

**Determination of 5'-leader sequences from radically disparate strains of porcine reproductive and respiratory syndrome virus reveals the presence of highly conserved sequence motifs**

Brief Report

**M. B. Oleksiewicz, A. Bøtner, J. Nielsen, and T. Storgaard**

Danish Veterinary Institute for Virus Research, Lindholm, Kalvehave, Denmark

Accepted January 8, 1999

**Summary.** We determined the untranslated 5'-leader sequence for three different isolates of porcine reproductive and respiratory syndrome virus (PRRSV): pathogenic European- and American-types, as well as an American-type vaccine strain. 5'-leader from European- and American-type PRRSV differed in length (220 and 190 nt, respectively), and exhibited only approximately 50% nucleotide homology. Nevertheless, highly conserved areas were identified in the leader of all 3 PRRSV isolates, which constitute candidate motifs for binding of protein(s) involved in viral replication. These comparative data provide *a priori* knowledge for mutational identification of virulence determinants in the 5' nontranslated part of the PRRSV genome.

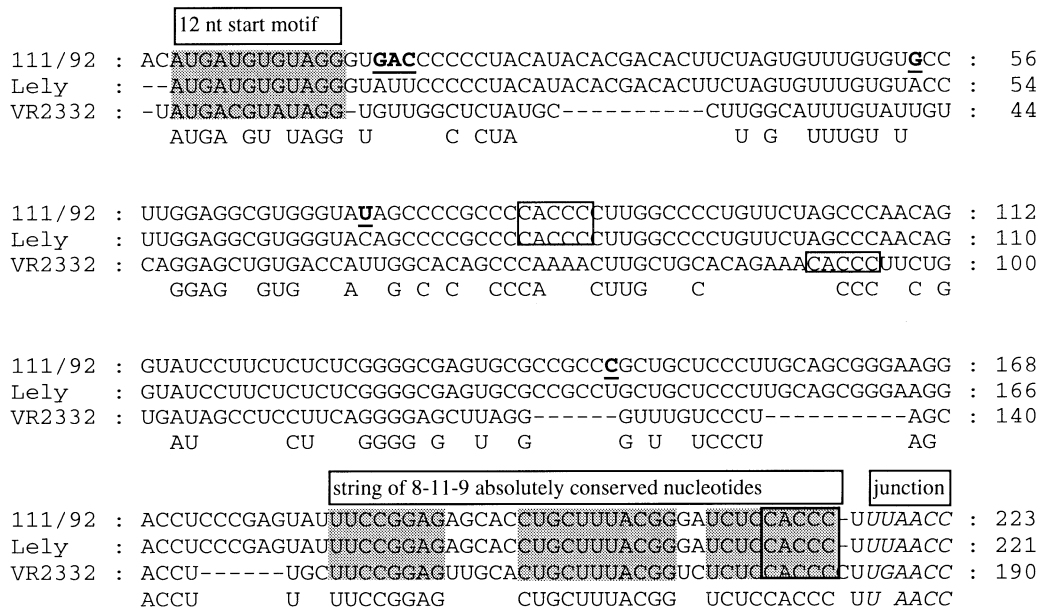
\*

Porcine reproductive and respiratory syndrome virus (PRRSV) is a recently emerged pathogen, which belongs to the *Arteriviridae* family of enveloped, plus-sense RNA viruses. PRRSV was first isolated in Western Europe, and shortly thereafter in North America (for a recent review see [20]). Despite the concurrent emergence of PRRSV-induced disease on the two continents, transatlantic PRRSV strains proved surprisingly divergent, with only 55–80% nucleotide homology in the structural genes [12]. In Denmark, the epidemic spread of indigenous, European-type PRRSV was attempted halted in 1996 by a large-scale vaccination campaign, using a “classical” cell culture-attenuated live vaccine, based on the pathogenic North American VR2332 isolate. Unfortunately, this live vaccine strain proved unexpectedly adept at reverting to virulence and spreading to non-vaccinated herds [4,10,17,21]. On a speculative note, many of the questions arising from the coexistence of European and American PRRSV

types in Denmark may impinge on a single viral genetic element: the PRRSV untranslated 5'-leader. For the closely related coronaviruses, the 5'-leader regulates viral genomic RNA replication, transcription and mRNA translation [9, 22, 23]. Furthermore, the 5'-leaders appears to be a highly mobile genetic element, able of high-efficiency recombination between related viruses [11], and 5'-leaders of other RNA viruses such as poliovirus and Venezuelan Equine Encephalitis virus contain attenuation/virulence determinants [5,8]. In short, examination of the PRRSV 5'-leader sequence was expected to provide basic knowledge about the molecular basis of PRRSV replication/virulence/attenuation, as well as improve our means for understanding and monitoring any future exchange of genetic material between coexisting PRRSV types.

*Determination of 111/92 and VR2332 leaders by 5' RACE:* The European-type Lelystad virus is the only PRRSV strain for which the 5'-leader sequence has been published [15]. We determined the 5'-leader sequences for 111/92, a pathogenic European-type PRRSV isolate from 1992 [3,10], and VR2332, a pathogenic American-type isolate [2], using the 5' RACE (rapid amplification of cDNA ends) kit from Life Technologies (Roskilde, Denmark). 5' RACE was carried out essentially according to the manufacturer's instructions. Cycle sequencing was done using fluorescent chain terminator chemistry (Perkin Elmer, Allerød, Denmark). A detailed description of the experimental protocol and primer sequences is available on request. The obtained 111/92 leader sequence was 96% similar to the Lelystad 5'-leader (Fig. 1). The VR2332 5'-leader was 30 nt shorter, and had only approximately 50% nucleotide homology to the European-type leaders. The GC% also differed slightly between European (59–61%) and American (52%) viruses. The 111/92 and VR2332 5'-leader sequences shown in Fig. 1 have been submitted to GenBank, and given the accession numbers AF094476 and AF094475, respectively.

*Identification of putative regulatory leader motifs:* The 5'-leader is thought to act as a landing pad for viral and cellular proteins involved in PRRSV replication. Yet, identification of regulatory motifs in arteriviral leaders has so far proven elusive [6]. We observed that American and European-type-PRRSV were apparently evolutionary sufficiently distant, yet adequately close, for highly conserved motifs to become apparent in the 5'-leader sequences (Fig. 1). A 12 nucleotide stretch containing a purine (A/G) and a pyrimidine (C/U) wobble position appeared to form a conserved "5'-leader start" motif (Fig. 1, shaded). Additionally, the 40 most 3'-terminal bases of the leader exhibited a high degree of conservation between European- and American-type PRRSV, and in this conserved area, a string of 8, 9 and 11 wholly invariable nucleotides appeared to be embedded (Fig. 1, shaded areas). Whereas these 5'-leader motifs are candidates for mediating the RNA-protein interactions thought important for regulating PRRSV replication, identification of protein partners, if any, will require further studies. Yet, we observed that the 9 nt motif contained a CACCC site, with another CACCC site occurring approximately 100 nt further upstream in American as well as Euro-



**Fig. 1.** Alignment of 5'-leaders of 111/92, Lelystad virus, and VR2332. 5'-leaders of 111/92 and VR2332 were determined by 5' RACE, and aligned to the Lelystad 5'-leader. Insertions/deletions are indicated by dashes. The bottom line shows completely conserved nucleotides. The 6 nucleotides that differ between 111/92 and Lelystad leaders are shown with underlined, bold lettering in the 111/92 sequence. Longer runs of nucleotides that are highly conserved between European and American PRRSV are shaded. CACCC motifs are boxed. The 6 nt junction sequence at the 3' end of the leader is italicized. The shown junction sequences are derived from the ORF 7 mRNA (111/92), and ORF 2 mRNA (VR2332), which were the subgenomic viral mRNA's targeted by 5' RACE. For Lelystad virus, the genomic junction sequence is shown. See text for details

pean leaders (Fig. 1, boxed nucleotides). CACCC motifs have been implicated in regulating Epstein-Barr virus and human T cell leukaemia virus replication through interaction with Sp1 and Krüppel-like transcription factors [7, 18], but their importance for non-retroviral RNA viruses is not known.

*Screening of 5'-leader for attenuation determinants:* The pathogenic North American VR2332 isolate underwent multiple cell culture passages to produce a live, attenuated PRRSV vaccine strain [10]. We determined the 5'-leader sequence of the PRRSV vaccine strain using 5' RACE. The vaccine 5'-leader sequence was completely identical to the VR2332 5'-leader sequence, except that the U preceding the 12 nt leader start motif in the VR2332 leader (Fig. 1) was absent in the vaccine leader (the vaccine leader not included in Fig. 1). However, the 5' RACE results for 111/92 showed substantial variability in the first 1–2 nucleotides preceding the conserved 12 nt leader start motif (not shown). Therefore, the significance of a single nucleotide difference between VR2332 and the vaccine in this variable position *vis-a-vis* attenuation can unfortunately not be decided from

our data. The 189/190 nucleotide identity between vaccine and VR2332 5'-leaders may indicate non-involvement of the 5'-leader in attenuation. Alternatively, the very high degree (99.7%, [10]) of conservation between the vaccine and VR2332 in the structural genes may in turn imply that attenuation conditions were not sufficiently harsh to affect the 5'-leader sequence.

*PCR amplification of subgenomic mRNAs:* Free 5'-leader has been hypothesized to prime subgenomic viral mRNA transcription [1]. 5'-leader priming of subgenomic viral mRNA transcription is thought to require interaction between a 6 nt long junction sequence found at the 3'-end of the free leader, and partly homologous junction sites upstream of the various open reading frames in the viral genome. Thus, examining the junction sites in subgenomic viral mRNAs is integral to understanding the possible function of the 5'-leader in regulating of viral transcription. To determine 5'-leader-mRNA body junctions, total intracellular RNA from PRRSV-infected cultures was reverse transcribed using random hexamers and the "Ready-To-Go" reverse transcription system (Pharmacia, Allerød, Denmark). cDNA was PCR-amplified using standard protocols. The forward PCR primers were specific for the extreme 5'-end of the leader sequences determined in this study (Fig. 1), and reverse PCR primers were situated 50–140 nt downstream of the start codon for each mRNA. Primer sequences are available on request. In virtually all cases, a single, predominant PCR product was observed, confirming previous reports that the PRRSV 5'-leader-mRNA junction mechanism is quite site-specific [16]. A highly notable exception was the ORF 7 (nucleocapsid) mRNA of VR2332, where the PCR band pattern indicated equal utilization of more than one junction site. Similar findings for VR2332 have recently been reported by others (see [20] and references therein). The sequence data are summarized in Table 1. For all subgenomic mRNAs of 111/92, the junction sequences as well as the length of untranslated sequence between the 5'-leader and the start codon of the downstream open reading frame (Table 1) corresponded to the published values for Lelystad virus [16]. Thus, the 6 nt differences in the 5'-leaders (Fig. 1), as well as the 7% nucleotide differences in the structural genes [10], had no influence on the junction-choice of these two European-type viruses. Also, the published lengths of untranslated sequence between the 5'-leader and downstream ORF in mRNAs 3 and 4 of different North American PRRSV isolates [14], and mRNAs 5, 6 and 7 of a Japanese PRRSV isolate [19], were the same as the values determined for VR2332 in this study. Thus, for North American PRRSV types, the junction-choice also appears to be well conserved, despite relatively high (up to 11%) nucleotide differences in the structural genes [13, 14]. The VR2332 junctions determined in this study for mRNAs 5, 6 and 7 differed from those determined in another study for a Japanese, American-type PRRSV isolate [19]. However, the VR2332 junctions determined in this study for mRNAs 3 and 4 were identical to those determined previously for different North American isolates [14]. These findings may reflect genetic differences between Japanese and North American PRRSV type [19].

**Table 1.** Organization of mRNA leader-body junctions in 111/92 and VR2332

mRNA number	111/92	VR2332
<b>2</b>	> <b>39</b>	> <b>21</b>
mRNA sequence>	<b>CUCCACCCG</b> * <u>UAAACC</u> * <u>CCGGCU</u>	<b>UCCACCCCU</b> * <u>UGAACC</u> * <u>AACUUU</u>
genomic sequence>	AGAGCUAGG* <u>UAAACC</u> * <u>CCGGCU</u>	CCUGUCAU* <u>UGAACC</u> * <u>AACUUU</u>
<b>3</b>	> <b>12</b>	> <b>84</b>
mRNA sequence>	<b>CUCCACCCU</b> * <u>UUGACC</u> * <u>GAUUUC</u>	<b>UCCACCCCU</b> * <u>GUAACC</u> * <u>AUAGUG</u>
genomic sequence>	AGGCCCAAG* <u>UUGACC</u> * <u>GAUUUC</u>	GGGUCAAU* <u>GUAACC</u> * <u>AUAGUG</u>
<b>4</b>	> <b>84</b>	> <b>6</b>
mRNA sequence>	<b>CUCCACCCU</b> * <u>UCAACC</u> * <u>GUAUCC</u>	<b>UCCACCCCU</b> * <u>UUCACC</u> * <u>UAGAAU</u>
genomic sequence>	GGACAGAAC* <u>UCAACC</u> * <u>GUAUCC</u>	GCAAUUGGU* <u>UUCACC</u> * <u>UAGAAU</u>
<b>5</b>	> <b>33</b>	> <b>41</b>
mRNA sequence>	<b>CUCCACCCU</b> * <u>UUAACC</u> * <u>AUCGCU</u>	<b>UCCACCCCU</b> * <u>UUAGCC</u> * <u>UGUCUU</u>
genomic sequence>	GUGGGCCAC* <u>ACAACC</u> * <u>AUCGCU</u>	GCAACUGUU* <u>UUAGCC</u> * <u>UGUCUU</u>
<b>6</b>	> <b>25</b>	> <b>17</b>
mRNA sequence>	<b>CUCCACCCY</b> * <u>UCAACC</u> * <u>CUUGAC</u>	<b>UCCACCCCU</b> * <u>AUAACC</u> * <u>AGAGUU</u>
genomic sequence>	GGUUAAGC* <u>UCAACC</u> * <u>CUUGAC</u>	GCAACCCCU* <u>AUAACC</u> * <u>AGAGUU</u>
<b>7</b>	> <b>10</b>	> <b>124</b>
mRNA sequence>	<b>CUCCACCCU</b> * <u>UUAACC</u> * <u>UCGUCA</u>	<b>UCCACCCCU</b> * <u>UUAACC</u> * <u>ACGCAU</u>
genomic sequence>	GAGGAGUGG* <u>UUAACC</u> * <u>UCGUCA</u>	CGGCAAUG* <u>AUAACC</u> * <u>ACGCAU</u>
<b>7'</b>	> <b>10</b>	> <b>10</b>
mRNA sequence>	Not applicable	<b>UCCACCCCK</b> * <u>UAAACC</u> * <u>UUGUCA</u>
genomic sequence>		AGGGAGUGG* <u>UAAACC</u> * <u>UUGUCA</u>

The 5' leader-mRNA junction sequence of the complete set of 6 subgenomic mRNAs (2–7) was determined for both PRRSV strains, as described in the text. The mRNA sequence immediately surrounding the junction between 5'-leader and mRNA body is shown aligned to the published genomic sequence. The 6 nt junction is sandwiched between\*, and the junction in the mRNA sequence is also italicized. The part of the mRNA sequence identical to the genomic sequence is underlined. The distance from the last nt of the junction to the first nt of the start codon is indicated following the > above each mRNA sequence. The nucleotides shown in bold lettering at the 5'-end of the mRNA sequence are part of the 9 nt conserved motif in the 5'-leader, see Fig 1. Y in mRNA 6 of 111/92 and K in mRNA 7 of VR2332 indicate C/U and U/G ambiguities, respectively, due to clonal sequence variation in the mRNA population [16]. The genomic site 5 for 111/92 (EMBL accession number AJ223078) and genomic site 2 for VR2332 (GenBank accession number U87392) were re-sequenced for the present study. For 111/92, no sequence information exists for genomic site 2, and published sequence from Lelystad virus (GenBank accession M96262) was substituted. mRNA 7' is a prevalent variant of VR2332 mRNA 7

In summary, we determined the 5'-leader sequences for three PRRSV isolates. This data provided the first insights into the degree of variability in this important, regulatory part of the PRRSV genome. Additionally, the new sequences allowed, for the first time, prediction of putative regulatory elements in the PRRSV 5'-leader, which may be an aid in developing second-generation, genetically engineered PRRSV vaccines. Finally, scrutiny of subgenomic mRNAs identified intriguing differences in the junction-choices for the ORF 7 (nucleocapsid) mRNA of European- and American-type PRRSV, which are currently under further investigation.

### Acknowledgements

P. Normann and F. Lihme are thanked for technical assistance. S. Alexandersen is thanked for helpful discussions. This work was partly supported by a grant from the Norma and Frode Jacobsen foundation, and grant 9600926 to TS from the Danish Agricultural and Veterinary Research council.

### Note added in proof

Since the submission of this manuscript, a study containing a partial 5'-leader sequence of VR2332 has appeared in print [Nelsen CJ et al. (1999) *J Virol* 73: 270–280]. The 20, 5'-terminal nucleotides missing in Nelsen et al.'s VR 2332 sequence are contained in our 5'-leader sequence.

### References

1. Baker SC, Lai MC (1990) An in vitro system for the leader-primed transcription of coronavirus mRNAs. *EMBO J* 9: 4 173–4 179
2. Benfield DA, Nelson E, Collings JE, Harris L, Goyal SM, Robison D, Christianson WT, Morrison RB, Gorcyca D (1992) Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). *J Vet Diagn Invest* 4: 127–133
3. Bøtner A, Nielsen J, Bille-Hansen V (1994) Isolation of porcine reproductive and respiratory syndrome (PRRS) virus in a Danish swine herd and experimental infection of pregnant gilts with the virus. *Vet Microbiol* 40: 351–360
4. Bøtner A, Strandbygaard B, Sørensen KJ, Have P, Madsen KG, Smedegaard Madsen E, Alexandersen S (1997) Appearance of acute PRRS-like symptoms in sow herds after vaccination with a modified live PRRS vaccine. *Vet Rec* 141: 497–499
5. Evans DMA, Dunn G, Minor PD, Schild GC, Cann AJ, Stanway G, Almond JW, Currey K, Maizel Jr JV (1985) Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome. *Nature* 314: 548–550
6. Hwang Y-K, Brinton MA (1998) A 68-nucleotide sequence, within the 3' noncoding region of simian hemorrhagic fever Virus negative-strand RNA binds to four MA 104 cell proteins. *J Virol* 72: 4 341–4 351
7. Jenkins TD, Opitz OG, Okano J, Rustgi AK (1997) Transactivation of the human keratin 4 and Epstein-Barr virus ED-L2 promoters by gut-enriched Krüppel-like factor. *J Biol Chem* 273: 10 747–10 754
8. Kinney RM, Chang G-J, Tsuchiya KR, Sneider JM, Roehrig JT, Woodward TM, Trent DW (1993) Attenuation of Venezuelan equine encephalitis virus strain TC -83 is encoded by the 5'-noncoding region and the E2 envelope glycoprotein. *J Virol* 67: 1 269–1 277
9. Liao C-L, Lai MC (1994) Requirement of the 5'-end genomic sequence as an upstream cis-acting element for coronavirus subgenomic mRNA transcription. *J Virol* 68: 4 727–4 737
10. Madsen KG, Hansen CM, Madsen ES, Strandbygaard B, Bøtner A, Sørensen KJ (1998) Detection of porcine reproductive and respiratory syndrome virus of the American type in Danish swine herds. *Arch Virol* 143: 1 683–1 700
11. Makino S, Lai MC (1969) High-frequency leader sequence switching during coronavirus defective interfering RNA replication. *J Virol* 63: 5 285–5 292
12. Meng XJ, Paul PS, Halbur PG, Lum MA (1995a) Phylogenetic analyses of the putative M (ORF 6) and N (ORF 7) genes of porcine reproductive and respiratory syndrome virus (PRRSV): implication for the existence of two genotypes of PRRSV in the U.S.A. and Europe. *Arch Virol* 140: 745–755

13. Meng XJ, Paul PS, Halbur PG, Morozov I (1995b) Sequence comparison of open reading frames 2 to 5 of low and high virulence United States isolates of porcine reproductive and respiratory syndrome virus. *J Gen Virol* 76: 3 181–3 188
14. Meng XJ, Paul PS, Morozov I, Halbur PG (1996) A nested set of six or seven subgenomic mRNAs is formed in cells infected with different isolates of porcine reproductive and respiratory syndrome virus. *J Gen Virol* 77: 1 265–1 270
15. Meulenbergh JJM, Bos-de Ruijter JNA, Van de Graaf R, Wensvoort G, Moormann RJM (1998) Infectious transcripts from cloned genome-length cDNA of porcine reproductive and respiratory syndrome virus. *J Virol* 72: 380–387
16. Meulenbergh JM, Meijer Ed, Moormann R (1993) Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. *J Gen Virol* 74: 1 697–1 701
17. Nielsen J, Bøtner A, Oleksiewicz MB (1998) Experimental inoculation of late-term pregnant sows with a field isolate of PRRS vaccine-like virus. *Proc. 15th Int. Pig Vet. Soc. Congr., Birmingham, U.K.*
18. Okumura K, Sakaguchi G, Takagi S, Naito K, Mimori T, Igarashi H (1996) Sp1 family proteins recognize the U5 repressive element of the long terminal repeat of human T cell leukemia virus type I through binding to the CACCC core motif. *Proc Natl Acad Sci USA* 271: 12 944–12 950
19. Saito A, Kanno T, Murakami Y, Muramatsu M, Yamaguchi S (1996) Characteristics of major structural protein coding gene and leader-body sequence in subgenomic mRNA of porcine reproductive and respiratory syndrome virus isolated in Japan. *J Vet Med Sci* 58: 377–380
20. Snijder EJ, Meulenbergh JJM (1998) The molecular biology of arteriviruses. *J Gen Virol* 79: 961–979
21. Sørensen KJ, Strandbygaard B, Bøtner A, Madsen ES, Nielsen J, Have P (1998) Blocking ELISA's for the distinction between antibodies against European and American strains of porcine reproductive and respiratory syndrome (PRRS) virus. *Vet Microbiol* 60: 169–177
22. Tahara SM, Dietlin TA, Bergmann CC, Nelson GW, Kyuwa S, Anthony RP, Stohlman SA (1994) Coronavirus translational regulation: leader affects mRNA efficiency. *Virology* 202: 621–630
23. Zhang X, Liao C-L, Lai MC (1993) Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in trans and in cis. *J Virol* 68: 4 738–4 746

Authors' address: Dr. T. Stargaard, Ministry of Food, Agriculture, and Fisheries, Danish Veterinary Institute for Virus Research, Lindholm, DK-4771 Kalvehave, Denmark.

Received October 12, 1998