VIROLOGICA SINICA, June 2009, 24 (3):179-186
DOI 10.1007/s12250-009-2982-y

CLC number: R373 Document code: A Article ID: 1674-0769 (2009) 03-0179-08

Cloning the Structure Genes and Expression the N Gene of Porcine Epidemic Diarrhea Virus DX*

Jian-qiang LI [1,2], Ji-xing LIU [2**], Xi LAN [2], Jie CHENG [1], Run WU [1], Zhong-Zi LOU[2]
Xiang-ping YIN[2], Xue-rui LI[2], Bao-yu LI[2], Bin YANG[2] and Zhi-yong LI[2]

(1. College of Veterinary Medicine, Gansu Agricultural University , Lanzhou 730070， China; 2. Key laboratory of Animal Virology of Ministry of Agriculture, State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046,China)

Abstract: The structure genes spike (S), nucleocapsid (N), membrane (M), small membrane (sM) of a porcine epidemic diarrhea virus (PEDV) strain DX isolated in Gansu province, North-west of China, were cloned, sequenced and compared with published sequences of PEDV strains. The nucleotide sequences encoding the entire S, sM, M and N genes open reading frame (ORF) of DX were 4 152, 231, 681 and 1 326 bases long respectively. There were transcription regulatory sequences (TRSs) upstream of the initiator ATG of the S, N and M genes. The amino acids sequences of S, M and N contained 30, 3 and 7 potential asparagine (N)-linked glycosylation sites. Homologous analysis and phylogenetic trees showed that DX had the closest relationship with strains LJB/06, JS-2004-2Z and CH/HLJH/06 that were also isolated from China and indicated the prevalence of some PEDV isolates in China were widespread since the JS-2004-2Z strain originated from the south of the China, and LJB/06 and CH/HLJH/06 were isolated from northeast China. The N gene was cloned using two primers which contained *Nco* I and *Bam*H I restriction enzyme sites and subcloned into expression vector pET30a. The recombinant plasmid was then transformed into *E.coli* Rossta. SDS-PAGE showed there was a protein of about 55kDa as expected and Western blot indicated the N protein had biological activity.

Key words: Cloning, Structure genes, N gene, Expression, PEDV

Porcine epidemic diarrhea virus (PEDV) causes acute infection in piglets 1-2 weeks old, and the disease is characterized by severe enteritis and diarrhea, frequently leading to death with mortality up to 90% (17). It was first reported in England in 1971 and has since been reported in many countries (4, 9, 13, 15); it was first reported in China in 1976 (3) and is now prevalent all over the country causing serious economic losses. PEDV is a in a number of coronavirus genuses and the family coronaviridae. The genome consists of a single molecule of positive-sense, single stranded RNA, 27-32kb in size, which is transcribed into

Received: 2008-07-31, Accepted: 2009-01-08
* Foundation item: National Basic Research Program (2004CCA00500); National High-tech Development Research Program of China (2006AA02Z440)
** Corresponding author.
Phone: +86-931-8342682, E-mail: liujixing@hotmail.com

several subgenomic mRNAs for the production of structure and non-structure proteins (6, 18).

The coronaviruses possesses four major structural proteins including the spike protein (S), the membrane protein (M), the small membrane protein (sM) and the phosphorylated nucleocapsid (N) protein. The S glycoprotein makes up the large surface projections of the virion and plays an important role in the attachment of viral particles to the receptor of the host cell with subsequent penetration to the cell by membrane fusion; it also stimulates induction of neutralizing antibodies in the host (1, 5). The M and sM proteins are essential for viral envelope formation and release, the M protein also can stimulate the production of interferon (IFN) (1,7). The N protein participates in transcription of the viral genome, the formation of the viral core, and packaging of viral RNA. In the early stage of PEDV infection, the pig produces high levels of antibodies against the protein N. Since the N protein is highly conservative in the coronaviruses, it has a good response and immunogenicity, so it is the best candidate protein for early diagnosis reagents and vaccine development.

Although some nucleotides sequences of PEDV isolated from China has been reported, the data of DX structure genes are useful for furthering the study of the molecular biology of PEDV strains that are prevalent in China, especially in the north-west. In this study, the structure genes have been cloned and the N protein has been expressed.

MATERIALS AND METHEODS

Viruses and RNA Extraction

The PEDV DX strain was collected from the feces of piglets suffering from severe diarrhea in Gansu north-west China. Total RNA was isolated from purified feces samples and extracted using a RNA extraction kit (Qiagene, Germany) following the manufacturer's instructions.

Cloning the structure genes

The RT-PCR amplifications were carried out using five primer sets (Table 1) and an RT-PCR amplification Kit (Toyobo, Japan). The products were ligated with the pMD18-T vector (TaKaRa) and transformed into the competent *E. coli* JM109. Positive clones were sequenced by the TaKaRa Biotechology (Dalian) Co.Ltd.

Phylogenetic analysis

Phylogenetic analysis was performed for the amino acids sequence of PEEV DX strain structure genes and compared to the PEDV reference strains retrieved from GenBank. The sequence data were aligned using

Table 1. Amplification primers for the S、sM、M and N genes

Primers	Position[a]	Sequence (5'-3')	Target gene	*length*[b]
S1S	20599	CATTTGTGGCTTTTCTAATC	S gene	2106
S1R	22705	AGCACCACTAGTGACATTCTT		
S2S	22663	GATTCTGGACAGTTGTTAGC		2182
S2R	24845	CTTCGAGACATCTTTGACAAC		
sMS	24581	ACCGAGTTGAGACATACA	sM gene	1121
sMR	25702	GGAATAGAACCGTTAGACAT		
MS	25625	AGTCTTACATGCGAATTGACC	M gene	765
MR	26390	AGCTGACAGAAGCCATAAAGT		
NS	26315	CCGAGTGCGGTTCTCACAGAT	N gene	1416
NR	27731	CATAGCCAGGATAAGCCGGTC		

[a] Position is relative to the PEDV CV777 strain. [b] Length of PCR products.

the Clustal W method (Lasergene program version 7.1, DNASTAR Inc. Madison, WI). Phylogenetic trees were constructed by the bootstrap N-J method in the DNASTAR program.

Construction of the expression plasmid

Two PCR primers of PEDV N, which contain specific restriction enzyme digestion sites depending on the multiple cloning sites contained in the expression vector pET30a were used. The sense primer was 5'- CGCCATGGCGCTTCTGTCAGTTTCA -3', the bases underlined are the enzyme digestion site of *Nco* I; the antisense primer was 5'-CGGGATCCTTA ATTTCCTGTGTCGGA-3', the bases underlined are the enzyme digestion site of *Bam*H I. The PCR product was ligated with the pET30a vector and the ligation product was transformed into Rosetta competent cells. The recombinant was named pET30a-pN and identified with both restriction enzyme digestion and PCR amplification.

SDS-PAGE and western blotting

The positive recombinant transformant was grown in LB media containing 100μg/mL Amp while shaking at 220 r/min at 37℃ and then induced with IPTG. Cells were harvested by centrifugation at 12000 r/min for 1 min. Total cellular pellets were analyzed by 10% SDS-PAGE and Western blotting.

RESULTS AND DISCUSSION

Sequencing analysis

Using RT-PCR, two overlapping products of the S gene of approximately 4.2 kb were amplified (Fig.1), the products of the N, M and sM genes of about 1.4 kb, 1.1kb and 0.75 kb respectively were also amplified (Fig.2). The complete nucleotide sequence of the PEDV DX strain structure genes has been deposited in GenBank under accession number is EU031893. Analysis of the structure gene of DX showed that all the structure genes had a low G+C content; this phenomenon had been observed in several enteric viruses.

The nucleotides sequence encoding the entire DX S gene was 4152 bases in length and coding for a protein of 1383 aa. Analysis showed the same lengths in the nucleotides and deduced amino acids sequences of the S gene in CV777, Br1/87, LZC, JS – 2004 – 2 and Chinju99. The S gene of two Korean isolates (accession numbers AF500215 and AF237764) and

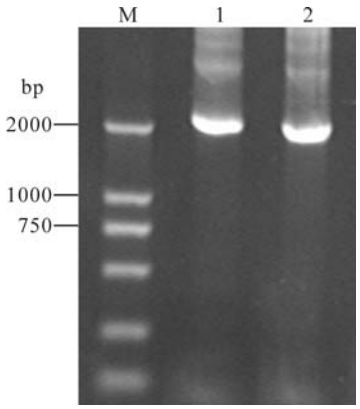

Fig. 1. PCR products of PEDV S gene.1, S1 product (2 182 bp); 2, S2 product (2 106 bp); M, Marker DL 2000.

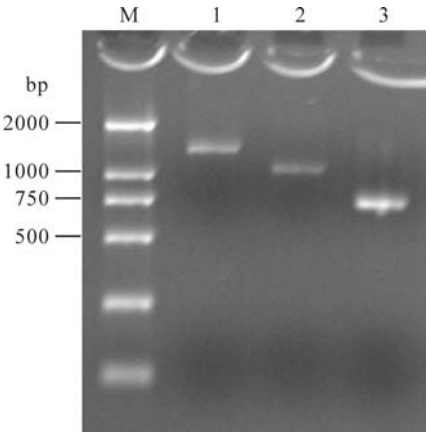

Fig. 2. PCR products of PEDV N, sM and M gene. 1, N product (1 416 bp); 2, sM product (1 121 bp); 3, M product (765 bp); M, Marker DL 2000.

the Chonju sample that was also from Korean had a ORF with 4161 nucleotides which encoded 1386 amino acids(10). This indicated that a group of PEDV strains in Korean existed that was different with other strains due to variation that was probably due to natural selective pressure. Like other PEDV strains, the sequence GUAAAC were found at 8 nucleotides upstream of the initiator ATG of the DX S gene, a similar sequence can also found in other coronavirus upstream of the initiator of the S gene. Common with other PEDV S protein sequences, the DX S protein has a GxCx motif and a conserved nonamer at the S1-S2 junctional region (Fig. 1). Between these two motifs there is a protein restriction site, such that the S protein can be cut into S1 and S2 in the course of virus of Coronavirus II infection which can enhance the virus infectivity. The DX S protein contains 30 potential asparagine (N)-linked glycosylation sites and a region of hydrophobic residues at positions 1327-1349 which is assumed to function as a membrane anchor similarly, the Chinju99 and Br1/87 contained 27 and 29 potential asparagine (N)-linked glycosylation sites and hydrophobic regions between positions 1327-1349 and 1322-1337 (19).

The results revealed that the structure N, M and sM genes had an ORF coding for a protein of 441, 226 and 57 amino acids respectively. Sequence analysis showed that all the reference strains in this study had the same lengths for the DX sequence for nucleotides and amino acids N, M and sM. The N and M protein of DX contained 7 and 4 potential asparagine (N)-linked glycosylation sites, the N protein of Chinju99 and LJB/03 had same number of potential asparagine (N)-linked glycosylation sites as DX (8,11). The sequence AGAAAC and AUAAAC were also found at 9 and 11 nucleotides respectively upstream of the initiator of the N and M gene of the PEDV DX strain, Coronaviruses have transcription regulatory sequences (TRSs) that include a highly conserved core sequence (CS, previously named the intergenic sequence [IS]) 5′-CUAAAC-3′, or a related sequence, depending on the coronavirus, at sites immediately upstream of most of the genes (14). Although TRSs are conserved among the TGEV (Transmissible gastroenteritis viurus) that also belong to the coronavirus group I, mutation frequently occurs in PEDV TRSs, especially in the first two nucleotides of the TRSs and no TRS was found upstream the sM gene of the DX strain.

Phylogenetic analysis

To analyze the phylogenetic relationships between DX and other PEDV strains isolated in various parts of the world, we constructed 3 neighbor-joining phylogenetic trees (Fig.2, Fig.3, Fig.4) using the structure gene amino acid sequences. The tree based on the S gene indicated that all the PEDV formed 3 groups: a group of Chinese samples, a group comprising Korean isolates and a group of English isolates, all the Chinese PEDV strains except LZC formed one group, the LZC strain and CV777 strain formed the English group. The tree based on the N gene showed that DX had a close relationship with LJB/06 and JS-2004-2Z formed the Chinese group. The tree based on the M gene showed that all the PEDV strains formed 6 groups. The DX had a close relationship with the 4 strains from Thailand and LJB/06 and JS-2004-2Z formed one group; other Chinese strains belonged to 4 other groups, indicating the prevalence of PEDV were widespread in China. All the trees showed that DX had a close relationship

with LJB/06, JS-2004-2Z and CH/HLJH/06. The structure gene homology analysis showed that DX were highly homologous with the LJB/06, JS-2004-2Z and CH/HLJH/06 (Table 2) sequences and indicated the prevalence the PEDV group in China was widespread because the JS-2004-2Z came from the south of the China but LJB/06 and CH/HLJH/06 were isolated from the northeast China.

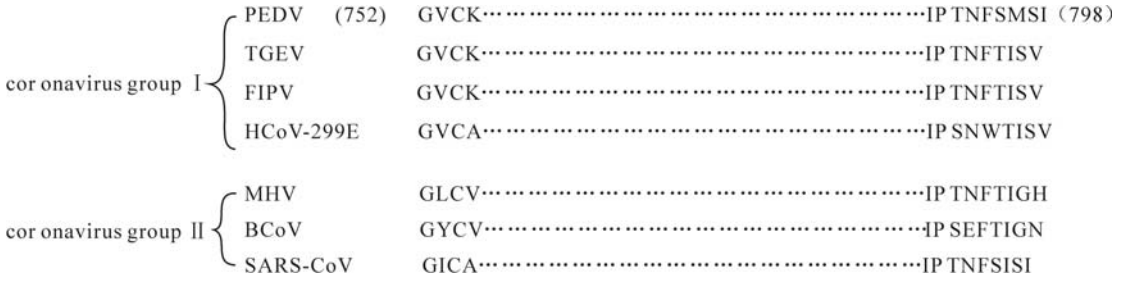

Fig. 3. Sequence analysis of GxCx motif and t he conserved nonamer region based on partial sequence of the Coronavirus spike protein. porcine transmissible gastroenteritis virus (TGEV, DQ200417), feline infectious peritonitis virus (FIPV, AY994055), human respiratory coronavirus 229E (HCoV-229E, DQ243986), murine hepatitis virus (MHV, NC-006852), bovine coronavirus (BCoV, NC003045), Severe Acute Respiratory Syndrome Virus (SARS/AAP41036)

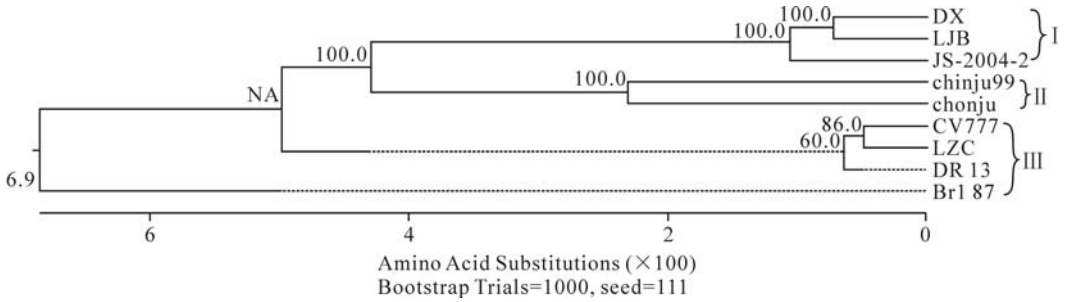

Fig. 4. Phylogenetic relationship of PEDV DX and other PEDV strains based on a comparison of S amino acid sequences. The GenBank accession number for the S genes of Chonju (Korean), Chinju99 (Korean), JS-2004-2Z (China), CV777 (England), Br1/87 (England), DR 13 (Korean) and LZC (China) are AF237764, AF237764 , AY653204, AF353511, Z25483, DQ 862099 and EF 185992 respectively.

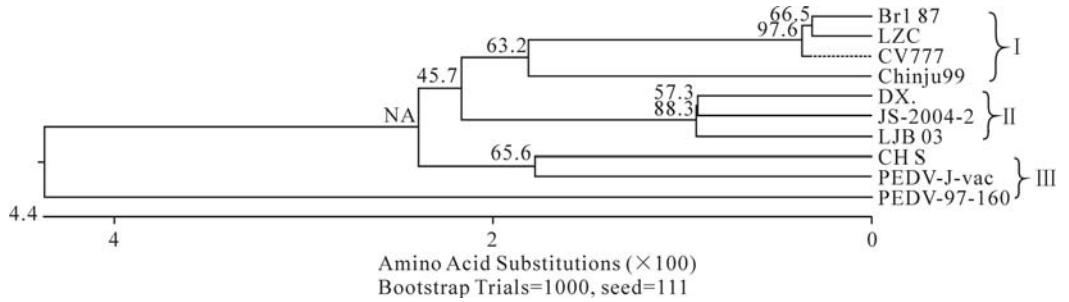

Fig. 5. Phylogenetic relationship of PEDV DX and other PEDV strains based on a comparison of N amino acid sequences. The GenBank accession number for the N genes of CH/S (China), Chinju99 (Korean), JS-2004-2Z (China), CV777 (England), LJB/03 (China), PEDV-J-vac (Japan), PEDV-97-160 (Korea) and LZC (China) are DQ35522, AF237764, AY653205, AF353511, DQ072726, EF 628251, EF628250 and EF 185992 respectively.

Table 2. Comparison of the nucleotides and amino acids homology of Structure Genes for strain PEDV DX compared to other strains (%)

Virus	S		sM		M		N	
	nt	aa	nt	aa	nt	aa	nt	aa
LZC	95.4	94.8	97.4	97.4	97.8	96.9	96.0	95.7
CV777	96.0	95.7	98.3	100	98.8	98.7	96.2	96.2
LJB/03	99.1	98.6	--	--	--	--	--	--
JS-2004-2	98.1	98.1	--	--	98.8	99.6	98.5	98.2
Chinjin99	92.9	91.3	97.8	98.7	98.5	98.2	95.4	95.7
QH	--	--	--	--	97.9	98.7	--	--
CH/JSX/06	--	--	--	--	99.4	98.6	--	--
J-Vac	--	--	--	--	--	--	94.6	93.1
CH/HLJH/06	--	--	--	--	99.6	99.6	--	--

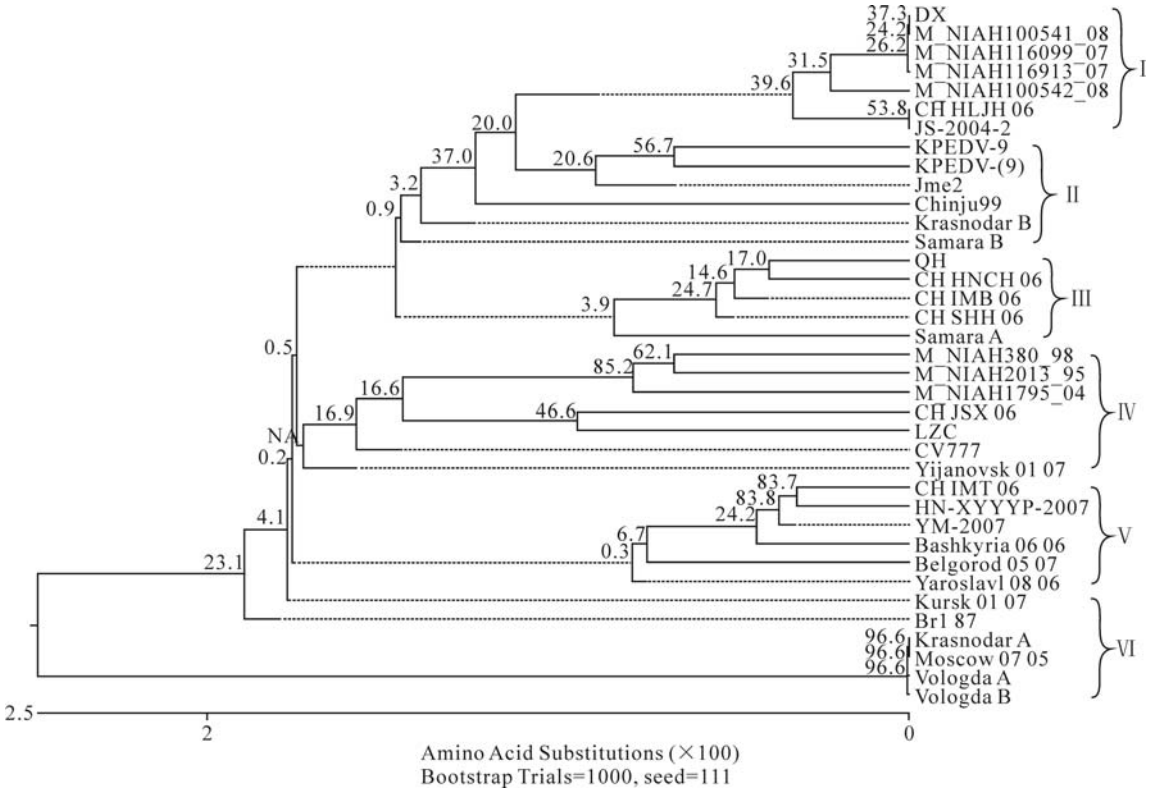

Fig. 6 Phylogenetic relationship of PEDV DX and other PEDV strains based on a comparison of M nucleotide sequences. The GenBank accession number for the M genes of QH (China), KPEDV-9 (Korean), JS-2004-2Z (China), CV777 (England), JMe2 (Japan), PEDV-97-160 (Korea) and LZC (China), CH/SHH/06 (China), CH/IMT/06 (China), CH/HNCH/06 (China), CH/IMB/06 (China), YM-2007(China), HN-XYYYP (China), M-NIAH380-98 (Thailand), M-NIAH2013-95 (Thailand), M-NIAH116913-07 (Thailand), M-NIAH100542-08(Thailand), M-NIAH116099-07 (Thailand), M-NIAH100541-08 (Thailand), M-NIAH1751-04 (Thailand), Moscow/07/05 (Russia), Vologda/A (Russia), Belgorod/05/07 (Russia), Yljanovsk/01/07 (Russia), Kursk/01/07 (Russia), Yaroslawl/08/06 (Russia), Aamara/B (Russia), Saman/A (Russia), Bashkyria/06/06 (Russia), Krasnodar/B (Russia), Krasnodar/A (Russia), Vologda/B (Russia) are DQ35522, AF237764, AY653205, AF353511, D89752, EU033964, EU033967, EF628251, EU033966, EU033965, EU033963, EU033962, EU302820, EU287429, EU581712, EU581711, EU542419, EU542418, EU542417, EU542416, UE542415, EU167541, EU179731, EU179730, EU179729, EU179728, EU179727, EU179726, EU179725, EU179724, EU179723, EU179722 and EU179721, respectively.

Expression of N gene

The recombinant plasmid pET30-PN was digested with *Nco*l I and *Bam*H I producing two fragments of 5 400bp (pET30a) and 1 400bp. PCR identification confirmed the presence of the 1 400bp fragment (Fig.6). To obtain the expressed recombinant protein, the time course expression of recombinant plasmid pET-PN was induced by IPTG with ranging from 0.5-1 mmol/L at 2, 3, 4, 5 and 6 h respectively. SDS- PAGE revealed that the best condition was the IPTG with 1 mmol/L at 6 h. The expressed proteins had a molecular weight of 55kDa as expected (Fig.7). Western blot showed the expressed protein of PN was able to bind immunologically to anti-PEDV serum by comparison with a positive control of native PEDV protein.

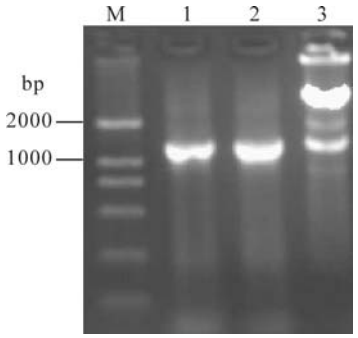

Fig. 7 Analysis of recombinant plasmid pET30-PN. 1, N gene RT-PCR product (1330 bp); 2, N product from recombinant plasmid (1330 bp); 3, Recombinant plasmid digested with the RE; M, a molecular weight Marker DL 2000

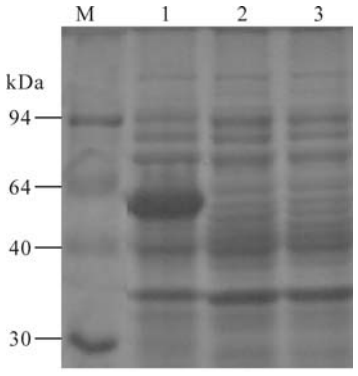

Fig. 8. Epression of recombinant N protein in E.coli by SDS-PAGE. 1, pET32-PN/Rossta induced with IPTG for 6 h; 2, pET32-PN/Rossta untreated with IPTG; 3, pET32-PN/Rossta induced with IPTG; M, Brilliant Blue Plus Pre-Stained Protein Standard

The N protein is a RNA binding protein, it contains basic amino acids such as Arginine and lysine and shows a high degree of alkalinity, it is also the only phosphorylated structure protein in the coronavirus. The Serine residue is a potential phosphorylation site (2). TheDX N protein contained 37 Arginine, 33 lysine and 36 Serine. The N protein is the most highly expressed protein in coronavirus infected cells and is about 45kDa-60kDa (12). The PEDV N protein is about 49kDa-58kDa in different expression systems (2). In this study, an N protein of about 50kDa (the expressed vector was abou 5kDa) was obtained using the *E. coli* system, the result is consistent with other *E. coli* expression system, although there are differences in the observed weights (16). Western blot results showed the expressed N protein reacted with the antibodies, which indicated that the expressed N protein has the biological activity. Our results provide a basis for further development of a diagnosis method.

References

1. **Bos E C, Luytjes W, vander Meulen H V, *et al.*** 1996. The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. **Virology.** 218: 52-60.
2. **Chen J F, Feng L, Shi H Y, *et al.*** 2007. Genetic variation and prolaryotic expression of N protein gene of porcine epidemic diarrhea virus CH/S strain. **Chin J Prev Vet Med,** 29: 856-860. (in Chinese)
3. **Cheng Q H, Niu X Y.** 1992. Investigation on the porcine epidemic diarrhea prevalent on Qinhai, **Chin Qinhai J Anim Vet Sci,** 22: 22-23. (in Chinese)
4. **DeBouck P, Callebaut P, Pensaert M.** 1982. Prevalence of the porcine epidemic diarrhea (PED) virus in the pig population of different countries, **Proc Int Congr Pig Vet**

Soc, 7: 53.

5. **Duarte M, Laude H.** 1994. Sequence of the spike protein of the Porcine epidemic diarrhea virus. **J Gen Virol,** 75: 1195-1200.
6. **Egberink H F, Ederveen J, Callebaut P.** 1988. Characterization of the structural proteins of porcine epidemic diarrhea virus, strain CV777. **Am J Vet Res,** 49: 1320-1324.
7. **Fan J H, Li Y J.** 2005. Cloning and sequence analysis of the M gene of Porcine epidemic diarrhea virus LJB/03**. Virus Genes.** 30: 69-73.
8. **Ge J W, Li B X, Tang L J, *et al.*** 2006. Cloning and sequence analysis of the N gene of porcine epidemic diarrhea virus LJB/03. **Virus Genes** 33: 215-219.
9. **Hofmann M, Wyler R.** 1987, Serologic study of the occurrence of epizootic viral diarrhea in swine in Switzerland, **Schweiz Arch Tierheilkd,** 129: 437.
10. **Kang T J, Seo J E, Kim D H, *et al.*** 2005. Cloning and sequence analysis of the Korean strain of spike gene of Porcine epidemic diarrhea virus and expression of its neutralizing epitope in plants . **Protein Expression and Purification.** 41: 378-383.
11. **Lee H G, Yeo S H.** 2003. Cloning and Sequence Analysis of the Nucleocapsid Gene of Porcine Epidemic Diarrhea Virus Chinju99. **Virus Genes.** 26: 207-212.
12. **Master P S**. 2006. The molecular biology of coronaviruses. **Adv Virus Res,** 66: 193-198
13. **Mostl K, Horvath E, Burki F.** 1990. Erhebungen uber porcine coronaviren in osterreich II porcine epidemic diarrhea virus (PEDV) der schweine. **Wien Tierarztl Monatsschr.** 77: 10-18.
14. **Penzes Z, Gonzalez J M, Calvo E, *et al.*** 2001. Complete Genome Sequence of Transmissible Gastroenteritis Coronavirus PUR46-MAD Clone and Evolution of the Purdue Virus Cluster. **Virus Genes,** 23: 105-118.
15. **Shibata I, Tsuda T, Mori M**. 2000. Isolation of porcine epidemic diarrhea virus in porcine cell cultures and experimental infection of pigs of different ages. **Vet Microbiol,** 72: 1732-1821
16. **Singh M.** 1999. A novel internal open reading frame product expressed from a polycistronic of porcine edpidemic diarrhea virus may not contribute to virus attenuation. **J Gen Virol,** 80: 1959-1963.
17. **Straw B E, Allaire S D, Mengeling W L, *et al*.** 2000. **In: Disease of Swine** (Translated by Zhao D M, Zhang Z Q, Shen J Z). 8th ed. **Beijing:The Chinese Agriculture University Press,** p181-187.
18. **Tzipori S, Spradbrow P B.** 1978. A cell culture vaccine against bovine ephemeral fever. **Aust Vet J Jul.** 54: 323-328.
19. **Yeo S G, Hernandez M, Krell P J.** 2003. Cloning and sequence analysis of the spike gene of Porcine epidemic diarrhea virus Chiju99 .**Virus Genes.** 26: 239-246.
20. **Wang Y, Zhang X.** 1999. The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 *in vitro* and *in vivo*. **Virology** 265: 96-109.