ABSTRACT

Title of Document: LONG- AND SHORT-DISTANCE RNA:RNA INTERACTIONS MODULATE VITAL VIRAL PROCESSES IN TWO CARMOVIRUSES

Maitreyi Chattopadhyay, Doctor of Philosophy, 2015

Directed By: Professor Anne E. Simon, Cell Biology and Molecular Genetics

Saguaro cactus virus (SCV) and *Turnip crinkle virus* (TCV) are two plant positive-strand [(+)-strand] RNA carmoviruses that serve as model systems for studying long- and short-distance RNA:RNA interactions required for essential viral processes in host cells. Our search for conserved 3' cap-independent translational elements (3' CITE) in the carmovirus genus led to the discovery of a *Panicum mosaic virus*-like 3'CITE (PTE) in seven carmoviruses, including SCV. We found that the SCV PTE participates in long-distance kissing-loop interactions with hairpins located in the p26 open reading frame (ORF) of the genomic RNA (gRNA) and in the 5' untranslated region (UTR) of the small subgenomic RNA (sgRNA2), and promotes efficient translation of both gRNA and sgRNA2 reporter mRNAs in plant protoplasts. For nearly all PTE-containing carmoviruses, a 5' hairpin that participates in a kissing-loop interaction with the PTE is found in conserved locations in the 5' ends of the gRNA and sgRNA2. We found that moving the 5' hairpin 27 nt closer to

the 5’ end had only a minimal effect on both gRNA and sgRNA2 translation. However, moving the 5’ hairpin further in the direction of the 5’ end had a detrimental effect.

The higher order structure of the TCV 3’ region is supported by a complex network of canonical and noncanonical medium- and short-distance RNA:RNA interactions that modulate many critical viral processes. We investigated several second-site mutations in the coat protein (p38) ORF that arose in response to a primary-site mutation (m21) located in critical 3’UTR hairpin H4. All tested second-site mutations were compensatory with a partial reversion of m21 (TCV-rev1). Selective 2’-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) structure probing indicated that these second-site mutations in the p38 ORF reside in a separate RNA domain (Domain2) and provided evidence for RNA:RNA interactions between Domain2 and 3’UTR-containing Domain1. These second-site mutations, however, were not compensatory in the absence of p38, the TCV silencing suppressor, or in the absence of functional Dicer-like proteins (DCLs) and RNA induced silencing complex (RISC). Our results indicate an unexpected connection between a 3’ primary-site mutation and the RNA silencing machinery in TCV.

LONG-AND SHORT-DISTANCE RNA:RNA INTERACTIONS MODULATE VITAL VIRAL PROCESSES IN TWO CARMOVIRUSES

By

Maitreyi Chattopadhyay

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Anne E. Simon, Chair
Professor Jonathan Dinman
Professor Jeffery DeStefano
Professor Kim Y. Green
Professor James N. Culver, Dean's Representative

Dedication

I would like to dedicate this dissertation to my daughter Abjini and my husband Somesh, for their immense support and encouragement throughout my graduate studies at UMD. I would also like to dedicate this dissertation to my parents Sunil and Jogomaya Bhattacharyya for their unconditional love and nurture.

Acknowledgements

I would like to thank Dr. Simon from the bottom of my heart for all her guidance, patience and support. I would also like to express my sincere gratitude to Dr. Dinman, Dr. DeStefano, Dr. Culver and Dr. Green for not losing their faith in me and their positive attitude. It would not be possible without two amazing post-docs of our lab, Vera and Feng. They are my very close friends, my biggest supporters and they are the treasure of our lab. They never hesitated to lend their hands to help me out through difficult times. Teresa, my friend and our former secretary, provided much needed comfort throughout my studies at UMD. I would also like to thank current and former members of Dr. Simon's lab including Micki, MyLe, Phillip and Megan, who made my life easier. I have met many amazing mentors throughout my journey as a graduate student at UMD, and especially I would like to thank Dr. Moctezuma, Dr. Heven Sze, and Dr. Zhongchi Liu for all their help.

Without my beloved husband Somesh, who has always supported me and encouraged me to be my best, I would not make it so far. My sweet daughter, Abjini is a blessing to me. I love her the most.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ΔgH1: Deletion of genomic hairpin1

3’343m: 3’343 mutant

3’CITE: Cap-independent translation element

AL : Asymmetric loop

AS: Asymmetric loop

BTE: Barley yellow dwarf- translational element

CarMV: Carnation mottle virus

CbMV: Calibrachoa mottle virus

CP: Coat protein

CrPV: Criket paralysis virus

DCL: Dicer-like proteins

DMSO: Dimethyl sulfoxide

dNTP: Deoxynucleotide tripphospahte

dpi: Days post inoculation

ds: Double stranded

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic acid

eIF2: Eukaryotic initiation factor 2

eIF4A: Eukaryotic initiation factor 4A

eIF4B: Eukaryotic initiation factor 4B

eIF4E: Eukaryotic initiation factor 4E

eIF4G: Eukaryotic initiation factor 4G

ELS: Elongated lower stem

FDMV: Foot and mouth disease virus

F-Luc: Firefly luciferase

GaMV: Galinsoga mosaic virus

GFP: Green florescence protein

gH1: Genomic hairpin 1

gH3: Genomic hairpin 3

GTP: Guanosine triphosphate

h: Hour

H1: Hairpin 1

H2: Hairpin2

H4: Hairpin 4

H5: Hairpin5

HCRSV: Hibiscus chlorotic ringspot virus

HCV: Hepatitis C virus

HnRSV: Honey suckle ringspot virus

Hpi: Hours post inoculation

IRES: Internal ribosome entry site

ISS: I-shaped structure

LS: Lower stem

Met-tRNA$^{Met}_{i}$: Methionine initiator t-RNA

mg: Milligram

min: Minute

miRNA: MicroRNA

ml: Milliliter

mRNA: Messenger RNA

NaCl: Sodium chloride

NMIA: N-methylisotoic anhydride

nt: Nucleotide

OH: Hydroxyl ladder

ORF: Open reading frame

PAPB: Ploy-A binding protein

PCM: Protoplast culture medium

PCR: Polymerase chain reaction

PEG: Polyethylene glycol

pH: Measure of acidity or alkalinity of a solution

PIM: Protoplasts isolation medium

Pmol: Picomole

PNK: Polynucleotide Kinase

Pr: Promoter

PSNV: Pea stem necrosis virus

PTE: Panicum mosaic virus like-translation enhancer

PV: Poliovirus

RISC: RNA induced silencing complex

RISC: RNA induced silencing complex

R-Luc: Renilla luciferase

RNA: Ribonucleic acid

satC: Satellite RNA C

SDS: Sodium dodecyl sulfate

sg5’137: subgenomic 5’ 137 nucleotide

sgH1: Subgenomic Hairpin 1

sgRNA: Subgenomic RNA

SHAPE: Selective 2′-hydroxyl acylation analyzed by primer extension

SiRNA: Small interfering RNA

T1: RNAse T1

TSS: tRNA-shaped structure

US: Upper stem

UTR: Untranslated region

UV: Ultraviolet

vsRNA: Viral small RNA

wt: Wild type

YSS: Y-shaped structure

Chapter 1: Diverse roles and interactions of *cis*-acting RNA elements in the viral life cycle

Introduction

A virus, by definition, is an obligate, intracellular parasite. The discovery of viruses as pathogens smaller than bacteria dates back to 1892 when Russian botanist Dmitri Ivanovsky observed that the agent causing tobacco mosaic disease could not be filtered by the porcelain filters capable of removing bacteria (1). Positive-strand RNA viruses, with 30 families, constitute the largest group of RNA viruses and contribute significantly to the global burden of infectious disease leading to loss of many human and animal lives as well as loss of economically important crops.

Infection by plant (+)-strand RNA viruses differs from that of animal viruses. Animal viruses enter host cells via receptor-mediated endocytosis or fusion with the cell's plasma membrane. Plant viruses enter cells following transmission by insect vectors or during mechanical injury. No cell surface receptors have been identified for plant virus entry. Uncoating of most animal viruses occurs by removal of the viral capsid protein either in an acidic environment or through digestion by proteases inside the endocytic vesicles of the host cell (2, 3). For many plant viruses, the uncoating event is poorly understood. Studies have found that *Tobacco mosaic virus* (TMV) undergoes cotranslational uncoating after it enters the cell cytoplasm (4). Following release from the capsid, the viral gRNA is translated. Most (+)-strand RNA viruses utilize non-canonical mechanisms of translation, a strategy that allows these viruses to compete with cellular mRNAs for the translational machinery (5).

Viral replication is a multi-step process that requires the viral encoded RNA-dependent RNA polymerase (RdRp), other viral proteins and numerous host proteins (6-11). Binding of viral-encoded proteins selectively to the viral RNA is proposed to switch the conformation of the viral RNA from a translationally competent template to a replicase competent template (12, 13). Replication takes place in a viral-induced, membrane-bound, multi-protein complex known as the viral replication complex (VRC). The viral RNA is transcribed into a negative-strand RNA [(-)-strand RNA], which is subsequently used as a template for the synthesis of (+)-strand RNA progeny (6, 7, 14-16). Many (+)-strand RNA viruses produce (+)- and (-)-strands asymmetrically, up to 100- fold more (+)-strand RNA progeny than (-)-strand (17). This asymmetric nature of viral RNA synthesis is a hallmark of (+)-strand RNA viruses and requires multiple host factors. For example, *Tomato bushy stunt virus* (TBSV) requires host glyceraldehyde 3-phosphate dehydrogenase (GAPDH) to regulate the asymmetry of RNA synthesis. Addition of purified recombinant GAPDH to the GAPDH-depleted yeast cell extracts results in increased (+)-strand viral RNA synthesis and asymmetric production of viral progeny (18). To express the 3' proximal ORFs, many (+)-strand RNA viruses generate sgRNAs during the late stages of infection, either via premature termination of transcription by the RdRp during (-)-strand RNA synthesis or internal initiation on the (-)-strand during (+)-strand RNA synthesis (19). These ORFs usually encode structural and movement proteins.

The newly synthesized (+)-strand RNAs are used either for a new round of translation and production of more replicase and structural proteins, or are encapsidated to produce viral particles. Encapsidation signals for many RNA viruses are *cis*-acting

RNA sequences that are located either in the 5’ or the 3’ end of the viral RNA (20, 21). Animal viruses either bud off from or lyse the host cell, while plant viral progeny move from cell-to-cell via plasmodesmata as a nucleoprotein particle that does not require the presence of viral coat protein (CP). However, long-distance movement of viral RNA through the phloem requires viral CP (22).

The genomes of (+)-strand RNA viruses are able to fold into complex higher-order structures via multiple short-and long-distance RNA:RNA interactions. These facilitate many basic functions such as translation, replication and evasion of host defenses (12, 23) (Table 1). Reports examining the RNA structural organization and tertiary interactions of the full-length gRNAs of TBSV, satellite of TMV (STMV) and *Human immunodeficiency virus* (HIV) revealed that *cis*-acting RNA secondary structural elements are organized into distinct structural domains of different shapes and sizes protruding from a central backbone (24-26). For TBSV, specific inter- and intra-domain long-distance RNA:RNA interactions regulate replication, sgRNA synthesis, cap-independent translation, and ribosome recoding (27-30). In addition, higher order genome structure can provide an escape from host defense responses. For example, extensive RNA secondary structures throughout a gRNA, known as genome-scale ordered RNA structure (GORS), correlates with virus persistence (31). The host RNA silencing pathway uses RISC and 20-25-nt small interfering RNA (siRNA) duplexes produced by host dicers or DCLs, to target complementary mRNAs for endonucleotide cleavage (32). It has been proposed that highly structured viral genomes can provide structural barriers for target site accessibility by the host silencing machinery (33-35).

Table 1: Names and abbreviations of (+)-strand RNA viruses that have been used in Chapter 1

Virus Name	Abbreviation	Family	Genus	Host
Barley yellow dwarf virus	BYDV	*LUTEOVIRIDAE*	Luteovirus	Plant
Blackcurrent reversion nepovirus	BRV	*POTYVIRIDAE*	Nepovirus	Plant
Cardamine chlorotic fleck virus	CCFV	*TOMBUSVIRIDAE*	Carmovirus	Plant
Carnation Italian ringspot virus	CIRV	*TOMBUSVIRIDAE*	Tombusvirus	Plant
Cauliflower mosaic virus	CaMV	*CAULIMOVIRIDAE*	Caulimovirus	Plant
Cricket paralysis virus	CrPV	*DICISTROVIRIDEA*	Cripavirus	Animal/ Insects
Dengue virus	DENV	*FLAVIVIRIDAE*	Flavivirus	Animal
Foot-and-mouth disease virus	FMDV	*PICORNAVIRIDAE*	Aphthovirus	Animal
Hepatitis C	HCV	*FLAVIVIRIDAE*	Hepacivirus	Animal
Human immunodeficiency virus	HIV	*RETROVIRIDAE*	Lentivirus	Animal
Maize necrotic streak virus	MNeSV	*TOMBUSVIRIDAE*	Tombusvirus	Plant
Melon necrotic spot virus	MNSV	*TOMBUSVIRIDAE*	Carmovirus	Plant
Panicum mosaic virus	PMV	*TOMBUSVIRIDAE*	Panicovirus	Plant
Pea enation mosaic virus	PEMV	*NOT ASSIGNED*	Umbravirus	Plant
Poliovirus	PV	*PICONAVIRIDEA*	Enterovirus	Animal
Red clover necrotic spot virus	RCNMV	*TOMBUSVIRIDAE*	Dianthovirus	Plant
Saguaro cactus virus	SCV	*TOMBUSVIRIDAE*	Carmovirus	Plant
Tobacco mosaic virus	TMV	*VIRGAVIRIDAE*	Tobamovirus	Plant
Tomato bushy stunt virus	TBSV	*TOMBUSVIRIDAE*	Tombusvirus	Plant
Turnip Crinkle virus	TCV	*TOMBUSVIRIDAE*	Carmovirus	Plant
West nile virus	WNV	*FLAVIVIRIDAE*	Flavivirus	Animal
Yellow fever virus	YFV	*FLAVIVIRIDAE*	Flavivirus	Animal

Long-range interactions that connect dispersed regions of the viral genome were first discovered in (+)-strand RNA bacteriophages and have subsequently been found in

various (+)-strand RNA viruses infecting eukaryotes (36, 37). All long-range interactions observed to date involve partially or fully complementary base pairing between interacting partners (kissing-loop interactions) and are usually located in the apical or bulge loop regions within RNA structures (5). Long-range interactions spanning >1000 nucleotides are less common than short-range local interactions (mostly H-type pseudoknots) that help naturally formed RNA secondary structural elements to combine and condense (38). These short-range RNA:RNA interactions are equally important for fundamental viral processes such as viral gene expression, and genome replication in many plant and animal (+)-strand RNA viruses (39).

Translation of viral proteins

Canonical translation

In the host, recruitment of the 43S preinitiation complex, which contains 40S ribosomal subunit, ternary complex eIF2–GTP–Met-tRNA$^{Met}_{i}$ and additional eukaryotic translation initiation factors, to the 5' end of most cellular mRNAs requires a 5' cap (m7G (5')ppp) and a 3' poly-A tail (40). The 5' cap is recognized by eIF4E, which is a part of the eIF4F complex that also contains eIF4A, eIF4B and eIF4G. eIF4G, a core scaffold protein, interacts with another initiation factor, eIF3, to recruit 43S preinitiation complex to the 5' end. The 43S preinitiation complex then scans in the 3' direction until encountering an initiation codon in an optimal Kozak context. Translation elongation begins after 60S subunit joins 40S (40-43). 5' bound eIF4G also interacts with poly-A binding protein (PABP), which circularizes the mRNA by binding to the 3' poly-A tail

thereby synergistically enhancing translational efficiency by placing terminating ribosomal subunits to the 5' end for recycling (44).

Non-canonical translation

As obligate parasites viruses rely on their hosts' translational machinery for translating viral proteins. However, many plant and animal viral RNAs lack a functional 5' cap and /or a poly-A tail. Some viruses like influenza, hanta and yeast L-A virus cap their viral RNAs by stealing or snatching the 5' cap from cellular mRNAs (45-48). Other viruses have adopted different strategies that do not require a 5' cap. Some plant and animal viral RNAs contain a small viral encoded protein (VPg) at the 5' end of their genome as a functional replacement for the 5' cap, and, which directly recruits translation initiation factors such as eIF4E and eIF3 (49, 50). Several (+)-strand RNA viruses, predominantly animal viruses and some cellular mRNA, which are regulated during stress, harbor internal ribosome entry site (IRESs) in their 5' UTRs as an alternative to the cap structure (51, 52). The IRESs are highly structured *cis*-acting RNA elements which recruit ribosomes or ribosomal subunits in the presence or absence of translational initiation factors to the vicinity of the initiation codon (53, 54). For example, IRESs of FMDV, PV and HCV require different combinations of translational initiation factors to recruit 43S preinitiation complex, but the CrPV IRES recruits 43S complex independent of any initiation factors (52). Plant viruses, which have short 5' UTRs, harbor RNA structural elements known as CITEs at the 3' ends that confer cap-independent translation (5). Many IRESs and CITEs require long-range RNA-RNA interactions connecting the 5' and 3' ends of the viral genomes to initiate translation of the virus genome (38).

IRES-mediated translation that requires long-distance RNA:RNA interactions

FMDV utilizes 5'IRES to recruit the 43S preinitiation complex directly to the initiation codon. Studies have found that a long-distance RNA:RNA interaction between the IRES and a region of the 3'UTR enhances IRES-mediated translation initiation (55). Additionally, the 3'UTR also engages in a long-distance RNA:RNA interaction with the apical loop of a 5' terminal hairpin, known as the S-region, to modulate viral replication. The 3'UTR does not participate in these two interactions simultaneously, suggesting they are mutually exclusive events in the viral life cycle (55).

The highly structured HCV IRES directly recruits the preinitiation complex to the initiation codon via its highly conserved domains II, IIId and IV (56, 57). Recent reports suggest that the HCV 3'UTR can stimulate IRES-dependent translation in cell culture (58, 59). The HCV 3'UTR interacts directly with both the 40S ribosomal subunit and eIF3, and may stimulate translation by transferring bound components to the 5' end via a long-distance interaction that occurs between the apical loop of helix IIId and a bulge loop of hairpin, 5BSL3.2, which is located in the 3'-proximal coding region (60). Interestingly, the same bulge sequence of 5BSL3.2 also participates in a short-range RNA:RNA interaction with an apical loop sequence of a nearby sequence (61). The terminal loop of 5BSL3.2 can base-pair with a 3′ structural element, SL2, located in the 3′ UTR, which may modulate genome replication (62). The interactions of 5BSL3.2 within the 3' end and across the genome likely coordinate a complex mechanism used by HCV to progress through its life cycle (63).

IRESs in plant viruses

Some plant viruses, including TEV have a poly-A tail but no 5' cap. The 5' end of

the genome is covalently linked to a small protein (VPg). While VPg binds to initiation factors eIF4E and eIFiso4E, it is not known if this interaction influences translation. Instead, an IRES in the 5'UTR can recruit 40S ribosomal subunits possibly by direct base pairing with the 18S rRNA. This interaction has been shown to be important for translational enhancement *in vivo* (64, 65). Interestingly, this IRES can recruit eIF4G directly or indirectly (66), and the proposed closed loop structure is maintained by an interaction between eIF4G and PABP bound at the 3' end. The 5'UTRs of RNA1 and RNA2 of BRV also have IRES activity. These IRESs can initiate translation internally due to proposed complementarity with the 18S rRNA (67, 68). Additionally, 5' and 3' translation elements of both of these RNAs can participate in long-distance RNA:RNA interactions (68, 69). In PFBV, an internal IRES is proposed to direct translation of the CP from the gRNA (70). This early expression of CP counteracts the host RNA silencing defense mechanism to establish an infection. The majority of the CP is produced from sgRNA2 during the later stages of infection.

CITE-mediated translation usually requires a long-distance RNA:RNA interaction

Many (+)-strand plant viruses harbor CITEs within or near the 3'UTR of their genomes to confer translation efficiently from the 5' end (5, 71, 72). These *cis*-acting CITEs are predominantly found in plant viruses belonging to the *Tombusviridae* and *Luteoviridae* families. To date, seven different types of CITEs have been identified in 6 genera of plant viruses (5) (Figure 1). Although 3' CITEs have diverse secondary structures, most serve a similar purpose of recruiting translation initiation factors or ribosomes in a structure and/or sequence-dependent manner and by communicating with the 5' end via long-distance kissing-loop interactions involving hairpins in the 5' UTR or

nearby coding region. This 5'-3' interaction facilitates the delivery of either 3' bound initiation factors or ribosomes to the 5' end, which results in ribosome entry at or near the 5' end of the genome (5, 71, 72), followed by scanning in the 3' direction to the initiation codon likely for the pioneer or subsequent round of translation. (5, 71-74).

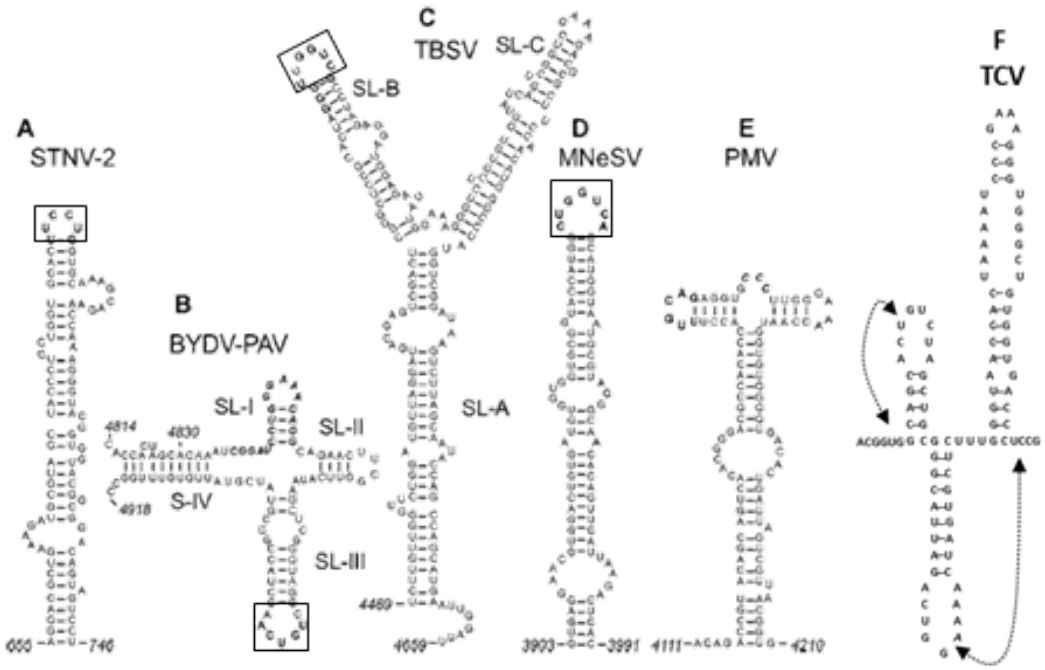


Figure 1: Sequences and secondary structures of 3' CITEs in six genera of plant viruses

(A) STNV TED, known as STNV-2. (B) BYDV BTE. (C) TBSV Y-shaped CITE. (D) I-shaped CITE of MNSV. (E) PTE of PMV. (F) T-shaped structure of TCV. Boxed bases are known or predicted to form long-distance RNA:RNA interaction with complementary sequence located at the 5' end of their genomes. Adapted with minor modification from (72).

The first CITE was discovered in the STNV 3'UTR, and is known as TED (translation enhancer domain) (Figure 1A). The TED of STNV, a long stem loop structure with multiple bulges, preferentially binds the eIF4F complex (76). Mutagenesis analyses have shown that a long-distance interaction between the apical loop of TED and

the apical loop of a 5’ terminal hairpin is not necessary for the TED function (77). It is possible that artificial constructs that were used for the *in vivo* study do not contain additional coding sequences of STNV necessary to support the function of the kissing-loop interaction. It has been proposed that a protein-mediated bridge of 5’UTR-43S-eIF2-eIF3-eIF4F-TED complex may form resulting in repositioning of the TED element to the 5’ end (75). The STNV CITE confers cap-independent translation *in vitro* when moved to the 5’UTR or even when located within a coding region (75) of an uncapped reporter mRNA, suggesting that the TED-mediated cap-independent translation is position independent.

The BTE-type (Barley yellow dwarf translational element) CITE of BYDV is one of the best characterized 3’ CITEs and can be found in all members of the luteovirus, dianthovirus, necrovirus and umbraviruses genera (5, 76) (Figure 1B). While BTEs vary in length, they share common features such as a multi-helix junction and a highly conserved 17 nt long sequence (77). The BTE preferentially binds to the eIF4G subunit of the eIF4F complex and facilitates cap-independent translation initiation by participating in a long-distance RNA:RNA interaction that takes place between the apical loop of a stable stem-loop (SL-III) of the BTE and a complementary loop (BCL) in the 5’UTR of the gRNA (71, 77, 78). While mutating a single base in either of the two interacting sequences reduces translation efficiency, a compensatory mutation designed to restore the 5’-3’ interaction restores the translational efficiency (71). The BTE can also direct translation of sgRNA1 of BYDV, which codes for CP, by participating in a long-distance kissing-loop interaction with a BCL located in the 5’ UTR of sgRNA1(79). BTE-mediated translation efficiency of an artificial reporter construct was not affected

when the BYDV 5'-3' RNA:RNA interacting sequences was substituted with DENV sequences that promote base-pairing between the 5' and 3' ends (80). This suggests that the BTE is functional as long as communication between the 5' and 3' ends is maintained. When the BTE was moved to the 5'UTR, the 5' interacting sequence was no longer required for cap-independent translation, indicating that a general role for the long-distance base pairing is to position the 3' CITE-bound eIF4G proximity to the 5' end (81). An upstream out-of-frame AUG or 5' extension of the 5'UTR, which increase the distance between the BCL and the 5' end, reduces translation efficiency, suggesting that the BTE relies on 5' end-dependent ribosome scanning and prefers that the BCL is located at the 5' end (80).

The Y-shaped 3'CITE (YSS) of TBSV and CIRV is conserved in all but two tombusviruses (Figure 1C). This type of CITE is characterized by a large Y-shaped branched structure consisting of three major helices. Compensatory mutational analyses have shown that a long-distance RNA:RNA interaction between the loop and stem of SL-B of the YSS and a T-shaped replication element in the 5'UTR is essential for translation enhancement (82, 83). The CIRV YSS interacts with eIF4F and its isoform. This CITE can be functionally replaced with heterologous CITEs such as the I-shaped and PTE-type 3' CITEs without any significant loss in translation or replication efficiency, suggesting that these 3' CITEs are interchangeable as long as they can maintain long-distance RNA:RNA interaction with the 5' end (84). Similar to the BTE, upstream out-of-frame AUGs reduce YSS-mediated translational efficiency, suggesting that translation initiation is dependent on 5' end ribosome scanning. However, unlike the BYDV BTE, initiation is not dependent on the position of the YSS-interacting partner at the 5' end (85).

The smallest CITEs of all, the ISS of MNeSV and MNSV (tombusvirus and carmovirus, respectively) has been shown to preferentially interact with the eIF4E subunit of eIF4F (Figure 1D). A model has been proposed suggesting that interaction between the 3' CITE and 5'UTR modulates delivery of the 3' bound initiation factors to the 5' end of the genome subsequently facilitating recruitment of the 43S preinitiation complex (73). It has been shown that the 5'UTR/3' I-shaped 3'CITE of MNSV and eIF4F form a complex *in vitro*, in support of this model. In addition, ribosome toe printing has demonstrated that while bound to eIF4F, the I-shaped CITE can simultaneously base pair with the 5' UTR and recruit ribosomes to the 5' end of the viral fragment (73). Direct genetic evidence for the interaction between the I-shaped CITE of MNSV and eIF4E has been shown in melon. A single amino acid change in eIF4E renders the I-shaped CITE less effective as a translational enhancer and makes melon resistant to MNSV infection. Interestingly, this resistance is overcome by compensatory mutations in the MNSV I-shaped CITE that correspondingly restores translational activity of the element (86, 87).

A unique CITE in TCV, forms an internal T-shaped structure (TSS) consisting of three hairpins and two pseudoknots (Figure 1F). The TCV TSS interacts with the 60S subunit or 80S ribosome *in vitro* independent of any translation factors (88) . For this CITE, no base pairing between the TSS and 5'UTR has been found. The 5'-3' synergistic enhancement of translation has been proposed to occur by bridging between 40S, which binds to the 5'UTR, and 60S, which binds the TSS (89). Within the known 16 carmoviruses, only CCFV has a 3'UTR region with the capacity to form a TCV-like TSS.

In seven other carmoviruses, a phylogenetically conserved PTE-like CITE was identified (90).

The PTE class of 3'CITEs was originally identified in the panicovirus PMV (91) and later in the umbravirus PEMV (92) (Figure 1E). Characteristics of the PTE structural components include a lower supporting stem followed by a bulged helix (G-rich bulge) and a three-way branched structure with two hairpins (Figure 1E). The PEMV PTE functions as a translational enhancer by recruiting eIF4E. A high-resolution structural model suggests that a pseudoknot that forms between a G-rich bulge and C-rich sequence at the three-helix junction of the PTE is critical for eIF4E recruitment to the PTE (92, 93). Unlike most other CITEs, the PEMV PTE does not participate in a long range RNA:RNA interaction with the 5' end. Instead, an adjacent upstream element, the kl-TSS, engages in a long-distance RNA:RNA interaction with a 5' proximal hairpin located in the p33 ORF thereby circularizing the viral genome (94). Interestingly, the same kl-TSS can also bind 60S ribosomal subunits. Binding to the ribosomal subunit and participation in the long-distance RNA:RNA interaction can occur simultaneously, suggesting that these two events are compatible (95). It is possible that the kl-TSS-mediated long-range RNA:RNA interaction is beneficial for PTE's activity, as PTE bound eIF4E can be positioned proximal to the 5' end via this interaction (95). Alternatively the eIF4E bound to PTE may help to recruit ribosomes to the kl-TSS.

Long-distance interactions that are required for ribosomal recoding

The genomes of (+)-strand RNA viruses often utilize ribosomal recoding via stop codon readthrough or frameshifting as strategies to translate C-terminal extended protein isoforms (96). The ribosomal frameshift occurs when ribosomes slip by one base in either

the +1 or -1 directions relative to codons in the zero reading frame. A heptanucleotide slippery sequence at the site of the frameshift and an RNA structural element, usually an H-type pseudoknot, positioned a few nucleotides downstream from the frameshifting site promote this process (96). In addition to these elements, BYDV requires a long-distance RNA:RNA interaction between an internal bulge of a stem loop adjacent to the frameshift site and the terminal loop (or nearby sequence) in a 3' terminal stem loop to facilitate -1 frameshifting. This frameshifting event produces the viral RdRp (97). Similar to BYDV, a small stable stem-loop structure in the 3′ UTR of RCNMV RNA1 is required for efficient − 1 frameshifting. The apical loop of this stem-loop structure promotes frameshifting via base-pairing with the bulge of an element located adjacent to the slippery site (98).

During stop-codon readthrough, ribosomes recode stop codons as sense codons and continue on with translation. This gives rise to a protein isoform with an extended C-terminus. Similar to ribosomal frameshifting, the efficiency of readthrough is regulated by RNA structural elements as well as sequences adjacent to the stop codon (99). In CIRV, stop codon readthrough that generates the viral RdRp requires a long-distance RNA:RNA interaction between an RNA bulge, known as the proximal readthrough element (PRTE), and a sequence in the 3' UTR known as the 3' distal readthrough element (DRTE). Interstingly, the DRTE is part of a RNA molecular switch coordianting translation and replication of the virus (29). The long-distance RNA:RNA interaction facilitating readthrough that expresses the viral RdRp is predicted for all genera of the *Tombusviridae* including the genus carmovirus. In TCV, the loop of a 3' terminal hairpin, known as the Pr, is involved in a similar long-distance interaction with a bulge

loop of a readthrough stimulatory element (RSE) located adjascent to the stop codon. This interaction facilitates readthough of the 5’ proximal ORF and generates the viral RdRp (29).

Long-distance RNA:RNA interactions modulate viral replication

The replication of (+)-strand RNA viral genomes occurs via synthesis of a complementary (-)-strand RNA, which in turn is used as a template for the production of progeny (+)-strand RNA genomes. The viral RdRp must initiates (-)-strand synthesis at the 3’ end of the viral genome (10). Multiple reports suggest that RNA sequences and structures that facilitate (-)-strand synthesis are not just limited to the 3’ end of the viral genome but can be located in both 5’ and 3’ ends and as well as in the viral coding region (10). Additionally, recent reports suggest that circularization of the viral genome via interaction between the 5′ and the 3′ end is necessary to initiate (-)-strand RNA synthesis (100, 101). Several members of the genus flavivirus, including DENV, WNV and YFV have complementary sequences at the 5’ and 3’ ends of their viral genomes and utilize genome circularization for replication (100). Three sets of long-distance RNA:RNA interactions between 5’-3’ upstream AUG region (UAR), 5’-3’ downstream AUG region (DAR) and 5’-3’ cyclization sequence in the DENV genome have been proposed to circularize the viral RNA genome and facilitate initiation of the (-)-strand synthesis (102, 103). RdRp binding assays have shown that RdRp is recruited by a 5’ terminal hairpin, SLA, and it has been proposed that genome circularization repositions the bound 5’ end RdRp to the 3’ end where it initiates (-)-strand synthesis (104).

Similar to the Flavivirus replication scheme, initiation of (-)-strand synthesis in TBSV is facilitated by an internal replication element known as RII that is located almost

3 kb upstream from the 3’ end. This internal replication element recruits viral RdRp to the viral genome (105). A long-distance RNA:RNA interaction between the upstream and downstream linker sequences within the TBSV genome has been proposed to facilitate transfer of the RII-recruited RdRp close to the 3’ end of the genome (28, 106). RII-like internal replication elements are well conserved across the genera in the *Tombusviridae* family. It is likely that other members might utilize similar mechanism for genome replication (28, 107).

Long-distance interactions required for sgRNA transcription

Many (+)-strand RNA viruses produce 3’ coterminal sgRNAs to enhance expression of 3’ proximal ORFs. Three different mechanisms have been postulated for the production of sgRNAs: internal initiation on a (-)-strand RNA template, which requires an internal subgenomic promoter; premature termination, which takes advantage of the prematurely terminated (-)-strand that is used as a template for sgRNA production; and discontinuous RNA synthesis, which utilizes the discontinued (-)-strand RNA as a template for sgRNA production (19). Some viruses that use premature termination or discontinuous RNA synthesis require long-distance RNA:RNA interactions to generate sgRNAs.

TBSV produces two sgRNAs, sgRNA1 and sgRNA2, via a premature termination mechanism during viral (-)-strand synthesis (108, 109). It has been proposed that premature termination of sgRNA synthesis is facilitated by forming a higher-order structure via long-distance RNA:RNA interactions between an activator and receptor sequence (AS-RS) located upstream of each sgRNA start site (109, 110). This higher-order structure causes the polymerase to stop and terminate synthesis of the (-)-strand

prematurely. The plant virus RCNMV, with its bisegmented genome, uses premature termination to transcribe sgRNA from RNA1. An intermolecular RNA:RNA interaction between RNA1 and RNA2 has been proposed to play a role in prematurely terminating (-)-strand synthesis from the (+)-strand template of RNA1 (111).

Using the TCV system to study medium- and short- range RNA: RNA interactions

TCV, a member of the genus carmovirus, has been used extensively as a model for studying diverse roles and interactions of *cis*-acting RNA structural elements during viral translation and replication. The 4054 nt TCV gRNA has a 63 nt 5'UTR followed by 5 overlapping ORFs (Figure 2A) (112, 113). The 5' proximal ORF encodes replicase-associated protein p28 and ribosomal-readthrough product p88, the viral RdRp (114). p8 and p9 movement proteins are translated from the 1.72 kb sgRNA, and p38, the TCV coat protein is translated from the 1.45 kb sgRNA (113). p38 is a multifunctional protein, forming virions, serving as an effector of virus resistance (115), and mediating suppression of virus accumulation by the virulent satellite RNA satC (116). p38 also functions as a silencing suppressor (117) and targets multiple components in the host RNA silencing pathway. It binds to double stranded RNA (dsRNA) in a size-independent manner and likely inhibits the processing of dsRNA to viral small (vs) RNAs (118, 119). In addition, it can directly interact with AGO1 and AGO2, which are the essential catalytic components of the RISC, and are major effectors of antiviral defense. Interaction of p38 with AGO1 is proposed to inhibit loading of host miRNAs and vsRNAs into RISC, which affects homeostasis of DCLs in *Arabidopsis thaliana,* via downregulation of DCL3 and DCL4 (120-123).

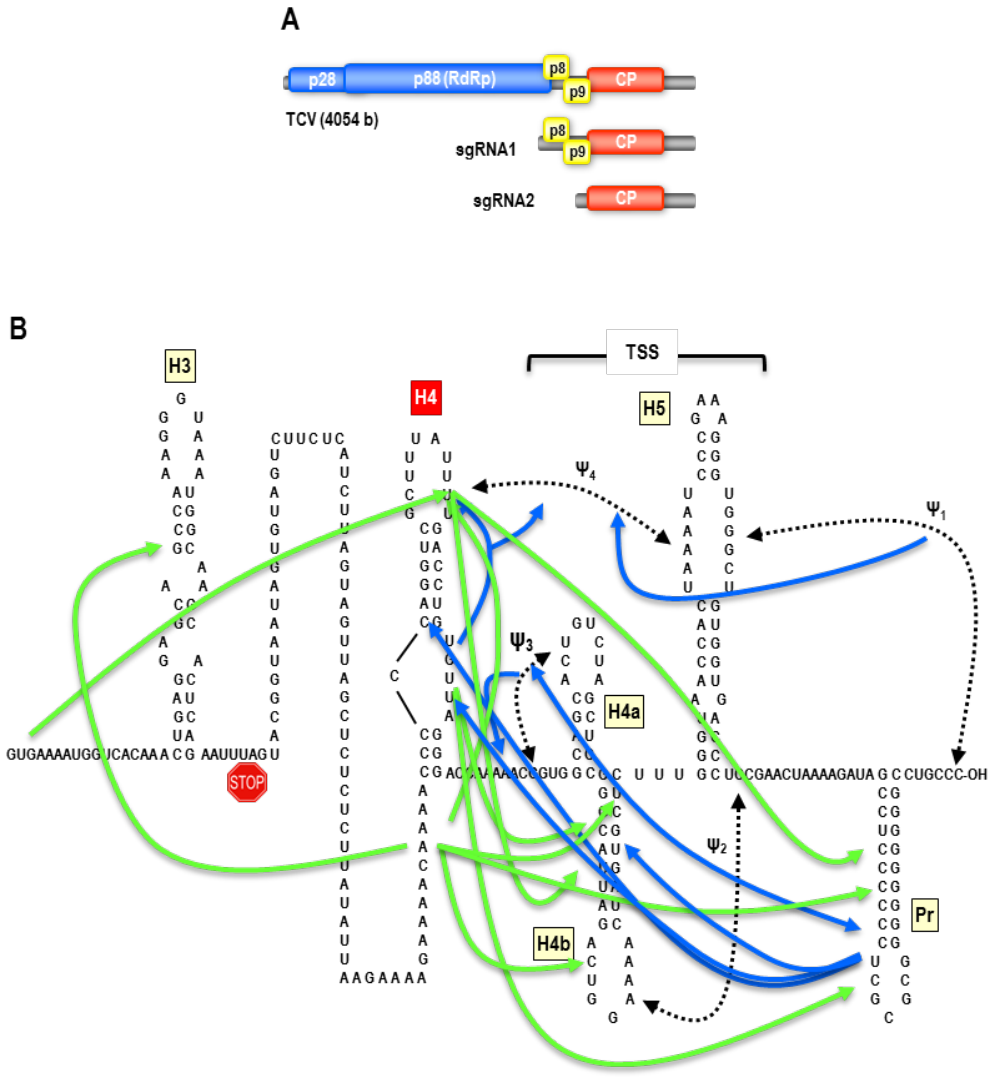


Figure 2: Complex interactions within the 3' region of TCV
(A) Genome organization of TCV. The gRNA and two sgRNAs are depicted.
(B) Canonical and non-canonical interactions near the 3' end of TCV. Interactions generated by examining structural changes in response to specific mutations are shown in blue lines. Interactions generated by the location of second-site mutations are shown in green. Location of the TSS 3'CITE is shown. Hairpin names are boxed.

TCV is associated with a small satellite RNA, satC, which is only 356 bases long and shares its 3' 166 bases with 3' sequences of TCV (124). Since satC is a noncoding RNA, it serves as a model for identifying homologous sequences and/or structures in TCV necessary for replication. Biochemical structure probing combined with genetic,

phylogenetic, and computational analyses have identified several key structural elements in the 3' terminal region of TCV, which includes the 3'UTR and part of the 3' proximal p38 ORF.

3' terminal tail, Ψ_1 and Pr hairpin

With the exception of GaMV, the 3' UTRs of all carmoviruses, including TCV, contain a very stable terminal hairpin at the 3' terminus, known as Pr. Pr and its adjacent 3' terminal tail have been identified as the core promoter for (-)-strand synthesis based on the analysis of a similar hairpin in satC. It was found that this core promoter of satC was able to direct transcription of an RNA template unrelated to TCV in the presence of purified RdRp *in vitro* (125). Mutational analyses of satC Pr showed that size and sequence of the Pr loop region could be flexible but stability of the stem is important for satC accumulation *in vitro* and *in vivo* (125, 126). The 3' terminal tail sequence of satC is 6 nt long and contains three terminal cytidylates residues. Deletion of the three terminal cytidylates resulted in an increased *in vitro* transcription of full-length satC and many aberrant products (127, 128). Solution structure probing of satC RNA containing this three base deletions supported genetic data that three consecutive guanylates in the large symmetrical loop (LSL) of H5 pairs with the three terminal cytidylates. Single point mutations that disrupt this potential interaction resulted in increased *in vitro* transcription while the compensatory mutation that restores this interaction resulted in wt levels of transcripts (127). These results indicated that the higher order structure that forms between the H5 LSL and the 3' terminus (pseudoknot Ψ_1) represses the levels of (-)-strand synthesis as well as facilitate 3' end recognition by the RdRp (127). Unlike the Pr hairpin of satC, TCV Pr cannot promote transcription of the (-)-strand without

additional upstream viral sequences *in vitro* (129, 130). The TCV Pr loop modulates ribosome recoding and RdRp synthesis by participating in the long-distance RNA: RNA interaction with the bulge loop of RSE. This is the only known long-distance RNA:RNA interaction in the TCV genome (29). Mutations in the TCV Pr loop reduced *in vitro* transcription of a 3' truncated fragment in the presence of purified recombinant RdRp and radioactively labelled nucleotides *in vitro. In vitro* electrophoretic mobility shift assays showed that certain Pr loop mutants have reduced affinity for RdRp binding (131). This suggests that the TCV Pr is also important for TCV replication. The Pr loop mutations enhanced translation of a luciferase reporter construct containing the 5'UTR and 3'400 nt of TCV, suggesting that Pr loop sequence has a positive effect on translation initiation (131).

H4a, H4b and H5 and two pseudoknots Ψ_2 and Ψ_3

mFold-predicted structure of the full-length satC gRNA revealed that the 3' terminal 140 bases of satC contains a series of hairpins (H5, H4b and H4a) in addition to Pr. Solution structure probing confirmed the existence of these hairpins in satC (132). Immediately upstream of Pr is phylogenetically conserved hairpin H5. H5 of TCV contains an upper stem (US), a lower stem (LS), a large symmetrical loop (LSL) and a stable apical loop. This H5 is similar to the SL3 hairpin of TBSV and is proposed to function as a RdRp chaperone in TCV (88, 133).

The MPGAfold folding program predicted two juxtaposed hairpins (H4a and H4b) located upstream of H5. The terminal loop of H4b forms a pseudoknot (Ψ_2) with a sequence just downstream of H5. This pseudoknot is phylogenetically conserved in four additional carmoviruses (88, 130) and required for efficient replication of TCV and satC

(134, 135). In TCV, the apical loop of H4a participates in an H-type pseudoknot (Ψ_3) with adjacent upstream sequence (88, 134). For satC the formation of this pseudoknot (Ψ_3) is important but the exact role is unknown (130).

The TSS, which functions as a 3'CITE, consists of Ψ_3, Ψ_2, H4a, H4b, and H5 (88, 89). Ψ_3 is an important part of the TSS as disruption of this pseudoknot resulted in reduced translation of a reporter construct *in vivo* (89). An A-rich linker region just upstream of Ψ_3 in TCV has been found to be critical for virus accumulation. Mutations in the A-rich region in TCV have been shown to negatively affect both transcription *in vitro* and translation of luciferase reporter constructs *in vivo* (13). Solution structure probing suggests that this A-rich region interacts with Ψ3 and stabilize it (13, 130). This is supported by single molecule force spectroscopy studies (M.T.Li and A.E.S unpublished).

A second-site mutational approach along with in-line structure probing analysis has revealed that the TCV 3' terminal elements participate in a complex network of short range, non-canonical RNA:RNA interactions. In addition to canonical interactions, these short range interactions support an integrated higher order structure of the TCV 3' end (136). In-line structure probing analysis showed that disruption of one element may alter the structure of the other if they are connected via RNA:RNA interactions. For example, a mutation in Ψ3 affected the structure of the adjacent adenylates and vice versa (13). Mutations that disrupted Ψ3 also altered the cleavage pattern in the Pr loop sequence, suggesting that the Pr loop sequence is likely associated with the Ψ3/H4a region of the TSS (13). Structure probing data suggested that the H5 LSL residues can participate in short-range RNA:RNA interactions simultaneously with the TCV 3' terminus sequence

as well as with a sequence in H4TL via Ψ1 and Ψ4 respectively (13). A single point mutation in the TCV 3' terminus sequence that disrupted Ψ1 affected the structure of non-Watson-Crick interactions across the H5 LSL and altered the structure of H4TL and H4AL sequences (13). Among the 3' elements, H4 occupies a central position in this network of structural interactions. H4 is located upstream of the TSS and is critical for viral accumulation through a yet unknown role in both replication and translation (Figure 2B) (13). In-line structure probing of a transcript containing a point mutation in H4AL altered the structure of the H4TL sequence and reduced the stability of Ψ4 (13). A three-base alteration in the H4TL sequence also affected the structure of H4AL, suggesting that non-canonical interactions connect these two critical H4 loops (13). Addition of purified, recombinant RdRp caused a widespread conformational change affecting nearly all elements throughout the 3' region, including the structure of H4AL, supporting evidence for an integrative, higher order structure in the region (13).

The higher order structure of the 3' end of the TCV genome was further supported by a genetic screening method. This method screens for second-site changes that compensate for detrimental primary-site mutations, and thus identifies putative interacting regions in RNA viruses. Second site mutations that arose in response to the primary site mutation in H4AL were found in the Pr stem and loop sequences (131). In-line structure probing showed that changes in the Pr loop sequence structurally altered a base in the H4 upper stem and three bases in the stem of H4b, suggesting that Pr loop and H4AL participate in a complex network of RNA:RNA interactions with each other and with H4b (13, 137). Specific mutations in H4TL and upstream A-rich region significantly reduced the flexibility of Pr loop residue in the presence of viral RdRp (136). Second-site

mutations that arose in response to primary-site mutations in H4TL and A-rich region were found in the Pr stem region, H4b stem and in the upper stem region of H3 (136). A number of many of these second-site changes in the TCV 3' end were able to compensate for primary-site mutations (136). Many second-site changes that arose in response to several different 3' UTR primary-site mutations were scattered throughout the 3'end region, although many were found in specific cluster, and most of them compensated for structural changes and restored virus accumulation. This suggests that the primary mutations were not disrupting a specific interaction, but rather were affecting a more general RNA conformation that was critical for virus viability (136). Second-site mutations in response to several different 3'UTR primary-site mutations were also located within the p38 ORF, but whether or not these distal changes were compensatory had not been investigated.

Thesis plan

Circularization of cellular and viral mRNAs is an important event prior to translation initiation. The PTE of PEMV is a weak 3' CITE as it does not engage in a required long-distance RNA:RNA interaction (92). In chapter I, I report that the PTE of SCV, a model carmovirus, confers cap-independent translation by participating in long-distance RNA:RNA interactions with hairpins located in the p26 ORF and in the 5'UTR of the small sgRNA and this interaction correlates with translation efficiency. Similar PTE in six additional carmoviruses have the potential to participate in long-distance RNA:RNA interactions with 5' hairpins located in comparable positions. In chapter II, I examine the hypothesis that conserved positioning of PTE-interacting hairpins is important for efficient PTE function as a translation enhancer.

Overall the 3’ end of TCV maintains a higher order structure with many canonical and non-canonical short- and medium-range interactions both within the 3’ end and upstream in the p38 ORF. In chapter III, I examine several p38 ORF second-site mutations that arose in response to primary mutations in H4AL and report the unexpected connection between 3’UTR primary-site mutations in TCV and the host RNA silencing machinery.

Chapter 2: Long-distance kissing-loop interactions between a 3' PTE and apical loops of 5' hairpins enhance translation of SCV[1]

Introduction

Most research on translation of cellular and viral RNAs concentrates on the 5'UTR, where assembled pre-initiation complexes composed of 40S ribosomal subunits, met-tRNA and translation factors including multisubunit eIF4F recognize the 5' cap, and then scan in the 3' direction for the first initiation codon in a good context (40, 41). Many animal (+)-strand RNA viruses lacking 5' caps contain IRESs in their 5'UTRs to confer cap-independent translation. Plant viruses have short 5'UTRs (10 to 200 nt) and the elements that enhance cap-independent translation have been found mainly in the 3'UTRs (5, 138). Most 3' CITEs recruit translation initiation factors or ribosomes in a structure and/or sequence dependent manner and engage in a long-distance kissing-loop interaction with a hairpin near the 5' end (5, 72). Delivery of either 3' bound ribosomes or initiation factors to the 5' end via a long-distance RNA:RNA interaction results in ribosome entry at or near the 5' end of the genome, followed by scanning in the 5' direction (71, 81, 94, 95).

In the carmovirus genus, TCV and CCFV have combinations of hairpins and pseudoknots that are required to adopt a TSS-like structure (88). The TCV TSS functions as a 3' CITE by binding to the P-site of 60S ribosomal subunits. (88, 89). For the TCV TSS, no complementary sequence near the 5' end of the genome has been found. The 5'-

[1]Modified from: Long-Distance Kissing Loop Interactions between a 3' Proximal Y-Shaped Structure and Apical Loops of 5' Hairpins Enhance Translation of Saguaro Cactus Virus. Maitreyi Chattopadhyay, Kerong Shi, Xuefeng Yuan, and Anne E. Simon. Virology. 2011 Aug 15; 417(1): 113–125.

3’ synergistic enhancement of translation has been proposed to occur by bridging between 40S, which binds to the 5’UTR, and 60S, which binds the TSS (89).

In contrast to TCV and CCFV, a phylogenetic search for conserved secondary structures in non-TSS carmoviruses found that 7 out 16 contain a PTE-type 3’ CITE that is similar to a translational enhancer described for PEMV (92) and PMV (91). Among the 7-PTE containing carmoviruses, an infectious clone was available only for SCV. Therefore, using SCV as a model carmovirus, in this chapter, I report that unlike the PEMV PTE, the SCV PTE engages in long-distance RNA:RNA interactions with 5’-proximal sequences in the gRNA and sgRNA2, and promotes cap-independent translation of reporter mRNAs in plant protoplasts.

Materials and methods

Generation of constructs

Single-luciferase reporter construct T7-Fluc was the progenitor for the translation constructs. The 5’ end of this construct contains a T7 promoter, followed by *Bam*HI and *Sac*I restriction sites into which 5’ gRNA or sgRNA fragments, generated by PCR, were inserted. SCV 3’ fragments were introduced downstream of the Fluc coding region using *Pml*I and *Ssp*I restriction sites. Oligonucleotide-mediated site-directed mutagenesis was used to generate point mutations. All constructs were verified by sequencing.

Small-scale plasmid DNA isolation

Bacterial cells (DH5α) from 6 ml overnight culture were collected in an Eppendorf tube and subjected to centrifugation at 13,000rpm for 20 sec. The supernatant was discarded and pellet was re-suspended in 1 volume (200 µl) of solution I [0.9%

glucose, 25 mM Tris (pH 8), 10 mM EDTA (pH 8)]. Following addition of 1 volume of solution II (1% SDS, 0.1 M NAOH) the tube was gently inverted several times and incubated at room temperature for 5 min. Upon addition of solution III (3 M KOAc, pH 4.8) the tube was vigorously mixed and centrifuged at 13,000 rpm for 5 min at room temperature. The supernatant was collected into a new tube. An equal volume of phenol/chloroform was added and the tube was mixed vigorously, followed by incubation at room temperature for 5 min. After centrifugation at 13,000 rpm for 10 min, the aqueous layer was precipitated by adding 1/10 volume of 3M NaOAc (pH 5.3) and 2.5 volumes of 100% ethanol at -80^{o}C for 30 min. After centrifugation at 13,000 rpm for 30 min at 4^{o}C the supernatant was discarded and the pellet was washed with 70% ethanol, air dried, and resuspended in 100 µl of autoclaved distilled water.

In vitro RNA synthesis using T7 polymerase

Eight micrograms of luciferase reporter constructs linearized with *Ssp*I were mixed with 6 µl of 100 mM DTT, 12 µl of ribonucleotide mix (5 mM of each nucleotide), 12 µl of T7 RNA polymerase buffer (125 mM NaCl, 40 mM $MgCl_2$, 10 mM spermidine, 200 mM Tris-HCL, pH 8.0), 0.5 µl of RNAase Out ribonuclease inhibitor (20 U/µl: New England Biolab), 3 µl of T7 RNA polymerase, and water to a final volume of 60 µl. The reaction mixture was incubated at 37^{o}C for an hour. The reaction was phenol extracted using phenol/chloroform in 1:1 volume, vortexed vigorously and then subjected to centrifugation at 13,000 rpm for 10 min at 4^{o}C. The aqueous layer was collected in a fresh tube without disrupting the interphase layer and was ethanol precipitated in 1/10 volume of 3M NaOAc (pH 5.3) and 100% of ethanol at -80^{o}C for 30 min. After centrifugation at 13,000 rpm for 30 min at 4^{o}C, the supernatant was decanted and the

pellet washed with 70% ethanol, air dried and resuspended in 100 µl of sterile distilled water.

Protoplast preparation, inoculations and measurement of luciferase activity

Protoplasts were generated from callus cultures of *Arabidopsis thaliana* ecotype Col-0. The calli were generated from sterilized seeds grown on MS agar plates containing Murashige-Skoog salts, 4.3 g/L, 3% sucrose, 1X vitamins/glycine, 0.5 mg/L kinetin, 0.5 mg/L 2,4 D, 1% agar, pH 5.8. 1X vitamins/glycine contains 1 mg/L nicotinic acid, 10 mg/L thiamine HCL, 1 mg/L pyridoxine, 0.1 mg/L myoinositol, and 4 mg/L glycine. The calli were incubated in a growth chamber and passaged every 21 days.

Protoplasts were prepared from calli beyond the 3rd passage. Calli were collected in a sterile Petri dish containing 10 ml cold 0.6 M mannitol and broken into pieces using a Pasteur pipet. Calli and mannitol were collected into a 50 ml tube and incubated at room temperature for 20 min with shaking. Calli were collected after centritfugation (Beckman GPR-type swinging bucket) at 2000 rpm for 5 min at room temperature. The collected calli were transferred into a sterile bottle containing 50 ml of freshly prepared protoplasts isolation medium (PIM), 0.5 g cellulase (12,000 U/g) and 0.1 g of pectinase (3150 U/g) (Calibiochem, La Jolla, CA) enzymes. PIM was made with 4.3 g/L MS salt, 0.1 M sucrose, 3 mM MES, 0.5 M Mannitol, 5 mM $CaCl_2$, 1 mg/L thiamine-HCl, 0.5 mg/L nicotinic acid and 0.1 g/L myo-inisitol, 0.2 mg/L 2,4-D, 0.2 mg/L kinetin and 25 mM KOH, pH 5.8. The bottle was then wrapped in aluminum foil and incubated at room temperature for 4 h with shaking at 100 rpm.

The turbid solution was filtered through a 53 µM nylon mesh (Small Parts), followed by centrifugation at 1000 rpm for 5 min at 4^{o}C. The supernatant was discarded

and the protoplasts were washed three times with cold 0.6 M mannitol to remove any residual PIM solution. After centrifugation, collected protoplasts were resuspended in 100 µl 0.6 M mannitol and counted using a microscope and hemicytometer. For each transfection, 7 x 10^6 cells were used. Protoplasts were transfected with 30 µg of uncapped luciferase reporter transcripts, 10 µg of uncapped transcripts containing a Renilla luciferase (Rluc) reporter as an internal control, 8 µl of 1 M $CaCl_2$ and 2.17 µl of 50% PEG (prepared in 50 mM Tris-HCl, pH 7.5) in a final volume of 430 µl. The mixture was incubated at 25^{o}C for 40 sec, followed by addition of 30 µl 0.6 M mannitol/ 1 mM $CaCl_2$ and incubation on ice for 15 min. The cells were collected by centrifugation at 1000 rpm for 5 min at 4^{o}C. After three washes with 20 ml of 0.6 M mannitol/1mM $CaCl_2$, protoplasts were resuspended in protoplasts culture medium (PCM, 4.3 g/L MS salt, 0.1 M sucrose, 3 mM MES, 0.4 M mannitol, 1 mg/L thiamine-HCl, 0.5 mg/L pyridioxine-HCl, 0.5 mg/L nicotinic acid and 0.1 g/L myoinositol, 0.2 mg/L kinetin and 25 mM KOH, pH 5.8) and incubated at 25^{o}C for 18 h in the light. For full-length virus protoplasts were kept in dark for 40 h.

Luciferase assays

Cells were collected after 18 h by centrifugation at 13,000 rpm for 30 sec. After removal of growth medium completely, the pallet was resuspended in 1X passive lysis buffer (Promega). The tubes were vigorously mixed and incubated at room temperature for 2 min, followed by centrifugation at 13,000 rpm for 5 min at 4^{o}C. Supernatant was collected in a new tube. Luciferase assay reagent (LAR) and stop and glow reagent (SGR) were prepared according to the manufacturer's instructions (Promega). The luminometer (Modulus™ Microplate Multimode Reader) was programmed for reading

96-well plates according to the protocol recommended by the manufacturer. Briefly, injector 1 was set to dispense 50 µl of LAR to read the Firefly luciferase activity and injector 2 was set to dispense 50 µl of SGB for reading the Renilla luciferase activity. For each measurement 15 µl of cell lysate was dispensed into each well of a 96-well plate. The plate was inserted into the luminometer and initiated to read. Firefly luciferase data were normalized to Renilla luciferase data and final luciferase activity was calculated by taking the average of three replicates.

Northern blotting using RNA gels

Five micrograms of total RNA extracted from protoplasts was mixed with an equal amount of 2X formamide loading buffer (prepared by mixing 800 µl of formamide and 200 µl of 10X formamide gel-loading buffer (50% (v/v) glycerol, 1 mM EDTA, pH 8.0, 0.25% bromophenol blue, 0.25% xylene cyanol). The mixture was heated at 65^{o}C for 5 min and snap cooled on ice followed by electrophoresed through a 1.2% non-denaturing agarose gel. The gel was then soaked in a 6% formaldehyde solution with gentle shaking for 1 h. The formaldehyde solution was decanted and the gel quickly washed with distilled water. The gel was soaked in a 10X SSC solution containing 0.15 M NaCl and 0.015 M sodium citrate for 30 min. RNA was transferred to a nitrocellulose membrane overnight by a capillary transfer method. The membrane was washed briefly with 10X SSC and then placed, upside down, on an UV light box (short wave) for crosslinking for 2 min. The membrane was dried at 80^{o}C for 5 min.

For detection of luciferase mRNAs, the membrane was pre-hybridized for at least 1 h at 42^{o}C in 30% (v/v) formamide pre-hybridization solution. Pre-hybridization solution contains 5X SSPE (20X SSPE [pH 7.4], 3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M

EDTA), 10X Denhardt's reagent (50X Denhardt's reagent contains 1% ficoll, 1% polyvinylpyrrolidone, 1% Bovine Serum Albumin), 0.2% SDS, 0.2 mg/ml freshly denatured salmon sperm DNA and 50% formamide.

During the pre-hybridization incubation, oligonucleotides were labeled at the 5' end using γ^{32}P-ATP and T4 polynucleotide kinase (NEB) according to the manufacturer's instructions. The kinase reaction was incubated for 30 min at 25^{o}C and was purified using Microspin G-25 columns (GE Health Care) following the manufacturer's instructions. The purified labelled oligonucleotides were then added to the pre-hybridization solution and incubated for 8 h at 42^{o}C. After hybridization, the prehybridization solution was decanted. The blot was washed in a high salt solution containing 6X SSPE and 0.1% SDS for 10 min, then washed in a low salt solution containing 0.1X SSPE and 0.1% SDS for 15 min at 42^{o}C. The membrane was checked for radioactive signal and covered with a plastic wrap and subjected to autoradiography.

In-line probing

RNA fragments subjected to in-line cleavage were generated by PCR. A T7 RNA polymerase promoter was incorporated at the 5' end of each fragment during PCR. In-vitro transcribed RNAs were purified by agarose gel electrophoresis, dephosphorylated using Antarctic phosphatase (NEB), labeled at the 5' end using γ^{32}P-ATP and T4 poly nucleotide kinase (NEB) and then re-purified through 5% polyacrylamide gels. Labeled fragments were denatured at 75^{o}C and slow cooled to room temperature. In-line cleavage reactions were performed at 25^{o}C by incubating 5 pmol of labeled RNAs in in-line cleavage buffer [50 mM Tris-HCl (pH 8.5), 20 mM $MgCl_2$] for 14 h. For competition assays, 10-fold molar excess of unlabelled fragments (50 pmol) were combined with the

labeled fragment (5 pmol), and the combined fragments were subjected to in-line cleavage as described above. RNA cleavage ladders were produced by incubating 5 pmol of 5' end-labeled RNA in a solution containing 1 µg yeast tRNA, 50 mM $NaHCO_3/Na_2CO_3$ (pH 9.2), and 1 mM EDTA for 5 min at 95°C. To generate RNase T1 digest ladders, 5 pmol of denatured 5' end-labeled RNAs were incubated in 1 ng yeast tRNA, 20 mM sodium citrate (pH 5.0), 1 mM EDTA, 7 M urea, and 1 U RNase T1 (Ambion) for 2 min at 25°C. All reactions were ethanol precipitated and resuspended in acrylamide gel loading buffer II (Ambion). Before electrophoresis, fragments were heated at 95°C for 2 min, and then snap cooled on ice. Electrophoresis was performed through 8 M urea, 8% or 10% denaturing polyacrylamide gels followed by autoradiography. At least two independent in-line probing assays were performed for each fragment.

Table 2: Constructs used in Chapter 2

Name	Description
Construct A	Single F-Luc flanked by 5'39 and 3'224
Construct B	Single F-Luc flanked by 5'39 and 3'400
Construct C	Single F-Luc flanked by 5'80 and 3' 224
Construct D	Single F-Luc flanked by 5'125 and 3'224
Construct E	Single F-Luc flanked by 5'125 and 3'400
Construct F	Single F-Luc flanked by 5'39 and 3' 224 nt random
Construct G	Single F-Luc flanked by 5'sgRNA2 137 and 3'224
Construct H	Single F-Luc flanked by 5'sgRNA2 137 and 3'400
Construct I	Single F-Luc flanked by 5'sgRNA2 137 and 3' 400 nt random
[3'343] 3537	Construct H with 3'343 bases
3591	Construct H with 3'289 bases
3650	Construct H with 3'230 bases
3704	Construct H with 3'176 bases
[3'127] 3753	Construct H with 3'127 bases
PTE H1 stem GC to GG	Construct H with G to G change at the 5' side
PTE H1 stem GC to CC	Construct H with G to C change at the 3' side
PTE H1 stem GC to CG	Construct H with GC to CG change
sgH1 apical loop GC to CG	Construct H with GC to CG change
PTE H1 apical loop GC to CG	Construct H with GC to CG change
CG +GC in construct H	A GC to CG in the apical loop of PTE H1+GC to CG in the apical loop of sgH1
gH3 apical loop C to G	Construct E with C to G change
gH3 apical loop	Construct E with GC to CG change

GC to CG	
PTE H1 apical loop G to C	Construct H with G to C change
CG+GC in construct E	A GC to CG in the apical loop of PTE H1+GC to CG in the apical loop of gH3

Table 3: Oligonucleotides used in Chapter 2

Name	Position	Sequence	Polarity
F1	3537-3558	CATGAC *CACGTG* CGAGTTCTGACTCAAAGTGCGG	+
F2	3591-3612	CATGAC *CACGTG* GTTCCATTGTTTCTACTACTGTA	+
F3	3650-3704	CATGAC *CACGTG* CTGTGAGGGGACCTACCCAC	+
F4	3704-3723	CATGAC *CACGTG* GGCGAGGTAGGGCAGAAGAG	+
Fluc-down-Ase1	3399-3417	GCAATAAACCAGCCAGCCG	-
SCV R	3879-3859	CATGAC *AATATT* GGGCGGGAGAGGTCCCTAAGG	-
SCV5'g-UTR-BamH1	1-23	CATG*GGATCC*GGGTAAGCTGCTGAATGCTT ATC	+
SCV5'g-UTR-Sac1	114-125	CATG *GAGCTC* ATGCGCCAATTGCAGGCCTG	-
C107G	95-122	CAGTGGGGCTGG<u>G</u>AGGCCTGCAATTG	+
C107G	95-122	CGCAATTGCAGGCCT<u>C</u>CCAGCCCCACTG	-
G3676C	3664-3685	ACCCACTGTGCT<u>C</u>CCACACAGG	+
G3676C	3664-3685	CCTGTGTGG<u>G</u>AGCACAGTGGG	-
G3671C	3654-3680	GAGGGGACCTACCCACT<u>C</u>TGCTGCCAC	+
G3671C	3654-3680	GTGGCAGCA<u>G</u>AGTGGGTAGGTCCCCTC	-
C3682G	3673-3692	GCTGCCACA<u>G</u>AGGAACTTCC	+
C3682G	3673-3692	GGAAGTTCCT<u>C</u>TGTGGCAGC	-
T7-5'gUTR	1-17	TAATACGACTCACTATAGGGTAAGCTGCTGAATGC	+
5'g-125-R	106-125	ATGCGCAATTGCAGGCCTG	-
GC107CG	95-122	CAGTGGGGCTG<u>GG</u>AGGCCTGCAATTG	+
GC3676CG	3664-3685	ACCCACTGTGCT<u>CG</u>CACACAGG	+
GC3676CG	3664-3685	CCTGTGTGCGAGCACAGTGGGT	-

Underlined letters indicate mutated bases; Italic letters are restriction enzyme sites;"+" and "-"indicate polarity of the oligonucleotides.

Results

SCV contains a 3' PTE that spans the CP termination codon

The genome organization and putative structure of the 3'UTR of SCV are shown in Figure 3. SCV, as with most carmoviruses, encodes five proteins that are translated from the gRNA and two sgRNAs. The viral replicase p86 is the ribosomal readthrough product of p26. sgRNA1 is a bi-cistronic mRNA for expression of p6 and p9, which are movement proteins. sgRNA2 codes for CP. The RNA structure in the SCV 3' region is based on computational predictions using mFold (139) as well as phylogenetic comparisons with other members of the carmovirus genus. The 3' terminal hairpin, designated "Pr" according to TCV terminology, is similar to 3' terminal hairpins found in 15 of 16 carmoviruses (131), and is the core promoter for transcription of complementary strands for the TCV satellite RNA, satC (125). The 3' penultimate hairpin, H5, also found in 15 of 16 carmoviruses (131), contains a large internal symmetrical loop with sequence identical to that of TCV. All carmoviruses that contain H5 are capable of forming a pseudoknot between the 3' side of their H5 large symmetrical loops and 3' terminal residues (140). Just upstream of H5 is hairpin H4b, and hairpins in similar locations are present in 14 of 16 carmoviruses (131). Unlike TCV, SCV does not likely contain sequence capable of forming the stable pseudoknot that connects the apical loop of TCV H4b to sequence adjacent to the 3' base of H5. Additionally, SCV does not contain a hairpin just upstream of H4b known as H4a (88). In TCV, H4a, H4b, the missing pseudoknot and an additional pseudoknot that links H4a to adjacent, upstream sequence, compose the TSS translational enhancer.

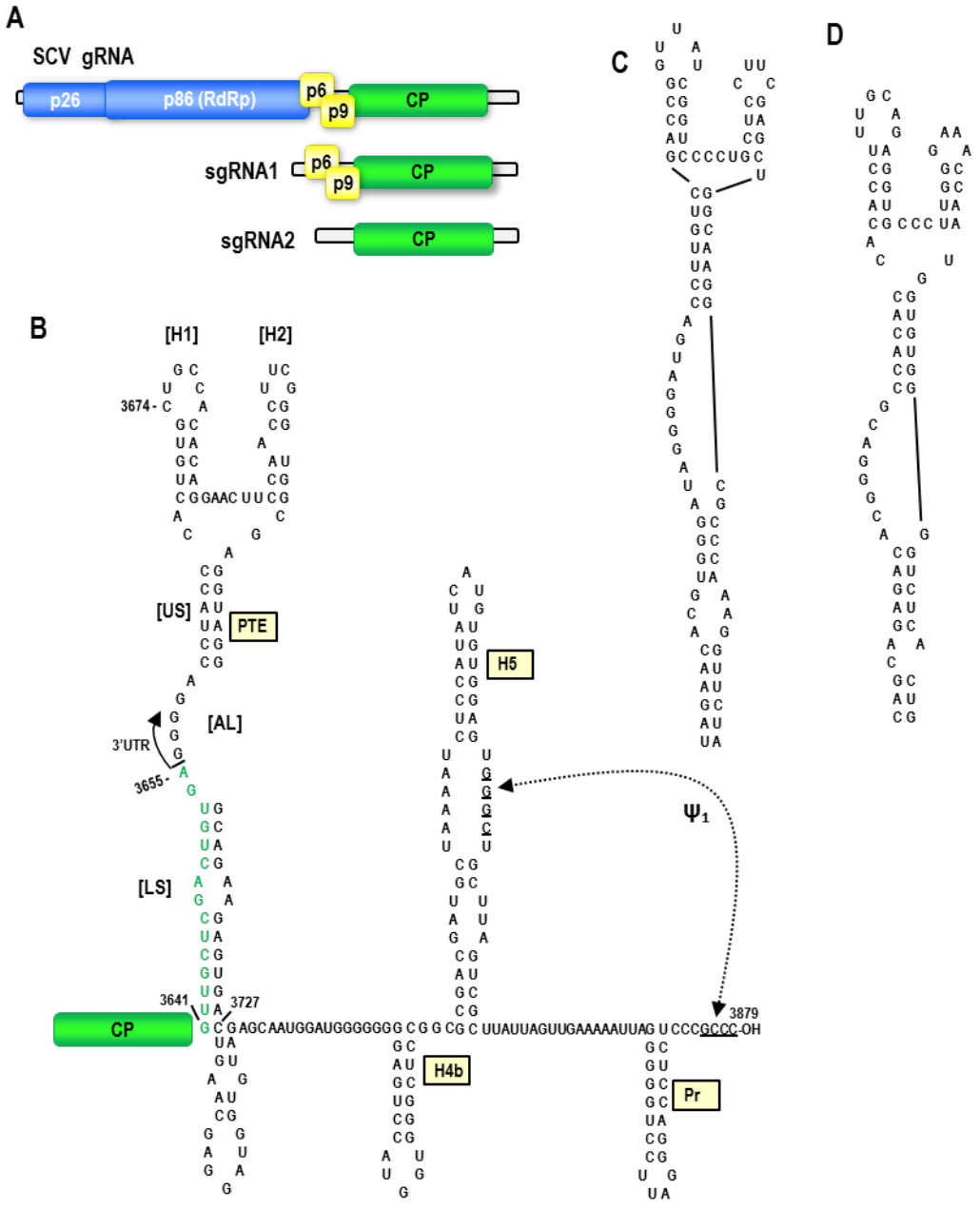


Figure 3: The SCV system

(A) Genome organization of SCV.

(B) Putative secondary structure in the 3'UTR of SCV. The PTE is partially within the CP ORF (green sequence). Pr, H5 and pseudoknot ψ_1 are conserved in carmoviruses. Most carmoviruses also have H4b. H5 and H4b in TCV are part of the TCV TSS. US, upper stem; AL, asymmetric loop; LS, lower stem.

(C) The similar PTE of PEMV, which is known to bind to translation initiation factor eIF4E.

(D) The PTE of PMV, previously determined to be a 3' translation enhancer.

The SCV, PEMV and PMV PTE consist of two hairpins (H1 and H2), a

supporting stem of 6 to 7 bp (US), a large asymmetric loop (AL) in the central region of the stem at the base of the hairpins that contains four consecutive guanylates flanked by adenylates, and a lower stem (LS) with a small symmetrical loop (Figure 3B-D). The PEMV and PMV PTE reside entirely within their 3'UTRs while the SCV PTE spans the termination codon of the CP ORF.

Efficient translation of SCV reporter constructs requires UTRs and coding sequences from the 5' and 3'UTRs

To determine if the SCV PTE is a translational enhancer and/or to identify other possible translation elements in SCV, we generated a single luciferase (FLuc) reporter construct containing either viral or random 5'- and 3'- sequences and assayed for luciferase activity in *Arabidopsis thaliana* protoplasts co-transfected with a control luciferase (RLuc) construct and luciferase levels assayed 18 h later. As shown in Figure 4A, construct A containing the precise SCV 3'UTR (includes only a partial PTE) and precise gRNA 5'UTR (39-nt) was 3.3-fold more efficient at translation compared with a construct containing the gRNA 5'UTR and a random 3' sequence (compare constructs A and F). Extension of the SCV 3' sequence to contain upstream coding sequence (400 nt total) that includes the full-sized PTE, did not significantly improve translation directed by the gRNA 5'UTR (compare constructs A and B). Since this result suggested that additional 5' sequence might be required for functioning of the PTE (or other 3' element) as a translational enhancer, the 5' sequence was extended into the p26 ORF to position 80 or 125. While extension of 5' sequence to position 80 did not improve translation of the reporter construct containing the 3' 400 nt (Figure 4A, construct C), extension to position 125 enhanced translation by 75-fold when the construct contained the extended 3' sequence and full-sized PTE (compare constructs A and E). Translation mediated by the

gRNA 5' 125 nt was enhanced by a more modest 4-fold over the exact 5'UTR when the precise 3'UTR was present (compare constructs A and D). These results suggest that efficient translation mediated by gRNA 5' sequences requires coding sequences at both ends of the viral genome.

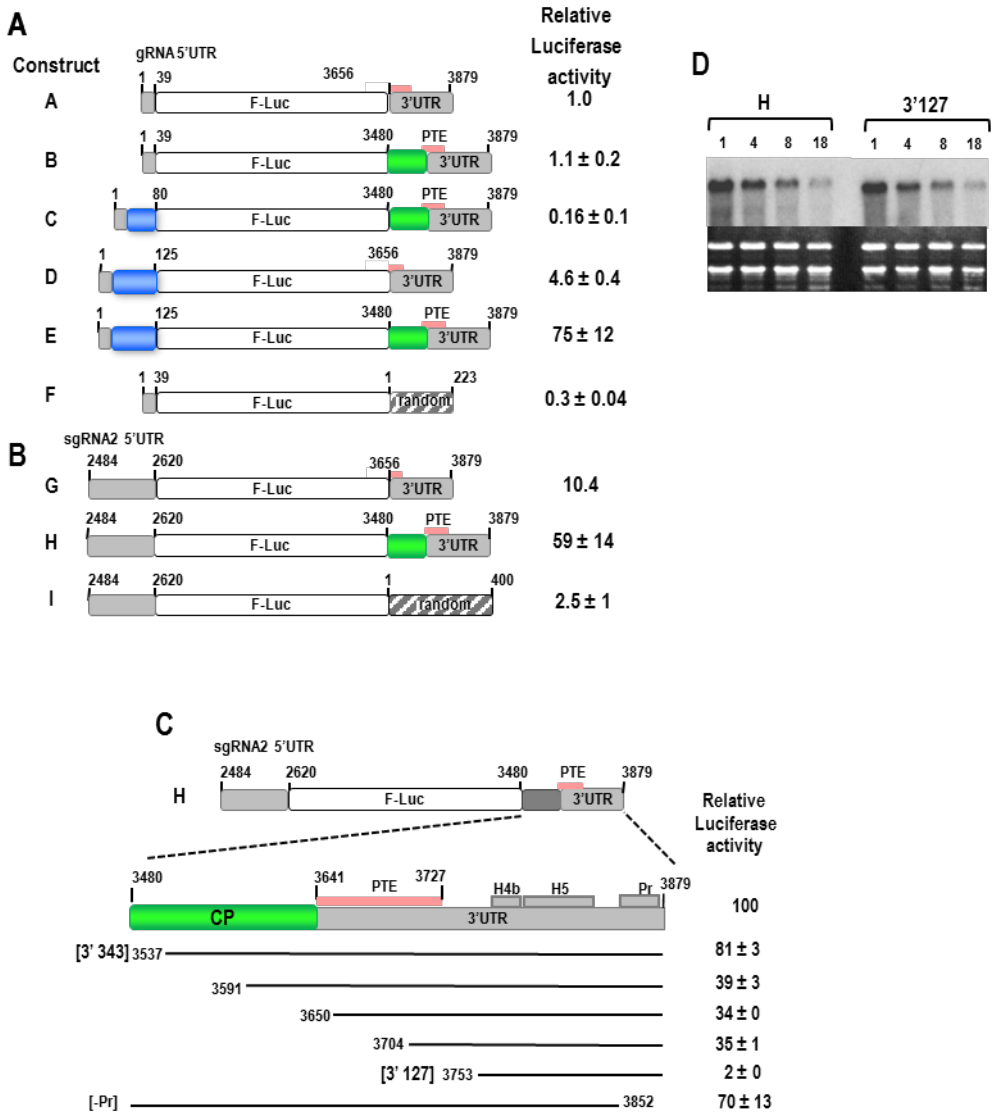


Figure 4: Identifying sequences that are required for translation in SCV
(A) gRNA constructs contained either the gRNA 5'UTR (1-39 nt) or extended 5' sequence (1-80 or 1-125 nt) at the 5' end and either the exact 3'UTR (positions 3656-3879), extended 3' sequence (positions 3480-3879; 400 nt), or 223 nt of random sequence at the 3' end. Short pink box denotes partial PTE element, with the larger pink box labeled "PTE" denoting the full-sized PTE spanning the 3'UTR junction. Light grey regions are SCV UTR sequences. Blue and green

regions are p26 and CP ORF sequences, respectively. Transcripts were inoculated into A. thaliana protoplasts along with a control R-Luc construct and luciferase levels assayed 18 h later. Standard deviation in three replicate experiments is shown.
(B) sgRNA2 constructs contained the exact 5'UTR (positions 2484-2620) at the 5' end and either the 3'UTR, 3'400 nt, or 400 random nt at the 3' end.
(C) Deletions were constructed in construct H, containing the 5'UTR of sgRNA2 and 3' 400 nt. Black lines represent remaining sequence. Identities of two fragments (3'343, 3'127) used in subsequent experiments are denoted. Positions of PTE, H4b, H5 and Pr are indicated. Standard deviation in three replicate experiments is shown.
(D) Stability of parental construct H and deletion construct 3'127 in protoplasts. Transcribed RNAs were transfected into protoplasts and total RNA isolated at the times (in hours) indicated above each lane. RNA was subjected to RNA gel blot analysis using oligonucleotide probes complementary to the luciferase coding region.

To investigate requirements for translation of the smaller of the two sgRNAs (sgRNA2), the precise sgRNA2 5'UTR (137 nt) was incorporated into reporter constructs containing either the exact SCV 3'UTR, the SCV 3'400 nt, or a random 400 nt sequence (Figure 4B). When combined with the 400 nt random 3' sequence, the sgRNA2 5'UTR was 8-fold better at directing translation than the gRNA 5'UTR with the random 223 nt sequence (compare constructs I and F). In the presence of the precise SCV 3'UTR, the sgRNA2 5'UTR was 10.4-fold more efficient at translation compared with the gRNA 5'UTR (compare constructs G and A). Extending the SCV 3'UTR to include a total of 400 nt improved translation directed by the sgRNA2 5'UTR by nearly 6-fold. These results suggest that the sgRNA2 5'UTR is a stronger promoter of translation than the gRNA 5'UTR in the absence of SCV 3' sequences. As with the gRNA 5' region, efficient translation directed by the sgRNA2 5'UTR requires the 3'UTR and upstream coding sequences.

To begin defining the SCV 3' sequences that enhance translation, deletions were generated in construct H, which contains the 5'UTR of sgRNA2 and the 3' 400 nt (positions 3480-3879) (Figure 4C). Deleting from the 5' end of the 3' sequence to position 3537 resulted in a 19% drop in translation while deleting to position 3591 caused

a 61% reduction. This reduced level of translation was essentially maintained when the deletions extended an additional 59 or 113 bases (to positions 3650 and 3704, respectively). Since these two latter deletions extended into the PTE without further reducing translation, a key translation element must either exist upstream of the PTE, or upstream sequences are required for proper PTE function. Deleting an additional 49 bases to position 3753 reduced translation to background levels suggesting that a critical element exists in the 3' region of the PTE or just downstream of the element. Deletion of only the Pr and short 3' terminal sequence in construct H reduces translation efficiency by 30%.

To rule out any possibility that RNA stability might be playing an indirect role in reducing translation efficiency, a construct containing 3' 127 nt, which was translated at a minimal level, and construct H, containing the sgRNA2 5'UTR and 3' 400 nt, were transfected into protoplasts. RNA levels were assayed between 1 and 18 hours post-transfection. No differences were found in the stability of the deletion-containing RNA transcripts compared with parental construct H transcripts (Figure 4D). This result indicates that a background level of translation due to deletion through the PTE is due to absence of required element(s) and not reduced stability of the transcripts.

A long-distance RNA:RNA interaction is important for sgRNA2 reporter mRNAs translation

Although the deletion analysis was inconclusive as to the involvement of the PTE in translation, the striking similarity between the predicted SCV PTE structure and the PEMV/PMV PTE translational enhancers suggested that the SCV PTE was a likely translation element. To first examine the validity of the predicted PTE secondary structure, in-line structure probing was used to evaluate the flexibility of PTE residues in

a fragment containing the 3' 343 nt (3’343), which supported a high level of translation when paired with the sgRNA2 5’UTR (see Figure 4C). In-line probing is based on self-cleavage of the RNA backbone, which requires in-line topology of the 2' hydroxyl, backbone phosphate and oxyanion leaving group. This topology only exists if the phosphate-sugar backbone can rotate about the C3’-O3’ and O3’P bonds, which does not occur if the base is overly constricted by canonical/non-canonical hydrogen bonding. Thus, the more flexible the nucleoside, the more backbone cleavage occurs. A typical in-line cleavage gel for the 3’343 fragment is shown in Figure 5A, lane 3. The cleavage pattern in the vicinity of the PTE was consistent with the predicted structure, with all but two of the flexible residues confined to predicted loop regions (Figure 5B). The lack of cleavages in the small symmetrical loop in the lower stem (LS) indicates that these residues are not flexible, which suggests that they may be constrained by non-Watson-Crick interactions across the stem.

To determine if the PTE is important for SCV translation, the 5' side PTE hairpin designated "H1" was disrupted by single mutations on both sides of the stem (Figure 5C). This hairpin was chosen as its structure and loop sequence are conserved in most other putative carmovirus PTE (see Figure 9A). Single alterations in the H1 stem of construct H, containing the sgRNA2 complete 5'UTR (sg5’137) and the 3' 400 nt, reduced translation to 41% or 12% of wt levels, while compensatory mutations that should reform the stem increased translation to 95% of wt levels (Figure 5C). To determine if the loop of PTE H1 plays a role in translation, two bases were altered from GC to CG in construct H (Figure 5C). These alterations reduced translation to 36% of wt. These

results indicate that the SCV PTE H1 stem and loop contribute to translational enhancement by the 3' region of SCV.

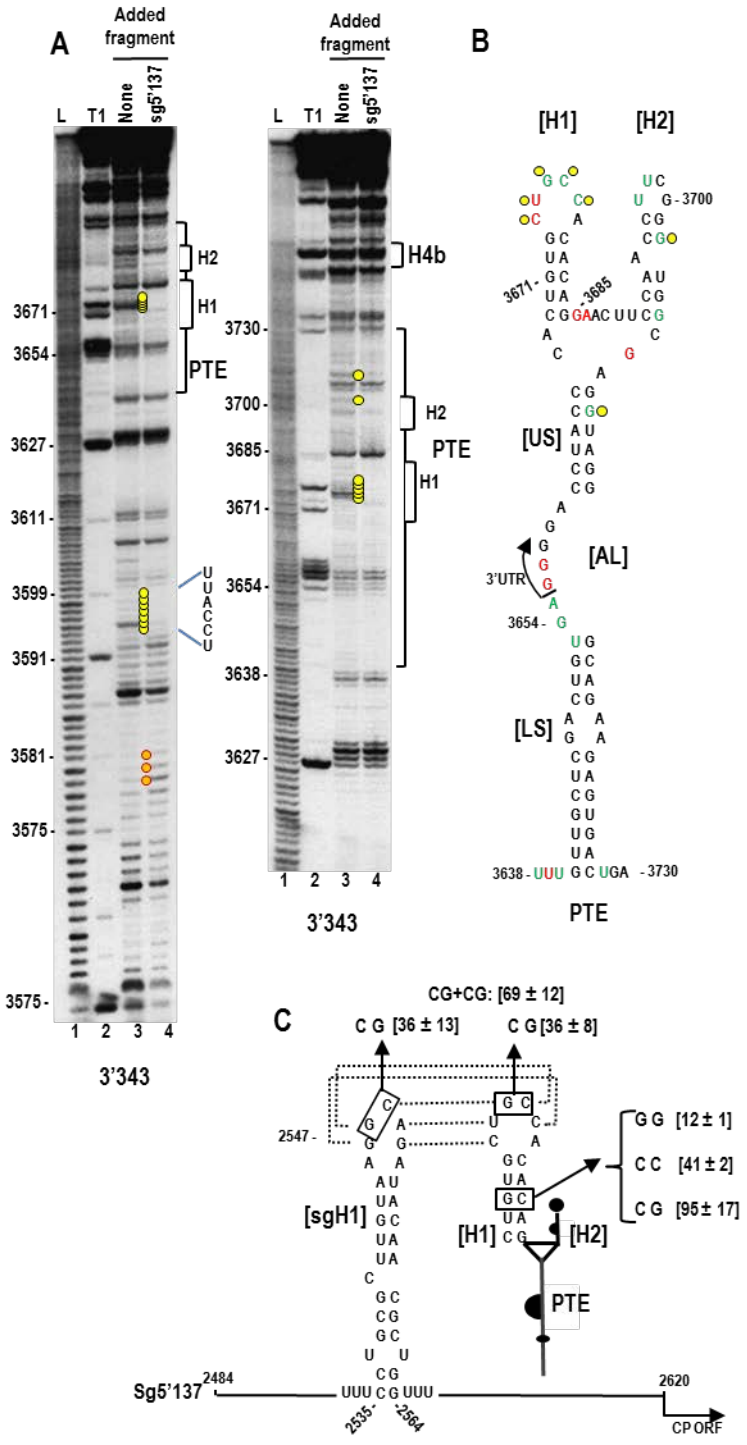


Figure 5: A kissing loop interaction exists between the 3' PTE and sgRNA2 5'UTR hairpin sgH1 that contributes to 3' translational enhancement

(A) In-line probing of fragment 3’343 in the absence and presence of the sgRNA2 5’UTR (5’137). The panel on the right is a longer run of the samples shown in the left panel. Fragment 3’343 was labeled at its 5’ end and allowed to self-cleave at 25ºC for 14 h. Intensity levels of the bands in lanes 3 and 4 are proportional to the flexibility of the residues. L, alkaline-generated ladder; T1, RNase T1 digest of partially denatured RNA. Note loss of flexibility in the PTE H1 loop in the presence of fragment 5’137, which is consistent with base-pairing between the PTE H1 apical loop and PTE-complementary sequence in the sgH1 apical loop (see C).
(B) Flexibility of residues in the 3’ PTE region. Data is from lane 3 in (A). Residues corresponding to high and low flexibility are shown by red and green colors, respectively. Yellow circles reflect residues that lose flexibility in the presence of fragment 5’137.
(C) Single and compensatory mutations generated in construct H that disrupt or reform the PTE H1 stem or the PTE RNA:RNA interaction. Levels of translation as a percentage of wt construct H are shown. Results are from three experiments with standard deviations given. Sequences that can putatively pair are connected by dotted lines.

One possible function for the PTE H1 in translation is through long-distance kissing-loop interactions with sequences in the 5' regions of the gRNA and sgRNA2. A hairpin (sgH1) located at position 2535 in the center of the sgRNA2 5'UTR contains apical loop sequences capable of forming a 5 bp interaction with the PTE H1 apical loop (Figure 5C). Converting the GC to CG in the sgH1 apical loop within construct H reduced translation to the same level as the GC alteration in the PTE H1 loop (36% of wt). When the sgH1 loop mutations were combined with the PTE H1 loop mutations, which were designed to be compensatory, translation was enhanced to 69% of wt (Figure 5C).

To provide additional support for this long-distance interaction, in-line probing was repeated for fragment 3'343 in the presence of the sgRNA2 5'UTR (fragment sg5’137). A direct RNA:RNA interaction between fragments should be discernable by loss of residue flexibility in the PTE H1 loop (since intermolecular base-pairing would significantly constrain residue flexibility). Addition of fragment sg5’137 to fragment 3'343 followed by in-line probing reduced the level of cleavage for all residues in the loop of PTE H1 (Figure 5A right, lane 4, five yellow adjacent circles). In addition,

residues in the stem of PTE H2 and the upper stem (US) that were slightly flexible in the in-line cleavage pattern of the PTE alone were no longer flexible in the presence of sg5'137 (Figure 5A, right, single yellow circles). Interestingly, additional alterations in the cleavage pattern were found upstream of the PTE (between positions 3578 to 3599, Figure 5A, left) when the two fragments were combined. The reasons for these additional cleavage alterations are not known, but may represent another RNA:RNA interaction.

We next examined the effect of the PTE on the cleavage pattern of sgH1 in fragment sg5'137 (Figure 6). In-line probing of sg5'137 produced a cleavage pattern that was consistent with the predicted sgH1 structure (Figure 6A, lane 3). Residues in the apical loop and upper stem region generated higher levels of cleavage indicating substantial flexibility, while residues in the remainder of the stem were susceptible to low levels of cleavage. The bulged cytidylate at position 2540 was significantly more flexible than adjacent base-paired stem residues. However, the symmetrical bulge in the lower stem contained minimally flexible residues, suggesting that the bulged uridylates are constrained by hydrogen bonding, possibly with each other across the stem.

Two regions of sg5'137 were strongly affected by addition of fragment 3'343 prior to cleavage: nearly the entire sgH1 hairpin and a small region upstream of the hairpin between positions 2505 and 2513 (Figure 6A, lane 4). Nearly every residue in the sgH1 stem that were previously susceptible to low but detectable cleavage, now exhibited background cleavage levels, indicating that the interaction between sg5'137 and 3'343 significantly reduces the flexibility of residues in the sgH1 stem. In addition, cleavages in sgH1 apical loop residues predicted to pair with the 3'PTE were reduced to

background levels (Figure 6A, lane 4, yellow circles). Only two residues in the upper portion of sgH1 were unchanged: the remaining unpaired adenylate at position 2546 in the apical loop, and the bulged cytidylate at position 2540 in the mid-stem region. This result is consistent with an RNA:RNA interaction between the 3' PTE H1 loop and the apical loop of sgH1, which supports the compensatory mutagenesis analysis (Figure 5C).

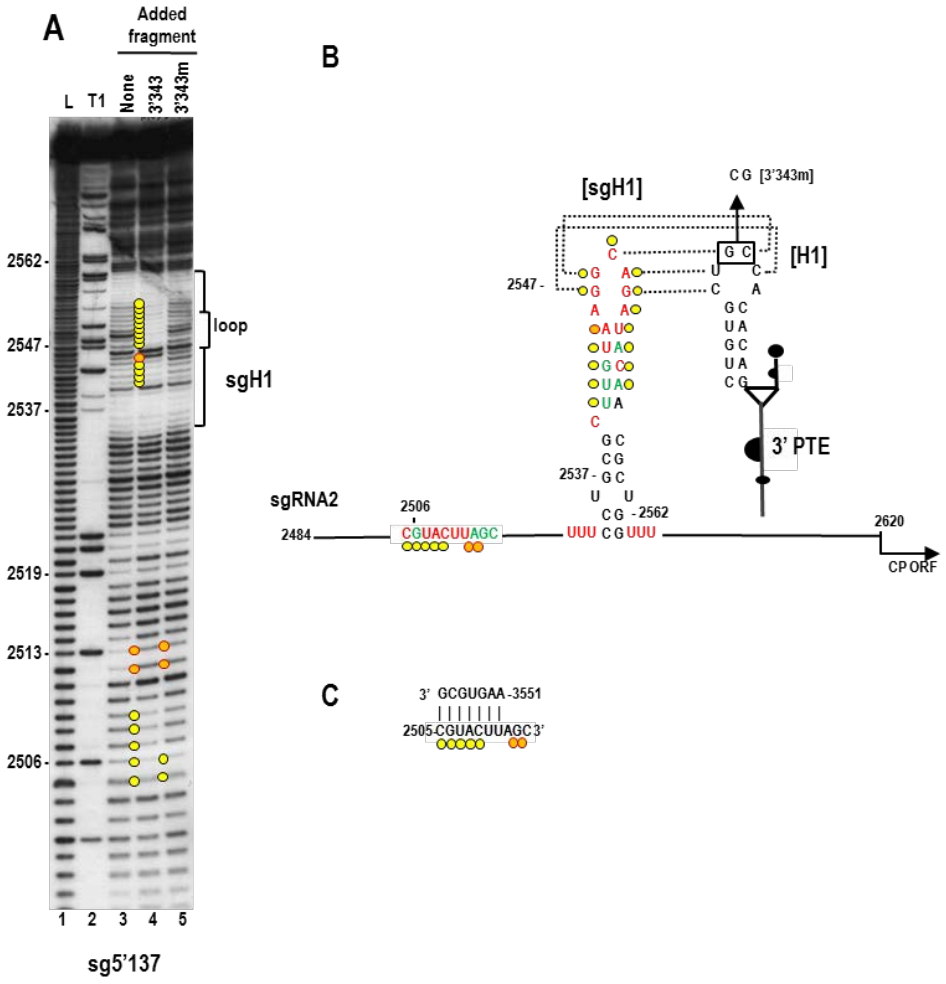


Figure 6: Effect of the PTE kissing loop interaction on the structure of the sgRNA2 5'UTR in the vicinity of sgH1

(A) In-line probing of the sgRNA2 5'UTR (5'137) in the absence and presence of fragment 3'343 and 3'343 with mutations disrupting the RNA:RNA interaction (3'343m). The 5'137 fragment was labeled at its 5' end and allowed to self-cleave at 25°C for 14 h in the absence and presence of 3'343 or 3'343m. L, alkaline-generated ladder; T1, RNase T1 digest of partially denatured RNA. Yellow and orange circles denote residues that lose or gain flexibility, respectively, in the presence of wt 3'343 or 3'343m. Note enhanced stability of the residues in the sgH1 stem and

loop when 5’137 and wt 3’343 are combined, which is consistent with base-pairing between sgH1 and PTE. Additional changes are also found in an upstream region.
(B) Flexibility of residues in the sgH1 region of the sgRNA2 5’UTR. Data is from (A). Residues corresponding to high and low flexibility are shown by red and green colors, respectively. Yellow and orange circles denote residues that lose or gain flexibility, respectively, in the presence of wt 3’343. Sequences that can putatively pair are connected by dotted lines. The mutations in 3’343m that disrupt the RNA:RNA interaction are shown. (C) Possible pairing between sequence at positions 3551-3557 in the 3’ region of SCV and sg5’137. Yellow and orange circles are as designated in (B).

Residues in positions 2505 to 2509 also exhibited reduced flexibility in the presence of 3'343 in all three replicates of this experiment, while residues in positions 2512 and 2513 displayed enhanced flexibility. This suggests either the existence of an additional RNA:RNA interaction or that the interaction between sgH1 and the PTE H1 generates an upstream conformational shift that alters the flexibility of these residues. To determine if the RNA:RNA interaction between sgH1 and PTE H1 is necessary for the upstream cleavage pattern alterations between positions 2505 and 2513, the in-line cleavage profile was determined for sg5’137 in the presence of fragment 3’343m, which contains a two base alteration in PTE H1 (GC to CG) that disrupts the RNA:RNA interaction. Our reasoning was that if the altered flexibility of residues 2505-2513 is unrelated to the PTE RNA:RNA interaction, then the cleavage pattern of these residues should be unaffected when this PTE interaction is prevented. As shown in Figure 6A lane 5, addition of 3’343m to labeled sg5’137 eliminates the cleavage pattern differences in sgH1 due to the PTE interaction. The cleavage pattern between residues 2505 and 2513, however, was not similarly restored to the wt pattern yet differed from the cleavage pattern found in the presence of wt 3'343. Residues in positions 2507 to 2509 no longer exhibited the reduced flexibility found in the presence of wt 3'343 while residues in positions 2505, 2506 and 2512 and 2513 retained their altered flexibility. This result

suggests that while the PTE:sgH1 interaction contributes to the changed cleavage pattern between positions 2505 and 2513, these changes are not due to a conformational shift in the region that results from this RNA:RNA interaction. Rather, it is likely that a second RNA:RNA interaction exists in this upstream region whose formation is partially dependent on the downstream PTE:sgH1 interaction. A search of the 3'343 nt region for sequence complementary to positions 2505 to 2509 revealed one possibility between residues 3551 and 3557 (Figure 6C).

A long-distance RNA:RNA interaction involving the PTE and a p26 ORF hairpin is important for the gRNA

A putative hairpin (gH3) located 94 nt from the 5' end of the gRNA in the p26 ORF contains a 6 nt apical loop that is fully complementary to the 6 nt loop of PTE H1 (Figure 7A). Hairpin gH3 is located within the portion of the 5' 125 nt fragment that substantially enhances translation of the reporter construct in the presence of the 3' 400 nt (Figure 4A). To investigate if a long-distance interaction takes place between PTE H1 and gH3, two base changes (GC to CG) were engineered into the gH3 apical loop that are complementary to the two base changes previously constructed in PTE H1. Mutations were generated in luciferase construct E, containing the 5' 125 nt of the gRNA and the 3' 400 nt (Figure 4A). Each of the alterations caused nearly identical decreases in luciferase levels, to 6.4 or 6.5% of wt construct E. Unexpectedly, the combined presumptive compensatory mutations did not substantially restore luciferase levels, reaching only 16.8% of wt levels. Since the possibility existed that this particular RNA:RNA interaction requires sequence specificity, or that the two base alteration disrupted the structure of gH3, the experiment was repeated with mutations in only one of the two adjacent residues in the apical loops of gH3 and PTE H1 (Figure 7A). Altering

only the guanylate in the PTE H1 apical loop to a cytidylate reduced translation of construct E to 35% of wt, while changing the cytidylate at position 107 in the apical loop of gH3 to a guanylate had an effect similar to the double mutation, with levels reduced to 4.2% of wt. Combining the two mutations enhanced translation to 279% of wt (Figure 7A). Altogether, these results support a long-distance RNA:RNA interaction between gH3 and PTE H1 that requires more sequence specificity in the interacting residues than needed for the PTE/sgH1 interaction described above.

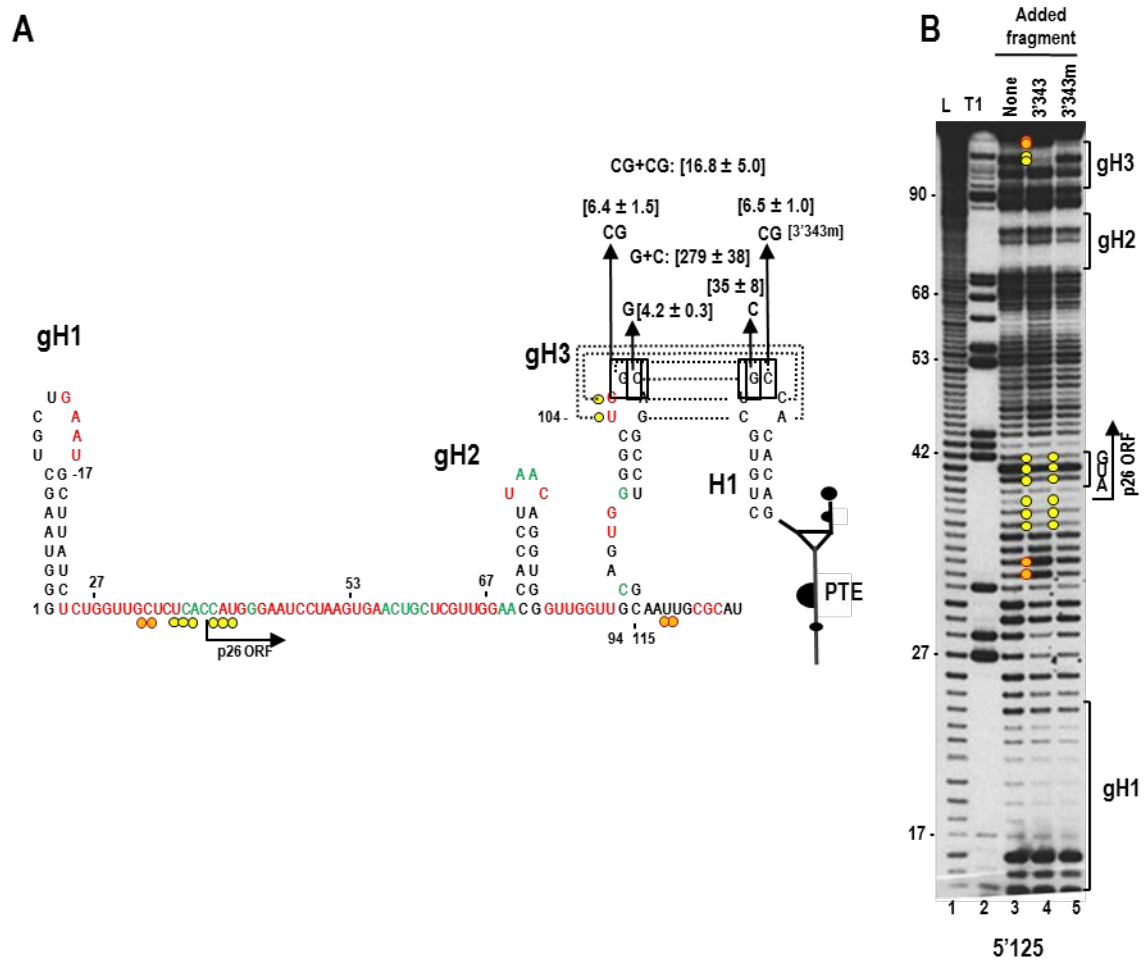


Figure 7: A kissing loop interaction between the 3' PTE and gRNA p26 ORF coding sequences contributes to 3' translational enhancement

(A) Putative structure of the 5' 125 nt of SCV. The secondary structure shown is based on mFold predictions and results of in-line probing (see B). The potential pairing between hairpin gH3 and

the PTE H1 is shown with dotted lines connecting the paired residues. Flexibility of residues in the 5' 125 nt region is from lane 3 in (B). Residues corresponding to high and low flexibility are shown by red and green colors, respectively. Yellow and orange circles reflect residues that lose or gain flexibility, respectively in the presence of fragment 3'343. Single and compensatory mutations generated in construct E that disrupt or reform the PTE RNA:RNA interaction are shown. Levels of translation as a percentage of wt construct E are given. Results are from three experiments with standard deviations given.
(B) In-line probing of labeled fragment 5'125 in the absence and presence of fragments 3'343 and 3'343m. L, alkaline-generated ladder; T1, RNase T1 digest of partially denatured RNA. Note that two gH3 loop residues lose flexibility and two residues near the base of gH3 show enhanced flexibility in the presence of fragment 3'343 only (lane 4, top), which is consistent with base-pairing between the apical loops of PTE H1 and gH3 (see A). Also note the presence of a second region of altered flexibility in the vicinity of the p26 initiation codon, which is retained in the absence of the PTE-mediated RNA:RNA interaction (lane 5) suggesting the presence of a second RNA:RNA interaction.

To determine if the PTE/gH3 RNA:RNA interaction is detectable by in-line probing, cleavage profiles were determined for the gRNA 5' 125 nt fragment (5'125) in the presence and absence of fragment 3'343. As with fragment sg5'137, a direct RNA:RNA interaction between gRNA gH3 and PTE H1 should be discernable by the loss of flexibility of residues in the 5'125 gH3 apical loop. The cleavage pattern of 5'125 was consistent with the presence of three hairpins: a 5' terminal hairpin (gH1) and two hairpins in the p26 coding region, including gH3 (Figure 7B lane 3). When fragment 3'343 was added prior to the cleavage reaction, the cleavage pattern of 5'125 was significantly altered in three regions: 1) the apical loop of gH3, where the highly flexible residues at positions 104 and 105 lost their flexibility, which is consistent with the predicted RNA:RNA interaction with PTE H1; 2) two residues at the 3' base of gH3, whose flexibility was substantially enhanced; and an upstream region (positions 32-42) that includes the initiation codon (Figure 7A and B).

To determine if the cleavage differences in all three locations were dependent on maintenance of the PTE/gH3 interaction, in-line probing of 5'125 was repeated in the presence of fragment 3'343m, which contains a two base alteration in the PTE H1 apical

loop. As shown in Figure 7B lane 5, the cleavage differences in the gH3 apical loop, the downstream adjacent region and the two residues at positions 32 and 33 were restored to their wt pattern, indicating that structure changes in these regions require the PTE/gH3 interaction. In contrast, the reduced flexibility of residues in positions 36 to 42 remained despite the loss of the PTE/gH3 interaction. This result strongly suggests the existence of a second RNA:RNA interaction that includes the initiation codon.

In-line probing was next conducted for fragment 3'343 in the presence and absence of fragment 5'125 (Figure 8). Addition of fragment 5'125 to 3'343 before cleavage significantly reduced the flexibility of all residues in the PTE H1 apical loop (Figure 8A, five adjacent yellow circles), consistent with the RNA:RNA interaction between PTE H1 and gRNA gH3. We also noticed very reproducible differences in the cleavage pattern in the apical loop of H4b when the 3' and 5' fragments were combined. The strong cleavages in the two guanylates in the H4b loop were reduced by approximately 50% in multiple independent experiments. In contrast, these H4b residues were never reduced in intensity when combined with fragment sg5'137 (see Figure 5A, right). Interestingly, the sequence of the lower stem and apical loop of H4b (5'CAUGGUG) is fully complementary to sequence at the initiation codon of the p26 ORF (5'CACCAUG) (Figure 8D).

Discussion

Long-distance protein-mediated or RNA:RNA interactions allow for the circularization of an RNA and for elements to function in locations distal to where they reside in the genome (39, 100). For the carmovirus TCV, ribosome-binding translation elements that overlap replication elements are found in the 3'UTR, allowing for a

conformational rearrangement of local RNA structures mediated by RdRp binding that coordinates the switch between replication and translation (13, 89). Repositioning of 3' proximal translation elements to the 5' end in TCV has been proposed to be facilitated by a bridge mediated by a single ribosome binding simultaneously to two distal elements (141), and not a more common long-distance RNA:RNA interaction found in several small plus-strand RNA plant viruses engaged in cap-independent translation (39, 142).

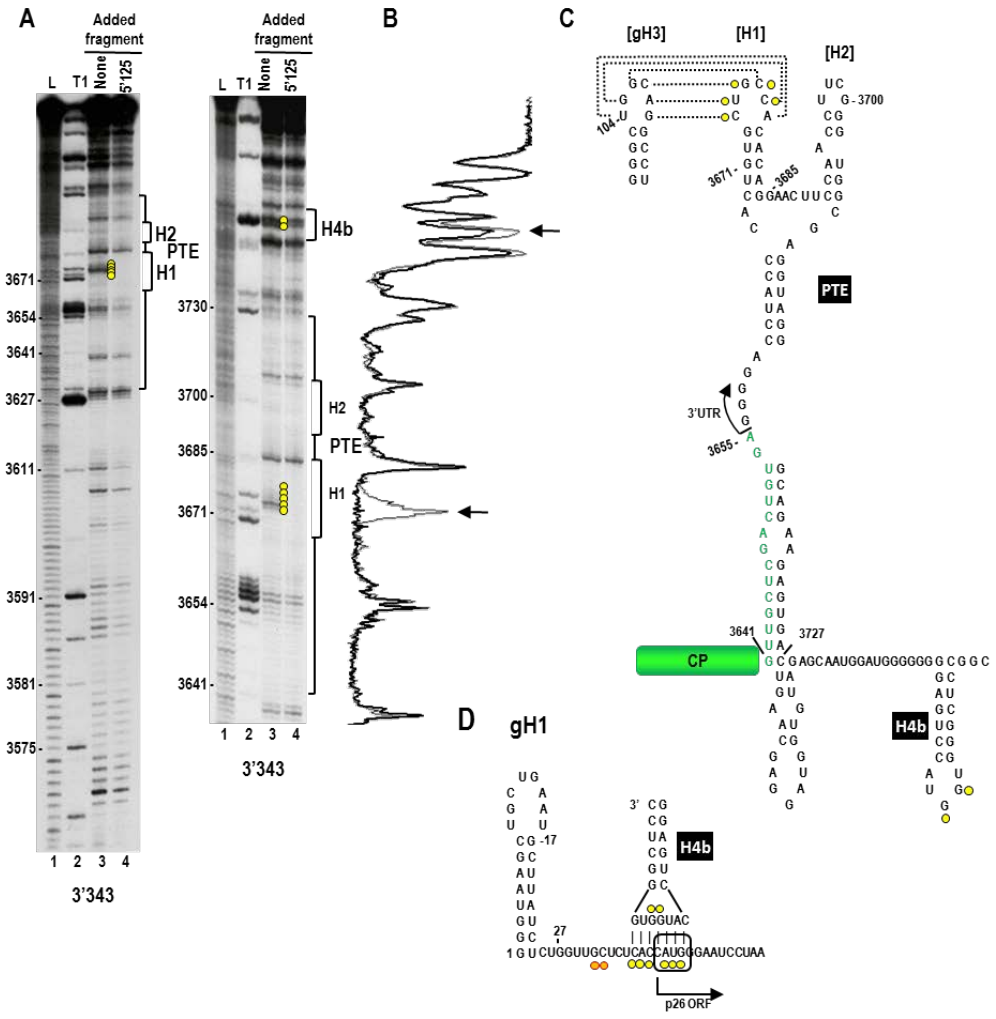


Figure 8: Two long-distance interactions occur between the 5' end of the SCV gRNA and the 3' region

(A) In-line probing of fragment 3'343 in the absence and presence of the gRNA fragment 5'125. The panel on the right is a longer run of the samples shown in the left panel. L, alkaline-

generated ladder; T1, RNase T1 digest of partially denatured RNA. Note loss of flexibility in the PTE H1 loop (lane 4, five consecutive yellow circles) in the presence of fragment 5'125, which is consistent with base-pairing between the PTE H1 apical loop and PTE-complementary sequence in hairpin gH3 (see C). Cleavages of two residues in the H4b apical loop were also reproducibly reduced in intensity in the presence of 5'125 but not 5'137.
(B) Densitometry tracing of the right side autoradiograph in A. Gray tracing is from lane 3 and black tracing is from lane 4. Arrows denote positions where the cleavage pattern differs between the two lanes.
(C) Location of cleavage pattern differences between lanes 3 and 4 in A. Yellow circles reflect residues that lose flexibility in the presence of fragment 5'125.
(D) Possible pairing between the apical loop of H4b and the region of reduced flexibility in the 5'125 fragment in the presence of 3'343. The start of the p26 ORF is indicated. The initiation codon is boxed.

To determine what aspects of translation in TCV are present in other carmoviruses, we conducted a phylogenetic search for conserved RNA structures in other carmoviruses. This led to the identification of the PTE-type element, which is present in seven carmoviruses. Comparison of the seven PTE structures reveals several conserved features (Figure 9A). All PTE have H1 hairpins containing a stem of 4 to 6 bp (5/7 with 5 bp) and a 6 nt loop. The sequences involved in the kissing-loop interactions are also highly conserved (UGCCA/UGGCA), and either sequence can be located in the 5' or 3' position (Figure 9A). The apical loop of the PTE H1 of PMV, a member of the same virus family (*Tombusviridae*), contains the similar sequence "UUGCAG". This conservation in sequence and structure likely reflects the important role of H1 in participating in the long-distance RNA:RNA interactions, and may explain why two base alterations in the kissing-loop interaction between PTE H1 and sgH1 were poorly compensatory (Figure 5).

The kissing-loop interaction between SCV PTE H1 and sgH1 caused a significant reduction in flexibility in the sgH1 stem. This reduced flexibility could be caused by enhanced stabilization of the hairpin stem, or it could result from conversion of an initial kissing-loop interaction into a more stable extended duplex by putative pairing of the

entire PTE H1 hairpin with the sgH1 stem (cUGUGCUGCCacACAG and cUUGUaaGGCAGaUACAa; putative paired bases are capitalized). This latter possibility would be similar to the two-step process for dimerization of HCV (143) and HIV (144), which require viral-encoded proteins to achieve the more stable complex (143). While this is an intriguing possibility, it should be noted that this extended pairing is not discernable for gH3, and that none of the other carmovirus putative interactions have this extended pairing capability.

The SCV PTE joins a growing list of 3' TEs that can pair with 5' elements to bridge the ends of gRNA and sgRNA (5, 39, 68, 69, 71, 73, 82). These elements fold into diverse 2D conformations including YSS (82, 85) and ISS (73). The YSS, conserved among all viruses in the tombusvirus genus (5), share some similar features with the carmoviral PTE including a branched stem upper structure, and the 5' side hairpin participating in the RNA:RNA interaction. The sequences engaged in the kissing-loop interactions are also similar in length and composition, consisting of between 5-7 bases and including the conserved sequence "CCA". The 5' gRNA sequences that interact with the tombusvirus 3'TE are also located in 5' terminal elements, similar to the location of the presumptive interacting sequences for some of the carmoviral PTE (Figure 9C).

The PEMV PTE H1, which is not involved in any discernable RNA:RNA interaction (92), contains sequence unrelated to the other PTE. Interestingly, the necessity for specific sequences in the H1 loop may differ depending on the location of the 5' complementary sequence. Compensatory alterations involving two adjacent bases restored nearly 70% of wt translation when assaying the interaction between the sgRNA2 sgH1 and the PTE, while the identical compensatory alterations with the gRNA gH3 only

restored 16% of wt translation levels. Single base compensatory alterations of the latter interaction produced substantially more translation than the wt interaction. We currently have no explanation for the enhanced translation of this mutant interaction.

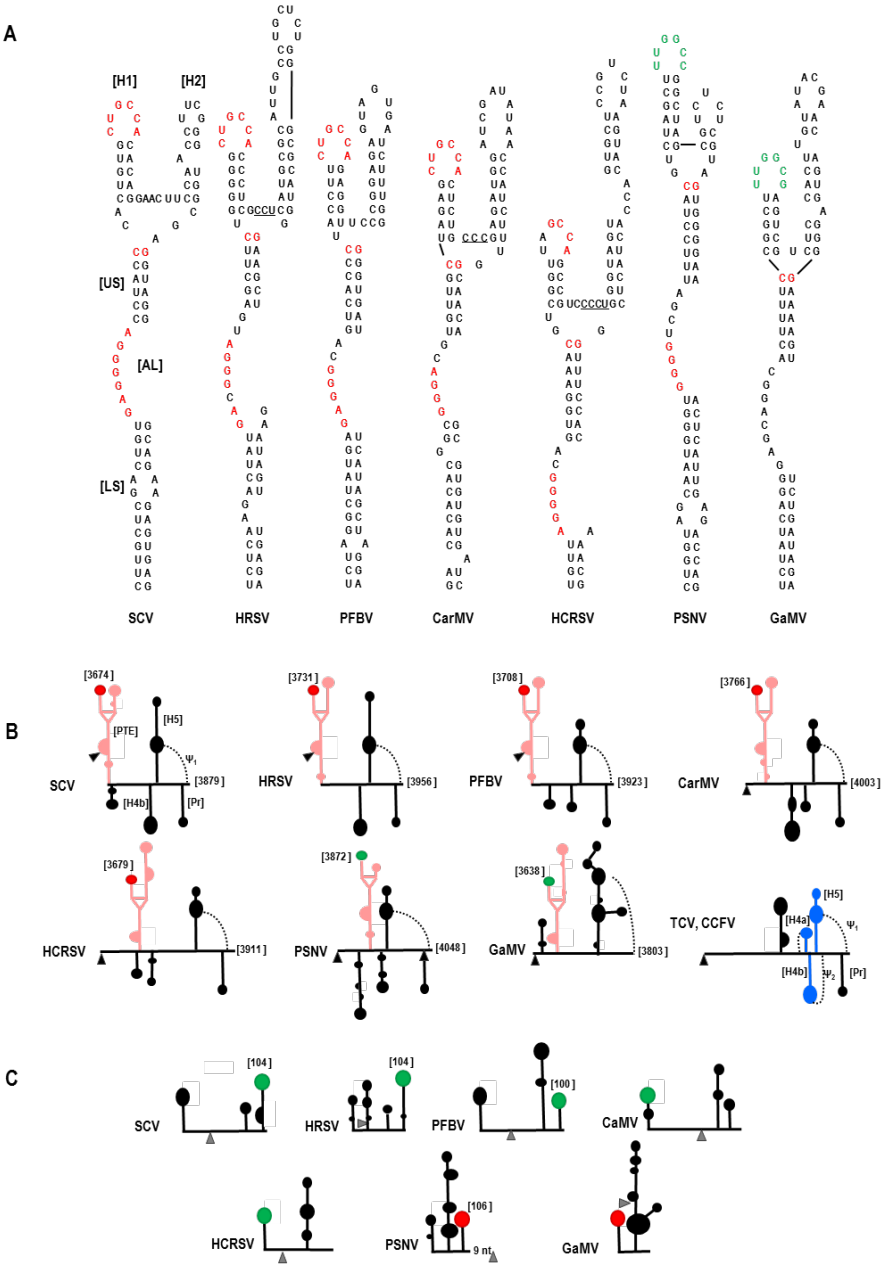


Figure 9: Comparison of PTE and PTE-interacting sequences in carmoviruses
(A) Sequences of PTE found in seven carmoviruses. Conserved sequences are colored alike. Note that the sequence in green contains the complement of the conserved GCCA sequence (UGGC).

(B) Location of the carmovirus PTE in the 3' regions of the viral genomes. Black triangle denotes location of the stop codon for the CP ORFs. The PTE is in pink, the conserved H1 sequence CUGCCA is denoted by a red terminal loop and the conserved H1 sequence UUGGCC/G is denoted by a green terminal loop. Positions of the conserved sequences within the viral genome is given. Note that all viruses except GaMV have Pr, Ψ_1, H5 and H4b and none have Ψ_2. Secondary structures are based on mFold predictions and tertiary structures are based on visual observations. The structure of the 3'UTR of TCV and CCFV is provided for reference.
(C) Location of PTE H1 complementary sequence near the 5' end of the gRNA. Note that all sequences are in apical loops of hairpins located either at the 5' terminus or positioned such that the loop sequence is 100 to 106 nt from the 5' end as denoted in the figure. The distance of these latter complementary sequences from the initiation codons (gray triangle) is highly variable. Secondary structures are based on mFold predictions. Sequences complementary to PTE H1 sequence CUGCCA is in red and UUGGCC/G is in green.

The H2 hairpins of the PTE are highly variable in sequence, length, and presence of interior bulges. The upper stem (US) of the PTE ranges from 6 to 8 bp, and all (including PMV and PEMV) have a C:G pair that closes the top of the stem. The lower stems (LS) of the PTEs are variable in length and sequence, with the majority containing small symmetric or asymmetric loops (Figure 9A). The PTE large asymmetrical loop (AL) contains between 6 and 8 nt, with nearly all accommodating at least three consecutive guanylates. These guanylate residues are required for eIF4E binding to the PEMV PTE *in vitro* and *in vivo*. The sequence between H1 and H2 in the PEMV PTE (CCCU) was also important for eIF4E binding and was proposed to participate in a pseudoknot with sequences in the asymmetric loop, although direct evidence for this interaction is lacking. While PMV and 3/7 carmovirus PTE contain sequence similar or identical to this conserved sequence between H1 and H2 (Figure 9A, underlined residues), this is not true for the remaining carmoviruses including SCV. A pseudoknot between the sequence connecting H1 and H2 and the asymmetric loop is also improbable for SCV (three possible paired bases would include two G:U pairs) and not discernable for two other carmoviruses, with only one or no bases between the two hairpins.

A previous report on translation of carmovirus *Hibiscus chlorotic ringspot virus* (HCRSV) identified a 6-nt sequence as important for translation (145). This sequence, GGGCAG, lies within the large asymmetric loop of the newly identified HCRSV PTE element (Figure 9A), which is critical for function of the PEMV PTE. The HCRSV report also concluded that a nearby hairpin was not important for translation. The alterations generated in this hairpin, now known to be H2 in its PTE, do not affect the structure of the H2 lower stem and thus are unlikely to have significantly affected the structure of the HCRSV PTE. It should also be noted that results presented in this report (145) relied on the premise that the 3' translational enhancer does not influence translation from the 5' end of the viral genome, a faulty premise given the current results on translation of SCV.

The location of PTE in the 3' region of these viruses is variable (Figure 9B). Three viruses (SCV, PFBV, HRSV) have PTE in identical locations that commence in the CP coding region and terminate in the 3'UTR, while the PTE of the remaining four viruses (as well as PEMV and PMV) reside wholly within their 3'UTR. The structure of the RNA between the PTE and the downstream conserved H5 hairpin is also variable as predicted by mFold (139). Interestingly, while 5/7 carmoviruses with PTE have an H4b-type hairpin just upstream of H5, none of these viruses appear capable of forming a stable pseudoknot (Ψ_2) between the H4b apical loop and sequence 3' of H5. In contrast, all non-PTE carmoviruses with H4b contain sequence capable of forming Ψ_2. This suggests that the H4b loop sequences in PTE carmoviruses are available for an alternative function(s). For SCV, this alterative function may be an RNA:RNA interaction with sequence that includes the p26 initiation codon. This possibility is based on finding that

the H4b apical loop and adjacent stem sequence can potentially form 7 bp with sequence in the vicinity of the p26 initiation codon. In addition, both the H4b loop and sequence in the vicinity of this initiation codon lose flexibility when 5' and 3' fragments are combined, even when the PTE/gH3 interaction is eliminated (Figure 7). Our results also suggest the possibility of at least one additional RNA:RNA interaction involving sequence between positions 2505 and 2509 in the 5'UTR of sgRNA2 (Figure 6A). Reduced flexibility of some of the residues in this location remained when the PTE H1 was combined with a mutant sg5'137 fragment that eliminates the RNA:RNA interaction (3'343m). This suggests that altered flexibility of these residues is partly, but not fully, dependent on the PTE H1/sgH1 interaction. The sequence in the 3' region with the most extensive complementarity resides between positions 3551 and 3557 (Figure 6C).

All carmoviral PTE contain putative long-distance interacting sequences in apical loops of hairpins located in both the 5' region of the gRNA (Figure 9C) and the sgRNA2 5'UTR. In the gRNA, putative interacting sequences were found in two locations: either in a 5' terminal hairpin (CaMV, GaMV, HCRSV [and also PMV]), or between 100 and 106 nt from the 5' end (SCV, HRSV, PFBV, PSNV). For viruses with short 5'UTR, the PTE H1 complementary sequence that is distal to the 5' end resides in the coding region of the 5' proximal ORF. For PSNV, which has a more extensive 5'UTR, the sequence is still within the 5'UTR. This suggests that the 5' sequence that interacts with PTE H1 is not randomly placed, but rather exists in a particular location necessary to fulfill its function in translation (Table 4).

For sgRNA2, the interacting sequences are all located in apical loops of hairpins between 74 and 90 nt upstream of the CP initiation codon, which is centrally positioned

within the sgRNA2 5'UTRs of carmoviruses whose sgRNA2 transcription start sites are known or estimated based on strong phylogenetic conservation of sequences just upstream of carmovirus sgRNA2 start sites (Table 5). In the next chapter how the position of the 5' PTE-interacting sequences influences the PTE-mediated translation of gRNA and sgRNA2 is discussed.

Table 4: Location of the PTE 5' kissing-loop sequences in carmoviruses

Virus	PTE kissing-loop sequence	Position of kissing-loop sequence in gRNA 5' region	Position of kissing-loop sequence in sgRNA2 relative to 5' end
SCV	CUGCCA	UGGCAG 104 nt, in ORF	AGGCAG 64 nt
PFBV	CUGCCA	UGGCAG 100 nt, in ORF	UGGCAG 63 nt
HnRSV	CUGCCA	UGGCAG 104 nt, in ORF	UGGCAG 62 nt*
PSNV	UGGC	GCCA 107 nt, in 5'UTR	GGCCAAC 65 nt
GaMV	UUGGCG	CGCCAA 5' Terminal HP	CGCCAA ND**
CarMV	CUGCCA	UGGCGG 5' Terminal HP	GGCAG 64 nt
HCRSV	UAGCCA	UGGC 5' Terminal HP	GGCUG 64 nt*

*****Estimated based on phylogenetic conservation of sgRNA2 start sites shown in Table 5.
** Start site for sgRNA2 could not be determined.

Table 5: Estimated sgRNA2 start site

Virus	Known or putative (*) sgRNA2 5' start site	gRNA 5' end
SCV	CCCCC ACU GGUAA	GGGUAA
CarMV	CCCCC GUU GGGUAA	GGGUAA
PFBV	CCCCC UCU GAUAAA	GGGAUA
PSNV	CCCCC A GGGAU	GGGGAU
TCV	CCCC GU GGGUAAU	GGUAAU
HnRSV	CCCCC GGU GAUUUU*	GGGGUUUU
HCRSV	CCC UCC GGGAAA*	GGGAAA

Chapter 3: Position of the 5'-PTE interacting sequences influences the PTE's function[2]

Introduction

Previous work by others has shown that the position of the 5' end sequence that participates in a long-distance kissing-loop interaction often modulates 3'CITE-mediated translational efficiency (80, 85). The BYDV BTE facilitates cap-independent translation initiation by participating in a long-distance RNA:RNA interaction between the apical loop sequence of the BTE and a complementary loop sequence of a hairpin (BCL) in the 5'UTR. In the BYDV gRNA, the 5' BCL is positioned 103-107 nt from the 5' end (81). Moving the 5' BCL 30-63 bases downstream from the 5' end by inserting sequences at the extreme 5' end of the 5 UTR had a significant negative impact on BTE-mediated cap-independent translation of the reporter mRNAs. However, moving the BCL to a site closer to the 5' end had only a modest negative effect on translation (80). In contrast to BYDV, a similar study in TBSV, which contains a YSS as its 3' CITE, indicated that repositioning the 5' interacting sequence relative to either the 5' end or the initiation codon had no significant negative effect on translation (85).

As mentioned in Chapter II, for several carmovirus gRNAs, 5' end sequences that participate in a known or predicted long-distance RNA:RNA interaction with the PTE H1 are found either in the loop of the 5' terminal hairpin gH1 or 100 to 106 nt downstream from the 5' end in the apical loop of hairpin gH3. In sgRNA2, the location of the

[2]Modified from: Position of the kissing-loop interaction associated with PTE-type 3'CITEs can affect enhancement of cap-independent translation. Maitreyi Chattopadhyay, Micki Kuhlmann, Kalyani Kumar, and Anne E. Simon. Virology 458-459 (2014) 43-52.

interacting sequence is even more specific: 62 to 65 nt from the 5’ end of sgRNA2 in the apical loop of hairpin sgH1 (Table 4) (90). The conserved placement of the PTE pairing partners for carmoviral PTE at one of two distinctive locations near the 5’ end of the gRNA and sgRNA2 suggests that proper positioning of the long-distance interaction is needed for PTE function.

In this chapter, I continue my analysis of the SCV PTE to examine the hypothesis that conserved positioning of PTE-interacting hairpins is important for PTE function as an efficient translation enhancer. Using reporter constructs assayed *in vivo*, I have determined that: 1) relocating 5’ partner hairpins at least 27 nt closer to the 5’ end does not substantially reduce function; 2) gH1, found at the 5’ terminus of most carmoviruses, impacts translation mediated by SCV 5’ sequences; 3) transferring the gH3 loop to gH1 allows for translation but at a significantly reduced level; and 4) movement of sgRNA2 hairpin sgH1 36 nt closer to the 5’ end can significantly impact translation despite only minor apparent changes to the structure of the hairpin. These results add to our understanding of the functioning of PTE-type translational enhancers and requirements for 3’CITE kissing-loop interactions.

Materials and methods

Generation of constructs

Chapter II described single luciferase reporter constructs 5’125 -Fluc-3’400 (E) and 5’137 -Fluc-3’400 (H), which were the parental constructs used for the alterations generated (90). Oligonucleotide-mediated site-directed mutagenesis was used to generate specific mutations in the 5’ 125 nt and the 5’ 137 nt regions. All mutations were verified

by sequencing. Plasmids were linearized with SspI and used as templates for *in vitro* transcription using T7 RNA polymerase.

In vivo translation in protoplasts

Please see Chapter 2.

SHAPE structure probing

Full-length SCV gRNA transcripts were used as the template to assay RNA structure in the PTE region. Transcripts synthesized from 5'137 -Fluc-3'400, 5' 137Δ27-Fluc-3'400 and 5'137Δ36-Fluc-3'400 were used to assay the structure of sgH1. Six pmoles of *in vitro* transcribed RNAs were heated at 65° C for 5 min and snap-cooled on ice for 2 min. Transcripts were folded in SHAPE Folding Buffer-2 (80 mM Tris-HCl pH 8, 11 mM $Mg(CH_3COO)_2$, 160 mM NH_4Cl) for 20 min at 37°C. Three pmoles of the folded RNA was combined with either NMIA or DMSO at a final concentration of 15 mM. RNA reaction mixtures were incubated for 35 min at 37°C (5 half-lives of NMIA) followed by ethanol precipitation. RNA was re-suspended in 8 µl of 0.5x TE buffer. Primer extension reactions were performed using ^{32}P-labeled oligonucleotides and Superscript III reverse transcriptase (Invitrogen) following the protocol as previously described (146). For SCV gRNAs, oligonucletides used were complementary to positions 3826-3851 and 3715-3738. For The 5'137 -Fluc-3'400 and derivatives, the oligonucleotide (5'-GTTTTTGGCGTCTTCCATGAGC-3') was complementary to sequence in the Fluc coding region. Products of the reverse transcription reaction were separated on 8% denaturing polyacrylamide gels and visualized using a phosphorimager.

Table 6: Constructs used in Chapter 3

Name	Description
Construct E	Single F-Luc flanked by 5'125 and 3' 400
Construct H	Single F-Luc is flanked by 5'sgRNA2 137 and 3'400
Δ61-69	Deletion of 9 bases in 5'125 nt in construct E
Δ52-69	Deletion of 18 bases in 5'125 nt in construct E
Δ43-69	Deletion of 27 bases in 5'125 nt in construct E
C_{106} to G	C to G change at position 106 in construct E
C_{106} to G + gH3 loop	gH1 loop replaced with gH3 with C to G change at position 106
ΔgH1	Deletion of gH1 in 5'125 nt in construct E
m1	Disruption of the gH1 stem in construct E
m2	Restoration of the gH1 stem in construct E
$m2_{AUA}$	AUA change to AUA in m2 in construct E
$gH1_{UUG}$	AUG change to UUG in the gH1 loop in construct E
$CbMV_{gH1}$	SCV gH1 is replaced with CbMV gH1 in construct E
$gH1_{CbMVL}$	SCV gH1 loop is replaced by CbMV gH1 loop sequence
Δ2525-2533	Deletion of 9 bases in 5'137 in construct H
Δ2516-2533	Deletion of 18 bases in 5'137 in construct H
Δ2507-2533	Deletion of 27 bases in 5'137 in construct H
Δ2498-2533	Deletion of 36 bases in 5'137 in construct H
Δ2489-2533	Deletion of 45 bases in 5'137 nt construct H
Δ2484-2533	Deletion of 49 bases in 5'137 nt construct H
Δ2498-2506	Deletion of internal 9 bases in 5'137 nt construct H

Table 7: Oligonucleotides used in Chapter 3

Name	Position	Sequence	Polarity
SCV g5' 125 m	G22A	CGATCA*GGATCC*TCTATTCGTGCTGAATCGAAT AGGCTGGTTGCTCTCACC	+
SCV sg2 del	24-50	AGAACAAATCATCGCCTGCGCTTGTAA	+
SCV sg2 del	24-50	TACAAGCGCAGGCGATGATTTGTT	-
SCV sg2 del	33-50	CATCGTACTTAGCTCCTGCGCTTGTAA	+
SCV sg2 del	33-50	TACAAGCGCAGGAGCTAAGTACGA	-
SCV sg2 del	42-50	TAGCTTTTGTGGTCCCTGCGCTTGTAAG	+
SCV sg2 del	42-50	TACAAGCGCAGGGACCACAAAAGCT	-
SCV g5'UTR DEL F2	24-69	ATCCAC*GAGCTC*ATGCGCCAATTGCAGGCCTG CCAGC	-
SCV g5'UTR DEL F2	24-69	CTCT*GACCAT*GACCACCTTAACAGGTGGGTTGG TTGCA	+
SCV g5'UTR DEL F1	24-69	GCAACCAACCCACCTGTTAAGGTGGTCATGGT CAGAGATAAGCA	-
SCV R7	3572-3590	CATGAC*AATATT*ACAGTACGGAGCTTGAGCAA	-
SCV R9	3679-3700	CATGAC*AATATT*AAACCACGTCAAGCTTGACGG	-
SCV R8	3731-3753	CATGAC*AATATT*TCCATTGCTCTACACCATCCTC	-
SCV g5'125 m	G22A	CGATCA*GGATCC*TCTATTCGTGCTGAATCGAAT AGGCTGGTTGCTCTCACC	+
SCV deletion	3595-3554	GGACTCCCCCCCCAACAGTACGGAGCTTG	-
SCV deletion	3595-3554	CAAGCTCCGTACTGTTGGGGGGGGAGTCC	+
SCV deletion	3515-3585	AGAAACAATGGAACAGTACTACCCCCACG	-
SCV deletion	3515-3585	CGTGGGGGTAGTACTGTTCCATTGTTTCT	+
SCV CUU CGU	3688	CGCCACCCGAAGGTTGGACGTTCC	-
SCV CUU CGU	3688	GCCACACAGGAACGTCCAACC	+
SCV GGG GTG	3658	AGGTCACCTCACAGTCGAGC	-
SCV GGG GTG	3658	GCTCGACTGTGAGGTGACCT	+

Underlined letters indicate mutated bases; Italic letters are restriction enzyme sites;"+" and "-"indicate polarity.

Results

The SCV PTE is supported by an extended base stem that is conserved in PTE-containing carmoviruses

The structure of the SCV PTE was previously determined using in-line structure probing of a 343 nt fragment that was co-terminal with the 3' end of the gRNA (see Chapter II) (90). This fragment, when located downstream of luciferase in a reporter construct containing the 5' 125 nt of SCV, was able to enhance translation by 75-fold compared with control constructs. For the current study, I employed SHAPE to examine the structure of the PTE within the full-length SCV gRNA. SHAPE interrogates the local backbone flexibility of the RNA by modifying 2'OH groups of unconstrained bases using the acylating agent, NMIA (146). The sites of modification are resolved by reverse transcriptase-mediated primer extension, which is impeded by 2' acylated nucleotides. An RNA structural map is generated in which flexibility of the nucleoside correlates with the level of SHAPE modification.

The SHAPE-derived structure of the PTE in the full-length gRNA (Figure 10B and C) was identical to the structure previously defined by in-line probing of a 3' co-terminal 343 nt fragment (90). All residues in the lower stem, including the A-A and G-A non-Watson-Crick pairs, were inflexible using both assays, denoting that a stable, paired lower stem also exists in the PTE within the full-length gRNA. The 5' side purines in the linker sequence between the two hairpins were flexible in both assays, whereas the adjacent pyrimidines were inflexible. Residues on the 3' side of the 3-way junction region were also flexible by both procedures. The different techniques and/or different sized templates also revealed differences in residue flexibility in the PTE and

surrounding sequences. For example, residues 3655-3658 in the 8 nt G domain were sensitive to NMIA in the full-length gRNA whereas residues 3653-3657 were susceptible to in-line cleavage within the shorter fragment. Furthermore, in the apical loop of PTE hairpin H2, different sets of residues were flexible in the two assays. The most striking difference was in the apical loop of PTE hairpin H1, which participates in the kissing-loop interaction. Five of six loop residues were flexible in the 343 nt fragment, whereas all loop residues were inflexible in the gRNA that contains the 5' partner sequences in gRNA hairpin gH3 and sgRNA hairpin sgH1 (Figure 10A and B). This latter result suggests that the PTE is paired with one of its interacting sequences in the full-length gRNA.

Previous mFold computational predictions (139) combined with in-line RNA structure probing suggested that the PTE was upstream of a small hairpin and long single-stranded region, followed by conserved carmovirus hairpin H4b (90). SHAPE probing of SCV gRNA did not support the predicted structure upstream of H4b. Rather, the SHAPE flexibility profile was consistent with the PTE branching off of a 3-way junction, which also contains a 5'side hairpin and an elongated base stem (ELS) (Figure 10A and B). Examination of the other six PTE-containing carmoviruses using mFold (139) revealed that a similar structural configuration was one of the predicted structures for all of these viruses (Figure 11). It is currently unknown if any of the presumptive conserved elements adjacent to the PTE contribute to translation or other viral function.

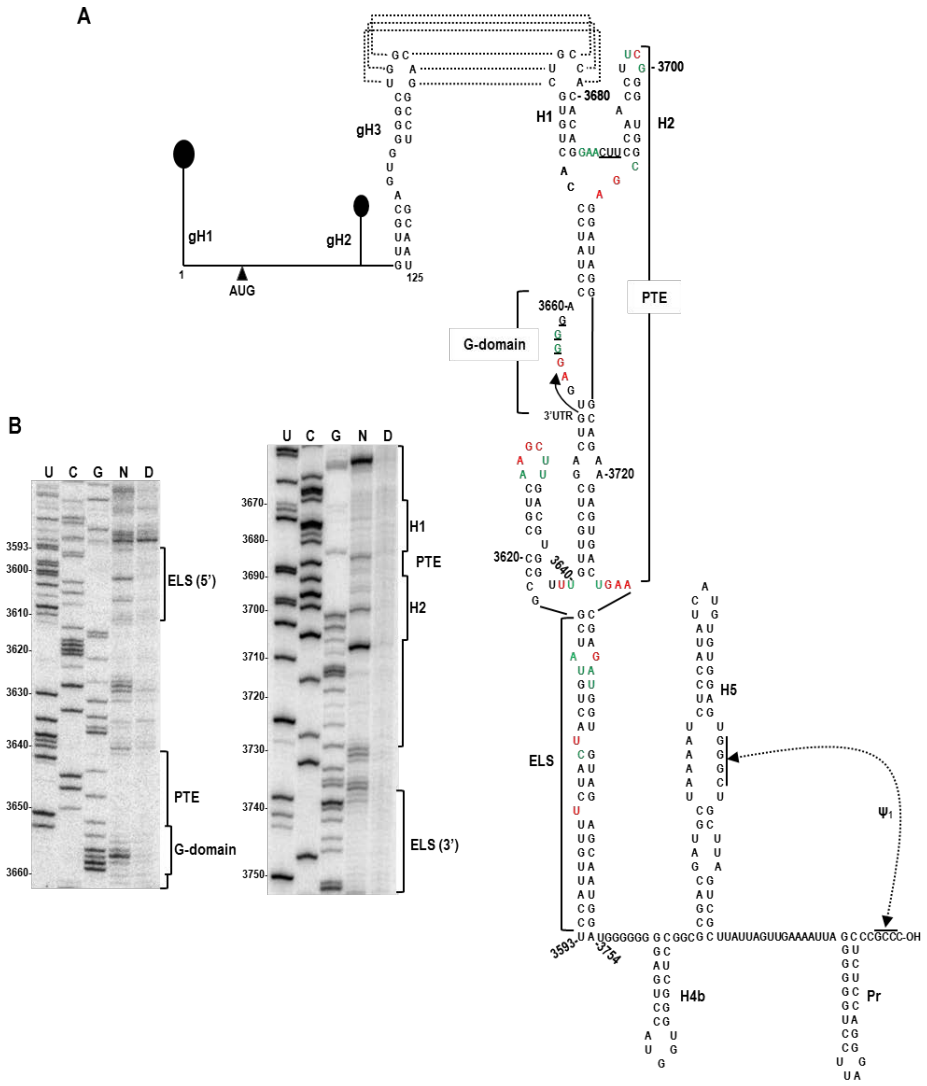


Figure 10: SHAPE predicted structure of the SCV PTE

(A) Secondary structure of the 5' and 3'ends of the SCV genomic RNA. The 5' 125 nt harbors three predicted hairpins. The initiation codon for p26 is shown. The 3' region of SCV contains the PTE, which starts in the CP ORF and terminates in the 3'UTR (arrow). Residues potentially involved in a pseudoknot necessary for eIF4E binding to the PEMV PTE are underlined. The long-distance kissing-loop interaction between gH3 and PTE H1 is shown with dotted lines. Residues in the PTE and surrounding regions corresponding to high and low reactivity to NMIA within full length SCV gRNA are denoted by red and green colors, respectively. ELS, elongated lower stem. Figure is slightly modified from the published version.

(B) Examples of SHAPE autoradiograms. Sequencing ladders for U, C, and G were generated by reverse transcription in the presence of dideoxy NTPs. Flexibility of the bases is proportional to the intensity of the bands in the NMIA lane (N). D, RNA treated with DMSO only. Positions of the 5' and 3' sides of the ELS, and conserved features in PTE are indicated to the right of each gel.

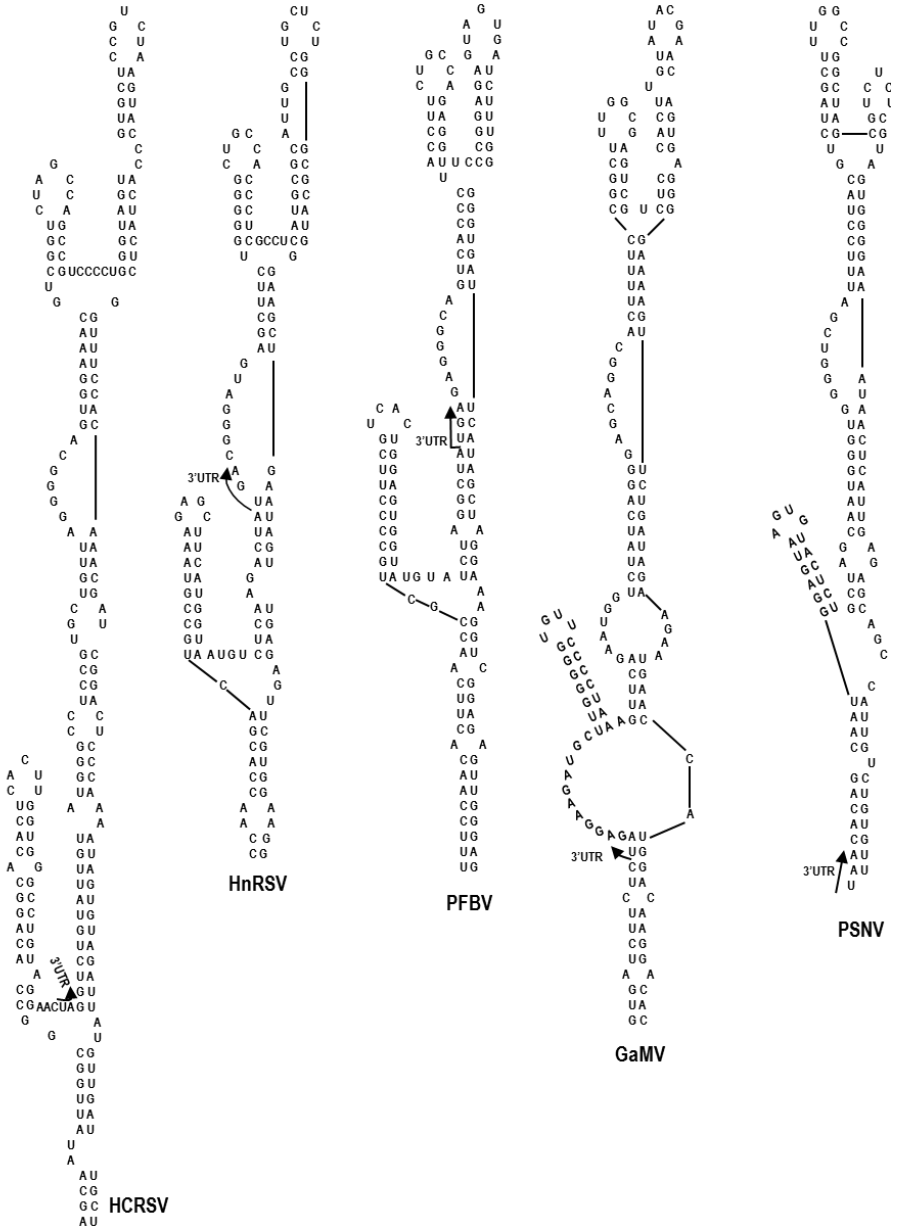


Figure 11: mFold-predicted secondary structures of the PTE region in other carmoviruses

Phylogenetic analysis revealed that the PTEs of PFBV, HnRSV and GaMV have structural configurations similar to the PTE of SCV. All four PTEs protrude from a 3-way junction containing an elongated lower stem and a short 5' side hairpin.

Efficient translation of the PTE does not require that the 5' gRNA interacting sequence be located at a specific position

In SCV, PFBV and HnRSV, the sequences at the 5' end of the gRNA that engage in the kissing-loop interaction with the PTE are similarly positioned 100 to104 nt from the 5' end, which for these viruses lies within their first ORF (Table 4). PSNV, which has an extended 5'UTR compared to other carmoviruses, contains a putative PTE partner sequence 107 nt from the 5' end within its 5'UTR. To determine if efficient translation enhancement requires specific positioning of the SCV PTE-interacting sequence relative to the 5' end of the gRNA, the location of gH3 was altered in luciferase reporter constructs containing the 5' 125 nt and 3' 400-nt of SCV (5'125-Fluc-3'400; Figure 12A). The 5' 125 nt of SCV gRNA contains the 39 nt UTR along with the beginning of the p26 coding region, which includes gH3 (Chapter II) (90). Wt parental and mutant constructs were co-inoculated with a control luciferase (RLuc) construct into Arabidopsis protoplasts, and luciferase activity was measured 18 h later. Deletion of 9 nt (positions 61-69) upstream of a small hairpin (gH2) adjacent to gH3 resulted in enhanced luciferase activity (130% of wt). Deletions of 18 and 27 nt resulted in only moderate reductions in luciferase activity (to 78% and 70% of wt, respectively) (Figure 12B). These results suggest that gH3 can be repositioned at least 27 nt closer to the 5' end in 5'125-Fluc-3'400 without a substantial loss in luciferase activity.

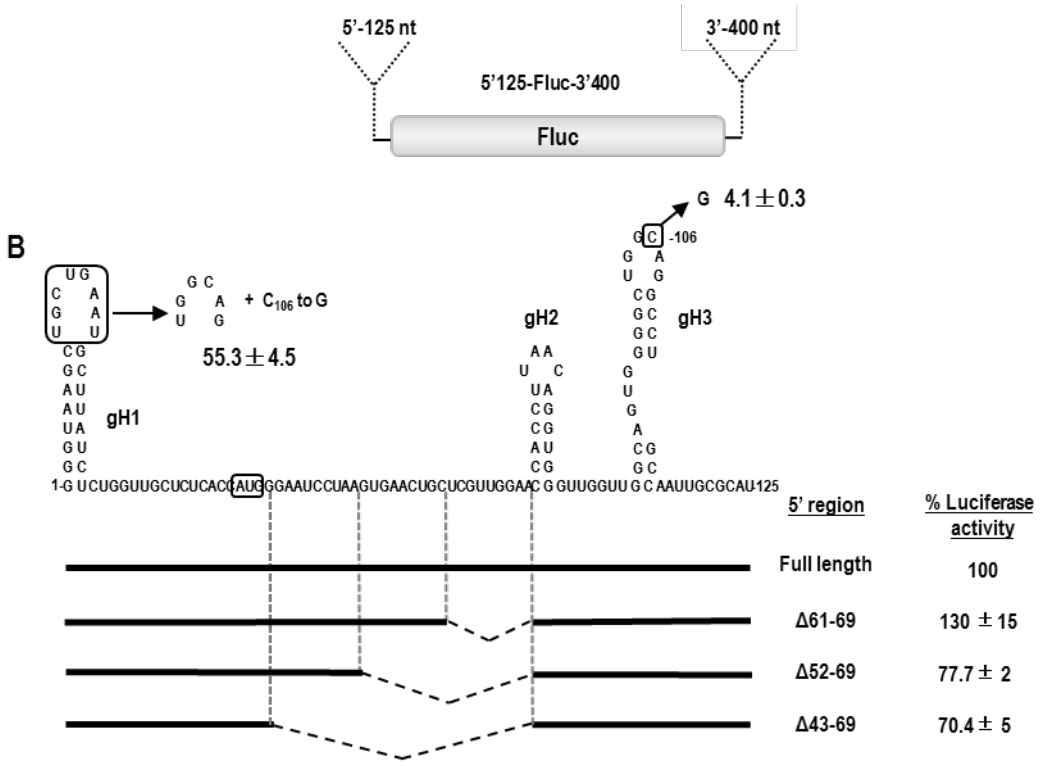


Figure 12: Effect of gH3 location on translation of reporter constructs
A) Parental reporter construct 5'125-Fluc-3'400 contains firefly luciferase (Fluc) flanked by the 5' 125 nt and 3' 400 nt of SCV gRNA.
(B) The position of gH3 was altered by deletions as well as replacement of gH1 loop sequence with that of gH3. The later alteration was combined with a single mutation in the gH3 loop (C_{106} to G) that disrupts the natural kissing-loop interaction. Deletions are denoted by hatched lines. Experimental constructs were assayed in *Arabidopsis* protoplasts co-inoculated with a control renilla luciferase reporter construct. Levels of luciferase activity are presented as a percentage of the activity of the parental construct. Results are from three independent experiments with standard deviations given. The initiation codon of the p26 ORF is boxed. Performed by Micki Khulman.

Most carmovirus gRNAs, including SCV, contain 5' terminal sequences that can fold into a stable hairpin (gH1). In GaMV, HCRSV and CarMV, the PTE-complementary sequence is located in the apical loop of gH1. To investigate the impact on translation when gH1 of SCV is altered to contain the PTE-interacting sequence, the apical loop of gH1 was replaced with the apical loop of gH3 (Figure 12B). In addition, a single point mutation (C_{106} to G) was engineered in the loop of gH3 to disrupt the natural

long-distance RNA:RNA interaction. This single mutation by itself reduced luciferase activity in protoplasts to 4.1% of wt (Figure 12B). When the construct also contained gH1 with the apical loop of gH3, luciferase activity improved 13-fold to 55% of wt. This suggests that moving the interacting sequence to the 5' terminus can support the required RNA:RNA interaction. However translation was reduced by nearly 50% from wt levels, suggesting either that a preference exists for the wt location of the interacting sequence, or that SCV hairpin gH1 plays a role in translation that is disrupted when its apical loop is altered.

gH1 is required for efficient translation of the SCV reporter construct

To determine if gH1 is important for SCV translation, the entire hairpin was deleted, producing construct ΔgH1. Luciferase activity of ΔgH1 was 57% of wt (Figure 13A), suggesting that the hairpin, or 24 nt in this location, improves translation. To determine the importance of the gH1 stem, the stem was disrupted by replacing the 5' side sequence with the 3' side sequence (construct m1). m1 luciferase activity (56% of wt) was similar to that of ΔgH1, suggesting that the hairpin is not acting as a simple spacer element. When compensatory mutations were included in the 3' side of gH1 to restore the stem (construct m2), luciferase activity decreased to 24% of wt (Figure 13A). One possibility for why the restored gH1 stem in construct m2 did not re-establish efficient translation was the inadvertent introduction of an out-of-frame AUG codon into the gH1 stem (Figure 13A). To determine if this upstream AUG reduced translation initiation from the p26 initiation codon, the AUG in the m2 gH1stem was mutated to AUA generating construct $m2_{AUA}$. Eliminating the AUG in m2 increased luciferase activity to 72% of wt, suggesting that the upstream out-of-frame AUG is interfering with

luciferase translation. This prompted an examination of a naturally occurring out-of-frame AUG located in wt gH1 for its effect on translation (Figure 13A, solid box). When the SCV gH1 AUG was altered to UUG (construct $gH1_{UUG}$), luciferase activity increased to 198% of wt, suggesting that the gH1 AUG was negatively impacting translation. However, a second possibility was that the alteration in the loop of gH1affected a function associated with gH1 that was independent of the presence of an initiation codon. To test for this possibility, the apical loop or entire gH1 sequence was replaced with gH1 sequences from carmovirus *Calibrachoa mottle virus* (CbMV), generating constructs $gH1_{CbMVAL}$ and $CbMV_{gH1}$, respectively. CbMV is predicted to contain a TED-like 3'CITE that engages in a long-distance kissing-loop interaction with the apical loop of a coding region hairpin located in a similar ORF position as the SCV PTE-interacting hairpin (110 nt from the 5' end) (5). As with SCV, CbMV also contains a 5' terminal gH1 hairpin but without an AUG (Figure 13B). Our reasoning was that if gH1 or its apical loop are important for translation, then the CbMV gH1 or CbMV gH1 apical loop replacements might provide effective substitutions. $gH1_{CbMVL}$ and $CbMV_{gH1}$ increased translation to 168% and 169% of wt, respectively (Figure 13C). These results suggest that CbMV gH1 can functionally replace SCV gH1 in our translation assays, and enhanced translation of constructs containing $gH1_{UUG}$, $gH1_{CBMVL}$ and $CBMV_{gH1}$ is due to the absence of an upstream AUG.

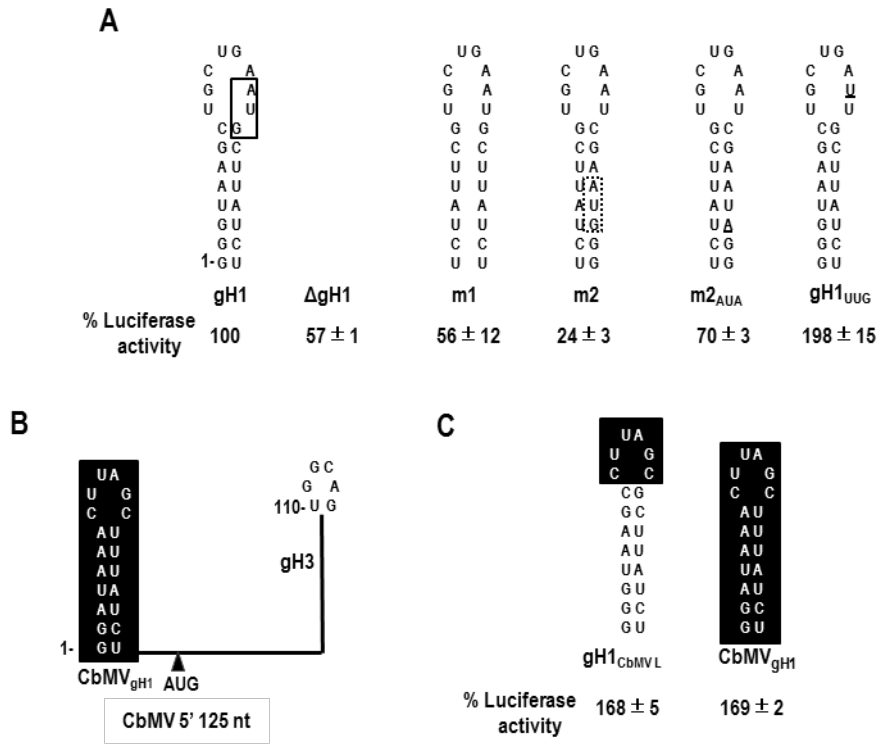


Figure 13: Mutational analysis of gH1
(A) Alterations were generated in 5'125-Fluc-3'400. A natural AUG located upstream and out-of-frame with the initiator AUG is boxed in a solid line in SCV gH1 (left). An out-of-frame AUG that was introduced into the stem in construct m2 when the 3' side sequence was replaced with the 5' side sequence is boxed with a dotted line. This AUG was altered to AU<u>A</u> in construct $m2_{AUA}$. $gH1_{UUG}$ contains a single base change (underlined) that eliminates the AUG in SCV gH1. Luciferase activity values were determined from three independent experiments and are presented as a percentage of the activity of the parental construct.
(B) Schematic representation showing the organization of predicted hairpins in the 5' 125 nt of CbMV. CbMV contains a putative I-shaped 3'CITE capable of forming a kissing-loop interaction with its gH3 that is very similar to the SCV kissing-loop interaction (Simon and Miller, 2013). $CbMV_{gH1}$ does not contain an AUG.
(C) Hairpins that were substituted for SCV gH1 in 5'125-Fluc-3'400. Results are from three independent experiments with standard deviations given. Performed by Micki Khulman and Maitreyi Chattopadhyay.

Since deletion of gH1 (construct ΔgH1) or compensatory mutations in the gH1 stem (construct $m2_{AUA}$) also removed any upstream out-of-frame initiation codons, the levels of luciferase activity produced by these constructs should more accurately be compared with that of $gH1_{UUG}$. Thus ΔgH1 reduced translation to 28% of $gH1_{UUG}$ and $m2_{AUA}$ only restored translation to 36% of $gH1_{UUG}$. In addition, since transferring the

PTE interacting sequence to gH1 (Figure 12) also eliminated the gH1 AUG, the level of luciferase acidity obtained by this construct should also be compared with that of $gH1_{UUG}$. Thus, transferring the RNA:RNA interaction site to the 5' terminal hairpin produced only 25% of the luciferase activity of $gH1_{UUG}$. These results together suggest that (1) SCV gH1 plays a role in translation; (2) the gH1 AUG impacts translation from the in-frame AUG in 5'125-Fluc-3'400; and (3) CbMV gH1, which shares structural but no sequence similarity with gH1 of SCV, can functionally replace SCV gH1.

Translation directed by the sgRNA2 5'UTR can be affected by altering the location of the sgH1

Carmoviral PTE also engage in a long-distance kissing-loop interaction with hairpin sgH1 located in the central region of their sgRNA2 5'UTR ((90); Figure 14A). All sgH1 interacting sequences in PTE-containing carmoviruses are found in the apical loops of diverse hairpins located in the center of the sgRNA2 5'UTR, with the interacting sequences occupying nearly identical positions relative to the 5' end (62 to 65 nt) (Table 4 and Table 5). In addition, the 5' terminal sequences of carmovirus sgRNA2 contain a carmovirus consensus sequence ($G_{2\text{-}3}$ A/U_3) similar or identical to the sequence found at the 5' end of their gRNA (147). To determine whether the location of sgH1 affects translation directed by the sgRNA2 5'UTR, sgH1 was moved closer to the 5' end by sequentially deleting upstream sequences in a parental luciferase construct containing the complete 5'UTR (137 nt) and 3' 400 nt (5'137-Fluc-3'400; Figure 14B). Deletions of 9, 18 or 27 nt (positions 2525-2533, 2516-2533 and 2507-2533) resulted in 80%, 121% and 94% of the wt parental luciferase activity, suggesting that relocating the interacting sequences at least 27 nt closer to the 5' end was not detrimental to translation of the reporter construct *in vivo*. In contrast, deletion of an additional 9 or 18 nt (positions

2489-2533 and 2498-2533) reduced luciferase activity by 4 to 5-fold. Deleting all upstream sequences (thus placing the kissing-loop hairpin at the 5' terminus) resulted in a further decrease in translation to 9% of wt luciferase levels.

The nearly 4-fold reduction in translation that occurred when constructs contained a 36-nt deletion compared with a 27-nt deletion could reflect either a positional requirement for sgH1 or the loss of a required element or both. To help determine if the decrease in luciferase activity was caused by disruption of a required element, a construct was generated that contained only the 9-nt deletion between positions 2498 and 2506. This 9-nt deletion caused a 32% loss in translation activity. Randomizing the entire sequence between the 5' end and sgH1 resulted in a similar loss of luciferase activity (33%). These results suggest that the 4-fold reduction in translation due to deletion of positions 2498-2533 could be due to a combination of effects: loss of an element that contributes to translation as well as an additional effect caused by the 36-nt deletion.

To determine if the 36-nt deletion caused an alteration in the structure of sgH3 (i.e., the deletion resulted in the loss of exposure of the sgH3 apical loop), RNA transcripts derived from parental construct 5'137-Fluc-3'400, and deletion constructs 5'137Δ27-Fluc-3'400 (deletion of positions 2507-2533) and 5'137Δ36-Fluc-3'400 (deletion of positions 2498-2533) were subjected to SHAPE analysis in the vicinity of sgH1. The pattern of residues susceptible to NMIA in sgH1in the wt luciferase construct (Figure 14D, left) was mainly consistent with the hairpin structure previously proposed using in-line probing (90). Residues in the apical loop participating in the kissing-loop interaction were more weakly modified by NMIA than the adjacent loop residues, suggesting that the long-distance interaction was present in a portion of the transcripts.

The bulged cytidylate at position 2540 was flexible according to SHAPE, whereas the U-U non-Watson Crick pair near the base of the stem was inflexible, suggesting that this nucleotide pair is forming non-canonical hydrogen bonds. The residues in the upper portion of the 5' side stem were also weakly modified.

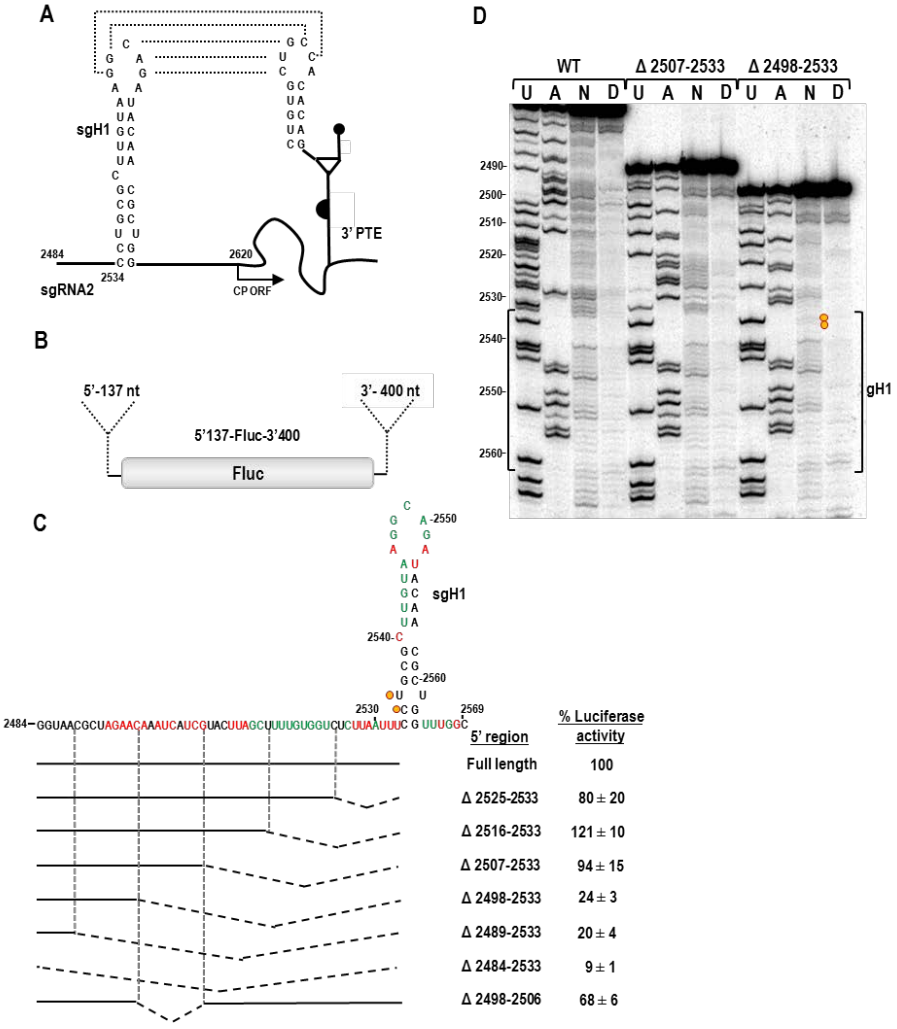


Figure 14: Effect of sgH1 location on translation of reporter constructs and sgH1 structure

(A) The long-distance kissing-loop interaction between sgH1 and the PTE is shown with dotted lines. Start of the CP ORF is indicated.

(B) sgRNA parental reporter construct 5'137-Fluc-3'400 contains Fluc flanked by the 5'UTR (137 nt) of sgRNA2 and the 3' 400 nt.

(C) Effect of deletions on translation of sgH1. Residues of sgH1 with strong or moderate reactivity to NMIA within 5'137-Fluc-3'400 transcripts are denoted by red and green colors, respectively. The position of sgH1 was altered by deleting upstream sequence. Deletions are denoted by hatched lines. Experimental constructs were assayed in Arabidopsis protoplasts co-

inoculated with a renilla luciferase reporter control construct. Levels of luciferase activity are presented as a percentage of the activity of the parental construct. Results are from three independent experiments with standard deviations given.
(D) SHAPE autoradiogram comparing sgH1 structure in parental wt transcripts, and transcripts containing deletions Δ2507-2533 or Δ2489-2533. U and A, ladder lanes; N, NMIA; D, DMSO. Position of sgH1 is bracketed. Orange circles denote residues with reproducible, slightly increased flexibility. Performed by Kalyani Kumar and Maitreyi Chattopadhyay.

The susceptibility to NMIA of sgH1 residues in RNA transcripts of parental construct 5'137-Fluc-3'400 and 5'137Δ27-Fluc-3'400 were indistinguishable (Figure 14D). Residue flexibility of sgH1 in 5'137Δ36-Fluc-3'400 was also very similar with two exceptions. Two residues, the bulged uridylate at position 2536 and the adjacent cytidylate at position 2535, showed a reproducible high level of modification by NMIA (Figure 14D, orange circles). The lack of discernible difference in flexibility of the apical loop suggests that the kissing-loop interaction with the PTE would still be available. Thus an alternative explanation, such as a positional effect, likely accounts for the reduced translational activity of 5'137Δ36-Fluc-3'400.

Discussion

At least four classes of 3'CITEs (PTE, TSS, ISS, TED-like) are associated with viruses in the carmovirus genus, with the PTE being the most common (5). The best studied PTE is the element associated with Umbravirus PEMV (92, 93). However, this PTE is exceptional in not being directly involved in a long-distance interaction. Thus, to better understand how PTE function in general, analyses of more canonical PTE are required. To this end, I examined the conformation of the SCV PTE and surrounding sequences within a full-length gRNA transcript and the consequences of altering the conserved placement of 5' sequences that interact with the PTE to bridge the 5' and 3' ends and circularize the translated RNAs.

The SHAPE-derived structure of the SCV PTE within full-length gRNA (Figure 10B) was consistent with the 3-way branched structure previously predicted by in-line structure probing of a shorter 3' terminal 343-nt fragment (90) and by SHAPE probing of a short fragment centered on the PTE (93). The major difference between the current and previous structures was the inflexibility of the 3H1 apical loop in the full-length structure. This finding strongly suggests that these residues in full-length transcripts are inflexible due to their base-pairing with one of its pairing partners (gH1 and sgH3). In the full-length fragment, the SHAPE flexibility profile of the surrounding sequences was most consistent with the PTE jutting out of a 3-way junction, which also contains a lower stem and short 5' side hairpin. Similar structures in this region were predicted for all PTE-containing carmoviruses, suggesting that this RNA conformation is phylogenetically conserved. However, it is not likely that this overall conformation is required for the PTE function. The PEMV PTE is known to be just downstream of a second 3'CITE, the kl-TSS ribosome-binding translation enhancer, and thus this region in PEMV also does not adopt a similar conformation. However, the PEMV PTE differs from the carmovirus PTE in that it does not participate in a long-distance interaction, a function associated with the adjacent kl-TSS (94).

For several carmovirus gRNAs, the PTE-interacting sequences are found either in the loop of 5' terminal hairpin gH1 or 100 to 107 nt from the 5' end also in the apical loop of hairpin gH3. In sgRNA2, the location of the interacting sequence is even more precise: 62 to 65 nt from the 5' end of sgRNA2 in the apical loop of hairpin sgH1. The PMV 5' sequence that is complementary with its PTE is also located in loop of a 5' terminal hairpin (91) as are the sequences that form kissing-loops with a number of other

3’CITEs in the *Luteoviridae* and *Tombusviridae* (39, 71, 148). Based on the strict positional conservation of 5’ sequences that partner with PTE in the critical long-distance kissing-loop interaction, our original hypothesis was that the position of the 5’ partner sequences relative to the 5’ end of the translated RNA was associated with PTE function as a translation enhancer. Moving SCV gH3 or sgH1 27 nt closer to the 5’ end of their respective reporter transcripts produced either only a modest 30% reduction in luciferase activity (gH3; Figure 12B), or had a negligible effect (sgH1; Figure 14B). This suggests that the strict conservation of PTE-interacting sequences may be either (1) an evolutionary remnant; (2) necessary for a function not directly connected with PTE-mediated translation; or (3) necessary for translation of the viral RNA and not the reporter construct. Interestingly, extending the 27-nt deletion to 36-nt by deleting positions 2498 to 2506 reduced translation mediated by the sgRNA2 5’UTR by 4-fold. Deletion of position 2498 to 2506 alone also reduced translation, but by only 32%, suggesting that the 4-fold reduction in translation resulting from the 36-nt deletion was not due simply to alteration of an important element in the region. The 36-nt deletion only slightly affected the structure near the base of sgH1 and did not alter the weak flexibility of the apical loop residues, indicating that the kissing-loop interaction involving these residues was likely still present. The 4-fold decrease in translation that occurred using the construct containing the 36 nt deletion thus may be a consequence of the altered position of sgH1 relative to the 5’ end.

When the entire sequence upstream of sgH1 was randomized, translation was reduced by a similar level as with the deletion of position 2498 to 2506, suggesting that an element contributing to translation may reside in that location. The 9-nt sequence

contains a portion of a CA-rich sequence (AACAAAUCA) and similar CA-rich motifs are found in 5′ UTRs of both gRNAs and sgRNAs of TCV and tombusvirus *Tomato bushy stunt virus* and in the 5′ UTRs of CarMV and CCFV sgRNA2 (149). Among the PTE-containing carmoviruses, a CA-rich motif that contributes additional length and reduced secondary structure is found in the HCRSV IRES element located in the 5'UTR of sgRNA2 (150). This single-stranded region has been postulated to provide a suitable landing site for ribosome docking with the viral mRNA (149). SHAPE structure probing of 5'137-Fluc-3'400 transcripts also indicated that this sequence is single-stranded (Figure 14B).

All carmoviruses, with the exception of the two viruses with ribosome-binding 3'CITEs (TCV and CCFV), have a predicted hairpin (gH1) at the 5' end of their gRNAs. A subset of the PTE-containing carmoviruses have the kissing-loop sequence within the apical loop of this hairpin, as does TGP-carmo isolate 1, which is predicted to have a TED-like 3'CITE (5). Transposition of the SCV PTE-interacting sequence from its distal location to the loop of gH1 resulted in significantly reduced translation, despite removal of an upstream out-of-frame AUG residing partially within the gH1 loop that reduced translation of the parental reporter construct by ~50% (Figure 13A). This result suggests that gH1 in carmoviruses may have a function in translation other than simply as a potential scaffold for the long-distance interaction. A separate role for SCV gH1 in translation is suggested by the inefficient translation that resulted when the stem of gH1 was disrupted, which was only partially restored by compensatory mutations and only when an inadvertent potential initiation codon within the stem was mutated. Although this result suggests a sequence-specific requirement for proper gH1 function, CbMV gH1

was able to effectively substitute for SCV gH1, despite having only limited sequence similarity. These results for the PTE differ from results obtained using similar reporter constructs and UTR sequences from BYDV luteovirus. The BYDV BTE 3'CITE also engages in a kissing-loop interaction with 5' sequences, which for the gRNA are located 104 nt from the 5' end, and for sgRNA1 are located in a 5'terminal hairpin. In this system, relocating the gRNA 5' interacting sequences to a 5'terminal hairpin had only a limited (26%) decrease in translation efficiency (80).

It has been proposed that the kissing-loop interactions that connect 3'CITEs to the 5' ends of their translated mRNAs function more efficiently when the 5' sequences are located proximal to the 5' end. This proximity is proposed to assist ribosomes in using the translation factor-bound 3'CITEs to dock with the template and finding the 5' terminus of the mRNA prior to scanning to the initiation codon (80). This hypothesis was based partially on finding that increasing the distance of the kissing-loop interacting sequence relative to the 5' end in BYDV reduced translation efficiency. Our results for the SCV PTE kissing-loop interaction do not currently support this hypothesis. With one exception (9-nt deletion in the gRNA), all deletions that repositioned the kissing-loop sequence closer to the 5' end had either a limited effect on translation or a strong negative effect on translation. In addition, sgRNA2 in PTE-containing carmoviruses must express the CP at high levels despite having the interacting sequence at least 63-nt distal to the 5' end. Movement of sgRNA2 sgH1 to the 5' end by deleting intervening sequences (which when randomized only reduced translation by 33%) reduced translation by over 10-fold (Figure 14C). Furthermore carmoviruses that have the gRNA kissing-loop sequence at the 5' end in gH1 or at least 100 nt distal to the 5' end in a downstream

hairpin likely require synthesis of comparable levels of their replication proteins. It seems likely that additional factors also dictate the evolutionary placement of the 5' kissing-loop interacting sequences for 3'CITEs.

Over the course of the current study, I tried to develop a SCV replication system. The goal was to assess the impact of mutations on viral replication and translation using the full-length virus. Previous attempts to develop an SCV replication system using Arabidopsis protoplasts and detection of viral RNA via northern blot analysis were not successful. Therefore, I attempted to detect SCV viral accumulation using a more sensitive assay: quantitative PCR. This method utilizes polymerase chain reaction chemistry to amplify viral RNA so that high enough concentrations of viral particles can be detected and quantified by fluorescence. *In vitro* transcribed full-length SCV was transfected into Arabidopsis protoplasts and cells were incubated at 25ºC for up to 48 h. Cells were harvested every 12 h. Total RNA was extracted and reverse transcribed into cDNA using M-MuLV Reverse Transcriptase (NEB) following manufacturer's instructions. The resultant cDNA was subjected to Taqman real time PCR assay using a LightCyler 480 real time PCR machine. As a negative control, an SCV full-length virus with a mutation in GDD domain of its RdRp was used. The GDD domain is responsible for coordinating the metal ions and nucleotide substrate selection during catalysis and its mutation renders the RdRp inactive (151-153). The levels of accumulation of wt SCV (1% to 20%) was similar to that of GDD mutant at each time point tested (Figure 15) suggesting that SCV does not accumulate in Arabidopsis protoplasts. Arabidopsis protoplasts have been shown to be infected by many different plant viruses, such as TCV, a carmovirus and PEMV, an umbravirus. Since the natural host of SCV is a saguaro

cactus, which is found in the desert, it is possible that the replication of SCV requires a higher temperature than 25ºC.

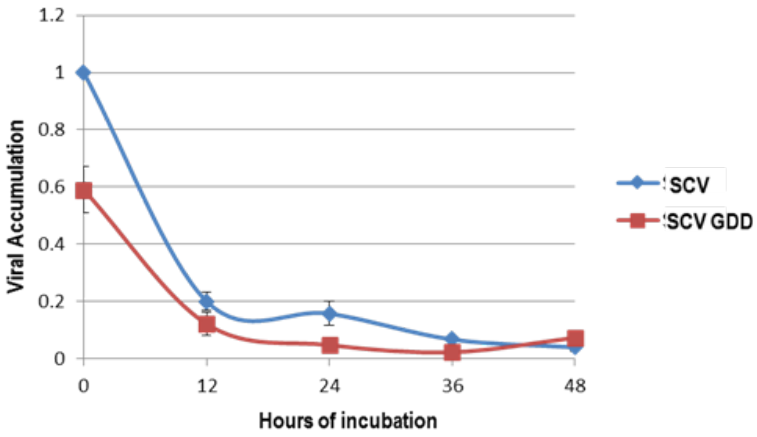


Figure 15: SCV replication in Arabidopsis protoplasts
Arabidopsis protoplasts (Col-0) were inoculated with full-length transcripts of SCV gRNA and GDD mutant and levels of gRNA accumulation was determined by quantitative real PCR every 12 h for 48 h. Values are from independent experiments conducted in triplicate.

Chapter 4: Second-Site mutations require host RNA silencing components and the virus silencing suppressor to compensate for structural defects in the 3'UTR

Introduction

The genomes of (+)-strand RNA viruses have evolved to fold into complex higher-order structures via multiple long- and short-range RNA:RNA interactions that facilitate basic functions including translation, replication, and evasion of host defenses (12, 23). Recent studies have suggested that the folded state of highly structured viral genomes can hinder the effectiveness of the host RNA silencing pathway (26, 33-35).

TCV is a small, (+)-strand RNA virus with five overlapping ORFs. The 3' proximal ORF, p38, encodes a multifunctional protein (Figure 2). In addition to its role in capsid formation, p38 functions as a silencing suppressor and targets multiple components in the host RNA silencing pathway (117, 154) that include binding to dsRNA of different sizes (155, 156) and interaction with Argonaut1 (AGO1) and AGO2 (120, 123, 157). AGO1 and AGO2 play major roles in RNA silencing processes as vital catalytic components of RISC, and function as major effectors of antiviral defense (121-123).

It has become clear that multiple short- and medium-range RNA:RNA interactions between various *cis*-acting structural elements make up the higher order structure of the TCV 3' end facilitating many basic functions of the virus in host cells (136), see Figure 2. The only known long-distance interaction in the TCV genome is between the apical loop of a 3' terminal hairpin, Pr, and a bulge loop of a hairpin (RSE).

This interaction modulates production of the TCV RdRp (29). Among the different elements in the 3' end, H4 plays a critical yet enigmatic role in TCV gRNA accumulation, and occupies a central role in a network of structural interactions connecting all 3' end elements (89, 131) (Figure 2). A 3 nt alteration in H4AL (UUA to ACU), known as TCV-m21, reduced TCV gRNA accumulation in protoplasts by ~80% (131), and reduced *in vivo* translation of a luciferase reporter construct containing the TCV 5'UTR and 400 3' residues by 77% (89). To determine how H4 participates in the structure and function of the gRNA, TCV containing important H4AL mutation, TCV-m21, was subjected to a genetic second-site mutation analysis to reveal any specific H4AL-interacting sequences within the 5' and 3' regions of the virus (131).

Turnip seedlings were inoculated with TCV-m21 and viral progeny were recovered after three passages. Fragments 5' terminal 1564 nt and the 3' terminal 900 nt were cloned from the progeny and examined for second-site mutations. As previously described (131), no second-site mutations were recovered in clones containing the 5' 1564 nt. Interestingly, the m21 primary-mutation site was partially reverted to "rev1" or "rev2" in all clones recovered containing 3' 900 nt and second-site mutations were found in the 3'UTR and upstream p38 ORF. Second-site mutations recovered in the 3'UTR included single residue changes in the stem of hairpin H4b, a single change in the loop of H4TL and three bases altered in the Pr loop (UCG→AA) (131). None of these locations contained sequence capable of canonical base-pairing with H4AL (131).

In this chapter, I investigated some selected second-site mutations in the p38 ORF, all of which were associated with rev1, the partial revertant of TCV-m21. TCV-rev1 differs from wtTCV at a single position in the H4AL sequence (A3899U) (Figure

16A and B). The second-site mutations examined were single missense mutations recovered from separate plants (G3561A, G→E; and U3329C, F→L). Another was a set of three second-site mutations present in the same clone (A3475G/A3709G/U3741C), which also produced an altered p38 (U3741C: V→A). Here I report that: 1) Second-site mutations in the p38 ORF reside in a separate upstream RNA domain (Domain2) from the primary-site mutation, which is in Domain1; 2) Domain1 and Domain2 are connected via RNA: RNA interactions; 3) The second-site mutations were compensatory only in the presence of p38, the TCV silencing suppressor, and in the presence of functional DCLs and RISC. Our analysis provides an unexpected connection between 3'UTR primary-site mutations and the RNA silencing machinery in TCV.

Materials and methods

Generation of constructs

Oligonucleotide-mediated site-directed mutagenesis was used to generate mutations in full-length TCV using parental constructs pTCV66, TCV-m21, TCV-rev1 and TCV CP-. PCR was performed using Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB) according to the manufacturer's protocol. PCR products were digested with Dpn1 for one hour at 37ºC, followed by transformation of DH5α cells. All mutations were subjected to regional sequencing to confirm the desired alteration. For agrobacterium infiltration, binary construct PZP-TCVp38 containing mutation G3561A or U3329C in the p38 ORF within vector pRTL2 was used (117). PZP-TCVp38 contains a 35S promoter, TEV translational enhancer, p38 ORF insert, and 35S terminator. Oligonucleotide-mediated site-directed mutagenesis was used to generate mutations.

The expression cassette was excised from the vector using Pst1 digestion and ligated into binary vector pPZP212 to produce constructs $p38_{G3561A}$ and $p38_{U3329C}$.

Isolation of second-site mutations

TCV-m21, containing a 3-nt alteration in H4AL (3897UUA to ACU) was used to generate the second-site mutations. Plants (Turnip cv Just Right) at the two true leaf stage were mechanically inoculated with *in vitro*-transcribed TCV-m21 RNA (2 μg for each of two leaves), as described previously (158). Total RNA was extracted at 21-days post inoculation (dpi) and used to re-inoculate seedlings (5 μg for each of the two leaves). The process was repeated a total of three times. Total RNA isolated from the third passage was used for RT-PCR amplification of fragments corresponding to the 5' 1564 nt or 3' 900 nt, which were cloned and subjected to sequencing. This was conducted by Xuefeng Yuan.

Protoplast preparation, inoculation and RNA gel blots

Please see Chapter 2.

SHAPE structure probing

Please See Chapter 3.

Agrobacterium infiltration

Agrobacterium infiltration was performed as described previously (117). Briefly, pZP212 binary vectors containing desired mutations in the p38 were introduced into agrobacterium strain C58C1 by electroporation. The transformed agrobacterium were grown on LB plates containing 50 μg/ml rifampicin, 50 μg/ml gentamycin and 100 μg/ml spectinomycin at 28^{o}C for 48 h. Individual colonies were grown at 28^{o}C overnight in 3 ml LB culture containing the above mentioned antibiotics. Individual agrobacterium

cultures carrying various constructs were pelleted at 5,000 rpm for 15 min and resuspended in inoculation buffer (10 mM morpholinepropanesulfonic acid (pH 5.5), 10 mM MgCl2, and 100 μM acetosyringone) to an optical density of 1.0 at 600 nm. Suspensions were incubated at room temperature for 2 hours and were mixed in equal volumes for co-infiltrations. The *Nicotiana benthamiana* plants constitutively expressing GFP (line 16c) were grown at 22^{o}C with a 12/12 h day/night cycle. Infiltration was carried out at the three-week-old stage with a 1ml needleless syringe. GFP fluorescence was observed under long wavelength ultraviolet lamp (Black Ray model B 100AP) and photographed using a Nikon D90 digital camera.

Detection of GFP

Total RNA from the infiltrated leaves was isolated by using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Infiltrated leaves were grounded into a fine powder under liquid nitrogen and were dissolved in 1 ml TRIzol reagent per 50-100 mg of tissue sample. The homogenized samples were incubated for 5 min at room temperature followed by addition of 200 μl of chloroform per 1 ml of TRIzol reagent. Tubes were shacked vigorously and incubated for 5 min at room temperature. The aqueous phase was collected after centrifugation at 13,000 rpm for 15 min at 4^{o}C. RNA precipitation was conducted in the presence of 1/10 volume of 3M NaOAc (pH 5.3) and 2.5 volume of 100% of ethanol at -80^{o}C for 30 min, followed by centrifugation at 13,000 rpm for 30 min at 4^{o}C. The pellet was washed with 70% ethanol, air dried and resuspended in 100 μl of sterile distilled water. RNA was transferred to a nitrocellulose membrane. The membrane was hybridized with [α-32P]-dCTP labeled DNA probes corresponding to the full-length ORF of GFP.

Protein gels and immunoblotting

Total protein was isolated from the saved interphase and organic phenol-chloroform phase by following the manufacture's recommendation. About 1.5 ml of isopropanol per 1 ml TRIzol were added and incubated for 10 min at room temperature. Total Protein was pelleted down by centrifugation at 12,000 rpm for 15 min at 4^{o}C. After three washes with wash solution (0.3 M guanidine hydrochloride in 95% ethanol), the pellet was air dried and resuspended in solution containing 1% SDS.

After quantifying total protein using the BCA Protein Assay Kit (Pierce), 3µg of protein were separated on 12% SDS-PAGE gel. Proteins were transferred onto Immunobilon-P transfer membranes (Millipore) at 10V for 30 min using a TransferBlot Semi-dry Transfer Cell (BioRad). For Western analysis, the membrane was wet in MeOH for 1 min, followed by treatment with Tris-Buffered Saline (TBST, 20 mM Tris-HCl, and 140 mM NaCl, 0.1% Tween (20 v/v, pH 7.5) containing 5% nonfat dry milk. After two 15 min washes with TBST, the membrane was incubated in 0.1% nonfat dry milk/TBST and 1:1500 diluted anti-CP antibody. The membrane was washed 3 times with TBST and then incubated in 0.1% nonfat dry milk/TBST and 1:30,000 diluted secondary antibody. The membrane was washed for 15 min with TBST and repeated three times. Super Signal West Pico Chemiluminescent Substrate Kit (Pierce) was used for chemiluminescent staining according to the manufacturer's instructions. The membrane incubated with the substrate was covered with plastic wrap and exposed to an X-ray film for 30 sec before developing the film.

Table 8: TCV constructs used in Chapter 4

Name	Description
pTCV66	Full-length TCV in pUC19
U3329C	U to C change at position 3329
G3561A	G to A change at position 3561A
A3475G	A to C change at position 3475
U3741C	U to C change at position 3741
A3475G/A3709G/U3741C	A to G change at position 3475, A to G change at position 3709 and U to C change at position 3741
m21	UUA to ACU change in the H4AL
rev1	UUA to UUU change in the H4AL
TCV CP-	R to stop change at position of 6th amino acid in the p38 ORF
TCV-CP/G3561A	G to E amino acid change at position 273 in the TCV p38 ORF
TCV-CP/U3329C	F to L amino acid change at position 196 in the TCV p38 ORF

Table 9: Oligonucleotides used in Chapter 4

Name	Position	Sequence	Polarity
U3329C	3329	GATTCTACTGACCGCCTTGTGGCGGATGG	+
U3329C	3329	CCATCCGCCACAAGGCGGTCAGTAGAATC	-
G3561A	3561	CAAGGGGACAGCTGAGTGGGAGCACGA	+
G3561A	3561	TCGTGCTCCCACTCAGCTGTCCCCTTG	-
Domain2 3130	3130-3157	GTACCTAAATCTGAGTGACGTGAATCGG	-
Domain2 3327	3327-3348	CTGATACCATCCGCCACAAAGC	-
Domain2 3395	3395-3414	GGCCAAGGAGCCAATGATGC	-
Domain2 3496	3496-3518	CGTCCTTAACACCTGCGAAGTCC	-
Domain2 3628	3628-3648	TTTTCTAGCCCCGAGACCGGC	-
Domain2 3707	3707-3730	CCTTTCTGCTACCTTCACTCCTGC	-
Domain 1 3871	3871-3894	CAGGTCAAAATAAAGCGACCTGGG	-
Domain1 3931	3931-3951	GTTTTCCAGTCTAATGCCCGCA	-
Domain1 3869	3869-3883	ATAAAGCGACCTGGGGG	-
Domain1 4035	4035-4054	GGGCAGGCCCCCCCCCCGCG	-

Underlined letters indicate mutated bases; Italic letters are restriction enzyme sites;"+" and "-"indicate polarity.

Results

Second-site mutations in the p38 ORF along with TCV-rev1 can partially compensate for reduced accumulation of TCV-m21

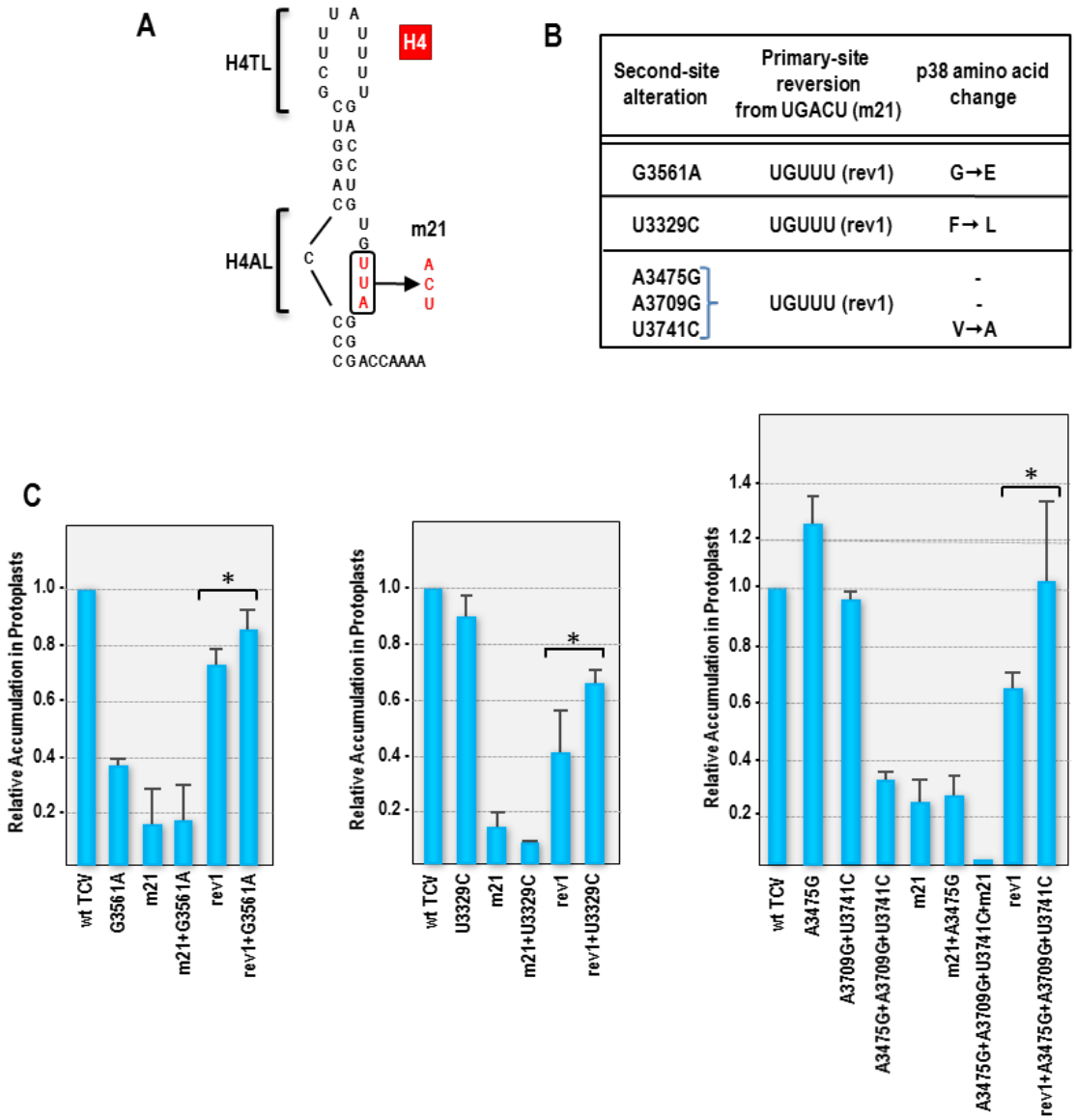
To determine whether any of these second-site alterations were compensatory in conjunction with the initial m21 alteration or the recovered rev1 variant, each second-site mutation was introduced into three full-length gRNA backbones and assayed for accumulation in *Arabidopsis thaliana* protoplasts (Figure 16C). These backbones were: 1) wtTCV, to determine if the second-site mutations by themselves affected TCV accumulation; 2) TCV-m21, to assess if the second-site changes improved the accumulation defect of m21 and thus may have arisen prior to the m21 partial reversion to rev1; and 3) TCV-rev1. Note that TCV-rev1 accumulation was quite variable in these experiments (performed by Maitreyi and Vera using the same starting material), although results were consistent for the individuals performing the experiments. This is highly unusual for TCV mutations and cannot currently be explained. Thus, for each second-site mutation, all constructs, including wtTCV, TCV-m21 and TCV-rev1, with and without second-site mutations, were assayed at the same time by the same individual. At least three independent experiments were performed for each alteration before final averaged values were obtained.

TCV gRNA that includes G3561A (TCV-G3561A) accumulated to 38% of wtTCV (Figure 16C, left panel), suggesting that the mutation negatively impacts either p38 function or gRNA structure. TCV-m21 accumulated to 17% of wtTCV and addition of G3561A did not appreciably improve accumulation. TCV-rev1 accumulated to 75% of wtTCV, indicating that rev1 partially reverses the defect caused by the m21 mutations

(Figure 16C, left panel). TCV gRNA containing both rev1 and G3561A (TCV-rev1+G3561A) improved accumulation of TCV-G3561A by over 2-fold to 88% of wtTCV levels. These results suggest that G3561A did not arise prior to the m21 partial reversion to rev1 and that the combination of rev1 (A3899U) and G3561A are synergistic for improved accumulation of TCV gRNA.

Second-site mutation U3329C had only a slight effect on TCV gRNA levels, with TCV-U3329C accumulating to 88% of wtTCV (Figure 16C, middle panel). As with G3561A, U3329C had no beneficial effect on accumulation of TCV-m21. TCV-rev1, in this set of experiments, accumulated to an average of 42% of wtTCV, and addition of U3329C enhanced accumulation by 1.6-fold. These results suggest that, as with G3561A, U3329C likely originated in the TCV-rev1 background and was able to partially compensate for the rev1 (A3899U) alteration.

The second-site mutation set A3475G/A3709G/U3741C was evaluated in wtTCV for A3475C alone, A3709G and U3741C together, and all three alterations combined. TCV+A3475C accumulated slightly better than wtTCV (125%), whereas TCV containing both A3709G and U3741C accumulated to a similar level as wtTCV. gRNA containing all three second-site alterations accumulated to only 30% of wtTCV (Figure 16C, right panel), suggesting that, all together these mutations are detrimental for gRNA accumulation. None of the mutation combinations improved accumulation of TCV-m21. TCV-rev1 accumulated to an average of 65% of wtTCV in these experiments, and addition of all three second-site mutations enhanced accumulation to wtTCV levels. This suggests that one or more of the alterations in this second-site mutation set is compensatory with rev1 (Figure 16C, right panel).



Second-site alteration	Primary-site reversion from UGACU (m21)	p38 amino acid change
G3561A	UGUUU (rev1)	G→E
U3329C	UGUUU (rev1)	F→ L
A3475G	UGUUU (rev1)	-
A3709G		-
U3741C		V→A

Figure 16: Second-site mutations located in the p38 ORF upstream of H4 are compensatory for virus accumulation

(A) Primary mutations in the H4AL that are the subject of the current study are.

(B) Selected second-site mutations, their effect on the sequence of the p38, and association with primary-site partial reversion (rev1) are shown.

(C) Effect of second-site mutations on accumulation of wtTCV, TCV-m21, and TCV-rev1 are shown. G3561A, left; U3329C, middle; A3475G, A3709G and U3741C, right. Arabidopsis protoplasts were inoculated with full-length transcripts and levels of gRNA determined by RNA gel blots at 40 hpi. Values are from independent experiments conducted in triplicate and standard deviations are shown. *=p<0.05, **=p<0.01, Student's t-test. G3561A data was generated by Vera Stupina.

Second-site mutations are located within a discrete RNA domain

We previously found that second-site alterations in the Pr loop that were isolated together with rev1 affected the structure of the 3'UTR (131). In addition, several second-site mutations in the 3'UTR that compensated for primary mutations in H4TL were compensatory for virus accumulation and eliminated RNA structural changes caused by the primary-site mutations (131). To determine if G3561A and U3329C affect local (p38 ORF) RNA structure and/or structure within the 3'UTR, the secondary structure of the p38 ORF was mapped using SHAPE structure probing combined with mFold computational predictions (139). The location of all flexible residues in the secondary structure of the wtTCV p38 ORF region is presented in Figure 17A. Whereas most of the flexible residues mapped to single-stranded terminal loops, bulged loops, and base-pairs next to loops, a notable exception was the stem flanking the 3' base of hairpin H2-2 (Stem1). Stem1 had a number of reactive bases mainly on one side of the stem that likely denote instability of a poorly stable structure (7/10 G:U or A:U pairs and a central A-A pair) and/or participation of that region in a higher order structure that cannot be currently be ascertained. G3561A is located near the base of H2-2; U3329C is located in one of the bulge loops of H2-4 near two other second-site changes; A3475G is within H2-3 and the two companion mutations are located downstream of this region (not shown, not done by Maitreyi). Eight of 11 second-site changes in progeny of TCV-m21 were predicted to weaken hairpin stems and none are in sequences that might canonically pair with H4AL. Thus, the compensatory nature of these second-site mutations likely involves a mechanism that does not require direct canonical base pairing with the rev1 H4AL sequence.

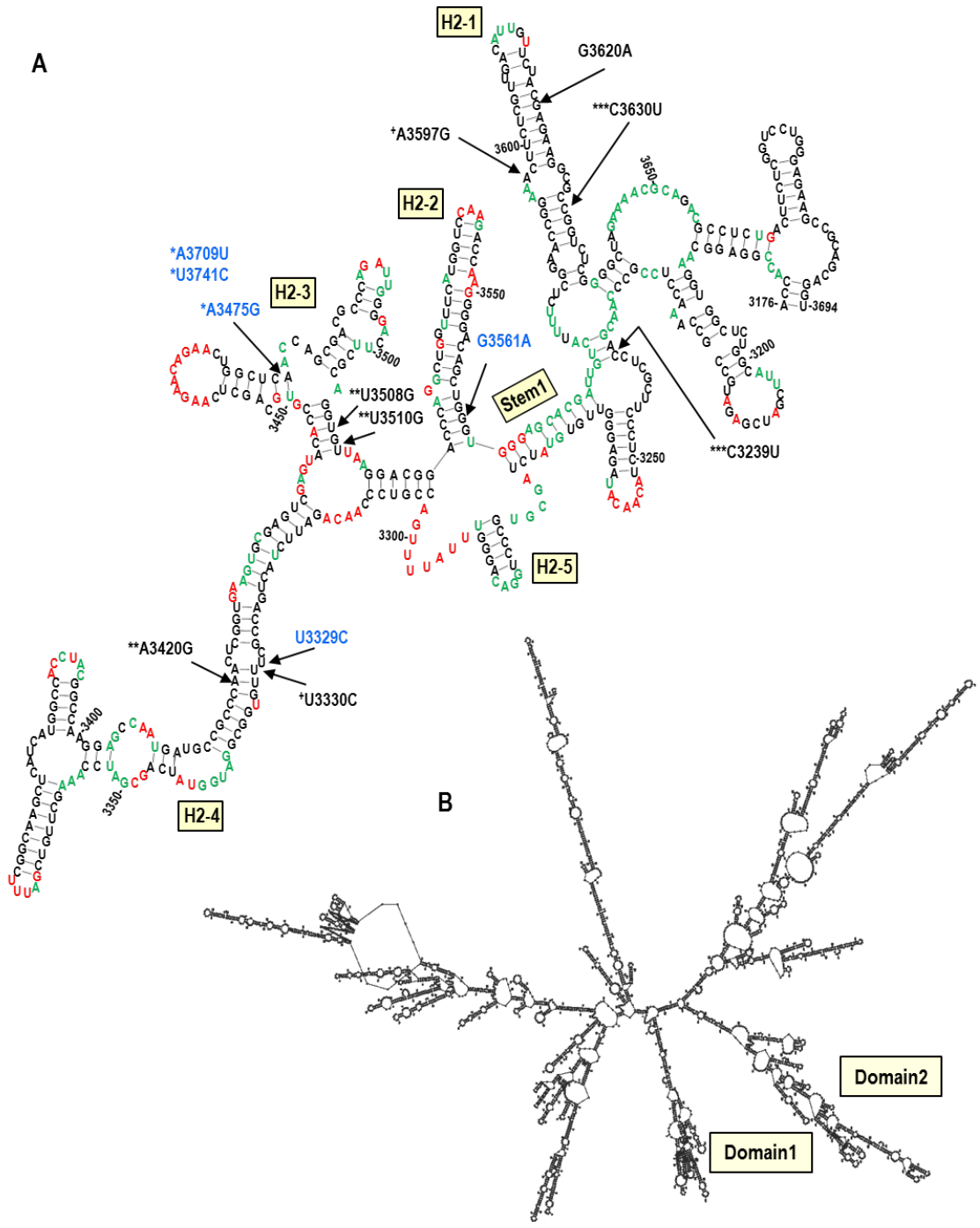


Figure 17: Location of second-site mutations in the structure of Domain2

(A) SHAPE-derived structure that includes most of the p38 ORF (2744-3799). Residues corresponding to high and low reactivity to NMIA are denoted by red and green colors, respectively. Names of hairpins and location of second-site mutations are indicated. Identical symbols denote association in the same clone. Second-site mutations investigated in this study are in blue.

(B) mFold-generated secondary structure of full-length TCV gRNA. Structural predictions used constraints according to SHAPE-derived flexibility information and included placing Stem1 at

the base of hairpin H2-2. Locations of Domain1 and Domain2 are shown. Performed by Micki Khulman and Christine R Szarko.

Secondary structure predictions of full-length TCV gRNA by mFold, which take into account the SHAPE data generated for the p38 ORF region, suggest that the genome is organized into multiple domains emanating from a central backbone, similar to the SHAPE-determined genome structure of TBSV and STMV (24, 25) (Figure 17B). This structure prediction places most of the p38 ORF into a domain (Domain2) that is separate from the domain (Domain1) containing the 3' region of TCV gRNA and known 3' hairpins including H4, the site of the primary mutations. Thus any direct RNA linkage between H4 and the second-site mutations would need to involve long-distance tertiary interactions.

Effect of second-site mutations on the RNA structure of Domain2

The effect of second-site mutations G3561A and U3329C on the structure of Domain2 was examined by SHAPE analysis of full-length TCV-G3561A and TCV-U3329C. G3561A, located near the base of hairpin H2-2, enhanced the flexibility of residues on both sides of the lower stem, supporting the existence of the lower H2-2 stem (Figure 18D, right). However, G3561A also enhanced residue flexibility only on the 5' side of the upper H2-2 stem, suggesting that the upper portion of the hairpin adopts a higher-order structure different from the structure shown in Figure 18D. G3561A also affected residue flexibility in the adjacent poorly organized Stem1 sequence, with 3564GGG showing reduced flexibility and A3567 becoming more flexible (Figure 18D).

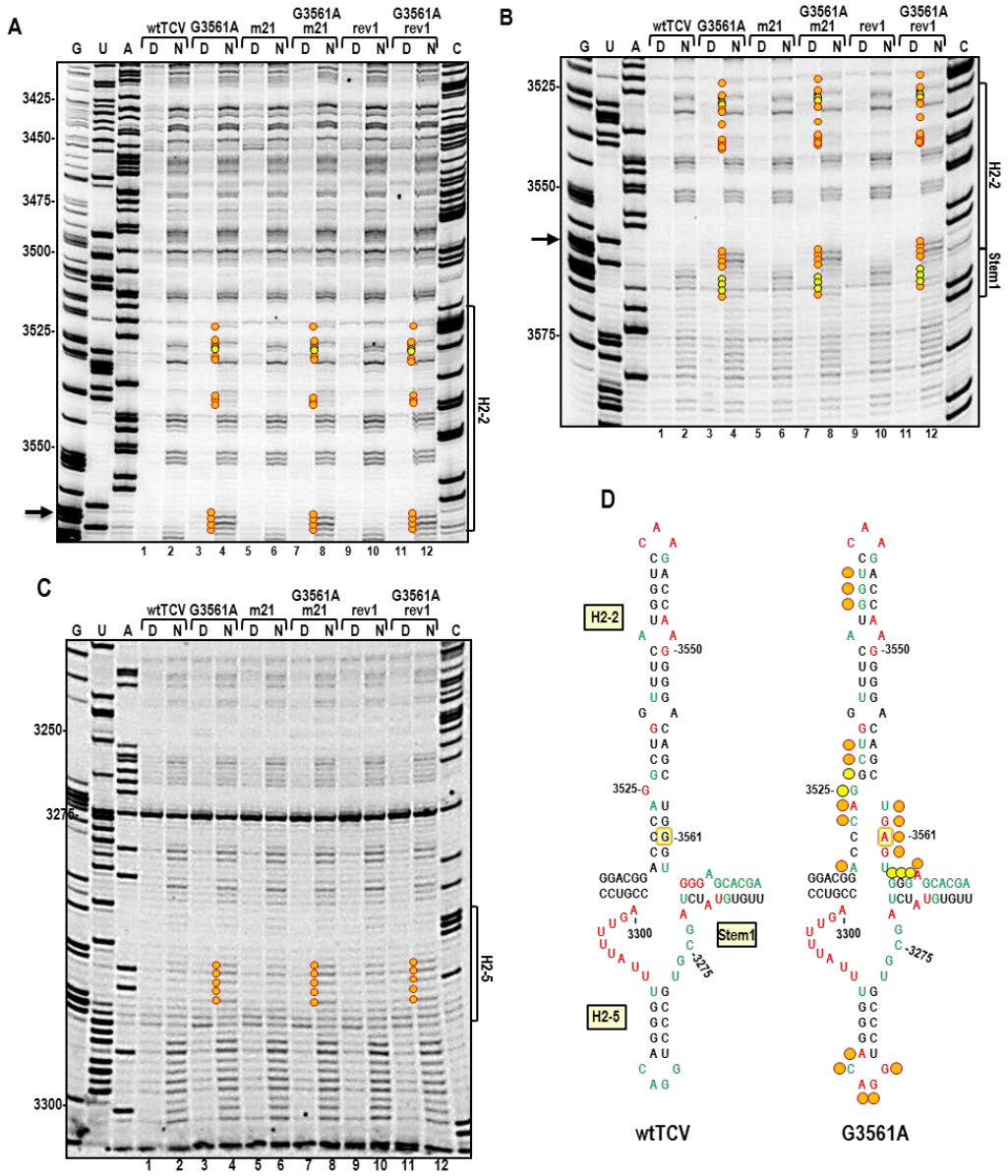


Figure 18: Effect of G3561A on the structure of resident hairpin H2-2 and other proximal regions

(A) SHAPE autoradiogram showing the effect of Domain1 and Domain2 mutations on the structure of the RNA in the vicinity of H2-2. RNAs subjected to SHAPE are designated above the lanes. A,G,C,U denote ladder lanes; N, NMIA; D, DMSO. Arrow denotes position of the G3561. Orange circles and yellow circles denote residues with increased and reduced flexibility, respectively, compared with wtTCV. No other changes in the flexibility of residues were found within Domain2 (data not shown). Performed by Micki Khulman and Christine R. Szarko.

(B) Shorter run of the samples shown in A.

(C) Region of hairpin H2-5 showing additional effects of G3561A.

(D) Mapping the flexibility of residues in wtTCV and TCV-G3561A. Strong and weak reactivity to NIMA is denoted by red and green colors, respectively. Only changes that were consistent in repeated experiments are denoted by orange (enhanced flexibility) and yellow (reduced flexibility) circles. G3561 and G3561A are boxed. Performed by Micki Khulman and Christine R. Szarko.

These changes in residue flexibility produced a profile that was more consistent with the predicted secondary structure of the stem, suggesting that G3561A is also disrupting higher-order structure in the region. No discernible differences were found within H2-3 and H2-4 (data not shown, not done by Maitreyi). However, G3561A enhanced the flexibility of residues in the loop of hairpin H2-5, which is predicted to be spatially proximal to H2-2 within Domain2 (Figure 18D). Enhanced flexibility in the base and upper 5' side of H2-2, adjacent Stem 1 and upstream H2-5 due to a single nucleotide change near the base of H2-2 suggests that tertiary interactions may connect H2-2, H2-5 and Stem1. Domain2 SHAPE structure profiles were also determined for TCV-m21, TCV-rev1, TCV-m21+G3561A and TCV-rev1+G3561A (Figure 18A, B and C). No changes in flexibility compared with wtTCV were discernable for TCV-m21 and TCV-rev1, and residue flexibility was very similar for all gRNAs containing G3561A (i.e., TCV-G3561A; TCV-m21+G3561A; TCV-rev1+G3561A).

The Domain2 structure profile was also determined for second-site mutation U3329C. Although U3329 is predicted to reside in a small internal loop, U3329 is not flexible as assayed by SHAPE in wtTCV gRNA. U3329C caused flexibility changes in three nearby residues on the 5'side of the H2-4 lower stem (G3323, C3325 and U3332; Figure 19C) and two spatially-proximal residues on the opposite side of the stem (A3420, A3421) (Figure 19C), supporting the existence of this stem. This suggests either that U3329 is involved in local interactions in the mid-portion of the H2-4 stem, and/or that

U3329C causes new interactions in the region. Residue flexibility was very similar for all gRNAs containing U3329C (TCV-U3329C, TCV-m21+U3329C and TCV-rev1+U229C). No other structural changes were found within Domain2. Altogether, these results suggest that the compensatory effects of G3561A and U3329C in combination with rev1 cannot be explained by any discernable common structural changes in Domain2.

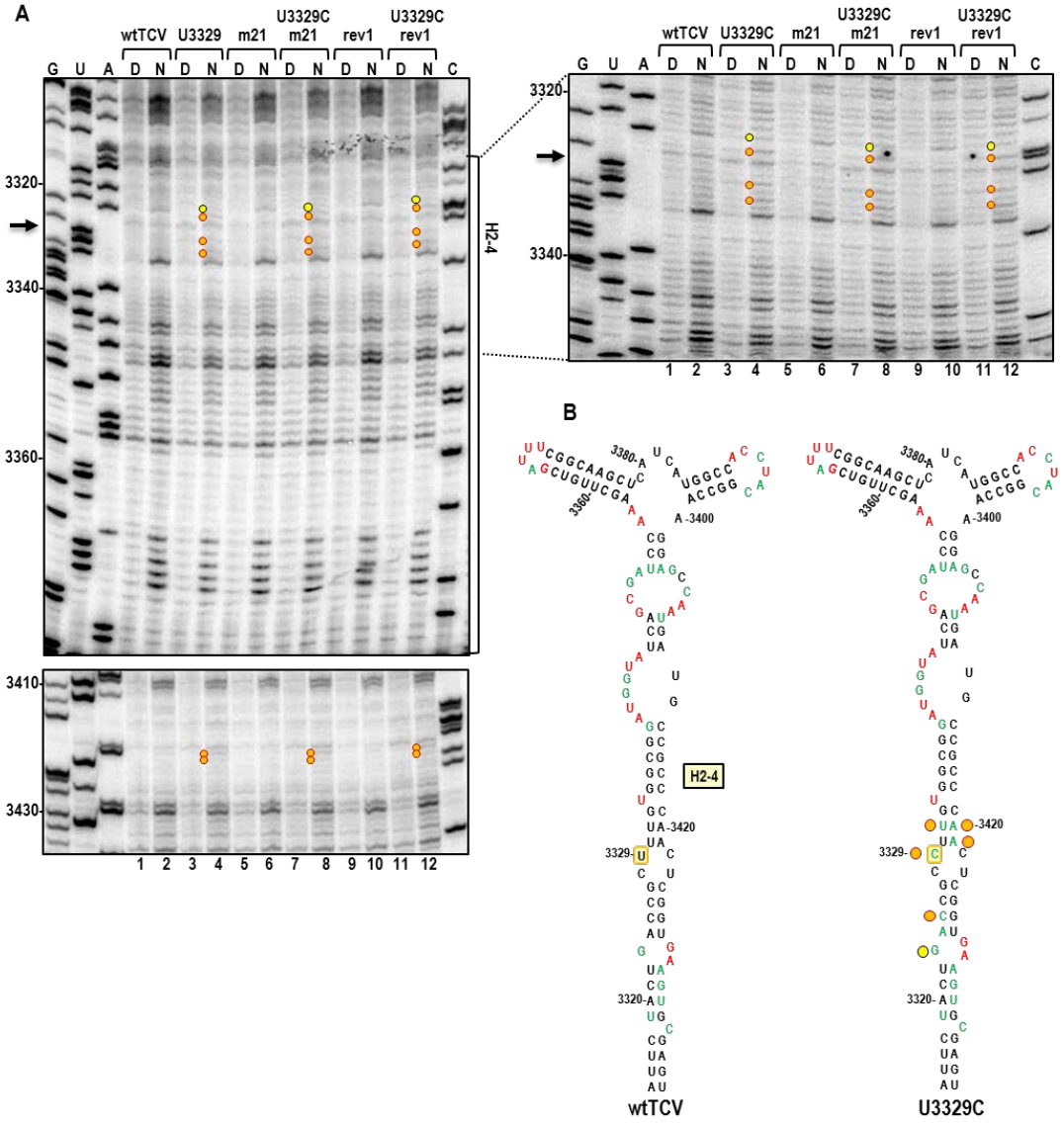


Figure 19: Effect of U3329C on the structure of resident hairpin H2-4

(A) SHAPE autoradiogram showing structural changes in the vicinity of H2-4 in response to U3329C with and without m21 or rev1. Arrow denotes location of U3329C. RNAs subjected to SHAPE are designated above the lanes. See legend to Figure 18.
(B) Comparison of residue flexibility in H2-4 between wtTCV and TCV-U3329C. Strong and weak reactivity to NIMA is denoted by red and green colors, respectively. Only changes that were consistent in repeated experiments are denoted by orange (enhanced flexibility) and yellow (reduced flexibility) circles. No other structural changes were found within Domain2.

G3561A affects the structure of the lower stem of hairpin H4

The structures of fragments containing portions of Domain1 were previously mapped by in-line probing, a technique that evaluates the flexibility of individual residues within short fragments (88, 136). In-line probing of the 3' region of Domain1 produced nearly identical results when using fragments that either began just upstream of H4, included the entire 3'UTR, or began just upstream of hairpin H3 within the p38 ORF (13, 141). For the current study, SHAPE was used to determine the structure of the 3' end region of Domain1 within the full-length gRNA. The majority of Domain1 residues were similarly flexible using either SHAPE or in-line probing, suggesting that sequences upstream of Domain1 do not interact substantially with most sequences in the domain (Figure 20A). Major differences with the in-line probing data were found, however, for two pyrimidine-rich sequences within the large unstructured region (USR) linking H3 and H4 (positions 3818-3827 and 3837-3846; boxed in Figure 20A). These sequences lacked flexibility only when full-length gRNA was probed, suggesting that they interact with sequences absent in 3' terminal fragments. In addition, several residues in H4TL and three of the five residues of H4AL exhibited different flexibility from the profile previously determined using 3'UTR fragments (compare H4 in Figure 20A and B), suggesting either that sequences upstream of the 3'UTR affect the structure of the H4 loops and/or that the different techniques have variable results for resides likely involved

in complex higher order structure (159).

Susceptibility of Domain1 residues to NMIA was determined for the gRNAs described in (Figure 20 and Figure 21). TCV-m21 had structural changes in H4AL (the local region containing the mutations), with reduced flexibility of G3894, U3898 and U3899 (Figure 20C, compare lanes 2 and 6). Although these residues are predicted to be unpaired, this result supports our earlier findings that H4AL participates in widespread higher order interactions in the 3' region. Rev1 partially restored the wt pattern in H4AL with the exception that the flexibility of U3898 and A3899 (the mutated residue in rev1) retained their reduced flexibilities (Figure 20C, compare lanes 2 and 10). Interestingly, Domain2 alteration G3561A generated a reproducible difference within the 3'UTR of TCV, increasing the flexibility of the H4 lower stem guanylates G3901 and G3902 (Figure 20C, compare lanes 2 and 4). This enhanced flexibility remained when G3561A was combined with m21 and rev1 (Figure 20C, compare lanes 4, 8 and 12). In all three repetitions of this experiment, the control DMSO lane also contained reverse transcriptase stops at residues G3901 and G3902 in the G3561A and G3561A+rev1 gRNAs (and in the G3561A+m21 gRNA in one experiment), which were absent from other lanes. Second-site mutation U3329C did not produce any discernable structural changes in the H4 region or elsewhere in Domain1 (Figure 21A and B), and m21 and rev1 produced similar local structural change in H4AL with or without U3329C. These results suggests that while a structural connection appears to exist between Domain 1 and Domain2, there are no discernable common RNA structural changes specifically associated with the synergism between rev1 and both G3561A and U3329C.

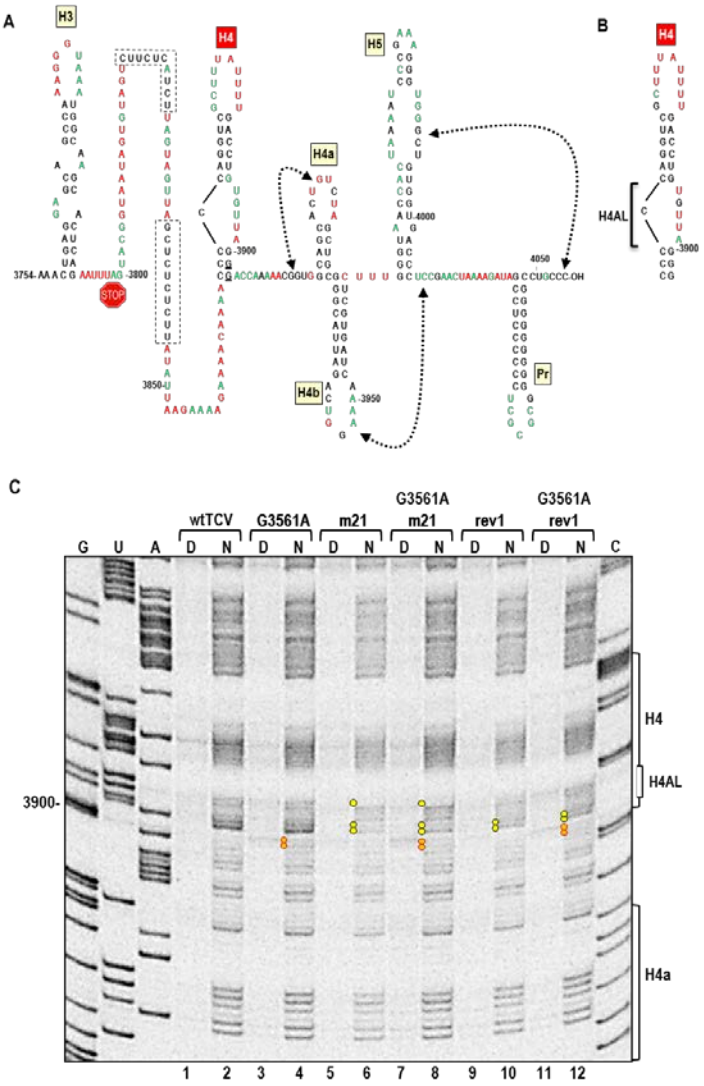


Figure 20: G3561A enhances the flexibility of the H4 lower stem

(A) Structure of Domain1 determined using SHAPE and full-length TCV gRNA. Residues corresponding to high and low reactivity to NMIA are denoted by red and green colors, respectively. Residues in hatched boxes differ significantly in their stability when assayed using SHAPE of full-length gRNA versus in-line probing of 3'UTR fragments (136). Residues whose flexibility changes in gRNA containing G3561A are underlined (G3901 and G3902).

(B) In-line probing results for H4 using a 3' terminal fragment beginning at position 3859 (13). Red residues are highly flexible and green residues are moderately flexible.

(C) SHAPE autoradiogram showing the effect of G3561A with and without m21 or rev1 on the structure of the RNA in the vicinity of H4 in Domain1. Location of hairpins H4 and H4a, and the H4 asymmetric loop (H4AL) are shown. Orange and yellow circles denote residues with enhanced flexibility and reduced flexibility, respectively. Performed by Vera Stupina.

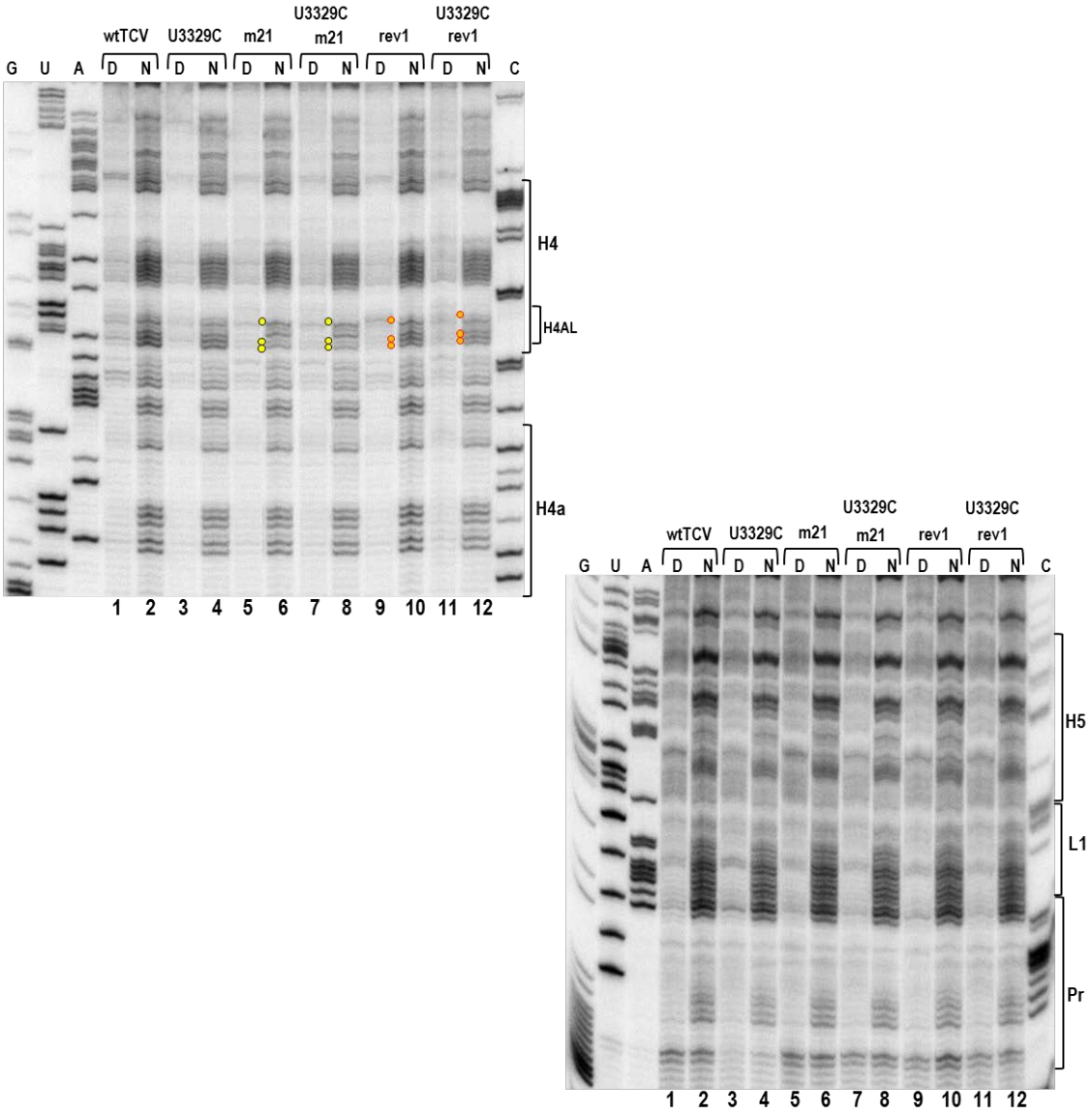


Figure 21: Second-site alteration U3329C did not affect the structure of the H4 lower stem

Second-site alteration U3329C did not affect the structure of the H4 lower stem. SHAPE autoradiogram showing no discernable flexibility changes due to U3329C on the structure of the RNA in the vicinity of H4 in Domain1.

Compensatory effects require a functional p38

p38 is a multifunctional protein that assembles into virions, serves as an effector of virus resistance, and mediates suppression of virus accumulation by the virulent satellite RNA satC (116). Importantly, p38 also functions as the TCV silencing

suppressor, a critical protein required to suppress the innate RNA silencing defense system of the host plant (117, 154). Both G3561A and U3329C, as well as U3741C from the triple mutation set A3475G/A3709G/U3741C are missense mutations and thus may alter one or more of the functional properties of p38. To determine if the beneficial effects of G3561A and U3329C on TCV-rev1 accumulation are connected with an altered p38, G3561A and U3329C were introduced into a TCV backbone (TCVCP-) with an engineered stop codon at the 6th codon position in the p38 ORF that eliminates detectable p38 synthesis (116). In the absence of p38, TCVCP- accumulates to only 10% of wtTCV in protoplasts with a functional RNA silencing system (116), indicating that p38 is required for efficient virus accumulation due at least in part to its role in suppressing silencing.

Whereas G3561A had a negative effect on wtTCV levels (Figure 16, right panel), it had an insignificant effect on TCVCP- accumulation (Figure 22A, left panel), suggesting that the negative effect of this alteration on wtTCV accumulation is due to a defect in p38 and not in the structure of the RNA. Rev1 reduced accumulation of TCVCP- to a similar extent as it did with wtTCV. G3561A did not enhance accumulation of TCVCP-/rev1, suggesting that the beneficial effect of G3561A on TCV-rev1 depends on functional p38.

As with wtTCV, U3329C did not negatively affect accumulation of TCVCP- (Figure 22A, right panel). TCVCP-/rev1 accumulated to 82% of TCVCP- levels in this set of experiments and, unexpectedly, addition of U3329C reduced TCVCP-/rev1 levels by 4-fold. This significant reduction in TCVCP-/rev1 levels by U3329C suggests that, in the absence of p38, U3329C negatively affects gRNA structure when combined with

A3899U. Since these two mutations are in separate RNA domains, this further supports possible structural connections between Domain1 and Domain2.

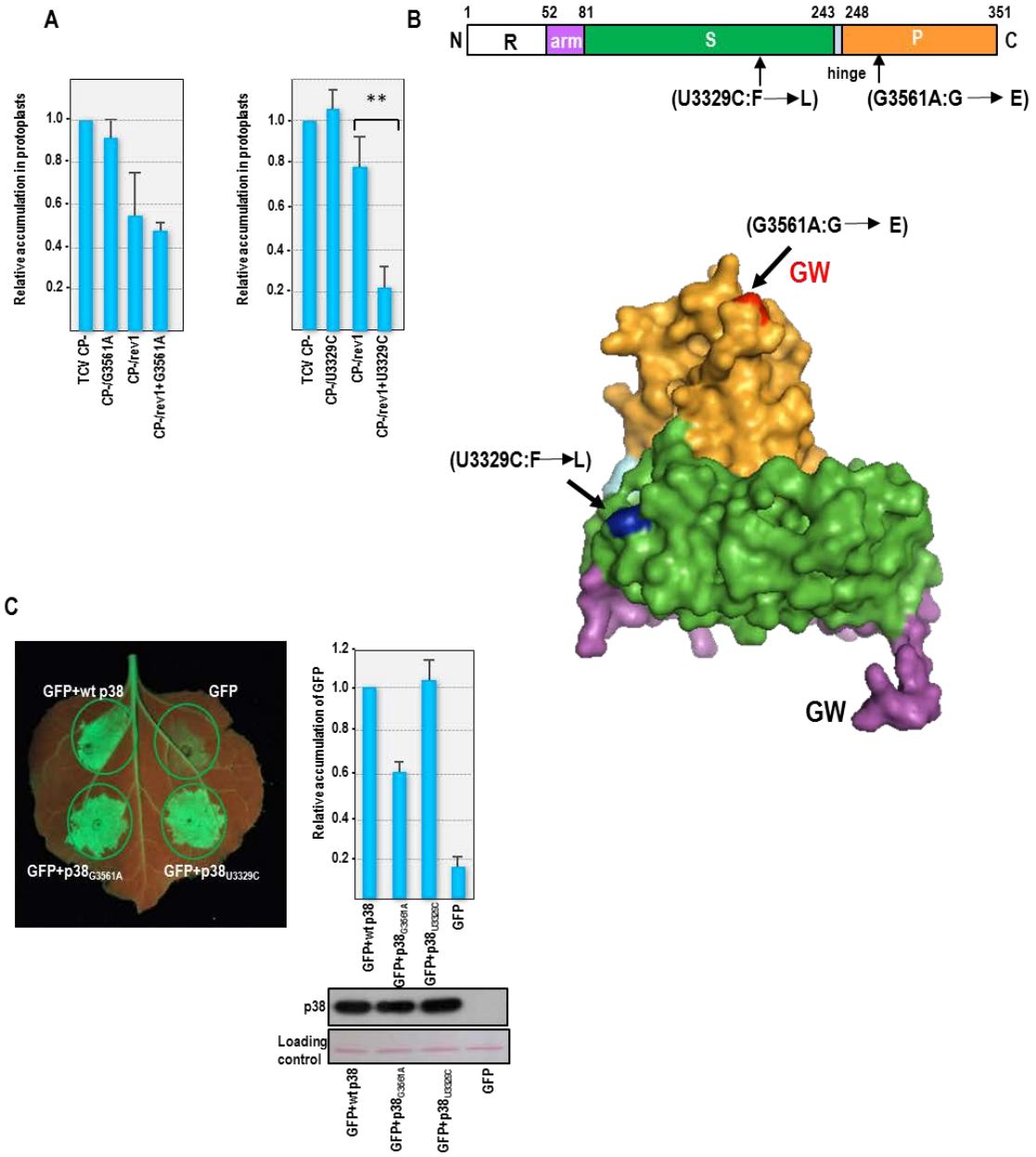


Figure 22: G3561A affects the silencing suppressor activity of the CP

(A) Accumulation of viral gRNA incapable of producing p38 (TCV CP-] in Arabidopsis protoplasts (Col-0). Left panel: accumulation of TCV-G3561A, TCV-rev1 and TCV-rev1+G3561A in the absence of p38. Right panel: accumulation of TCV-U3329C, TCV-rev1 and TCV-rev1+U3329C in the absence of p38.

(B) Top, structural domains of p38. R, RNA binding domain; A, Arm; S, Surface domain; H, hinge; P, protruding domain. G561A is located in the P domain and U3329C is located in the S domain. Bottom, Pymol-generated 3-D structure of p38 showing the location of the GW motifs

and the second-site mutations. The N-terminal GW motif is located in an unresolved region in p38. G3561A alters the C-terminal GW motif (red). U3329C modifies a surface non-polar residue (blue).
(C) Effect of second-site mutations on silencing suppressor activity of p38. Left, expression of a reporter GFP transgene with and without mutant and wt p38 in N. benthamiana leaves. Photographs were taken at 5-days post-infiltration under ultraviolet light to show the green fluorescence of GFP and the red fluorescence of chlorophyll. Right, accumulation of GFP RNA in leaves assayed by RNA blotting using a [α-^{32}P] dCTP-labeled DNA probe corresponding to full-length GFP ORF. To detect levels of p38, equal amount of total protein were separated by 12% SDS-PAGE and subjected to immunoblotting using a polyclonal anti-p38 antibody. Performed by Feng Gao and Maitreyi Chattopadhyay.

G3561A alters one of two GW motifs in the CP AGO binding platform

Since the positive effects of G3561A and U3329C on rev1 were dependent on production of p38, one possibility was that the mutations affect the silencing suppressor activity of p38. p38 contains two GW motifs located proximal to the N- and C-termini of the protein (120), both of which are necessary for efficient binding to AGO1 and AGO2 (120, 123), and for binding to synthetic 19-nt duplex RNAs (160). The N-terminal GW motif is conserved in nearly all carmoviruses and the C-terminal motif is unique to TCV. The C-proximal GW motif at position 273 is located on the surface of the protrusion (P) domain of p38, while the N-proximal GW motif at position 25 is in the RNA-binding (R) domain that is structurally unsolved due to disorganization of the protein in this region (Figure 22B). In addition to the GW motifs, R74, E122, R130 and R137 have been reported to be implicated in p38 silencing suppressor activity (119, 157). However, R74 disrupts several p38 functions including virion formation, indicating that (at least) this defect causes a general disruption of p38 structure (116). In contrast, R130 did not affect virion formation in *dcl2-dcl3-dcl4* triple mutant plants (157).

Figure 22B shows the location of G3561A and U3329C in the 3-D structure of p38. Strikingly, G3561A replaces a non-polar glycine R group in the C-proximal GW motif with the negatively charged glutamic acid R group (Figure 22), thus likely

disrupting the interaction of this GW motif with AGO1/AGO2 and dsRNAs. U3329C, which causes a conservative phenylalanine to leucine alteration at position 196, maps to the surface of p38 in a region of the S domain that is distal to the GW motifs. There are currently no specific activities of p38 associated with this region (S.C. Harrison, personal communication).

To determine if G3561A or U3329C affect silencing suppressor activity of p38, co-transient expression assays were conducted by infiltrating *Nicotiana benthamiana* line 16c with Agrobacteria carrying a reporter GFP transgene, or the transgene together with either wt p38, $p38_{G3561A}$ or $p38_{U3329C}$. Leaves infiltrated with the GFP transgene alone showed a marked decreased in GFP expression after 5 dpi due to rapid induction of host RNA silencing (Figure 22C). In contrast, co-infiltration with wt p38 suppressed RNA silencing allowing for GFP expression and detection. When leaves were infiltrated with GFP along with $p38_{G3561A}$, a 40% reduction in GFP mRNA accumulation was observed, supporting previous reports on the importance of at least the C-proximal GW motif for efficient RNA silencing (120). Since G3561A only negatively impacts TCV accumulation when p38 is synthesized, this suggests that the principal effect of G3561A is to reduce silencing suppressor activity of p38. In contrast, U3329C had no detectable effect on silencing suppressor activity of p38 (Figure 22C). Therefore, these results suggest that G3561A and U3329C enhance accumulation of TCV-rev1 using different mechanisms, however both mechanisms require p38.

The compensatory effects of G3561A and U3329C require DCL proteins

Of the four DCLs found in Arabidopsis, DCL4 is the key enzyme in the plant's

anti-viral defensive arsenal, targeting structured, imperfect duplexes in viral gRNA or double-stranded replicative intermediates (161). In *dcl4* mutant backgrounds, DCL2 becomes the principal anti-viral factor, and may also have an additional role in production of secondary vsRNAs that are generated following host RdR synthesis of long viral dsRNAs using DCL-generated primers (162). In the absence of p38, TCV accumulates to wt levels in *dcl2-dcl4* plants but not in plants with defects at single loci (120). In addition, TCV gRNA in the absence of p38 can accumulate to wt levels in *dcl2-dcl4* protoplasts (116, 163), demonstrating that RNA silencing is active in callus-derived single cells.

If G3561A enhancement of TCV-rev1 levels is due to reduced silencing suppressor activity of $p38_{G3561A}$, then G3561A should have no positive effect if cells cannot synthesize DCLs that target viral RNAs. To test this prediction, the effects of G3561A and U3329C on TCV-rev1 accumulation were assessed in Arabidopsis *dcl2-dcl4* protoplasts. TCV-G3561A accumulated in *dcl2-dcl4* cells to 85% of wtTCV, supporting the suggestion that the negative effect of G3561A on wt protoplasts is primarily due to reduced silencing suppressor activity of $p38_{G3561A}$ (Figure 23A, left panel). Similar to results using wt protoplasts, G3561A had no effect on TCV-m21 accumulation. As predicted, G3561A also did not enhance levels of TCV-rev1 in *dcl2-dcl4* protoplasts, suggesting that DCL2/DCL4 (one or both) are necessary for G3561/rev1 synergy (Figure 23B, top panel).

TCV-U3329C accumulated to near wtTCV levels in *dcl2-dcl4* cells (Figure 23A, bottom panel) and U3329C had no effect on TCV-m21 levels. Addition of U3329C to TCV-rev1 reduced accumulation by 55% in *dcl2-dcl4* cells, similar to the results obtained

when U3329C was included with TCVCP-/rev1 in wt protoplasts. These results suggest that, as with G3561A, DCL2/DCL4 are necessary for U3329C to have a positive effect on TCV-rev1, and, in the absence of the DCLs, the combination of U3329C and A3899U (the rev1 alteration) is strongly negative for TCV accumulation.

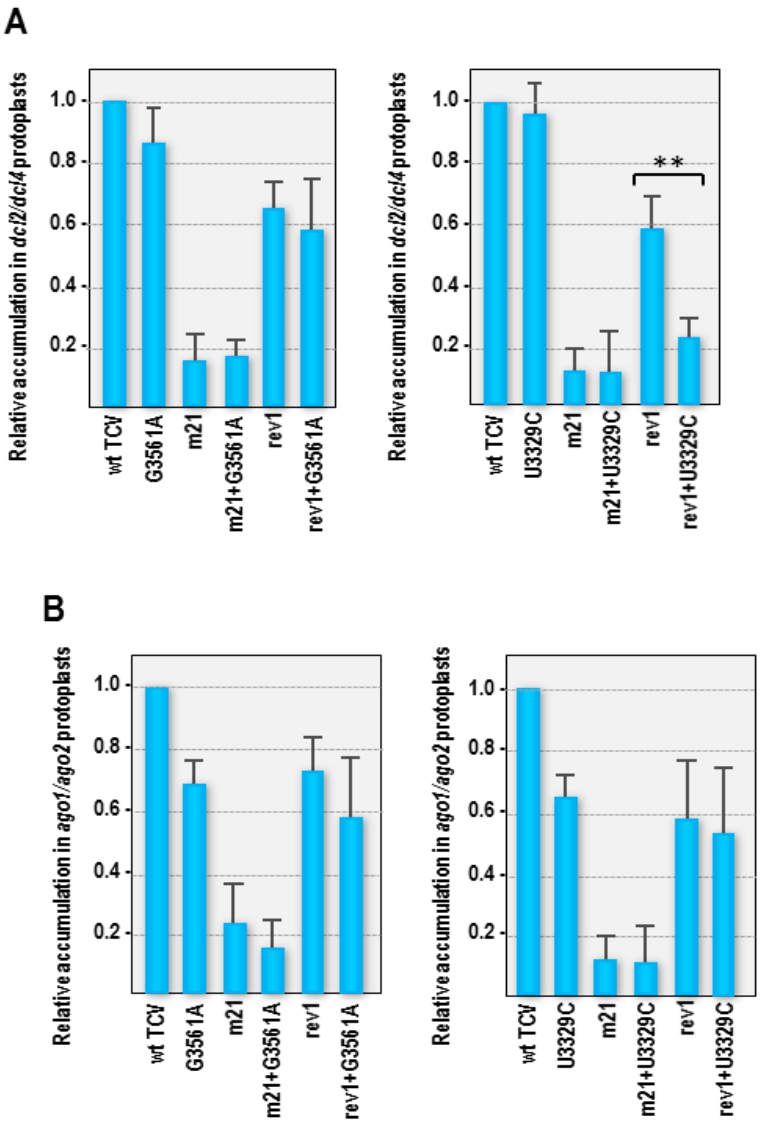


Figure 23: Compensatory effect of the second-site mutations requires DCL2/DCL4 and AGO1/AGO2

(A) Effect of second-site mutations on accumulation of wtTCV, TCV-m21 and TCV-rev1 in *dcl2-dcl4* protoplasts. Left, G3561A; right, U3329C.

(B) Effect of second-site mutations on accumulation of wtTCV, TCV-m21 and TCV-rev1 in *ago1-ago2* protoplasts. Left, G3561A; right, U3329C. Viral RNA levels were assayed at 40 hpi by RNA gel blots. Values are from independent experiments conducted in triplicate and standard

deviations are shown. *=p<0.05, **=p<0.01 Student's t-test.

The compensatory effects of G3561A and U3329C require AGO1/AGO2

Active DCLs do not by themselves limit virus infection (122), suggesting that additional cleavage events mediated by RISC are necessary to clear viruses from plants. AGO slicer enzymes are the active components within RISC, and AGO proteins are critical for antiviral defense and are popular targets for virus silencing suppressors, including p38 (161). AGO1 and AGO2, which are two of the 10 AGO proteins in Arabidopsis, act synergistically to control virus infections and both are targeted by p38 (120, 121)

To determine if AGO1/AGO2 are needed for TCV-rev1 compensation by G3561A and U3329C, TCV containing the second-site changes in combination with the rev1 alteration were assessed in *ago1/ago2* protoplasts. TCV-G3561A and TCV-U3229C accumulated in *ago1/ago2* protoplasts to 68% and 65% of wtTCV, respectively and neither second-site mutation significantly affected levels of TCV-m21 or TCV-rev1 (Figure 23B). Altogether, these results suggest that neither G3561A nor U3329C are compensatory with rev1 in the absence of p38, or when cells lack important components of RNA silencing.

Discussion

The principal goals of this study were to investigate whether second-site changes that arose in the p38 ORF compensated for primary mutations in the 3'UTR, and if so, what mechanism(s) might be responsible for such compensation. Primary-site mutations were constructed in H4AL because the internal (and terminal) loops of H4 have a central

role in a network of structural interactions that connect 3' elements (136), see Figure 2. Such interactions were previously detected by in-line probing, a technique that requires the use of short fragments (13). In-line probing of 3'UTR and extended 3'UTR fragments showed that: *i.* mutations generated in H4AL and H4TL disrupt RNA structure in several downstream regions (13, 131, 136); *ii*. mutations in the 3' terminal Pr hairpin and in the pseudoknot that connects 3' terminal residues to hairpin H5 (Ψ_1) alter the flexibility of residues in H4AL (131); and *iii*. flexibilities of H4TL and H4AL residues were altered upon RdRp binding to 3'UTR fragments as part of a wide-spread conformational shift (13). These findings suggest a critical role for H4 in supporting the structure of the 3'region, which is likely required for 3' end-associated activities such as translation mediated by the TSS 3'CITE (89), the long-distance interaction involved in ribosome recoding that extends the p28 ORF to synthesize the p88 RdRp (29), and RdRp recognition of the promoter for complementary-strand synthesis (13). Attempts to detect structural changes caused by H4AL alterations m21 and rev1 outside of the H4AL region using SHAPE and full-length TCV gRNA were not successful (Figure 21). Likewise, SHAPE analysis of the TBSV genome could only account for two of six known long-distance inter- and intra-domains interactions (24). Therefore, SHAPE analysis using NMIA may not be effective for ascertaining disturbances in subtle tertiary interactions (159).

m21 primary-site alterations were detrimental for virus accumulation in plants and protoplasts, and all second-site changes that arose in response to m21 were associated with partial reversions including rev1, which leaves only a single base alteration in H4AL (U3899A) (131). Surprisingly, TCV-rev1 accumulation was quite variable, ranging from

40% to 65% of wtTCV in Col-0 protoplasts in independent experiments (Figure 16B), and 58% to 75% of wtTCV in different mutant protoplast cells (Figure 23). This highly unusual variation in levels of a TCV mutant transcript does not have a definitive explanation. One possibility is that the structure of the transcript when inoculated into protoplasts differed in different preparations and experiments (conducted by Vera Stupina and myself), affecting early translation/replication events.

Second-site mutations accumulating in progeny of TCV-m21 transcripts were found in several locations proximal to H4 (131), as well as throughout a 400 nt region within the upstream p38 ORF (this report). SHAPE chemical probing revealed that this upstream region comprised a domain (Domain2) that includes most of the p38 ORF (Figure 17). When combined with the rev1 alteration, all second-site mutations tested enhanced levels of gRNA accumulation compared with those of TCV-rev1 and/or wt TCV containing the second-site mutations assayed in the same experiment (Figure 16). None of the mutations compensated for m21, suggesting that they arose subsequent to the conversion of m21 to rev1. U3329C was located very near two other second-site changes present on both sides of Domain2 hairpin H2-4 (Figure 17), suggesting a possible role for regional RNA structure disruption in the U3329C compensatory mechanism. Primary mutations in the 3' terminal Pr hairpin loop, a region that also affects the higher order structure of the 3'UTR (34), gave rise to second-site mutations that cluster in the same vicinity on both sides of the H2-4 stem as well as in Domain1 locations very similar to second-site mutations of m21 (X. Yuan and A.E.Simon, unpublished results). We are currently testing if other clustered second-site mutations in Domain1 and Domain2 are compensatory in the presence of rev1 or Pr terminal loop mutations.

When translation of p38 was blocked, U3329C caused a 4-fold decrease in levels of gRNA with the A3899U rev1 alteration, supporting the hypothesis that structural alterations caused by U3329C on Domain2 impact the structure of Domain1. Although these proposed structural alterations could not be detected by SHAPE, G3561A consistently altered the flexibility of two guanylates in the H4 lower stem, supporting a structural connection between Domain1 and Domain2 (Figure 20A and C). This particular structural change was G3561A-specific, it did not correlate with compensation in general by alterations in Domain2. RNA:RNA interactions connecting domains protruding from a central backbone is consistent with observations of viral RNA genomes by CryoEM, which show condensation of domains in the presence of magnesium, suggesting that tertiary interactions connect secondary-structure elements within and between RNA domains (164).

Since G3561A reduced gRNA accumulation by 62%, but had no detrimental effect in the absence of P38, the G3561A G→E amino acid alteration negatively impacts p38. Several pieces of evidence suggest that this alteration reduces silencing suppressor activity of p38 and that this reduction is responsible for the compensatory property of the mutation. These include: *i.* $p38_{G3561A}$ has 40% less silencing suppressor activity than wt p38 (Figure 22C); *ii.* the G→A alteration changes the C-terminal GW motif, known to be important for binding to AGOs and dsRNA; *iii.* G3561A had no compensatory effect in the absence of p38; *iv.* G3561A had no compensatory effect if cells lacked one target of the p38 silencing suppressor (*ago1/ago2*); *v.* G3561A had no compensatory effect if cells lacked RNA silencing (*dcl2/dcl4*).

A significant question is how reducing the silencing suppressor activity of p38 compensates for rev1-associated structural changes in the 3'UTR of TCV. Any mechanism must account for compensation not occurring in the absence of key RNA silencing components. One possibility is that weakening silencing suppressor activity of p38 allows for the enhanced production of one or more vsRNAs. Such a vsRNA could benefit the virus by correcting structural defects through complementary pairing. This hypothesis would account for the lack of compensation in the absence of DCL2/DCL4 and AGO1/AGO2, which are responsible for the production of primary and, presumptively, secondary vsRNAs (119, 122, 165). An alternative explanation for the lack of compensation in the *ago1/ago2* background is that lack of AGO1 affects homeostasis of DCLs, including reduced accumulation of DCL4, which should affect primary vsRNAs (120). Several studies have reported that virus- or host-derived small RNAs regulate virus accumulation in host cells (166, 167). For example, accumulation of HCV in Huh7 cells requires an endogenous microRNA (mi122) (166) and *West Nile virus* (WNV) encodes a miRNA-like vsRNA (KUN-miR-1) that facilitates viral replication in mosquito cells (168). Similar to WNV, a microRNA-like vsRNA autoregulates *Dengue virus 2* replication in mosquito cells (169). However, enhanced production of a vsRNA does not account for the lack of compensation in the absence of CP, which would be expected to produce more vsRNAs. However, the absence of the silencing suppressor severely decreases virus levels (Figure 22A), which may negate any benefit from enhanced production of a beneficial vsRNA.

The conservative F to L amino acid alteration in TCV-U3329C had no negative effect on TCV accumulation (Figure 16B), and did not reduce silencing suppressor

activity of the CP (Figure 22C). These results suggest that the compensatory mechanism for U3329C differs from that of G3561A. Despite this difference, the compensatory effect of U3329C on rev1 was also negated in the absence of CP, or in *dcl2/dcl4* or *ago1/ago2* protoplasts. These findings suggest that, as with G3561A, the compensatory mechanism also requires active RNA silencing. Since the negative effect of combining rev1 with U3329C is mitigated in cells with functional DCL2/DCL4 (i.e., with functional DCLs, the combination of the two mutations is compensatory, not strongly inhibitory), a DCL-produced vsRNA may also play a role in the compensatory effect of U3329C. One possibility is if combined structural changes by rev1 and U3329C create enhanced opportunity for interaction with a beneficial vsRNA. In the absence of the vsRNA (in *dcl2/dcl4* protoplasts), or when there is no silencing suppressor, these combined structural changes must be detrimental, since they result in a 2-fold or 4-fold decrease in virus levels, respectively. In conclusion, this is the first example, to our knowledge, of a requirement for RNA silencing components for second-site changes to be compensatory for mutations in an untranslated region of a virus. As a future experiment deep sequencing might be employed to detect the population of small RNAs that might be important for TCV accumulation.

Chapter 5: Overall conclusions

The genomes of (+)-strand RNA viruses have evolved to integrate multiple canonical and non-canonical long- and short-range RNA:RNA interactions forming higher order structures. These can promote multiple functions in host cells, ranging from translation initiation to evasion of host defense responses. This dissertation sheds light on the role of multiple long- and short-range RNA:RNA interactions that connect dispersed regions of the viral genome and modulate vital viral processes in two carmoviruses, SCV and TCV. These two viruses are small, uncapped, non-polyadenylated (+)-strand RNA viruses. Due to their small size and limited genome capacity SCV and TCV serve as excellent model systems for studying structure-function relationships pertaining to viral viability in host cells.

Translation initiation is one of the earliest events during a viral infection. The TCV 3' CITE, the TSS, binds 80S ribosome and 60S ribosomal subunits and likely communicates with the 5' end via a single ribosome bridge. The PTE of SCV enhances translation using a different mechanism, which also differs from the mechanism used by the PEMV PTE. As a translational enhancer, PEMV PTE binds to eIF4E, but does not directly communicate with the 5'end. For communicating with the 5' end, the PEMV PTE relies on an upstream CITE, the kl-TSS. Here we report that the SCV PTE utilizes long-distance RNA:RNA interactions to confer cap-independent translation of its gRNA and sgRNA2 (Figure 5 and Figure 7). The SCV PTE joins the growing list of 3'CITEs that can bridge the ends of viral genomes, suggesting that a common feature of most of these 3'CITEs is to deliver 3' bound ribosomes or initiation factors via a long-distance RNA:RNA interaction to the 5' end for translation initiation (73, 81, 94).

For nearly all known PTE-containing carmoviruses, the 5’ PTE-interacting sequence is found in specific locations in the gRNA and sgRNA2 (90). This raises the question about the importance of the location of the PTE-interacting sequence relative to the 5’ end. Our analyses of strict positional conservation of the 5’ PTE-interacting sequence in the 5’ end led to the discovery that the process of translation initiation in SCV requires the 5’ PTE-interacting sequence to be positioned some distance away from the 5’ ends of its gRNA and sgRNA2 reporter transcripts (Figure 12 to Figure 14). During these analyses we also identified a conserved CA-rich translational enhancer motif (between positions 2498-2506) in the 5’ UTR of sgRNA2 and determined that deletion of this region partially contributed to the reduced translational activity of the sgRNA2 reporter transcripts (Figure 14).

The transposition of the 5’PTE-interacting sequence from the gH3 apical loop to the apical loop of gH1 was not favorable for translation of the gRNA reporter transcripts, suggesting that gH1 plays a separate role in translation other than serving as a scaffold element for the long-distance interaction (Figure 13). The exact role of gH1 as a translational element in SCV is unknown. One possibility is that gH1 may recruit translation initiation factors or ribosomes to facilitate translation initiation. Certain viruses have shown to utilize their 5’ terminal hairpin for similar purposes (170, 171). The apical loop sequence of gH1 harbors a naturally occurring AUG. Mutating this AUG or introducing an upstream AUG out-of-frame reduces PTE-mediated translation efficiency suggesting that the PTE-mediated cap-independent translation relies on 5’-end-dependent scanning rather than on internal ribosome entry (Figure 13). A similar 5’-end-dependent-scanning mechanism is reported for TCV, TBSV and BYDV (80, 85, 141).

Our analyses in chapter 4 have provided evidence for long-distance RNA:RNA interactions between Domain1 and Domain2 in the TCV genome. G3561A, a Domain2 second-site mutation, caused local structural changes in Domain2 and also affected the structure of Domain1 by altering the flexibility of the H4 lower stem region (Figure 20). The existence of possible structural connections between Domain1 and Domain2 was further supported by the finding that another Domain2 second-site mutation U3329C, in combination with rev1, impacts the structure of Domain1 (Figure 22, left panel). Since none of these second-site changes are in sequences that might canonically pair with the primary-site mutation in Domain1, potential long-distance RNA:RNA interactions between Domain1 and Domain2 likely occur non-canonically.

Previously, it was found that several second-site mutations in Domain1 were compensatory for virus accumulation and corrected the structural defects caused by the primary-site mutations likely via re-establishing RNA:RNA interactions within Domain1. Our analyses do not support the hypothesis that long-distance RNA:RNA interactions between Domain1 and Domain2 are responsible for the compensatory effects exerted by G3561A and U3329C. Instead, we found that a novel mechanism involving host RNA silencing machinery is responsible for that effect. G3561A and U3329C, were able to partially compensate TCV-rev1 only in the presence of the TCV silencing suppressor p38, functional DCLs and RISC (Figure 16, Figure 22 and Figure 23). G3561A reduces p38's silencing suppressor activity (Figure 22), suggesting that an increased production of one or more vsRNAs enhances the accumulation of G3561A in conjunction with rev1, likely by correcting the structural defects in Domain1, which are associated with the primary-site mutation. The second-site mutation U3329C, in combination with rev1,

likely creates a synergistic structural change making Domain1 more accessible to the beneficial effect of low levels of vsRNA. These findings suggest that we need to consider host RNA silencing components as a possible contributor of the compensatory effect. We will investigate if any additional second-site changes within Domain1 and Domain2 also utilize active RNA silencing components to exert the compensatory effect during viral accumulation.

Conventionally, we associate host RNA silencing as a negative regulator for viruses. But under certain circumstances it has been found that a small subset of beneficial virus-derived siRNAs or host-derived microRNAs often facilitates viral propagation in host cells (173-175). If our hypothesis that virus or host-derived small RNAs are responsible for correcting the structural defects, which are associated with the primary-site mutation, leading to an enhanced viral accumulation in host cells is proven to be correct, it will provide a novel insight into the beneficial role of small RNAs in host cells.

Overall, the second-site mutations in the p38 ORF of TCV have revealed that the interconnectivity of the 3'UTR extends even further upstream in Domain2. Our analyses of long- and short-distance RNA:RNA interactions in TCV and SCV demonstrated that the dispersed regions of the viral genomes are connected and likely maintain a global higher order structure that coordinates multiple functions.

Appendix

Domain1 SHAPE:

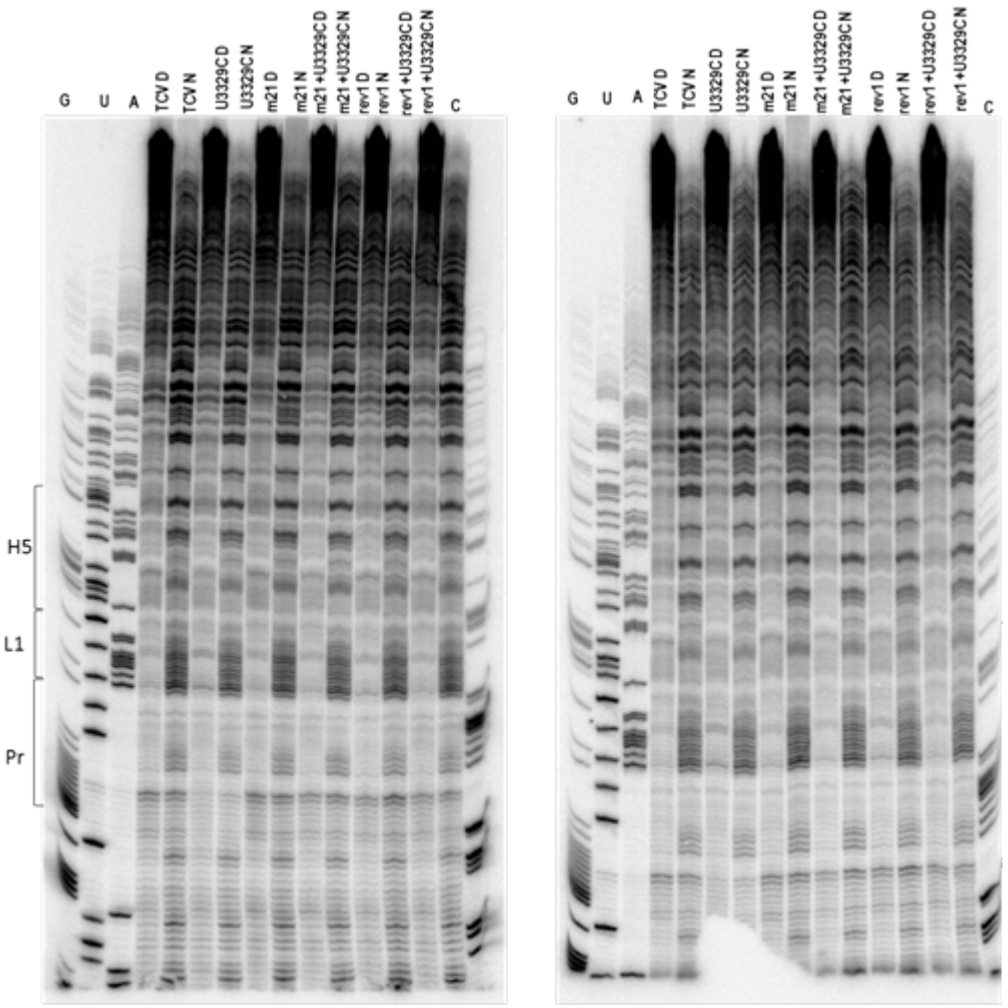


Figure A.1: Structure of Domain1 determined using SHAPE and full-length TCV gRNA. Location of hairpins H5 and Pr and linker region (L1) are shown. Left is a shorter run of samples. Right is a longer run.

Domain2 SHAPE:

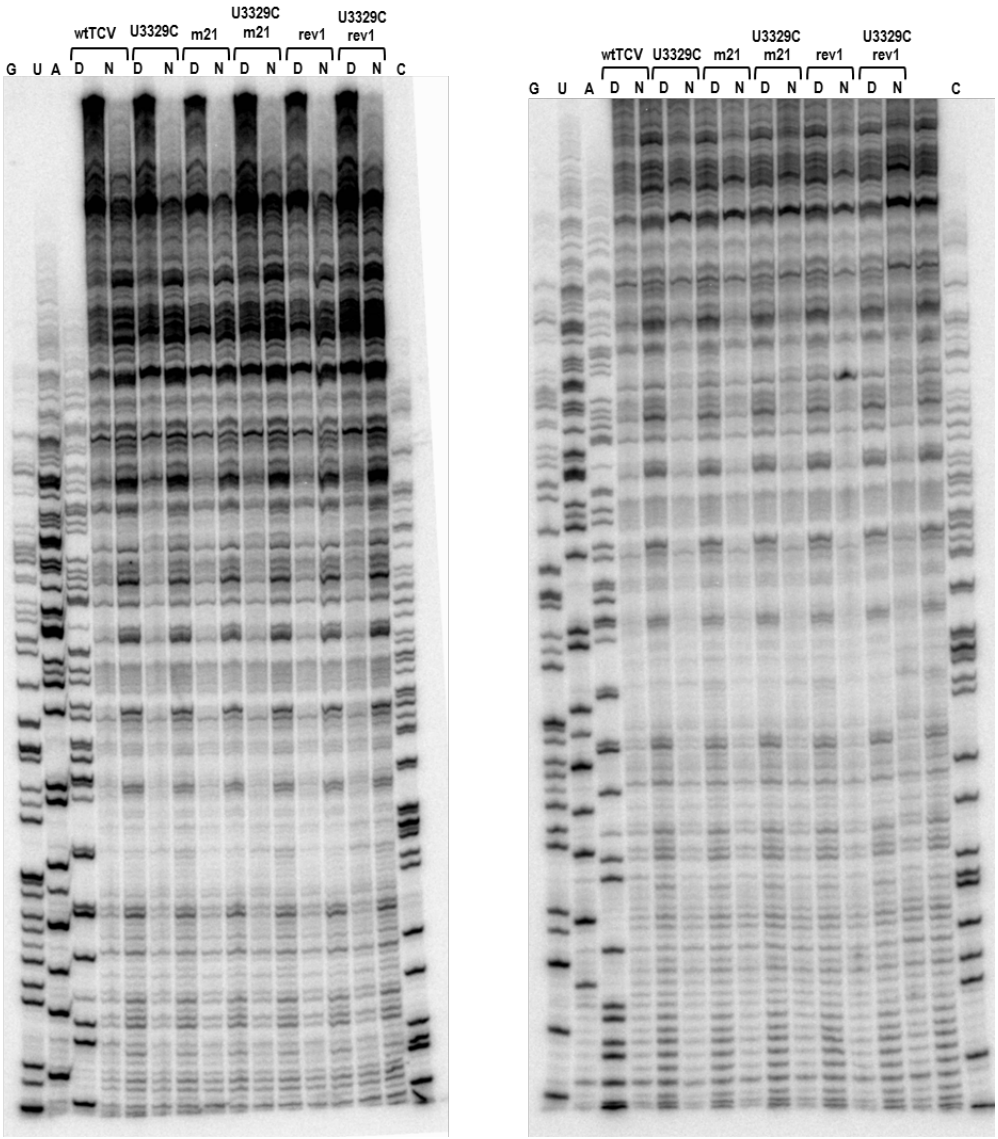


Figure A.2: SHAPE autoradiogram showing structural changes in the vicinity of H2-4 in response to U3329C with and without m21 or rev1. Left is a shorter run and right is a longer run of samples.

Bibliography

1. **Zaitlin M, Palukaitis P.** 2000. ADVANCES IN UNDERSTANDING PLANT VIRUSES AND VIRUS DISEASES. Annu Rev Phytopathol **38:**117-143.

2. **Flint M, Logvinoff C, Rice CM, McKeating JA.** 2004. Characterization of infectious retroviral pseudotype particles bearing hepatitis C virus glycoproteins. J Virol **78:**6875-6882.

3. **Greber UF, Singh I, Helenius A.** 1994. Mechanisms of virus uncoating. Trends Microbiol **2:**52-56.

4. **Shaw JG, Plaskitt KA, Wilson TM.** 1986. Evidence that tobacco mosaic virus particles disassemble contranslationally in vivo. Virology **148:**326-336.

5. **Simon AE, Miller WA.** 2013. 3' cap-independent translation enhancers of plant viruses. Annu Rev Microbiol **67:**21-42.

6. **Ahlquist P, Noueiry AO, Lee WM, Kushner DB, Dye BT.** 2003. Host factors in positive-strand RNA virus genome replication. J Virol **77:**8181-8186.

7. **Kushner DB, Lindenbach BD, Grdzelishvili VZ, Noueiry AO, Paul SM, Ahlquist P.** 2003. Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. Proc Natl Acad Sci U S A **100:**15764-15769.

8. **Nagy PD, Pogany J.** 2008. Multiple roles of viral replication proteins in plant RNA virus replication. Methods Mol Biol **451:**55-68.

9. **Nagy PD, Barajas D, Pogany J.** 2012. Host factors with regulatory roles in tombusvirus replication. Curr Opin Virol **2:**691-698.

10. **Nagy PD, Pogany J.** 2012. The dependence of viral RNA replication on co-opted host factors. Nat Rev Microbiol **10:**137-149.

11. **Pathak KB, Pogany J, Nagy PD.** 2011. Non-template functions of the viral RNA in plant RNA virus replication. Curr Opin Virol **1:**332-338.

12. **Gamarnik AV, Andino R.** 1998. Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev **12:**2293-2304.

13. **Yuan X, Shi K, Meskauskas A, Simon AE.** 2009. The 3' end of Turnip crinkle virus contains a highly interactive structure including a translational enhancer that

is disrupted by binding to the RNA-dependent RNA polymerase. RNA **15:**1849-1864.

14. **Nagy PD, Pogany J.** 2006. Yeast as a model host to dissect functions of viral and host factors in tombusvirus replication. Virology **344:**211-220.

15. **Ahlquist P.** 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. Science **296:**1270-1273.

16. **Noueiry AO, Ahlquist P.** 2003. Brome mosaic virus RNA replication: revealing the role of the host in RNA virus replication. Annu Rev Phytopathol **41:**77-98.

17. **Buck KW.** 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. Adv Virus Res **47:**159-251.

18. **Huang TS, Nagy PD.** 2011. Direct inhibition of tombusvirus plus-strand RNA synthesis by a dominant negative mutant of a host metabolic enzyme, glyceraldehyde-3-phosphate dehydrogenase, in yeast and plants. J Virol **85:**9090-9102.

19. **Sztuba-Solińska J, Stollar V, Bujarski JJ.** 2011. Subgenomic messenger RNAs: mastering regulation of (+)-strand RNA virus life cycle. Virology **412:**245-255.

20. **Sasaki J, Nagashima S, Taniguchi K.** 2003. Aichi virus leader protein is involved in viral RNA replication and encapsidation. J Virol **77:**10799-10807.

21. **Frolova E, Frolov I, Schlesinger S.** 1997. Packaging signals in alphaviruses. J Virol **71:**248-258.

22. **Mothes W, Sherer NM, Jin J, Zhong P.** 2010. Virus cell-to-cell transmission. J Virol **84:**8360-8368.

23. **Barton DJ, Morasco BJ, Flanegan JB.** 1999. Translating ribosomes inhibit poliovirus negative-strand RNA synthesis. J Virol **73:**10104-10112.

24. **Wu B, Grigull J, Ore MO, Morin S, White KA.** 2013. Global organization of a positive-strand RNA virus genome. PLoS Pathog **9:**e1003363.

25. **Archer EJ, Simpson MA, Watts NJ, O'Kane R, Wang B, Erie DA, McPherson A, Weeks KM.** 2013. Long-range architecture in a viral RNA genome. Biochemistry **52:**3182-3190.

26. **Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-716.

27. **Lin HX, White KA.** 2004. A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. EMBO J **23:**3365-3374.

28. **Wu B, Pogany J, Na H, Nicholson BL, Nagy PD, White KA.** 2009. A discontinuous RNA platform mediates RNA virus replication: building an integrated model for RNA-based regulation of viral processes. PLoS Path **5**.

29. **Cimino PA, Nicholson BL, Wu B, Xu W, White KA.** 2011. Multifaceted regulation of translational readthrough by RNA replication elements in a Tombusvirus. PLoS Path **7:**e1002423.

30. **Jiwan SD, White KA.** 2011. Subgenomic mRNA transcription in Tombusviridae. RNA Biol **8:**287-294.

31. **Simmonds P, Tuplin A, Evans DJ.** 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. RNA **10:**1337-1351.

32. **Pumplin N, Voinnet O.** 2013. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. Nat Rev Microbiol **11:**745-760.

33. **Schubert S, Grünweller A, Erdmann VA, Kurreck J.** 2005. Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. J Mol Biol **348:**883-893.

34. **Sagan SM, Nasheri N, Luebbert C, Pezacki JP.** 2010. The efficacy of siRNAs against hepatitis C virus is strongly influenced by structure and target site accessibility. Chem Biol **17:**515-527.

35. **Whisnant AW, Bogerd HP, Flores O, Ho P, Powers JG, Sharova N, Stevenson M, Chen CH, Cullen BR.** 2013. In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. MBio **4:**e000193.

36. **van Himbergen J, van Geffen B, van Duin J.** 1993. Translational control by a long range RNA-RNA interaction; a basepair substitution analysis. Nucleic Acids Res **21:**1713-1717.

37. **Klovins J, Berzins V, van Duin J.** 1998. A long-range interaction in Qbeta RNA that bridges the thousand nucleotides between the M-site and the 3' end is required for replication. RNA **4:**948-957.

38. **Nicholson BL, White KA.** 2014. Functional long-range RNA-RNA interactions in positive-strand RNA viruses. Nat Rev Microbiol **12:**493-504.

39. **Miller WA, White KA.** 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. Ann Rev Phytopath **44:**447-467.

40. **Pestova T. V. LJR, Hellen C. U.** 2007. Translational Control in Biology and Medicine, p NY: 87–128. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

41. **Kozak M.** 1980. Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. Cell **22:**7-8.

42. **Kozak M.** 1999. Initiation of translation in prokaryotes and eukaryotes. Gene **234:**187-208.

43. **Marintchev A, Wagner G.** 2004. Translation initiation: structures, mechanisms and evolution. Q Rev Biophys **37:**197-284.

44. **Wells SE, Hillner PE, Vale RD, Sachs AB.** 1998. Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell **2:**135-140.

45. **Plotch SJ, Bouloy M, Ulmanen I, Krug RM.** 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell **23:**847-858.

46. **Mir MA, Panganiban AT.** 2008. A protein that replaces the entire cellular eIF4F complex. EMBO J **27:**3129-3139.

47. **Reguera J, Weber F, Cusack S.** 2010. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. PLoS Pathog **6:**e1001101.

48. **Fujimura T, Esteban R.** 2011. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. Proc Natl Acad Sci U S A **108:**17667-17671.

49. **Daughenbaugh KF, Fraser CS, Hershey JWB, Hardy ME.** 2003. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. EMBO J **22:**2852-2859.

50. **Goodfellow I, Chaudhry Y, Gioldasi I, Gerondopoulos A, Natoni A, Labrie L, Laliberte JF, Roberts L.** 2005. Calicivirus translation initiation requires an interaction between VPg and eIF4E. EMBO Rep **6:**968-972.

51. **Hellen CU, Sarnow P.** 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev **15:**1593-1612.

52. **Tuplin A.** 2015. Diverse roles and interactions of RNA structures during the replication of positive-stranded RNA viruses of humans and animals. J Gen Virol.

53. **Hertz MI, Thompson SR.** 2011. Mechanism of translation initiation by Dicistroviridae IGR IRESs. Virology **411:**355-361.

54. **Martinez-Salas E, Pineiro D, Fernandez N.** 2012. Alternative Mechanisms to Initiate Translation in Eukaryotic mRNAs. Comp Funct Genomics **2012:**Article ID 391546.

55. **Serrano P, Pulido MR, Sáiz M, Martínez-Salas E.** 2006. The 3' end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA-RNA interactions with the 5' end region. J Gen Virol **87:**3013-3022.

56. **Fraser CS, Doudna JA.** 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. Nat Rev Microbiol **5:**29-38.

57. **Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A.** 1992. Internal ribosome entry site within hepatitis C virus RNA. J Virol **66:**1476-1483.

58. **Ito T, Tahara SM, Lai MM.** 1998. The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. J Virol **72:**8789-8796.

59. **Yi M, Lemon SM.** 2003. 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. J Virol **77:**3557-3568.

60. **Bai Y, Zhou K, Doudna JA.** 2013. Hepatitis C virus 3'UTR regulates viral translation through direct interactions with the host translation machinery. Nucleic Acids Res **41:**7861-7874.

61. **Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, Evans DJ.** 2008. A hepatitis C virus cis-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B. J Virol **82:**9008-9022.

62. **Palau W, Masante C, Ventura M, Di Primo C.** 2013. Direct evidence for RNA-RNA interactions at the 3' end of the Hepatitis C virus genome using surface plasmon resonance. RNA **19:**982-991.

63. **Shetty S, Stefanovic S, Mihailescu MR.** 2013. Hepatitis C virus RNA: molecular switches mediated by long-range RNA-RNA interactions? Nucleic Acids Res **41:**2526-2540.

64. **Basso J, Dallaire P, Charest PJ, Devantier Y, Laliberté JF.** 1994. Evidence for an internal ribosome entry site within the 5' non-translated region of turnip mosaic potyvirus RNA. J Gen Virol **75 (Pt 11):**3157-3165.

65. **Niepel M, Gallie DR.** 1999. Identification and characterization of the functional elements within the tobacco etch virus 5' leader required for cap-independent translation. J Virol **73:**9080-9088.

66. **Gallie DR.** 2001. Cap-independent translation conferred by the 5' leader of tobacco etch virus is eukaryotic initiation factor 4G dependent. J Virol **75:**12141-12152.

67. **Karetnikov A, Keränen M, Lehto K.** 2006. Role of the RNA2 3' non-translated region of Blackcurrant reversion nepovirus in translational regulation. Virology **354:**178-191.

68. **Karetnikov A, Lehto K.** 2007. The RNA2 5' leader of Blackcurrant reversion virus mediates efficient in vivo translation through an internal ribosomal entry site mechanism. J Gen Virol **88:**286-297.

69. **Karetnikov A, Lehto K.** 2008. Translation mechanisms involving long-distance base pairing interactions between the 5' and 3' non-translated regions and internal ribosomal entry are conserved for both genomic RNAs of Blackcurrant reversion nepovirus. Virology **371:**292-308.

70. **Fernández-Miragall O, 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.

71. **Guo L, Allen EM, Miller WA.** 2001. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. Mol Cell **7:**1103-1109.

72. **Miller WA, Wang Z, Treder K.** 2007. The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. Biochem Soc Trans **35:**1629-1633.

73. **Nicholson BL, Wu B, Chevtchenko I, White KA.** 2010. Tombusvirus recruitment of host translational machinery via the 3' UTR. RNA **16:**1402-1419.

74. **Maquat LE, Tarn WY, Isken O.** 2010. The pioneer round of translation: features and functions. Cell **142:**368-374.

75. **Meulewaeter F, Danthinne X, Van Montagu M, Cornelissen M.** 1998. 5'- and 3'-sequences of satellite tobacco necrosis virus RNA promoting translation in tobacco. Plant J **14:**169-176.

76. **Wang Z, Kraft JJ, Hui AY, Miller WA.** 2010. Structural plasticity of Barley yellow dwarf virus-like cap-independent translation elements in four genera of plant viral RNAs. Virology **402:**177-186.

77. **Wang S, Browning KS, Miller WA.** 1997. A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. EMBO J **16:**4107-4116.

78. **Treder K, Kneller EL, Allen EM, Wang Z, Browning KS, Miller WA.** 2008. The 3' cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the eIF4G subunit to initiate translation. RNA **14:**134-147.

79. **Allen E, Wang S, Miller WA.** 1999. Barley yellow dwarf virus RNA requires a cap-independent translation sequence because it lacks a 5' cap. Virology **253:**139-144.

80. **Rakotondrafara AM, Polacek C, Harris E, Miller WA.** 2006. Oscillating kissing stem-loop interactions mediate 5' scanning-dependent translation by a viral 3'-cap-independent translation element. RNA **12:**1893-1906.

81. **Guo L, Allen E, Miller WA.** 2000. Structure and function of a cap-independent translation element that functions in either the 3' or the 5' untranslated region. RNA **6:**1808-1820.

82. **Fabian MR, White KA.** 2004. 5'-3' RNA-RNA interaction facilitates cap- and poly(A) tail-independent translation of tomato bushy stunt virus mrna: a potential common mechanism for tombusviridae. J Biol Chem **279:**28862-28872.

83. **Fabian MR, Na H, Ray D, White KA.** 2003. 3'-Terminal RNA secondary structures are important for accumulation of tomato bushy stunt virus DI RNAs. Virology **313:**567-580.

84. **Nicholson BL, Zaslaver O, Mayberry LK, Browning KS, White KA.** 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.

85. **Fabian MR, White KA.** 2006. Analysis of a 3'-translation enhancer in a tombusvirus: a dynamic model for RNA-RNA interactions of mRNA termini. RNA **12:**1304-1314.

86. **Nieto C, Morales M, Orjeda G, Clepet C, Monfort A, Sturbois B, Puigdomènech P, Pitrat M, Caboche M, Dogimont C, Garcia-Mas J, Aranda MA, Bendahmane A.** 2006. An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. Plant J **48:**452-462.

87. **Truniger V, Nieto C, González-Ibeas D, 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.

88. **McCormack JC, Yuan X, Yingling YG, Kasprzak W, Zamora RE, Shapiro BA, Simon AE.** 2008. Structural domains within the 3' untranslated region of Turnip crinkle virus. J Virol **82:**8706-8720.

89. **Stupina VA, Meskauskas A, McCormack JC, Yingling YG, Shapiro BA, Dinman JD, Simon AE.** 2008. The 3' proximal translational enhancer of Turnip crinkle virus binds to 60S ribosomal subunits. RNA **14:**2379-2393.

90. **Chattopadhyay M, Shi K, Yuan X, Simon AE.** 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.

91. **Batten JS, Desvoyes B, Yamamura Y, Scholthof KBG.** 2006. A translational enhancer element on the 3'-proximal end of the Panicum mosaic virus genome. FEBS Lett **580:**2591-2597.

92. **Wang Z, Treder K, Miller WA.** 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.

93. **Wang Z, Parisien M, Scheets K, Miller WA.** 2011. The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. Structure **19:**868-880.

94. **Gao F, Kasprzak W, Stupina VA, Shapiro BA, Simon AE.** 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.

95. **Gao F, Gulay SP, Kasprzak W, Dinman JD, Shapiro BA, Simon AE.** 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.

96. **Firth AE, Brierley I.** 2012. Non-canonical translation in RNA viruses. J Gen Virol **93:**1385-1409.

97. **Barry JK, Miller WA.** 2002. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. Proc Natl Acad Sci U S A **99:**11133-11138.

98. **Tajima Y, Iwakawa HO, Kaido M, Mise K, Okuno T.** 2011. A long-distance RNA-RNA interaction plays an important role in programmed -1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. Virology **417:**169-178.

99. **Harrell L, Melcher U, Atkins JF.** 2002. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. Nucleic Acids Res **30:**2011-2017.

100. **Villordo SM, Gamarnik AV.** 2009. Genome cyclization as strategy for flavivirus RNA replication. Virus Res **139:**230-239.

101. **Kovalev N, Pogany J, Nagy PD.** 2014. Template role of double-stranded RNA in tombusvirus replication. J Virol **88:**5638-5651.

102. **Alvarez DE, Lodeiro MF, Ludueña SJ, Pietrasanta LI, Gamarnik AV.** 2005. Long-range RNA-RNA interactions circularize the dengue virus genome. J Virol **79:**6631-6643.

103. **Alvarez DE, Lodeiro MF, Filomatori CV, Fucito S, Mondotte JA, Gamarnik AV.** 2006. Structural and functional analysis of dengue virus RNA. Novartis Found Symp **277:**120-132; discussion 132-125, 251-123.

104. **Filomatori CV, Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, Gamarnik AV.** 2006. A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev **20:**2238-2249.

105. **Pogany J, White KA, Nagy PD.** 2005. Specific binding of tombusvirus replication protein p33 to an internal replication element in the viral RNA is essential for replication. J Virol **79:**4859-4869.

106. **Pathak KB, Pogany J, Xu K, White KA, Nagy PD.** 2012. Defining the roles of cis-acting RNA elements in tombusvirus replicase assembly in vitro. J Virol **86:**156-171.

107. **Nicholson BL, Lee PK, White KA.** 2012. Internal RNA Replication Elements are Prevalent in Tombusviridae. Front Microbiol **3:**279.

108. **Choi IR, White KA.** 2002. An RNA activator of subgenomic mRNA1 transcription in tomato bushy stunt virus. J Biol Chem **277:**3760-3766.

109. **Choi IR, Ostrovsky M, Zhang G, White KA.** 2001. Regulatory activity of distal and core RNA elements in Tombusvirus subgenomic mRNA2 transcription. J Biol Chem **276:**41761-41768.

110. **Wang S, Mortazavi L, White KA.** 2008. Higher-order RNA structural requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. J Virol **82:**3864-3871.

111. **Sit TL, Vaewhongs AA, Lommel SA.** 1998. RNA-mediated trans-activation of transcription from a viral RNA. Science **281:**829-832.

112. **Carrington JC, Heaton LA, Zuidema D, Hillman BI, Morris TJ.** 1989. The genome structure of turnip crinkle virus. Virology **170:**219-226.

113. **Hacker DL, Petty IT, Wei N, Morris TJ.** 1992. Turnip crinkle virus genes required for RNA replication and virus movement. Virology **186:**1-8.

114. **White KA, Skuzeski JM, Li W, Wei N, Morris TJ.** 1995. Immunodetection, expression strategy and complementation of turnip crinkle virus p28 and p88 replication components. Virology **211:**525-534.

115. **Kong Q, Oh JW, Simon AE.** 1995. Symptom attenuation by a normally virulent satellite RNA of turnip crinkle virus is associated with the coat protein open reading frame. Plant Cell **7:**1625-1634.

116. **Manfre AJ, Simon AE.** 2008. Importance of coat protein and RNA silencing in satellite RNA/virus interactions. Virology **379:**161-167.

117. **Qu F, Ren T, Morris TJ.** 2003. The coat protein of turnip crinkle virus suppresses posttranscriptional gene silencing at an early initiation step. J Virol **77:**511-522.

118. **Mérai Z, Kerényi Z, Kertész S, Magna M, Lakatos L, Silhavy D.** 2006. Double-stranded RNA binding may be a general plant RNA viral strategy to suppress RNA silencing. J Virol **80:**5747-5756.

119. **Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O.** 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science **313:**68-71.

120. **Azevedo J, Garcia D, Pontier D, Ohnesorge S, Yu A, Garcia S, Braun L, Bergdoll M, Hakimi MA, Lagrange T, Voinnet O.** 2010. Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. Genes Dev **24:**904-915.

121. **Harvey JJ, Lewsey MG, Patel K, Westwood J, Heimstädt S, Carr JP, Baulcombe DC.** 2011. An antiviral defense role of AGO2 in plants. PLoS One **6:**e14639.

122. **Wang XB, Jovel J, Udomporn P, Wang Y, Wu Q, Li WX, Gasciolli V, Vaucheret H, Ding SW.** 2011. The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in Arabidopsis thaliana. Plant Cell **23:**1625-1638.

123. **Zhang X, Zhang X, Singh J, Li D, Qu F.** 2012. Temperature-Dependent Survival of Turnip Crinkle Virus-Infected Arabidopsis Plants Relies on an RNA Silencing-Based Defense That Requires DCL2, AG02, and HEN1. J Virol **86:**6847-6854.

124. **Simon AE, Howell SH.** 1986. The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3' end of the helper virus genome. EMBO J **5:**3423-3428.

125. **Song C, Simon AE.** 1995. Requirement of a 3'-terminal stem-loop in in vitro transcription by an RNA-dependent RNA polymerase. J Mol Biol **254:**6-14.

126. **Stupina V, Simon AE.** 1997. Analysis in vivo of turnip crinkle virus satellite RNA C variants with mutations in the 3'-terminal minus-strand promoter. Virology **238:**470-477.

127. **Zhang G, Zhang J, Simon AE.** 2004. Repression and derepression of minus-strand synthesis in a plus-strand RNA virus replicon. J Virol **78:**7619-7633.

128. **Guan H, Simon AE.** 2000. Polymerization of nontemplate bases before transcription initiation at the 3' ends of templates by an RNA-dependent RNA polymerase: an activity involved in 3' end repair of viral RNAs. Proc Natl Acad Sci U S A **97:**12451-12456.

129. **Sun X, Simon AE.** 2006. A cis-replication element functions in both orientations to enhance replication of Turnip crinkle virus. Virology **352:**39-51.

130. **Simon AE.** 2015. 3'UTRs of carmoviruses. Virus Res.

131. **Yuan X, Shi K, Young MY, Simon AE.** 2010. The terminal loop of a 3' proximal hairpin plays a critical role in replication and the structure of the 3' region of Turnip crinkle virus. Virology **402:**271-280.

132. **Zhang J, Stuntz RM, Simon AE.** 2004. Analysis of a viral replication repressor: sequence requirements for a large symmetrical internal loop. Virology **326:**90-102.

133. **McCormack JC, Simon AE.** 2004. Biased hypermutagenesis associated with mutations in an untranslated hairpin of an RNA virus. J Virol **78:**7813-7817.

134. **Guo R, Lin W, Zhang J, Simon AE, Kushner DB.** 2009. Structural plasticity and rapid evolution in a viral RNA revealed by in vivo genetic selection. J Virol **83:**927-939.

135. **Zhang J, Zhang G, McCormack JC, Simon AE.** 2006. Evolution of virus-derived sequences for high-level replication of a subviral RNA. Virology **351:**476-488.

136. **Yuan X, Shi K, Simon AE.** 2012. A Local, Interactive Network of 3 ' RNA Elements Supports Translation and Replication of Turnip Crinkle Virus. Journal of Virology **86:**4065-4081.

137. **Zhang G, Zhang J, George AT, Baumstark T, Simon AE.** 2006. Conformational changes involved in initiation of minus-strand synthesis of a virus-associated RNA. RNA **12:**147-162.

138. **Dreher TW, Miller WA.** 2006. Translational control in positive strand RNA plant viruses. Virology **344:**185-197.

139. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl Acids Res **31:**3406-3415.

140. **Zhang J, Zhang G, Guo R, Shapiro BA, Simon AE.** 2006. A pseudoknot in a preactive form of a viral RNA is part of a structural switch activating minus-strand synthesis. J Virol **80:**9181-9191.

141. **Stupina VA, Yuan X, Meskauskas A, Dinman JD, Simon AE.** 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.

142. **Kneller EL, Rakotondrafara AM, Miller WA.** 2006. Cap-independent translation of plant viral RNAs. Virus Res **119:**63-75.

143. **Shetty S, Kim S, Shimakami T, Lemon SM, Mihailescu MR.** 2010. Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. RNA **16:**913-925.

144. **Laughrea M, Jetté L.** 1994. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. Biochemistry **33:**13464-13474.

145. **Koh DC, Wong SM, Liu DX.** 2003. Synergism of the 3'-untranslated region and an internal ribosome entry site differentially enhances the translation of a plant virus coat protein. J Biol Chem **278:**20565-20573.

146. **Wilkinson KA, Merino EJ, Weeks KM.** 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc **1:**1610-1616.

147. **Guan HC, Carpenter CD, Simon AE.** 2000. Analysis of cis-acting sequences involved in plus-strand synthesis of a turnip crinkle virus-associated satellite RNA identifies a new carmovirus replication element. Virology **268:**345-354.

148. **Shen RZ, Miller WA.** 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.

149. **Qu F, Morris TJ.** 2000. Cap-independent translational enhancement of turnip crinkle virus genomic and subgenomic RNAs. J Virol **74:**1085-1093.

150. **Koh DC, Liu DX, Wong SM.** 2002. A six-nucleotide segment within the 3' untranslated region of hibiscus chlorotic ringspot virus plays an essential role in translational enhancement. J Virol **76:**1144-1153.

151. **Jablonski SA, Morrow CD.** 1995. Mutation of the aspartic acid residues of the GDD sequence motif of poliovirus RNA-dependent RNA polymerase results in enzymes with altered metal ion requirements for activity. J Virol **69:**1532-1539.

152. **Vázquez AL, Alonso JM, Parra F.** 2000. Mutation analysis of the GDD sequence motif of a calicivirus RNA-dependent RNA polymerase. J Virol **74:**3888-3891.

153. **Wang Y, Xiao M, Chen J, Zhang W, Luo J, Bao K, Nie M, Li B.** 2007. Mutational analysis of the GDD sequence motif of classical swine fever virus RNA-dependent RNA polymerases. Virus Genes **34:**63-65.

154. **Thomas CL, Leh V, Lederer C, Maule AJ.** 2003. Turnip crinkle virus coat protein mediates suppression of RNA silencing in Nicotiana benthamiana. Virology **306:**33-41.

155. **Merai Z, Kerenyi Z, Kertesz S, Magna M, Lakatos L, Silhavy D.** 2006. Double-stranded RNA binding may be a general plant RNA viral strategy to suppress RNA silencing. J Virol **80:**5747-5756.

156. **Perez-Canamas M, Hernandez C.** 2015. Key Importance of Small RNA Binding for the Activity of a Glycine-Tryptophan (GW) Motif-containing Viral Suppressor of RNA Silencing. J Biol Chem **290:**3106-3120.

157. **Cao M, Ye X, Willie K, Lin J, Zhang X, Redinbaugh MG, Simon AE, Morris TJ, Qu F.** 2010. The capsid protein of Turnip crinkle virus overcomes two separate defense barriers to facilitate systemic movement of the virus in Arabidopsis. J Virol **84:**7793-7802.

158. **Kong QZ, Wang JL, Simon AE.** 1997. Satellite RNA-mediated resistance to turnip crinkle virus in Arabidopsis involves a reduction in virus movement. Plant Cell **9:**2051-2063.

159. **Bindewald E, Wendeler M, Legiewicz M, Bona MK, Wang Y, Pritt MJ, Le Grice SF, Shapiro BA.** 2011. Correlating SHAPE signatures with three-dimensional RNA structures. RNA **17:**1688-1696.

160. **Pérez-Cañamás M, Hernández C.** 2015. Key importance of small RNA binding for the activity of a glycine-tryptophan (GW) motif-containing viral suppressor of RNA silencing. J Biol Chem **290:**3106-3120.

161. **Csorba T, Kontra L, Burgyán J.** 2015. viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. Virology **479-480C:**85-103.

162. **Parent JS, Bouteiller N, Elmayan T, Vaucheret H.** 2015. Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. Plant J **81:**223-232.

163. **Qi Y, Zhong X, Itaya A, Ding B.** 2004. Dissecting RNA silencing in protoplasts uncovers novel effects of viral suppressors on the silencing pathway at the cellular level. Nucleic Acids Res **32:**e179.

164. **Gopal A, Zhou ZH, Knobler CM, Gelbart WM.** 2012. Visualizing large RNA molecules in solution. RNA **18:**284-299.

165. **Chapman EJ, Carrington JC.** 2007. Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet **8:**884-896.

166. **Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P.** 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science **309:**1577-1581.

167. **Ding SW, Voinnet O.** 2007. Antiviral immunity directed by small RNAs. Cell **130:**413-426.

168. **Hussain M, Torres S, Schnettler E, Funk A, Grundhoff A, Pijlman GP, Khromykh AA, Asgari S.** 2012. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res **40:**2210-2223.

169. **Hussain M, Asgari S.** 2014. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A **111:**2746-2751.

170. **Ben-Asouli Y, Banai Y, Hauser H, Kaempfer R.** 2000. Recognition of 5'-terminal TAR structure in human immunodeficiency virus-1 mRNA by eukaryotic translation initiation factor 2. Nucleic Acids Res **28:**1011-1018.

171. **Simoes EA, Sarnow P.** 1991. An RNA hairpin at the extreme 5' end of the poliovirus RNA genome modulates viral translation in human cells. J Virol **65:**913-921.

172. **Conrad KD, Niepmann M.** 2014. The role of microRNAs in hepatitis C virus RNA replication. Arch Virol **159:**849-862.

173. **Sullivan CS, Ganem D.** 2005. MicroRNAs and viral infection. Mol Cell **20:**3-7.

174. **Barth S, Pfuhl T, Mamiani A, Ehses C, Roemer K, Kremmer E, Jäker C, Höck J, Meister G, Grässer FA.** 2008. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. Nucleic Acids Res **36:**666-675.

175. **Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D.** 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature **435:**682-686.