Arch Virol (1992) 126: 159–169

Archives of Virology

The fusion protein gene of phocine distemper virus: nucleotide and deduced amino acid sequences and a comparison of morbillivirus fusion proteins

M. D. Curran, Y. J. Lü*, and B. K. Rima

Division of Genetic Engineering, School of Biology and Biochemistry, The Queen's University of Belfast, Belfast, Northern Ireland

Accepted January 27, 1992

Summary. The nucleotide sequence of the gene encoding the fusion protein of phocine distemper virus has been determined. The mRNA is 2206 nucleotides in length and contains one major open reading frame (ORF) of 1893 nucleotides encoding a potential protein of 631 amino acid residues. However, analogy with canine distemper virus (CDV) suggests that translation of the F protein starts at the sixth AUG codon in the mRNA sequence which is located at position 461, resulting in an F_0 protein of exactly the same size (537 aa) as that of CDV. The overall homology at nucleotide level between the CDV and PDV F genes is 66%. The homology between the two F proteins of these respective viruses is 83%.

Introduction

Phocine distemper virus (PDV) is the causative agent of an epizootic which killed more than 18000 seals in north western Europe in 1988 [17, 20, 21]. PDV belongs to the serologically cross-reactive morbillivirus genus of which the human virus, measles virus (MV) is the type species. Other members include rinderpest (RPV), peste-des-petits ruminants (PPRV), and canine distemper virus (CDV). PDV has been found to be most closely related to CDV by studies of cross-reactivity with monoclonal antibodies (Mab) [9, 19] and by cross protection studies [22].

In order to study the relationship between PDV and CDV further, we have started an analysis of the nucleotide sequences of a number of PDV genes. Comparisons of Mab binding has shown that all PDV proteins except the

* Permanent address: Department of Biochemistry, Chongquing University of Medical Sciences, Chonguing, The Peoples Republic of China.

attachment protein H are antigenically similar [9, 19] to those of CDV. Nevertheless, the mobility of the proteins in SDS-PAGE [25] and preliminary nucleotide sequence comparisons on the basis of small parts of four genes of PDV including the 5′ part of the F mRNA [10] have indicated that PDV is distinct from CDV. Recently, we have determined that the gene order, transcription map and intergenic sequences of CDV and PDV are identical [11]. The gene order is that of a typical morbillivirus [3] which places on the negative stranded genomic template the genes in the order 3′-N-P/V/C-M-F-H-L-5′ where the N gene encodes the nucleocapsid protein; the P/V/C gene encodes the phosphoprotein (P), the non-structural C protein in an overlapping reading frame and, via edited transcripts the V protein [7]; the M gene encodes the matrix protein; the F gene the fusion glycoprotein and the H gene the attachment protein called H in analogy to the protein of MV which carries the haemagglutination activity and finally, the L gene encodes the large replicase-protein. The H genes of PDV and CDV are very different, displaying only 70% homology at the nucleotide level, although in the amino acid sequence of the encoded proteins they are more (75%) conserved [11].

In this paper we compare the size and nucleotide sequence of the fusion gene of PDV with that of CDV and compare the fusion proteins and coding strategies of the F genes of PDV, CDV, MV and RPV. The data show that PDV and CDV form a distinct subgroup in the morbilliviruses.

Materials and Methods

Viruses and cells

The Ulster/88 strain of PDV was employed throughout this study. It was isolated from kidney tissue of a diseased seal [17]. The virus was propagated in Vero cells using Eagle's medium supplemented with 2% newborn calf serum. Cells infected with this strain did not fuse. At 3 days post infection, the cell layers were scraped into the medium, sonicated and centrifuged at 3000 × **g** to remove cell debris. Virus stocks were kept at − 70 °C.

Labelling of viral RNA

Vero cells were infected at an m.o.i. of 1–3 $TCID_{50}$ per cell and incubated for 3 days. The medium was then replaced by phosphate free medium containing 20 μg/ml actinomycin D. After 1 h of phosphate-starvation, 1 mCi per ml ^{32}P-orthophosphate (carrier-free; Amersham, U.K.) was added to the medium and allowed to label the newly synthesized species of RNA for 4 h. Then, the cell monolayers were rinsed 3 × with ice cold PBS and lysed using guanidinium isothiocyanate and RNA was extracted and purified according to the method of Chirgwin and coworkers [8]. RNA was analysed on a 1% MOPS-formaldehyde agarose gel and visualized by autoradiography of the dried gel.

cDNA cloning and sequence analysis

Viral nucleocapsids were isolated from PDV infected Vero cells by the method of Udem and Cook [27]. RNA was purified from the nucleocapsids by phenol extraction and converted to double stranded (ds) cDNA by the method of Gubler and Hoffman [14] using random hexanucleotides to prime reverse transcription. The ds cDNA was cloned

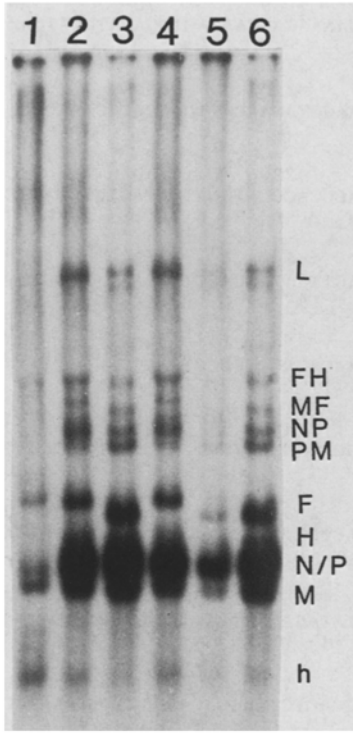

Fig. 1. Analysis of radiolabelled virus induced RNA species. Vero cells infected with MV (*2* and *4*), CDV (*3* and *6*), PDV (*5*) or mock infected (*1*) were labelled with ^{32}P-orthophosphate as described. The positions of transcripts of the N, P/C/V, M, F, M and L genes as well as readthrough dicistronic transcripts such as PM, NP, MF, and FH are indicated. *h* A prominent host band present and not masked, albeit reduced, in virus infected cells. The conditions of phosphate starvation appeared to mitigate against genome replication

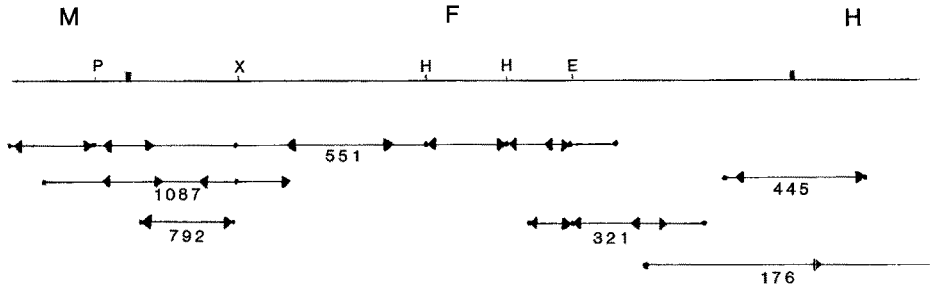

Fig. 2. A restriction enzyme map and sequencing strategy for the F gene of PDV. The cDNA clones used to determine the sequence are represented by fine lines and correspondingly numbered. Restriction sites used for subcloning and sequence analysis are also indicated: *P Pst* 1; *X Xba* 1; *H Hinc*II; *E Eco*RI. Arrows denote the direction and extent of sequencing and dots the starts of sequences

into Blue script (Stratagene) phagemid vectors either by direct insertion of flush ended cDNA into *Sma* 1 cut, dephosphorylated vector DNA or after extension with *Eco*RI adaptors and insertion into *Eco*RI cut, dephosphorylated vector DNA.

The inserts of selected clones were excised and either labelled by the oligopriming method of Feinberg and Vogelstein [13] for the hybridisation studies, or for sequencing by the dideoxy chain termination method they were subcloned into M13 tg130/131 phage. Sequences were analysed using the microgenie (Beckmann) suite of programmes.

```
        10         20         30         40         50         60         70         80
AGGGTCAAGGAACTGTCAAACCACCGGGCCTACCAAGACCAACAACCTAAGACAAAACAGCAATACTCGAGATCTATCAA

        90        100        110        120        130        140        150        160
AACTCAGAAATCTCCCAACTATCACCAGCCCCAAGAATCACACCTGAACCCCAAGGTACAACATTCAGCACCCACACACT

       170        180        190        200        210        220        230        240
GGTGTCATCAGGTTCCGAATGACTCGAGTCAAGAAACTCCCAGTGCCAACAAACCCACCCATGCATCATAGTCTCGATTC
                     M  T  R  V  K  K  L  P  V  P  T  N  P  P  M  H  H  S  L  D  S

       250        260        270        280        290        300        310        320
ACCATTTCTAAACCCTGAACATGCTACAGGGAAGATCAGTATTACCGATGACACTTCATCCCAGCTGACAAATTTCCTTT
  P  F  L  N  P  E  H  A  T  G  K  I  S  I  T  D  D  T  S  S  Q  L  T  N  F  L

       330        340        350        360        370        380        390        400
ATCATAAATACCATAAAACCACCATCAATCACCTATCTAGAACCGTTTCCGGAACAGACCCTCCCTCAGCCAAGCTAAAC
Y  H  K  Y  H  K  T  T  I  N  H  L  S  R  T  V  S  G  T  D  P  P  S  A  K  L  N

       410        420        430        440        450        460        470        480
AAGTTCGGATCACCTATTTTGAGTACATACCAGATCCGGTCAGCATTATGGTGGATAGCAATGGTCATCCTTGTTCACTG
 K  F  G  S  P  I  L  S  T  Y  Q  I  R  S  A  L  W  W  I  A  M  V  I  L  V  H  C

       490        500        510        520        530        540        550        560
TGTGATGGGTCAGATACACTGGACCAACTTATCAACAATTGGAATTATCGGGACTGATAGTTCCCATTATAAGATTATGA
  V  M  G  Q  I  H  W  T  N  L  S  T  I  G  I  I  G  T  D  S  S  H  Y  K  I  M

       570        580        590        600        610        620        630        640
CCAGGTCTAGTCATCAATATTTGGTCTTGAAATTAATGCCTAATGTATCAATTATAGACAATTGTACTAAAGCAGAATTG
T  R  S  S  H  Q  Y  L  V  L  K  L  M  P  N  V  S  I  I  D  N  C  T  K  A  E  L

       650        660        670        680        690        700        710        720
GATGAATATGAAAAGTTGCTAAATTCAGTATTAGAACCAATCAATCAGGCATTGACTCTCATGACAAAGAATGTAAAATC
 D  E  Y  E  K  L  L  N  S  V  L  E  P  I  N  Q  A  L  T  L  M  T  K  N  V  K  S

       730        740        750        760        770        780        790        800
ATTACAATCATTGGGATCAGGTAGAAGACAGAGGCGTTTTGCAGGGGTAGTAATTGCAGGTGCAGCTCTAGGTGTAGCAA
  L  Q  S  L  G  S  G  R  R  Q  R  R  F  A  G  V  V  I  A  G  A  A  L  G  V  A

       810        820        830        840        850        860        870        880
CTGCTGCCCAAATTACTGCAGGAGTTGCTTTATATCAATCTAATCTCAATGCTCAAGCAATTCAGTCTCTAAGAGCAAGC
T  A  A  Q  I  T  A  G  V  A  L  Y  Q  S  N  L  N  A  Q  A  I  Q  S  L  R  A  S

       890        900        910        920        930        940        950        960
CTTGAACAATCTAATAAAGCCATTGATGAAGTTAGACAGGCGTCGCAGAACATAATCATCGCTGTTCAAGGGGTTCAAGA
 L  E  Q  S  N  K  A  I  D  E  V  R  Q  A  S  Q  N  I  I  I  A  V  Q  G  V  Q  D

       970        980        990       1000       1010       1020       1030       1040
TTATGTTAACAATGAGATTGTCCCTGCTTTGCAGCATATGTCATGTGAGCTAATTGGACAGAGGCTAGGCCTAAAACTTC
  Y  V  N  N  E  I  V  P  A  L  Q  H  M  S  C  E  L  I  G  Q  R  L  G  L  K  L

      1050       1060       1070       1080       1090       1100       1110       1120
TTCGTTATTATACAGAACTTTTATCTGTATTTGGACCAAGTTTACGGGATCCTATTTCAGCAGAAATCTCAATACAAGCA
L  R  Y  Y  T  E  L  L  S  V  F  G  P  S  L  R  D  P  I  S  A  E  I  S  I  Q  A

      1130       1140       1150       1160       1170       1180       1190       1200
TTGAGCTATGCCTTGGGAGGGGAGATTCATAAGATACTTGAGAAACTAGGGTATTCAGGGAATGACATGGTTGCTATCCT
 L  S  Y  A  L  G  G  E  I  H  K  I  L  E  K  L  G  Y  S  G  N  D  M  V  A  I  L

      1210       1220       1230       1240       1250       1260       1270       1280
AGAAACCAAAGGTATCAGAGCCAAAATAACTCATGTTGACTTATCAGGGAAGTTTATAGTCCTAAGTATTTCTTACCCGA
  E  T  K  G  I  R  A  K  I  T  H  V  D  L  S  G  K  F  I  V  L  S  I  S  Y  P

      1290       1300       1310       1320       1330       1340       1350       1360
CTTTGTCTGAGGTTAAGGGGGTTGTCGTACATAGATTGGAGGCAGTTTCTTATAATATAGGGTCACAGGAATGGTACACT
T  L  S  E  V  K  G  V  V  V  H  R  L  E  A  V  S  Y  N  I  G  S  Q  E  W  Y  T

      1370       1380       1390       1400       1410       1420       1430       1440
ACTGTCCCTAGATATGTTGCAACGAATGGTTACTTGATCTCCAATTTTGACGAGTCGTCCTGTGTGTTCGTCTCCGAATC
 T  V  P  R  Y  V  A  T  N  G  Y  L  I  S  N  F  D  E  S  S  C  V  F  V  S  E  S

      1450       1460       1470       1480       1490       1500       1510       1520
GGCAATCTGTAGTCAGAATTCCTTGTATCCCATGAGCCCTATCTTGCAACAGTGTCTCAGAGGAGAAACTGCATCATGTG
  A  I  C  S  Q  N  S  L  Y  P  M  S  P  I  L  Q  Q  C  L  R  G  E  T  A  S  C
```

```
      1530       1540       1550       1560       1570       1580       1590       1600
CTAGAACTCTTGTCTCAGGGACATTAGGCAACAAGTTCATCTTATCAAAGGGAAACATTATTGCCAATTGTGCTTCTATC
A  R  T  L  V  S  G  T  L  G  N  K  F  I  L  S  K  G  N  I  I  A  N  C  A  S  I

      1610       1620       1630       1640       1650       1660       1670       1680
TTGTGTAAATGTCATAGCACCAGCAAGATCATCAACCAAAGTCCTGACAAATTGTTAACCTTCATTGCCTCCGATACTTG
 L  C  K  C  H  S  T  S  K  I  I  N  Q  S  P  D  K  L  L  T  F  I  A  S  D  T  C

      1690       1700       1710       1720       1730       1740       1750       1760
CTCATTAGTGGAGATAGACGGGGTTACTATTCAAGTTGGGAGTAGGCAGTATCCTGATGTTGTGTATGCGAGCAAAGTTA
  S  L  V  E  I  D  G  V  T  I  Q  V  G  S  R  Q  Y  P  D  V  V  Y  A  S  K  V

      1770       1780       1790       1800       1810       1820       1830       1840
TCCTCGGTCCGGCAATATCGCTAGAAAGATTAGATGTTGGGACAAACTTAGGCAGTGCTCTGAAGAAATTAGATGATGCT
I  L  G  P  A  I  S  L  E  R  L  D  V  G  T  N  L  G  S  A  L  K  K  L  D  D  A

      1850       1860       1870       1880       1890       1900       1910       1920
AAGGTATTGATAGAATCCTCTGATCAAATCCTTGACACTGTTAAAAATTCGTATTTAAGCTTAGGCACCCTCATTGCACT
 K  V  L  I  E  S  S  D  Q  I  L  D  T  V  K  N  S  Y  L  S  L  G  T  L  I  A  L

      1930       1940       1950       1960       1970       1980       1990       2000
CCCGGTATCTATCGGATTAGGCCTAATTTTATTACTGTTGATTTGTTGCTGTAAAAAGCGATATCAACATCTATTTAGTC
  P  V  S  I  G  L  G  L  I  L  L  L  L  I  C  C  C  K  K  R  Y  Q  H  L  F  S

      2010       2020       2030       2040       2050       2060       2070       2080
AATCTACTAAGGTTGCCCCTGTATTCAAGCCTGATTTGACTGGGACTTCAAAATCTTATGTCAGATCTCTATGAGATATC
Q  S  T  K  V  A  P  V  F  K  P  D  L  T  G  T  S  K  S  Y  V  R  S  L  *

      2090       2100       2110       2120       2130       2140       2150       2160
AGTGCCACTTGTTTTGACCAAACTCTTAATGCAAGCTGGCACATCAGCAATTCATTTTTGTCTATCACTGATCACTTGTT

      2170       2180       2190       2200
CCCAGAGACATAATAAATACGATATGGTCCCTACTATTAAAGAAAACTT
```

Fig. 3. Nucleotide sequence and deduced protein sequence of the ORF encoding the F protein of PDV: The positive (antigenomic) sequence is given. Translation is proposed to start at the underlined methionine (M) residue

Results

Size of F mRNA transcripts

A distinctive difference between CDV and the other morbilliviruses is the size of the transcript that encodes the F protein [1]. In MV and RPV these are 2372 and 2359 nucleotides in length respectively (excluding poly A tails) [23, 16, 26], whilst in CDV the 5′ untranslated region of the mRNA is approximately 120 nucleotides shorter [2]. In order to assess the size of the F mRNA and to confirm the general similarities between PDV and CDV on one hand and MV on the other, Vero cells infected with these three morbilliviruses were labelled with ^{32}P-orthophosphate in the presence of actinomycin D, when cpe (cell rounding) was near maximal. Figure 1 shows a denaturing gel analysis of the induced RNA species. Vero cells always produce a number of actinomycin D resistant RNA species but the virus induced bands stood out over the background profile. The PDV mRNAs including the F mRNA were found to be similar in size to those of CDV. Northern blot analysis has confirmed the designation of the various RNA species observed [11]. Thus PDV and CDV

from a distinct subgroup in the morbillivirus genus characterised by different sized F mRNAs.

Selection of cDNA clones of the F gene

cDNA clones were picked at random from the genomic library and those with large inserts were hybridised to Northern blots of PDV infected cell RNA to establish the gene order and physical and transcriptional map of PDV [11]. A putative F clone p 792 which hybridised to a band of the same size as that of the F mRNA in Fig. 1 was used to select further cDNA clones from the library until the physical map extended into the M and H genes. Restriction maps were prepared from the putative F gene clones and these were used to direct the sequence analysis. The position of the clones and sequencing strategy are shown in Fig. 2. The cDNA clones were sequenced in both directions and almost everywhere two or more independent copies of the F gene were available.

Sequence analysis

Our preliminary data on clone p 792 had shown very little homology between PDV and CDV [10]. This appeared to be a clone representing the 5′ end of the F mRNA. The sequences of all the cDNA clones were merged into a contiguous DNA sequence represented in Fig. 3 in the positive-antigenome-sense. The almost perfect conservation of transcription signals and particularly the intergenic trinucleotide CTT [11] allowed a prediction of the F gene transcript size to be made from the sequence of the cDNA clones. The 2206 nt (nucleotides) mRNA (excluding the poly A tail) contains one major open reading frame (ORF) starting at position 178. This encodes a protein of 631 amino acids and terminates with a TGA codon at position 2074.

Discussion

The size of the CDV F mRNA is smaller than those of MV [23] and RPV [16, 26] and this has been a distinguishing characteristic for CDV [1, 2]. From its RNA profile in a denaturing gel, PDV appears to have a similar sized F mRNA to CDV, borne out by sequence analysis which demonstrated it to be exactly the same in length as the F mRNA of CDV [2].

The untranslated and probable coding sequences of the F mRNA are similarly arranged in PDV and CDV but the start of the ORF which encodes the F protein varies between the various morbilliviruses (see Fig. 4). In MV, translation of the F_0 protein probably starts at the second AUG codon at position 578 and in RPV at positon 587 (the third AUG codon). In the case of PDV and CDV the potential exists for a 95 and 126 amino acid N-terminal extension to the F_0 protein, assuming that F_0 protein synthesis starts at the AUG codon at position 461 as suggested earlier [2]. Analysis of the N terminal extensions of CDV and PDV and the ORF preceding the RPV F_0-ORF which starts at position 587 (Fig. 4) revealed no significant homologies in these proteins nor

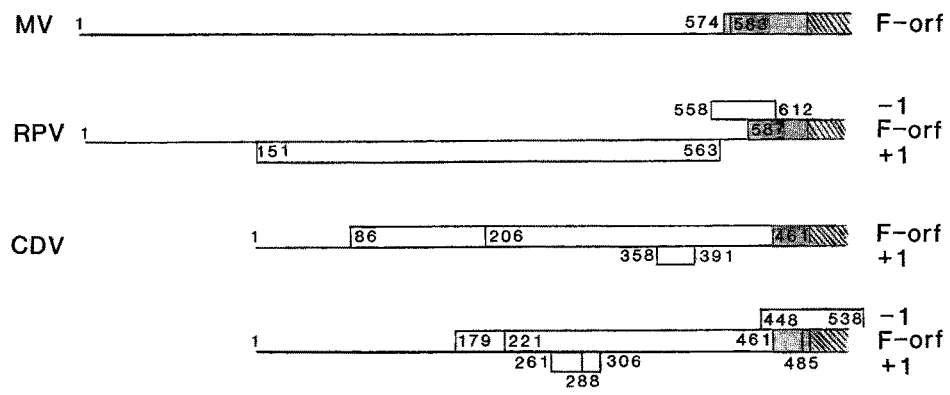

Fig. 4. Arrangement of open reading frames and AUG codons at the 5′ end of the morbillivirus F mRNAs. Data are from [2, 3, 16, 23, 26]. The various viruses are aligned on the codon encoding the Q residue likely to be at the N terminus of the F_2 proteins. Hatched areas indicate the F_2 protein after the Q residue. The shaded area before it represents the probable signal peptide. The starts and stops on the − 1 and + 1 reading frames are also indicated

were similarities in hydrophobicity plots observed (data not shown). In CDV the three AUG codons before the putative F_0-ORF start at position 461 can be removed without effect on the synthesis of the F_0 protein in translation in vitro. When the AUG at position 461 is removed, F_0 protein is no longer synthesised [12]. However, in vivo the removal of the preceding AUG codons reduces expression of the F_0 protein [3, 12]. Similar results have been observed for RPV [3, 12] and MV [15], so that it appears that the large 5′ untranslated region which contains several AUG codons aids expression of the protein in vivo but is not important in translation in vitro. In the case of PDV, synthesis of the F_0 protein would be predicted to be initiated at the sixth AUG codon in the mRNA, which is the third in the F-ORF (Fig. 4).

Figure 5 shows an alignment of the fusion proteins of morbilliviruses. The numbering starts with the methionine predicted to start the protein with a small signal sequence in front of the sequence QIHW, which is completely conserved among the morbilliviruses and QI throughout the paramyxoviruses [24]. The Q residue has been shown to be at the blocked N terminus of the F protein of Sendai virus [4].

The overall percentage homology at the nucleotide level between the PDV and CDV sequences is 66%, but this masks a rather uneven distribution. The homology in the first 460 nucleotides (before the predicted AUG start codon) is 43% ; the 3′ untranslated 132 nucleotides are 39% homologous and the coding region is 77% homologous between the two viruses. The region encoding the membrane anchor is the only major variable region within the ORF encoding the F_0 protein. Table 1 shows the overall homologies at the protein level between the various morbilliviruses. Clearly, PDV and CDV sequences are highly con-

```
PDV                                          [MTRVKKLPVPTNPPMHHSLDSPFLNPEHA
CDV [MHRGIPKSSKTQTHTQQDRPPQPSTELEETRTSRARHSTTSAQRSTHYDPRTSDRPVSYT
                                            -*--       -   *-   -       -
PDV  TGKISITDDTSSQLTNFLYHKYHKTTINHLSRTVSGTDPPSAKLNKFGSPILSTYQIRSA
CDV  MNRTRSRKQTSHRLKNIPVHGNHEATIQHIPESVSKGARSQIERRQPNAINSGSHCTWLV
       -      -**  * *   *  *  **-*-  -**       -              --     -
PDV  LWWIA]
CDV  LWCLG]
     ** -
PDV                MVILVHCVMGQIHWTNLSTIGIIGTDSSHYKIMTRSSHQYLVLKLMP  47
CDV                MASLFLCSKAQIHWDNLSTIGIIGTDNVHYKIMTRPSHQYLVIKLIP  47
RPV       MKILFATLLVVTTPHLVTGQIHWGNLSKIGVVGTGSASYKVMTGSSHQTLVIKLMP  56
MV   MGLKVNVSAIFMAVLLTLQTPTGQIHWGNLSKIGVVGIGSASYKVMTRSSHQSLVIKLMP  60
                     -         **** !** **--*      **-**  *** **-**-*
PDV  NVSIIDNCTKAELDEYEKLLNSVLEPINQALTLMTKNVKSLQSLGSGRRQRR           99
CDV  NASLIENCTKAELGEYEKLLNSVLEPINQALTLMTKNVKPLQSLGSGRRQRR           99
RPV  NITAIDNCTKTEIEEYKRLLGTVLQPIKVALNAITKNIKPIRSSTTSRRHRR          108
MV   NITLLNNCTRVEIAEYRRLLRTVLEPIRDALNAMTQNIRPVQSVASSRRHKR          112
     !-----!**- *- ** -** -**-**  ** --* *--*- *  - ** -*
PDV  FAGVVIAGAALGVATAAQITAGVALYQSNLNAQAIQSLRASLEQSNKAIDEVRQASQNII 159
CDV  FAGVVLAGVALGVATAAQITAGIALHQSNLNAQAIQSLRTSLEQSNKAIEAIREATQETV 159
RPV  FAGVALAGAALGVATAAQITAGIALHQSMMNTQAIESLKASLETTNQAIEEIRQAGQEMI 168
MV   FAGVVLAGAALGVATAAQITAGIALHQSMLNSQAIDNLRASLETTNQAIEAIRQAGQEMI 172
     ****-***-**************-**-** -* ***- *- *** -* **- -*-* *- -
PDV  IAVQGVQDYVNNEIVPALQHMSCELIGQRLGLKLLRYYTELLSVFGPSLRDPISAEISIQ 219
CDV  IAVQGVQDYVNNELVPAMQHMSCELVGQRLGLRLLRYYTELLSIFGPSLRDPISAEISIQ 219
RPV  LAVQGVQDYINNELVPAMGQLSCDIVGQKLGLKLLRYYTEILSLFGPSLRDPISAEISIQ 228
MV   LAVQGVQDYINNELIPSMNQLSCDLIGQKLGLKLLRYYTEILSLFGPSLRDPISAEISIQ 232
     -********-***--* -  -**---**-***-*******-**-****************
PDV  ALSYALGGEIHKILEKLGYSGNDMVAILETKGIRAKITHVDLSGKFIVLSISYPTLSEVK 279
CDV  ALIYALGGEIHKILEKLGYSGSDMIAILESRGIKTKITHVDLPGKFIILSISYPTLSEVK 279
RPV  ALSYALGGDINKILEKLGYSGSDLLAILESKGIKAKITYVDIESYFIVLSIAYPSLSEIK 288
MV   ALSYALGGDINKVLEKLGYSGGDLLGILESRGIKARITHVDTESYFIVLSIAYPTLSEIK 292
     ** *****-* *-******** *-- ***--**- -**-**    **-*** **-***-*
PDV  GVVVHRLEAVSYNIGSQEWYTTVPRYVATNGYLISNFDESSCVFVSESAICSQNSLYPMS 339
CDV  GVIVHRLEAVSYNIGSQEWYTTVPRYIATNGYLISNFDESSCVFVSESAICVYNSLYPMS 339
RPV  GVIIHRLEGVSYNIGSQEWYTTVPRYVATQGYLISNFDDTPCAFTPEGTICSQNALYPMS 348
MV   GVIVHRLEGVSYNIGSQEWYTTVPKYVATQGYLISNFDESSCTFMPEGTVCSQNALYPMS 352
     **--**** ***************-*-**-********-- * *  *  -*  * *****
PDV  PILQQCLRGETASCARTLVSGTLGNKFILSKGNIIANCASILCKCHSTSKIINQSPDKLL 399
CDV  PLLQQCIRGDTSSCARTLVSGTMGNKFILSKGNIVANCASILCKCYSTSTIINQSPDKLL 399
RPV  PLLQECFRGSTRSCARTLVSGSIGNRFILSKGNLIANCASILCKCYTTGSIISQDPDKIL 408
MV   PLLQECLRGSTKSCARTLVSGSFGNRFILSQGNLIANCASILCKCYTTGTIINQDPDKIL 412
     *-**-* ** * *********- **-**** **--**********--*  **!* ***-*
PDV  TFIASDTCSLVEIDGVTIQVGSRQYPDVVYASKVILGPAISLERLDVGTNLGSALKKLDD 459
CDV  TFIASDTCPLVEIDGATIQVGGRQYPDMVYEGKVALGPAISLDRLDVGTNLGNALKKLDD 459
RPV  TYIAADQCPIVEVDGVTIQVGSREYPDAVYLHKIDLGPPISLEKLDVGTNLGNAVTKLEK 468
MV   TYIAADHCPVVEVNGVAIQVGSRRYPDAVYLHRIDLGPPISLERLDVGTNLGNAIAKLED 472
     *-** * * -**--*- **** *-***-**  -- *** ***--******** *- **-
PDV  AKVLIESSDQILDTVKNSYLSLGTLIALPVSIGLGLILLLLICCCKKRYQHLFSQSTKVA 519
CDV  AKVLIDSSNQILETVRRSSFNFGSLLSVPILSCTALALLLLIYCCKRRYQQTLKQMTKVD 519
RPV  AKDLLDSSDLILETIKGASVTNTGHILVGAGLIAVVGILIVTCCCRKRSNDSKVSTVILN 528
MV   AKELLESSDQILRSMKGLSSTSIVYILIAVCLGGLIGIPALICCCRGRCNKKGEQVGMSR 532
     ** *--**- ** ---          - -       - - -   **- * -
PDV  PVFKPDLTGTSKSYVRSL                                           537
CDV  PAFKPDLTGTSKSYVRSL                                           537
RPV  PGLKPDLTGTSKSYVRSL                                           546
MV   PGLKPDLTGTSKSYVRSL                                           550
     *  ***************
```

Fig. 5. Alignment of the fusion protein of MV [23], CDV [2], RPV [26], and PDV (this work). The N terminal extensions of the F-ORF (see Fig. 4) are given as unnumbered residues. The residue are numbered from the methionine residue most likely to be used to initiate translation. Asterisks indicate conserved residues present in all morbilliviruses. Dashes indicate that only conservative replacements have occurred. / Potential N linked glycosylation sites

Table 1. Percentages of identical amino acid residues in the F proteins of morbilliviruses

	PDV	CDV	RPV	MV
PDV	100	83	56	57
CDV		100	64	66
RPV			100	77
MV				100

served in this, one of the most conserved genes in the paramyxo- and morbilliviruses [24].

PDV shares the three potential N-linked glycosylation sites of the other morbilliviruses in the F_2 part of the protein, and with CDV only, a potential glycosylation site in F_1, although no carbohydrate label has been found associated with the F_1 of the Onderstepoort strain of CDV [6]. The probable site for the cleavage and biological activation of the F_0 protein into F_2 and F_1 is conserved and the fusion related domain at the N terminus of F_1 is also highly conserved and so are the leucine residues that can form a potential zipper in the external domain proximal to the membrane anchor [5]. A number of conserved C residues at the cytoplasmic side of the membrane anchor may be involved in acylation of the F proteins but this has not yet been demonstrated for any of the morbilliviruses. The C terminal pentadekapeptide of the cytoplasmic domain is totally conserved within lytically growing morbilliviruses. The remainder is less conserved although all viruses contain paired basic residues in the region.

The morbilliviruses provide a unique group with which to study the effect of host range separation on comparative evolution of a set of serologically related paramyxoviruses. It is clear that the group itself consists of a subgroup of distemper viruses, PDV and CDV and the MV-RPV-PPRV group. The levels of homology in the F proteins of the various virus species are very high indeed. This reflects the high levels of conservation between all paramyxoviruses in the protein. The differences between PDV and CDV are so large as to be outside the scope of mutants and particularly the different arrangement in the 5′ untranslated region indicates that the speciation has taken place here. Thus, PDV appears to be a separate circulating morbillivirus which is primarily but not exclusively associated with pinnipeds [19]. Nucleotide sequence data of other sea mammal morbilliviruses [18] will throw further light on the evolutionary relationship between the viruses in the distemper subgroup.

Note added in proof. After completing the work presented in this paper, Kövamees et al. [17 a] reported the nucleotide sequence of the F gene of another isolate of phocine distemper. A comparison with our sequence reveals a remarkable level of conservation, with only five

nucleotide transitions at positions 280 (T→C), 365 (G→A), 373 (A→G), 526 (T→C) and 1031 (C→T). With the exception of the change at 365 which leads to a conservative replacement of a V→I at amino acid 63 in the protein sequence the changes are all silent. Since this amino acid change occurs upstream of the predicted translation initiation codon the F protein sequences of these two isolates of PDV are identical. It is noteworthy to highlight the absence of changes in the 3′ noncoding region and the level of conservation in the putative 95 amino acid N-terminal extension to the F_0 protein. Furthermore no changes occurred in the sequence upstream of this ORF.

Acknowledgements

This work was supported by NERC grant number GR 3/7159 A. We thank Ms. C. Lyons for expert advice and help in the growth of the virus and Ms. D. McKay for typing the manuscript. The nucleotide sequence data in this paper have been submitted to the DDBJ/EMBL/Genbank database and are accessible under number D10371.

References

1. Barrett T, Underwood B (1985) Comparison of messenger RNAs induced in cells infected with each member of the morbillivirus group. Virology 145: 195–199
2. Barrett T, Clarke DK, Evans SA, Rima BK (1987) The nucleotide sequence of the gene encoding the F protein of canine distemper virus: a comparison of the deduced amino acid sequence with other paramyxoviruses. Virus Res 8: 373–386
3. Barrett T, Subbarao SM, Belsham GJ, Mahy BWJ (1991) The molecular biology of the morbilliviruses. In: Kingsbury DW (ed) The paramyxoviruses. Plenum Press, New York, pp 83–102
4. Blumberg BM, Giorgi C, Rose K, Kolakofsky D (1985) Sequence determination of the Sendai virus fusion protein gene. J Gen Virol 66: 317–331
5. Buckland R, Wild TF (1989) Leucine zipper motif extends. Nature 338: 547
6. Campbell JJ, Cosby SL, Scott JK, Rima BK, Martin SJ, Appel M (1980) A comparison of measles and canine distemper virus polypeptides. J Gen Virol 48: 149–159
7. Cattaneo R, Kaelin K, Baczko K, Billeter MA (1989) Measles virus editing provides an additional cysteine rich protein. Cell 56: 759–764
8. Chirgwin JM, Przybyla AE, MacDonald R, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources rich in ribonuclease. Biochemistry 18: 5294–5299
9. Cosby SL, McQuaid S, Duffy N, Lyons C, Rima BK, Allan GM, McCullough SJ, Kennedy S, Smyth JA, McNeilly F, Craig C, Örvell C (1988) Characterisation of a seal morbillivirus. Nature 336: 115–116
10. Curran MD, O'Loan D, Rima BK, Kennedy S (1990) Nucleotide sequence analysis of phocine distemper virus reveals its distinctness from canine distemper virus. Vet Rec 127: 430–431
11. Curran MD, O'Loan D, Kennedy S, Rima BK (1992) Molecular characterization of phocine distemper virus: gene order and sequence of the gene encoding the attachment (H) protein. J Gen Virol 73: 1189–1194
12. Evans SA, Belsham GJ, Barrett T (1990) The role of the 5′ nontranslated regions of the fusion protein mRNAs of canine distemper virus and rinderpest virus. Virology 177: 317–323
13. Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction fragments to high specific activity. Anal Biochem 137: 266–269
14. Gubler V, Hoffmann BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269
15. Hasel KW, Day S, Millward S, Richardson CD, Bellini WJ, Greer PA (1987) Char-

acterization of cloned measles virus mRNAs by in vitro transcription, translation and immuno precipitation. Intervirology 28: 26–39
16. Hsu D, Yamanaka M, Miller J, Dale B, Grubman M, Yilma T (1988) Cloning of the fusion gene of rinderpest virus: comparative sequence analysis with other morbilliviruses. Virology 166: 149–153
17. Kennedy S, Smyth JA, McCullough SJ, Allan GM, McNeilly F, McQuaid S (1988) Confirmation of cause of recent seal deaths. Nature 335: 404
17a. Kövamees J, Blixenkrone-Möller M., Sharma B, Örvell C, Norrby E (1991) The nucleolide sequence and deduced amino acid composition of the haemagglutinin and fusion proteins of the morbillivirus phocid distemper virus. J Gen Virol 72: 2959–2966
18. McCullough SJ, McNeilly F, Allan GM, Kennedy S, Smyth JA, Cosby SL, McQuaid S, Rima BK (1991) Isolation and characterisation of a porpoise morbillivirus. Arch Virol 118: 247–252
19. Örvell C, Blixenkrone-Möller M, Svansson V, Have P (1990) Immunological relationships between phocid and canine distemper virus studied with monoclonal antibodies. J Gen Virol 71: 2085–2092
20. Osterhaus ADME, Vedder EJ (1988) Identification of virus causing recent seal deaths. Nature 335: 20
21. Osterhaus ADME, Groen J, de Vries P, UytdeHaag FGCM, Klingeborn B, Zarnke R (1988) Canine distemper virus in seals. Nature 335: 403–404
22. Osterhaus ADME, Uytdehaag FGCM, Visser IKG, Vedder EJ, Reynders PJM, Kuiper J, Brugge HN (1989) Seal vaccination success. Nature 337: 21
23. Richardson CD, Hull D, Greer P, Hasel K, Berkovich A, Englund G, Bellini WJ, Rima BK, Lazzarini RA (1986) The nucleotide sequence of the mRNA encoding the fusion protein of measles virus (Edmonston strain): a comparison of fusion proteins from several different paramyxoviruses. Virology 155: 508–523
24. Rima BK (1989) Comparison of amino acid sequences of the major structural proteins of the paramyxo- and morbilliviruses. In: Kolakofsky D, Mahy BWJ (eds) Genetics and pathogenicity of negative strand viruses. Elsevier, Amsterdam, pp 254–263
25. Rima BK, Cosby SL, Duffy N, Lyons C, O'Loan D, Kennedy S, McCullough SJ, Smyth JA, McNeilly F (1990) Humoral immune responses in seals infected with phocine distemper virus. Res Vet Sci 149: 114–116
26. Tsukiyama K, Yoshikawa Y, Yamanouchi K (1988) Fusion glycoprotein (F) of rinderpest virus: entire nucleotide sequence of the F mRNA, and several features of the F protein. Virology 164: 523–530
27. Udem SA, Cook KA (1984) Isolation and characterisation of measles virus intracellular nucleocapsid RNA. J Virol 49: 57–65

Authors' address: B. Rima, Division of Genetic Engineering, School of Biology and Biochemistry, The Queen's University of Belfast, Belfast BT 9 7 BL, Northern Ireland.

Received November 27, 1992