The asymmetric trimeric ring structure of the nucleocapsid protein of *Tospovirus*

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

Graduate School of Life Sciences, Hokkaido University, Sapporo, Hokkaido, Japan[a]; Faculty of Advanced Life Science, Hokkaido University, Sapporo, Hokkaido, Japan[b]

Running title: Crystal structure of Tospovirus N protein

#Co-corresponding authors

To whom correspondence may be addressed:

my: yao@castor.sci.hokudai.ac.jp

kk: kkomoda@iastate.edu or komoda.k@gmail.com

*Present Address: Keisuke Komoda, Department of Plant Pathology and Microbiology, Iowa State University, Ames, USA; Keitaro Yamashita, RIKEN SPring-8 center, Hyogo, Japan

ABSTRACT

Tomato spotted wilt virus (TSWV), belonging to the genus *Tospovirus* of the family *Bunyaviridae*, causes significant economic damage in 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: 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 characters are similar to those of N proteins of other viruses in the family *Bunyaviridae*. Based on these observations, we discuss possibilities of 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 their worldwide losses estimated to be in excess of US$1 billion annually. Despite such an importance, any proteins of Tospoviruses have not been elucidated so far. Among TSWV-encoded proteins, N protein is required for assembling the viral genomic RNA into viral ribonucleoprotein (vRNP), and that is involved in various steps of their life cycle, 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 N proteins of the family *Bunyaviridae*.

INTRODUCTION

Tomato spotted wilt virus (TSWV), a type member of the genus *Tospovirus*, is one of the most devastating pathogens known for 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). *Bunyaviridae* consists of five genera (*Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus*, and *Tospovirus*) (5). Except for *Tospovirus*, all genera belonging 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 well as 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 complexes (vRNP) 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, N protein is 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, 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 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-, pentamer-, or hexamer-ring) (14–17). For the genus *Nairovirus*, some oligomerization states (monomer and superhelical organization) 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, N protein of the genus *Hantavirus* has been reported as a hexamer-ring in crystals (29).

In this study, we elucidated the last unknown structure of Tospovirus N protein in five genera of *Bunyaviridae*. X-ray crystal structure analysis showed that TSWV N protein forms non-planar 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. Recombinant TSWV N protein (residues M1–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_A, NP_B, and NP_C (Fig. 1C). The electron density clearly showed all atomic positions except for the first methionine plus artificially fused tag of all NP molecules and the C terminal eleven residues of NP_A.

Each N protein molecule consists of three parts: an N-terminal arm (residues M1–N33, termed N arm), a C-terminal arm (residues S222–A258, termed C arm), and a central core domain (residues F34–S221) (Fig. 1A). The N arm is formed by one alpha helix (α1: residues K8–T16) with flexible strands (residues M1–T7 and Q17–N33). The core domain is formed by eleven alpha helices (α2–α12) and two beta strands (β1 and β2). The C arm is formed by one bent alpha helix (α13: residues V230–F246) wedged between two flexible strands (residues S222–S229 and G247–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 TSWV N protein was distorted (Fig. 1C), corresponding to the result of a self-rotation search which found no local three-fold axis (30). Despite the lack of a symmetry axis, the central core domains of the three molecules were identical and superposed very well with the root-mean-square deviation (RMSD) of 0.280 Å (NP_A–NP_B), 0.455 Å (NP_B–NP_C), and 0.293 Å (NP_C–NP_A) for 188 Cα atoms. When we superimposed the core domains of NP_A, NP_B, and 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 (NP_A–NP_B, NP_B–NP_C, and NP_C–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 by the arms' flexibility. In the N arm, the 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, 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, the 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 N protein and several types of nucleic acid molecules of different nucleotide (nt) lengths (10–30 nts 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 absorbance of this peak at 260 and 280nm (A260/A280) was increased from 0.6 (apo form) to 1.2–1.5 (complex with nucleic acids) (data not shown). We obtained the best dataset from crystals of N protein bound with 25 deoxythymidine nucleotides (N-DNA complex) and with 19 nts of viral RNA sequences (see Material 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 apo form as a search model (Table 1, 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 N protein of each trimeric ring are different. As expected, the arm-to-core interactions of these N-nucleic acid complexes are exactly the same as that 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 on the presence or absence of nucleic acid was 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–N33 and the C arm bent at residues P224–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 nts and 16 (4 + 7 + 5) out of the 25 DNA nts were determined, respectively (Fig. 3A and B). Although the difference of the resolution was small, the final structural refinement showed that electron density map at the cleft of N-DNA complex was much clearer than that of N-RNA complex (Fig. 3A and B). So we analyzed the details of NP and nucleic acids interaction using N-DNA complex. In this structure, amino acid residues contributing to the interaction with DNA were depicted in stick form (Fig. 3C-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_A 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_B 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_C 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* (Fig. 5). Though we could not address the exact number of nucleotides which bind 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 18−24 nts (that is 6−8 nts per N protein protomer) based on the electron density map of N-DNA complex.

Structural comparison with other N proteins in *Bunyaviridae*. Whereas, many structures of N proteins have been reported from 4 of the 5 genera in the family *Bunyaviridae*, there has been no structural detail of the N protein for any member of the *Tospoviruses* until this study. Then, we compared the structures of TSWV N protein with those of representative viruses belonging to other genera in the family *Bunyaviridae*. Although the amino acid sequence

similarity of N proteins was low among five genera, topological similarity has been reported among *Phlebovirus*, *Hantavirus*, *Orthobunyavirus*, and *Tospovirus* (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 N arm (*Phlebovirus*), or mobile subdomains (*Nairovirus*) (33). Among them, current structure (TSWV (*Tospovirus*)) is most similar to that of *Schmallenberg virus* (SBV) (RMSD = 4.3 Å for 110 Ca atoms) from *Orthobunyavirus,* which is the phylogenetically closest genus with *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 two genera (Fig. 5 and 4B). Namely, the TSWV N proteins formed a trimeric ring, whereas most Orthobunyavirus N proteins formed a tetrameric ring in crystal (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 crystal structure of 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 TSWV N protein (35). Their model of protomer agreed with our structure, especially in core domains. Although they argued that TSWV N proteins formed various multimers, their result of Blue Native PAGE showed that the most abundant band was N protein trimer, and it is consistent with our structure.

In our TSWV N protein structures, both 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 N-terminal (residues M1–L39, L42–F56) and C-terminal (residues S233–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

200 reduce N–N homotypic interactions drastically. Here we provide structural evidence explaining
201 why these two residues F242 and F246 are important for N–N interactions; these residues are
202 located at the center of the C arm-to-core interaction (Fig. 2C). Thus, the arm-to-core interaction
203 is very important for N–N oligomerization, although the amino acid sequences of both TSWV N
204 protein arms show low similarity with those of other genera in the family *Bunyaviridae* (Fig. 5).
205
206 **Possible encapsidation models of the TSWV N proteins.** Negative-strand RNA viruses
207 (NSV) can be divided into two orders; the order Mononegavirales, also known as
208 non-segmented NSV (nsNSV), and Multinegavirales, also known as segmented NSV (sNSV).
209 The vRNP structures of nsNSV have been revealed as generally linear with a relatively rigid
210 helical conformation by a cryoelectron microscopy (cryo-EM) reconstruction (37–40). On the
211 other hand, the crystal structures of N proteins of *respiratory syncytial virus* (RSV) and
212 *Vesicular stomatitis virus* (VSV) have also been determined as a 10-protomer ring (PDB code:
213 2WJ8, 2GIC) (38, 41), while that of rabies virus was an 11-protomer ring (PDB code: 2GTT) (42).
214 Also in the case of sNSV, the vRNP structures of the *Bunyaviridae* family have been considered
215 to form a helix, although the electron microscopy (EM) images of them were irregular and
216 flexible architectures (34). Actually, N proteins of CCHFV that belongs to the genus *Nairovirus*
217 showed the helical organization even in the crystal, implying that this organization reflects their

218 vRNP structure in the living cells (20). The N proteins of the other three genera of the family
219 *Bunyaviridae* have been reported to form smaller ring structures (3 to 6-mer) than those of
220 nsNSVs in crystals (14–17, 22–29). In most of the reports, it has been suggested that their N
221 protein rings of the genus *Phlebo*-, *Hanta*-, and *Orthobunyavirus* were rearranged to form coiled
222 vRNPs when it enwraps a long genomic RNA on the basis of their 2D EM images (26, 27, 43).
223 According to this model, Tospovirus vRNP may also form a helical structure in the native state
224 by rearrangement of the trimer ring structure.

225 Although N protein of Influenza virus formed a trimer structure (but not a ring) in some
226 crystals, it has been determined that the vRNP structure of them was a peculiar double-helical
227 structure with two anti-parallel strands by cryo-EM reconstruction (44, 45). Therefore, we
228 cannot exclude the third possibility that the vRNP structure in the native state is quite different
229 from the N protein formation in crystals, like a vRNP of Influenza virus. The further study to
230 elucidate the vRNP structure of Tospoviruses in solution is indispensible.

231

MATERIALS AND METHODS

Protein expression, purification, crystallization and data collection of N protein apo form, N-DNA, and N-RNA complexes.

Cloning, expression, purification, and crystallization of the TSWV N protein have been described previously (30). Briefly, TSWV N protein fused with expression tag of pET28a was expressed in Escherichia coli strain B834(DE3) + pRARE2 cells, and purified with a HisTrap HP column (GE Healthcare) and succeeding 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^{-1}, flash-cooled and stored at 193K for further experiments. The 25-poly-dT-nucleotide was purchased from Sigma-Aldrich Japan. The 19nt 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′-GAGAGCAAUUGUGUCAGAA-3′) was purified using a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) with a buffer [20 mM HEPES NaOH (pH7.5), 100 mM NaCl, 10% Glycerol]. To form N-nucleic acid complexes, purified N protein was mixed in a 1:1 molar ratio with DNA or RNA in a buffer [10 mM HEPES NaOH (pH7.5), 100 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$ 0.5 mM EDTA, 2 mM DTT] and incubated at 30°C for 30 min. After incubation, the N-nucleic acid complexes were purified 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 of the N protein without nucleic acids were previously described (30). The crystallization conditions of the N-DNA and N-RNA complexes were slightly modified as follows: 0.1 M HEPES (pH7.5), 25% (w/v) PEG1000 for the N-DNA complex. 0.1 M HEPES-NaOH (pH7.5), 25% (w/v) PEG1000, 0.1 M glycine for the N-RNA complex. We have tested various DNA and RNA strands for preparing a crystal of complex, and these nucleic acid molecules gave us the best resolution of crystals as a complex with N protein. The crystal size and shape of the N-nucleic acid complexes were similar to 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 of 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 Se-SAD method using SHELXC/D/E (47) and the initial model was built 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-work and R-free factors were converged to 21.0% and 25.5%, respectively (49–51). The structure of the N protein in complex with the 19nt 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-ID 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 of 3–5 residue lengths were built manually and fitted using NAFIT (53). After several cycles of refinement with “phenix.refine” with noncrystallographic symmetry restraints and manual fitting with COOT, the R-work and R-free factors were converged to 28.2% and 31.7%, respectively. The structure of N-DNA complex was determined at 3.0 Å in the same way. DNA models of 4-7 residue lengths were built manually and fitted using NAFIT. Finally, R-work and R-free factors were converged to 24.5% and 29.0%, respectively. Moreover, to confirm the trimer of N protein structure formation, a self-rotation search was performed using the *CCP*4 program *POLARRFN* (54). The structures of the N protein apo form and N-DNA/N-RNA complex were depicted using PyMOL and APBS plugin (55, 56).

Data availability

285 The atomic coordinates and the structure factors have been deposited in the Protein Data Bank
286 under accession no. 5IP1 for the SeMet TSWV N protein structure, 5IP2 for the N-RNA complex
287 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 Dr. 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.

FUNDING

This work was supported by a Grant-in Aid for Young Scientists (Start-up) [Number 22880001] from the Japan Society for the Promotion of Science (JSPS), Japan.

REFERENCES

1. **King AMQ**, **Adams MJ**, **Carstens EB**, **Lefkowitz EJ**. 2012. Virus taxonomy : classification and nomenclature of viruses : ninth report of the International Committee on Taxonomy of Viruses. Elsevier.

2. **German TL**, **Ullman DE**, **Moyer JW**. 1992. Tospoviruses: diagnosis, molecular biology, phylogeny, and vector relationships. Annu Rev Phytopathol **30**:315–48.

3. **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**:938–54.

4. **Walter CT**, **Barr JN**. 2011. Recent advances in the molecular and cellular biology of bunyaviruses. J Gen Virol **92**:2467–84.

5. **Elliott RM**. 1990. Molecular biology of the Bunyaviridae. J Gen Virol **71 (Pt 3)**:501–22.

6. **Adkins S**. 2000. Tomato spotted wilt virus-positive steps towards negative success. Mol Plant Pathol **1**:151–7.

7. **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.

8. **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.

9. **van Knippenberg I**, **Goldbach R**, **Kormelink R**. 2002. Purified tomato spotted wilt virus particles support both genome replication and transcription in vitro. Virology **303**:278–86.

10. **Komoda K**, **Ishibashi K**, **Kawamura-Nagaya K**, **Ishikawa M**. 2014. Possible involvement of eEF1A in Tomato spotted wilt virus RNA synthesis. Virology **468**–**470**:81–7.

11. **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**:121–30.

12. **Kormelink R**, **Storms M**, **Van Lent J**, **Peters D**, **Goldbach R**. 1994. Expression and subcellular location of the NSM protein of tomato spotted wilt virus (TSWV), a putative viral movement protein. Virology **200**:56–65.

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**:2373–8.

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**:11769–74.

15. **Ferron F**, **Li Z**, **Danek EI**, **Luo D**, **Wong Y**, **Coutard B**, **Lantez V**, **Charrel R**, **Canard B**, **Walz T**, **Lescar J**. 2011. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. PLoS Pathog **7**:e1002030.

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**:19208–13.

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**:6025–37.

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**:10914–23.

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**:5046–51.

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**:12294–303.

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**:11740–9.

22. **Ariza A**, **Tanner SJ**, **Walter CT**, **Dent KC**, **Shepherd DA**, **Wu W**, **Matthews S V**, **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.

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**:5593–601.

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**:1129–36.

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**:9048–53.

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**:9054–9.

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**:7246–51.

28. **Zheng W**, **Tao YJ**. 2013. Genome encapsidation by orthobunyavirus nucleoproteins. Proc Natl Acad Sci U S A **110**:8769–70.

29. **Olal D**, **Daumke O**. 2016. Structure of the Hantavirus Nucleoprotein Provides Insights into the Mechanism of RNA Encapsidation. Cell Rep **14**:2092–9.

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.

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 **90**:1048–1061.

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**.

33. **Reguera J**, **Cusack S**, **Kolakofsky D**. 2014. Segmented negative strand RNA virus nucleoprotein structure. Curr Opin Virol **5**:7–15.

34. **Ruigrok RWH**, **Crépin T**, **Kolakofsky D**. 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr Opin Microbiol **14**:504–510.

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**:3950–61.

36. **Kainz M**, **Hilson P**, **Sweeney L**, **Derose E**, **German TL**. 2004. Interaction Between Tomato spotted wilt virus N Protein Monomers Involves Nonelectrostatic Forces Governed by Multiple Distinct Regions in the Primary Structure. Phytopathology **94**:759–65.

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**:490–8.

38. **Tawar RG**, **Duquerroy S**, **Vonrhein 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**:1279–83.

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**:689–93.

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**:1734–8.

41. **Green TJ**, **Zhang X**, **Wertz GW**, **Luo M**. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. Science **313**:357–60.

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**:360–3.

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.

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**:1634–7.

45. **Moeller A**, **Kirchdoerfer RN**, **Potter CS**, **Carragher B**, **Wilson IA**. 2012. Organization of the influenza virus replication machinery. Science **338**:1631–4.

46. **Kabsch W**. 2010. XDS. Acta Crystallogr D Biol Crystallogr **66**:125–32.

47. **Sheldrick GM**. 2010. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. Acta Crystallogr Sect D Biol Crystallogr **66**:479–485.

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**:1171–9.

49. **Adams PD**, **Afonine P V.**, **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 Sect D Biol Crystallogr **66**:213–221.

50. **Emsley P**, **Lohkamp B**, **Scott WG**, **Cowtan K**. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66**:486–501.

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.

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**:1171–9.

54. **Collaborative Computational Project, Number 4**. 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr **50**:760–3.

55. **DeLano WL**. 2014. The PyMOL Molecular Graphics System, Version 1.7.0.0 Schrödinger, LLC.

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.

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.

58. **Gouet P, Jones JD, Hoffman NE, Nussaume L.** ESPript:analysis of multiple sequence alignments in PostScript. Bioinformatics **15**:305-308.

463 **Table 1. Data collection and refinement statistics**

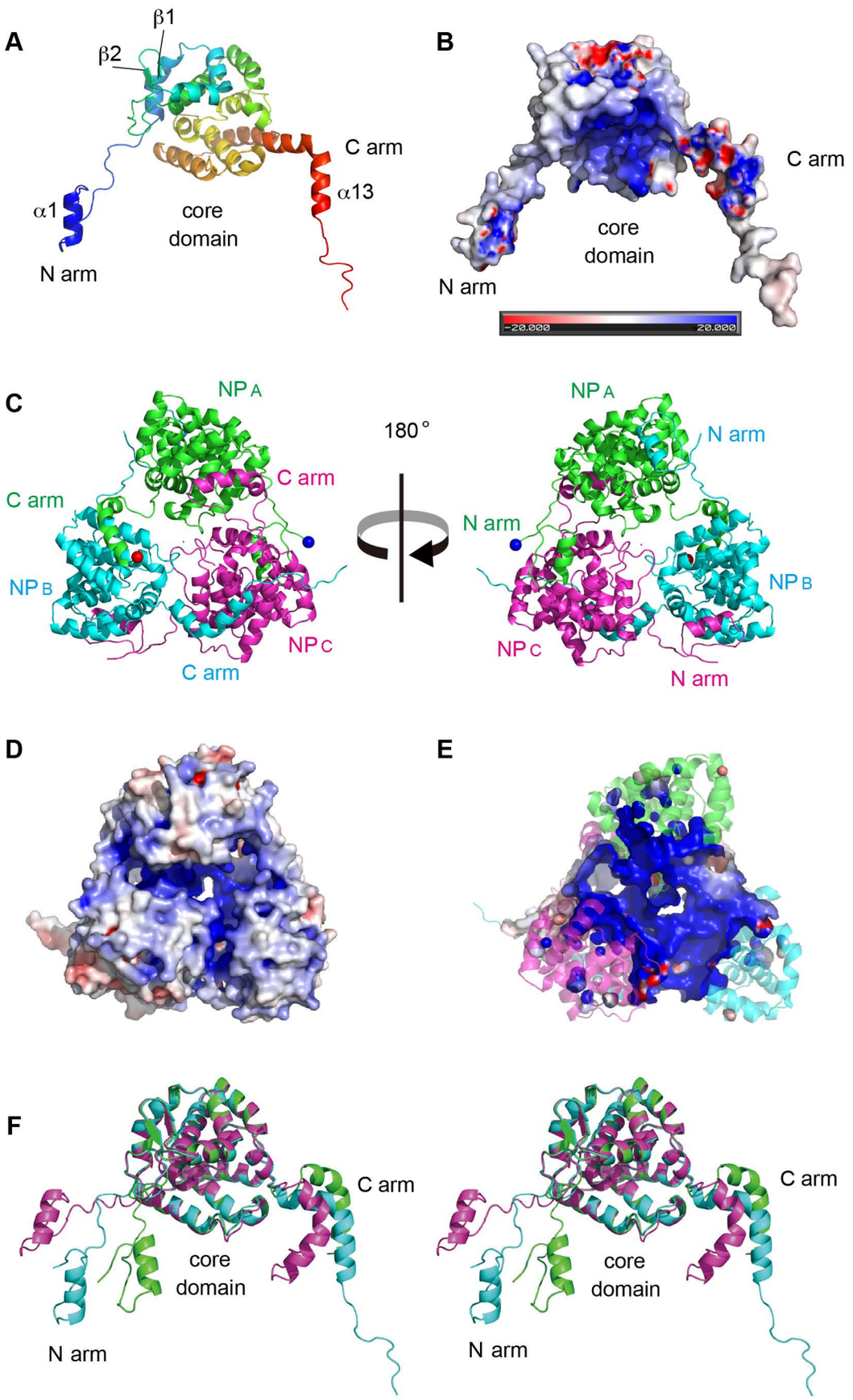
	N protein (SeMet)	N-RNA complex	N-DNA complex
Data collection			
Wavelength (Å)	0.97906	0.98000	1.00000
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	66.2, 96.0, 71.8	68.2, 94.7, 70.9	68.9, 86.5, 66.3
α , β , γ (°)	90, 113.4, 90	90, 112.6, 90	90, 112.1, 90
Resolution (Å) [a]	50–2.70 (2.87–2.70)	50–3.3 (3.5–3.3)	50–3.00 (3.18–3.00)
R_{meas} (%)[a,b]	9.7 (76.3)	5.2 (94.8)	8.0 (99.6)
$\langle I / s(I) \rangle$ [a]	10.8 (2.13)	16.5 (1.75)	17.1 (2.38)
Completeness (%)[a]	92.3 (98.5)	99.0 (97.8)	99.4 (98.0)
No. of unique reflections	41185 (7154)	12585 (1985)	14540 (2273)
Redundancy[a]	3.8 (3.8)	3.7 (3.7)	7.6 (7.6)
Refinement			
No. of reflections	41171	12581	14534
R_{work}/R_{free} (%)	21.04/25.52	28.23/31.66	25.05/29.13
No. of atoms			
Protein	5965	5439	5666
RNA/DNA	—	220	328

Averaged *B*-factors (Å^2)			
Protein	77.7	135.4	103.8
RNA/DNA	—	227.9	130
RMSD from ideal			
Bond lengths (Å)	0.003	0.002	0.005
Bond angles (°)	0.783	0.531	0.938
Ramachandran analysis [c]			
Favored (%)	96.95	98.22	95.82
Outliers (%)	0.27	0	0

[a]Values in parentheses are for the highest resolution shell

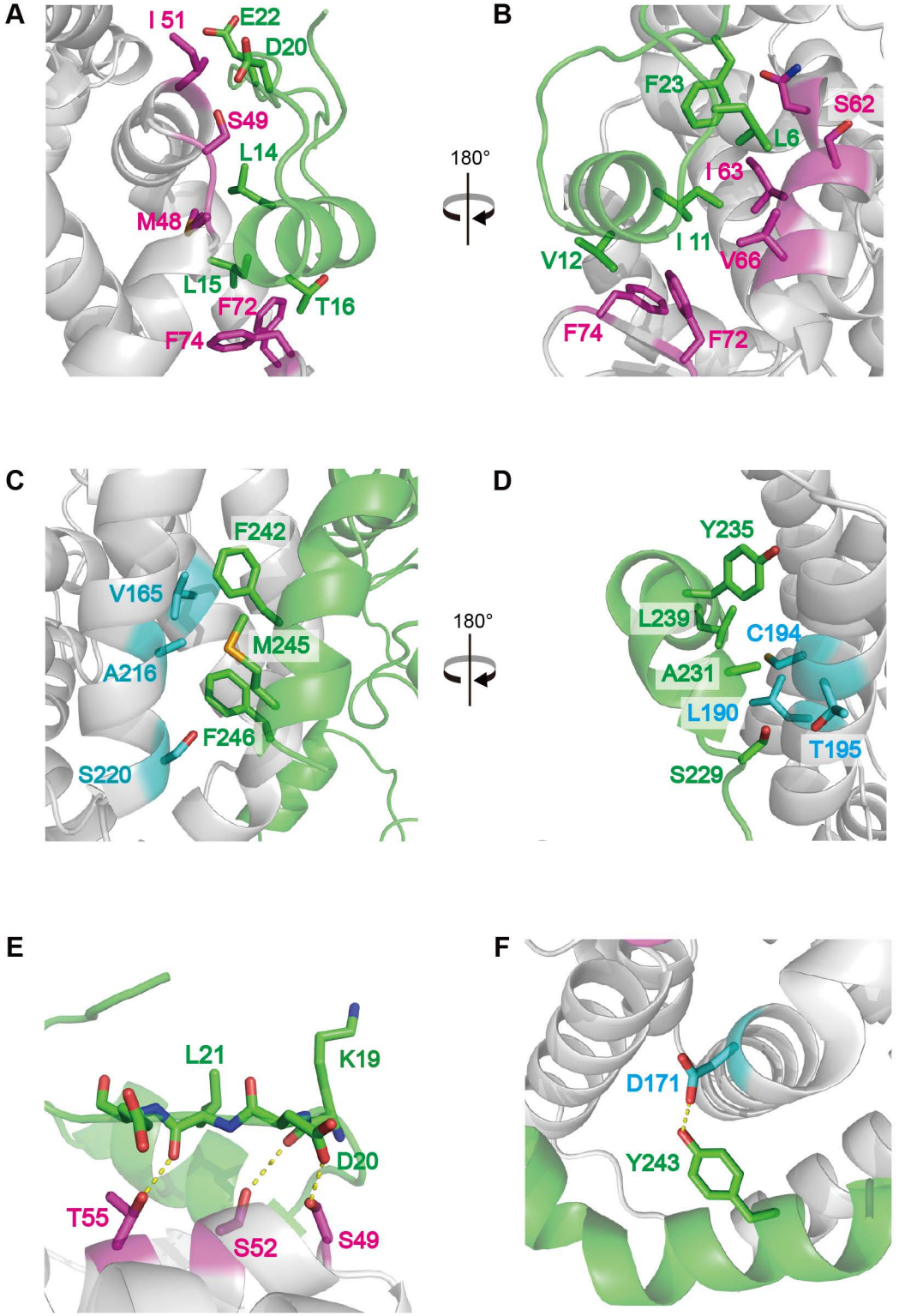
[b]R_{meas} is the redundancy-independent R_{sym}

[c]Ramachandran analysis was performed using MolProbity in the PHENIX package



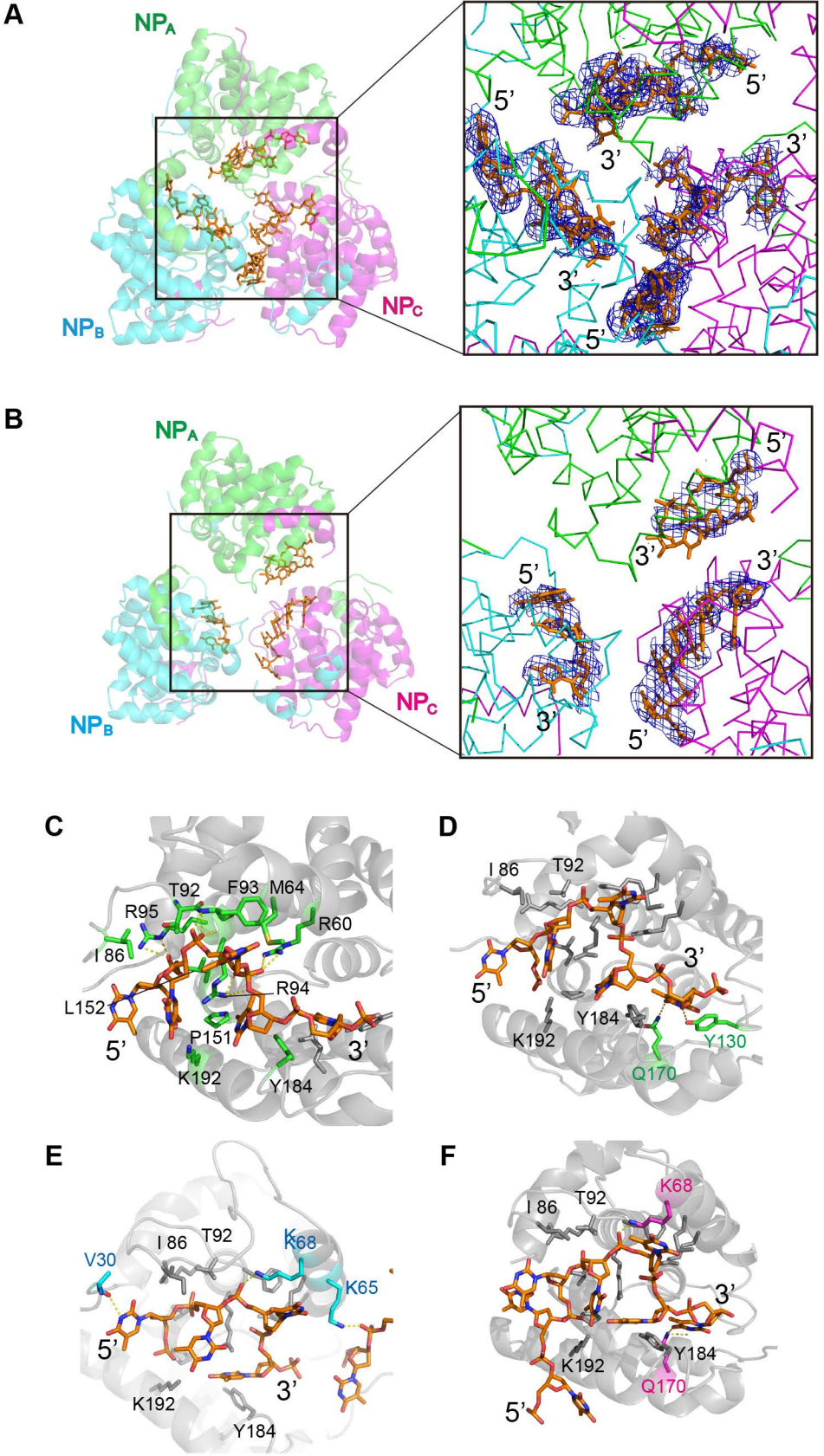
Komoda et al. **FIG 1**

FIG 1 Crystal structure of the TSWV N protein. (A) Structure of the TSWV N protein protomer (NP$_B$). 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$_B$ 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_BT and +20 k_BT, where k_B is Boltzmann's constant and T is temperature. (C) Cartoon representation of the TSWV N protein trimeric ring. NP$_A$, NP$_B$, and NP$_C$ are depicted in green, cyan, and magenta, respectively. The N and C termini of NP$_A$ 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 panel D at the level of the positively charged cleft. (F) Stereo view of the superposition of NP$_A$, NP$_B$, and NP$_C$. Each N protein molecule is shown using the same color as that described above for panel C.



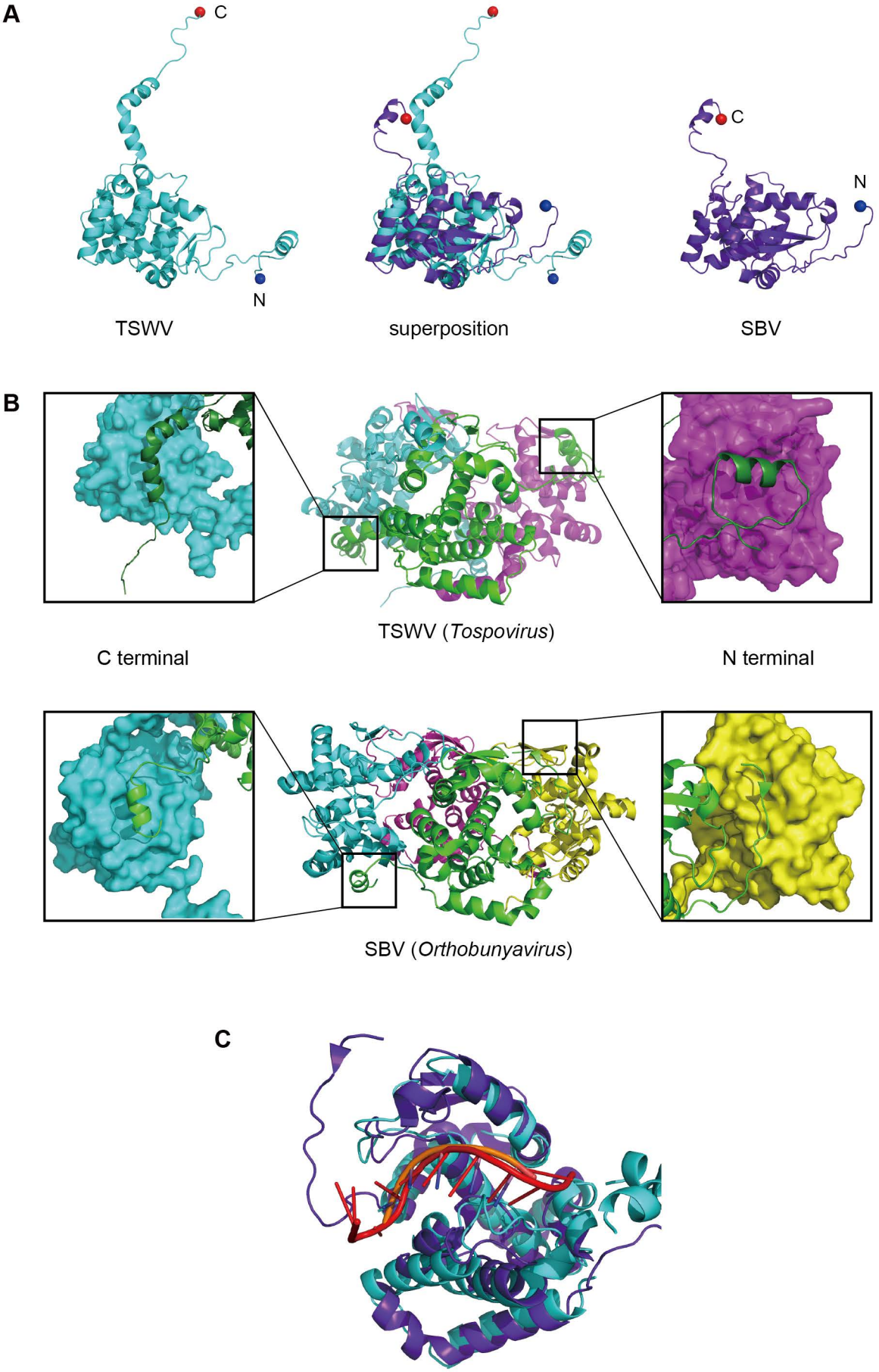
Komoda et al. **FIG 2**

FIG 2 Arm-to-core interactions of the TSWV NP trimer ring. (A) Interactions of the NP$_A$ N arm (green) and the NP$_C$ core domain (magenta). We depicted the representative hydrophobic residues that are involved in the interaction as sticks on one side of N arm helix α1 bound to a hydrophobic patch on the core domain formed by residues from α3, α4, and β1. (B) Opposite view of the image shown in panel A after 180° rotation. (C) Interactions of the NP$_A$ C arm (green) and the NP$_B$ core domain (cyan). We depicted the representative hydrophobic residues that are involved in the interaction as sticks on one side of the C arm helix α13 bound to a hydrophobic patch on the core domain formed by residues from α11 and α12. (D) Opposite view of the image shown in panel C after 180° rotation. (E) Magnified view of the hydrogen bonds of the NP$_A$ N arm (green) and the NP$_C$ core domain (magenta). (F) Magnified view of the hydrogen bonds between the NP$_A$ C arm (green) and the NP$_B$ core domain (cyan). Amino acid residues forming the interaction are displayed as stick models; dashed lines indicate hydrogen bonds.



Komoda et al. **FIG 3**

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 2Fo-Fc map of DNA contoured at 1.0 σ (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 Fig. 1C. Shown is a composite-omit 2Fo-Fc map of RNA contoured at 1.0 σ (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], F: NP_C [F]) are depicted in stick form in green (NP_A), cyan (NP_B), and pink (NP_C).



Komoda et al. **FIG 4**

FIG 4 Comparison of N protein structures of TSWV (*Tospovirus*) and SBV (*Orthobunyavirus*). (A) Protomer structure 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_A 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 SBV N-RNA complex (purple, SBV N; red, RNA).

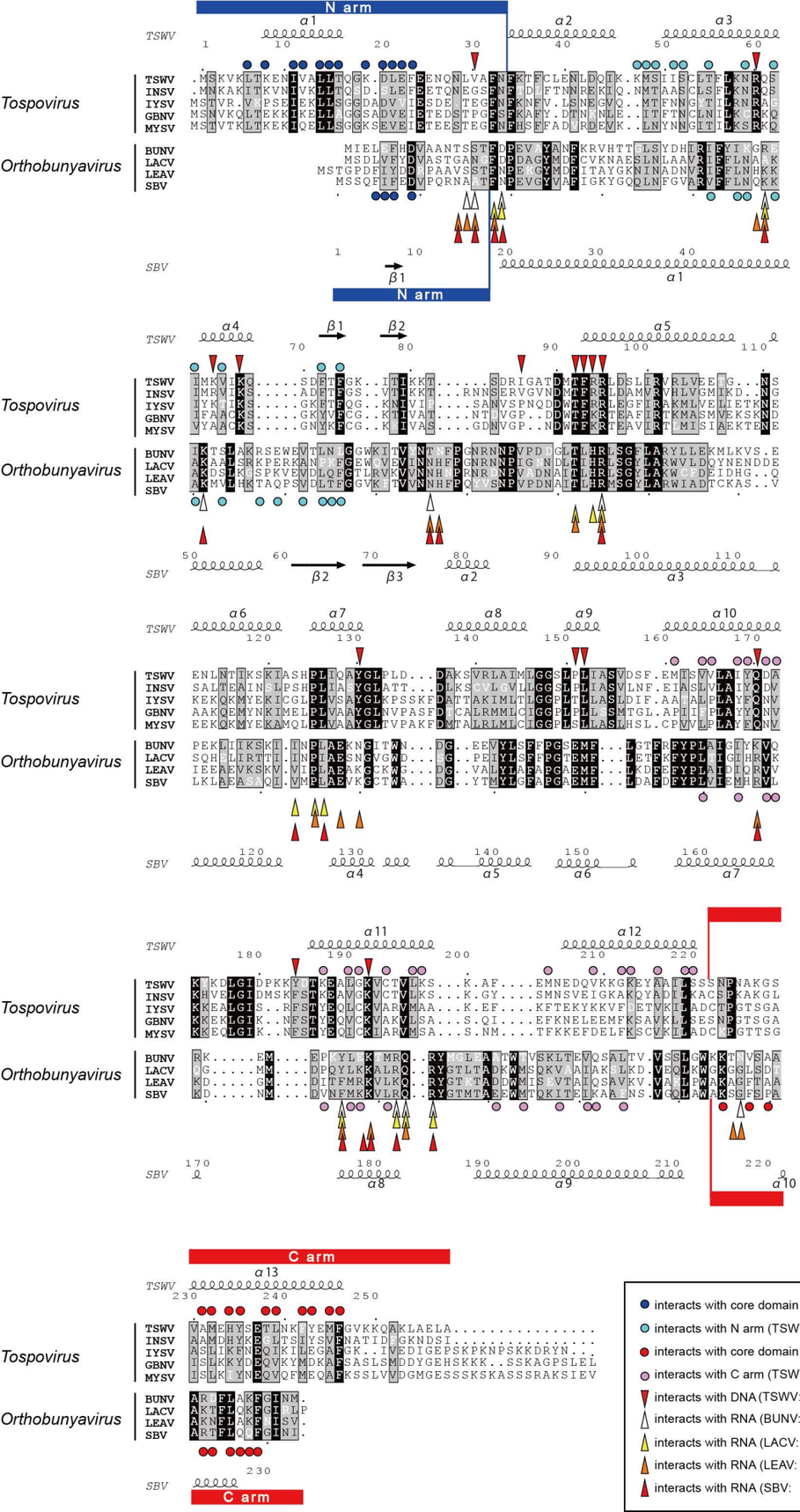


FIG 5 Amino acid sequence alignment of N proteins among tospoviruses and orhobunyaviruses. Key residues for nucleic acid binding are marked with downward arrowheads (red, TSWV) (the residues interact with DNA molecules) and upward arrowheads (white, *Bunyamwera virus* (BUNV); yellow, LACV; orange, *Leanyer 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 (TSWV and SBV). Key residues in 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 α-helix is depicted by a coil, and the β-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).