Identification of major differences in the nucleocapsid protein genes of a Québec strain and European strains of porcine reproductive and respiratory syndrome virus

Helmi Mardassi, Samir Mounir and Serge Dea*

Centre de recherche en virologie, Institut Armand-Frappier, 531 Boulevard des Prairies, Université du Québec, Laval, Québec, Canada H7N 4Z3

The sequence of the 3'-terminal region of the genome of Québec reference strain IAF-exp91 of porcine reproductive and respiratory syndrome virus (PRRSV) was investigated by analysis of four cDNA clones. The 3'-terminal 530 nucleotides (nt) encompassed a large open reading frame with a coding capacity of 123 amino acids (34, 13, 649). The predicted protein was extremely basic and hence was considered to correspond to the nucleocapsid (N) protein gene. When compared to the homologous sequences of two reference Netherlands strains (Lelystad and isolate 10) of PRRSV, the IAF-exp91 N protein was found to be five amino acids shorter and displayed a high degree of divergence. Overall, IAF-exp91 strain showed identities of 63% and 59% with both reference European strains at the nucleotide and amino acid level, respectively. Two amino acid stretches, STAPM and SQGAS, present respectively at the N- and C-terminal regions of the N protein of European strains, were missing in the IAF-exp91 N protein sequence. The 3'-terminal non-coding region (151 nt) of the IAF-exp91 strain was 22 nt longer than that of the European strains. The aligned nucleotide sequence of this non-coding region exhibited an overall identity of 59% with that of the European strains. The Québec reference strain of PRRSV appeared to be related more closely to equine arteritis virus and lactate dehydrogenase-elevating virus than are the two European strains of the virus. Preliminary data obtained by reverse transcription-PCR experiments, using specific or common oligonucleotide primers, suggested that this approach could be useful for distinguishing between PRRSV strains from different geographic origins.

Porcine reproductive and respiratory syndrome (PRRS) was first described in the United States (Poison et al., 1990) and Canada (Bilodeau et al., 1991) in 1986 to 1987. The disease was initially characterized by severe reproductive failure in sows of any parities (late-term abortions, increased numbers of stillborn, mummified and weakborn pigs, increased preweaning mortality), and respiratory problems affecting pigs of all ages, but mainly unweaned piglets. In 1990, a similar disease was observed in Germany (Lindhaus & Lindhaus, 1991) and rapidly swept through Western European countries in 1991 (Baron et al., 1992; Wensvoort et al., 1992b).

The aetiological agent of the syndrome (PRRSV) has been definitively identified as a small spherical enveloped virus, 50 to 65 nm in diameter, with a central isometric nucleocapsid of approximately 25 to 30 nm (Benfield et al., 1992; Wensvoort et al., 1992b). The viral genome is a positive-stranded polyadenylated RNA of about 15 kb, which generates in infected cells a 3'-coterinal nested set of six subgenomic mRNAs (Meulenberg et al., 1993; Conzelmann et al., 1993). The genomic RNA contains at least eight open reading frames (ORFs) organized similarly to those of equine arteritis virus (EAV) and lactate dehydrogenase-elevating virus (LDV) genomic RNAs. The latter two viruses have been proposed as members of the genus Arterivirus, family Togaviridae (Plagemann & Moennig, 1992; Conzelmann et al., 1993). European isolates of PRRSV appear to belong to the same serotype, whereas antigenic variability has been demonstrated among American isolates, and between American and European isolates (Wensvoort et al., 1992a). Here we report the sequence analysis of the 3'-terminal 530 nucleotides (nt) of the reference Québec strain IAF-exp91 of PRRSV (Dea et al., 1992).

IAF-exp91 was propagated in primary cultures of porcine alveolar macrophages (PAM) prepared as described by Wensvoort et al. (1991). After complete degeneration of the monolayers, supernatant fluids were clarified by centrifugation at 5000 g for 20 min, followed by ultracentrifugation through a cushion of 30% sucrose...

The nucleotide sequence data reported here will appear in the EMBL and GenBank nucleotide sequence databases under accession number U02095 (N-PRRS-IA).
OTYRAFY ATG CCA ANT ACC AAC GGC AGA CAG CAA AGG AAA GAG GAT 52
MPNQNHGQRKKGDF 15
GCC AGA CCA GTC ANR CAG CTD TCT TCG ANR GCT GAA YRC YRC CAC CAG CAC GTC
QC PRV RLOCLWMLNGKFAQ 32
CA6 ANC GNF TUC ACA GTG AGA GCA GDD AGG AAA ANR ANR AAA AGC 124
QDM SROGSPOKENKRD 49
CCG GAA ANR AGC CAT TTT CCT CTA GAA ACT GAC GAT GTC CAT CAC GAC CAT 205
PSPPEPLATEDDVKKR 56
TVC AAC GCT ANR CAG CAA CCC TGG TGT CTO TCN ATC ANC GCA GCT TTT 256
PPTCEBOLLLSSTIQTAF 73
AAN NTG GCT GGA ACT TGC GCT CAT GAA ATG AGA NTG ANR TAC 307
NQGAGCTCTLSDSQSYS 90
OUT GGT GNO TTT AAT TTG CCT GTG ATC CAT ANC GSP GTC GSP CTT ATC GAC 358
A V R P F PA T B B D V A R T 117
ACA GCA TCA CCC GCA GCA GTT CGC AGA CAC CAC 408
Q Q R P N 123
AAAAAAA@U@A AAAAAAAA 130
ACGAACATTGTCGAAATTGCATCACACACGCCATGCATAGCTCGAGGGC 485
OYUGGGGGTMCXKSWKYTPGQ0GQAQQCATGGCGGGGAATATTTTTTTTTTTTTTTTTTTTT T 530

Fig. 1. Nucleotide sequence of the 3'-terminal 530 nt of the IAF-exp91 strain of PRRSV, including the N protein gene. The predicted amino acid sequence is shown below the nucleotide sequence. The asterisk indicates the termination codon. The poly(A) tail of 34 nt was obtained from one of four cDNA clones analysed.

(w/v) for 3 h at 100000 g. The viral pellets were resuspended in TNE buffer (50 mm-Tris–HCl pH 8.0, 150 mm-NaCl and 1 mm-EDTA) and viral genomic RNA was isolated according to the method of Chomczynski & Sacchi (1987). To obtain the 3' end of the viral genome, a cDNA library was constructed according to Gubler & Hoffman (1983), using oligo(dT)12–18 as a primer in the first-strand synthesis reaction, and the second-strand mix was obtained from the Gibco BRL cDNA synthesis kit with modifications. Phenol–chloroform extractions were omitted and replaced by Sephaglas BandPrep purification of DNA (Pharmacia) which was also used to purify cDNA for ligation into the pSPORT1 vector instead of column size fractionation. The cDNA was selected according to size by preparative agarose gel electrophoresis. Sequencing of four randomly selected cDNA clones was performed on both strands by the dideoxynucleotide chain termination method (Sanger et al., 1977) using T7 DNA polymerase (Pharmacia) in an Automated Laser Fluorescent DNA sequence analyser (Pharmacia LKB). No mismatched bases, additions or deletions were found between the four cDNA clones. Sequence analyses were performed on an Apple Macintosh computer with the MacVector 3.5 (International Biotechnologies) and GeneWorks 2.2 (IntelliGenetics) sequence analysis programs.

The nucleotide sequence of the 3'-terminal 530 nt of the IAF-exp91 strain of PRRSV [excluding the poly(A) tail] and its predicted amino acid sequence are shown in Fig. 1. This genomic region encompassed a large ORF (nt 8 to 379) encoding a polypeptide of 123 amino acids with a predicted M, of 13649, consistent with the estimated M, of the nucleocapsid (N) protein determined previously by SDS–PAGE analyses (Mardassi et al., 1994). As previously described for Lelystad virus (LV) and isolate 10 of PRRSV, two Netherlands reference strains (Meulenberg et al., 1993; Conzelmann et al., 1993), the N protein of the IAF-exp91 strain was extremely basic; 26% of the N-terminal half of the polypeptide consisted of Arg, Lys and His residues. As shown in Fig. 2, the N protein of IAF-exp91 was compared with that of the LV strain, with the homologous regions exhibiting only 63% and 59% identity at the nucleotide and amino acid levels, respectively. Such relatively high divergence resulted from a number of nucleotide substitutions, insertions or deletions (data not shown), making the IAF-exp91 N protein five amino acids shorter than that of the LV strain. The nucleotide substitutions appeared randomly distributed, but were more frequent in the first half of the N gene. Four amino acid stretches, located at amino acid positions 16 to 24, 50 to 58, 78 to 89, and 110 to 118 in the IAF-exp91 sequence, were shared by the Québec and the reference European strains (Fig. 2). The unique potential N-glycosylation site observed in the N protein of both European strains (Meulenberg et al., 1993; Conzelmann et al., 1993) was absent from the IAF-exp91 strain. Two amino acid stretches, STAPM and SQGAS, situated respectively at the N-terminal and C-terminal regions of the N protein of both European isolates were missing from the IAF-exp91 N protein (Fig. 2). The amino acid identity between the IAF-exp91 and the LV strains increases to 70% when only the sequences located between these two stretches were compared. Interestingly, these two amino acid stretches have not been identified in the amino acid sequences of the N proteins of LDV and EAV for which the amino acid identities with the IAF-exp91 strain were estimated at 49% and 23%, respectively. Consequently, the N protein of the IAF-exp91 isolate appeared to be more closely related to those of LDV and EAV than to those of the two European strains (Meulenberg et al., 1993; Conzelmann et al., 1993).

The IAF-exp91 non-coding region downstream of the stop codon of the N gene was 151 nt in length, thus exceeding the European strains' non-coding region by 37 nt. Twenty-two of these nucleotides were absent from the first half of the homologous non-coding region of the LV strain (Fig. 3). An overall identity of 59% was found by comparing the non-coding region of the IAF-exp91 strain to that of the LV strain. The consensus sequence C-C-G-G/A-A-A-T-T-poly(A) at the 3' end of LDV and PRRSV isolate 10 genomes (Conzelmann et al., 1993) is also present in the genome of IAF-exp91. Despite the high genomic variability observed in the non-coding regions of the Québec and European PRRSV strains, the presence of this conserved sequence suggests that it may have functional significance.

To establish further the differences between PRRSV strains from different geographic origins, reverse tran-
Table 1. **RT-PCR** oligonucleotide primers deduced from the sequence of the IAF-exp91 genome

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Position</th>
<th>Product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1008PS</td>
<td>5' TAAATATGCCAATAAACACAC 3'</td>
<td>3-22</td>
<td>468</td>
</tr>
<tr>
<td>1009PR</td>
<td>5' TAOGTGACCTAQAAGGCAAC 3'</td>
<td>450-470</td>
<td></td>
</tr>
<tr>
<td>1010PLS</td>
<td>5' ATGCCAGCAGCTCACAATCA 3'</td>
<td>51-69</td>
<td>433</td>
</tr>
<tr>
<td>1011PLR</td>
<td>5' TGGCCCTATTAGATAGG 3'</td>
<td>464-483</td>
<td></td>
</tr>
</tbody>
</table>
Taken together, results from RT-PCR experiments previously identified by nucleotide sequence analysis. The absence of 37 nt at the Y end of the LV genome further confirmed the existence of relatively high genomic diversity among the N genes of the European and Quebec reference strains of PRRSV.

Using sera from convalescent pigs in indirect immunofluorescence or immunoperoxidase tests, it has been shown that antigenic variations exist between European and North American isolates of PRRSV (Wensvoort et al., 1992a). These findings are in part supported by the variability observed at the level of the N protein, as demonstrated in the present study, between the prototype Quebec strain IAF-exp91, and two Netherlands strains of PRRSV (Fig. 2). This antigenic variability between North American and European isolates of PRRSV has been recently confirmed in our laboratory using monoclonal antibodies EP147, VO17 and SDOW17 directed against the N protein of the prototype American strain ATCC-VR2332. Indeed, two of these (EP147 and VO17) failed to react with the homologous protein of the LV strain in radioimmunoprecipitation experiments, but they all strongly immunoprecipitated the N protein of the IAF-exp91 strain (Mardassi et al., 1994). These results suggested that our reference strain is closely related to the American strain, thus confirming previous serological studies using sera from experimentally infected pigs, as well as sera collected from different pig farms in Quebec that have experienced typical outbreaks of PRRS (Dea et al., 1992). Despite its internal location in the virion, the N protein of PRRSV undergoes a high degree of antigenic variability compared to those of other enveloped positive-stranded RNA viruses such as coronaviruses (Lapps et al., 1987; Williams et al., 1992). The extent of genomic variation observed between the Canadian and European strains of PRRSV confirms the need for further investigation of genomic heterogeneity among strains of this virus before techniques that rely upon nucleic acid hybridization can be applied effectively as diagnostic procedures. Interestingly, results obtained in RT-PCR experiments, using either specific or common oligonucleotide primers, suggested that this approach could be useful for distinguishing between PRRSV strains from different geographical origins.

119R10 primer with its corresponding sequence in IAF-exp91 cDNA revealed a three base mismatch at its 3' end which could be an explanation for the inability to amplify such a region in the IAF-exp91 strain. In contrast, when primers 1008PS and 1009PR which represent the IAF-exp91 sequence in a region displaying differences from the LV strain were chosen, amplification of the N gene was possible for the Quebec strain, but not for the LV strain (Fig. 5, lanes 1 and 2). As expected, the use of oligonucleotide primers 1010PLS and 1011PLR which display total complementarity with sequences of the LV strain (common primer) resulted in the amplification of both the Quebec and the LV strains of PRRSV (Fig. 5, lanes 4 and 5). The genomic fragment generated by these common oligonucleotide primers migrated faster for LV RNA than that amplified from the IAF-exp91 strain. These results are thus in agreement with the absence of 37 nt at the 3' end of the LV genome previously identified by nucleotide sequence analysis.

Taken together, results from RT-PCR experiments further confirmed the existence of relatively high genomic diversity among the N genes of the European and Quebec reference strains of PRRSV.

This report was taken in part from a dissertation to be submitted by H.M. to the Center for Research in Virology, Institut Armand-Frappier, University of Quebec, in partial fulfillment of the requirements for the Ph.D. degree. We thank Ms L. Wilson and J. Roger for technical assistance. The collaboration of Drs D.A. Benfield and E. Nelson, South Dakota State University, in providing the monoclonal antibodies to the ATCC-VR2332 strain of PRRSV was greatly appreciated. This study was partly supported by the Conseil de Recherches en Pêche et Agro-Alimentaire du Québec, la Fédération des Producteurs de porcs du Québec, and Vetpharm Research Inc., London, Ontario.

References


(Received 16 July 1993; Accepted 8 October 1993)