Supporting Information

Raymond et al. 10.1073/pnas.1213553109

SI Materials and Methods

Plasmid Construction. All plasmids were generated using standard molecular cloning techniques and confirmed by sequencing. The plasmid encoding Rift Valley fever virus (RVFV) N (pIPER1) has been described (1). To construct the pTOSVN plasmid, the full-length TOSV gene encoding N was amplified using primers 5′-agattggtggcATGTCAGACGAGAATTATCGC and 5′-gaggagagtttagacattaCTTGCCAACCTTGGCG (IDT). The PCR product was processed with T4 DNA polymerase (Promega) and dGTP (Invitrogen) to generate an overhang for ligation-independent cloning (LIC) and then incubated with processed LIC-SUMO vector (pETHSUL), which encodes a His_6-SUMO fusion protein and was provided by P. Loll (Drexel University, Philadelphia, PA) (2). The mixture was transformed into *E. coli* XL1-Blue–competent cells for ligation and amplification of the plasmid.

Production and Purification of Recombinant Proteins. Recombinant RVFV N and TOSV N were purified under native or denaturing conditions, as described previously (1). Briefly, recombinant proteins were purified by Ni-affinity chromatography, the His_6-SUMO tag was removed by SUMO hydrolase, and proteins were further purified by gel filtration. RNA-free proteins were generated by Ni-affinity purification, denaturation, refolding, tag cleavage, and gel filtration. Proteins were stored at −80 °C in 20 mM Tris (pH 7.8), 500 mM NaCl, and 5% (vol/vol) glycerol.

N Oligomeric State in Solution. When purified under native conditions with extensive ribonuclease treatment, recombinant RVFV N was in a complex with endogenous *E. coli* RNA. The final gel-filtration purification step (Superdex200) yielded two fractions: ∼90% of the "natively purified" N was a multimer of apparent molecular weight 100 kDa and an A_{260} /A_{280} ratio of 1.32, and ∼10% was in a monomeric state with an A_{260}/A_{280} ratio of 0.75 (1). Thus, both the natively purified monomer and multimer were contaminated with nucleic acid, because the theoretical A_{260}/A_{280} ratio for pure N protein is ∼0.57 (3). Single-particle EM analysis of the multimeric species revealed a heterogeneous mixture of oligomers with four to seven subunits. SDS/PAGE analysis of both fractions showed a single polypeptide with a molecular mass of 27 kDa, corresponding to N (1). Under denaturing conditions (8 M urea) followed by refolding, recombinant RVFV N purified predominantly as a monomer with less than 5% in a dimeric state (1). Refolded N was essentially free of RNA, as judged by the A_{260}/A_{280} ratio of 0.56 (1).

Fluorescence Polarization Binding Experiments. Fluorescence polarization (FP) experiments were performed using natively purified or refolded, RNA-free N proteins and PAGE-purified, 5′-6-carboxyfluorescein (6-FAM) fluorescein-labeled single-stranded RNA or DNA oligomers (IDT). Before FP experiments, the N proteins were dialyzed against FP buffer [20 mM Tris (pH 7.8), 50 mM NaCl]. FP measurements were performed on a SpectraMax M5 (Molecular Devices) using 96-well black microplates (Fluotrac 200; Greiner Bio One) and a reaction mixture containing 10 mM Tris (pH 7.8), 25 mM NaCl, and 5 nM RNA or DNA. N was serially diluted 10-fold into FP buffer, and 45 μL of each dilution was added to 45 μL of 10 nM labeled RNA or DNA. All experiments were done in duplicate with appropriate controls. Binding constants (K_d) were determined by fitting FP data by nonlinear least-squares regression to FP = $B_{max} \times$ [N]/(K_d + [N]), where B_{max} is the FP signal at maximum binding (4) in KaleidaGraph (version 4.1; Synergy Software). RNA binding to the RVFV recombinant N multimer did not reach saturation, so B_{max} was approximated as the average millipolarization (mP) values for fully bound RNA-free N and recombinant N monomer.

Electron Microscopy. Negative-stained samples for EM were prepared as described (1). RVFV N-DNA complexes were produced for screening by 1-h incubation of RNA-free N with PAGE-purified DNA oligomers at a ratio of eight N to one DNA poly-thymidine oligomer. Imaging was performed at room temperature with a Tecnai T12 transmission electron microscope operated at an acceleration voltage of 120 kV. Images of RVFV N-DNA_{35} and N-DNA_{40} were recorded on a mounted Gatan US4000 CCD camera at a magnification of 71,138× and a defocus value of ∼1.5 μm. All images were binned (2 × 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. For 2D reference-free alignment and classification of particles, 3,770 particles of RVFV N-DNA_{35} were selected from micrographs using EMAN Boxer (5). Reference-free alignment, classification into ninety-nine groups, and class averaging were performed with EMAN (5).

Crystallization of N Proteins. RVFV N-RNA_{28}, N-RNA_{35}, and N-DNA_{30} complexes were produced by 1-h incubation of RNA-free N with PAGE-purified nucleic acid at a ratio of eight N to one RNA poly-uracil oligomer or DNA poly-thymidine oligomer. Complexes were separated from excess N by gel filtration. Before crystallization, all samples were dialyzed against crystallization buffer [20 mM Tris (pH 7.8), 250 mM NaCl] and maintained at a concentration of ∼10 mg/mL. All crystals were grown by hanging-drop vapor diffusion, cryoprotected by a 10-s soak in well solution augmented with cryoprotectant, harvested into loops, and flash-cooled by plunging into liquid N_2.

RVFV N_4-RNA_{28} crystallized at 20 °C in ∼7 d from a 1:1 mixture of N-RNA_{28} and well solution containing 18% (wt/vol) PEG 3350, 350 mM NaCl, and 100 mM Bis-Tris (pH 5.5). RVFV N_5-RNA_{35} crystallized at 20 °C in ∼4 wk from a 1:1 mixture of N-RNA_{35} and well solution containing 28% (wt/vol) PEG 3350 and 280 mM $(NH_4)_2SO_4$. RVFV N_6-RNA_{35} crystallized at 20 °C in ∼4 wk from a 1:1 mixture of N-RNA_{35} and well solution containing 24% (wt/vol) PEG 3350 and 350 mM $(NH_4)_2SO_4$. RVFV N-DNA_{30} crystallized at 20 °C in ∼4 d from a 1:1 mixture of N-DNA_{30} and well solution containing 15% (wt/vol) PEG 3350, 350 mM NaCl, and 100 mM Bis-Tris (pH 5.5). RVFV N crystallized at 20 °C in ∼7 d from a 1:1 mixture of protein and well solution containing 15% (wt/vol) PEG 3350 and 150 mM $(NH_4)_2SO_4$. Recombinant, natively purified TOSV N crystallized at 4 °C in ∼7 d from a 1:1 mixture of protein and well solution containing 23% (vol/vol) isopropanol, 280 mM $MgCl_2$, and 100 mM Hepes (pH 7.5). Cryoprotectants were 10% (vol/vol) glycerol for RVFV N_5-RNA_{35}, RVFV N_6-RNA_{35}, RVFV N_6-DNA_{30}, and RVFV N_6; 15% (vol/vol) glycerol for RVFV N_4-RNA_{28}; and 30% (vol/vol) MPD for TOSV N_6.

Data Collection and Structure Determination. Diffraction data were collected at 100 K on GM/CA beamlines 23ID-B (RVFV N_4-RNA_{28}, N_5-RNA_{35}, N_6-RNA_{35} and N_6) and 23ID-D (RVFV N_6-DNA_{30} and TOSV N_6) at the Advanced Photon Source, Argonne National Laboratory at λ = 1.0332 Å. Diffraction images were indexed and integrated using iMOSFLM (6), and data were scaled using SCALA (7), both part of the CCP4 suite (8). A truncated version of monomeric RVFV N [PDB ID code 3LYF

(1); residues 35–245] was used as a probe for molecular replacement with PHASER (9). Modeling was completed manually using Coot (10) and RCrane (11). Refinement was performed using BUSTER (12) with noncrystallographic symmetry (NCS) restraints and translation-libration-screw (TLS) parameterization of molecular motion (13). Individual isotropic thermal parameters were refined for structures with d_{min} less than 3.0 Å, and grouped isotropic thermal parameters were refined for those with d_{min} greater than 3.0 Å. Ramachandran analysis and structure validation were performed with MolProbity (14). Data and model quality are summarized in Table S2. PyMOL was used for structural alignments and to generate figures (15). The APBS plugin in PyMOL was used to calculate electrostatic surface potentials (16). The range of motion for the N helical arm was calculated with the DynDom server (17).

Nucleic Acid Electron Density. In all N–nucleic acid structures, the multimer unit is a ring of N subunits. The nucleic acids used in crystallization were homooligomers (poly-U or poly-dT) that lacked terminal phosphates. Positive difference density corresponding to nucleic acid was observed in the RNA-binding slot for structures of reconstituted N–nucleic acid in maps calculated with phases from molecular replacement with an RVFV N probe. The ends of the nucleic acid oligomers were distributed randomly within the multimers, so electron density was continuous around the RNA-binding slot with no visible termini (Fig. S4). In the N_4-RNA_{28} and N_5-RNA_{35} structures, random binding would result in an undetectable reduced occupancy at the phosphate positions by 1/28 and 1/35, respectively, because each N subunit bound seven nucleotides. In the N_6-RNA_{35} structure, nucleotides were observed in 36 positions because each N subunit bound 6 nt.

DNA Binding in the RNA-Binding Slot. DNA and RNA bound somewhat differently in the RNA-binding slot. The number of DNA nucleotides bound to each N subunit varied. Density, although continuous, was visible for only 28 of 30 nt. Four subunits of the N_6-DNA_{30} hexamer had no nucleotide in position 2. In the other two subunits, B2 occupied a different position within the RNA-binding slot than in the N-RNA complexes and did not stack with Phe33. At two N-N subunit interfaces, the N_6-DNA_{30} hexamer lacked nucleotides at both positions 6 and 7, whereas in the N-RNA hexamer and pentamer structures, B6 and B7 stack with B1′. There were also significantly fewer interactions with the DNA sugar-phosphate backbone in all subunits of the N_6-DNA_{30} hexamer than with the RNA sugar-phosphate backbone in the N-RNA complexes. Only three of the eight backbone interactions seen in the N-RNA structures were observed in the N_6-DNA_{30} hexamer: Lys67 and the Asn66 main-chain amide to P5 and Lys70 to P1′. Despite these differences, N had equal affinities for RNA and DNA (Table S1 and Fig. S1). Density was visible for only 28 nt in N_6-DNA_{30}. Because the hexamer sits on a crystallographic twofold, there are two gaps in electron density corresponding to the chain ends.

Volume Calculations. RNP volume calculations were based on average protein and RNA partial specific volumes (18) and on a tripartite genome with a total of 11,980 nt, one N per 7 nt and one RdRp per genome segment. The calculated total volume for a tripartite encapsidated genome is $\sim 61 \times 10^6$ Å^3. The calculated volume for a tripartite encapsidated genome with a ratio of 1:4:4 large:medium:small segments (19) is $\sim 148 \times 10^6$ Å^3. The average inner volume of the RVFV virus particle, $\sim 150 \times 10^6$ Å^3, is from an EM reconstruction of RVFV (20).

1. 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 USA* 107(26):11769–11774.
2. Weeks SD, Drinker M, Loll PJ (2007) Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr Purif* 53(1):40–50.
3. Glasel JA (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques* 18(1):62–63.
4. Motulsky HJ, Neubig RR (2010) Analyzing binding data. *Curr Protoc Neurosci* Chapter 7:Unit 7.5.
5. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128(1):82–97.
6. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271–281.
7. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72–82.
8. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
9. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4): 658–674.
10. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
11. Keating KS, Pyle AM (2010) Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc Natl Acad Sci USA* 107(18):8177–8182.
12. Bricogne G, et al. (2010) BUSTER (Global Phasing, Cambridge, UK), Version 2.9.
13. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62(Pt 4): 439–450.
14. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375-83.
15. Delano WL (2007) The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).
16. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.
17. Lee RA, Razaz M, Hayward S (2003) The DynDom database of protein domain motions. *Bioinformatics* 19(10):1290–1291.
18. Voss NR, Gerstein M (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J Mol Biol* 346(2):477–492.
19. Gauliard N, Billecocq A, Flick R, Bouloy M (2006) Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis. *Virology* 351(1):170–179.
20. Freiberg AN, Sherman MB, Morais MC, Holbrook MR, Watowich SJ (2008) Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J Virol* 82(21):10341–10348.

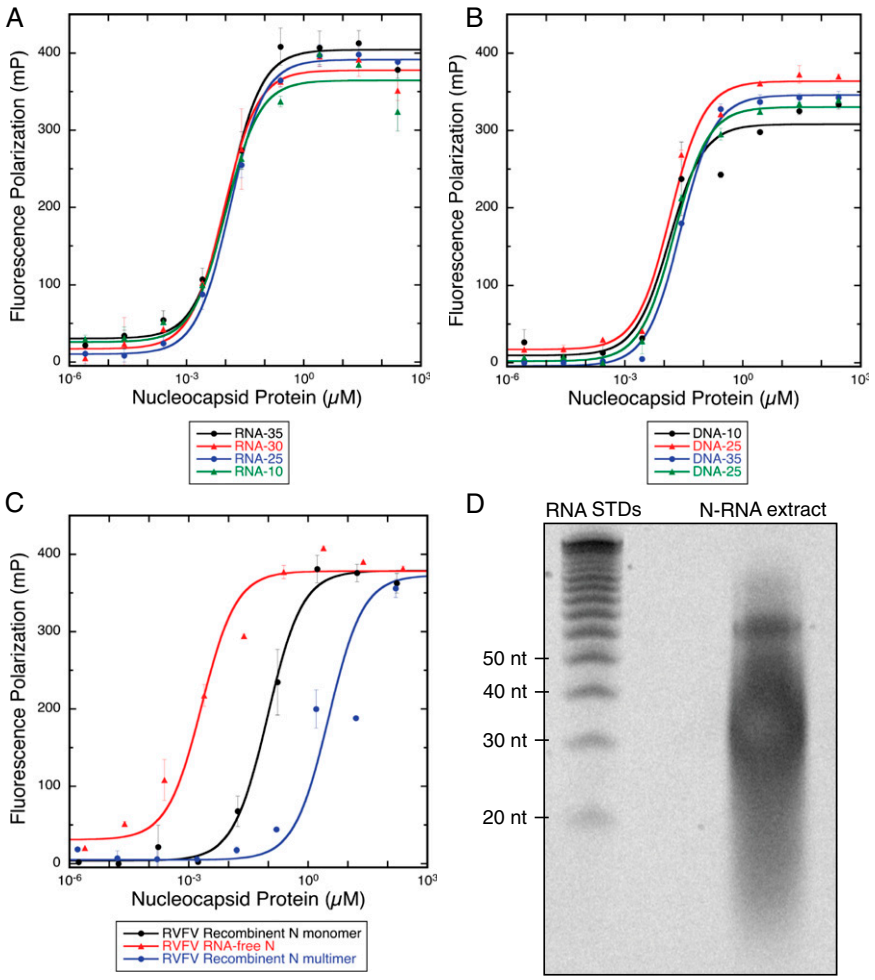

Fig. S1. Analysis of N binding to nucleic acid. (*A*–*C*) Fluorescence polarization measurement of RNA-free N binding to RNA (*A*), RNA-free N binding to DNA (*B*), and RNA_{25} binding to N preparations with varying RNA contamination (*C*), as detected from the ratio of absorbances at 260 and 280 nm (recombinant monomer, A_{260}/A_{280} = 0.75; recombinant multimer, A_{260}/A_{280} = 1.32; RNA-free N, A_{260}/A_{280} = 0.56). (*D*) Analysis of RNA extracted from recombinant, natively purified RVFV N multimer. After extensive ribonuclease digestion, RNA was extracted from the N multimer with an equal-volume mixture of phenol and chloroform. The extracted RNA was separated on a denaturing urea-polyacrylamide gel and visualized using SYBR Green. Single-stranded RNA standards are shown in the first lane.

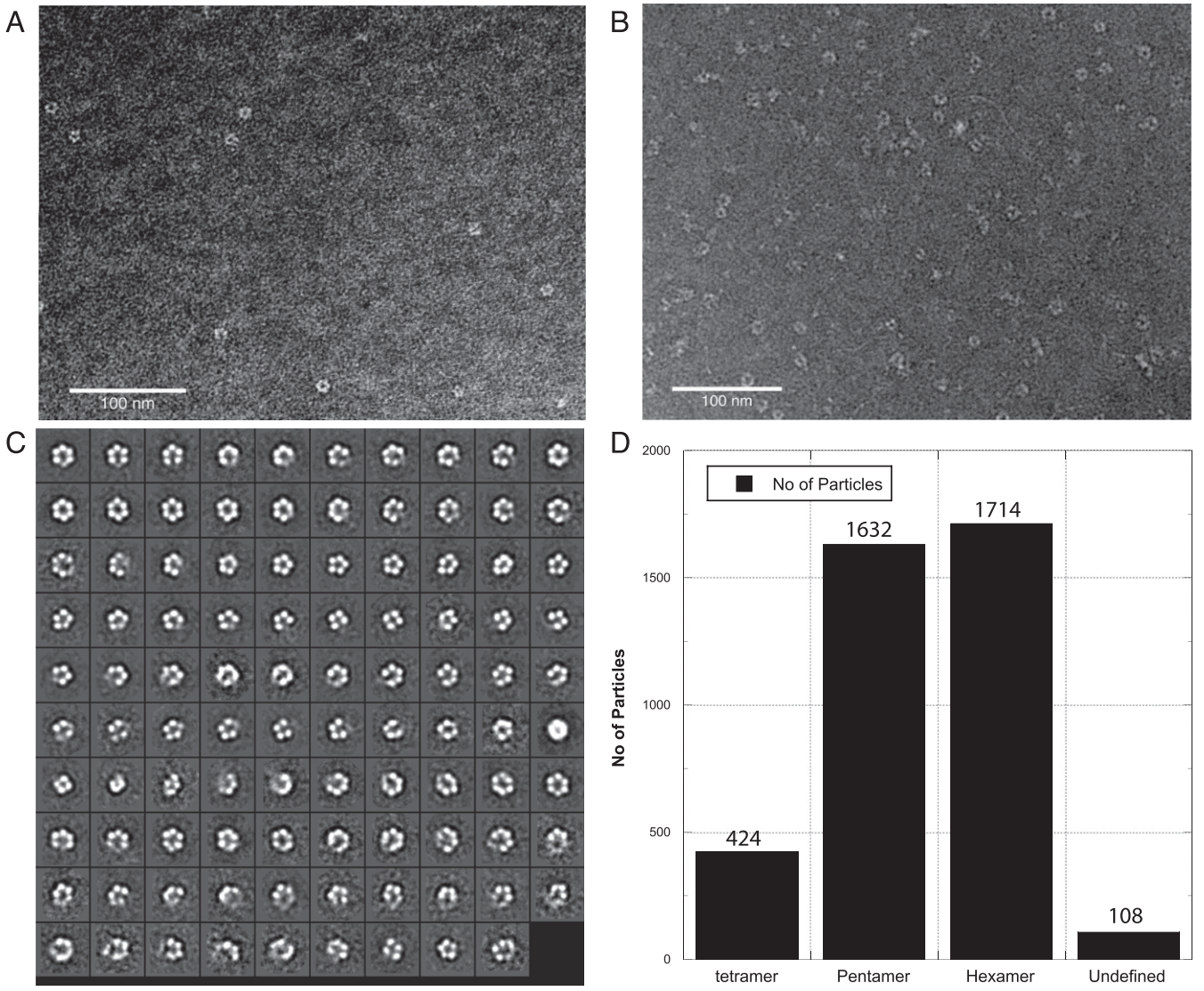

Fig. S2. EM of RVFV N-DNA. (*A* and *B*) Comparative EM fields of negatively stained N-DNA_{35} (*A*) and N-DNA_{40} (*B*) complexes. (C) Reference-free classifications of 3770 N-DNA_{35} particles in 99 classes. (*D*) Histogram of oligomeric state distribution of N-DNA_{35} particles.

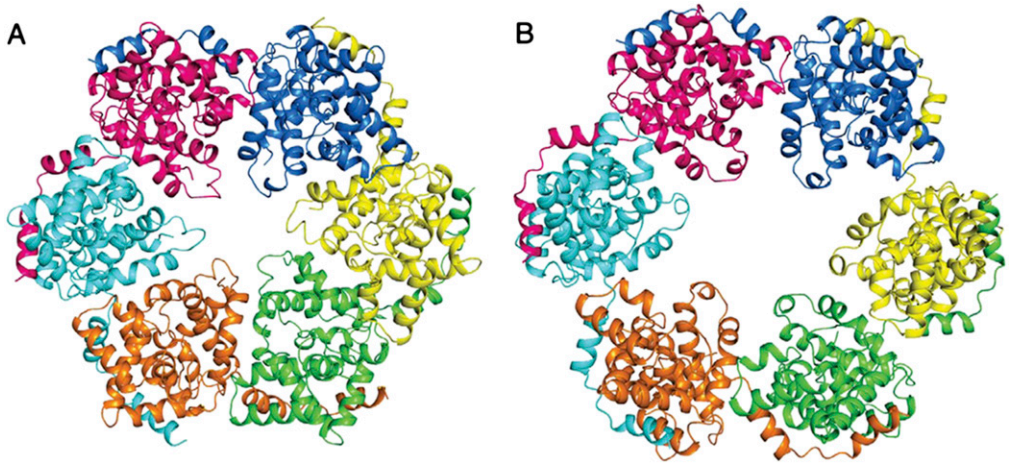

Fig. S3. Structures of phlebovirus N. (*A*) RVFV N_6. (*B*) TOSV N_6. N subunits are rendered in ribbon form with contrasting colors. The helical arm of each subunit wraps around the neighboring subunit on the outside of the multimer. Note that both the RVFV and the TOSV N hexamer deviate from perfect sixfold symmetry.

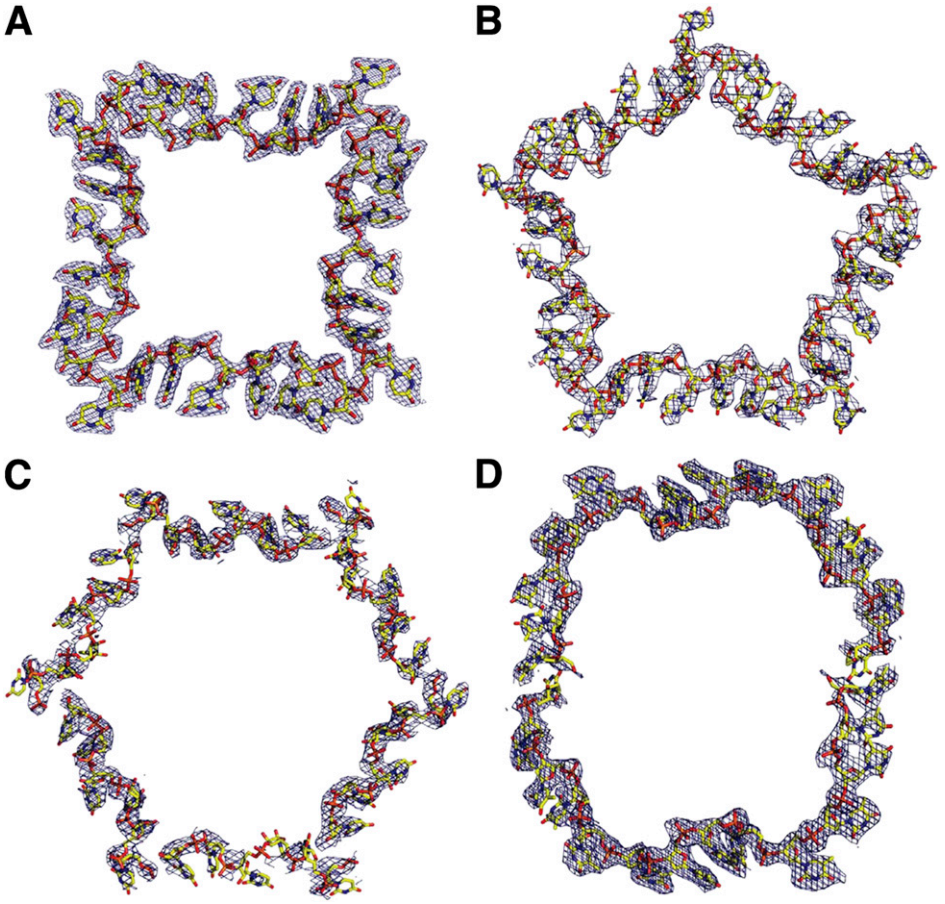

Fig. S4. Electron density of nucleic acid bound to N. Electron density ($2F_o$–F_c contoured at 1σ) showing nucleic acid in the RNA-binding slot in crystal structures of N_4-RNA_{28} at 2.15 Å (*A*), N_5-RNA_{35} at 3.90 Å (*B*), N_6-RNA_{35} at 3.40 Å (*C*), and N_6-DNA_{30} at 2.70 Å (*D*). Nucleic acids are rendered in stick form with yellow C atoms.

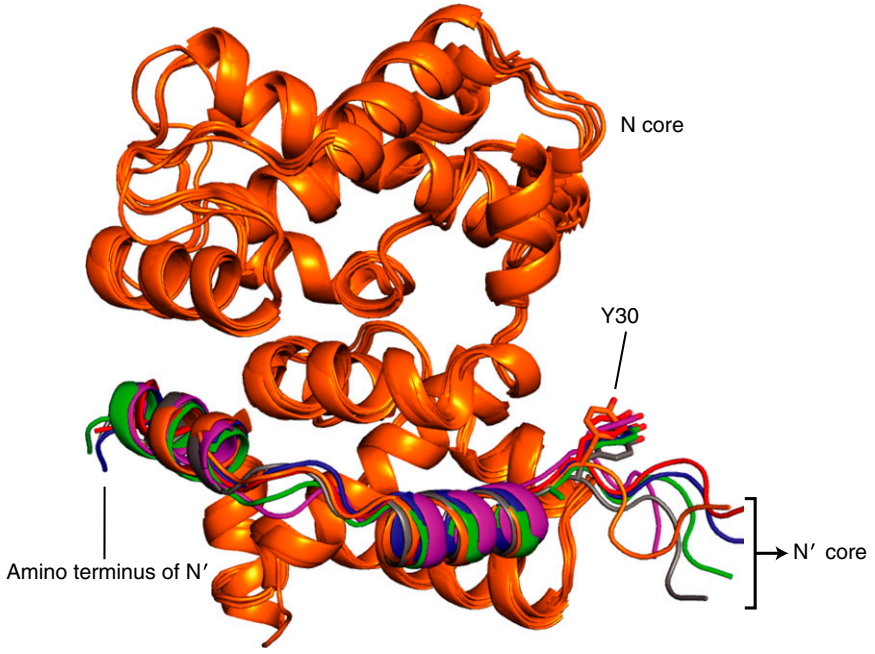

Fig. S5. Superposition of N cores with the associated helical arms from neighboring subunits. The N cores are rendered in ribbon form and colored orange. The arm–core interface buries more than 1,300 $Å^2$ of surface area per subunit in all structures, regardless of RNA content. The helical arms include residues 5–30 and are colored by structure. Orange, RVFV N_6-RNA_{35}; green, RVFV N_5-RNA_{35}; blue, RVFV N_4-RNA_{28}; gray; RVFV N_6, magenta, TOSV N_6; red, RVFV N_6-DNA_{30}.

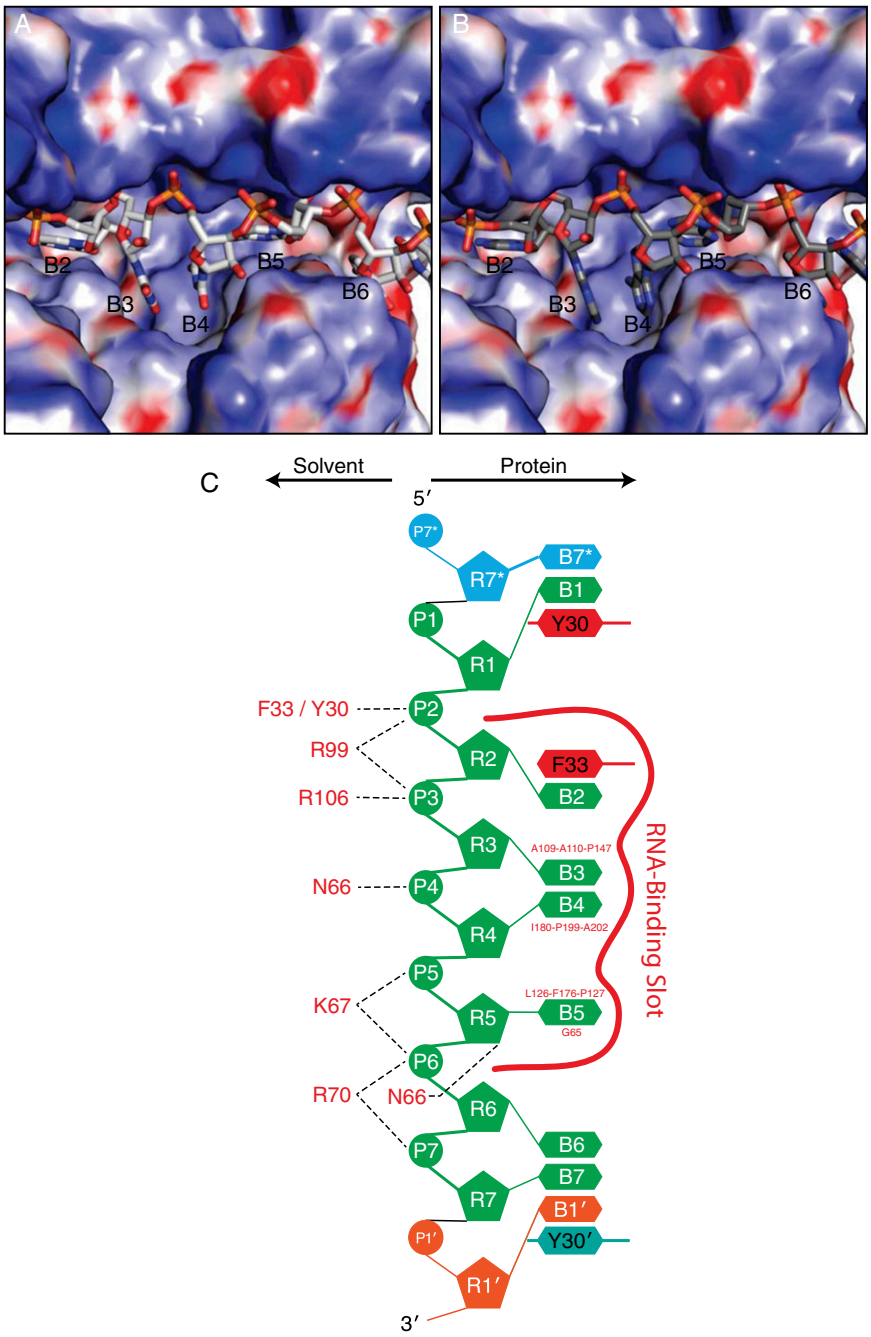

Fig. S6. Nucleotides in the RVFV N RNA-binding slot. (*A* and *B*) Observed uracil nucleotides from RVFV N_4-RNA_{28} (*A*) and modeled adenine nucleotides (*B*) are rendered in stick form. The surface potential from −14 kT in red to +14 kT in blue is shown for the RNA-binding slot. (*C*) Schematic diagram of RVFV N interactions with RNA. P, R, and B represent RNA phosphate, ribose, and base, respectively. Amino acids from adjacent N subunits are shown in red and cyan. RNA nucleotides in blue with an asterisk label (*) and in orange with a prime symbol label (′) interact with adjacent N subunits.

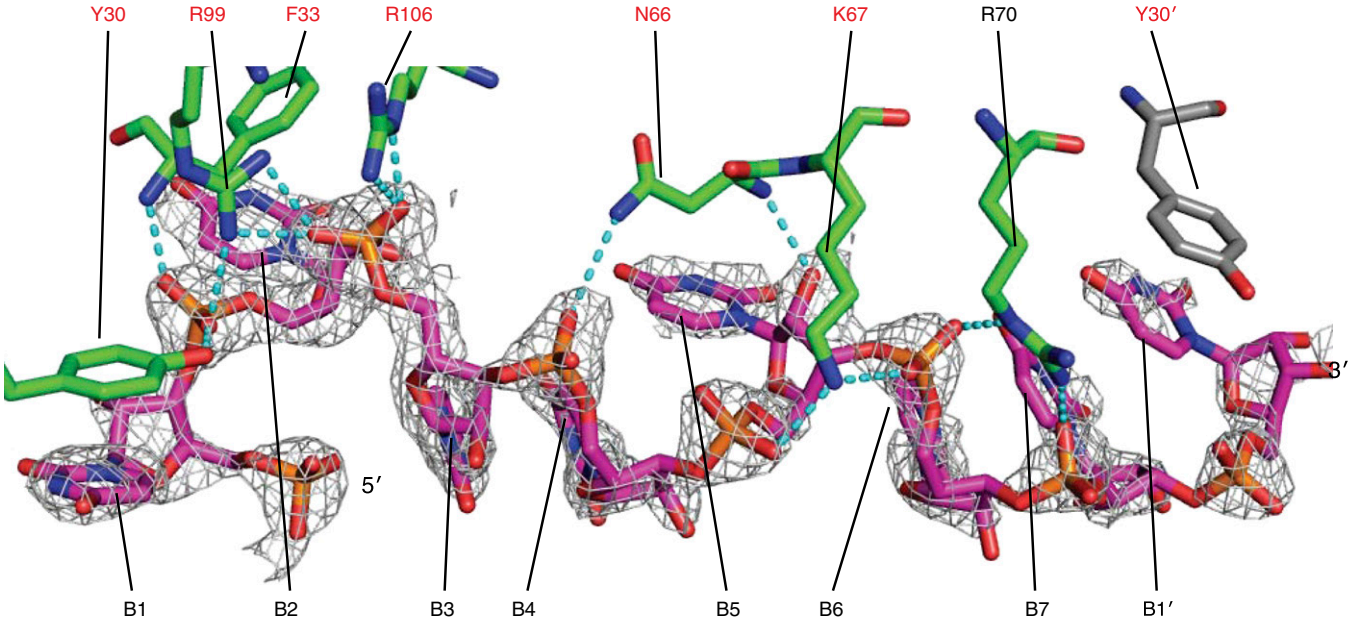

Fig. S7. Unbiased electron density for RNA bound to phlebovirus N. Amino acids and RNA nucleotides are rendered in stick form. The carbon atoms in amino acids from adjacent subunits are colored green and gray. The 2.15-Å F_o–F_c omit electron density contoured at 3σ shows RNA bound to an N subunit in RVFV N_4-RNA_{28}. Amino acids labeled in red are invariant in phlebovirus N proteins. RNA bases 1–7 are labeled B1 to B7; B1′ is B1 of the adjacent N subunit. Dashed lines represent hydrogen bonds.

Table S1. N binding to nucleic acid oligomers

Type	Name	Sequence	Length	Protein	A_{260}/A_{280}*	K_d (nM)
RNA	U_{25}	U_{25}	25	RVFV Recombinant N monomer†	0.75	96 ± 14
RNA	U_{25}	U_{25}	25	RVFV Recombinant N multimer†	1.32	>4,800
RNA	U_{25}	U_{25}	25	RVFV RNA-free N‡	0.56	13 ± 1
RNA	U_{30}	U_{30}	30	RVFV RNA-free N	0.56	9 ± 2
RNA	U_{35}	U_{35}	35	RVFV RNA-free N	0.56	11 ± 2
RNA	RNA_{10}	AAC GCU UCC C	10	RVFV RNA-free N	0.56	10 ± 3
DNA	T_{10}	dT_{10}	10	RVFV RNA-free N	0.56	13 ± 3
DNA	T_{25}	dT_{25}	25	RVFV RNA-free N	0.56	14 ± 3
DNA	T_{35}	dT_{35}	35	RVFV RNA-free N	0.56	26 ± 3
DNA	DNA_{25}	$d(ACC)_8$ A	25	RVFV RNA-free N	0.56	18 ± 3
DNA	T_{25}	dT_{25}	25	TOSV RNA-free N	0.56	6 ± 2
RNA	U_{25}	U_{25}	25	TOSV RNA-free N	0.56	14 ± 3

*Absorbance ratio of purified N preparation used for binding measurement.
†Recombinant RVFV N yielded both monomeric and multimeric species after digestion with ribonuclease.
‡RNA-free N was generated by denaturation and refolding.

Table S2. Crystallographic summary

	RVFV N_4-RNA_{28}	RVFV N_5-RNA_{35}	RVFV N_6-RNA_{35}	RVFV N_6-DNA_{30}	RVFV N_6	TOSV N_6
Diffraction data						
Space group	C2	P1	P1	$P6_422$	$P3_121$	P1
Unit cell lengths (Å), a, b, c	77.6, 193.2, 77.4	79.8, 93.6, 124.7	91.6, 173.3, 172.9	108.6, 108.6, 261.3	107.1, 107.1, 258.4	50.3, 93.7, 95.5
Unit cell angles (°), α, β, γ	90.0, 108.9, 90.0	101.7, 90.3, 114.2	119.9, 99.3, 90.1	90.0, 90.0, 120.0	90.0, 90.0, 120.0	67.9, 85.9, 87.9
X-ray source	APS 23 ID-B	APS 23 ID-B	APS 23 ID-B	APS 23 ID-D	APS 23 ID-B	APS 23 ID-D
Wavelength (Å)	1.0332	1.0332	1.0332	1.0332	1.0332	1.0332
d_{min} (Å)	2.15 (2.27–2.15)*	3.90 (4.11–3.90)	3.40 (3.58–3.40)	2.70 (2.85–2.70)	3.10 (3.27–3.10)	2.75 (2.90–2.75)
Unique reflections	55,733	28,961	117,933	25,771	32,046	39,945
R_{merge}†	0.11 (0.58)	0.21 (0.46)	0.19 (0.51)	0.11 (0.78)	0.09 (0.69)	0.08 (0.47)
Average I/σ_I	9.3 (2.3)	2.8 (1.6)	4.0 (1.7)	10.0 (2.0)	13.1 (2.4)	6.8 (1.7)
Completeness (%)	95.5 (77.9)	99.1 (99.1)	94.7 (92.3)	99.6 (98.5)	91.9 (89.7)	95.2 (94.2)
Average redundancy	3.7 (3.3)	2.0 (2.0)	1.9 (1.9)	4.4 (4.3)	6.2 (5.0)	1.9 (1.9)
Refinement						
Data range (Å)	31.48–2.15	41.6–3.90	57.42–3.40	53.66–2.70	46.4–3.10	50.22–2.75
Reflections	55716	28958	117932	25749	31,959	39,944
R/R_{free}‡	0.180/0.219	0.228/0.248§	0.224/0.240§	0.189/0.230	0.215/0.256	0.218/0.238
rms deviations						
Bond lengths (Å)	0.009	0.008	0.008	0.010	0.009	0.009
Bond angles (°)	0.93	0.87	0.85	1.06	1.02	0.99
Average B factor (Å^2)						
Protein	39.6	56.8	70.5	60.2	122.5	83.0
Water	37.2	NA	NA	45.8	NA	28.4
DNA/RNA	36.0	36.8	117.2	81.6	NA	NA
Ramachandran¶						
Allowed (%)	99.38	98.90	98.80	99.58	99.10	99.50
Outlier (%)	0.62	1.10	1.20	0.42	0.90	0.50
No. of atoms						
Protein	7,636	19,093	67,752	5,698	11,404	11,221
Water	545	0	0	56	0	18
DNA/RNA	559	1400	4320	277	NA	NA
ASU content	2× 1/2(N_4-RNA_{28})	2×(N_5-RNA_{35})	6×(N_6-RNA_{36})	N_3-DNA_{13}	N_6	N_6
PDB ID code	4H5P	4H5O	4H6F and 4H6G	4H5Q	4H5M	4H5L

APS, Advanced Photon Source.

*Values in parenthesis are for the outermost shell of data.

†$R_{merge} = \sum|I_i - \langle I\rangle|/\sum I_i$, where I_i is the intensity of the ith observation, and $\langle I\rangle$ is the mean intensity. Sums are taken over all reflections.

‡$R = \sum||F_o| - |F_c||/\sum|F_o|$. R_{free} is calculated for a 5% subset of the data.

§For structures with 10 or more N subunits in the asymmetric unit, reflections in the R_{free} set were chosen in thin shells.

¶Calculated with MolProbity (14).