Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, purification, crystallization and preliminary X-ray crystallographic study of the nucleocapsid protein of *Tomato spotted wilt virus*

Keisuke Komoda,[a,b] Masanori Narita,[a] Isao Tanaka[a,b] and Min Yao[a,b]*

[a]Graduate School of Life Sciences, Hokkaido University, Sapporo 060-0810, Japan, and [b]Faculty of Advanced Life Science, Hokkaido University, Sapporo 060-0810, Japan

Correspondence e-mail: yao@castor.sci.hokudai.ac.jp

Received 13 March 2013
Accepted 13 May 2013

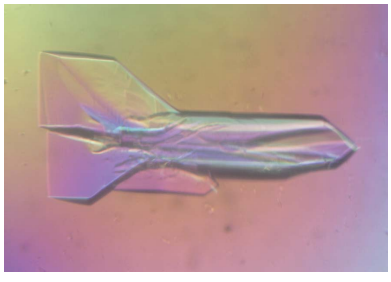
Tomato spotted wilt virus (TSWV), which causes severe damage to various agricultural crops such as tomato, pepper, lettuce and peanut, is a negative-stranded RNA virus belonging to the *Tospovirus* genus of the *Bunyaviridae* family. Viral genomic RNA molecules are packaged in the form of ribonucleoprotein complexes, each of which contains one viral RNA molecule that is coated with many nucleocapsid (N) proteins. Here, the expression and crystallization of TSWV N protein are reported. Native and selenomethionine-substituted crystals of N protein belonged to the same space group $P2_1$. Their unit-cell parameters were a = 66.8, b = 97.2, c = 72.0 Å, β = 112.8° and a = 66.5, b = 96.3, c = 72.1 Å, β = 113.4°, respectively.

1. Introduction

Tomato spotted wilt virus (TSWV) is classified as the type member of the genus *Tospovirus*, which includes *Impatiens necrotic spot virus* (INSV) and *Watermelon silver mottle virus* (WSMoV), within the family *Bunyaviridae*. The *Bunyaviridae* family has five genera, only one of which consists of plant-infecting viruses (King *et al.*, 2011); the others (*Orthobunyavirus*, *Hantavirus*, *Nairovirus* and *Phlebovirus*) are animal viruses, several of which are the causative agents of serious human disease (Elliott, 1990). TSWV has a worldwide distribution and causes severe damage to various agricultural crops, including tomato, pepper, lettuce and peanut. With an estimated economic loss owing to TSWV in excess of US$1 billion annually in 1994, control of TSWV is urgently required (Goldbach & Peters, 1994).

The TSWV genome consists of small (S), medium (M) and large (L) RNA segments. The S RNA segment encodes the nucleocapsid (N) and NSs proteins, the M RNA segment encodes glycoproteins (Gn and Gc) and NSm proteins, and the L RNA segment encodes the RNA-dependent RNA polymerase (RdRp) (Adkins, 2000). The numerous N proteins encapsidate viral genome RNAs and, in association with a few RdRps, constitute vRNP (viral ribonucleoprotein) complexes (Uhrig *et al.*, 1999; Richmond *et al.*, 1998). These vRNP complexes perform genome RNA replication and viral mRNA transcription. Moreover, N protein has been shown to interact with Gn/Gc protein and it has therefore been suggested that the N protein plays a crucial role in viral encapsidation (Ribeiro *et al.*, 2009). Overall, the N protein is an important factor in the TSWV life cycle.

Recently, the structures of N proteins from three animal virus genera of the *Bunyaviridae* family have been reported and several encapsidation mechanisms of viral genome RNAs have been suggested based on the formation of different oligomers of N proteins (Raymond *et al.*, 2010, 2012; Ferron *et al.*, 2011; Guo *et al.*, 2012; Carter *et al.*, 2012; Wang *et al.*, 2012; Dong *et al.*, 2013; Niu *et al.*, 2013; Li *et al.*, 2013; Reguera *et al.*, 2013). Interestingly, the sequence and structure of N proteins are highly conserved within a genus but differ between genera. For a full understanding of the genome encapsidation mechanism of the genus *Tospovirus*, the structure of the N protein from TSWV is indispensable, although the domains or amino-acid residues contributing to RNA binding or oligomer formation have been suggested by biochemical analyses (Richmond *et al.*, 1998; Uhrig *et al.*, 1999). Here, we report the crystallization and preliminary X-ray crystallographic analysis of the TSWV N protein to gain



further insight into its oligomeric state and the mechanism of RNA binding.

2. Materials and methods

2.1. Cloning, expression and purification

TSWV-infected leaves of *Pericallis* × *hybrida* (MAFF No. 260050) were obtained from the National Institute of Agrobiological Sciences Genebank, Tsukuba, Ibaraki, Japan. Total RNA was isolated from TSWV-infected leaves with RNAiso Plus (Takara Bio, Japan) according to the manufacturer's protocol. The purified total RNA was used as a template to synthesize the first-strand of cDNA of the S RNA segment. To obtain the full cDNA template of the S RNA segment, the primers TSWV-SP (5′-AGAGCAATCGTGTCAATTTTGTGTTCATACCTTAACACTC-3′) and TSWV-SN (5′-AGAGCAATTGTGTCAGAATTTTGTTCATAATCAAACCT-3′) were designed with reference to the sequence of TSWV CPNH9 strain (GenBank D00645.1). Reverse transcription (RT) was performed using primer TSWV-SP and ReverTra Ace reverse transcriptase (Toyobo, Japan) according to the manufacturer's protocol. The RT product was then PCR-amplified using KOD -Plus- polymerase (Toyobo, Japan) with primer pairs of TSWV-SP and TSWV-SN, and cloned into pCR-Blunt II-TOPO vector (Life Technologies, California, USA). After identifying the sequence of this cDNA template, we amplified the N-protein ORF (258 amino acids, molecular mass 29 kDa, theoretical pI of 9.5) by PCR using KOD -Plus- polymerase (Toyobo, Japan) with primer pairs FP (5′-AATTCATATGTCTAAGGTTAAGCTCACTAAGG-3′) and RP (5′-AATTCTCGAGTTAAGCAAGTTCTGCAAGTTTTGCC-3′). This PCR product was digested with *Nde*I/*Xho*I and inserted into the expression vector pET28a fused with extra amino acids (**MNHK**HHHHHHSSGENLYFQGH from the N-terminus to the C-terminus) including a D-box (MNHK), a His_6 tag (HHHHHH) and a TEV cleavage site (ENLYFQG) to the N-terminus of the N-protein ORF. To confirm oligomeric formation of the N protein, we prepared a double mutation F242A/F246A (Phe242 and Phe246 were substituted with Ala). The mutant variant F242A/F246A was produced using the expression plasmid of the wild-type N protein as a template with specific primer pairs mutFP (5′-GCTTATGAAATGGCTGGGGTTAAAAAACAGGCAAAACTTG-3′) and mutRP (5′-CTTGTTAAGAGTTTCACTGTAATGTTCCAT-3′).

The resulting plasmid was transformed into *Escherichia coli* strain B834 (DE3) + pRARE2 cells by electroporation. For the preparation of native N protein, the transformed cells were cultivated in Luria broth (LB) containing 25 μg ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol. For the preparation of selenomethionine-substituted (SeMet) N proteins, we used medium containing 25 μg ml^{-1} kanamycin and 34 μg ml^{-1} chloramphenicol, 1500 μg ml^{-1} adenine, 480 μg ml^{-1} thymine, 1980 μg ml^{-1} guanosine, 1500 μg ml^{-1} uracil, 120 μg ml^{-1} of each amino acid (except methionine) and 75 μg ml^{-1} selenomethionine. Transformed cells were cultivated at 310 K until the optical density at 600 nm (OD_{600}) reached about 0.6 and gene expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M*. The cells were cultured at 293 K for a further 18 h, harvested by centrifugation and resuspended in lysis buffer [50 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 10 m*M* $MgCl_2$, 10%(*v*/*v*) glycerol]. The resuspended cells were sonicated at 277 K and cleared by ultracentrifugation for 30 min at 40 000*g* at 277 K to remove cell debris.

The cleared cell lysate was loaded onto a HisTrap HP column (GE Healthcare) pre-equilibrated with His buffer *A* (lysis buffer containing 20 m*M* imidazole). The column was then washed with His buffer *A* and the bound protein was eluted with a linear gradient of 0–50% buffer *B* (lysis buffer containing 500 m*M* imidazole). The N-terminal His_6 tag was not removed because TEV protease was not able to digest the His_6 tag perfectly. Next, eluted fractions were applied onto a HiTrap Heparin HP column (GE Healthcare) to remove the endogenous nucleotides with heparin buffer *A* [50 m*M* Tris–HCl pH 7.5, 1 m*M* $MgCl_2$, 10%(*v*/*v*) glycerol] and the bound protein was eluted with a linear gradient of 0–100% buffer *B* [50 m*M* Tris–HCl pH 7.5, 1 m*M* $MgCl_2$, 2 *M* NaCl, 10%(*v*/*v*) glycerol]. After heparin column purification, the A_{280}:A_{260} ratio of a sample fraction had increased from 0.7 to 1.6, indicating that most of the host-derived nucleic acids had been removed. Finally, the collected fractions were concentrated and subjected to size-exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with 20 m*M* Na HEPES pH 7.5, 100 m*M* NaCl, 10%(*v*/*v*) glycerol. At each step in the purification procedure, fractions were analysed by SDS–PAGE and appropriate fractions were pooled. The final eluted protein was concentrated to 5 mg ml^{-1}, flash-cooled and stored at 193 K for further crystallization.

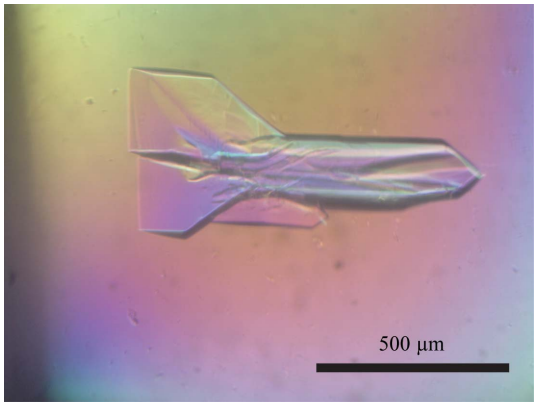


Figure 1
Crystals of SeMet N protein obtained by the hanging-drop vapour-diffusion method.

2.2. Crystallization

Crystallization screening of native N protein was performed using commercial kits (The JCSG Core Suites I, II, III and IV, and The PEGs and PEGs II Suites, Qiagen) by the sitting-drop vapour-diffusion method in 96-well plates. A 0.75 μl drop of the protein sample (5 mg ml^{-1}) was mixed with an equal volume of reservoir solution and the mixture was equilibrated against 75 μl reservoir solution at 293 K. Crystals of the N protein were obtained from condition No. 77 of The JCSG Core I Suite [0.1 *M* citric acid, 10%(*w*/*v*) PEG 6000, final pH 5.0], condition No. 49 of The JCSG Core Suite II [0.2 *M* sodium chloride, 0.1 *M* sodium/potassium phosphate pH 6.2, 20%(*w*/*v*) PEG 1000], condition No. 22 of The JCSG Core Suite III (0.1 *M* Tris pH 8.5, 2.4 *M* ammonium sulfate, final pH 8.0) and condition No. 15 of The PEGs II Suite [0.1 *M* Na HEPES pH 7.5, 25%(*w*/*v*) PEG 1000]. For optimization, the concentration of individual components and pH were adjusted and Additive Screen HT (Hampton Research) was also used under the above conditions *via* the hanging-drop vapour-diffusion method in 24-well plates. The best crystal was grown using 0.1 *M* Na HEPES pH 7.5, 25%(*w*/*v*) PEG 1000, 10 m*M* betaine hydrochloride. The crystals of the SeMet derivative of the N protein were obtained in 100 m*M* Na HEPES pH 7.5, 25%(*w*/*v*) PEG 1000, 50 m*M* NaCl by the hanging-

Table 1
Data for native and SeMet N protein crystals.

Values in parentheses are for the outer resolution shell.

	Native N protein	SeMet N protein
Beamline	BL-1A, Photon Factory	BL41XU, SPring-8
Wavelength (Å)	1.0000	0.9790
Detector	Quantum 270	Rayonix MX225HE
Crystal-to-detector distance (mm)	360	280
Rotation range per image (°)	1	1.4
Total rotation range (°)	180	360
Exposure time per image (s)	2	1
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å,°)	a = 66.8, b = 97.2, c = 72.0, β = 112.8	a = 66.5, b = 96.3, c = 72.1, β = 113.4
Mosaicity (°)	0.97–1.46	0.56–1.77
Resolution range (Å)	50.0–3.25 (3.37–3.25)	50.0–2.70 (2.80–2.70)
Total No. of reflections	46995	169351
No. of unique reflections	13251 (1330)	22739 (2268)
Completeness (%)	99.4 (99.7)	99.9 (99.9)
Multiplicity	3.5 (3.6)	7.4 (7.4)
$R_{\rm merge}$† (%)	8.1 (58.9)	12.9 (68.6)
$\langle I\rangle/\langle\sigma\ (I)\rangle$	14.2 (1.5)	15.3 (2.2)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.

drop vapour-diffusion method in 24-well plates. The largest crystal obtained under this condition was used for data collection (Fig. 1).

2.3. Data collection

For data collection at 100 K, the crystals of N protein were transferred into cryoprotectant solution containing 10% glycerol (native) or 5.6% sucrose (SeMet) and flash-cooled under a stream of liquid nitrogen. Diffraction data for the native N protein were collected using a Quantum 270 detector on BL-1A at the Photon Factory synchrotron facility, Tsukuba, Japan, while an Se-SAD data set was collected for the SeMet derivative of the N protein using a CCD detector (Rayonix MX225HE) on BL41XU at SPring-8, Harima, Japan. An X-ray wavelength of 0.9790 Å was used corresponding to the maximum f'' for the Se-SAD data set based on the fluorescence spectrum of the Se atom. All data sets were indexed, integrated, scaled and merged using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

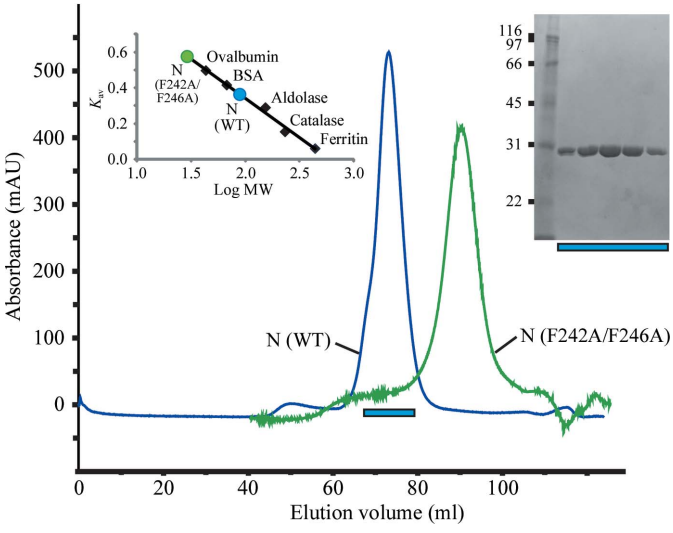


Figure 2
Elution profile of the TSWV N protein on HiLoad 16/60 Superdex 200. Left inset, the standard curve was obtained using the proteins ferritin (440 kDa), catalase (232 kDa), aldolase (153 kDa), bovine serum albumin (67 kDa) and ovalbumin (43 kDa). From its elution volume, purified native N protein (wild type; WT) was estimated at 87 kDa, which is nearly equal to the size of a trimer. We also analysed the double mutant F242A/F246A, which lost the oligomeric interaction (Uhrig *et al.*, 1999). The estimated size of this N protein mutant is about 30 kDa, which is nearly equal to the size of a monomer. $K_{\rm av} = (V_{\rm e} - V_{\rm o})/(V_{\rm t} - V_{\rm o})$, where $V_{\rm e}$ is the elution volume of the protein, $V_{\rm o}$ is the void volume and $V_{\rm t}$ is the bed volume. The elution pattern of the SeMet N protein was similar to that of the WT. Right inset, SDS–PAGE of the eluted fractions of WT. The left lane contains molecular-mass markers (labelled in kDa).

3. Results and discussion

In the final step of purification by size-exclusion chromatography using a Superdex 200 16/60 column (GE Healthcare), we estimated that the N protein formed trimers in solution, which was confirmed by analysis relative to molecular-weight standards (Fig. 2).

Both native and SeMet N protein crystals belonged to space group $P2_1$, with unit-cell parameters a = 66.8, b = 97.2, c = 72.0 Å, β = 112.8° and a = 66.5, b = 96.3, c = 72.1 Å, β = 113.4°, respectively (Table 1). For tag-fused N protein with molecular weight 31 452 Da, the Matthews equation (Matthews, 1968) indicated that there are three molecules with a $V_{\rm M}$ of 2.25 Å^3 Da^{-1} (solvent content = 45%) per asymmetric unit. The result of a self-rotation search with *POLARRFN* (Winn *et al.*, 2011) did not show obvious local threefold symmetry. The initial phases were obtained by the Se-SAD method using *SHELXC*/*D*/*E* (Sheldrick, 2008). The value of d''/sig was estimated to be larger than 0.61 from 3.1 Å, and 24 of 30 Se sites were determined with CC(all) = 50.8 and CC(weak) = 31.6. After phase improvement the pseudo-free CC volume showed that the inverted Se sites were correct. Initial model building was carried out with the program *ARP*/*wARP* (Langer *et al.*, 2008) after phase improvement. Refinement of this model is currently in progress in our laboratory.

This work was supported by a Grant-in-Aid for Young Scientists (Start-up; No. 22880001) from the Japan Society for the Promotion of Science (JSPS), Japan. The X-ray experiments were performed on BL41XU at SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI; proposal No. 2011A1272) and on BL-1A at the Photon Factory. We thank Dr Yoshikazu Tanaka for assistance with data collection and Dr Masayuki Ishikawa and the National Institute of Agrobiological Sciences for allowing the use of the TSWV cDNA clones.

References

Adkins, S. (2000). *Mol. Plant Pathol.* **1**, 151–157.
Carter, S. D., Surtees, R., Walter, C. T., Ariza, A., Bergeron, É, Nichol, S. T., Hiscox, J. A., Edwards, T. A. & Barr, J. N. (2012). *J. Virol.* **86**, 10914–10923.
Dong, H., Li, P., Elliott, R. M. & Dong, C. (2013). *J. Virol.* **87**, 5593–5601.
Elliott, R. M. (1990). *J. Gen. Virol.* **71**, 501–522.
Ferron, F., Li, Z., Danek, E. I., Luo, D., Wong, Y., Coutard, B., Lantez, V., Charrel, R., Canard, B., Walz, T. & Lescar, J. (2011). *PLoS Pathog.* **7**, e1002030.
Goldbach, R. & Peters, D. (1994). *Semin. Virol.* **5**, 113–120.
Guo, Y., Wang, W., Ji, W., Deng, M., Sun, Y., Zhou, H., Yang, C., Deng, F., Wang, H., Hu, Z., Lou, Z. & Rao, Z. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 5046–5051.
King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. (2011). Editors. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses.* London, Waltham, San Diego: Academic Press.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Li, B., Wang, Q., Pan, X., Fernández de Castro, I., Sun, Y., Guo, Y., Tao, X., Risco, C., Sui, S.-F. & Lou, Z. (2013). *Proc. Natl Acad. Sci. USA*, doi:10.1073/pnas.1222552110.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Niu, F., Shaw, N., Wang, Y. E., Jiao, L., Ding, W., Li, X., Zhu, P., Upur, H., Ouyang, S., Cheng, G. & Liu, Z.-J. (2013). *Proc. Natl Acad. Sci. USA*, doi:10.1073/pnas.1300035110.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Raymond, D. D., Piper, M. E., Gerrard, S. R., Skiniotis, G. & Smith, J. L. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 19208–19213.
Raymond, D. D., Piper, M. E., Gerrard, S. R. & Smith, J. L. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 11769–11774.
Reguera, J., Malet, H., Weber, F. & Cusack, S. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 7246–7251.
Ribeiro, D., Borst, J. W., Goldbach, R. & Kormelink, R. (2009). *Virology*, **383**, 121–130.
Richmond, K. E., Chenault, K., Sherwood, J. L. & German, T. L. (1998). *Virology*, **248**, 6–11.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Uhrig, J. F., Soellick, T. R., Minke, C. J., Philipp, C., Kellmann, J. W. & Schreier, P. H. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 55–60.
Wang, Y., Dutta, S., Karlberg, H., Devignot, S., Weber, F., Hao, Q., Tan, Y. J., Mirazimi, A. & Kotaka, M. (2012). *J. Virol.* **86**, 12294–12303.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.