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

Graphical Abstract

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)

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.
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,6 Nikolaus Grigorieff,6 Stephen C. Harrison,3,4 and Sean P.J. Whelan1,*

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 methylase 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 reassociate 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 signals in the viral genome both require N (Morin et al., 2012). Segments in the center and at the C-terminal end of the L-P complexes are a mixture of single and dimeric L species, in which the two L molecules associate as the process continues (Albertini et al., 2006; 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).

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 approxi- mate. 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, Rfree and Rwork 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; methyltransferase 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.

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.
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
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.
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 λ3 (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 λ3 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
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

---

**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 (Liuzzi 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.
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-ligated 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 λ3 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 pyrophosphorylase 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(55–106) 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 λ3 (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 λ3 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
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 5. Domain Reorganization
(A) Projection angle matching between class averages of negatively stained complexes of VSV-L and P protein (top row) [Rahmeh 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 (map_{observed} − map_{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.
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 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 β1 and αA. An SAH molecule (green) is derived from a superposition of its complex with dengue virus NS5 MT (PDB: 1l9k). 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: 1l9k; 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.
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 determine 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 rein-
sert 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 elonga-
tion proceeds might have two consequences. First, the active
site of the PRINTase 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 displace-
ment might also induce rearrangement of the rest of the capping
machinery. A large-scale reorganization of this kind could ac-
count 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, muta-
tions 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 capp-
ed, 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 hy-
per-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 elf4E–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. Methyl-
ation 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 tran-
siently 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 counter-
part 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 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., 2009).

**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\(\AA^{-2}\). 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 CTFIND3 (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 \(\sim 10\) 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 3x 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 \(\sim 7\) 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\(\AA^{-2}\)) stack and a low-dose (12 e\(\AA^{-2}\)) 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 Å\(^2\) 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 (I3 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 Cx positions will have larger errors than in the rest of the model: 1159–1171; 1210–1226, 1308–1334, 1387–1396; 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-5337, 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

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 Excellent for Bioderence 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


