Structural characterization of a new subclass of panicum mosaic virus-like 3′ cap-independent translation enhancer

Philip Z. Johnson[1], Wojciech K. Kasprzak[2], Bruce A. Shapiro[3] and Anne E. Simon ⓘ[1],*

[1]Department of Cell Biology and Molecular Genetics, University of Maryland - College Park, College Park, MD 20742, USA, [2]Basic Science Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA and [3]RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

ABSTRACT

Canonical eukaryotic mRNA translation requires 5′cap recognition by initiation factor 4E (eIF4E). In contrast, many positive-strand RNA virus genomes lack a 5′cap and promote translation by non-canonical mechanisms. Among plant viruses, PTEs are a major class of cap-independent translation enhancers located in/near the 3′UTR that recruit eIF4E to greatly enhance viral translation. Previous work proposed a single form of PTE characterized by a Y-shaped secondary structure with two terminal stem-loops (SL1 and SL2) atop a supporting stem containing a large, G-rich asymmetric loop that forms an essential pseudoknot (PK) involving C/U residues located between SL1 and SL2. We found that PTEs with less than three consecutive cytidylates available for PK formation have an upstream stem-loop that forms a kissing loop interaction with the apical loop of SL2, important for formation/stabilization of PK. PKs found in both subclasses of PTE assume a specific conformation with a hyperreactive guanylate (G*) in SHAPE structure probing, previously found critical for binding eIF4E. While PTE PKs were proposed to be formed by Watson–Crick base-pairing, alternative chemical probing and 3D modeling indicate that the Watson–Crick faces of G* and an adjacent guanylate have high solvent accessibilities. Thus, PTE PKs are likely composed primarily of non-canonical interactions.

INTRODUCTION

Eukaryotic mRNAs typically contain a 5′m⁷GpppN cap and a 3′poly(A) tail that function to attract translation initiation factors and ribosomes to the 5′end of the template.

Canonical translation begins with recognition of the 5′cap by initiation factor 4E (eIF4E), and association of poly(A)-binding protein with the poly(A) tail. These initiation factors bind to the core scaffolding protein eIF4G to form part of the eIF4F initiation factor complex that bridges the 5′ and 3′ends of the mRNA (1,2) to facilitate recycling of ribosomes (3,4). The 43S ribosome preinitiation complex, composed of the 40S ribosomal subunit and ternary complex (Met-tRNA, eIF2 and GTP), is then recruited to the 5′end of the mRNA, scans the transcript in the 5′ to 3′ direction until encountering a start codon in a favorable context, recruits the 60S ribosomal subunit to form the 80S ribosome, and translation initiates after an amino-acylated tRNA enters the ribosome A-site.

Positive-strand (+)RNA viruses have genomes that directly serve as templates for the host translation machinery upon cell entry. Many (+)RNA viruses lack a 5′cap and 3′poly(A) tail to differentiate their templates from host mRNAs, and thus must recruit the translational machinery by non-canonical mechanisms (5,6). To promote efficient translation, the genomes of these (+)RNA viruses form higher order structures that recruit all or a subset of the host translation machinery (e.g. eIF4E, eIF4G, eIF4F, ribosomal subunits). Many such structures have been identified at both the 5′ and 3′ends of plant and animal (+)RNA viral genomes, and understanding their mode of action has led to the elucidation of non-canonical mechanisms of viral gene regulation at the level of translation. A high prevalence of related non-canonical mechanisms that regulate translation of a subset of host mRNAs has also been revealed by transcriptome-wide studies (7).

Plant (+)RNA viruses that lack a 5′cap and 3′poly(A) tail generally contain structures promoting translation in the 3′ region of their genome, which are termed 3′cap-independent translation enhancers or 3′CITEs (8,9). 3′CITEs are classified by their secondary structures, three-dimensional (3D) shapes, conserved RNA motifs, and translation factor(s) with which they interact. For example,

---

*To whom correspondence should be addressed. Tel: +1 301 405 8975; Email: simona@umd.edu

barley yellow dwarf virus-like translation elements (BTEs), which are found in various viruses in the *Tombusviridae* and *Luteoviridae* families (9–11), have a signature 17-nt motif that forms a short stem-loop (SL) implicated in the binding of eIF4G (12). The three subclasses of BTEs have different numbers of additional stem-loops atop a base-paired stem, as well as other general and class-specific motifs and conserved structural features uncovered by a recent analysis (13). In contrast, other types of 3′CITEs, such as the T-shaped structure (TSS) found in members of the carmovirus and umbravirus genera within the *Tombusviridae*, lack conserved motifs but contain conserved structural elements (three hairpins and two pseudoknots). These elements fold into a tRNA-like 3D shape that binds to 60S ribosomal subunits or 80S ribosomes by occupying the P-site (14–16).

Most 3′CITEs deliver bound translation factors/complexes to the 5′end of the viral genome via long-distance base-pairing between a 5′ proximal hairpin and a hairpin within or proximal to the 3′CITE (9). The 3′ hairpins that participate in these long-distance RNA:RNA interactions in carmoviruses and umbraviruses commonly contain the motif 5′UGCCA or its complement (5′UGGCA) in their apical loops. For a minority of 3′CITEs, long-distance base-pairings are not discernible, suggesting that alternative mechanisms are required to position the 3′CITE near the 5′end (9,17).

In addition to structural diversity among different 3′CITEs, there are also subclasses of particular 3′CITEs. For example, BTEs were recently divided into three subclasses based on their number of terminal SLs and other conserved features (13). Nearly all reports have examined BTEs in the subclass with three terminal SLs, which has left the other two subclasses and their unique functional and structural elements to be characterized. There is also a lack of information about the structure and function of 3′CITEs *in situ* (i.e. within the full-length viral genomic RNA or subgenomic RNA), as nearly all reports involve the study of 3′CITE fragments with the assumption that they contain the complete functional enhancer. However, elements associated with 3′CITE function and/or regulation can be located outside of the region included in these fragments, as was recently shown for the BTE of umbravirus tobacco bushy top virus (18). A number of 3′CITEs are also located adjacent to coding sequence, if not partly within coding sequence, such as the translation enhancer domain (TED) of satellite tobacco necrosis virus (19,20). However, most studies analyzing 3′CITE function use reporter constructs containing only the 3′UTR and thus any elements residing in the adjacent upstream coding sequence would be overlooked.

Panicum mosaic virus-like translation enhancers (PTEs) are a major class of 3′CITE that bind strongly to the cap-binding initiation factor eIF4E (21–23). PTEs across genera in the *Tombusviridae* have been subjected to extensive investigation over the past 15 years including chemical RNA structure probing, mutational analyses, eIF4E binding, and involvement in viral translation *in vitro* and *in vivo* (21–25). Structurally, all PTEs have a Y-shaped secondary structure consisting of a three-way junction with two upper SLs (SL1 and SL2) atop a supporting stem of 6–7 bp, followed by a

large, G-rich asymmetric loop and a lower supporting stem of different sizes (Figure 1C). Sequential guanylates (usually 3–4) within the asymmetric loop participate in a pseudoknot (PK) involving C/U-rich residues between SL1 and SL2. This PK is proposed as a common feature of all PTEs and is critical for binding to eIF4E, which correlates with efficient enhancement of viral translation (21,22). Additionally, most PTEs contain 5′UGCCA, or a close variant, in the apical loop of SL1 that engages in the long-distance base-pairing with the apical loop of a 5′ proximal hairpin or other unpaired sequence (24). An exception is the PTE of pea enation mosaic virus 2 (PEMV2; umbravirus), which uses a hairpin in a neighboring structure (known as the kl-TSS) to connect with a 5′ proximal, coding region hairpin (26).

A key feature of all PTEs is the presence of a particular guanylate within the G-rich asymmetric loop that is hyper-reactive to SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) structure probing reagents. SHAPE is a standard technique for helping to determine the structure of RNA molecules, and involves reacting the folded RNA with a modifying reagent (e.g. NMIA, 1M7) that acylates the 2′-hydroxyls of residues independent of their nitrogenous base identity (27). Reverse transcription stalls in the RNA template at the residue immediately 3′ to the residue that is modified by NMIA. Paired residues typically have low reactivities with the modifying reagent and unpaired residues have high reactivities, allowing the putative structure of an RNA molecule to be inferred. Hyperreactivity of a critical guanylate in SHAPE probing of isolated PTE fragments correlated with formation of the PK and ability to enhance translation (21,22). Footprinting of PTE fragments bound to eIF4E also directly implicated the hyperreactive guanylate (denoted here as G*) in eIF4E binding (22). Given the hyperreactivity of G*, it was proposed to be unpaired and highly accessible to the solvent, despite the presence of available cytidylate pairing partners between SL1 and SL2 (22).

Previous work focused on PTEs containing PKs with the potential to form three or more G:C pairs, such as the PTE of PEMV2 (Figure 1C). However, nearly half of all known PTEs, particularly those in virus members of the three carmovirus genera, do not have three consecutive cytidylates between SL1 and SL2 and thus have seemingly weaker PKs, usually requiring one or more G:U wobble pairs (22,24) (Figure 1D). In addition, Watson–Crick paired PKs are not discernible for the PTEs of galinsoga mosaic virus (GaMV) and Bermuda grass latent virus (BGLV) (see Figure 2). Little is known about the stability or functionality of PTEs containing these seemingly weak or indiscernible PKs.

The possible presence of a third stem loop (SL3) just upstream of many PTEs with seemingly weak PKs was noted for a number of alphacarmoviruses and closely related GaMV, but its involvement, if any, in PTE function was not examined (25). For this report, we have identified a kissing loop (KL) interaction between the apical loops of SL3 and PTE SL2 in alphacarmovirus pelargonium flower break virus (PFBV) that is conserved in nearly all PTEs with upstream SL3, and found that this KL is critical for G* hyperreactivity and thus formation of the PK. Through alternative chemical probing techniques, we found the Watson–
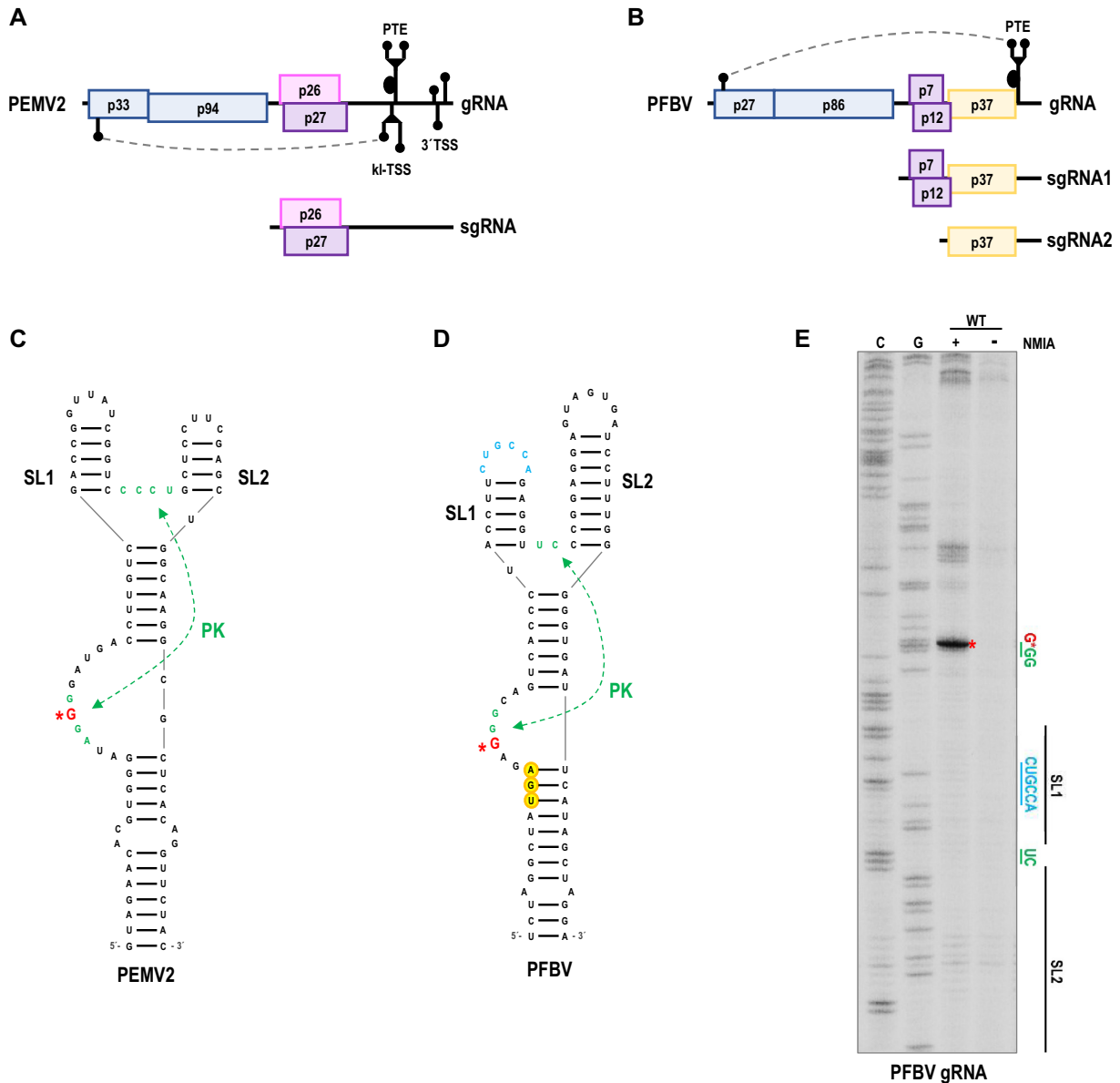
Figure 1. PEMV2 and PFBV genome organizations and PTEs. (**A**) PEMV2 gRNA and subgenomic (sg)RNA. p33 and its –1 ribosomal frameshift extension product p94 (the RNA dependent RNA polymerase [RdRp]) are expressed from the gRNA (48,49). Movement proteins p26 and p27 are expressed from the sgRNA by a non-leaky scanning mechanism (50). The 3′CITEs of PEMV2 (present in both gRNA and sgRNA) are presented on the gRNA and are composed of the PTE, the upstream kissing loop T-shaped structure (kl-TSS), and the 3′TSS (16,51). Dashed line denotes long-distance base-pairing between the kl-TSS and a 5′ proximal, coding region hairpin (26). (**B**) PFBV gRNA and sgRNAs 1 and 2. p27 is expressed from the gRNA along with its ribosomal readthrough product p86 (the RdRp) (52). Movement proteins p7 and p12 are expressed from sgRNA1 (53). p37 (capsid protein) is expressed from sgRNA2 by a 5′ scanning dependent mechanism and also from the gRNA via an internal ribosome entry site (IRES) immediately upstream of the p37 ORF (54,55). The location of the PTE (present in both gRNA and sgRNAs) is shown on the gRNA and includes a portion of the p37 ORF. Dashed line denotes putative long-distance base-pairing between PTE SL1 loop sequence and a 5′ coding region hairpin (24). (**C, D**) Previously defined structures of the PEMV2 and PFBV PTEs (21,22). The hyperreactive G* is in red. Note that the PEMV2 G* residue was mislabeled in these prior publications as the 3′ adjacent guanylate. (**E**) SHAPE probing of the gRNA region containing the PFBV PTE. Regions corresponding to specific locations in the secondary structure are indicated at right. Red asterisks denote the location of G*. (+) +NMIA; (–) DMSO only. The difference in band intensities between the (+) and (–) lanes for a residue indicates the reactivity of a residue with NMIA. C and G are sequencing lanes. Reverse transcription stalls on the RNA template at the residue immediately 3′ to the residue that is modified by NMIA, and thus residues in the SHAPE lanes migrate one residue lower than the corresponding residue in the sequencing lanes.
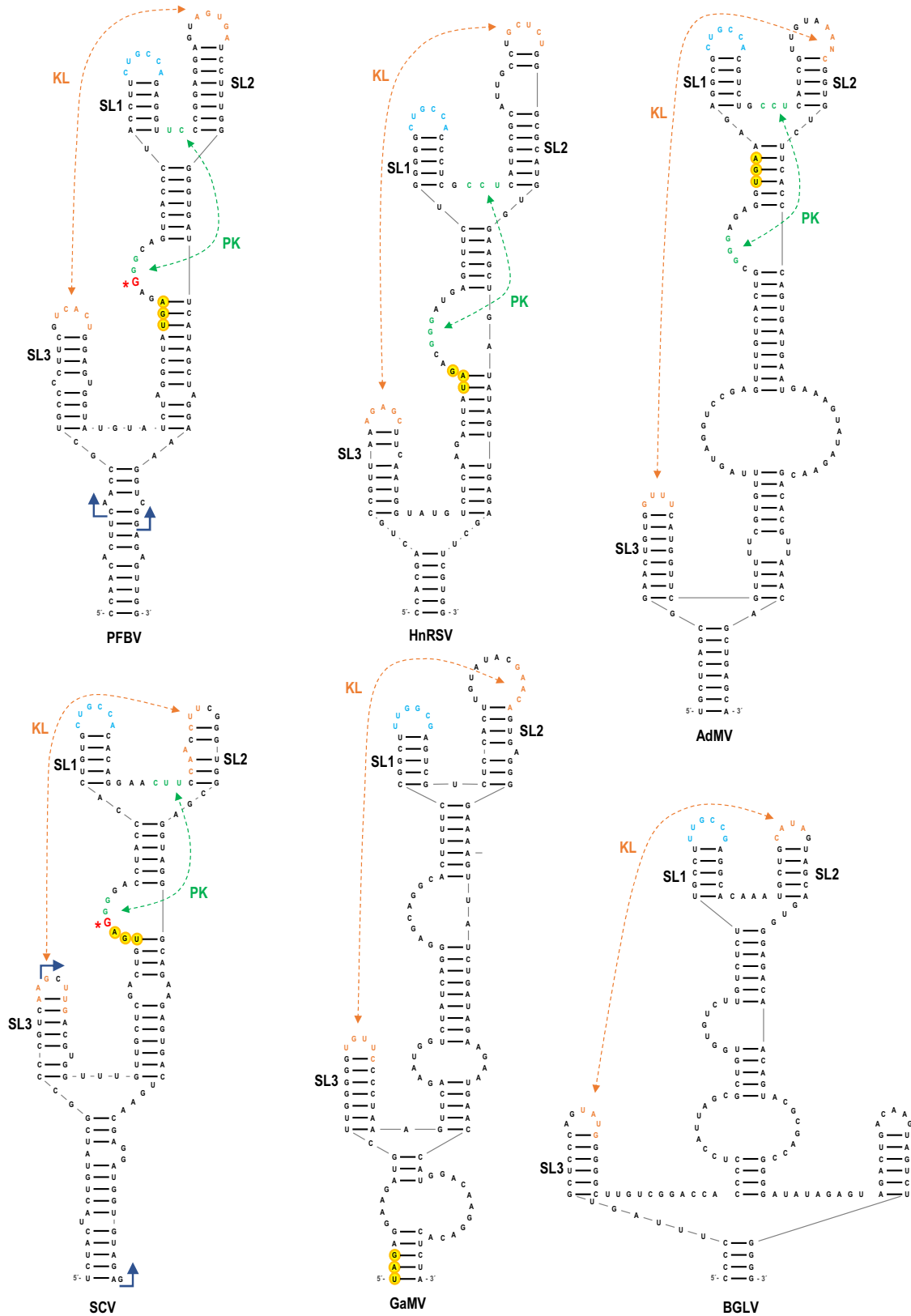
Figure 2. PTEs with putative upstream SL3. Complementary bases between SL3 and SL2 are in orange. Possible PK residues are in green. Hyperreactive G* residues previously identified (22) are in red. SL1 loops containing 5′UGCCA or related motifs known or proposed to participate in long-distance base-pairing are in light blue (24,26). Residues in yellow are termination codons for the upstream ORF (if present in the PTE). Blue arrows denote boundaries of the PTE fragments used in a previous report (22). Pairing in AdMV is presumptive based on the only available sequence having an unspecified residue (N) in the SL2 apical loop. HnRSV, honey suckle ringspot virus; AdMV, adonis mosaic virus. See text for additional abbreviations.

Crick faces of the PFBV (and PEMV2) G* residues to be highly accessible to the solvent and unpaired. Surprisingly, the Watson–Crick face of the guanylate immediately downstream of G* for both the PFBV and PEMV2 PTEs showed even higher accessibility and is also unpaired, suggesting that the PKs of the PFBV and PEMV2 PTEs are composed primarily of non-canonical interactions, in contrast to the model previously proposed for the PEMV2 PTE (22). Our findings show that PTEs with seemingly strong and weak PKs, and with and without a KL, share similar conformations involving their G* and nearby residues that likely play a central role in how PTEs bind to eIF4E.

MATERIALS AND METHODS

Constructs

Plasmid pUC19-PFBV contains the full-length PFBV genome (NC_005286) downstream of a T7 promoter and was synthesized by GenScript (Piscataway, NJ). Plasmid pUC19-PEMV2 contains the full-length PEMV2 RNA2 genome downstream of a T7 promoter (26). PFBV and PEMV2 mutants were generated by QuikChange one-step site-directed mutagenesis (28) using custom oligonucleotide primers (Integrated DNA Technologies) (Supplemental Table S1). Resultant PCR products were subjected to DpnI digestion and transformed into chemically competent DH5α *E. coli* cells. Mutations were confirmed by sequencing (Eurofins Genomics). The JIN-PTE-WT fragment (D86123; positions 3738–3880) and JIN-PTEmPK fragment were purchased as gBlocks Gene Fragments from Integrated DNA Technologies.

In vitro transcription of full-length viral genomes

pUC19-PFBV and pUC19-PEMV2 were linearized with SrfI and SmaI (New England BioLabs), respectively. Linearized plasmids were used as templates for *in vitro* transcription using bacteriophage T7 RNA polymerase.

In vitro transcription of PTE fragment RNA

PFBV PTE and JIN-PTE fragments were PCR amplified using custom oligonucleotide primers (Integrated DNA Technologies) (all oligonucleotide primer sequences are available upon request). Forward primers contained an upstream T7 promoter. For *in vitro* translation *trans*-inhibition assays, the PFBV PTE fragment amplified was positions 3619–3781, and for SHAPE structure probing, the PFBV PTE fragment amplified was positions 3619–3792 and the probe annealing site 5′GGTCATAGCTGTTTCCT was added to the 3′end. Resulting PCR products were used as templates for *in vitro* transcription using T7 RNA polymerase.

Purification, quantification and quality assessment of *in vitro* transcribed RNA

The PFBV PTE RNA fragment for trans-inhibition assays was purified using a Monarch RNA Cleanup Kit (New England BioLabs; #T2040L). All other *in vitro* transcribed RNAs were purified by lithium chloride precipitation. For all assays, *in vitro* transcribed RNA was quantified using a DeNovix DS-11 FX spectrophotometer and its quality assessed by ethidium-stained agarose gel electrophoresis.

In vitro translation and *trans*-inhibition assays

Five microliters of wheat germ extracts (Promega), 0.8 μl of 1 mM amino acid mix (Met–), 0.47 μl of 1 M potassium acetate, 0.25 μl of Murine RNase inhibitor (New England BioLabs), and 0.5 μl [5 μCi] of ^{35}S-methionine were mixed with 0.5 pmol of RNA transcripts. For trans-inhibition assays, 20 pmol of PTE fragment transcripts were also added. Total volumes of translation reaction mixtures were raised to 10 μl with distilled, deionized water (ddH$_2$O). Translation reactions were incubated at 25°C for 30 min and then resolved by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The gel was dried and exposed to a phosphorimager screen and scanned using an Amersham Typhoon Biomolecular Imager (serial #: 96510600). ImageJ software was used to produce gel images and quantify band intensities by boxing bands and using the ImageJ gel analysis tools to measure the total intensities of boxed bands while subtracting out local background signal beneath bands (see for a step-by-step guide to the band quantitation technique: https://www.sybil-fp7.eu/node/95) (29).

RNA folding

RNA transcripts (12 pmol) were denatured at 65°C for 3 min then snap-cooled on ice for 2 min. Unless otherwise stated, RNA was subjected to folding in buffer containing 80 mM Tris–HCl (pH 8.0), 11 mM Mg(CH$_3$COO)$_2$ and 160 mM NH$_4$Cl. When performing SHAPE structure probing under low Mg^{2+} conditions, the 11 mM Mg(CH$_3$COO)$_2$ in the default folding buffer was omitted and replaced with 22 mM NaCl, and for PFBV PTE analysis, 10 mM EDTA was also added. RNA transcripts were incubated in the folding buffer at 37°C for 10 min. Folded RNAs were split evenly into two aliquots: (+) sample to be modified and (–) negative modification control.

SHAPE modification of RNA

As described previously (27), *N*-methylisatoic anhydride (NMIA) [dissolved in dimethyl sulfoxide (DMSO)] was added to a final concentration of 15 mM to the (+) folded RNA aliquot, while an equal volume of DMSO was added to the (–) negative modification control. Folded RNAs were then incubated at 37°C for 30 min, followed by ethanol precipitation.

CMCT modification of RNA

CMCT (dissolved in ddH$_2$O) was added to a final concentration of 75 mM to the (+) folded RNA aliquot and an equal volume of ddH$_2$O was added to the (–) negative modification control, as previously described (30). Folded RNAs were then incubated at 25°C for 30 min, followed by ethanol precipitation.

DMS modification of RNA

0.025 volumes of DMS were added to the (+) folded RNA aliquot and an equal volume of ddH$_2$O was added to the (–) negative modification control, as previously described (30). RNAs were incubated at 25°C for 90 s, followed by ethanol precipitation, after which samples were resuspended in 0.5 M Tris–HCl (pH 8.0), 0.1 M NaBH$_4$, and incubated on ice in the dark for 30 min. Samples were ethanol precipitated and resuspended in 1 M aniline-acetate buffer (created by mixing 10 ml of aniline (99.5%), 93 ml of ddH$_2$O, and 6 ml of glacial acetic acid), followed by incubation at 60°C in the dark for 15 min and then ethanol precipitation.

Reverse transcription of modified RNA

Oligonucleotides labeled at their 5′ends with [γ-^{32}P]ATP were used to perform primer extension reactions on modified RNA using SuperScript III reverse transcriptase (Invitrogen), as previously described (27). The oligonucleotides used targeted the following locations in RNA transcripts: full-length PFBV gRNA, 3781–3802; full-length PEMV2 gRNA, 3909–3929; PFBV PTE fragment, the annealing site 5′GGTCATAGCTGTTTCCT included at the 3′end; JIN-PTE fragment, 3862–3880. Primer extension products were resolved on an 8% denaturing polyacrylamide gel. The gel was dried and exposed to a phosphorimager screen and scanned using an Amersham Typhoon Biomolecular Imager. ImageJ software was used to produce gel images and quantify band intensities by boxing bands and using the ImageJ gel analysis tools to measure the total intensities of boxed bands while subtracting out local background signal beneath bands (29). Loop intensity values are the summation of band intensities comprising the entire loop. Reactivities of G* and the SL3 and SL2 loops were determined by subtracting control (–) lane band intensities from NMIA-treated (+) lane band intensities and normalized to invariant bands just upstream of the PTE structures. Dideoxynucleotide-based Sanger methods were used to generate the sequencing ladders used to identify the positions of bands on the gel.

3D structure predictions and coarse-grained simulations

RNA sequences and Watson–Crick base-pairing constraints were input to the 3D RNA structure prediction program RNAComposer (31). Default prediction parameters were used. Predicted 3D models were selected for downstream analysis from among the 10 highest ranking predictions. Coarse-grained simulations using SimRNA (32) were also used to generate 3D models of PTEs, using input RNA sequences and Watson–Crick base-pairing constraints. Default simulation parameters were used [500 steps; 1% lowest energy frames cutoff for clustering]. Generated 3D models were selected for downstream analysis from among the five highest ranking final models. Sequences for PFBV, PEMV2 and JINRV PTE models were the same as used in wet lab experiments and were comprised of the following fragments: PFBV, 3629–3781; PEMV2, 3815–3912; JINRV, 3757–3849. Other PTE sequence accession numbers (and fragments) were as follows: PMV, NC_002598 (4114–4200);

CMMV, EU081018 (3979–4068); CarMV, X02986 (3736–3824); TPAV, JX848617.1 (3983–4063).

Molecular dynamics

Amber 18 employing the OL3 force field for RNA (leaprc.RNA.OL3) was used to run explicit solvent, particle mesh Ewald MD simulations of the 3D models of PTEs with a TIP3P water box (33–36). PTE models were neutralized with Na$^+$ ions, and Na$^+$/Cl$^-$ ion pairs were added to attain a net salt concentration of 0.15 M. Equilibration was performed in 12 stages including energy minimization, heating, and short dynamics with harmonic restraints applied to the solute (RNA) and gradually relaxed over the course of the process from the value of 200.0 down to 0.01. No restraints on the RNA were used in the final 2.0 ns of the equilibration. Production runs (without any harmonic restraints) used the NPT approach (constant system size, pressure and temperature) employing Berendsen thermostat (37) and were each 500 ns in duration (excluding equilibration), performed at a temperature of 300 K and pressure of 1.0 Pa. Simulations employed explicit solvent periodic boundary conditions, 2 fs time steps, the SHAKE algorithm to constrain all hydrogen bonds in the system, and a 9 Å non-bonded interactions cut-off. For more protocol details, see (38). Hydrogen bond occupancies (including those comprising Watson–Crick base-pairs) were measured over the course of production runs using the Amber 18 utility CPPTRAJ (39). All MD simulation data and 3D models are available upon request.

RNA structure drawing

All 2D drawings of RNA structures were created using the RNA2Drawer web app (available at https://rna2drawer.app) (40).

RESULTS

SHAPE probing of the PFBV PTE in full-length genomic RNA

The PEMV2 PTE, which was the subject of nearly all foundational research on PTEs, is centrally located within the genomic (g)RNA's large 702 nt 3′UTR (Figure 1A). Just upstream of the PTE is the ribosome-binding kl-TSS 3′CITE that participates in a long-distance interaction with a coding region hairpin near the 5′end of the gRNA (26). In contrast, the PFBV PTE is located at the junction of the p37 coat protein ORF and the 236 nt 3′UTR, with the apical loop of SL1 providing the connecting sequence for the long-distance interaction with a hairpin located at position 94 within the p27 ORF (Figure 1B) (9). The PTEs of PEMV2 and PFBV have strong and weak PKs, respectively, based on the strength of their putative PK base-pairing. The PEMV2 PTE contains the sequence 5′CCCU between stem-loops SL1 and SL2 and 5′AUAGG*GGAUGA in the asymmetric loop (G* is the hyperreactive G; Figure 1C). In contrast, the PFBV PTE has only 5′UC between SL1 and SL2 for pairing with asymmetric loop sequence 5′GAG*GGCA (21,22). However, two cytidylates downstream of the 5′UC that are

predicted to pair with guanylates at the base of SL2 could be available for PK pairing (Figure 1D).

To gain information about the properties of PTEs in their natural locations within viral gRNAs, the PFBV PTE was subjected to SHAPE probing using full-length gRNA transcripts (Figure 1E). As with SHAPE conducted using the isolated PFBV PTE fragment (22), the PFBV PTE contained the same hyperreactive G* within the asymmetric loop, and the reactivities of other residues were consistent with the previously defined Y-shaped secondary structure. However, unlike SHAPE of isolated fragments, the PTE SL1 loop residues were only weakly reactive, supporting their involvement in long-distance base-pairing in the gRNA.

PFBV gRNA contains a SL upstream of the previously defined PTE structure that participates in a kissing-loop interaction with SL2

PTEs with seemingly weak or indiscernible PKs (i.e. fewer than three consecutive cytidylates between SL1 and SL2) are situated downstream from a putative SL shown in a previous publication (25) that is designated here as SL3. In the previous publication, SHAPE probing of full-length alphacarmovirus saguaro cactus virus was consistent with SL3 being present and with both SL3 and the Y-shaped PTE structure positioned atop a lower stem (Figure 2). Examination of SL3 terminal loops revealed diverse sequences that are complementary with sequences in the PTE SL2 apical loops (Figure 2, in orange), suggesting that a KL might connect SL2 with SL3. To investigate if this KL exists for the PFBV PTE and if so, whether it is important for PTE function, SL3 and SL2 loop sequences were mutated such that base-pairing (if present) could reform when both alterations were present (Figure 3A). Wild-type (WT) PFBV gRNA (gPTE-WT) along with mutant constructs gPTE-SL3sm and gPTE-SL2sm were translated *in vitro* in wheat germ extracts (WGE), and levels of p27 synthesized were determined (Figure 3B). Translation of gPTE-SL3sm and gPTE-SL2sm was reduced to 33% and 27% of gPTE-WT, respectively, and both mutations together (gPTE-SL3/2sm) restored gRNA translation to 64% of gPTE-WT. This result supports a KL interaction between SL3 and SL2 that enhances the functionality of the PFBV PTE.

SHAPE structure probing was also conducted for gPTE-WT, gPTE-SL3sm, gPTE-SL2sm and gPTE-SL3/2sm (Figure 3C). As also shown in Figure 1E, the SHAPE profile of gPTE-WT indicated that both SL3 and SL2 loop sequences had weak NMIA reactivity, suggesting that these residues were not single-stranded. In contrast, the SHAPE profiles of gPTE-SL3sm and gPTE-SL2sm displayed an average of 6.3- to 12.3-fold increases in reactivities of residues in the SL loop containing the mutation and in the loop of the proposed KL partner sequence, supporting the existence of a KL between these residues (Figure 3C). Hyperreactivity of G* was also reduced for gPTE-SL3sm and gPTE-SL2sm by 7.1- and 5.0-fold, respectively, suggesting that the KL is important for formation/stabilization of an active PTE structure. The SHAPE profile for the compensatory gPTE-SL3/2sm mutant revealed partial restoration of G* reactivity (2.7- to 3.86-fold higher than for gPTE-SL3sm and

gPTE-SL2sm) but only a slight reduction, if any, in the reactivity of the two loop sequences, suggesting that the KL was not reforming efficiently. One possibility for lack of efficient SHAPE complementation was if one or both of the altered loop sequences had established unintentional off-target pairing, such as may have occurred between the SL2sm loop and the 3′ side of the SL2 stem, which is consistent with enhanced reactivity of residues on the 5′ side of the SL2 stem (Figure 3C, Supplemental Figure S1, red line).

When PTE fragments are added *in trans* to *in vitro* translation reactions, they inhibit translation by sequestering eIF4E in the reaction mixture (21). To determine if SL3 and SL2 mutations interfere with eIF4E sequestration, WT and mutant PFBV PTE fragments (PTE-WT, PTE-SL3sm, PTE-SL2sm, PTE-SL3/2sm) were added to WGE along with PFBV gRNA template (Figure 3D). Forty-fold molar excess of the PTE-WT fragment inhibited gRNA translation by 90%, whereas the same molar excess of PTE-SL3sm and PTE-SL2sm inhibited translation by 48% and 56%, respectively. When the compensatory PTE-SL3/2sm was added, translation inhibition increased to 64%. Altogether, these results suggest that the PFBV PTE contains a KL that is necessary for efficient function and that the KL is required for hyperreactivity of G* in SHAPE probing.

The PFBV PTE contains a PK along with the KL

Previous studies found that hyperreactivity of G* correlated with formation of the PK between the G-rich asymmetric loop and C/U-rich sequence between SL1 and SL2 in several viruses including PEMV2 (21,22). As described above, hyperreactivity of G* was significantly reduced when the KL was disrupted. This could occur if the KL is required for efficient formation/stabilization of the PK, or alternatively, the KL might replace the PK in the structure of these PTE. To distinguish between the two possibilities, the single cytidylate between SL1 and SL2 was altered to an adenylate in the PFBV gRNA that should reduce or eliminate formation of a potential PK. Additionally, to prevent possible usage of the adjacent downstream cytidylates (that likely pair with guanylates at the base of SL2), the terminal base-pair of SL2 was changed from a C:G to an A:U, altogether generating construct gPTE-Cm (Figure 4A). When subjected to translation in WGE, gPTE-Cm was a weaker template than gPTE-SL2sm, generating only 10% of the p27 levels produced using gPTE-WT (Figure 4B). When a PTE fragment containing the Cm mutations (PTE-Cm) was added *in trans* to WGE, translation of the PFBV gRNA template was only reduced 9% compared to 90% reduction by a fragment containing the WT PTE (PTE-WT; Figure 4C). These results corroborate previous work demonstrating the importance of the PK for PTE activity (21,22). SHAPE probing of gPTE-Cm showed increased reactivity of the residues located between SL1 and SL2, and reduced hyperreactivity of G* (Figure 4D), consistent with the WT PFBV PTE containing a PK in the context of full-length gRNA. In addition, the gPTE-Cm SHAPE profile showed increased reactivities for SL3 and SL2 loop residues, strongly suggesting that the KL is also disrupted when the PK does not form in the gRNA. However, when SHAPE was used to probe the PTE-Cm fragment (Figure 4E), reactivities of residues
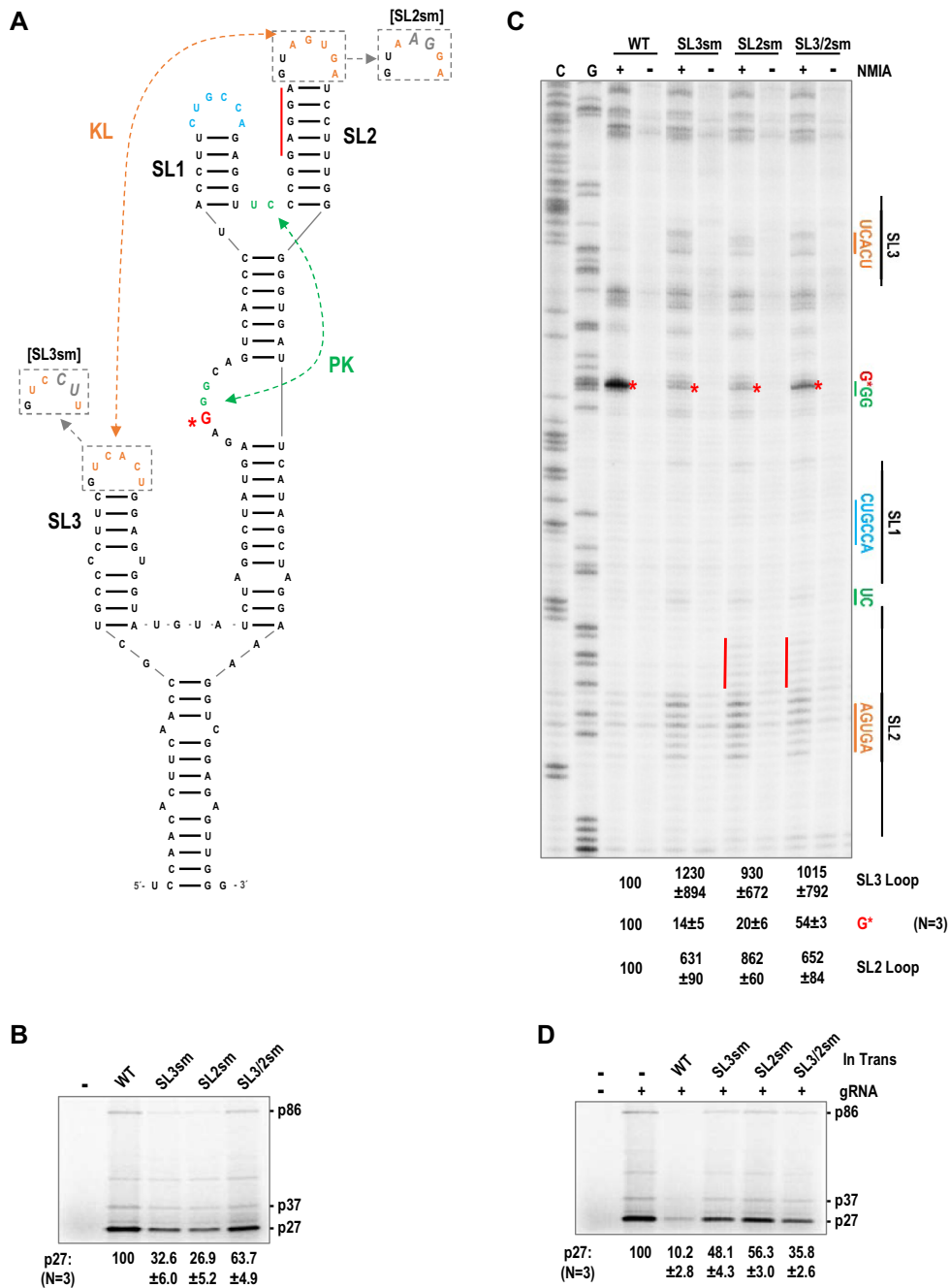
Figure 3. The PFBV PTE contains a KL between the apical loops of SL3 and SL2. (**A**) PFBV PTE showing locations of SL3sm and SL2sm mutations. Red line denotes residues with increased SHAPE reactivity in the stem of SL2 in gPTE-SL2sm, which may result from mutated loop residues now complementary to residues across the SL2 stem. (**B**) Translation of WT and mutant PFBV gRNAs in WGE. Locations are shown for p27 and readthrough protein p86, and p37, which can be synthesized from the gRNA by an unstructured IRES (54,55). Numbers denote mean levels of p27 and standard deviations from three independent experiments. SL3/2sm, PTE contains both SL3 and SL2 mutations. (**C**) SHAPE probing of the PTE region in PFBV gRNA. See Figure 1E legend for explanation. Average reactivity measurements for G* and SL3 and SL2 loops are given below the gel with standard deviations from three independent experiments. SL loop reactivities were measured as the summation of band reactivities comprising the entire loop and were normalized to invariant residues outside of the PTE region. (See Supplemental Figure S1 for the three replicate gels.) Red line denotes residues with inadvertent increased reactivity in the SL2 stem with the SL2sm mutation. (**D**) *Trans*-inhibition of PFBV gRNA translation in WGE by addition of WT or mutant PFBV PTE fragments (40-fold molar excess). Mean levels of p27 and standard deviations from three independent experiments are given.
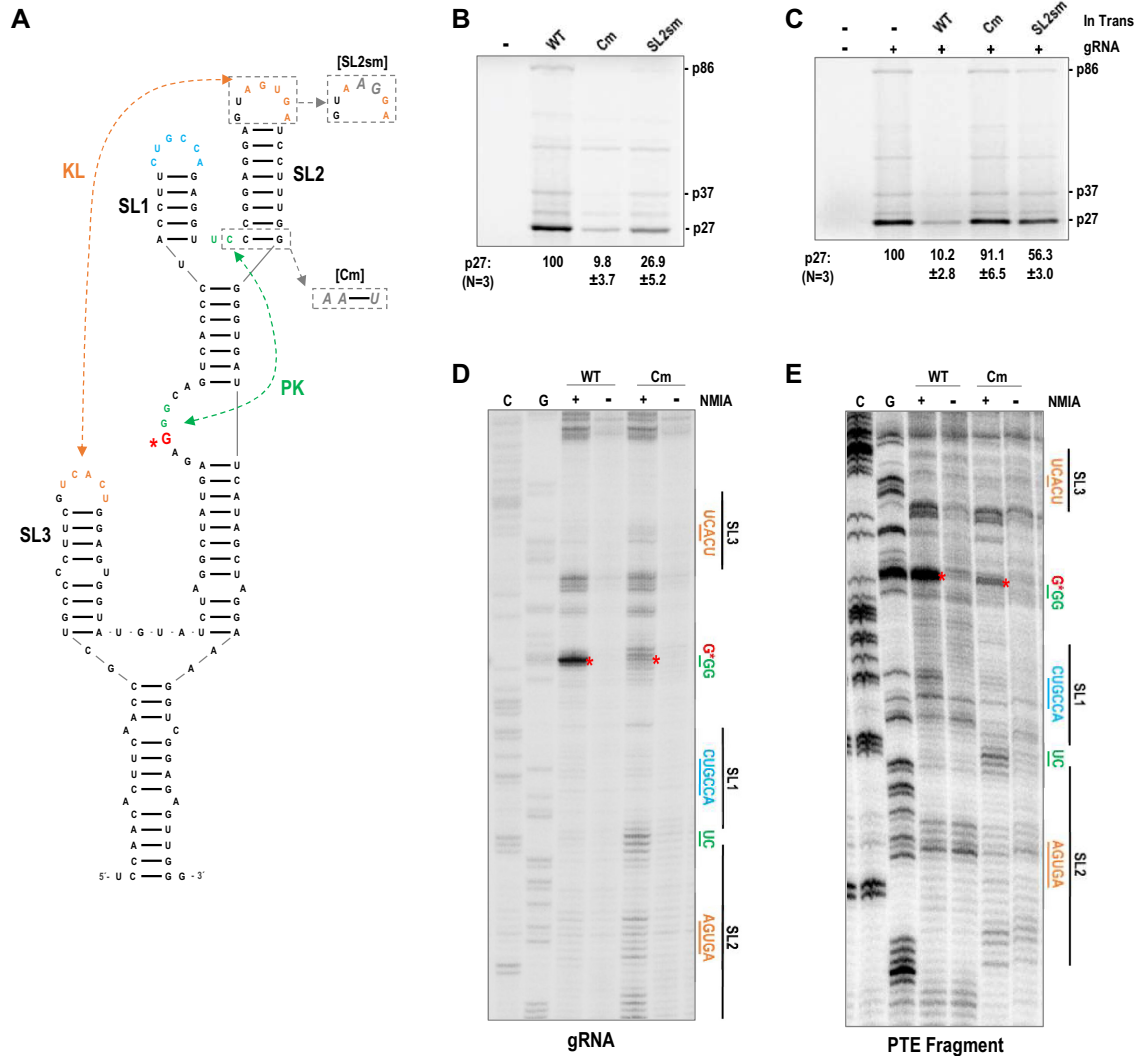
Figure 4. PFBV PTE contains both a KL and PK. (**A**) PFBV PTE with mutations used in this figure. (**B**) Translation of WT and mutant PFBV gRNAs in WGE. (**C**) Trans-inhibition of PFBV gRNA translation in WGE by WT or mutant PFBV PTE fragments (40-fold molar excess). For (B) and (C), means of p27 synthesis and standard deviations are given for three independent experiments. (**D**) SHAPE probing of the PTE region in PFBV gRNA. (**E**) SHAPE probing of isolated PFBV PTE fragment. See legend to Figure 1E for details.

in the loops of SL3 and SL2 remained low, suggesting that in the absence of other viral sequences, the KL remained stable when the PK was disrupted. G* hyperreactivity was also reduced in PTE-Cm, and reactivities of residues in the spacer between SL1 and SL2 increased, consistent with disruption of the PK in the fragment. Altogether, these results support a PFBV PTE structure that contains both the PK and the KL.

To determine if increasing the number of cytidylates between SL1 and SL2 in the PFBV PTE obviates the need for the KL by allowing a seemingly stronger PK to form, the pairing partners for the two cytidylates at the base of SL2 were changed from guanylates to adenylates and the stability of SL2 (in the absence of these two base-pairs) was increased by converting a G:U to a G:C in the stem (gPTE-wcPKm; Figure 5A). These alterations should result

in three consecutive cytidylates (5′UCCC) between SL1 and SL2 being available for PK pairing, similar to the 5′CCCU found between SL1 and SL2 in the PEMV2 PTE and other PTEs that lack a discernible SL3. SHAPE probing of gPTE-wcPKm indicated that these alterations eliminated the hyperreactive G* and increased reactivity of the sequence between SL1 and SL2 (Figure 5B), suggesting that the PK did not form effectively. Addition of the KL-disrupting SL3sm mutation to gPTE-wcPKm did not lead to a restoration of G*, indicating that the presence of a KL was not preventing the PK from forming. Altogether, these results suggest that for PTEs with seemingly weak PKs (i.e. fewer than three consecutive cytidylates between SL1 and SL2): (i) Both the PK and the KL exist; (ii) the KL appears necessary for efficient formation/stabilization of the PK and (iii) the KL can form in the absence of the PK in isolated PTE fragments.
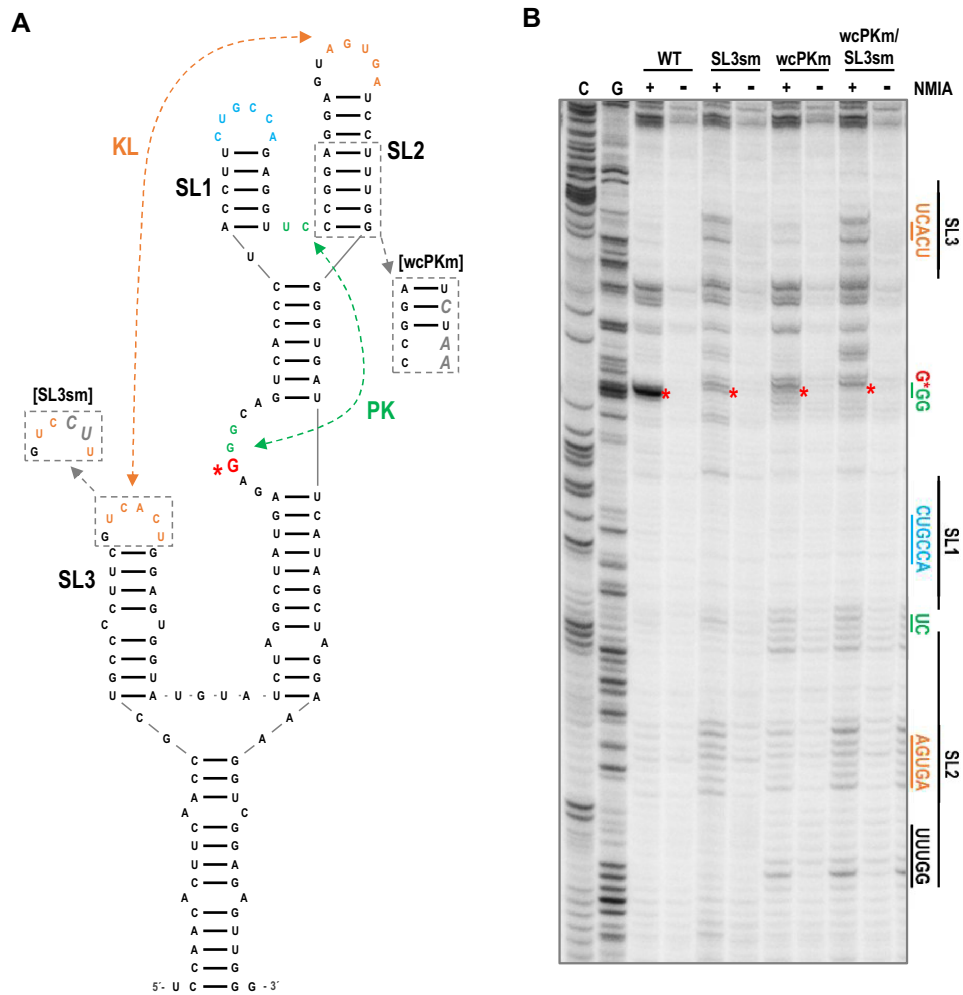
Figure 5. Three consecutive cytidylates between SL1 and SL2 cannot replace requirement for the KL. (**A**) PFBV PTE showing location of mutations used in this experiment. wcPKm mutations were designed to increase the number of cytidylates between SL1 and SL2 and stabilize the SL2 stem. (**B**) SHAPE probing of the PTE region in PFBV gRNA.

Mg^{2+} is not critical for PTE folding in the PFBV gRNA

Complexing with Mg^{2+} often significantly promotes RNA folding by supporting the formation of tertiary interactions such as PKs and KLs. It was previously determined that Mg^{2+} was required for correct folding of a PFBV PTE fragment, as determined by complete loss of the hyperreactive G* when Mg^{2+} was absent from the folding buffer (22). Although SL3 was not identified in this study, the fragment used in this report contained SL3 and several base-pairs in the lower stem (see Figure 2, top left, arrows). To test if Mg^{2+} is also required for proper folding of the PFBV PTE within the full-length gRNA, gPTE-WT was subjected to folding in normal folding buffer [(+)Mg^{2+}] and in folding buffer without Mg^{2+} to which 10 mM EDTA was also added to sequester any Mg^{2+} carried over from synthesis of the RNA [(–)Mg^{2+}]. SHAPE probing of gPTE-WT in the absence of Mg^{2+} revealed that the reactivities of SL1 loop residues increased markedly (Figure 6A; see also Supplemental Figure S2 for three replicate gels), suggesting that the long-distance interaction with the 5′end is disrupted in the absence of Mg^{2+}. In contrast with the previous study

using isolated PTE fragments (22), gPTE-WT G* retained most of its hyperreactivity (79%) in the absence of Mg^{2+} (Figure 6A), whereas direct disruption of the PK (gPTE-Cm) in the absence of Mg^{2+} reduced G* by an additional 3.6-fold (Figure 6B; see Supplemental Figure S3 for the three replicate gels). Similarly, lack of Mg^{2+} increased reactivities of residues in the loops of SL3 and SL2 by 4.1- and 8.9-fold, respectively (Figure 6A), while direct disruption of the KL (gPTE-SL2sm) in the absence of Mg^{2+} increased reactivities of these loop residues by an additional 2.2- and 1.9-fold, respectively (Figure 6B). These results suggest that both G* and the KL were not completely disrupted in the absence of Mg^{2+} when the PTE was examined within the full-length gRNA. These data imply that the PFBV PTE within the gRNA is not as dependent on Mg^{2+} for folding as it is as an isolated fragment.

G* and the adjacent 3′ guanylate are not Watson–Crick base-paired in the PFBV PTE PK

PTE G* residues were previously proposed to be unpaired and highly accessible to the solvent to account for their
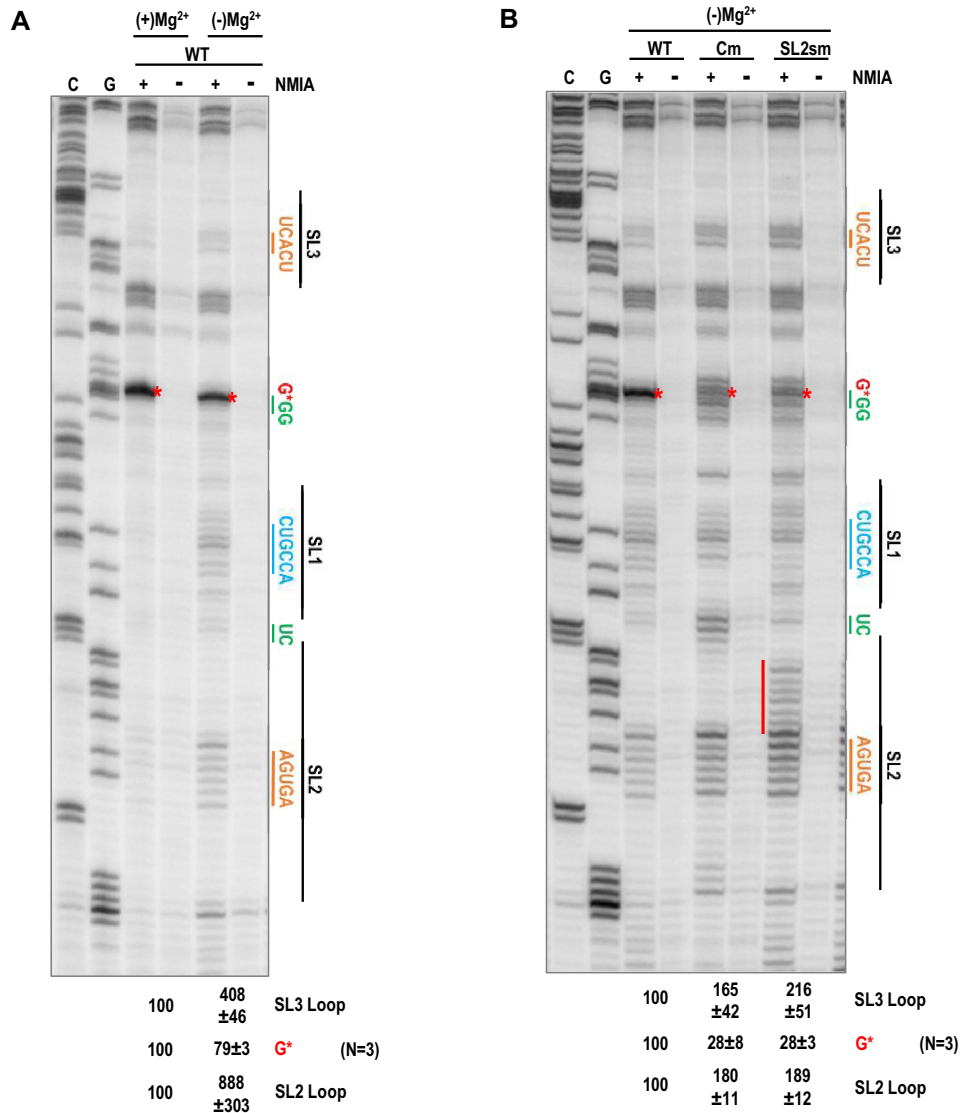
Figure 6. SHAPE probing of the PFBV PTE in the absence of Mg^{2+}. Mg^{2+} was either added to the folding buffer as usual [$(+)Mg^{2+}$], or absent from the folding buffer together to which 10 mM EDTA was also added [$(-)Mg^{2+}$]. Values below the gels reflect average densitometry tracings of G* and SL3 and SL2 loop reactivities with standard deviations from three independent trials for each experiment. Loop reactivities were measured as the summation of residue reactivities comprising the loop. (**A**) Comparison of $(+)Mg^{2+}$ and $(-)Mg^{2+}$ for gPTE-WT. (**B**) SHAPE probing of gPTE-Cm and gPTE-SL2sm. Red line denotes residues with inadvertent increased reactivity in the SL2 stem with the SL2sm mutation.

hyperreactivity to SHAPE reagents despite the presence of available Watson–Crick pairing partners in the PK (22). To test this hypothesis, we used the modifying reagent 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluenesulfonate (CMCT) to probe the PTE RNA structures of gPTE-WT, gPTE-Cm and gPTE-SL2sm. CMCT reacts with N-1 of guanylates and (more efficiently) with N-3 of uridylates (30,41,42), which reside on the Watson–Crick faces of their respective residues (Figure 7B). Therefore, CMCT only reacts with residues that are not engaged in Watson–Crick base-pairs. If the Watson–Crick face of G* is highly solvent accessible and not Watson–Crick base-paired, it would be expected to react strongly with CMCT.

As shown in Figure 7C, gPTE-WT G* had high reactivity with CMCT, though G* did not show hyperreactivity, in contrast to its hyperreactivity with SHAPE probing

reagents. This indicates that the Watson–Crick face of G* has high accessibility and suggests that G* is not Watson–Crick base-paired in the gPTE-WT PK (see Supplemental Figure S4 for full-length gel images of the three replicate probings.)

Unexpectedly, the guanylate immediately 3′ of G* (denoted G•; Figure 7A) was more reactive than G* with CMCT, indicating that the Watson–Crick face of G• is also unpaired and may have greater solvent accessibility. For gPTE-Cm and gPTE-SL2sm, which contain disrupted PKs, both G* and G• had reduced CMCT reactivities (Figure 7C). This result suggests that G* and G• gained alternative pairing/interacting partners for their Watson–Crick faces when the PK was disrupted. Interestingly, the SL2sm mutation (but not the Cm mutation) increased CMCT reactivity of the 5′GU immediately downstream of the asymmetric
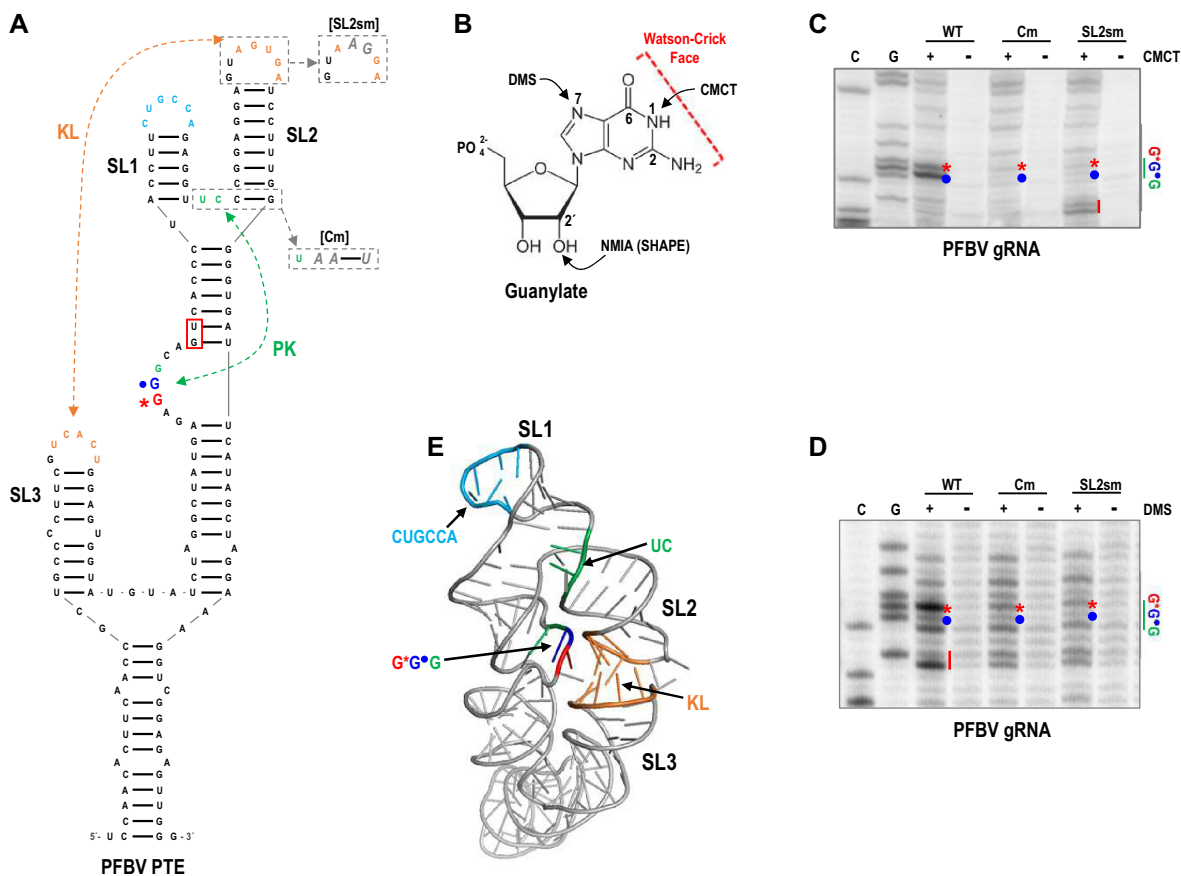
Figure 7. Chemical probing of the PFBV PTE with alternative modifying reagents. (**A**) PFBV PTE showing mutations used in this figure. SHAPE hyperreactive G* is in red and adjacent G• is in blue. Residues boxed in red showed differential reactivity with CMCT and DMS with the Cm and/or SL2sm mutations. (**B**) Guanylate showing locations that react with either SHAPE probing reagents (e.g., NMIA) (2′OH), DMS (N-7), and CMCT (N-1 on the Watson–Crick face). (**C**) CMCT probing of PTE region of PFBV WT and mutant gRNAs. (**D**) DMS probing of PTE region of PFBV WT and mutant gRNAs. Red asterisk, location of SHAPE hyperreactive G*; Blue dot, location of G•. Red line denotes residues boxed in (A). (**E**) Snapshot from molecular dynamics of the PFBV PTE. The initial model was generated by RNAComposer for which only the Y-shaped secondary structure and 5 bp of the KL were specified. The KL was stable throughout the three 500 ns molecular dynamic runs. Although the PK residues (G* in red, G• in blue and all others in green) (22) came in proximity during the runs, they never formed a clear PK interaction.

loop (Figure 7A, boxed in red; Figure 7C, red line), suggesting that the postulated alternative pairing disrupts this GU base-pairing across the PTE stem.

We also probed gPTE-WT, gPTE-Cm and gPTE-SL2sm with dimethyl sulfate (DMS). DMS reacts with N-7 of guanylates, N-1 of adenylates, and N-3 of cytidylates (30) (Figure 7B). N-1 of adenylates and N-3 of cytidylates reside on the Watson–Crick faces of their residues, while guanylate N-7 points into, and is accessible from the major groove of an A-form RNA helix. Therefore, DMS can react with guanylates engaged in Watson–Crick base-pairs and primarily reports on the presence or absence of non-canonical interactions involving guanylate N-7. gPTE-WT G* was highly reactive with DMS, indicating that G* N-7 in the WT PTE is accessible and not involved in a non-canonical interaction (Figure 7D). In contrast, G• had moderate to low DMS reactivity, suggesting that for the WT PTE, G• N-7 is possibly involved in a non-canonical interaction.

When gPTE-Cm and gPTE-SL2sm were probed with DMS, there were major reductions in G* reactivity and minor reductions in G• reactivity, suggesting that in the ab-

sence of the PK, G* N-7 may engage in non-canonical interaction(s). For both gPTE-Cm and gPTE-SL2sm, reduced reactivity was found for the guanylate (but not the uridylate) of the 5′GU immediately downstream of the asymmetric loop (Figure 7A, red box; Figure 7D, red line). This suggests that the conformation of these residues changed, as was also detected with CMCT probing of gPTE-SL2sm (Figure 7C). Altogether, these results suggest that: (i) PFBV PTE G* and G• residues are not Watson–Crick paired in the PK and have highly accessible Watson–Crick faces (with the Watson–Crick face of G• being significantly more accessible) and (ii) the PK of the PFBV PTE is likely composed of primarily non-canonical interactions given that both G* and G• are not Watson–Crick paired.

Modeling the PFBV PTE

We endeavored to generate a stable 3D model of the PFBV PTE containing both the PK and KL by initially producing models with the KL and without the PK using the 3D structure prediction program RNAComposer (31). The in-

put included the PTE secondary structure and 5 bp for the KL, but no PK. Molecular dynamics simulations (500 ns runs) suggested that the KL and overall secondary structure of the model were stable with an average occupancy of the KL base-pairs at nearly 100%. However, while the sequences proposed to participate in the PK came into proximity with each other during the runs (Figure 7E, molecular dynamics snapshot, and Supplemental Figure S5, minimum distance measurements between the sides of the PK), no clear PK interaction formed.

PFBV PTE models were also generated containing both the KL and PK where the PK was composed of varying combinations of Watson–Crick base-pairs between the three consecutive G residues in the asymmetric loop and the 5′UC or extended 5′UCCC residues between SL1 and extending into the base of SL2. However, molecular dynamics simulations of all these PFBV PTE models indicated that the KL was maintained but the PK was not stable, further suggesting that the PFBV PTE PK forms by non-canonical interactions in agreement with the CMCT probing (Figure 7C). It should also be noted that any modeling of PTEs and their PKs should maintain the accessibility of the asymmetric loop for possible interaction with eIF4E as previously proposed (22). Although our modeling of the PFBV PTE remains incomplete and lacks a clear PK interaction (Figure 7E), it does suggest that the KL can form while maintaining the accessibility of the asymmetric loop.

G* and G• are not Watson–Crick base-paired in the PEMV2 PTE PK

SHAPE probing of the WT PEMV2 PTE (seemingly strong PK/no SL3) within the gRNA (gPEMV-PTE-WT) agreed with the previously defined Y-shaped secondary structure (21,22) and confirmed the hyperreactivity of the same G* residue (Figure 8, A and B; note that in these publications, the G designated as hyperreactive was mislabeled on the PEMV2 PTE structure based on the SHAPE data presented). Mutation of two C residues between SL1 and SL2 that are proposed to participate in the PK (construct gPEMV-PTEm2; Figure 8A) reduced SHAPE reactivity of G* and increased reactivity of residues between SL1 and SL2 (CCCU), similar to PFBV gPTE-Cm. CMCT probing of gPEMV-PTE-WT produced a profile very similar to PFBV gPTE-WT, with G* exhibiting high CMCT reactivity and G• having higher CMCT reactivity despite the presence of available Watson–Crick pairing partners between SL1 and SL2 for both G* and G• (Figure 8C; see Supplemental Figure S6 for full-length gel images of the three replicate probings). The two guanylates neighboring G* and G• in the asymmetric loop also exhibited moderate to low reactivities with CMCT. G* and G• reactivities for CMCT in gPEMV-PTEm2 were also similar to PFBV gPTE-Cm, with large reductions in CMCT reactivity for G* and G•, denoting possible alternative pairing/interacting partners.

DMS reactivity of G* for gPEMV-PTE-WT showed moderate reactivity that was comparable to the reactivities of the other G residues in the asymmetric loop including G• (Figure 8D). Reactivity levels were only slightly reduced in the absence of the PK (gPEMV-PTEm2), suggesting that G* and G• N-7 remained unpaired in the absence of the PK.

These results reveal that the CMCT reactivity patterns for G* and G• are similar for both PFBV and PEMV2 PTEs within the gRNAs, suggesting a shared conformation involving G*, G• and the PK despite these two PTEs differing in the seeming strength of their PKs and the presence or absence of SL3 and the KL interaction.

Mg^{2+} is required to form the PK in full-length PEMV2 gRNA

As shown in Figure 6, the PFBV PTE within the gRNA is mostly stable in the absence of Mg^{2+}. To investigate if a PTE that forms in the absence of a KL is similarly stable in the absence of Mg^{2+} within the gRNA, gPEMV-PTE-WT was subjected to folding in the absence of Mg^{2+}, and the SHAPE profile compared with that of gPEMV-PTEm2. Contrary to results found for the PFBV PTE, G* hyperreactivity in gPEMV-PTE-WT was nearly abolished in the absence of Mg^{2+} (Figure 8E). These results suggest that the PEMV2 PTE is less stable than the PFBV PTE within the gRNA, inferring that the KL may offer increased stability to a PTE.

The absence of a Watson–Crick base-pair involving G• correlates with an A-C mismatch in the PK of the JINRV PTE

CMCT probing of PFBV and PEMV2 PTEs (Figures 7C and 8C) indicate that G•, in addition to G*, lacks a Watson–Crick base-pairing partner, despite the presence of potential C/U partners between SL1 and SL2. This finding suggests that the unpaired G• residue just downstream of G* might not need to be a guanylate to form the PK required for a functional PTE for some viruses. While most PTEs contain three consecutive guanylates in their asymmetric loops, the PTE in betacarmovirus Japanese iris necrotic ring virus (JINRV) is an exception. The JINRV PTE has an asymmetric loop consisting of 5′GCG*<u>A</u>GAUG (the normal G• position is underlined) for pairing with 5′UGACCC between SL1 and SL2 (Figure 9A). As with other PTEs with three cytidylates between SL1 and SL2, the JINRV PTE does not contain a discernible SL3 or KL, and thus the PK must exist in a stable conformation with 5′G*AG in its asymmetric loop. SHAPE probing of the JINRV PTE fragment (Figure 9B; JINRV gRNA was not available) was consistent with its previously proposed Y-shaped secondary structure (22) and confirmed the hyperreactivity of the G* residue shown in Figure 9A. Altering the 3′ terminal cytidylate between SL1 and SL2 to an adenylate (generating JIN-PTEmPK) significantly reduced G* reactivity (Figure 9B), suggesting that the PK involves at least this cytidylate. The asymmetric loop adenylate is in the position corresponding to G• in PFBV and PEMV2, resulting in an A-C mismatch and a PK that may also be formed by primarily non-canonical interactions.

Coarse-grained simulation using SimRNA (32) was used to attempt the generation of a 3D model of the JINRV PTE and its PK, and the stability of the model was further tested by molecular dynamics. The SimRNA input explicitly specified the Y-shaped secondary structure and a single Watson–Crick base-pair in the PK between G* and the C residue just upstream of SL2 (i.e. just downstream of the A-C mismatch). All other PK elements (if any) in the model generated by SimRNA were allowed to form passively. Though
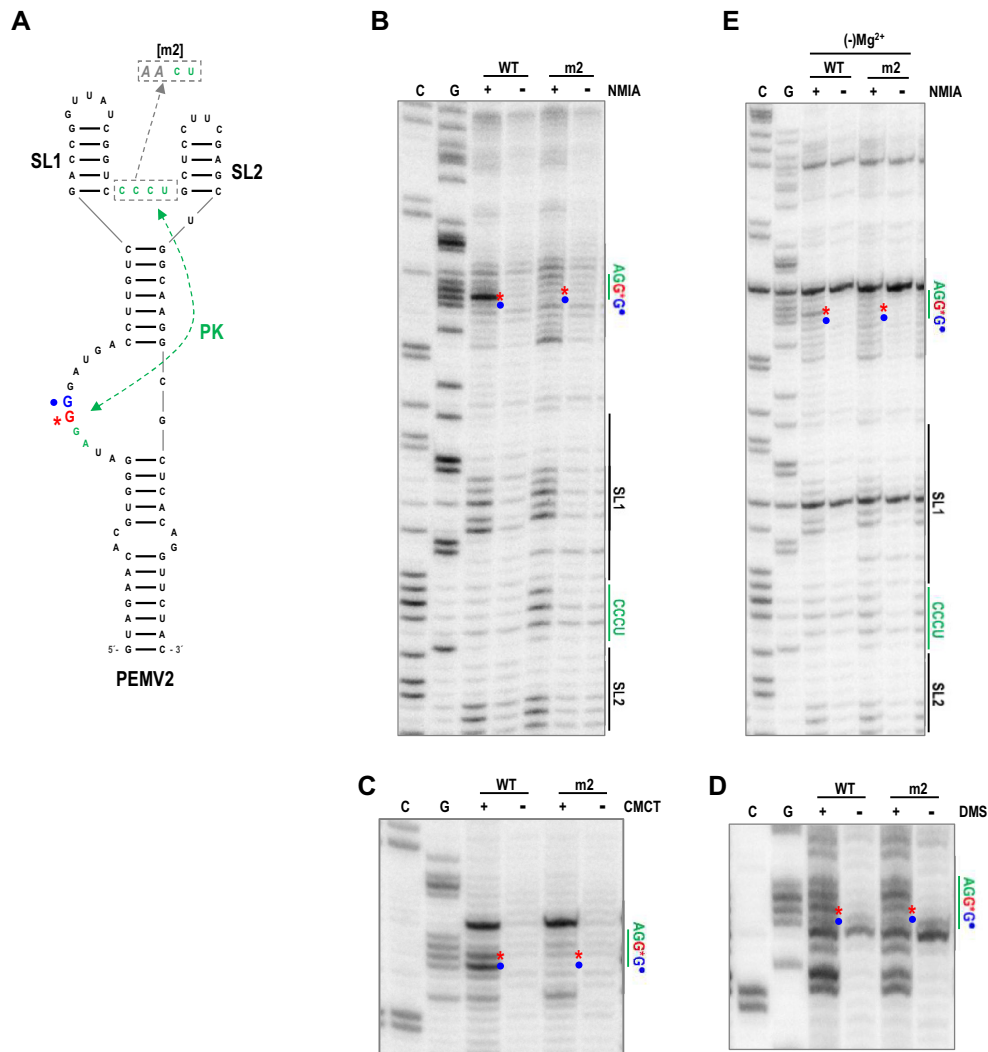
Figure 8. PEMV2 PTE in the gRNA also has unpaired G* and G• residues. (A) PEMV2 PTE with mutations used in this figure. The m2 mutation was previously shown to disable the PEMV2 PTE by disrupting the PK (21). The SHAPE hyperreactive G* and adjacent G• are denoted. (B) SHAPE probing under normal conditions (i.e. with Mg²⁺ in the folding buffer). Red asterisk denotes G* and blue dot denotes G•. Note that SL1 loop residues are reactive as they do not participate in any long-distance interaction. (C) Probing with CMCT. (D) Probing with DMS. (E) SHAPE probing in which Mg²⁺ was withheld from the folding buffer [(–)Mg²⁺]. No EDTA was added to the folding buffer in this experiment, making these conditions less stringent than those used to probe the PFBV PTE shown in Figure 6.

CMCT probing of the PFBV and PEMV2 PTEs suggests that PTE G* residues are not Watson–Crick paired (Figures 7C and 8C), specifying a Watson–Crick base-pair involving G* in the PK of the JINRV PTE allowed for formation of a PK interaction in the initial model produced by SimRNA with the potential to be further refined. In three 500 ns molecular dynamics runs using the Amber suite (36), however, the Watson–Crick base-pair in the PK involving G* was highly stable as measured by occupancies close to 100% (Figure 9C), preventing further refinement of the model towards one possessing an unpaired G*.

Modeling PTEs that have three cytidylates between SL1 and SL2 and no SL3

We also generated 3D models of the PEMV2 PTE and PTEs from PMV, cocksfoot mild mottle virus (CMMV), car-nation mottle virus (CarMV), and thin paspalum asymptomatic virus (TPAV) by coarse-grained simulation using SimRNA. All of these PTEs have no discernible SL3 and contain three cytidylates between SL1 and SL2 available for base-pairing in the PK (Figure 10A). The input to SimRNA contained only the Y-shaped secondary structures. Thus, in all these models, the PK formed passively in coarse-grained simulation using SimRNA. In molecular dynamics using the Amber suite, the PKs of all PTE models were stable (Figure 10B). Interestingly, in all PTE models produced by SimRNA, G* and G• residues were Watson–Crick paired in the PK, and in all molecular dynamics runs, G* had Watson–Crick base-pair occupancy near 100% and, with the single exception of one molecular dynamics run of the TPAV PTE (Figure 10C), G• also had Watson–Crick base-pair occupancy near 100%. These results disagree with CMCT probing of the
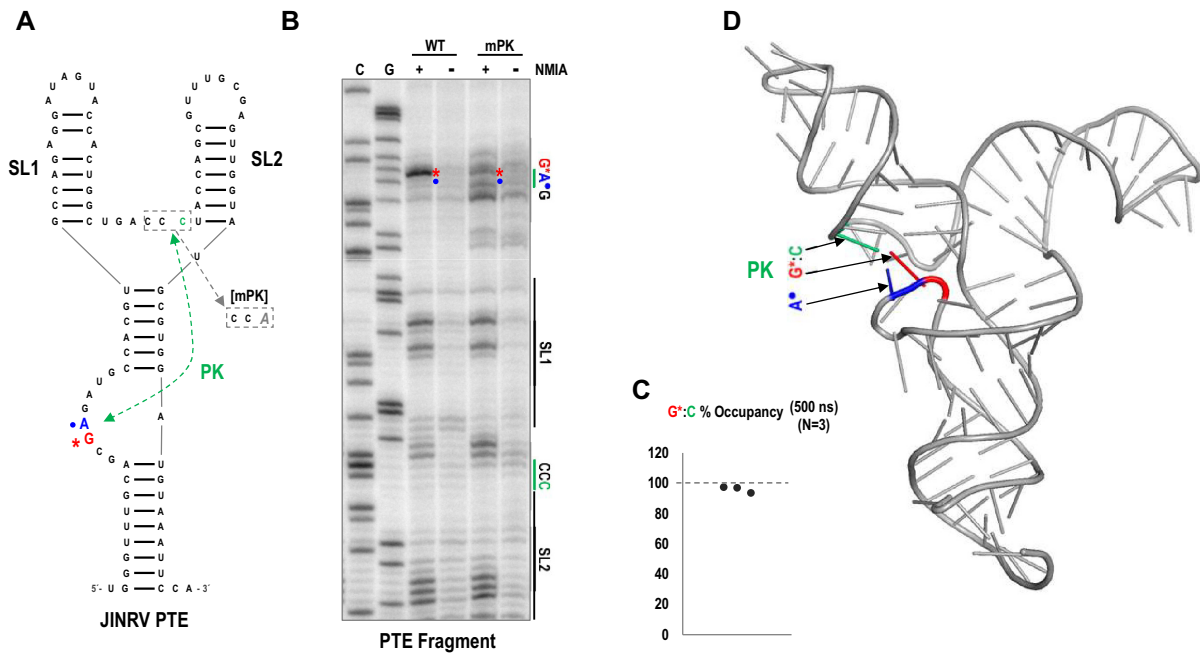
Figure 9. SHAPE probing and modeling of the JINRV PTE. (**A**) JINRV PTE with the mutations used in this figure. (**B**) SHAPE probing of the JINRV PTE fragment shown in (A). (**C**) Molecular dynamics of the JINRV PTE. Initial model generated by coarse-grained simulation using SimRNA contained the Y-shaped secondary structure and the Watson–Crick base-pair between G* and the cytidylate immediately 5′ of SL2 in the PK. All other interactions were allowed to form passively. Each dot in the graph indicates occupancy of the Watson–Crick base-pair involving G* in one of the three 500 ns molecular dynamics runs. (**D**) Snapshot from molecular dynamics in (C). Residue colorings are as in (A).

PFBV and PEMV2 PTEs (Figures 7C and 8C), which showed high CMCT reactivities for G* and G• residues that strongly suggested no Watson–Crick pairing for G* and G•.

To remove the Watson–Crick base-pair involving G• (which had higher CMCT reactivity than G*) and perhaps obtain more accurate PK conformations in our PTE models, we mutated G• to an adenylate *in silico* using PyMOL and repeated molecular dynamics on the mutated models (Figure 10D). For all PTEs except for the TPAV PTE, the PKs of these altered models remained stable, suggesting that a base-pair involving G• is indeed unnecessary for PK stability. However, Watson–Crick base-pair occupancies for G* remained close to 100% for all but the TPAV PTE. Furthermore, the PK of the altered TPAV PTE was noticeably less stable with base-pair occupancies <50%, suggesting that the TPAV PTE may be unable to accept an adenylate replacing G•. Interestingly, the PK of the WT TPAV PTE remained stable when the occupancy of its G• base-pair was only 52% in one run of molecular dynamics (Figure 10B). At several times during this run, G• was no longer Watson–Crick base-paired (Figure 10C), possibly revealing a PK conformation that better agrees with the CMCT probing of the PFBV and PEMV2 PTEs (Figures 7C and 8C). Considering that all PKs were required to form passively in WT PTEs and contained Watson–Crick paired G* and G• residues, more detailed modeling with manual intervention and further constraints may be necessary to more reliably produce models with stable PKs comprised of non-Watson–Crick paired G* and G• residues.

DISCUSSION

PTE structures with fewer than three cytidylates between SL1 and SL2 have an upstream SL3

3′CITEs are critical components of translation regulation that allow many plant RNA viruses to compete with ongoing cap-dependent host mRNA translation, which is necessary as plant viruses must exit from a living cell during systemic infection of the host. By understanding the structure of 3′CITEs and how they attract the host translation machinery, new facets of cap-independent translation have been uncovered that may also be relevant for translation of a subset of host mRNAs. Most studies on the structure and function of 3′CITEs use isolated fragments or reporter constructs that separate the presumptive complete element from the remainder of the viral genome. By examining the PFBV PTE *in situ*, we discovered an additional upstream stem-loop (SL3) whose apical loop forms a KL with the apical loop of the PTE SL2. SL3s in similar locations relative to core PTE structures and with apical loops capable of forming a KL with SL2 correlate with PTEs that do not contain three cytidylate residues between SL1 and SL2 for pairing with the G-rich asymmetric loop (Figure 2).

We determined that the KL was important for the formation and/or stabilization of the PFBV PTE PK, which involves residues in the G-rich asymmetric loop and C/U residues between SL1 and SL2. In the absence of the KL, the PK is disrupted and SHAPE hyperreactivity of G* and the high CMCT reactivities of G* and G• are not present (Figures 3C, 7C, 8B and C). However, it should be noted that mutations designed to directly eliminate PFBV PK for-
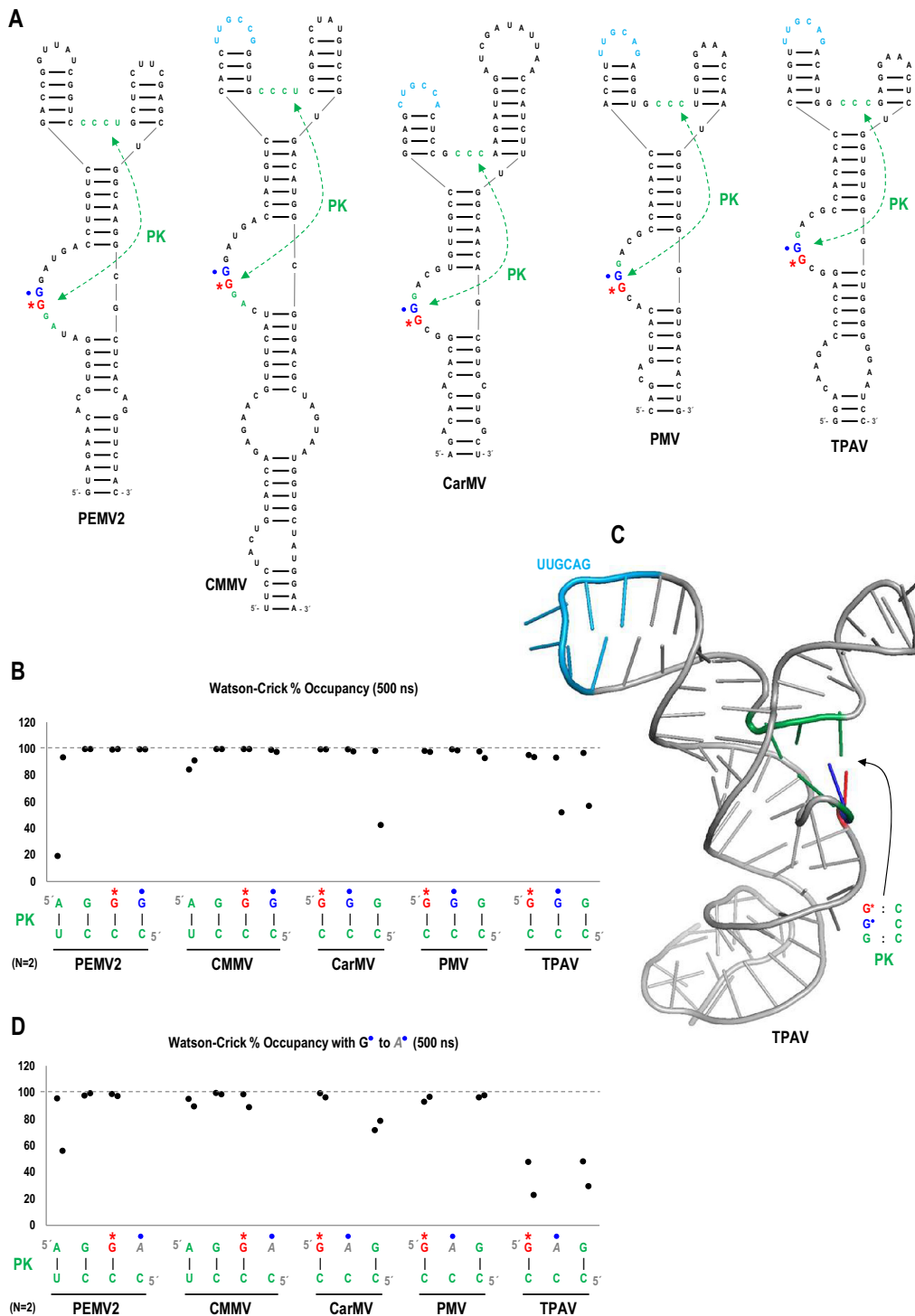
Figure 10. Modeling PTEs with no discernible SL3 and three cytidylates between SL1 and SL2. (**A**) Secondary structures of the PTEs used for this analysis. PEMV2 (umbravirus); CMMV, cocksfoot mild mosaic virus (panicovirus); CarMV, carnation mottle virus (alphacarmovirus); PMV, panicum mosaic virus (panicovirus); TPAV, thin paspalum asymptomatic virus (panicovirus). SHAPE hyperreactive G* previously identified are in red (22) and G• are in blue. SL1 loops containing 5′UGCCA or related motifs known or proposed to participate in long-distance base-pairing are in light blue (24,26). (**B**) Molecular dynamics of the PTEs shown in (A). Initial models were generated by coarse-grained simulation using SimRNA with only the Y-shaped secondary structures specified. All PKs in the models generated by SimRNA formed passively in coarse-grained simulation. Each dot in the graph indicates the occupancy of a Watson–Crick base-pair in the PK in one 500 ns molecular dynamics run. (**C**) Timepoint from molecular dynamics of the TPAV PTE at which the base-pair involving G• was broken, while the base-pair involving G* was entirely present (with all three hydrogen bonds comprising a guanylate-cytidylate base-pair) at this timepoint. The base-pair involving the guanylate immediately 3′ to G• (in green) was also partially broken at this timepoint being comprised of only the hydrogen bond involving guanylate N-2 and cytidylate O-2. (**D**) Molecular dynamics of PTEs with G• mutated to an adenylate. Initial models for these molecular dynamics were timepoints from the molecular dynamics of WT PTEs shown in (B) with the G• mutation introduced using the mutagenesis wizard in PyMOL.

mation (gPTE-Cm) resulted in translation levels 2.7-fold less than mutations that directly disrupted the KL (gPTE-SL2sm). In addition, PFBV PTE-Cm was less able to interfere with translation when added *in trans* (Figure 4B and C). These results suggest that some functional PTE still exist in the absence of the KL and that the KL primarily serves an accessory role to stabilize the PK. Efforts to validate the presence of the KL *in vivo* were not successful as we were unable to infect our standard Arabidopsis protoplast system with RNA transcripts synthesized from the full-length construct obtained using sequence deposited in GenBank.

Comparing the structures of the PFBV PTE within the gRNA and in isolated fragments revealed three major structural differences: (i) In isolated fragments, but not within the gRNA, the KL formed independently of the PK (Figure 4D and E); (ii) the PFBV PTE fragment tested previously (Figure 2; PFBV; blue arrows) required Mg^{2+} in the folding buffer to fold correctly (22), whereas the PFBV PTE in the context of the gRNA was much less reliant on Mg^{2+} in the folding buffer (Figure 6) and (iii) residues in SL1 were less reactive in SHAPE probing in the gRNA than in the isolated fragment, consistent with these residues engaging in a long-distance RNA:RNA interaction with a complementary sequence in the loop of a 5′ region hairpin in the gRNA (Figure 4D and E). The structural differences described in (i) and (ii) suggest that additional as yet unknown elements in gRNAs may interact with the PFBV PTE structure that reduce its dependence on Mg^{2+} for folding and destabilize the KL when the PK is absent. The ability of the KL to form in the absence of the PK, at least in the isolated fragment, also suggests a folding pathway for the PFBV PTE in which the KL forms before the PK.

G* and G• accessibilities are consistent with SHAPE chemistry

Previous models of the PEMV2 PTE suggested that G* is unpaired and highly accessible to the solvent based on its hyperreactivity to SHAPE reagents (21,22). In CMCT probing of the PFBV and PEMV2 PTEs (Figures 7C and 8C), G* indeed had high reactivity with CMCT suggestive of a highly accessible, unpaired Watson–Crick face. However, G* was not hyperreactive with CMCT and had lower CMCT reactivity than the adjacent, downstream G•, despite G• having much lower SHAPE reactivity than G*. Higher CMCT reactivity suggests that G• may have greater solvent accessibility than G*, which is consistent with the findings of previous studies addressing the mechanisms of SHAPE chemistry and the causes of residue hyperreactivity (43–46), as well as large-scale analyses of residue accessibility (47). These studies found that SHAPE hyperreactivity has minimal correlation with the solvent accessibility of a residue and that hyperreactive residues can even be base-paired. SHAPE reactivity typically measures the flexibility of a residue to adopt various conformations, including rare conformations that promote the modification reaction. Because of this, unpaired residues typically have high SHAPE reactivities due to their flexibility in sampling different conformations. However, hyperreactivity in SHAPE probing was proposed to signify that a residue is constrained in one of the rare conformations that promote the modification

reaction, or by the presence of a nearby group that catalyzes the modification reaction. Thus, solvent accessibility may not directly correlate with SHAPE reactivity, as suggested by the relative CMCT reactivities of PTE G* and G• residues.

PTEs with and without SL3 share a similar PK conformation

We have determined that the PTE PKs of PFBV and PEMV2 share similar conformations of G* and G•. PFBV and PEMV2 G* and G• residues had high CMCT reactivities, suggesting that both G* and G• are not Watson–Crick paired with another residue. The JINRV PTE has an adenylate at the G• location, causing an A–C mismatch that may be analogous to G• as well. PFBV and PEMV2 G• residues also had significantly higher CMCT reactivity than G*, suggesting that G• may have higher solvent accessibility and that this non-Watson–Crick paired residue may bind the cap-binding site of eIF4E, as was also proposed previously for G* (22). It remains uncertain, though, if an adenylate at the G• location, such as in the JINRV PTE, may bind the cap-binding site in lieu of a guanylate. In prior work that used 3D modeling to dock eIF4E to the residue identified as the hyperreactive G in PEMV2 (22), the hyperreactive G was incorrectly designated on the PTE structure, and was actually the residue identified here as G•. Previous mutation of G• to an adenylate in a luciferase reporter construct was also the least detrimental of all the alterations made to G• and the three adjacent guanylates, retaining 25.6% of translation levels *in vitro* and ≈40% *in vivo* relative to the unmutated construct (21). In addition, they found that an adenylate replacing G• did not affect the hyperreactive G* in SHAPE probing, suggesting that the PK remained intact, while replacing any of the three adjacent guanylates with an adenylate abolished G* hyperreactivity. When taking into account the misidentification of G*, this previous study further suggests that certain PTE structures can accept an adenylate at the G• location and remain functional and that an adenylate may possibly occupy the cap-binding site of eIF4E. Our inability to generate models of PTEs with both the KL and PK suggests that unusual conformations of nucleotides exist in the PK that likely contribute to its overall structure and ability to bind to eIF4E, which will only be revealed by more in-depth studies.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Jackson,R.J., Hellen,C.U. and Pestova,T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, **11**, 113–127.
2. Browning,K.S. and Bailey-Serres,J. (2015) Mechanism of cytoplasmic mRNA translation. *Arabidopsis Book*, **13**, e0176.
3. Amrani,N., Ghosh,S., Mangus,D.A. and Jacobson,A. (2008) Translation factors promote the formation of two states of the closed-loop mRNP. *Nature*, **453**, 1276–1280.
4. Wells,S.E., Hillner,P.E., Vale,R.D. and Sachs,A.B. (1998) Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell*, **2**, 135–140.
5. Firth,A.E. and Brierley,I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.*, **93**, 1385–1409.
6. Miras,M., Miller,W.A., Truniger,V. and Aranda,M.A. (2017) Non-canonical translation in plant RNA viruses. *Front. Plant Sci.*, **8**, 494.
7. Shatsky,I.N., Dmitriev,S.E., Andreev,D.E. and Terenin,I.M. (2014) Transcriptome-wide studies uncover the diversity of modes of mRNA recruitment to eukaryotic ribosomes. *Crit. Rev. Biochem. Mol. Biol.*, **49**, 164–177.
8. Nicholson,B.L. and White,K.A. (2011) 3' Cap-independent translation enhancers of positive-strand RNA plant viruses. *Curr. Opin. Virol.*, **1**, 373–380.
9. Simon,A.E. and Miller,W.A. (2013) 3' cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.*, **67**, 21–42.
10. Guo,L., Allen,E.M. and Miller,W.A. (2001) Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. *Mol. Cell*, **7**, 1103–1109.
11. Wang,S. and Miller,W.A. (1995) A sequence located 4.5 to 5 kilobases from the 5' end of the barley yellow dwarf virus (PAV) genome strongly stimulates translation of uncapped mRNA. *J. Biol. Chem.*, **270**, 13446–13452.
12. Wang,S., Browning,K.S. and Miller,W.A. (1997) A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. *EMBO J.*, **16**, 4107–4116.
13. Ilyas,M., Du,Z. and Simon,A.E. (2021) Opium poppy mosaic virus has an Xrn-resistant, translated subgenomic RNA and a BTE 3' CITE. *J. Virol.*, **95**, e02109-20.
14. McCormack,J.C., Yuan,X., Yingling,Y.G., Kasprzak,W., Zamora,R.E., Shapiro,B.A. and Simon,A.E. (2008) Structural domains within the 3' untranslated region of turnip crinkle virus. *J. Virol.*, **82**, 8706–8720.
15. Stupina,V.A., Meskauskas,A., McCormack,J.C., Yingling,Y.G., Shapiro,B.A., Dinman,J.D. and Simon,A.E. (2008) The 3' proximal translational enhancer of turnip crinkle virus binds to 60S ribosomal subunits. *RNA*, **14**, 2379–2393.
16. Gao,F., Kasprzak,W.K., Szarko,C., Shapiro,B.A. and Simon,A.E. (2014) The 3' untranslated region of pea enation mosaic virus contains two T-shaped, ribosome-binding, cap-independent translation enhancers. *J. Virol.*, **88**, 11696–11712.
17. Stupina,V.A., Yuan,X., Meskauskas,A., Dinman,J.D. and Simon,A.E. (2011) Ribosome binding to a 5' translational enhancer is altered in the presence of the 3' untranslated region in cap-independent translation of turnip crinkle virus. *J. Virol.*, **85**, 4638–4653.
18. Wang,D.Y., Yu,C.M., Liu,S.S., Wang,G.L., Shi,K.R., Li,X.D. and Yuan,X.F. (2017) Structural alteration of a BYDV-like translation element (BTE) that attenuates p35 expression in three mild tobacco bushy top virus isolates. *Sci. Rep.*, **7**, 4213.
19. Danthinne,X., Seurinck,J., Meulewaeter,F., Van Montagu,M. and Cornelissen,M. (1993) The 3' untranslated region of satellite tobacco necrosis virus RNA stimulates translation in vitro. *Mol. Cell. Biol.*, **13**, 3340–3349.
20. Timmer,R.T., Benkowski,L.A., Schodin,D., Lax,S.R., Metz,A.M., Ravel,J.M. and Browning,K.S. (1993) The 5' and 3' untranslated regions of satellite tobacco necrosis virus RNA affect translational efficiency and dependence on a 5' cap structure. *J. Biol. Chem.*, **268**, 9504–9510.
21. Wang,Z., Treder,K. and Miller,W.A. (2009) Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. *J. Biol. Chem.*, **284**, 14189–14202.
22. Wang,Z., Parisien,M., Scheets,K. and Miller,W.A. (2011) The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. *Structure*, **19**, 868–880.
23. Kraft,J.J., Peterson,M.S., Cho,S.K., Wang,Z., Hui,A., Rakotondrafara,A.M., Treder,K., Miller,C.L. and Miller,W.A. (2019) The 3' untranslated region of a plant viral RNA directs efficient cap-independent translation in plant and mammalian systems. *Pathogens*, **8**, 28.
24. Chattopadhyay,M., Shi,K., Yuan,X. and Simon,A.E. (2011) Long-distance kissing loop interactions between a 3' proximal Y-shaped structure and apical loops of 5' hairpins enhance translation of saguaro cactus virus. *Virology*, **417**, 113–125.
25. Chattopadhyay,M., Kuhlmann,M.M., Kumar,K. and Simon,A.E. (2014) Position of the kissing-loop interaction associated with PTE-type 3'CITEs can affect enhancement of cap-independent translation. *Virology*, **458**, 43–52.
26. Gao,F., Kasprzak,W., Stupina,V.A., Shapiro,B.A. and Simon,A.E. (2012) A ribosome-binding, 3' translational enhancer has a T-shaped structure and engages in a long-distance RNA-RNA interaction. *J. Virol.*, **86**, 9828–9842.
27. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
28. Liu,H. and Naismith,J.H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.*, **8**, 91.
29. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH image to imagej: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
30. Ziehler,W.A. and Engelke,D.R. (2001) Probing RNA structure with chemical reagents and enzymes. *Curr Protoc Nucleic Acid Chem*, **Chapter 6**, Unit 6.1.
31. Biesiada,M., Purzycka,K.J., Szachniuk,M., Blazewicz,J. and Adamiak,R.W. (2016) Automated RNA 3D structure prediction with RNAComposer. *Methods Mol. Biol.*, **1490**, 199–215.
32. Boniecki,M.J., Lach,G., Dawson,W.K., Tomala,K., Lukasz,P., Soltysinski,T., Rother,K.M. and Bujnicki,J.M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic. Acids. Res.*, **44**, e63.
33. Cheatham,T.E. III, Miller,J.L., Fox,T., Darden,T.A. and Kollman,P.A. (1995) Molecular dynamics simulations on solvated biomolecular systems: the particle mesh ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.*, **117**, 4193–4194.
34. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh ewald: an N·log(N) method for ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
35. Essmann,U., Perera,L. and Berkowitz,M.L. (1995) A smooth particle mesh ewald method. *J. Chem. Phys.*, **103**, 8577–593.
36. Case,D.A., Ben-Shalom,I.Y., Brozell,S.R., Cerutti,D.S., Cheatham,T.E. III, Cruzeiro,V.W.D., Darden,T.A., Duke,R.E., Ghoreishi,D., Gilson,M.K. *et al.* (2018) *Amber 2018*. University of California, San Francisco.
37. Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
38. Kim,T., Kasprzak,W.K. and Shapiro,B.A. (2017) Protocols for molecular dynamics simulations of RNA nanostructures. *Methods Mol. Biol.*, **1632**, 33–64.

39. Roe, D.R. and Cheatham, T.E., 3rd. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
40. Johnson,P.Z., Kasprzak,W.K., Shapiro,B.A. and Simon,A.E. (2019) RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA Biol*, **16**, 1667–1671.
41. Metz,D.H. and Brown,G.L. (1969) The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and model nucleotides. *Biochemistry*, **8**, 2312–2328.
42. Metz,D.H. and Brown,G.L. (1969) The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. II. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and transfer ribonucleic acid. *Biochemistry*, **8**, 2329–2342.
43. McGinnis,J.L., Dunkle,J.A., Cate,J.H. and Weeks,K.M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
44. Mlýnský,V. and Bussi,G. (2018) Molecular dynamics simulations reveal an interplay between SHAPE reagent binding and RNA flexibility. *J. Phys. Chem. Lett.*, **9**, 313–318.
45. Vicens,Q., Gooding,A.R., Laederach,A. and Cech,T.R. (2007) Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA*, **13**, 536–548.
46. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
47. Yang,Y., Li,X., Zhao,H., Zhan,J., Wang,J. and Zhou,Y. (2017) Genome-scale characterization of RNA tertiary structures and their functional impact by RNA solvent accessibility prediction. *RNA*, **23**, 14–22.
48. Demler,S.A., Rucker,D.G. and de Zoeten,G.A. (1993) The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *J. Gen. Virol.*, **74**, 1–14.
49. Gao,F. and Simon,A.E. (2016) Multiple Cis-acting elements modulate programmed -1 ribosomal frameshifting in pea enation mosaic virus. *Nucleic Acids Res.*, **44**, 878–895.
50. Gao,F., Alekhina,O.M., Vassilenko,K.S. and Simon,A.E. (2018) Unusual dicistronic expression from closely spaced initiation codons in an umbravirus subgenomic RNA. *Nucleic Acids Res.*, **46**, 11726–11742.
51. Du,Z., Alekhina,O.M., Vassilenko,K.S. and Simon,A.E. (2017) Concerted action of two 3' cap-independent translation enhancers increases the competitive strength of translated viral genomes. *Nucleic Acids Res.*, **45**, 9558–9572.
52. Rico,P. and Hernández,C. (2004) Complete nucleotide sequence and genome organization of pelargonium flower break virus. *Arch. Virol.*, **149**, 641–651.
53. Rico,P. and Hernández,C. (2009) Characterization of the subgenomic RNAs produced by pelargonium flower break virus: identification of two novel RNAs species. *Virus Res.*, **142**, 100–107.
54. Fernández-Miragall,O. and Hernández,C. (2011) An internal ribosome entry site directs translation of the 3'-gene from pelargonium flower break virus genomic RNA: implications for infectivity. *PLoS One*, **6**, e22617.
55. May,J., Johnson,P., Saleem,H. and Simon,A.E. (2017) A sequence-independent, unstructured internal ribosome entry site is responsible for internal expression of the coat protein of turnip crinkle virus. *J. Virol.*, **91**, e02421-16.