The complete pathway for co-transcriptional mRNA maturation within a large protein of a non-segmented negative-strand RNA virus

Minako Ogino1, Todd J. Green2 and Tomoaki Ogino*1, *

1Department of Medical Microbiology and Immunology, College of Medicine and Life Sciences, University of Toledo, Toledo, OH 43614, USA
2Department of Microbiology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA

*To whom correspondence should be addressed. Tel: +1 419 383 4591; Email: tomoaki.ogino@utoledo.edu

Abstract

Non-segmented negative-strand (NNS) RNA viruses, such as rabies, Nipah and Ebola, produce 5'-capped and 3'-polyadenylated mRNAs resembling higher eukaryotic mRNAs. Here, we developed a transcription-elongation-coupled pre-mRNA capping system for vesicular stomatitis virus (VSV, a prototypic NNS RNA virus). Using this system, we demonstrate that the single-polypeptide RNA-dependent RNA polymerase (RdRp) large protein (L) catalyzes all pre-mRNA modifications co-transcriptionally in the following order: (i) 5'-capping (polyribonucleotidylation of GDP) to form a GpppA cap core structure, (ii) 2'-O-methylation of GpppA into GpppAm, (iii) guanine-N7'-methylation of GpppAm into m7GpppAm (cap 1), (iv) 3'-polyadenylation to yield a poly(A) tail. The GDP polyribonucleotidylyltransferase (PRNTase) domain of L generated capped pre-mRNAs of 18 nucleotides or longer via the formation of covalent enzyme-pre-mRNA intermediates. The single methyltransferase domain of L sequentially methylated the cap structure only when pre-mRNAs of 40 nucleotides or longer were associated with elongation complexes. These results suggest that the formation of pre-mRNA closed loop structures in elongation complexes via the RdRp and PRNTase domains followed by the RdRp and MTase domains on the same polypeptide is required for the cap 1 formation during transcription. Taken together, our findings indicate that NNS RNA virus L acts as an all-in-one viral mRNA assembly machinery.

Graphical abstract

Introduction

In the nucleus of eukaryotic cells, pre-mRNA processing events, such as 5'-capping, splicing, and 3'-polyadenylation, occur co-transcriptionally by various mRNA processing factors associated with Pol II elongation complexes. The 5'-cap core structure (GpppN-') is formed on pre-mRNAs of 19 nucleotides (nt) or longer by the dual-functional mRNA capping enzyme with the RNA 5'-triphosphatase (RTPase) and guanylyltransferase (GTase) activities (1–6). Once pre-mRNAs reach approximately 30 nt long, the cap core is methylated using S-adenosyl-l-methionine (SAM) as a methyl donor by mRNA (guanine-N7')-methyltransferase (G-N7'-MTase) to form the cap 0 structure (m7GpppN) (2,3,7). The cap 0 structure is required for various aspects of mRNA metabolism, such as stability, splicing, polyadenylation, export, and translation (8). In higher eukaryotes, the ribose moiety of the first transcribed nucleotide within the cap is methylated by mRNA (nucleoside,2'-O-)methyltransferase (N1,2'-O-MTase) into the cap 1 structure (m7GpppNm) (6,9). N1,2'-O-methylation of the cap not only makes Pol II transcripts more stable (10), but also avoids cellular innate immune responses against these self-mRNAs (11,12).
Eukaryotic viruses have evolved divergent strategies to generate stable, translatable, and non-immunogenic mRNAs in host cells. Vaccinia virus, a cytoplasmic large DNA virus, belonging to the *Poxviridae* family is known to produce mRNAs with a cap I structure and poly(A) tail by virus-encoded enzymes (13). Similar to the eukaryotic Pol II system, vaccinia virus DNA-dependent RNA polymerase recruits viral mRNA capping enzyme (D1–D12 complex) with the RiboTase, GTase and G-N^2^-MTase activities during an early stage of transcription elongation to generate the cap core structure on pre-mRNAs of 31 nt or longer (14,15). The timing of the respective cap methylation reactions with vaccinia virus G-N^2^-MTase and N1-2^'-O-MTase (VP39) during transcription is unknown.

Non-segmented negative-strand (NNS) RNA viruses including significant human pathogens, such as rabies virus (RABV, *Rhabdoviridae*), Nipah virus (NiV, *Paramyxoviridae*), human respiratory syncytial virus (HRSV, *Pneumoviridae*), and Ebola virus (EBOV, *Filoviridae*), generate mRNAs with the cap I and poly(A) tail to exploit the translation machinery in infected host cells (16). However, how their transcription systems synthesize extensively modified mRNAs in the cytoplasm of host cells remains elusive.

Vesicular stomatitis virus (VSV), a prototypic rhabdovirus, has served as a paradigm for studying molecular mechanisms of transcription and replication in NNS RNA viruses (16). The VSV genome is arranged as follows: 3' leader (*Le*) region, genes encoding the nucleocapsid (N), phospho- (P), matrix (M), glyco- (G) and large (L) proteins, and a 5' trailer (*Tr*) region. The genome is wrapped with the N proteins to form the N-RNA template for transcription and replication (17,18). The catalytic L and co-factor P proteins form an RNA-dependent RNA polymerase (RdRp) complex that conducts RNA synthesis, mRNA capping, cap methylation, and polyadenylation (16).

At the beginning of transcription, the VSV RdRp initiates primer-independent (de novo) synthesis of the uncapped leader RNA (LeRNA) of 47 nt at the 3'-end of the genome (19–23). After LeRNA synthesis, the RdRp transcribes the internal genes into five mRNAs in a sequential and polar manner by the stop-start mechanism (24–26). The gene-start sequence of each gene serves as a transcription initiation signal to produce pre-mRNAs with the 5'-triphasphorylated AACAG mRNA-start sequence, which in turn acts as a capping signal (27,28). Following cap addition, the 5'-cap is sequentially methylated at the adenosine-1'-2'-O (A1-2'-O) position followed by the G-N^2^-position to form the cap I (29). Through a polymerase stuttering mechanism on the poly(U)-stretch in the gene-end sequence, the RdRp generates a 3'-poly(A) tail on the mRNA (30,31).

The mechanism of mRNA capping with rhabdoviral L proteins is distinct from that with the eukaryotic mRNA capping enzyme (28,32–34). The GDP polyribonucleotidyltransferase (PRNTase, EC 2.7.7.88) domain of L transfers 5'-monophospho-RNA (pRNA) from 5'-triphaspho-RNA (pppRNA) to GDP via a covalent enzyme (L)-pRNA intermediate to yield the cap on RNA (29,32), whereas eukaryotic GTase transfers GMP from GTP to 5'-diphospho-RNA (pppRNA) via a covalent enzyme-GMP intermediate (35,36). However, previous studies provided no direct evidence that the L-pRNA intermediate formation and subsequent pRNA transfer to GDP occur co-translationally.

The PRNTase domain possesses five conserved motifs (A–E), in which motif D (also called the HR motif) includes a nucleophilic histidine residue (H1227 for VSV) responsible for the G-pRNA intermediate formation (16,32,34,37). Capping-defective mutations in the PRNTase motifs of VSV L trigger premature termination of N mRNA synthesis at a cryptic signal, resulting in the release of a 5'–pppRNA of 40 nt (37,38). Earlier studies showed that uncapped (11–42 nt) and capped (23–41 nt) short RNAs are generated as abortive transcripts from the N gene-start sequence (26,39–42), suggesting that pre-mRNA capping takes place as an early event during mRNA elongation. However, it has been technically challenging to demonstrate that capped short transcripts synthesized *in vitro* serve as precursors of mature mRNAs. Furthermore, since the VSV-associated MTase activities cannot methylate the cap structure of exogenously added VSV mRNAs (43), it has been believed that these reactions are tightly coupled to mRNA synthesis. Thus, unresolved longstanding questions are when and how the single MTase domain of L catalyzes cap methylation during transcription.

Although considerable biochemical and structural insights have been gained into our understanding of the roles of NNS RNA virus L proteins in mRNA biosynthesis, the mechanisms of pre-mRNA modifications coupled to RNA chain elongation remain largely unknown. In this study, we developed a transcription-coupled pre-mRNA capping system for VSV to elucidate how the enzymatic domains of L coordinate and carry out mRNA maturation during the dynamic transcription cycle. Here, we present evidence that NNS RNA viruses have evolved a unique mRNA assembly machinery, which conducts sequential mRNA modifications during mRNA chain elongation.

Materials and methods

**Nucleic acids**

Nucleotides (ATP, GTP, CTP, UTP and GDP) were purchased from Trilink Biotechnologies or Sigma-Aldrich. [α-32P]GTP (3000 Ci/mmol) and [γ-32P]ATP (3000 Ci/mmol) were obtained from PerkinElmer/Revitix. [α-32P]GTP was prepared from [α-32P]GTP with the nucleoside 5'-triphosphatase activity of vaccinia virus capping enzyme (Cellscript) as described previously (44). mGpppA and GpppA were from New England Biolabs. 5'-Triphosphorylated oligo-RNAs (purified by HPLC, checked by mass spectrometry), such as pppN10 (5'-pppAAC AGU AAU C), pppN20 (5'-pppAAC AGU AAU CAA AAU GUC UG) and pppN30 (5'-pppAAC AGU AAU CAA AAU GUC UGU UAC AGU CAA), were obtained from Trilink Biotechnologies or Bio-Synthesis. 5'-Hydroxyl oligo-RNAs (purified by HPLC, checked by mass spectrometry), such as HO-N10 (5'-AAC AGU AAU C), HO-N20 (5'-AAC AGU AAU CAA AAU GUC UG) and HO-N30 (5'-AAC AGU AAU CAA AAU GUC UGU UAC AGU CAA), were obtained from Sigma-Aldrich. Oligo-DNAs for mutagenesis or RNase protection assays were from Sigma-Aldrich.

**Preparation of the VSV transcription factors**

Ribonucleoprotein (RNP), N-RNA and L-P RdRp complexes were prepared from native (strain Mudd-Summers) or recombinant VSV (rVSV) particles as described previously (28). The recombinant VSV P (rP) and L (rL) proteins were expressed...
as His-tagged forms in Sf21 insect cells using the Bac-to-Bac baculovirus expression system (Invitrogen) and purified as described previously (28,32,44,45). The D714A and HR-RH (H1227R and R1228H) mutant L proteins were prepared as described previously (32). To generate L proteins with a single amino acid point mutation in the MTase domain, PCR-mediated site-directed mutagenesis was performed using Q5 High-Fidelity DNA Polymerase and KLD Enzyme Mix (New England Biolabs) according to the manufacturer’s protocol essentially as described previously (23,46). The D1671V mutant L gene (47) was obtained from Dr. Sue A. Moyer (University of Florida). The names of the point mutant includes the original amino acid (one-letter code) at the indicated position followed by the replacement amino acid.

Generation of rVSVs
rVSVs were generated using the reverse genetics system with the pVSFL-2 plasmid (48) or its N-gene-mutant plasmids as described previously (49). rVSVs were isolated from single plaques and amplified as described previously (23,38,49). N-RNA complexes were purified from the plaque-isolated rVSVs. Genomic RNAs were purified from the N-RNA complexes and used as templates to generate first-strand cDNAs with random primers as described previously (23,38,49). A region encompassing genome nucleotide positions 1–550 was amplified from each first-strand cDNA with forward (5′-ACG AAG ACA AAC AAA CCA TTA TTA TCA TTA AAA GG) and reverse (5′-GAT TTG TCA GCC CAT CCA TGA G) primers. The resulting PCR products were sequenced with the above forward primer.

Preparation of recombinant vaccinia virus capping enzyme
The vaccinia virus D1 and D12 open reading frames were amplified from the vaccinia virus genome [strain vTF7-3 (Wr), ATCC VR-2153] by PCR and cloned into modified pFascBac1 (Invitrogen) vectors encoding C-terminal His-tag and Strep-tag II, respectively. Recombinant baculoviruses expressing the respective proteins were generated using the Bac-to-Bac baculovirus expression system (Invitrogen). The D1 and D12 subunits of vaccinia virus capping enzyme were co-expressed in Sf21 insect cells, which had been co-infected with baculoviruses expressing the respective proteins. A complex of the subunits was purified using Ni-NTA agarose resin (Qiagen) followed by Strep-Tactin Superflow Plus resin (Qiagen).

Protein analyses
Protein concentrations were determined by a Bradford protein assay (Bio-Rad) using bovine γ-globulin as a standard. Purified proteins were analyzed by electrophoresis in polyacrylamide gels (acylamide-bisacrylamide ratio of 30:0.8) containing SDS (SDS-PAGE) followed by staining with Coomassie Brilliant Blue or One-Step Blue (Biotium). Alternatively, gels with a low degree of crosslinking (acylamide-bisacrylamide ratio of 30:0.4) were used for separating P from N (44).

In vitro synthesis of cap-labeled mRNAs
In vitro transcription was performed with N-RNA (0.4 μg of protein), the recombinant P protein (rp, 0.075 μg), and the recombinant L protein [rL, 0.15 μg, wild-type (WT) or mutant] at 30°C for 1 h (or indicated periods) in 25 μl of a transcription buffer [50 mM Tris–HCl (pH 8.0), 5 mM MgCl2, 50 mM NaCl, 2 mM 1,4-dithiothreitol (DTT), 0.2 mg/ml bovine serum albumin (BSA)] containing 2 mM ATP, 0.5 mM CTP, 0.1 mM UTP, 0.1 mM GTP, 0.1 μM [α-32P]GDP (1–3 × 103 cpm/mmol) in the presence or absence 5-adenosyl-1-methionine (SAM, 30 μM unless otherwise mentioned). After treatment with ribonuclease (RNase) H (Roche) in the presence of oligo(dT), 32P-labeled transcripts were extracted with phenol/chloroform/isoamy alcohol, precipitated with ethanol, treated with calf intestine alkaline phosphatase (CIAP), and analyzed by electrophoresis in a 5% polyacrylamide gel containing 7 M urea (urea–PAGE) followed by autoradiography as detailed in the protocol described by Ogino (44). Alternatively, 32P-labeled transcripts were purified using the Direct-zol RNA MicroPrep kit (Zymo Research Corporation) according to the manufacturer’s protocol. To analyze cap structures, transcripts were treated with 1 unit of shrimp alkaline phosphatase (SAP, Roche or Affymetrix) at 37°C for 30 min. After heat-inactivation of SAP at 65°C for 15 min, transcripts were extracted with phenol/chloroform/isoamy alcohol, precipitated with ethanol, and digested with nuclease (Nase) P1 (Sigma-Aldrich) as described previously (44). Nase P1-resistant products (cap structures) were analyzed by thin-layer chromatography (TLC) on a polyethyleneimine (PEI) cellulose plate (Millipore Sigma) followed by autoradiography as described previously (44). Marker compounds were visualized under UV light at a wavelength of 254 nm. Cap-Clip acid pyrophosphatase (Clebscript) was used to cleave the pyrophosphate linkages in the cap.

In vitro transcription-coupled pre-mRNA capping
To assemble transcription initiation complexes, N-RNA (0.4 or 0.8 μg of protein) was pre-incubated with rP (0.075 or 0.15 μg) and rL (0.15 or 0.3 μg) at 30°C for 15 min in 23 μl of the transcription buffer containing 2 mM ATP and 0.5 mM CTP in the presence or absence of 30 μM SAM. After adding 2 μl of a mixture of UTP (final concentration: 0.1 mM), GTP (final concentration: 0.1 mM), and [α-32P]GDP (final concentration: 0.1 μM, 3–9 × 103 cpm/mmol), the reaction mixture was incubated for 1 min (or the periods indicated in figures) to incorporate [α-32P]GDP into the cap formed on short pre-mRNAs. Alternatively, cap-labeling of short transcripts was performed with a limiting concentration (1 μM), rather than the saturating concentration (0.1 mM), of GTP for 3 min (or the periods indicated in figures). Transcription was terminated by adding RNA extraction buffer (10 mM Tris–HCl [pH 7.5], 1 mM EDTA, 0.5% SDS, 0.2 M NaCl, 7 M urea). Transcripts were purified as described above and analyzed by 20% urea–PAGE followed by autoradiography. Pre-mRNAs were eluted from the gel and their cap structures were analyzed as described previously (44).

Post-transcriptional capping of 5′-triphosphorylated pre-mRNAs
Short pre-mRNAs were generated with cold GDP, instead of [α-32P]GDP, as described above. 5′-Triphosphorylated pre-mRNAs (pppAACAG-started RNAs) were selectively capped with rL in the presence of [α-32P]GDP under the conditions optimized for transcription-uncoupled capping as described previously (38).
Pulse-chase transcription

As described above, transcription initiation complexes were pre-assembled in the presence or absence of 30 μM SAM, and [α-32P]GDP-labeling of short transcripts was performed in the presence of the saturating concentration (0.1 mM) of GTP for 1 min (pulse). After adding 2 μl of 1.25 mM GDP (final concentration: 0.1 mM), the reaction mixture was further incubated for the periods indicated in the figures to elongate the pre-mRNAs (chase). Alternatively, pulse [α-32P]GDP-labeling of short transcripts was performed with a limiting concentration (1 μM), rather than the saturating concentration, of GTP, for 3 min. After adding 2 μl of a mixture of GTP and GDP (final concentration: 0.1 mM each), transcripts were chased for the times indicated in the figures. In some experiments when pulse-chase transcription was carried out in the absence of SAM, the chase reaction mixtures received SAM (final concentration: 30 μM) at the time points indicated in the figures and then were further incubated for 3 min. Transcription was terminated by adding RNA extraction buffer. Transcripts were purified as described above and analyzed by 20% or 5% urea-PAGE followed by autoradiography. The caps formed on transcripts were analyzed as described above.

Separation of free transcripts from RNP-associated transcripts

Pulse-labeling of pre-mRNAs with [α-32P]GDP was performed with N-RNA (0.8 μg of protein), rP (0.15 μg), and rL (0.3 μg) in the presence of the limiting concentration (1 μM) of GTP as described above. After adding the excess concentrations of GTP and GDP, [α-32P]GDP-labeled pre-mRNAs were chased for the indicated periods. The reactions were stopped by adding 25 μl of ice-cold 2× stopping buffer [40 mM Tris–HCl (pH 8.0), 20 mM EDTA, 0.6 M NaCl, and 0.2% Triton X-100, 2 mM DTT]. The reaction mixtures and rinses of the tube inner walls with 10 μl of 1× stopping buffer (total volume: ∼60 μl) were sequentially overlaid onto 200-μl glycerol cushions [20 mM Tris–HCl (pH 8.0), 0.5 mM EDTA, 50 mM NaCl, 20% glycerol] in 1.5-ml ultracentrifuge tubes (Beckman) and centrifuged in the Beckman TL-A 55 fixed-angle rotor at 110000 × gav (average centrifugal force) for 30 min at 4°C. Two hundred μl of the supernatants were collected from the top of the solutions. Afterward, 200 μl of 1× stopping buffer was overlaid onto the remaining glycerol cushions and then removed. These wash steps were repeated to remove free 32P-labeled nucleotides. After removing the remaining cushions, pellets were dissolved in TRizol LS Reagent (Invitrogen). 32P-Labeled transcripts were purified from the supernatant and pellet fractions using the Direct-zol RNA MicroPrep kit (Zymo Research Corporation) according to the manufacturer’s protocol.

Co-transcriptional L-pRNA intermediate formation

Transcription initiation complexes were pre-assembled with N-RNA (N gene mutant A+19C, 1.6 μg of protein), rP (0.3 μg), and rL (WT or mutant, 0.6 μg) at 30°C for 15 min in 22 μl of the transcription buffer containing 2 mM ATP and 0.5 mM CTP in the absence of 30 μM SAM. After adding 3 μl of a mixture of UTP (final concentration: 0.1 mM) and [α-32P]GTP (final concentration: 1 μM, 1–2 × 10^3 cpm/fmol), the reaction mixture was incubated for 3 min. After stopping the reactions, the RNP complexes were precipitated through glycerol cushions by ultracentrifugation as detailed above. The resulting pellets were dissolved in a modified SDS-PAGE sample buffer [200 mM Tris–HCl (pH 8.0), 4 mM EDTA, 4% SDS, 0.02% bromophenol blue (BPB), 7 M urea, 20 mM DTT], and immediately analyzed by 7.5% SDS-PAGE (acylamide-bisacrylamide ratio of 30:4) followed by autoradiography. After autoradiography, proteins in the gels were stained as described above.

Identification of transcripts covalently linked to the L protein during transcription

RNP complexes (3 or 6 μg of protein) purified from rVSV particles [WT or N gene mutant 1 (A+19C)] were subjected to the co-transcriptional L-pRNA intermediate formation as described above. After stopping the reactions, the RNP complexes were precipitated through a glycerol cushion by ultracentrifugation and analyzed by SDS-PAGE followed by autoradiography as described above. Since putative L-pRNA intermediates formed during transcription seemed to be heterogeneous, large gel blocks containing multiple bands were excised. Alternatively, only a major band was excised from the gel when N gene mutant 1 RNP was used. After rinsing the gel pieces with water, they were incubated in a 1 M hydroxyamine (pH 7.5) solution containing 0.1% SDS at 37°C and 1000 rpm in the ThermoMixer (Eppendorf) for 30 min to cleave the phosphoamide bond between L and pRNA. After the incubation, pRNAs detached from L were eluted from the gel. After brief centrifugation, the reaction mixtures (supernatants) were collected. Incubation of the gel with hydroxyamine was repeated. After adding sodium acetate (pH 5.5) to the combined reaction mixtures at a final concentration of 0.3 M, the pRNAs were precipitated with 2.5 volumes of ethanol in the presence of 20 μg of glycerol. The resulting pellet was rinsed with cold 70% ethanol, dried in a SpeedVac concentrator, and dissolved in water or formamide loading dye solution.

The purified 32P-labeled RNAs (∼300 cpm) were mixed with an oligo-DNA probe (N1-18_R, 5'-GAC ATT TTG ATT ACT GTT or N1-18_F, AAC AGT AAT CAA AAT GTC) in a 10-μl hybridization buffer (20 mM Tris–HCl, 150 mM NaCl, 1 mM EDTA). The mixtures were heated at 80°C for 5 min and then cooled to 37°C at the ramp rate of 0.1°C/s. The mixtures were further incubated with or without 0.5 units of RNase T2 (Invitrogen) at 37°C for 15 min. Alternatively, the purified RNAs were treated with or without 1 unit of SAP at 37°C for 20 min. After heat-inactivation of SAP at 65°C for 15 min, the RNAs were digested with 0.5 units of XRN-1 (New England Biolabs) at 37°C for 20 min. The digested RNAs were analyzed by 20% urea-PAGE followed by autoradiography.

Pre-mRNA cap methylation with elongation complexes

Pulse-labeling of pre-mRNAs with [α-32P]GDP was performed with N-RNA (0.8 μg of protein), rP (0.15 μg), and rL (0.3 μg) in the presence of the limiting concentration (1 μM) of GTP as described above. After adding the excess concentrations of GTP and GDP, [α-32P]GDP-labeled pre-mRNAs were chased for 20 s. After stopping the reactions, RNP complexes were precipitated through glycerol cushions as described above. The resulting pellets were suspended in 23 μl of a buffer [50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 2 mM DTT, 0.2 mg/ml BSA] and preincubated at 30°C for 1 min. After adding 2 μl of a mixture of MgCl₂ (final concentration: 5 mM), SAM (final concentration: 30 μM), and 4 NTPs or
3'-deoxy-NTPs (Trilink Biotechnologies, final concentration: 0.1 mM each), the reaction mixture was further incubated for 3 min. The cap-labeled RNAs and their Nase P1 digests were analyzed as described above.

RNA cap methylation with free enzymes
Cap-labeled transcripts were synthesized by the pulse-chase transcription system as described above and separated from the genome by 5% urea–PAGE. After autoradiography, a small block of the gel containing cap-labeled transcripts was excised and chopped into pieces. Cap-labeled transcripts were eluted from the gel pieces and recovered by ethanol precipitation as described previously (44). The cap-labeled transcripts and 30 μM SAM were incubated with RL (0.3 μg) and rP (0.15 μg) in the presence or absence of the N-RNA (0.8 μg of protein) in 25 μl of the transcription buffer at 30°C for 3 min. Our recombinant vaccinia virus capping enzyme (0.1 μg) was used as a positive control. The cap-labeled transcripts were extracted with phenol/chloroform/isoamyl alcohol and recovered by ethanol precipitation. Nase P1 digests of the cap-labeled transcripts were analyzed by PEI-cellulose TLC (44).

The pppN10 oligo-RNA was capped with vaccinia virus capping enzyme (Cellscript) in the presence of [α-32P]GTP and used as a substrate for cap methylation under the same conditions as those reported by Rahmeh et al. (50).

Preparation of marker RNAs
5'-Triphosphorylated oligo-RNAs (pppN10, pppN20, and pppN30) were capped with vaccinia virus (Cellscript) or fass-tovirus (New England Biolabs) capping enzyme in the presence of [α-32P]GTP to generate cap-labeled RNAs (GpppN10, GpppN20, and GpppN30) according to the manufacturer’s protocols. 5'-Hydroxy oligo-RNAs (HO-N10, HO-N20, and HO-N30) were phosphorylated with T4 polynucleotide kinase (Roche) in the presence of [γ-32P]ATP to yield 5’-32P-labeled RNAs (pN10, pN20, and pN30). Alkaline ladders of 5’-32P-labeled RNAs were generated as described previously (38) and treated with the 3'-phosphatase and 2’,3’-cyclic phosphodiesterase activities of T4 polynucleotide kinase (Roche or New England Biolabs). 32P-Labeled Ambion Century-Plus and Decade RNA markers (Life Technologies/Thermo Fisher Scientific) were prepared according to the manufacturer’s protocols. 32P-Labeled cap structures (GpppA, GpppAm, m1GpppA, and m1GpppAm) were synthesized with vaccinia virus capping enzyme and N1-2'-O-MTase (VP39) (Cellscript) as described previously (44).

Modeling of the L protein elongation complex with eighteen nucleotide pre-mRNA
Coordinates of VSV L in the putative post-initiation state were obtained from the ModelArchive [ID: ma-ecf5b (46)]. In the prior model of the pre-mRNA, the emerging RNA strand was built as a 12-mer. This was extended by aligning an idealized strand of RNA with the helically stacked RNA at the emerging 5'-terminus in Coot (51). RNA was then fused and renumbered. The 5'-nucleotide was converted to the ATP to incorporate the triphosphate moiety, and the first five ribonucleotides were modeled as the AAGAC. For an alternate model with the initial ATP bound to the PRNTase, the cap structure was docked with AutoDock Vina (52). The cap was split between the G and A moieties. The A with triphosphate was fused to the nucleotides 2–18 of the pre-mRNA. The completed models were subjected to the molecular dynamics energy minimization routine in YASARA (53). Structural images were generated using the PyMOL software (http://www.pymol.org/).

Results
Development of an in vitro transcription system with the recombinant VSV L protein to analyze co-transcriptional mRNA processing
The apparent K_m for GDP (0.03 μM) as the pRNA acceptor in RNA capping with VSV L (54) is 3–4 orders of magnitude lower than GDP concentrations in mammalian cells (55), suggesting that VSV L can form the cap with pre-existing GDP without GDP hydrolysis in infected cells. Furthermore, the K_m for GDP in capping is significantly lower than that for GTP (22 μM) as the GMP donor in VSV RNA synthesis (56). These observations prompted us to examine whether L can directly incorporate GDP into the cap core during transcription. In vitro transcription reconstituted with the recombinant L and P proteins (RL and rP) and native N-RNA (Supplementary Figure S1A) was performed for 0.5–2 h in the presence of 0.1 μM [α-32P]GDP and 0.1 mM GTP, which seem to be enough to achieve maximum reaction velocities for capping and RNA synthesis, respectively (Figure 1A). After removing poly(A) tails, transcripts were analyzed by denaturing PAGE (Figure 1B) N, P/M and G RNAs synthesized in the presence of [α-32P]GDP were found to be 32P-labeled efficiently. Nuclease (Nase) P1 digestion of the 32P-labeled transcripts (Figure 1C, lane 1) released a single product (co-migrating with a GpppA cap standard) (lane 2), which was resistant to calf intestine alkaline phosphatase (CIAP) (lane 3). Cap-Clip acid pyrophosphatase digested the Nase P1-resistant product into [α-32P]GMP (lane 4), which was further hydrolyzed into inorganic phosphate (P_i) with CIAP (lane 5). These results indicate that the 5’-terminal cap core structure on mRNAs can be specifically labeled with [α-32P]GDP.

Cap-labeled mRNAs synthesized by wild-type (WT) VVL with (Figure 1D, lane 3) or without (lane 1) 30 μM SAM migrated as a broad smear, while those treated with RNase H in the presence of oligo(dT) were separated into three discrete bands corresponding to N1 P/M, and G mRNAs (lanes 2 and 4) as in panel B, indicating that these mRNAs were polyadenylated. Consistent with the previous finding obtained using an in vitro transcription system with detergent-disrupted VSV particles (29), single cap methylation at the A/2'-O position in the reconstituted reactions was observed at lower concentrations (0.03–1 μM, Figure 1E, lanes 3–6) of SAM, whereas methylation at both the A/2'-O and G-N= position was observed at higher concentrations (3–100 μM, lanes 7–10) of SAM.

To confirm whether the co-transcriptional methylation reactions are catalyzed by intrinsic activities of RL, we prepared L proteins with a single point mutation of conserved amino acid residues in its MTase domain (Supplementary Figure S2). Like other 2'-O-MTases, the VSV MTase domain contains a catalytic K–D–K–E tetrad sequence (K1651A–D1762–K1795–E1833) (50, 57). Although we could not obtain three K1651A, K1795A or E1833A of the previously reported K–D–K–E mutants (50, 57) in satisfactory yield and purity due to their low expression levels and solubility, we identified K1795R (Supplementary Figure S1B) as an ideal
Figure 1. Development of a co-transcriptional 5'-end processing assay for VSV RL. (A) In vitro transcription was performed with N-RNA and recombinant L (RL) and P (RP) in the presence of [α-32P]GDP with or without S-adenosyl-L-methionine (SAM) (left). [α-32P]GDP is incorporated into the mRNA cap core (GpppA), which is methylated at the A1-2'-O and G-N' positions into the cap 1 (m7GpppAm). An asterisk indicates a 32P-labeled phosphate. The RNA-dependent RNA polymerase (RdRp) L-P complex transcribes the genome into the leader RNA (LeRNA) and 5'-capped and 3'-polyadenylated mRNAs (right, for detail, see text). (B) mRNAs were synthesized in the presence of [α-32P]GDP for the indicated periods, digested with RNase H in the presence of oligo(dT), and analyzed by 5% urea-PAGE followed by autoradiography. The positions of deaminated mRNAs (N, P/M and G) and gel origin (ori.) are shown on the left. (C) mRNAs synthesized for 1 h as in panel B were digested with nuclease (Nase) P1, calf intestine alkaline phosphatase (CIAP), and/or Cap-Clamp acid pyrophosphatase. The digests were analyzed by PEI-cellulose TLC followed by autoradiography. The positions of the indicated compounds are shown on the right. (D) In vitro transcription was performed with wild-type (WT) RL with the RdRp, GDP polynucleotide transferase (PRNase), and methyltransferase (MTase) domains or its mutant [K1795R, methylation-defective; HR-RH (H1227R-R1228H), capping-defective; D714A, transcription-defective] in the presence or absence of 30 μM SAM for 1 h as in panel B. After the treatment with or without RNase H in the presence of oligo(dT), mRNAs were analyzed as in panel B. (E) mRNAs were synthesized with RL (WT or mutant) in the presence (lanes 2–9: 0.01, 0.03, 0.1, 0.3, 1, 3, 10 and 30 μM, respectively; lanes 10–13: 100 μM) or absence of SAM as in panel D and digested with Nase P1. The digests were analyzed as in panel C. ‘MC’ lane shows a mixture of the indicated 32P-labeled cap structures. The graph shows relative amounts of GpppA, GpppAm, m7GpppA, and m7GpppAm, whose total amounts were set to 100%. Symbols and error bars represent the means and standard deviations, respectively, of three independent experiments (n = 3).

methyltransferase-defective mutant that maintains the solubility and RNA synthesis activity (Figure 1D, lanes 5 and 6) but loses both the methylation activities (Figure 1E, lane 11). R1640A also exhibited similar phenotypes to K1795R, whereas other methyltransferase-defective mutants D1671V, D1762A and D1762E, (47, 57, 58) showed transcription activities lower than that of WT (Supplementary Figure S2C). Interestingly, E1764A and E1764E mutations were found to diminish the G-N'-MTase activity but had a moderate or little effect on the 2'-O-MTase activity (Supplementary Figure S2C). Taken together, these results suggest that the MTase domain of RL is responsible for both the co-transcriptional cap methylation activities. Furthermore, neither the capping-defective [HR-RH, H1227R-R1228H (32)] nor transcription-defective [D714A (59)] L mutant generated cap-labeled mRNAs (Figure 1D, lanes 7 and 8; Figure 1E, lanes 12 and 13), indicating that both the RdRp and PRNase activities of RL are essential for their generation. Therefore, we demonstrated that the transcription system with [α-32P]GDP can monitor all the enzymatic reactions required for the mature mRNA synthesis.

The VSV pre-mRNAs of 21–23 nt synthesized by a fast-transcribing RdRp have an unmethylated cap core structure

To investigate a minimum length of pre-N mRNAs that can be capped co-transcriptionally in the presence of saturating concentrations of 4 NTPs and SAM, we analyzed short transcripts synthesized in the presence of [α-32P]GDP for short periods (15–180 s) (Figure 2A). The RdRp tended to produce larger amounts of longer transcripts when incubated for longer periods (lanes 3–5). Interestingly, cap-labeled transcripts of 21–23 nt initially appeared in transcripts synthesized for 30 s and were also consistently detected in transcripts synthesized for 60 and 180 s. The transcripts of 21–23 nt possessed the cap core (GpppA) (Figure 2B).

A pulse-chase experiment (Figure 2C, left scheme) demonstrated that the short transcripts (including the 21–23 nt RNAs) pulse-labeled with [α-32P]GDP for 60 s can be chased in the presence of an excess concentration of cold GDP into longer transcripts in a time-dependent manner (Figure 2C, right pane, lanes 2–4). Furthermore, their cap was methylated into m7GpppAm during the 3-min chase period.
Figure 2. Nascent transcripts of 21–23 nt, but not of 13–15 nt, are capped with fast-transcribing rL. (A) N-RNA was pre-incubated with rL and rP in the presence of ATP, CTP, and SAM. After adding saturating concentrations of UTP, GTP, and [α-32P]GDP, the reaction mixtures were incubated for the indicated periods. Cap-labeled transcripts were analyzed by 20% urea–PAGE followed by autoradiography. Marker RNAs are as follows: i, a mixture of 5′-capped oligo-RNAs with the VSV N mRNA start sequence (10 nt, GpppN10; 20 nt, GpppN20; 30 nt, GpppN30); ii, a hydroxide cleavage ladder of GpppN30, 3′-dephosphorylated. The positions of the gel origin (ori.) and bottom (bot.) are indicated. (B) The transcripts of 21–23 nt synthesized in the presence of [α-32P]GDP for 60 s (see panel A, lane 4) were purified by urea–PAGE and digested with Nase P1. The digests were analyzed as in Figure 1E. (C) Nascent transcripts were pulse-labeled with [α-32P]GDP for 60 sec as in panel A. After adding excess cold GDP, cap-labeled transcripts were chased for the indicated periods, and analyzed as in panel A. Marker lane iii indicates Ambion Decade Markers. (D) Cap-labeled nascent transcripts before and after chasing for 180 s (see panel C, lanes 1 and 4) were digested with Nase P1. The digests were analyzed as in Figure 1E. (E) Nascent transcripts were synthesized in the presence of [α-32P]GDP (left scheme) or cold GDP (right scheme) for 60 s as in panel A. The latter RNAs were post-transcriptionally capped with [α-32P]GDP by rL. (F) Co-transcriptionally (lane 1) or post-transcriptionally (lane 2) capped transcripts were analyzed as in panel A. Marker RNAs are as follows: iv, a mixture of GpppN10 and GpppN20; v, a hydroxide cleavage ladder of GpppN30, 3′-dephosphorylated.
The minimum length of pre-mRNA required for co-transcriptional capping is 18 nt

To investigate whether L can cap pre-mRNAs shorter than 21 nt, we performed in vitro transcription with [α-32P]GDP in the presence of a limiting concentration (1 μM) of GTP for 1–9 min, under which conditions the RdRp was expected to pause or terminate transcription at C residues within the N gene (Mudd-Summers strain, GenBank accession no.: MN164438) (Figure 3A). As expected, the VSV RdRp generated ladders of capped RNAs (bands 1–7: 19, 25, 30, 32, 40, 49, and 55 nt, respectively; bands 8–16: ~80–~150 nt) at the expected positions (Figure 3B, lanes 3–6). The shortest capped transcript consistently produced for 3–9 min was a 19-nt RNA (band 1, lanes 3–6; Supplementary Figure S4A and B, lane 1). Since the mobility of 5’-monophosphorylated (and possibly 3’-monophosphorylated) RNAs of the indicated sizes with unknown sequences generated by the Ambion Decade Markers System (Figure 3B, M lane iii) is different from that of capped VSV N pre-mRNAs of the same sizes, we determined the length (19 nt) of the shortest capped N pre-mRNA using capped synthetic oligo-RNAs with the VSV N mRNA-start sequences (10 and 20 nt) and capped RNA ladders [generated by partial digestion of the capped 20-nt RNA with Nase P1 or alkaline followed by the 3’-phosphatase activity of T4 polynucleotide kinase (60)] as accurate size markers (Supplementary Figure S4). Post-transcriptional capping of transcripts (synthesized in the presence of cold GDP) with Rl also showed that the RdRp co-transcriptionally produced a 5’-triphosphorylated pre-mRNA of 15 nt by transcriptional pausing/termination at position +16 (see Figure 3A) in the presence of a limiting concentration of GTP (Supplementary Figure S4A and B, lane 2). We also confirmed that native VSV RNPs produce the 19-nt transcript as the shortest capped RNA under the GTP-limiting conditions (Supplementary Figure S5).

To analyze the cap methylation status of transcripts synthesized in the presence of the limiting concentration of GTP for 9 min, they were eluted from respective gel slices, digested with Nase P1, and then analyzed by PEI-cellulose TLC (Figure 3C). The transcripts of 19–32 nt (bands 1–4) possessed the cap core. The cap on transcripts of 40–70 nt (bands 5–7) and ~80–120 nt (bands 8–13) was partially and almost completely methylated at the N1–2’-O position, respectively, whereas the cap on transcripts of ~80–~120 nt (bands 9–13) and ~150 nt or longer (bands 14–16) were partially and almost fully methylated at the G-N7 position, respectively, into the cap 1 (m7GpppG). These results indicate that co-transcriptional A2′-O-methylation of the cap precedes G-N7-methylation.

To determine the minimum length of pre-mRNA required for capping at one-nucleotide resolution, we generated engineered recombinant N-RNA templates (tN-RNA) with an additional C residue between positions +17 and +19 in the N gene using the reverse genetics system (Figures 3D and Supplementary Figure S6). Transcription was performed with rN-RNA (WT or mutant), Rl, and rp for 3 min in the presence of the limiting concentration of GTP and 0.1 μM [α-32P]GTP to generate co-transcriptionally capped transcripts (Figure 3E). Furthermore, transcripts were synthesized in the presence of cold GDP under the same conditions, and then post-transcriptionally capped with Rl in the presence of [α-32P]GTP to detect 5’-pppACAG-RNAs (Figure 3F). The RdRp produced a 5’-capped 19-nt transcript as the smallest capped RNA (Figure 3E, lane 1) and 5’-triphosphorylated 15-nt transcript from WT rN-RNA (Figure 3F, lane 1) as from native N-RNA (Figures 3B and Supplementary Figure S4). The RdRp strongly paused/terminated at position +19 on the N gene mutant 1 genome most probably due to the presence of the tandem C residues at positions +19 and +20, generating 5’-capped (Figure 3E, lane 2) and 5’-triphosphorylated (Figure 3F, lane 2) 18-nt transcripts with higher band intensities than expected. The RdRp yielded 5’-triphosphorylated (Figure 3F, lanes 3 and 4), but not 5’-capped (Figure 3E, lanes 3 and 4), 17-nt and 16-nt transcripts from the genomes of N gene mutants 2 and 3, respectively. We also confirmed that, under UTP-limiting conditions, the RdRp produces 5’-capped 18-nt and 5’-triphosphorylated 14-nt transcripts from native N-RNA (Supplementary Figure S7). Thus, the 5’-ends of pre-mRNAs shorter than 18 nt seem unlikely to gain access to the PRNTase active site.

The co-transcriptionally capped 18-nt transcript synthesized from N gene mutant 1 rN-RNA contained the cap core (Figure 3G, lane 1) and could be chased into longer transcripts in the presence of excess concentrations of GDP and GTP (Figure 3H, lanes 2–4). We also confirmed that its cap could be methylated at both the positions in the cap 1 during the 3-min chase period (Figure 3I, lanes 2).

The L protein forms covalent intermediates with pre-mRNAs for capping at an early stage of mRNA elongation

The N gene mutant 1 template paused the RdRp predominantly at position +19 in the presence of the limiting concentration of GTP to generate the capped and 5’-triphosphorylated 18-nt pre-mRNA (Figure 3E and F). This finding prompted us to investigate whether the PRNTase domain transiently forms a covalent capping intermediate with the 18-nt pre-mRNA. We preincubated N gene mutant 1 rN-RNA with rp and Rl. (WT or mutant) in the presence of ATP and CTP (Figure 4A). After adding the saturating concentration (0.1 mM) of UTP and the limiting concentration (1 μM) of [α-32P]GTP, RNA synthesis was performed without external GDP for 3 min under the same conditions as in Figure 3E and F. Assembled RNPs were precipitated through a glycerol cushion by ultracentrifugation and analyzed by SDS-PAGE followed by autoradiography (Figure 4B). SDS-resistant
Figure 3. The minimum length of pre-mRNA required for co-transcriptional capping with Rl is 18 nt. (A) When transcription is performed with [α-32P]GDP in the presence of a limiting concentration (1 μM) of GTP for short periods, the RdRp is expected to pause/terminate transcription at C residues within the N gene, resulting in the synthesis of 5'-triphosphorylated or capped pre-mRNAs with the indicated lengths. (B) N-RNA was pre-incubated with Rl and RdRp in the presence of ATR, CTP, and SAM as in Figure 2A. After adding the limiting concentration of GTP and saturating concentrations of UTP and [α-32P]GDP as in the presence of GTP, the reaction mixtures were further incubated for the indicated periods. Transcripts were analyzed as in Figure 2A. (C) Nascent transcripts were synthesized in min as in panel B. Individual transcripts (bands 1–16, see panel B) were purified by urea-PAGE and digested with Nase P1. The digests were analyzed as in Figure 1E. (D) The RdRp is expected to produce extra transcripts (N-pre-mRNAs) of 18, 17, and 16 nt from end-RNAs of N gene mutants 1, 2, and 3, respectively, in the presence of the limiting concentration of GTP as in panel A. The latter RNAs were post-transcriptionally capped with [α-32P]GDP by Rl as in Figure 2E. Transcripts were analyzed as in Figure 2A. (G) The nascent transcript of 18 nt synthesized from N gene mutant 1 N-RNA (see panel E, lane 2) was purified by urea-PAGE and digested with Nase P1. The digests were analyzed as in Figure 1E. (H) Nascent transcripts from N gene mutant 1 N-RNA were pulse-labeled with [α-32P]GDP for 3 min as in panel E. After adding excess concentrations of cold GDP and GTP, cap-labeled nascent transcripts were chased for the indicated periods and analyzed as in Figure 2A. (I) Cap-labeled nascent transcripts before and after chasing for 180 s (see panel H, lanes 1 and 4) were digested with Nase P1. The digests were analyzed as in Figure 1E.
Figure 4. rL forms covalent complexes with nascent pre-mRNAs during RNA chain elongation. (A–C) N gene mutant 1 rN-RNA and rP were pre-incubated with or without rL (WT or mutant) in the presence of ATP and CTP as in Figure 2A. After adding the saturating concentrations of UTP and the limiting concentration of [α-32P]GTP, the reaction mixtures were further incubated for 3 min. RNPs were precipitated through a glycerol cushion by ultracentrifugation. The resulting pellets were analyzed by 7.5% SDS-PAGE followed by autoradiography (B). The area containing putative L-pRNA intermediates is shown on the right. After autoradiography, the gel was stained with One-Step Blue (C). M lane shows marker proteins with the indicated molecular masses. The positions of the viral proteins are indicated on the right. (D and E) RNPs purified from rVSV particles (WT or N gene mutant 1) were subjected to the co-transcriptional L-pRNA formation assay and analyzed as in panel A (D, upper left). Radioactive bands were excised from the gel and incubated with hydroxylamine (pH 7.5) to cleave the phosphoamide bond between pRNA and the active site histidine residue (H1227) of the L PRNtase domain (lower). The resulting pRNAs recovered from the supernatant fractions were analyzed by 20% urea–PAGE followed by autoradiography (E). Marker RNAs are as follows: vii, a mixture of 5′-monophosphorylated oligo-RNAs with the VSV N mRNA start sequence (10 nt, pN18; 20 nt, pN16; 30 nt, pN16); viii, a hydroxide cleavage ladder of pN18, 3′-dephosphorylated. (F) Based on the N gene-start sequence of the rVSV genome, a sequence of 16-nt pRNA (pRNA16) detached from L with hydroxylamine was predicted (top). 32P-Labeled phosphates are shown in red. Purified pRNA18 was subjected to an RNase T2 protection assay with the indicated oligo-DNA probes (middle, see panel G) or an XRN-1 (5′-monophosphate-dependent 5′-3′ exoribonuclease 1) digestion assay (bottom, see panel H). SAP indicates shrimp alkaline phosphatase. (G) Isolated pRNA18 was annealed with the indicated oligo-DNA probe (see panel F) and digested with RNase T2, a base non-specific endonuclease. The digests were analyzed as in panel E. (H) Isolated pRNA18 was pre-treated with or without SAP, and then digested with XRN-1. The digests were analyzed as in panel E.
high-molecular-weight radioactive products were detected when WT rL was used (lane 2). As in the absence of rL (lane 1), the transcription-defective D714A (lane 3) and capping defective HR-RH (lane 4) mutants were inert, suggesting that L–pRNA intermediates could be formed with the RdRp and PRNTase activities during transcription. Consistent with the ability of the methylation-defective K1795R mutant to synthesize capped mRNAs (see Figure 1D and E), it formed the putative intermediates (Figure 4B, lane 5). It should be noted that the major and minor products (Figure 4B, lanes 2 and 5) migrated slower than unreacted rL (Figure 4C, lanes 2 and 5) slightly and significantly, respectively. These results suggest that rL formed covalent intermediates with pre-mRNAs of various lengths (e.g. 18 nt and longer), whose negative charges might retard the migration of rL by repelling negatively charged SDS.

To identify pre-mRNAs covalently linked to L during transcription, we used RNPs purified from WT or N gene mutant 1 rVSV particles that exhibited much stronger transcription activity than those reconstituted with rL, rP and rN-RNA (WT or N gene mutant 1). When WT RNPs were subjected to the co-transcriptional L–pRNA formation assay (Figure 4A), multiple faint bands were observed (Figure 4D, lane 1). In contrast, when N gene mutant 1 RNPs were used, lower and upper bands were detected as major and minor products, respectively (lane 2). In-gel treatment of putative intermediates with hydroxylamine (pH 7.5) was carried out to cleave the phosphoamidate bond between L and pRNA (32). The resulting pRNAs were analyzed by urea-PAGE followed by autoradiography (Figure 4E). When putative intermediates formed with WT RNPs were subjected to the hydroxylamine treatment, multiple products including small RNAs co-migrating with 19-nt and 21-nt marker pRNAs were recovered (lane 1). On the other hand, when N gene mutant 1 RNPs were used, an RNA co-migrating with the 18-nt marker pRNA (pRNA18) was observed as a major product (the smallest RNA) (lane 2). Furthermore, longer pRNAs with estimated lengths of 40, 49, and 52 nt were detected as minor products in both samples (lanes 1 and 2).

To demonstrate that pRNA18 contains residues 1–18 of N mRNA (Figure 4F, top), it was subjected to an RNase T2 protection assay (Figure 4F, middle). When pRNA18 was annealed with an oligo-DNA probe with the negative-sense N gene-start sequence encompassing nucleotide positions 1 to +18 [N1-18(−)], it became resistant to RNase T2, a base non-specific endonuclease (Figure 4G, lane 3). In contrast, pRNA18 annealed with an oligo-DNA probe with the positive-sense sequence [N1-18(+)] was completely degraded with RNase T2 (lane 4) as in the absence of a probe (lane 2). Furthermore, RNase T1, RNase A or RNase T2 cleaved pRNA18 at predicted positions, generating expected products (Supplementary Figure S8).

To confirm the presence of a 5′-monophosphate group on pRNA18, it was subjected to a digestion assay with XRNR1, a 5′-monophosphate-dependent 5′-to-3′-exonuclease (Figure 4F, bottom). pRNA18 was degraded with XRNR1 (Figure 4H, lane 3), whereas shrimp alkaline phosphatase-treated pRNA18, with a 5′-hydroxyl group (lane 2) was resistant to XRNR1 (lane 4). Therefore, pRNA18 detached from the putative intermediate was identified as the 5′-monophosphorylated N pre-mRNA with residues 1–18.

The 5′-capped short transcripts are precursors of mature N mRNA

As outlined in Figure 5A, we performed the pulse-chase transcription assay to investigate whether the capped short transcripts in length from 19 to ∼80 nt observed in Figure 3B (see lane 3) can be elongated to produce full-length N mRNA (Figure 5B and C). These capped transcripts (Figure 5B, lane 1), except for the 40-nt transcript (lanes 2–6), were found to be elongated to generate heterogeneous RNAs longer than 1.3 kilo-nt by 7–9 min (Figure 5C, lanes 5 and 6). Transcripts chased for 9 and 11 min migrated as a broad smear and their elongation appeared to be halted (Figure 5C, lanes 6 and 7; Figure 5D, lane 1). These transcripts could be digested with RNase H in the presence of oligo(dT) into a 1.3-kilo-nt RNA that co-migrated with deadenylated N mRNA (Figure 5D, lane 2), indicating that they were co-transcriptionally 3′-polyadenylated. We found that the capped short transcripts at time point 0 are mostly unmethylated (Figure 5E, lane 1), whereas those ranging from 0.5 to 0.6 kilo-nt at the time point of 3 min are mostly methylated at both the positions (lane 6). Again, A1-2′-O-methylation of the cap preceded G-N′-methylation (lanes 2–5). These results indicate that the capped short transcripts are precursors of mature N mRNA.

To investigate whether longer pre-mRNAs serve as substrates for cap methylation by transcribing L, SAM was added to the chase reaction mixtures lacking SAM at different time points (3, 7 and 11 min), and the reaction mixtures were subsequently incubated for additional 3 min (Figure 5A, lower arrows). As a negative control, the reaction mixture received the same volume of water, instead of SAM, after a 3-min chase and further incubated for 3 min (Figure 5F, lane 1). When the reaction mixture received SAM after a 3-min chase, the cap on pre-mRNAs of ∼0.6 kilo-nt was efficiently methylated at both the positions during the following 3-min incubation (lane 2). However, the efficiency of cap methylation during the 3-min incubation decreased significantly if the reaction received SAM at the time point of 7 min (lane 3). At this time point, transcripts included the full-length N mRNA and its 3′-polyadenylated forms (Figure 5C, lane 5). Cap methylation was rarely observed if SAM was added at the time point of 11 min (Figure 5E, lane 4), at which mRNA 3′-polyadenylation seemed to be completed (Figure 5C, lane 7).

To investigate whether capped transcripts are associated with elongation complexes, the chase reaction mixtures at respective time points were separated into supernatant and pellet (RNP) fractions by ultracentrifugation (Figure 5G). At time point 0, capped transcripts shorter than 100 nt were detected in the supernatant and/or pellet fractions (Figure 5H, lanes 1 and 2), whereas transcripts longer than ∼100 nt at the time points of 20 sec, 1 min, and 3 min were mostly in the pellet fraction (lanes 3–6; Figure 5I, lanes 1–4), suggesting that middle-phase elongation complexes are highly stable. It is noteworthy that the capped 40-nt transcript was recovered in the supernatant fractions (Figure 5H, lanes 2, 4 and 6), suggesting that this transcript was released as a dead-end product. Non-polyadenylated transcripts of 1.3 kilo-nt were still associated with elongation complexes at the time point of 7 min (Figure 5I, lane 8), whereas polyadenylated transcripts were detected in the supernatant fraction at the time point of 11 min (lane 8). These results indicate that the sequential cap methylation reactions occur on pre-mRNAs associated with
The L protein can methylate the cap only when pre-mRNAs are associated with elongation complexes

To investigate whether concomitant RNA chain elongation is essential for cap methylation of pre-mRNAs, elongation complexes associated with cap-labeled pre-mRNAs of ~80 to ~200 nt were generated using the pulse-chase transcription system (chase time, 20 sec), precipitated by ultracentrifugation, and subsequently subjected to cap methylation with SAM in the presence of 4 NTPs or 3′-deoxy-NTPs (chain terminators) for 3 min (Figure 6A). In the presence of 4 NTPs, the precipitated elongation complexes were able to elongate pre-mRNAs (Figure 6B, lane 1) and methylate the cap core into the cap 1 (Figure 6C, lane 1), although part of them appeared to be inactivated during precipitation. In the presence of 3′-deoxy-NTPs, the elongation complexes were found to catalyze the cap methylation reactions without RNA chain elongation (Figure 6B and C, lane 2), indicating that continuous mRNA chain elongation is not necessary for cap methylation.

To examine whether L can post-transcriptionally methylate the cap on free pre-mRNAs, the cap-labeled pre-mRNAs of ~80 to ~200 nt were purified by denaturing PAGE to remove the genomic RNA (Figure 6D) and used as substrates for cap methylation with SAM for 3 min under the same conditions as those for transcription (Figure 6E). However, Rl was not able to methylate the externally added cap-labeled pre-mRNAs even in the presence of P with (lane 3) or without (lane 2) N-RNA. On the other hand, the cap of the pre-mRNAs was methylated by the G-N7-MTase activity of
Figure 6. L can methylate the cap only on pre-mRNAs associated with elongation complexes.  (A–C) Cap-labeled transcripts were synthesized in the presence of [α-32P]GDP and chased in the presence of excess cold GDP and NTPs for 20 sec as in Figure 5. After stopping the reactions, RNPs were precipitated through a glycerol cushion by ultracentrifugation, and subjected to cap methylation with SAM in the presence of 4 NTPs or 3′-deoxy-NTPs (chain terminators) for 3 min (A). Cap-labeled transcripts (B) and their Nase P1 digests (C) were analyzed by 20% urea–PAGE and PEI-cellulose TLC, respectively, followed by autoradiography. (D and E) Cap-labeled transcripts were synthesized as in panel A and purified by urea–PAGE to remove the VSV genome. Purified cap-labeled transcripts were re-analyzed by 20% urea–PAGE followed by autoradiography (D). Purified cap-labeled transcripts were subjected to cap methylation with the indicated factors in the presence of SAM for 3 min under the same conditions as those for mRNA chain elongation. Nase P1 digests of the products were analyzed as in panel C (E). Recombinant vaccinia virus capping enzyme (VacV rCE) was used as a positive control. (F and G) A cap-labeled 10-nt oligo-RNA with the VSV N mRNA-start sequence was incubated with native L-P complex, rL, or VacV rCE in the presence of SAM for 3 h under the reported conditions for the MTase activities associated with rL (50) (F). Nase P1 digests of the products were analyzed as in panel C (G).
vaccinia virus capping enzyme (see Supplementary Figure S1C; Figure 6E, lane 4).

Since Rahme et al. (50) reported that their L alone could sequentially methylate the cap core on a pre-mRNA of 10 nt or longer into the cap 1, we examined whether our native and recombinant L proteins can methylate the cap on a 10-nt chemically synthesized oligo-RNA with the VSV N mRNA-start sequence (GpppAACAGUAACUC) under the reported conditions (Figure 6F). Unexpectedly, 2.5 μg of a native L-P RdRp complex (see Supplementary Figure S1A, Figure 6G, lane 2), as well as 2 μg of our L (lane 3), was not able to methylate the cap even if incubated for 3 h, although the native L-P complex could methylate the cap on pre-mRNAs within 3 min during mRNA chain elongation (Supplementary Figure S9). In contrast, only 1 ng of the recombinant vaccinia virus capping enzyme was enough to catalyze G-N'-methylation of the externally added capped RNA substrate (Figure 6G, lane 4). Therefore, in contrast to the previous observation (50), we conclude that L can methylate pre-mRNAs only when associated with elongation complexes.

Discussion

The newly developed transcription assay with [α-32P]GDP allowed us to reveal the complete pathway of mRNA maturation by VSV L. Using transcription- [D714A (59)], capping-[HR-RH (32,38)] and methylation- [K179SR and the others (this study)] defective L mutants, we demonstrated that all the activities required for mature mRNA synthesis reside within L (Figures 1 and Supplementary Figure S2). We discovered that pre-mRNA capping and cap methylation reactions are sequentially catalyzed by the PRNTase and MTase domains of L at the early and middle stages of mRNA elongation, respectively (Figures 2–5). The technological achievements of this study include the development of the method to capture covalent complexes of L with pre-mRNA (L–pRNA intermediates for capping) formed during transcription (Figure 4). This study also provides strong evidence that L can methylate the cap only on pre-mRNAs associated with elongation complexes (Figures 5 and 6). Based on these results, we propose a new model for the entire process of co-transcriptional mRNA maturation by L proteins of rhabdoviruses (Figure 7A) and, by extension, other NNS RNA viruses.

Once L is associated with the transactivation domain of P to form a pre-initiation L-P complex (61), it can interact with N-RNA via the N-RNA binding domain of P and catalyze terminal de novo initiation to generate LeRNA (21,23,49,62,63). (Figure 7B). During this step, the tryptophan residue (W1167) on the dual functional priming-capping (PC) loop extended from the PRNTase domain plays a critical role in terminal de novo initiation (23). We have proposed that the indole side chain of the tryptophan residue π-stacks with the adenine ring of initiator ATP to stabilize the terminal initiation complex formed at the extreme 3'-end of the genome (Figure 7C, left (23,64). Immediately after terminal de novo initiation, the PC loop may be retracted from the RdRp active site cavity to open a putative transcript exit channel for elongating LeRNA (46).

After LeRNA synthesis, the same RdRp reinitiates de novo mRNA synthesis at the internal N gene-start sequence (Figure 7D). The tryptophan residue on the PC loop is no longer required for internal de novo initiation (23). During an early stage of mRNA chain elongation, the 5’-pppAACAG sequence of pre-mRNAs seems to be extruded from the RdRp domain of L to gain access to the PRNTase domain (Figure 7E). The PRNTase domain specifically recognizes the AACAG sequence on pre-mRNAs to form covalent L–pRNA intermediates (28,32). We confirmed that the capping-defective HR-RH mutant is not able to form the intermediates co-transcriptionally (Figure 4B). Note that the HR-RH mutant is known to generate 5’-triphosphorylated, prematurely terminated N transcript of 40 nt and polyadenylated full-length N mRNA as major and minor products, respectively (38). In addition to the active site amino acid residues (e.g. H1227 and R1228 in motif D, T1157 in motif B, W1188 in motif C) in the PRNTase domain (32,37), a conserved TxΨ motif (T1161-x11163) on the PC loop is also required for the intermediate formation (23). The biochemical data in this study indicate that at least 18 nt in pre-mRNA length is required for the formation of the intermediate followed by the transfer of pRNA to GDP (Figures 3 and 4).

To evaluate the length of pre-mRNA required to reach the PRNTase active site, we modeled the pre-mRNA based on a prior model of the putative L protein elongation complex [ModelArchive ID: ma-efc5b (46)], which provided the initial 12 nucleotides of the pre-mRNA (ModelArchive ID: ma-f4x26, Model 1) (Figures 7C (right), F, and Supplementary Figure S10). Through simple extension of the emerging 5’-helical RNA segment of the pre-mRNA, 18 nucleotides can reach from the polymerase active site to the pocket containing the PRNTase active site. Following energy minimization, the terminal triphosphate moiety is within 12.2 and 5.4 angstroms of H1227 and R1228, respectively. The large distance observed between H1227 and the RNA is due to the histidine side chain pointing in the opposing direction. With simple rotational adjustment of the side chain rotamers or for the triphosphate, these distances can be lowered to 8.5 and 4.6 or 3.8 (with the side chain flip) or <3 Å. These measurements are based on the built rigid helical RNA, though it is expected, when bound, that the 5’ triphosphates and the terminal nucleotide can be rearranged to engage the active site residues more closely during catalysis. Likewise, the L protein local structure would undergo some induced conformational changes. To approximate what a bound state might look like, we docked the cap structure to the L protein structure using methods similar to those in (16). The adenosine moiety was extracted, triphosphorylated, fused to nucleotides 2–18 of the pre-mRNA, and minimized (Model Archive ID: ma-yytxxd, Model 2). In this model (Supplementary Figure S11), R1228 is situated 2.83 angstroms from the terminal ATP. H1227 is posited 7.6 angstroms away, again with the rotamer flipped opposite the RNA. With a rotamer flip, the side chain sits 4.7 Å away. Based on these observations and our experimental data, it is reasonable to assume that 18 nucleotides is a sufficient number necessary to access the capping site and the minimum length for VSV pre-mRNA capping.

Since longer transcripts (21, 40, 49–52 nt) linked to L during transcription were also observed (Figure 4), the timing of the L–pRNA intermediate formation and/or the following release of the cap structure from the PRNTase active site seems to be flexible during elongation. In the presence of the saturating concentrations of 4 NTPs, L was found to cap pre-mRNAs of 21–23 nt (Figure 2). However, our attempts to capture intermediates linked to pre-mRNAs of 21–23 nt were unsuccessful due to the difficulty in efficiently labeling these short transcripts in the presence of the
Figure 7. A model of mRNA assembly by VSV L. (A) The pathway for sequential synthesis of LeRNA (light brown curved line) and \( N \) mRNA (light blue curved line) by VSV RdRp (L-P; pale blue) from the 3′-terminal leader (Le) region (brown thick line) and the \( S \) gene (dark blue thick line), respectively, on the genome is depicted according to the single-entry, stop-start transcription model. The cap core (red oval sphere), A-′-2′-O-methyl group (light green sphere), \( G-N^\prime \)-methyl group (green sphere), and poly(A) tail (magenta curved line) are added to pre-mRNAs during elongation. Bent arrows indicate transcription initiation signals. (B) (middle) induces a conformation rearrangement in L (left) with flexible three C-terminal domains [connector, yellow; MTase, orange; C-terminal (CT), brown] into a compact structure (pre-initiation complex). The N-terminal superdomain of L is composed of the N-terminal (gray), RdRp (fingers, blue; palm, red; thumb, green), bridge putative template-exit channel (46), sky blue), and PRNTase (pale green) (subdomains). The terminal de novo initiation complex is formed with initiator (ATP) and incoming (CTP) nucleotides on the 3′-UG-end of the genome to initiate LeRNA synthesis. A tryptophan residue (W1167, circled W) on the dual-functional priming-capping (PC) loop stabilizes the complex. Active site amino acid (AA) residues in the RdRp (D605 and D714) and PRNTase (H1227 and R1228) domains are labeled by circled letters. To show the inside of the RdRp domain, the fingers subdomain is depicted as a transparent object with a dashed outline (right). (C) A comparison of the VSV L protein in the terminal initiation (left, Model Archive ID: ma-54323, AAs 1228-1228) and elongation (right) complexes is shown. L residues in the priming-loop and flanking AAs are shown in cartoon model with key AA residues shown as sticks and labeled. D605 and D714 are found in the polymerase active site along with Mg and Mn ions (purple and darkened purple spheres, respectively). The model on the left shows the 3′-end of the initiation sequence (yellow sticks) with initiator (ATP) and incoming (CTP) nucleotides, while the model on the right shows an equivalent four nucleotides within the genomic template and the modeled 18-nt pre-mRNA (white sticks). The priming loop undergoes a large conformational change to accommodate egress of the growing pre-mRNA. (D) After releasing LeRNA, the internal de novo initiation complex is formed on the UU sequence in the \( N \) gene-start sequence to initiate \( N \) mRNA synthesis. (E) When the length of pre-mRNA reaches 18 nt or longer, its 5′-pppAACAG-end gains access to the PRNTase active site to form the cap core on pre-mRNA via the formation of the covalent L-pRNA intermediate followed by the pre-mRNA transfer to GTP. (F) Close-up view of the putative VSV L transcription elongation complex (Model Archive ID: ma-54323) is shown. Subdomains of L (shown in PyMol cartoon) are colored as in other panels, with residues previously identified as essential for efficient capping shown as magenta sticks. Nucleotide numbers and the 5′-triphosphate of the pre-mRNA are labeled. The 5′-triphosphate of nt-1 is adjacent to the PRNTase active site. (G and H) When the length of pre-mRNA reaches 40 nt or longer, the cap core is sequentially methylated using SAM as a methyl donor at the A-′-2′-O position (G) followed by the G-N^\prime position by the MTase domain to form the cap 1 (m^7GpppAm). The MTase domain has a 2′-O-MTase motif, called K-D-K-E catalytic tetrad (K1651–D1762–K1795–E1833). (I) When the transcribing RdRp domain encounters the U-rich gene-end sequence, a poly(A) tail is added to the 3′-end of full-length \( N \) mRNA by the polymerase stuttering mechanism to process it into a mature form.
saturating concentration (0.1 mM) of \([\alpha-^{32}\text{P}]\text{GTP}\) with a 100-fold lower specific radioactivity than that of the limiting concentration (1 \(\mu\text{M}\)) of \([\alpha-^{32}\text{P}]\text{GTP}\). Nevertheless, a possibility exists that the RdRp domain of L may remain paused at positions +22 to +24 within the N gene even in the presence of saturating levels of all 4 NTPs to provide the PRNTase domain with an opportunity for the L-pRNA intermediate formation. The VSV RdRp was reported to release large amounts of 5'-triposphorylated abortive transcripts (11–14 nt) (39–41). When small subsets of pre-mRNAs reach 18–23 nt long, they are accepted as substrates for the PRNTase domain (Figures 2–4). Once pre-mRNAs are capped, capped pre-mRNAs can be efficiently elongated in elongation complexes (Figures 2, 3 and 5) (37,38). It also should be noted that post-transcriptional capping of free pre-mRNAs with rL was observed at pH 5.8, but not at pH 8.0 which is optimal for in vitro transcription (28,38), suggesting that capping is tightly coupled with transcription elongation under the conditions for transcription.

During mRNA chain elongation, the cap on pre-mRNAs longer than 40 nt is sequentially methylated at the \(\text{A}_1-\text{O}\) position (Figure 7G) followed by the \(\text{G}-\text{N'}\) position (Figure 7H) by the MTase domain of L (Figures 3 and 5). The MTase domain could methylate the cap on pre-mRNAs associated with the elongation complex, but not on free pre-mRNAs, with or without elongation (Figure 6), suggesting that the formation of a closed loop between pre-mRNA and the RdRp and MTase domains may facilitate the MTase reactions. Finally, full-length mRNA is polyadenylated at its 3'-end by polymerase stalling on the poly(U)-stretch in the gene-end sequence (Figure 7I). Once 3'-polyadenylated mature mRNAs are released from elongation complexes, they no longer act as substrates for the MTase domain (Figure 5).

Our conclusions are at odds with the previous report by Tekes et al. (65) that the capping and two methylation reactions take place simultaneously on a 31-nt transcript, but not on a 30-nt transcript when the VSV RdRp was stalled/terminated on a genetically modified VSV genome by omitting one of NTPs during a 5-h incubation, although it remains unknown whether the 31-nt transcript is a precursor of mRNA or is capped and methylated co-transcriptionally. In addition, Rahme et al. (50) reported that a large amount (2 \(\mu\text{g}, \sim 8 \text{ pmol}\)) of rL alone can sequentially methylate approximately 10 fmol of the cap on the 10-nt VSV mRNA thereafter to the cap 1 during a 2-h incubation. However, our multiple attempts to detect such MTase activities of our rL as well as native L complex or RNPs using externally added capped RNA substrates have been unsuccessful, although our native and recombinant L proteins have been proven to be fully active in catalyzing the cap 1 formation within 3 min during transcription. Thus, our results strongly suggest that methylation of the cap on pre-mRNAs associated with L engaged in elongation complexes proceeds more efficiently by orders of magnitude versus those on free pre-mRNAs by an apo or preinitiation form of L.

Takanishi et al., 2024, this study demonstrates that VSV, a prototypic NNS RNA virus, mRNA modifications are catalyzed by the single-polypeptide RdRp L protein co-transcriptionally and proceed during mRNA chain elongation in the following order: (i) 5'-capping \(\rightarrow\) (ii) \(\text{N}_1-\text{O}-\text{methylation}\) \(\rightarrow\) (iii) \(\text{G}-\text{N'}-\text{methyla-}

**References**


