Arch Virol (1995) 140: 1313–1319

Archives of Virology

Sequence analysis of open reading frames (ORFs) 2 to 4 of a U.S. isolate of porcine reproductive and respiratory syndrome virus

Brief Report

I. Morozov, X.-J. Meng, and **P. S. Paul**

Veterinary Medical Research Institute and the Department of Microbiology, Immunology and Preventive Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, U.S.A.

Accepted February 23, 1995

Summary. The sequence of ORFs 2 to 4 of a U.S. isolate of porcine reproductive and respiratory syndrome virus (PRRSV), ATCC VR2385, was determined by analysis of a cDNA λ library. The cDNA clones containing PRRSV specific sequences were selected using a VR2385 ORF 4 specific PCR probe and sequenced. The ORFs 2, 3 and 4 overlapped each other and encoded polypeptides with predicted M_r of 29.5 kDa (ORF 2), 28.7 kDa (ORF 3) and 19.5 kDa (ORF 4), respectively. No overlap was found between ORFs 4 and 5, and instead there was a 10 bp sequence which separated these two ORFs. The nucleic acid homology with corresponding ORFs of the European PRRSV isolate Lelystad virus (LV) was 65% for ORF 2, 64% for ORF 3 and 66% for ORF 4. Comparison of the ORF 4 sequences of VR2385 with that of another U.S. isolate MN-1b revealed only 86% amino acid sequence homology and the presence of deletions in the ORF 4 of MN-1b. Our results further strengthen the observation that there is sequence variation between US and European PRRSV isolates.

*

Porcine reproductive and respiratory syndrome virus (PRRSV) belongs to the newly proposed virus family *Arteriviridae*, which also includes equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) and simian hemorrhagic fever virus (SHFV). Porcine reproductive and respiratory syndrome (PRRS) was first described in the U.S. in 1987 [9]. A similar disease referred to as porcine epidemic abortion and respiratory syndrome (PEARS) was then reported in Europe [17]. PRRSV was first isolated in Europe and is believed to be widespread in swine population around the world [4, 21, 22]. All European isolates of PRRSV are antigenically and genetically related, whereas there are antigenic variations between US and European isolates as well as among US isolates [1, 16, 21]. The complete

nucleotide sequence of the genome of LV has been determined [14], but until recently limited information was available about the molecular structure of the genome of North American isolates of PRRSV [10–13]. We have previously reported the cloning and sequencing of the ORFs 5 to 7 of a U.S. isolate of PRRSV VR2385 of high virulence [12]. The 3' end of the genome of the VR2385 and the other U.S. PRRSV isolates showed a striking difference when compared to the European isolates [13]. In this study, we report on the cloning and sequencing of the ORFs 2–4 of the U.S. isolate VR2385.

For sequencing and characterization of the viral genome of VR2385 a cDNA λ library was constructed. The CRL11171 cells were infected with VR2385 virus at a M.O.I. of 0.1 and the total RNA from infected cells was isolated at 24 h post infection by using a guanidinium thiocyanate method [18]. Polyadenylated RNA was enriched, reverse transcribed and cloned into the λ ZAP vector using the Uni-Zap cDNA cloning kit (Stratagene, La Jolla, CA). A PCR probe generated by ORF 4 specific primers DP585 (5'GCTTTGCTGTCCTCCAAG 3') and DP586 (5'GATGCCTGACACATTGCC 3') [11] were used to screen the library. Plaques that hybridized with the probe were isolated and purified. The phagemids containing viral cDNA inserts were rescued by in vitro excision using ExAssist helper phage and *E. coli* SOLR cells (Stratagene, LaJolla, CA). Several recombinant phagemids with virus specific cDNA inserts with sizes ranging from 2.3 to 3.9 kb were selected and sequenced by Sanger's dideoxynucleotide chain termination method [19] with an automated DNA sequencer (Applied Biosystems, Foster City, CA). Universal, reverse and specific internal primers were used to determine the sequence. At least 3 independent clones representing sequence of the ORFs 2 to 4 were sequenced. The sequencing data was assembled and analyzed using Mac Vector (International Biotechnologies, Inc., CT) and GeneWorks (IntelliGenetics, CA) computer programs. The nucleotide sequence reported in this paper has been deposited in the GenBank with the accession number U20788.

Analysis of the nucleotide sequence identified three partially overlapping ORFs. The ORF 2 extended from nucleotide 28 to 795, ORF 3 from 651 to 1412, and ORF 4 from 1196 to 1729. There was an overlap of 144 bp between ORFs 2 and 3, and 216 bp between ORFs 3 and 4. Surprisingly, no overlap was found between ORFs 4 and 5. The start codon of ORF 5 was located 10 bp downstream of the stop codon of ORF 4. However, the ATG start codon of ORF 5 and TGA stop codon of ORF 4

Table 1. Characteristics of VR2385 and LV ORFs 2, 3, and 4

ORF	VR2385			LV		
	size (bp)	predicted Mr of product (kDa)	potential *N*-glycosylation sites	size (bp)	predicted Mr of product (kDa)	potential *N*-glycosylation sites
2	768	29.5	2	747	28.4	2
3	762	28.7	7	795	30.6	7
4	534	19.5	4	549	19.3	4

overlapped by only 1 bp in LV [5, 14]. The sequence at the region of the ORF 4 and ORF 5 junction of LV is ATATGA. We sequenced the corresponding region of 5 additional independent clones of VR2385 and in all cases the sequence of this region of VR2385 was ATTTGA. The point mutation from A to T in VR2385 and

a

```
VR2385 ORF 2  MKWGLC--K----AFLTKLAN-FLWMLSRSSWCPLLISLYFWPFCLASPSQVGWWSFASDWFAPRYSVRALPFTL   68
LV ORF 2      MQWGHCGVKSASCSWTPSLSSLLVWLI-------LPFSL---PYCLGSPSQDGYWSFFSEWFAPRFSVRALPFTL   65

VR2385 ORF 2  SNYRRSYEAFLSQCQVDIPTWGTKHPLGMLWHHKVSTLIDEMVSRRMYRIMEKAGQAAWKQVVSEATLSRISSLD   143
LV ORF 2      PNYRRSYEGLLPNCRPDVPQFAVKHPLGMFWHMRVSHLIDEMVSRRIYQTMEHSGQAAWKQVVGEATLTKLSGLD   140

VR2385 ORF 2  VVAHFQHLAAIEAETCKYLASRLPMLHHLRMTGSNVTIVYNSTLNQVFAVFPTPGSRPKLHDFQQWLIAVHSSIF   218
LV ORF 2      IVTHFQHLAAVEADSCRFLSSRLVMLKNLAV--GNVSLQYNTTLDRVELIFPTPGTRPKLTDFRQWLISVHASIF   213

VR2385 ORF 2  SSVAASCTLFVVLWLRVPMLRTVFGFRWLGAIFLSNSR   256
LV ORF 2      SSVASSVTLFIVLWLRIPALRYVFGFHWPTAT--HHSS   249
```

b

```
VR2385 ORF 3  MANSCTFLYIFLCCSFLYSFCCAVVAGSNATYCFWFPLVRGNFSFELTVNYTVCPPCLTRQAAAEAYEPGRSLWC   50
LV ORF 3      MAHQCARFHFFLCGFICYLVHSALASNSSSTLCFWFPLAHGNTSFELTINYTICMPCSTSQAARQRLEPGRNMWC   50

VR2385 ORF 3  RIGHDRCGEDDHDELGFVVPSGLSSEGHLTSAYAWLAFLSFSYTAQFHPEIFGIGNVSRVYVDIKHQFICAVHDG   150
LV ORF 3      KIGHDRCEERDHDELLMSIPSGYDNL-KLEGYYAWLAFLSFSYAAQFHPELFGIGNVSRVFVDKRHQFICAEHDG   149

VR2385 ORF 3  QNTTLPHHDNISAVFQTYYQHQVDGGNWFHLEWLRPFFSSWLVLNVSWFLRRSPASHVSVRVFQTSRPTPPQRQA   225
LV ORF 3      HNSTVSTGHNISALYAAYYHHQIDGGNWFHLEWLRPLFSSWLVLNISWFlRRSPVSPVSRRIYQILRPTRPRLPV   224

VR2385 ORF 3  LLSSKTSV--ALGIATRPLRRFA--------KS--LSAARR   254
LV ORF 3      SWSFRTSIVSDLTGSQQRKRKFPSESRPNVVKPSVLPSTSR   265
```

c

```
VR2385 ORF 4  MAASLLFLLVGFKCLLVSQAFACKPCFSSSLSDIKTNTTAAAGFAVLQDISCLRHR--NSASEAIR--KVPQCRT    71
LV ORF 4      MAAATLFFLAGAQHIMVSEAFACKPCFSTHLSDIETNTTAAAGFMVLQDINCFRPHGVSAAQEKISFGKSSQCRE    75
MN1b ORF4     MAAATLFLTVGFKCLLVSQAFA-----ANHVADIKTNTTAAASFAVLQDISCLRHR--NSASEAIR--KIPQCRA    66

VR2385 ORF 4  AIGTPVYITVTANVTDENYLHSSDLLMLSSCLFYASEMSEKGFKVVFGNVSGIVAVCVNFTSYVQHVKEFTQRSL   146
LV ORF 4      AVGTPQYITITANVTDESYLYNADLLMLSACLFYASEMSEKGFKVIFGNVSGVVSACVNFTDYVAHVTQHTQQHH   150
MN1b ORF4     AIGTPVYITITANVTDENYLHSSDLLMLSSCLFYASEMSEKGFEVVFGNVSGIVAVCVNFTSYVQHVREFTQLLL   141

VR2385 ORF 4  VVDH-VRLLHFMTPETMRWATVLACLFTILLAI   178
LV ORF 4      LVIDHIRLLHFLTPSAMRWATTIACLFAILLAI   183
MN1b ORF4     DRVRLLHFM---TPETMRWATVLACLFAILLAI   171
```

Fig. 1. Alignment of predicted amino acid sequence of ORFs 2 (**a**), 3 (**b**), and 4 (**c**) of PRRSV VR2385 and LV. The ORF 4 sequence of another U.S. isolate MN-1B was also included in the alignment. The sequences for LV were reported by Meulenberg et al. [14], and the ORF 4 sequence for MN-1b was reported by Kwang et al. [10]

probably some other unidentified changes in this region of VR2385 made the ORF 5 ATG start codon 10 bp downstream of the stop codon of ORF 4, and a 10 bp non-coding region appeared in the ORF 4 and 5 junction of VR2385.

The characteristics of ORFs 2 to 4 of VR2385 are summarized in Table 1. The ORF 2 encodes a 256 amino acid polypeptide with a predicted size of 29.5 kDa. The carboxy and amino terminus of the predicted protein are hydrophobic (data not shown) and there are two potential N-glycosylation sites in the ORF 2 protein. ORF 3 encodes a protein of 254 amino acids and contains 7 potential N-glycosylation sites. The amino terminus of the ORF 3 protein is extremely hydrophobic. ORF 4 encoded a 178 amino acid protein with a predicted size of 19.5 kDa. The amino and carboxy termini and 4 regions within the protein are highly hydrophobic. Comparison of the nucleotide sequences of VR2385 and LV showed extensive variations. Nucleotide sequence identity between VR2385 and LV is 65% for ORF 2, 64% for ORF 3 and 66% for ORF 4. Alignment of the predicted amino acid sequences of ORFs 2– 4 of VR2385 and LV is presented in Fig. 1. Amino acid identity between VR2385 and LV is 58% for ORF 2, 56% for ORF 3 and 67% for ORF 4. We also compared the sequence of VR2385 ORF 4 with that of MN-1b, another US isolate of PRRSV [10]. The ORF 4 of VR2385 is 21 bp longer and shares an 88% nucleotide sequence homology with MN-1b. The amino acid homology between the ORF 4 of VR2385 and MN-1b is 86% (Fig. 1c). Several deletions were found in the ORF 4 of MN-1b compared to VR2385.

The ORFs 6 and 7 of PRRSV are predicted to encode the viral membrane glycoprotein and the viral nucleocapsid protein, respectively [12, 13]. Analysis of predicted amino acid sequences encoded by ORFs 2–5 of LV, LDV and EAV showed that all of these proteins share features of membrane associated proteins [5, 6, 8, 14]. The EAV ORF 5 product was identified as the main envelope glycoprotein

Table 2. Potential leader-mRNA junction regions in the genome of VR2385 and leader-mRNA junction regions of LV[a]

	VR2385		Lelystad virus	
ORF	sequence	position[b]	sequence	position
1	–	–	UUAACC	–
2	UGAACC	20	UAAACC	38
3	GUAACC	83	UUGACC	11
	CCAACC	35		
4	UUGACC	230	UCAACC	83
	CAGACC	44		
5	UUGACC	99	ACAACC	32
	GAGACC	64		
6	GUAACC	17	UCAACC	24
7	UAAACC	9	UUAACC	9

[a]Sequence for VR2385 ORFs 2–4 is presented in the study, ORFs 5–7 was reported by Meng et al. [12] and LV ORFs 1–7 was reported by Meulenberg et al. [14]

[b]Distance in nucleotides between proposed junction motif and AUG start codon of downstream ORF

[7]. Our data indicates that the proteins encoded by ORFs 2–4 of VR2385 possess characteristics similar to those of LV and probably are envelope or membrane associated glycoproteins because of their hydrophobicity and presence of potential glycosylation sites. Further work is necessary to determine the roles of these proteins. The variability found in the ORF 4 sequence between the two U.S. isolates correlate with the findings that ORF 4 protein of the MN-1b expressed in *E. coli* reacted with only 65% of PRRSV positive sera by Western blot analysis [10].

A nested set of subgenomic mRNA is formed during replication of PRRSV and other members of the arterivirus group [5, 6, 8, 12, 14]. All subgenomic mRNAs contain a common leader sequence derived from the 5' noncoding region of the viral genome. The site of the leader-mRNA junction is similar and located upstream of the start codon of each ORF. The consensus leader-mRNA junction sequence of the six subgenomic mRNAs of LV was determined to be (U/A)(C/U/A)(A/G)ACC [15]. Similar sequences were also found as leader-mRNA junction regions for LDV [3]. The potential leader-mRNA junction motifs of ORFs 2 to 4 of VR2385 was proposed and compared with those of LV (Table 2). The last four nucleotides of the motif for ORFs 1, 2, 4, 5, 6 and 7 in LV are AACC, and for ORF 3 is GACC. The AACC motif has been found upstream of ORFs 6 and 7 of VR2385 [12] as well as ORFs 2 and 3. There are two potential junction regions for ORF 3, 83 bp and 35 bp upstream of the ORF 3 start codon, respectively (Table 1). No AACC motif was found upstream of VR2385 ORFs 4 and 5. However, the sequences UUGACC and CAGACC upstream of ORF 4, UUGACC and GAGACC upstream of ORF 5, may be the leader-mRNA junction regions for the mRNAs 4 and 5 of VR2385. Multiple potential leader-mRNA junction sites suggest that polymorphism of subgenomic mRNAs may exist among PRRSV isolates. Experiments to determine the exact locations of leader-mRNA junction regions are now in progress.

The sequence variations observed in this study between a U.S. and a European PRRSV isolate, as well as between two North American PRRSV isolates, indicates the heterogenetic nature of PRRSV isolates and the need for further characterization of additional PRRSV isolates. Whether this genetic variation between VR2385 and LV reflects the observed difference in virulence needs to be further studied.

Acknowledgements

This work was supported by a grant No. 94-02092 from the National Research Initiative Competitive Grants program of the U.S. Department of Agriculture, and in part by a grant from the Solvay Animal Health, Inc., Mendota Heights, MN. The authors would like to thank Drs. Pat Halbur and Melissa Lum for their expert help throughout this project, and Dr. Harold Hills at the Nucleic Acid Facility, Iowa State University for his assistance in sequence analysis.

References

1. Bautista EM, Goyal SM, Yoon IJ, Joo HS, Collins JE (1993). Comparison of porcine alveolar macrophages and CL 2621 for the detection of porcine reproductive and respiratory syndrome (PRRS) virus and anti-PRRS antibody. J Vet Diagn Invest 5: 163–165

2. Benfield DA, Nelson E, Collins JE, Harris L, Goyal SM, Robison D, Christianson WT, Morrison RB, Gorcyca D, Chladek D (1992) Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332). J Vet Diagn Invest 5: 612–614
3. Chen Z, Kuo L, Rowland RRR, Even C, Faaberg KS, Plagemann PGW (1993) Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of seven subgenomic mRNAs. J Gen Virol 74: 643–660
4. Collins JE, Benfield DA, Christianson WT, Harris L, Hennings JC, Shaw DP, Goyal SM, McCullough S, Morrison RB, Joo HS, Gorcyca D, Chladek D (1992) Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. J Vet Diagn Invest 14: 117–126
5. Conzelmann K-K, Visser N, Van Woensel P, Thiel H-J (1993) Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the Arterivirus group. Virology 193: 329–339
6. Den Boon JA, Snijder EJ, Chirnside ED, De Vries AAF, Horzinek MC, Spaan WJM (1991) Equine arteritis virus is not a togavirus but belongs to the coronavirus-like superfamily. J Virol 65: 2910–2920
7. de Vries AAF, Chirnside ED, Horzinek MC, Rottier PJM (1992) Structural proteins of equine arteritis virus. J Virol 66: 6294–6303
8. Godeny EK, Chen L, Kumar SN, Methven SL, Koonin EV, Brinton MA (1993) Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). Virology 194: 585–596
9. Hill H (1990) Overview and history of mystery swine disease (swine infertility and respiratory syndrome). In: Proceedings of the Mystery Swine Disease Committee Meeting, Denver, Colorado, pp 29–31. Denver, Livestock Conservation Institute
10. Kwang J, Kim HS, Joo HS (1994) Cloning, expression, and sequence analysis of the ORF 4 gene of porcine reproductive and respiratory syndrome virus MN-1b. J Vet Diagn Invest 6: 293–296
11. Mardassi H, Mounir S, Dea S (1994) Identification of major differences in the nucleocapsid protein genes of a Quebec strain and European strains of porcine reproductive and respiratory syndrome virus. J Gen Virol 75: 681–685
12. Meng X-J, Paul PS, Halbur PG (1994) Molecular cloning and nucleotide sequencing of the 3'-terminal genomic RNA of porcine reproductive and respiratory syndrome virus. J Gen Virol 75: 1795–1801
13. Meng X-J, Paul PS, Halbur PG, Lum MA (1995) Phylogenetic analysis 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 U.S. and Europe. Arch Virol 140: 745–755
14. Meulenberg JJM, Hulst MM, De Meijer EJ, Moonen PLJM, Den Besten A, De Kluyver EP, Wensvoort G, Moormann RJM (1993a) Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. Virology 192: 62–72
15. Meulenberg JJM, De Meijer EJ, Moormann RJM (1993b) Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. J Gen Virol 74: 1697–1701
16. Nelson EA, Christopher-Hennings J, Drew T, Wensvoort G, Collins JE, Benfield DA (1993) Differentiation of U.S. and European isolates of porcine reproductive and respiratory syndrome virus by monoclonal antibodies. J Clin Microbiol 31: 3184–3189
17. Paul PS, Halbur PG, Meng X-J (1993) Porcine reproductive and respiratory syndrome: an overview. J Clin Vet Med 11: 19–28

18. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
19. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
20. Terpstra C, Wensvoort G, Pol JMA (1991) Experimental reproduction of porcine epidemic abortion and respiratory syndrome (mystery swine disease) by infection with Lelystad virus: Koch's postulates fulfilled. Vet Q 13: 131–136
21. Wensvoort G, De Kluyver EP, Luijtze EA, Den Besten A, Harris L, Collins JE, Christinanson WT, Chladek D (1992) Antigenic comparison of Lelystad virus and swine infertility and respiratory syndrome virus. J Vet Diagn Invest 4: 134–138
22. Wensvoort G, Terpstra C, Pol JMA, Ter Laak EA, Bloemraad M, de Kluyver EP, Kragten C, van Buiten L, den Besten A, Wagenaar F, Broekhuijsen JM, Moonen PLJM, Zetstra T, de Boer EA, Tibben HJ, de Jong MF, Van't Veld P, Groenland GJR, van Gennep JA, Votes MT, Verheijden JHM, Braamskamp J (1991) Mystery swine disease in the Netherlands: the isolation of Lelystad virus. Vet Q 13: 121–130

Authors' address: Dr. P. S. Paul, Veterinary Medical Research Institute, College of Veterinary Medicine, Iowa State University, 1802 Elwood Drive, Ames, Iowa 50011, U.S.A.

Received December 20, 1994