

VIRUS 00713

The cloning and sequencing of the virion protein genes from a British isolate of porcine respiratory coronavirus: comparison with transmissible gastroenteritis virus genes

Paul Britton, Karen L. Mawditt and Kevin W. Page *

Division of Molecular Biology A.F.R.C., Institute for Animal Health, Compton Laboratory, Compton, U.K.

(Received 28 June 1991; revision received 9 August 1991; accepted 12 August 1991)

Summary

Previous analysis of porcine respiratory coronavirus (PRCV) mRNA species showed that mRNAs 2 and 3 were smaller than the corresponding transmissible gastroenteritis virus (TGEV) mRNA species (Page et al. (1991) *J. Gen. Virol.* 72, 579–587). Sequence analysis showed that mRNA 3 was smaller due to the presence of a new putative RNA-leader binding site upstream of the PRCV ORF-3 gene. However, this observation did not explain the deletion observed in PRCV mRNA 2. Polymerase chain reaction (PCR) was used to generate cDNA from the 3' coding region of the putative polymerase gene to the poly (A) tail of PRCV for comparison to the equivalent region from TGEV. The PRCV S protein was found to consist of 1225 amino acids, which had 98% similarity to the TGEV S protein. However, the PRCV S gene contained a 672 nucleotide deletion, corresponding to 224 amino acids (residues 21 to 245 in TGEV S protein), 59 nucleotides downstream of the S gene initiation codon. The PRCV genome from the ORF-3 gene to the poly (A) tail was sequenced for comparison to TGEV in order to identify other potential differences between the two viruses. Four ORFs were identified that showed 98% similarity to the TGEV ORF-4, M, N and ORF-7 genes. No other deletions or any PRCV specific sequences were identified.

Correspondence to: P. Britton, Division of Molecular Biology, A.F.R.C., Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire, RG16 0NN, U.K.

* *Present address:* Institute of Virology, University of Glasgow, Church Street, Glasgow, G11 5JR.

Porcine respiratory coronavirus; Transmissible gastroenteritis virus; Coronavirus; Spike; Membrane protein; Nucleoprotein

Introduction

Porcine respiratory coronavirus (PRCV) appeared about 1984 and rapidly spread throughout the pig population in several, if not all, European countries where it now persists enzootically (Pensaert et al., 1986) and was recently isolated in the United States (Wesley et al., 1990a). The virus, initially isolated in Belgium (Pensaert et al., 1986) and in Britain (Brown and Cartwright, 1986), produced serological responses following infection of pigs that could not be distinguished from transmissible gastroenteritis virus (TGEV), by available diagnostic tests. The most striking difference between the two viruses was seen in their pathology. PRCV grows principally in the respiratory tract, producing mild or no clinical signs (Pensaert et al., 1986; Pensaert, 1989; O'Toole et al., 1989; Cox et al., 1990), although, Van Nieuwstadt and Pol (1989) found that a Dutch isolate of PRCV intranasally inoculated into SPF pigs caused a fatal pneumonia. In contrast, although TGEV can grow in the respiratory tract, the virus preferentially grows in the enterocytes covering the tips of the villi in the small intestine, causing diarrhoea and dehydration resulting in high morbidity and mortality in neonatal pigs.

PRCV has been classified as a coronavirus, a group of enveloped viruses with a positive-stranded RNA genome, belonging to the family *Coronaviridae*. TGEV (Britton et al., 1986; Jacobs et al., 1986) and PRCV (Britton et al., 1990) infected cells, in addition to the genomic RNA (mRNA 1), have six species of subgenomic mRNA (mRNAs 2–7) which form a 3' co-terminal "nested" set. The TGEV (Garwes and Pocock, 1975) and PRCV (Britton et al., 1990) virions contain three major structural polypeptides; a surface glycoprotein (spike (S) with a monomeric relative molecular mass (M_r) 200,000, a glycosylated integral membrane protein (M), observed as a series of polypeptides between M_r 28,000–31,000 and a basic phosphorylated nucleoprotein (N) of M_r 45,000 associated with the viral genomic RNA. However, the PRCV S protein appears to have a slightly lower M_r than TGEV on polyacrylamide gels (Britton et al., 1990).

PRCV was neutralised in vitro by antisera prepared against TGEV and the majority of monoclonal antibodies (MAbs) raised against any of the TGEV virion proteins cross reacted with PRCV (Sánchez et al., 1990). However some MAbs, raised against the S protein of either the virulent British strain FS772/70 (Garwes et al., 1988) or the avirulent Purdue strain of TGEV (Callebaut et al., 1988), did not recognise PRCV. Callebaut et al. (1988) showed that the MAbs which did not react with PRCV, produced from the Purdue strain of TGEV, were derived from three separate antigenic sites and mapped between amino acid residues 17 and 325 (Correa et al., 1990). However, MAbs derived from one of these epitopes on the

Purdue-115 strain did not react with TGEV strains FS772/70 and Miller due to a point mutation in the S protein sequences.

The molecular characterisation of a British isolate of PRCV has been undertaken with the aim of identifying differences between the PRCV and TGEV genomes which may be linked to the different tropisms and pathogenicities of the two viruses. Initial work found that: (1) the putative leader RNA sequences present on the viral RNA species, postulated to be involved in transcription of the mRNA species, were identical for TGEV and PRCV (Page et al., 1990); (2) two of the PRCV mRNA species, 2 and 3, were smaller than the corresponding TGEV species (Britton et al., 1990; Page et al., 1991); (3) sequencing studies on the PRCV genome, corresponding to the 5'-end of TGEV mRNA 3, identified several small deletions (84 nucleotides) resulting in the loss of the potential TGEV ORF-3a gene (Page et al., 1991). However, the 84 nucleotides deleted did not account for the size difference, about 600 nucleotides, observed between TGEV and PRCV mRNA 2 nor did the deletions explain the differential binding of the MAbs. In this paper we present sequence data of the PRCV S gene to identify any variation to the TGEV S gene that would account for the size difference of mRNA 2 and the differential reaction of the MAbs and the sequence of the PRCV genome from the ORF-3 gene to the 3' poly (A) tail to identify any other variations in the PRCV genome.

Materials and Methods

Preparation of viral RNA

Viral RNA was isolated from LLC-PK1 cells infected with PRCV strains 86/137004 or 86/135308 as described by Britton et al. (1987) and Page et al. (1990).

Preparation of oligonucleotide primers

Oligonucleotides used for PCR amplifications were synthesised by the phosphoramidite method on an Applied Biosystem 381A DNA synthesizer. These were derived from published TGEV sequence data (Britton et al., 1988a; 1988b; 1989; Page et al., 1990; Britton and Page, 1990) and are listed in Table 1. Fig. 1 shows the position of the oligonucleotides on the TGEV viral genome.

Cloning of PCR generated fragments

Synthesis of first-strand cDNA was carried out in 30 μ l samples containing 5 μ g total RNA (isolated from virus infected cells), 40 U RNasin (Promega), 50 mM Tris-HCl pH8.3, 3 mM MgCl₂, 75 mM KCl, 10 mM DTT, 2.5 mM dNTPs and primed with 160 ng of oligos 55, 14, 17, 41, 51, 25 or 85 using 23 U of avian myeloma virus (AMV) reverse transcriptase (Super-RT; Anglian Biotech) at 42 °C

Table 1

Sequence of oligonucleotide primers used for PCR amplifications

OLIGONUCLEOTIDE	SEQUENCE	SENSE
oligo 55	5'-AGTAACACAACACTCTTA-3'	-
oligo 32	5'-TGTTGCCATTAAATCA-3'	+
oligo 14	5'-CGTGACGTTACCAGTGC-3'	-
oligo 18	5'-AGATTGCTATTAGTAAG-3'	+
oligo 17	5'-ACATACTAAGTCAGCTA-3'	-
oligo 13	5'-CAGTGCTACACCTAGATTCATG-3'	+
oligo 41	5'-TTTTCAATAGGTTTCGTA-3'	-
oligo 76	5'-AAACGTAAGTATCGTTCAG-3'	+
oligo 51	5'-CTGTCCTTCCTAAATTGCAACACACCATGCATAGC-3'	-
oligo 52	5'-GGCCTTGGTATGTGTGGCTACTAATAGGC-3'	+
oligo 25	5'-GGTATGTTATTACTCTTC-3'	-
oligo 60	5'-GTGTCGGCATCTTAATG-3'	+
oligo 85	5'-TTTTTGTATATCACTATC-3'	-
oligo 75	5'-CCTTTTAAAGTAAAGTGAGT-3'	+

Note: all the above oligonucleotide sequences were derived from the sequence of TGEV strain FS772/70 except oligo 75 which corresponds to the 5'-end of the leader RNA sequence from TGEV FS772/70, PRCV 86/137004 and PRCV 86/135308.

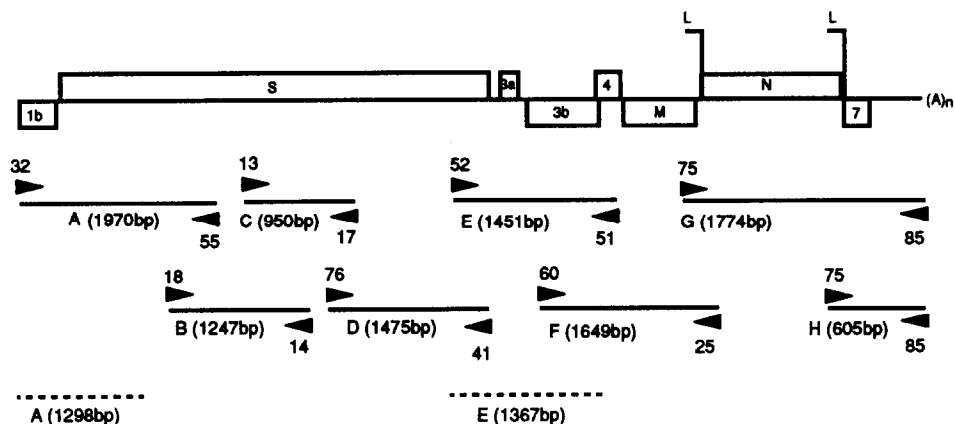


Fig. 1. Schematic diagram of the TGEV/PRCV genome showing the position of the oligonucleotides used for PCR cloning. The arrow heads show the position and the orientation of the oligonucleotides. The boxes show the positions of the TGEV/PRCV genes. The lines show the sizes of the PCR amplified fragments expected from the TGEV sequence and the dotted lines show the sizes of the PRCV fragments if different from the equivalent TGEV fragments. The L denotes the position of the putative leader RNA sequences, upstream of the N and ORF-7 genes, on mRNA species 6 and 7 from which oligo 75 was used for PCR amplifications.

for 90 min. The ssDNA was PCR amplified, using the oligonucleotide primers shown in Table 1, following a protocol supplied with the AmpliTaq™ kit (Perkin-Elmer-Cetus) in a Techne PHC-1 programmable thermal cycler using 35 cycles of 94 °C for 1 min, 40 °C for 2 min and 72 °C for 3 min with a final elongation step of 72 °C for 9 min. The PCR generated cDNA fragments were purified by agarose gel electrophoresis and isolated from the gel using Geneclean™. The cDNA was 5'-phosphorylated using T₄ polynucleotide kinase (Gibco, Bethesda Research Laboratories) and any incomplete ends repaired using Klenow fragment (Pharmacia), prior to ligation into *Sma*I-cut dephosphorylated pUC13 (Pharmacia). The resulting plasmids were transformed into *E. coli* strain JM105 and ampicillin resistant transformants directly analysed by a modification of the method of Güssow and Clarkson (1989). The transformants were grown in 2 ml of X2 LB containing 100 µg ml⁻¹ ampicillin of which 400 µl samples were centrifuged, the cell pellets resuspended in 500 µl H₂O and boiled for 5 min. The cell debris was centrifuged for 5 min and 10 µl aliquots of the supernatants PCR amplified, using universal and reverse primers, for 30 cycles of 94 °C for 1 min, 55 °C for 2 min, 72 °C for 2 min with a final elongation step at 72 °C for 9 min. The reaction products were analysed on 1% agarose gels. Plasmid DNA was isolated from transformants containing PRCV cDNA fragments of the expected length as outlined in Fig. 1.

Sequencing of cloned PCR fragments

PRCV cDNA was cut from plasmids, using *Bam*HI and *Eco*RI, and ligated into *Eco*RI and *Bam*HI digested M13 mp18 and 19 phage vectors. The PRCV cDNA was sequenced, from the M13 ssDNA templates using the Sequenase™ (United States Biochemical Corporation) protocol and oligonucleotide primers derived from TGEV sequence data. Each cDNA fragment was sequenced several times in both directions to eliminate any ambiguous data.

Data handling and analysis

A sonic digitizer (Graf/Bar; Science Accessories Corporation) was used to read data into a Elonex PC-286 microcomputer and data were analysed on a MicroVAX 3600 using the computer programs of Staden (1982), the University of Wisconsin Genetics Computer Group (UWGCG; Devereux et al., 1984) and CLUSTAL (Higgins and Sharp, 1988).

Results

Cloning of PRCV RNA

The PRCV (86/137004) genome upstream of the S gene to the poly (A)-tail was cloned by PCR amplification. Eight cDNA fragments of 1298 bp (A), 1247 bp (B),

950 bp (C), 1475 bp (D), 1367 bp (E), 1649 bp (F), 1774 bp (G) and 605 bp (H) were amplified from PRCV RNA using appropriate primers (Fig. 1). The sizes of fragments B, C, D, F, G and H were as expected from the TGEV sequence data, however, fragments A and E were 1298 bp and 1367 bp in contrast to the TGEV 1970 bp and 1451 bp fragments (Fig. 1). The size difference, 84 bp, in PRCV fragment E was due to the several small deletions reported by Page et al. (1991).

End of Oligo 32

```

AATCAGCGAATTAGTGGGAATAATATTTATGAATGATTCAAGATTGGAGTTGGAGTGTGTTTGTACAGCTGTAACACAGCTCATCATCAGAGGCTTCTGATTGGTGTAA 120
I T E F S W N K Y L Y E L I Q R F E Y N T V F C T S V N T S S S E G F L I G V N 40
D

CTACTTAGGACCATCTGTCGACAAGCAATAGATAGGGAATATATGCAATGCCAATATATATTTGGAGAAATCTCAACATGGCCTATCATCAACTCAGCTCAGACATCC 240
Y L G F S Y D K A I V D G N I H M A N T A N T I M A L S H N S V L D T C 80

TAAATTCAGGTGCTGTGTAACAACGCACTTATTGTAAATTTAAAGAAAAAAGAAATGAATGAATGGTTATTGGATTACTAAGGAAGGGTAAGTGCTCATTAGAAAAATGGTAAGTT 360
K F K C R C N N A L I V N L K E K E L N E M V I G L L R K G K L L I R N N G K L 120

ACTAAACTTTGGTAACCACTTTATTAACACACCATGAAAAAATATTGTGGTCTGGTGTGAATGCCATTGATTATGGAGACAAGTTTCCACTCTCCGTAGTTTCCAATTGCACGTAT 480
L N F G N H F I N T P * 131
V (11)

Spike M K K L F V V L V V M P L I Y G D K F P T S V V S N C T D 29
CAATGTGCTAGTTATGTGGCTAATGTGTTTACTACACAGCAGGAGGCTTTATACCATCAGATTTAGTTTTAAATAATGGTTCATCCTCACTAATAGTCCACGTTGGTTAGTGCCAAA 600
Q C A S Y V A N V F T T Q P G G F I P S D F S F N N W F I L T N S S T L V S G K 69
TTAGTTACCAACAGCCTCTATAGTTAAATGCTTATGGCAGTCCCTAGCTTTGAAGAAGCAGCTTCTACATTTGTGTTGAAGGTGCTGACTTTGATCAATGTAATGGTGTCTGTTTA 720
L V T K Q P L L V N C L W P V P S F E E A A S T F C F E G A D F D Q C N G A V L 109
AATACTGCTAGACGCTAGTTAGTTTAACTCTAATTTTACTACAATGTACAATCAGGTAAAGGGTGCTACAGTGTGTTTCATTGAACACACCGGTGCTGACTCTTGAATCTCAGTGT 840
N N T V D V I R F N L N F T T N V Q S G K G A T V F S L N T T G G V T L E I S C 149
TATAATGATACAGTGAGTGATTGCGAGTTTTCAGTTACGGTGAATTCGGTCGGCGTAACTAATGGACCAGGTACTGTGTAGTACTCTATAATGGACAGCTCTTAAGTATCTAGGA 960
Y N D T V S D S S F S S Y G E I P F G V T N G P R Y C Y V L Y N G T A L K Y L G 189

oligo 18
ACATTACCACCTAGTGCAAGGAGATGCTATTAGTAAGTGGGGCCATTTTATATTAATGGTTACAATTTCTTACGACATTTCCATTGATTGTATATCTTTTAATTGACTACTGGT 1080
T L P P S V K E I A I S K W G H F Y I N G Y N F F S T F P I D C I S F N L T T G 229
GATAGTGAGCTTCTGGCAATAGCTTACACATCGTACACTGAGCAATAGTACAAAGTTGAACACAGCTATTACAATGTGACGATTTGTAATAGTTATGTTAATAACATTAATGC 1200
D S D V F W T I A Y T S Y T E A L V Q V E N T A I T N V T Y C N S Y V N N I K C 269

oligo 55
TCTCAACTTACTGCTAATTTGAATATGGATTTTCTCTGTTCTTCAAGTGAAGTTGGTCTGTGTAAGAGTGTTGTTGTTACTACCTAGCTTCTGACACATACCATTTGTAACATA 1320
S Q L T A N L N N G F Y P V S S S E V G S V N K S V V L L P S F L T H T I V N I 309
ACTATTGCTCTGGTATGAAGCGTAGTGGTTATGGTCAACCCATAGCCTCAACGCTAAGTAACATTACACTACCAATGCGAGATAACAACACCGGTGCTGACTGTGTTGCTTCTGACAA 1440
T I G L G M K R S G Y G Q P I A S T T L S N I T L P H Q D N N T D V Y C V R S D Q 349
TTTTCAGTTATTGTCATTCTACTTGCACAAAGTGCTTTATGGGCAATGTTTAAAGCGAACTGCACGGAGCTTTTAGATGCCACAGCTGTTATAAAAACTGGTACTGTGCTTTCTCA 1560
F S V Y V H S T C K S A L W D N V F K R N C T D V L D A T A V I K T G T C P F S 389
TTTGATAAATTGAACAACTACTTTAACTTTTAAAGAAATCTGTTTGTGCTGTGAGTCCTGTTGGTGCTAATTGTAAGTTTGAATGTAGTGCCTGCTGACAGAACCAATGATCAGTTGTGTAGA 1680
F D K L N N Y L T F N K F C L S L S P V G A N C K F D V A A R T R T N D Q F V R 429

oligo 13
AGTTTGTATCTAATATATGAAGAAGGAGACGATAGTTGGTGTACCGCTGACAACTAGTGTTTGCACGATTGTGTCAGTGCTACACCTAGATTGCTGCACAGATTACAATATATAGTT 1800
S L Y V I Y E E G D S I V G V P S D N S G L H D L S V L H L D S C T D Y N I Y G 469
AGAATCGTGTGTTGATATTAGACAACTAACAGGACGCTACTTAGTGGCTATATTACACATCAGTATCGGTGATTGCTTAGGTTTAAAAATGTTAGTATGCTGTTACTACTCT 1920
R T G V G I I R Q T N R T L L S G L Y Y T S L S G D L L G F K N V S D G V I Y S 509
GTAAACGCTGTGATGTTAGCGCAACAGCAGCTATTATTAGTGGTGCCATGTTGGGGCTATCACTTCCATTAAACAGAAATGTTAGCTTCAACAGATTGGCAANTAAACACCTAAATTT 2040
V T P C D V S A Q A I D G A I V G A I T S I N S E L L A L T H W T I T P N F 549
TATTACTACTCTATATATAATACACAAATGATAAGACTCGTGCCACTCCAATTGGCAGTAATGACGTTGATTGTGAACCTGTCATACCTATTCTAACATAGTGTGTTGTAAAAATGGT 2160
Y Y Y S I Y N Y T N D K T R G T P I G S N D V D C E P V I T Y S N I G V C K N G 589

oligo 14
GCTTTGGTTTATTATTAACGTCACACATTTCTGATGACAGCTGCAACCAATTAGCAGTGGTAACGTCACGATACCTACTAACTTTACTATATCCGTGCAAGTGAATACATACAGTTTAC 2280
A L V F I N V T H S D G D V Q P I S T G N V T I P T N F T I S V Q V E Y I Q V Y 629

ACTACACAGGTGCAATAGACTGTTCAGATATTGTTGTAATGCCAACCTAGGTGTAAACAACTGTTAAACAACTACGTTTCTGATGTCAAACATTTAGCAAGCACTTGCATGAGGTT 2400
T T P V S I D C S R Y V C N G N P R C N K L L T Q Y V S A C Q T I E Q A L A M G 669

```

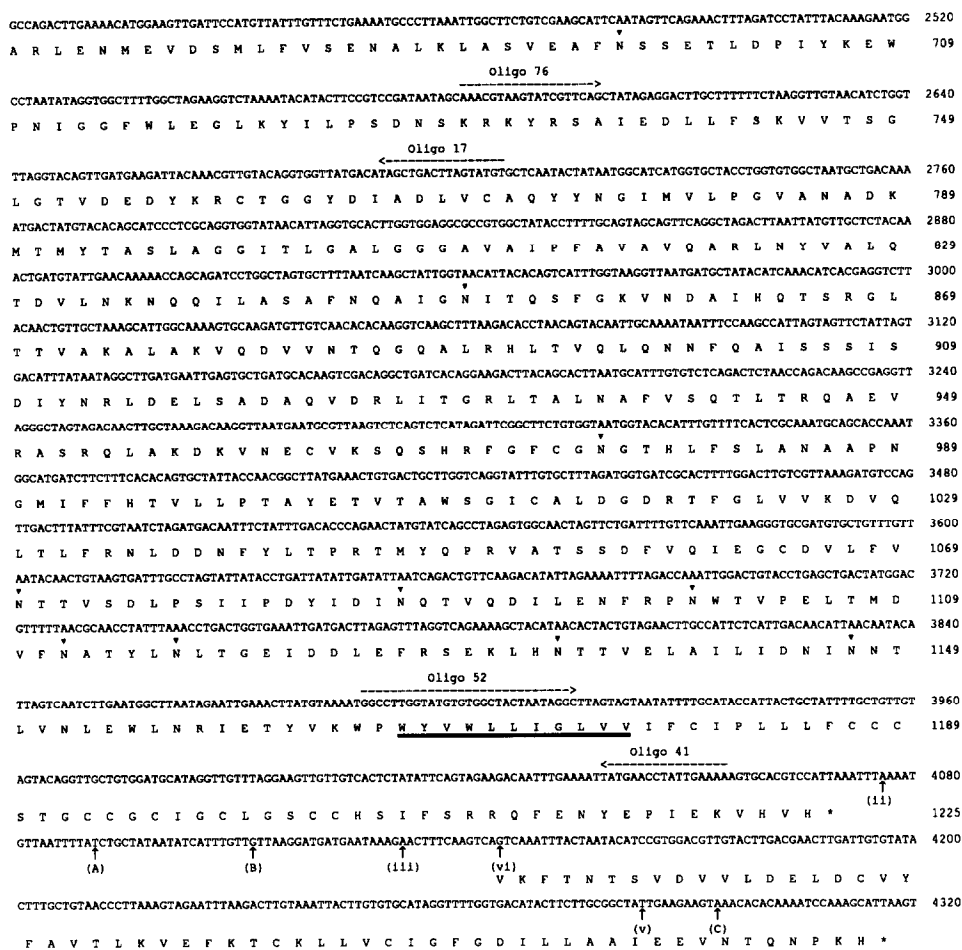


Fig. 2. The nucleotide and deduced amino acid sequences of the carboxyl-terminus of the 1b polymerase and S genes, including the ORF-3a pseudogene, from PRCV 86/137004. The horizontal arrows show the position and orientation of the primers used to generate the PCR fragments. The ACTAAAC sequence upstream of the S gene is identified with a thick line. Amino acids below the PRCV polymerase sequences are substitutions found for the FS772/70 strain of TGEV. The double underlined sequence at the beginning of the S gene is the predicted N-terminal signal sequence. Potential N-glycosylation sites (NXT or NXS) are identified with a black triangle. The double underlined sequence at the carboxyl-terminus of the S protein shows the position of the potential transmembrane domain. The numbered vertical arrows shown below the PRCV nucleotide sequence indicate the positions of the deletions found when compared to the TGEV FS772/70 sequence, corresponding to (i) 672, (ii) 9 (this deletion is also found in TGEV strains Purdue-115 and Miller), (iii) 13, (iv) 22 and (v) 36 nucleotides (Page et al., 1991). The lettered vertical arrows below the PRCV sequence indicate insertions, corresponding to (A) 3, (B) 16 and (C) 29 nucleotides (Page et al., 1991), found in either TGEV Purdue-115, (B) and (C), or both Purdue-115 and Miller, (A), that are not found in either TGEV FS772/70 or PRCV. These sequence data will appear in the EMBL/GenBank/DBJ Nucleotide Databases under the accession number X60089.

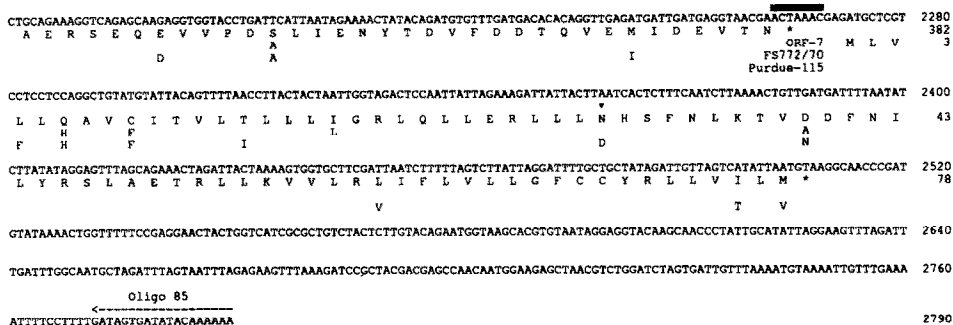


Fig. 3. The nucleotide sequence of the PRCV (86/137004) genome from the 3'-end of ORF-3 gene to the poly (A) tail. The amino acid sequences corresponding to the carboxyl-terminus of ORF-3, ORF-4, M, N, and ORF-7 gene products are shown below the nucleotide sequences. The horizontal arrows show the position and orientation of the primers used to generate the PCR fragments. The position of oligo 75 is not shown as it corresponds to the 5'-end of the leader RNA sequence and was used to generate PCR fragments from the two smallest mRNA species, mRNAs 6 and 7. The position of oligo 60 is not shown as it corresponded to a region within PRCV ORF-3, not shown in this figure, described by Page et al. (1991). Amino acid substitutions found on the TGEV FS772/70 (Britton et al., 1988a; 1988b) and Purdue-115 (Laude et al., 1987; Rasschaert et al., 1987) sequences are shown below the PRCV sequences. Positions of the putative RNA-leader binding sites CTAAAC for ORF-4 and ACTAAAC for M, N and ORF-7 genes are shown as thick lines above the nucleotide sequence. The predicted N-terminal membrane signal sequence of the M protein is double underlined. The thick lines below the M amino acid sequence show the positions of the predicted transmembrane domains. The number sign shows the position of the amino acid deleted from the ORF-4 gene in TGEV FS772/70. The black triangle shows the position of potential N-glycosylation sites. The nucleotide sequence data reported in this paper for PRCV strain 86/137004 have been submitted to the EMBL/Genbank/DBJ nucleotide sequence databases and have been assigned the accession number X60056.

The observation that the PRCV fragment A was about 600 bp smaller than the equivalent TGEV fragment would account for the observed difference in the size of PRCV mRNA 2 (Britton et al., 1990; Page et al., 1991). The PRCV cDNA fragments A to H were cloned into pUC13 and the corresponding plasmids pPR137-7 (A), pPR137-9 (B), pPR137-11 (C), pPR137-13 (D), pKP-1 (E), pPR137-5 (F), pPR137-1 (G) and pPR137-3 (H) were used for DNA sequencing.

Sequencing of PRCV cDNA

The PRCV cDNA from the above plasmids, except pKP-1, was subcloned into M13 vectors and ssDNA templates sequenced. The PRCV cDNA from pKP-1 was sequenced as described by Page et al. (1991).

The sequence of 4320 bp (Fig. 2), derived from the PRCV cDNA in fragments A-D and part of E, is shown in Fig. 2. The sequence was translated in all six reading frames. An ORF of 3678 bp, nucleotides 394-4071, corresponding to a gene product of 1225 amino acids preceded by the potential RNA polymerase-leader complex binding site, ACTAAAC, 26 bp upstream of the initiation site (Fig. 2) was identified as the PRCV S gene. Comparison of the PRCV and TGEV S genes identified a deletion of 672 nucleotides, corresponding to 224 amino acids

(TGEV residues 21 to 245), 59 nucleotides downstream of the PRCV S gene initiation codon (Figs. 2 and 4). The rest of the deduced amino acid sequence of the PRCV S protein showed 98% similarity to the TGEV S protein. The 5'-end region of the PRCV S gene from a second British isolate of PRCV, strain 86/135308, was cloned and sequenced and found to have an identical deletion. The first 20 amino acids of the PRCV S protein were identical to those found on TGEV of which the first 16 fulfil the criteria for a eukaryotic signal sequence with the potential cleavage site between the glycine (16) and aspartic acid (17) residues (Fig. 2). This cleavage site has been confirmed, for the avirulent Purdue strain of TGEV, by N-terminal amino acid sequencing of the S protein isolated from virions (Rasschaert and Laude, 1987). Assuming that the PRCV signal sequence is cleaved the PRCV S protein would comprise a polypeptide of 1209 amino acids with a M_r 132,897 compared to 1433 amino acids with a M_r 158,160 for TGEV S protein.

An incomplete ORF of 397 bp was identified, nucleotides 2–397 (Fig. 2), which consisted of 131 amino acids that terminated at a TGA stop codon. The deduced amino acid sequence of this ORF had 99.2% similarity with 97.7% identity to an ORF found upstream of the TGEV FS772/70 S gene, previously identified as the C-terminal end of the TGEV ORF-1b polymerase subunit, because of its homology to the ORF-1b polymerase subunits of infectious bronchitis virus (IBV; Boursnell et al., 1987) and mouse hepatitis virus (MHV; Bredenbeek et al., 1990).

A sequence of 57 amino acids, corresponding to nucleotides 4145–4318 (Fig. 2), identified downstream of the PRCV S gene that had no ACTAAAC site or initiation codon had 71.9% similarity with 70.2% identity to part of the TGEV FS772/70 ORF-3a gene indicating that they formed part of a pseudogene corresponding to TGEV ORF-3a. The 174bp PRCV ORF-3a pseudogene had the first three amino acids deleted and contained a 36 base deletion which resulted in the loss of the last 13 amino acids when compared to TGEV ORF-3a. However, an in frame fusion at nucleotide 4284 (Fig. 2) resulted in 11 extra amino acids at the C-terminal end of the PRCV pseudogene product which were identical to the amino acids found at the end of ORF-3a from the Miller strain of TGEV (Wesley et al., 1989). Work by Page et al. (1991) has shown that PRCV has no mRNA species with an ORF equivalent to TGEV ORF-3a at the 5'-end indicating that a gene product equivalent to TGEV ORF-3a will not be produced by PRCV though part of this potential TGEV gene appears as a pseudogene in the PRCV genome.

Sequence data of 2790 bp from the PRCV cDNA fragments F–H is shown in Fig. 3. Four complete ORFs, nucleotides 59–304, 318–1103, 1119–2264 and 2273–2506, corresponding to gene products of 82, 262, 382 and 78 amino acids (Fig. 3). The nucleotide sequence of the PRCV cDNA showed 98% similarity with the equivalent region from TGEV. The amino acid sequences of the four ORFs were almost identical to the ORF-4, M, N and ORF-7 gene products of TGEV (Fig. 3). The initiation codon of ORF-4 was preceded by the hexameric sequence, CTAAAC, and the initiation codons of the M, N and ORF-7 genes were preceded by the heptameric sequence, ACTAAAC.

The potential product of the PRCV ORF-4 gene is very similar to the equivalent TGEV gene product (Fig. 2) and showed 96% identity to the FS772/70

(Britton et al., 1989) and Purdue-115 (Rasschaert et al., 1987) but 100% identity to the virulent Miller (Wesley et al., 1989) strains of TGEV. The isoleucine at residue 55 of the Purdue-115 and Miller TGEV sequences, found to be deleted from the FS772/70 sequence, was present in the PRCV sequence. There were three amino acid substitutions between the PRCV and FS772/70 ORF-4 sequences and four amino acid substitutions between the PRCV and Purdue-115 ORF-4 sequences, none of which occur between the two TGEV sequences.

Analysis of the PRCV M protein amino acid sequence by the method of von Heijne (1986) identified a potential membrane signal sequence with the cleavage site between the Gly¹⁶ and Glu¹⁷ residues (Fig. 3). A similar M protein membrane signal sequence was identified in the TGEV (Laude et al., 1987; Britton et al., 1988b) and on the M protein from the antigenically related coronavirus feline infectious peritonitis virus (FIPV; Vennema et al., 1991).

Comparison of the PRCV N protein and ORF-7 sequences to those of TGEV FS772/70 (Britton et al., 1988a) and Purdue-115 (Kapke and Brian 1986; Rasschaert et al., 1987), Fig. 3, showed that there were no deletions or insertions within the PRCV genes. No deletions were found within the 3' non-coding region in contrast to the two and five base deletions observed in the non-coding region of the RM4 isolate of PRCV (Rasschaert et al., 1990). The N protein of PRCV contained eight or 14 amino acid substitutions when compared to the FS772/70 and Purdue-115 sequences respectively (Fig. 3) of which six were identical between the TGEV strains. The PRCV sequence contained the octameric sequence, GGAAGAGC, at the 3'-end of the genome, upstream of the poly (A) site, conserved in all coronavirus sequences to date.

Discussion

In this study the 3'-end of the genome from a British isolate of PRCV was cloned and sequenced. The region analysed consisted of 7822 nucleotides and extended from the 3'-end of the 1b subunit of the PRCV polymerase gene to the start of the poly(A) tail. Previous work (Britton et al., 1990; Page et al., 1991) had shown that two of the PRCV mRNA species, mRNA 2 and mRNA 3, were smaller than the corresponding TGEV species. We concluded that PRCV mRNA 3 was smaller due to the creation of a new putative RNA-leader binding site upstream of a gene equivalent to the TGEV ORF-3b gene with the loss of the putative RNA binding site upstream of TGEV ORF-3a. The new putative RNA-leader binding site resulted in a PRCV mRNA 3 species, about 200 nucleotides smaller than the equivalent TGEV mRNA, with a gene equivalent to TGEV ORF-3b at the 5'-end. The study also identified several small deletions, 84 nucleotides in total, in the PRCV genome corresponding to the 5'-end of TGEV mRNA 3. Two of the deletions were in the area corresponding to the potential TGEV gene, ORF-3a, resulting in the loss of this potential gene in PRCV, although part of the gene remained as a pseudogene. This observation was confirmed on a second British isolate (Page et al., 1991) and by Rasschaert et al. (1990) for a French isolate of

PRCV indicating that it may be a common feature of PRCV. However, these small deletions could not account for the smaller size of PRCV mRNA 2. This was due to a major deletion of 672 nucleotides, corresponding to 244 amino acids, equivalent to amino acid residues 5 to 229 on the TGEV S protein following cleavage of the membrane signal (Fig. 4). The deletion was shown to be present on two British isolates of PRCV and a French isolate (Rasschaert et al., 1990) indicating that this deletion may also be a common feature of PRCV. A cDNA probe, pD24, corresponding to amino acids 66 to 154 of the S protein from TGEV (Miller strain) reacted with the Miller and Purdue-115 strains of TGEV but not to FIPV, feline enteric coronavirus (FECV), canine coronavirus (CCV) or PRCV (strain ISU-1) isolated from a pig in Indiana, U.S.A. (Bae et al., 1991). On the other hand a cDNA probe, pE21, corresponding to amino acids 325 to 451 of the Miller S protein, reacted with both TGEV strains, PRCV (ISU-1), FIPV, FECV and CCV. The region of the Miller S gene corresponding to probe pD24 was within the region deleted from PRCV implying that the American isolate of PRCV may also contain the S gene deletion. The sequence data presented in this study showed that the gene encoding the PRCV ORF-1b polymerase subunit is directly upstream of the S gene, in a similar position as observed for TGEV, with no intervening ORFs, in contrast to bovine coronavirus (BCV) and MHV where other genes or pseudogenes are present. Therefore the results presented in this paper and from previous work indicate that the order of viral genes in PRCV was 5'-[1b-S-3-4-M-N-7]-3' whereas TGEV has the gene order 5'-[1b-S-3a-3b-4-M-N-7]-3'. The sequence data also showed that PRCV had only diverged by about 4% when compared to TGEV sequence data.

Previous comparisons of the S protein sequences from two virulent TGEV strains, FS772/70 (Britton and Page, 1990) and Miller (Wesley 1990), and one avirulent strain, Purdue-115 (Jacobs et al., 1987; Rasschaert and Laude, 1987), identified a six base insert in the genomes of the virulent strains resulting in two extra amino acids with an extra potential N-glycosylation site and an amino acid substitution at residue 384, serine to phenylalanine, which prevented the binding of MAbs belonging to site C (Delmas et al., 1990) or site D (Correa et al., 1988; 1990; Posthumus, 1990). The British and French PRCV isolates were shown to contain both the six nucleotide insert and point mutation. Comparison of S protein sequences from the antigenically related coronaviruses, TGEV and FIPV, showed 94% similarity except for the first 267 amino acids which had only 30% similarity (Jacobs et al., 1987). The deletion within the PRCV S protein was in the same

Fig. 4. Alignment of the S protein amino acid sequences from TGEV strains FS772/70 (Britton and Page, 1990), Miller (Wesley 1990) and Purdue-115 (Rasschaert and Laude, 1987), PRCV 86/137004 (this paper) and FIPV 79-1146 (De Groot et al., 1987) by the CLUSTAL program. The plusses are padding characters for optimal alignment of the sequences. The asterisks below the sequences show identical amino acids in all five viruses and the minus is used if the amino acids match from four viruses. The thick line above the sequences shows the position of the heptad repeat identified by Rasschaert and Laude (1987) for TGEV and the row of arrow heads shows the position of the potential transmembrane domain also present in all the virus sequences.

FS772/70 MKKLFVVLVVM++++PLIYGDNFPCKSLNTRTIGNHWNLIETFLNYSRSLSPNSDVLDGYFTVPQWFNCIRNNSNDLYVTLENLKALY++DYATEN
Miller MKKLFVVLVVM++++PLIYGDNFPCKSLNTRTIGNHWNLIETFLNYSRSLSPNSDVLDGYFTVPQWFNCIRNNSNDLYVTLENLKALY++DYATEN
Purdue MKKLFVVLVVM++++PLIYGDNFPCKSLNTRTIGNHWNLIETFLNYSRSLSPNSDVLDGYFTVPQWFNCIRNNSNDLYVTLENLKALY++DYATEN
PCV MKKLFVVLVVM++++PLIYGDNFPCKSLNTRTIGNHWNLIETFLNYSRSLSPNSDVLDGYFTVPQWFNCIRNNSNDLYVTLENLKALY++DYATEN
FIPV MKKLFVVLVVM++++PLIYGDNFPCKSLNTRTIGNHWNLIETFLNYSRSLSPNSDVLDGYFTVPQWFNCIRNNSNDLYVTLENLKALY++DYATEN
* * * * *

FS772/70 STWNHQRNLNVVNGVPYSITVTTTRFNFSAEGLI++CICKGSPPTTTTESSLTCNWSGECRLNHH++FPICPSNSEANCMBLYGLQWFADEVAY
Miller STWNHQRNLNVVNGVPYSITVTTTRFNFSAEGLI++CICKGSPPTTTTESSLTCNWSGECRLNHH++FPICPSNSEANCMBLYGLQWFADEVAY
Purdue STWNHQRNLNVVNGVPYSITVTTTRFNFSAEGLI++CICKGSPPTTTTESSLTCNWSGECRLNHH++FPICPSNSEANCMBLYGLQWFADEVAY
PCV STWNHQRNLNVVNGVPYSITVTTTRFNFSAEGLI++CICKGSPPTTTTESSLTCNWSGECRLNHH++FPICPSNSEANCMBLYGLQWFADEVAY
FIPV STWNHQRNLNVVNGVPYSITVTTTRFNFSAEGLI++CICKGSPPTTTTESSLTCNWSGECRLNHH++FPICPSNSEANCMBLYGLQWFADEVAY
* * * * *

FS772/70 LHGASYRISFENQMSGTVTLDGMRRATLETAGTLVDLMWFNPVY++++DVSYYRVNNKNGTTVSNCND++QCASVANVFTTPQGGFIPSDFSFNHN
Miller LHGASYRISFENQMSGTVTLDGMRRATLETAGTLVDLMWFNPVY++++DVSYYRVNNKNGTTVSNCND++QCASVANVFTTPQGGFIPSDFSFNHN
Purdue LHGASYRISFENQMSGTVTLDGMRRATLETAGTLVDLMWFNPVY++++DVSYYRVNNKNGTTVSNCND++QCASVANVFTTPQGGFIPSDFSFNHN
PCV LHGASYRISFENQMSGTVTLDGMRRATLETAGTLVDLMWFNPVY++++DVSYYRVNNKNGTTVSNCND++QCASVANVFTTPQGGFIPSDFSFNHN
FIPV LHGASYRISFENQMSGTVTLDGMRRATLETAGTLVDLMWFNPVY++++DVSYYRVNNKNGTTVSNCND++QCASVANVFTTPQGGFIPSDFSFNHN
* * * * *

FS772/70 FLLNSSLTSLVSGKLVTKQPLLVNCLMVPVSFEAAASTCFCEGAGFDQCNQAVLNNVTVDVIRFNLLNTTNVQSGKGATVSLNTTGGVLEISCYNDTVSD
Miller FLLNSSLTSLVSGKLVTKQPLLVNCLMVPVSFEAAASTCFCEGAGFDQCNQAVLNNVTVDVIRFNLLNTTNVQSGKGATVSLNTTGGVLEISCYNDTVSD
Purdue FLLNSSLTSLVSGKLVTKQPLLVNCLMVPVSFEAAASTCFCEGAGFDQCNQAVLNNVTVDVIRFNLLNTTNVQSGKGATVSLNTTGGVLEISCYNDTVSD
PCV FLLNSSLTSLVSGKLVTKQPLLVNCLMVPVSFEAAASTCFCEGAGFDQCNQAVLNNVTVDVIRFNLLNTTNVQSGKGATVSLNTTGGVLEISCYNDTVSD
FIPV FLLNSSLTSLVSGKLVTKQPLLVNCLMVPVSFEAAASTCFCEGAGFDQCNQAVLNNVTVDVIRFNLLNTTNVQSGKGATVSLNTTGGVLEISCYNDTVSD
* * * * *

FS772/70 SSFSYSGEIPFGVTDGPRYCVLYNGTALKYGLTLPSPVKEIAISKMGHFIYNGYFFSTFIDCISFNLTTGDSGVFWTIAITSYTEALQVQENATIK
Miller SSFSYSGEIPFGVTDGPRYCVLYNGTALKYGLTLPSPVKEIAISKMGHFIYNGYFFSTFIDCISFNLTTGDSGVFWTIAITSYTEALQVQENATIK
Purdue SSFSYSGEIPFGVTDGPRYCVLYNGTALKYGLTLPSPVKEIAISKMGHFIYNGYFFSTFIDCISFNLTTGDSGVFWTIAITSYTEALQVQENATIK
PCV SSFSYSGEIPFGVTDGPRYCVLYNGTALKYGLTLPSPVKEIAISKMGHFIYNGYFFSTFIDCISFNLTTGDSGVFWTIAITSYTEALQVQENATIK
FIPV SSFSYSGEIPFGVTDGPRYCVLYNGTALKYGLTLPSPVKEIAISKMGHFIYNGYFFSTFIDCISFNLTTGDSGVFWTIAITSYTEALQVQENATIK
* * * * *

FS772/70 VTYCNSYVNNIKCSQITANLNGGFPYVSSEVGVNKS SVLLPFTYHTIVNITIGLQKRSYGGOPIASTLSNITLPMOHNTOVYCRSDQSVYVHS
Miller VTYCNSYVNNIKCSQITANLNGGFPYVSSEVGVNKS SVLLPFTYHTIVNITIGLQKRSYGGOPIASTLSNITLPMOHNTOVYCRSDQSVYVHS
Purdue VTYCNSYVNNIKCSQITANLNGGFPYVSSEVGVNKS SVLLPFTYHTIVNITIGLQKRSYGGOPIASTLSNITLPMOHNTOVYCRSDQSVYVHS
PCV VTYCNSYVNNIKCSQITANLNGGFPYVSSEVGVNKS SVLLPFTYHTIVNITIGLQKRSYGGOPIASTLSNITLPMOHNTOVYCRSDQSVYVHS
FIPV VTYCNSYVNNIKCSQITANLNGGFPYVSSEVGVNKS SVLLPFTYHTIVNITIGLQKRSYGGOPIASTLSNITLPMOHNTOVYCRSDQSVYVHS
* * * * *

FS772/70 TCKSALMNDVFNKRCNTDVLDAVATIKGTGCFPSFDKLNLYLTNFKFCLSLSPVGANCKFDVAARTRNDQVRSLYVIEEGDNIQVGPSONSGHLDLS
Miller TCKSALMNDVFNKRCNTDVLDAVATIKGTGCFPSFDKLNLYLTNFKFCLSLSPVGANCKFDVAARTRNDQVRSLYVIEEGDNIQVGPSONSGHLDLS
Purdue TCKSALMNDVFNKRCNTDVLDAVATIKGTGCFPSFDKLNLYLTNFKFCLSLSPVGANCKFDVAARTRNDQVRSLYVIEEGDNIQVGPSONSGHLDLS
PCV TCKSALMNDVFNKRCNTDVLDAVATIKGTGCFPSFDKLNLYLTNFKFCLSLSPVGANCKFDVAARTRNDQVRSLYVIEEGDNIQVGPSONSGHLDLS
FIPV TCKSALMNDVFNKRCNTDVLDAVATIKGTGCFPSFDKLNLYLTNFKFCLSLSPVGANCKFDVAARTRNDQVRSLYVIEEGDNIQVGPSONSGHLDLS
* * * * *

FS772/70 LHLSDCTDYNIGRSGVGIIRQNTRTLSSGLYYTSLSGDLGFKNVSDGVIYSVTPCDVSAQAVIDGIVAGTISINSELLGLTHWTTNPFYYSIYN
Miller LHLSDCTDYNIGRSGVGIIRQNTRTLSSGLYYTSLSGDLGFKNVSDGVIYSVTPCDVSAQAVIDGIVAGTISINSELLGLTHWTTNPFYYSIYN
Purdue LHLSDCTDYNIGRSGVGIIRQNTRTLSSGLYYTSLSGDLGFKNVSDGVIYSVTPCDVSAQAVIDGIVAGTISINSELLGLTHWTTNPFYYSIYN
PCV LHLSDCTDYNIGRSGVGIIRQNTRTLSSGLYYTSLSGDLGFKNVSDGVIYSVTPCDVSAQAVIDGIVAGTISINSELLGLTHWTTNPFYYSIYN
FIPV LHLSDCTDYNIGRSGVGIIRQNTRTLSSGLYYTSLSGDLGFKNVSDGVIYSVTPCDVSAQAVIDGIVAGTISINSELLGLTHWTTNPFYYSIYN
* * * * *

FS772/70 YTNDRTRGTALDSDNDVCEPVIYTSNIGVCKNGALVFINVTHSDGDVQPISTGNVITPTNFTISVQVEYIQVYTPVSDCSRYVGNPNRCKNLLTQVY
Miller YTNDRTRGTALDSDNDVCEPVIYTSNIGVCKNGALVFINVTHSDGDVQPISTGNVITPTNFTISVQVEYIQVYTPVSDCSRYVGNPNRCKNLLTQVY
Purdue YTNDRTRGTALDSDNDVCEPVIYTSNIGVCKNGALVFINVTHSDGDVQPISTGNVITPTNFTISVQVEYIQVYTPVSDCSRYVGNPNRCKNLLTQVY
PCV YTNDRTRGTALDSDNDVCEPVIYTSNIGVCKNGALVFINVTHSDGDVQPISTGNVITPTNFTISVQVEYIQVYTPVSDCSRYVGNPNRCKNLLTQVY
FIPV YTNDRTRGTALDSDNDVCEPVIYTSNIGVCKNGALVFINVTHSDGDVQPISTGNVITPTNFTISVQVEYIQVYTPVSDCSRYVGNPNRCKNLLTQVY
* * * * *

FS772/70 SACQTEQALAMGARLENMVEDSMLFVSENALKLASVEAFNSSETLDP IYKEWPNIGGSMLEGLKYLILPSNHSKRYKSAIEDLFLDKVVTSGLTGVED
Miller SACQTEQALAMGARLENMVEDSMLFVSENALKLASVEAFNSSETLDP IYKEWPNIGGSMLEGLKYLILPSNHSKRYKSAIEDLFLDKVVTSGLTGVED
Purdue SACQTEQALAMGARLENMVEDSMLFVSENALKLASVEAFNSSETLDP IYKEWPNIGGSMLEGLKYLILPSNHSKRYKSAIEDLFLDKVVTSGLTGVED
PCV SACQTEQALAMGARLENMVEDSMLFVSENALKLASVEAFNSSETLDP IYKEWPNIGGSMLEGLKYLILPSNHSKRYKSAIEDLFLDKVVTSGLTGVED
FIPV SACQTEQALAMGARLENMVEDSMLFVSENALKLASVEAFNSSETLDP IYKEWPNIGGSMLEGLKYLILPSNHSKRYKSAIEDLFLDKVVTSGLTGVED
* * * * *

FS772/70 YKRCTGGYDIADLVCAQYNGIIMVLPVANADKMTMTYASLAGGITLGAIGGGGAVAIPFAVAVQARINVALQTDVLNKNQQLILASAFNQAIGNITQSGF
Miller YKRCTGGYDIADLVCAQYNGIIMVLPVANADKMTMTYASLAGGITLGAIGGGGAVAIPFAVAVQARINVALQTDVLNKNQQLILASAFNQAIGNITQSGF
Purdue YKRCTGGYDIADLVCAQYNGIIMVLPVANADKMTMTYASLAGGITLGAIGGGGAVAIPFAVAVQARINVALQTDVLNKNQQLILASAFNQAIGNITQSGF
PCV YKRCTGGYDIADLVCAQYNGIIMVLPVANADKMTMTYASLAGGITLGAIGGGGAVAIPFAVAVQARINVALQTDVLNKNQQLILASAFNQAIGNITQSGF
FIPV YKRCTGGYDIADLVCAQYNGIIMVLPVANADKMTMTYASLAGGITLGAIGGGGAVAIPFAVAVQARINVALQTDVLNKNQQLILASAFNQAIGNITQSGF
* * * * *

FS772/70 KVNDAIHQTSRGLATVAKALAKVQDVVNTQQAALSHLTQVLQNNFQATSSSIDINRKLDELISADAHVRLITGRALTALNAFVSQTLTROAEVRSRGLA
Miller KVNDAIHQTSRGLATVAKALAKVQDVVNTQQAALSHLTQVLQNNFQATSSSIDINRKLDELISADAHVRLITGRALTALNAFVSQTLTROAEVRSRGLA
Purdue KVNDAIHQTSRGLATVAKALAKVQDVVNTQQAALSHLTQVLQNNFQATSSSIDINRKLDELISADAHVRLITGRALTALNAFVSQTLTROAEVRSRGLA
PCV KVNDAIHQTSRGLATVAKALAKVQDVVNTQQAALSHLTQVLQNNFQATSSSIDINRKLDELISADAHVRLITGRALTALNAFVSQTLTROAEVRSRGLA
FIPV KVNDAIHQTSRGLATVAKALAKVQDVVNTQQAALSHLTQVLQNNFQATSSSIDINRKLDELISADAHVRLITGRALTALNAFVSQTLTROAEVRSRGLA
* * * * *

FS772/70 KDKVNECVRSQSRQFGCGNGTHLFLSANAAPNGMIFHTVLLPTAYETVTAWAGICALDGGRTGLVVKDQVLTFLRNLDDKFLYTPRTMYQPRVATSS
Miller KDKVNECVRSQSRQFGCGNGTHLFLSANAAPNGMIFHTVLLPTAYETVTAWAGICALDGGRTGLVVKDQVLTFLRNLDDKFLYTPRTMYQPRVATSS
Purdue KDKVNECVRSQSRQFGCGNGTHLFLSANAAPNGMIFHTVLLPTAYETVTAWAGICALDGGRTGLVVKDQVLTFLRNLDDKFLYTPRTMYQPRVATSS
PCV KDKVNECVRSQSRQFGCGNGTHLFLSANAAPNGMIFHTVLLPTAYETVTAWAGICALDGGRTGLVVKDQVLTFLRNLDDKFLYTPRTMYQPRVATSS
FIPV KDKVNECVRSQSRQFGCGNGTHLFLSANAAPNGMIFHTVLLPTAYETVTAWAGICALDGGRTGLVVKDQVLTFLRNLDDKFLYTPRTMYQPRVATSS
* * * * *

FS772/70 DFVQIEGCDVLFVNATVSDLPSIPDYIDINQTVQDILENFRPNMTVPETLDFINATYLNLTGEIDDLFRSEKLNHTVEALILIDININTLVNLEWL
Miller DFVQIEGCDVLFVNATVSDLPSIPDYIDINQTVQDILENFRPNMTVPETLDFINATYLNLTGEIDDLFRSEKLNHTVEALILIDININTLVNLEWL
Purdue DFVQIEGCDVLFVNATVSDLPSIPDYIDINQTVQDILENFRPNMTVPETLDFINATYLNLTGEIDDLFRSEKLNHTVEALILIDININTLVNLEWL
PCV DFVQIEGCDVLFVNATVSDLPSIPDYIDINQTVQDILENFRPNMTVPETLDFINATYLNLTGEIDDLFRSEKLNHTVEALILIDININTLVNLEWL
FIPV DFVQIEGCDVLFVNATVSDLPSIPDYIDINQTVQDILENFRPNMTVPETLDFINATYLNLTGEIDDLFRSEKLNHTVEALILIDININTLVNLEWL
* * * * *

FS772/70 NRIETVYKWPYMYLLIGLVVFCIPLLLFCCSTGCCGICGLGSCCHSICSRROFENYPIEKVHH
Miller NRIETVYKWPYMYLLIGLVVFCIPLLLFCCSTGCCGICGLGSCCHSICSRROFENYPIEKVHH
Purdue NRIETVYKWPYMYLLIGLVVFCIPLLLFCCSTGCCGICGLGSCCHSICSRROFENYPIEKVHH
PCV NRIETVYKWPYMYLLIGLVVFCIPLLLFCCSTGCCGICGLGSCCHSICSRROFENYPIEKVHH
FIPV NRIETVYKWPYMYLLIGLVVFCIPLLLFCCSTGCCGICGLGSCCHSICSRROFENYPIEKVHH
* * * * *

N-terminal region where the FIPV sequence differed from TGEV (Fig. 4) indicating that this region of the S protein may be involved in the different tropisms observed for the three viruses.

Fleming et al. (1986) indicated that the MHV S protein encoded the determinants required for binding to susceptible cells. Comparison of the S protein amino acid sequences from MHV strains, JHM (Schmidt et al., 1987) and A59 (Luytjes et al., 1987), identified an 89 amino acid deletion in the JHM sequence. Comparison of the S protein sequence from MHV-4 and several neuro-attenuated variants of MHV identified a polymorphic region with respect to deletions ranging from 142 to 159 amino acids (Parker et al., 1989). The S protein from MHV-4 had an insert of 141 amino acids when compared to MHV JHM. These observations indicate that deletions in the S protein of coronaviruses may be a natural way of altering the tropism and concurrently the pathogenicities of the viruses. The exact mechanism for the deletion events is not known, there is no evidence for repeat sequences or secondary structures allowing the jumping of the polymerase. There is evidence that some coronaviruses may undergo recombination events and this mechanism cannot be ruled out for the introduction of specific deletions within areas of the genomes.

The potential TGEV ORF-3a gene has been identified in three different strains of TGEV, Purdue-115 (Rasschaert et al., 1987), FS772/70 (Britton et al., 1989) and Miller (Wesley et al., 1989); however, the C-terminal end was found to differ between the strains. Due to the deletions in the PRCV genome a gene equivalent to TGEV ORF-3a was present as a pseudogene, but interestingly, the C-terminal end of the gene was the same as the Miller strain. An avirulent small plaque variant of the Miller strain was shown to have 462 nucleotides deleted eliminating ORF-3a (Wesley et al., 1990b). These observations indicate that ORF-3a is not needed for propagation of TGEV *in vitro* and *in vivo* but whether it plays some role in the pathogenicity of TGEV has yet to be elucidated.

Comparison of the ORF-4 genes showed three substitutions between PRCV and TGEV (FS772/70) and four different substitutions for TGEV (Purdue-115). The amino acid, residue 55 of Purdue-115, found to be deleted in the FS772/70 strain of TGEV (Britton et al., 1989), was present in PRCV. No product has been assigned to this gene for TGEV. It is interesting that the mRNA₄ species, encoding this gene, is the lowest abundance mRNA in all TGEV and PRCV strains examined to date, which might be due to the loss of the adenosine residue at the 5'-end of the putative RNA-leader binding site.

Previous comparison of the M protein sequences of the TGEV FS772/70 (Britton et al., 1988b) and two independently published sequences of Purdue-115, determined by Kapke et al. (1987) and Laude et al. (1987), revealed 96% similarity with 11 and 12 amino acid substitutions respectively. Comparison of the M protein sequences between PRCV and these TGEV strains showed 97% homology, however, some amino acid substitutions between PRCV and one strain of TGEV were not present on the other (Fig. 3), indicating that the degree of divergence of the M proteins of PRCV and TGEV was no greater than that observed between two strains of TGEV. There were 5 amino acid substitutions (residues 10, 14, 33, 44

and 198) on the PRCV sequence not found on either TGEV strain. Two of these (residues 10 and 14) were within the predicted PRCV signal sequence and two (residues 33 and 44) within the N-terminal domain of the molecule, postulated to be outside the virion envelope, and the fifth (residue 198) was within the C-terminal domain. Although one of the PRCV substitutions was within a predicted N-glycosylation site (glycine (33) in PRCV and serine in TGEV) this should not affect the addition of N-glycans to the asparagine residue at amino acid 32.

The PRCV ORF-7 gene product contained four or nine amino acid substitutions when compared to the FS772/70 and Purdue-115 sequences respectively (Fig. 3). These observations together with the comparisons of the number of amino acid substitutions between PRCV and the other TGEV genes indicated that PRCV was more closely related to the virulent strains of TGEV. The ORF-7 gene product has been detected in TGEV (Garwes et al., 1989) and PRCV (Britton et al., 1990) infected cells using antisera raised against a synthetic oligopeptide, derived from the TGEV ORF-7 sequence, indicating that this gene product has a function in viral replication. Both the TGEV and PRCV ORF-7 gene product sequences have a high similarity to the penultimate ORF in FIPV (De Groot et al., 1988) indicating that this gene may be indicative for the TGEV family of coronaviruses.

The data presented in this paper along with the data presented by Page et al. (1990), showing that the putative leader RNA sequence for TGEV and PRCV are identical, indicate that the regions of PRCV sequenced so far show good homology to TGEV. The observed differences between PRCV and TGEV genomes consisted of deletions and point mutations with no sequences unique to PRCV being identified. As these differences were no greater than between two different strains of TGEV eg Purdue-115 and FS772/70 the results suggest that PRCV is a variant of TGEV.

Acknowledgements

We would like to thank Dr D. Pocock for helpful comments on the manuscript.

References

- Bae, I., Jackwood, D.J., Benfield, D.A., Saif, L.J., Wesley, R.D. and Hill, H. (1991) Differentiation of transmissible gastroenteritis virus from porcine respiratory coronavirus and other antigenically related coronaviruses by using cDNA probes specific for the 5' region of the S glycoprotein gene. *J. Clin. Microbiol.* 29, 215-218.
- Bournsnel, M.E., Brown, T.D., Foulds, I.J., Green, P.F., Tomley, F.M. and Binns, M.M. (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* 68, 57-77.
- Bredenbeek, P.J., Pachuk, C.J., Noten, A.F.H., Charite, J., Luytjes, W., Weiss, S.R. and Spaan, W.J.M. (1990) The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59. A highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucl. Acids Res.* 18, 1825-1832.

- Britton, P. and Page, K.W. (1990) Sequence of the S gene from a virulent British field isolate of transmissible gastroenteritis virus. *Virus Res.* 18, 71–80.
- Britton, P., Garwes, D.J., Millson, G.C., Page, K., Bountiff, L., Stewart, F. and Walmsley, J. (1986) Towards a genetically-engineered vaccine against porcine transmissible gastroenteritis virus. In: Magnien, E. (Ed.), *Biomolecular Engineering in the European Community (Final Report)*; pp. 301–313. Nijhoff, The Netherlands.
- Britton, P., Garwes, D.J., Page, K. and Walmsley, J. (1987) Expression of porcine transmissible gastroenteritis virus genes in *E. coli* as β -galactosidase chimaeric proteins. In: M.M.C. Lai and S.A. Stohlman (Eds.), *Coronaviruses, Advances in Experimental Medicine and Biology*, Vol. 218, pp. 55–64. Plenum, New York, NY.
- Britton, P., Carmenes, R.S., Page, K.W., Garwes, D.J. and Parra, F. (1988a) Sequence of the nucleoprotein from a virulent British field isolate of transmissible gastroenteritis virus and its expression in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 2, 89–99.
- Britton, P., Carmenes, R.S., Page, K.W. and Garwes, D.J. (1988b) The integral membrane protein from a virulent isolate of transmissible gastroenteritis virus: molecular characterization, sequence and expression in *Escherichia coli*. *Mol. Microbiol.* 2, 497–505.
- Britton, P., Lopez Otin, C., Martin Alonso, J. and Parra, F. (1989) Sequence of the coding regions from the 3.0 kb and 3.9 kb mRNA subgenomic species from a virulent isolate of transmissible gastroenteritis virus. *Arch. Virol.* 105, 165–178.
- Britton, P., Garwes, D.J., Page, K. and Stewart, F. (1990) Molecular aspects of the relationship of transmissible gastroenteritis virus (TGEV) with porcine respiratory coronavirus (PRCV). In: D. Cavanagh and T.D.K. Brown (Eds.), *Coronaviruses and Their Diseases, Advances in Experimental Medicine and Biology*, Vol. 276, pp. 441–446. Plenum, New York, NY.
- Brown, I. and Cartwright, S. (1986) New porcine coronavirus? *Vet. Rec.* 119, 282–283.
- Callebaut, P., Correa, I., Pensaert, M., Jimenez, G. and Enjuanes, L. (1988) Antigenic differentiation between transmissible gastroenteritis virus of swine and a related porcine respiratory coronavirus. *J. Gen. Virol.* 69, 1725–1730.
- Correa, I., Jimenez, G., Sune, C., Bullido, M.J. and Enjuanes, L. (1988) Antigenic structure of the E2 glycoprotein from transmissible gastroenteritis coronavirus. *Virus Res.* 10, 77–93.
- Correa, I., Gebauer, F., Bullido, M.J., Sune, C., Baay, M.F.D., Zwaagstra, K.A., Posthumus, W.P.A., Lenstra, J.A. and Enjuanes, L. (1990) Localisation of antigenic sites of the E2 glycoprotein of transmissible gastroenteritis coronavirus. *J. Gen. Virol.* 71, 271–279.
- Cox, E., Pensaert, M.B., Callebaut, P. and Van Deun, K. (1990) Intestinal replication of a porcine respiratory coronavirus closely related antigenically to the enteric transmissible gastroenteritis virus. *Vet. Microbiol.* 23, 237–243.
- De Groot, R.J., Maduro, J., Lenstra, J.A., Horzinek, M.C., Van der Zeijst, B.A.M. and Spaan, W.J.M. (1987) cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* 68, 2639–2646.
- De Groot, R.J., Andeweg, A.C., Horzinek, M.C. and Spaan, W.J.M. (1988) Sequence analysis of the 3'-end of the feline coronavirus FIPV79–1146 genome: comparison with the genome of porcine coronavirus TGEV reveals large insertions. *Virology* 167, 370–376.
- Delmas, B., Rasschaert, D., Godet, M., Gelfi, J. and Laude, H. (1990) Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein S.J. *Gen. Virol.* 71, 1313–1323.
- Devereux, J., Haeblerli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12, 387–395.
- Fleming, J.O., Trousdale, M.D., El Zaatari, F.A., Stohlman, S.A. and Weiner, L.P. (1986) Pathogenicity of antigenic variants of murine coronavirus JHM selected with monoclonal antibodies. *J. Virol.* 58, 869–875.
- Garwes, D.J. and Pocock, D.H. (1975) The polypeptide structure of transmissible gastroenteritis virus. *J. Gen. Virol.* 29, 25–34.
- Garwes, D.J., Stewart, F., Cartwright, S.F. and Brown, I. (1988) Differentiation of porcine coronavirus from transmissible gastroenteritis virus. *Vet. Rec.* 122, 86–87.

- Garwes, D.J., Stewart, F. and Britton, P. (1989) The polypeptide of M_r 14,000 of porcine transmissible gastroenteritis virus: gene assignment and intracellular location. *J. Gen. Virol.* 70, 2495–2499.
- Güssow, D. and Clackson, T. (1989) Direct clone characterization from plaques and colonies by the polymerase chain reaction. *Nucl. Acids Res.* 17, 4000.
- Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237–244.
- Jacobs, L., Van der Zeijst, B.A.M. and Horzinek, M.C. (1986) Characterization and translation of transmissible gastroenteritis virus mRNAs. *J. Virol.* 57, 1010–1015.
- Jacobs, L., De Groot, R., Van der Zeijst, B.A.M., Horzinek, M.C. and Spaan, W. (1987) The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV). *Virus Res.* 8, 363–371.
- Kapke, P.A. and Brian, D.A. (1986) Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* 151, 41–49.
- Kapke, P.A., Tung, F.Y.C., Brian, D.A., Woods, R.D. and Wesley, R. (1987) Nucleotide sequence of the porcine transmissible gastroenteritis coronavirus matrix protein. In: M.M.C. Lai and S.A. Stohman (Eds.), *Coronaviruses, Advances in Experimental Medicine and Biology*, Vol. 218, pp. 117–122. Plenum, New York, NY.
- Laude, H., Rasschaert, D. and Huet, J.-C. (1987) Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68, 1687–1693.
- Luytjes, W., Sturman, L.S., Bredenbeek, P.J., Charite, J., Van der Zeijst, B.A., Horzinek, M.C. and Spaan, W.J. (1987) Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site. *Virology* 161, 479–487.
- O'Toole, D., Brown, I., Bridges, A. and Cartwright, S.F. (1989) Pathogenicity of experimental infection with "pneumotropic" porcine coronavirus. *Res. Vet. Sci.* 47, 23–29.
- Page, K.W., Britton, P. and Boursnell, M.E.G. (1990) Sequence analysis of the leader RNA of two porcine coronaviruses: transmissible gastroenteritis virus and porcine respiratory coronavirus. *Virus Genes* 4, 289–301.
- Page, K.W., Mawditt, K.L. and Britton, P. (1991) Sequence comparison of the 5'-end of mRNA 3 from transmissible gastroenteritis virus and porcine respiratory coronavirus. *J. Gen. Virol.* 72, 579–587.
- Parker, S.E., Gallagher, T.M. and Buchmeier, M.J. (1989) Sequence analysis reveals extensive polymorphism and evidence of deletions within the E2 glycoprotein gene of several strains of murine hepatitis virus. *Virology* 173, 664–673.
- Pensaert, M.B. (1989) Transmissible gastroenteritis virus (respiratory variant). In: M.C. Horzinek and M.B. Pensaert (Eds.), *Virus Infections of Vertebrates*, Vol. 2, pp. 154–165. Elsevier, Amsterdam.
- Pensaert, M.B., Callebaut, P.E. and Vergote, J. (1986) Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Vet. Q.* 8, 257–260.
- Posthumus, W.P.A., Lenstra, J.A., Schaaper, W.M.M., Van Nieuwstadt, A.P., Enjuanes, L. and Melen, R.H. (1990) Analysis and simulation of a neutralizing epitope of transmissible gastroenteritis virus. *J. Virol.* 64, 3304–3309.
- Rasschaert, D. and Laude, H. (1987) The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68, 1883–1890.
- Rasschaert, D., Gelfi, J. and Laude, H. (1987) Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. *Biochimie* 69, 591–600.
- Rasschaert, D., Duarte, M. and Laude, H. (1990) Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. *J. Gen. Virol.* 71, 2599–2607.
- Sánchez, C.M., Jimenez, G., Laviada, M.D., Correa, I., Sune, C., Bullido, M.J., Gebauer, F., Smerdou, C., Callebaut, P., Escribano, J.M. and Enjuanes, L. (1990) Antigenic homology among coronaviruses related to transmissible gastroenteritis virus. *Virology* 174, 410–417.
- Schmidt, I., Skinner, M. and Siddell, S. (1987) Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *J. Gen. Virol.* 68, 47–56.
- Staden, R. (1982) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucl. Acids Res.* 10, 4731–4751.

- Van Nieuwstadt, A.P. and Pol, J.M.A. (1989) Isolation of a TGE virus-related respiratory coronavirus causing fatal pneumonia in pigs. *Vet. Rec.* 124, 43–44.
- Vennema, H., De Groot, R.J., Harbour, D.A., Horzinek, M.C. and Spaan, W.J.M. (1991) Primary structure of the membrane and nucleocapsid protein genes of feline infectious peritonitis virus and immunogenicity of recombinant vaccinia viruses in kittens. *Virology* 181, 327–335.
- Von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* 14, 4683–4690.
- Wesley, R.D. (1990) Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequence of the upstream polymerase gene of transmissible gastroenteritis virus (Miller strain). In: D. Cavanagh and T.D.K. Brown (Eds.), *Coronaviruses and Their Diseases, Advances in Experimental Medicine and Biology*, Vol. 276, pp. 301–306. Plenum, New York, NY.
- Wesley, R.D., Cheung, A.K., Michael, D.D. and Woods, R.D. (1989) Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNA species between the peplomer and matrix protein genes. *Virus Res.* 13, 87–100.
- Wesley, R.D., Woods, R.D., Hill, H.T. and Biwer, J.D. (1990a) Evidence for a porcine respiratory coronavirus, antigenically similar to transmissible gastroenteritis virus, in the United States. *J. Vet. Diagn. Invest.* 2, 312–317.
- Wesley, R.D., Woods, R.D. and Cheung, A.K. (1990b) Genetic basis for the pathogenesis of transmissible gastroenteritis virus. *J. Virol.* 64, 4761–4766.