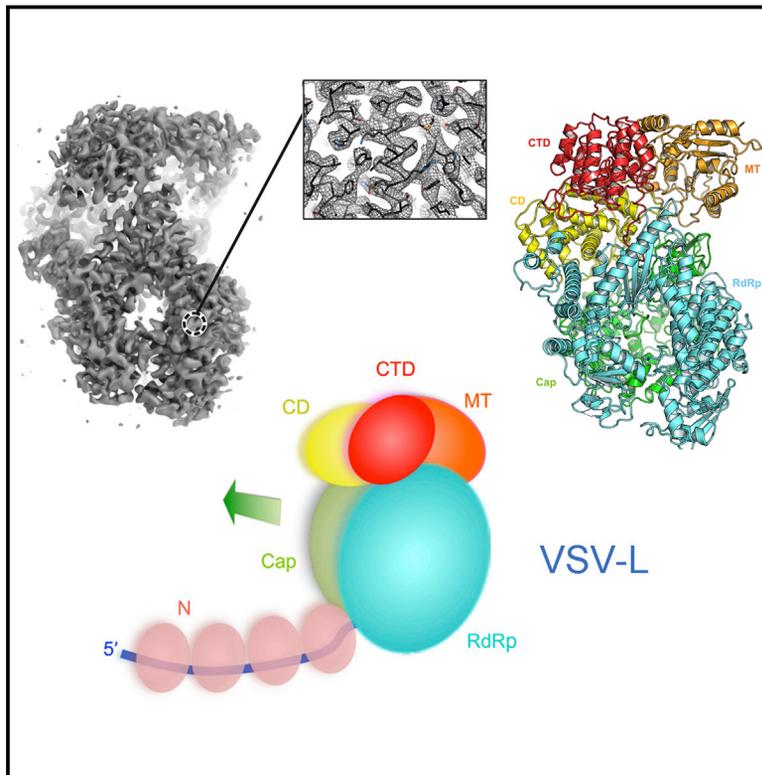


Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy

Graphical Abstract



Authors

Bo Liang, Zongli Li, Simon Jenni, ..., Nikolaus Grigorieff, Stephen C. Harrison, Sean P.J. Whelan

Correspondence

sean_whelan@hms.harvard.edu

In Brief

The vesicular stomatitis virus (VSV) L protein is the prototype of the single-chain RNA-dependent RNA polymerase, 5' capping enzyme, and methyltransferase in all non-segmented, negative-strand RNA viruses. The structure of VSV-L is now determined by electron cryomicroscopy at 3.8 Å resolution.

Highlights

- Vesicular stomatitis virus L protein structure from cryo-EM at 3.8 Å resolution
- Full de novo chain trace: RdRp, capping, MTase, and two structural domains
- P protein locks L in an initiation competent state with all five domains fixed
- Homology with other NNS virus polymerases (e.g., Ebola virus, RSV)



Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy

Bo Liang,^{1,3,6} Zongli Li,^{2,4,6} Simon Jenni,^{3,6} Amal A. Rahmeh,¹ Benjamin M. Morin,¹ Timothy Grant,⁵ Nikolaus Grigorieff,⁵ Stephen C. Harrison,^{3,4} and Sean P.J. Whelan^{1,*}

¹Department of Microbiology and Immunobiology

²Department of Cell Biology

³Department of Biological Chemistry and Molecular Pharmacology

⁴Howard Hughes Medical Institute

Harvard Medical School, Boston, MA 20115, USA

⁵Howard Hughes Medical Institute, Janelia Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA

⁶Co-first author

*Correspondence: sean_whelan@hms.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2015.06.018>

SUMMARY

The large (L) proteins of non-segmented, negative-strand RNA viruses, a group that includes Ebola and rabies viruses, catalyze RNA-dependent RNA polymerization with viral ribonucleoprotein as template, a non-canonical sequence of capping and methylation reactions, and polyadenylation of viral messages. We have determined by electron cryomicroscopy the structure of the vesicular stomatitis virus (VSV) L protein. The density map, at a resolution of 3.8 Å, has led to an atomic model for nearly all of the 2109-residue polypeptide chain, which comprises three enzymatic domains (RNA-dependent RNA polymerase [RdRp], polyribonucleotidyl transferase [PRNTase], and methyltransferase) and two structural domains. The RdRp resembles the corresponding enzymatic regions of dsRNA virus polymerases and influenza virus polymerase. A loop from the PRNTase (capping) domain projects into the catalytic site of the RdRp, where it appears to have the role of a priming loop and to couple product elongation to large-scale conformational changes in L.

INTRODUCTION

The non-segmented negative-strand (NNS) RNA viruses include some of the most lethal human and animal pathogens, including Ebola virus and rabies virus. Their multifunctional, large (L) polymerase proteins, carried within the virions (Baltimore et al., 1970), have biochemical properties that distinguish them from most other RNA polymerases of viruses or of their hosts. In addition to their RdRp activity (Emerson and Wagner, 1973), NNS RNA virus L proteins catalyze an unusual sequence of mRNA capping reactions (Hercyk et al., 1988), and the RdRp itself polyadenylates the viral message (Hunt et al., 1984). A nucleocapsid (N) protein sheath coats the genomic RNA, and the viral polymerase uses this N-RNA complex as template, rather than uncoated RNA.

Our understanding of RNA synthesis in NNS RNA viruses comes principally from studies of vesicular stomatitis virus (VSV)—an enveloped, bullet-shaped, rhabdovirus, closely related to rabies virus. VSV causes an acute disease of livestock. The VSV L protein does not bind the N-RNA template directly but requires a cofactor, the viral phosphoprotein (P), as a bridge (Green and Luo, 2009). Delivery of the N-RNA-L-P complex into the cytoplasm when the virus enters a cell initiates infection. Transcription starts at the 3' end of the genome and produces a triphosphate 47-nucleotide leader RNA, followed by sequential transcription of five capped and polyadenylated mRNAs (Abraham and Banerjee, 1976; Ball and White, 1976; Whelan and Wertz, 2002). At each gene junction, L terminates synthesis of the upstream gene, adding a polyadenylate (polyA) tail by iterative transcription of a U7 tract (Barr and Wertz, 2001; Barr et al., 1997; Stillman and Whitt, 1997). It then transcribes the downstream gene. Approximately 30% attenuation occurs at each successive gene, with dissociation of the template (Iverson and Rose, 1981). Replication also starts at the 3' end of the genome, but encapsidation by newly synthesized N accompanies synthesis of the nascent RNA strand, and in this mode, L ignores all the *cis*-acting signals that dictate sequential transcription of mRNAs (La Ferla and Peluso, 1989; Patton et al., 1984; Peluso and Moyer, 1983).

The various enzymatic activities of L are tightly linked. A GDP polyribonucleotidyl transferase (PRNTase) adds the cap structure (Ogino and Banerjee, 2007) when the nascent RNA chain length has reached 31 nucleotides, as shown by artificially stalling transcription at various chain lengths (Tekes et al., 2011). The unconventional mechanism of cap addition proceeds through a covalent adduct between a histidine residue on L (H1227) and the monophosphate nascent RNA, which is transferred onto a GTP-derived GDP acceptor (Li et al., 2008; Ogino and Banerjee, 2007). Failure to cap results in the premature termination of transcription, which links cap addition to RdRp processivity (Li et al., 2008; Stillman and Whitt, 1999; Wang et al., 2007). Subsequent methylation is first at the 2'O position on the ribose of the first nucleotide and then on the N7 of the capping guanylate—opposite to the typical order of modification (Rahmeh et al., 2009). Failure to methylate, either by exogenously manipulating the concentration of S-adenosyl

homocysteine (SAH) to compete with the S-adenosyl methionine methyl donor or by mutating critical catalytic residues in the methylase domain, can result in hyper-polyadenylation of mRNA transcripts, linking the methylation activity to the RdRp (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). Both cap addition and subsequent methylation require specific sequence elements within the first 10 nt of the mRNA (Wang et al., 2007). The uncapped leader lacks those signals (Li et al., 2008; Ogino and Banerjee, 2007).

During RNA synthesis, a few subunits of N dissociate from the template RNA for access to the catalytic site of RdRp and then re-associate as the process continues (Albertini et al., 2006; Green et al., 2006). P may help coordinate these events (Green and Luo, 2009). There are nine nucleotides associated with each N subunit in the ribonucleoprotein (RNP), but we do not yet know the number of transiently dissociated N subunits (and hence the length of the RNA segment inserted into the polymerase). Full polymerase processivity and correct recognition of the *cis*-acting signals in the viral genome both require N (Morin et al., 2012).

Images from negative-stain electron microscopy (EM) of purified VSV-L show a ring-like “core” decorated by a set of three variably oriented globular appendages (Rahmeh et al., 2010). Truncations of L map the RdRp to the ring-like domain. Formation of a complex with P causes the globular appendages to reorganize into a compact tail, wrapped onto one side of the ring (Rahmeh et al., 2010). Complex formation with P enhances RdRp initiation and processivity, but P itself has no enzymatic activity. A fragment of P comprising residues 35–106 is sufficient to induce the conformational rearrangement (Rahmeh et al., 2012). Segments in the center and at the C-terminal end of the P polypeptide chain mediate dimerization and N-RNA binding, respectively (Ding et al., 2006; Green and Luo, 2009). Images from negative-stain EM of purified L-P complexes are a mixture of single and dimeric L species, in which the two L molecules have variable relative orientation (Rahmeh et al., 2010).

We have determined the structure of a complex of L with P(35–106) by electron cryomicroscopy (cryo-EM). Into a density map at 3.8 Å resolution, we have built a nearly complete model of the 2109-residue polypeptide chain. The result shows that even for a fully asymmetric structure with molecular mass <250 kDa, the features in a density map from single-particle cryo-EM can be well enough defined for a complete *de novo* chain trace. We distinguish five domains: three—an RNA-dependent RNA polymerase (RdRp), an mRNA capping domain, and a methyltransferase domain—with assigned enzymatic activity and two—a connector between the capping and methyltransferase domains and a C-terminal domain—that appear to have largely organizational roles. We suggest that the conformation of the protein in the complex we have examined corresponds to an initiation state, ready to accept the 3' end of a template. Elongation beyond one or two nucleotides will require a large-scale domain reorganization.

RESULTS

Cryo-EM Structure Determination

We recorded images from VSV-L bound with P(35–106), from grids prepared for cryo-EM as described in [Experimental](#)

[Procedures](#). All data were taken on an FEI F20 microscope with a Gatan K2 Summit detector. Following two-dimensional classification with IMAGIC (van Heel et al., 1996) and TIGRIS (<http://tigris.sourceforge.net>) and calculation of an initial three-dimensional reference density with EMAN2 (Tang et al., 2007), we used FREALIGN (Lyumkis et al., 2013) for refinement and three-dimensional classification (see [Experimental Procedures](#)). [Figure 1](#) shows images, summarizes stages of the analysis, and illustrates the final 3.8 Å resolution density map.

Domain Organization of VSV-L

We traced the polypeptide chain from residue 35 to the carboxy terminus, residue 2109, leaving out poorly ordered linker segments 1335–1357 and 1558–1597 and a short, poorly ordered loop 1840–1849. Some segments of ~10–15 residues between 1100 and 1334 and between 1358 and 1557 are in poor density, and the chain trace in those regions is approximate. Because the beginning and end of the segment and “puddles” of density leave little uncertainty about the overall course of the segment in question, we have kept those residues in the model. We list the specific segments in question in the [Experimental Procedures](#). We paid considerable attention to correct stereochemistry during model building, but to adjust both the fit and the stereochemistry beyond the capacity of visual inspection, we carried out one round of refinement, by calculating structure factors from the map and using both amplitudes and phases in the target function, with secondary-structure hydrogen bonds imposed as restraints. After some minor adjustments and one more cycle, R_{free} and R_{work} were 29.6 and 26.2, respectively ([Figure S1](#); [Table S1](#)).

We have assigned residues to domains and linkers as follows: RdRp, 35–865; capping domain, 866–1334; linker 1 1335–1357; connector domain, 1358–1557; linker 2, 1558–1597; methyltransferase, 1598–1892; C-terminal domain, 1893–2109 ([Figures 2](#) and [S2](#); [Data S1](#)). The boundaries between the capping domain and the connector domain and between the latter and the methyltransferase are evident from the disordered linkers that intervene. The boundary between the RdRp and the capping domain corresponds to a previously identified tryptic cleavage site. Although the interface between the two domains is relatively extensive and well packed, we have shown that fragment 1–860 can be expressed independently (Rahmeh et al., 2010). The boundary between the connector domain and the methyltransferase domain also corresponds to two independently stable fragments: 1–1593 and 1594–2109. Between the methyltransferase domain and the C-terminal domain is an extended but relatively polar interface. Negative-stain electron microscopy of a fragment that includes residues 1594–2109 shows that stain can penetrate between the two domains and that their connection might be flexible (Rahmeh et al., 2010).

There are well defined secondary-structure features in the density for each of the domains, leading us to believe that poor density for certain segments reflects local disorder, rather than failure of three-dimensional classification algorithms to detect overall variation in position or orientation of any particular domain with respect to the others.

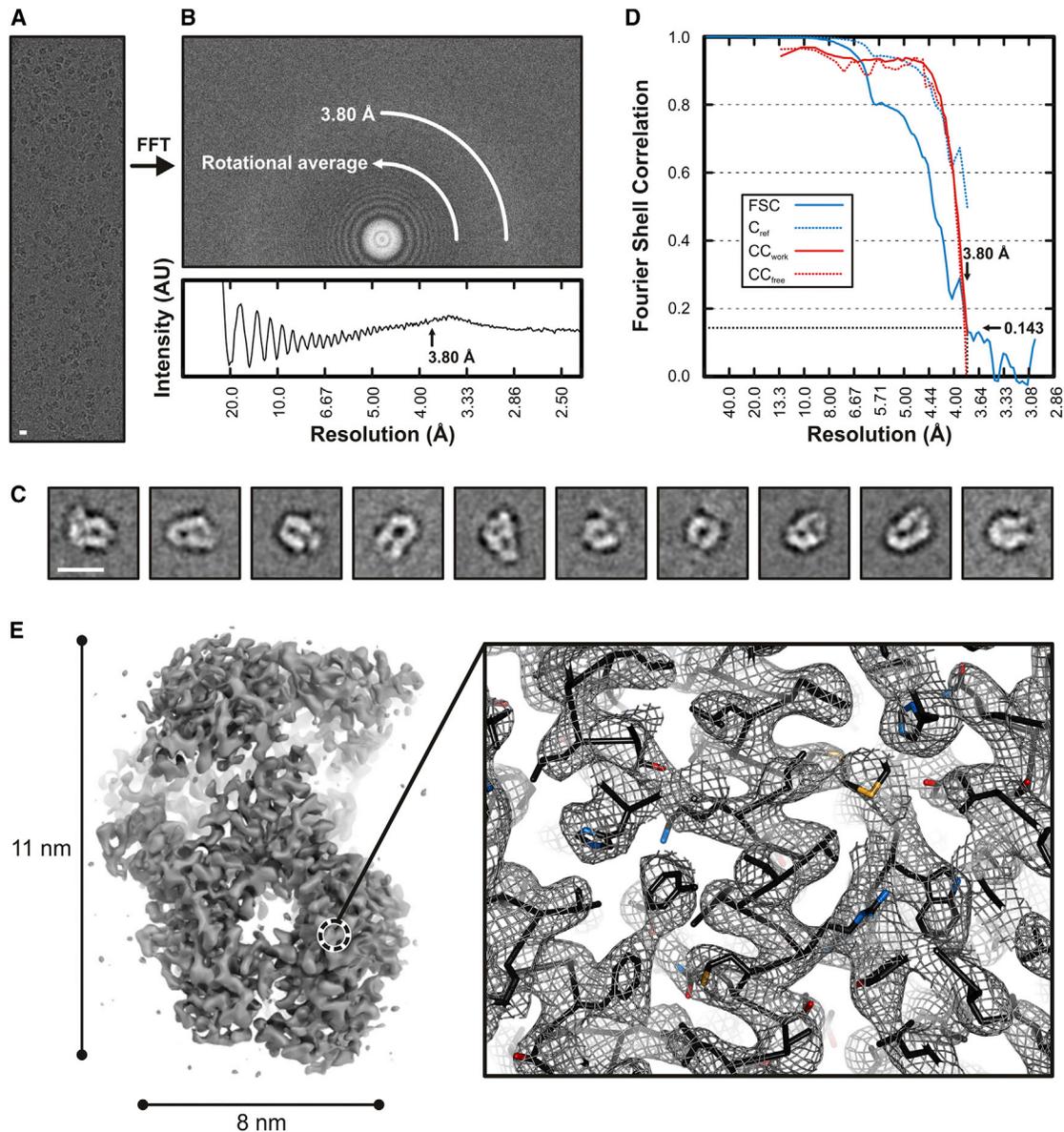


Figure 1. Electron Cryomicroscopic Reconstruction of VSV-L at 3.8 Å Resolution

(A) Raw image of VSV-L particles in vitreous ice recorded at 1.8 μm defocus. Scale bar, 10 nm.

(B) Power spectrum of the image shown in (A), with plot of the rotationally averaged intensity versus resolution. Arrow indicates the spatial frequency corresponding to 3.8 Å resolution.

(C) Representative class averages. Scale bar, 10 nm.

(D) Fourier shell correlation analysis: FSC, correlation between the half-set three-dimensional reconstructions (solid blue line); C_{ref} , estimated correlation between the final map and a perfect reference map containing no errors, calculated from FSC (dotted blue line) (Rosenthal and Henderson, 2003); CC_{work} (solid red line) and CC_{free} (dotted red line), correlation between the final map and refined model for working and test set of structure factors, respectively.

(E) Left: overview of VSV-L reconstruction. In the view shown, the particle (241 kDa) is ~ 110 Å long and 80 Å wide. Right: close-up view of a representative region in the polymerase domain (RdRp). The volume shown in close-up is from the protein interior, not on the RdRp surface. Density is shown as gray mesh; polypeptide-chain backbone of the refined model, as black ribbon; side-chain atoms, as sticks (carbons, black; nitrogen, blue; oxygen, red; sulfur, orange). Note the continuous backbone density, α -helical grooves and resolution of bulky side chains—features that allowed building and stereochemical refinement of the atomic model.

See also Figure S1 and Table S1.

RdRp

The RdRp has at its core a right-hand, “fingers-palm-thumb” structure (residues 360–865) common to a very large group

of RNA and DNA polymerases (Figure 3). The catalytic site on the palm is in a deep channel between the fingers and thumb subdomains (extended from the palm as if in a

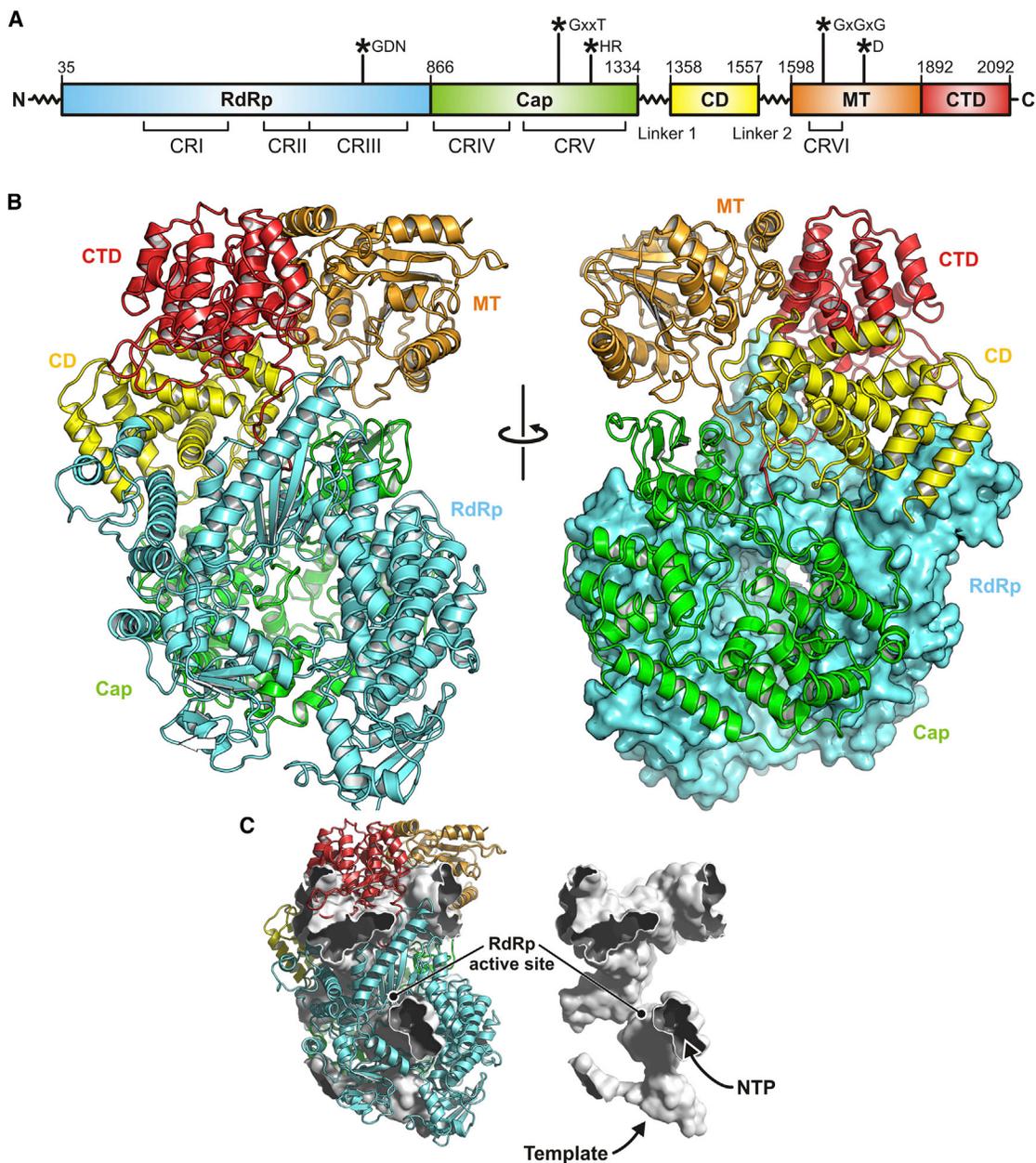


Figure 2. Structure of VSV-L

(A) Domain organization of VSV-L. The polymerase domain (RdRp) is in cyan; capping domain (Cap), green; connector domain (CD), yellow; methyltransferase (MT), orange; C-terminal domain (CTD), red. Amino-acid residue numbers indicate functional domain boundaries. Flexible linkers 1 and 2 connect Cap to CD and CD to MT domain, respectively. Conserved regions within L proteins of non-segmented negative-strand (NNS) RNA viruses are labeled CR I–VI. Asterisks indicate the position of active site residues.

(B) Ribbon diagram of VSV-L polypeptide chain; domains colored as in (A).

(C) Substrate channels and internal cavities of VSV-L, depicted as white surface enclosed by the structure in ribbon representation. In this orientation, the entrance to the template channel leading to the active site faces down; the channel runs between the RdRp and capping domains. Nucleotides can access the RdRp active site through the channel in the foreground.

See also [Figure S2](#).

loose hand grip). Appended to the core on the N-terminal side is a globular region (residues 1–359) that closes the channel on one end and reinforces the relatively slender thumb subdomain.

From the appearance of VSV-L in negative-stain electron microscopy, and in particular from the size and staining of a “doughnut-like” part, we suggested that the RdRp might be similar in cage-like structure to the dsRNA virus polymerases

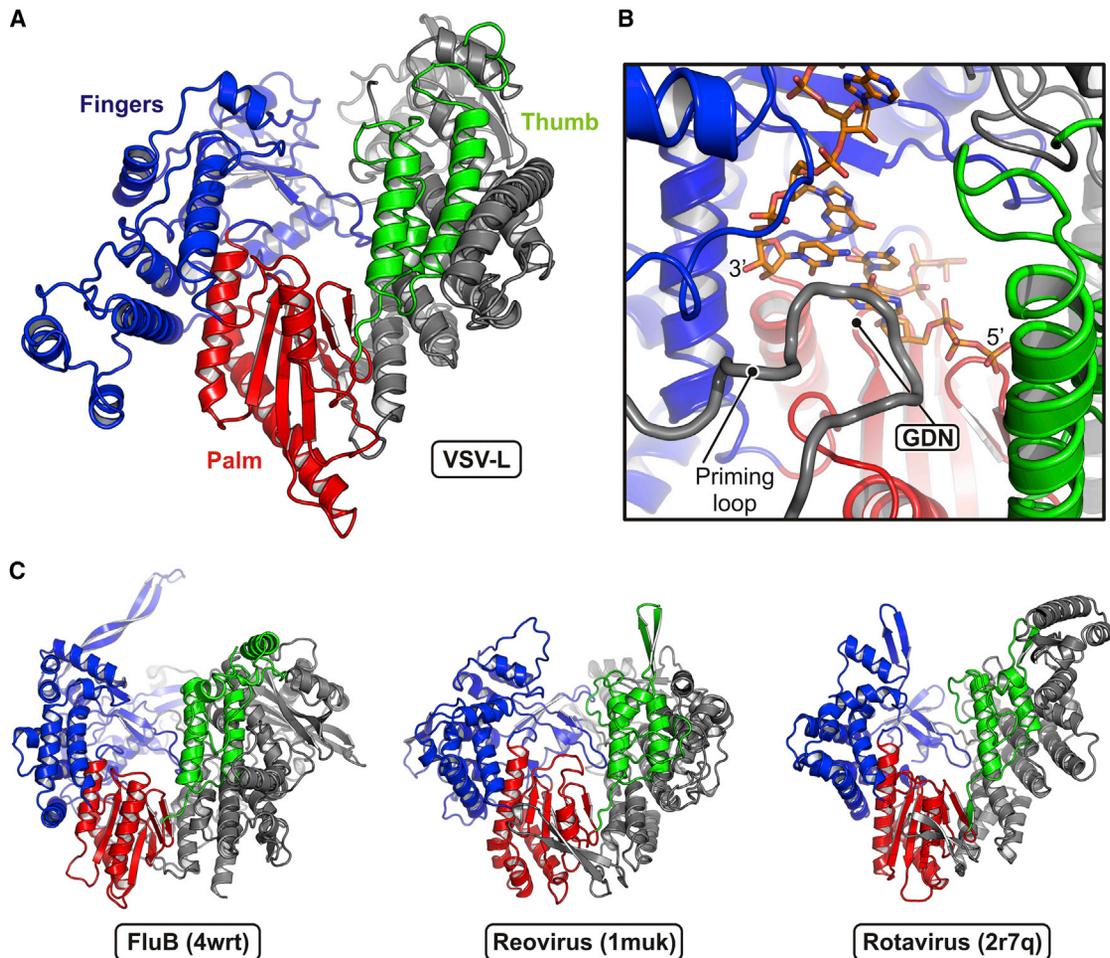


Figure 3. Polymerase RdRp Domain

(A) Structure of the RdRp domain. Residues 35–865 are shown in ribbon representation in conventional orientation (viewed from inside the surrounding “cage” and upside down with respect to the view in Figure 2). The palm subdomain is in red; the fingers, blue; the thumb, green. The N-terminal region is gray.

(B) Close-up view of the active site. Palm, fingers and thumb are colored as in (A). The GDN active site motif is at a β -hairpin in the palm domain. A model for the positions of the template RNA strand and two nucleotides is derived from the reovirus λ 3 initiation complex (PDB: 1n1h) after superposition on VSV-L RdRp. The priming loop (residues 1157–1173) intruding from the capping domain (gray) positions the initial nucleotide of the transcript.

(C) Similarity of the VSV-L RdRp domain to those of other viral polymerases. Structures of influenza virus B polymerase (PDB: 4wrt; PA residues 248–716 and PB1 residues 1–616), reovirus λ 3 (PDB: 1muk; residues 2–890) and rotavirus VP1 (PDB: 2r7q; residues 2–778) are shown with the same orientation and coloring scheme as the VSV-L RdRp in (A).

(Rahmeh et al., 2010). Those enzymes have their catalytic sites at the center of an enclosed cavity, connected to the exterior by four channels, for template entrance, template exit, transcript exit, and NTP access (Lu et al., 2008; Tao et al., 2002). Comparison of the chain trace with their structures shows that this suggestion was correct, with one modification. The dsRNA virus RdRps have a C-terminal “bracelet” domain that encircles the exit path for the template and includes a site for binding the methyl G cap on the non-template, plus-sense strand (Lu et al., 2008; Tao et al., 2002). In VSV L, the capping domain, which has no structural similarity to the bracelet domain of the dsRNA virus RdRps, occupies the corresponding space. That is, residue 865, which we have taken as the end of the RdRp, is at the C terminus of the thumb.

We compared the positions of secondary structural elements in VSV L, reovirus λ 3 (Tao et al., 2002), rotavirus VP1 (Lu et al., 2008), and the heterotrimeric influenza virus polymerase (Pflug et al., 2014; Reich et al., 2014). The secondary structural elements with correspondences in the three other polymerases extend from about residue 107 in VSV-L to the end of the RdRp domain (Figure 3). The analogous parts of reovirus λ 3 encompass residues 150–860 (approximately); those of rotavirus VP1, residues 135–750; those of human influenza virus B polymerase, residues 415 to the C terminus (714) of the PA subunit, and residues 8–586 of the PB1 subunit. The homology thus extends from the middle of PA into PB1. The region in common between VSV-L and influenza virus PB1 corresponds to the fingers-palm-thumb core structure, and the

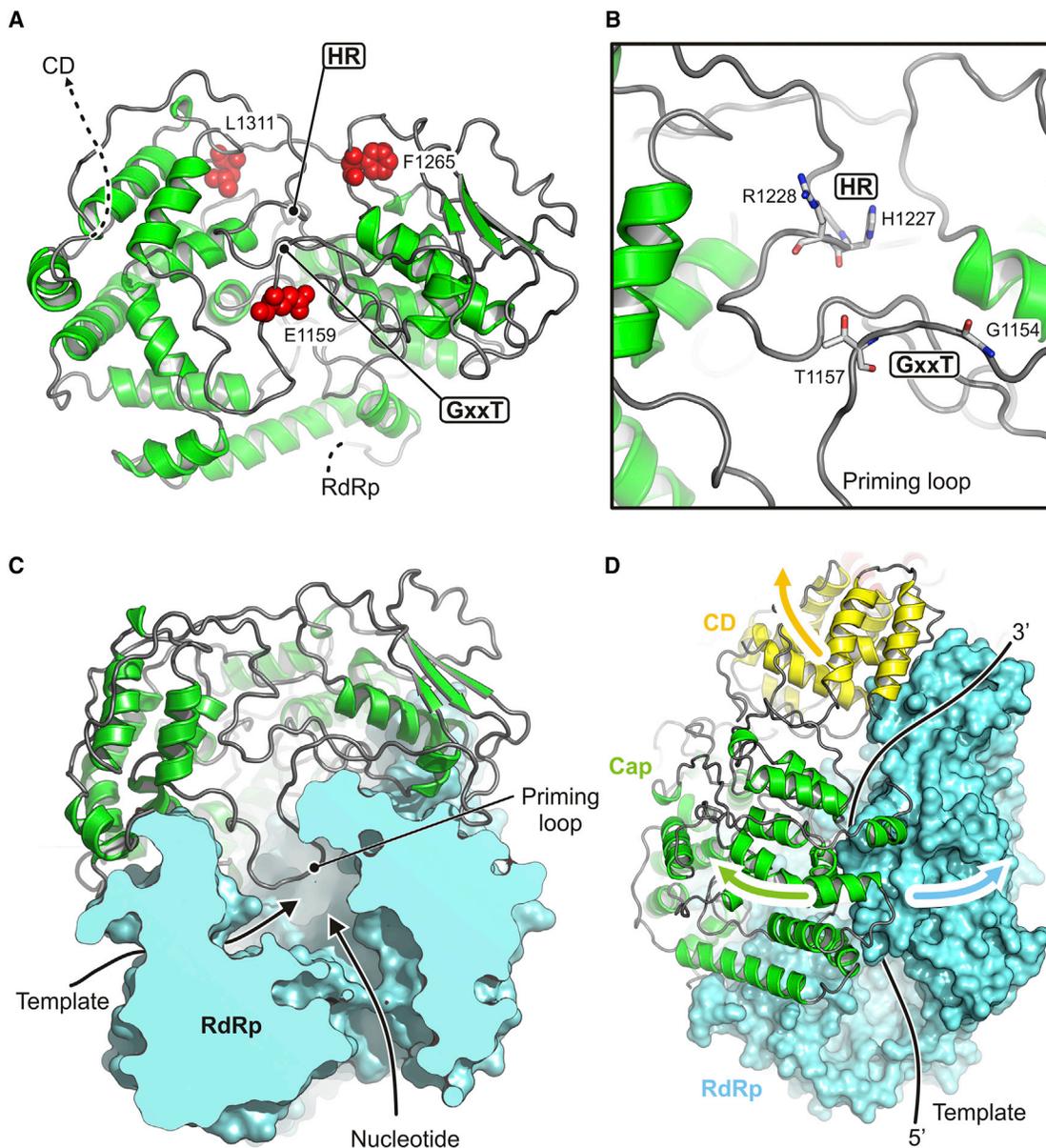


Figure 4. Capping Domain

(A) Structure of the capping domain. Residues 866–1334 are in ribbon representation. Motifs GxxT and HR are sites of guanosine nucleotide binding and of covalent RNA attachment, respectively. Residues corresponding to positions of inhibitor-resistance mutations in human RSV polymerase (Liu *et al.*, 2005) are shown as red spheres.

(B) Close-up of the active site.

(C) Configuration of the priming loop in VSV-L. Only the RdRp and capping domains are shown. The priming loop (residues 1157–1173) protrudes from the capping domain into the active site of the RdRp domain.

(D) Proposed domain shifts to allow transcript elongation and eventual template release.

region shared with PA is a large part of the RdRp N-terminal domain.

Capping Domain

Unlike the corresponding host-cell process, the capping reaction of NNS RNA viruses proceeds from a covalent linkage between the 5' end of the RNA and a histidine residue, with attack

on that linkage by a guanosine nucleotide. The enzyme is thus a polyribonucleotidyl transferase (PRNTase) rather than a guanylyl transferase. Two conserved motifs—GxxT and HR, separated by ~70 residues—mark the catalytic site (Figures 4A and 4B). The former participates in guanosine nucleotide binding; the latter is the site of covalent RNA attachment. The domain has no structural homologs that we could detect with standard

search methods. The largely α -helical, N-terminal half (residues 866–1100), which abuts the polymerase domain, is well ordered. The C-terminal half (1100–1334) has several poorly defined segments, including the loop that bears the HR sequence. Despite uncertain definition of side chains, the separation of the two conserved sites is ~ 10 Å, appropriate if a GTP bound at the former is to attack the histidine-liganded 5' phosphate of the nascent RNA at the latter (Figure 4B). Positions corresponding to sites in human respiratory syncytial virus (hRSV) of resistance mutations to a small-molecule capping inhibitor (Liuzzi et al., 2005) impinge on the active site from three sides (Figure 4A); their locations, and the relatively poor definition of the active site in the map, suggest that activation of the domain, perhaps by binding the 5' end of the nascent message, induces a conformational rearrangement, similar to the domain closures seen in many enzymes when they bind their substrate. Two candidate Zn sites, one with clear density where two Cys (residues 1120 and 1123) and two His (1294 and 1296) ligand the likely Zn ion, and one with three Cys (residues 1081, 1299, and 1302) and a Glu (1108), contribute structural integrity to the capping domain. The sites are close to each other and well outside the catalytic center. In both cases, the liganding residues are present as a conserved set in most NNS virus L proteins and absent as a set in the others.

A loop between residues 1157 (the threonine of the GxxT motif) and 1173 projects back into the cavity of the polymerase domain (Figures 3B, 4C, and S3). The poorly ordered tip of this loop occupies the same position as the priming loop in the reovirus polymerase (Tao et al., 2002). The loop in VSV-L polymerase domain that corresponds to the $\lambda 3$ priming loop is shorter than its reovirus homolog, and the capping-domain loop projects over it. Neither polymerase requires a polynucleotide primer to initiate, and the priming loop in the reovirus polymerase supports the initiating nucleoside triphosphate. As elongation proceeds, the tip of the loop recedes to make room for the dsRNA replication product or for the short double-stranded region just upstream of the newly added nucleotide during transcription (Tao et al., 2002). This loop, which contacts the minor groove of the nascent product, may also enhance fidelity, by retarding elongation of mismatches detected by poor minor-groove geometry and allowing more time for ATP-based pyrophosphorolysis of the mismatch. The position of the likely priming loop of VSV-L on the capping domain, adjacent to the GxxT residues, suggests coupling of capping to initiation of polymerization.

Connector Domain

The connector domain is a bundle of eight helices (Figure 2B); it appears to have largely an organizational role in positioning or spacing the catalytic domains. Disordered linkers, 23 and 40 residues long, respectively, lead into and out of the connector domain. The endpoints of these linkers in well-defined density show that they must occupy an extended groove between the capping and connector domains; the groove also extends into the interface between the capping and methyltransferase domains (Figure 5). Strong, low resolution density features fill this groove, but they are not sharp enough to suggest particular linker conformations (Figure 5B). The location of P indicated by

negative-stain electron microscopy (Figures 5A and S4; Table S2) leads us to suggest that the groove also holds some or all of the P fragment present in the L-P complex we have imaged. Because P(35–106) locks the smaller domains of L into a fixed configuration, it is plausible that it might do so by stabilizing folded structures for the two linker segments—gluing them down, so to speak, alongside the connector domain (Figure 5B).

Methyltransferase Domain

The methyltransferase domain has the structure characteristic of many other domains that catalyze transfer of a methyl group from S-adenosyl methionine (Figures 6 and S5). It methylates both the ribose O2' and the guanosine N7 (Rahmeh et al., 2009). Most of the domain superposes extremely well on the flavivirus methyltransferases, also dual specificity enzymes (Egloff et al., 2002; Ray et al., 2006; Zhou et al., 2007). Evidence for functional feedback from the VSV-L methyltransferase to the RdRp comes from the observation that addition of S-adenosyl homocysteine, which inhibits methylation, leads to hyperpolyadenylation; mutations that prevent methyl transfer have a similar effect (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). The methyltransferase domain contacts both the connector and the capping domains, but it has no direct contact with the RdRp. Moreover, there is no obvious “tunnel” that would allow the 5' end of the transcript to move from the catalytic site of the capping domain to the catalytic site of the methyltransferase domain. We conclude, as we discuss in greater detail below, that the L protein probably undergoes a substantial conformation change following initiation of polymerization and that the inter-domain communication we see in this structure is relevant to formation of the first one or two phosphodiester bonds, but not to subsequent elongation and 5'-end modification.

C-Terminal Domain

Like the connector domain, the C-terminal domain, which terminates in a ~ 25 -residue long C-terminal “arm,” appears to have an essentially organizational role (Figure 2B). It is largely an α -helical bundle, but a projecting, almost beak-like, β -hairpin supported by a second interhelical loop, imparts a noticeable asymmetry. The C-terminal arm, a feature that appears from sequence alignments to be conserved among NNS viral polymerases, but variable in length, extends back against the RdRp, augmenting the β -hairpin that bears the catalytic Asp-Asn sequence at its tip, and terminates at the three-way junction of the capping, connector, and methyltransferase domains, where it has one or more contacts with each. The arm thus contributes to closing the multi-domain structure we see in the L-P complex, further stabilized by the phosphoprotein, P.

Template Channel

Superpositions of related positions in reovirus $\lambda 3$ (Tao et al., 2002) and rotavirus VP1 (Lu et al., 2008) have allowed us to model a bound template and a template-primer-NTP complex, because $\lambda 3$ was catalytically active in the crystals studied used to determine the structure and VP1 in its crystals incorporated template in a sequence-specific register (Tao et al., 2002). The template entrance channel is at the interface between the

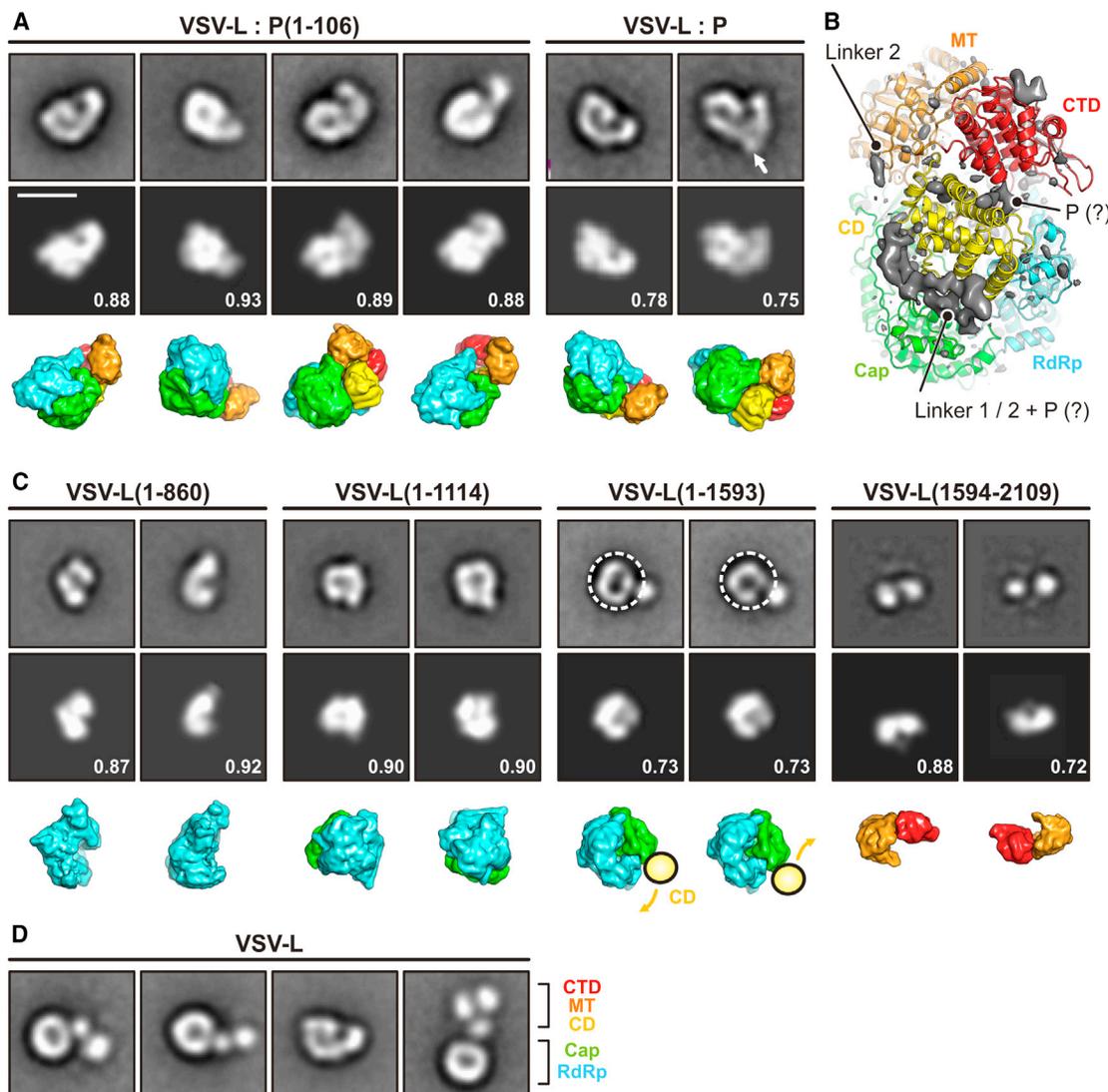


Figure 5. Domain Reorganization

(A) Projection angle matching between class averages of negatively stained complexes of VSV-L and P protein (top row) (Rahme et al., 2010, 2012) and projections calculated from the model (middle row). The bottom row shows the model in the same orientation with the individual domains colored as in Figure 2. Numbers are correlation coefficients between model and negative-stain class averages. VSV-L:P(1–106) corresponds to the structure determined here. In the panel for VSV-L:P, an arrow indicates additional density observed in some class averages that we attribute to the bound P dimer. Scale bar, 10 nm.

(B) Difference density map ($\text{map}_{\text{observed}} - \text{map}_{\text{model}}$) calculated to 5 Å resolution and shown together with the model. The map shows density present in the image reconstruction that could not be fit with a molecular model. Strong density—presumably from linkers 1 and 2, which enter and leave this density at defined points, and with potential contribution from P (tentative assignment indicated by “?”)—lines the groove between the capping and connector domains. We have not attempted to interpret the small, low-resolution feature at the upper right.

(C) Projection angle matching of VSV-L fragments. For VSV-L(1–1557), the negative-stain class averages suggest a conformationally variable connection between the connector domain (CD) and the polymerase (RdRp) and capping domain (Cap). We therefore selected only the “doughnut” part of the image and aligned residues 35–1334.

(D) Full-length VSV-L without P. CD, MT, and CTD extend in variable orientation from the RdRp-Cap doughnut.

See also Figure S4 and Table S2.

RdRp and the capping domain (Figure 2C). Polar and especially basic residues project into the groove from both sides. As in all polymerases of this family, the template runs across a “fingers loop” (residues 523–545 in VSV-L) and twists sharply to present the templating base to the catalytic center. A hydrophobic residue in the loop (Phe541 in VSV-L) bears on the templating

base to enforce correct base pairing with the incoming nucleoside triphosphate. For initiation at the 3' end of the viral RNA (either for replication or for transcription of leader RNA), the priming nucleoside triphosphate will rest against the loop from the capping domain described above (Figure 3B). Any further elongation, after forming the initial phosphodiester bond, will

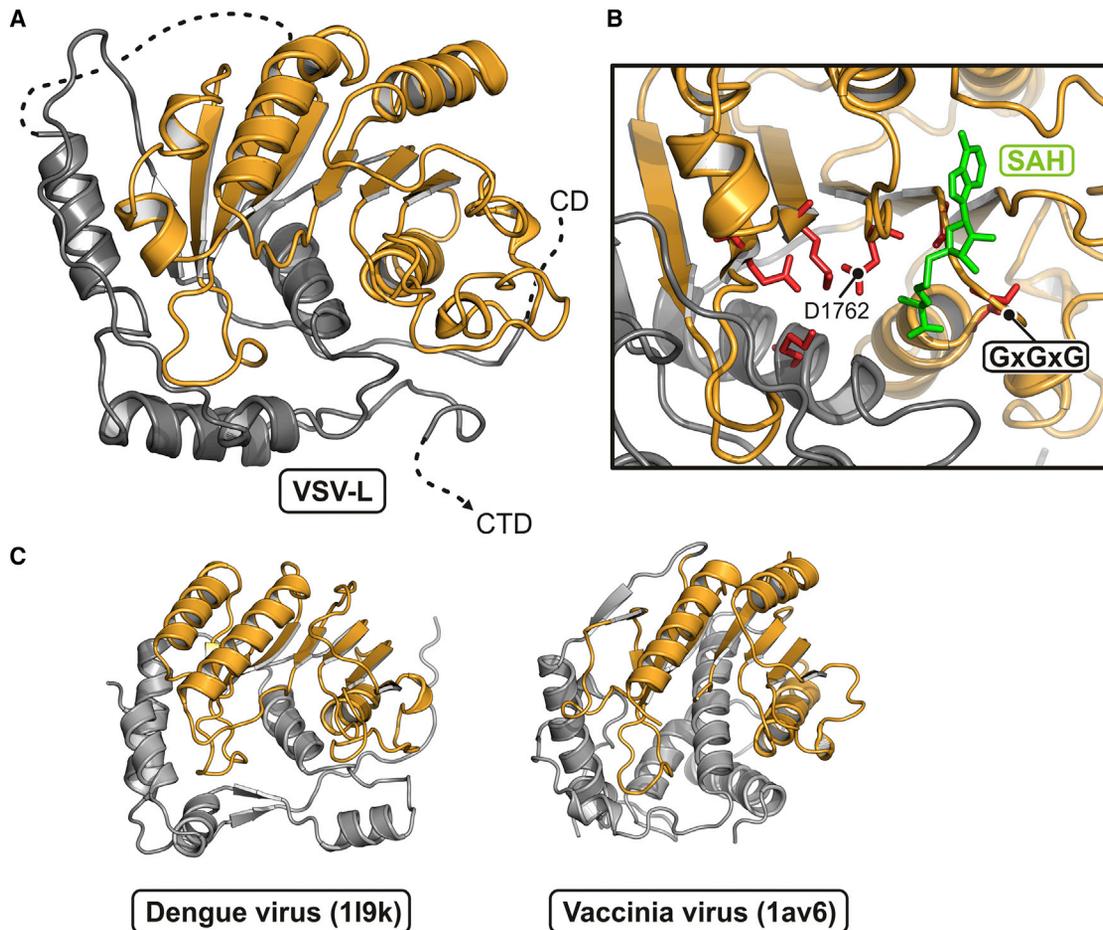


Figure 6. Methyltransferase Domain

(A) Structure of the methyltransferase: residues 1598–1892 in ribbon representation. The consensus fold of the S-adenosyl methionine-dependent methyltransferase subdomain is in orange. The N-terminal and C-terminal regions are in gray.

(B) Close-up of the active site. The SAM/SAH binding-site motif, GxGxG, is between $\beta 1$ and αA . An SAH molecule (green) is derived from a superposition of its complex with dengue virus NS5 MT (PDB: 119k). Residues that participate in the methyltransferase activity are in red.

(C) Comparison of VSV-L MT domain with other viral AdoMet-dependent methyltransferases. Structures of dengue virus NS5 MT (PDB: 119k; residues 7–267) and vaccinia virus VP39 MT (PDB: 1av6; residues 3–297) are shown in the same orientation and color scheme as in (A).

See also [Figure S5](#).

require this loop to move, and substantial elongation will almost certainly require displacement of the entire capping domain ([Figure 4D](#)). Indeed, we suggest that to accommodate transcriptional elongation, the entire array of smaller domains may reorganize.

Domain Reorganization

The configurations of VSV-L we have characterized in published work by negative-stain electron microscopy illustrate the potential for large-scale domain reorganization ([Rahmeh et al., 2010](#)). Images of L alone show a core “doughnut,” which admits stain at its center, three globular appendages, in apparently variable positions and orientations with respect to each other and to the core. Addition of P, or of the peptide, residues 35–106, that we have used to stabilize the complex studied here, locks the appendages in place ([Rahmeh et al., 2010](#),

[2012](#)). Many of the projections of this locked structure resemble a figure “6.” Class averages from these images agree extremely well with projections of the structure we describe ([Figure 5A](#)), as do class averages from images of four different L fragments ([Figure 5C](#)). One of these (1–860) corresponds precisely to the RdRp. Another, 1–1121, includes the RdRp and the largely helical, N-terminal half of the capping domain. The tryptic cleavage that initially generated that fragment is in a surface loop. The last fragment previously imaged by negative staining comprises the methyltransferase and C-terminal domains ([Figure 5C](#)).

These comparisons show that we need to modify our initial assignment of the three globular appendages to the capping domain, the methyltransferase domain, and one unassigned domain ([Rahmeh et al., 2010](#)). The capping domain is part of the doughnut, and the appendages correspond to the

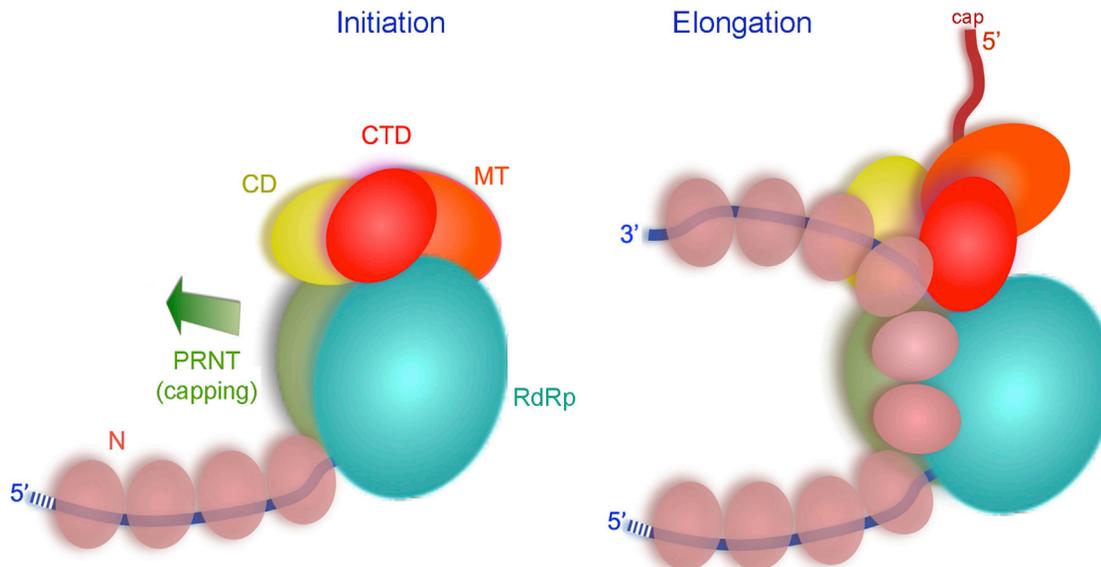


Figure 7. Model for Transcription of an RNP Complex by VSV-L

Left: initiation complex, with domains organized as in the structure described here. Viral genomic RNA is in blue, N protein is in beige, and domains of VSV-L are in the colors used in Figure 2. Arrow shows direction of capping domain displacement required for the transition to an elongation complex. Right: elongation requires both displacement of the capping domain (with likely accompanying reorganization of the CD, MT, and CTD) and displacement of two to three N subunits from the template residues looped into the polymerase. The N subunits are shown linked as a continuous chain, as suggested by their structure (Green et al., 2006). The emerging transcript is in red.

connector, methyltransferase, and C-terminal domains, respectively (Figure 5D). The linkers between the capping domain and the connector and between the connector and the methyltransferase clearly allow the latter two to move away from the rest of the molecule; good definition in negative stain for the third globular appendage suggests that in the unlocked structure, the C-terminal arm also pulls away from the RdRp. Many of its interactions, as it inserts back against the rest of the molecule in the structure we have determined, are indeed with the connector and methyltransferase domains.

Images of negatively stained complexes of L with inact, dimeric P often show two, linked, figure-“6” L molecules, but occasionally the P dimer does not recruit a second L and appears as a surface feature on the hook of the “6” (Figure 5A). Comparison of the L structure with these projections is consistent with our proposal that the interaction with P(35–106) that stabilizes the “6” conformation is with the linker segments at either end of the connector domain (Figure 5B).

DISCUSSION

Cryo-EM

High-resolution cryo-EM structure determination has until recently relied on either high symmetry or large size—for example, icosahedral viruses, which have both, or ribosomes, which are large enough to produce reasonable contrast for getting started with iterative determination of particle orientations and centers (Grigorieff and Harrison, 2011). Developments in cryo-EM during the past 5 years have now allowed us to deter-

mine the molecular structure of an asymmetric protein of total mass <250 kDa. Dose fractionation (“movies”), enabled by use of a direct electron detector, and refinement and maximum-likelihood classification procedures (Lyumkis et al., 2013), implemented in FREALIGN, were crucial for achieving a resolution adequate to build an atomic model.

Sequential Transcription

A de novo initiation event with ATP as initiating nucleotide appears to start synthesis of each mRNA transcript. We interpret our structure as that of an early initiation state, representing an L-P complex ready for loading onto the end of the template to synthesize leader RNA. During the transition to elongation the priming loop—contributed by the capping domain—must shift out of the way to accommodate the product (Figure 7). Inspection further suggests that after addition of only a few more nucleotides, the capping domain as a whole must withdraw from tight contact with the RdRp to allow further elongation, as there do not appear to be clear exit channels for transcript and template. Upon termination of a transcript, the polymerase reinitiates on the next gene, but the efficiency of producing the succeeding transcript is only ~70% (Iverson and Rose, 1981). The template entrance channel in VSV-L is at the interface of the capping domain and the RdRp, and dissociation of the template will be straightforward (when initiation or early elongation aborts) if that interface opens as suggested. Otherwise, the entire template would have to thread through the active site and emerge through another channel. Transcription of the downstream gene probably requires reestablishing the inter-domain

contact seen in our structure, so that the priming loop can reinsert into the active site of the RdRp domain for subsequent de novo initiation.

Coupling of Capping, Polymerase, and Methyltransferase Activities

Displacement of the capping domain from the RdRp as elongation proceeds might have two consequences. First, the active site of the PRNTase might reorganize (e.g., by “domain closure”) into a better ordered configuration than the one we see in the present structure. Second, because the capping domain faces both the connector and methyltransferase domains, its displacement might also induce rearrangement of the rest of the capping machinery. A large-scale reorganization of this kind could account for some of the observed functional crosstalk between the capping and polymerase activities.

A cap is added only when the length of a transcript has reached 31 nucleotides (Tekes et al., 2011). In vitro, very short (up to 5 nt) transcripts can be capped in *trans* by L, but this process is inefficient and fails completely with longer transcripts (Li et al., 2008; Ogino and Banerjee, 2007). Conversely, mutations in L that disrupt cap addition cause premature termination (Li et al., 2008, 2009). Mutations in the specific, *cis*-acting signals at the 5' end of the nascent strand, which are absent from the leader RNA, also block cap addition and result in premature termination of that transcript (Li et al., 2008; Ogino and Banerjee, 2007; Stillman and Whitt, 1997; Wang et al., 2007). The precision of the 31-nt requirement suggests that the reorganized structure that allows elongation is a well-defined state, rather than a loosely ordered one.

Reorganization of the capping machinery can also account for why mRNA cap methylation requires no additional chain length (Tekes et al., 2011). In the configuration represented by the structure we have determined, the catalytic sites for capping and methylation are distant from each other. If the smaller domains move away from the polymerase core, the capped, nascent RNA could probably release from the capping enzyme and gain immediate access to the methylase domain. Methylation in *trans* can occur under some circumstances, but previous work has shown that transcripts stalled at a chain length of 31 nt are fully methylated—presumably in *cis*—by the stalled L (Tekes et al., 2011).

Inhibition of mRNA cap methylation by high concentrations of S-adenosyl homocysteine can result in hyper-polyadenylation, demonstrating a linkage between the methylase and RdRp domains (Galloway and Wertz, 2008; Li et al., 2009; Rose et al., 1977). The RdRp domain of L carries out polyadenylation by iterative transcription of a gene-end U tract element. Some, but not all, mutations that inhibit methylation result in the hyper-polyadenylation phenotype (Galloway and Wertz, 2008), indicating that the crosstalk mechanism is not a readout of cap modification but probably a consequence of interactions between domains and between the protein and the nascent transcript.

The cap methylase of VSV participates in both ribose 2'O and guanine-N7 methylation reactions. The preferred substrate for all other ribose 2'O methyltransferases is 7mGpppN and like other proteins that recognize the mRNA cap structure—such

as eIF4E-2'O methylases—those enzymes position the ribose in the active site by π - π stacking interactions with the 7mG RNA. The order of cap methylation in VSV is reversed. Methylation of 2'O precedes and facilitates subsequent methylation of guanine-N7. The absence of aromatic residues that could participate in such interactions with a 7mGpppN RNA in the VSV methyltransferase is consistent with this altered reaction sequence.

The N Protein

The template for polymerase is not naked RNA, but a complex in which the template RNA is encased within the nucleocapsid protein sheath (Figure 7). In that complex the RNA bases are not accessible to the RdRp of L, and the N protein must transiently dissociate from the RNA for the RdRp to proceed (Green et al., 2006). The structure of L allows us to estimate that 20–25 nt of the template strand are threaded through the polymerase domain. Accordingly, because each molecule of N covers 9 nt of RNA, two or three molecules of N must be displaced from the template strand at any one time. Adjacent N subunits in the RNP interact stably, embracing each other through N- and C-terminal extensions (Green et al., 2006). Thus, looping out of template RNA need not entail dissociation of N from the RNP coil. Indeed, if we consider the linked chain of N subunits as the analog of a cRNA strand, then the displaced N is the counterpart of a looped-out plus-sense strand during transcription by the related polymerases of dsRNA viruses. This N-protein bridge could account for the precision of the 31-nt length of nascent transcript required for cap addition, perhaps by creating a defined spacing between the RdRp and the popped-out capping domain. The N protein influences L activity, as recognition of the *cis*-acting signals in the genome requires it and as its presence influences incorporation by L of substituted nucleotide analogs (Morin and Whelan, 2014). It may also be necessary for capping.

The P Protein

P is an adaptor that engages both the N-RNA template complex and the L protein. A small, globular domain at the C-terminal end of P (residues 195–265) interacts with the N-RNA complex (Green and Luo, 2009). This domain could in principle move from one subunit to the next as polymerization proceeds. The structure we describe here contains only part of the N-terminal region of P. Although it is poorly ordered in our density map, we suggest that P(35–106) may occupy some of the strong, low resolution density features between the capping and connector domains, locking in the linker segments at both ends of the latter. Depending on the chain polarity and on the flexibility of intervening segments, the C-terminal domain of P could lie near the opening through which RNA enters the active site of the RdRp and thus, through its interaction with N, be part of the process that feeds template rapidly through the polymerase channel.

Homologies and Comparisons

Homologous L proteins include those of rabies, Ebola, measles, and respiratory syncytial viruses. Alignment of their sequences (Data S1) shows the same overall arrangement of the various domains, identifies the active site residues of the protein for

RdRp, PRNTase, and methyltransferase activities and suggests domain boundaries for expressing various fragments of the proteins from VSV and related viruses. All NNS RNA viruses have a polymerase complex that comprises the enzymatic subunit, L, and an equivalent of the VSV phosphoprotein, P. In some cases, additional viral proteins (VP24 in the case of the filoviruses, M2-1 in the case of respiratory syncytial virus) are necessary for full polymerase processivity. The three-dimensional interconnections among domains in the VSV polymerase suggest that binding of these accessory proteins to any of the smaller domains might influence large-scale rearrangements and affect enzyme processivity.

Parts of L are structurally similar to the heterotrimeric polymerase of influenza virus. The major differences are the capping machineries, reflecting the distinction between influenza-virus cap-snatching and NNS RNA-virus cap synthesis (Reich et al., 2014). A structural feature of the influenza virus polymerase that is also absent from the VSV polymerase is the long PB1 arm, which is probably required for a step of mRNA transcription (Pflug et al., 2014; Reich et al., 2014).

The VSV-L structure allows preliminary interpretation of mechanisms for inhibitors of its homologs. For measles virus, resistance to a non-nucleoside analog inhibitor that blocks gene expression maps to the polymerase domain (Krumm et al., 2014). The positions of those mutations, mapped onto VSV-L, flank the GDNQ motif at the RdRp catalytic site, suggesting an allosteric mechanism, like that of the non-nucleoside analog reverse transcriptase inhibitors for HIV. A compound active against RSV targets the capping domain. The locations of positions in VSV-L that correspond to sites of resistance mutations in RSV-L are consistent with our proposal that a domain closure accompanies activation of the capping activity after polymerization has commenced (Liuzzi et al., 2005).

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

We expressed N-terminally 6xHis-tagged L protein from the Indiana strain of VSV in Sf21 insect cells and purified the protein as described previously (Li et al., 2008; Rahmeh et al., 2010). We expressed VSV-P(35–106) in *E. coli*, purified the fragment, and combined it with VSV-L, as described in the Supplemental Experimental Procedures.

Electron Microscopy

Images of VSV-L:P(35–106) were recorded with a Tecnai F20 electron microscope (FEI) operated at 200 kV, using UCSF Image4 (courtesy Yuemin Li, UCSF) to collect movies on a K2 Summit direct detector (Gatan), operated in superresolution mode with dose fractionation. For each 6-s exposure, we collected 30 frames of 200 ms each, with a total electron dose of 31 e/Å². Full details are in the Supplemental Experimental Procedures.

Image Processing

From 1,272 movies, we picked 356,611 particles, carried out two-dimensional classification with multivariate statistical analysis (MSA) in IMAGIC (van Heel et al., 1996) and K-means classification in TIGRIS (<http://tigris.sourceforge.net>), determined defocus with CTFFIND3 (Mindell and Grigorieff, 2003), calculated class averages with full contrast transfer function (CTF) correction, and selected good class sums as references for particle alignment and subsequent re-classification, iterating twice. An initial model, calculated with EMAN2 (e2initialmodel.py) (Tang et al., 2007) used 292 class averages. Refinement and three-dimensional classification (three classes) in FREALIGN (Lyumkis

et al., 2013) of the 292 CTF-corrected class averages yielded one class with ~10 Å resolution, which we used as an initial reference for refinement and classification of the full particle stack.

We used FREALIGN to refine and 3D classify (three classes) the full-dose particle stack, initially using 3× binned images (high resolution limit 10 Å) for 11 cycles of refinement and classification, followed by 160 cycles with unbinned images. At this point, the resolution was ~7 Å. We extracted the best set of 155,443 particles and carried out 100 more cycles of refinement and classification (three classes), extending the resolution to 6 Å. We then extracted 74,940 particles with the best scores from the two best classes. A final seven cycles of refinement of angles and shifts (using a 6 Å reference model) alternated between the full-dose (31 e/Å²) stack and a low-dose (12 e/Å²) stack. The final map (3.8 Å resolution at FSC = 0.143 criterion) was calculated from the low-dose images and a B-factor of -500 Å² was applied.

Model Building

We traced the polypeptide chain of the RdRp using the programs O (Jones et al., 1991) and Coot (Emsley et al., 2010), with the reovirus and rotavirus polymerases (λ3 and VP1, respectively) as connectivity guides, and traced the methyltransferase domain, with the consensus fold of S-adenosylmethionine (SAM)-dependent transferases as guide. For the capping domain, connector domain, and C-terminal domain, with no known homologs, we built initial models to fit the density, confirming and adjusting connectivity by reference to the amino-acid sequence. Side-chain density was strong enough in secondary structure elements of each domain to establish sequence register. We checked and corrected the entire structure with O. For the following segments of the capping and connector domains, the density did not allow confident assignment of backbone stereochemistry, and Cα positions will have larger errors than in the rest of the model: 1159–1171; 1210–1226; 1308–1334; 1387–1395; 1512–1516; 1534–1541 (see Figure S6 and the Supplemental Experimental Procedures for further details).

Structure Refinement

We refined the structure to improve the fit and to optimize stereochemistry, by calculating structure factors from the final map (Figure S1) and using them as input to standard crystallographic refinement procedures in PHENIX (Adams et al., 2010). We flagged 4% of the structure factors for cross validation and estimated figures of merit from the phase angle difference between the two half-set reconstructions used to estimate FSC. We carried out several rounds of individually restrained positional and B-factor refinement, including one round of torsion-angle simulated annealing and real-space refinement. We applied secondary structure restraints throughout the refinement and Ramachandran restraints in the final round. We analyzed the final model with MolProbity (Chen et al., 2010). Refinement and model statistics are in Table S1, and a complete account of the refinement protocol is in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession numbers for the cryo-EM map and the refined atomic coordinates reported in this paper are Electron Microscopy Data Bank (EMDB): EMD-6337, and Protein Data Bank (PDB): 5a22, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, two tables, and one dataset and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.06.018>.

AUTHOR CONTRIBUTIONS

B.L. and Z.L. designed experiments and acquired and analyzed data. S.J. developed strategies for computational data analysis and analyzed data. A.R. and B.M. helped design experiments and interpret data. N.G. and T.G. contributed software developments and image analysis. S.C.H. and S.P.J.W. conceived and co-directed the project and contributed to the

experimental design. B.L., Z.L., S.J., N.G., S.C.H., and S.P.J.W. all participated in data interpretation and in writing and revising the manuscript.

ACKNOWLEDGMENTS

We thank Robin Ross for protein expression and purification and Thomas Walz for advice and encouragement. The research was supported by NIH grants GM62580 (to S.C.H.), AI059371 (to S.P.J.W.), and AI057159 (New England Regional Center of Excellence for Biodefense and Emerging Infectious Diseases). N.G. and S.C.H. are Investigators in the Howard Hughes Medical Institute.

Received: April 6, 2015

Revised: May 11, 2015

Accepted: May 22, 2015

Published: July 2, 2015

REFERENCES

- Abraham, G., and Banerjee, A.K. (1976). Sequential transcription of the genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* **73**, 1504–1508.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W., and Ruigrok, R.W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360–363.
- Ball, L.A., and White, C.N. (1976). Order of transcription of genes of vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* **73**, 442–446.
- Baltimore, D., Huang, A.S., and Stampfer, M. (1970). Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion. *Proc. Natl. Acad. Sci. USA* **66**, 572–576.
- Barr, J.N., and Wertz, G.W. (2001). Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J. Virol.* **75**, 6901–6913.
- Barr, J.N., Whelan, S.P., and Wertz, G.W. (1997). cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J. Virol.* **71**, 8718–8725.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21.
- Ding, H., Green, T.J., Lu, S., and Luo, M. (2006). Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **80**, 2808–2814.
- Egloff, M.P., Benarroch, D., Selisko, B., Romette, J.L., and Canard, B. (2002). An RNA cap (nucleoside-2'-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21**, 2757–2768.
- Emerson, S.U., and Wagner, R.R. (1973). L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus. *J. Virol.* **12**, 1325–1335.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
- Galloway, S.E., and Wertz, G.W. (2008). S-adenosyl homocysteine-induced hyperpolyadenylation of vesicular stomatitis virus mRNA requires the methyltransferase activity of L protein. *J. Virol.* **82**, 12280–12290.
- Green, T.J., and Luo, M. (2009). Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl. Acad. Sci. USA* **106**, 11713–11718.
- Green, T.J., Zhang, X., Wertz, G.W., and Luo, M. (2006). Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357–360.
- Grigorieff, N., and Harrison, S.C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr. Opin. Struct. Biol.* **21**, 265–273.
- Hercyk, N., Horikami, S.M., and Moyer, S.A. (1988). The vesicular stomatitis virus L protein possesses the mRNA methyltransferase activities. *Virology* **163**, 222–225.
- Hunt, D.M., Smith, E.F., and Buckley, D.W. (1984). Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein. *J. Virol.* **52**, 515–521.
- Iverson, L.E., and Rose, J.K. (1981). Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell* **23**, 477–484.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
- Krumm, S.A., Yan, D., Hovingh, E.S., Evers, T.J., Enkirch, T., Reddy, G.P., Sun, A., Saindane, M.T., Arrendale, R.F., Painter, G., et al. (2014). An orally available, small-molecule polymerase inhibitor shows efficacy against a lethal morbillivirus infection in a large animal model. *Sci. Transl. Med.* **6**, 232ra252.
- La Ferla, F.M., and Peluso, R.W. (1989). The 1:1 N-NS protein complex of vesicular stomatitis virus is essential for efficient genome replication. *J. Virol.* **63**, 3852–3857.
- Li, J., Rahmeh, A., Morelli, M., and Whelan, S.P. (2008). A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J. Virol.* **82**, 775–784.
- Li, J., Rahmeh, A., Brusic, V., and Whelan, S.P. (2009). Opposing effects of inhibiting cap addition and cap methylation on polyadenylation during vesicular stomatitis virus mRNA synthesis. *J. Virol.* **83**, 1930–1940.
- Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y., Lagacé, L., et al. (2005). Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylation by viral RNA-dependent RNA polymerase. *J. Virol.* **79**, 13105–13115.
- Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T., and Harrison, S.C. (2008). Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure* **16**, 1678–1688.
- Lyumkis, D., Brilot, A.F., Theobald, D.L., and Grigorieff, N. (2013). Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388.
- Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347.
- Morin, B., and Whelan, S.P. (2014). Sensitivity of the polymerase of vesicular stomatitis virus to 2' substitutions in the template and nucleotide triphosphate during initiation and elongation. *J. Biol. Chem.* **289**, 9961–9969.
- Morin, B., Rahmeh, A.A., and Whelan, S.P. (2012). Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J.* **31**, 1320–1329.
- Ogino, T., and Banerjee, A.K. (2007). Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol. Cell* **25**, 85–97.
- Patton, J.T., Davis, N.L., and Wertz, G.W. (1984). N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J. Virol.* **49**, 303–309.
- Peluso, R.W., and Moyer, S.A. (1983). Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system. *Proc. Natl. Acad. Sci. USA* **80**, 3198–3202.
- Pflug, A., Guilligay, D., Reich, S., and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360.
- Rahmeh, A.A., Li, J., Kranzusch, P.J., and Whelan, S.P. (2009). Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. *J. Virol.* **83**, 11043–11050.

- Rahmeh, A.A., Schenk, A.D., Danek, E.I., Kranzusch, P.J., Liang, B., Walz, T., and Whelan, S.P. (2010). Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proc. Natl. Acad. Sci. USA* *107*, 20075–20080.
- Rahmeh, A.A., Morin, B., Schenk, A.D., Liang, B., Heinrich, B.S., Brusic, V., Walz, T., and Whelan, S.P. (2012). Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proc. Natl. Acad. Sci. USA* *109*, 14628–14633.
- Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T.S., Zhou, Y., Li, H., and Shi, P.Y. (2006). West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. *J. Virol.* *80*, 8362–8370.
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunnardi, T., Nanao, M., Ruijgrok, R.W., and Cusack, S. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* *516*, 361–366.
- Rose, J.K., Lodish, H.F., and Brock, M.L. (1977). Giant heterogeneous polyadenylic acid on vesicular stomatitis virus mRNA synthesized in vitro in the presence of S-adenosylhomocysteine. *J. Virol.* *21*, 683–693.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* *333*, 721–745.
- Stillman, E.A., and Whitt, M.A. (1997). Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. *J. Virol.* *71*, 2127–2137.
- Stillman, E.A., and Whitt, M.A. (1999). Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. *J. Virol.* *73*, 7199–7209.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* *157*, 38–46.
- Tao, Y., Farsetta, D.L., Nibert, M.L., and Harrison, S.C. (2002). RNA synthesis in a cage—structural studies of reovirus polymerase lambda3. *Cell* *111*, 733–745.
- Tekes, G., Rahmeh, A.A., and Whelan, S.P. (2011). A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing. *PLoS Pathog.* *7*, e1002073.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol.* *116*, 17–24.
- Wang, J.T., McElvain, L.E., and Whelan, S.P. (2007). Vesicular stomatitis virus mRNA capping machinery requires specific cis-acting signals in the RNA. *J. Virol.* *81*, 11499–11506.
- Whelan, S.P., and Wertz, G.W. (2002). Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome. *Proc. Natl. Acad. Sci. USA* *99*, 9178–9183.
- Zhou, Y., Ray, D., Zhao, Y., Dong, H., Ren, S., Li, Z., Guo, Y., Bernard, K.A., Shi, P.Y., and Li, H. (2007). Structure and function of flavivirus NS5 methyltransferase. *J. Virol.* *81*, 3891–3903.

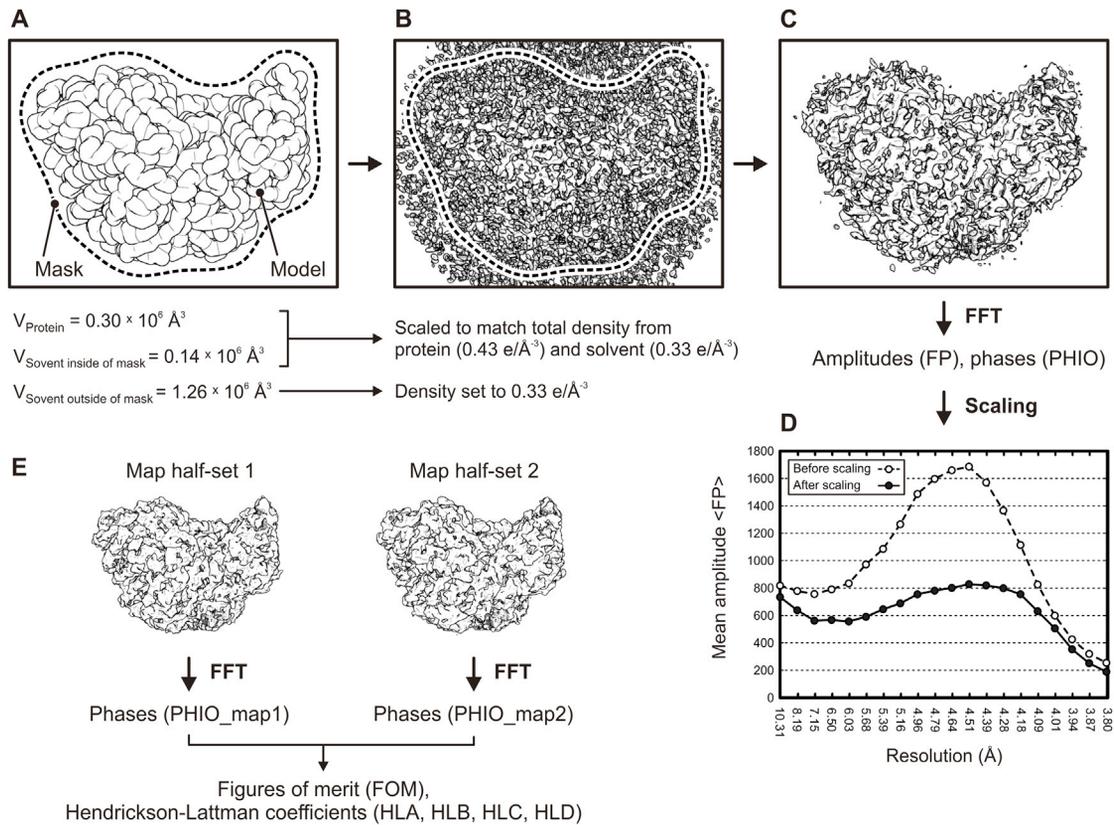


Figure S1. Preparation of Structure Factors for Refinement, Related to Figure 1

(A–D) Preparation of Fourier coefficients from the experimental reconstruction.

(A) A mask is generated around the model.

(B) Density outside the mask is flattened, and density inside the mask is put on absolute scale.

(C) Amplitudes and phases are calculated from the flattened map by Fourier transformation (FFT).

(D) Scaling of amplitudes before refinement.

(E) Estimation of figures of merits from phase-angle differences between the two half-set reconstructions.

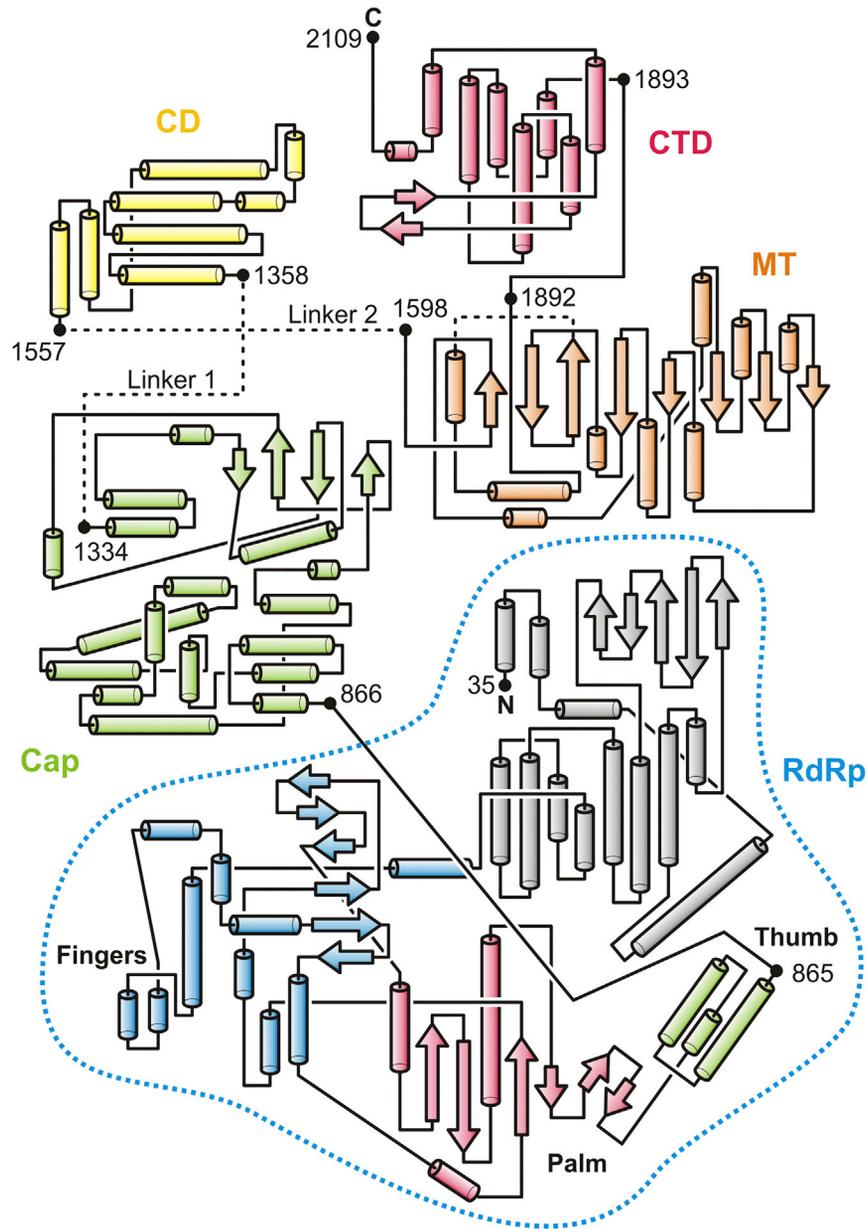


Figure S2. Secondary Structures, Related to Figure 2

Secondary structure diagram of VSV-L. Secondary structure elements along the VSV-L sequence are shown as cylinders and arrows for α helices and β strands, respectively. Domains are colored as in Figure 2, with the exception of the polymerase domain (RdRp), which is colored as in Figure 3, with the palm domain in red, the fingers in blue and the thumb in green. Domain boundaries are indicated by the corresponding residue numbers.

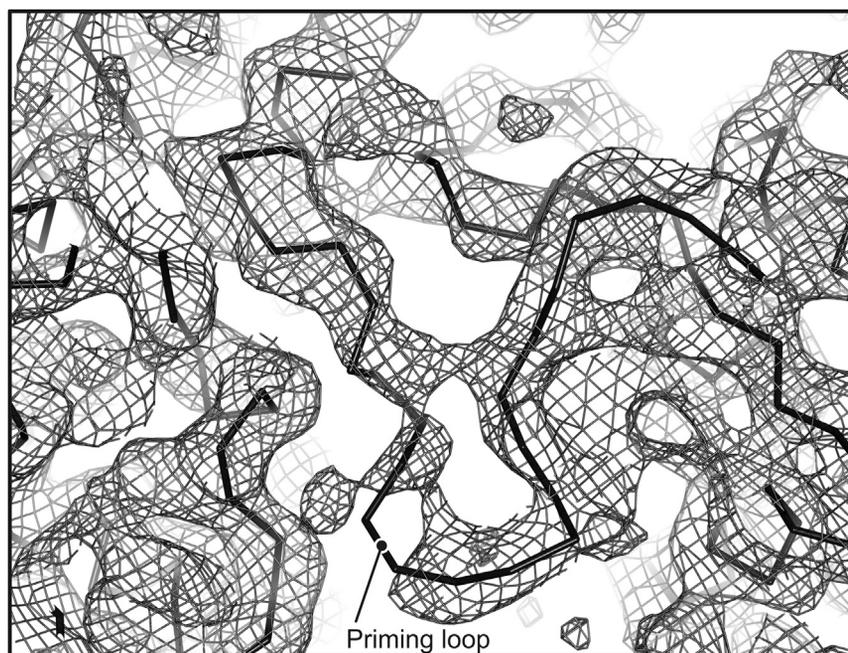
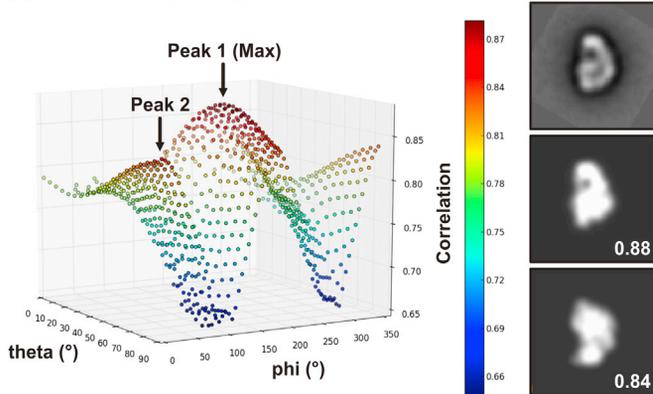


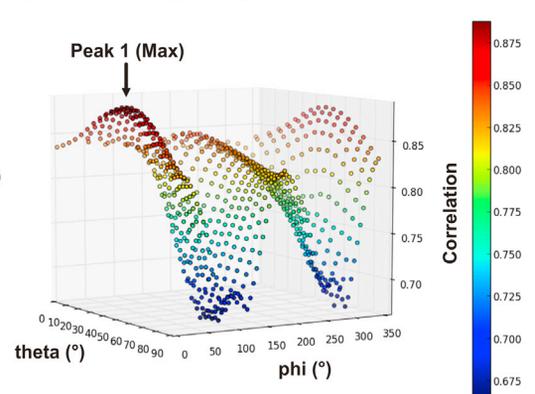
Figure S3. Density around the Priming Loop, Related to Figure 4

Contours from a 5 Å resolution density map outline the flexible priming loop that projects from the capping domain. Heavy black lines show the C α trace for the loop and adjacent polypeptide chain.

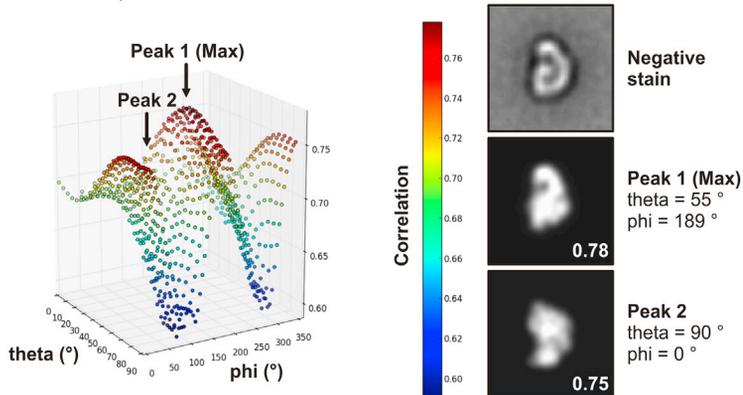
A VSV-L : P(1-106), class 1



B VSV-L : P(1-106), class 3



C VSV-L : P, class 1



D VSV-L : P, class 2

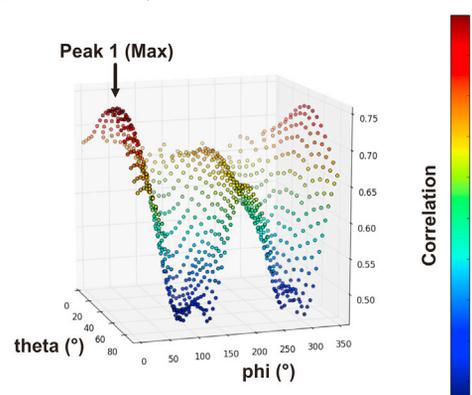


Figure S4. Correlation of Images of Negatively Stained VSV-L and Fragments with Projected Views of Model-Based Density, Related to Figure 5

For each panel, an image from negative-stain electron microscopy (Rahmeh et al., 2012) was correlated with projections of a density map calculated from the molecular model of VSV-L, using routines in SPIDER (Shaikh et al., 2008). The images for (A)–(D) are those in the first, third, fifth and sixth panels, respectively, in the top row of Figure 5A. For (A) and (C), which have a subsidiary maximum in the correlation plot, we show the projected views for both peaks. Angular coordinates as defined in SPIDER.

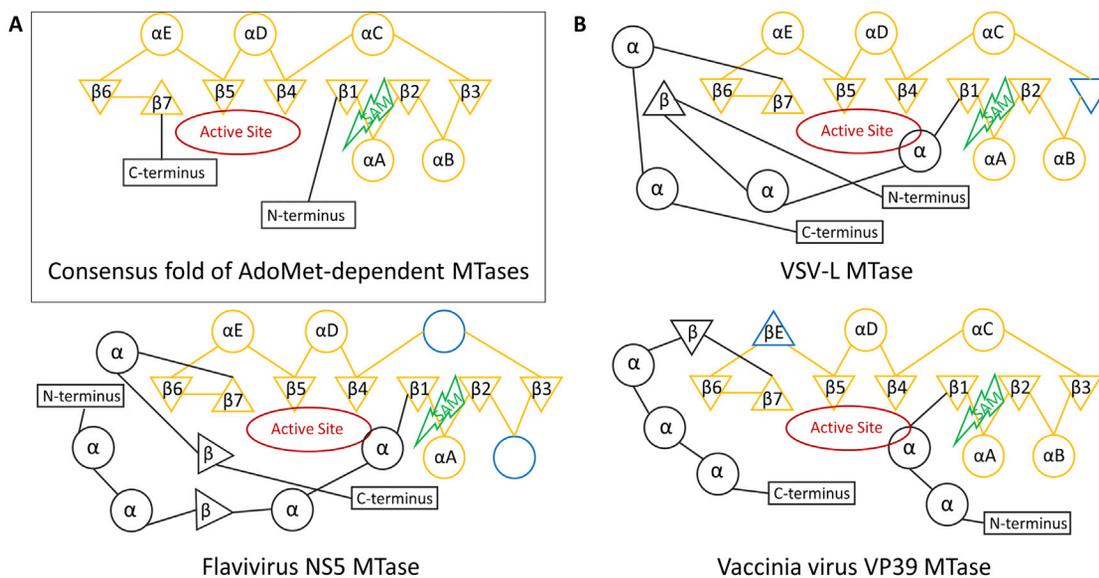


Figure S5. Secondary Structure Diagram of the Methyltransferase Domain, Related to Figure 6

(A) Diagram of the consensus fold for AdoMet-dependent methyl transferases; α helices, β strands, and termini are represented by circles, triangles and rectangles, respectively. The consensus AdoMet-dependent methyl transferase fold is in orange, other regions in gray (as in Figure 6). The positions of SAM/SAH and the active site are in green and red, respectively.

(B) Diagram of VSV-L MTase, vaccinia virus VP39 MTase, and flavivirus NS5 MTase, in the same color scheme as (A). Insertions or deletions of α helices or β strands are in blue.

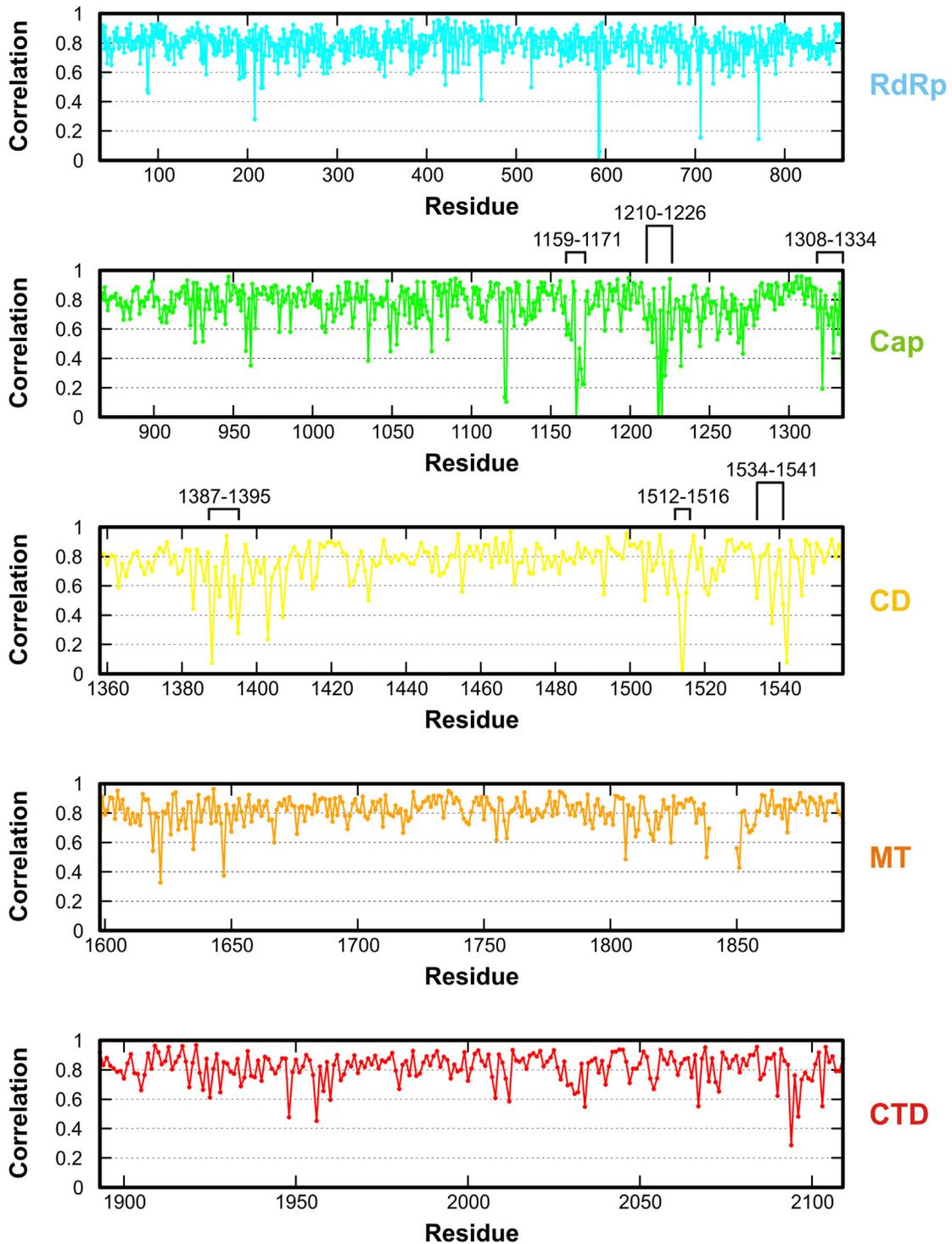


Figure S6. Correlation of Model and Density, Related to Experimental Procedures

Residue-by-residue correlation, with each panel corresponding to a domain, color coded as in Figure 2. Residue numbers above the plots for the capping and connector domains indicate segments of poor density. The gap in the methyltransferase plot corresponds to a short loop omitted from the model. Calculated with CNS (Brunger, 2007).

Cell

Supplemental Information

Structure of the L-Protein of Vesicular Stomatitis

Virus from Electron Cryomicroscopy

**Bo Liang, Zongli Li, Simon Jenni, Amal A. Rahmeh, Benjamin M. Morin, Tim Grant,
Nikolaus Grigorieff, Stephen C. Harrison, and Sean P.J. Whelan**

Supplemental Experimental Procedures

Protein expression and purification

For insect-cell expression of VSV-L, we used a baculovirus vector created with pFastBac Dual (Invitrogen). We placed L under control of the polyhedrin promoter and green fluorescent protein (GFP) under control of the P10 promoter (to visualize expression, which correlated well with expression of L). Sf21 cells were infected, incubated at 27 °C for 60-72 hours, and harvested as cell pellets by centrifugation followed by a phosphate buffered saline (PBS) wash. Following lysis by sonication and removal of cell debris by centrifugation, the L protein was purified by Ni-nitrilotriacetic acid (NTA) chromatography followed by Hi-Trap S and size-exclusion chromatography, as described (Li et al., 2008; Rahmeh et al., 2010). VSV P, residues 35-106 with an N-terminal 6xHis-tag followed by a tobacco-etch virus (TEV) protease recognition motif, was expressed in Rosetta BL21 (DE3) *E. coli* cells grown in LB medium containing 100 µg/mL ampicillin. We induced protein expression with 0.8 mM IPTG at optical density 0.8, and incubated overnight at 18 °C. The P(35-106) fragment was first purified by Ni-NTA agarose chromatography followed by tag removal by incubating with TEV protease overnight at 4 °C. A second round of Ni-NTA chromatography separated cleaved P(35-106) from uncleaved product. The cleaved proteins were dialyzed against 25 mM Tris pH7.4, 250 mM NaCl, 1 mM DTT. Purified VSV-L and P(35-106) were incubated overnight at 4 °C in a molar ratio of 1:4, and the complex was isolated on a Superdex 200 gel filtration column in 25 mM HEPES pH 7.4, 250 mM NaCl, 6 mM MgSO₄, 0.5 mM TCEP.

Electron microscopy

We screened the purified VSV-L:P complex for homogeneity by examining negatively-stained samples on a Philips CM10 electron microscope (EM). For cryo preparation, we applied 3.5 µL of protein at ~0.35 mg/mL to a Quantifoil R1.2/1.3 Cu grid (400 mesh) (Quantifoil, Germany) that had been glow discharged at 40 mA for 30 s. Grids were plunge-frozen with an FEI Vitrobot Mark I, with the following settings: 65 % humidity, offset -3, blot time 2 s, drain time 1 s. Images were recorded with liquid-nitrogen cooling on a Tecnai F20 EM (FEI) with a CT3500 cryo-specimen holder (Gatan); the microscope was operated at 200 kV; the defocus range was 0.9-2.3 µm. We used a semi-automated acquisition program, UCSFImage4 (courtesy Yueming Li, UCSF) to record movies with a K2 Summit direct detector (Gatan), operated in super-resolution mode with dose fractionation. The nominal magnification was 29,000x, corresponding to a calibrated magnification of 40,410x on the sensor plane of the camera. The beam intensity was set to 8 e/pixel/s. During a 6 s exposure, we collected 30 frames of 200 ms each for a total electron dose of 31 e/Å². Frames were binned over 2x2 pixels, yielding a pixel size of 1.24 Å, and aligned to each other using the program dosefgpu_driftcorr (Li et al., 2013).

Image processing

From 1272 movies we picked a total of 356,611 particles by hand from 6x binned images. We carried out two-dimensional classification with multivariate statistical analysis (MSA) in IMAGIC (van Heel et al., 1996) and K-means classification in TIGRIS (<http://tigris.sourceforge.net>). Image defocus was determined with CTFFIND3 (Mindell and Grigorieff, 2003), and class averages were calculated with full contrast transfer function (CTF) correction. We selected good class sums as references for particle alignment and subsequent re-classification, iterating twice. An initial model, calculated with EMAN2 (e2initialmodel.py (Tang et al., 2007)) used 292 class averages. Refinement and three-dimensional classification (3 classes) in FREALIGN

(Lyumkis et al., 2013) of the 292 CTF-corrected class averages resulted in one class with about 10 Å resolution, which we used as an initial reference for refinement and classification of the full particle stack.

FREALIGN was also used for refinement and three-dimensional classification (3 classes) of the full-dose particle stack, initially using 3x binned images (high resolution limit 10 Å). After 11 cycles of refinement and classification, the computation switched to unbinned particles. After 160 cycles, the resolution had gradually extended to 7 Å, at which point we extracted the best set of 155,443 particles. A further 100 cycles of refinement and classification (3 classes) extended the resolution to 6 Å. We then extracted 74,940 particles with the best scores from the two best classes. A final 7 cycles of refinement of angles and shifts (using a 6 Å reference model) alternated between the full-dose (31 e/Å²) stack and a low-dose (12 e/Å²) stack. The final map (3.8 Å resolution at FSC=0.143 criterion) was calculated from the low-dose images.

Model building

We traced the polypeptide chain of the RdRp, ring-like domain using the programs O (Jones et al., 1991) and Coot (Emsley et al., 2010). We used Coot to place standard poly-alanine α -helices into evident helical density features, which were usually well enough defined to determine polarity, and connected the helices with poly-alanine loops, following strong density. We used the reovirus and rotavirus polymerases (λ 3 and VP1, respectively) as connectivity guides, having established correspondence of helical segments over a span of about 600 total residues. We built the methyltransferase domain following a similar strategy, guided by the consensus fold of S-adenosylmethionine (SAM)-dependent transferases. The capping domain, connector domain, and C-terminal domain have no known homologs; we relied on the density to build initial models, confirming and adjusting connectivity subsequently by reference to the amino-acid sequence. Side-chain density was strong enough in secondary structure elements of each domain to establish the sequence register. Secondary structure prediction (www.predictprotein.org) helped locate principal helices and strands. We checked and corrected the entire structure with O, using the lego-loop provision to rebuild many of inter-secondary-structure loops and adjusting side-chain torsion angles to fit density. For the following segments of the capping and connector domains, the density did not allow confident assignment of backbone stereochemistry, and C α positions will have larger errors than in the rest of the model: 1159-1171; 1210-1226; 1308-1334; 1512-1518; 1534-1541 (see Fig. S6).

Structure refinement

We fine-sampled the density map on a grid with 0.72 Å spacing and transferred the density (and model) into a P1 cell (a=112 Å, b=143 Å, c=106 Å, with 90° angles) using MAPROT (Stein et al., 1994) from the CCP4 suite (Winn et al., 2011). We solvent flattened this map (Fig. S1) by calculating a mask around the model with a probe radius of 3.9 Å and setting grid points outside the masked region to a constant value corresponding to 0.33 e/Å³. Density within the mask was set to an absolute scale by determining a scale factor assuming 0.33 e/Å³ and 0.43 e/Å³, respectively for solvent and protein within the mask (determining the "dry" protein volume from its mass and partial specific volume; the "hydration" calculated this way is about 0.3 w/w). Map and mask operations were carried out with MAPMAN (Kleywegt and Jones, 1996).

We calculated amplitudes (FP) and phases (PHIO) from the solvent-flattened map. Although we did not use amplitude standard deviations (SIGFP) in any of our calculations, we supplied dummy values (SIGFP = 0.1 FP) to satisfy input requirements of certain programs. We flagged 4 % of the structure factors as a cross validation set for calculating R_{free}. We estimated figures

of merit (FOM) from the phase angle difference between the two half-set reconstructions and calculated Hendrickson-Lattman coefficients from PHIO and FOM.

We refined the structure against amplitudes and phases, including data to a minimum Bragg spacing of 3.8 Å. We applied procedures in PHENIX (Adams et al., 2010), using a protocol with several rounds of individually restrained positional and B-factor refinement, including one round of torsion-angle simulated annealing and real-space refinement. We determined appropriate weights for the experimental terms in the target function by monitoring R_{free} , model geometry, and B-factor statistics. We applied secondary structure restraints throughout the refinement and Ramachandran restraints in the final round. We analyzed the final model with MolProbity (Chen et al., 2010). Refinement and model statistics are in Table S1.

Figure preparation

Figures were prepared with PyMol (Schrödinger, LLC) and POV-Ray (www.povray.org). Sequences were aligned with MAFFT (Kato and Standley, 2013) and displayed with ESPript (Robert and Gouet, 2014).

Table S1. Refinement and Model Statistics, Related to Figure 1.

Refinement	
Space group	P1
Cell dimensions \square	
a, b, c (Å)	112.0, 143.0, 106.0
α , β , γ (°)	90.0, 90.0, 90.0
Asymmetric unit composition	
Number of residues	2004
Non-hydrogen atoms	16077
Resolution (Å)	143.1 – 3.80 (3.87 – 3.80)
Number of reflections (work / free)	62328 (2490) / 3570 (140)
$R_{\text{work}} / R_{\text{free}}$ (%)	26.2 / 29.6 (85.6 / 78.3)
$CC_{\text{work}} / CC_{\text{free}}$ ^b	0.88 / 0.86 (0.18 / 0.05)
Phase angle difference (°)	33.2 (73.0)
Wilson B factor (Å ²)	60.2
Model statistics	
B factors	
Average $\langle B \rangle$ (Å ²)	112.4
$\langle B_i - B_j \rangle$ (Å ²)	25.3
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.547
Ramachandran angles	84.1 % / 13.9 % / 2.0 % (favored / allowed / outliers)
MolProbity clash score ^c	13.05, 97 th percentile (N = 37, 3.00 – 999.00 Å)
MolProbity score ^c	2.79, 93 rd percentile (N = 342, 3.25 – 4.20 Å)

^a Highest resolution shell is shown in parenthesis.

^b $CC = \Sigma(F_{\text{map}} F_{\text{model}}^*) / \{\Sigma(|F_{\text{map}}|^2) \Sigma(|F_{\text{model}}|^2)\}^{1/2}$, correlation between experimental and model structure factors for working and test set, respectively.

^c 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

Table S2. Matching Statistics of VSV-L with Class Averages from Negative Stain Data ^a, Related to Figures 5 and S4.

	Cross correlation coefficient				Z score ^b	3D-plot
	Max	Mean	Min	Sigma		
VSV-L : P(41-106)						
Class 1	0.88	0.78	0.65	0.05	1.89	Figure S4A
Class 2 □	0.93	0.89	0.85	0.02	2.32	
Class 3	0.89	0.80	0.67	0.06	1.67	Figure S4B
Class 4	0.88	0.83	0.78	0.02	2.50	
VSV-L : P						
Class 1	0.78	0.70	0.59	0.04	1.81	Figure S4C
Class 2	0.75	0.61	0.47	0.07	1.88	Figure S4D
VSV-L(35-860)						
Class 1	0.87	0.80	0.73	0.03	2.59	
Class 2	0.92	0.84	0.74	0.04	1.84	
VSV-L(35-1114)						
Class 1	0.90	0.86	0.82	0.02	2.59	
Class 2	0.90	0.86	0.80	0.02	2.05	
VSV-L(35-1557)						
Class 1	0.73	0.67	0.60	0.02	2.46	
Class 2	0.73	0.64	0.57	0.04	2.37	
VSV-L(1598-2109)						
Class 1	0.83	0.72	0.58	0.07	1.42	
Class 2	0.72	0.62	0.52	0.05	1.95	

^a Cross correlation coefficients were calculated for all the class averages from negative stain data shown in Figure 5, pairing each to 799 reference images (all possible orientations sampled at a 5° angular grid) of 2D projections from the model 3D volume. Then the mean, maximum, minimum and standard deviation were calculated.

^b Z score is defined as number of standard deviations (Sigma) of Max above Mean.

Supplemental References:

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* 2, 2728-2733.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30, 772-780.

Kleywegt, G.J., and Jones, T.A. (1996). xDIMPAN and xDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta cryst. D* 52, 826-828.

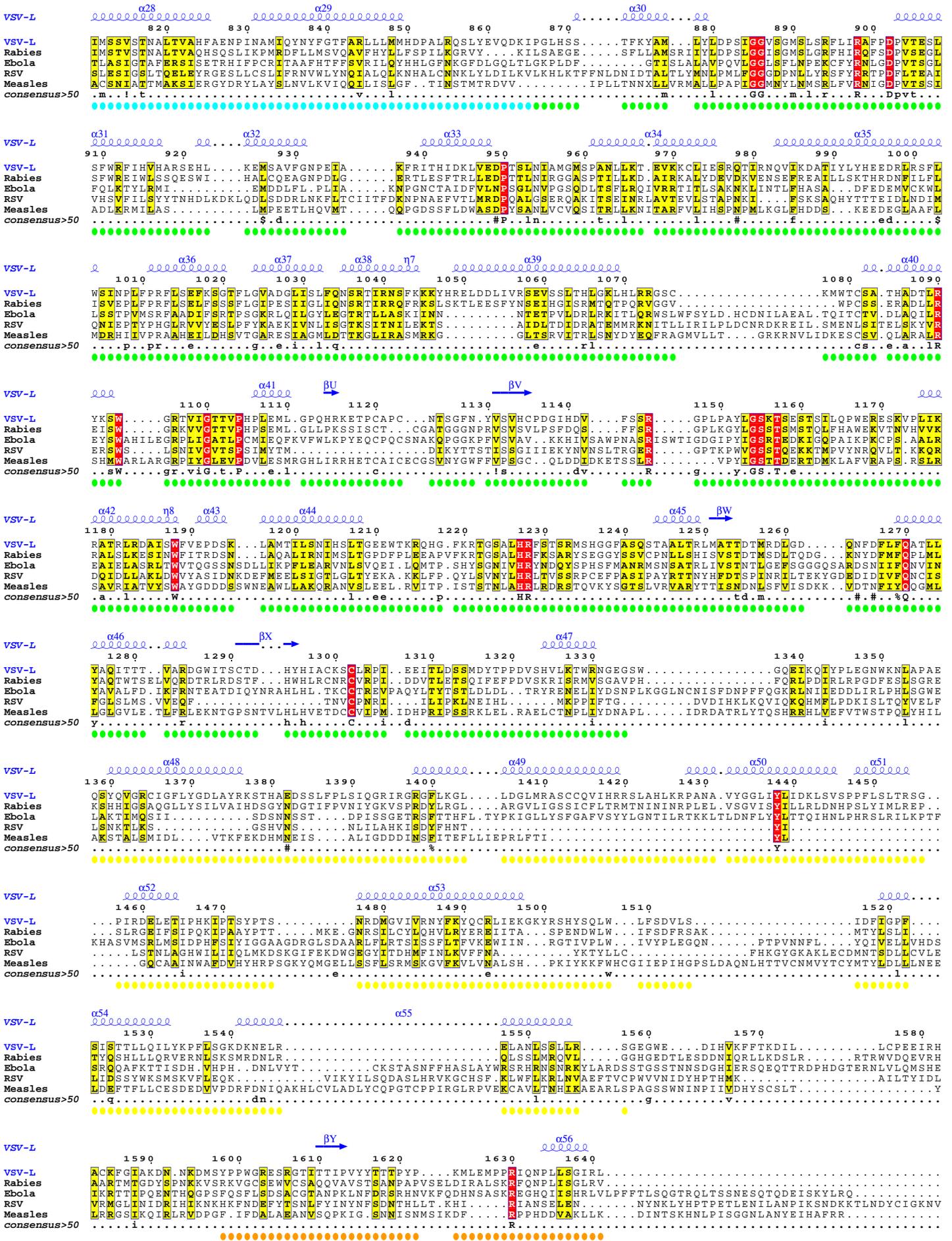
Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. methods* 10, 584-590.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids res.* 42, W320-324.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Feith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* 3, 1941-1974.

Stein, P.E., Boodhoo, A., Armstrong, G.D., Cockle, S.A., Klein, M.H., and Read, R.J. (1994). The crystal structure of pertussis toxin. *Structure* 2, 45-57.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta cryst. D* 67, 235-242.



VSV-L

VSV-LGQLP.TGAH.YK.I.R.S.L.H.G.M.G.T.H
 RabiesV.Q.N.A.T.G.A.H.Y.K.L.K.P.I.L.D.D.L.N.V.F
 EbolaR.F.T.G.I.V.S.M.H.Y.K.L.D.E.V.L.W.E.I.E.S.F
 RSVT.G.C.K.I.S.I.E.Y.L.K.D.L.K.I.K
 MeaslesI.G.L.N.S.S.A.C.Y.K.A.V.E.L.L.S.L
 consensus>50y.k.....i.l.....

VSV-L

VSV-LY.R.D.F.L.S.....C.D.D.C.S.G.C.M.T.A.L.L.R.R.E.N.V.H.S.R.G.I.F.N.S.L.L.E.L.S.G.S.V.M..R.G.A.S.P.E.P.P...S.A.L.E.T.L.G.G.D.K.S.R.C.V.N.G.E.T.C.W.E.Y.P.S.D.C.D.P.R.T.W.D.Y.F.L.R.L.K.A.G.L.G.
 RabiesP.S.L.C.L.V.....V.C.D.C.S.G.I.S.R.A.V.L.N.M.F.P.D.A.K.L.V.F.N.S.L.E.V.N.D.L.M.A..S.G.T.H.P.L.P.P...S.A.I.M.R.G.G.N.D.I.V.S.R.V.I.D.L.D.S.I.W.E.K.P.S.D.L.R.N.L.A.T.W.K.Y.F.Q.S.V.Q.K.O.V.N.
 EbolaK.S.A.V.T.L.....A.E.C.D.C.A.C.L.L.L.L..I.Q.K.Y.Q.V.K.T.L.F.N.T.L.A.T.E.S.I.E.S.E.I.V.S.G.M.T.T.P.R.M.L.L.P.V.M.S.K.F.H.N.D.Q.I.E.I.I.L.N..N.S.A.S.Q.I.T.D.I.T.N.P.T.W.F.K.D.Q.R.A.R.L.P.
 RSVD.P.N.C.I.A.....F.I.C.E.A.C.N.L.L.L.R.T.V.V.L.H.P.D.I.R.Y.I.Y.R.S.L.K.D.C.N.D...H.S.L.P.E.F..L.R.L.Y.N.G.H.I.N..I.D.Y.G.E..N.L.T.I.P.A.T.D.A.T.N.N.I.H.W.S.Y.L.H.I.K.F.A.
 MeaslesI.R.R.C.L.E.P.G.E.D.G.L.F.L.C.E.C.S.S.M.L.V..T.Y.K.E.I.L.K.L.N.K.C.F.Y.N.S.G.V.S.A.N.S.R.S.G.O..R.E.L.A.P.Y.P.S.E.V.G.L.V.E.H.R.M.G.V.G.N.I.V.K.V.L.F....N.G.R.P.E.V.T.W.V.G.S.V.D.C.F.N.F.L.V.S.N.I.P.T
 consensus>50G.#.G.m.....e.....%nsl.....P.....i.....n.....#.d.d.....f.....

VSV-L

VSV-LL.Q.D.L.I.V.M.D.M.V.R.D.S.S.T.S.L.K.I.E.T.N.V.R.N.Y.V.H.R.I.L.D.E.O.G...V.L.I.Y.K.T.Y.G.T.Y.I.C.E.S.E.K.N.A.V.T.I.L.G.P.M.F.K.T.V.D.L.V.Q.T.E.F...S.S.Q.T.S.P.V.Y.M.V.C.K.G.L.K.K.L.I.D.E.P.N.P
 RabiesM.S.Y.D.L.I.L.C.D.A.V.V.T.D.I.A.S.I.N.R.I.T.L.M.S.D.F.A.L.S.I.D.G.P.L...Y.L.V.F.K.T.Y.G.T.M.L.V.N.P.N.Y.K.A.I.O.H.L.S.R.A.F.P.S.V.T.G.F.I.T.O.V..T.S.F.S.S.L.Y.L.R.F.S.K.R.G.K.F.F.R.D.A.E.Y.L
 EbolaK.O.V.E.V.I.T.M.D.A.F.T.E.N.I.N.R.S.K.I.L.Y.A.V.Y.K.L.L.H.H.I.D.P.S.V.L.K.A.V.V.L.K.V.F.L.S.D.T.E.G.M.L.W.L.N.D.N.L.A.P.F.A.T.G.Y.L.I.K.P.I.T..S.S.A.R.S.S.P.W.Y.L.C.L.T.N.F.L.S.T.R.K....
 RSVE.P.I.S.L.F.V.C.D.A.L.S.V.T.V.N.W.S.K.I.I.L.E.W.S.K.H.V.K.C.K.Y.C.S.S.V.N.K.C.M.L.I.V.K.Y.H.A...Q.D.D.I.D.P.K.L.D.N.I.T.I.L.K.T.Y.V.C.L.G.S.K.L.G.S.P.V.L.T.I.G.F.A.N.I...P.P
 MeaslesS.S.V.G.F.H.S.D.I.L.T.P.N.K.D.T.I.K.K.L.E.L.A.A.I.L.S.M.A.L.L.L.G.K.I.G..S.I.L.V.I.K.L.M.P.F.S.G.D.F.V.Q.G.F.I.S.V.Y.G.S.H.Y.R.E.V.N.L.V.P.R.Y..S.N.F.I.S.T.S.Y.L.V.M.A.D.L.K.A.N.R.L.M...
 consensus>50i.i.i.D.E.....k.....l.l.k.....e.....f.v.....s.....s.E.Y\$.....

VSV-L

VSV-LD.W.S.S.I.N.E.S.W.K.N.L.Y.A.F.Q...S.S.E.Q.F.A.R.A.K.K.V.S.T.Y.F.T.L.T.G.I.P.S.Q.F.I.P.D.P.P.V.N.I.E.T.M.L.Q.I.F.G.V.P.T.G.V.S.H.A.A.L.K.S.S.D.R.P.A.D.L.L.T.I.S.L.F.Y.M.A.I.I.S.Y...Y.N.I.N.H.I.R.V.G
 RabiesT.S.S.T.L.R.E.M.S.L.V.L.F.N.C.S...S.P.K.S.E.M.Q.R.A.R.S.L.N.Y.Q.D.L.V.R.G.F.P.E.E.I.S.N.P.Y.N.E.M.I.I.T.L.I.D.S.D.V.E.S.F.L.V.H.K.M.V.D.D.L.E.L.O.R.G.T.L.S.K.V.A.I.I.A.I.M.I.V.F.S.N.R.V.F.N.V.S.K.P.L.T.D
 EbolaV.F.N.V.V.Q.N.A.K.L.I.L.S.R.T.K.N.F.I.M.P.K.K.A.D.K.E.S.I.D.A.N.I.K.S.L.I.P.F.L.C.Y.P.I.T.K.K.G.N.T.A.L.S.K.L.S.V.S.V.G.D.I.L.S.Y.S.I.A.G.R.N.E...Y.F.S.N.K.L.I.N
 RSVV.F.N.V.V.Q.N.A.K.L.I.L.S.R.T.K.N.F.I.M.P.K.K.A.D.K.E.S.I.D.A.N.I.K.S.L.I.P.F.L.C.Y.P.I.T.K.K.G.N.T.A.L.S.K.L.S.V.S.V.G.D.I.L.S.Y.S.I.A.G.R.N.E...Y.F.S.N.K.L.I.N
 MeaslesS.S.V.G.F.H.S.D.I.L.T.P.N.K.D.T.I.K.K.L.E.L.A.A.I.L.S.M.A.L.L.L.G.K.I.G..S.I.L.V.I.K.L.M.P.F.S.G.D.F.V.Q.G.F.I.S.V.Y.G.S.H.Y.R.E.V.N.L.V.P.R.Y..S.N.F.I.S.T.S.Y.L.V.M.A.D.L.K.A.N.R.L.M...
 consensus>50p.....e.....l.....y.n...n.....

VSV-L

VSV-LP.I.P.P.N.P.P.S.D.G.I.A.Q.N.V.G.I.A.T.T.G.I.S.F.W.L.S.L.M.E...K.D.I.P.L.Y.Q.O.C.L.A.V.I.Q.O.S.F.P.I.R.W.E.A.V.S.V.K.G.G.Y.K.Q.K.W.S.T.R.G.D.G.L.P.K.D.T.F.I.S.D.S.L.A.P.T.G.N.W.T.R.S.L.E.L.V...R.N
 RabiesP.S.F.Y.P.P.S.D.P.K.I.L.R.H.F.N.I.C.S.T.M.M.Y.L.S.T.A.L...G.D.V.P.S.F.A.R.L.H.D.L.Y.N.R.P.I.T.Y.Y.F.R.K.Q.V.I.R.G.N.V.Y.L.S.W.S.W.S.N.D.T.S.V.F.K.R.V.A.C.N.S.S.L.S.L.S.H.W.I.R.L.I.Y.K.I...V.K
 EbolaH.N.G.T.W.Q.A.E.F.K.L.P.E.L.I.S.V.C.N.R.F.Y.H.I.R.D.C.N.C.E.E.R.F.L.V.Q.T.L...Y.L.H.R.M.Q.D.S.E.V.K.L.I.E.R.L.T.G.L.L.S...L.F...P.D.G.L.Y.R.F...D
 RSVH.L.Y.M.V.E.S.T.Y.P.P.L.S.E...L.L.N.S.L.T.N.E.L.K.K.L.I.K.I.T.G.S.L...L.Y.N.P.H...N
 MeaslesH.L.L.Y.S.G.N.R.K.L.I.N.K.F.I.Q.N.L.K.S.G.Y.L.I.L.D...L.H.Q.N.T...F.V.K.N.L.S.K.S.E.K.Q.I.I...L.T.G.G.L.K.R.E.W.V.P.K.V.T.V.K.E.T.K.R.E.W.Y.K.L.L.G.Y.S.A.L.I.K.D
 consensus>50p.....h.....i.l.y.r.e.l.a.r.f.k.d.....n.q.r.s.q.q.g.m.f.h.a.y.p.v.l.v.s.s.r.q.r.e.l.l.s.r.i.t.r.k.f...w.g

VSV-L

VSV-LQ.V.R.L.N.P.F.N.E.I.L.F.N.Q.L.C.R.T.V.D.N.H.L.K.W.S.N...L.R.R.N.T.G.M.I.E.W.I.N.R.R.I.S.K.E.D.R.S.I.L...M.L.K.S.D.L.H.E.E.N.S.W...R.D
 RabiesT.R.R.L.V.G.S.I.K.D.L.S.R.F.V.E.R.H.L.H.R.Y.N.R.W.I.T...L.E.D.I.R.S.R.S.L.L.D.V.S.C.L...L.E.D.I.R.S.R.S.L.L.D.V.S.C.L...L.E.D.I.R.S.R.S.L.L.D.V.S.C.L...L.E.D.I.R.S.R.S.L.L.D.V.S.C.L...
 EbolaH.N.G.T.W.Q.A.E.F.K.L.P.E.L.I.S.V.C.N.R.F.Y.H.I.R.D.C.N.C.E.E.R.F.L.V.Q.T.L...Y.L.H.R.M.Q.D.S.E.V.K.L.I.E.R.L.T.G.L.L.S...L.F...P.D.G.L.Y.R.F...D
 RSVH.L.Y.M.V.E.S.T.Y.P.P.L.S.E...L.L.N.S.L.T.N.E.L.K.K.L.I.K.I.T.G.S.L...L.Y.N.P.H...N
 MeaslesH.L.L.Y.S.G.N.R.K.L.I.N.K.F.I.Q.N.L.K.S.G.Y.L.I.L.D...L.H.Q.N.T...F.V.K.N.L.S.K.S.E.K.Q.I.I...L.T.G.G.L.K.R.E.W.V.P.K.V.T.V.K.E.T.K.R.E.W.Y.K.L.L.G.Y.S.A.L.I.K.D
 consensus>50l.....e.....#.....d