Supporting Information

Guo et al. 10.1073/pnas.1407122111

SI Materials and Methods

Cloning, Expression, and Purification. The DNA sequence corresponding to nucleotides 86–1261 from Orsay virus RNA2 (GenBank accession no. HM030971.2) was cloned into the vector pET28a. This sequence encodes the putative capsid protein (CP) ORF, 392 residues in total, starting from 138 residues upstream of the first Met to the first predicted stop codon. A six-histidine tag was added to the C terminus of the Orsay virus CP to facilitate protein purification. Recombinant full-length CP was expressed in *Escherichia coli* Rosetta cells, and the cell pellets were sonicated in a lysis buffer containing 50 mM Tris·HCl (pH 8.0), 300 mM NaCl, and 10% (vol/vol) glycerol. Full-length CP was first purified by Ni-NTA affinity (Qiagen), heparin affinity, and a Superdex 200 gel-filtration column (GE Healthcare). To purify the virus-like particles (VLPs) only, cell lysates were precipitated by adding 20% (wt/vol) ammonium sulfate, and the VLPs were collected by centrifugation at 15,000 × *g* for 30 min. The pellets were solubilized in lysis buffer and subjected to ultracentrifugation in the CsCl gradient. After centrifugation at 38,000 rpm in a Beckman SW Ti 45 rotor for 6 h, the lower band at ∼1.3 g/cm^3 was collected and dialyzed against the storage buffer containing 50 mM Tris·HCl (pH 7.5), 500 mM NaCl, 10% (vol/vol) glycerol, 1 mM $CaCl_2$, and 1 mM $MgCl_2$.

To identify a stable core of the Orsay virus CP, the oligomer fraction from the Superdex 200 gel-filtration chromatography was treated with trypsin at a 200:1 mass ratio for 30 min at 37 °C. SDS/PAGE of the digested sample showed a fragment with a slightly reduced molecular weight. The identity of the resulting smaller fragment was determined by N-terminal sequencing and Western blot against the 6×His tag at the C terminus. An N-terminally truncated construct, $CP_{42–391}$, which exactly matches the trypsin-digested CP fragment, was cloned, expressed, and purified in the same manner as described for the full-length CP.

VLP Nucleic Acid Extraction. For both full-length CP and $CP_{42–391}$, 0.25 mL of a 3 mg/mL VLP sample was mixed with 0.75 mL of phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol) and incubated at room temperature for 5 min. The lysate was supplemented with 0.2 mL of chloroform and the mixture was shaken vigorously for 15 s and incubated at room temperature for another 15 min. After centrifugation at 12,000 × *g* for 15 min at 4 °C, the resulting colorless upper aqueous phase was transferred to a fresh tube. RNA was precipitated from the aqueous phase by mixing with 0.5 mL of isopropanol and incubating at room temperature for 10 min. A white pellet became visible after centrifugation at 12,000 × *g* for 8 min. The RNA pellet was then washed twice with 1 mL of 75% ethanol by vortexing and subsequent centrifugation at 12,000 × *g* for 5 min. After briefly air drying for 5 min, RNA was dissolved in 50 μL diethylpyrocarbonate water by incubating for 15 min at 60 °C.

Fluorescence Polarization Assays. 5′-Fluorescein–labeled RNA oligomers of simple AC repeats were purchased (Thermo Fisher Scientific and Sigma-Aldrich). Binding solutions contained 50 mM NaCl, 50 mM Tris (pH 7.5), and fluoresceinated RNA at 5 nM. CP dimer samples were titrated into the binding solution until the millipolarization reading stabilized. Experiments were carried out at 20 °C. The data were analyzed as described (1).

Electron Cryomicroscopy. An aliquot of 2.5 μL purified Orsay VLPs was applied to a 400-mesh Quantifoil R 1.2/1.3 copper grid, blotted for 2 s, and then rapidly plunge-frozen in liquid ethane with an FEI Vitrobot Mark III. The grid was frozen with continuous carbon and images were collected directly thereafter. Images were taken at ∼25 e/Å^2 with a JEM-3200FSC electron cryomicroscope (JEOL) operated at 300 kV and liquid-N_2 temperature. The microscope is equipped with a field emission gun and an in-column omega energy filter to remove electrons with larger than 25 eV. After an initial screening, 112 frames were recorded on a direct detection device using a Direct Electron 4K × 3K DE-12 camera. Images were taken with a 1.5-s exposure, a nominal magnification of 25,000× with a sampling rate of 2.15 Å per pixel, and the defocus was set to 1.5 μm.

The 112 images were imported into EMAN2 (2) and evaluated, and frames with a high amount of contamination/ice were removed. A total of 5,824 particles were boxed out (300 × 300 pixels) in a semiautomated routine from the raw images. Damage and drift correction was done on the individual particles that showed that a minimal drift (<2 Å) had occurred on the complete dataset. The corrected particles were imported back into EMAN2 and the contrast transfer function (CTF) was determined automatically followed by manual correction. The defocus was estimated to be between 1.3 and 4 μm. Various subsets of particles were removed based on poor CTF parameters, resulting in 4,332 particles to be used in the reconstruction. This dataset was then split in half and two small subsets (∼200 particles) were extracted from both halves of the data and binned by 4×, and two independent initial models were generated with EMAN2. The remainder of the particles was added to the two datasets and then refined independently (not binned) with icosahedral symmetry enforced. Angular step size was reduced after 15 iterations, and after 25 iterations of refinement the datasets converged to a final density map. A low-pass filter was then applied to the capsid protein portion of the density map and a soft-edge mask removed the nonicosahedral portion of the density map (RNA). The gold standard resolution assessment resulted in a final Fourier shell correlation of 6.9-Å resolution between the two independent density maps, using the 0.143 criterion (3). The final density map was generated by combining the two datasets and performing one last round of refinement. The map has been deposited in the EMDataBank (accession no. EMD-5952).

Crystallization, Data Collection, and Data Processing. The crystals of full-length CP and $CP_{42–391}$ VLPs were obtained under similar conditions using the hanging-drop vapor-diffusion method. The reservoir contained 0.7–1.05 M sodium malonate and 0.25–1% (vol/vol) Jeffamine ED-2001 in a 0.1 M Hepes buffer at pH 7.0. The droplet was made by mixing 1.5 μL of the protein sample at a concentration of 3.5 mg/mL with 0.5 μL of the reservoir solution. Diamond-shaped crystals appeared within 4 d at room temperature. VLP crystals were cryoprotected using mother liquor containing 20% (vol/vol) glycerol. The diffraction data were collected from single crystals at the Life Sciences Collaborative Access Team at the Advanced Photon Source. Images were collected using an oscillation angle of 0.2° at the wavelengths 1.13 Å and 0.98 Å for full-length CP and $CP_{42–391}$, respectively. The data were processed using HKL2000 (4) (Table S1).

Structure Determination and Refinement. Self-rotation function was calculated for full-length CP using the program GLRF (5), and the results confirmed the icosahedral symmetry of the VLP and showed that the three twofold icosahedral symmetry axes of the VLP coincided with the three orthogonal unit-cell axes. The program Phaser (6) was used to calculate a molecular replacement solution, and the 9-Å resolution cryo-EM reconstruction was

placed into the crystal unit cell in a manner consistent with self-rotation function. The noncrystallographic symmetry (NCS) mask was created from a 25-Å cryo-EM map. Using 15-fold NCS averaging, phases were gradually extended from 9-Å to 3.25-Å resolution using the programs RAVE (7) and CCP4 (8), with a step size corresponding to one reciprocal lattice interval and 10 cycles of averaging for each phase extension step. The final map was sharpened using $B = -150$ Å^2. Atomic models were built manually using Coot (9). Refinements were conducted using REFMAC5 (10) and CNS (11) with 15-fold NCS constraints. Main-chain dihedral angles were analyzed by Ramachandran plots calculated using Coot and PROCHECK (12).

For the $CP_{42–391}$ dataset, molecular replacement was performed using Phaser. The full-length CP coordinates (i.e., chains A, B, and C) were used as the initial model with the polypeptide region preceding residue 42 truncated to avoid model bias. In addition, a set of NCS operators was applied to the model coordinates to generate a quarter of the viral capsid corresponding to the exact content of a crystallographic asymmetric unit. The initial solution gave an R factor of 0.40 after rigid-body refinement. Afterward, model adjustment and refinement were performed using Coot, REFMAC5, and CNS as described above.

Cartoon and ribbon diagrams were generated with PyMOL (www.pymol.org), and cryo-EM images were prepared using Chimera (13). The program CNS was used to calculate buried surface areas. The coordinates have been deposited in the Protein Data Bank (PDB ID codes 4NWV for full-length CP and 4NWW for $CP_{42–391}$).

Analytical Ultracentrifugation and Data Analysis. Sedimentation velocity experiments were conducted using a Beckman Coulter Optima XL-A analytical ultracentrifuge. The optical density value of the sample was obtained against a water reference at 230 nm. The sedimentation velocity runs were performed at 36,000 rpm using a Beckman AN60 rotor at 10 °C for 15 h. One hundred and fifty scans were obtained for the sample at a radial step size of 30 μm. Data analysis was performed using UltraScan III software version 2.0 (14) (www.ultrascan.uthscsa.edu). Time-invariant noise and radially invariant noise were subtracted from the sedimentation velocity data by two-dimensional spectrum analysis (15). Monte Carlo analysis of genetic algorithm optimization was also performed (16).

Cross-Linking. The CP oligomer sample was dialyzed in reaction buffer containing 20 mM Hepes (pH 7.5) and 300 mM NaCl, with the protein concentration adjusted to 0.1 mg/mL. A freshly prepared solution of glutaraldehyde was added to the mixture with a final concentration of 0.05% (vol/vol) and incubated at 37 °C for 5 min or at room temperature for 20 min. Cross-linking was quenched by adding both Tris·HCl (pH 8.0) and glycine to 0.2 M. The protein product was then precipitated using trichloroacetic acid/acetone and resolved by SDS/PAGE.

1. Zheng W, Olson J, Vakharia V, Tao YJ (2013) The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. *PLoS Pathog* 9(9):e1003624.
2. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
3. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9(9):853–854.
4. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, Volume 276: *Macromolecular Crystallography, Part A*, eds Carter CW, Jr, Sweet RM (Academic, New York), pp 307–326.
5. Tong L, Rossmann MG (1997) Rotation function calculations with GLRF program. *Methods Enzymol* 276:594–611.
6. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4): 658–674.
7. Jones TA (1992) A set of averaging programs. *Molecular Replacement*, eds Dodson EJ, Gover S, Wolf W (SERC Daresbury Lab, Warrington, UK), pp 91–105.
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. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
10. Vagin AA, et al. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 No 1):2184–2195.
11. Brünger AT, et al. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5): 905–921.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283–291.
13. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
14. Demeler B (2005) UltraScan: A comprehensive data analysis software package for analytical ultracentrifugation experiments. *Modern Analytical Ultracentrifugation: Techniques and Methods*, eds Scott DJ, Harding SE, Rowe AJ (R Soc Chem, London), pp 210–229.
15. Brookes E, Cao W, Demeler B (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *Eur Biophys J* 39(3):405–414.
16. Brookes E, Demeler B (2006) Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments. *Analytical Ultracentrifugation VIII, Progr. Colloid Polym. Sci.*, eds Wandrey C, Coelfen H (Springer, New York), Vol 131, pp 78–82.

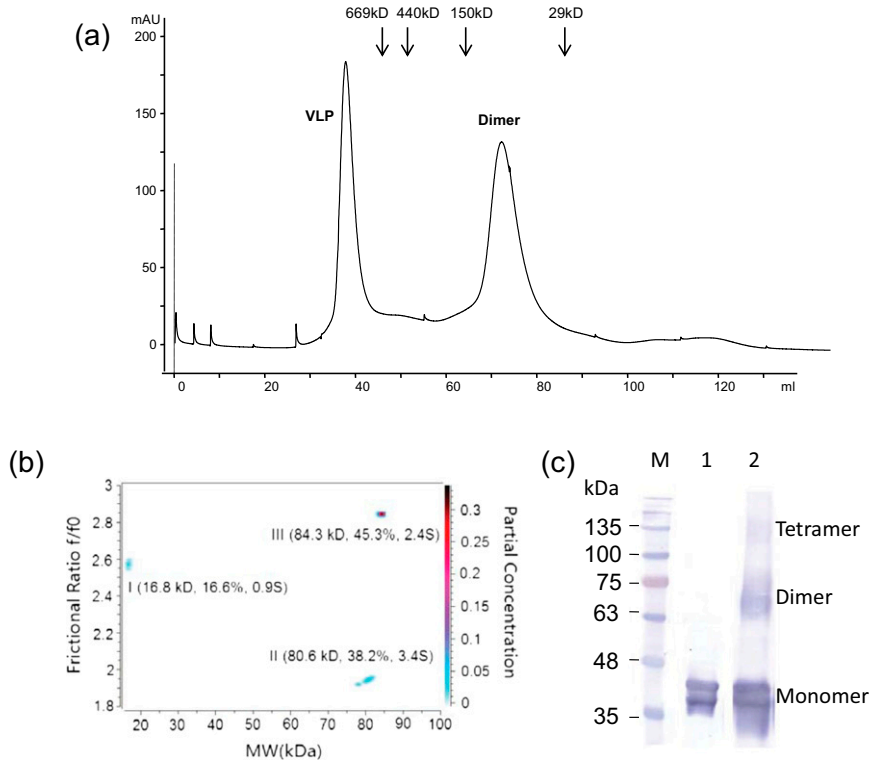

Fig. S1. Recombinant Orsay CP. (*A*) Gel-filtration chromatogram. mAU, milli absorbance units. (*B*) Analytical ultracentrifugation analysis for the Orsay dimer sample. Three molecular species were identified. The most populated one, III, had a molecular mass of 84.3 kD and a sedimentation coefficient of 2.4S, and counted for 45.3% of the total population. The second–most-populated species, II, had a molecular mass of 80.6 and likely corresponded to the partially truncated CP without the N-terminal arm. (*C*) Nonreducing SDS/PAGE of cross-linked oligomers. Lane M, marker; lane 1, native CP; lane 2, CP after glutaraldehyde treatment. The N-terminal region of the Orsay CP could be easily digested, thus producing a double band at the monomer position.

```
Orsay          1  NKNNTHKSNRKGKPVKQPSSHKRNNNPSRSVAPVAKANALRTTSSNSILLKGCDRIVTVVDASTYDAGSAI   71
LeBlanc        1  NKHRTHKSNRRGKPVKQPHS--RRSNRNGNTAPVARANPLRMTTSNSIILKGCDRIVTVVDASTLSAGDPL   69
Santeuil       1  NKNNKKNISRKLKPAKQQPPRKRNTQRRSALAPIARSNQLRQTNGNSIMLKGVDRIVTVKDATTLTPGKSV   71

cons           1  **:..:: .*: **.**  .  *..:     **:*::* ** *..***:*** ****** **:*  .*..:   71


Orsay         72 VSIPITPDIAYRLGSTARTFQRIKYRSLKFRVNAQCATTTAGGYVAGFVKDAADVLPTGTASIPYLMSNTGS  143
LeBlanc       70 VKLMITPDIAYRLGATAKTFQRIKYKTMRFRVNAQCATTTAGGYVAGFVKDAGDELPTGDRAVPYLMSNTGS  141
Santeuil      72 VSLLITPLIADRLRAASQMFQKIKYHKLVFRVNSQCSTMTAGGYVAGFVKDPADEIPSESVAVQYLMSNAGS  143

cons          72 *.: *** ** ** :::: **:***:.: ****:**:* ***********..* :*:   :: *****:**  143


Orsay        144 FTQPWWKSTVHNVKIPQKLFYTEAPTRGADAVREYCPGQFHVLVDSKPSQICPVTVDLEWVVELHDATFRKE  215
LeBlanc      142 FTQPWWKSTVHNVKLPQKMFYTEKPAVGSDAVREYSPGTFYVVVDSKPSQVCPVTVDLEWAVELHDATYREV  213
Santeuil     144 YTQPWWRSTVHNIRASQRMFFTDKPALGADAIREYSPGRFYVIVDSTPSQTCPVTVDLEWTVELREATFKLP  215

cons         144 :*****:*****:: .*::*:*: *: *:**:***.** *:*:***.*** *********.***::**::    215


Orsay        216 SDQTAISAIVADHTLNVYGLPATSNRVGHILISPIGQTPKDLTPTRFATFFGFLPDDKFCVRIPTPVDVVLT  287
LeBlanc      214 EAVTTISGISADYSLNVFGIPATSVLAGRILITPKGSNPRPLTIERFRTFFGFAPDLMFCVRLPRPEDVVLT  285
Santeuil     216 DTTTVISSIRPELSLSTYGMPNTSPMAGHVVVKPSGETPRWLTPARFQTMFGFAPDEDMCATIPTPVQCVLT  287

cons         216 .  *.**.* .: :*..:*:* **  .*::::.* *..*: **  ** *:*** **  :*. :* * : ***  287


Orsay        288 GDNVYQSVEATHIRAYLVNGGLGIDFHLAAYNDTTHTIQPIIPTLWNVYDVTGAVTAPFTSAIYDNHVWTHK  359
LeBlanc      286 GDNVYQGVKATHFMAYTADGL--LDFHLAAFNDVTQSIIPVRPSDWRVYSVDGSSSTPFTSAIYENTTWSSD  355
Santeuil     288 GENAYQTVTATHFSAVLTDSAV-QSFMLSAFNDVTQDIQPINH-TWQIMDALGQIVTPFIAVIYDSAIWSAT  357

cons         288 *:*.** * ***: *  .:.    .* *:*:**.*: * *:    *.: .. *   :** :.**:.  *:    359


Orsay        360 DKFVPVSFQDEPIPGTVFDYLYPR-SYSLPSSN  391
LeBlanc      356 DKFVPVTISDVPIPGAVFRFLYPSVGYSLPSSN  388
Santeuil     358 DSYTITTRTGKPVPLDVFRFLHPN-ATVLPPEN  389

cons         360 *.:. .:  . *:*  ** :*:*  .  **..*  392
```

Fig. S2. Multisequence alignment for the three nematode viral CPs. Residues conserved in Le Blanc and Santeuil but different in Orsay are highlighted in red.

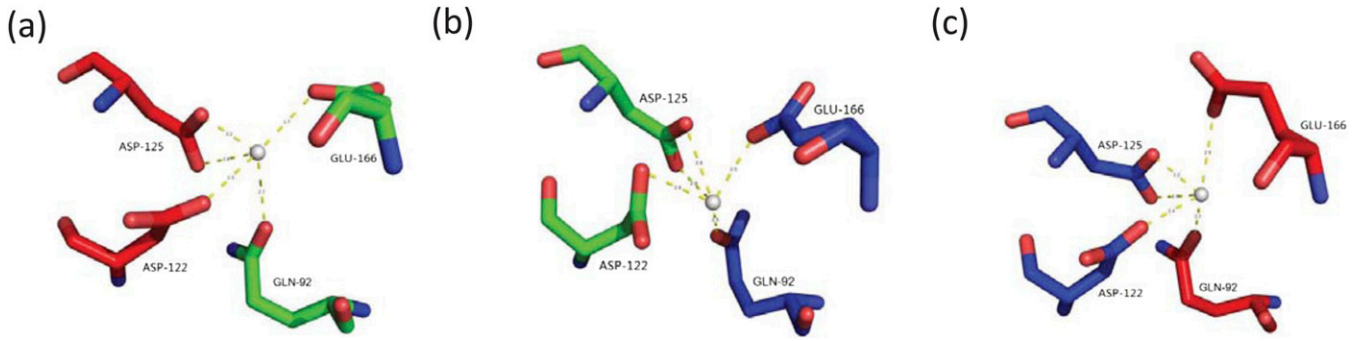

Fig. S3. Calcium in Orsay CP. (*A*) Calcium between subunits A and B. (*B*) Calcium between subunits B and C. (*C*) Calcium between subunits A and C. Calcium ions are indicated by gray spheres. Coordinating ligands from subunits A, B, and C are colored in red, green, and blue, respectively.

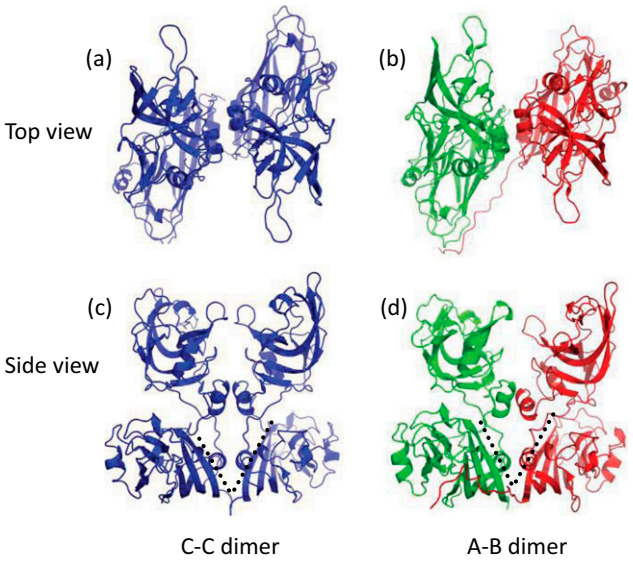

Fig. S4. Comparison of the Orsay C–C (*Left*) and A–B dimers (*Right*). Both top and side views are shown. The dotted lines in the two side views highlight a CP bending angle that is narrower in the A–B dimer than in the C–C dimer.

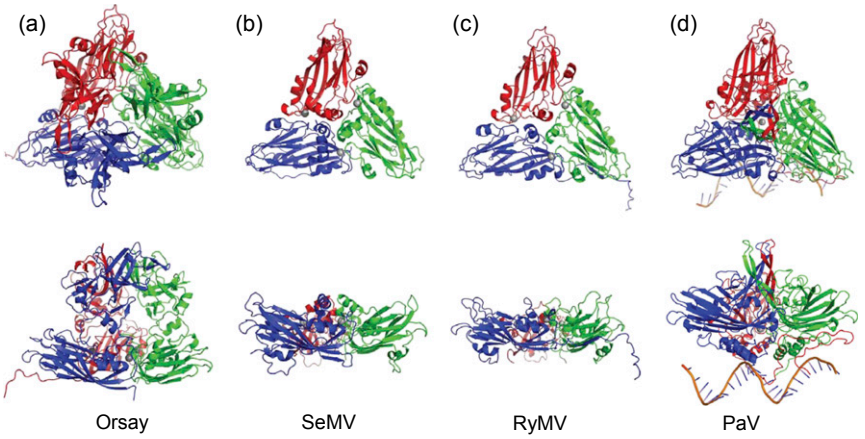

Fig. S5. Extended N-terminal arms in $T = 3$ RNA viruses. (*A*) Orsay asymmetric unit (ASU). (*B*) Sesbania mosaic virus (SeMV) ASU. (*C*) Rice yellow mottle virus (RyMV) ASU. (*D*) Pariacoto virus (PaV) ASU. (*Upper*) Viewed along the quasi-threefold axis from the outside of the VLP. (*Lower*) Viewed from the side. A, B, and C subunits are colored in red, green, and blue, respectively. Calcium ions are indicated by gray spheres.

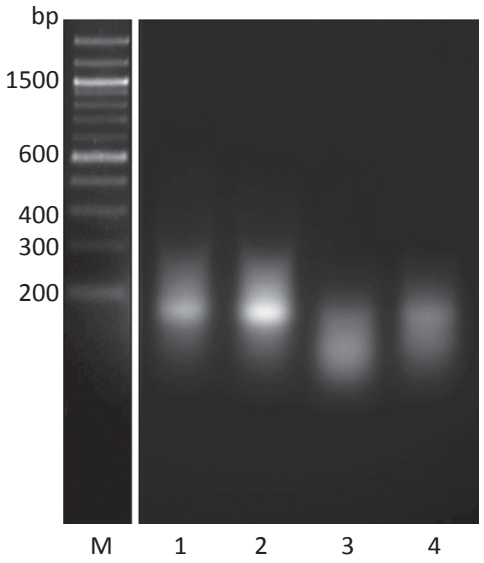

Fig. S6. Nucleic acids in Orsay VLP. An RNA sample extracted from Orsay VLP was treated with different nucleases. Lane 1, freshly extracted RNA; lane 2, RNA treated with DNase I; lane 3, RNA treated with RNase V1; lane 4, RNA treated with RNase T1.

Table S1. X-ray data statistics

Parameters	$CP_{1–391}$	$CP_{42–391}$
Data collection		
Space group	I222	I222
Unit-cell dimensions, Å	a = 402.2, b = 369.9, c = 410.5	a = 404.9, b = 375.1, c = 412.2
Resolution, Å	50–3.25 (3.31–3.25)	30–3.75 (3.81–3.75)
Total no. of frames	380	150
Total no. of reflections	1,152,775	322,993
Unique reflections	410,231	204,023
I/σ	8.6 (1.8)	3.7 (1.1)
Redundancy	3.0 (2.7)	1.9 (1.8)
Completeness, %	87.8 (58.3)	64.3 (68.0)
R_{merge}, %	19.0 (72.0)	24.7 (66.8)
Phase extension		
Averaging *R* factor	0.34	—
Correlation coefficient	0.84	—
Refinement		
R_{work}, %	27.8	30.6

The numbers in parentheses are for the highest-resolution shell.

Table S2. Pairwise amino acid sequence identity for the three nematode-infecting viruses

	CP, %					
Virus pair	Overall	N-terminal peptide	S domain	P domain	δ, %	RdRP, %
Orsay–Santeuil	52	47	62	44	37	45
Orsay–Le Blanc	68	64	79	61	39	36
Santeuil–Le Blanc	53	42	67	44	39	37

Three viral proteins are used for pairwise comparison: capsid protein, δ protein, and RNA-dependent RNA polymerase (RdRP). For CP, sequence identities are calculated for individual domains including the N-terminal peptide, the shell (S) domain, and the protrusion (P) domain.

Table S3. Buried surface areas in Orsay capsid protein

		Subunits related by symmetry axis (1,000 Å^2)							
		q3			q2	i2	i5	q6	
Capsid	Domains	A1B1	B1C1	A1C1	A1B5	C1C2	A1A5	C1B5	C1B2
$CP_{1–391}$									
	N peptide	0	0	0.5	0.7	0	~0	0	0
	S domain	1.8	1.8	2.1	1.6	1.4	1.7	1.2	1.5
	P domain	1.3	1.4	1.4	0.9	0.8	0.1	0.1	0.1
	N+S+P	2.9	3.0	3.8	3.1	1.8	1.6	1.1	1.4
$CP_{42–391}$									
	S+P	2.8	2.9	3.0	2.2	1.6	1.6	1.1	1.3

See figure below for the numbering of different Orsay CPs in a $T = 3$ icosahedron. Definition of symmetry axes: q3, quasi-threefold; i5, icosahedral fivefold; q2, quasi-twofold; q6, quasi-sixfold; i2, icosahedral twofold.