The 3'UTR of Pea Enation Mosaic Virus Contains Two T-Shaped, Ribosome-Binding Cap-Independent Translation Enhancers

Feng Gao¹, Wojciech K. Kasprzak², Christina Szarko¹, Bruce A. Shapiro³††, and Anne E. Simon¹†

¹Department of Cell Biology and Molecular Genetics
University of Maryland College Park
College Park, MD, 20742

²Basic Science Program,
Leidos Biomedical Research, Inc.
Frederick National Laboratory for Cancer Research,
Frederick, MD 21702

³Basic Research Laboratory
National Cancer Institute,
Frederick, MD 21702

†Corresponding Author
Phone: 301-405-8975
Fax: 301-314-7920
Email: simona@umd.edu

††Co-corresponding Author
Phone: 301-846-5536
Fax: 301-846-5598
Email: shapirbr@mail.nih.gov

JVI Accepts, published online ahead of print on 6 August 2014
J. Virol. doi:10.1128/JVI.01433-14
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
Abstract

Many plant viruses without 5′ caps or 3′ polyA tails contain 3′ proximal, cap-independent translation enhancers (3′CITEs) that bind to ribosomal subunits or translation factors thought to assist in ribosome recruitment. Most 3′CITEs participate in a long-distance kissing-loop interaction with a 5′ proximal hairpin to deliver ribosomal subunits to the 5′ end for translation initiation. *Pea enation mosaic virus* (PEMV) contains two adjacent 3′CITEs in the center of its 703-nt 3′UTR, the ribosome-binding, kissing-loop T-shaped structure (kl-TSS) and eIF4E-binding Panicum mosaic virus-like translation enhance (PTE). We now report that PEMV contains a third, independent 3′CITE located near the 3′ terminus. This 3′CITE is composed of three hairpins and two pseudoknots, similar to the TSS 3′CITE of Carmovirus *Turnip crinkle virus* (TCV). As with the TCV TSS, the PEMV 3′TSS is predicted to fold into a T-shaped structure that binds to 80S ribosomes and 60S ribosomal subunits. A small hairpin (kl-H) upstream of the 3′TSS contains an apical loop capable of forming a kissing-loop interaction with a 5′ proximal hairpin and is critical for accumulation of full-length PEMV in protoplasts.

Although the kl-H and 3′TSS are dispensable for translation of a reporter construct containing the complete PEMV 3′UTR in vitro, deleting the normally required kl-TSS and PTE 3′CITEs and placing the kl-H and 3′TSS proximal to the reporter termination codon restores translation to near WT levels. This suggests that PEMV requires three 3′CITEs for proper translation and that additional translation enhancers may have been missed if reporter constructs were used in
3’CITE identification.

**Importance**

The rapid life-cycle of viruses requires efficient translation of viral-encoded proteins. Many plant RNA viruses contain 3’ cap-independent translation enhancers (3’CITEs) to effectively compete with ongoing host translation. Since only single 3’CITEs have been identified for the vast majority of individual virus, it is widely accepted that this is sufficient for a virus’s translational needs. *Pea enation mosaic virus* possesses a ribosome-binding 3’CITE that can connect to the 5’ end through an RNA:RNA interaction and an adjacent eIF4E-binding 3’CITE. We report the identification of a third 3’CITE that binds weakly to ribosomes and requires an upstream hairpin to form a bridge between the 3’ and 5’ ends. Although both ribosome-binding 3’CITEs are critical for virus accumulation in vivo, only the CITE closest to the termination codon of a reporter ORF is active, suggesting that artificial constructs used for 3’CITE identification may underestimate the number of CITEs that participate in translation.
Introduction

Protein biosynthesis is principally regulated at the stage of translation initiation, allowing for rapid and efficient post-transcriptional control of gene expression (1-3). The 5' m7GpppN cap and 3' poly(A) tail at the termini of most eukaryotic mRNAs are required to recruit the ribosome with the assistance of numerous eukaryotic initiation factors (eIFs). The 5' cap structure is recognized by eIF4E, a subunit of the eukaryotic translation initiation factor complex eIF4F (4). eIF4G, another subunit of eIF4F, is the core scaffolding protein that simultaneously interacts with eIF4E and 3'terminal-bound poly(A)-binding protein to bridge the cap and poly(A) tail (5). Bridging the ends of an mRNA is generally thought to facilitate reinitiation by post-termination ribosomes (5, 6), although circularized polyribosomes on a mRNA template are also visible in the absence of a 5' cap and 3' poly(A) tail (7). The 43S ribosome preinitiation complex, which is recruited to the 5' end of mRNA via eIF4G and additional eIFs, transits the template in the 3' direction until recognizing an initiation codon in the proper context. After the 60S ribosomal subunit joins the preinitiation complex, the 80S ribosome is formed and translation initiation commences.

Many positive-strand RNA viruses do not possess a cap structure and/or a 3’ poly(A) tail but instead have evolved a variety of effective non-canonical mechanisms to directly recruit eIFs or ribosomes to the viral RNA for translation initiation (8). Most animal RNA viruses have replaced the 5' cap with a highly structured, cis-acting element known as an internal ribosome entry site (IRES), which is located in the 5'UTR and may overlap the initiation codon and extend into the translated ORF. Viral IRESs are capable of recruiting ribosomes directly or via the assistance of a subset of eIFs (9, 10). In contrast, most uncapped and non-polyadenylated plant
viruses contain cap-independent translation enhancers (CITEs) in their 3’UTRs that can extend into nearby coding regions (11, 12). Currently, 3’CITEs have been predominantly characterized in genomes of viruses throughout the family Tombusviridae, in the Luteovirus genus of the family Luteoviridae and in the genus Umbravirus (no family assignment). 3’CITEs adopt diverse secondary structures, such as I-shaped, Y-shaped, T-shaped, or radiate multiple helices from a central hub (12). Most 3’CITEs stimulate cap-independent translation by recruiting either eIFs or ribosomes directly and then re-locating the bound eIFs/ribosomes to the 5’ end via long-distance, kissing-loop interactions with a hairpin loop located in the 5’UTR or nearby coding region (13-15). Since nearly all of these 3’CITEs are functional only in cis, it is generally assumed that they assist in recycling ribosomes to the 5’ end following translation termination.

Several classes of 3’ CITEs have been shown to bind preferably to different components of the eIF4F complex to efficiently initiate translation in host cells. The Barley yellow dwarf virus (BYDV)-like translation element (BTE) binds to the eIF4G subunit of eIF4F and does not require eIF4E for translation enhancement in vitro (16). The Panicum mosaic virus-like translational enhancer (PTE) in Pea enation mosaic virus (PEMV) RNA2 binds to eIF4E for efficient translation enhancement in vitro (17). The Satellite tobacco necrosis virus (STNV) translation enhancer domain (TED), the Y-shaped 3’CITE of Carnation Italian ringspot virus (CIRV) and the I-shaped 3’ CITE of Maize necrotic streak virus (MNeSV) have been shown to bind to eIF4F (18-20).

3’CITEs that bind directly to ribosomes/ribosomal subunits were first discovered in the Carmovirus Turnip crinkle virus (TCV). The TCV 3’ CITE, consisting of three hairpins and two pseudoknots, forms a T-shaped structure (TSS) that binds to the P-site of 80S ribosomes and 60S subunits and binding is important for efficient translation enhancement (21). In contrast to most
3’CITEs, no long-distance RNA:RNA interaction connects the TCV TSS with the 5’ end, leading to a ribosome-bridging model whereby TSS-bound 60S ribosomal subunits join with 40S subunits bound to a pyrimidine-rich sequence in the 5’UTR to recycle 60S subunits and enhance translation (22). A similar set of hairpins and pseudoknots was also predicted for Carmoviruses Cardamine chlorotic fleck virus (CCFV) and Japanese iris necrosis ringspot virus (JINRV) (23), suggesting that this type of TSS 3’CITE may also exist in other viruses.

A second type of ribosome-binding 3’CITE was recently discovered in the central 3’UTR of PEMV RNA2. PEMV is composed of two taxonomically distinct, single-stranded, plus-sense viral RNAs, Luteovirus PEMV RNA1 and Umbravirus PEMV RNA2 (24). The 4252 nt genome of PEMV RNA2 lacks a 5’ cap and 3’ poly(A) tail and relies on PEMV RNA1-encoded capsid protein for encapsidation and transmission, but not infection of single cells (25). The carmovirus-like RNA-dependent RNA polymerase (RdRp) (94 kDa) encoded by PEMV RNA2 (referred to now as “PEMV” in this report) is synthesized following a -1 ribosomal frameshift that extends the length of p33, the product of ORF1 (25). Two overlapping proteins (p26 and p27) are expressed from ORF3 and ORF4 via at least one subgenomic RNA (26, 27). PEMV contains two adjacent 3’CITEs in the center of its 703-nt 3’UTR: eIF4E-binding PTE and the ribosome-binding, kissing-loop TSS (kl-TSS) (13, 28). Unlike PTEs in Carmoviruses (12), the PEMV PTE is not associated with any long-distance RNA:RNA interaction and instead likely functions to assist in ribosomes binding to the adjacent, upstream kl-TSS (28). The kl-TSS is a two hairpin, three-way branched element that forms a 3-D T-shaped structure similar to the three hairpin/two pseudoknot TCV TSS. Unlike the TCV TSS, the kl-TSS binds to 40S ribosomal subunits in addition to 60S and 80S ribosomes and can engage in a simultaneous long-distance kissing-loop interaction with a 5’ end coding sequence hairpin (13, 28). It has been proposed...
that the kl-TSS binds to ribosomal subunits following translation termination to facilitate
reinitiation at the 5’ end.

With the possible exception of the paired PTE/kl-TSS elements of PEMV and an isolate
of Melon necrotic spot virus (MNSV) that acquired an additional 3’CITE by recombination (29),
only single 3’CITEs have been discovered in 3’ regions of plant viruses, leading to the general
assumption that a single element is normally sufficient for all necessary translation enhancement.

We now report that PEMV contains a second T-shaped structure downstream from the kl-
TSS/PTE, located in the same position in the genome as the TCV TSS. The 3’ proximal PEMV
TSS (3’TSS) is composed of a similar combination of three hairpins and two pseudoknots as the
TCV TSS and also binds to 80S and 60S ribosomes, but with lower affinity. A hairpin upstream
of the PEMV 3’TSS (kl-H) can participate in a long-distance interaction with the same 5’
proximal coding region hairpin that is utilized by the kl-TSS in reporter constructs in vivo and in
vitro, and together they can effectively substitute for the kl-TSS/PTE if placed proximal to the
translation termination site. The kl-H and 3’TSS are critical for accumulation of PEMV in vivo
but the kl-H does not connect with the 5’ proximal coding region hairpin as it does in the reporter
construct. This suggests that the kl-H/3’TSS must enhance translation of a critical factor other
than the 5’ proximal ORF.

MATERIALS AND METHODS

In-line structure probing. In-line probing was performed as previously described (23). Briefly,
PEMV TSS RNA transcripts were 5’ end-labeled with [γ-32P]-ATP and purified by
electrophoresis through 5% denaturing polyacrylamide gels. End-labeled RNA was heated at
75°C and allowed to slow cool to room temperature. For in-line probing, 5 pmol of end-labeled
RNA was allowed to self-cleave at 25°C for 14 h in 1x in-line probing buffer (50 mM Tris-HCl pH 8.5, 20 mM MgCl₂). Reaction products were resolved on an 8% denaturing polyacrylamide gel. At least three independent in-line probing assays were performed and only reproducible differences are described.

**Accumulation of PEMV in protoplasts.** Protoplasts were prepared from callus cultures of *Arabidopsis thaliana* (ecotype Col-0) and were transformed with PEMV gRNA transcripts using polyethylene glycol as previously described (30). First-strand cDNA was synthesized from total RNA prepared from infected protoplasts using MMLV reverse transcriptase (Invitrogen) and subjected to Taqman® real-time PCR assays using a LightCyler® 480 real-time PCR system (Roche Applied Science), as previously described (13). Raw data was analyzed using LightCyler® 480 real-time PCR system software. Each sample was assayed in triplicate in every experiment and at least three independent experiments were performed.

**Purification of 80S ribosomes and 40S/60S ribosomal subunits from Arabidopsis thaliana protoplasts and filter binding assays.** Plant 80S ribosomes and 40S/60S ribosomal subunits were purified as previously described (31). Briefly, Arabidopsis protoplasts prepared from callus cultures were lysed, centrifuged and the resulting supernatants were subjected to centrifugation through a 25% glycerol cushion. Ribosome pellets were resuspended in storage buffer (50 mM HEPES-KOH pH 7.6, 5 mM Mg(CH₃COO)₂, 50 mM NH₄Cl, 25% glycerol, 1 mM DTT) and subjected to salt-washing to remove any associated tRNAs and translation factors, as described previously for yeast ribosomes (22). For purification of 40S/60S ribosomal subunits, purified
80S ribosomes were subjected to sucrose gradient centrifugation and 40S- or 60S-containing fractions applied to Amicon ultra (100k) columns (Millipore) for buffer exchange. Purified 40S/60S ribosomal subunits were stored in storage buffer (2 to 10 pmol/µl) at -80°C. Filter binding assays were performed as described previously (21).

SHAPE structure probing. SHAPE structure probing was performed essentially as previously described (32). Briefly, full length PEMV gRNA was synthesized in vitro using T7 RNA polymerase and subjected to phenol-chloroform extraction and ethanol precipitation. RNA was denatured at 95°C for 3 min and snap-cooled on ice for 2 min. The RNA was then folded in SHAPE folding buffer (80 mM Tris-HCl pH 8.0, 11 mM Mg(CH3COO)2, 160 mM NH4Cl) at 37°C for 20 min. The folded RNA was divided into two aliquots and each aliquot was incubated with either 15 mM N-methylisatoic anhydride (NMIA) or DMSO at 37°C for 40 min. Oligonucleotides labeled at their 5’ ends with [γ-32P]-ATP were used for primer extension reactions with SuperScript® III Reverse Transcriptase (Invitrogen) as previously described (32). Six oligonucleotides (positions 3595-3614, 3830-3849, 4018-4037, 4121-4140, 4235-4252 and 4302-4318) were used to probe the structure of the 3’UTR of PEMV. To provide structural data for the 3’ terminus of the 3’UTR, 214 additional nucleotides from the pUC19 vector located downstream of the 3’ end were included in the in vitro synthesis of full length PEMV gRNA to generate a 4466 nt RNA (used only to structurally probe the 3’ terminus). Reaction products were then subjected to electrophoresis through 8% denaturing polyacrylamide gels and radioactive bands were visualized with a phosphorimager. Intensity of individual bands were inspected visually and assigned values of weak/moderate and moderate strong/strong. The nucleotide positions were identified by reference to the sequencing ladder generated from SHAPE structure probing.
unmodified RNA by Sanger methods. RNA secondary structural maps were generated from structure probing results and the best fitting Mfold predictions (33).

**In vitro translation and trans inhibition assays.** In vitro translation and trans inhibition assays were performed as previously described (28). Briefly, 3 pmol of in vitro synthesized RNA transcripts from designated translation reporter constructs were used for a 15 μl translation reaction using wheat germ extracts (WGE) (Promega) according to the manufacturer’s instructions. Luciferase activity was measured using a luciferase assay reporter system (Promega) and a Modulus microplate multimode reader (Turner BioSystems). For trans inhibition assays, 30 pmol of competitor RNAs were added to the translation reactions prior to incubation.

**In vivo translation assays.** In vivo translation assays were performed as previously described (13). Briefly, 30 μg of uncapped parental and mutant Firefly luciferase transcripts and 10 μg of Renilla luciferase control transcripts were introduced into 5 x 10^6 A. thaliana (ecotype Col-0) protoplasts using polyethylene glycol-mediated transformation. Luciferase activity was measured 18-h later using a luciferase assay reporter system (Promega) and a Modulus microplate multimode reader (Turner BioSystems).

**Molecular Modeling.** Three-dimensional models are based on the 3’TSS SHAPE-probed secondary structure information for the 82-nt fragment spanning positions 4128 through 4209.
Two models of the 3-D structure were built; one with a 3-bp pseudoknot $\psi_3$ (residues 4128, 4129, 4130 pairing with 4140, 4139, 4138, respectively) and one with 4 bp (adding pair 4131-4137). RNAComposer (34, 35) web server (http://euterpe.man.poznan.pl/Home) was first used to build the preliminary models and RNA2D3D (36) was used to edit certain structural collisions in those models. The corrected models (two variants for 3-bp $\psi_3$ and one for 4-bp $\psi_3$) were energy minimized, then placed in explicit solvent boxes, equilibrated and subjected to molecular dynamics simulations (MD). Since the initial models and a standard MD protocol showed potential problems with the stability of $\psi_3$, several MD protocols with some restraints meant to temporarily stabilize the base pairs in $\psi_3$ were run (see MD Simulation Protocol for details). Multiple 100 ns long MD simulations were conducted.

Molecular Dynamics Simulation Protocol. Amber 12 with the ff10 Cornell force field for RNA was used to perform MD simulations, using the Particle Mesh Ewald (PME) summation method to calculate the electrostatic interactions (37-39). The energy-minimized 3′TSS models (82 nt, 2,625 atoms) were solvated in a TIP3P water box with a buffer zone of 15.0 Å. The RNA was neutralized with 81 Na$^+$ ions. Additional Na$^+/\text{Cl}^-$ ion pairs were added to reach a relative salt concentration of 0.1 M (varying from 64 to 69, depending on the 3′TSS model used). The total sizes of the systems ranged from 90,560 to 97,272 atoms. A multi-step equilibration protocol was used that started with solvent equilibration (minimization, heating and short dynamics stages), while the RNA that was being subjected to motion restraints (holding) decreased at every stage. The last phase of the equilibration was performed for 2.0 ns without and with (separate equilibrations) distance restraints placed on the hydrogen bonds of the 3-bp or 4-bp $\psi_3$ (depending on the initial model). The entire system was equilibrated at 300K using the Berendsen thermostat (40). Other parameters included use of periodic boundary condition, a cut-
off of 9Å for the non-bonded interactions and SHAKE applied to all hydrogen bonds in the system. Pressure was maintained at 1.0 Pa utilizing the Berendsen algorithm (40). Production runs of 100 ns duration (excluding the equilibration phase) were conducted without and with the $\psi_3$ base-pair restraints lasting 10 and 25 ns of the total production run.

**Results**

The PEMV 3’ proximal region contains a series of hairpins and pseudoknots similar to those present in several Carmoviruses. Examination of the 3’ terminus of PEMV revealed sequences that could potentially fold into a 3’ terminal hairpin (Pr), an upstream long linker region (L1), and an adjacent hairpin (H5). This configuration of elements is very similar to those found in identical locations in all members of the Carmovirus genus (23). Complementary sequences in the PEMV H5 apical loop (5’ GGGC) and at the 3’ terminus (3’ OH-CCCG) also suggested that a pseudoknot ($\psi_1$) connects these two motifs, similar to pseudoknots found in all members of the Tombusviridae (41, 42). Three Carmoviruses, TCV, Cardamine chlorotic fleck virus (CCFV), and Japanese iris necrosis ringspot virus (JINRV) also contain juxtaposed hairpins H4a and H4b just upstream of H5, with two additional pseudoknots ($\psi_2$ and $\psi_3$) connecting their apical loops with nearby single-stranded sequences (23) (Fig. 1A). Disrupting any of these hairpins and pseudoknots in TCV results in significantly reduced levels of virus accumulation in vivo (23). Visual inspection of sequences upstream of H5 in PEMV revealed that a similar configuration of H4a, H4b, $\psi_2$, and $\psi_3$ was also a possibility (Fig. 1B). Comparable TCV-like structural organization was also possible for Umbraviruses *Carrot mottle virus* (CMoV) and *Tobacco bushy top virus* (TBTV) (Fig. 1C and D). In contrast, *Carrot mottle mimic*
virus (CMoMV) may only contain H5, Pr and ψ1, and all elements were absent from the 3’ end of Groundnut rosette virus (data not shown).

A 170 nt 3’ terminal PEMV fragment was subjected to in-line probing to determine if an experimentally-derived structure was consistent with H4a, H4b, H5, ψ2 and ψ3. In-line probing interrogates the flexibility of each nucleotide within an RNA fragment, as only flexible (i.e., non-base paired) nucleotides can position their 2’ oxygen, backbone phosphorus and adjacent 5’ oxygen in the in-line configuration that is needed for nucleophilic attack of the 2’ oxygen on the phosphorus and subsequent backbone cleavage. The level of cleavage at each nucleotide in a radiolabeled fragment is detected following denaturing gel electrophoresis and autoradiography, and is directly proportional to the flexibility of the nucleotide within an averaged population of structures.

The cleavage pattern for the 3’ region of PEMV is presented in Fig. 2B and typical in-line cleavage gels are shown in Fig. 2A. The flexibility profiles for Pr and H4b residues were consistent with hairpins in these locations, with inflexible stem residues and highly flexible loop residues. In the apical loop of H4b, residues 4161GCA showed reduced cleavages compared with the remaining apical loop residues, consistent with their participation in ψ2. L1 was also highly flexible with the exception of putative ψ2 residues 4207UGC. The flexibility of H5 residues were mainly consistent with the predicted hairpin. All apical loop residues were highly flexible with the exception of 4183GGG, which are proposed to participate in ψ1. The residues on the 5’ side of the lower H5 stem were weakly flexible, similar to the pattern found for the same H5 residues in a fragment containing the TCV 3’ region that was subjected to in-line probing (23). Unlike H5 and H4b, the flexibility pattern for PEMV H4a within this fragment was not consistent with a hairpin, with 3’ side "stem" residues exhibiting considerable flexibility. These results suggest
that this PEMV fragment contains H4b, H5, Pr, ψ1 and ψ2.

Since it is possible that RNA structures exist in the full-length virus in vivo that are not present or discernible in the averaged population of in vitro folded molecules (and vice versa), single and compensatory mutations were generated in each putative hairpin and pseudoknot in full-length PEMV and virus levels assayed for in Arabidopsis thaliana protoplasts using qPCR (Fig. 3). Altering 4150CG in the H4b stem to GC (construct M1) reduced viral RNA levels to 27% of WT and combining these mutations with compensatory alterations across the stem (construct M2) restored accumulation to 82% of WT (Fig. 3B). The C to G transversion at position 4162 (c4162G) in the H4b apical loop that was designed to disrupt ψ2 also reduced accumulation to 27% of WT. Disruption of the putative partner residue in the L1 linker region (G4308C) had a similar negative effect on virus accumulation (27% of WT), whereas the two compensatory mutations together restored PEMV accumulation to 94% of WT. These results support the presence of H4b and ψ2 and their importance for PEMV accumulation in vivo.

Disruption of either the 5’ or 3’ sides of the H5 stem (M4 and M5) reduced accumulation to 34% and 21% of WT, respectively. The compensatory alteration that combined these two mutations (M6), however, did not restore PEMV accumulation. Likewise, altering the stem of H4a (construct M7) reduced accumulation to 24% of WT, which was not restored by inclusion of the compensatory mutations (construct M8) (Fig. 3B). In contrast, disrupting ψ3 by altering either the loop of H4a (g4138C) or the partner residue upstream of H4a (c4130G) reduced accumulation to 30 and 19% of WT, respectively, whereas the two compensatory mutations together restored accumulation to 83% of WT.

The inability of H4a and H5 compensatory mutations to restore accumulation could reflect either (1) the absence of these structures; (2) a requirement for specific sequences within
the stems; or (3) the stem-disrupting alterations caused stable, inadvertent structural changes that were not corrected by the compensatory mutations. To distinguish among these possibilities, in-line probing was conducted for the 3’ fragment of PEMV in the presence of the single and compensatory stem alterations in H4b, H5 and H4a (Fig. 3D). Disrupting the H4b stem (M1) caused significant changes in the flexibility of residues throughout the H4a, H4b and L1 region (Fig. 3D, compare lane 3 with lane 4), suggesting that this hairpin is part of an integrated 3-D structure. The compensatory mutations in the stem of H4b (M2) restored the WT flexibility pattern, indicating that the H4b structure refolded properly. Altering either side of the upper stem of H5 (M4, M5) disrupted the structure in that region (Fig. 3D, lanes 6 and 7). The compensatory alterations (M6) restored the structure of H5 to its WT configuration (Fig. 3D, lane 8), indicating that a hairpin is present at this location. Therefore, the compensatory M6 alterations that did not restore virus accumulation may either disrupt a necessary sequence requirement or affect the structure of the hairpin within the full-length virus.

In-line probing of the fragment containing H4a alteration M7 revealed significant loss of flexibility in the putative apical loop of H4a and the apical loop of H4b, as well as alterations in H5, and L1. Inclusion of H4a alterations designed to be compensatory (M8) did not restore the WT structure, but rather caused additional flexibility changes in H4a and H4b. One possible explanation is that the M7 alterations created conditions allowing for inadvertent pairing between the apical loops of H4a (5’ CUAG) and H4b (3’GAUU), and this pairing was maintained in M8. Although these results do not demonstrate the existence of H4a, the presence of all other elements that together with H4a form a TSS in TCV (H4b, H5, ψ2, and ψ3), along with conservation of a hairpin at the H4a location in CMoV and TBTV, suggest that H4a may be present in PEMV.
SHAPE-derived flexibility profile of H4a residues in full-length PEMV is consistent with a hairpin in this location. In a further attempt to provide evidence for the presence of H4a in PEMV, selective 2’OH acylation analyzed by primer extension (SHAPE) was used to probe the structure of the 3’UTR of PEMV within full-length gRNA. To probe the structure of the 3’ terminal region only, 214 additional plasmid-derived nucleotides were present at the 3’ end of full-length PEMV gRNA. SHAPE interrogates the flexibility of residues in full-length RNAs by their level of conjugation to NMIA, which restricts primer extension by reverse transcriptase. By prohibiting Mfold (33) from pairing moderate and strong NMIA-reactive nucleotides, a secondary structure emerged that was a good fit with most of the SHAPE structural data (Fig. 4A). The structure contains several known elements within the 3’UTR, including the kl-TSS and adjacent PTE 3’CITEs, and a hairpin at position 3635 of unknown function that is conserved in all Umbraviruses (A.E. Simon, unpublished). The region of the structure least supported by the SHAPE data lies just upstream of this conserved hairpin (positions 3612 to 3636). Originally, Mfold predictions suggested that the conserved hairpin lower stem should contain three additional basepairs (A:U, C:G, C:G). The poor support for these pairings by the SHAPE data could be explained by the presence of an H-type pseudoknot connecting the loop of the adjacent hairpin with at least two of these residues, a structure predicted by pknotsRG (44).

At the 3’ end of PEMV, the structures of Pr, H5 and H4b are compatible with the SHAPE-derived data, with flexible bases restricted to apical loops. Residues in the 5’ side lower stem of H5 were no longer weakly flexible according to SHAPE (Fig. 4B), suggesting that the weak cleavages at these residues detected by in-line probing were a function of either the different procedure, or that in-line probing used a shorter fragment that affected the structure of...
the hairpin. A major difference was found in the flexibility of residues within L1, most of which were no longer flexible in the full-length gRNA, suggesting a possible interaction between these residues and upstream sequences not present in the fragment used for in-line probing. Searches of upstream sequences for a likely pairing partner, however, did not reveal any obvious candidates.

Importantly, the flexibility pattern of residues within H4a as assayed by SHAPE in full-length PEMV was now consistent with a hairpin at this location. Unlike the in-line probing data, the upper three residues in the 3' side of the H4a stem were not flexible in the SHAPE assay, suggesting that these residues are paired across the stem. The SHAPE data also suggests that the putative base pair $\psi_{33}G:U_{146}$ at the base of the stem may not be stable and that $\psi_3$ residue $A_{4127}$ is not pairing with $U_{4141}$ in the averaged population.

In the SHAPE-derived structure of the 3' region of PEMV, the 3'UTR together with 73 upstream residues were predicted to form a fold-back RNA domain. Since the parental reporter construct (5'89+3U) used to identify the kl-TSS as a 3'CITE contained only the precise 3'UTR downstream of the firefly luciferase reporter ORF, the possibility existed that adding 73 nt of upstream coding sequence would enhance translation by improving the formation of important structures in the 3'UTR. However, addition of these 73 residues to the reporter construct had no significant effect on luciferase activity (data not shown), suggesting that the entire domain is not necessary for proper translation enhancement in the reporter construct.

**The PEMV 3' region is predicted to form a TSS.** In TCV, H4a, H4b, H5, $\psi_3$ and $\psi_2$ fold into a tRNA-shaped structure (TSS) that binds to 80S ribosomes and 60S ribosomal subunits and functions as a 3'CITE (22, 23). To determine if the region flanked by $\psi_3$ and $\psi_2$ in PEMV also
adopts a 3-D T-shaped structure, molecular models were built for this region of PEMV and their stability evaluated with the aid of MD simulations (see Materials and Methods for details). The smaller size of the PEMV H4a hairpin loop (relative to that of TCV’s) made it difficult to accommodate a 4 bp $\psi_3$. Therefore a 3 bp $\psi_3$ model was also constructed. In MD simulations, the two variants of the latter model converged to nearly identical T-shaped structures reminiscent of the TCV TSS and similar in size to a tRNA, as illustrated for one model in Fig. 5A and 5C. All helices (H4a, H4b and H5) and $\psi_2$ were stable in all MD protocols employed. MD simulations also revealed a lower stability for $\psi_3$ base-pairs than was previously observed for the TCV TSS (23). An MD protocol employing $\psi_3$ base pair restraints in the equilibration phase and during the first 25 ns of the production run was necessary to maintain a stable 3 bp $\psi_3$ in the remainder of the 100 ns simulation (Fig. 5B). Restraints applied for only the first 10 ns of the production MD yielded an average structure maintaining the second and third base pairs of $\psi_3$, however the first two base pairs opened by the end of the run. The PEMV TSS model with 4 bp in $\psi_3$ also had problems maintaining all base pairs despite a 25 ns-long restrained phase of the production MD run (the first and second base pairs opened intermittently after ~72 ns of the run; data not shown). While it is possible that the quality of the initial models (built from a database of relatively large fragments fit together) affected the MD results, the long equilibration protocol and the extended stabilization with base-pair restraints applied to $\psi_3$ should have overcome any initial structural strain. Considering that no special measures were necessary to maintain the stability of the TCV $\psi_3$, these results suggest that PEMV $\psi_3$ is less stable. This structural weakness may be contributing to the experimentally observed weaker ribosome binding relative to the TCV TSS (see below).
The PEMV 3’TSS binds to 80S ribosomes and 60S ribosomal subunits. TCV TSS binding to 80S ribosomes and 60S ribosomal subunits is important for its function as a 3’CITE (22). The kl-TSS, which also binds to 80S ribosomes and 60S subunits, although binding is strongest to 60S subunits (13). To determine if the 3’TSS of PEMV is a ribosome-binding element, filter binding assays were conducted using Arabidopsis ribosomes purified from callus-derived protoplasts (31). The initial fragment used in the assay (S1) extended from just before ψ3 to just after ψ2 (Fig. 6A). Fragment S1 bound 80S ribosomes with a Kd = 2.4 μM, which is 5-fold weaker than the TCV TSS and PEMV kl-TSS (Fig. 6B). To determine if a larger fragment might improve binding (i.e., might better support the TSS structure), the fragment was extended by 7-nt on the 5’ side and 6-nt on the 3’ side. This fragment (S2) was a weaker template for 80S ribosome binding (Kd = 3.1 μM). Using a larger fragment that was previously used for in-line probing (S3) further weakened binding (Kd = 3.6 μM).

We also subjected all three PEMV fragments to 60S subunit binding (Fig. 6B). Binding to 60S by fragments S1 and S2 was slightly stronger compared with 80S (Kd = 1.8 and 2.0 μM, respectively), as was previously found for the TCV TSS and PEMV kl-TSS. Fragment S3 binding to 60S was very weak (Kd = 5.4 μM), suggesting that the TSS structure (i.e., H4a or ψ3) may not be properly forming.

Since PEMV fragment S1 achieved the best binding to 80S and 60S, this fragment was also subjected to 40S filter binding. Similar to the TCV TSS, fragment S1 was unable to bind specifically to 40S subunits. Altogether, these results suggest that the PEMV 3’TSS can bind to 80S and 60S ribosomes, similar to the TCV TSS. However, reduced binding compared with the TCV TSS could reflect either weaker binding by the element, a need for additional factors to
bind more efficiently, or that the 3’TSS structure is not forming properly within the truncated
fragments, as suggested by the in-line probing data for H4a, and the MD simulation data for ψ3.

The 3’TSS and upstream hairpin kl-H comprise a functional 3’CITE. In PEMV, the
adjacent kl-TSS and PTE are 3’CITEs thought to function together to efficiently attract
ribosomes to the kl-TSS (28). As described above, the kl-TSS can bind to ribosomes and
simultaneously engage in a long-distance kissing-loop interaction between its 5’ side hairpin
(3H1) and a stable coding region hairpin (5H2) near the gRNA 5’ end (28). Major questions
therefore are whether the PEMV 3’TSS functions as a third 3’CITE and if so, whether an
additional kissing-loop interaction connects the 3’TSS with the 5’ end. It should be noted that
lack of a specific connection between the 3’TSS and the 5’ end would not abrogate the
hypothesis that the 3’TSS is an independent 3’CITE since no discernible RNA:RNA interaction
connects the TCV TSS with its 5’ end (22).

To begin answering these questions, we searched the PEMV 3’UTR for a second
sequence that might putatively engage in a kissing-loop interaction with either of the hairpins in
the 5’89 nt of PEMV (5H2 or upstream hairpin 5H1; Fig. 7A) or other nearby downstream
sequences. One possibility, the apical loop of hairpin kl-H, contains the same five kissing-loop
residues (UCGCC) found in the apical loop of kl-TSS hairpin 3H1, the hairpin that interacts with
5H2 (Fig. 7A). Discounting the structures branching off of the central “backbone” of the 3’UTR,
the kl-H is located approximately 30-nt positions upstream of the 3’TSS (Fig. 4A), suggesting
that the kl-H might also be capable of engaging in a long-distance interaction with 5H2 to
position the 3’TSS (and bound ribosome/ribosomal subunit) proximal to the gRNA 5’ end (Fig.
7A).
Using a reporter construct containing firefly luciferase flanked by the 5’ 89-nt and 3’UTR of PEMV (5’89+3U), we previously reported that disrupting the kissing-loop interaction between 3H1 and 5H2 reduced translation in protoplasts and wheat germ extracts (WGE) by at least 4-fold whereas compensatory mutations restored translation to greater than WT levels (13, 28) (Fig. 7 A and B). To determine if the kl-H contributes to translation, mutations were generated in the kl-H loop in 5’89+3U that would disrupt any kissing-loop interaction with 5H2 (Fig. 7A).

Translation of 5’89+3U containing this alteration (kl-Hm) in WGE was slightly enhanced compared with the parental construct (116% of WT) (Fig. 7B), suggesting that any interaction between kl-H and 5H2 is not contributing positively to translation. Mutations in H4b of the 3’TSS that disrupt and reform the hairpin structure (M1 and M2 alterations in the H4b stem) produced only a 15% and 20% reduction in luciferase activity (Fig. 7B). These results suggest that neither the kl-H nor the 3’TSS are important for translation of the parental reporter construct in WGE.

In 5’89+3U, the kl-TSS/PTE are located spatially proximal to the 3’ end of the luciferase ORF (see Fig. 4A). If the purpose of a TSS is to recycle ribosomes back to the 5’ end following translation termination, then proximity to where ribosomes terminate translation may be an important factor for TSS function. To determine if the kl-H and 3’TSS might function as translational enhancers if they are located proximal to the 3’ end of the luciferase ORF, the sequence between the 5’ end of the 3’UTR and the kl-H was deleted, generating construct 5’89+3Umini. In addition to placing the kl-H and 3’TSS proximal to the luciferase ORF termination site, this deletion eliminated both the kl-TSS and PTE, which are critical elements for translation of 5’89+3U. Luciferase activity of 5’89+3Umini was 84% of 5’89+3U levels in WGE (Fig. 7D), indicating that the kl-TSS/PTE were no longer needed for translation.
Mutations in the kl-H (kl-Hm) and 3’TSS (M1) that previously had little effect on translation of 5’89+3U now reduced translation of 5’89+3Umini in WGE by 78% and 65%, respectively. The compensatory mutations in H4b (M2) enhanced translation to 84% of 5’89+3Umini levels, strongly suggesting that the 3’TSS is being utilized as a translation enhancer. Mutation in the 5H2 loop in 5’89-U3mini reduced translation to 18% of 5’89+3Umini levels, suggesting that a long-distance interaction with this hairpin is still important in 5’89-U3mini. Combining the mutations in 5H2 and kl-H that were designed to be compensatory restored translation to 84% of 5’89-U3mini levels, strongly suggesting that a kissing-loop interaction was forming between 5H2 and kl-H in WGE.

To confirm these results using a more natural in vivo system, Arabidopsis protoplasts were transformed with RNA transcripts of parental (5’89+3U and 5’89+3Umini) and mutant constructs and luciferase activity was measured 18-h later. As in WGE, mutations disrupting the kl-H or 3’TSS had no significant effect on translation of 5’89+3U (Fig. 8A). In contrast, mutating the kl-H in 5’89+3Umini reduced translation to 28% of the parental construct (Fig. 8B). When the kl-H mutations were combined with 5H2 mutations that restored the putative long-distance RNA:RNA interaction, luciferase activity increased to 90% of the parental construct. Disrupting the 3’TSS by mutating H4b (construct M1) reduced luciferase activity to similar levels as the kl-H mutation (to 27% of the parental construct) and the compensatory H4b M2 mutations increased luciferase activity to 76% of the parental construct (Fig. 8B). All together, these results strongly suggest that when placed proximal to where ribosomes are terminating translation, the 3’TSS and kl-H function to enhance translation in WGE and in protoplasts in the absence of the kl-TSS and PTE. In addition, a long-distance RNA:RNA interaction with the 5’ end is likely important for kl-H function in enhancing translation.
The kl-H is important for PEMV accumulation in vivo but not for interaction with the 5' end of the genome. As shown in Fig. 2, mutations in 3'TSS hairpins and pseudoknots strongly reduced accumulation of PEMV in vivo. To determine if the kl-H also impacts virus accumulation in vivo, a single base alteration was generated in the kl-H loop (kl-Hm1; Fig. 9A) in full-length PEMV gRNA and virus levels were assayed for in protoplasts. kl-Hm1 reduced PEMV viral RNA to near background (PEMV-GDD) levels (Fig. 9B), indicating that this hairpin is essential for PEMV viability in protoplasts.

We previously demonstrated that single base alterations in hairpin loops that disrupted the kl-TSS/5H2 interaction in PEMV gRNA reduced accumulation to background levels, and the mutations together that were designed to be compensatory restored accumulation to 88% of wt (13). Given this previous result, it was unlikely that the critical function of the kl-H was to also interact with 5H2 in the viral gRNA, since this interaction would be disrupted in the virus containing the kl-TSS/5H2 compensatory alterations. In support of this hypothesis, PEMV containing kl-Hm1 and 5H2m1, which would disrupt the kl-TSS/5H2 interaction but allow the kl-H/5H2 interaction (Fig. 9A) accumulated at background levels (Fig. 9B). Furthermore, virus with all three mutations that would restore the kl-TSS/5H2 and kl-H/5H2 interactions also accumulated at near background levels (Fig. 9B). These results strongly suggest that the 3’CITE activity of the kl-H/3’TSS supports translation of a factor other than ORF1.

The kl-H and 3’TSS inhibit translation in trans. Addition of a 10-fold molar excess of the entire 3’UTR or a fragment containing only the kl-TSS and PTE to WGE programmed with 5’89+3U reduced luciferase activity by 75% compared with levels obtained in the absence of any
added fragments (28). The kl-TSS alone reduced translation by 57% and the PTE by 15%. To determine if the kl-H and 3’TSS could also inhibit 5’89+3U translation when added in trans, fragments containing either the kl-H or 3’TSS or both were added in a 10-fold molar excess to WGE programmed with 5’89+3U (Fig. 10A). Addition of transcripts containing both elements reduced translation by 66%, whereas the individual kl-H and 3’TSS reduced translation by 52% and 33%, respectively (Fig. 10). This reduction in translation mediated by the 3’TSS was also negated when the fragment contained the H4b M1 mutations. These results suggest that the kl-H is nearly as inhibitory as the kl-TSS alone, possibly due to interference with the cis long-distance kissing-loop interaction. The 3’TSS inhibited translation approximately 2-fold better than the PTE alone but only 50% as effective as was previously found for the TCV TSS (28). The ability to trans-inhibit translation of 5’89+3U supports the hypothesis that the kl-H and 3’TSS also function as elements important for translation of PEMV.

Discussion

Termination of translation in eukaryotes occurs when a translating ribosome encounters a stop codon in the A-site. A ternary complex containing release factors and GTP binds to the vacant A-site through tRNA mimicry, prompting GTP hydrolysis and binding of the ATPase ABCE1/Rli1p. After peptide release, ATP hydrolysis triggers ribosome subunit dissociation (45, 46), leading to two presumptive fates: i. Separated subunits diffuse into the cytosol and combine with initiation factors for reinitiation on the same or a different template (45, 47); and ii. inefficient subunit dissociation by ABCE1/Rli1p allows ribosomes to continue scanning through the 3’UTR until release by rescue factors DoM24 and Hbs1 (48). In the latter scenario, continued association with the 3’UTR might favor reinitiation on the same template due to
canonical bridges between the poly(A) tail and the 5’ cap. Our discovery of T-shaped, ribosome-binding elements in the 3’UTR of TCV (TSS) and PEMV (kl-TSS) suggest an additional fate for terminating ribosomes: iii. ribosomes or separated subunits remain with the template by interacting with internal ribosome-binding structures located proximal to the termination site (13, 21, 22). In the case of the kl-TSS, ribosomes/subunits are then transferred to the 5’ end through the kl-TSS’s long-distance kissing-loop interaction that mimics the closed loop topology of canonical mRNAs (28). Maintaining template association with ribosomal subunits following translation termination should enhance protein synthesis by increasing the rate of re-initiation, as was recently shown for a 40S binding element in the 3’UTR of Hepatitis C virus (49), which would be positioned proximal to the 5’ IRES through nearby long-distance RNA:RNA interactions (50).

Although efficient 3’CITEs had been identified for PEMV (the kl-TSS and adjacent PTE), visual inspection of the virus’s 3’ terminal region suggested the presence of hairpins and pseudoknots (H4a, H4b, H5, ψ2, ψ3) that in TCV form a well-studied TSS (23). These structures in PEMV were located upstream of a 3’ terminal hairpin (the Pr) and extended linker sequence, elements also found in similar locations in TCV and other Carmoviruses. A PEMV Pr hairpin was consistent with both in-line probing and SHAPE structure mapping data (Figs. 2 and 4). As with the Pr of TCV and most other Carmoviruses, the PEMV Pr apical loop contains sequence that is complementary with a bulged loop in a hairpin just downstream from the ribosome-recoding site at the end of the p33 ORF, an interaction that is required for ribosome readthrough in TCV or -1 frameshifting in PEMV (51). ψ1 connecting the 3’ terminal 4 residues and the loop of hairpin H5 is conserved in all Carmoviruses and Tombusviruses (52), and both SHAPE of the full-length gRNA and in-line probing of a 3’ fragment suggests that this pseudoknot also forms
in PEMV (Figs. 2 and 4). The presence of H5 was confirmed structurally by in-line probing and SHAPE, and mutations disrupting the H5 upper stem in full-length gRNA reduced virus accumulation in protoplasts. H5 is also conserved in all Carmoviruses (53), and is critical for accumulation of a non-translated TCV satellite RNA, satC (54). TCV H5 has been proposed to function as an RdRp chaperone (55) independent of its role in the TCV TSS (23) and the analogous hairpin in Tombusvirus *Tomato bushy stunt virus* binds to viral replicase proteins and has been proposed to also function as a replication repressor (41, 53). The existence and importance of H4b in PEMV was supported both genetically and structurally, and single and compensatory mutations also indicate the presence and importance of ψ2 and ψ3 (Fig. 3). H4a was consistent with the SHAPE-derived structural data (Fig. 4A) but not in-line probing of a 3’ fragment (Fig. 2), and mutations designed to be compensatory did not restore the WT structure in the fragment as assayed by in-line probing (Fig. 3). However, the conservation of all these TCV-like 3’ elements in two other Umbraviruses (Fig. 1) suggests that all are likely present in the PEMV gRNA and altering any of these elements reduced PEMV accumulation, indicating that they are important for PEMV viability.

PEMV H4a, H4b, H5, ψ2, ψ3 combine to fold into a T-shaped structure (3’TSS) with structural similarities to the TCV TSS and PEMV kl-TSS (Fig. 5C). There was no identifiable structural similarity with a recently determined high resolution structure for the Cricket paralysis virus tRNA-like element that functions as an IRES (56). Although the model and SHAPE data are mainly consistent, we did note that the lower base-pair in H4a (4133G:U4146) was stable in the MD simulations, but the residues showed flexibility by SHAPE (Fig. 4). Similar to the TCV TSS, the PEMV 3’TSS binds to 80S ribosomes and 60S subunits but not to 40S subunits, although binding was about 4-fold weaker than that of the TCV TSS. There are many possible
reasons why ribosome binding to the PEMV 3’TSS was less robust compared with the TCV TSS: i. the limited stability of ψ formation revealed by structure modeling and MD simulations suggests that fewer RNA transcripts synthesized in vitro might adopt the proper folding necessary for ribosome binding; ii. efficient ribosome binding to the PEMV 3’TSS may require additional factors not present in the purified components; or iii. the function of the PEMV 3’TSS may necessitate weaker ribosome binding.

Although the TCV TSS is not associated with any obvious RNA bridge connecting it to the 5’ end (22), we propose that the kl-H hairpin performs this function for the PEMV 3’TSS. The kl-H apical loop was critical for PEMV accumulation in vivo, and a fragment containing the kl-H was able to partially repress translation of a reporter construct, most likely by interfering with the kl-TSS kissing-loop interaction with hairpin 5H2 (Fig. 10). In the absence of the kl-TSS/PTE and when the kl-H (and 3’TSS) are positioned proximal to the 3’ end of the luciferase ORF, a kissing-loop interaction between the kl-H and 5H2 was important for efficient translation in WGE (Fig. 7D) and protoplasts (Fig. 8B). Altogether, these results indicate that the kl-H/3’TSS can function as a 3’CITE for translation of the reporter mRNA, and thus likely comprise an additional 3’CITE for PEMV. As was previously reported for the TCV TSS and the PEMV kl-TSS/PTE (13), inhibition of 5’89+3U translation by fragments containing the kl-H and 3’TSS added in trans to WGE implies that the elements enhance translation only in cis. Thus, translation enhancement mediated by the kl-H/3’TSS and the kl-TSS/PTE likely impact translation re-initiation and not the pioneer round of translation initiation.

An important question is why a virus would contain two (presumably) independently functioning, ribosome-binding 3’CITEs and why would both be associated with a motif that could potentially connect with the same 5’ hairpin (5H2)? Some possibilities for why a virus
might have multiple 3’CITEs includes:

i. redundancy to support a critical aspect of translation;

ii. ribosomes could be recycled from different translation termination sites; iii. ribosomes could
be recycled to different initiation sites, such as the sgRNA initiation site; and iv. ability to
overcome translation-based resistance in different hosts (29). Since mutations in the kl-TSS that
disrupt the kissing-loop interaction are highly detrimental despite maintaining the potential kl-
H/5H2 interaction, the answer is unlikely to be redundancy of a single function. In addition,
compensatory alterations between the kl-TSS and 5H2 apical loops in full-length PEMV gRNA
resulted in near WT levels of PEMV accumulation in protoplasts (13), despite the mutations in
5H2 disrupting a putative interaction with the kl-H. Furthermore, since alteration of the kl-H
apical loop in full-length PEMV alone or in combination with mutations that maintain the kl-
TSS/5H2 interaction and possible kl-H/5H2 interaction, still produced background levels of
virus in protoplasts (Fig. 9), it is unlikely that the function of the kl-H in PEMV is to form a
kissing-loop interaction with 5H2. Rather, the kl-H may connect the 3’TSS to an alternative
interacting sequence. We previously found that nearly all Carmovirus kissing-loops between
their 3’CITEs and 5’ ends involve the same sequence motifs (5’YGCCA and 5’UGGCR) (12)
and thus the presence of these motifs in both the kl-TSS and kl-H may reflect an unknown
requirement for these particular sequences and not necessarily interaction with the same 5’
hairpin (5H2).

A hairpin just downstream from the translation start site of PEMV p26/p27 can also
potentially pair with either the kl-H or kl-TSS (2869\text{GGAGCUAkcucGcGcGUCU}; putative stem
is underlined and kl-H/kl-TSS complementary sequence is in bold italic). An additional possible
interacting sequence exclusive to the kl-H is located just downstream at 2866\text{CGGCGA}. In
addition to possibly enhancing translation of the subgenomic RNA template for p26/p27,
connecting the kl-H to either of these sequences might allow for putative internal ribosome entry site in the gRNA that could be used for translation of p26/p27 prior to sgRNA synthesis, as was found for the coat protein of Carmovirus *Pelargonium flower break virus* (57). Since mutations in kl-H reduced PEMV accumulation in protoplasts to background levels, both of these latter possibilities require that p26 and/or p27 be critical for PEMV accumulation in single cells (or that the kl-H is important for PEMV replication). The p27 orthologue of the Umbravirus *Groundnut rosette virus* (GRV) functions as a cell-to-cell movement protein (27, 58), while the p26 orthologue is a multifunctional protein that transits to the nucleolus through its association with fibrillarin and is then exported to cytoplasmic granules for binding to the gRNA and long-distance movement (59, 60). p26 is also important for stabilization of the gRNA via its cooperative binding (61), which may be critical in the apparent absence of an encoded umbravirus silencing suppressor (27). The importance of these products for PEMV accumulation in protoplasts and possible translational regulation mediated by the kl-H/3’TSS is currently being investigated.

An important question is whether PEMV is unique in having three 3’ CITEs or whether additional 3’CITEs remain to be discovered in other cap-independently translated plant viruses. Umbravirus TBTV has a BTE-like element in the same location as the kl-TSS (62) and also contains a similar collection of hairpins and pseudoknots at the 3’ end that form the PEMV 3’TSS (Fig. 1C), suggesting that it too has multiple 3’CITEs. Since 3’CITEs have mainly been elucidated using reporter constructs, proximity of a 3’CITE to the reporter termination codon may have guided identification, suggesting that additional translation enhancers may remain to be elucidated.
Acknowledgements

This work was supported by NSF (MCB 1157906) to AES. This publication has been funded in part with federal funds from the Frederick National Laboratory for Cancer Research, National Institutes of Health, under Contract HHSN 261200800001E to WKK. This research was supported in part by the Intramural Research Program of the National Institutes of Health, Center for Cancer Research to BAS. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Figure Legends

Fig. 1. Conserved structural features at the 3’ ends of TCV and three Umbraviruses. A. Top, genome organization of Carmovirus TCV. p28 is a replication-required protein and its readthrough product p88 is the viral RdRp. p8 and p9 movement proteins are expressed from one sgRNA and the capsid protein (CP) is expressed from a second sgRNA. Bottom, secondary and tertiary structure elements at the 3’ terminus of TCV. All 15 Carmoviruses contain Pr, H5 and ψ1. Most Carmoviruses contain H4b and fewer contain both H4b and ψ2 (63). Only TCV, CCFV and JINRV contain all hairpins and pseudoknots shown (23). The boxed, shaded region denotes that hairpins H4a, H4b, H5 and pseudoknots ψ2 and ψ3 fold into a 3-D T-shaped structure (TSS). B. Top, Genome organization of PEMV. p33 is a putative replication-required protein and its -1 ribosomal frameshift protein is the RdRp. p26 and p27 are movement required proteins expressed from the single sgRNA. Bottom, possible structural organization at the 3’ terminus of PEMV based on the structures found in TCV. C and D. Phylogenetically conserved
elements at the 3’ end of TBTV (C) and CMoV (D).

Fig. 2. In-line probing of the 3’ end of PEMV. A. Susceptibility of residues at the 3’ end of PEMV to in-line cleavage. The 170-nt 3’ terminal fragment was radiolabeled at the 5’ end and incubated at 25°C for 14 h followed by denaturing gel electrophoresis. Location of different putative 3’ elements are indicated to the right of each autoradiogram. L, OH’ treated ladder; T1, partial RNase T1 digest to denote location of guanylates; I, in-line cleavage of the fragment. Intensity of each band is proportional to the flexibility of the residue at that location. B. Susceptibility of residues in the putative structure of PEMV 3’ region to in-line cleavage. Darker triangles denote stronger cleavage.

Fig. 3. Effect of alterations in 3’ elements on PEMV accumulation in protoplasts. A. Single and compensatory mutations generated in the 3’ hairpins and pseudoknots that may comprise a TSS. Names of the alterations are in brackets. B. Relative levels of full-length WT PEMV and PEMV containing alterations in 3’ hairpins accumulating in Arabidopsis protoplasts at 24 h after inoculation. Standard deviations from three independent experiments are shown. GDD, PEMV non-replicating control with an altered GDD RdRp active site motif. C. Relative levels of full-length PEMV accumulating in Arabidopsis protoplasts containing alterations in ψ₂ or ψ₃. D. In-line probing of PEMV 3’ terminal fragment containing alterations in hairpins. Residues that are more susceptible to cleavage in fragments containing mutations are denoted by a solid circle. Open circle denotes residues showing reduced cleavage levels. Location of hairpins and linker region are indicated on the right.
Fig. 4. Proposed secondary structure for the 3′UTR of PEMV. SHAPE was conducted on full-length PEMV gRNA as described in Materials and Methods. A. Residues with moderately high and high reactivity to NMIA are colored red and residues with low or moderate reactivity are colored green. The locations of the 3′TSS, kl-TSS, and PTE are shown. kl-H is a hairpin identified as also engaging in a long-distance interaction (see text). Single asterisk denotes a hairpin that is sequence, structure, and positionally conserved in all sequenced Umbraviruses (A. E. Simon, unpublished). Double asterisk denotes a potential pseudoknot predicted by pknotsRG (44). Termination codon for p27 is boxed in yellow. B. Portion of one of the SHAPE phosphorimages used for the structural prediction. C, G are ladder lanes of cytidylates and guanylate positions. N, NMIA-treated sample; D, control DMSO-treated sample denoting reverse transcriptase stops in the absence of NMIA. Green bars denote regions in H4a or H5 that are susceptible to in-line cleavages but not reactive with NMIA.

Fig. 5. 3-D model and molecular dynamics simulation of the PEMV 3′TSS. The model with a 3-bp ψ3 is shown (see text for details). A. Energy-minimized average structure based on the 100 ns long MD simulation. B. The all-atom RMSD (82 nt; 2,625 atoms), measured relative to the first structure of the MD simulation is 5.6 Å (black). RMSD values plotted in red were calculated for all atoms of the 6 nts involved in ψ3 (3 bp long in this model, 191 atoms). The blue vertical line at the 25 ns point in the MD indicates when restraints on the first three base pairs were released (see text for details). The low mean RMSD and standard deviation (0.9 ±0.1 Å) of ψ3 illustrates the stability of the base-pairs after the restraints were lifted. C. Comparison of the 3-D structure models of the TCV TSS (left), PEMV 3′TSS (center), and PEMV kl-TSS (13) (right), all shown in red, aligned with the tRNAPhe structure (PDB #: 1EHZ) (gray). The TSS and kl-TSS 5′ and 3′
positions are labeled in red, while the tRNA’s 5’ and 3’ positions, the anticodon loop (AC) and
the acceptor stem (AA) are labeled in black over the most exposed tRNA (right).

Fig. 6. Ribosome binding to the 3’TSS. Filter-binding assays were conducted using three
different sized fragments that contain the 3’TSS (S1, S2, S3). A. Locations of the end-points of
the fragments used for filter binding. B. One to 60 pmol of [32P] 5’-end labeled fragments were
combined with 15 pmol of salt-washed ribosomes or separated subunits purified from
Arabidopsis protoplasts. Standard deviations are shown for three experiments. Asterisks denote
that the values for the TCV TSS (31) and PEMV kl-TSS (13) were previously published and are
presented here for comparison.

Fig. 7. 3’TSS and kl-H function as 3’CITEs in WGE when proximal to the luciferase reporter
termination codon. A. Location of mutations in 3’TSS, kl-H and 5’89. The putative kissing-loop interaction between hairpin 5H2 in the 5’89 nt of PEMV and the kl-H is shown. Sequences
shared between the kl-H and hairpin 3H1 of the kl-TSS (which is known to interact with 5H2)
are boxed. B. Relative luciferase activity in WGE programmed with 5’89+3U containing
mutations in the kl-H and 3’TSS. Asterisks denote data previously published (13) and presented
here for comparison. Data is from three independent experiments performed in triplicate and
standard deviations are shown. C. Diagram of 5’89+3Umini. This construct contains a deletion
that removes the kl-TSS and PTE and places the kl-H just downstream of the luciferase ORF. D.
Relative luciferase activity of 5’89+3Umini and 5’89+3Umini that contains mutations in the kl-
H and 3’TSS shown in A.
3'TSS and kl-H function as 3'CITEs in protoplasts when proximal to the luciferase reporter termination codon. A. Relative luciferase activity in protoplasts at 18 h following transformation with 5′89+3U transcripts containing mutations in the kl-H and 3'TSS that were described in Fig. 7A. Asterisks denote data previously published (13) and presented here for comparison. Data is from three independent experiments performed in triplicate and standard deviations are shown. B. Relative luciferase activity in protoplasts of 5′89+3Umini and 5′89+3Umini containing mutations in the kl-H and 3’TSS.

The kl-H is critical for gRNA accumulation but does not connect with 5′ proximal hairpin 5H2. A. Location of mutations in 5H2, kl-TSS (13) and kl-H. The known kissing-loop interaction between 5H2 and the kl-TSS is shown. Sequences shared between the kl-H and hairpin 3H1 of the kl-TSS are boxed. B. PEMV accumulation in protoplasts at 24 hpi as measured by qPCR. GDD, PEMV containing mutations in the RdRp active site that abrogate PEMV replication. Data is from three independent experiments performed in triplicate and standard deviations are shown.

kl-H and 3′TSS repress translation in trans. A. Diagram of the 3′UTR fragments added to WGE containing 5′89+3U. B. Relative luciferase activity of 5′89+3U in the presence of the fragments indicated. None, no fragments added; satC, 356 nt untranslated satellite RNA of TCV used as a control; 3′TSS M1, 3′TSS fragment containing mutations that disrupt hairpin H4b (see Fig. 3A). Values represent three independent experiments and standard deviation is
shown.

References


the major nucleolar protein fibrillarin is required for systemic virus infection. Proc Natl Acad Sci USA 104:11115-11120.


Figure 1
Figure 2
Figure 5

Average Structure; MD of 100 ns

RMSD (Å) vs Time (ns)

TOTAL (82 nt)   MEAN RMSD: 5.6 ± 1.0 Å
Ψ3 (6 nt)            MEAN RMSD: 0.9 ± 0.1 Å

Figure 5
Figure 6

RNA secondary structure and related data.

**A** shows the RNA secondary structure, with some sequences labeled (e.g., S1, S2, S3).

**B** shows a bar graph with ribosome binding Kd (μM) values for different conditions.

Key:
- **S1**
- **S2**
- **S3**
- **TCV TSS**
- **40S**
- **60S**
- **80S**

**Ribosome binding Kd (μM)**

- 1.0
- 2.0
- 3.0
- 4.0
- 5.0

**Conditions**
- **Non-specific**
- **Specific**
- **PEMV kl-TSS**
Figure 9