  E  T  S  C  G  A  V  D  E   102

4601 AATAGCCGCACTATAAACGAATCCCTAGTTCCGGCCAATCCGGTACCACGAACTCCAGTTCCATCTGGTGGCTTTGTCCTTACTATCGGTCGATGTTGCC 4700
 101  F  L  R  V  I  F  S  D  R  T  G  A  L  G  T  G  R  V  G  T  G  D  P  P  K  T  R  V  I  P  R  H  Q   69

4701 GAGGAAGAATTAACATGGGTTTGGCAAAACGGAATAGGTCTGCAGCTCTGGCGATTATGGGCACACCCACATCATCCTGTATTTGTTCCATACATTGCTT 4800
  68  R  P  L  I  L  M  P  K  A  F  R  F  L  D  A  A  R  A  I  I  P  V  G  V  D  D  Q  I  Q  E  M  C  Q  K  35

4801 TATAAGGAATATCCATAAAGTAGATGCAGCATCTCTAGATCTTCCTGGCAATCGATCGCATTCATCTAGAAGTGTGACTATAGTTATCATGGACACACCC 4900
  34   I  L  F  I  W  L  T  S  A  A  D  R  S  R  G  P  L  R  D  C  E  D  L  L  T  V  I  T  I  M  S  V  G   2

4901 ATCTTCACCTCCACCAATAATCTTTTTTATTGTTAATAACTGGGCCGGTCTGATCTCCAAATCTTATACTCTGGTAGAATATGAAACAGGGTTAAAACTA 5000
US2  1  M
US3  1                 M  S  S  T  P  E  A  E  T  M  E  C  G  I  S  S  S  K  V  H  D  S  K  T  N  T  T  Y  G   29
(PK) 5001 GGTAATAGACTGGATGTCTTCGACTCCGGAGGCAGAAACGATGGAATGTGGCATTTCTTCGTCGAAAGTACACGACTCTAAAACTAATACTACCTACGGA 5100

  30  I  I  H  N  S  I  N  G  T  D  T  T  L  F  D  T  F  P  D  S  T  D  N  A  E  V  T  G  D  V  D  D  V  K   63
5101 ATTATACATAACAGCATCAATGGTACGGATACGACGTTGTTTGATACTTTTCCCGACAGTACCGATAACGCGGAAGTGACGGGGGATGTGGACGATGTGA 5200

  64     T  E  S  S  P  E  S  Q  S  E  D  L  S  P  F  G  N  D  G  N  E  S  P  E  T  V  T  D  I  D  A  V  S   96
5201 AGACTGAGAGCTCTCCCGAGTCCCAATCTGAAGATTTGTCACCTTTTGGGAACGATGGAAATGAATCCCCCGAAACGGTGACGGACATTGATGCAGTTTC 5300

  97  A  V  R  M  Q  Y  N  I  V  S  S  L  P  P  G  S  E  G  Y  I  Y  V  C  T  K  R  G  D  N  T  K  R  K   129
5301 AGCTGTGCGAATGCAGTATAACATTGTTTCATCGTTACCGCCCGGATCTGAAGGGTATATCTATGTTTGTACAAAGCGTGGGGATAATACCAAGAGAAAA 5400

 130  V  I  V  K  A  V  T  G  G  K  T  L  G  S  E  I  D  I  L  K  K  M  S  H  R  S  I  I  R  L  V  H  A  Y   163
5401 GTCATTGTGAAAGCTGTGACTGGTGGCAAAACCCTTGGGAGTGAAATTGATATATTAAAAAAAAATGTCTCACCGCTCCATAATTAGATTAGTTCATGCTT 5500

 164     R  W  K  S  T  V  C  M  V  M  P  K  Y  K  C  D  L  F  T  Y  I  D  I  M  G  P  L  P  L  N  Q  I  I   196
5501 ATAGATGGAAATCGACAGTTTGTATGGTAATGCCTAAATACAAATGCGACTTGTTTACGTACATAGATATCATGGGACCATTGCCACTAAATCAAATAAT 5600

 197     T  I  E  R  G  L  L  G  A  L  A  Y  I  H  E  K  G  I  I  H  R  D  V  K  T  E  N  I  F  L  D  K  P   229
5601 TACGATAGAACGGGGTTTGCTTGGAGCATTGGCATATATCCACGAAAAGGGTATAATACATCGTGATGTAAAAACTGAAAATATATTTTTGGATAAACCT 5700

 230  E  N  V  V  L  G  D  F  G  A  A  C  K  L  D  E  H  T  D  K  P  K  C  Y  G  W  S  G  T  L  E  T  N  S   263
5701 GAAAATGTAGTATTGGGGGACTTTGGGGCAGCATGTAAATTAGATGAACATACAGATAAACCCAAATGTTATGGATGGAGTGGAACTCTGGAAACCAATT 5800

 264     P  E  L  L  A  L  D  P  Y  C  T  K  T  D  I  W  S  A  G  L  V  L  F  E  M  S  V  K  N  I  T  F  F   296
5801 CGCCTGAACTGCTTGCACTTGATCCATACTGTACAAAAACTGATATATGGAGTGCAGGATTAGTTCTGTTTGAGATGTCAGTAAAAAAATATAACCTTTT 5900

 297  G  K  Q  V  N  G  S  G  S  Q  L  R  S  I  I  R  C  L  Q  V  H  P  L  E  F  P  Q  N  N  S  T  N  L   329
5901 TGGCAAACAAGTAAACGGCTCAGGTTCTCAGCTGAGATCCATAATTAGATGCCTGCAAGTCCATCCGTTGGAATTTCCACAGAACAATTCTACAAACTTA 6000

 330  C  K  H  F  K  Q  Y  A  I  Q  L  R  H  P  Y  A  I  P  Q  I  I  R  K  S  G  M  T  M  D  L  E  Y  A  I   363
6001 TGCAAACACTTCAAGCAGTACGCGATTCAGTTACGACATCCATATGCAATCCCTCAGATTATACGAAAGAGTGGTATGACGATGGATCTTGAATATGCTA 6100

 364     A  K  M  L  T  F  D  Q  E  F  R  P  S  A  Q  D  I  L  M  L  P  L  F  T  K  E  P  A  D  A  L  Y  T   396
6101 TTGCAAAAATGCTCACATTCGATCAGGAGTTTAGACCATCTGCCCAAGATATTTTAATGTTGCCTCTTTTTACTAAAGAACCCGCTGACGCATTATACAC 6200

 397  I  T  A  A  H  M  *  402
6201 GATAACTGCCGCTCATATGTAAACACCCGTCAAAAATAACTTCAATGATTCATTTTATAATATATACTACGCGTTACCTGCAATAATGACAACATTCGAA 6300
```

FIG. 2—Continued

SORF4
```
                                        M  A  P  S  G  P  T  P  Y  S  H  R  P  Q  I  K  H  Y  G  T  F  S  D        23
   1
6301 GTCTTTGAAGATTCGCAGACCTTTTTTGCGAATGGCACCTTCGGGACCTACGCCATATTCCCACAGACCGCAAATAAAGCATTATGGAACATTTTCGGAT     6400
                                                                          ......

  24 C  M  R  Y  T  L  N  D  E  S  K  V  D  D  R  C  S  D  I  H  N  S  L  A  Q  S  N  V  T  S  S  M  S  V      57
6401 TGCATGAGATATACTCTAAACGATGAGAGTAAGGTAGATGATAGATGTTCAGACATACATAACTCCTTAGCACAATCCAATGTTACTTCAAGCATGTCTG    6500

  58 M  N  D  S  E  E  C  P  L  I  N  G  P  S  M  Q  A  E  D  P  K  S  V  F  Y  K  V  R  K  P  D  R  S        90
6501 TAATGAACGATTCGGAAGAATGTCCATTAATAAATGGACCTTCGATGCAGGCAGAGGACCCTAAAAGTGTTTTTTATAAAGTTCGTAAGCCTGACCGAAG    6600

  91 R  D  F  S  W  Q  N  L  N  S  H  G  N  S  G  L  R  R  E  K  Y  I  R  S  S  K  R  R  W  K  N  P  E       123
6601 TCGTGATTTTTCATGGCAAAATCTGAACTCCCATGGCAATAGTGGTCTACGTCGTGAAAAATATATACGTTCCTCTAAGAGGCGATGGAAGAATCCCGAG    6700

 124 I  F  K  V  S  L  K  C  E  S  I  G  A  G  N  G  I  K  I  S  F  S  F  F  *  147
6701 ATATTTAAGGTATCTTTGAAATGTGAATCAATTGGCGCTGGTAACGGAATAAAAATTTCATTCTCATTTTTCTAACATTATAATATATCAGATCGTTTCT    6800
                                                                        ......

6801 TATATACTTATTTTCATCGTCGGGATATGACTAACGTATACTAAGTTACAAGAAACAACTGCTTAACGTCGAACATAACGGAAATAAAAATATATATAGC    6900
```

US6
(gD)
```
                                     M  N  R  Y  R  Y  E  S  I  F  F  R  Y  I  S  S  T  R  M  I        20
   1
6901 GTCTCCTATAACTGTTATATTGGCACCTTTTAGAGCTTCGGTATGAATAGATACAGATATGAAAGTATTTTTTTTAGATATATCTCATCCACGAGAATGA    7000

  21 L  I  T  C  L  L  L  G  T  G  D  M  S  A  M  G  L  K  K  D  N  S  P  I  I  P  T  L  H  P  K  G  N       53
7001 TTCTTATAATCTGTTTACTTTTGGGAACTGGGGACATGTCCGCAATGGGACTTAAGAAAGACAATTCTCCGATCATTCCCACATTACATCCGAAAGGTAA    7100

  54 E  N  L  R  A  T  L  N  E  Y  K  I  P  S  P  L  F  D  T  L  D  N  S  Y  E  T  K  H  V  I  Y  T  D       86
7101 TGAAAACCTCCGGGCTACTCTCAATGAATACAAAATCCCGTCTCCACTGTTTGATACACTTGACAATTCATATGAGACAAAACACGTAATATATACGGAT    7200

  87 N  C  S  F  A  V  L  N  P  F  G  D  P  K  Y  T  L  L  S  L  L  L  M  G  R  R  K  Y  D  A  L  V  A  W    120
7201 AATTGTAGTTTTGCTGTTTTGAATCCATTTGGCGATCCGAAATATACGCTTCTCAGTTTACTGTTGATGGGACGACGCAAATATGATGCTCTAGTAGCAT    7300

 121 F  V  L  G  R  A  C  G  R  P  I  Y  L  R  E  Y  A  N  C  S  T  N  E  P  F  G  T  C  K  L  K  S  L      153
7301 GGTTTGTCTTGGGCAGAGCATGTGGGAGACCAATTTATTTACGTGAATATGCCAACTGCTCTACTAATGAACCATTTGGAACTTGTAAATTAAAGTCCCT    7400

 154 G  W  D  R  R  Y  A  M  T  S  Y  I  D  R  D  E  L  K  L  I  I  A  A  P  S  R  E  L  S  G  L  Y        186
7401 AGGATGGTGGGATAGAAGATATGCAATGACGAGTTATATCGATCGAGATGAATTGAAATTGATTATTGCAGCACCCAGTCGTGAGCTAAGTGGATTATAT    7500

 187 T  R  L  I  I  I  N  G  E  P  I  S  S  D  I  L  L  T  V  K  G  T  C  S  F  S  R  R  G  I  K  D  N  K    220
7501 ACGCGTTTAATAATTATTAATGGAGAACCCATTTCGAGTGACATATTACTGACTGTTAAAGGAACATGTAGTTTTTCGAGACGGGGGATAAAGGATAACA    7600

 221 L  C  K  P  F  S  F  F  V  N  G  T  T  R  L  L  D  M  V  R  T  G  T  P  R  A  H  E  E  N  V  K  Q      253
7601 AACTATGCAAACCGTTCAGTTTTTTTGTCAATGGTACAACACGGCTGTTAGACATGGTGCGAACAGGAACCCCGAGAGCCCATGAAGAAAATGTGAAGCA    7700

 254 W  L  E  R  N  G  G  K  H  L  P  I  V  V  E  T  S  M  Q  Q  V  S  N  L  P  R  S  F  R  D  S  Y  L     286
7701 GTGGCTTGAACGAAATGGTGGTAAACATCTACCAATCGTCGTCGAAACATCTATGCAACAAGTCTCAAATTTGCCGAGAAGTTTTAGAGATTCATATTTA    7800

 287 K  S  P  D  D  D  K  Y  N  D  V  K  M  T  S  A  T  T  N  N  I  T  T  S  V  D  G  Y  T  G  L  T  N  R   320
7801 AAATCACCTGACGACGATAAATATAATGACGTCAAAATGACATCGGCCACTACTAATAACATTACCACCTCCGTGGATGGTTACACTGGACTCACTAATC    7900

 321 P  E  D  F  E  K  A  P  Y  I  T  K  R  P  I  I  S  V  E  E  A  S  S  Q  S  P  K  I  S  T  E  K  K     353
7901 GGCCCGAGGACTTTGAGAAAGCACCATACATAACTAAACGACCGATAATCTCTGTCGAGGAGGCATCCAGTCAATCACCTAAAATATCAACAGAAAAAAA    8000

 354 S  R  T  Q  I  I  I  S  L  V  V  L  C  V  M  F  C  F  I  V  I  G  S  G  I  W  I  L  R  K  H  R  K     386
8001 ATCCCGAACGCAAATAATAATTTCACTAGTTGTTCTATGCGTCATGTTTTGTTTCATTGTAATCGGGTCTGGTATATGGATCCTTCGCAAACACCGCAAA    8100

 387 T  V  M  Y  D  R  R  R  P  S  R  R  A  Y  S  R  L  *  403
8101 ACGGTGATGTATGATAGACGTCGTCCATCAAGACGGGCATATTCCCGCCTATAACACGTGTTTGGTATGGGCGTGTCGCTATAGTGCATAAGAAGTTGAC    8200
```

US7
(gI)
```
                                        M  Y  V  L  Q  L  L  F  W  I  R  L  F  R       14
   1
8201 TACATTGATCAATGACATTATATAGCTTCTTTGGTCAGATAGACGGCGTGTGTGATTGCGATGTATGTACTACAATTATTATTTTGGATCCGCCTCTTTC    8300

  15 G  I  W  S  I  V  Y  T  G  T  S  V  T  L  S  T  D  Q  S  A  L  V  A  F  R  G  L  D  K  M  V  N  V       47
8301 GAGGCATCTGGTCTATAGTTTATACTGGAACATCTGTTACGTTATCAACGGACCAATCTGCTCTTGTTGCGTTCCGCGGATTAGATAAAATGGTGAATGT    8400

  48 R  G  Q  L  L  F  L  G  D  Q  T  R  T  S  S  Y  T  G  T  T  E  I  L  K  W  D  E  E  Y  K  C  Y  S      80
8401 ACGCGGCCAACTTTTATTCCTGGGCGACCAGACTCGGACCAGTTCTTATACAGGAACGACGGAAATCTTGAAATGGGATGAAGAATATAAATGCTATTCC    8500

  81 V  L  H  A  T  S  Y  M  D  C  P  A  I  D  A  T  V  F  R  G  C  R  D  A  V  V  Y  A  Q  P  H  G  R  V   114
8501 GTTCTACATGCGACATCATATATGGATTGTCCTGCTATAGACGCCACGGTATTCAGAGGCTGTAGAGACGCTGTGGTATATGCTCAACCTCATGGTAGAG    8600

 115 Q  P  F  P  E  K  G  T  L  L  R  I  V  E  P  R  V  S  D  T  G  S  Y  Y  I  R  V  S  L  A  G  R  N     147
8601 TACAACCTTTTCCCGAAAAGGGAACATTGTTGAGAATTGTCGAACCCAGAGTATCAGATACAGGCAGCTATTACATACGTGTATCTCTCGCTGGAAGAAA    8700

 148 M  S  D  I  F  R  M  V  V  I  I  R  S  S  K  S  W  A  C  N  H  S  A  S  S  F  Q  A  H  K  C  I  R     180
8701 TATGAGCGATATATTTAGAATGGTTGTTATTATAAGGAGTAGCAAATCTTGGGCCTGTAATCACTCTGCTAGTTCATTTCAGGCCCATAAATGTATTCGC    8800

 181 Y  V  D  R  M  A  F  E  N  Y  L  I  G  H  V  G  N  L  L  D  S  D  S  E  L  H  A  I  Y  N  I  T  P  Q   214
8801 TATGTCGACCGTATGGCCTTTGAAAATTATCTGATTGGACATGTAGGCAATTTGCTGGACAGTGACTCGGAATTGCATGCAATTTATAATATTACTCCCC    8900

 215 S  I  S  T  D  I  N  I  V  T  T  P  F  Y  D  N  S  G  T  I  Y  S  P  T  V  F  N  L  F  N  N  N  S     247
8901 AATCCATTTCCACAGATATTAATATTGTAACGACTCCATTTTACGATAATTCGGGAACAAATTTATTCACCTACGGTTTTTAATTTGTTTAATAACAATTC    9000

 248 H  V  D  A  M  N  S  T  G  M  W  N  T  V  L  K  Y  T  L  P  R  L  I  Y  F  S  T  M  I  V  L  C  I     280
9001 CCATGTCGATGCAATGAATTCGACTGGTATGTGGAATACCGTTTTAAAATATACCCTTCCAAGGCTTATTTACTTTTCTACGATGATTGTACTATGTATA    9100

 281 I  A  L  A  I  Y  L  V  C  E  R  C  R  S  P  H  R  R  I  Y  I  G  E  P  R  S  D  E  A  P  L  I  T  S   314
9101 ATAGCATTGGCAATTTATTTGGTCTGTGAAAGGTGCCGCTCTCCCCATCGTAGGATATACATCGGTGAACCAAGATCTGATGAGGCCCCACTCATCACTT    9200

 315 A  V  N  E  S  F  Q  Y  D  Y  N  V  K  E  T  P  S  D  V  I  E  K  E  L  M  E  K  L  K  K  K  V  E     347
9201 CTGCAGTTAACGAATCATTTCAATATGATTATAATGTAAAGGAAACTCCTTCAGATGTTATTGAAAAGGAGTTGATGGAAAAACTGAAGAAGAAAGTCGA    9300
```

FIG. 2—Continued

```
        348   L  L  E  R  E  E  C  V  *  355
        9301  ATTGTTGGAAAGAGAAGAATGTGTATAGGTTTGAGAAACTATTATAGGTAGGTGGTACCTGTTAGCTTAGTATAAGGGGAGGAGCCGTTTCTTGTTTTAA   9400

US8       1                                                              M  C  V  F  Q  I  L  I  I  V  T  T   12
(gE)    9401  AGACACGAACACAAGGCCGTAAGTTTTATATGTGAATTTTGTGCATGTCTGCGAGTCAGCGTCATAATGTGTGTTTTCCAAATCCTGATAATAGTGACGA   9500

         13   I  K  V  A  G  T  A  N  I  N  H  I  D  V  P  A  G  H  S  A  T  T  T  I  P  R  Y  P  P  V  V  D  G   45
        9501  CGATCAAAGTAGCTGGAACGGCCAACATAAATCATATAGACGTTCCTGCAGGACATTCTGCTACAACGACGATCCCGCGATATCCACCAGTTGTCGATGG   9600

         46   T  L  Y  T  E  T  W  T  W  I  P  N  H  C  N  E  T  A  T  G  Y  V  C  L  E  S  A  H  C  F  T  D  L   78
        9601  GACCCTTTACACCGAGACGTGGACATGGATTCCCAATCACTGCAACGAAACGGCAACAGGCTATGTATGTCTGGAAAGTGCTCACTGTTTTACCGATTTG   9700

         79   I  L  G  V  S  C  M  R  Y  A  D  E  I  V  L  R  T  D  K  F  I  V  D  A  G  S  I  K  Q  I  E  S  L  S   112
        9701  ATATTAGGAGTATCCTGCATGAGGTATGCGGATGAAATCGTCTTACGAACTGATAAATTTATTGTCGATGCGGGATCCATTAAACAAATAGAATCGCTAA   9800

        113   L  N  G  V  P  N  I  F  L  S  T  K  A  S  N  K  L  E  I  L  N  A  S  L  Q  N  A  G  I  Y  I  R  Y   145
        9801  GTCTGAATGGAGTTCCGAATATATATTCCTATCTACGAAAGCAAGTAACAAGTTGGAGATACTAAATGCTAGCCTACAAAATGCGGGTATCTACATTCGGTA   9900

        146   S  R  N  G  T  R  T  A  K  L  D  V  V  V  V  G  V  L  G  Q  A  R  D  R  L  P  Q  M  S  S  P  M  I   178
        9901  TTCTAGAAATGGGACGAGGACTGCAAAGCTGGATGTTGTTGTGGTTGGCGTTTTGGGTCAAGCAAGGGATCGCCTACCCCAAATGTCCAGTCCTATGATC   10000

        179   S  S  H  A  D  I  K  L  S  L  K  N  F  K  A  L  V  Y  H  V  G  D  T  I  N  V  S  T  A  V  I  L  G  P   212
        10001 TCATCCCACGCCGATATCAAGTTGTCATTAAAAAAACTTTAAAGCATTAGTATATCACGTGGGAGATACTATCAATGTCTCGACGGCGGTTATACTAGGAC   10100

        213   S  P  E  I  F  T  L  E  F  R  V  L  F  L  R  Y  N  P  T  C  K  F  V  T  I  Y  E  P  C  I  F  H  P   245
        10101 CTTCTCCGGAGATATTCACATTGGAATTTAGGGTGTTGTTCCTCCGTTATAATCCAACGTGCAAGTTCGTCACGATTTATGAACCTTGTATATTTCACCC   10200

        246   K  E  P  E  C  I  T  T  A  E  Q  S  V  C  H  F  A  S  N  I  D  I  L  Q  I  A  A  A  R  S  E  N  C   278
        10201 CAAAGAACCAGAGTGTATTACTACTGCAGAACAATCGGTATGTCATTTCGCATCCAACATTGACATTCTGCAGATAGCCGCCGCACGTTCTGAAAATTGT   10300

        279   S  T  G  Y  R  R  C  I  Y  D  T  A  I  D  E  S  V  Q  A  R  L  T  F  I  E  P  G  I  P  S  F  K  M  K   312
        10301 AGCACAGGGTATCGTAGATGTATTTATGACACGGCTATCGATGAATCTGTGCAGGCCAGATTAACATTCATAGAACCAGGAATTCCTTCCTTTAAAATGA   10400

        313   D  V  Q  V  D  D  A  G  L  Y  V  V  V  A  L  Y  N  G  R  P  S  A  W  T  Y  I  Y  L  S  T  V  E  T   345
        10401 AAGATGTCCAGGTAGACGATGCTGGATTGTATGTGGTTGTGGCTTTATACAATGGACGTCCAAGTGCATGGACTTACATTTATTTGTCAACGGTGGAAAC   10500

        346   Y  L  N  V  Y  E  N  Y  H  K  P  G  F  G  Y  K  S  F  L  Q  N  S  S  I  V  D  E  N  E  A  S  D  W   378
        10501 ATATCTTAATGTATATGAAAACTACCACAAGCCGGGATTTGGGTATAAATCATTTCTACAGAACAGTAGTATCGTCGACGAAAATGAGGCTAGCGATTGG   10600

        379   S  S  S  S  I  K  R  R  N  N  G  T  I  I  Y  D  I  L  L  T  S  L  S  I  G  A  I  I  I  V  I  V  G  G   412
        10601 TCCAGCTCGTCCATTAAACGGAGAAATAATGGTACTATCATTTATGATATTTTACTCACATCGCTATCAATTGGGGCGATTATTATCGTCATAGTAGGGG   10700

        413   V  C  I  A  I  L  I  R  R  R  R  R  R  R  T  R  G  L  F  D  E  Y  P  K  Y  M  T  L  P  G  N  D  L   445
        10701 GTGTTTGTATTGCCATATTAATTAGGCGTAGGAGACGACGTCGCACGAGGGGGTTATTCGATGAATATCCCAAATATATGACGCTACCAGGAAACGATCT   10800

        446   G  G  M  N  V  P  Y  D  N  T  C  S  G  N  Q  V  E  Y  Y  Q  E  K  S  A  K  M  K  R  M  G  S  G  Y   478
        10801 GGGGGGCATGAATGTACCGTATGATAATACATGCTCTGGTAACCAAGTTGAATATTATCAAGAAAAGTCGGCTAAAATGAAAAGAATGGGTTCGGGTTAT   10900

        479   T  A  W  L  K  N  D  M  P  K  I  R  K  R  L  D  L  Y  H  *  497
        10901 ACCGCTTGGCTAAAAAATGATATGCCGAAAATTAGGAAACGCTTAGATTTATACCACTGATATGTACATATTTAAACTTAATGGGATATAGTATATGGAC   11000

        11001 GTCTATATGACGAGAGTAAATAAACTGACAATGCAAATGAAGCTGATCTATATTGTGCTTTATATTGGGACAAACCACTCGCACAAGCTCATTCAACACA   11100
        11101 TCCACTCTTGGACAGCTTCATGTTAAAATAAACTGTAAATCATTCAATGATAATGGGAGAAGAATGTGAGCAAGGATCCATGGTGTCTGCTTTTTATAGA   11200
        11201 TACTACCGCAATGCTACATATAAaataaaaatatacctctacccaaaaatgggcggtatgagatgcacggggaaaatacgcagctg   11286
              TRs---->
```

FIG. 2—Continued

ORFs (beginning from the first ATG codon) are shown relative to the nucleotide sequence in Fig. 2. Since their gene products are yet to be characterized, in order to simplify identification, 7 have been named (Fig. 1B, Table 1) based on homology (see below) to HSV-1-encoded U_S ORFs (McGeoch et al., 1985). When appropriate, the letters MDV will preface the homolog's name to indicate the ORF's origin. The four nonalphaherpesvirus-related ORFs have been arbitrarily named SORFs 1, -2, -3 and -4 (unique Short region Open Reading Frame). Due to the A–T-rich nature of MDV, there are numerous TATA-like sequences for transcriptional initiation, more than are likely to have functional relevance. Therefore, these sites have not been highlighted in Fig. 2. All of the ORFs contain potential AUG (ATG) codons in a favorable context for translational initiation (purine in −3 position and/or guanine in +4 position; Kozak, 1989). Such a context is lacking for the first methionine codons of SORF2 and US3 (PK), but is observed among secondary methionine codons which correspond to amino acid positions 8 and

10, respectively (Fig. 2). Potential polyadenylation signals, according to HSV-1 usage patterns (McGeoch, 1991), are identified in Fig. 2; ORF locations, lengths (AA), and predicted molecular masses of the putative translational products are outlined in Table 1.

Database- and computer-assisted homology comparisons

Using the computer program FASTA or TFASTA (Pearson and Lipman, 1988), each of the 11 predicted amino acid sequences was screened against the SwissProt protein database or GenBank/EMBL nucleic acid databases, respectively, in addition to recently published pseudorabies virus (PRV) (van Zijl et al., 1990) and equine herpesvirus-1 (EHV-1) (Colle et al., 1992; Elton et al., 1991; Holden et al., 1992a,b; Telford et al., 1992) S segment gene sequences not present in these databases. Optimized FASTA/TFASTA scores greater than 100 were initially considered as potential candidates possessing a

TABLE 1

Summary of MDV U$_s$ ORF Data

Name	ORF start	ORF stop Codons	M$_r$[a]	FASTA scores with				Properties/references
				HSV-1	VZ	PRV	EHV-1	
SORF1	331	62	89	10.1	—	—	—	—	—
SORF2	521	1060	179	20.1	—	—	—	—	Homologous to FPV ORF4[c]
US1	1227	1766	179	20.4	101	160	218	209	Regulatory protein[d]
US10	2056	2697	213	23.6	134	147	—	260	Virion protein[e]
SORF3	3863	2805	351	40.6	—	—	—	—	—
US2	4902	4090	270	29.7	335	—	168	355	—
US3	5014	6222	402	44.7	611	616	563	551	Protein kinase[f]
SORF4	6332	6775	147	16.8	—	—	—	—	—
US6	6943	8154	403	42.6[b]	211	—	279	246	Membrane glycoprotein D (gD)[g]
US7	8261	9328	355	38.3[b]	145	228	188	242	Membrane glycoprotein I (gI)[h]
US8	9467	10960	497	53.7[b]	192	376	217	402	Membrane glycoprotein E (gE)[h]

[a] Predicted, in absence of posttranslational modifications.
[b] Based on sequences that follow the predicted signal peptide cleavage site.
[c] FASTA = 237; Tomley et al. (1988).
[d] Holden et al. (1992a); Jackers et al. (1992); Sears et al. (1985).
[e] Holden et al. (1992b); McGeoch et al. (1988).
[f] Leader and Purves (1988); Purves et al. (1991); Zhang and Leader (1990).
[g] Campadelli-Fiume et al. (1990); Johnson and Spear (1989); Long et al. (1992); Peeters et al. (1993).
[h] Dubin et al. (1991); Johnson et al. (1988); Whealy et al. (1993); Zsak et al. (1992).

significant degree of amino acid similarity. The results of this analysis are in Table 1; the scores obtained are comparable to those of previously established S region homologies. Further evidence for homology was derived from dot matrix analyses and % similarity/identity analysis (using GAP; data not shown). Homologies between MDVs and their alphaherpesvirus S region counterparts ranged between 40 and 60% similarity and between 20 and 40% identity (data not shown). Apart from MDV US3, 6 ORFs (MDV US1, -10, -2, -6, -7 and -8) were found to be exclusively homologous to alphaherpesvirus S segment genes; in contrast, SORF1, -3, and -4 failed to show statistically significant homology with any sequences in either of the two databases. On the other hand, using SORF2 as a probe for FASTA analysis, a FASTA score of 237 was obtained, indicating homology to an uncharacterized fowlpox virus (FPV) ORF (e.g., FPV ORF4; Tomley et al., 1988). Upon alignment, these sequences were found to exhibit 67% similarity and 42% identity over the 100 AA aligned (Fig. 3). Like other US3 homologs, MDV's counterpart exhibits homology to the serine–threonine protein kinase superfamily (Hanks et al., 1988), as evidenced by a relatively large number of FASTA scores between 150 and 250. Nevertheless, these scores were three- to four-fold lower than those obtained between US3 homologs of HSV, VZV, and PRV (Table 1). The US3 gene family of herpesvirus protein kinases appear to define a distinct subfamily within the serine–threonine protein kinase superfamily; it is thought that related cellular counterparts exist and await future characterization (Hanks et al., 1988). Homologies to HSV-2 US2, -3, -6, and -7 are not

presented in this report, inasmuch as their ORFs exhibit greater than 70% identity to their HSV-1 counterparts (McGeoch et al., 1987) and result in homologies with MDV that basically resemble those with HSV-1. MDV US6 exhibits demonstrable homology to HSV-2 US4 (FASTA = 100) and its PRV counterpart, gX (FASTA = 90). This is consistent with earlier findings suggesting duplication and divergence of S component glycoprotein genes from common precursors (McGeoch, 1990).

Analysis of MDV glycoproteins, gD, gI, and gE

In comparing the gB homologs of seven different herpesviruses included in the alpha-, beta-, and gammaherpesvirus subfamilies, there is complete conservation of 10 cysteine residues (Ross et al., 1989). Alphaherpesvirus S component glycoproteins have also been found to contain similar patterns of conserved cysteine residues (McGeoch, 1990). HSV-1 US6 (gD) contains seven cysteine residues; six appear critical for correct folding, antigenic structure, and extent of oligosaccharide processing (Wilcox et al., 1988; Long et al., 1992). Not only are these same six cysteines conserved among gD homologs of HSV-2 (McGeoch et al., 1987), PRV (Petrovskis et al., 1986a), EHV-1 (Audonnet et al., 1990; Flowers et al., 1991; Telford et al., 1992), BHV-1 (Tikoo et al., 1990), and simian herpes B virus (SHBV; Bennett et al., 1992), but they are conserved by the MDV gD homolog as well (data not shown). Similar cysteine conservation patterns apply to alphaherpesvirus US7 (gI) and US8 (gE) homologs (McGeoch, 1990) and their MDV counterparts (data not shown).

```
MDV SORF2    83 LEASFHKGQEEL...LEYCEALYLPQPVKMEIVGIVDDVPCLATGMQLLI 129
                ::  ::  :|||     |:|| ||||||| :|:|:|:::|    :     |:|
FPV ORF4      1 MDRNINLPEEELKYIKECCEVLYLPQPTRMDIIGVMNDSD.ISWNENLII 49


MDV SORF2   130 LVAEGGEVYAYEEDTLHKLATSFSEFLEIGVKSLGREVYHCGEYIEQVVH 179
                |: |:| :| |::: | |:|  : || |||: || ||||| | |  :
FPV ORF4     50 LMSEDGKIYVYDDEALYKVADTMEEFSEIGLINLGNEVYHCREDIKPLPE 99
```

FIG. 3. Homology between MDV SORF2 and FPV ORF4. GAP (UWGCG) analysis aligning area conserved between SORF2 and fowlpox virus ORF4 (Tomley et al., 1988). Amino acid numbers (with respect to predicted 5' ATG) of aligned sequences are listed at the beginning and end of each line. Bars and double dots identify identical and similar amino acid matches, respectively. The area aligned was 67% similar, 42% identical; a FASTA score of 237 was obtained.

Careful inspection of the N-terminal regions of the MDV gD, gI, and gE homologs has revealed that all contain the three basic building blocks of signal peptide sequences: a basic, positively charged N-terminal region (n-region), a central hydrophobic region (h-region), and a more polar terminal region (c-region) that seems to define the cleavage site (von Heijne, 1985). Figure 2 shows the likely position of these sites (von Heijne, 1986). Also included are the locations of other characteristic features of membrane glycoproteins, namely, the presence of potential N-glycosylation sites (i.e., N-X-S/T) and putative hydrophobic transmembrane and charged cytoplasmic domains near the C-terminal end. Like other gI-homologs, MDV's counterpart contains a relatively long cytoplasmic domain. However, in contrast to the other gD homologs, MDV gD's signal peptide contains a longer n-region (18 residues) that is unusually highly charged (+4; Fig. 2) considering an overall mean value of +1.7 among eukaryotes, which generally does not vary with length (von Heijne, 1986). Although a methionine codon exists directly before the hydrophobic h-region at position 6997 in Fig. 2 (as in PRV's gD homolog; Petrovskis et al., 1986a), the scanning model for translation (Kozak, 1989) favors usage of the more 5'-proximal initiation codon (at position 6943, Fig. 2).

Comparison of MDV sequences to those previously published

Comparison of sequences of the "virulent" GA strain of MDV (Fig. 2) with those derived from a 5.5-kbp region of the "very virulent" RB1B strain of MDV (Ross et al., 1991) has revealed over 99% identity at both the nucleic acid and the predicted amino acid levels. One difference results in an extension of 5 additional amino acids at the 5' end of the GA US6/gD (M-N-R-Y-R) relative to its RB1B sequence counterpart (ORF5; Ross et al., 1991). In addition to the 5-AA extension, the next four positions in the GA strain (Y-E-S-I) would differ from the corresponding RB1B positions (M-K-V-F). Differences in these signal sequences could account for differences in gD processing between these two strains. The predicted amino acid sequence of the GA US2 (Fig. 2) is identical to that

published in a recent report (Cantello et al., 1991), except for the presence of an alanine in place of an arginine at position 143. This minor difference is due to the inversion of a guanine and a cytosine relative to each other in the two GA sequences.

The RB1B counterpart of SORF4 (ORF4 in their report) was recently proposed to be a probable homolog of HSV-1 gG (Ross and Binns, 1991). It is tempting to propose such a homology, given their similar locations relative to other Us region genes. We have further tested this proposed homology by similarly aligning these two sequences with GAP, following repeated shuffles of either of the two sequences while maintaining length and composition (using the /RANdomizations command line option for 100 randomizations). This analysis was performed twice (each time with one of the two sequences shuffled). In doing so, we failed to find a significant difference in homology score ratios between the actual versus randomized alignments (1.12 ± 0.07). In some cases, the homologies of the randomized alignments actually exceeded the proposed MDV ORF4/HSV-1 gG alignment. Therefore we do not consider the proposed homology to be statistically significant. In fact, when using the type of stringency in the above example, equally significant homologies are encountered with almost any given protein database search (data not shown). While we cannot absolutely rule out that the two sequences are evolutionarily related, any functional homology would appear absent, since the supposed MDV gG homolog lacks hydrophobic domains representing signal peptide and transmembrane domain regions. Thus, it would appear that, at the very least, selection pressure for the maintenance of a common glycoprotein function appears to have been lost in this case.

DISCUSSION

New findings

In this report, we have characterized the sequences of 3 new MDV Us region ORFs homologous to HSV US1 (ICP22), US8 (gE), and US10; 2 new MDV-specific Us ORFs (SORF1 and -3); a fowlpox virus homolog (SORF2);

and a complete HSV US7-homologous sequence. This extends upon the sequence analysis of a 5255-bp segment located in the U$_S$ region of the RB1B strain (Ross et al., 1991; Ross and Binns, 1991). Between two oncogenic serotype 1 strains, the "very virulent" RB1B- and the "virulent" GA strain, only minor sequence differences were found over their common 5.3-kbp region. With completion of the entire 11,160-bp U$_S$ sequence (GA strain), we have precisely determined the IR$_S$–U$_S$ and U$_S$–TR$_S$ junctions (Fig. 2); these were somewhat of a surprise, since previous workers (Fukuchi et al., 1985) using the same MDV strain as ours (GA), mapped the IR$_S$–TR$_S$ junction to a different fragment located 2–3 kb upstream of the correct location.

Alphaherpesvirus S regions are characterized by a set of homologs which are specific to members of this taxonomic subfamily (Davison and Taylor, 1987; McGeoch, 1990). The identification of seven alphaherpesvirus S region homologs in this study is consistent with MDV bearing a closer relation to alphaherpesviruses than gammaherpesviruses (Buckmaster et al., 1988). Failure to identify CpG dinucleotide deficiencies among MDV U$_S$ region sequences is further consistent with this proposal (Efstathiou et al., 1990; Honess et al., 1989).

Potential importance of alphaherpesvirus S component differences in determining biological divergence and pathogenesis

Since MDV has been traditionally regarded as a gammaherpesvirus, much of the previous work interpreting MDV's properties has proceeded by analogy with the association between EBV and B-cells (Wen et al., 1988, for example). Because of the closer genetic relationship between MDV and other alphaherpesviruses, we agree with others (Lawrence et al., 1990) that the lymphotropic properties of MDV and HVT are unlikely to be determined by molecules homologous to those of EBV.

Upon further examination, more parallels exist between the "lymphotropic" MDV and the "neurotropic" alphaherpesviruses than previously appreciated. Lymphotropism (and epitheliotropism) is probably common to all herpesviruses and is largely responsible for the widespread dissemination of HSV and VZV in cases involving neonatal and immunocompromised patients, often resulting in death (Nahmias and Roizman, 1973; Grose, 1982). These infections are characterized by a biphasic viremia similar to that observed in MDV-infected chickens; in the absence of maternal antibodies, young chickens can often die from an early mortality syndrome lacking any tumor involvement (Jakowski et al., 1970; Witter et al., 1980). A biphasic viremia has also been established for VZV infections involving immunocompetent patients as well (Grose, 1981; Ozaki et al., 1986). With respect to T-cell tropism, MDV and HSV are similar; replication of each is restricted to activated, Ia-bearing T-cells

(Braun et al., 1984; Calnek, 1986). Like MDV, equine herpesvirus-1, an alphaherpesvirus, can also establish latent infections in T-lymphocytes (Welch et al., 1992). This lends support to an earlier proposal characterizing EHV-1 as a T-lymphotropic herpesvirus (Scott et al., 1989). In addition to latent T-lymphocyte infections, MDV also appears to establish latent infections in both Schwann and satellite cells (Pepose et al., 1981) like VZV (Croen et al., 1988). Such complexities suggest that a biologically based classification system is overly simplistic, potentially misleading, and guided by biases that are dictated by the manner in which these viruses are studied.

To account for the different biological expressions that exist, a renewed focus on molecular differences between MDV and other alphaherpesviruses may be in order. In this regard, the MDV U$_S$ region (and adjoining repeats) may be particularly important. Fifty-three of the 55 unique long (U$_l$) region genes of HSV-1 possess an equivalent in VZV (McGeoch et al., 1988); a considerable number of these are related to beta- and gammaherpesvirus genes as well (29 of 67 EBV genes are counterparts to VZV U$_l$ genes; Davison and Taylor, 1987). In contrast, alphaherpesvirus S components are specific for members of this taxonomic subfamily and appear to represent their most divergent coding region (Davison and Wilkie, 1983; Davison and McGeoch, 1986; Telford et al., 1992). In comparing MDV with other alphaherpesviruses, significant divergence also extends to the U$_l$-flanking repeat regions (Buckmaster et al., 1988) which are known to be expressed in tumor cells (Jones et al., 1992; Schat et al., 1989; Sugaya et al., 1991). A comparison of the genetic organization of selected alphaherpesvirus S segment genes is presented in Fig. 4. It is based on previously published reports on EHV-1 Ab4p field isolate strain (Telford et al., 1992), HSV-1 (McGeoch et al., 1985), VZV (Davison and Scott, 1986), and PRV (Petrovskis et al., 1986a,b; Petrovskis and Post, 1987; van Zijl et al., 1990; Zhang and Leader, 1990); other alphaherpesvirus S segment genes corresponding to BHV-1 (Tikoo et al., 1990); EHV-1 Ky-A-cell culture strains (Audonnet et al., 1990; Breeden et al., 1992; Colle et al., 1992; Elton et al., 1991; Flowers et al., 1991; Flowers and O'Callaghan, 1992; Holden et al., 1992a,b), EHV-4 (Cullinane et al., 1988; Nagesha et al., 1993), HSV-2 (McGeoch et al., 1987), SHBV (Bennett et al., 1992; Killeen et al., 1992), and simian varicella virus (SVV; Fletcher and Gray, 1993) have been described elsewhere. Despite obvious similarities, there are marked differences in (i) gene content, organization, and localization; (ii) sequence conservation; and (iii) positioning of IR$_S$–U$_S$ and U$_S$–TR$_S$ junctions. Nevertheless, these overall gene layouts are consistent with a model to account for the divergence of alphaherpesviruses from a common ancestor by a number of homologous and semi-homologous recombination events which result in expansion or contraction of the inverted repeat regions and a concomitant loss or gain of U$_S$ gene(s) (Davison and
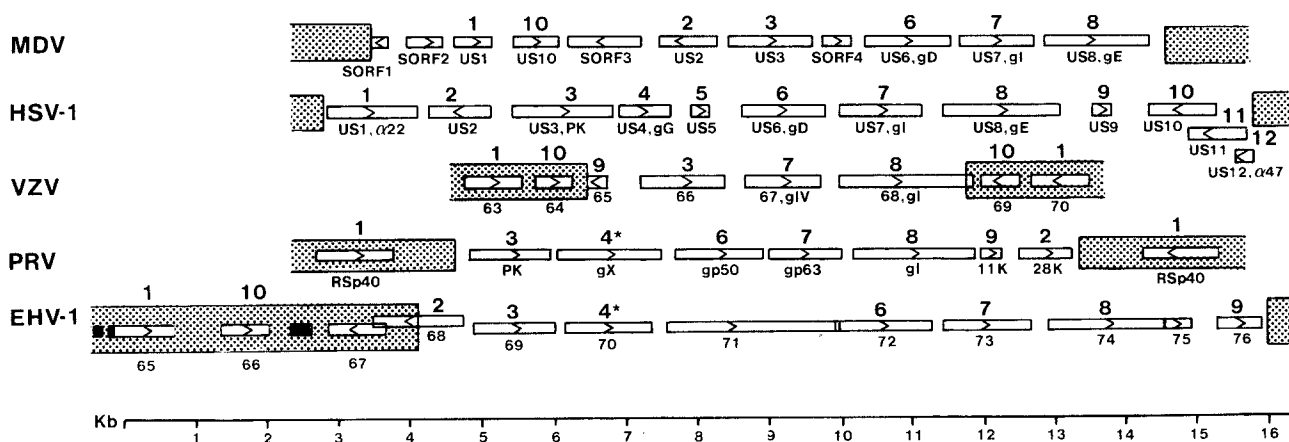
FIG. 4. Comparison of MDV and alphaherpesvirus S region genes. Based on published S region ORFs (Davison and Scott, 1986; McGeoch et al., 1985; Petrovskis et al., 1986a,b; Petrovskis and Post, 1987; Telford et al., 1992; van Zijl et al., 1990; Zhang and Leader, 1990). Numbers above boxes refer to homologs based on relation to HSV-1 U$_S$ ORF nomenclature (McGeoch et al., 1985). Polypeptide designations common to each system are listed below each of those boxes where applicable. Larger, stippled boxes refer to identified IR$_S$, TR$_S$, and/or R$_S$ regions. Bolded areas identify repeat sequences present in the EHV-1 Ab4p field isolate strain. Asterisks refer to homologs which show relatedness to HSV-1 US4, rather than HSV-1 US4.

McGeoch, 1986). In the case of VZV, homologs of six HSV-1 U$_S$ region genes are missing (US2, US4, US5, US6, US11, US12). Unlike all other alphaherpesviruses thus far analyzed (Fig. 4; Fletcher and Gray, 1993; Killeen et al., 1992), MDV appears to lack a US9 homolog. The HSV-1 US9 gene is known to encode a differentially phosphorylated 12- to 20-kDa tegument protein which becomes associated with nucleocapsids at or soon after their formation in the nuclei of infected cells (Frame et al., 1986). A recent study has suggested that PRV's US9 homolog has a function associated with envelopment at the nuclear membrane (Pol et al., 1991). Lacking such a homolog might contribute to MDV's characteristic inability to become stably enveloped in tissues other than the feather follicle epithelium.

Presence of MDV-specific and fowlpox virus-homologous genes

MDV contains at least 3 ORFs unrelated to any others presently described (SORFs 1, -3, and -4; Fig. 4). Other S component ORFs have been identified that are specific to a given alphaherpesvirus and/or its common-host relative. Such genes have been identified in HSV-1/HSV-2 (US11 and US12; Davison and McGeoch, 1986; McGeoch et al., 1985) and in EHV-1/EHV-4 (ORF67/IR6, ORF 71/EUS4, and ORF 75; Telford et al., 1992; Colle et al., 1992; Breeden et al., 1992; Nagesha et al., 1993). Further sequence analysis of other alphaherpesvirus S regions will be necessary to determine whether such genes are truly unique to these herpesviruses and whether they confer a species-specific growth advantage.

SORF3, located in the EcoRI-O subfragment (Fig. 1B), specifies a 351-AA MDV-specific ORF. Considering its location, preliminary transcriptional mapping of the other

genes mapping in EcoRI-O (e.g., MDV US1 and −10; P. Brunovskis, unpublished observations) and previously reported data (Schat et al., 1989), it appears possible that SORF3 may code for the 1.1-kb A$_j$ transcript, one of four immediate-early transcripts consistently identified in all MDV tumor cell lines tested (Schat et al., 1989).

A major surprise from this work was finding a FPV-related ORF. We are not aware of any other examples of such conservation across virus family lines, except a few cases that include cellular counterparts as well. MDV's FPV homolog, SORF2, was found to be 67% similar and 42% identical (over 100 AA) to FPV ORF4 (Tomley et al., 1988). With a FASTA score of 237 and the alignment in Fig. .3, the level of conservation is more striking than that generally characterizing alphaherpesvirus S region homologies (Fig. 4). Interestingly, compared with FPV ORF4, SORF2 contains an amino-terminal extension of 82 AA; conversely, ORF4 carries a carboxy-terminal extension of 41 AA. The block of conserved sequences may encode one or more functional domains that have independently evolved following host cell acquired gene transfer. On the other hand, it is intriguing to consider the possibility of virus–virus gene transfer. Individual cells have been found to be simultaneously infected by MDV and FPV (Tripathy et al., 1975). Given the different modes of replication for MDV and FPV (e.g., nuclear vs. cytoplasmic) such a possibility could point to a possibly novel form of gene transfer.

MDV U$_S$ region genes as potential determinants for pathogenesis and tissue tropism

Recent studies have shown that 11 of 12 open reading frames contained in the HSV-1 U$_S$ region are dispensable for growth in vitro (Longnecker et al., 1987; Roizman and

Sears, 1990). These, and other "dispensable" genes appear to specify functions for optimal survival, maintenance, and dissemination among the host (and its population at large), rather than the presence of functions necessary for replication (Longnecker et al., 1987; Roizman and Sears, 1990). The significant divergence of alphaherpesvirus S components may reflect this region's capacity for determining distinct tissue and host cell growth potentialities. Previous results have suggested that the product of HSV US1 (ICP22) encodes a determinant for tissue tropism, since its function appears to be dispensable for growth in some cell lines, but not others (Sears et al., 1985). Considering the extensive genetic divergence among a cluster of different glycoprotein homologs, each potentially subject to glycosylation, phosphorylation, palmitylation, myristylation, and/or sulfation (Grose, 1990; Spear, 1984), a potentially large window exists for the creation of multiply distinct virus–cell interactions which can affect host range, tissue tropism, invasiveness, and cell–cell spread. Previous results have demonstrated that "nonessential" alphaherpesvirus glycoproteins encode functions associated with virulence (Lomniczi et al., 1984; Meignier et al., 1988; Mettenleiter et al., 1988; Roizman and Sears, 1990). This may reflect their ability to promote the infection and spread of virus in vivo (Lomniczi et al., 1984; Longnecker et al., 1987; Mettenleiter et al., 1988; Pol et al., 1991; Card et al., 1992). Consistent with this proposal is the observation that a specific deletion of PRV gI (homolog of HSV gE) and/or PRV gp63 (homolog of HSV gI) was found to reduce the spread of infection in both rat (Card et al., 1992; Whealy et al., 1993) and pig (Kimman et al., 1992) central nervous systems. This defect could reflect the inability of PRV gI mutants to promote cell–cell spread (Zsak et al., 1992).

If MDV U$_S$ region genes specify virulence determinants, these could indirectly affect oncogenic potential by affecting any number of critical events which precede tumor induction. Previous studies have shown that oncogenic potential appears to be directly correlated with cell-associated viremia levels and the capacity to cause immunosuppression (Calnek and Witter, 1991). The sequence of events leading to transformation include (i) an initial lytic growth phase in B-cells, which is thought to cause activation and expansion of T-cells; (ii) a latent growth phase involving infected T-cells; (iii) a second wave of lytic infection, coincident with permanent immunosuppression; and (iv) oncogenic transformation (Calnek, 1986). Attenuated MDV strains (derived from oncogenic serotype 1 strains), as well as nononcogenic MDV and HVT stains (serotypes 2 and 3, respectively), are deficient in inducing the early cytolytic infection of B-cells in chickens, suggesting that their cell tropisms differ from those of oncogenic strains (Schat et al., 1985; Shek et al., 1982). This is reflected in evidence that attenuation of MDV leads to a marked reduction in infectivity and/or replication in lymphocytes (Schat et al., 1985).

In conclusion

The current herpesvirus classification system has been described as "simple, fortuitously appropriate and defective" (Roizman, 1990b). It has been further suggested that "the delineation and evolutionary relatedness of genes responsible for biological properties may be a more significant criterion for both evolutionary relatedness and classification than the arrangement and evolution of genes conserved throughout the family Herpesviridae" (Roizman, 1990b, 1992). While such a view is currently open to debate, inasmuch as alphaherpesvirus-specific U$_S$ regions specify a cluster of "dispensable" functions thought to be important biological property determinants, our sequence provides a foundation for further studies to resolve the apparent discrepancy between MDV's genetic and biologic properties.

Note added in proof. After submission of this paper, we learned of an article describing the nucleotide sequence of an 8.9-kb region of the MDV GA strain comprising 72% (8020 bp) of the MDV U$_S$ region, along with 905 bp of flanking IR$_S$ sequences (Sakaguchi et al., Virus Genes 6(4) 365–378). Their published IR$_S$–U$_S$ and U$_S$–TR$_S$ junction sites were identical to ours. Moreover, their sequence was found to be 99.9% identical to ours. However, the few differences that exist presumably account for their failure to identify the HSV US1 homolog of MDV and the FPV ORF4 homolog (SORF2) in MDV. The US1 error has recently been corrected in a published erratum (Virus Genes 7, 109). A missing base in their sequence would lead to the premature termination of a smaller ORF, with only half of the homology depicted in our Fig. 3. Another missing base in their sequence would result in the premature termination of the US2 ORF near amino acid 95. Our US2 ORF is consistent with that of Cantello et al., 1991. Also, recently Zelnick et al. (J. Gen. Virol. 74, 2151–2162) described the nucleotide sequence and gene organization of MDV's vaccine virus, HVT. HVT was found to contain a similar ORF organization, with the exception that it lacks MDV SORF1, SORF2, and SORF4 homologs.

REFERENCES

AUDONNET, J.-C., WINSLOW, J., ALLEN, G., and PAOLETTI, E. (1990). Equine herpesvirus type 1 unique short fragment encodes glycoproteins with homology to herpes simplex virus type 1 gD, gI and gE. J. Gen. Virol. 71, 2969–2978.

BENNETT, A. M., HARRINGTON, L., and KELLY, D. C. (1992). Nucleotide sequence analysis of genes encoding glycoproteins D and J in simian herpes B virus. J. Gen. Virol. 73, 2963–2967.

BINNS, M. M., and ROSS, N. L. J. (1989). Nucleotide sequence of the Marek's disease virus (MDV) RB-1B A antigen gene and the identifi-

cation of the MDV A antigen as the herpes simplex virus-1 glycoprotein C homologue. *Virus Res.* **12**, 371–382.

BRAUN, R. W., TEUTE, H. K., KIRCHNER, H., and MUNK, M. (1984). Replication of herpes simplex virus in human T lymphocytes: characterization of the viral target cell. *J. Immunol.* **132**, 914–919.

BREEDEN, C. A., YALAMANCHILI, R. R., COLLE, C. F., and O'CALLAGHAN, D. J. (1992). Identification and transcriptional mapping of genes encoded at the IR/Us junction of equine herpesvirus type 1. *Virology* **191**, 649–660.

BUCKMASTER, A. E., SCOTT, S. D., SANDERSON, M. J., BOURSNELL, M. E. G., ROSS, N. L. J., and BINNS, M. M. (1988). Gene sequence and mapping data from Marek's disease virus and herpesvirus of turkeys: Implications for herpesvirus classification. *J. Gen. Virol.* **69**, 2033–2042.

CALNEK, B. W. (1986). Marek's disease, a model for herpesvirus oncology. *Crit. Rev. Microbiol.* **12**, 293–320.

CALNEK, B. W., and WITTER, R. L. (1991). Marek's disease. *In* "Diseases of Poultry" (B. W. Calnek, Ed.), pp. 342–385. Iowa State Univ. Press, Ames.

CAMPADELLI-FIUME, G., QI, S., AVITABILE, E., FOA-TOMASI, L., BRANDIMARTI, R., and ROIZMAN, B. (1990). Glycoprotein D of herpes simplex virus encodes a domain which precludes penetration of cells expressing the glycoprotein by superinfecting herpes simplex virus. *J. Virol.* **64**, 6070–6079.

CANTELLO, J. L., ANDERSON, A. S., FRANCESCONI, A., and MORGAN, R. W. (1991). Isolation of a Marek's disease virus (MDV) recombinant containing the *lacZ* gene of *Escherichia coli* stably inserted within the MDV US2 gene. *J. Virol.* **65**, 1584–1588.

CARD, J. P., WHEALY, M. E., ROBBINS, A. K., and ENQUIST, L. W. (1992). Pseudorabies virus envelope glycoprotein gI influences both neurotropism and virulence during infection of the rat visual system. *J. Virol.* **66**, 3032–3041.

CEBRIAN, J., KASCHKA-DIERICH, C., BERTHELOT, N., and SHELDRICK, P. (1982). Inverted repeat nucleotide sequences in the genomes of Marek's disease virus and the herpesvirus of the turkey. *Proc. Natl. Acad. Sci. USA* **79**, 555–558.

CHEN, X., SONDERMEIJER, P. J. A., and VELICER, L. F. (1992). Identification of a unique Marek's disease virus gene which encodes a 38-kilodalton phosphoprotein and is expressed in both lytically infected cells and latently infected lymphoblastoid tumor cells. *J. Virol.* **66**, 85–94.

CHURCHILL, A. E., PAYNE, L. N., and CHUBB, R. C. (1969). Immunization against Marek's disease using a live attenuated virus. *Nature* **221**, 744–747.

COLLE, C., III, FLOWERS, C. C., and O'CALLAGHAN, D. J. (1992). Open reading frames encoding a protein kinase, homolog of glycoprotein gX of pseudorabies virus, and a novel glycoprotein map within the unique short segment of equine herpesvirus type 1. *Virology* **188**, 545–557.

CROEN, K. D., OSTROVE, J. M., DRAGOVIC, L. J., and STRAUS, S. E. (1988). Patterns of gene expression are different for varicella-zoster and herpes simplex viruses. *Proc. Natl. Acad. Sci. USA* **85**, 9773–9777.

CULLINANE, A. A., RIXON, F. J., and DAVISON, A. J. (1988). Characterization of the genome of equine herpesvirus 1 subtype 2. *J. Gen. Virol.* **69**, 1575–1590.

DAVISON, A. J., and WILKIE, N. M. (1983). Location and orientation of homologous sequences in the genomes of five herpesviruses. *J. Gen. Virol.* **64**, 1927–1942.

DAVISON, A. J., and McGEOCH, D. J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *J. Gen. Virol.* **67**, 597–611.

DAVISON, A. J., and SCOTT, J. E. (1986). The complete DNA sequence of varicella-zoster virus. *J. Gen. Virol.* **67**, 1759–1816.

DAVISON, A. J., and TAYLOR, P. (1987). Genetic relations between varicella-zoster virus and Epstein–Barr virus. *J. Gen. Virol.* **68**, 1067–1079.

DEVEREAUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.

DUBIN, G., SOCOLOF, E., FRANK, I., and FRIEDMAN, H. M. (1991). Herpes simplex virus type 1 Fc receptor protects infected cells from antibody-dependent cellular cytotoxicity. *J. Virol.* **65**, 7046–7050.

EFSTATHIOU, S., HO, Y. M., HALL, S., STYLES, C. J., SCOTT, S. D., and GOMPELS, U. A. (1990). Murine herpesvirus 68 is genetically related to the gammaherpesviruses Epstein–Barr virus and herpesvirus saimiri. *J. Gen. Virol.* **71**, 1365–1372.

ELTON, D. M., HALLIBURTON, I. W., KILLINGTON, R. A., MEREDITH, D. M., and BONASS, W. A. (1991). Sequence analysis of the 4.7-kb *Bam*HI–*Eco*RI fragment of the equine herpesvirus type-1 short unique region. *Gene* **101**, 203–208.

FLETCHER, T. M., III, and GRAY, W. L. (1993). DNA sequence and genetic organization of the unique short (U$_S$) region of the simian varicella virus genome. *J. Gen. Virol.* **193**, 762–773.

FLOWERS, C. C., EASTMAN, E. M., and O'CALLAGHAN, D. J. (1991). Sequence analysis of a glycoprotein D gene homolog within the unique short segment of the EHV-1 genome. *Virology* **180**, 175–184.

FLOWERS, C. C., and O'CALLAGHAN, D. J. (1992). The equine herpesvirus type 1 (EHV-1) homolog of herpes simplex virus type 1 US9 and the nature of a major deletion within the unique short segment of the EHV-1 KyA strain genome. *Virology* **190**, 307–315.

FRAME, M. C., McGEOCH, D. J., RIXON, F. J., ORR, A. C., and MARSDEN, H. S. (1986). The 10K virion phosphoprotein encoded by gene US9 from herpes simplex virus type 1. *Virology* **150**, 321–332.

FUKUCHI, K., SUDO, M., LEE, Y.-S., TANAKA, A., and NONOYAMA, M. (1985). Structure of Marek's disease virus DNA: Detailed restriction enzyme map. *J. Virol.* **51**, 102–109.

GIBBS, C. P., NAZERIAN, K., VELICER, L. F., and KUNG, H.-J. (1984). Extensive homology exists between Marek's disease herpesvirus and its vaccine virus, herpesvirus of turkeys. *Proc. Natl. Acad. Sci. USA* **81**, 3365–3369.

GROSE, C. (1981). Variation on a theme by Fenner: The pathogenesis of chickenpox. *Pediatrics* **68**, 735–737.

GROSE, C. (1982). Varicella-zoster infections: Chickenpox (varicella) and shingles (zoster). *In* "Human Herpesvirus Infections" (R. Glaser and T. Gotlieb-Stematsky, Eds.), pp. 85–150. Dekker, New York.

GROSE, C. (1990). Glycoproteins encoded by varicella-zoster virus: Biosynthesis, phosphorylation, and intracellular trafficking. *Annu. Rev. Microbiol.* **44**, 59–80.

HANKS, S. K., QUINN, A. M., and HUNTER, T. (1988). The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52.

HOLDEN, V. R., YALAMANCHILI, R. R., HARTY, R. N., and O'CALLAGHAN, D. J. (1992a). ICP22 homolog of equine herpesvirus 1: Expression from early and late promoters. *J. Virol.* **66**, 664–673.

HOLDEN, V. R., YALAMANCHILI, R. R., HARTY, R. N., and O'CALLAGHAN, D. J. (1992b). Identification and characterization of an equine herpesvirus 1 late gene encoding a potential zinc finger. *Virology* **188**, 704–713.

HONESS, R. W., GOMPELS, U. A., BARRELL, B. G., CRAXTON, M., CAMERON, K. R., STADEN, R., CHANG, Y.-N., and HAYWARD, G. S. (1989). Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *J. Gen. Virol.* **70**, 837–855.

IGARASHI, T., TAKAHASHI, M., DONOVAN, J., JESSIP, J., SMITH, M., HIRAI, K., TANAKA, A., and NONOYAMA, M. (1987). Restriction enzyme map of herpesvirus of turkey DNA and its collinear relationship with Marek's disease virus DNA. *Virology* **157**, 351–358.

JACKERS, P., DEFECHEREUX, P., BAUDOUX, L., LAMBERT, C., MASSAER, M., MERVILLE-LOUIS, M.-P., RENTIER, B., and PIETTE, J. (1992). Characterization of regulatory functions of the varicella-zoster virus gene 63-encoded protein. *J. Virol.* **66**, 3899–3903.

JAKOWSKI, R. M., FREDRICKSON, T. N., CHOMIAK, T. W., and LUGINBUL, R. E. (1970). Hematopoietic destruction in Marek's disease. *Avian Dis.* **14**, 374–385.

JOHNSON, D. C., FRAME, M. C., LIGAS, M. W., CROSS, A. M., and STOW, N. D. (1988). Herpes simplex virus immunoglobulin G Fc receptor activity

depends on a complex of two viral glycoproteins, gE and gI. *J. Virol.* 62, 1347–1354.

JOHNSON, R. M., and SPEAR, P. G. (1989). Herpes simplex virus glycoprotein D mediates interference with herpes simplex virus infection. *J. Virol.* 63, 819–827.

JONES, D., LEE, L., LIU, J.-L., KUNG, H.-J., and TILLOTSON, J. K. (1992). Marek's disease virus encodes a basic-leucine zipper gene resembling the *fos/jun* oncogenes that is highly expressed in lymphoblastoid tumors. *Proc. Natl. Acad. Sci. USA* 89, 4042–4046.

KILLEEN, A. M., HARRINGTON, L., WALL, L. V., and KELLY, D. C. (1992). Nucleotide sequence analysis of a homologue of herpes simplex virus type 1 gene US9 found in the genome of simian herpes B virus. *J. Gen. Virol.* 73, 195–199.

KIMMAN, T. G., DE WIND, N., OEI-LIE, N., POL, J. M. A., BERNS, A. J., and GIELKENS, A. L. J. (1992). Contribution of single genes within the unique short region of Aujeszky's disease virus (suid herpesvirus type 1) to virulence, pathogenesis and immunogenicity. *J. Gen. Virol.* 73, 243–251.

KOZAK, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* 108, 229–241.

LAWRENCE, G. L., CHEE, M., CRAXTON, M. A., GOMPELS, U. A., HONESS, R. W., and BARRELL, B. G. (1990). Human herpesvirus 6 is closely related to human cytomegalovirus. *J. Virol.* 64, 287–299.

LEADER, D. P., and PURVES, F. C. (1988). The herpesvirus protein kinase: A new departure in protein phosphorylation? *Trends Biochem. Sci.* 13, 244–246.

LOMNICZI, B., WATANABE, S., BEN-PORAT, T., and KAPLAN, A. S. (1984). Genetic basis of the neurovirulence of pseudorabies virus. *J. Virol.* 52, 198–205.

LONG, D., WILCOX, W. C., ABRAMS, W. R., COHEN, G. H., and EISENBERG, R. J. (1992). Disulfide bond structure of glycoprotein D of herpes simplex virus types 1 and 2. *J. Virol.* 66, 6668–6685.

LONGNECKER, R., CHATTERJEE, S., WHITLEY, R. J., and ROIZMAN, B. (1987). Identification of a herpes simplex virus 1 glycoprotein gene within a gene cluster dispensable for growth in cell culture. *Proc. Natl. Acad. Sci. USA* 84, 4303–4307.

MCGEOCH, D. J. (1990). Evolutionary relationships of virion glycoprotein genes in the S region of alphaherpesvirus genomes. *J. Gen. Virol.* 71, 2361–2367.

MCGEOCH, D. J. (1991). Correlation between HSV-1 DNA sequence and viral transcription maps. *In* "Herpesvirus Transcription and Its Regulation" (E. K. Wagner, Ed.), pp. 29–47. CRC Press, Boca Raton, FL.

MCGEOCH, D. J., DOLAN, A., DONALD, S., and RIXON, F. J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *J. Mol. Biol.* 181, 1–13.

MCGEOCH, D. J., MOSS, H. W. M., MCNAB, D., and FRAME, M. C. (1987). DNA sequence and genetic content of the *Hind*III L region in the short unique component of the herpes simplex virus type 2 genome: Identification of the gene encoding glycoprotein G, and evolutionary comparisons. *J. Gen. Virol.* 68, 19–38.

MCGEOCH, D. J., DALRYMPLE, M. A., DAVISON, A. J., DOLAN, A., FRAME, M. C., MCNAB, D., PERRY, L. J., SCOTT, J. E., and TAYLOR, P. (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.

MEIGNIER, B., LONGNECKER, R., MAVROMARA-NAZOS, P., SEARS, A. E., and ROIZMAN, B. (1988). Virulence of and establishment of latency by genetically engineered deletion mutants of herpes simplex virus 1. *Virology* 162, 251–254.

METTENLEITER, T. C., SCHREURS, C., ZUCKERMAN, F., BEN-PORAT, T., and KAPLAN, A. S. (1988). Role of glycoprotein gIII of pseudorabies virus in virulence. *J. Virol.* 62, 2712–2717.

NAGESHA, H. S., CRABB, B. S., and STUDDERT, M. J. (1993). Analysis of the nucleotide sequence of five genes at the left end of the unique short region of the equine herpesvirus 4 genome. *Arch. Virol.* 128, 143–154.

NAHMIAS, A. J., and ROIZMAN, B. (1973). Infection with herpes-simplex virus 1 and 2. *N. Engl. J. Med.* 289, 781–789.

OZAKI, T., ICHIKAWA, T., MATSUI, Y., KONDO, H., NAGAI, T., ASANO, Y., YAMANISHI, K., and TAKAHASHI, M. (1986). Lymphocyte-associated viremia in varicella. *J. Med. Virol.* 19, 249–253.

PEARSON, W. R., and LIPMAN, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.

PEETERS, B.; POL, J., GIELKENS, A., and MOORMANN. (1993). Envelope glycoprotein gp50 of pseudorabies virus is essential for virus entry but is not required for viral spread in mice. *J. Virol.* 67, 170–177.

PEPOSE, J. S., STEVENS, J. G., COOK, M. L., and LAMPERT, P. W. (1981). Marek's disease as a model for the Landry–Guillain–Barre Syndrome: Latent viral infection in nonneuronal cells is accompanied by specific immune responses to peripheral nerve and myelin. *Am. J. Pathol.* 103, 309–320.

PETROVSKIS, E. A., TIMMINS, J. G., ARMENTROUT, M. A., MARCHIOLI, C. C., YANCEY, R. J., JR., and POST, L. E. (1986a). DNA sequence of the gene for pseudorabies virus gp50, a glycoprotein without N-linked glycosylation. *J. Virol.* 59, 216–223.

PETROVSKIS, E. A., TIMMINS, J. G., and POST, L. E. (1986b). Use of lambda-gt11 to isolate genes for two pseudorabies virus glycoproteins with homology to herpes simplex virus and varicella-zoster virus glycoproteins. *J. Virol.* 60, 185–193.

PETROVSKIS, E. A., and POST, L. E. (1987). A small open reading frame in pseudorabies virus and implications for evolutionary relationships between herpesviruses. *Virology* 159, 193–195.

POL, J. M. A., WAGENAAR, F., and GIELKENS, A. (1991). Morphogenesis of three pseudorabies virus strains in porcine nasal mucosa. *Intervirology* 32, 327–337.

PURVES, F. C., SPECTOR, D., and ROIZMAN, B. (1991). The herpes simplex virus 1 protein kinase encoded by the U$_S$3 gene mediates posttranslational modification of the phosphoprotein encoded by the U$_L$34 gene. *J. Virol.* 65, 5757–5764.

PURVES, F. C., and ROIZMAN, B. (1992). The U$_L$13 gene of herpes simplex virus 1 encodes the functions for posttranslational processing associated with phosphorylation of the regulatory protein a22. *Proc. Natl. Acad. Sci. USA* 89, 7310–7314.

ROIZMAN, B. (1990a). Whither herpesviruses? *In* "Immunobiology and Prophylaxis of Human Herpesvirus Infections" (C. Lopez, Ed.), pp. 285–291. Plenum, New York.

ROIZMAN, B. (1990b). Herpesviridae: A brief introduction. *In* "Fields Virology" (B. N. Fields *et al.*, Eds.), 2nd ed., Vol. 2, pp. 1787–1793. Raven Press, New York.

ROIZMAN, B. (1992). The family Herpesviridae: An update. *Arch. Virol.* 123, 425–449.

ROIZMAN, B., CARMICHAEL, L. E., DEINHARDT, F., DE-THE, G., NAHMIAS, A. J., PLOWRIGHT, W., RAPP, F., SHELDRICK, P., TAKAHASHI, M., and WOLF, K. (1981). Herpesviridae: Definition, provisional nomenclature, and taxonomy. *Intervirology* 16, 201–217.

ROIZMAN, B., and SEARS, A. E. (1990). Herpes simplex viruses and their replication. *In* "Fields Virology" (B. N. Fields *et al.*, Eds.), Vol. 2, pp. 1795–1841. Raven Press, New York.

ROSS, L. J. N., SANDERSON, M., SCOTT, S. D., BINNS, M. M., DOEL, T., and MILNE, B. (1989). Nucleotide sequence and characterization of the Marek's disease virus homologue of glycoprotein B of herpes simplex virus. *J. Gen. Virol.* 70, 1789–1804.

ROSS, L. J. N., and BINNS, M. M. (1991). Properties and evolutionary relationships of the Marek's disease virus homologous of protein kinase, glycoprotein D and glycoprotein I of herpes simplex virus. *J. Gen. Virol.* 72, 939–947.

ROSS, L. J. N., BINNS, M. M., and PASTOREK, J. (1991). DNA sequence and organization of genes in a 5.5 kbp *Eco*RI fragment mapping in the unique short segment of Marek's disease virus (strain RB1B). *J. Gen. Virol.* 72, 949–954.

SANGER, F. S., NICKLIN, S., and COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

SCHAT, K. A., CALNEK, B. W., FABRICANT, J., and GRAHAM, D. L. (1985). Pathogenesis of infection with attenuated Marek's disease virus strains. *Avian Pathol.* **14**, 127–146.

SCHAT, K. A., BUCKMASTER, A., and ROSS, L. J. N. (1989). Partial transcriptional map of Marek's disease herpesvirus in lytically infected cells and lymphoblastoid cell lines. *Int. J. Cancer* **44**, 101–109.

SCOTT, S. D., ROSS, N. L. J., and BINNS, M. M. (1989). Nucleotide and predicted amino acid sequences of the Marek's disease virus and turkey herpesvirus thymidine kinase genes; Comparison with thymidine kinase genes of other herpesviruses. *J. Gen. Virol.* **70**, 3055–3065.

SEARS, A. E., HALLIBURTON, I. W., MEIGNIER, B., SILVER, S., and ROIZMAN, B. (1985). Herpes simplex virus 1 mutant deleted in the a22 gene: Growth and gene expression in permissive and restricted cells and establishment of latency in mice. *J. Virol.* **55**, 338–346.

SHEK, W. R., SCHAT, K. A., and CALNEK, B. W. (1982). Characterization of non-oncogenic Marek's disease virus-infected and turkey herpesvirus-infected lymphocytes. *J. Gen. Virol.* **63**, 333–341.

SPEAR, P. G. (1984). Glycoproteins specified by herpes simplex viruses. *In* "Fields Virology" (B. N. Fields *et al.,* Eds.), pp. 315–356. Plenum, New York.

SUGAYA, K., BRADLEY, G., NONOYAMA, M., and TANAKA, A. (1991). Latent transcripts of Marek's disease virus are clustered in the short and long repeat regions. *J. Virol.* **64**, 5773–5782.

TELFORD, E. A. R., WATSON, M. S., MCBRIDE, K., and DAVISON, A. J. (1992). The DNA sequence of equine herpesvirus-1. *Virology* **189**, 304–316.

TIKOO, S. K., FITZPATRICK, D. R., BABIUK, L. A., and ZAMB, T. J. (1990). Molecular cloning, sequencing, and expression of functional bovine herpesvirus 1 glycoprotein gIV in transfected bovine cells. *J. Virol.* **64**, 5132–5142.

TOMLEY, F., BINNS, M., CAMPBELL, J., and BOURSNELL, M. (1988). Sequence analysis of an 11.2 kilobase, near-terminal, *Bam*HI fragment of fowlpox virus. *J. Gen. Virol.* **69**, 1025–1040.

TRIPATHY, D. N., SELLS, D. M., and HANSON, L. E. (1975). Natural pox and herpes as a dual viral infection in chickens. *Avian Dis.* **19**, 75–81.

VAN ZIJL, M., VAN DER GULDEN, H., DE WIND, N., GIELKENS, A., and BERNS, A. (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.

VON HEIJNE, G. (1985). Signal sequences: The limits of variation. *J. Mol. Biol.* **184**, 99–105.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683–4690.

WELCH, H. M., BRIDGES, C. G., LYON, A. M., GRIFFITHS, L., and EDINGTON, N. (1992). Latent equid herpesviruses 1 and 4: Detection and distinction using the polymerase chain reaction and cocultivation from lymphoid cells. *J. Gen. Virol.* **73**, 261–268.

WEN, L.-T., TANAKA, A., and NONOYAMA, M. (1988). Identification of Marek's disease virus nuclear antigen in latently infected lymphoblastoid cells. *J. Virol.* **62**, 3764–3771.

WHEALY, M. E., CARD, J. P., ROBBINS, A. K., DUBIN, J. R., RZIHA, H.-J., and ENQUIST, L. W. (1993). Specific pseudorabies virus infection of the rat visual system requires both gI and gp63 glycoproteins. *J. Virol.* **67**, 3786–3797.

WILCOX, W. C., LONG, D., SODORA, D. L., EISENBERG, R. J., and COHEN, G. H. (1988). The contribution of cysteine residues to antigenicity and extent of processing of herpes simplex virus type 1 glycoprotein D. *J. Virol.* **62**, 1941–1947.

WITTER, R. L., SHARMA, J. M., and FADLY, A. M. (1980). Pathogenicity of variant Marek's disease virus isolants in vaccinated and unvaccinated chickens. *Avian Dis.* **24**, 210–232.

ZHANG, G., and LEADER, D. P. (1990). The structure of the pseudorabies virus genome at the end of the inverted repeat sequences proximal to the junction with the short unique region. *J. Gen. Virol.* **71**, 2433–2441.

ZSAK, L., ZUCKERMANN, F., SUGG, N., and BEN-PORAT, T. (1992). Glycoprotein gI of pseudorabies virus promotes cell fusion and virus spread via direct cell-to-cell transmission. *J. Virol.* **66**, 2316–2325.