Structure and Function of the N-Terminal Domain of the Vesicular Stomatitis Virus RNA Polymerase

Shihong Qiu,^a Minako Ogino,^b Ming Luo,^c Tomoaki Ogino,^b Todd J. Green^a

Department of Microbiology, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA^a; Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, Cleveland, Ohio, USA^b; Department of Chemistry, Georgia State University, Atlanta, Georgia, USA^c

ABSTRACT

Viruses have various mechanisms to duplicate their genomes and produce virus-specific mRNAs. Negative-strand RNA viruses encode their own polymerases to perform each of these processes. For the nonsegmented negative-strand RNA viruses, the polymerase is comprised of the large polymerase subunit (L) and the phosphoprotein (P). L proteins from members of the *Rhabdoviridae*, *Paramyxoviridae*, and *Filoviridae* share sequence and predicted secondary structure homology. Here, we present the structure of the N-terminal domain (conserved region I) of the L protein from a rhabdovirus, vesicular stomatitis virus, at 1.8-Å resolution. The strictly and strongly conserved residues in this domain cluster in a single area of the protein. Serial mutation of these residues shows that many of the amino acids are essential for viral transcription but not for mRNA capping. Three-dimensional alignments show that this domain shares structural homology with polymerases from other viral families, including segmented negative-strand RNA and double-stranded RNA (dsRNA) viruses.

IMPORTANCE

Negative-strand RNA viruses include a diverse set of viral families that infect animals and plants, causing serious illness and economic impact. The members of this group of viruses share a set of functionally conserved proteins that are essential to their replication cycle. Among this set of proteins is the viral polymerase, which performs a unique set of reactions to produce genome- and subgenome-length RNA transcripts. In this article, we study the polymerase of vesicular stomatitis virus, a member of the rhabdoviruses, which has served in the past as a model to study negative-strand RNA virus replication. We have identified a site in the N-terminal domain of the polymerase that is essential to viral transcription and that shares sequence homology with members of the paramyxoviruses and the filoviruses. Newly identified sites such as that described here could prove to be useful targets in the design of new therapeutics against negative-strand RNA viruses.

Nonsegmented negative-strand RNA viruses (NNSVs), including measles, Ebola, and rabies viruses, are causative agents of serious human and animal diseases. The genomes of these viruses encode three functionally conserved proteins (the nucleocapsid protein [N], the large polymerase subunit [L], and the phosphoprotein [P]) that are essential to the unconventional mechanisms that these viruses employ for transcription and replication. These processes require two viral complexes, the RNA-dependent RNA polymerase (RdRP) (a complex between the L protein and the P protein [1, 2]) and the genomic template. The viral genomic RNA of NNSVs does not exist as naked RNA but rather is encapsidated by the nucleocapsid protein (N). The nucleocapsid, rather than the naked RNA, serves as the active template for transcription and replication (1, 3). This requirement underlines the novel mechanisms used by the NNSV to perform transcription and replication, including special roles of the polymerase complex.

Our knowledge of the N and P proteins has been enhanced by the publication of numerous high-resolution structures over the last decade. In particular, many of these structures have centered on the study of the N and P proteins from vesicular stomatitis virus (VSV), a prototypic rhabdovirus that has served as a model system to investigate the molecular mechanisms employed by NNSVs during transcription and replication. These molecular snapshots have given insight into the structure of N and how it sequesters RNA (4–6), how N associates with P prior to encapsidation of RNA and capsid assembly (7), how P associates with the nucleocapsid during transcription (5), and the structure of P

alone (8, 9). These structures have complemented decades of molecular biology and biochemical studies.

Though the RdRPs of the NNSV are a complex between L and P, the L protein harbors all of the enzymatic functions (10), including polynucleotide synthesis and the multistep process of *de novo* 5' mRNA cap formation, a process that is distinct from that seen with eukaryotic enzymes (11, 12). The L proteins of the NNS RNA viruses share primary sequence homology (13) in six conserved regions (CRs) of the protein, namely, CRs I to VI (Fig. 1A). The functions of some of these regions have been identified through active site sequence homology and mutagenesis studies. These regions include CRIII (RdRP [14]), CRV (mRNA capping enzyme, polyribonucleotidyltransferase [PRNTase] [11, 12, 15]), and CRVI (methyltransferase [16–19]). Electron microscopy (EM) reconstructions gave the first insight into the VSV L structure, showing that its shape consists of a central ring-like structure

Received 8 September 2015 Accepted 15 October 2015

Accepted manuscript posted online 28 October 2015

Citation Qiu S, Ogino M, Luo M, Ogino T, Green TJ. 2016. Structure and function of the N-terminal domain of the vesicular stomatitis virus RNA polymerase. *J Virol* 90:715–724. doi:10.1128/JVI.02317-15.

Editor: D. S. Lyles

Address correspondence to Tomoaki Ogino, tomoaki.ogino@case.edu, or Todd J. Green, tgreen@uab.edu.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.

FIG 1 Crystal structure of the L protein N-terminal domain (L_{NTD}) from vesicular stomatitis virus. (A) The locations of the six conserved regions (I to VI) are illustrated along the horizontal line corresponding to the VSV L primary sequence. Regions of known function are indicated by labels below their respective CRs. The region corresponding to crystal structure shown in panel B is boxed and shaded in gray. (B) The L_{NTD} is shown in a ribbon illustration. α -Helices and β -strands are numbered in sequential order from the N terminus to the C terminus (labeled N and C). Coloring is shown in a rainbow representation from blue to red. (C) The ribbon representation is rotated by 180° around the y axis with respect to the orientation represented in panel B. Subdomains I and II are colored cyan and green, respectively. (D) The L_{NTD} is shown as a solvent-accessible surface, with amino acid residues colored according to electrostatic potential (contoured from -0.5 V [red] to $+0.5$ V [blue]). This figure was generated with CCP4MG (46).

connected to a large appendage (20, 21), corresponding to the polymerase and capping domains, respectively. These studies and others showed that L is inherently flexible and undergoes large conformational changes, especially in the appendage, when bound to VSV P (20–22).

Due to the enormous size of the protein (2,109 amino acids) and the observed flexibility, the intact L protein has not been crystallized. As an alternative approach to gain high-resolution details of L, we have produced multiple recombinant fragments of the protein. One such fragment corresponding to amino acid residues 37 to 379, covering most of CRI (residues 220 to 410), has been crystallized, and the crystal structure is reported here at 1.8-Å resolution. The structure shows that conserved residues in this domain predominantly cluster in a single area of the domain among members of *Mononegavirales*. Residues having high or strict homology were serially mutated, and the effects of site mutation on transcription and mRNA capping were analyzed. Our results show that many of the amino acids are essential for viral transcription but not for mRNA capping. These findings and a comparison of the NNSV L protein N-terminal domain (L_{NTD}) with those of other viral polymerases are presented here.

MATERIALS AND METHODS

Cloning and protein production. The segment of the L gene corresponding to amino acid residues 37 to 379 was amplified from a cDNA clone of the VSV L gene (Indiana strain) via PCR performed with the following primers (Integrated DNA Technologies [IDT]): 5'-GACTAGCTAGCGC TGATTACAATTTGAATTCTCCTC and 3'-CACCGCTCGAGTTACTT CATGGTTACTTGGGAATGTAATTTTTC. This gene fragment and vector pET-28b were each digested with NheI and XhoI restriction enzymes, purified, and ligated together to yield plasmid pET-L37-379. DNA sequencing in the UAB Sequencing Core Facility confirmed the presence of positive clones.

The pET-L37-379 plasmid was transformed into *Escherichia coli* strain BL21(DE3) CodonPlus competent cells (Stratagene). Transformants were grown at 37°C in LB broth (Fisher) with shaking in the presence of kanamycin and chloramphenicol overnight. Cultures were then diluted 1:100 in fresh LB with antibiotics and shaken at 37°C to an A_{600} of 0.5. Cultures were cooled to 18°C, and protein expression was induced at 18°C with IPTG (isopropyl- β -D-thiogalactopyranoside) (final concentration, 0.5 mM) for 20 h. Bacteria were harvested by centrifugation, resuspended in 50 mM Tris buffer (pH 7.9) containing 500 mM NaCl, sonicated, and centrifuged at $15,000 \times g$ for 30 min. Soluble L_{37-379} was purified by Ni affinity chromatography. The resulting protein was pooled, concentrated, and further purified by size exclusion chromatography on an S-200 column (GE) in buffer consisting of 20 mM Tris, 0.5 M ammonium bicar-

TABLE 1 X-ray diffraction and refinement statistics^a

Parameter	Result for indicated crystal type ^a	
	Native	SeMet
Wavelength (Å)	0.97932	0.97625
Resolution range (Å)	37.24–1.798 (1.863–1.798)	49.78–2.294 (2.376–2.294)
Space group	P 21 21 21	P 1 21 1
Unit cell	64.79 77.23 136.53 90 90 90	64.78 136.80 66.50 90 90.17 90
Total no. of reflections	367,180	403,838
No. of unique reflections	63,955 (6,159)	58,643 (5,350)
Multiplicity	5.8	7
Completeness (%)	99.28 (97.19)	97.27 (89.69)
Mean I/σ(I)	19.12 (3.53)	20.65 (3.04)
Mosaicity (°)	0.287	0.858
Twin fraction		0.39
Twin law		h -k -l
Wilson B-factor	21.51	30.25
R-merge	0.066 (0.416)	0.096 (0.445)
R-meas	0.073 (0.459)	0.113 (0.525)
R-pim	0.030 (0.189)	0.058 (0.276)
No. of reflections used for R-free analysis	2,000	
R-work	0.1735 (0.2047)	
R-free	0.1998 (0.2352)	
No. of nonhydrogen atoms	5,609	
Macromolecules	4,944	
Ligands	17	
Water	648	
No. of protein residues	609	
RMS value of bonds (Å)	0.007	
RMS value of angles (Å)	0.96	
Ramachandran favored (%)	99	
Ramachandran allowed (%)	1	
Ramachandran outliers (%)	0	
Clash score	5.12	
B-factor		
Avg	32.1	
Macromolecules	31.1	
Ligands	38	
Solvent	39.4	

^a Values in parentheses correspond to the highest-resolution data shell.

bonate, 50 mM arginine, 50 mM glutamate, and 0.5 M NaCl. The final purified protein was concentrated for crystallization trials. To produce selenomethionine (Se-Met)-substituted L_{37–379} protein, the pET-L37-379 plasmid was transformed into *E. coli* methionine-auxotrophic strain B834. Bacteria were grown overnight in LB broth. Bacteria were pelleted and resuspended in SelenoMethionine Medium Complete (Molecular Dimension). Cultures were grown and induced according to the protocols of the manufacturers. Purification of Se-Met protein followed the same protocols used for nonlabeled protein.

Crystallization. L_{37–379} was screened for crystallization conditions by the use of commercially available kits. Diffraction-quality crystals were obtained by the hanging drop vapor diffusion method using 100 mM Tris (pH 8.0), 1 M ammonium formate, and 4% Jeffamine T403 with 0.05% B-octylglucopyranoside. Crystals of Se-Met-substituted protein were obtained under similar conditions by the use of seeding techniques with native protein crystals. Crystals were cryocooled under crystallization conditions using a reaction mixture supplemented with 20% polyethylene glycol (PEG) 400.

Data collection and structure determination and refinement. X-ray diffraction data were collected with native and Se-Met-substituted pro-

tein crystals at the Advanced Photon Source using SER-CAT beamline 22-ID. Raw intensity data were processed with the HKL2000 package (23). Analysis of the selenium substructure solution, phasing, solvent flattening, phase extension, and an initial trace of the protein structure were performed with programs in the SHELX package (24) with the aid of the HKL2MAP interface (25). Further model building and real-space refinement were carried out with the aid of COOT (26). Crystals of the Se-Met-substituted protein belong to primitive monoclinic space group P2₁, with a twin fraction of 0.39, and contain four molecules in the asymmetric unit. A single copy of the L_{37–379} structure was used as a molecular replacement search model for phase analysis of the native diffraction data set. Native crystals belong to primitive orthorhombic space group P2₁2₁2₁ and diffracted X-rays to beyond 1.8-Å resolution. COOT was used to complete model building. Refinement of the structure was carried out with PHENIX (27). The analyzed data and crystallographic refinement statistics are listed in Table 1. All ribbon and surface illustrations of protein structures were prepared with PyMOL (28).

Sequence alignments. The amino acid sequences corresponding to the L proteins of Hendra virus (O89344), Nipah virus (Q997F0), Newcastle disease virus (P11205), simian virus 5 (SV5; now known as parain-

FIG 2 Sequence conservation in the N-terminal domain of the L proteins from member families of the *Mononegavirales*. Sequences of members from the *Paramyxoviridae* (Hendra virus [Q89344], Nipah virus [Q997F0], Newcastle disease virus [NDV; P11205], SV5 [PIV5; Q88434], Sendai virus [P06447], measles virus [P12576], and mumps virus [P30929]), *Filoviridae* (Marburg virus [P35262] and Ebola virus [Q05318]), and the *Rhabdoviridae* (VSV [Q98776]) were aligned with Clustal Omega (30). Residues that show absolute conservation (*) are shaded in red, while residues with strong conservation (:) are shaded in yellow. Single dots (.) indicate weak conservation. UniProt codes for each protein are noted in parentheses following each virus name and are noted in the figure. Secondary structure elements of the VSV L_{NTD} are shown above the VSV sequence.

fluenza virus 5 [Q88434], Sendai virus [P06447], measles virus [P12576], mumps virus [P30929], Marburg virus [P35262] and Ebola virus [Q05318], and VSV [Q98776] were downloaded from UniProt (29). The N-terminal sequences of these viruses were aligned with Clustal Omega (30).

In vitro transcription. His-tagged mutant L proteins were expressed in insect cells using a baculovirus expression system and purified as described previously (31). *In vitro* VSV transcription was carried out with recombinant L protein (0.15 µg), recombinant P protein (0.05 µg), and N-RNA complex (0.4 µg), as described previously (11, 31, 32). Internally ³²P-labeled leader RNA and deadenylated mRNAs were analyzed by electrophoresis in polyacrylamide gels containing 7 M urea (urea-PAGE; 20% and 5%, respectively) followed by autoradiography as described previously (32).

In vitro RNA capping. *In vitro* oligonucleotide RNA capping was performed with the recombinant VSV L protein (60 ng) using [α-³²P]GDP and pppAACAG oligonucleotide RNA as the substrates (11, 12, 31). After treatment of the reaction mixtures with calf intestine alkaline phosphatase, RNA products were purified and digested with nuclease P1 as described previously (31). The resulting digests containing the ³²P-labeled GpppA cap structure were analyzed by polyethyleneimine-cellulose thin-layer chromatography (TLC) followed by autoradiography (31).

Structural alignments. Structural alignments were performed using combinatorial extension (CE) (33) and the RCSB Web interface (34). The VSV L_{NTD} (residues 43 to 371) was structurally aligned with the reovirus λ3 N-terminal domain (Protein Data Bank [PDB] code 1MUK; residues 2 to 379 [35]), the influenza A virus PA C-terminal domain (PDB code 3CM8; PA_C residues 257 to 716; PB1 residues 1 to 25 [36]), and the La Crosse orthobunyavirus (LACV) polymerase PA_C-like domain (PDB code 5AMQ; residues 270 to 755 [37]). Flexible alignments were performed with FATCAT (38).

Protein structure accession number. Structure factors and final refined coordinates have been deposited in the PDB with accession code 5CHS.

RESULTS

X-ray crystal structure of the L protein N-terminal domain. Many N- and C-terminal truncations of the L N-terminal region were cloned, expressed, purified, and subjected to crystallization trials. In this study, the L protein N-terminal region covering residues 37 to 379 was crystallized, and the structure was solved by the single isomorphous replacement using the anomalous diffraction method and selenium as an anomalous scatterer at a resolu-

FIG 3 Conserved amino acids found in the L_{NTD} of members of the *Rhabdoviridae*, *Paramyxoviridae*, and *Filoviridae* are localized predominantly to a single location on the surface of the domain. (A) A surface representation of the L_{NTD} is shown (light cyan), with amino acids that are strictly conserved shaded in red and those that are strongly conserved shaded in yellow. The conservation and shading follow the sequence alignment in Fig. 2. (B) The boxed region from panel A is shown in a ribbon-and-stick model. Stick models are shown for conserved residues and for residues that were mutated for analysis in mRNA capping and transcription reactions, with the sites of mutation labeled. Stick colors reflect the level of conservation found in the sequence alignment of Fig. 2. Label colors (black, completely or nearly completely abolished; magenta, moderate change; green, no significant change) reflect the effect each mutation has on the transcription levels represented in Fig. 4.

tion of 1.8 Å. The full model covers residues between amino acids serine 43 and lysine 371. The refined structure shows that there are two copies of the protein in the asymmetric unit: chain B (residues 43 to 133 and residues 148 to 371) and chain A (residues 43 to 59, residues 72 to 132, and residues 148 and 363). The two chains align with an overall root mean square (RMS) value of 0.430 Å for the 294 common residues.

The model of the VSV L N-terminal domain (L_{NTD}) (Fig. 1B and C) contains two subdomains. Subdomain I is comprised of a five-stranded antiparallel β-sheet (β1 to β5; residues 189 to 227) sandwiched between a three-α-helix bundle (helices α1 to α3; residues 47 to 81) and a lone helix, α7. Subdomain II is comprised of nine antiparallel α-helices (α4 to α6 and α8 to α13) that have a flattened rather than globular arrangement. The structure of subdomain I begins with a short extended stretch of amino acids (residues Ser43 to Ile46), folds into a series of three helices (α1 to α3 [residues Ser47 to Asn58, Ser64 to Ser68, and Asp72 to Cys81]), and then extends upward to helix α4 (Thr88 to Leu99). This helix swaps over to the second subdomain (II). A loop then leads down to helix α5 (Ser107 to Phe130). This helix extends from the main body of subdomain II and is stabilized by the non-crystallographic mate in the asymmetric unit (also through helix α5). A break in the electron density leaves residues Trp134 to Trp147 unresolved. Antiparallel to helix α5 is helix α6 (Asp149 to Asn172). Helix α7 (Glu176 to Phe186) follows and swaps back

to subdomain I. This helix is positioned along with helices α1 to α3 to bookend a five-stranded antiparallel β-sheet (β1, Lys189 to Ser193; β2, Gly196 to Val204; β3, Leu207 to Ser213; β4, Trp216 to Phe219; β5, Asp223 to Asp227). The remaining residues make up the rest of subdomain II. Helices α8 to α13 are all arranged in an antiparallel orientation. The residue boundaries of these helices are as follows: α8, Arg228 to Val249; α9, Glu258 to Gln278; α10, Asn280 to Glu303; α11, Pro312 to Ile328; α12, Arg330 to Ser341; α13, Val345 to Phe355. Subsequently to helix α13, residues 356 to 371 have an extended conformation that lines a crevasse created by helices α4 and α8 to α10 before ascending to interact with the N-terminal residues of subdomain I. The N and C termini of the domain come together, forming hydrophobic interactions between residues Leu45, Ile46, Phe362, and Ile363.

The overall shape of the domain generates two distinct surfaces, one of which is convex and the other slightly concave. The convex surface of the NTD has two regions of electronegative charge due to the presence of two clusters of positively charged residues (in subdomain I, Arg184, Arg191, Arg192, Arg201, and Arg203 and Lys187, Lys189, Lys220, and Lys221; in subdomain II, Arg269, Arg277, Arg302, Arg330, and Arg333 and Lys93, Lys273, Lys296, Lys299, Lys319, Lys327, and Lys342) (Fig. 1D). In contrast, on the concave surface there are two patches of electronegative surface generated by Asp48, Asp49, and Asp284 and by Glu290, Glu324, and Glu368. Finally, helix α5 is also highly

charged, having residues Asp117 and Asp125 and residues Glu115, Glu119, Glu121, and Glu128.

Homology among L proteins from members of the *Mononegavirales*. The L proteins among members of the *Mononegavirales* share primary sequence homology predominantly in six regions along the length of the protein. To visualize this homology in the L_{NTD}, covering CRI, the amino acid sequence of the L_{NTD} from VSV (a rhabdovirus) was aligned with the corresponding regions of seven members of the *Paramyxoviridae* and two members of the *Filoviridae* (Fig. 2). Residues that were found to have strict or strong sequence homology were then mapped to the structure of the VSV L_{NTD} (Fig. 3). The conserved residues, including 8 (Ser43, Pro44, Met233, Asp236, Glu290, Gly359, His360, and Pro361) of the 11 strictly conserved residues, predominantly cluster to a single location on one surface of the protein. Arg356 sits among these residues and shares strict homology among members of the *Paramyxoviridae* and is as strongly conserved as Lys among members of the *Filoviridae* (residue 373 and 376 in Ebola virus and Marburg virus L, respectively). Two additional residues with strict conservation, Arg241 and Asp272, are below the surface of this area and form a salt bridge in the interior core of the domain. The adjacent CRII (39) and CRIII have been shown to be essential for RdRP activity; however, the role of CRI in VSV is unknown.

Secondary structure predictions were performed for the sequences of this group of L proteins in order to determine if the homology extended beyond their primary sequences (data not shown). Analyzing by linear prediction, each of the L_{NTD} is predicted to contain two helical regions (similar to VSV α -helices α 1 to α 7 and α 8 to α 13) that flank a single β -strand region (similar to VSV β -strands β 1 to β 5). Insertions in sequences of the viruses of the *Paramyxoviridae* and *Filoviridae* lead to a slightly longer N-terminal region of the domain, whereas VSV has an extension of approximately 30 residues at the extreme N terminus compared to the other viruses. These features were also observed in the primary sequence alignment (Fig. 2). Overall, the secondary structures are similar, especially in the regions that share greater sequence homology (e.g., VSV residues 140 to 371). This region of the L_{NTD}, which includes CRI of NNSV L proteins (residues 220 to 410), more closely aligns with the three-dimensional structures of other viral polymerases as discussed below.

Generation of L mutants and evaluation of function. To identify amino acid residues in the NTD that may be critical for transcription, we performed alanine-scanning mutagenesis of this domain in the full-length VSV L protein. We mutated amino acid residues from position 38 to position 46 (DYNLNSPLI; the underlined residues are not observed in the structure) and from position 356 to position 362 (RHWGHPPF; a residue in helix α 13 is underlined). These two sets of residues are found on surface-oriented loops located in close proximity in the crystal structure (see Fig. 3) and contain (D/E)xL ζ S(P/A) Ψ (Ψ) (ζ , hydrophilic; Ψ , aliphatic) and (R/K)x Ω GHP (Ω , aromatic) motifs conserved in NNSV L proteins (see Fig. 2). Furthermore, we introduced mutations in other highly conserved amino acid residues (D236, R241, D272, and E290) and in some nonconserved charged amino acid residues (D284, K287, E303, and D346). Most of the mutants as well as the wild-type (WT) L protein were expressed in large quantities in insect cells by the use of a baculovirus expression system, whereas R241A and D272A were expressed very poorly (data not shown). This is likely due to instability of the L_{NTD}, since R241 and

FIG 4 Conserved amino acid residues in the NTD of the VSV L protein are required for RNA synthesis but not for mRNA capping. (A) The wild-type (WT) and mutant L proteins (1 μ g) were analyzed by 7.5% SDS-PAGE followed by staining with Coomassie brilliant blue. Lanes 1 and 12 show marker proteins with the indicated masses. (B and C) The L proteins were subjected to *in vitro* transcription reactions reconstituted with the P protein and N-RNA template. After removal of poly(A) tails from mRNAs, 32 P-labeled transcripts were analyzed by 20% (B) or 5% (C) urea-PAGE followed by autoradiography. Lanes 1 and 12 indicate that no L protein is present. The positions of the leader RNA (Le; 47 nt), mRNAs (G [1,700 nt], N [1,300 nt], and P/M [800 nt]) and the gel origins (ori.) are indicated on the left. The graphs show relative RNA synthesis activities of the WT (defined as 100%) and mutant L proteins. Columns and error bars represent the means and standard deviations, respectively, of the results from three independent experiments. (D) The L proteins were subjected to *in vitro* capping reactions with [α - 32 P]GDP and pppAACAG oligonucleotide RNA. Calf intestine alkaline phosphatase and nuclease P1-resistant products were analyzed by polyethyleneimine-cellulose TLC followed by autoradiography. Lanes 1 and 12 indicate that there was no L protein present. The positions of standard GpppA cap analogue, GMP, GDP, and GTP are shown on the left.

FIG 5 Alignment of crystal structures of the VSV L_{NTD} with those of other viral RNA polymerase proteins. The VSV L_{NTD} was structurally aligned with the reovirus $\lambda 3$ N-terminal domain (PDB code 1MUK [35]) (A and D), the influenza A virus PA C-terminal domain (PA_C) (PDB code 3CM8 [36]) (B and E), and the La Crosse orthobunyavirus (LACV) polymerase (PDB code 5AMQ [37]) PA_C -like domain (C and F). In all figures, the VSV L_{NTD} is shown in shades of green. Aligned secondary structure elements are shown in darker shades and others in lighter shades. The two alignments shown on the left are rotated 180° with respect to each other. The images on the right show only the closely aligned segments of each protein. Models in panels A to C were aligned as single rigid bodies. Models in panels D to F were aligned with flexible fitting.

D272, found on α -helices 8 and 9, form a salt bridge across the internal core of the α -helical bundle that makes up subdomain II. For each of the well-expressed mutants, we purified and verified their purity by SDS-PAGE (Fig. 4A).

To measure RNA synthesis activities of these mutants, we reconstituted transcription reactions with the highly purified WT or mutant L protein, the recombinant P protein, and the native N-RNA complex as described previously (11, 31, 32). The resulting ^{32}P -labeled 47-nucleotide (nt) leader RNA (Fig. 4B) and deadenylated mRNAs (Fig. 4C) were analyzed by 20% and 5% urea-PAGE, respectively, followed by autoradiography. As reported previously (31, 32), the WT L protein synthesized the 47-nt leader RNA from the 3' end of the genome (Fig. 4B, lane 2) and mRNAs from the internal genes (Fig. 4C, lane 2). In the alanine mutations of the loop preceding helix $\alpha 1$, the L41A mutation nearly completely abolished both leader RNA synthesis and mRNA synthesis (lane

6), while Y39A (lane 4), L45A (lane 10), and I46A (lane 11) significantly reduced the transcription activity. Although Y39 is conserved only in rhabdoviral L proteins, L41 and one or two aliphatic amino acid residues located 4 amino acids downstream of L41 (L45 and I46 in the VSV L protein) are strikingly conserved in the NNSV L proteins, suggesting critical roles of these aliphatic residues in transcription by NNSV L proteins. In contrast, mutations in other residues of this loop had no effect or modest effects on transcription (lanes 3, 5, 7, 8, and 9).

Furthermore, the D236A (lane 14) and E290A (lane 17) mutations remarkably diminished and abolished, respectively, the transcription activity, whereas mutations of nonconserved charged residues D284A (lane 15), K287A (lane 16), E303A (lane 18), and D346A (lane 19) in this region affected the transcription activity only modestly or not at all. The R356A mutation (lane 20) subverted the transcription activity completely, while the H357A

TABLE 2 Combinatorial extension and flexible alignment analysis of viral RNA polymerases against L_{NTD} coordinates

Analysis and PDB code	Protein name	Residues	RMS superposition	Z score	No. of residues aligned	Region aligned	VSV region aligned
Rigid body alignment with CE							
1MUK	Reovirus λ3	2–379	4.5	4.42	176	157–379	112–363
3CM8	Influenza A virus PA _C	257–716 ^a , 1–16 ^b	5.13	4.74	166	414–714 ^a , 1–16 ^b	99–350
5AMQ	LACV polymerase	270–755	4.9	4.42	189	470–754	110–357
Flexible body alignment with FATCAT							
1MUK	Reovirus λ3	2–379	3.18		206	95–379	61–363
3CM8	Influenza A virus PA _C	257–716 ^a , 1–16 ^b	3.35		210	328–710 ^a , 1–16 ^b	44–349
5AMQ	LACV polymerase	270–755	3.75		259	355–759	45–365

^a PA.

^b PB1.

(lane 21), W358A (lane 22), and P361A (lane 25) mutations in the extended loop following helix α13 significantly reduced the transcription activity as reported for H360A (lane 24) (32). In contrast, the G359A (lane 23) and F362A (lane 26) mutations showed modest effects on transcription.

Similarly to the L proteins with mutations in the RdRP domain (e.g., D714A in the RdRP active site) (32), all the transcription-defective mutants identified in this study exhibited RNA capping activities to extents similar to that seen with the WT L protein (Fig. 4D). Taken together, these results indicate that highly conserved amino acid residues (e.g., L41, D236, E290, R356, W358, H360, and P361) in the L_{NTD} are specifically required for transcription catalyzed by the RdRP domain but not for capping catalyzed by the PRNTase domain.

Homology of L_{NTD} with other NSV polymerases. Rahmeh et al. (20) previously suggested that the structures of the VSV polymerase domains are reminiscent of the cage-like structure observed for reovirus λ3, a double-stranded RNA (dsRNA) virus. To explore similarity with other viral polymerases, we first aligned the structure of the VSV L_{NTD} with the reovirus λ3 (35) and observed that the L_{NTD} aligns with the NTD of λ3. Subsequently, the influenza A PA C-terminal domain (PA_C) (36, 40) and the La Crosse orthobunyavirus (LACV) polymerase PA_C-like domain (37), both of which share homology with λ3, were each aligned with the VSV L_{NTD}. In all three cases, when the whole L_{NTD} was aligned as a rigid body, helical subdomain 2 aligned more closely with the other viral polymerases than the elements of subdomain 1 (Fig. 5A to C). In total, a minimum of 166 residues (influenza A PA_C) and a maximum of 189 residues (LACV) aligned with the VSV L_{NTD} from the rigid body analysis. However, flexible alignments allowed more overlay between both domains (Fig. 5D to F), raising the minimum number of residues aligned to 206 (reovirus λ3) and the maximum number to 259 (LACV). The latter represents 82.2% of the possible 315 residues in the VSV L_{NTD}. This analysis confirms that each of these three viral domains shares topological similarity with the VSV L_{NTD}, confirming previous suggestions (20) and extending the homology shown in reference 21. Statistical values from the alignments are shown in Table 2.

DISCUSSION

In summary, we have determined the first X-ray crystal structure of the N-terminal domain of a NNSV L protein (from VSV), showing that the structure has topological similarity to structures of polymerases from other RNA viruses (influenza A virus, LACV,

and reovirus). The L proteins among the NNSVs share sequence homology. Homology in the L_{NTD} region, CRL, localizes mainly to a single surface area of the domain. We have produced wild-type L protein and L protein harboring mutations to individual residues within this region. Each purified L protein was evaluated by *in vitro* assay for RNA synthesis activity and mRNA capping, and the results showed that residues clustered in this region were important for transcription but not mRNA capping.

Interestingly, not only are amino acid residues identified as essential for transcription located in close proximity on the same molecular surface in the crystal structure but some of the side chains of these residues (e.g., D236 [helix α8], E290 [helix α10], R356 at the conclusion of helix α13, and H360) are oriented in the same direction (see Fig. 3B), suggesting that these residues are involved in binding of an as-yet-unidentified ligand(s) (e.g., template, substrate, other polymerase domains). The counterparts of R356 and H360 in the L protein of Sendai virus, a prototypic paramyxovirus, have been shown to be required for efficient tran-

FIG 6 Proximity of the VSV L_{NTD} to other polymerase elements. The VSV L_{NTD} is shown aligned with the viral polymerase of reovirus (A), influenza virus (B), and LACV (C). In each case, the VSV cartoon is colored pale cyan, with residues determined here to be important for transcription displayed in a black stick model. The λ3 reovirus is colored according to the shades used in reference 35, while the influenza polymerase complex and the LACV polymerase are colored according to the shades used in reference 37. Specifically, thumb domains are colored green, while finger regions are shaded in blue. Viral RNA (vRNA) sequences are shown in stick models in panels B and C.

scription (41), suggesting general roles of these residues in transcription of NNSV genomes. Importantly, all the transcription-defective mutations were found to negatively affect mRNA synthesis to extents similar to that seen with leader RNA synthesis. Furthermore, these transcription-defective mutants did not synthesize increased amounts of prematurely terminated transcripts (data not shown). The RdRP complex has been suggested to enter from the 3' end of the genomic RNA to synthesize leader RNA followed by five mRNAs from the internal genes by a stop-start mechanism (42, 43). Thus, one possibility is that the mutations in the L_{NTD} abolish or diminish a very early stage (e.g., initiation complex formation) of leader RNA synthesis, thereby impairing leader RNA synthesis-dependent transcription of the internal genes into mRNAs.

The NTDs of the viral polymerase structures described above form an affixing point for the finger and thumb domains, thus completing the cylindrical enclosure of the polymerase enzyme (21, 35, 37, 44, 45). Based on alignments of the VSV L_{NTD} with the complete polymerase structures of reovirus, LACV, and influenza virus, the fingertips and thumbs are in near proximity to the site of high conservation among members of the *Mononegavirales* (Fig. 6). This site is also adjacent to the binding sites for the viral RNAs bound to the influenza virus and LACV polymerase structures. Thus, this conserved area of L could be involved in RNA binding, polymerase organization, and/or other unknown functions. Additional ligand structures and further biochemical analysis of molecular interactions may aid in understanding the role of these important residues.

ACKNOWLEDGMENTS

Portions of this research were performed at the South East Regional Collaborative Access Team (SER-CAT) facility at the Advanced Photon Source (APS), Argonne National Laboratory. Use of the APS was supported by the U.S. Department of Energy (USDOE), Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the USDOE, Office of Biological and Environmental Research, and by the National Institutes of Health (NIH), National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

T.J.G., T.O., and M.L. designed the project. T.J.G., S.Q., T.O., and M.O. performed experiments. T.J.G., M.L., and T.O. prepared the manuscript. All of us discussed the results and commented on the manuscript.

The work is supported in part by NIH grants 1R56AI101087 and 1R01AI116738 (T.J.G.) and R01AI093569 (T.O.). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

FUNDING INFORMATION

HHS | NIH | National Institute of Allergy and Infectious Diseases provided funding to Todd J. Green under grant numbers 1R56AI101087 and 1R01AI116738. HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) provided funding to Tomoaki Ogino under grant number R01AI093569.

REFERENCES

- Emerson SU, Wagner RR. 1972. Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T viruses. *J Virol* 10:297–309.

- Emerson SU, Yu Y. 1975. Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J Virol* 15:1348–1356.
- Naito S, Ishihama A. 1976. Function and structure of RNA polymerase from vesicular stomatitis virus. *J Biol Chem* 251:4307–4314.
- Green TJ, Zhang X, Wertz GW, Luo M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313:357–360. <http://dx.doi.org/10.1126/science.1126953>.
- Green TJ, 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 U S A* 106:11713–11718. <http://dx.doi.org/10.1073/pnas.0903228106>.
- Green TJ, Rowse M, Tsao J, Kang J, Ge P, Zhou ZH, Luo M. 2011. Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J Virol* 85:2714–2722. <http://dx.doi.org/10.1128/JVI.01927-10>.
- Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA, Jr, Jensen MR, Blackledge M, Ruigrok RW, Jamin M. 2011. Structure of the vesicular stomatitis virus N(0)-P complex. *PLoS Pathog* 7:e1002248. <http://dx.doi.org/10.1371/journal.ppat.1002248>.
- Ding H, Green TJ, Lu S, Luo M. 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* 80:2808–2814. <http://dx.doi.org/10.1128/JVI.80.6.2808-2814.2006>.
- Ribeiro EA, Jr, Favier A, Gerard FC, Leyrat C, Brutscher B, Blondel D, Ruigrok RW, Blackledge M, Jamin M. 2008. Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* 382:525–538. <http://dx.doi.org/10.1016/j.jmb.2008.07.028>.
- Emerson SU, Wagner RR. 1973. L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus. *J Virol* 12:1325–1335.
- Ogino T, Banerjee AK. 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol Cell* 25:85–97. <http://dx.doi.org/10.1016/j.molcel.2006.11.013>.
- Ogino T, Yadav SP, Banerjee AK. 2010. Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase. *Proc Natl Acad Sci U S A* 107:3463–3468. <http://dx.doi.org/10.1073/pnas.0913083107>.
- Poch O, Blumberg BM, Bougueleret L, Tordo N. 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol* 71(Pt 5):1153–1162. <http://dx.doi.org/10.1099/0022-1317-71-5-1153>.
- Sleat DE, Banerjee AK. 1993. Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase. *J Virol* 67:1334–1339.
- Li J, Rahmeh A, Morelli M, Whelan SP. 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. <http://dx.doi.org/10.1128/JVI.02107-07>.
- Li J, Fontaine-Rodriguez EC, Whelan SP. 2005. Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. *J Virol* 79:13373–13384. <http://dx.doi.org/10.1128/JVI.79.21.13373-13384.2005>.
- Li J, Wang JT, Whelan SP. 2006. A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc Natl Acad Sci U S A* 103:8493–8498. <http://dx.doi.org/10.1073/pnas.0509821103>.
- Ogino T, Kobayashi M, Iwama M, Mizumoto K. 2005. Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. *J Biol Chem* 280:4429–4435. <http://dx.doi.org/10.1074/jbc.M411167200>.
- Rahmeh AA, Li J, Kranszus PJ, Whelan SP. 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. <http://dx.doi.org/10.1128/JVI.01426-09>.
- Rahmeh AA, Schenk AD, Danek EI, Kranszus PJ, Liang B, Walz T, Whelan SP. 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proc Natl Acad Sci U S A* 107:20075–20080. <http://dx.doi.org/10.1073/pnas.1013559107>.
- Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SP. 2 July 2015, posting date. Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* <http://dx.doi.org/10.1016/j.cell.2015.06.018>.
- Rahmeh AA, Morin B, Schenk AD, Liang B, Heinrich BS, Brusic V, Walz T, Whelan SP. 2012. Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proc*

- Natl Acad Sci U S A 109:14628–14633. <http://dx.doi.org/10.1073/pnas.1209147109>.
23. Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode, p 307–326. In Carter C Jr, Sweet R (ed), *Macromolecular crystallography: part A*, vol 276. Academic Press, New York, NY.
 24. Sheldrick GM. 2008. A short history of SHELX. *Acta Crystallogr A* 64: 112–122. <http://dx.doi.org/10.1107/S0108767307043930>.
 25. Pape T, Schneider TR. 2004. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* 37: 843–844. <http://dx.doi.org/10.1107/S0021889804018047>.
 26. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <http://dx.doi.org/10.1107/S0907444910007493>.
 27. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213–221. <http://dx.doi.org/10.1107/S0907444909052925>.
 28. DeLano WL. 2002. The PyMOL Molecular Graphics System. <http://www.pymol.org/>.
 29. The UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res* 43:D204–D212. <http://dx.doi.org/10.1093/nar/gku989>.
 30. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <http://dx.doi.org/10.1038/msb.2011.75>.
 31. Ogino T. 2013. In vitro capping and transcription of rhabdoviruses. *Methods* 59:188–198. <http://dx.doi.org/10.1016/j.ymeth.2012.05.013>.
 32. Ogino T. 2014. Capping of vesicular stomatitis virus pre-mRNA is required for accurate selection of transcription stop-start sites and virus propagation. *Nucleic Acids Res* 42:12112–12125. <http://dx.doi.org/10.1093/nar/gku901>.
 33. Shindyalov IN, Bourne PE. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739–747. <http://dx.doi.org/10.1093/protein/11.9.739>.
 34. Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE. 2010. Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26:2983–2985. <http://dx.doi.org/10.1093/bioinformatics/btq572>.
 35. Tao Y, Farsetta DL, Nibert ML, Harrison SC. 2002. RNA synthesis in a cage—structural studies of reovirus polymerase lambda3. *Cell* 111:733–745. [http://dx.doi.org/10.1016/S0092-8674\(02\)01110-8](http://dx.doi.org/10.1016/S0092-8674(02)01110-8).
 36. He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z, Liu Y. 2008. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* 454: 1123–1126. <http://dx.doi.org/10.1038/nature07120>.
 37. Gerlach P, Malet H, Cusack S, Reguera J. 2015. Structural insights into bunyavirus replication and its regulation by the vRNA promoter. *Cell* 161:1267–1279. <http://dx.doi.org/10.1016/j.cell.2015.05.006>.
 38. Ye Y, Godzik A. 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19(Suppl 2):ii246–ii255.
 39. Smallwood S, Eason CD, Feller JA, Horikami SM, Moyer SA. 1999. Mutations in conserved domain II of the large (L) subunit of the Sendai virus RNA polymerase abolish RNA synthesis. *Virology* 262:375–383. <http://dx.doi.org/10.1006/viro.1999.9933>.
 40. Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, Nagata K, Tame JR, Park SY. 2008. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* 454:1127–1131. <http://dx.doi.org/10.1038/nature07225>.
 41. Chandrika R, Horikami SM, Smallwood S, Moyer SA. 1995. Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication. *Virology* 213:352–363. <http://dx.doi.org/10.1006/viro.1995.0008>.
 42. Emerson SU. 1982. Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome. *Cell* 31:635–642. [http://dx.doi.org/10.1016/0092-8674\(82\)90319-1](http://dx.doi.org/10.1016/0092-8674(82)90319-1).
 43. Testa D, Chanda PK, Banerjee AK. 1980. Unique mode of transcription in vitro by vesicular stomatitis virus. *Cell* 21:267–275. [http://dx.doi.org/10.1016/0092-8674\(80\)90134-8](http://dx.doi.org/10.1016/0092-8674(80)90134-8).
 44. Pflug A, Guilligay D, Reich S, Cusack S. 2014. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516:355–360. <http://dx.doi.org/10.1038/nature14008>.
 45. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crepin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516:361–366. <http://dx.doi.org/10.1038/nature14009>.
 46. McNicholas S, Potterton E, Wilson KS, Noble ME. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* 67:386–394. <http://dx.doi.org/10.1107/S0907444911007281>.