Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions

Denis Rasschaert, Mariela Duarte and Hubert Laude*

Laboratoire de Virologie et d'Immunologie Moléculaires, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy-en-Josas, Domaine de Vilvert, 78350 Jouy-en-Josas, France

The genome organization of porcine respiratory coronavirus (PRCV), a newly recognized agent which has a close antigenic relationship to the enteropathogenic transmissible gastroenteritis virus (TGEV), was studied. Genomic RNA from cell-cultured PRCV (French isolate RM4) was used to produce cDNA clones covering the genomic 3' end to the start of the spike (S) glycoprotein gene (7519 nucleotides). Six open reading frames (ORFs) were identified that allowed the translation of three coronavirus structural proteins and three putative non-structural (NS) peptides, homologous to TGEV ORFs designated NS3-1, NS4 and NS7. Pairwise alignment of PRCV nucleotide and amino acid sequences with sequence data available for three TGEV strains revealed a 96% overall homology. However, the genome of PRCV exhibited two important distinctive features. The first was that the S gene lacked 672 nucleotides in the 5' region and encoded a truncated form of the S polypeptide, and secondly, the first NS ORF downstream of the S gene was predicted to be non-functional as a consequence of a double deletion. The significance of genomic deletions with respect to tissue tropism and evolution of coronaviruses is discussed.

Introduction

A new porcine respiratory coronavirus, tentatively designated PRCV, which has a close antigenic relationship to the enteropathogenic transmissible gastroenteritis virus (TGEV), suddenly emerged in 1983 to 1984 and spread within less than 2 years in most if not all European countries, where it now persists enzootically (Pensaert et al., 1986; Brown & Cartwright, 1986; Jestin et al., 1987; Have, 1990). A cytopathic agent has been isolated in cell culture and pathogenesis studies have shown that it replicates at a high titer in the respiratory tract, but to a very low extent in the gut (Pensaert et al., 1986). The significance of this virus as a pathogen is still unclear. Although it was first considered as non-pathogenic (Pensaert et al., 1986), subsequent investigations have linked PRCV with field outbreaks of respiratory disease (Jestin et al., 1987) and with pneumonia lesions following an experimental infection (Duret et al., 1988; O'Toole et al., 1989; Van Nieuwstadt & Pol, 1989). Conventional serological tests do not distinguish between PRCV- and TGEV-infected animals. However, antigenic dissimilarities were evident on examining the heterologous reactivity of monoclonal antibodies (MAbs). PRCV was found to be non-reactive towards several non-neutralizing anti-TGEV MAbs directed against the S (spike) or M (membrane) proteins (Callebaut et al., 1988; Garwes et al., 1988); neutralization-mediating epitopes unique to TGEV have been identified also (Laude et al., 1988). Conversely, anti-PRCV S-specific MAbs have been shown to differentiate between the two viruses (O. Deriabine, R. L'Hairdon, J. Gelfi & H. Laude, unpublished data). Examination of the structural polypeptide pattern of PRCV virions (TLM83 strain) shown by immunoblotting did not reveal any significant size discrepancy as compared to the TGEV Purdue strain (Callebaut et al., 1988).

The molecular characterization of this newly recognized agent has been undertaken in the laboratory with the aim of gaining information with respect to two important questions: what is the phylogenetic origin of PRCV?, and which gene(s) control(s) the expression of the enteric tropism of TGEV? In the present study part of the genetic information of PRCV was determined and compared to available TGEV sequence data. We showed that although the known part of their genomes are nearly identical, one major difference exists in both a structural and a non-structural gene.

Methods

Virus and cells. The French isolate RM4 of PRCV at four passages in the swine testicle (ST) cell line (Duret et al., 1988), supplied by Rhône-Mérieux Laboratoire IFFA, was used as a virus source. The virus was...
adapted to the pig kidney (PDS) cell line (three passages), and then plaque-purified. Infected cultures were maintained in Dulbecco's MEM plus 5% foetal calf serum and incubated for 40 h at 38 °C. Infectivity titres measured in ST cells averaged 10^6 p.f.u./ml.

**RNA extraction.** Clarified supernatant from the infected PDS cell culture medium was used as a starting material. The virions were semi-purified by ultracentrifugation through a 10 ml 25% glycerol cushion (Beckman rotor SW28, 60 min at 25000 r.p.m.). Pellets were treated with proteinase K plus 2% SDS and RNA was phenol-extracted as described (Rasschaert et al., 1987). RNA material was precipitated overnight by 2 M-LiCl at −20 °C then washed twice with 80% ethanol and resuspended in double-distilled water.

**cDNA cloning.** The strategy and protocols were essentially as reported (Rasschaert et al., 1987). Briefly, purified RNA was copied by reverse transcriptase (Mooney murine leukaemia virus, Bethesda Research Laboratories) using the TGEV-specific reverse primer pE2, located 47 nucleotides downstream of the S gene (5' CATCATCCT-TAACAAAATTCTCTAGCAGAA 3'). RNase T2-treated cDNA-RNA hybrids were dC-tailed and inserted into PstI-cut dG-tailed pBR322. Transfection of Escherichia coli RR1 cells and screening of clones by colony hybridization with TGEV cDNA clones pTG2.18 and pTG2.26 (Rasschaert et al., 1987), were performed using standard methods (Sambrook et al., 1989). Two additional clones (designated pTVR) were obtained by insertion of PCR-amplified (see below) cDNA fragments into the Smal site of the pTZ vector (Pharmacia) and transformation of E. coli NM522 cells.

**Polymerase chain reaction (PCR).** PCR-amplified fragments were obtained using cDNA-RNA heteroduplexes as templates and 20-mer oligonucleotides derived from the PRCV genome delimited using appropriate am- plifiers derived from partial sequencing. Restriction modifications. ST cell monolayers were infected or mock-infected by 2 M-LiCl at −20 °C then washed twice with 80% ethanol and resuspended in double-distilled water.

**DNA sequencing and analysis**

The insert of each of the clones specified in Fig. 1 was sequenced directly on plasmid supercoiled DNA and in both directions. Translation of the resulting 7519 nucleotide long sequence allowed the identification of six ORFs in the viral sense strand (Fig. 1; nucleotide and amino acid sequences displayed in Fig. 2). The largest three were predicted to encode the typical coronavirus structural peptides S, M and N. The other three were found to be homologous to TGEV ORFs encoding putative non-structural polypeptides, and were designated NS3-1, NS4 and NS7 as discussed later. The repeated sequence 5' CTAAAC 3' found upstream of each ORF, assumed to act as a signal for the transcription of subgenomic RNA in coronaviruses (review: Spaan et al., 1988), is identical to that of TGEV (Rasschaert et al., 1987). The extreme 3' non-coding clones to detect the S or the M and nucleocapsid (N) genes, respectively (Rasschaert et al., 1987). The pTVR12 and pTVR26 clones arose from non-specific priming (Fig. 1). This appears to be the case also for the clones covering the S gene, as the TGEV reverse primer used (derived from Purdue virus sequence) was found to match imperfectly with the PRCV sequence, in particular at the 3' end. The clones pTVR10 and pTVR25 were obtained subsequently through PCR amplification of portions of the PRCV genome delimited using appropriate amplimers derived from partial sequencing. Restriction analysis of the pVR and pTVR clones permitted the construction of a continuous map covering approximately 7.5 kb of the PRCV genomic 3' region.

**Results**

**Generation and mapping of the PRCV cDNA clones**

PRCV genomic RNA extracted from semi-purified virions was reverse-transcribed after priming with a TGEV-specific sequence downstream of the S gene. Heteroduplex RNA-DNA material was then introduced into the pBR322 vector and cloned into E. coli. The resulting clones were probed against two TGEVcDNA

![Fig. 1. Restriction and gene maps of PRCV genome 3' region. The positions of the cDNA clones used for sequencing are shown. pTVG.26 and pTVG.218 (dotted lines) are TGEV cDNA clones that were used for screening. Enzyme sites: BglII (Bg), HindIII (Hi), HpaI (Hp), PstI (Ps), PvuII (Pvu), XhoI (Xb).](image-url)
region comprised an 8 nucleotide sequence conserved in coronaviruses, and ended in a poly(A) stretch.

Pairwise alignment of the PRCV sequence data with those available for the three different TGEV strains Purdue 115 (Laude et al., 1987; Rasschaert & Laude, 1987; Rasschaert et al., 1987), FS772/70 (Britton et al., 1988a, b; 1989) and Miller (Wesley et al., 1989; Wesley, 1990) revealed a very high level of homology. Thus, 96% overall homology was found between PRCV and TGEV Purdue strains at both the nucleotide and amino acid level. However, several features unique to the PRCV genome were apparent. The most striking are three deletions affecting ORFs: (i) a 672 base in-frame deletion (position marked 1 in Fig. 2), located 59 bases downstream of the initiation codon of the S gene and (ii) two deletions which result in the alteration of the ORF situated immediately downstream of the S gene in the TGEV genome, a deletion of 38 bases (at position 2 in Fig. 2) removing both the CTAAAC sequence and the ATG initiation codon, and a deletion of 36 bases (at position 3) removing part of the body sequence. Moreover, two short deletions of two and six bases (positions 4 and 5) were seen in the PRCV 3' non-coding region by comparison to the Purdue and FS772/70 strain sequences (not yet available for the Miller strain). To summarize, the observed differences between PRCV and TGEV genomes consisted of deletions and point mutations only; no sequence unique to PRCV was found.

Estimation of $M_r$ for PRCV S protein

The above sequence data predicted an $M_r$ of 135K for the PRCV S apoprotein instead of the 158K for TGEV S. An experiment was designed to compare the relative size of PRCV and Purdue virus polypeptides by SDS-PAGE (Fig. 3). Both the PRCV S species and its intracellular precursor S' exhibited an increased mobility as compared to TGEV species S (220K) and S' (175K), formerly named E2 and E'2 (Laude et al., 1986). An $M_r$ of 190K was estimated for the PRCV S polypeptide. The two viruses showed no significant difference in $M_r$ for either polypeptide N or M.

Discussion

In this study we cloned and analysed the 3' region of the genomic RNA of a French isolate of PRCV, a newly recognized porcine pneumotropic coronavirus which is antigenically closely related to TGEV but which does not induce intestinal lesions. The sequenced region, about 7500 nucleotides long, stretched from the beginning of the S glycoprotein gene up to the 3' poly(A) tail and probably encompasses all the viral genes except for the polymerase gene. Thus a substantial part of the genetic information of PRCV was available for comparison to TGEV. It appeared that the genomes of the two viruses had not only a closely related organization, but exhibited also a very low level of nucleotide divergence (4%). However, as schematized in Fig. 4, PRCV and TGEV differed by two striking features.

First, the region homologous to the TGEV ORF NS3a contained no functional ORF (see Fig. 2 and Results). Additional detailed sequence data are needed to ascertain whether this is a common feature of PRCV isolates. However this view is supported by the recent report that the British isolate had the same ORF altered (P. Britton, et al., IV International Coronavirus Symposium, Cambridge, July 1989). The NS3a ORF (NS3 for the Miller strain) has been shown to be transcribed for three different TGEV strains and has the potential to encode a non-structural polypeptide which remains to be characterized (Rasschaert et al., 1987; Britton et al., 1989; Wesley et al., 1989). The efficient propagation of PRCV both in vitro and in vivo suggests that the expression of this gene is not an absolute requirement. Incidentally, alteration of an ORF encoding a putative NS polypeptide is not unprecedented in TGEV; the NS3b ORF from the Purdue strain propagated in our laboratory was reported to have a mutated first ATG codon and, at least in some clones, a 13 base frameshift deletion in the 3' region (Rasschaert et al., 1987) when compared to the sequence from another laboratory (Kapke et al., 1988).

Secondly, the genome of PRCV was found to encode a truncated S protein. Analysis of the S gene nucleotide sequence established from four cDNA clones predicted a 1225 long amino acid precursor polypeptide (see Fig. 2) instead of 1447 (Purdue virus; Rasschaert et al., 1987) and 1449 amino acids (Miller and FS772/70 strains; Wesley, 1990; P. Britton, personal communication). The difference involves the omission in the PRCV S sequence of a continuous stretch of 224 residues starting at position 5 from the N terminus (Fig. 5). The sequence unique to TGEV S includes the 180 residue long segment where the antigenic site D has been mapped (Delmas et al., 1990). This finding is consistent with the observed lack of reactivity of Purdue site D-specific MAbs towards three different PRCV isolates (Laude et al., 1988). Further, it strengthens the view that the truncation of S protein, directly confirmed for the French isolate (Fig. 3), is a typical feature of this virus. Also of interest is the observation that one of the four MAbs defining site D (78.17) retained a significant reactivity towards PRCV, suggesting that the relevant epitope involves additional residues outside the TGEV-specific domain. Alignment of PRCV and TGEV S sequences revealed in addition 33 to 43 positions of non-identity, depending upon the PRCV–TGEV strain pair (Fig. 5).
Porcine coronavirus nucleotide sequence

Fig. 2. Sequence of the extreme 3' 7519 nucleotides of the PRCV genome and deduced amino acid sequences of the major ORFs. The positions of the deletions relative to three TGEV strains (see Results) are indicated by the framed numbers 1 to 5 above the nucleotide sequence. The framed letters A and B show the positions of two sequence segments unique to the Purdue strain (see Discussion). The positions of the deletions relative to three TGEV strains (see Results) are indicated by the framed numbers 1 to 5 above the nucleotide sequence.
tides. Immunoprecipitated material was resolved in a 9 to 20% acrylamide gradient SDS-PAGE gel. Lane 1, mock-infected cells; lane 2, PRCV-infected cells (10^5 c.p.m.); lane 3, TGEV-infected cells (10^6 c.p.m.). *M* of marker proteins are indicated in the left margin.

In particular, one change, Ser to Phe, was present within the nine residue stretch expressing C epitopes (Delmas et al., 1990); this correlates with the failure of two site C-specific MAbs to recognize the PRCV S antigen (Laude et al., 1988). By contrast to site D- and C-specific MAbs, all site A- and B-specific MAbs exhibited a strong cross-reactivity towards PRCV S, despite the existence of 12 amino acid changes in the region stretching from positions 300 to 512, where the relevant epitopes have been mapped (Delmas et al., 1990). This further supports our earlier assumption that biologically crucial determinants are expressed in this domain of the S molecule (Delmas et al., 1986).

The nature of the events responsible for the genomic diversity between PRCV and TGEV remains an open question. Both homologous and heterologous RNA recombinations are proposed to be a driving force for coronavirus evolution (see review by Spaan et al., 1988). Thus the possibility cannot be formally excluded that the sequences unique to TGEV result from insertional events. However, we believe that the idea that PRCV is derived from TGEV through successive deletion events is the most plausible. Recent data indicate that deletions might be responsible for the observed polymorphism of the S protein of mouse hepatitis virus (MHV). In particular, the S genes of three selected MAb-resistant variants of MHV type 4 have been shown to encode truncated S proteins with a large deletion of 142 to 159 amino acids mapping approximately in the N-terminal two-thirds of the S1 subunit of S protein (Parker et al., 1989). Besides, the S protein genes of two MHV strains, JHM (Schmidt et al., 1987) and A59 (Luytjes et al., 1987), harbour a sequence deleted in the same region relative to that of MHV-4. Examination of the TGEV nucleotide sequence did not reveal any feature indicative of a jumping of the RNA polymerase complex such as a site of strong secondary structure, or repeated sequences in the vicinity of the deletion site. The only noticeable feature was the presence of a heptanucleotide sequence AGTTTCC immediately adjacent to the unique sequence of the TGEV S gene which is repeated 10 nucleotides downstream; however, no simple mechanism underlying a possible deletion process could be reconstructed on this basis.

Another conclusion drawn from the comparison of PRCV and of TGEV strains sequences was a closer relationship of PRCV to the fully virulent Miller and FS772/70 strains than to the high passage Purdue strain, of which several sequence features were not seen in PRCV: (i) a six nucleotide deletion in the S gene, encoding the residues Asn and Asp at positions 151 and 152 (asterisks in Fig. 5) and (ii) two short insertions of 16 and 29 nucleotides located upstream and at the 3' end of the NS3a ORF, respectively (corresponding positions in PRCV sequence noted A and B in Fig. 2).

Fig. 4. Compared gene organization of PRCV and of three TGEV strains. The genes are designated according to the latest recommendations of the coronavirus study group (Cavanagh et al., 1990). In particular, the second NS ORF downstream of the S gene is referred to as NS3-1 in the case of the Miller strain to acknowledge the fact that its transcription involves a distinct subgenomic RNA species (Wesley et al., 1989), in contrast to NS3b ORFs of Purdue and FS772/70 viruses. The PRCV ORF named NS3-1 is preceded by an intergenic sequence homologous to that of the Miller ORF NS3-1. The TGEV ORFs NS3a, NS3b, NS4 and NS7 were formerly named X2a, X2b, X1 and X3, respectively (Rasschaert et al., 1987). The diagram was drawn according to the sequence data cited in the Results section.
Porcine coronavirus nucleotide sequence

224 as unique to TGEV

Fig. 5. Alignment of the S protein sequences of PRCV and two TGEV strains. The position of the segment omitted in the PRCV sequence is shown. Amino acid changes relative to Miller and to Purdue strains are indicated above and beneath the PRCV sequence, respectively. The asterisks mark two amino acids absent in the Purdue S protein. The location of the major TGEV antigenic sites C (open box) and A and B (filled box) is given (Delmas et al., 1990).
transferred ORFs revealed a greater overall homology between PRCV and Miller or FS772/70 strains than between PRCV and the Purdue strain (Table 1). A further point to mention with respect to PRCV/TGEV sequence comparison is that the TGEV strains for which sequence data are available were isolated a long time before the PRCV outbreak. Therefore, one cannot exclude the possibility that the observed divergence in terms of point mutations reflects, at least in part, the natural drift of TGEV.

No conclusion can be reached about the genome modification(s) controlling the phenotypic difference observed between PRCV and TGEV, namely the impaired multiplication of the former in the digestive tract. It is tempting to speculate that the truncation of the S gene is relevant as it is well known that the spike glycoproteins play a pivotal role in conferring tissue specificity to enveloped RNA viruses. In this respect, mutants of MHV encoding a point-mutated or a truncated S protein have been shown to be neuroattenuated for mice, thus indicating that at least one virulence determinant is associated with this protein (Dalziel et al., 1986; Fleming et al., 1986; Gallagher et al., 1990). The omission of protein (about 30K) in the proximal region of PRCV S might affect, directly or by induced distant conformational change, a putative domain involved in the recognition of TGEV receptors on the enterocytes. Studies aiming to examine the pathogenicity for piglets together with those reported on MHV allow the proposition that genomic deletion as well as recombination might be a major source of diversity among the coronavirus family. The studies on PRCV provide the first indication that a deleted coronavirus may emerge not only as the result of experimental selection but also spontaneously in the field and, furthermore, may have great epizootiological impact.

We wish to thank Drs C. Dumont and J. P. Soulebot for kindly providing the Rhône-Mérieux PRCV isolate and I. Gehl for her expert assistance in producing PRCV stocks. Thanks are also due to A. M. Spie for typing the manuscript and to K. Rerat for revising the English style. M. Duarte (C.I.V.E.T., Pando, Uruguay) holds a fellowship from the French Government.

References


Table 1. Pairwise sequence homology between PRCV and TGEV polypeptides

<table>
<thead>
<tr>
<th>PRCV predicted polypeptides</th>
<th>TGEV strain</th>
<th>Number of amino acids</th>
<th>aa%</th>
<th>aa%</th>
<th>aa%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Purdue</td>
<td>Miller</td>
<td>FS772/70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>1225</td>
<td>44</td>
<td>97-3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M</td>
<td>262</td>
<td>13</td>
<td>95</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>N</td>
<td>382</td>
<td>14</td>
<td>96-3</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>NS3-1</td>
<td>244</td>
<td>75</td>
<td>97-1</td>
<td>4</td>
<td>98</td>
</tr>
<tr>
<td>NS4</td>
<td>82</td>
<td>3</td>
<td>96-3</td>
<td>4</td>
<td>95-1</td>
</tr>
<tr>
<td>NS7</td>
<td>78</td>
<td>7</td>
<td>91</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

* Number of divergent amino acids (aa).
† Percentage homology.
‡ Sequence data not available.
§ Compared with full-length NS3b ORF (Kapke et al., 1988).


