PNAS
www.pnas.org

**Supplementary Information for:**

Structure of a rabies virus polymerase complex from electron cryo-microscopy

Joshua A. Horwitz, Simon Jenni, Stephen C. Harrison, Sean P.J. Whelan

**Corresponding authors**:

Stephen C. Harrison harrison@crystal.harvard.edu

Sean P.J. Whelan sean_whelan@hms.harvard.edu

**This PDF file includes:**

Figures S1 to S14

Table S1

Legend for Movie S1

SI References

**Other supplementary materials for this manuscript include the following:**

Movie S1

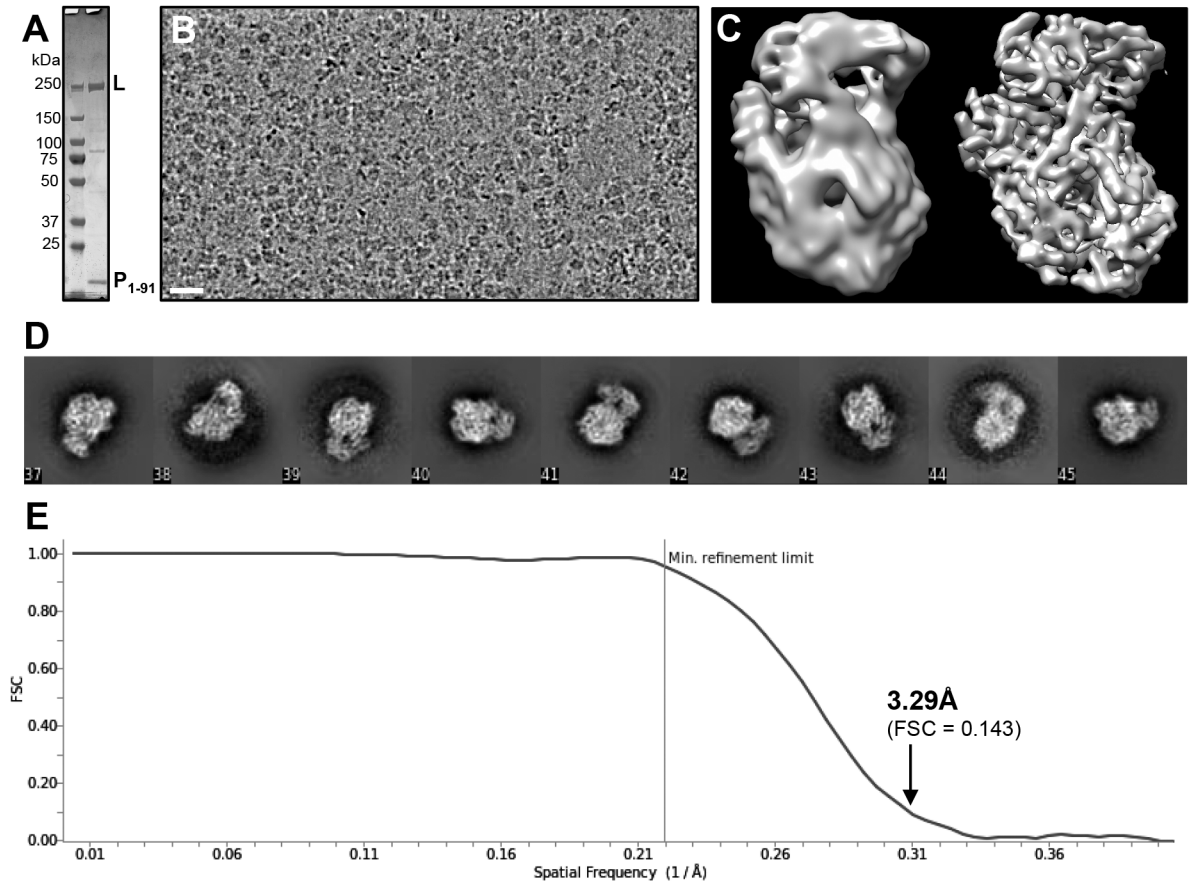


**Figure S1: Cryo-EM reconstruction.** (A) Coomassie-stained SDS-PAGE gel showing the purified RABV L-$P_{1-91}$ preparation used for cryo-EM. (B) Representative cryo-EM micrograph of RABV L-$P_{1-91}$ complexes from (A) immobilized in vitreous ice and imaged at 300 kV. Scale bar: 25nm. (C) Ab-initio 3D reference (left) and 6.7 Å reconstruction (right) from a preliminary 200 kV dataset. (D) Representative 2D class averages from the final 44.5K particle dataset (Fig. S2) used to obtain the 3.3 Å reconstruction. (E) Fourier shell correlation curve for the final 3.3 Å map (cisTEM: (1)).

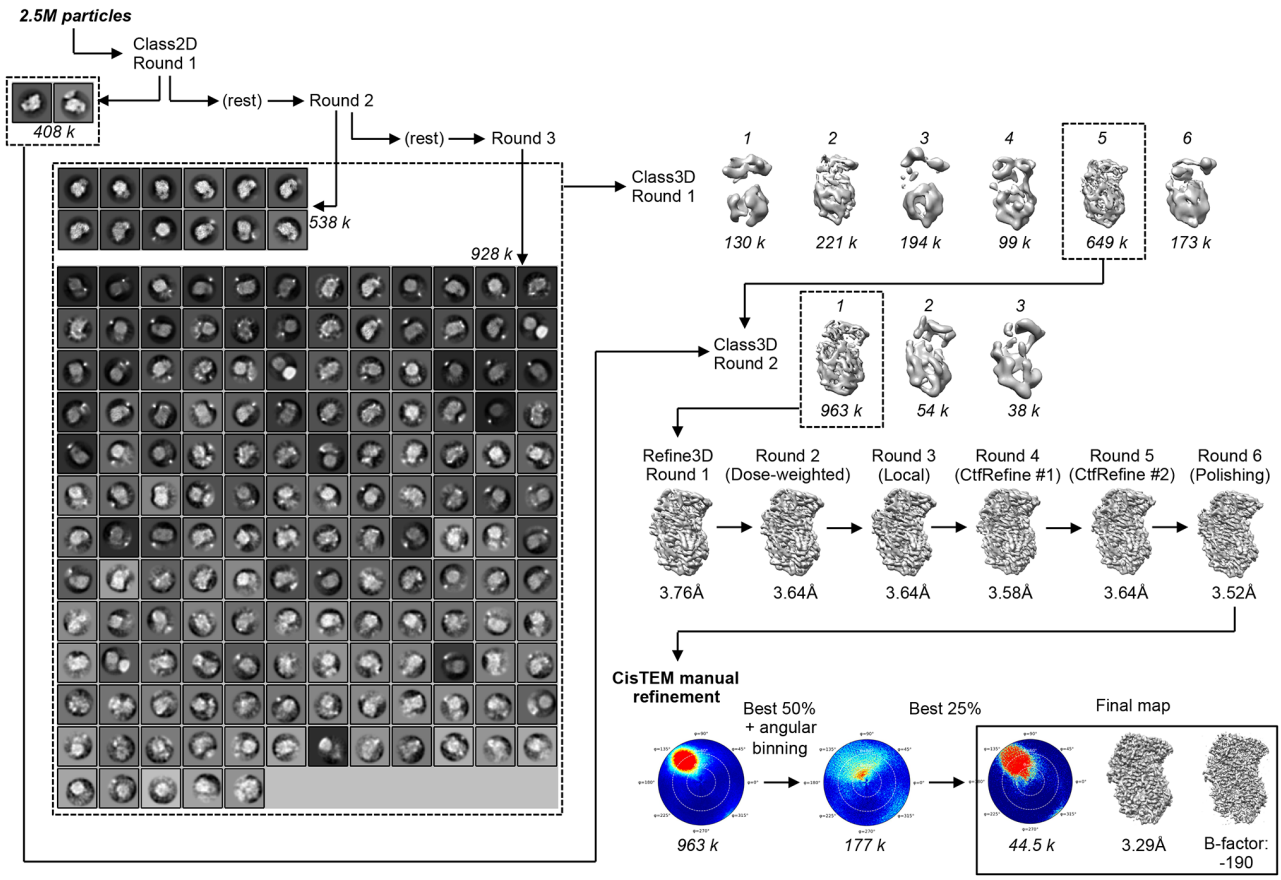


**Figure S2: Particle classification.** From 5.3 M initial particles after auto-picking and manual cleaning, particle images from each of the three 300 kV datasets were binned 3x and subjected to an initial round of 2D classification in RELION-2 (2) to remove bad particles (not shown). The remaining 2.5 M particles were subjected to a round of coarse (25 Å resolution cutoff) 2D classification (Class2D Round 1). The two classes with the highest occupancy, containing a large portion of the particles adopting the preferred orientation we observed, were separated from the remaining particles and retained for later use. In a second round of 2D classification with the remaining particles, 12 good classes totaling 538 k particles were selected and again separated from the rest, to be used later. The remaining particles were subjected to a third round of 2D classification, from which all but the very worst classes were kept to maximize retention of particles with rare views, totaling 928 k particles. These particles were pooled with the 538 k particles selected in round 2, and the combined 1.47 M particles subjected to 3D classification (Class3D Round 1). The best class among six was chosen for further classification, containing 649 k particles. These particles were merged with

the additional 408 k preferred-orientation particles separated in the first round of 2D classification, and the combined 1.06 M particles subjected to a round of 3D classification (Class3D Round 2). A single class predominated, with 963 k particles. We extracted unbinned particles images of this class for several rounds of 3D refinement using automated Refine3D. After an initial round using non-dose-weighted particles, subsequent rounds used dose-weighted particles, iterative per-particle CTF-refinement, and Bayesian polishing in RELION 3 (3). Although modest gains in reported resolution were observed, the map remained visibly anisotropic even after particle polishing. We then imported the entire polished stack of 963 k particles into cisTEM (1) and carried out several rounds of manual refinement, in which particles were removed on the basis of the ‘SCORE’ parameter and by angular flattening. Lower right: Heat-maps (cisTEM), showing the distribution of Euler angles at the outset and after initial and final particle pruning, and two representations of the final map, with an average resolution of 3.29 Å (FSC = 0.143 criterion), reconstructed from 44,500 particles.

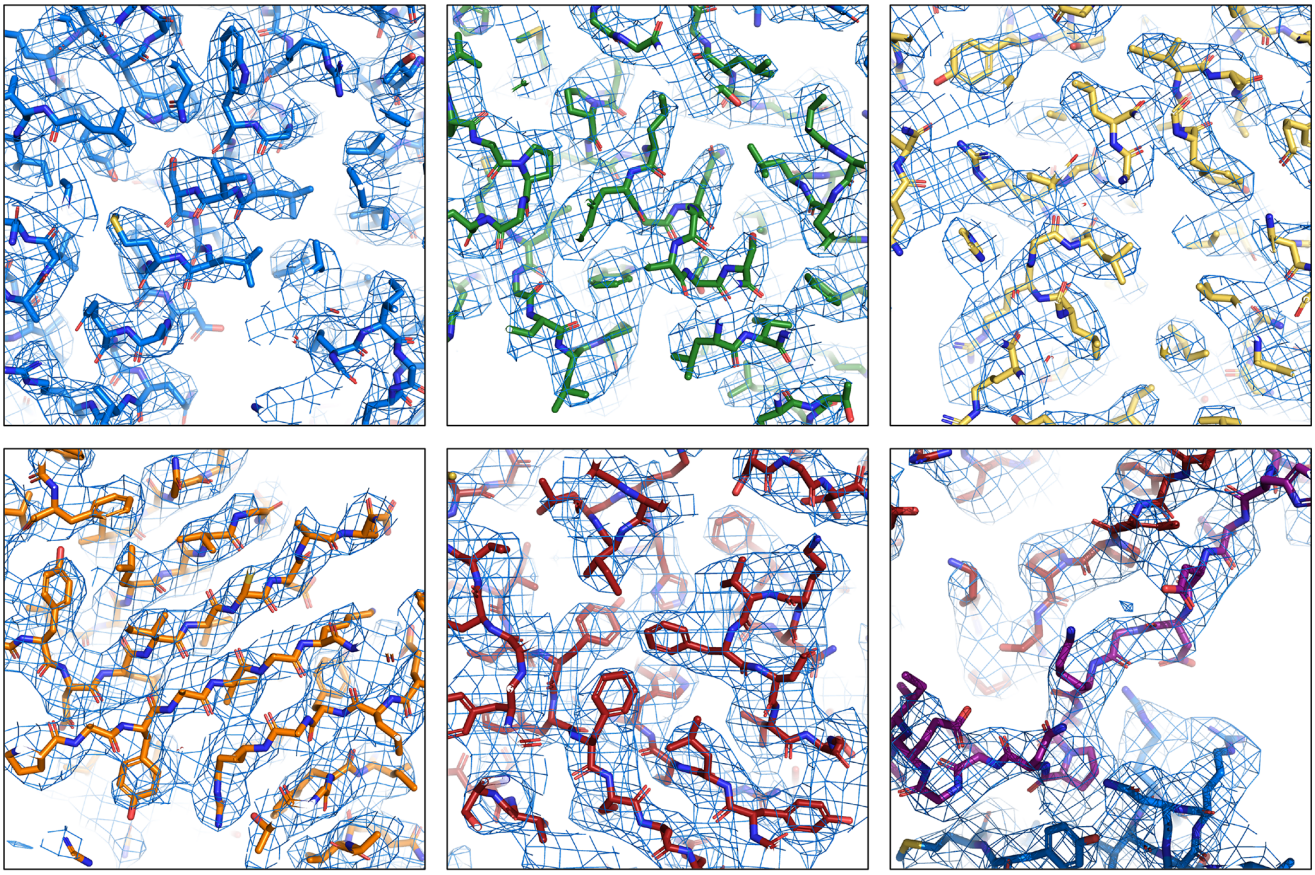

**Figure S3: Details of the 3.3 Å resolution map.** Representative densities showing visible side chains for each of the five domains of L (top row and first two panels of second row) and a segment of P (last panel, lower right). Atomic model is in stick representation. Carbon atoms colored by domain, as in Fig. 2C; oxygen, red; nitrogen, blue; sulfur, yellow.

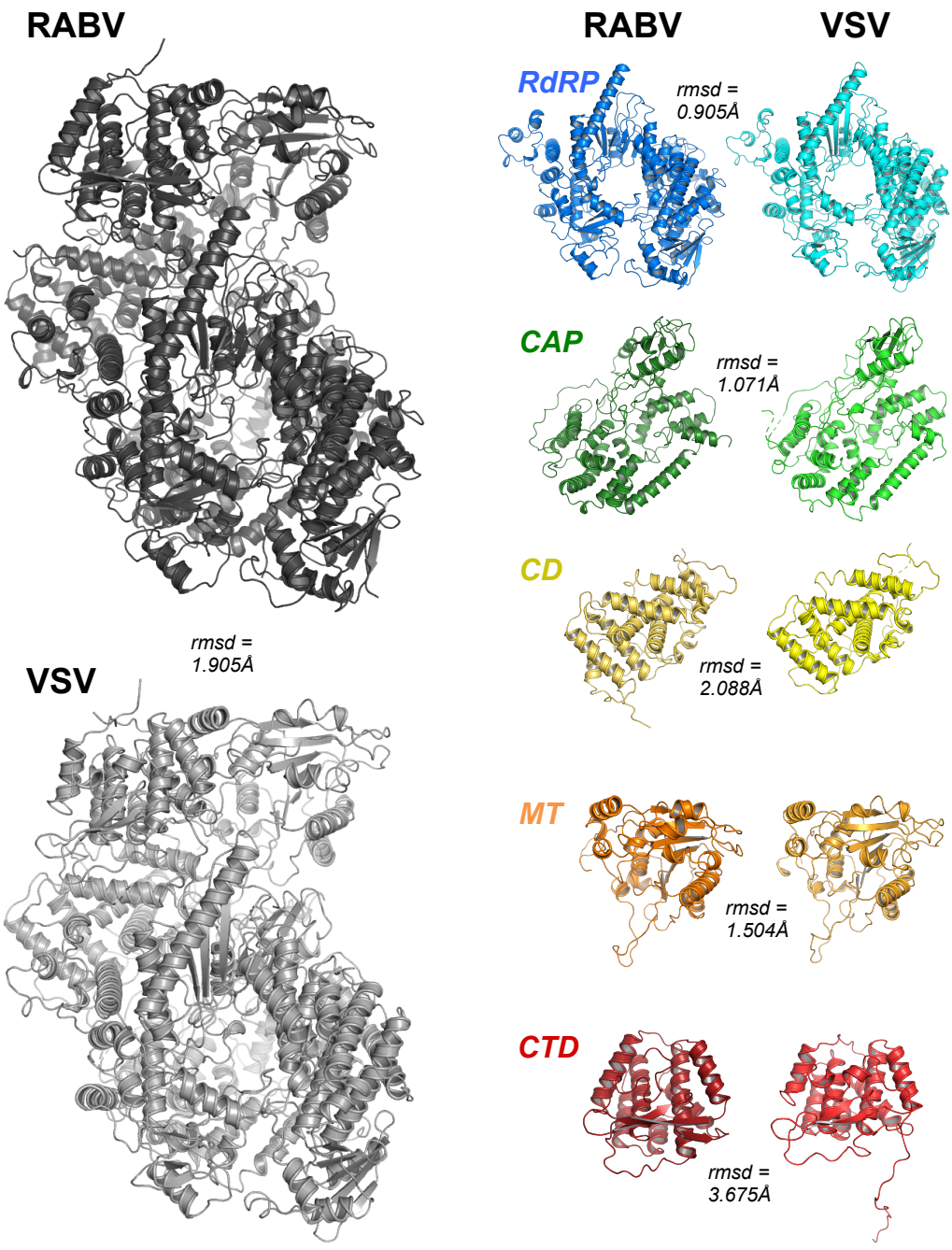


**Figure S4: Alignments of RABV and VSV L and their domains**. *Left:* RABV (dark gray) and VSV (light gray) L-P structures (PDB ID 6UEB and 6U1X, respectively) in cartoon representation with the RMSD after iterative whole-structure alignment of Cα atoms with outlier elimination in PyMOL 2.0 (Schrödinger, LLC) . *Right:* independently aligned domains of RABV and VSV L in cartoon representation (colored, as in Fig. 2C) with their respective RMSDs.

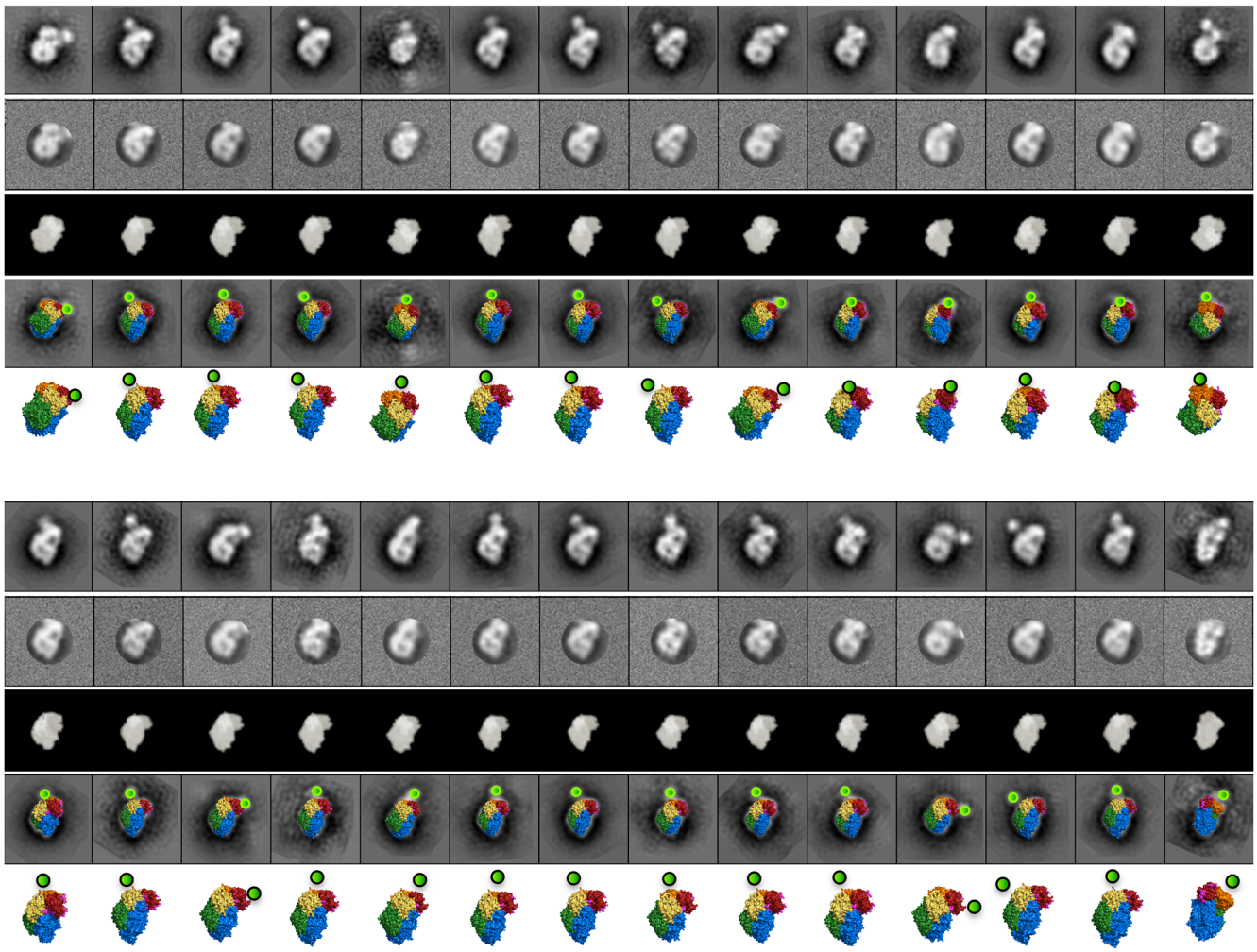

**Figure S5: L-GFP-$P_{1-91}$ projection matching.** Twenty-eight negative-stain EM 2D class averages of RABV L in complex with GFP-$P_{1-91}$ for which a clear feature corresponded to the globular GFP moiety, after alignment to a 2D reference to standardize their orientation in the X/Y plane (top rows). The 2D class images were noise-masked to remove as much of the GFP signal as possible without losing information from L (second rows). Corresponding matched projections of RABV L were obtained as described in Methods (third, fourth and fifth rows). We then made a best approximation for the location of the GFP globule relative to each projection of RABV L (fourth rows) and scaled the resulting models for ease of visualization (fifth rows). The preferred orientation of the particles is evident from the low diversity of projections represented; the preferred view helps visualize the wide arc defining the N terminus of P in RABV L, as the GFP globule occupies many different positions about several near-identical projections of L.

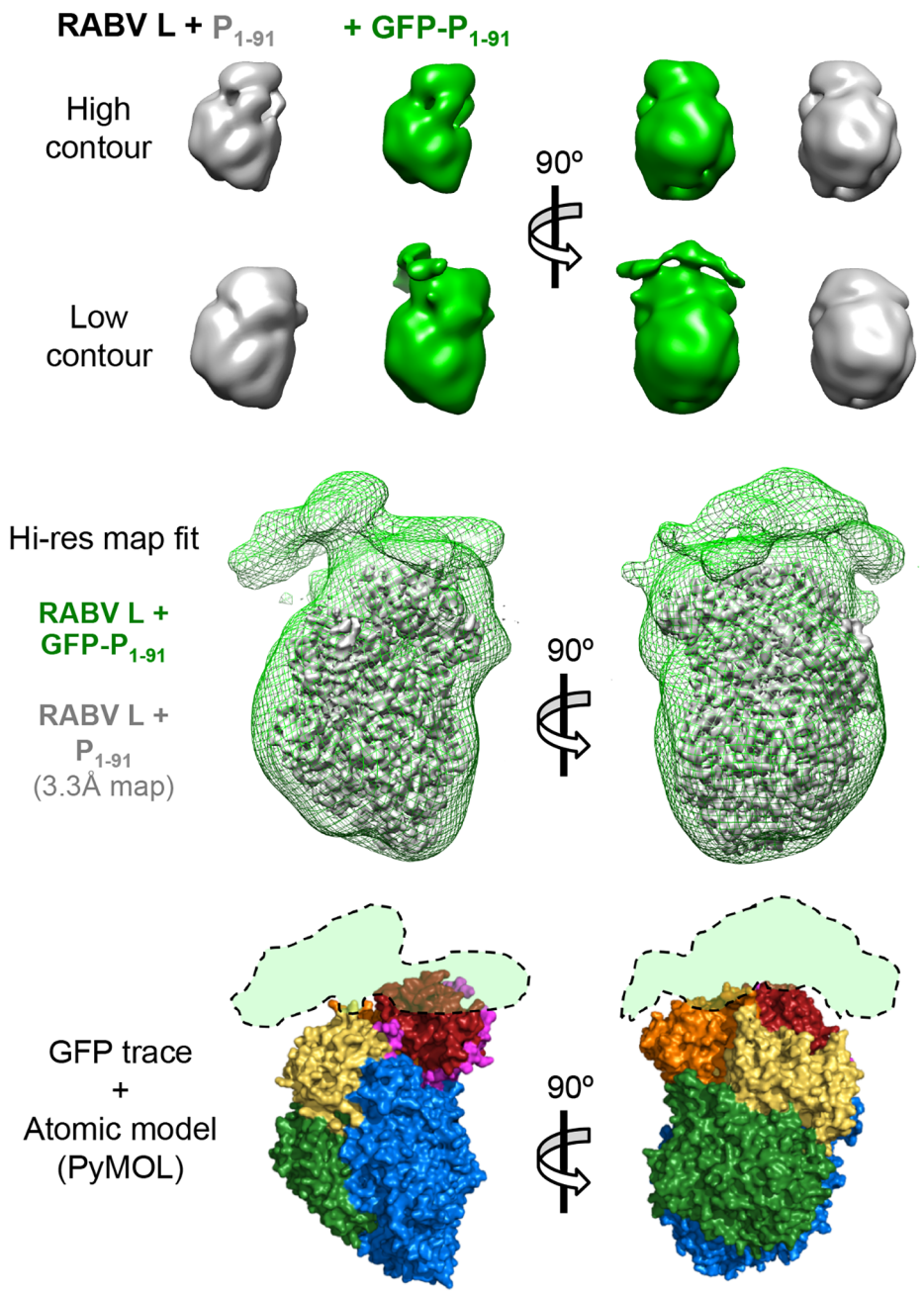


**Figure S6: L-GFP-$P_{1-91}$: comparison with high-resolution structure.** *Top,* Low-resolution negative-stain EM density maps are shown for RABV L in complex with either $P_{1-91}$ (gray) or GFP-$P_{1-91}$ (green) at high and low contour levels. *Middle*, the high-resolution RABV L-$P_{1-91}$ cryo-EM density map fit (with UCSF Chimera ) into the map of negatively stained L-GFP-$P_{1-91}$ complex. The L-GFP-$P_{1-91}$ map is shown at low contour to illustrate the diffuse density for the GFP globule about the top of RABV L. *Bottom*, a crude trace of the GFP density, superposed on the atomic model of RABV L corresponding to the high-resolution map fit above. This image is reproduced in Fig. 4A.

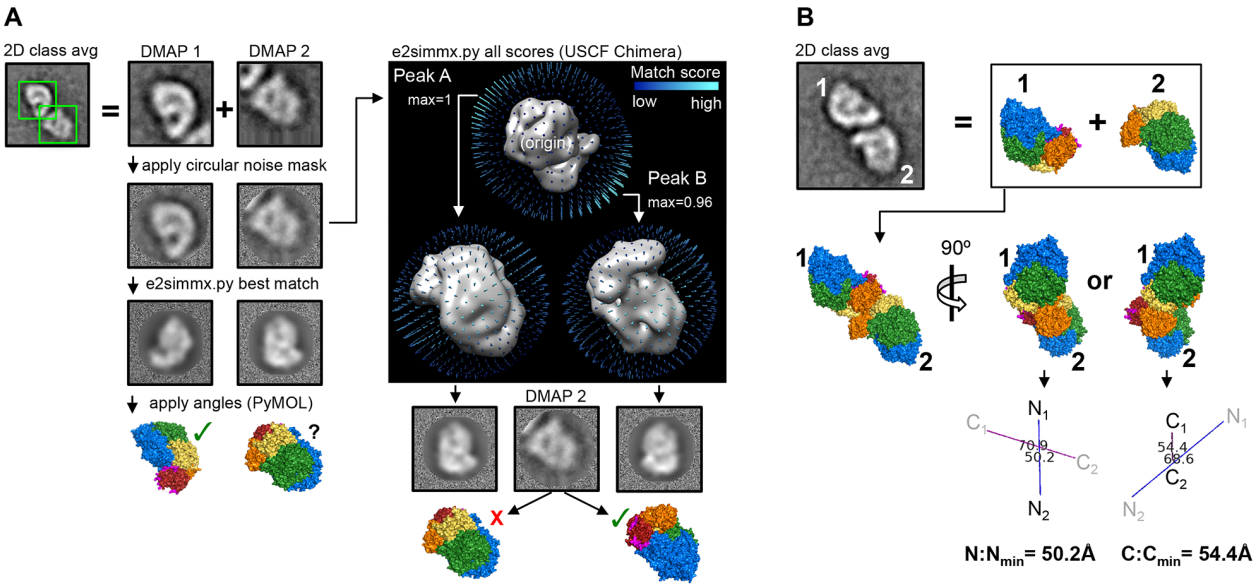


**Figure S7: L-$P_{FL}$ projection-matching.** (A) Projection matching of L-$P_{FL}$ dimers, as described in Methods, illustrating a case in which one of the partners resulted in an incorrect top projection match. After application of the top projection-matched angles for each dimer partner in PyMOL (Schrödinger, LLC), the resulting RABV L model images were inspected for correctness. We generated 3-dimensional heat maps for the e2simmx.py projection scores in the coordinate system of the EM density map to visually evaluate the peaks. When obvious, a secondary or alternate peak was selected instead of the top match. (B) Projection-matched models for each monomer (DMAP) in a dimer pair 2D class average were arranged in the X/Y plane in PyMOL to a best-approximation with the original 2D class average. To estimate distances between dimer partners, we translated one of the models along the Z axis such that the distance was minimized between Cα atoms of either the C-terminal or the N-terminal residues of P. This optimization yielded two measurements for each dimer, which we refer to as the $C:C_{min}$ or $N:N_{min}$ distance.

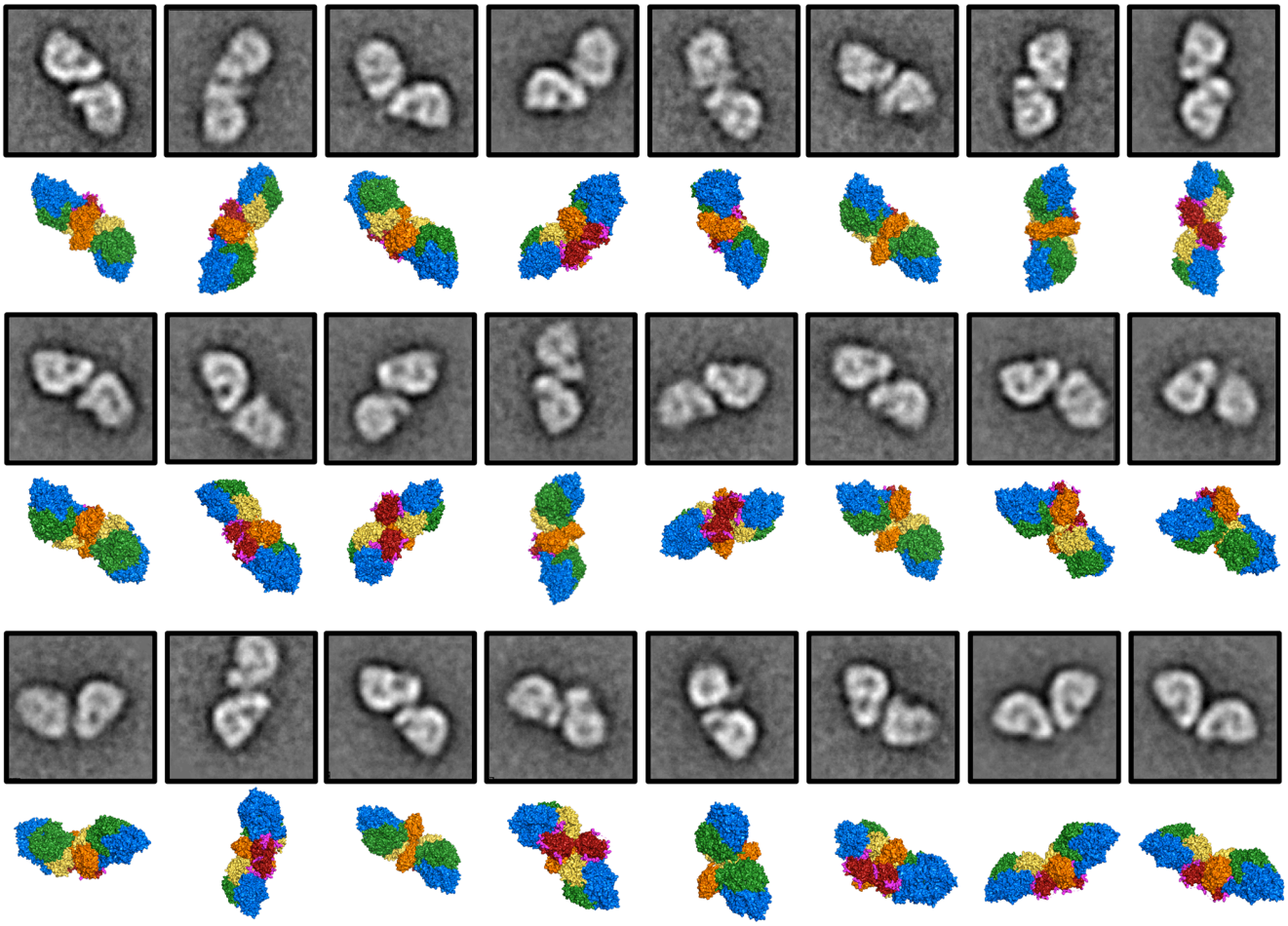

**Figure S8: L-$P_{FL}$ projection-matched dimers.** Negative-stain EM 2D class averages (top rows) and projection-matched RABV L models (bottom rows) are shown for the 24 dimer pairs (of an initial 29 pairs, not shown) for which we could match both DMAPs.

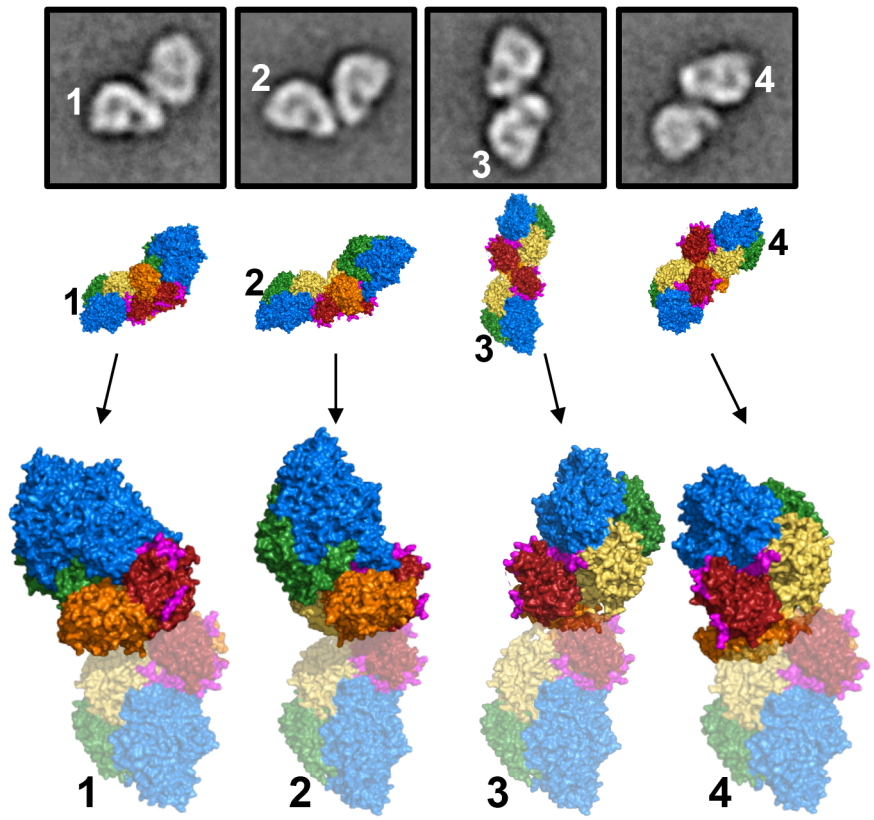


**Figure S9: Arrangement of L-$P_{FL}$ dimers**. *Top,* Four projection-matched dimers are shown (as in Figs. 4B and S7) for which one of the dimer partners in each shares a common projection with the others. The partners in each dimer sharing the common projection are listed 1-4. *Bottom*, the four projected dimers were rotated about the Z axis such that the common projections were aligned with each other. The common projections are shown partially transparent to show that the partners in each dimer adopt different positions with respect to the common projection.

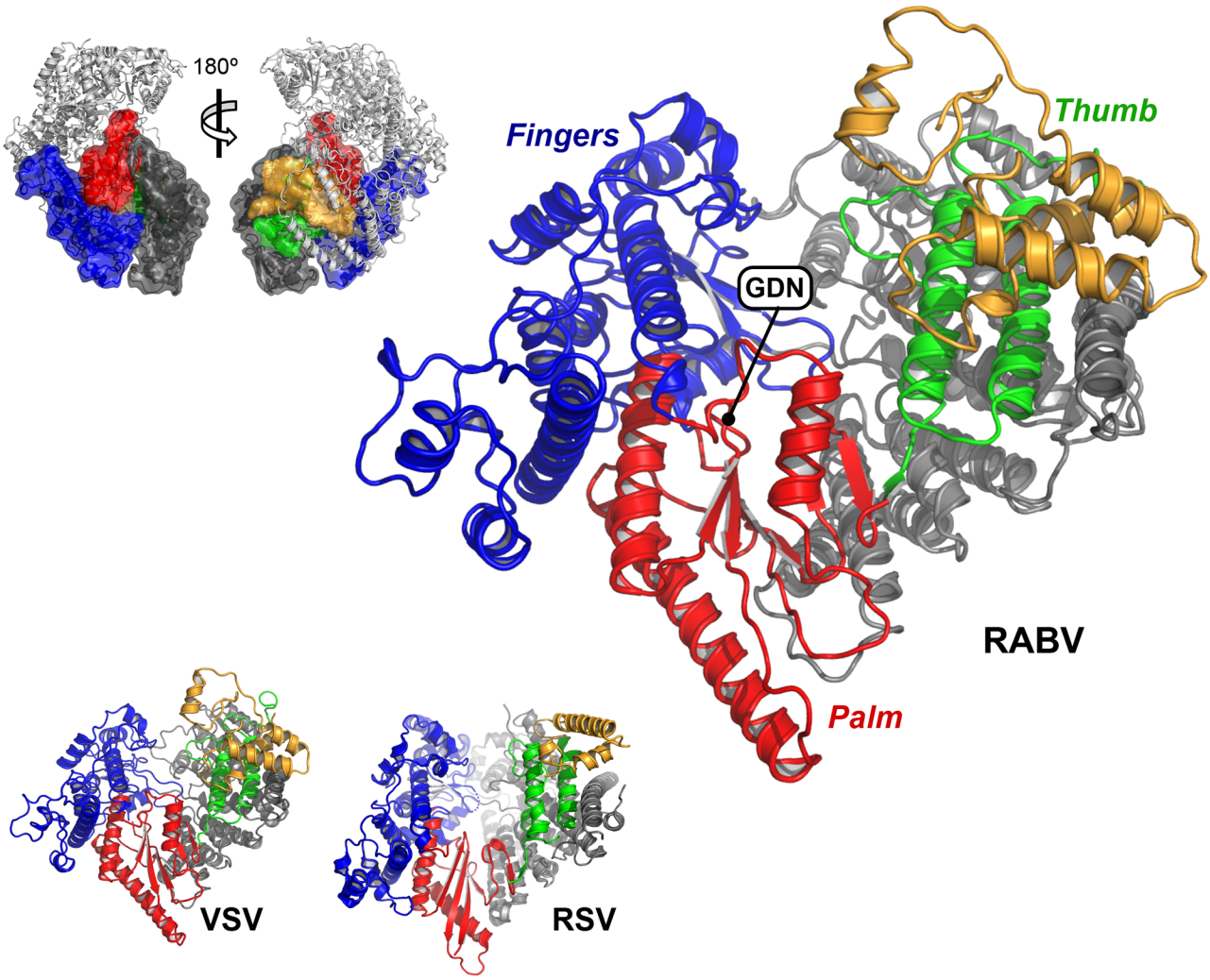


**Figure S10: The RdRp.** *Right and bottom:* the RABV (PDB ID 6UEB), VSV (PDB ID 6U1X) and RSV (PDB ID 6PZK) RdRps adopt the familiar “right hand” configuration of nucleotide polymerases, and are shown in cartoon representation: fingers, blue; palm, red; thumb, green; N-terminal extension, gray; and C-terminal extension, orange. *Top-left:* subdomains of the RABV RdRp (surfaces, colored as at right) are highlighted against the complete RABV L-P structure (gray cartoon) to illustrate their positions in L relative to the orientation shown at right.

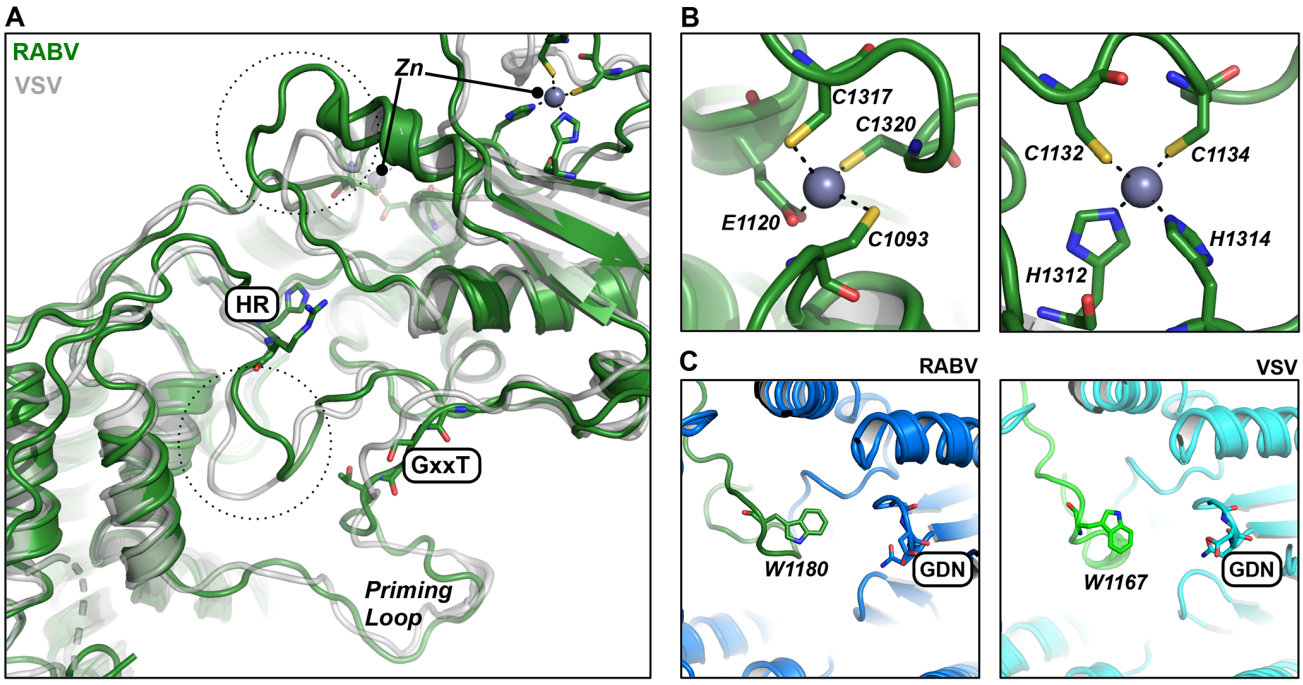


**Figure S11: Capping enzyme (CAP).** (A) RABV (green) and VSV (gray) CAP domains superimposed. Catalytic residues in the active site (HR and GxxT motifs) and residues participating in Zinc ion coordination are shown for RABV in stick representation. Black dashed circles highlight loops near the active site with different conformations. (B) Zn coordinating residues in RABV CAP (green sticks). Dashed black lines indicate the side-chain atoms involved in metal coordination. (C) RdRp (blue) and CAP (green) domains of RABV (left) and VSV L (right) show similar insertion of the poorly ordered priming loop from CAP into the RdRP active site. The conserved tryptophan from CAP important for transcription initiation from the 3‘ end of the viral genome is in the same location in both structures.

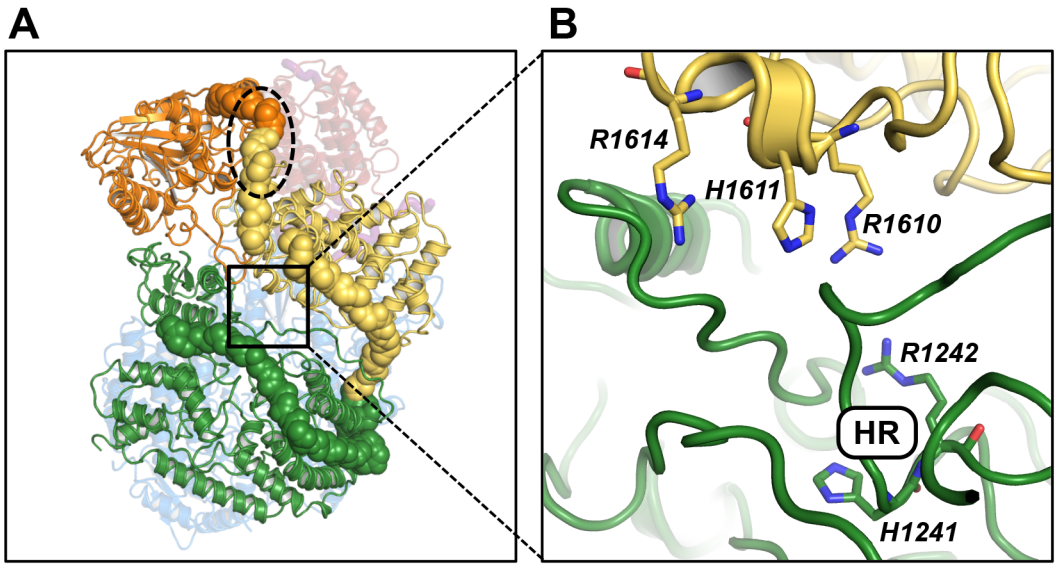


**Figure S12: Connector domain (CD).** (A) Cartoon representation of RABV CAP (green), CD (yellow), and MT (orange) with long linkers between CAP:CD and CD:MT shown as spheres for main-chain atoms. Black dashed oval, region of the CD:MT linker disordered in the density map. Boxed window and (B), CAP active site at the interface with the CD showing a set of conserved basic residues opposite the HR motif from CAP.

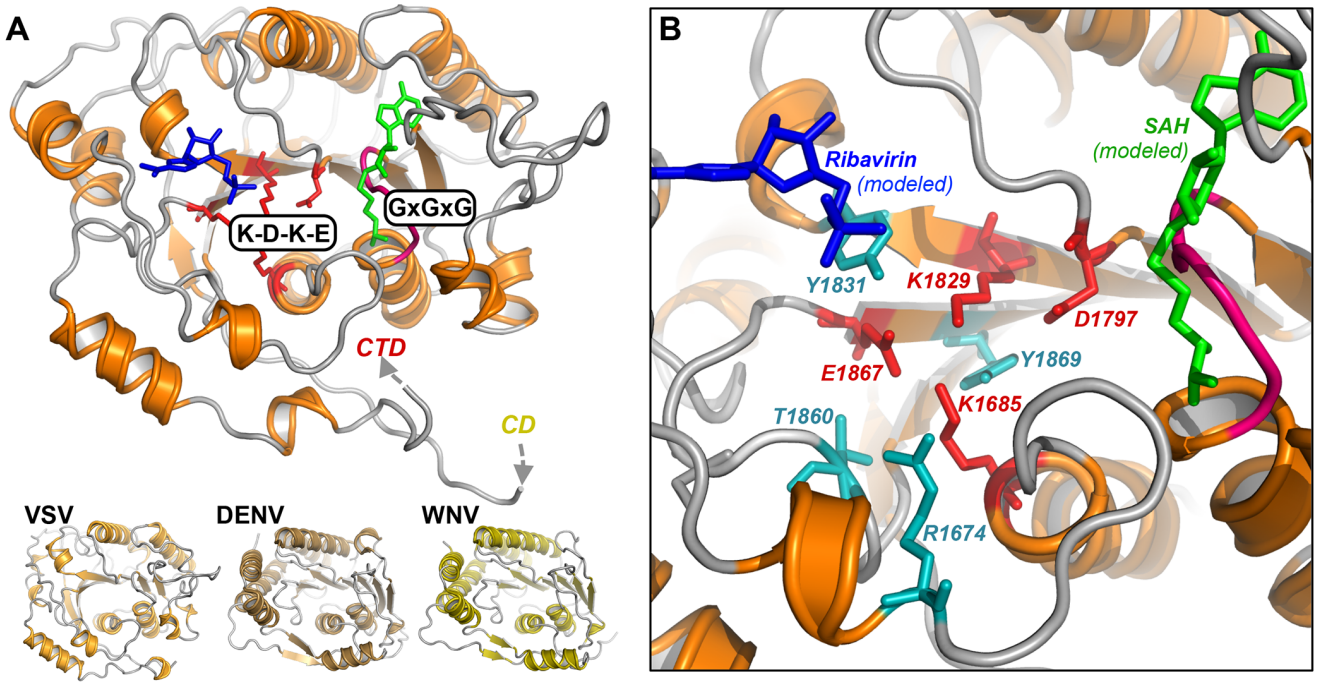


**Figure S13: Methyltransferase (MT).** (A) RABV (top) and three other dual-function viral MT domains (bottom). Catalytic tetrad residues are in red stick representation; GxGxG motif in pink. The positions of ribavirin (blue) and SAH (green) present in the DENV crystal structure are modeled against the RABV MTase to illustrate likely binding areas for the GTP cap and SAM, respectively. (B) Close-up of the RABV MTase active site, as in (A), showing conserved residues (teal, sticks) that may support catalysis.

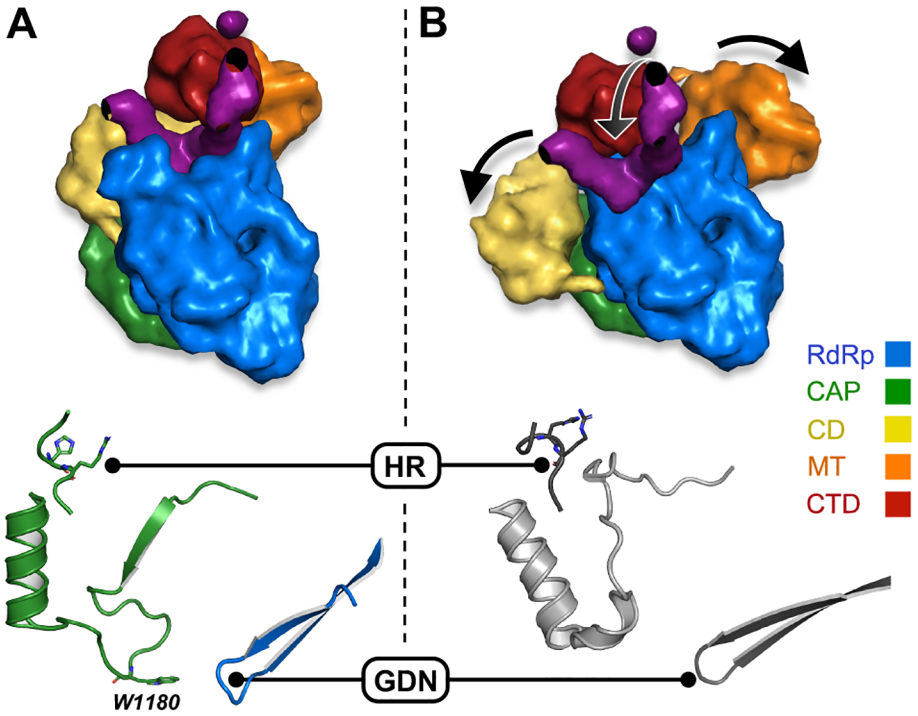


**Figure S14: Transition from initiation to elongation.** (A) RABV L-P initiation state captured in the cryo-EM structure (top, surface representation) with the priming loop positioned opposite the RdRp GDN motif and distant from the Cap HR motif (bottom, cartoon with sticks, colored as in Fig. 2C). (B) Proposed re-arrangement of RABV L-P domains during RNA elongation (top), based on observed flexibility in Fig. S4; bottom, model of the priming loop position in the RNA elongation state, from the recent cryo-EM structure of RSV L-P (gray cartoon with sticks: PDB ID 6PZK).

## Table 1. Cryo-EM data collection and validation statistics

| | |
|---|---|
| ***Data collection and processing*** | |
| Microscope | Tecnai F30 Polara |
| Detector | Gatan K2 Summit |
| Magnification | 31,000X |
| Pixel size (Å) | 1.234 |
| Voltage (kV) | 300 |
| Exposure ($e^-/Å^2$) | 72 |
| Defocus range (μm) | 0.6-3.5 |
| Final particles | 44,500 |
| Symmetry | C1 |
| Resolution (Å) | 3.29 |
| ***Model composition, refinement and validation*** | |
| Non-hydrogen atoms | 17219 |
| RMSD bonds (Å) | 0.005 |
| RMSD angles (º) | 1.345 |
| Ramachandran: | |
| Favored (%) | 99.44 |
| Allowed (%) | 0.47 |
| Outliers (%) | 0.09 |
| Rotamer outliers (%) | 0.00 |
| Clash score | 6.65 |
| MolProbity score | 1.37 |
| ***Model-to-map fit*** | |
| Cross-correlation coefficient (mask) | 0.75 |
| Cross-correlation coefficient (volume) | 0.76 |
| Main-chain | 0.76 |
| Side-chain | 0.71 |

**Movie S1: RABV L domains explore multiple conformations.** Multi-body refinement in RELION 3.0, with four masked regions (RdRP+CAP; CD; MT; CTD) and with an expanded dataset of 2.4M particles, describes a broad range of conformations for the C-terminal domains of RABV L. **Movies S1a,b, and c** illustrate conformations explored along the three eigenvectors accounting for the greatest variance in the data (of 24 calculated). Each movie is composed of ten "frames", with each frame a reconstruction from $1/10^{th}$ of the particles populating the given eigenvector. The frames in each movie were aligned to the RdRP domain of RABV L to highlight the different positions of the C-terminal domains (CD, MT, CTD) with respect to each other and the RdRP/Cap domains. The movies are not aligned to one another, but are instead oriented to best illustrate the range of conformations for each eigenvector.

**References:**


1. Grant T, Rohou A, & Grigorieff N (2018) cisTEM, user-friendly software for single-particle image processing. *Elife* 7:pii: e35383.
2. Kimanius D, Forsberg BO, Scheres SH, & Lindahl E (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* 5:pii: e18722.
3. Zivanov J*, et al.* (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7:pii: e42166.