Journal of General Virology (1992), 73, 885–893. *Printed in Great Britain*

Sequence analysis of the genes encoding the nucleocapsid protein and phosphoprotein (P) of phocid distemper virus, and editing of the P gene transcript

Merete Blixenkrone-Möller,[1]* Bhaskar Sharma,[2]† Tamas M. Varsanyi,[2] Aizhung Hu,[2] Erling Norrby[2] and Jan Kövamees[2]

[1]*Institute of Veterinary Microbiology, Laboratory for Virology and Immunology, The Royal Veterinary and Agricultural University, Bülowsvej 13, Frederiksberg C, DK1870, Denmark and* [2]*Department of Virology, School of Medicine, Karolinska Institute, Stockholm S-105 21, Sweden*

The nucleotide and deduced amino acid sequences of two genes of phocid distemper virus (PDV) were determined by cDNA cloning and sequencing. The long open reading frame of the gene encoding the nucleocapsid (N) protein is presented. As with other morbilliviruses, the phosphoprotein (P) gene of PDV was found to be located after the 5′ end of the N gene and before the 3′ end of the matrix protein gene. The P gene was shown to have the capacity to encode three distinct proteins, P, V and C, in analogy to other morbilliviruses. The results presented provide evidence for editing of the PDV P mRNA transcript by insertion of G residues. When the nucleotide and deduced amino acid sequences of the N, P, V and C genes were aligned with corresponding sequences of other established members of the morbillivirus genus, compelling homology was found between PDV and canine distemper virus (CDV), whereas there was markedly less similarity between PDV and measles virus or rinderpest virus. On the basis of the alignments presented, the estimated amino acid sequence similarity between the N and P genes of PDV and CDV was 84% and 76%, respectively. These differences at the genomic level indicate that the viruses are two separate entities.

Introduction

Phocid distemper virus (PDV), a newly recognized morbillivirus, has recently been identified as the primary infectious pathogen causing devastating epizootics among harbour seals (*Phoca vitulina*) in northwestern Europe (Osterhaus *et al.*, 1988; Kennedy *et al.*, 1988*b*; Kennedy, 1990).

The morbillivirus genus within the Paramyxoviridae includes other highly contagious mammal pathogens, namely canine distemper virus (CDV), rinderpest virus (RPV), peste des petits ruminants virus and measles virus (MV). Morbillivirus candidates infecting porpoises and dolphins have also been described recently (Domingo *et al.*, 1990; Trudgett *et al.*, 1991; Kennedy *et al.*, 1988*a*). Their relationship to the established morbilliviruses remains to be clarified.

The morbillivirus particle is a single helical nucleocapsid containing a negative-sense ssRNA genome of 15 to 16 kb complexed with protein. The nucleocapsid (N) protein is the major component of the nucleocapsid core, although two other proteins, the phosphoprotein (P) and large (L) protein, are associated with it. Three virus proteins are associated with the envelope, namely two integral membrane glycoproteins, the fusion (F) protein and the attachment protein haemagglutinin (H), and the internal matrix (M) protein.

The morbillivirus genome contains six transcriptional units or genes encoding, from the 3′ to the 5′ end, the N, P/C/(V), M, F, H and L proteins, which are separated by intergenic regions (for review see Kingsbury, 1990; Cattaneo *et al.*, 1989*a*). The role of the N protein in transcription and replication is unclear; however, it protects and allows the polymerase complex to move along the genome and antigenome. The P and L proteins are known to play a crucial role in viral RNA synthesis (Hamaguchi *et al.*, 1983).

P mRNA editing has been shown to occur in cells infected with MV (Cattaneo *et al.*, 1989*a*) and other paramyxoviruses (Kondo *et al.*, 1990; Ohgimoto *et al.*, 1990; Paterson & Lamb, 1990; Pelet *et al.*, 1991; Southern *et al.*, 1990; Takeuchi *et al.*, 1990; Thomas *et al.*, 1988; Vidal *et al.*, 1990*a*, *b*). Furthermore, recent

† Permanent address: Division of Virology, Indian Veterinary Research Institute, Mukteswar-Kumaon, Nainital 263138, India.

studies on MV and Sendai virus have suggested that the RNA editing activity is virus-encoded (Horikami & Moyer, 1991; Vidal *et al.*, 1990*a*).

Comparative genomic studies have placed PDV in the morbillivirus genus (Mahy *et al.*, 1988; Curran *et al.*, 1990; Bostock *et al.*, 1990; Haas *et al.*, 1991) and, in studies employing monoclonal antibodies, PDV exhibits antigenic characteristics most closely related to CDV of the known morbilliviruses (Cosby *et al.*, 1988; Örvell *et al.*, 1990). Recent data on the genes encoding the F and H proteins of PDV substantiate the close relationship between the two distemper viruses within the morbillivirus genus (Kövamees *et al.*, 1991*b*).

To date, there is no indication of a recent epizootiological link between PDV and other morbilliviruses, and the search for clues concerning the origin and spread of PDV poses intriguing questions about its phylogenetic relationship to other morbilliviruses. We present here nucleotide sequences of the N and P genes of PDV, and compare the predicted protein sequences with corresponding data for CDV, RPV and MV. Furthermore, our results provide evidence for editing of PDV P mRNA transcripts by G residue insertion.

Methods

Viruses and cells. PDV was isolated from diseased harbour seals in Danish waters (Blixenkrone-Möller *et al.*, 1989, 1992). The 8th passage of the cell culture-adapted PDV isolate and the Convac vaccine strain of CDV (Örvell, 1980) were grown in monolayer cultures of Vero cells in Eagle's MEM supplemented with 1% foetal calf serum and antibiotics. The viruses were plaque-purified three times before isolation of RNA from infected cell cultures.

Isolation of mRNA. Vero cells grown in roller bottles were infected at a multiplicity of 0·01 p.f.u./cell. Actinomycin D (5 μg/ml) was added to the infected cell cultures 24 h before harvesting. The PDV- and CDV-infected cultures were harvested at 4 and 3 days post-infection respectively, and mRNA was extracted from the cells and purified as described by Varsanyi *et al.* (1991).

Isolation of genomic RNA. Vero cell cultures infected with PDV were harvested when c.p.e. was extensive. Genomic RNA isolation from the virus particles was performed as described previously (Kövamees *et al.*, 1990).

PDV and CDV cDNA libraries. The construction of PDV and CDV cDNA libraries from reverse-transcribed oligo(dT)-primed poly(A)$^+$ RNA extracted from infected Vero cells has been described previously (Kövamees *et al.*, 1991*a*, *b*; Gubler & Hoffman, 1983). An *Eco*RI/*Not*I adaptor was ligated to the double-stranded cDNA, and the constructs were ligated into pT7/T3α-19 (BRL). *Escherichia coli* strain DH5α was used for transformation.

Assignment of cDNA clones to the N and P protein genes of PDV. Using a synthetic oligonucleotide (Rozenblatt *et al.*, 1985, nucleotides 1024 to 1046), we isolated an N gene-specific clone with an insert of 1·4 kb from the CDV cDNA library (Woods, 1984). The specificity of the ^{32}P-labelled cDNA was confirmed by Northern blot analysis (Sambrook *et al.*, 1989) before it was cross-hybridized to the PDV-derived cDNA library in 30% formamide at 37 °C (low stringency conditions) (Tsukiyama *et al.*, 1987; Maniatis *et al.*, 1982). To identify P gene-specific clones in the PDV cDNA library we performed hybridization as described by Woods (1984) using a ^{32}P-labelled oligonucleotide probe based on sequences at the 3′ end of the P mRNA of an Irish PDV isolate (M. D. Curran, D. O'Loan, B. K. Rima & S. Kennedy, unpublished results). The specificity of the large hybridizing clones (N10, P19), later determined to contain the entire coding region and parts of the non-coding regions of the PDV N and P mRNAs, was confirmed by Northern blot analysis (Sambrook *et al.*, 1989).

Sequencing strategies. Nucleotide sequences of the N and P genes of PDV were determined from two N gene-specific and three P gene-specific plasmids, one of which also contained sequences of the M gene (B. Sharma, unpublished results). Sequencing was carried out by primer extension. Denatured plasmid DNA (Zhang *et al.*, 1988) was sequenced by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) using the T7 Sequencing Kit (Pharmacia). The 5′ non-coding termini of the N and P mRNAs, which were not included in the plasmids, were determined by mRNA sequencing as described by Kövamees *et al.* (1990) using 10 μg purified mRNA as template and avian myeloblastosis virus (AMB) reverse transcriptase. Walking 'upstream' from the P gene into the N gene was achieved by sequencing bicistronic N–P mRNA. Nucleotide sequences of the genomic 5′ ends of the two genes were confirmed by RNA sequencing using purified genomic RNA (5 μg) as template (Kövamees *et al.*, 1990). For sequence analysis the Microgenie (Beckman) and DNA Strider programs (Marck, 1988) were used.

Primer extension method. The editing of P mRNA in PDV was studied by using primer extension according to a principle introduced by Driscoll *et al.* (1989) and Pelet *et al.* (1991). Briefly, 9 μg poly(A)$^+$-selected RNA from virus-infected and mock-infected Vero cells (negative control) was annealed to a 5′ end-labelled oligonucleotide primer (35-mer) complementary to the P mRNA sequence downstream of the editing region (nucleotides 752 to 786; Fig. 3) and extended with 2 units/μl AMV reverse transcriptase in the presence of 0·5 mM dCTP, dGTP, dTTP and ddATP. The primer extension permits the synthesis of a product that would terminate at the first U residue (T in the DNA form) after the editing site. A synthetic RNA was transcribed from a cDNA copy of the V mRNA of PDV using T7 polymerase and the SP6 Transcription Kit (Amersham). As a positive control for the primer extension reactions, this synthetic RNA template was included in the reactions. The products were fractionated by electrophoresis on a 10% polyacrylamide–7 M-urea gel, and analysed by autoradiography and densitometric scanning.

Results and Discussion

The nucleotide sequence of the N gene is shown in the antigenome sense in Fig. 1 and compared to the sequence of the CDV N gene reported by Rozenblatt *et al.* (1985). The long open reading frame (ORF) of the PDV N mRNA starts with CCACAAUGG, a moderate consensus sequence for ribosomal binding (Kozak, 1986). The predicted amino acid sequence of the PDV N protein is aligned with those for CDV, RPV and MV determined previously (Kamata *et al.*, 1991; Rozenblatt *et al.*, 1985; Fig. 2). The N mRNA of PDV has 52 non-coding nucleotides at the 5′ end and 53 at the 3′ end, excluding the poly(A) tail. The size of these non-coding areas is the same as those determined for CDV, RPV and MV. From

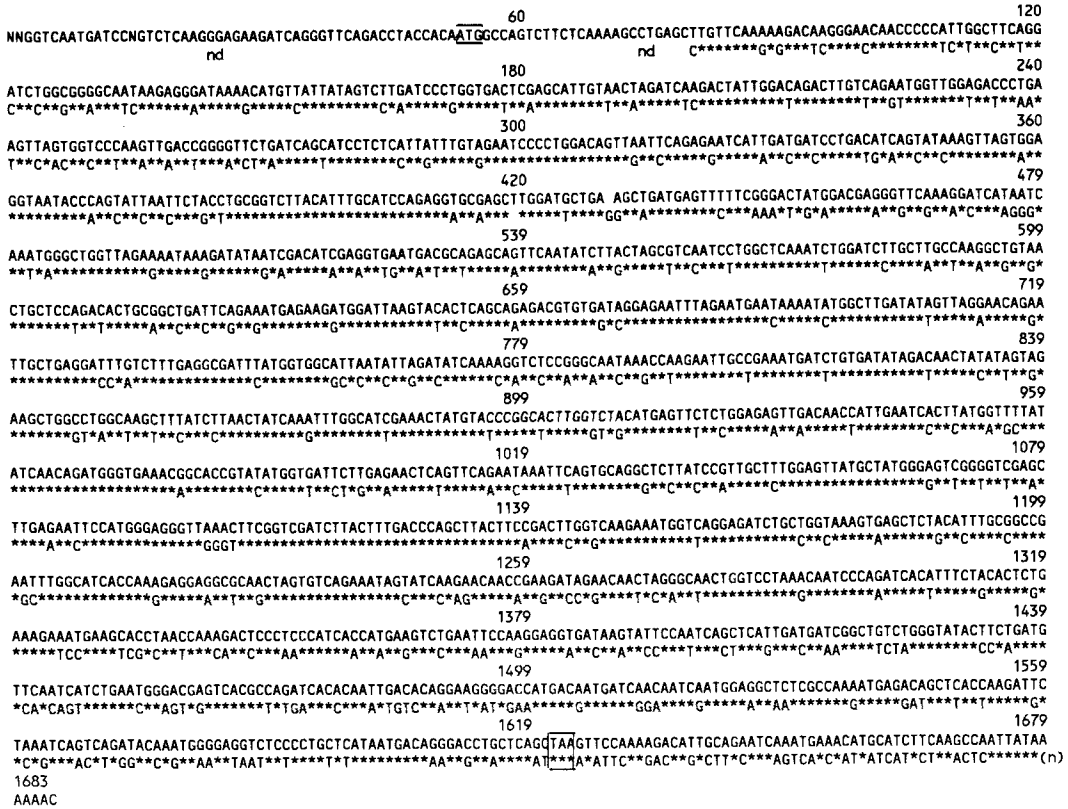

Fig. 1. Comparison of the nucleotide sequences of the PDV N gene (top line) and that of CDV (bottom line) (Rozenblatt *et al.*, 1985). The sequences are displayed in the antigenome sense. The positions of initiation and termination codons are boxed. Asterisks represent nucleotide identity; nd, sequence not determined.

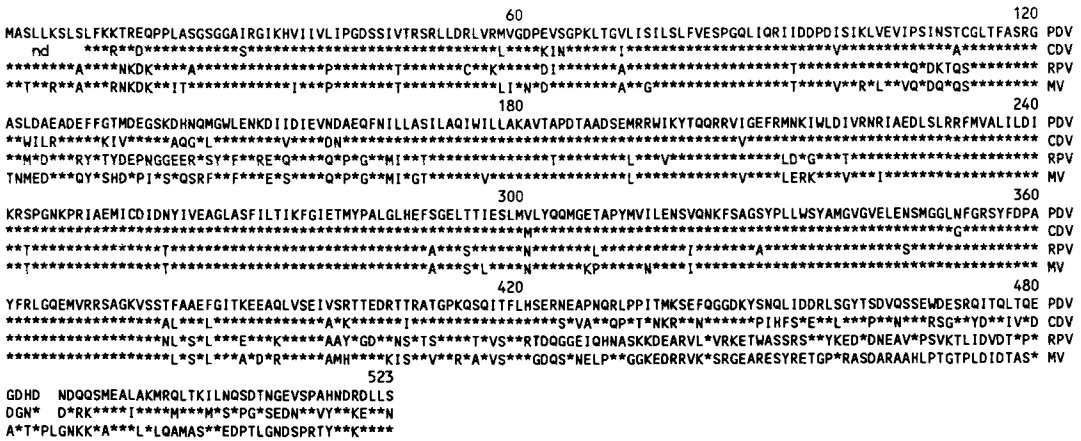

Fig. 2. Comparison of the amino acid sequences of the N proteins of PDV, CDV (Rozenblatt *et al.*, 1985), RPV (Kamata *et al.*, 1991) and MV (Rozenblatt *et al.*, 1985; Cattaneo *et al.*, 1989*b*). Asterisks represent amino acid identity.

the alignment in Fig. 1, the overall identity between the nucleotide sequences of the coding regions of the PDV and CDV N genes was estimated to be 77%, compared to 66% and 64% when aligned to the sequences of RPV and MV, respectively (data not shown).

The nucleotide sequence of the PDV P gene and its alignment from the first AUG with the CDV P gene (Barrett *et al.*, 1985) is shown in Fig. 3. As in other morbilliviruses, the P gene was determined to be located after the 5′ end of the N gene and before the 3′ end of the

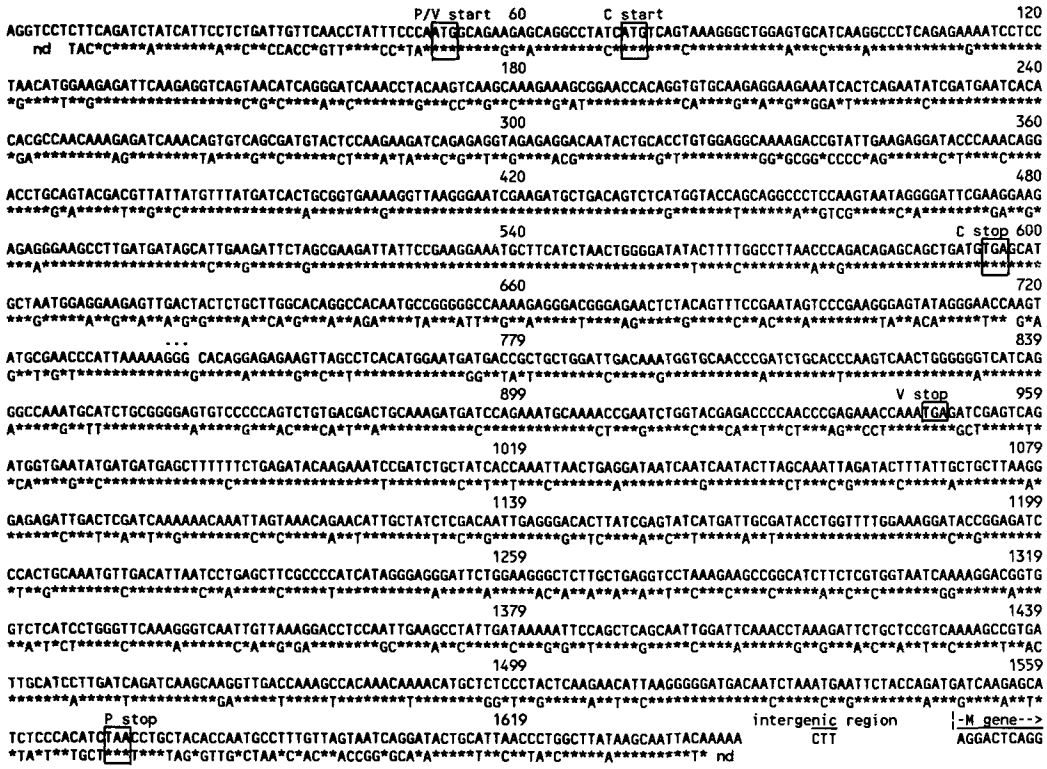

Fig. 3. Nucleotide sequence of the PDV P gene (top line) compared with that of the CDV P gene (bottom line) (Barrett *et al.*, 1985). The sequences are displayed in the antigenome sense. Initiation and stop codons of the P, V and C proteins are boxed. The editing region is marked with dots and is in accordance with Cattaneo *et al.* (1989*a*). Owing to strong secondary structure preceding the editing site, repeated attempts failed to determine the nucleotide at position 695 unambiguously. However, the presence of a C residue has been confirmed independently (B. Rima & T. Barrett, personal communication).

M gene. It was also shown to have coding capacity for three distinct proteins, P, V and C, by analogy to those described for MV (Cattaneo *et al.*, 1989*a*). The amino acid sequences of the predicted P, V and C proteins of PDV are aligned with those of CDV and MV in Fig. 4. Forty-nine nucleotides were identified in the non-coding 5′-terminal region of the PDV P mRNA; that of MV P mRNA has been reported to contain 59 nucleotides (Bellini *et al.*, 1985). The non-coding area at the 3′ end of the PDV P mRNA comprised 66 nucleotides, identical to the corresponding regions in CDV and MV (Barrett *et al.*, 1985; Bellini *et al.*, 1985). The nucleotide sequence identity between the P proteins of PDV and CDV was estimated to be 79%.

In this study we identified cDNA copies of the V and P mRNA, as well as copies which were in a third reading frame. The V mRNA was represented in plasmids which contained four G residues in an area in which mRNA editing is thought to occur (Fig. 3; nucleotides 739 to 741) (Cattaneo *et al.*, 1989*a*). We sequenced this region in six plasmids and the inserts were shown to vary in the numbers of G residues at this predicted editing site. Thus, plasmids with three, four, five and six G residues were isolated. Sequencing the genome in the editing region of the P gene revealed only three C residues, at nucleotides 739 to 741 (Fig. 3), which is similar to that for MV (Cattaneo *et al.*, 1989*a*, *b*; Bellini *et al.*, 1985). The PDV sequence for the area of the presumed frameshift, UUAAAAAGGGCACAG (nucleotides 732 to 746; Fig. 3), is identical to the conserved consensus sequence for the presumed frameshift in MV and RPV (Cattaneo *et al.*, 1989*a*; Barrett *et al.*, 1991).

The editing region of PDV P mRNA was analysed by primer extension (see Methods) and the length of the extension products should depend on the number of non-templated G residues, if any, inserted at the editing site. Autoradiographs of the sequencing gel clearly showed that the primer extension products varied in size (Fig. 5*a*, *b*). Densitometric scanning of the bands (Fig. 5*c*) indicated that more than half of the mRNA-derived products represented exact copies of the P gene in the editing region, whereas more than one-third of the products had one additional G residue in the editing region, and one-tenth had two or more G residues. The frequency and distribution of G nucleotide insertion found in PDV corresponded roughly to that found in MV using a similar method (Horikami & Moyer, 1991). Bands representing primer extension products with two

Fig. 4. (*a*) Comparison of the amino acid sequences of the P proteins of PDV, CDV (Barrett *et al.*, 1985) and MV (Bellini *et al.*, 1985; Cattaneo *et al.*, 1989*b*). The residues at the end of the N-coterminal parts of the P and V proteins are boxed. (*b*) Amino acid sequence of the C-terminal 68 amino acids of the V protein of PDV after the shift in reading frame caused by insertion of one extra G residue in the editing region (nucleotides 739 to 741; see Fig. 3) and alignment with corresponding sequences of the MV V protein (Bellini *et al.*, 1985; Cattaneo *et al.*, 1989*a*) and the proposed CDV V protein (Cattaneo *et al.*, 1989*a*), deduced from the nucleotide sequence of the P gene published by Barrett *et al.* (1985). (*c*) Comparison of the amino acid sequences of the C proteins of PDV, CDV (Barrett *et al.*, 1985) and MV (Bellini *et al.*, 1985). Asterisks represent amino acid identity.

G residue insertions appeared by densitometry to form 7% of the PDV P mRNA transcripts. The double G residue insertion would result in a shift into the third reading frame, with a stop codon only a few codons downstream of the editing site. As suggested for Sendai virus, a truncated form of P protein was predicted from these mRNAs (Vidal *et al.*, 1990*a*). By alignment of the coding region of the CDV P gene with that of PDV one nucleotide deletion was seen before the editing site and four residues in the editing region (Fig. 3). However, as previously suggested by Cattaneo *et al.* (1989*a*), the editing mechanism allowing the P gene to express at least two proteins with identical N-terminal segments probably applies to all morbilliviruses, including PDV and CDV.

In PDV the reading frame encoding the C protein, designated first in the Sendai virus system (Lamb *et al.*, 1976), starts with the second AUG at positions 72 to 74 and extends to the stop codon at positions 594 to 596 (Fig. 3). Start codons for both the P and C reading frames fitted only partially to the optimal consensus sequence for ribosomal binding (Kozak, 1986). Within the first 300 nucleotides of the long ORF encoding the P protein another AUG is present at positions 125 to 127. This codon better fits the consensus sequence; however from electrophoretic studies the size of the PDV P protein was estimated to be equal to that of other morbilliviruses (Rima *et al.*, 1990), indicating that the first AUG is utilized for transcription initiation. Two additional AUG codons were present within the first 300 nucleotides of the long ORF encoding the C protein. However, these start codons are probably not used, as they are in a no more favourable context for ribosomal binding than the first AUG (Lamb & Paterson, 1991).

The long PDV N gene ORF encodes an amino acid sequence of 523 residues with a calculated M_r of 58250, which agrees with estimates for the size of the CDV N protein from SDS–PAGE (Örvell, 1980). (The published nucleotide sequence of the CDV N gene is incomplete and, when compared to those of PDV, RPV and MV, the deduced amino acid sequence lacks nine N-terminal amino acid residues.)

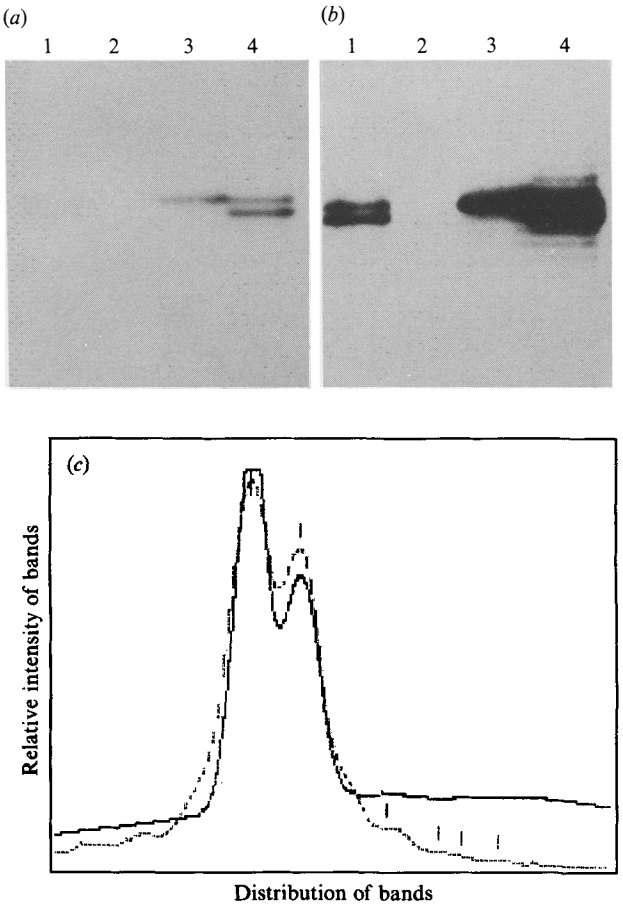

Fig. 5. (*a* and *b*) Autoradiographs of sequencing gels to illustrate the difference in size between PDV mRNA transcripts in the presumed editing region of the P mRNA. Lanes 1 and 4, PDV mRNA-derived products transcribed by primer extension starting downstream from the editing region and stopping at nucleotide position 733 (see Fig. 3). (Three times as much material was loaded in lane 4 as lane 1.) Lane 2, negative control using Vero cell mRNA as template; lane 3, positive control using a synthetic RNA template made from a cDNA copy representing V mRNA. Panel (*a*) is underexposed for clarity; panel (*b*) is overexposed to show differences in size of more than one nucleotide. (*c*) Densitometric scanning of the autoradiographs indicates that there is a bias towards one G insertion.

The N-terminal domain of the PDV N protein, representing three-quarters of the molecule (residues 1 to 397), exhibited high similarity to that of CDV (92%), whereas moderate similarity (60%) was estimated for the C-terminal part (residues 398 to 523). The two amino acid residues missing from the PDV and CDV N proteins seem to be missing from this least conserved part of the protein, as judged by alignment with the RPV and MV N proteins. The N proteins of PDV and CDV shared 84% overall sequence identity, whereas both displayed about 69% and 66% identity with the N proteins of RPV and MV, respectively. On the other hand, the N proteins of RPV and MV exhibited a considerable level of sequence identity (74%). Together, the levels of amino acid sequence identity between the morbillivirus N proteins seem to separate the members of the morbillivirus genus into two closely related subgroups, namely the distemper virus subgroup (PDV and CDV) and the RPV/MV subgroup.

The results presented further indicate an overall similarity in the structure of the N protein of the two distemper viruses, which is consistent with the relatively moderate epitopic variation of the N protein structural component of the two distemper viruses (Örvell *et al.*, 1990).

The central region of the N molecule (residues 169 to 211; Fig. 2) showed identical or conserved amino acids in all morbilliviruses, which might reflect a functional importance of the area involved. Furthermore, the central region of the N protein contains clusters of sequence which are homologous in members of the paramyxovirus and morbillivirus genera, confirming their phylogenetic relationship (Morgan, 1991). Different features of the paramyxovirus N protein have provided evidence for the existence of two domains within this structural unit, namely an N-terminal domain comprising about two-thirds of the molecule, which interacts directly with RNA, and a C-terminal domain, which lies at the surface of the assembled nucleocapsid (Buckland *et al.*, 1989; Rima, 1983).

The predicted amino acid sequence of the PDV P protein comprises 507 residues, which is the same size as that determined for CDV (Barrett *et al.*, 1985) and MV (Bellini *et al.*, 1985; Cattaneo *et al.*, 1989*b*). The calculated M_r of 54784 is in close agreement with those calculated for the P protein of CDV (54936) and MV (53900) (Barrett *et al.*, 1985; Bellini *et al.*, 1985). As with other paramyxoviruses, the presence of numerous potential post-translational phosphorylation sites is a characteristic feature of the PDV P protein, which increases the overall negative charge of the protein and apparently contributes to the difference in size estimates from the amino acid sequence and from migration in SDS–polyacrylamide gels (Rima, 1983). Glycine and proline residues are moderately conserved between PDV and CDV, 36 of 46 glycine residues and 19 of 26 proline residues having conserved positions. Together with the conserved position of two of four cysteine residues and similarities in hydrophilicity plots a distinct relationship between the P proteins of PDV and CDV is predicted.

The amino acid sequence of the PDV P protein exhibited two major regions of high similarity to that of CDV, located at amino acid residues 112 to 190 and 3306 to 507 (Fig. 4*a*), and smaller areas of similarity were scattered along the whole protein. The overall amino acid sequence identity between the P proteins of PDV

and CDV was estimated to be 76%; both PDV and CDV P proteins were approximately 45% similar to that of MV. Interspersed regions of high sequence identity were preferentially located towards the C-terminal half of the P protein, which is apparently required for binding to nucleocapsids (Ryan & Kingsbury, 1988; Huber *et al.*, 1991).

The predicted M_rs of the C (20323) and V (33562) proteins of PDV are in the range of those predicted for MV and CDV (Bellini *et al.*, 1985; Cattaneo *et al.*, 1989*a*, *b*; Barrett *et al.*, 1985). Sequence identity between the C proteins of the two distemper viruses and MV (Fig. 4*c*) was closely similar to those estimated for the P protein. In general, the P and C coding regions appear to be the least conserved of all paramyxovirus coding regions (Morgan, 1991). This is in contrast to the sequence of the 68 amino acid long C-terminal region of the V protein, which is highly conserved in PDV, CDV and MV (Fig. 4*b*). This part of the PDV V protein has a typical cysteine-rich region, in which all eight cysteine residues are conserved in CDV. The conservation of this unique part of the V protein between the morbilliviruses and paramyxoviruses might indicate that it has a specific function (Thomas *et al.*, 1988; Cattaneo *et al.*, 1989*a*). The cysteine-rich part resembles a metal-binding domain that could be involved in nucleic acid binding, protein–protein interactions or stabilizing oligomers (Thomas *et al.*, 1988). Recent studies have indicated that the V protein, unlike the P protein, does not bind effectively to MV nucleocapsid structures (Huber *et al.*, 1991). However, the precise functions of the P gene and its various encoded proteins are far from understood.

The nucleotide sequence of 245 residues near the 5′ end of the P gene mRNA of a German PDV isolate has been reported recently (Haas *et al.*, 1991). When compared to sequences in the present study, only three residues differed. It is interesting to note that the adaptation of the virus to growth in a continuous cell line apparently does not involve mutations in this area of the viral genome as identical sequences were obtained from cDNA copies of the cell culture-adapted German PDV isolate, and from reverse transcription and PCR amplification of RNA from tissues of naturally infected seals (Haas *et al.*, 1991). The deduced amino acid sequences of the N and P proteins of an Irish PDV isolate (M. D. Curran, personal communication) displayed only small differences to the sequences obtained from the Danish PDV isolate in our study. These amino acid sequence homologies between different isolates of PDV are consistent with the antigenic homogeneity found between different isolates of PDV (Blixenkrone-Möller *et al.*, 1992).

Besides the properties shared at the molecular and antigenic levels, the close relationship between PDV and CDV includes similar *in vivo* biological properties such as pathogenesis and clinical signs in susceptible hosts (Blixenkrone-Möller *et al.*, 1989; Visser *et al.*, 1990). Furthermore, the broad host range of CDV seems to overlap that of PDV (Blixenkrone-Möller *et al.*, 1990; Visser *et al.*, 1990).

In summary, our comparative genomic data indicate that (i) judging from the differences found between PDV and CDV these viruses are separate entities, (ii) PDV and CDV are closely related and exhibit similar relationship to RPV and MV respectively and (iii) members of the morbillivirus genus evolved into the distemper virus subgroup and the MV/RPV subgroup.

The origin of PDV remains unknown, but retrospective serological investigations have revealed that morbillivirus infections were present in aquatic mammals several years before the present epizootic (Dietz *et al.*, 1989; V. Svansson, M. Blixenkrone-Möller, K. Skirnisson, J. Geraci & P. Have, unpublished results; N. Markussen & P. Have, unpublished results).

Comparative genomic investigations of uncharacterized morbilliviruses from aquatic and terrestrial mammals might reveal further valuable information for defining the evolutionary relationships between morbilliviruses, and about molecular factors influencing their host range.

We thank Mrs V. S. Fredriksson, Mrs A. Foo, T. Danneman and E. Schach for excellent technical assistance, and P. Stålhandske, S. Löfdahl, T. Barrett. B. Rima and E. Lund for scientific discussions. This project was supported by The Swedish Medical Research Council, The Danish Agricultural and Veterinary Research Council and The Department of Biotechnology, Government of India.

References

BARRETT, T., SHRIMPTON, S. B. & RUSSELL, S. E. H. (1985). Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase-associated (P) protein mRNA. *Virus Research* **3**, 367–372.

BARRETT, T., SUBBARAO, S. M., BELSHAM, G. J. & MAHY, B. W. J. (1991). The molecular biology of the morbilliviruses. In *The Paramyxoviruses*, pp. 83–102. Edited by D. W. Kingsbury. London: Plenum Press.

BELLINI, W. J., ENGLUND, G., ROZENBLATT, S., ARNHEITER, H. & RICHARDSON, C. D. (1985). Measles virus P gene codes for two proteins. *Journal of Virology* **53**, 908–919.

BLIXENKRONE-MÖLLER, M., SVANSSON, V., HAVE, P. & BÖTNER, A. (1989). Infection studies in mink with seal-derived morbillivirus. *Archives of Virology* **106**, 165–170.

BLIXENKRONE-MÖLLER, M., SVANSSON, V., ÖRVELL, C. & HAVE, P. (1990). Phocid distemper virus – a threat to terrestrial mammals? *Veterinary Record* **127**, 263–264.

BLIXENKRONE-MÖLLER, M., SVANSSON, V., APPEL, M., KROGSRUD, J., HAVE, P. & ÖRVELL, C. (1992). Antigenic relationships between field isolates of morbilliviruses from different carnivores. *Archives of Virology* (in press).

BOSTOCK, C. J., BARRETT, T. & CROWTHER, J. R. (1990). Characterization of the European seal morbillivirus. *Veterinary Microbiology* **23**, 351–360.

BUCKLAND, R., GIRAUDON, P. & WILD, F. (1989). Expression of measles virus nucleoprotein in *Escherichia coli*: use of deletion mutants to locate the antigenic sites. *Journal of General Virology* **70**, 435–441.

CATTANEO, R., KAELIN, K., BACZKO, K. & BILLETER, M. A. (1989*a*). Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**, 759–764.

CATTANEO, R., SCHMID, A., SPIELHOFER, P., KAELIN, K., BACZKO, K., TER MEULEN, V., PARDOWITZ, J., FLANAGAN, S., RIMA, B., UDEM, S. & BILLETER, M. A. (1989*b*). Mutated and hypermutated genes of persistent measles viruses which caused lethal human brain diseases. *Virology* **173**, 415–425.

COSBY, S. L., MCQUAID, S., DUFFY, N., LYONS, C., RIMA, B. K., ALLAN, G. M., MCCULLOUGH, S. J., KENNEDY, S., SMYTH, J. A., MCNEILLY, F., CRAIG, C. & ÖRVELL, C. (1988). Characterisation of a seal morbillivirus. *Nature, London* **336**, 115–116.

CURRAN, M. D., O'LOAN, D., RIMA, B. K. & KENNEDY, S. (1990). Nucleotide sequence analysis of phocine distemper virus reveals its distinctness from canine distemper virus. *Veterinary Record* **127**, 430–431.

DIETZ, R., HANSEN, C. T., HAVE, P. & HEIDE-JØRGENSEN, M.-P. (1989). Clue to seal epizootic? *Nature, London* **338**, 627.

DOMINGO, M., FERRER, L., PUMAROLA, M., MARCO, A., PLANA, J., KENNEDY, S., MCALISKEY, M. & RIMA, B. K. (1990). Morbillivirus in dolphins. *Nature, London* **348**, 21.

DRISCOLL, D. M., WYNNE, J. K., WALLIS, S. C. & SCOTT, J. (1989). An *in vitro* system for the editing of apolipoprotein B mRNA. *Cell* **58**, 519–525.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HAAS, L., SUBBARAO, S. M., HARDER, T., LIESS, B. & BARRETT, T. (1991). Detection of phocid distemper virus RNA in seal tissues using slot hybridization and the polymerase chain reaction amplification assay: genetic evidence that the virus is distinct from canine distemper virus. *Journal of General Virology* **72**, 825–832.

HAMAGUCHI, M., YOSHIDA, T., NISHIKAWA, K., NARUSE, H. & NAGAI, Y. (1983). Transcriptive complex of Newcastle disease virus. I. Both the L and P proteins are required to constitute an active complex. *Virology* **128**, 105–117.

HORIKAMI, S. M. & MOYER, S. A. (1991). Synthesis of leader RNA and editing of the P mRNA during transcription by purified measles virus. *Journal of Virology* **65**, 5342–5347.

HUBER, M., CATTANEO, R., SPIELHOFER, P., ÖRVELL, C., NORRBY, E., MESSERLI, M., PERRIARD, J.-C. & BILLETER, M. A. (1991). Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. *Virology* **185**, 299–308.

KAMATA, H., TSUKIYAMA, K., SUGIYAMA, M., KAMATA, Y., YOSHIKAWA, Y. & YAMANOUCHI, K. (1991). Nucleotide sequence of cDNA to the rinderpest virus mRNA encoding the nucleocapsid protein. *Virus Genes* **5**, 5–16.

KENNEDY, S. (1990). A review of the 1988 European seal morbillivirus epizootic. *Veterinary Record* **8**, 563–567.

KENNEDY, S., SMYTH, J. A., CUSH, P. F., MCCULLOUGH, S. J., ALLAN, G. M. & MCQUAID, S. (1988*a*). Viral distemper now found in porpoises. *Nature, London* **336**, 21.

KENNEDY, S., SMYTH, J. A., MCCULLOUGH, S. J., ALLAN, G. M. & MCNEILLY, F. (1988*b*). Confirmation of recent seal deaths. *Nature, London* **335**, 404.

KINGSBURY, D. W. (1990). Paramyxoviridae and their replication. In *Virology*, 2nd edn., pp. 945–962. Edited by B. N. Fields & D. M. Knipe. New York: Raven Press.

KONDO, K., BANDO, M., TSURUDOME, M., KAWANO, M., NISHIO, M. & ITO, Y. (1990). Sequence analysis of the phosphoprotein (P) genes of human parainfluenza type 4A and 4B viruses and RNA editing at transcript of the P genes: the number of G residues added is imprecise. *Virology* **178**, 321–326.

KÖVAMEES, J., RYDBECK, R., ÖRVELL, C. & NORRBY, E. (1990). Hemagglutinin-neuramidase (HN) amino acid alterations in neutralization escape mutants of Kilham mumps virus. *Virus Research* **17**, 119–130.

KÖVAMEES, J., BLIXENKRONE-MÖLLER, M. & NORRBY, E. (1991*a*). The nucleotide and predicted amino acid sequence of the attachment protein of canine distemper virus. *Virus Research* **19**, 223–234.

KÖVAMEES, J., BLIXENKRONE-MÖLLER, M., SHARMA, B., ÖRVELL, C. & NORRBY, E. (1991*b*). The nucleotide sequence and deduced amino acid composition of the haemagglutinin and fusion proteins of the morbillivirus phocid distemper virus. *Journal of General Virology* **72**, 2959–2966.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. *Cell* **44**, 283–292.

LAMB, R. A. & PATERSON, R. G. (1991). The nonstructural proteins of paramyxoviruses. In *The Paramyxoviruses*, pp. 181–214. Edited by D. W. Kingsbury. London: Plenum Press.

LAMB, R. A., MAHY, B. W. J. & CHOPPIN, P. W. (1976). The synthesis of Sendai virus polypeptides in infected cells. *Virology* **69**, 116–131.

MAHY, B. W. J., BARRETT, T., EVANS, S. A., ANDERSON, E. C. & BOSTOCK, C. J. (1988). Characterization of a seal morbillivirus. *Nature, London* **336**, 115.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

MARCK, C. (1988). DNA strider: a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research* **16**, 1829–1836.

MORGAN, E. M. (1991). Evolutionary relationships of paramyxovirus nucleocapsid-associated proteins. In *The Paramyxoviruses*, pp. 163–179. Edited by D. W. Kingsbury. London: Plenum Press.

OHGIMOTO, S., BANDO, M., KAWANO, M., OKAMOTO, K., KONDO, K., TSURUDOME, M., NISHIO, M. & ITO, Y. (1990). Sequence analysis of P gene of human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two non-templated G residues. *Virology* **177**, 116–123.

ÖRVELL, C. (1980). Structural polypeptides of canine distemper virus. *Archives of Virology* **66**, 193–206.

ÖRVELL, C., BLIXENKRONE-MÖLLER, M., SVANSSON, V. & HAVE, P. (1990). Immunological relationships between phocid and canine distemper virus studied with monoclonal antibodies. *Journal of General Virology* **71**, 2085–2092.

OSTERHAUS, A. D. M. E., GROEN, J., DE VRIES, P., UYTDEHAAG, F. C. G. M., KLINGEBORN, B. & ZARNKE, R. (1988). Canine distemper virus in seals. *Nature, London* **335**, 403–404.

PATERSON, R. G. & LAMB, R. A. (1990). RNA editing by G-nucleotide insertion in mumps virus P-gene mRNA transcripts. *Journal of Virology* **64**, 4137–4145.

PELET, T., CURRAN, J. & KOLAKOFSKY, D. (1991). The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *EMBO Journal* **10**, 443–448.

RIMA, B. K. (1983). The proteins of morbilliviruses. *Journal of General Virology* **64**, 1205–1219.

RIMA, B. K., COSBY, S. L., DUFFY, N., SMYTH, J. A. & MCNEILLY, F. (1990). Humoral immune responses in seals infected by phocine distemper virus. *Research in Veterinary Science* **49**, 114–116.

ROZENBLATT, S., EIZENBERG, O., BEN-LEVY, R., LAVIE, V. & BELLINI, W. J. (1985). Sequence homology within the morbilliviruses. *Journal of Virology* **53**, 684–690.

RYAN, K. W. & KINGSBURY, D. W. (1988). Carboxy-terminal region of Sendai virus P protein is required for binding to viral nucleocapsids. *Virology* **167**, 106–112.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. New York: Cold Spring Harbor Laboratory.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SOUTHERN, J. A., PRECIOUS, B. & RANDALL, R. E. (1990). Two nontemplated nucleotide additions are required to generate the P mRNA of parainfluenza virus type 2 since the RNA genome encodes protein V. *Virology* **177**, 388–390.

TAKEUCHI, K., TANABAYASHI, M., HISHIYAMA, M., YAMADA, Y. K., YAMADA, A. & SUGIURA, A. (1990). Detection and characterization of mumps virus V protein. *Virology* **178**, 247–253.

THOMAS, S. M., LAMB, R. A. & PATERSON, R. G. (1988). Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891–902.

TRUDGETT, A., LYONS, C., WELSH, M. J., DUFFY, N., MCCULLOUGH, S. J. & MCNEILLY, F. (1991). Analysis of a seal and a porpoise morbillivirus using monoclonal antibodies. *Veterinary Record* **128**, 61.

TSUKIYAMA, K., SUGIYAMA, M., YOSHIKAWA, Y. & YAMANOUCHI, K. (1987). Molecular cloning and sequence analysis of the rinderpest virus mRNA encoding the hemagglutinin protein. *Virology* **160**, 48–54.

VARSANYI, T. M., KÖVAMEES, J. & NORRBY, E. (1991). Molecular cloning and sequence analysis of human parainfluenza type 2 virus mRNA encoding the fusion glycoprotein. *Journal of General Virology* **72**, 89–95.

VIDAL, S., CURRAN, J. & KOLAKOFSKY, D. (1990*a*). Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity. *Journal of Virology* **64**, 239–246.

VIDAL, S., CURRAN, J. & KOLAKOFSKY, D. (1990*b*). A stuttering model for paramyxovirus P mRNA editing. *EMBO Journal* **9**, 2017–2022.

VISSER, I. K., KUMAREV, V. P., ÖRVELL, C., DE VRIES, P., BROEDERS, H. V. J., VAN DE BILDT, M. W. G., GROEN, J., TEPPEMA, J. S., BURGER, M. C., UYTDEHAAG, F. G. C. M. & OSTERHAUS, A. D. M. E. (1990). Comparison of two morbilliviruses isolated from seals during outbreaks of distemper in North-western Europe and Siberia. *Archives of Virology* **111**, 149–164.

WOODS, D. (1984). Oligonucleotide screening of cDNA libraries. *Focus* **6**, 1–3.

ZHANG, H., SCHOLL, R., BROWSE, J. & SOMMERVILLE, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Research* **16**, 1220.

(*Received 19 July 1991; Accepted 16 December 1991*)