Cloning and sequence analysis of the phosphoprotein gene of rinderpest virus

M. D. Baron,* M. S. Shaila† and T. Barrett

AFRC Institute for Animal Health, Pirbright Laboratory, Ash Road, Woking, Surrey GU24 0NF, U.K.

We have cloned several cDNAs derived from the P gene of rinderpest virus. One of these, derived from a bicistronic N-P mRNA, has been sequenced in its entirety. Sequencing of a section of the others, and comparison with the genome sequence, showed that P gene transcripts, as for other morbilliviruses, were variable; non-templated Gs could be added at a site resembling the normal stop transcription site. Primer extension analysis showed that about half the transcripts were edited. Sequences of the P, C and V proteins encoded by the normal and edited transcripts were compared with those of other morbilliviruses and with those of the more distantly related paramyxoviruses.

Rinderpest is an economically highly important disease affecting cattle and wild bovids; it is endemic throughout large parts of Africa, Asia and the Middle East. The causative agent rinderpest virus (RPV) belongs to the Morbillivirus genus of the family Paramyxoviridae, and is thus related to measles virus (MV) and the canine and phocid distemper viruses (CDV and PDV). All the morbilliviruses are related serologically, and available sequence data show a high degree of similarity (for review see Barrett et al., 1989).

The Paramyxoviridae all possess an ssRNA genome of negative polarity; in the morbilliviruses this contains six genes, encoding the surface glycoproteins F and H, the nucleocapsid (N) protein and the envelope matrix (M) protein (for which the RPV sequence has been published), and the polymerase proteins L and P. The gene order is N–P–M–F–H–L (3′ to 5′ on the genome). As part of our investigations into virulence factors in RPV we are cloning all the genes of the vaccine (RBOK) strain. Here we report the sequence of the P gene of RPV, and compare this sequence with corresponding sequences of other morbilliviruses and members of the related paramyxovirus subgroup of the Paramyxoviridae.

Poly(A)⁺ RNA from RBOK-infected Vero cells was isolated and cDNA synthesized by standard methods. A cDNA library was constructed in λgt11 and the non-amplified library screened with a previously isolated partial clone of the P gene of the Kuwait strain of RPV (T. Barrett, unpublished results). The probe was labelled with biotin using the random-priming method (Feinberg & Vogelstein, 1983, 1984) and plaque lifts hybridized at high stringency (5× SSPE, 65 °C). The biotinylated probe was detected with streptavidin–alkaline phosphatase and the Photogene reagent (Life Technologies) essentially as previously described (Pandey et al., 1992).

Positive clones were plaque-purified and phage DNA isolated as described (Windle, 1988). Two large inserts (> 2.2 kb) were subcloned into pBluescriptKS(⁺) (Stratagene) and the ends sequenced. Both were derived from bicistronic mRNAs and began at the same point in the RBOK N gene, continuing through the intergenic sequence and the whole of the P gene. One of these (P14) was subcloned into M13tg130 and sequenced completely on both strands. The DNA sequence is shown in Fig. 1 in the anti-genome (message) sense, together with the derived amino acid sequence of the coding regions.

The intergenic sequence could be identified by the presence of the CTT motif that has been found to be highly conserved in all morbillivirus intergenic regions (Crowley et al., 1988; M. D. Curran et al., 1992). The conserved CTT marks the start of transcription, which begins at the next base; the first bases of the transcript (AGGA) are also conserved (Crowley et al., 1988).

The intergenic motif is preceded by the sequence TTATAAAAA, also found at the end of the P mRNA. This sequence is identical to the corresponding sequence in MV and is thought to be involved in normal mRNA transcription termination and polyadenylation. The sequence as a whole showed a high degree of similarity to that of MV (66% identity); the coding region of the P gene is closer to that of MV (69% identity) than to
those of either CDV or PDV (57% and 58% identity respectively). This confirms previous observations that the distemper viruses form one branch of the morbillivirus genus, and MV and RPV form another (Curran & Rima, 1992).

The sequence at the editing site (Fig. 1, boxed) contained one extra G relative to the MV sequence. It has previously been shown for simian virus 5 (SV5; Thomas et al., 1988), for MV (Cattaneo et al., 1989) and for many of the paramyxoviruses that the P gene is capable of producing several proteins. In MV, the P gene encodes the full-length P protein; cotranscriptional insertion of a non-templated G into a percentage of mRNAs leads to the production of an amino-coterminal form of the P protein. This confirms previous observations that the distemper viruses form one branch of the morbillivirus genus, and MV and RPV form another (Curran & Rima, 1992).

The sequence at the editing site (Fig. 1, boxed) contained one extra G relative to the MV sequence. It has previously been shown for simian virus 5 (SV5; Thomas et al., 1988), for MV (Cattaneo et al., 1989) and for many of the paramyxoviruses that the P gene is capable of producing several proteins. In MV, the P gene encodes the full-length P protein; cotranscriptional insertion of a non-templated G into a percentage of mRNAs leads to the production of an amino-coterminal form of the P protein. This confirms previous observations that the distemper viruses form one branch of the morbillivirus genus, and MV and RPV form another (Curran & Rima, 1992).
Fig. 3. Alignment of (a) P protein sequences and (b) C protein sequences of morbilliviruses. Conserved residues are boxed. The nucleotide and amino acid sequences were obtained from the EMBL database. RVF, this paper; MV (accession number K01711); CDV (accession number M32418); PDV (accession number X65512). Alignment and marking of conserved regions in this and other figures used the program ALIEN and the UWGCG package (Devereux et al., 1984).
et al., 1985) and shown recently to occur in PDV (Blixenkrone-Möller et al., 1992; Curran & Rima, 1992). It was therefore expected that the RPV P gene would also encode P, V and C proteins.

It was clear from the analysis of the reading frames of the P14 clone that this clone encoded proteins homologous to the MV C and V. In the absence of the fourth G at the editing site, the open reading frame (ORF) would have followed the sequence marked P in Fig. 1. We therefore isolated a further 10 full-length P clones and sequenced the editing site region of all these, together with that of the other bicistronic cDNA cloned in the first round. Of these 11 clones, three contained 3 G residues (encode P), seven contained four Gs (encode V) and one contained seven Gs (also encodes V). To show that this variation was not the result of variation in the genome, we isolated RPV genomic RNA from purified virus, cDNA was synthesized by priming upstream of the ORF, ligated into pT7Blue(U) (AMS Biotechnology) and 12 separate clones were sequenced. In all cases the sequence contained only three Gs at the editing site, which is identical for MV, RPV and PDV.

In order to examine the extent of mRNA editing in infected cells, we used the primer extension method (Driscol et al., 1989; Pelet et al., 1991). A primer (5’GGTCGATTTACGCTCTGTG3’) that was the complement of the sequence immediately downstream of the editing site (1362 to 1381 in Fig. 1) was 5’-labelled with 32P and hybridized to poly(A)+ RNA from RBOK-infected cells. For controls, the same primer was hybridized to uninfected Vero cell mRNA or to RNA transcribed in vitro from clones containing three or four Gs at the editing site. Hybridization, primer extension and analysis were as described (Driscol et al., 1989) except that a 12% acrylamide gel was used for analysing the reaction products. As seen in Fig. 2, mRNA from RPV-infected cells contains approximately equal amounts of transcripts containing one extra G or none; no significant band was found corresponding to the addition of two Gs, even when large amounts of RNA were analysed (Fig. 2, lane 5), indicating that the control of editing is very tight. In this respect the one clone found containing four additional Gs must be considered very unusual.

Studies on MV (Cattaneo et al., 1989) have also indicated that approximately equal numbers of P- and V-type mRNAs are present in infected cells, and again mRNAs with two extra Gs were not found. This is in contrast to PDV, where about 10% of transcripts had two Gs and the ratio of P to V transcripts was approximately 5:3 (Blixenkrone-Möller et al., 1992). This difference cannot be ascribed to the sequence at the editing site, which is identical for MV, RPV and PDV. The LPMV sequence was taken from Blixenkrone-Möller et al. (1992).

Alignment of the four morbillivirus P gene sequences currently available (Fig. 3a) shows that the carboxy-terminal half is far more conserved than the amino-terminal part. Studies on MV (Huber et al., 1991) have shown that it is the carboxy-terminal 40% that interacts with the N protein, and it is interesting to note that it is precisely this region, following the section that overlaps with the V-specific ORF, which is most conserved. The function of the amino-terminal part of the protein has yet to be established; it has been suggested, by analogy with the paramyxoviruses, that this section interacts directly with the L protein. However, comparison of the morbillivirus P proteins with the paramyxovirus P proteins reveals no homology. Indeed, the paramyxovirus P proteins themselves fall into at least two subgroups with P proteins that are shorter than those of the morbilliviruses [mumps virus, SV5, human parainfluenza virus (hPIV) types 2 and 4, Newcastle disease virus (NDV) and La Piedad-Michoacan-Mexico virus (LPMV)] and those which are considerably longer [Sendai virus, hPIV type 1 and 3, bovine PIV (bPIV) type 3], the P proteins of each group being non-homologous with the other. The C protein of MV has been found to colocalize
with the N protein in infected cells (Bellini et al., 1985), but there is little information about its function. The protein is usually classified as non-structural, because it has never been detected in purified virions. Alignment of the C protein sequences of the morbilliviruses shows only the highly conserved region previously identified by Rima and co-workers (Curran & Rima, 1992) (Fig. 3b). Only the ‘long P protein’ paramyxoviruses, and LPMV, also have C-encoding ORFs, and these C proteins also show no similarity to the morbillivirus C proteins. Recent work has suggested that the C protein of Sendai virus represses mRNA transcription, and therefore may be involved in the switch from mRNA transcription to genome replication (J. Curran et al., 1992). However, given the lack of any homology between Sendai and morbillivirus C proteins, it is not possible to assume that all the C proteins have a similar function. Exhaustive searching of the protein sequence databases (Collins & Coulson, 1987) failed to find any known protein with homology either to the C protein as a whole or to the long completely conserved domain.

However, a different picture emerges when the V-specific sequences of morbilliviruses and paramyxoviruses are considered. Not all paramyxoviruses appear to possess or express the cysteine-rich V domain; in hPIV type 3 the V ORF is conserved, but the editing event required to access it does not seem to occur (Galinski et al., 1992), whereas in the hPIV type 1 the V ORF is interrupted by stop codons (Matsuoka et al., 1991). The editing scheme in the ‘short P gene’ group is reversed, the genome encoding V, with two Gs being inserted to produce P mRNA (Berg et al., 1992; Ohgimoto et al., 1990; Paterson & Lamb, 1990; Pelet et al., 1991; Thomas et al., 1988). Despite this, alignment of the cysteine-rich sections (Fig. 4) shows that the positions of the cysteine residues, and a number of others, are conserved in all 12 proteins where V protein expression is proven or likely. The function of this domain is unclear. The spacing of the cysteine residues does not match any of the known metal-binding domains, although there is some similarity to metallothionine. The grouping of cysteines 3 to 7 is also homologous to the tat protein of human immunodeficiency virus (a ‘zinc finger’ domain), but the V proteins all lack the conserved histidine residue that forms part of such a sequence. MV V has been found distributed throughout the cytoplasm in infected cells, and does not colocalize with the P protein (Wardrop & Briedis, 1991). It has been found in neither MV (Gombart et al., 1992) nor Sendai virus (Curran et al., 1991b) virions, but has been found in mumps virus (Takeuchi et al., 1992). Work on Sendai virus has led to the suggestion that the V protein inhibits replication by binding to and inactivating the L protein (Curran et al., 1991a); if this is its general function, the V domain may be involved in targeting the L protein away from the nucleocapsid/replication centres.

References


(Received 7 July 1992: Accepted 29 September 1992)