Cloning and Sequence Analysis of the Spike Gene of Porcine Epidemic Diarrhea Virus Chinju99*

SANG-GEON YEO,^{1,†} MERCEDES HERNANDEZ,² PETER J. KRELL³ & ÉVA NAGY⁴

¹*College of Veterinary Medicine, Institute of Animal Medicine, Gyeongsang National University, Chinju 660-701, Republic of Korea*

²*Department of Biology, Malaspina University College, Nanaimo, Canada, V9R 5S5*

³*Department of Microbiology*

⁴*Department of Pathobiology, University of Guelph, Guelph, Canada, N1G 2W1*

Received January 22, 2003; Accepted February 4, 2003

Abstract. The spike (S) gene of the porcine epidemic diarrhea virus (PEDV) Chinju99 which was previously isolated in Chinju, Korea was cloned and sequenced to aid in the development of genetically engineered vaccines and diagnostic reagents against PEDV. The nucleotide sequence encoding the entire S gene open reading frame (ORF) of Chinju99 was 4152 bases long encoding 1383 amino acids. It consisted of 1001 adenine (24.1%), 849 cytosine (20.4%), 877 guanine (21.1%) and 1425 thymine (34.3%) residues. The Chinju99 S ORF nucleotide sequence was 94.5% homologous with that of the Br1/87 and CV777 strains, respectively. The Chinju99 S protein had 92.8% amino acid identity with that of Br1/87 and CV777, respectively. The amino acid sequence contained 27 potential sites for asparagine (N)-linked glycosylation and there was a stretch of highly hydrophobic residues at position 1325–1350.

Key words: nucleotide sequence, PEDV, S gene

Introduction

Since porcine epidemic diarrhea (PED) was first reported in Belgium [1] and the UK [2] in 1978, it has occurred in many countries and causes economic losses in pig farming [3–6]. Pigs in all age groups are susceptible to the disease and show acute diarrhea and dehydration, leading to death with high mortality, especially up to 90% in 1–2 weeks old piglets [7–9]. PED is clinically and pathologically similar to transmissible gastroenteritis (TGE), which obscures the differentiation of the two diseases. Immunity is,

therefore, important for the prevention and control of the disease. Moreover, maternal antibodies derived from immunized sows are the sole source for immunity to the disease in highly susceptible neonates during the first few weeks after birth [7,9]. However, no effective vaccines have been developed for the mucosal immunity in the gut of pigs where the etiologic agent causes severe enteritis [9].

The etiologic agent, porcine epidemic diarrhea virus (PEDV), belongs to the family *Coronaviridae* of the order *Nidovirales* and has a genome of positive sense, ss-RNA [7,10]. The subgenomic mRNAs, which are transcribed from the genome, produce viral proteins such as the spike (S, 180–220 kDa), membrane (M, 27–32 kDa) and nucleocapsid (N, 55–58 kDa) proteins [10–12]. Recently, Pensaert [9] suggested a molecular mass of 85–135 kDa for the S protein. Among the proteins, S, a glycoprotein peplomer on the viral surface, plays an important role

*Nucleotide sequence data reported is available in the GenBank database under the accession no. AY167585.

† Author for all correspondence.

E-mail: sgyeo@nongae.gsnu.ac.kr

in the attachment of viral particles to the receptors of host cells with subsequent penetration into the cells by membrane fusion. The S glycoprotein also stimulates induction of neutralizing antibodies in the host [13]. The S gene is, therefore, an important target for cloning and expression in the development of genetically engineered vaccines. Duarte et al. [14] reported a sequence of 1741 nucleotides in the region between the S and N genes of the British, Br1/87 and Belgian, CV777 strains of PEDV. Duarte and Laude [13] also reported a sequence of 4162 nucleotides including the coding region of the Br1/87 S gene. More recently, the complete S gene sequence has been reported within the complete genome sequence of PEDV CV777 [15].

PED occurs frequently in Korea since the virus was first isolated there [6], and developmental efforts should target accurate diagnosis and control of the disease. Studies on the nucleotide sequences of the S gene of Korean PEDV isolates have not yet been reported. In the present study, DNA clones of the full-length S gene open reading frame (ORF) of PEDV isolated in Chinju, Korea (PEDV Chinju99) were constructed. The nucleotide and deduced amino acid sequences of the S gene were determined, and further analyzed and aligned with those of other PEDVs to aid in the production of genetically engineered vaccines and diagnostic reagents.

Materials and Methods

Virus

PEDV Chinju99, first isolated by the Virology Laboratory of Gyeongsang National University College of Veterinary Medicine, Chinju, Korea from

the intestinal tissues of piglets suffering from severe diarrhea (data not shown), was used for cloning of the S gene. The virus was propagated in Vero cells grown in minimal essential medium (MEM) containing streptomycin (100 µg/ml), penicillin (100 U/ml) and trypsin (10 µg/ml) in a 5% CO₂ incubator at 37°C following the methods of Hofmann and Wyler [16].

Extraction of Viral RNA

The medium was removed at 15 h post-infection at the early stage of cytopathic effects such as rounding, degeneration and syncytia formation. Cells were lysed by Trizol[®] reagent (Invitrogen, USA) at 2 ml per tissue culture flask (25 cm²) and homogenized by passing the cell lysate several times through a pipette. Viral RNA was extracted from the homogenate following the manufacturer's suggestions and dissolved in diethyl pyrocarbonate-treated distilled water.

Primers used for cDNA Synthesis

Pairs of sense and antisense primers were designed and aligned based on the nucleotide sequence of the Br1/87 S gene [13] from the GenBank data base (National Center for Biotechnology Information, USA). These primers were used to generate cDNA for the S gene. The nucleotide sequences and relative position of the primers are shown in Table 1 and Fig. 1, respectively.

cDNA Synthesis for the S Gene

Synthesis of the first-strand cDNA for the S gene was carried out by reverse transcription (RT) using

Table 1. Primers used in cDNA synthesis for PEDV S gene

Primers	Nucleotide Sequences	Mers	% GC	Strand
SF1	5'ACGTAAACAAATGAGGTC 3'	18	39.0	+
SF2	5'TACCTTCTACTGTCAGGG 3'	18	50.0	+
SF3	5'ACTCTGGTGTCATGGTAC 3'	18	48.6	+
SR1	5'CTGTTCGTGACTCAGAAG 3'	18	50.0	—
SR2	5'CTCGCTTGAACAGCATAG 3	18	50.0	—
SR3	5'ACATCACTGCACGTGGAC 3'	18	61.0	—

Fig. 1. Construction of cDNA clones for the full-length S gene of PEDV Chinju99 strain by RT-PCR using sense (SF) and antisense (SR) primers: diagrammatic representation of the S gene of viral RNA (solid bar) and S gene ORF (long open rectangle) show primer-binding sites (small open rectangles); three DNA fragments amplified by PCR and cloned into pUC19 vector are denoted as recombinant DNA clones Sfrag1, Sfrag2 and Sfrag3.

Superscript II[®] reverse transcriptase reagent kit (Invitrogen) following the manufacturer's suggestions. The viral RNA was mixed with 1 µl of 100 pM of the appropriate antisense primer, 4 µl of 5X first-strand buffer, 1 µl of 10 mM dNTP mixture, 2 µl of 0.1 M DTT, 1 µl of RNase inhibitor (40 U/µl), 1 µl of reverse transcriptase (200 U/µl) and brought to 20 µl with distilled water. The reaction mixture was incubated for 50 min at 42°C, and the reaction was stopped by heat at 70°C for 15 min. RNase H (1 U) treatment was done for 20 min at 37°C to degrade the RNA template.

The ds-cDNA for the S gene was synthesized by polymerase chain reaction (PCR) using a reagent kit (Perkin-Elmer, USA). A 10 µl portion of the first-strand cDNA template was added to 5 µl of 10X PCR buffer, 4 µl of 25 mM MgCl₂, 1 µl of 10 mM dNTP mixture, 1 µl of each 100 pM sense and antisense primers, 1 µl of *Taq* DNA polymerase (5 U/µl) and brought to 50 µl with distilled water. The PCR was carried out in a thermocycler (Perkin-Elmer) following the program of 2 min at 94°C and 30 cycles of 1 min at 94°C, 1 min at 45–50°C depending on the primers and 1 min at 72°C, and a final extension at 72°C for 5 min. The PCR

products were resolved by electrophoresis in 1% agarose gels.

Cloning of cDNA

The PCR-generated S gene ds-cDNAs were blunt-ended with Klenow enzyme (2 U) and 1 µl of 0.5 mM dNTPs (Invitrogen) in 20-µl reaction volume and cloned into the *Sma*I site of pUC19 plasmid DNA by ligation using T4 DNA ligase (1 U) (Invitrogen) [17]. The recombinant plasmid DNAs were transformed into competent *Escherichia coli* DH5α cells by heat shock for 45 s at 42°C. After adding SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 20 mM MgSO₄, 20 mM glucose), the tube was shaken for 1 h at 220 rpm, 37°C. The transformed cells were plated onto Luria Bertani (LB) agar (Invitrogen) containing ampicillin (50 µg/ml), X-gal (40 µg/ml) and isopropylthio-β-galactoside (20 µg/ml) (Invitrogen) and incubated at 37°C, overnight. Transformed colonies were cultured in LB broth with ampicillin (50 µg/ml) by shaking at 220 rpm, at 37°C, overnight, and were subjected to DNA extraction by alkaline-lysis, restriction enzyme

Chinju99	SF1											
Br1/87	ACGTAAACAA	ATGAGGTCTT	TAATTTATTT	TTTGTGTTTC	GTACCAGTAC	TTCCAACACT	CAGCCTACCA	CAAGATGTCT	ATAGGTGCTC	AGCTAACACT	100	
			C C G C T					A C	CA GT CT		100	
Chinju99	AACTTTAGGC	GGTTCTTTTC	AAAATTTAAT	GTTCAGGCGC	CTGCAGTCGT	TGTAAGGCGC	GGTTATCTAC	CTAGTGTGAA	ACACGCTGGT	TGGTACTGTG	200	
Br1/87				A C	C TT T C		A	CTCTT A C			200	
Chinju99	CTGCCGACAT	CCAAACTGCT	AGTGGCGTTC	ATGGTATCTT	CCTTAGCCAT	ATTAGAGGTG	GTGATGGCTT	TGAGATTGGC	ATTTCGCAAG	AGCCGTTTGA	300	
Br1/87	GCA A G TG			T T C T C	CGATTC	G					300	
Chinju99	CCCTAGTGGT	TACCAGCTTT	ATTTACATAA	GGCCACTAAC	GGTAACACTA	ATGCTACTGC	GCGACTGCGT	ATTTGCCAGT	TTCCGACGAG	TAAACATTG	400	
Br1/87	T			T		(T)* A	C		GAT A		400	
Chinju99	GGCCCCACTG	CTAATGATGT	TACAACAGGT	CGTAATTGCC	TATTTAACAA	AGCCATTCCA	GCTCATATGA	GTGATGGAAA	ACATAGTGTG	GTCCGCATAA	500	
Br1/87		T		C C			T C		G T		500	
Chinju99	CATGGGATAA	TGATCGTGTC	ACTGTTTTTG	CTGACAAGAT	CTATCATTTT	TATCTTAAAA	ATGATTGGTC	CCGCGTTGCG	ACAAACTGTT	ACAACAGTGG	600	
Br1/87									GAT TC CA		600	
Chinju99	AGGTTTGCT	ATGCAATATG	TTTACGAACC	TATTTACTAC	ATGCTTAATG	TTACTAGTGC	TGGTGAGGAT	GGTATTTCCT	ATCAACCCCTG	TACAGCTAAT	700	
Br1/87	A		TAC CC			A	C		G		700	
Chinju99	TGCATTGGTT	ATGCTGCCAA	TGTATTTGCC	ACTGAGTCCA	ATGGCCACAT	ACCAGAAGGT	TTTAGTTTTA	ATAATTGGTT	TCITTTTGCC	AATGATTCCA	800	
Br1/87	C C			T	T				A C		800	
Chinju99	CTTTGTTTCA	TGGTAAAGTG	GTTTCCAACC	AACCTTTGTT	GGTCAACTGT	CTTTGGGCCA	TTCTTAAGAT	TTACGACTA	GGCCACTTTT	TCTCATTTAA	900	
Br1/87	G			C	T T			T A		C	900	
Chinju99	TCAAACGATG	GATGGCGTTT	GTAACGGAGC	TACTGCGTAT	CGTGCCCCAC	AGGCTCTGAG	GTTTAATATT	AATGACACTT	CTGTCATTCT	TGCCGAAGGC	1000	
Br1/87	C		T G T		G			C CG		T	1000	
Chinju99	TCAATTGTAC	TTGACTATGC	TTTAGGAACA	AATCTTTCTT	TTGTTTGCAG	TAATTCTCTA	GATCCTCACA	AAGCCATCTT	CTCCATACCT	CTGGGTGCTA	1100	
Br1/87							TT T		TG		1100	
Chinju99	CCCAAGTACC	CTACTATTGC	TTTCTTAAAG	TTGATACTTA	CAACTCCACT	GTCTATAAAT	TTCTGCTGTG	TTTACCTTCT	ACTGTCAGGG	AAATTGTGAT	1200	
Br1/87	TG		G			T	CTTG	(C)*			1200	
Chinju99	CACAAAGTAG	GGTGATGTTT	ATGTCAACGG	GTTTGCGTAC	TTGCACTCGG	GTTTGTGGGA	TGCTGTGACA	ATTAATTTC	CTGGCCATGG	CACCTAACGAT	1300	
Br1/87	C T		T		T			T	T G		1300	
Chinju99	GACGCTCAG	GTTTCTGGAC	CATAGCATCG	ACTAATTTTA	TTGATGCACT	CATCGAGGTT	CAAGCAACTG	CCATTGAGCG	TATCTTTTAT	TGTGATGATC	1400	
Br1/87	T			G			G T				1400	
Chinju99	CTGTTAGCCA	ACTCAAGTGT	TCTCAGGTTT	CTTTTGACCT	TGATGATGGT	TTTTACCCTA	TTTCTTCTAG	AAACCTTCTG	AGTCACGAAC	AGCCAATTTC	1500	
Br1/87			G		C	C					1500	
Chinju99	TTTTGTTACT	TTGCCATCAT	TTAATGATCA	TTCTTTTGTT	AATATTACTG	TATCTGCGGC	TTTTGGTGAT	TCTGGTGGTG	CCAACCTCGT	TGCATCTGAC	1600	
Br1/87						C		G CT A A	T		1600	
Chinju99	ACTACTATCA	ATGGGTTTAG	TTCTTTCTGT	GTGACACTA	GACAATATAC	CATTAGACTG	TTTTACAACG	TTACAAACAG	TTATGGTTAT	GTGCTAAAT	1700	
Br1/87					T	C	T T				1700	
Chinju99	CACAGGATAG	CAATTGCCCT	TTACCTTTGC	AATCTGTTAA	TGATTACCTG	TCTTTTAGCA	AATTTTGTGT	TTCAACCAGC	CTTTTGGCTG	GTGCTTGATC	1800	
Br1/87		T T									1800	
Chinju99	CATAGATCTT	TTTGGTTACC	CTGCTTTGCG	TAGTGGTGTG	AATTACAGGT	CCCTCTATTT	TCAATTGACA	AAAGGTGAGT	GGATTACTTG	GGCACCTAAA	1900	
Br1/87				G TG C	T				T G CA		1900	
Chinju99	CCCTTAAACG	GCCTTTTAAA	GGGGGGTTTT	ATGACTCTGG	ATGTGTGTAC	CAAGTATACT	ATCTATGGCT	TTAAAGGTGA	GGGTATTATT	ACCCTTACAA	2000	
Br1/87	AC TG A	TA CAC G	C TTTC								2000	
Chinju99	ATTCTAGCAT	TTTGGCAGGT	GTTTATTATA	CATCTGATTC	TGGACAGTTG	TTAGCCCTTA	AGAATGTCAC	TAGTGGTGCT	GTTTATTCTG	TCACGCCATG	2100	
Br1/87											2100	
Chinju99	TTCTTTTCA	GAGCAGGCTG	CATATGTTAA	TGATGATATA	GTGGGTGTTA	TTTCTAGTTT	GTCTAACTCC	ACTTTTAACA	ATACTAGGGA	GTTGCCTGGT	2200	
Br1/87											2200	
Chinju99	TTCTTCTACC	ATTCTAATGA	CGGCTCCAAT	TGTACAGAGC	CTGTGTGGGT	GTATAGTAAC	ATAGGTGTTT	GTAATCTGG	CAGTATTGGC	TATGTTCCAT	2300	
Br1/87											2300	

Fig. 2. Nucleotide sequence of the Chinju99 S gene and comparison of the sequence with Br1/87 (EMBL accession No. Z25483) and CV777 (GenBank accession No. AF353511) strains: start codon ATG and stop codon TGA are underlined; only the 229 nucleotides of Br1/87 and CV777 which mismatched the Chinju99 sequence are included; *, four nucleotides of CV777 which mismatched those of Br1/87 are shown in the Br1/87 sequence; regions corresponding to the six primers used for cloning are underlined and labeled above the sequence as SF1-3 and SR1-3.

Chinju99	CTCAGTATGG CCAAGTCAAG ATTGCACCCA CGGTACTTGG GAATATTAGT ATTCCCACCA ACITTAGTAT GAGTATTAGA ACAGAATATT TACAGCTTTA	2400
Br1/87		2400
Chinju99	CAACACGCCT GTTAGTGTG ATTGTGCTAC ATATGTTTGT AATGGTAACT CTCGTTGTAA ACAATTACTT ACCCAGTACA CTGCAGCATG TAAGACCATA	2500
Br1/87		2500
Chinju99	GAGTCAGCAT TAGAACTCAG CGCTAGGCTT GAGTCTGTTG AAGTTAACTC TATGCTTACT ATTTCTGAAG AGGCCCTACA GTTAGCTACC ATCAGTTTCT	2600
Br1/87		2600
Chinju99	TTAATGGTGA TGGGTATAAC TTTACTAATG TGTGGGGTGT TTCTGTGTAC GATTCTCAAA GTGGCAGGGT GATACACGAA AGGTCCTTTA TTGAAGACCT	2700
Br1/87		2700
Chinju99	GCTTTTAAAT AAGGTGGTTA CTAATGGCCT TGGTACTGTT GATGAAGACT ATAAGCGCTG TTCTAATGGT CGCTCTGTGG CTGATAAAGT CTGTGCGCAG	2800
Br1/87		2800
Chinju99	TATTACTCTG GTCTCATGGT ACTACCTGGG GTTGTGACG CTGAGAAGCT TCACATGTAT AGTGCCTCTC TCATCGGTGG TATGGCGCTA GGAGGTTTAA	2900
Br1/87		2900
Chinju99	CTGCCGCAGT GGCATTGCCT TTTAGCTATG CTGTTCAAGC GAGACTCAAT TATCTTGCTT TACAGACGGA TGTTTTACAG CGGAACCAGC AACTGCTTGC	3000
Br1/87		3000
Chinju99	TGAGTCTTTT AACTCTGCTA TTGTAATAT AACCTCAGCC TTGCGAGTG TTAAGAGGC TATCAGTCAA ACTTCTAAGG GTTTGAATAC TGTGGCTCAT	3100
Br1/87		3100
Chinju99	GCGTCTACTA AAGTTCAAGA GGTGTTAAT TCGCAGGGTT CAGCTTTGAC TCAACTTACC ATACAGCTGC AACATACTT CCAAGCCATT TCTAGTTCTA	3200
Br1/87		3200
Chinju99	CTGATGACAT TTAATCCCGA CTGGACATTC TTTCAGCCGA TGTTCAGGTT GATGCTCTCA TCACCGGCAG ATTATCAGCA CTTAATGCTT TTGTTGCTCA	3300
Br1/87		3300
Chinju99	AACCCTCACT AAGTATACTG AGGTTTCAGG TAGCAGGAAG CTAGCACAGC AAAAGGTAA TGAGTGCCTC AAATCGCAAT CTCAGCGTTA TGGTTTTTGT	3400
Br1/87		3400
Chinju99	GGTGGTGATG GCGAGCACAT TTTCTCTCTT GTACAGGCCG CACCCAGGG CCTGCTGTTT TTACACACAG TACTGTGACC GGGTGATTTT GTAAATGTTA	3500
Br1/87		3500
Chinju99	TTGCCATCGC AGGCTTATGC GTTAATGGTG ATATTGCCTT GACTCTACGT GAGCCTGGCT TAGTCTTGGT TACGCATGAA CTTCAAACCT ATACTGCGAC	3600
Br1/87		3600
Chinju99	GGAATATTTT GTTTCATCGC GAGGTATGTT TGAACCTAGA AAACCTACCG TTAGTGATTT TGTTCAAATT CAGAGTTGTG TGGTCACCTA TGTCAATCTG	3700
Br1/87		3700
Chinju99	ACTAGCGACC AACTACCAGA TGTAATCCCA GATTACATCG ATGTTAACAA AACACTTGAT GAGATTTTAG CTCTCTGCC CAATAGAACT GGTCCAAATC	3800
Br1/87		3800
Chinju99	TTCCCCTAGA TGTTTTTAAAT GCCACTTATC TTAATCTCAC TGGTGAAATT GCAGATCTAG AGCAGCGTTC AGAGTCTCTC CGTAATACTA CAGAAGAGCT	3900
Br1/87		3900
Chinju99	CCGAAGTCTT ATATACAATA TCAACAACAC ACTTGTGTGAC CTGAGTGGC TCAACCGAGT TGAGACTTAT ATCAAGTGGC CGTGGTGGGT TTGTTGATT	4000
Br1/87		4000
Chinju99	ATTTTATTG GTCTCATCTT TGTTGTGTCA TTATTAGTGT TCTGCTGCAT TTCCACGGGT TGTGTGGAT GCTGCGGTTG CTGCGGTGCT TGTTTTTCAG	4100
Br1/87		4100
Chinju99	GTTGTGGAG GGGTCCTAGA CTTCAACCTT ACGAAGGTTT TGAAAAGGTC CACGTGCAGT GA	4162
Br1/87		4162

Fig. 2. (Continued)

digestion and electrophoresis through 1% agarose gels for identification of recombinant DNA clones.

Nucleotide Sequencing

Nucleotide sequencing was done for the S gene-recombinant DNA clones using Dye Terminator Cycle Sequencing kit (Perkin-Elmer) by the automatic

sequencer (ABI prism 377, Advanced Biotechnologies, USA).

Analysis of Sequences of Nucleotides and Amino Acids

The sequences of nucleotides and deduced amino acids were analyzed by ClustalW, version 1.82 using

Chinju99	MRSLIYFLLF	VPVLPTLSLP	QDVYRCSANT	NFRFFSKFN	VQAPAVVLG	GYLPSVKHAG	WYCAADIQTA	SGVHGIFLSH	IRGGHGFIEIG	ISQEPFDPSPG	100
Br1/87	W L L		T QST			MNSSS	GTG E		Y DS Q		100
CV777	W L L		T QST			MNSSS	GTG E		Y DS Q		100
Chinju99	YQLYLHKATN	<u>GNTNATARLR</u>	ICQFPSSKTL	GPTANDVTG	RNCLFNKAIP	AHMSDGKHSV	VGITWDNDRV	TVFADKIYHF	YLKNDWSRVA	TNCYNSGGCA	200
Br1/87			DN V			Y R DI				R RRS	200
CV777		I	DN V			Y R DI				R RRS	200
Chinju99	MQVVEPIYY	<u>MLNVT</u> SAGED	GISYQPCAN	CIGYAANVFA	TESNGHIPEG	FSFNWFLLS	<u>NDSTLFHGKV</u>	VSNQPLLVC	LWAIPKIYGL	GHFFSFNQTM	300
Br1/87	T		Y E T		D		L		L	Q H	300
CV777	T		Y E T		D		L		L	Q H	300
Chinju99	DGVCGATAY	RAPQALRFNI	<u>NDTSVILAEG</u>	SIVLHTALGT	<u>NLSFVCSNSS</u>	DPHKAIFSIP	LGATQVPYYC	FLKVDTYNST	VYKFPVAVLS	TVREIVITKY	400
Br1/87	AVD E					L A E			L S		400
CV777	AVD E					L A E			L P		400
Chinju99	GDVYVNGFGY	LHLGLLDAVT	<u>INFTHGHTND</u>	DVSGFWTIAS	TNFDALIEV	QATAIQRIY	CDDPVSQLKC	SQVSDLDG	FYPISRNLL	SHEQPISFVT	500
Br1/87			Y D		V	G S		A			500
CV777			D		V	G S		A			500
Chinju99	LPSFNDHSFV	<u>NITVSAAFGD</u>	SGGANLVASD	TTINGFSSFC	VDTRQYITRL	FYNVTNSYGY	VSKSQDSNCP	FTLQSVNDYL	SFSKFCVSTS	LLAGACTIDL	600
Br1/87		G LSS			F T						600
CV777		G LSS			F T						600
Chinju99	FGYPAFGSGV	NYRSLYQFT	KGEWITWAPK	PLNGLLKGGF	MTLDVCTKYT	IYGFKGEGII	<u>TLTNSS</u> LAG	VYTSDSGQL	LAFKNVTSGA	VYSVTPCSFS	700
Br1/87		KLT	L GT	E ITDVS							700
CV777		KLT	L GT	E ITDVS							700
Chinju99	EQAAVNDI	VGVISSLSNS	<u>TFNNTRELPG</u>	FFYHSNDGSN	<u>CTEPLVLYSN</u>	IGVCKSGSIG	YVPSQYGQVK	IAPTVTGNIS	<u>IPTNFMS</u> IR	TEYLQLYNTP	800
Br1/87											800
CV777											800
Chinju99	VSDCATYVC	NGNSRCKQLL	TQYTAACKTI	ESALELSARL	ESVEVNSMLT	ISEEALQLAT	ISSFNGDGYN	<u>FTNVLGVS</u> VY	DSQSGRVIHE	RSFIEDLLFN	900
Br1/87				Q				A PA	VQK V		900
CV777				Q				A PA	VQK V		900
Chinju99	KVVTNGLGTV	DEDYKRCNSG	RSVADKVCAQ	YYSGLMVLPG	VVDAEKLHMY	SASLIGGMAL	GGFTAVALP	FSYAVQARLN	YLALQTDVLQ	RNQQLLAESF	1000
Br1/87			L V				I A				1000
CV777			L V				I A				1000
Chinju99	<u>NSAIGNITSA</u>	FASVKEAISQ	TSKGLNTVAH	ALTKVQEVVN	SQGSALTQLT	IQLQHNFAQI	SSSTDDIYSR	LDILSADVQV	DRLITGRLSA	LNAFVAQTLT	1100
Br1/87		E			N V		I	L			1100
CV777		E			N V		I				1100
Chinju99	KYTEVQASRK	LAQKQVNECV	KSQSQRYGFC	GGDGEHIFSL	VQAAPQGLLF	LHTVLVPGDF	VNVIAIAGLC	VNGDIALTLR	EPGLVLVTHE	LQTYTATEYF	1200
Br1/87							L	E	F		1200
CV777							L	E	F		1200
Chinju99	VSSRRMFEPK	KPTVSDFVQI	<u>QSCVVTYVNL</u>	<u>TSDQLPDVIP</u>	DYIDVNKTLD	EILASLPNRT	<u>GNPLPLDVFN</u>	<u>ATYLNLTGEI</u>	ADLEQRSESL	<u>RNTTEELRSL</u>	1300
Br1/87			E				S				1300
CV777			E				S				1300
Chinju99	<u>IYNINNTLVD</u>	LEWLNRVETY	IKWPPWVWLI	IFIGLIFVVS	LLVFCISTG	CCGCCGCCGA	CFSGCWRGPR	LQPYEGFEKV	HVQ*		1383
Br1/87	N			V V			C	A			1383
CV777	N			V V			C	A			1383

Fig. 3. Comparison of the amino acid sequence deduced from Chinju99 S gene with that of Br1/87 and CV777 strains; only amino acids of Br1/87 and CV777 which mismatched those of Chinju99 are included; *, translation termination; 27 potential sites for asparagine (N)-linked glycosylation are underlined.

data available from GenBank and the European Molecular Biology Laboratory (EMBL). S gene nucleotide and amino acid sequences of Chinju99 were compared with Br1/87 [14] (EMBL accession No. Z25483) and CV777 (GenBank accession No. AF353511).

Results

Cloning of the S Gene

In the production of ds-cDNA of the Chinju99 S gene, three overlapping DNA fragments were amplified by

Fig. 4. Hydropathicity of Chinju99 amino acid sequence: a stretch of highly hydrophobic residues was found at position 1325–1348 (>1.6 on the Kyte–Doolittle scale).

RT-PCR using appropriate pairs of sense (SF) and antisense (SR) primers. The DNAs, designated as Sfrag1 (1.5 kb), Sfrag2 (1.8 kb) and Sfrag3 (1.4 kb), were each cloned into pUC19 vector DNAs (Fig. 1), and sequenced.

Analysis of S Gene Nucleotide and Amino Acid Sequences

The nucleotide sequence encoding the entire Chinju99 S gene was 4162 bases in length and contained a single 4152 bases long ORF starting with an initiator ATG at nucleotide 11 and ending with a terminator TGA at 4160. The sequence GUAAAC, found at 8 nucleotides upstream of the ATG, was considered part of the leader sequence preceding the ORF. The coding region of the gene had 229 nucleotide mismatches compared to Br1/87 and CV777, respectively (Fig. 2). It consisted of 1001 adenine (24.1%), 849 cytosine (20.4%), 877 guanine (21.1%) and 1425 thymine (34.3%) nucleotides, and a GC content of 41.5%. The gene showed 94.5% nucleotide sequence homology to that of Br1/87 and CV777, respectively.

The putative translation product of the PEDV Chinju99 S gene consisted of 1383 amino acids starting at the ATG at nucleotide 11 and ending at the terminator TGA at nucleotide 4160. There were 27 potential asparagine (N)-linked glycosylation sites recognized in the protein. The Chinju99 S protein, had 100 and 99 amino acid mismatches compared to those of Br1/87 and CV777, respectively (Fig. 3).

There was also a stretch of highly hydrophobic residues at position 1325–1348 (>1.6 on the Kyte–Doolittle scale) (Fig. 4). The Chinju99 S protein showed 92.8% amino acid sequence identity with that of Br1/87 and CV777, respectively.

Discussion

The S gene of the Korean Chinju99 isolate of PEDV was cloned as a series of three overlapping cDNA clones, and the resulting sequence data revealed a single large ORF of 4152 nucleotides which encodes 1383 amino acids to the terminator TGA. Previous studies showed the same lengths for the nucleotide and deduced amino acid sequences of the S gene in Br1/87 [13] and CV777. A single ORF of 4149 nucleotides was identified with the potential to produce a coronavirus S protein [18]. The Chinju99 S gene also had a sequence GUAAAC at 8 nucleotides upstream of the initiator ATG as previously recognized in Br1/87 [13]. This is a common hexameric motif in coronaviruses and similar to the motifs XUA(A/G)AC found adjacent to other PEDV ORFs [12]. The hexameric motifs were suggested as the site for transcription of the subgenomic mRNAs [18].

The Chinju99 S protein contained 27 potential N-linked glycosylation sites and a region of hydrophobic residues at position 1325–1348 was assumed to function as a membrane anchor. Similarly, the

Br1/87 S protein [13] had 29 potential N-linked glycosylation sites and a hydrophobic stretch at 1322–1337. Consequently, the conformational features of the Chinju99 S protein are probably highly conserved with those of Br1/87. Nevertheless, the Chinju99 S gene showed 94.5% nucleotide sequence homology and 92.8% amino acid sequence identity with those of both Br1/87 and CV777, respectively, although the ratios of adenine (24.0–24.1%), cytosine (20.4%), guanine (21.1%) and thymine (34.3–34.4%) were identical or very similar among the three strains. Accordingly, the Chinju99 S protein had only minimal differences in overall structure compared to that of Br1/87 and CV777. Nevertheless the minor differences in the structural features of the S protein could help in elucidating aspects related to both molecular pathogenesis and antigenic structures of PEDV isolates in relation to prevention and control of the disease.

This is the first published report on the nucleotide sequences of the S gene of any Korean PEDV isolate. In the present study, the complete nucleotide sequence of the Chinju99 S gene was determined, and its amino acid composition was analyzed in comparison to other PEDVs. The nucleotide sequence of the Chinju99 S gene can form the basis for further studies on the development of genetically engineered vaccines and diagnostic reagents for PEDV isolates in Korea.

Acknowledgement

This study was supported by a grant (No. 981-0613-065-2) from the Korea Science and Engineering Foundation (KOSEF), Ministry of Science, Korea.

References

1. Pensaert M.B., and Debouck P., *Arch Virol* 58, 243–247, 1978.
2. Chasey D., and Cartwright S.F., *Res Vet Sci* 25, 255–256, 1978.
3. Turgeon D.C., Morin M., Jolette J., Higgins R., Marsolais G., and DiFranco E., *Canadian Vet J* 21, 100–101, 1980.
4. Horvath I., and Mocsari E., *Arch Virol* 68, 103–113, 1981.
5. Takahashi K., Okada K., and Ohshima K., *J Jpn Vet Sci* 45, 829–832, 1983.
6. Kwon C.H., Kwon B.J., Jung T.S., Kee Y.J., Hur D.H., Hwang E.K., Rhee J.C., and An S.H., *Korean J Vet Res* 33, 249–254, 1993.
7. Timoney J.F., Gillespie J.H., Scott F.W., and Barlough J.E., *Hagan and Bruner's Microbiology and Infectious Diseases of Domestic Animals*, Comstock Publishing Associates, Ithaca, 1988, pp. 897–898.
8. Bridgen A., Duarte M., Tobler K., Laude H., and Ackermann M., *J Gen Virol* 74, 1795–1804, 1993.
9. Pensaert M.B., in Straw B.E., D'Allaire S., Mengeling W.L., Taylor D.I. (eds), *Diseases of Swine*, The Iowa University Press, Ames, IA, 1999, pp. 179–185.
10. Murphy F.A., Gibbs E.P.J., Horzinek M.C., and Studdert M.J., *Veterinary Virology*, Academic Press, San Diego, 1999, pp. 496–501.
11. Egberink H.F., Ederveen J., Callebaut P., and Horzinek M.C., *Am J Vet Res* 49, 1320–1324, 1988.
12. Knuchel M., Ackermann M., Muller H., and Kihm U., *Vet Microbiol* 32, 117–134, 1992.
13. Duarte M., and Laude H., *J Gen Virol* 75, 1195–1200, 1994.
14. Duarte M., Tobler K., Bridgen A., Rasschaert D., Ackermann M., and Laude H., *Virology* 198, 466–476, 1994.
15. Kocherhans R., Bridgen A., Ackermann M., and Tobler K., *Virus Genes* 23, 137–144, 2001.
16. Hofmann M., and Wyler R., *J Clin Microbiol* 26, 2235–2239, 1988.
17. Ausubel F., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A., and Struhl K., *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 1992.
18. Lai M.M.C., *Ann Rev Microbiol* 44, 303–333, 1990.