https://doi.org/10.1093/nar/gkad649

Crystal structure of a cap-independent translation enhancer RNA

Anna Lewicka [1,†], Christina Roman [1,†], Stacey Jones [1], Michael Disare [2], Phoebe A. Rice [1] and Joseph A. Piccirilli [1,2,*]

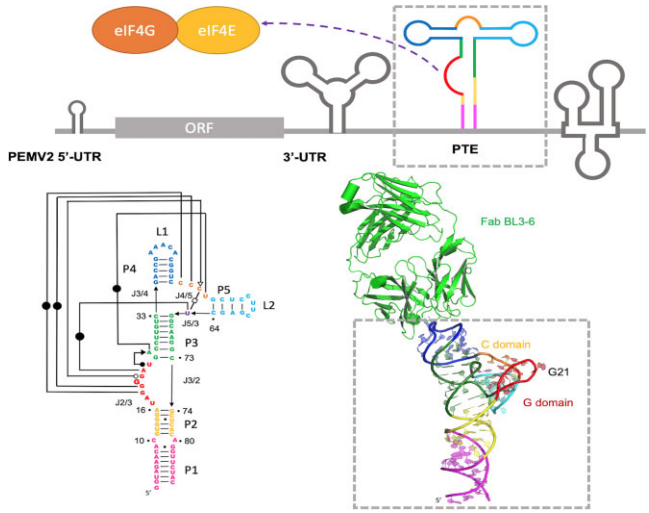
[1]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA and [2]Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA

Received April 20, 2023; Revised July 14, 2023; Editorial Decision July 19, 2023; Accepted July 28, 2023

ABSTRACT

In eukaryotic messenger RNAs, the 5′ cap structure binds to the translation initiation factor 4E to facilitate early stages of translation. Although many plant viruses lack the 5′ cap structure, some contain cap-independent translation elements (CITEs) in their 3′ untranslated region. The PTE (Panicum mosaic virus translation element) class of CITEs contains a G-rich asymmetric bulge and a C-rich helical junction that were proposed to interact via formation of a pseudoknot. SHAPE analysis of PTE homologs reveals a highly reactive guanosine residue within the G-rich region proposed to mediate eukaryotic initiation factor 4E (eIF4E) recognition. Here we have obtained the crystal structure of the PTE from Pea enation mosaic virus 2 (PEMV2) RNA in complex with our structural chaperone, Fab BL3–6. The structure reveals that the G-rich and C-rich regions interact through a complex network of interactions distinct from those expected for a pseudoknot. The motif, which contains a short parallel duplex, provides a structural mechanism for how the guanosine is extruded from the core stack to enable eIF4E recognition. Homologous PTE elements harbor a G-rich bulge and a three-way junction and exhibit covariation at crucial positions, suggesting that the PEMV2 tertiary architecture is conserved among these homologs.

GRAPHICAL ABSTRACT



INTRODUCTION

In eukaryotes, processed messenger RNA (mRNA) is covalently modified with a 5′ *N*7-methylguanine triphosphate cap and 3′ poly(A) tail, which serve to identify the mRNA by the eukaryotic translation initiation factors (1). Protein translation begins when eukaryotic initiation factor 4E (eIF4E) recognizes the 5′ cap structure. This eIF4E recognition is a key first step in a series of interactions with other initiation factors that trigger ribosome assembly at the AUG start site (2). The 5′ m^7GpppN cap serves as a crucial beacon that distinguishes mRNA from other cellular RNAs. The cap structure is added co-transcriptionally in the nucleus once the nascent mRNA is ∼25–30 nucleotides (nt) long (3,4). This presents a problem for parasitic viruses, whose mRNA is generated in the cytoplasm. These viral mRNAs necessarily lack the 5′ cap structure needed for cap-dependent translation. Instead of encoding capping enzymes in their small genomes, viruses have evolved

*To whom correspondence should be addressed. Tel: +1 773 702 9312; Email: jpicciri@uchicago.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

cap-independent translation mechanisms (5). In some cases, this cap-independent translation mechanism involves covalently modifying viral RNAs with the Vpg protein (short for virus protein, genome linked) (6). In other cases, structured RNA elements in the 5′ or 3′ untranslated regions (UTRs) bind and hijack the host cell translation initiation machinery (5,7).

Internal ribosome entry sites (IRESs) are perhaps the most common example of cap-independent translation enhancers (CITEs) found in viral RNAs. IRESs are large (500–1000 nt) multidomain structured RNAs located within the 5′ UTR of viral genomes (8). IRESs function as *cis*-acting elements that recruit the initiation factors and/or the ribosomal subunits directly to the start of the open reading frame (8).

IRESs are notably less common in plant viruses than in animal virus RNAs (8–10); instead, these viruses more often deploy an alternative cap-independent translation mechanism (8). Recently, structured RNA translation elements were discovered in the 3′ UTRs of plant viruses (11). These 3′ CITEs are still being identified in plant viruses through deep sequencing and phylogenetic analysis. Although CITEs have been identified and characterized in viruses belonging to the *Tombusviridae* family and the *Luteovirus* genus, they are expected to be utilized more broadly in biology (12–15).

Like IRES elements, CITEs are also structured, *cis*-acting, autonomous RNA elements that induce translation of viral RNA genomes by recruiting various host translation initiation factors or their components. However, CITEs are generally smaller than IRESs, usually only spanning 100–200 nt in length and generally only encompass a single domain (12). Their complex tertiary structures are capable of binding host factors that facilitate translation of viral genes (12). CITEs often function independently, with one CITE per 3′ UTR. However, two instances of virus harboring multiple CITEs have recently been discovered (12). In the first case, Pea enation mosaic virus 2 (PEMV2) RNA, studied here, was found to use three CITEs (16): one to bind eIF4E, another to recruit the 40S ribosomal subunit and a third to circularize the RNA, each functioning independently of one another (16). In the second case of a virus harboring multiple CITEs, horizontal gene transfer between two co-infecting viruses led to the acquisition of a second functional CITE (17). This finding raises the possibility that CITEs are readily exchanged between viruses, enabling rapid viral evolution to enable immune evasion and expansion of the host range (17,18).

There are seven distinct classes of CITEs; each adopts a distinct, well-conserved structure and employs unique mechanisms of host factor recruitment (12). These classes include the following: translation enhancer domain (19); Barley yellow dwarf virus-like translation enhancer (11,20–24); Y-shaped structure CITE (25); T-shaped structure (TSS) CITE (26,27); I-shaped structure (10,17); and the recently discovered CITE, CABYV-Xinjiang-like translation element (17).

The class of CITE studied here is the Panicum mosaic virus translation element (PTE), first identified in Panicum mosaic virus (PMV; *Panicovirus*, *Tombusviridae*) and then subsequently in PEMV2 (28,29). PTEs have been discovered in panicoviruses, auresviruses, carmoviruses and umbraviruses (12). They adopt a structure that binds plant, and in one case mammalian, eIF4E (10). PTEs cannot bind any other eIF4 proteins like eIF4G or eIF4(Iso)E, but PTE affinity for eIF4E is enhanced in the presence of eIF4G (Figure 1B) (29). These PTE elements are among the most well-characterized CITEs biochemically, making them ideal targets for structural studies. A PTE's ability to induce cap-independent translation or compete with the native cap structure relates directly to its affinity for eIF4E as determined by Wang *et al.* (29).

Homology analysis, secondary structure computational predictions and chemical structure probing with SHAPE have shown that PTEs adopt a T-shaped secondary structure (28). A basal stem is made up of three helical segments, helix 1 (H1), helix 2 (H2) and helix 3 (H3), separated by small internal bulges (Figure 1A) (28). Between H2 and H3, there is an asymmetrical G-rich bulge (J2/3) on the 5′ side, known as the G domain (28). At the top of H3, a three-way junction branches the structure into stem–loop 1 (SL1) and stem–loop 2 (SL2). SL1's 3′ end is connected to SL2's 5′ end by a C-rich bulge (JSL1/SL2), known as the C domain (28). The lengths of SL1 and SL2 vary across PTEs (28). In almost all PTEs with the exception of PEMV2, loop 1 (L1) shares sequence complementarity with a hairpin (5H2) in the 5′ UTR (28). The interaction between L1 and 5H2 circularizes the RNA and helps recruit the preinitiation complex in a manner analogous to eIF4E and poly(A)-binding protein (11).

Generally, the G domain bulge contains a G-rich region that exhibits apparent complementarity to the C-rich region in the C domain; mutations to either of these tracts eliminate or severely reduce PTE-induced translation and eIF4E binding (Figure 1A) (28,29). The C domain is largely unreactive to SHAPE structure probing (28), leading to the hypothesis that the G and C domains form a pseudoknot in the PTE (28). Curiously, the C and G domains from different PTEs cannot be mutationally transferred into other PTEs without loss of function, suggesting that their eIF4E binding activity is dependent on the structural context of their respective PTEs (28). The G domain of every PTE examined by SHAPE probing contains a single G that is 'hyper-reactive' in the presence of magnesium ions, suggesting that it occupies a solvent-exposed, flexible position in the structure (28). In the footprinting experiments where PTEs were treated with RNase T1 or 1-methyl-7-nitroisatoic anhydride (1M7), this so-called hyper-reactive G showed reduced modification in the presence of increasing concentrations of eIF4E suggesting that this residue resides within the eIF4E binding interface. Mutation of either of the two tryptophans (W62 and W108) in the cap-binding pocket drastically reduces eIF4E's affinity ($K_d = 58$ nM) for the PTE (28,29), consistent with the hypothesis that eIF4E binds the PTE using its cap-binding pocket. Based on these observations, tertiary models of the PEMV2 and PMV PTEs were made, in which the SHAPE hyper-reactive G of the G domain flips out from the presumed pseudoknot and docks into m^7G cap-binding pocket of eIF4E (28,29).

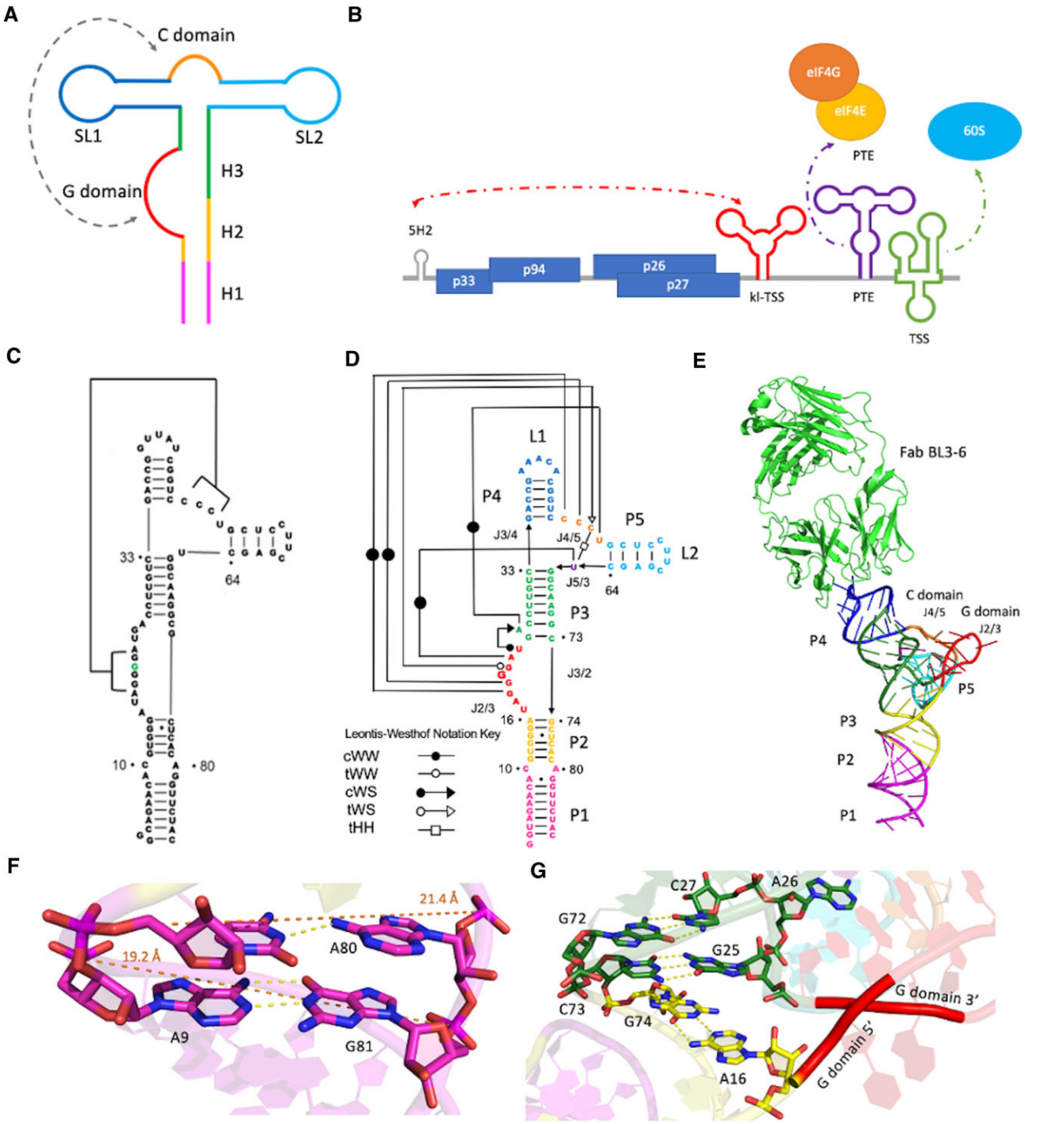


Figure 1. Genome organization of PEMV2 RNA and PTE secondary and tertiary structure scheme. (**A**) General secondary structure of PTE class of CITEs; regions are colored to match coloring of regions used throughout. (**B**) Genomic organization of PEMV2 illustrating the relative positions of the 3′ UTR CITEs, their binding partners and the 5′ proximal hairpin 5H2 kissing-loop interaction with kissing-loop TSS (kl-TSS). p33 and −1 ribosomal frameshifting product p94 are expressed from the gRNA. p26 and p27 are expressed from the overlapping open reading frames on the sgRNA, adapted from (18). (**C**) PEMV2 PTE secondary structure determined by Wang *et al.* (29). (**D**) PEMV2 PTE secondary structure determined by the crystal structure; interactions are denoted using the Leontis–Westhof notation (31) and pair and junction nomenclature (29). (**E**) Overall arrangement of PEMV2 PTE and Fab BL3–6 in the crystal structure. (**F**) The final two base pairs of P1. (**G**) P2 and P3 helical stack region illustrating the relative rotation of P3 (green) to P2 (yellow).

In this work, we focus on PEMV2 RNA, which contains three CITEs in its 3′ UTR: one PTE and two TSS elements (Figure 1B) (16,30,31). The kl-TSS element contributes to genome circularization during translation by binding a small hairpin structure in the 5′ UTR, thus alleviating the PEMV2 PTE's role in genome circularization (16,30). Interestingly, PEMV2 PTE is the only PTE found to not use its L1 to hybridize to a 5′ UTR hairpin to circularize the RNA (11). The other TSS element in the PEMV2 3′ UTR binds the 60S ribosome subunit. Even in the context of the full-length 3′ UTR containing all three CITEs, mutations to critical regions of the PTE element reduce eIF4E affinity and translation efficiency (28,29). Mutations that alter the sequence of either loop 1 or 2 do not alter eIF4E affinity or translation inhibition; however, mutating residues in the G and C domains reduces translation efficiency by 95% (29). Mutation of the 'hyper-reactive' G in particular decreases eIF4E binding and substantially reduces translation efficiency (29).

Underscoring the importance of understanding viral replication mechanisms, viral epidemics and pandemics have always threatened food security, and a growing global population only magnifies their devastating effects (32). Additionally, the proliferation of genetically modified crops drives monoculture farming, a practice that leaves crops particularly susceptible to disease due to an absence of natural genetic variation (32). Climate change also increases plant's susceptibility to viral infection by upregulating abiotic stress response gene expression pathways (33). Despite the increasing threat that plant viruses pose, study of their molecular biology lags behind that of human pathogens, which leaves society less prepared to respond to agricultural disasters (32). As crucial components to many plant viral replication mechanisms, CITEs are an excellent target to study using structural biology with the long-term goal of developing strategies to enable plants to evade viruses and improve agricultural resilience (12,34,35). Understanding the structural mechanism of translation factor recruitment of CITEs could guide the development of genetically modified crops that use mutant transcription factors resistant to CITE hijacking (12,34,35).

Using the extensive biochemical characterization as a guide, we generated a crystallization construct of the PEMV2 PTE to determine its structure experimentally. To crystallize RNAs, which often can be recalcitrant to crystallization, we applied chaperone-assisted RNA crystallography (36). By mutating a solvent-exposed loop to a Fab-binding epitope, we can create a complex of Fab and RNA element, which may crystallize more readily than RNA alone (36). The Fab not only facilitates RNA crystallization but also serves as a model for determining X-ray phases through molecular replacement. While considerable evidence exists to support the PTE secondary structure and the long-range, putative pseudoknot interaction between the C and G domains, it was unknown how the PTE forms a structure that competes with the native 5′ cap structure for binding eIF4E with high affinity. Our crystal structure agrees with the biochemical model that a single G of the G domain projects into solvent poised to engage eIF4E. However, we find that the PTE adopts an unexpected fold to accomplish this.

MATERIALS AND METHODS

Construct design

To create a version of the PEMV2 PTE element likely to fold into the native conformation, we began with a construct that represented nucleotides 3820–3907 in the PEMV2 viral genome (29). This region was identified as the minimal sequence necessary to induce cap-independent translation (28). To create a continuous helix at the base of the construct, residue C3822 was mutated to a U to form a base pair with A3906. This construct was also designed to contain an extra G at the 5′ end to improve transcriptional efficiency in an *in vitro* transcription reaction. To construct a version of the PEMV2 PTE element that can bind our crystallization chaperone, Fab BL3–6 (37), and its mutant K170A, one of the solvent-exposed RNA loops needed to be mutated to the Fab-binding sequence without disruption of the native fold or function. Previous mutational assays have determined that both loops 1 and 2 can tolerate mutations without affecting eIF4E binding activity. Therefore, we anticipated that loop 1 or loop 2 can tolerate a mutation to the Fab-binding epitope, an AAACA pentaloop closed by a GC base pair (29). However, only the loop 1 mutant was able to bind Fab efficiently. Therefore, crystallization trials were only attempted using the mutated L1 construct; this construct is henceforth referred to as PEMV2 BL3–6.

RNA transcription and purification

Single-stranded DNA templates and primers for polymerase chain reaction (PCR) and transcription were ordered from IDT encoding the transcription template for each RNA construct with a T7 promoter (Supplementary Table S1). Forward primers were ordered matching the T7 promoter region and the first two nucleotides of the reverse primers contained a 2′*O*-methyl modification to avoid 3′ end untemplated additions by T7 polymerase. Transcription template DNA was amplified into double-stranded DNA using PCR. RNA was transcribed from the purified PCR product using an *in vitro* transcription reaction as follows: 50 pmol/ml DNA template was incubated for 3 h at 37°C in buffer containing 40 mM Tris–HCl (pH 8.0), 2 mM spermidine, 10 mM NaCl, 25 mM $MgCl_2$, 10 mM DTT, 40 U/ml RNase inhibitor, 5 U/ml thermostable inorganic pyrophosphatase, 5 mM of each NTP and 50 μg/ml T7 RNA polymerase. Reactions were halted by addition of RNase-free DNase I at 5 U/ml and incubation at 37°C for 30 min. RNA was purified on a 10% denaturing polyacrylamide gel in 0.5× TBE running buffer. The RNA was visualized with UV shadowing, extracted and eluted into 10 mM Tris (pH 8.0), 2 mM EDTA and 300 mM NaCl buffer via overnight incubation at 4°C. The eluted RNA was then concentrated and exchanged into double-distilled H_2O using a 10K Amicon filter and stored at −80°C until further use.

Fab purification

The Fab BL3–6 expression vector or its mutant version Fab BL3–6 K170A (available upon request) was transformed into 55 244 chemically competent cells (www.atcc.org) and grown on LB plates supplemented with carbenicillin at

100 μg/ml. Nine colonies from the plates were chosen and inoculated to a starter culture with 100 μg/ml carbenicillin, which was grown at 30°C for 8 h. Once the starter culture reached an OD 600, 15 ml of starter culture was used to inoculate 1 l of 2× YT media and grown for 24 h at 30°C. The cells were then pelleted via centrifugation at room temperature, and the cell pellet was resuspended in 1 l of freshly prepared CRAP-Pi media supplemented with 100 μg/ml carbenicillin. The cells were set to grow for 24 h at 30°C, harvested via centrifugation at 4°C and frozen at −20°C. Frozen cell pellets were lysed in phosphate-buffered saline (PBS) supplemented with 0.4 mg/ml lysozyme and 0.01 mg/ml DNase I. After 30 min, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 0.5 mM. After 30 min, the mixture containing cellular debris and lysate was centrifuged for 45 min, 12 000 rpm, with a rotor type JLA 16.250 (Beckman) at 4°C. Lysate was transferred to new sterile bottles and centrifuged again for 15 min, 12 000 rpm, at 4°C. Supernatant was filtered through 0.45-μm filters into a sterile bottle (Millipore Sigma, www.sigmaaldrich.com), and Fab proteins were purified using the AKTAxpress fast protein liquid chromatography purification system (Amersham, www.gelifesciences.com) as described previously (37,38). The lysate in PBS (pH 7.4) was loaded into a protein A column, and the eluted Fab in 1 M acetic acid was buffer exchanged back into the PBS (pH 7.4) using 30-kDa cutoff Amicon filter and loaded into a protein G column. The Fab was eluted from protein G column in 0.1 M glycine (pH 2.7) and then buffer exchanged into 50 mM NaOAc and 50 mM NaCl buffer (pH 5.5) and loaded into a heparin column. Finally, the eluted Fab in 50 mM NaOAc and 2 M NaCl (pH 5.5) was dialyzed back into 1× PBS (pH 7.4), concentrated and analyzed by 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis using Coomassie Blue R250 staining for visualization. Aliquots of Fab samples were tested for RNase activity using the RNaseAlert kit (Ambion, www.thermofisher.com). The aliquots of Fab samples were flash frozen in liquid nitrogen and stored at −80°C until further use.

Wheat eIF4E expression

The gene for wheat eIF4E (GenBank Z12616.2) was cloned into a pET21a vector to include an N-terminal 6× histidine tag. The resulting plasmid sequence was verified and transformed into BL21 Rosetta DE3 *Escherichia coli.* Cells were grown in LB supplemented with 100 μg/ml carbenicillin at 37°C first in a 50 ml small-scale culture overnight, which was then used to induce a 1 l culture. The 1 l culture was grown at 37°C, until log phase was achieved. Cells were induced for protein expression using 100 mM IPTG and grown for 3 h before harvesting through centrifugation at 6000 rpm for 10 min at 4°C in 1 l batches.

Wheat eIF4E purification

Cell pellets were lysed using lysis buffer (25 mM HEPES, pH 7.6, 100 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 0.4 mg/ml lysozyme, 0.1 mg/ml DNase I, 1 mM PMSF and 0.5% Triton X-100) and a single freeze–thaw cycle, followed by an incubation at room temperature for 1 h with 0.4 mg/ml lysozyme and 0.1 mg/ml DNase. Cell debris was removed from the lysis by a 45-min centrifugation at 12 000 rpm. Clarified lysate was filtered and applied to a Qiagen His trap column, washed with 5 CV running buffer (50 mM HEPES, pH 7.6, 2 mM $MgCl_2$, 10% glycerol, 10 mM imidazole, 100 mM KCl, pH 7.5) and eluted with running buffer supplemented with 250 mM imidazole. Fractions containing high UV absorbance were pooled and buffer exchanged into running buffer and applied to the Ni^+ column once more. The column was then washed with running buffer supplemented with additional 20 mM imidazole and 500 mM NaCl for 30 CV followed by a second wash with 30 mM imidazole and 1 M NaCl for 40 CV. Finally, protein was eluted using running buffer supplemented with 250 mM imidazole. Fractions with a high UV absorbance were collected and pooled; the purified protein was then buffer exchanged into storage buffer containing 50 mM HEPES (pH 7.6), 2 mM $MgCl_2$, 10% glycerol and 100 mM KCl (pH 7.5), and flash frozen with liquid nitrogen at 0.12 mg/ml aliquots for further use in electrophoretic mobility shift assays (EMSAs).

Electrophoretic mobility shift assay

To facilitate refolding, purified RNA constructs in double-distilled H_2O were heated to 90°C for 1 min, then cooled on ice for 2 min and held at room temperature for 3 min. Refolding buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 mM KCl) was added, and the RNA was then incubated at 50°C for 10 min, followed by a 5-min incubation on ice.

Refolded RNA was then mixed with either eIF4E storage buffer as a negative control or a 1.5 molar excess of RNase-free eIF4E and incubated on ice for 15 min to establish equilibrium state binding. eIF4E RNA complexes were separated by gel electrophoresis in a 10% polyacrylamide gel made in 0.5× TBE buffer supplemented with 11 mM $MgCl_2$. The gel was loaded while running to avoid complex dissociation and was run at 4°C for 1 h. The gel was stained with ethidium bromide and visualized via UV light and photographed, then stained in Coomassie, destained and photographed a second time.

Crystallization

The crystallization construct was analyzed by EMSA under nondenaturing conditions. Refolded wild-type and BL3–6 minimal RNA constructs migrate at the same positions in the native gel supplemented with an effective concentration of 10 mM $MgCl_2$, consistent with retention of the wild-type fold. In the presence of a stoichiometric amount of eIF4E, both constructs shifted quantitatively to the bound state. The same shift did not occur for a non-PTE RNA construct, indicating that this complex is not the result of nonspecific RNA binding activity by eIF4E. We conclude that the BL3–6 construct folds and functions in a way that closely resembles the wild-type PEMV2 PTE such that it retains the native eIF4E binding activity.

Validated RNA construct in ultrapure H_2O was heated to 90°C for 1 min, then cooled on ice for 2 min and held at room temperature for 3 min. Refolding buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 mM KCl) was added, and the RNA was then incubated at 50°C for 10 min, followed by

a 5-min incubation on ice to facilitate refolding. Four hundred eighty micrograms of refolded RNA was mixed with a 1.1 molar equivalent of Fab BL3–6 K170A (referred to here as Fab BL3–6 to form a complex of RNA and Fab) in buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 mM KCl). The complex was then concentrated to 6 mg/ml final concentration of RNA (80 μl). The 100 nl + 100 nl hanging drop crystal trials were set in commercially available crystallization kits from Hampton Research and Jena Bioscience and allowed to grow for 2−3 weeks at room temperature. Crystals grew in 5 mM $MgCl_2$, 50 mM HEPES (pH 7.0) and 25% PEG 550, and were further optimized with an additive screen resulting in a final condition of 5 mM $MgCl_2$, 50 mM HEPES (pH 7.0), 25% PEG 550 and 2% benzamidine hydrochloride. Crystals were then looped and frozen in liquid nitrogen.

Diffraction data collection

Diffraction data were collected at APS beamline 24-ID-E. All of the datasets were then integrated and scaled using its on-site RAPD automated programs (https://rapd.nec.aps.anl.gov). Crystals of PEMV2 and Fab were small, so data were collected in two 180° halves with a pause between the two to take a snap diffraction to confirm that the crystal was still in the beam. The last 300 images from the diffraction set were removed due to radiation damage and the remaining images were merged together into a final dataset that was then processed in the Phenix software suite (39).

Crystallographic data processing

A dataset was collected out to 2.1 Å, but the data were highly anisotropic. After anisotropy correction with Xia2 dials server (40), the dataset was truncated to 2.75 Å. The structure was solved using molecular replacement of the Fab model in Phenix. Only one copy of the RNA and Fab was discovered in the $P\ 1\ 2_1\ 1$ space group. Using the initial phases from the molecular replacement solution, the RNA was able to be built into the emerging density after multiple rounds of refinement using Coot and phenix.refine (41–43). Simulated annealing and composite omit maps were applied to the model during refinement and building to combat accumulating model bias as identified by a growing gap between the R_{work} and R_{free} values.

RESULTS

Structural analysis

The global architecture of the PEMV2 BL3–6 crystal structure. The final structure was solved to 2.75 Å with an R_{work} of 25% and an R_{free} of 29%. Additional statistics are reported in Table 1. The coordinates for this structure have been deposited in the Protein Data Bank (PDB) under accession code 8SH5. The secondary structure determined by the crystal structure closely resembles the secondary structure previously predicted (Figure 1C and D), but differs in important ways (Figure 1D and E) (28,29). Beginning at the 5′ end, P1 forms through pairing between residues 2–9 and residues 81–88 in our numbering, with the first G of the construct numbering position 1 (Figure 1D). P1 ends in an uncommon A9–G81 Watson–Crick–Franklin (WCF) pair (Figure 1D and E). This purine–purine pair locally widens the helical diameter and increases the separation between adjacent C10 and A80 nucleotides. Nevertheless, their WCF faces point toward one another and the nucleobases remained in the helical stack (Figure 1F).

Stacked above P1, P2 forms between residues 11–16 and 74–79. J2/3, known as the G domain, was predicted to include 11 residues on the 5′ side (residues 16–26) and two residues on the 3′ side (J3/2; residues 72–73) (28,29). However, in the structure, some of the residues in these regions engage in pairing interactions with each other (Figure 1D). At the top of P2, A16 and G74 use their WCF faces to form a purine–purine pair that extends and widens the helix. At the bottom of P3, G25 forms a WCF pair with C73, which extends the helix and consequently subsumes A26 into P3 as a bulged nucleotide (Figure 1D and G). These interactions leave no unpaired nucleotides in J3/2 and effectively shorten the region considered to be in the G domain by three nucleotides. P3 stacks upon P2 with a ∼90° twist angle between the final pair of P2 and the initial pair of P3 (A16–G74 in P2 and C25–G73 in P3; Figure 1G). This P2/P3 stacking forces the 3′ and 5′ ends of the G domain (J2/3) close together with the backbones of each end crossing over one another (Figure 1G).

Aside from the secondary structure adjustments around the asymmetric bulge, the remaining secondary structure (C27 to G72) matches the biochemically derived secondary structure (Figure 1C) (28,29). The top of P3 leads into the P4L1 stem–loop, which stacks upon P3, followed by the P5L2 stem–loop. Together, P3, P4L1 and P5L2 form the

Table 1. Crystallographic data collection and refinement statistics

PEMV2 PTE and Fab BL3–6	
Wavelength	0.9792
Resolution range	59.42–2.75 (2.848–2.75)
Space group	$P\ 1\ 2_1\ 1$
Cell dimensions	
a, *b*, *c* (Å)	63.346, 65.46, 89.817
α, *β*, *γ* (°)	90, 110.29, 90
Total reflections	69 129 (6664)
Unique reflections	18 060 (1760)
Multiplicity	3.8 (3.7)
Completeness (%)	97.27 (97.01)
Mean $I/\sigma(I)$	14.45 (2.35)
Wilson *B*-factor	61.71
R-merge	0.1914 (0.5717)
R-meas	0.2188 (0.6634)
CC1/2	0.979 (0.793)
CC*	0.995 (0.941)
Reflections used in refinement	17 640 (1752)
Reflections used for *R*-free	1600 (162)
R-work	0.2517 (0.4800)
R-free	0.2970 (0.5793)
CC(work)	0.942 (0.683)
CC(free)	0.762 (0.516)
RMS(bonds)	0.002
RMS(angles)	0.59
Ramachandran favored (%)	96.33
Ramachandran allowed (%)	3.67
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.27
Clashscore	5.97
Average *B*-factor	97.47

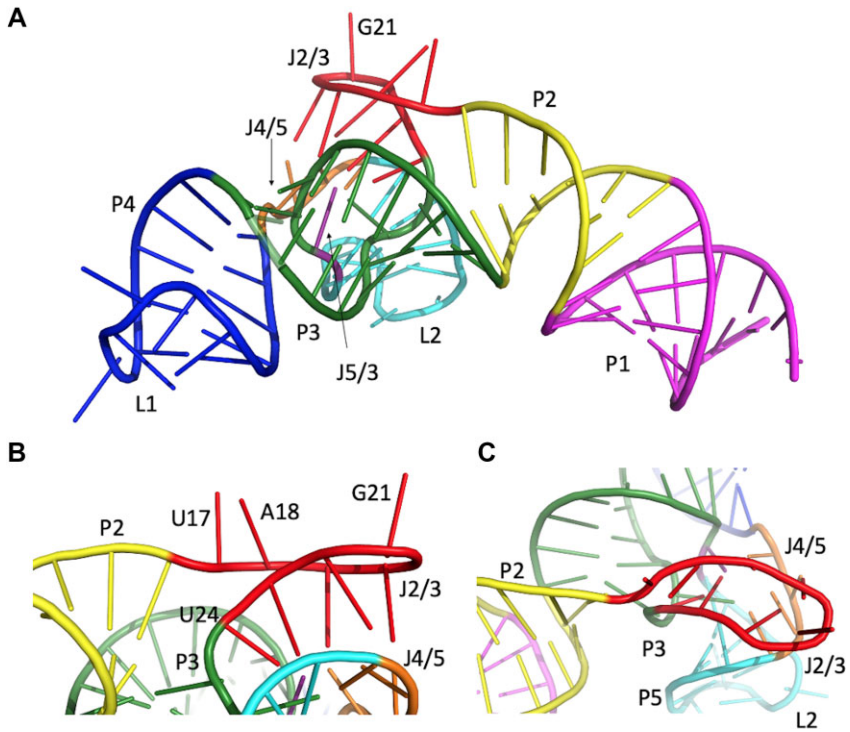


Figure 2. (**A**) Overall arrangement of PEMV2 PTE in the crystal structure. (**B**) G domain position in the PEMV2 PTE structure. (**C**) The image shown in panel (B) but rotated by 90°.

T-shaped region at the top of the secondary structure (Figure 1D). P4L1 encompasses paired nucleotides 34–38 and 44–48 and loop nucleotides 39–43 (L1). P5L2 encompasses paired nucleotides 53–56 and 61–64 and loop nucleotides 57–60 (L2). An internal bulge, J4/5, joins the 3′ end of P4L1 and the 5′ end of P5L2, known as the C domain, which encompasses residues 49–52. The J5/3 region contains a single residue, U65, as predicted (28,29).

The P3/P4/P5 three-way junction. The lengths of the joining regions connecting P3, P4 and P5 and the relative orientation of the helices define this region as an A-type three-way junction (44). In an A-type three-way junction comprising paired regions A, B and C (PA, PB and PC) with the first two helices coaxially stacked, the JA/B region typically contains 0–4 residues, JB/C typically 3–9 residues and JC/A typically 1–3 residues but less than JB/C. PEMV2 conforms to these ranges with J3/4 = 0, J4/5 = 4 and J5/3 = 1 (Supplementary Figure S1A and B). An additional defining feature of an A-type junction pertains to the orientation of the third helix (P5 in PEMV2), which usually runs perpendicular to the central axis of the two coaxially stacked helices (44) as is the case for PEMV2 (Supplementary Figure S1C and D). Compared to other junction types, interactions between the junction and the joining regions are the least extensive in A type. This feature extends to PEMV2's junction, where we observe only two interactions, both of them involving the P3 phosphodiester at G67 (Supplementary Figure S2) with OP1 forming a hydrogen bond with C49's (J4/5) N4 exocyclic group and OP2 forming a ribose–phosphate zipper (45) interaction with the U65's (J5/3) 2′OH (Supplementary Figure S2). We also observe additional ribose–phosphate zipper interactions in close proximity to the junction (Supplementary Figure S3) involving the phosphates of A26 and C27 at the base of P3 and the 2′OHs of C54 and C64, respectively, in the minor groove of P5 (Supplementary Figure S3).

G and C domains form a structured motif containing base pairs formed by parallel strands. The crystal structure of the PEMV2 PTE has an overall bent shape, with the G domain emerging from the middle of the stack at the P2–P3 junction (Figures 1D and E and 2A). The G domain has a compressed, elongated loop conformation, with one side facing the PEMV2 PTE structural elements and the other side being exposed to solvent (Figures 1E and 2B). This positioning is facilitated by long-range interactions with the C domain (J4/5) and J5/3, enabled by the arrangement of the three-way junction. The C domain is sandwiched between the G domain on one side and the P5 stem–loop on the other side (Figure 2). The long-range interactions between the G domain and the three-way junction form an unusual architecture that involves parallel strand WCF base pairing within a unique stacked scaffold. This scaffold consists of four columns of nucleobases: A (G19, A23 and U24); B (G20 and G22); C (C49, U65 and A26); and D (C50, C51 and U52). Together, these nucleotides form four layers of stacked nucleobase planes (Table 2 and Figure 3 and 4) buttressed by the P5 helix.

At the 5′ end of the G domain, U17 and A18 flip out from the P2–P3 helical stack into solution (Figure 3C and E). The electron density for the phosphates of these nucleotides is clearly defined as illustrated in Figure 3C, but the nucleobases lack clear electron density, suggesting free

Table 2. G domain and the three-way junction scaffold organized into four nucleobase stacking columns (A, B, C and D) and four stacking layers (1–4)

	Nucleobase stacking column			
Stacking layer	A	B	C	D
1		G20		C50
2	G19		C49	
3	A23	G22	U65	C51
4	U24		A26	U52

rotation about the glycosidic bond in these residues. The phosphate backbones at the 5′ and 3′ ends of the G domain are in close proximity. The OP1 atoms of U17 and A18 are within 3.0 and 4.2 Å of U24:OP1, respectively. Additionally, a ribose–phosphate zipper interaction occurs between A18:2′OH and A23:OP1 (Figure 3F). An analogous interaction could occur between A18:OP2 and A23:2′OH as these atoms reside just outside of hydrogen bond distance (3.8 Å; Figure 3F). Together, these interactions presumably restrain the backbone at A18 and U17 (Figure 3C, E and F), consistent with the well-defined electron density.

Four layers of stacked nucleobase planes. Moving beyond U17 and A18 through the G domain in the 5′ to 3′ direction, G19 and G20 form the first two base pairs with the C49 and C50 of the C domain, respectively; these two pairs constitute the first two stacking layers of the core scaffold (Table 2, Figures 3C and 4A, B and D, and Supplementary Figure S4). Instead of forming a standard antiparallel A-form helix, in which G19 would pair with C50 and G20 would pair with C49, we observe unusual parallel base pairing for G19–C49 and G20–C50, with their glycosidic bonds arranged in *cis* to one another (Figure 4A and B). In each pair, one sugar is C2′-*endo* (G20 and C49) and the other is C3′-*endo* (G19 and C50). Each of these four nucleobases rotates away from its sugar in the *anti*-conformation. To achieve the parallel *cis* WCF pairing, the backbone is flattened and extended compared to a standard antiparallel helix, creating a wider minor groove into which downstream residues can pack and stack (Figure 4A and Supplementary Figure S4). The G20–C50 (layer 1) and G19–C49 (layer 2) pairs form the uppermost two layers of the core scaffold (Figure 4A, B and D). In addition to stacking, a possible hydrogen bond between C50:N4 and C49:O2 located 2.7 Å (Figure 4B and Supplementary Figure S4) away may contribute to this unique nucleobase arrangement. Due to this uncommon parallel pairing, we have chosen not to refer to the G and C domain structure as a pseudoknot (46).

Following G20, the G domain strand acutely inverts on itself, folding back over to position the plane of the G22 nucleobase 3.4 Å from O4′ of G20 (Figure 4D and E). Underscoring the acuteness of the turn, the ribose C1′ atoms of G20, G21 and G22 subtend an angle of 47° (Supplementary Figure S5). This abrupt turn flips the G21 nucleobase out of the stacked core toward solvent (Figure 4D and E, and Supplementary Figure S5). Despite the high SHAPE reactivity of G21, its electron density is clearly defined (Figure 3A–C and Supplementary Figure S5), likely reflecting a crystal contact between G21 and a pocket in the heavy chain scaffold from the symmetry mate Fab molecule (Supplementary Figure S6).

The backbone reversal due to the acute turn at G21 places G22 and A23 below G20 and G19, respectively, where they interact with C51 and U65, respectively, from the three-way junction, forming the third stacking layer of the scaffold (Figure 4D and F). G22 forms a *trans* WCF/sugar edge interaction with C51, the third residue of the C domain, and A23 forms a *cis* WCF base pair with U65, the single J5/3 nucleotide, which inserts into the stacking scaffold with C49 above and A26 below. The two pairs in this layer also interact with each other through a hydrogen bond between the exocyclic amine of C51 and the O4 keto group of U65 (Figure 4F). In contrast to the nearly planar arrangement of the first two C domain residues (C49 and C50), which face the opposing G domain strand to form the parallel base pairs, C51 turns inward toward the RNA's structural core, positioning its nucleobase plane below C50 (Figures 3B and 4D). The nucleobase planes of both G22 and C51 tilt with respect to the stacking axis, allowing this non-canonical GC pair to make additional interactions with the minor groove edge of the G19–C49 pair in the preceding layer. G22 engages in a *cis* sugar edge/sugar edge interaction with G19, and C51 directs its imino nitrogen and exocyclic amine within hydrogen bonding distance of the amino group of G19 and the keto group of C49, respectively (Figure 4C and Supplementary Figure S7).

In the fourth layer of the stacking scaffold, the final nucleotide of the G domain, U24, participates in a base triple interaction with the *cis* WCF pair formed by A26, the bulged nucleotide at the base of P3, and U52, the last nucleotide of the C domain (Figure 4D and H). Figure 4 illustrates this complex network of hydrogen bonds and cross-strand stacks involving the G and C domains, P3, P5 and J5/3. Additionally, two nucleotides from this fourth stacking layer make interactions with nucleotides from the third stacking layer: U24's O4′ ether oxygen and U52's O2 keto group reside within hydrogen bonding distance of A23's 2′OH and N6 exocyclic amine, respectively (Figure 4G and Supplementary Figure S8). Beneath the U24–A26–U52 base triple, P5 forms a platform for the scaffold stack starting with a simple GC pair between G53 and C64 (Figure 4D and Supplementary Figure S9). This base pair helps orient the base planes of all the layers that stack above it.

Mutational analysis of PTE's structurally inferred scaffold interaction network

PEMV2 PTE function correlates directly with eIF4E binding activity (28,29); in wheat germ extract, addition of refolded PTE in isolation inhibits translation of a luciferase reporter construct; the magnitude of translation inhibition correlates with the affinity the exogenous PTE because it competes for the available eIF4E (28,29). Therefore, we used an EMSA to detect eIF4E binding and investigated the effect of PTE single mutations expected to disrupt or double mutations that could potentially restore the structurally inferred scaffold interaction network. We found that constructs containing mutations expected to disrupt the core migrated slower in the gel relative to wild-type RNA, suggesting that the tertiary interactions of the folded motif

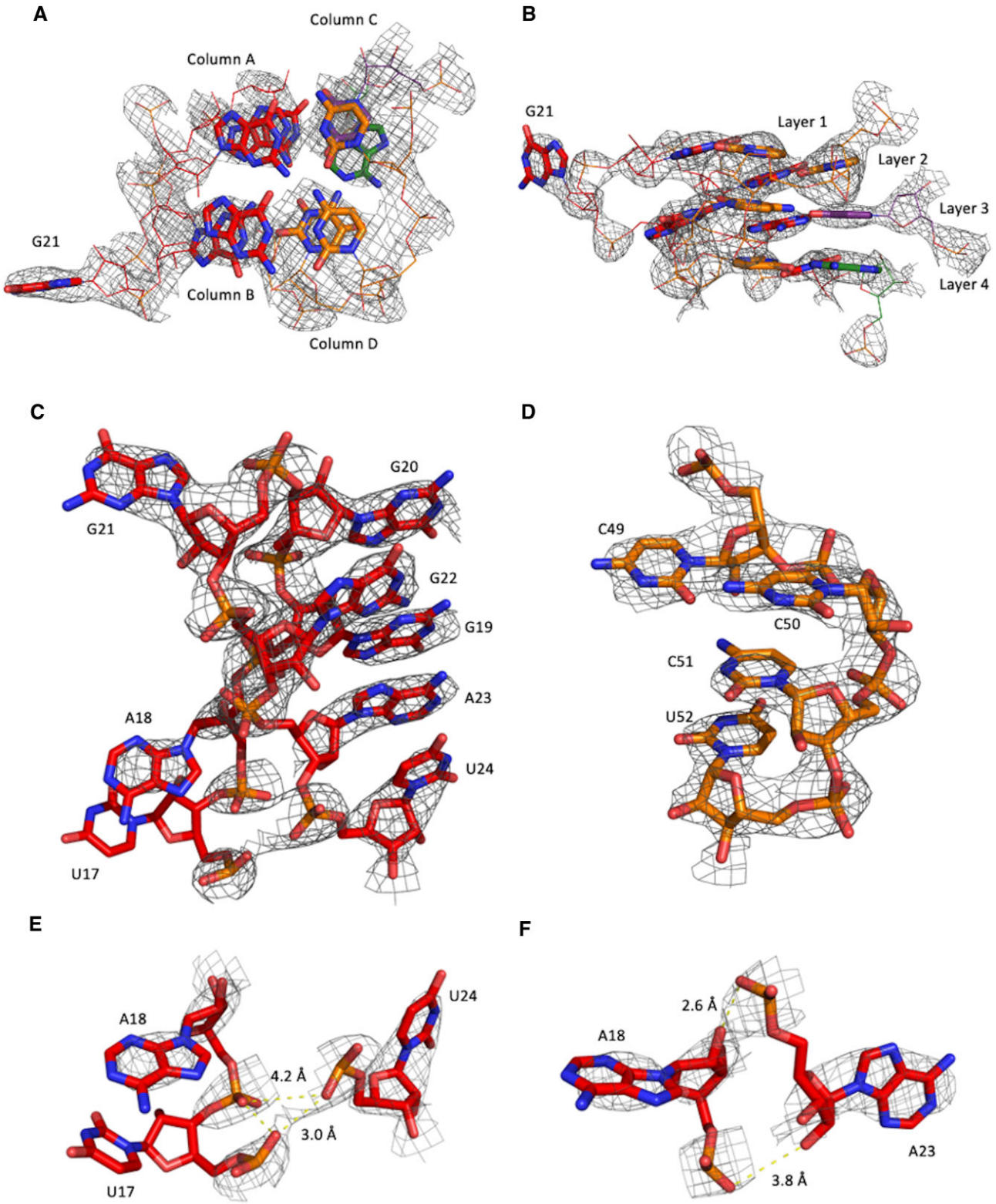


Figure 3. Core motif created by long-range interactions between the G domain and the C domain including the associated 3-4-5 three-way junction. 2Fo–Fc electron density map shown as gray mesh, contoured at 2.0σ. (**A**) Stacked scaffold consisting of four columns of nucleobases: A (G19, A23 and U24); B (G20 and G22); C (C49, U65 and A26); and D (C50, C51 and U52), forming four layers of stacked nucleobase planes. (**B**) The image shown in panel (A) but rotated to show layer 1 (G20, C50), layer 2 (G19, C49), layer 3 (A23, G22, U65, C51) and layer 4 (U24, A26, U52). (**C**) G domain overall crystal structure. (**D**) C domain overall crystal structure. (**E**) U17, A18 and U24 phosphate backbone distances. (**F**) A18:2′OH-A23:OP1 ribose–phosphate zipper.

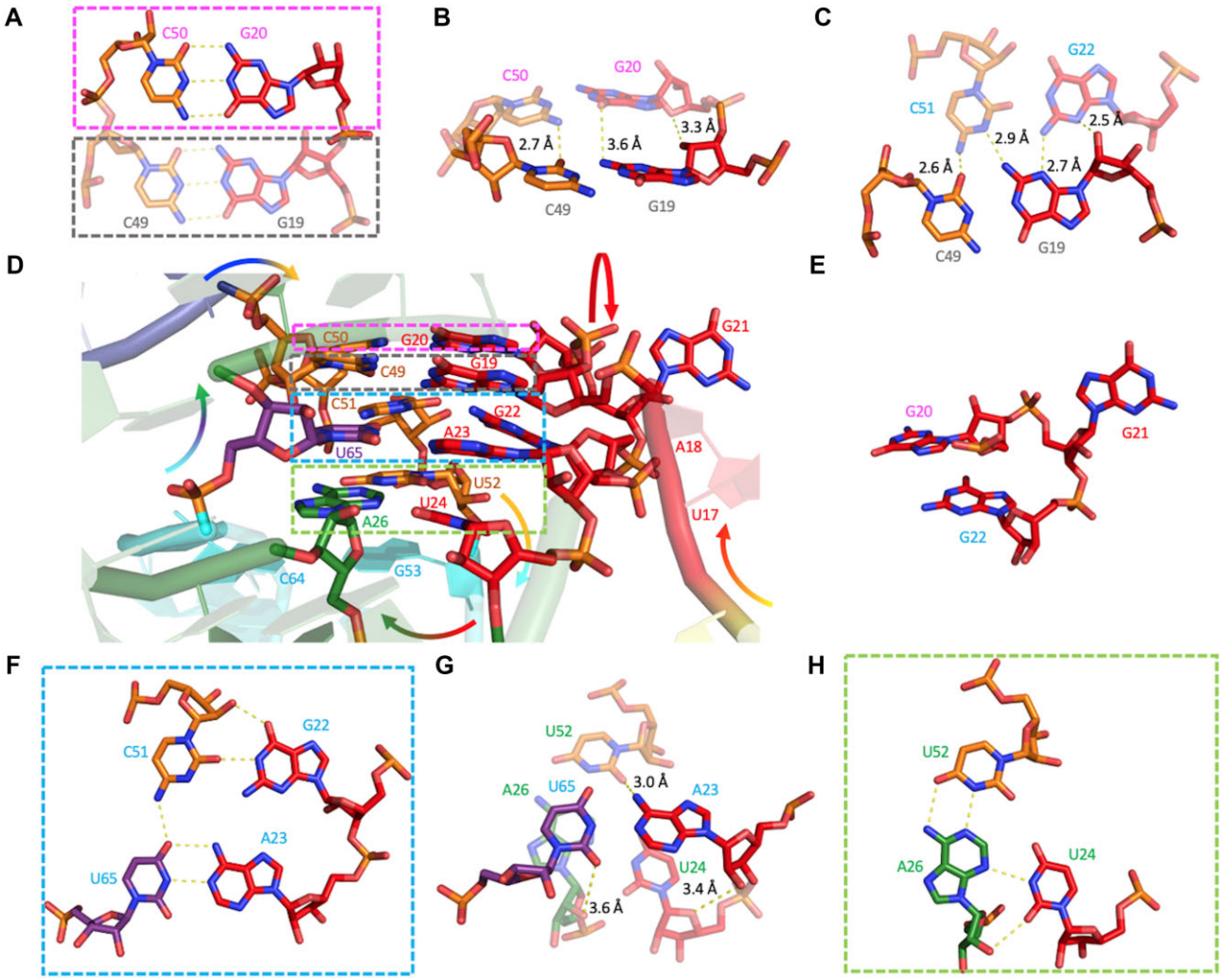


Figure 4. Base pairing interactions within the PTE core. Dashed colored rectangles represent each of four layers; residue denotation color is analogous to its layer, except for panel (D). (**A**) Layer 1 (G20 and C50, pink rectangle) and layer 2 (G19 and C49, gray rectangle). (**B**) The image shown in panel (A) but rotated by 90° to show cross-layer interactions between layers 1 and 2: G19:N2 (layer 2) interacts with G20:O6 (layer 1); G19:2′OH (layer 2) interacts with G20:O4′ (layer 1); and C50:N4 (layer 1) interacts with C49:O2 (layer 2). (**C**) Cross-layer interactions between layers 2 and 3: hydrogen bond interaction between C49:O2 (layer 2) and C51:N4 (layer 3); G19 (layer 2) interacts with C51:N3 (layer 3) by its N2, with G22:N2 by its N3 and with G22:N3 by its 2′OH. (**D**) Structured motif containing base pairs formed by parallel strands with each stacking layer represented by a colored rectangle. Residue color corresponds to its structured region. Colored arrows represent 5′ to 3′ RNA direction of the corresponding structural region. (**E**) G20, G21 and G22 undergo an abrupt phosphate backbone turn. (**F**) Layer 3 (G22, A23, C51, U65, blue rectangle). (**G**) Cross-layer interactions between layers 3 and 4: U24:O4′ and U52:O2 (both in layer 4) reside within hydrogen bonding distance of A23:2′OH and N6 (layer 3), respectively; U65:O2 (layer 3) interacts with A26:2′OH. (**H**) Layer 4 (U24, A26, U52, green rectangle).

render the RNA more compact compared to an unfolded core (47). The effects of these mutations on the native fold are detected as a shift in the position of the free RNA band in an EMSA relative to wild-type RNA. Binding to eIF4E manifests as a shift in electrophoretic mobility of the RNA in the presence of eIF4E relative to RNA mobility in the absence of eIF4E due to the change in size and charge to mass ratio. Previous studies have revealed that mutations to crucial residues in the C and G domains often eliminate eIF4E binding entirely instead of simply reducing affinity (28,29). Therefore, in this assay excess eIF4E was used to allow detection of weakened binding.

To determine whether the identity of first two residues of the G domain U17 and A18 contributes to eIF4E binding, these residues were mutated to C and U, respectively, in a double mutant construct. This U17C:A18U double mutant PEMV2 PTE RNA in the absence of eIF4E had the same electrophoretic mobility as wild type, suggesting that the native fold was not altered by these mutations (Figure 5 and Supplementary Figure S11, lane U17C:A18U '−'). This mutant also retains eIF4E binding capacity (Figure 5 and Supplementary Figure S11, lane U17C:A18U '+'). These results suggest that the residue identities of U17 and A18 do not play a role in eIF4E recruitment. Nevertheless, backbone atoms of U17 and A18 may contribute to folding through interactions with A23 and U24.

To probe the role of the first two pairs of the C and G domains, each C was mutated to a U individually and the

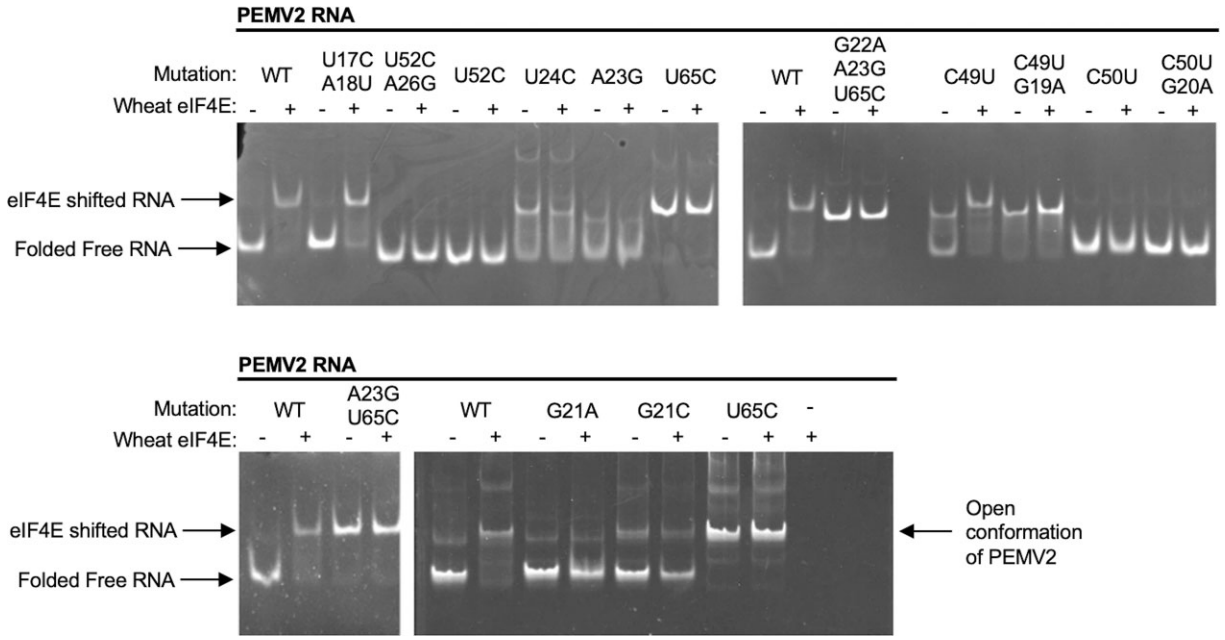


Figure 5. Analysis of eIF4E complex formation with PEMV2 constructs by EMSA. Gel and running buffer contained 10 mM $MgCl_2$. RNA was visualized using ethidium bromide. Protein was visualized using Coomassie staining of this gel (Supplementary Figure S11). Mutations present in each tested construct are indicated above their respective lanes. Lanes labeled with a '−' contain RNA alone, while lanes labeled with a '+' contain 1.5-fold excess of recombinantly expressed wheat eIF4E in addition to the refolded RNA.

corresponding compensatory G to A mutations were made in double mutants. In the absence of eIF4E, C49U migrates as two bands, suggesting two conformations (Figure 5 and Supplementary Figure S11, lane C49U '−'). One band migrates with a mobility similar to the wild-type native fold and the other migrates as the open conformation, possibly reflecting disruption of the interaction between C49's exocyclic amine with the phosphodiester of G67 (Supplementary Figure S2). Despite the misfolded fraction of the C49U mutant, in the presence of eIF4E the RNA–protein complex migrates as a single band, and Coomassie staining of the gel shows eIF4E co-migrating with the RNA (Figure 5 and Supplementary Figure S11, lane C49U '+'). Possibly, eIF4E can bind to both conformations and the respective complexes migrate as a single species. Alternatively, the two RNA conformations may interconvert, and eIF4E binding to the compact conformation exclusively shifts the equilibrium to draw the RNA into complex formation. These observations suggest that C49U may form favorable interactions with G19 through formation of a wobble pair, which could help G19 to maintain native interactions with G22 and C51 (Supplementary Figure S7). Making the compensatory G19A mutation in the C49U background more strongly favors the open conformation represented by the upper band and abolishes eIF4E binding. While G19A would be able to use its sugar edge to maintain the interaction with G22, the absence of the C2 exocyclic amine precludes the interaction with C51 (Supplementary Figure S7).

In the second base pair, the C50U mutant quantitatively retains the compactness of the native fold based upon electrophoretic mobility (Figure 5 and Supplementary Figure S11, lane C50U '−'), suggesting an even greater tolerance to wobble pairing than the C49U mutant. Nevertheless, in contrast to C49U, C50U eliminates eIF4E binding activity (Figure 5 and Supplementary Figure S11, lane C50U '+'). The compensatory G20A mutation also retains the native-like electrophoretic mobility but does not recover eIF4E binding (Figure 5 and Supplementary Figure S11, lane C50U:G20A '+'). This result suggests that eIF4E recruitment may be more complex than simply recognizing the exposed G in the folded PTE structure, possibly involving direct recognition of the G20–C50 pair or other features of the scaffold surface. Despite some covariation of pairing partners among close homologs of PEMV2 PTE, the base pair corresponding to G20–C50 remains strictly conserved. Intriguingly, the Hoogsteen edge of the G20–C50 pair engages in a base triple crystal contact with the WCF face of G1 (Supplementary Figure S10) underscoring its potential for macromolecular recognition.

To perturb the quadruple base interaction involving G22, A23, C51 and U65, we mutated U65 to C and its partner A23 to G separately (Figure 5 and Supplementary Figure S11, lanes U65C '−' and A23G '−'). The U65C mutant migrates as a single slower moving band compared to wild type and does not bind eIF4E, presumably reflecting the expected disruption to interactions with A23 and C51. The A23G mutant migrates as two bands; the major band migrates with the mobility of the native conformation and the minor one with the mobility of the open conformation. Despite the apparent ability to maintain the native fold, the A23G mutant does not bind eIF4E. Although A23G maintains some compaction, the compensatory mutation A23G U65C disrupts this (Figure 5 and Supplementary Figure S11, lane G22A:A23G:U65C '−'). These results establish the structural importance of A23 and U65 pair

beyond their ability to pair. Consistent with the apparent specific requirement of an A–U base pair, the sequence and SHAPE-derived secondary structures of PTE homologs from other viruses indicate that uridine predominates as the J5/3 nucleotide.

To perturb the U52–A26–U24 base triple at the base of the motif, we constructed a U52C mutant, which would be expected to disrupt the U52–A26 base pair within the base triple. This construct migrates similarly to wild type (Figure 5, lane U52C '−'), but we observe no mobility shift in the presence of eIF4E (Figure 5 and Supplementary Figure S11, lane U52C '+'). A second compensatory mutation A26G in the context of U52C gives results similar to those for the U52C single mutation alone (Figure 5 and Supplementary Figure S11, lanes U52C:A26G '−' and '+'). Possibly, the presence of four consecutive Cs in the C domain (residues 49–52) of the U52C variant results in pairing with the four Gs in the G domain (residues 19–22) to form an antiparallel pseudoknot. This alternate pairing mode would be expected to retain and possibly enhance compaction and eliminate eIF4E binding through alteration of the structure, including sequestration of the critical G21 residue.

To disrupt the U24–A26 WCF–sugar edge interaction within the triple, we constructed the U24C mutant. In the EMSA, this mutation drastically alters the electrophoretic signature of the PTE, creating three conformers, one migrating close to the native conformation, one migrating like the open conformation observed with other mutants and one significantly slower migrating conformation that could involve multimerization (Figure 5 and Supplementary Figure S11, lane U24C '−'). None of these conformations appears to bind eIF4E (Figure 5 and Supplementary Figure S11, lane U24C '+').

To probe the role of G21 in eIF4E binding, we constructed two mutants G21A and G21C. PTE constructs bearing these mutations retained compactness (Figure 5 and Supplementary Figure S11, lanes G21A '−' and G21C '−') but lacked the ability to bind eIF4E (Figure 5 and Supplementary Figure S11, lanes G21A '+' and G21C '+'). These data together suggest that the PTE interacts with eIF4E via G21 using the same binding pocket as m^7GTP (28,29). Collectively, the current mutational analysis supports the functional relevance of the crystal structure and illustrates the specific and complex network of hydrogen bonds at work stabilizing the unusual conformation of the G domain.

DISCUSSION

The PTE architecture creates an eIF4E binding surface

Our structure of the PTE from PEMV2 shows that G21 within the G domain is flipped out into solution, consistent with computational modeling and SHAPE hyper-reactivity (28). The 2′-*endo* sugar pucker at G21, which renders the 2′OH sterically accessible, may also contribute favorably to the enhanced reactivity with SHAPE (6,30). Supporting the unusual conformation of the G domain is a complex network of stacking and hydrogen bonding interactions between joining regions of the three-way junction and the G domain bulge that pinch and flip the G domain backbone, leaving no space for G21 to insert into the RNA stack. The clear and complete density corresponding to G21 suggests a well-ordered and relatively static position in the crystal, likely reflecting the crystal contact between G21 and the Fab heavy chain (Supplementary Figure S6).

Biochemical data and associated computational modeling suggest that binding of the PTE to eIF4E involves docking of the exposed G21 into the m^7G cap-binding pocket of eIF4E. eIF4E binding protects only the G domain in RNase T1 probing experiments and only the hypermodifiable G in 1M7 SHAPE probing experiments; mutation of the two tryptophans (W62 and W108) in the cap-binding pocket drastically reduces eIF4E's affinity. Moreover, the G21A mutation in the PTE significantly reduces translation (28,29). Our eIF4E binding analysis showed that mutations that shift the PTE fold to the open conformation abolished eIF4E binding, presumably by altering the position or dynamic character of the flipped out G21 and the surrounding surface. These mutations could not be recovered by compensatory mutations, suggesting that the complex network of hydrogen bonds creates site-specific requirements for residue identity. Even mutants that retained a degree of compactness analogous to wild type, including mutation of G21 itself, lose eIF4E binding capacity, indicating that a compact fold, while necessary, is not sufficient to create a productive eIF4E binding interaction. Possibly, docking of the exposed G into the cap-binding pocket alone is not sufficient for eIF4E binding, and additional constraints on PTE conformation or interactions between the PTE surface and eIF4E are necessary.

Biophysical characterization of eIF4E binding to m^7GpppG reveals a binding free energy of −10.8 kcal/mol, reflecting contributions from base stacking, hydrogen bonding and electrostatic interactions (48). Pi stacking interactions between the methylated guanine base and the two tryptophans above and below make the major binding contribution (−4.9 kcal/mol), aided by hydrogen bonding interactions between the binding pocket and the WCF face of the guanine. The methylation mark helps to orient and add additional positive charge character to the nucleobase, estimated to strengthen the stacking interaction by roughly −3 kcal/mol compared to corresponding unmodified cap analogs. Additional free energy contributions come from electrostatic interactions between the protein and the three phosphates of the cap structure (48). Consistent with the importance of electrostatic interactions, increasing monovalent cation concentration attenuates binding, whereas appending the cap structure to an RNA oligonucleotide strengthens binding (49).

PEMV2 PTE binds significantly better to wheat eIF4E than does GMP or GTP (28,48). If PEMV2's interaction with eIF4E were solely mediated by G21 insertion into the binding pocket, its affinity would likely most closely resemble that of GMP ($K_d \sim 5$ mM). Therefore, considering the biophysical properties of eIF4E cap–analog interactions, it is likely that the PEMV2 core fold creates additional interactions with eIF4E. Beyond the expected stacking interactions between eIF4E's tryptophans and hydrogen bonding interactions with G21's WCF face, additional interactions could consist of hydrogen bonding and electrostatic interactions involving exposed surface of the PEMV2 core, such as the G20–C50 pair.

Homologous PTE sequences have the potential to adopt the PEMV2-like fold

The motif described here for PEMV2 involves tertiary interactions between an asymmetric bulge (the G domain) and a type A three-way junction. Among these tertiary interactions are edge-to-edge interactions of bases in *cis* or *trans* orientation of the glycosidic bonds. Two of four WCF pairs involve a parallel arrangement of the strand backbones; another interaction includes the bulged nucleotide near the base of P3 and engages in an additional interaction with a G domain nucleotide to form a base triple; we also observe a Hoogsteen/Hoogsteen interaction, between a residue in the C domain and J5/3. The secondary structures of PEMV2 homologs from other viruses have the potential to conform to this general strategy for organizing the G domain through interactions with a three-way junction to expose and present a guanosine residue to eIF4E. This organizational strategy includes coaxial stacking of P3 and P4 together with parallel double-strand interactions between the C and G domains that topologically constrain the length of P3. Additionally, the lengths of the joining regions of the three-way junction should conform to the type A specifications. Published secondary structure models for the PTE homologs can be slightly adjusted to conform to the motif described here for PEMV2 (28) via alterations analogous to those revealed by our tertiary structure for PEMV2.

In this exercise, the secondary structure models for some homologs required additional rearrangements to create junction and paired region lengths that match PEMV2's A-type junction (Table 3 and Figure 6). The paired regions and junctions of PTEs belonging to PMV, CMMV and TPAV easily accommodated the changes inferred from our PEMV2 secondary structure with no additional adjustments needed to match the conditions of the PTE three-way junction (Table 3) (28). For PMV PTE, G4184 was made to pair with C4137, which increases the length of P3 to 8 bp and subsumes G4138 into P3 as a bulged residue. Consequently, the G domain and J3/2 become shorter by two and one nucleotide(s), respectively. This adjusted secondary structure readily conforms to the pairing rules outlined above. The analogous operations for CMMV and TPAV, in which C4052 and G4050 are made to pair with G4003 and C4003, respectively, render the corresponding P3 helices 8 bp in length and subsume A4004 and G4004 into P3 as bulged nucleotides, respectively. These adjusted secondary structures also readily conform to the pairing rules derived from PEMV2. In the case of CarMV PTE, the final base pair (C3748 and G3813) of P2 was broken and G3813 was instead made to pair with C3753, increasing the length of P3 to 8 bp and subsuming G3754 into P3 as a bulged nucleotide. Additionally the pair between C3814 and G3747 was broken and instead C3814 was paired with G3746. Other PTEs required more extensive rearrangements but could still meet the constraints defined by the PEMV2 fold. PFBV was altered by pairing U3718 with A3703 and breaking the C3722–G3742 base pair so that J4/5 now consists of the sequence UCCC, leaving G3781 as the single residue in J5/3. At the base of P3, the G3695–U3790 pair was revised so that U3790 could pair with A3694. Further study is required to ascertain whether these newly proposed secondary structures, revised to allow PEMV2-like pairing, reflect the actual structures. Nevertheless, the revised structures remain consistent with the reported chemical probing sensitivity.

Table 3. PTE homologs' revised A-type junction motif comparison

	PTE homologs					
Structured element	PEMV2	PMV	CMMV	TPAV	CarMV	PFBV
P2	6[a] (5)	5	7	6	4[a] (5)	8
G domain	8[a] (11)	6[a] (8)	8[a] (10)	6[a] (8)	6	6[a] (7)
P3	8[a] (7)	8[a] (7)	8[a] (7)	8[a] (7)	8[a] (7)	8
J3/4	0	0	0	0	0	0[a] (1)
C domain	4	4	4	4	4	4
J5/3	1	1	1	1	1	1[a] (0)
J3/2	0[a] (2)	0[a] (1)	0[a] (1)	0[a] (1)	0	0

[a]The length of this region was altered from the predictions made in (28).

Three of the biochemically characterized PTE homologs were too different from PEMV2 to create a secondary structure that matches PEMV2's three-way junction components. HCRSV contains an additional bulge in J3/2 absent in the other PTEs considered here. JINRV has a shorter P3 and a longer J4/5, which makes the sites of G and C domain interactions more ambiguous. SCV PTE is even more divergent from the PEMV2 structure with a 6-bp P3, a 1-nt J3/4, a 6-nt J4/5 and a 3-nt J5/3. These differences could substantially alter the alignment of the G and C domain pairing residues from the PEMV2 structure model, making this set of interactions difficult to infer from the PEMV2 structure. Therefore HCRSV, JINRV and SCV were excluded from the covariation analysis that follows.

In the context of these adjusted PTE secondary structure models, we then asked whether the homologs could make long-range pairing interactions between the G domain and the three-way junction analogous to those observed in PEMV2. We found that the PTEs that could adopt a PEMV2-like three-way junction also show covariation between predicted interacting residues, consistent with the PEMV2 model and supporting the functional relevance of our structure. This includes the parallel base pairing between J2/3 and J4/5.

To create the eIF4E binding site in a PEMV2-like manner, an internal G of the G domain must be extruded from the core stack via an acute direction reversal of the phosphodiester backbone. In PEMV2, there are two G domain–C domain pairs upstream of the eIF4E binding G and two pairs downstream of the eIF4E binding G. In each of the PTE homologs, we see that the first of the two residues of the C domain covaries with a residue near the start of the G domain. Additionally, in accord with the parallel pairing observed for PEMV2, the covarying positions in the homologs are followed by a G residue in the G domain and a C residue in the C domain. These two parallel pairs are therefore likely retained in each of the five PTEs capable of forming the same three-way junction as PEMV2. In some cases, this structure-guided modeling predicts that the extruded G residue resides one nucleotide downstream of the position originally annotated as hypermodified by SHAPE analysis (Figure 6). This discrepancy possibly results from ambiguity in assigning the position of reverse transcriptase stops

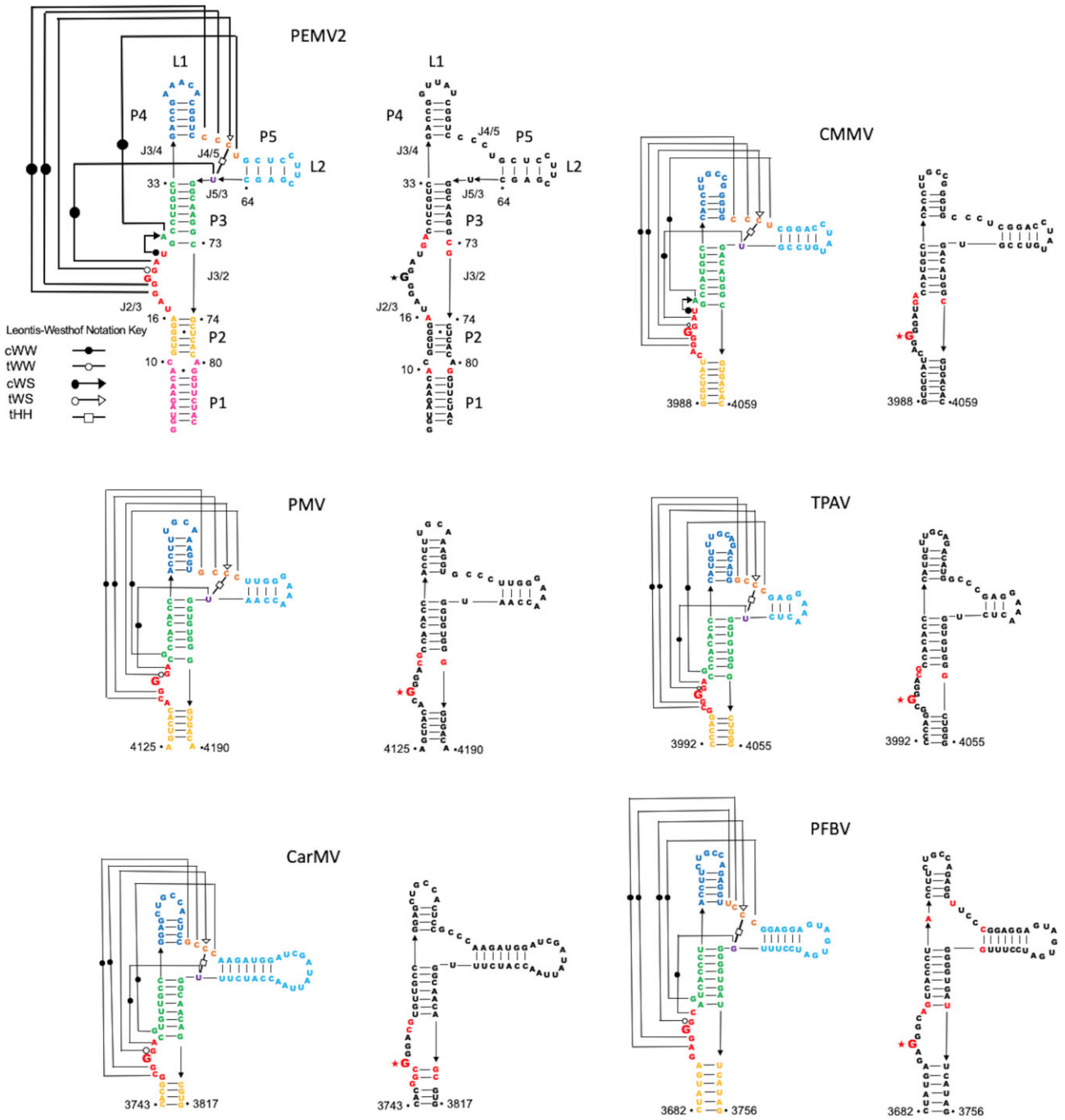


Figure 6. PEMV2 long-distance interaction network superimposed on the PTE homologs' revised secondary structures. Homologous PTE sequences adjusted to conform to the PEMV2 fold are in color and shown with predicted tertiary interactions. PTE secondary structures proposed in (28) are shown in black, where red bases indicate newly revised structure predictions. Black star indicates the same location of 'hyper-reactive' G in both depicted predictions, and red star indicates different locations of 'hyper-reactive' G between depicted predictions.

via DPAGE (28). More distant PTE homologs like SCV with a longer C domain bulge may use different residues in the C domain to organize the G domain conformation.

Continuing to view the PEMV2 structure as though it represents a general arrangement of PTEs, the residue immediately downstream of the bulged G undergoes a backbone reversal that directs its minor grove edge toward the minor grove of the first parallel base pair and its WCF edge toward the sugar edge of the third residue of the C domain. Like PEMV2, the homologs contain G and C at these respective positions, suggesting that the G–C WCF–sugar edge interaction observed in PEMV2 may be present in all of these homologs. However, in PEMV2 both of these nucleotides interact with the minor grove edge of the G19 in the first parallel base pair, whereas some homologs have a C–G as the first parallel base pair, suggesting that

alternative hydrogen bonding would have to occur to maintain the scaffold's close-packed structure.

In PEMV2, the second residue downstream from the hypermodifiable G forms a WCF base pair with the J5/3 residue. These positions in homologous PTEs maintain WCF complementarity: PMV A–U, CMMV A–U, TPAV A–U, CarMV A–U and covary in PFBV C–G, strongly suggesting that the homologs preserve this pairing feature.

The final residues of the G domain alternate between P3 and the scaffold structure. U24 forms a base triple, sugar edge interaction with the bulged nucleotide in P3, A26, which flips out of the P3 stack and intercalates into the scaffold stack to pair with the last residue of the C domain. Concomitantly, G25 forms the first base pair of P3 with the residue originally attributed to J3/2. In PMV, CMMV, TPAV, JINRV and CarMV, after the proposed secondary structure rearrangements, the J3/2 residue within each PTE can pair with its respective second-to-last residue of the G domain to start P3, leaving the last residue of the G domain free, as a bulge in P3, to pair with the final residue of J4/5. In the published secondary structure for PFBV (28), the P3 helix contains eight base pairs and no bulge near its base. However, we suggest that U3750 and G3695 can be unpaired from each other and made to pair with A3694 and C3721, respectively, to adopt the form of PEMV2 (Figure 6). In CarMV's secondary structure map, the bulged G3746 can be modeled to pair with the proximal C3814, which would move G3747 and C3748 into the G domain and enable G3813 to pair with C3753 at the start of P3. Revising the secondary structures of the PTE homologs to match features of PEMV2 reveals a conserved pattern of long-distance pairing interactions.

Two PTE subclasses emerge from the proposed secondary structure alterations that differ in the number of G domain residues and in the presence of the capacity for the base triple interaction involving the P3 bulged nucleotide. From this perspective, CMMV most resembles PEMV2 with the same C domain sequence, CCCU and eight nucleotides in the G domain (after the proposed secondary structure adjustments), including four consecutive Gs and a U that forms the base triple. The other subclass includes PMV, TPAV and CarMV. These PTEs have a distinct C domain sequence, GCCC and two fewer nucleotides in their G domains (after the proposed secondary structure adjustments), which include three consecutive Gs and lack the nucleotide for base triple formation.

CONCLUSIONS

Presented here is a high-resolution crystal structure of a 3′ CITE, which, to the best of our knowledge, represents a new RNA structural motif. The structure supports the hypothesis that the PEMV2 PTE uses a flipped G in the G domain to bind eIF4E; nevertheless, our mutational analysis suggests that architectural features in addition to the flipped G are likely required for forming stable eIF4E interactions. In accordance with the crystallograpically defined secondary structure of PEMV2, homologous PTE sequences harbor a G-rich bulge and can accommodate topologically similar three-way junction. Moreover, their sequences accommodated revisions to their secondary structures and exhibited covariation at crucial positions, suggesting that the PEMV2 tertiary architecture is conserved in these homologs. Notably, this pattern of conservation would have been difficult to identify with phylogeny and chemical probing data alone, underscoring the value of combining these data with high-resolution structure analysis.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the PDB (https://www.rcsb.org/) under accession number 8SH5.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Allison Brink, Wojciech Gogacz, Daniel Krochmal and Matt Moore for comments on the manuscript. We thank Deepak Koirala for assistance with construct design early on in the work. We thank Kay Perry for help in data collection and valuable advice during initial phases of solving tertiary structure. We also thank Bryan Dickinson and Jon Staley for advice and suggestions through the course of the work. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beamline is funded by an NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

FUNDING

National Institute of General Medical Sciences [GM102489 and GM149336 to J.A.P.]; Howard Hughes Medical Institute (to C.R.); National Institutes of Health [T32GM008720 to C.R. and M.D.]; National Science Foundation [DGE-1746045 to M.D.]; University of Chicago [R25GM066522 to S.J.]. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Browning,K.S. and Bailey-Serres,J. (2015) Mechanism of cytoplasmic mRNA translation. *Arabidopsis Book*, **13**, e0176.
2. Sonenberg,N. and Hinnebusch,A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731–745.
3. Shatkin,A.J. and Manley,J.L. (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.*, **7**, 838–842.
4. Moteki,S. and Price,D. (2002) Functional coupling of capping and transcription of mRNA. *Mol. Cell*, **10**, 599–609.
5. Jaafar,Z.A. and Kieft,J.S. (2019) Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.*, **17**, 110–123.

6. Miyoshi,H., Suehiro,N., Tomoo,K., Muto,S., Takahashi,T., Tsukamoto,T., Ohmori,T. and Natsuaki,T. (2006) Binding analyses for the interaction between plant virus genome-linked protein (VPg) and plant translational initiation factors. *Biochimie*, **88**, 329–340.
7. Bonilla,S.L., Sherlock,M.E., MacFadden,A. and Kieft,J.S. (2021) A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. *Science*, **374**, 955–960.
8. Martinez-Salas,E., Francisco-Velilla,R., Fernandez-Chamorro,J. and Embarek,A.M. (2018) Insights into structural and mechanistic features of viral IRES elements. *Front. Microbiol.*, **8**, 2629.
9. Zhang,J., Roberts,R. and Rakotondrafara,A.M. (2015) The role of the 5′ untranslated regions of Potyviridae in translation. *Virus Res.*, **206**, 74–81.
10. 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.
11. Simon,A.E. and Miller,W.A. (2013) 3′ Cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.*, **67**, 21–42.
12. Truniger,V., Miras,M. and Aranda,M.A. (2017) Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). *Front. Plant Sci.*, **8**, 2047.
13. Terenin,I.M., Andreev,D.E., Dmitriev,S.E. and Shatsky,I.N. (2013) A novel mechanism of eukaryotic translation initiation that is neither m^7G-cap-, nor IRES-dependent. *Nucleic Acids Res.*, **41**, 1807–1816.
14. Shatsky,I.N., Terenin,I.M., Smirnova,V.V. and Andreev,D.E. (2018) Cap-independent translation: what's in a name?*Trends Biochem. Sci.*, **43**, 882–895.
15. Weingarten-Gabbay,S., Elias-Kirma,S., Nir,R., Gritsenko,A.A., Stern-Ginossar,N., Yakhini,Z., Weinberger,A. and Segal,E. (2016) Comparative genetics: systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*, **351**, aad4939.
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. Miras,M., Sempere,R.N., Kraft,J.J., Miller,W.A., Aranda,M.A. and Truniger,V. (2014) Interfamilial recombination between viruses led to acquisition of a novel translation-enhancing RNA element that allows resistance breaking. *New Phytol.*, **202**, 233–246.
18. 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.
19. Meulewaeter,F., Van Montagu,M. and Cornelissen,M. (1998) Features of the autonomous function of the translational enhancer domain of satellite tobacco necrosis virus. *RNA*, **4**, 1347–1356.
20. Wang,Z., Kraft,J.J., Hui,A.Y. and Miller,W.A. (2010) Structural plasticity of Barley yellow dwarf virus-like cap-independent translation elements in four genera of plant viral RNAs. *Virology*, **402**, 177–186.
21. 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.
22. Mizumoto,H., Tatsuta,M., Kaido,M., Mise,K. and Okuno,T. (2003) Cap-independent translational enhancement by the 3′ untranslated region of red clover necrotic mosaic virus RNA1. *J. Virol.*, **77**, 12113–12121.
23. Shen,R. and Miller,W.A. (2004) The 3′ untranslated region of tobacco necrosis virus RNA contains a Barley yellow dwarf virus-like cap-independent translation element. *J. Virol.*, **78**, 4655–4664.
24. Meulewaeter,F., van Lipzig,R., Gultyaev,A.P., Pleij,C.W.A., Van Damme,D., Cornelissen,M. and van Eldik,G. (2004) Conservation of RNA structures enables TNV and BYDV 5′ and 3′ elements to cooperate synergistically in cap-independent translation. *Nucleic Acids Res.*, **32**, 1721–1730.
25. Nicholson,B.L., Zaslaver,O., Mayberry,L.K., Browning,K.S. and White,K.A. (2013) Tombusvirus Y-shaped translational enhancer forms a complex with eIF4F and can be functionally replaced by heterologous translational enhancers. *J. Virol.*, **87**, 1872–1883.
26. 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.
27. Zuo,X., Wang,J., Yu,P., Eyler,D., Xu,H., Starich,M.R., Tiede,D.M., Simon,A.E., Kasprzak,W., Schwieters,C.D. *et al.* (2010) Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3′ UTR of turnip crinkle virus. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 1385–1390.
28. 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.
29. 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.
30. Gao,F., Gulay,S.P., Kasprzak,W., Dinman,J.D., Shapiro,B.A. and Simon,A.E. (2013) The kissing-loop T-shaped structure translational enhancer of Pea enation mosaic virus can bind simultaneously to ribosomes and a 5′ proximal hairpin. *J. Virol.*, **87**, 11987–12002.
31. Gao,F. and Simon,A.E. (2017) Differential use of 3′CITEs by the subgenomic RNA of Pea enation mosaic virus 2. *Virology*, **510**, 194–204.
32. Jones,R.A.C. (2021) Global plant virus disease pandemics and epidemics. *Plants*, **10**, 233.
33. Javed,T., Shabbir,R., Ali,A., Afzal,I., Zaheer,U. and Gao,S.J. (2020) Transcription factors in plant stress responses: challenges and potential for sugarcane improvement. *Plants*, **9**, 491.
34. Shahzad,R., Jamil,S., Ahmad,S., Nisar,A., Amina,Z., Saleem,S., Zaffar Iqbal,M., Muhammad Atif,R. and Wang,X. (2021) Harnessing the potential of plant transcription factors in developing climate resilient crops to improve global food security: current and future perspectives. *Saudi J. Biol. Sci.*, **28**, 2323–2341.
35. Truniger,V., Nieto,C., González-Ibeas,D. and Aranda,M. (2008) Mechanism of plant eIF4E-mediated resistance against a *Carmovirus* (*Tombusviridae*): cap-independent translation of a viral RNA controlled *in cis* by an (a)virulence determinant. *Plant J.*, **56**, 716–727.
36. Huang,H., Suslov,N.B., Li,N.S., Shelke,S.A., Evans,M.E., Koldobskaya,Y., Rice,P.A. and Piccirilli,J.A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.*, **10**, 686–691.
37. Koldobskaya,Y., Duguid,E.M., Shechner,D.M., Suslov,N.B., Ye,J., Sidhu,S.S., Bartel,D.P., Koide,S., Kossiakoff,A.A. and Piccirilli,J.A. (2011) A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol.*, **18**, 100–107.
38. Ye,J.D., Tereshko,V., Frederiksen,J.K., Koide,A., Fellouse,F.A., Sidhu,S.S., Koide,S., Kossiakoff,A.A. and Piccirilli,J.A. (2008) Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 82–87.
39. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 213–221.
40. Winter,G. (2010) Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.*, **43**, 186–190.
41. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**, 352–367.
42. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
43. Williams,C.J., Headd,J.J., Moriarty,N.W., Prisant,M.G., Videau,L.L., Deis,L.N., Verma,V., Keedy,D.A., Hintze,B.J., Chen,V.B. *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.
44. Lescoute,A. and Westhof,E. (2006) Topology of three-way junctions in folded RNAs. *RNA*, **12**, 83–93.
45. Ulyanov,N.B. and James,T.L. (2010) RNA structural motifs that entail hydrogen bonds involving sugar–phosphate backbone atoms of RNA. *New J. Chem.*, **34**, 910–917.

46. Staple,D.W. and Butcher,S.E. (2005) Pseudoknots: RNA structures with diverse functions. *PLoS Biol.*, **3**, e213.
47. Murphy,F.L., Wang,Y.H., Griffith,J.D. and Cech,T.R. (1994) Coaxially stacked RNA helices in the catalytic center of the *Tetrahymena* ribozyme. *Science*, **265**, 1709–1712.
48. Niedzwiecka,A., Marcotrigiano,J., Stepinski,J., Jankowska-Anyszka,M., Wyslouch-Cieszynska,A., Dadlez,M., Gingras,A.C., Mak,P., Darzynkiewicz,E., Sonenberg,N. *et al.* (2002) Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5′ cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol.*, **319**, 615–635.
49. Slepenkov,S.V., Darzynkiewicz,E. and Rhoads,R.E. (2006) Stopped-flow kinetic analysis of eIF4E and phosphorylated eIF4E binding to cap analogs and capped oligoribonucleotides: evidence for a one-step binding mechanism. *J. Biol. Chem.*, **281**, 14927–14938.