Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of a proteinase-K-resistant domain within the phosphoprotein of vesicular stomatitis virus (Indiana)

Haitao Ding,[a] Todd J. Green[a] and Ming Luo[a,b]*

[a]Center for Biophysical Sciences and Engineering, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA, and [b]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

Correspondence e-mail: mingluo@uab.edu

Received 12 May 2004
Accepted 24 September 2004

Several stable domains of the phosphoprotein (P) of vesicular stomatitis virus (Indiana) were identified by limited proteolysis of purified recombinant P protein expressed in *Escherichia coli*. The proteinase-K-resistant domain could be crystallized using ammonium sulfate as a precipitant and ethylene glycol as an additive. The crystals belong to space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 74.50$, $c = 156.84$ Å. X-ray diffraction data were collected to 2.75 Å resolution at a synchrotron-radiation source.

1. Introduction

Vesicular stomatitis virus (VSV), a member of the *Rhabdoviridae* family, contains a non-segmented negative-strand RNA genome within the mature virion (Banerjee, 1987; Banerjee & Barik, 1992). There are two major serotypes of VSV: New Jersey and Indiana (Martinez *et al.*, 2003). The VSV genome is 11 161 nucleotides long and encodes five virus-specific proteins: the nucleocapsid protein N, the phosphoprotein P, the large multifunctional polymerase subunit L, the matrix protein M and the glycoprotein G in the order 3′ N-P-M-G-L 5′ (Banerjee, 1987). Of these, the M protein (Gaudier *et al.*, 2002) is inside the membrane envelope and the G protein (Schloemer & Wagner, 1975) is outside. The G protein recognizes the host receptor on the surface of the cellular membrane and is responsible for the induction of fusion between viral and cellular membranes during entry. The N protein encapsidates the viral genome to form the ribonucleoprotein complex and associates with the P and L proteins to participate in transcription and replication in virus infection (Banerjee, 1987; Gupta *et al.*, 2003; Qanungo *et al.*, 2004).

The P protein, which acts as the polymerase cofactor, is an essential component in both transcriptase and replicase complexes (Qanungo *et al.*, 2004). The P protein can also bind the newly synthesized N protein to form specific complexes that are required for encapsidating the viral RNA (Das & Banerjee, 1992). Several functional domains of the P protein have been studied previously (Fig. 1): the phosphorylation of the acidic region (Hudson *et al.*, 1986), which causes the oligomerization of P, at the amino-terminus of the P protein by casein kinase-II could activate the P protein for transcription (Gao & Lenard, 1995*a*; Gao *et al.*, 1996; Chen *et al.*, 1997), but the phosphorylation is not required for replication (Pattnaik *et al.*, 1997). The other P-protein domain binds the L protein and the phosphorylation pattern of serine residues within this domain by the L-protein-associated protein kinase modulates the association of the P protein with the L protein and N-RNA template (Chattopadhyay & Banerjee, 1987; Gao & Lenard, 1995*b*; Chen *et al.*, 1997). The highly basic C-terminal 21 residues of the P protein are required for binding to the soluble N protein as well as to the N-RNA template (Hudson *et al.*, 1986; Takacs *et al.*, 1993; Green *et al.*, 2000).

In this study, the full-length P protein of VSV (Ind) as well as several stable domains which were obtained from spontaneous

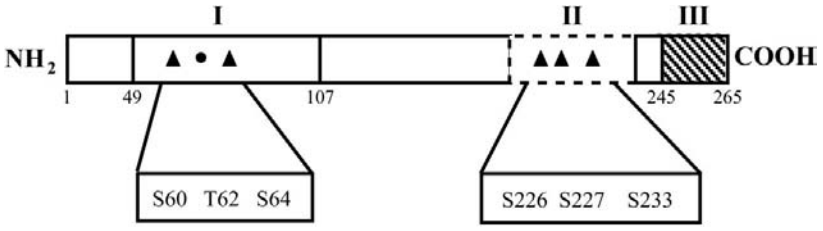

Figure 1
Functional domains of the P protein of VSV (Ind) are shown. Domain I represents the N-terminal negatively charged domain and the solid triangles and dot represent the serines and theronine that are the potential phosphorylation sites. Domain II represents the L-protein-binding domain and the solid triangles represent the serines that are the possible phosphorylation sites. Domain III represents the highly basic C-terminal domain.

degradation or limited proteolysis were expressed, purified and subjected to crystallization screening. Cloning of a P gene fragment encoding a proteinase-K-resistant domain of the P protein (amino acids 107–177, referred to as P3) led to the production of a soluble recombinant protein in *Escherichia coli* at high levels. Purified P3 protein was very stable on storage and has been crystallized. The P3 domain is located within the P hinge region (Pattnaik *et al.*, 1997), for which no structural or biochemical information is available. The X-ray structure of the oligomerization domain of Sendai virus, a prototypical respirovirus of the negative-strand RNA-virus superfamily, has been solved. This domain was shown to have a homotetrameric coiled-coil structure (Tarbouriech *et al.*, 2000). Since the P3 protein has no sequence similarity to the Sendai virus domain, its structure will help to obtain insights into the functions of VSV P protein in viral transcription and replication.

2. Experimental

2.1. Stable-domain identification and cloning

Proteinase-K digestion of the gel-filtration-purified full-length recombinant P protein was carried out using 0.01 unit soluble proteinase-K (Fisher Biotech, USA) per milligram of substrate protein at 310 K for 22 h, rotated in a shaker. Trypsin digestion of the purified full-length recombinant P protein was carried out using 0.2 U immobilized trypsin (Sigma-Aldrich, USA) per milligram of substrate protein at 310 K for 22 h, rotated in a shaker. Sequence information for the fragments of the spontaneously degraded P protein and proteinase-resistant fragments was acquired by protein N-terminal sequencing and mass-spectrometric determination of protein molecular mass (Ding, Qiu, Li *et al.*, 2003). The Polymerase Chain Reaction (PCR) amplified genes for the domains identified above were subcloned into the pET28b vector (Novagen, USA) using *Nde*I–*Bam*HI enzyme-restriction sites.

2.2. Expression and purification

The recombinant plasmids were transformed into *E. coli* BL21 (DE3) strain and the cells were grown in Luria–Bertani broth (Fisher Scientific, Germany) at 310 K until the OD_{600} reached 0.7. Expression of the recombinant proteins was then induced with 1 m*M* isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 5 h at 310 K. The deposited cells were suspended in binding buffer (20 m*M* Tris–HCl, 5 m*M* imidazole, 500 m*M* NaCl pH 7.9) and were disrupted by sonication on ice (Ding, Qiu, Bunzel *et al.*, 2003). The cell debris was removed by centrifugation for 1 h at 15 000 rev min^{-1} and 277 K. The supernatants were loaded onto an $NiSO_4$-charged affinity column (Amersham Biosciences, Sweden) and the recombinant proteins were washed with 50 m*M* imidazole in binding buffer and eluted out with strip buffer (binding buffer containing 100 m*M* EDTA). The size and purity of the proteins were checked by SDS–PAGE. The cleavage of the N-terminal six-histidine tag was conducted at 295 K overnight using 1 U thrombin (Sigma-Aldrich, USA) per 16.7 mg protein in the strip buffer and a fusion peptide Gly-Ser-His remained at the N-terminus of the proteins after the cleavage. The resulting sample was then purified by HiLoad 16/60 Superdex 200 or HiLoad 16/60 Superdex 75 (120 ml, Amersham Biosciences, Sweden) in 20 m*M* HEPES, 150 m*M* NaCl pH 8.0 buffer using the method described in Ding, Qiu, Bunzel *et al.* (2003). Fractions containing the proteins were pooled and exchanged into a buffer consisting of 10 m*M* HEPES, 0.01%(*w*/*v*) sodium azide pH 7.5 prior to use in crystallization screening.

2.3. Crystallization

Crystallization conditions were initially screened using commercially available screening kits (Hampton Research/Emerald Biostructures, USA). Experiments were set up using the hanging-drop vapour-diffusion method in 24-well VDX plates (Hampton Research, USA) at 291 K. 1 µl protein solution mixed with 1 µl of the commercial buffer was equilibrated against 600 µl of the commercial buffer in the reservoir. Crystallization was also manually screened by mixing selected precipitants, additives and detergents with the protein drops and equilibrating against 1000 µl of prepared reservoir solutions in 24-well VDX plates at 291 K.

2.4. Data collection and processing

X-ray diffraction data were collected at SER-CAT beamline 22-ID, Advanced Photon Source, Argonne National Laboratory using a MAR 225 CCD area detector. Data were processed with *HKL*2000 (Otwinowski & Minor, 1997).

3. Results and discussion

The bacterially expressed full-length P protein is not stable. It is degraded in the course of expression and nickel-affinity purification. Based on sequencing information, the N-terminal six-histidine tag of the P protein had been removed prior to the addition of thrombin to cleave off the six-histidine tag (Fig. 2*a*). The elution profile of the P protein from the Superdex-200 column indicated a certain degree of aggregation of the P protein, but the exact number of P proteins in the aggregated complex could not be determined because the elution peaks of the P protein contain mixed products from various degradations (data not shown).

The gel-filtration-purified full-length P protein and its proteinase-digested products were separated by gel electrophoresis. Bands corresponding to the P protein or its fragments (Fig. 2*b*) were sequenced from the

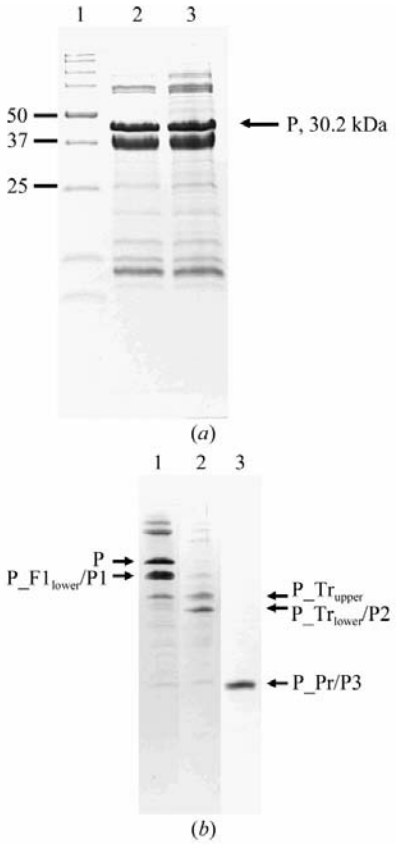

Figure 2
Bacterially expressed P protein and its stable domains were analyzed by Coomassie brilliant blue R-250 stained SDS–15% polyacrylamide gel. (*a*) The N-terminal six-histidine tag of the P protein is not stable. Lane 1, standard molecular-weight markers (kDa). Lane 2, the recombinant P protein purified by a nickel-affinity column from cell-lysis supernatants; the protein-sequencing result indicated the first five residues of the P band to be Gly-Ser-His-Met-Asp. Lane 3, proteins in lane 2 treated with thrombin. (*b*) Identification of stable domains within the P protein. Lane 1, gel-filtration-purified bacterially expressed full-length P protein; lane 2, proteins in lane 1 digested with trypsin; lane 3, proteins in lane 1 digested with proteinase-K.

N-terminus for at least 15 residues and the results all match the P-protein sequence perfectly. The first five residues of the P band are Gly-Ser-His-Met-Asp and the P_Fl$_{lower}$ band has the same N-terminal sequence. In band P_Tr$_{upper}$ the N-terminal sequence is Ser-Tyr-Ser-Arg-Leu, in band P_Tr$_{lower}$ it is Ser-Asn-Tyr-Glu-Leu and in P_Pr it is Asp-Trp-Lys-Gln-Pro. Mass spectrometry showed that the molecular mass of the protein in band P is 30 314, that in band P_Fl$_{lower}$ is 20 350, that in band P_Tr$_{lower}$ is between 16 414 and 16 666 and that in band P_Pr is 8198. No reliable results were obtained for band P_Tr$_{upper}$. These data suggest that the protein in band P is the full-length P protein and the sequences of P_Fl$_{lower}$, P_Tr$_{lower}$ and P_Pr all terminate in a narrow region close to the C-terminal end, but with different starting positions at the N-terminal end (Fig. 3). It seems that the P protein has at least two separated parts connected by a hinge region which is protease-hypersensitive, one part of the protein at the N-terminal end and the other at the C-terminal end. Furthermore, the C-terminal part is not very stable in full-length P, as it was eliminated in all spontaneously degraded fragments.

Based on the analysis shown above, four fragments of the P protein, of which P4 is the C-terminal domain, were subcloned (Fig. 3), expressed and purified for crystallization screening. The recombinant P1 and P2 proteins aggregated to a certain degree in the gel-filtration buffer, as inferred from their gel-filtration profiles. The P1 protein was not stable, as indicated by the presence of degradation products. The P2 protein was eluted from the Superdex-200 column as two peaks of apparent molecular weights 83.2 and 60.3 kDa; for comparison, the molecular weight of P2 is 16.4 kDa. The yield of the recombinant P3 protein in bacterial expression was very high. P3 has a molecular weight of 8.3 kDa and eluted at about 20 kDa from the Superdex-75 column in the gel-filtration buffer. P4 protein with a calculated molecular weight of 9.6 kDa eluted at about 12 kDa from the Superdex-75 column. P3 and P4 proteins were very stable when stored at 277 K.

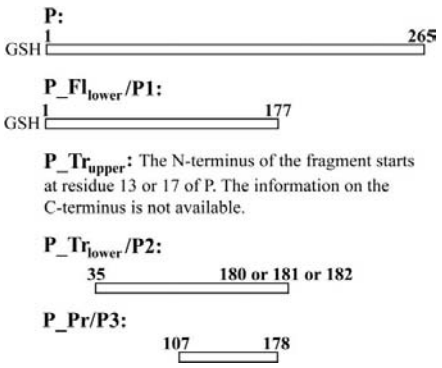

Figure 3
Diagrammatic representations of P-protein domains resulting from spontaneous degradation or limited proteolysis. The range of each domain within the P protein is indicated. Corresponding to the digestion map of the P protein, four fragments of the P protein, labelled P1 (amino acids 1–177), P2 (amino acids 35–177), P3 (amino acids 107–177) and P4 (amino acids 183–265) were subcloned.

Purified P1 and P2 were concentrated to 8.15 and 7.5 mg ml^{-1}, respectively, using Amicon ultracentrifugal filters with a 10 kDa cutoff membrane (Millipore, USA) and initial crystallization trials were performed using commercial screening kits including Hampton Research Crystal Screen (Jancarik & Kim, 1991), Natrix (Scott *et al.*, 1995), PEG/Ion Screen and Emerald Biostructures Wizard I. Purified P3 and P4 were concentrated to 22.5 and 13.5 mg ml^{-1}, respectively, using filters with a 5 kDa cutoff and initial crystallization trials were performed using screening kits including Hampton Research Crystal Screen, Crystal Screen 2 (Jancarik & Kim, 1991; Cudney *et al.*, 1994), Natrix and PEG/Ion Screen. Purified P3 and P4 were also concentrated to 88.8 and 49.4 mg ml^{-1}, respectively, for crystal screening using Hampton Research Crystal Screen and Natrix. Crystal screenings for P1, P2 and P4 are currently in progress. For P3, potassium fluoride and ammonium sulfate were tested as salt precipitants to induce crystallization when it was realised that the P3 protein contains three aspartic acid and eight glutamic acid residues within a 71-amino-acid peptide. Needle-shaped crystals were obtained in about 10 d by mixing 1 µl protein at 11 mg ml^{-1} with 1 µl reservoir solution consisting of 2.3 *M* ammonium sulfate. Optimization of this condition was carried out by varying the protein concentration and adding different chemical additives. Tetragonal shaped crystals were grown in about two weeks by mixing 1 µl protein at 5.5 mg ml^{-1} with an equal volume of reservoir solution containing 2.48 *M* ammonium sulfate and 7%(*v*/*v*) ethylene glycol. Refinement in order to grow diffraction-quality crystals was carried out by adding low concentrations of *n*-octyl-β-D-glucopyranoside (β-OG) to drops consisting of 1 µl protein at 22 mg ml^{-1} mixed with 1.5 µl 0.05%(*w*/*v*) β-OG (diluted with double-distilled water) and 2.5 µl reservoir solution. The reservoir solution contained 2.4815 *M* ammonium sulfate and 7%(*v*/*v*) ethylene glycol. Crystals grew to their full size in about 2–3 weeks (Fig. 4). A single crystal was cryoprotected in a solution containing 2.492 *M* ammonium sulfate, 15%(*v*/*v*) ethylene glycol and 0.01%(*w*/*v*) β-OG and flash-cooled in liquid nitrogen. A native data set was collected to 2.75 Å resolution using synchrotron radiation at a wavelength of 0.99997 Å, 0.25° oscillations and a crystal-to-detector distance of 270 mm. After indexing and density integration, systematic absences indicate that the crystal belongs to the tetragonal space group $P4_12_12$ or $P4_32_12$. The unit-cell parameters are $a = b = 74.50$, $c = 156.84$ Å (Table 1).

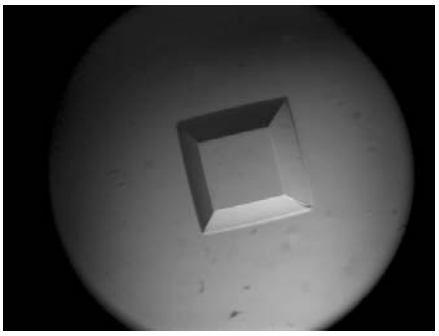

Figure 4
Native crystal of P3 protein; the maximum dimension of the crystal is about 0.4 mm.

Table 1
Data-collection and processing statistics of the P3 crystal.

Values in parentheses refer to the outer resolution shell.

Space group	$P4_12_12$ or $P4_32_12$
Unit-cell parameters (Å)	$a = b = 74.50$, $c = 156.84$
Resolution range (Å)	50.00–2.75 (2.85–2.75)
Average mosaicity (°)	0.563
No. observations	180521
No. unique reflections	11237
Completeness (%)	92.3 (58.3)
Average $I/\sigma(I)$	20.7 (5.6)
R_{sym}†	0.048 (0.228)

† $R_{sym} = \sum |I_{obs} - I_{avg}| / \sum I_{obs}$.

Data were collected at beamline 22-ID (or 22-BM) in the facilities of the South East Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org/main.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science under Contract No. W-31-109-Eng-38. The work was supported in part by a NIH grant (AI50066) to ML.

References

Banerjee, A. K. (1987). *Microbiol. Rev.* **51**, 66–87.

Banerjee, A. K. & Barik, S. (1992). *Virology*, **188**, 417–428.
Chattopadhyay, D. & Banerjee, A. K. (1987). *Cell*, **49**, 407–414.
Chen, J. L., Das, T. & Banerjee, A. K. (1997). *Virology*, **228**, 200–212.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Das, T. & Banerjee, A. K. (1992). *Cell Mol. Biol.* **38**, 17–26.
Ding, H., Qiu, S., Bunzel, R. J., Luo, D., Arabashi, A., Lu, S., Symersky, J., Nagy, L. A., DeLucas, L. J., Li, S. & Luo, M. (2003). *Acta Cryst.* D**59**, 1106–1108.
Ding, H., Qiu, S., Li, S., Symersky, J., Green, T. J. & Luo, M. (2003). *Acta Cryst.* D**59**, 1323–1326.
Gao, Y., Greenfield, N. J., Cleverley, D. Z. & Lenard, J. (1996). *Biochemistry*, **35**, 14569–14573.
Gao, Y. & Lenard, J. (1995*a*). *EMBO J.* **14**, 1240–1247.
Gao, Y. & Lenard, J. (1995*b*). *J. Virol.* **69**, 7718–7723.
Gaudier, M., Gaudin, Y. & Knossow, M. (2002). *EMBO J.* **21**, 2886–2892.
Green, T. J., MacPherson, S., Qiu, S., Lebowitz, J., Wertz, G. W. & Luo, M. (2000). *J. Virol.* **74**, 9515–9524.
Gupta, A. K., Shaji, D. & Banerjee, A. K. (2003). *J. Virol.* **77**, 732–738.
Hudson, L. D., Condra, C. & Lazzarini, R. A. (1986). *J. Gen. Virol.* **67**, 1571–1579.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Martinez, I., Rodriguez, L. L., Jimenez, C., Pauszek, S. J. & Wertz, G. W. (2003). *J. Virol.* **77**, 8039–8047.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pattnaik, A. K., Hwang, L., Li, T., Englund, N., Mathur, M., Das, T. & Banerjee, A. K. (1997). *J. Virol.* **71**, 8167–8175.
Qanungo, K. R., Shaji, D., Mathur, M. & Banerjee, A. K. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 5952–5927.
Schloemer, R. H. & Wagner, R. R. (1975). *J. Virol.* **16**, 237–240.
Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *J. Mol. Biol.* **250**, 327–332.
Takacs, A. M., Das, T. & Banerjee, A. K. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 10375–10379.
Tarbouriech, N., Curran, J., Ruigrok, R. W. & Burmeister, W. P. (2000). *Nature Struct. Biol.* **7**, 777–781.