Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the structural domain of the nucleocapsid N protein from porcine reproductive and respiratory syndrome virus (PRRSV)

Danny N. P. Doan* and Terje Dokland

Institute of Molecular and Cell Biology, 30 Medical Drive, Singapore 117609, Singapore

Correspondence e-mail: doan@imcb.nus.edu.sg

Received 31 March 2003
Accepted 9 June 2003

The structural domain of the PRRSV nucleocapsid N protein was overexpressed in *Escherichia coli* and purified to homogeneity. Crystals of the expressed protein, designated His-NΔ_{57}, were obtained by hanging-drop vapour diffusion using PEG 3350 as precipitant at pH 6.5. A native data set from a frozen crystal was collected to 2.7 Å resolution using synchrotron radiation. The crystals belong to space group $P3_121$ or $P3_221$, with unit-cell parameters a = 44.41, c = 125.05, and contain a dimer in the asymmetric unit.

1. Introduction

Porcine reproductive and respiratory syndrome (PRRS) is an economically important disease in pigs, whose symptoms include severe respiratory disorders in piglets and abortion in sows (Dea *et al.*, 2000). PRRS is caused by a virus, often referred to as PRRS virus or PRRSV, which belongs to the family *Arterividae*, order *Nidovirales*. The virus was first identified in North America but has since been detected worldwide in almost all swine-producing countries. The mature virion contains a linear positive-strand RNA genome of approximately 15 kbp encoding eight open reading frames (ORFs; Conzelmann *et al.*, 1993; Mardassi *et al.*, 1995; Meulenberg *et al.*, 1993, 1995, 1996). ORF 1a and ORF 1b code for non-structural proteins required for replication of the viral genome. ORFs 2 to 4 code for minor structural glycoproteins (GP_{2a}, GP_{2b}, GP_3 and GP_4). Major structural proteins including the 25 kDa envelope glycoprotein (E or GP_5), the 19 KDa non-glycosylated membrane protein (M) and the 15 kDa nucleocapsid protein (N) are encoded by ORFs 5, 6 and 7, respectively. The RNA genome is enclosed by a spherical nucleocapsid consisting of the N protein, which is further enveloped by a lipid membrane together with associated GP_2, GP_3, GP_4, GP_5 and M proteins (Benfield *et al.*, 1992; Mardassi *et al.*, 1994; Dea *et al.*, 2000).

The PRRSV N protein is the most abundant protein expressed in infected cells, and comprises up to 40% of total proteins in the mature virion (Mardassi *et al.*, 1994). The protein consists of 123 and 128 amino acids in the North American and the European strains, respectively (Mardassi *et al.*, 1995; Meulenberg *et al.*, 1995). The protein was suggested to be a disulfide-linked homodimer (Mardassi *et al.*, 1996) and its amino-acid sequence can be divided into the RNA-binding domain and the structural domain. The N-terminal half of the protein, rich in basic amino-acid residues such as lysine and arginine, may facilitate interaction of the N protein with the RNA genome. On the other hand, the last C-terminal 11 residues are necessary to maintain conformational integrity of the protein tertiary structure (Wootton *et al.*, 1998, 2001).

In this report, we describe the cloning, expression, purification, crystallization and preliminary diffraction data of the structural domain of the N protein from a PRRS virus strain (North American isolate VR-2332). In brief, the C-terminal half of the protein (residues 58–123) was fused in frame at the N-terminus with a stretch of Met-Ala-His_6 sequence and overexpressed in *E. coli*. The recombinant His-tagged protein, designated His-NΔ_{57}, was purified by affinity and ion-exchange chromatographies and crystallized. A native data set from a frozen crystal was collected to 2.7 Å resolution using synchrotron radiation.

2. Methods

2.1. Molecular cloning and protein expression

DNA fragment coding for the structural domain of the PRRSV N protein (amino-acid residues 58–123) was cloned by polymerase chain reaction (PCR) amplification using the plasmid pGEX-PRRSV7 as template, pfu polymerase (Promega), a forward primer (5′-GATATACCATGGCGCACCATCACCATC-ACCATACTGAAGATGATGTC-3′) and a reverse primer (5′-CTATAGGATCCTCATG-CTGAGGGTGATGC-3′). The PCR product

was digested with restriction enzymes *Nco*I and *Bam*HI and the fragment was inserted into pET14b vector and transformed into *Escherichia coli* BL21 (DE3) pLysS (Novagen). DNA sequencing was performed to verify that the cloned DNA contained no incorrect bases or mismatches.

Cells were grown at 310 K in LB medium supplemented with 100 µg ml^{-1} ampicillin until OD_{600} was about 0.5. After addition of isopropyl-β-D-1-thiogalactoside (IPTG) to a final concentration of 1 m*M*, cells were allowed to grow for another 3 h. Cell culture was harvested by centrifugation at 1000*g* for 5 min and the cell pellet collected and stored at 253 K.

2.2. Purification of His-NΔ_{57} protein

The cell pellet from a 1 l culture expressing His-NΔ_{57} protein was resuspended in 20 m*M* phosphate buffer pH 6.8 containing 2 m*M* DTT, passed through a French pressure cell (Spectronic Instruments) and centrifuged at 16 000*g* for 15 min. Ammonium sulfate was added to the supernatant to 65% saturation and the protein solution was incubated at 277 K for 1 h with constant stirring and centrifuged at 16 000*g* for 15 min. The pellet was resuspended in 50 ml 20 m*M* phosphate buffer pH 6.80 containing 500 m*M* NaCl and 10 m*M* imidazole and centrifuged to remove insoluble material. The protein solution was applied onto a 5 ml metal-affinity column pre-charged with nickel chloride (Hitrap Chelating HP, Amersham Bioscience). After washing the column with 50 ml 20 m*M* phosphate buffer pH 6.8 containing 500 m*M* NaCl and 10 m*M* imidazole, the His-tagged protein was eluted with a gradient of 10–500 m*M* imidazole prepared in the same solution. Fractions containing the His-tagged protein were pooled, dialyzed against 2 l 20 m*M* HEPES pH 7.0 containing 2 m*M* DTT and loaded onto a 1 ml ion-exchange column (EconoPac high Q, BioRad). After washing with 20 m*M* HEPES pH 7.0 containing 2 m*M* DTT, the His-tagged protein was eluted with a gradient of 0–1 *M* NaCl prepared in the same buffer. Fractions containing the highest amount of protein were pooled, dialyzed to remove NaCl, concentrated to 20 mg ml^{-1} and stored in small aliquots at 193 K (Fig. 1).

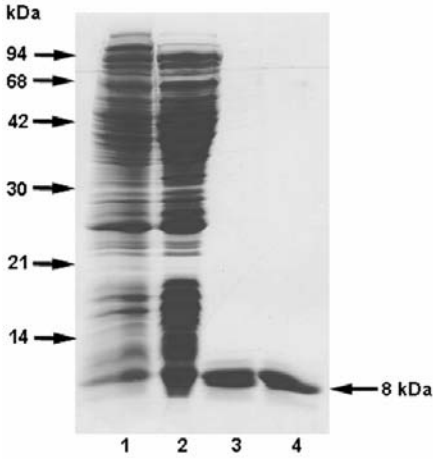

Figure 1
Purification of the recombinant His-NΔ_{57}. Protein samples were separated by 15% SDS–PAGE: lane 1, crude extract; lane 2, after ammonium sulfate precipitation; lane 3, after affinity purification; lane 4, after ion-exchange chromatography.

Table 1
Crystallographic parameters and data-collection statistics.

Values in parentheses correspond to the outermost resolution shell (2.8–2.7 Å).

X-ray source	BW7A, EMBL Hamburg
Wavelength (Å)	0.9202
Temperature (K)	100K
Resolution (Å)	20.0–2.7 (2.8–2.7)
Space group	$P3_121$ or P_221
Unit- cell parameters (Å)	a = 44.41, c = 125.05
Matthews coefficient (Å^3 Da^{-1})	2.2
Solvent content (%)	43
Total observations	81837
Independent reflections	4370
Completeness (%)	99.9 (100.0)
$I/\sigma(I)$	37.5 (9.4)
R_{merge}	8.8 (32.1)

The molecular weight of the purified His-NΔ_{57} was determined using a MALDI–TOF mass spectrometer (API 300 MS/MS, Applied Biosystems) at the Proteomics and Protein Centre, Department of Biological Sciences, National University of Singapore.

2.3. Crystallization

Crystallization was performed using a 24-well plate (Hampton Research) by hanging-drop vapour diffusion at 288 K with 1 ml of the mother-liquor solution containing 100 m*M* MES pH 6.5 and 8%(*w*/*v*) PEG 3350. The protein solution was mixed with an equal volume of the mother liquor to a final volume of 10 µl and a protein concentration of 10 mg ml^{-1} (McPherson, 1999). Hexagonal rod-shaped crystals appeared after 3 d and continued to grow for up to two weeks.

2.4. Data collection and processing

Protein crystals were soaked for 5 min in the reservoir solution containing an additional 10%(*v*/*v*) glycerol as cryoprotectant and quickly frozen in liquid N_2. Diffraction data (λ = 0.9202 Å) were collected at 100 K using a MAR CCD 165 image-plate detector at the BW7A beamline, EMBL Hamburg Outstation, German Synchrotron Research Centre (DESY). The crystal was rotated through 360° with an oscillation angle of 1° per frame and an exposure of 2 min per image. Diffraction data were processed and scaled using *DENZO* (Otwinowski & Minor, 1997). Data collection statistics are given in Table 1.

3. Results and discussion

The first half of the PRRSV N protein is rich in basic amino-acid residues and has been implicated in binding to viral RNA. This region is likely to be highly flexible and might prevent the protein from being crystallized. Therefore, the recombinant His-NΔ_{57} protein was expressed as a truncated protein with the first N-terminal 57 amino-acid residues removed. To facilitate protein purification by affinity chromatography, the expressed protein also contains a stretch of six histidines fused at the N-terminus. The purified recombinant protein was readily crystallized in buffer solutions of pH 5.0 and higher and in the presence of polyethylene glycol (PEG) of different molecular weights including 1500, 3350, 4000, 6000, 8000 and 20 000. However, crystal growth appeared to be optimal at pH 6.5 with PEG 3350.

The protein crystals have the appearance of hexagonal rods with an average length of 0.26 mm and a diameter of 0.07 mm (Fig. 2). Preliminary diffraction data indicate that the crystals belong to space group $P3_121$ or $P3_221$, with unit-cell parameters a = 44.41, c = 125.05 Å. This gives a Matthews coefficient value of 2.2 Å^3 Da^{-1} and a solvent content of 43% with a dimer in the asym-

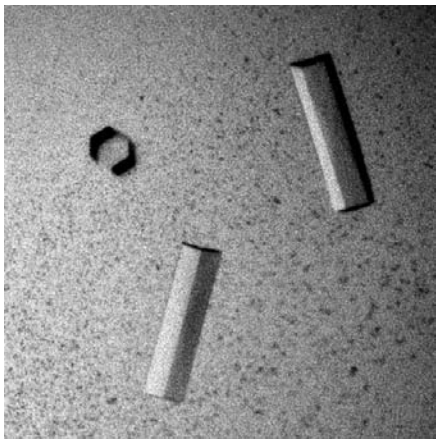

Figure 2
Crystals of the His-tagged NΔ_{57}. The crystals formed hexagonal rods with average length 0.28 mm and diameter 0.07 mm.

metrical unit (Matthews, 1968). This finding is in agreement with previous biochemical data which suggested that the PRRSV N protein is a homodimer linked by disulfide bridges (Mardassi *et al.*, 1996).

Prior to crystallization, the mass of the purified protein was verified by MALDI-TOF mass spectrometry. The determined mass (8136 Da) was in agreement with the theoretical molecular weight (8124 Da). This indicates that the *E. coli* expressed His-NΔ_{57} protein is not phosphorylated. However, phosphorylation has been reported in PRRSV nucleocapsid protein isolated from infected mammalian cells or from mature virions (Wootton *et al.*, 2002). Thus, either phosphorylation does not occur at sites in the structural domain of N protein (Wootton *et al.*, 2002) or phosphorylation is not required for the correct folding of the truncated protein.

Soaking protein crystals in heavy-atom solutions resulted in either no diffraction (mercury acetate or mercury chloride) or no incorporation of heavy atoms (sodium bromide, potassium tetracyanoaurate, potassium tetracyanoplatinate, sodium hexachloroiridate or samarium choride). We are in the process of trying alternative approaches such as sulfur SAD phasing or SeMet phasing. Since there are no Met residues in the amino-acid sequence of the truncated PRRSV nucleocapsid protein, site-directed mutagenesis of the cDNA needs to be employed so that SeMet can be incorporated into the expressed protein.

We thank Dr Jimmy Kwang for providing the pGEX-PRRSV7 plasmid and Ms Haoying Xu for performing the DNA sequencing. We are grateful for Drs Paul Tucker, Ehmke Pohl and Cristofer Enroth at the EMBL Hamburg Outstation, German Synchrotron Research Centre (DESY) for their excellent support during data collection.

References

Benfield, D. A., Nelson, E., Collins, J. E., Harris, L., Goyal, S. M., Robison, D., Christianson, W. T., Morrisson, R. B., Gorcyca, D. & Chladek, D. (1992). *J. Vet. Diagn. Invest.* **4**, 127–133.

Conzelmann, K. K., Visser, N., Van Woensel, P., & Thiel, H. J. (1993). *Virology*, **193**, 329–339.

Dea, S., Gagnon, C. A. Mardassi, H., Pirzadeh, B. & Rogan, D. (2000). *Arch. Virol.* **145**, 659–688.

McPherson, A. (1999). *Crystallization of Biological Macromolecules.* New York: Cold Spring Harbor Laboratory Press.

Mardassi, H., Athanassious, R., Mounir, S. & Dea, S. (1994). *Can. J. Vet. Res.* **58**, 55–64.

Mardassi, H., Massie, B. & Dea, S. (1996). *Virology*, **221**, 98–112.

Mardassi, H., Mounir, S. & Dea, S. (1995). *Arch. Virol.* **140**, 1405–1418.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Meulenberg, J. J. M., Hulst, M., De Meijer, E. J., Moonen, P. L. J., Petersen-Den Besten, A., De Kluyer, E. P., Wensvoort, G. & Moormann, R. J. M. (1993). *Virology*, **192**, 62–72.

Meulenberg, J. J. M., Petersen-Den Besten, A., De Kluyer, E. P., Moormann, R. J. M. Schaaper, W. M. M. & Wensvoort, G. (1995). *Virology*, **206**, 155–163.

Meulenberg, J. J. M. & Petersen-Den Besten, A. (1996). *Virology*, **225**, 44–51.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Wootton, S. K., Koljesar, G., Yang, L., Yoon, K.-J. E. & Yoo, D. (2001). *Clin. Diagn. Lab. Immunol.* **8**, 598–603.

Wootton, S. K., Nelson, E. A. & Yoo, D. (1998). *Clin. Diagn. Lab. Immunol.* **5**, 773–779.

Wootton, S. K., Rowland, R. R. R. & Yoo, D. (2002). *J. Virol.* **76**, 10569–10576.