



# Asymmetric Trimeric Ring Structure of the Nucleocapsid Protein of *Tospovirus*

Keisuke Komoda,<sup>a,b</sup> Masanori Narita,<sup>a</sup> Keitaro Yamashita,<sup>a\*</sup> Isao Tanaka,<sup>b</sup>  
Min Yao<sup>a,b</sup>

Graduate School of Life Sciences, Hokkaido University, Sapporo, Hokkaido, Japan<sup>a</sup>; Faculty of Advanced Life Science, Hokkaido University, Sapporo, Hokkaido, Japan<sup>b</sup>

**ABSTRACT** *Tomato spotted wilt virus* (TSWV), belonging to the genus *Tospovirus* of the family *Bunyaviridae*, causes significant economic damage to several vegetables and ornamental plants worldwide. Similar to those of all other negative-strand RNA viruses, the nucleocapsid (N) protein plays very important roles in its viral life cycle. N proteins protect genomic RNAs by encapsidation and form a viral ribonucleoprotein complex (vRNP) with some RNA-dependent RNA polymerases. Here we show the crystal structure of the N protein from TSWV. Protomers of TSWV N proteins consist of three parts: the N arm, C arm, and core domain. Unlike N proteins of other negative-strand RNA viruses, the TSWV N protein forms an asymmetric trimeric ring. To form the trimeric ring, the N and C arms of the N protein interact with the core domains of two adjacent N proteins. By solving the crystal structures of the TSWV N protein with nucleic acids, we showed that an inner cleft of the asymmetric trimeric ring is an RNA-binding site. These characteristics are similar to those of N proteins of other viruses of the family *Bunyaviridae*. Based on these observations, we discuss possibilities of a TSWV encapsidation model.

**IMPORTANCE** Tospoviruses cause significant crop losses throughout the world. Particularly, TSWV has an extremely wide host range (>1,000 plant species, including dicots and monocots), and worldwide losses are estimated to be in excess of \$1 billion annually. Despite such importance, no proteins of tospoviruses have been elucidated so far. Among TSWV-encoded proteins, the N protein is required for assembling the viral genomic RNA into the viral ribonucleoprotein (vRNP), which is involved in various steps of the life cycle of these viruses, such as RNA replication, virus particle formation, and cell-to-cell movement. This study revealed the structure of the N protein, with or without nucleic acids, of TSWV as the first virus of the genus *Tospovirus*, so it completed our view of the N proteins of the family *Bunyaviridae*.

**KEYWORDS** X-ray crystallography, bunyavirus, negative-strand RNA virus, nucleoprotein, plant viruses, tospovirus

**T**omato spotted wilt virus (TSWV), a type member of the genus *Tospovirus*, is one of the most devastating known pathogens of vegetables and ornamental plants (e.g., tomato, potato, lettuce, pepper, cyclamen, and impatiens) (1). TSWV can infect more than 1,000 different plant species, including dicots and monocots, and is transmitted by insects (thrips) (2). TSWV was recently nominated as one of the top 10 most important plant viruses based on scientific/economic importance (3).

The genus *Tospovirus* belongs to the family *Bunyaviridae*, a very large family of negative-strand RNA viruses comprising more than 350 serologically distinct viruses (4). The family *Bunyaviridae* consists of five genera (*Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus*, and *Tospovirus*) (5). Except for the genus *Tospovirus*, all genera belonging

Received 29 June 2017 Accepted 27 July 2017

Accepted manuscript posted online 2 August 2017

**Citation** Komoda K, Narita M, Yamashita K, Tanaka I, Yao M. 2017. Asymmetric trimeric ring structure of the nucleocapsid protein of *Tospovirus*. *J Virol* 91:e01002-17. <https://doi.org/10.1128/JVI.01002-17>.

**Editor** Anne E. Simon, University of Maryland, College Park

**Copyright** © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Keisuke Komoda, [komoda.k@gmail.com](mailto:komoda.k@gmail.com), or Min Yao, [yao@castor.sci.hokudai.ac.jp](mailto:yao@castor.sci.hokudai.ac.jp).

\* Present address: Keitaro Yamashita, RIKEN SPring-8 Center, Hyogo, Japan.

to the family *Bunyaviridae* are animal viruses, including human and livestock pathogens such as *Rift Valley fever virus* (RVFV) (*Phlebovirus*), *Severe fever with thrombocytopenia syndrome virus* (SFTSV) (*Phlebovirus*), *Crimean-Congo hemorrhagic fever virus* (CCHFV) (*Nairovirus*), *La Crosse virus* (LACV) (*Orthobunyavirus*), and *Hantaan virus* (HANV) (*Hantavirus*) (5).

As for all other bunyaviruses, TSWV virions contain single-strand, tripartite RNA genomes implementing negative-sense or ambisense coding strategies (5). A small (S) RNA segment encodes the nucleocapsid (N) protein and the NSs protein. A medium (M) RNA segment encodes two surface glycoproteins (Gn and Gc) and the NSm protein. A large (L) RNA segment encodes the RNA-dependent RNA polymerase (LP [L protein]) (6). The N protein interacts with genomic RNAs to form viral ribonucleoprotein (vRNP) complexes that work as a functional template for genomic RNA replication and viral mRNA transcription (7–10). In addition, the TSWV N protein interacts with Gn and Gc glycoproteins and appears to be involved in the formation of the virion (11). Moreover, the N protein was also reported to interact with NSm cell-to-cell movement proteins and plays a role in spreading viral genomic RNA to adjacent cells (12, 13). Thus, the TSWV N protein is a multifunctional protein, and the elucidation of the N protein structure will provide an important basis for further analyzing its biological function in diverse aspects of the TSWV life cycle.

Among members of the family *Bunyaviridae*, the N protein structures of four out of five genera except for *Tospovirus* have been elucidated to date. For the genus *Phlebovirus*, the N proteins of RVFV and *Toscana virus* were elucidated in several different oligomerization states (dimer and tetramer ring, pentamer ring, or hexamer ring) (14–17). For the genus *Nairovirus*, some oligomerization states (monomer and superhelical organizations) of N proteins have been reported (18–21). N proteins of several viruses of the genus *Orthobunyavirus* were determined as tetramer ring structures in crystals (22–28). Finally, the N protein of the genus *Hantavirus* has been reported to be a hexamer ring in crystals (29).

In this study, we elucidated the last unknown structure of tospovirus N proteins of five genera of the *Bunyaviridae*. X-ray crystal structure analysis showed that the TSWV N protein forms a nonplanar trimeric ring structure with or without nucleic acids. Our data may provide further insight into the mechanism of vRNP formation in the family *Bunyaviridae*.

## RESULTS

**Structure of the apo-TSWV N protein.** The recombinant TSWV N protein (residues M1 to A258) was prepared and crystallized as reported previously (30). The structure of the TSWV N protein was determined by a single-wavelength anomalous diffraction method using selenium atoms as anomalous scatterers (Se-SAD) (Table 1). The asymmetric unit contains three TSWV N protein molecules, denoted NP<sub>A</sub>, NP<sub>B</sub>, and NP<sub>C</sub> (Fig. 1C). The electron density clearly showed all atomic positions except for the first methionine plus the artificially fused tag of all NP molecules and the C-terminal 11 residues of NP<sub>A</sub>.

Each N protein molecule consists of three parts: an N-terminal arm (residues M1 to N33, termed the N arm), a C-terminal arm (residues S222 to A258, termed the C arm), and a central core domain (residues F34 to S221) (Fig. 1A). The N arm is formed by one alpha helix ( $\alpha$ 1, residues K8 to T16) with flexible strands (residues M1 to T7 and Q17 to N33). The core domain is formed by 11 alpha helices ( $\alpha$ 2 to  $\alpha$ 12) and two beta strands ( $\beta$ 1 and  $\beta$ 2). The C arm is formed by one bent alpha helix ( $\alpha$ 13, residues V230 to F246) wedged between two flexible strands (residues S222 to S229 and G247 to A258) (Fig. 1A). These N and C arms interact with the core domains of adjacent N protein molecules to form a trimeric ring structure (Fig. 1C), and positively charged clefts are located inside the ring (Fig. 1B, D, and E).

This trimeric ring of the TSWV N protein was distorted (Fig. 1C), corresponding to the result of a self-rotation search that found no local 3-fold axis (30). Despite the lack of a symmetry axis, the central core domains of the three molecules were identical and

**TABLE 1** Data collection and refinement statistics

Parameter	Value <sup>a</sup>		
	N protein (SeMet)	N-RNA complex	N-DNA complex
Data collection statistics			
Wavelength (Å)	0.97906	0.98000	1.00000
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	66.2, 96.0, 71.8	68.2, 94.7, 70.9	68.9, 86.5, 66.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 113.4, 90	90, 112.6, 90	90, 112.1, 90
Resolution (Å)	50–2.70 (2.87–2.70)	50–3.3 (3.5–3.3)	50–3.00 (3.18–3.00)
$R_{\text{meas}}$ (%) <sup>b</sup>	9.7 (76.3)	5.2 (94.8)	8.0 (99.6)
$\langle I/s(I) \rangle$	10.8 (2.13)	16.5 (1.75)	17.1 (2.38)
Completeness (%)	92.3 (98.5)	99.0 (97.8)	99.4 (98.0)
No. of unique reflections	41,185 (7,154)	12,585 (1,985)	14,540 (2,273)
Redundancy	3.8 (3.8)	3.7 (3.7)	7.6 (7.6)
Refinement statistics			
No. of reflections	41,171	12,581	14,534
$R_{\text{work}}/R_{\text{free}}$ (%)	21.04/25.52	28.23/31.66	25.05/29.13
No. of atoms			
Protein	5,965	5,439	5,666
RNA/DNA		220	328
Average <i>B</i> -factors (Å <sup>2</sup> )			
Protein	77.7	135.4	103.8
RNA/DNA		227.9	130
RMSD from ideal			
Bond length (Å)	0.003	0.002	0.005
Bond angle (°)	0.783	0.531	0.938
Ramachandran plot (%) <sup>c</sup>			
Favored regions	96.95	98.22	95.82
Outliers	0.27	0	0

<sup>a</sup>Values in parentheses are for the highest-resolution shell.

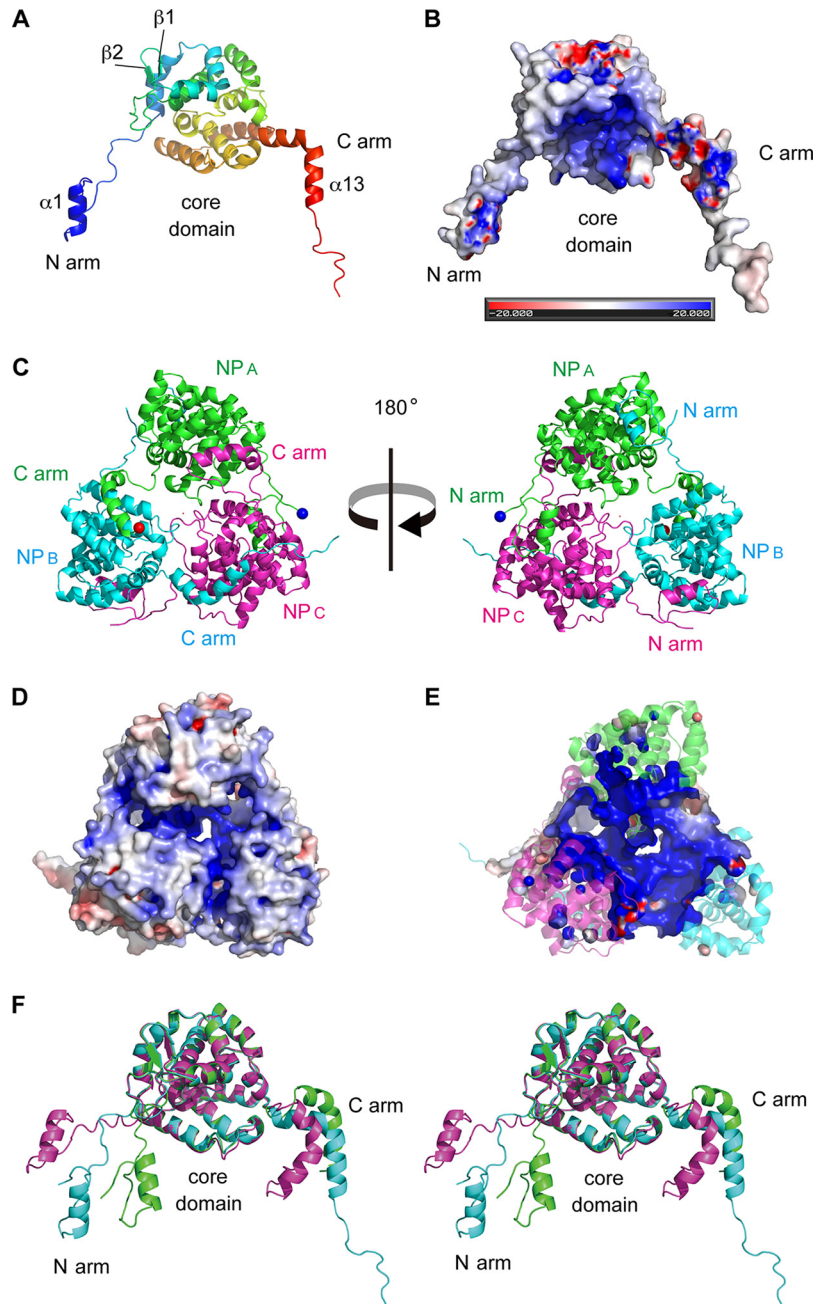
<sup>b</sup> $R_{\text{meas}}$  is the redundancy-independent  $R_{\text{sym}}$  value.

<sup>c</sup>Ramachandran analysis was performed by using MolProbity in the PHENIX package.

superposed very well, with root mean square deviations (RMSDs) of 0.280 Å ( $\text{NP}_A\text{-NP}_B$ ), 0.455 Å ( $\text{NP}_B\text{-NP}_C$ ), and 0.293 Å ( $\text{NP}_C\text{-NP}_A$ ) for 188  $\text{C}_\alpha$  atoms. When we superimposed the core domains of  $\text{NP}_A$ ,  $\text{NP}_B$ , and  $\text{NP}_C$ , the N and C arms of each molecule were extended in different directions, indicating that the distortion of the trimer is caused by the flexible arm structures (Fig. 1F). In addition, the absence of conserved interactions between the surfaces of the core domains of adjacent N protein members of the ring appears to be important for trimer distortion.

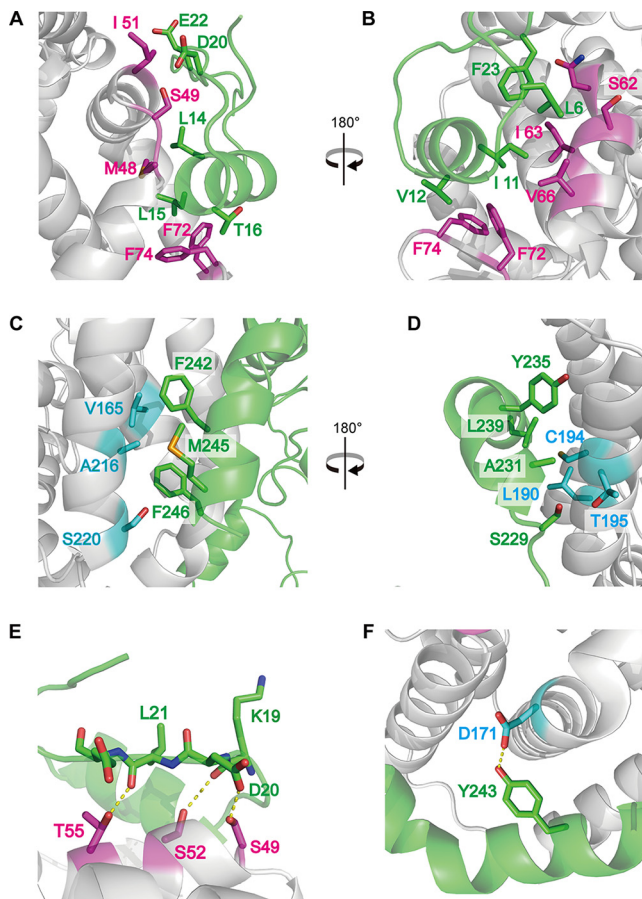
Figure 2 shows the detailed arm-to-core interactions of the N protein trimeric ring. Both N-arm-to-core and C-arm-to-core interactions are conserved among all combinations ( $\text{NP}_A\text{-NP}_B$ ,  $\text{NP}_B\text{-NP}_C$ , and  $\text{NP}_C\text{-NP}_A$ ) of N protein molecules, whereas the interactions of core domains are not conserved. This suggests that the distortion of the trimeric ring was not caused by differences in arm-to-core interactions but was caused by the arms' flexibility. In the N arm, residues L6, K8, I11, V12, L14, L15, T16, D20, E22, and F23 interact with the hydrophobic patch of the core domain containing residues K47, M48, S49, I51, S52, T55, N59, S62, I63, V66, F72, F74, and G75 by van der Waals forces (Fig. 2A and B). In addition, the main chain atoms of K19 and L21 and the side chain atoms of D20 form hydrogen bonds with the side chain atoms of S52, T55, and S49, respectively (Fig. 2E). In the C arm, residues A231, M232, H234, Y235, T238, L239, F242, Y243, M245, and F246 interact with another hydrophobic patch of the core domain containing residues M161, V164, V165, I168, Y169, A172, K187, L190, G191, C194, L197, K198, M204, V209, G212, K213, A216, L219, and S220 by van der Waals forces (Fig. 2C and D). In addition, the side chain atoms of D171 form hydrogen bonds with the side chain atoms of Y243 (Fig. 2F).

**Structure of the TSWV N protein with nucleic acids.** To elucidate the structure of the TSWV N protein in complex with nucleic acids, we assembled different complexes that consist of the N protein and several types of nucleic acid molecules of different



**FIG 1** Crystal structure of the TSWV N protein. (A) Structure of the TSWV N protein protomer (NP<sub>B</sub>). The N-terminal arm (N arm), central core domain (core domain), and C-terminal arm (C arm) are indicated. (B) Analysis of the electrostatic surface of TSWV NP<sub>B</sub> viewed from the same direction as in panel A. Blue and red represent positive and negative potentials, respectively. The color scale ranges between  $-20 k_B T$  and  $+20 k_B T$ , where  $k_B$  is Boltzmann's constant and  $T$  is temperature. (C) Cartoon representation of the TSWV N protein trimeric ring. NP<sub>A</sub>, NP<sub>B</sub>, and NP<sub>C</sub> are depicted in green, cyan, and magenta, respectively. The N and C termini of NP<sub>A</sub> are marked with blue and red spheres, respectively. (D) Analysis of the electrostatic surface of the TSWV N protein trimer viewed from the same direction as in panel C (right). (E) Transverse section of the structure in panel D at the level of the positively charged cleft. (F) Stereo view of the superposition of NP<sub>A</sub>, NP<sub>B</sub>, and NP<sub>C</sub>. Each N protein molecule is shown in the same color as that described above for panel C.

nucleotide lengths (10- to 30-nucleotide [nt] RNA or DNA) and purified these complexes by size exclusion chromatography (SEC). All complexes were eluted at almost the same position as the N protein apo form by SEC, but the ratio of the absorbance of this peak at 260 to that at 280 nm ( $A_{260}/A_{280}$ ) was increased from 0.6 (apo form) to 1.2 to 1.5



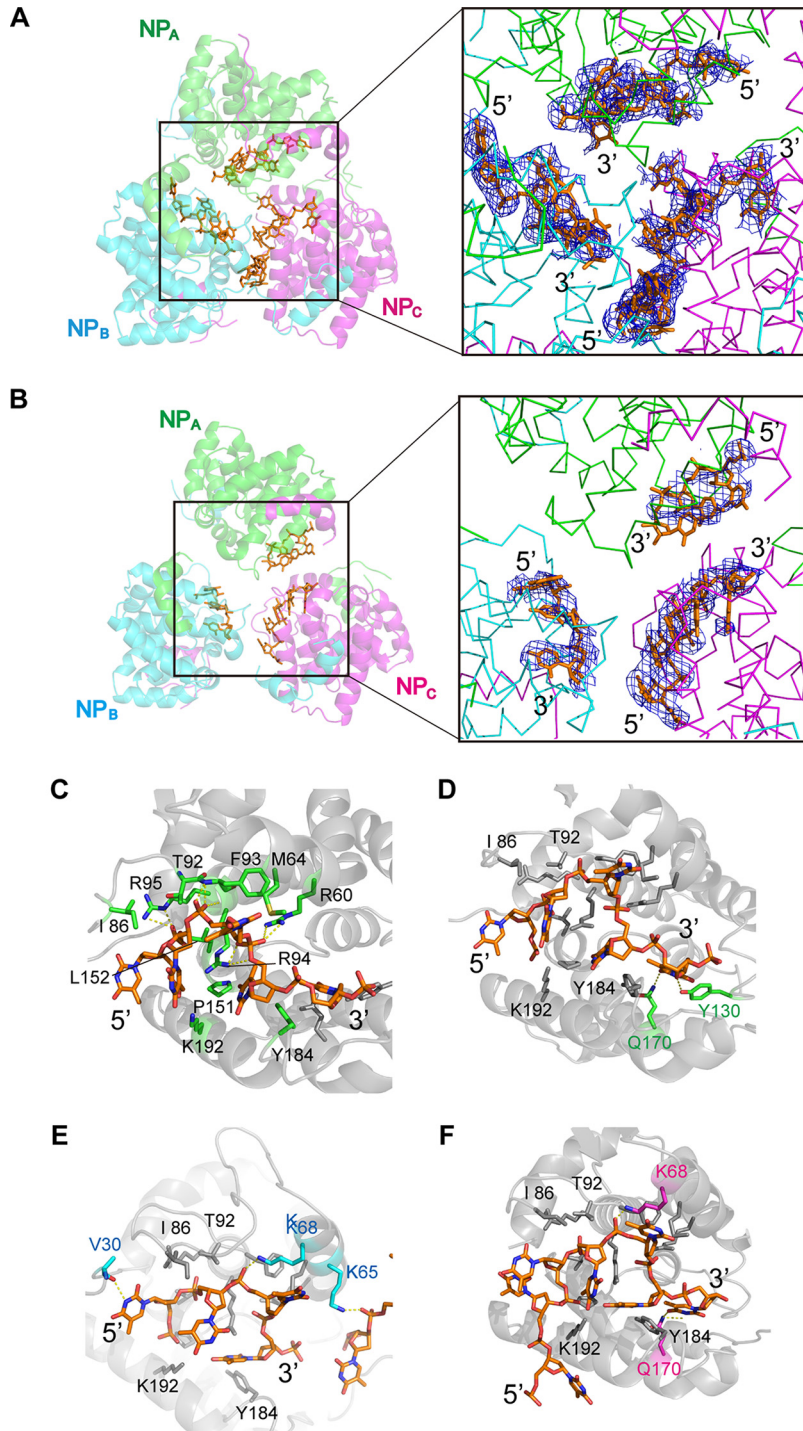
**FIG 2** Arm-to-core interactions of the TSWV NP trimer ring. (A) Interactions of the NP<sub>A</sub> N arm (green) and the NP<sub>C</sub> core domain (magenta). We depicted the representative hydrophobic residues that are involved in the interaction as sticks on one side of N arm helix  $\alpha$ 1 bound to a hydrophobic patch on the core domain formed by residues from  $\alpha$ 3,  $\alpha$ 4, and  $\beta$ 1. (B) Opposite view of the image shown in panel A after 180° rotation. (C) Interactions of the NP<sub>A</sub> C arm (green) and the NP<sub>B</sub> core domain (cyan). We depicted the representative hydrophobic residues that are involved in the interaction as sticks on one side of C arm helix  $\alpha$ 13 bound to a hydrophobic patch on the core domain formed by residues from  $\alpha$ 11 and  $\alpha$ 12. (D) Opposite view of the image shown in panel C after 180° rotation. (E) Magnified view of the hydrogen bonds of the NP<sub>A</sub> N arm (green) and the NP<sub>C</sub> core domain (magenta). (F) Magnified view of the hydrogen bonds between the NP<sub>A</sub> C arm (green) and the NP<sub>B</sub> core domain (cyan). Amino acid residues forming the interaction are displayed as stick models; dashed lines indicate hydrogen bonds.

(complex with nucleic acids) (data not shown). We obtained the best data set from crystals of the N protein bound with 25 deoxythymidine nucleotides (N-DNA complex) and with 19 nt of viral RNA sequences (see Materials and Methods) (N-RNA complex) at 3.0-Å and 3.3-Å resolutions, respectively. The complex structures were solved with molecular replacement using the monomer structure of the apo form as a search model (Table 1 and Fig. 3A and B).

Although the distorted trimeric ring structures formed by N proteins with DNA/RNA were similar to the trimeric ring formed without nucleic acids, the relative positions of the N proteins of each trimeric ring are different. As expected, the arm-to-core interactions of these N-nucleic acid complexes are exactly the same as those of the N protein apo form, whereas the interactions of core domains are not conserved (data not shown). The structural difference of each trimeric ring distortion in the presence or absence of nucleic acids also appeared in flexible arms. In particular, we found that by measuring the torsion angles of the polypeptide chains, the N arm bent at residues A31 to N33, and the C arm bent at residues P224 to A226.

Both DNA and RNA were bound at the central cleft of the core domain. In the positively charged cleft of the core domains, 11 (3 + 3 + 5) out of the 19 RNA





**FIG 3** TSWV N protein trimer with nucleic acids. (A) Structure of the N protein-DNA complex. DNA molecules are displayed in an orange stick model. The color of each N protein subunit is the same as that described in the legend of Fig. 1C. Shown is a composite-omit  $2F_o - F_c$  map of DNA contoured at  $1.0 \sigma$  (right). (B) Structure of the N-RNA complex. RNA molecules are displayed in an orange stick model. The color of each N protein subunit is the same as that described in the legend of Fig. 1C. Shown is a composite-omit  $2F_o - F_c$  map of RNA contoured at  $1.0 \sigma$  (right). (C to F) Magnified views of the interactions with DNA in each N protein subunit. DNA molecules are displayed in an orange stick model. (C) Residues that commonly interact with the DNA molecule in three N protein subunits are depicted in stick form in green. (D to F) Residues that interact with the DNA molecule in only one or two N protein subunits ( $NP_A$ , [D],  $NP_B$  [E], and  $NP_C$  [F]) are depicted in stick form in green ( $NP_A$ ), cyan ( $NP_B$ ), and pink ( $NP_C$ ).

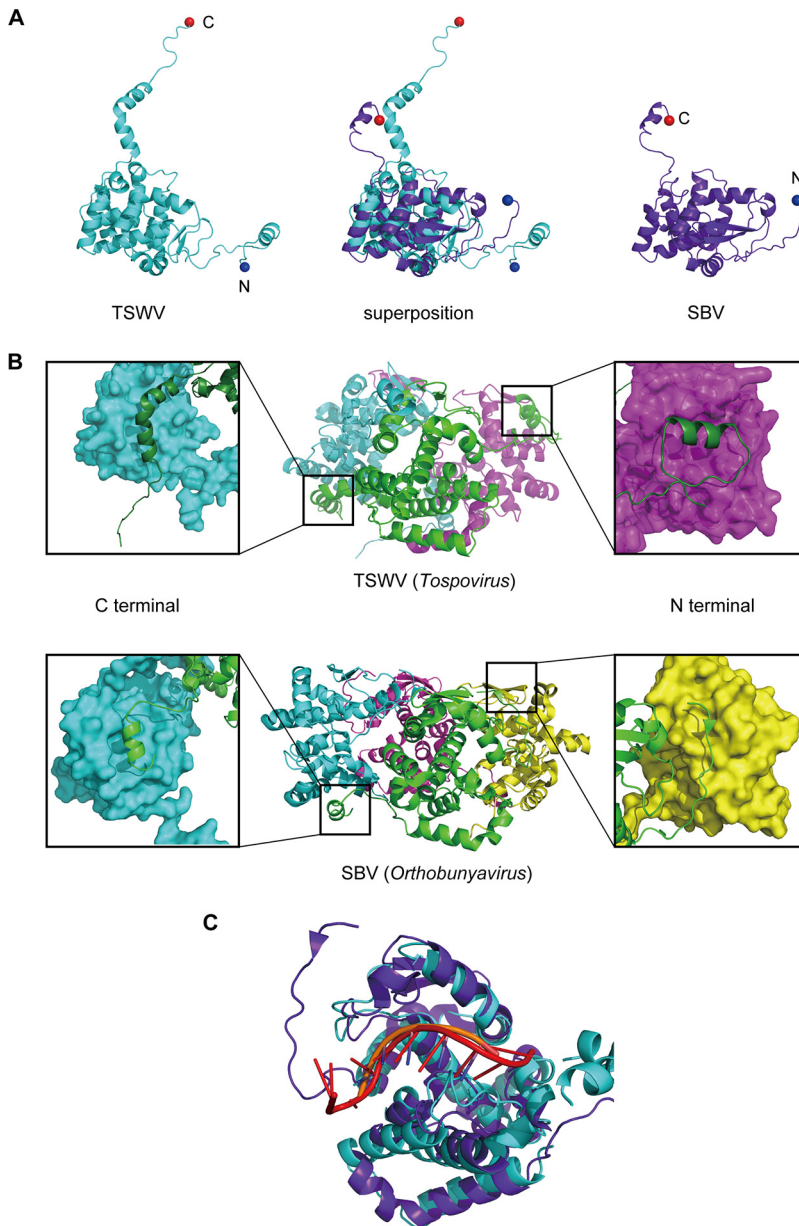
nucleotides and 16 (4 + 7 + 5) out of the 25 DNA nucleotides were determined, respectively (Fig. 3A and B). Although the difference in the resolutions was small, the final structural refinement showed that the electron density map at the cleft of the N-DNA complex was much clearer than that of the N-RNA complex (Fig. 3A and B). Thus, we analyzed the details of the interaction of NP and nucleic acids using the N-DNA complex. In this structure, amino acid residues contributing to the interaction with DNA were depicted in a stick form (Fig. 3C to F). Amino acid residues contributing to common interactions with DNA in the three molecules are depicted in Fig. 3C. The common interactions with the phosphates of the DNA molecule involved the side chains of R60, R94, and R95 and the main chains of F93 and R94. On the other hand, residues M64, I86, T92, R94, P151, L152, Y184, and K192 interacted with the DNA molecule by van der Waals forces (Fig. 3C). In addition, in the NP<sub>A</sub> molecule, the main chain of V30 interacted with the DNA base, and the side chains of K65 and K68 interacted with the DNA phosphates (Fig. 3D). In the NP<sub>B</sub> molecule, the side chain of K68 interacted with the DNA phosphate, and the side chain of Q170 interacted with the DNA base (Fig. 3E). In the NP<sub>C</sub> molecule, the side chains of Y130 and Q170 interacted with the DNA bases (Fig. 3F). Residues R60, K68, T92, F93, R94, R95, Y130, L152, Q170, and K192 are highly conserved in tospoviruses, while residues T92, R95, and K192 are conserved within both tospoviruses and orthobunyaviruses (see Fig. 5). Although we could not address the exact number of nucleotides that bind the N protein trimer due to the poor electron density map of nucleic acids at the N protein-N protein interfaces, the length range is estimated to be about 18 to 24 nt (that is, 6 to 8 nt per N protein protomer) based on the electron density map of the N-DNA complex.

#### **Structural comparison with other N proteins of members of the *Bunyaviridae*.**

Whereas many structures of N proteins have been reported for 4 of the 5 genera of the family *Bunyaviridae*, there has been no structural details of the N proteins of any member of the tospoviruses until this study. Next, we compared the structures of the TSWV N protein with those of representative viruses belonging to other genera of the family *Bunyaviridae*. Although the amino acid sequence similarity of N proteins among the five genera was low, topological similarity among *Phlebovirus*, *Hantavirus*, *Orthobunyavirus*, and *Tospovirus* has been reported (31, 32). All N proteins contain a largely globular domain (i.e., core domain in this study) with a positively charged groove for RNA binding and also have extensions such as N and C arms (*Orthobunyavirus*, *Hantavirus*, and *Tospovirus*), only an N arm (*Phlebovirus*), or mobile subdomains (*Nairovirus*) (33). Among them, the current structure (TSWV [*Tospovirus*]) is most similar to that of *Schmallenberg virus* (SBV) (RMSD = 4.3 Å for 110 C<sub>α</sub> atoms) of the *Orthobunyavirus* genus, which is the phylogenetically closest genus to *Tospovirus* (Fig. 4A). Although the structures of the core domain and manners of RNA binding seemed to be similar (Fig. 4A and C), the sequences and structures of both arms and the oligomeric states were clearly different between the two genera (Fig. 4B and 5). Namely, the TSWV N proteins formed a trimeric ring, whereas most orthobunyavirus N proteins formed a tetrameric ring in crystals (Fig. 4B).

## **DISCUSSION**

**The asymmetric trimeric ring structure of the TSWV N protein.** N proteins encapsidate viral genomic RNA to form vRNP complexes, and vRNP works as a template for genomic RNA replication and mRNA transcription (33). The appearance of vRNP complexes varies among different families and genera of negative-strand RNA viruses, according to the structure and oligomeric form of each N protein (34). To the best of our knowledge, this is the first report of a crystal structure of a tospovirus N protein in the apo, N-RNA binding, and N-DNA binding forms. TSWV N proteins form a distorted trimeric ring, displaying a unique oligomeric structure among all negative-strand RNA viruses (Fig. 1C). Recently, Li et al. reported the predicted structure of the TSWV N protein (35). Their model of the protomer agreed with our structure, especially in core domains. Although they argued that TSWV N proteins formed various multimers, their

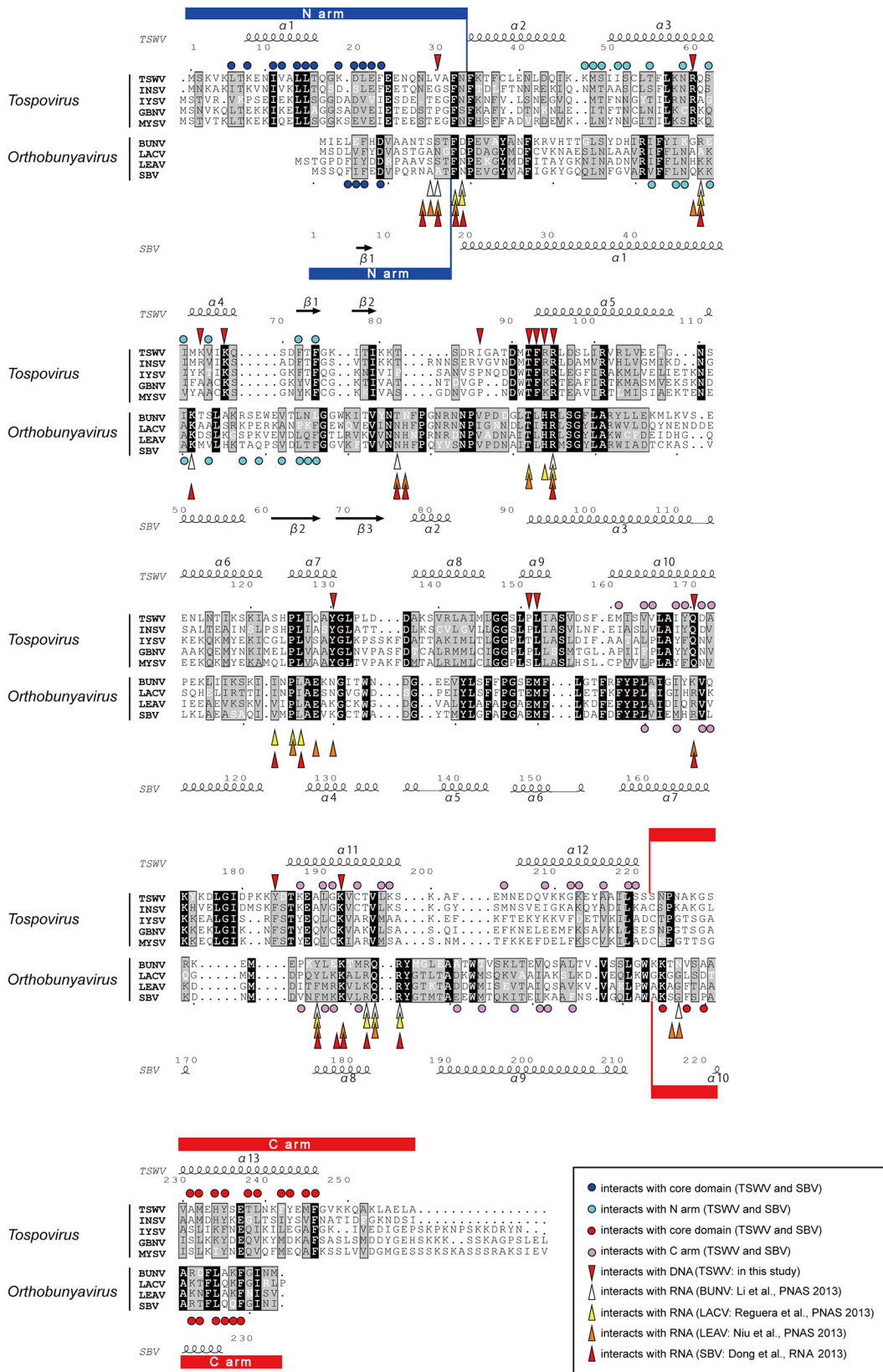


**FIG 4** Comparison of N protein structures of TSWV (*Tospovirus*) and SBV (*Orthobunyavirus*). (A) Protomer structures of the TSWV N protein (left) and the SBV N protein (PDB accession no. 4JNG) (right) and the superposition of both structures (middle). (B) Magnified views of the arm-to-core interaction within the TSWV (top) and SBV (bottom) N proteins are shown for each side. In the magnified views, NP<sub>A</sub> is shown as a green cartoon, whereas the adjacent N protein molecule is depicted as a molecular surface. (C) Superposition of the TSWV N-RNA complex (cyan, TSWV N; orange, RNA) and the SBV N-RNA complex (purple, SBV N; red, RNA).

results of blue native PAGE showed that the most abundant band was the N protein trimer, and this is consistent with our structure.

In our TSWV N protein structures, both the N and C arms interacted with the adjacent N proteins to form the trimeric ring (Fig. 2). These results were consistent with the previous observation that both the N-terminal (residues M1 to L39 and L42 to F56) and C-terminal (residues S233 to V248) regions of the TSWV N protein are important for the N-N homotypic interaction in the yeast two-hybrid system (8, 36). The F242A/F246A double mutant was previously reported to reduce N-N homotypic interactions drastically. Here we provide structural evidence explaining why these two residues, F242





**FIG 5** Amino acid sequence alignment of N proteins among tospoviruses and orthobunyaviruses. Key residues for nucleic acid binding are marked with downward arrowheads (red, TSWV) (the residues interact with DNA molecules) and upward arrowheads (white, *Bunyamvera virus* (BUNV); yellow, LACV; orange, *Leanyer virus* (LEAV); red, SBV). Key residues in the N arm for the interactions with the core domain are marked with dark blue circles, and key residues in the core domain for the interactions with the N arm are marked with light blue circles.

(Continued on next page)

and F246, are important for N-N interactions; these residues are located at the center of the C-arm-to-core interaction (Fig. 2C). Thus, the arm-to-core interaction is very important for N-N oligomerization, although the amino acid sequences of both TSWV N protein arms show low similarity to those of other genera of the family *Bunyaviridae* (Fig. 5).

**Possible encapsidation models of the TSWV N proteins.** Negative-strand RNA viruses (NSV) can be divided into two orders: the orders *Mononegavirales*, also known as nonsegmented NSV (nsNSV), and *Multinegavirales*, also known as segmented NSV (sNSV). The vRNP structures of nsNSV have been revealed to be generally linear with a relatively rigid helical conformation by cryo-electron microscopy (cryo-EM) reconstruction (37–40). On the other hand, the crystal structures of N proteins of *Respiratory syncytial virus* (RSV) and *Vesicular stomatitis virus* (VSV) have also been determined to be a 10-protomer ring (Protein Data Bank [PDB] accession no. 2WJ8 and 2GIC) (38, 41), while that of rabies virus was an 11-protomer ring (PDB accession no. 2GTT) (42). Also, in the case of sNSV, the vRNP structures of the *Bunyaviridae* family have been considered to form a helix, although the EM images of these structures showed irregular and flexible architectures (34). Actually, N proteins of CCHFV, which belongs to the genus *Nairovirus*, showed a helical organization even in the crystal, implying that this organization reflects their vRNP structure in living cells (20). The N proteins of the other three genera of the family *Bunyaviridae* have been reported to form smaller ring structures (3- to 6-mers) than those of nsNSVs in crystals (14–17, 22–29). In most previous reports, it was suggested that the N protein rings of the genera *Phlebovirus*, *Hantavirus*, and *Orthobunyavirus* were rearranged to form coiled vRNPs when they enwrap a long genomic RNA on the basis of their two-dimensional (2D) EM images (26, 27, 43). According to this model, tospovirus vRNP may also form a helical structure in the native state by rearrangement of the trimer ring structure.

Although the N protein of influenza virus formed a trimer structure (but not a ring) in some crystals, it was determined that its vRNP structure was a peculiar double-helical structure with two antiparallel strands by cryo-EM reconstruction (44, 45). Therefore, we cannot exclude the possibility that the vRNP structure in the native state is quite different from the N protein formation in crystals, like vRNP of influenza virus. Further study to elucidate the vRNP structure of tospoviruses in solution is indispensable.

## MATERIALS AND METHODS

**Protein expression, purification, crystallization, and data collection for the N protein apo form and N-DNA and N-RNA complexes.** The cloning, expression, purification, and crystallization of the TSWV N protein were described previously (30). Briefly, the TSWV N protein fused with the expression tag of pET28a was expressed in *Escherichia coli* strain B834(DE3)/pRARE2 cells and purified with a HisTrap HP column (GE Healthcare), followed by purification on a HiTrap Heparin HP column. Next, eluted fractions were subjected to size exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare). The final eluted protein was concentrated to 5 mg ml<sup>-1</sup>, flash-cooled, and stored at 193 K for further experiments. The 25-poly(dT) nucleotide was purchased from Sigma-Aldrich Japan. The 19-nt RNA (containing the TSWV 5'-terminal sequence of the S RNA genome) was synthesized from a DNA template by *in vitro* transcription using T7 RNA polymerase. The transcription product (5'-GAGAGCAAUUGUGUC AGAA-3') was purified by using a HiLoad 16/60 Superdex 75 prep-grade column (GE Healthcare) with buffer (20 mM HEPES-NaOH [pH 7.5], 100 mM NaCl, 10% glycerol). To form N-nucleic acid complexes,

### FIG 5 Legend (Continued)

interactions with the N arm are marked with light blue circles (TSWV and SBV). Key residues in the C arm for the interactions with the core domain are marked with red circles, and key residues in the core domain for the interactions with the C arm are marked with pink circles (TSWV and SBV). For each genus, strictly conserved residues are depicted in white characters on a black background, and highly conserved residues are depicted as black characters on a gray background. At the top and bottom of the sequences, a schematic representation of the secondary structure elements of TSWV NP and SBV NP are shown, and every 10 residues of TSWV NP and SBV NP are indicated with a dot. Basically, the alignment was generated by using ClustalW (57). The  $\alpha$ -helix is depicted by a coil, and the  $\beta$ -strand is depicted by an arrow. The figure was generated by using ESPript (58) and modified manually based on the structural information. The N arm is underscored in blue, and the C arm is underscored in red. GenBank database accession numbers are as follows: AB889601 for TSWV NP, AFH88369 for *Impatiens necrotic spot virus* (INSV) NP, ACA09432 for *Iris yellow spot virus* (IYSV) NP, NP\_619701 for *Groundnut bud necrosis virus* (GBNV) NP, ABY60853 for *Melon yellow spot virus* (MYSV) NP, NP\_047213 for BUNV NP, P04873 for LACV NP, AEA02984 for LEAV NP, and CCF55031 for SBV NP (24–27).

purified N protein was mixed in a 1:1 molar ratio with DNA or RNA in buffer (10 mM HEPES NaOH [pH 7.5], 100 mM NaCl, 20 mM KCl, 5 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 2 mM dithiothreitol [DTT]) and incubated at 30°C for 30 min. After incubation, the N-nucleic acid complexes were purified by using a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare) with buffer (20 mM HEPES-NaOH [pH 7.5], 100 mM NaCl, 10% glycerol). The crystallization conditions for the N protein without nucleic acids were previously described (30). The crystallization conditions for the N-DNA and N-RNA complexes were slightly modified as follows: 0.1 M HEPES (pH 7.5) plus 25% (wt/vol) polyethylene glycol (PEG 1000) for the N-DNA complex and 0.1 M HEPES-NaOH (pH 7.5), 25% (wt/vol) PEG 1000, and 0.1 M glycine for the N-RNA complex. We have tested various DNA and RNA strands for the preparation of crystals of a complex, and these nucleic acid molecules gave us the best resolution of crystals as a complex with the N protein. The crystal size and shape of the N-nucleic acid complexes were similar to those of the N protein alone. The X-ray diffraction data sets for the selenomethionine-substituted (SeMet) N protein and N-DNA complex were collected on beamline BL41XU, SPring-8, Japan. The data set for the N-RNA complex was collected on beamline BL-17A, Photon Factory, Japan. Every data set was indexed, integrated, and scaled with XDS (46).

**Structure determination and refinement.** The phase determination of the SeMet N protein was previously described (30). Briefly, the initial phases were determined by the Se-SAD method using SHELXC/D/E (47), and the initial model was built by using ARP/wARP (48). After several cycles of refinement with “phenix.refine” and “autoBUSTER” with noncrystallographic symmetry restraints and manual fitting with COOT, the  $R_{\text{work}}$  and  $R_{\text{free}}$  factors were converged to 21.0% and 25.5%, respectively (49–51). The structure of the N protein in complex with the 19-nt RNA was determined at a resolution of 3.3 Å by the molecular replacement (MR) method with Phaser, using the core domain of the SeMet-substituted TSWV N protein monomeric structure (PDB accession no. 5IP1) as a search model (52). It should be noted that we could not get a correct MR solution with a search model of the N protein trimer because of the significantly different arrangement of the trimeric structure. RNA models with lengths of 3 to 5 residues were built manually and fitted by using NAFIT (53). After several cycles of refinement with phenix.refine with noncrystallographic symmetry restraints and manual fitting with COOT, the  $R_{\text{work}}$  and  $R_{\text{free}}$  factors were converged to 28.2% and 31.7%, respectively. The structure of the N-DNA complex was determined at a 3.0-Å resolution in the same way. DNA models with lengths of 4 to 7 residues were built manually and fitted by using NAFIT. Finally,  $R_{\text{work}}$  and  $R_{\text{free}}$  factors were converged to 24.5% and 29.0%, respectively. Moreover, to confirm N protein trimer structure formation, a self-rotation search was performed by using the CCP4 program POLARRFN (54). The structures of the N protein apo form and the N-DNA/N-RNA complex were depicted by using PyMOL and APBS plugin (55, 56).

**Data availability.** The atomic coordinates and the structure factors have been deposited in the Protein Data Bank under accession no. 5IP1 for the SeMet TSWV N protein structure, 5IP2 for the N-RNA complex structure, and 5IP3 for the N-DNA complex structure.

## ACKNOWLEDGMENTS

The X-ray experiments were performed at BL41XU, SPring-8, with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (proposal no. 2011A1272) and at BL-17A, Photon Factory. We thank Yoshikazu Tanaka for assistance with data collection and Masayuki Ishikawa and the National Institute of Agrobiological Sciences for allowing us the use of TSWV cDNA clones.

This work was supported by a grant-in-aid for young scientists (startup) (grant no. 22880001) from the Japan Society for the Promotion of Science (JSPS), Japan.

## REFERENCES

- King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed). 2011. Virus taxonomy. Classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
- German TL, Ullman DE, Moyer JW. 1992. Tospoviruses: diagnosis, molecular biology, phylogeny, and vector relationships. *Annu Rev Phytopathol* 30: 315–348. <https://doi.org/10.1146/annurev.py.30.090192.001531>.
- Scholthof KB, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P, Hemenway C, Foster GD. 2011. Top 10 plant viruses in molecular plant pathology. *Mol Plant Pathol* 12:938–954. <https://doi.org/10.1111/j.1364-3703.2011.00752.x>.
- Walter CT, Barr JN. 2011. Recent advances in the molecular and cellular biology of bunyaviruses. *J Gen Virol* 92:2467–2484. <https://doi.org/10.1099/vir.0.035105-0>.
- Elliott RM. 1990. Molecular biology of the Bunyaviridae. *J Gen Virol* 71(Part 3):501–522. <https://doi.org/10.1099/0022-1317-71-3-501>.
- Adkins S. 2000. Tomato spotted wilt virus-positive steps towards negative success. *Mol Plant Pathol* 1:151–157. <https://doi.org/10.1046/j.1364-3703.2000.00022.x>.
- Richmond KE, Chenault K, Sherwood JL, German TL. 1998. Characterization of the nucleic acid binding properties of tomato spotted wilt virus nucleocapsid protein. *Virology* 248:6–11. <https://doi.org/10.1006/viro.1998.9223>.
- Uhrig JF, Soellick TR, Minke CJ, Philipp C, Kellmann JW, Schreier PH. 1999. Homotypic interaction and multimerization of nucleocapsid protein of tomato spotted wilt tospovirus: identification and characterization of two interacting domains. *Proc Natl Acad Sci U S A* 96:55–60. <https://doi.org/10.1073/pnas.96.1.55>.
- van Knippenberg I, Goldbach R, Kormelink R. 2002. Purified tomato spotted wilt virus particles support both genome replication and transcription in vitro. *Virology* 303:278–286. <https://doi.org/10.1006/viro.2002.1632>.
- Komoda K, Ishibashi K, Kawamura-Nagaya K, Ishikawa M. 2014. Possible involvement of eEF1A in Tomato spotted wilt virus RNA synthesis. *Virology* 468–470:81–87. <https://doi.org/10.1016/j.virol.2014.07.053>.
- Ribeiro D, Borst JW, Goldbach R, Kormelink R. 2009. Tomato spotted wilt virus nucleocapsid protein interacts with both viral glycoproteins Gn and Gc in planta. *Virology* 383:121–130. <https://doi.org/10.1016/j.virol.2008.09.028>.
- Kormelink R, Storms M, Van Lent J, Peters D, Goldbach R. 1994. Expres-

- sion and subcellular location of the NSM protein of tomato spotted wilt virus (TSWV), a putative viral movement protein. *Virology* 200:56–65. <https://doi.org/10.1006/viro.1994.1162>.
13. Soellick T, Uhrig JF, Bucher GL, Kellmann JW, Schreier PH. 2000. The movement protein NSM of tomato spotted wilt tospovirus (TSWV): RNA binding, interaction with the TSWV N protein, and identification of interacting plant proteins. *Proc Natl Acad Sci U S A* 97:2373–2378. <https://doi.org/10.1073/pnas.030548397>.
  14. Raymond DD, Piper ME, Gerrard SR, Smith JL. 2010. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proc Natl Acad Sci U S A* 107:11769–11774. <https://doi.org/10.1073/pnas.1001760107>.
  15. Ferron F, Li Z, Daneke EI, Luo D, Wong Y, Coutard B, Lantze V, Charrel R, Canard B, Walz T, Lescaur J. 2011. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog* 7:e1002030. <https://doi.org/10.1371/journal.ppat.1002030>.
  16. Raymond DD, Piper ME, Gerrard SR, Skiniotis G, Smith JL. 2012. Phleboviruses encapsidate their genomes by sequestering RNA bases. *Proc Natl Acad Sci U S A* 109:19208–19213. <https://doi.org/10.1073/pnas.1213553109>.
  17. Olal D, Dick A, Woods VL, Liu T, Li S, Devignot S, Weber F, Saphire EO, Daumke O. 2014. Structural insights into RNA encapsidation and helical assembly of the Toscana virus nucleoprotein. *Nucleic Acids Res* 42:6025–6037. <https://doi.org/10.1093/nar/gku229>.
  18. Carter SD, Surtees R, Walter CT, Ariza A, Bergeron É, Nichol ST, Hiscox JA, Edwards TA, Barr JN. 2012. Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *J Virol* 86:10914–10923. <https://doi.org/10.1128/JVI.01555-12>.
  19. 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. Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proc Natl Acad Sci U S A* 109:5046–5051. <https://doi.org/10.1073/pnas.1200808109>.
  20. Wang Y, Dutta S, Karlberg H, Devignot S, Weber F, Hao Q, Tan YJ, Mirazimi A, Kotaka M. 2012. Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. *J Virol* 86:12294–12303. <https://doi.org/10.1128/JVI.01627-12>.
  21. Wang W, Liu X, Wang X, Dong H, Ma C, Wang J, Liu B, Mao Y, Wang Y, Li T, Yang C, Guo Y. 2015. Structural and functional diversity of nairovirus-encoded nucleoproteins. *J Virol* 89:11740–11749. <https://doi.org/10.1128/JVI.01680-15>.
  22. Ariza A, Tanner SJ, Walter CT, Dent KC, Shepherd DA, Wu W, Matthews SV, Hiscox JA, Green TJ, Luo M, Elliott RM, Fooks AR, Ashcroft AE, Stonehouse NJ, Ranson NA, Barr JN, Edwards TA. 2013. Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic Acids Res* 41:5912–5926. <https://doi.org/10.1093/nar/gkt268>.
  23. Dong H, Li P, Elliott RM, Dong C. 2013. Structure of Schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. *J Virol* 87:5593–5601. <https://doi.org/10.1128/JVI.00223-13>.
  24. Dong H, Li P, Böttcher B, Elliott RM, Dong C. 2013. Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism. *RNA* 19:1129–1136. <https://doi.org/10.1261/rna.039057.113>.
  25. Li B, Wang Q, Pan X, Fernández de Castro I, Sun Y, Guo Y, Tao X, Risco C, Sui SF, Lou Z. 2013. Bunyamwera virus possesses a distinct nucleocapsid protein to facilitate genome encapsidation. *Proc Natl Acad Sci U S A* 110:9048–9053. <https://doi.org/10.1073/pnas.1222552110>.
  26. Niu F, Shaw N, Wang YE, Jiao L, Ding W, Li X, Zhu P, Upur H, Ouyang S, Cheng G, Liu ZJ. 2013. Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. *Proc Natl Acad Sci U S A* 110:9054–9059. <https://doi.org/10.1073/pnas.1300035110>.
  27. Reguera J, Malet H, Weber F, Cusack S. 2013. Structural basis for encapsidation of genomic RNA by La Crosse orthobunyavirus nucleoprotein. *Proc Natl Acad Sci U S A* 110:7246–7251. <https://doi.org/10.1073/pnas.1302298110>.
  28. Zheng W, Tao YJ. 2013. Genome encapsidation by orthobunyavirus nucleoproteins. *Proc Natl Acad Sci U S A* 110:8769–8770. <https://doi.org/10.1073/pnas.1306838110>.
  29. Olal D, Daumke O. 2016. Structure of the hantavirus nucleoprotein provides insights into the mechanism of RNA encapsidation. *Cell Rep* 14:2092–2099. <https://doi.org/10.1016/j.celrep.2016.02.005>.
  30. Komoda K, Narita M, Tanaka I, Yao M. 2013. Expression, purification, crystallization and preliminary X-ray crystallographic study of the nucleocapsid protein of Tomato spotted wilt virus. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 69:700–703. <https://doi.org/10.1107/S174430911301302X>.
  31. Guo Y, Wang W, Sun Y, Wang X, Wang X, Liu P, Shen S, Li B, Lin J, Deng F, Wang H, Lou Z. 2016. Crystal structure of the core region of hantavirus nucleocapsid protein reveals the mechanism for ribonucleoprotein complex formation. *J Virol* 90:1048–1061. <https://doi.org/10.1128/JVI.02523-15>.
  32. Lima RN, Faheem M, Alexandre J, Gonçalves R, Polêto MD, Verli H, Melo FL, Resende RO. 2016. Homology modeling and molecular dynamics provide structural insights into tospovirus nucleoprotein. *BMC Bioinformatics* 17:489. <https://doi.org/10.1186/s12859-016-1339-4>.
  33. Reguera J, Cusack S, Kolakofsky D. 2014. Segmented negative strand RNA virus nucleoprotein structure. *Curr Opin Virol* 5:7–15. <https://doi.org/10.1016/j.coviro.2014.01.003>.
  34. Ruigrok RWH, Crépin T, Kolakofsky D. 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 14:504–510. <https://doi.org/10.1016/j.mib.2011.07.011>.
  35. Li J, Feng Z, Wu J, Huang Y, Lu G, Zhu M, Wang B, Mao X, Tao X. 2015. Structure and function analysis of nucleocapsid protein of tomato spotted wilt virus interacting with RNA using homology modeling. *J Biol Chem* 290:3950–3961. <https://doi.org/10.1074/jbc.M114.604678>.
  36. Kainz M, Hilson P, Sweeney L, Derosé E, German TL. 2004. Interaction between Tomato spotted wilt virus N protein monomers involves non-electrostatic forces governed by multiple distinct regions in the primary structure. *Phytopathology* 94:759–765. <https://doi.org/10.1094/PHYTO.2004.94.7.759>.
  37. Schoehn G, Iseni F, Mavrakis M, Blondel D, Ruigrok RW. 2001. Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol* 75:490–498. <https://doi.org/10.1128/JVI.75.1.490-498.2001>.
  38. Tawar RG, Duquerroy S, Vohrner C, Varela PF, Damier-Piolle L, Castagné N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eléouët J-F, Rey FA. 2009. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* 326:1279–1283. <https://doi.org/10.1126/science.1177634>.
  39. Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH. 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327:689–693. <https://doi.org/10.1126/science.1181766>.
  40. Bakker SE, Duquerroy S, Galloux M, Loney C, Conner E, Eléouët J-F, Rey FA, Bhella D. 2013. The respiratory syncytial virus nucleoprotein-RNA complex forms a left-handed helical nucleocapsid. *J Gen Virol* 94:1734–1738. <https://doi.org/10.1099/vir.0.053025-0>.
  41. Green TJ, Zhang X, Wertz GW, Luo M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313:357–360. <https://doi.org/10.1126/science.1126953>.
  42. Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313:360–363. <https://doi.org/10.1126/science.1125280>.
  43. Shepherd DA, Ariza A, Edwards TA, Barr JN, Stonehouse NJ, Ashcroft AE. 2014. Probing bunyavirus N protein oligomerisation using mass spectrometry. *Rapid Commun Mass Spectrom* 28:793–800. <https://doi.org/10.1002/rcm.6841>.
  44. Arranz R, Coloma R, Chichón FJ, Conesa JJ, Carrascosa JL, Valpuesta JM, Ortín J, Martín-Benito J. 2012. The structure of native influenza virion ribonucleoproteins. *Science* 338:1634–1637. <https://doi.org/10.1126/science.1228172>.
  45. Moeller A, Kirchdoerfer RN, Potter CS, Carragher B, Wilson IA. 2012. Organization of the influenza virus replication machinery. *Science* 338:1631–1634. <https://doi.org/10.1126/science.1227270>.
  46. Kabsch W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132. <https://doi.org/10.1107/S0907444909047337>.
  47. Sheldrick GM. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66:479–485. <https://doi.org/10.1107/S0907444909038360>.
  48. Langer G, Cohen SX, Lamzin VS, Perrakis A. 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171–1179. <https://doi.org/10.1038/nprot.2008.91>.
  49. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd



- JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221. <https://doi.org/10.1107/S0907444909052925>.
50. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <https://doi.org/10.1107/S0907444910007493>.
51. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek C, Roversi P, Sharff A, Smart OS, Vonrhein C, Womack TO. 2011. BUSTER, version 2.11.5. Global Phasing Ltd, Cambridge, United Kingdom.
52. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674. <https://doi.org/10.1107/S0021889807021206>.
53. Yamashita K, Zhou Y, Tanaka I, Yao M. 2013. New model-fitting and model-completion programs for automated iterative nucleic acid refinement. *Acta Crystallogr D Biol Crystallogr* 69:1171–1179. <https://doi.org/10.1107/S0907444913007191>.
54. Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763. <https://doi.org/10.1107/S0907444994003112>.
55. DeLano WL. 2014. The PyMOL molecular graphics system, version 1.7.0.0. Schrödinger, LLC, New York, NY.
56. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037–10041. <https://doi.org/10.1073/pnas.181342398>.
57. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
58. Gouet P, Jones JD, Hoffman NE, Nussaume L. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308. <https://doi.org/10.1093/bioinformatics/15.4.305>.