REGULATION OF VIRAL SUBGENOMIC mRNA-MEDIATED GENE EXPRESSION IN THREE RNA PLANT VIRUSES

TAMARI CHKUASELI

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO
CANADA

June 2023

ABSTRACT

RNA virus genomes encode different viral proteins that are essential for establishing infections in their hosts. Expression of a subset of these proteins occurs from viral genome-derived subgenomic (sg) mRNAs that are transcribed during infections. The regulation of sg mRNA transcription and translation ensures that requisite amounts of each of these viral proteins are produced at specific stages of the infectious cycle. Higher-order RNA structures present within viral genomes and sg mRNAs are commonly used as molecular switches to achieve the necessary control. The goal of my dissertation was to investigate the structure and function of RNA elements involved in regulating sg mRNA-mediated gene expression in three plus-strand RNA plant viruses: carnation Italian ringspot virus (CIRV), pea enation mosaic virus 2 (PEMV2), and PEMV1. Structural and functional analyses, both in vitro and in cell infections, allowed for the delineation of distinct regulatory RNA elements in each virus. The RNA structures involved in activating sg mRNA transcription were investigated in CIRV and PEMV2. Activation of sg mRNA1 transcription in CIRV requires the formation of a large, complex, intragenomic higher-order RNA structure that assembles via a multistep folding pathway involving six long-distance RNA-RNA base-pairing interactions. In contrast, PEMV2 sg mRNA transcription involves a small RNA stem-loop that contains a self-complementary palindromic loop sequence. Transcription is activated by viral genome dimerization via an intergenomic kissing-loop interaction involving pairing of the palindromic sequences. In PEMV1, the RNA structure required to promote translation readthrough of a C-terminally extended capsid protein from its sg mRNA was investigated. A complex, non-contiguous RNA structure assembled by sequential formation of three long-distance RNA-RNA interactions was found to be required for this recoding event. Collectively, these results have uncovered several distinct regulatory RNA structures involved in controlling different aspects of sg mRNA-mediated viral gene expression and provide novel insights into RNA-based regulation in plus-strand RNA viruses.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDICES

LIST OF TABLES

CHAPTER 3

LIST OF ABBREVIATIONS

3'CITE:	3' cap-independent translational enhancer
AS:	activator sequence
ATS:	attenuation structure
bp:	base pair
BTE:	barley yellow dwarf virus-like translational enhancer
C-:	carboxyl
CE:	core element
CIRV:	carnation Italian ringspot virus
CMoMV:	carrot mottle mimic virus
CMoV:	carrot mottle virus
COMRADES:	cross-linking of matched RNAs and deep sequencing
CP-RTD:	capsid protein-readthrough domain
CP:	capsid protein
DE:	distal element
DL:	downstream linker
DRTE:	distal readthrough element
eIF:	eukaryotic initiation factor
EMSA:	electrophoretic mobility shift assay
eRF:	eukaryotic release factor
FS:	frameshift
g:	genome
HIV-1:	human immunodeficiency virus-1
IRES:	internal ribosome entry site
kb:	kilo base
kDa:	kilo Dalton
kl:	kissing-loop
LD:	large domain
mRNA:	messenger RNA
nm:	nanometer
nt:	nucleotide
OPMV:	opium poppy mosaic virus
ORF:	open reading frame

PAGE:	polyacrylamide gel electrophoresis
PasUV1:	pastinaca umbravirus 1
PCR:	polymerase chain reaction
PEMV:	pea enation mosaic virus
PLRV:	potato leafroll virus
poly(A):	polyadenylated
Pr:	promoter
PRTE:	proximal readthrough element
PT:	premature termination
PTE:	panicum mosaic virus-like translational enhancer
RACE:	rapid amplification of cDNA ends
RCNMV:	red clover necrotic mosaic virus
RdRp:	RNA-dependent RNA polymerase
RNA:	ribonucleic acid
rRNA:	ribosomal RNA
RS:	receptor sequence
RT:	programmed stop codon/translational readthrough
RTSL:	readthrough SL
SARS-CoV:	severe acute respiratory syndrome coronavirus
SbaVA:	strawberry-associated virus A
SDS-PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE:	spacer element
seq:	sequence
sg:	subgenomic
SHAPE:	selective 2'-hydroxyl acylation analyzed by primer extension
SL:	stem-loop
SPLASH:	sequencing of psoralen crosslinked, ligated, and selected hybrids
TBSV:	tomato bushy stunt virus
TGEV:	transmissible gastroenteritis virus
tRNA:	transfer RNA
TRS:	transcription regulatory sequence
TSS:	T-shaped structure
UL:	upstream linker
UTR:	untranslated region

VPg:	viral protein genome-linked
wge:	wheat germ extract
wt:	wildtype

CHAPTER 1

INTRODUCTION

1.1 Plant-infecting plus-strand RNA viruses

Viruses are molecular parasites known to infect all cellular life forms and are divided into different classes based on nucleic acid type and the format of their genomes (Koonin et al., 2020). The largest proportion of eukaryote-infecting viruses possess plus-strand RNA genomes (Koonin et al., 2020). Among these, are important pathogens of animals and plants, such as coronaviruses and potyviruses, respectively (Ahlquist, 2006; Newburn and White, 2015; Koonin et al., 2020). Plus-strand RNA viruses are characterized by virions consisting of single-stranded, coding-sensed RNA genomes that are packaged into icosahedral or helical protein shells (termed capsids), with some surrounded by lipid envelopes (**Figure 1, top**) (Ahlquist, 2006; Newburn and White, 2015; Wolf et al., 2018; Koonin et al., 2020). Being messenger-sensed, these viral RNA genomes are directly translatable upon entering a host cell, and this allows for production of essential early proteins, such as viral RNA-dependent RNA polymerases (RdRps) that catalyze viral RNA synthesis (Ahlquist, 2006; Wolf et al., 2018).

To initiate their infectious cycle and induce disease, virions of plus-strand RNA viruses must first enter their host cells (**Figure 1, (i)**). Entry can occur in different ways depending on the virus and the host cell, such as through cell membrane fusion or endocytosis for animal viruses or through breaching of the cell wall and membrane for plant viruses (Ahlquist, 2006; White, 2011; Newburn and White, 2015). Once inside the cell, the virion disassembles to release its plus-strand genome into the cytosol (**Figure 1, (ii)**) (Ahlquist, 2006; Newburn and White, 2015). There, eukaryotic translation machinery translates RdRp and other replication-associated proteins from the genome to assemble a viral replicase complex, which generally also includes host proteins (**Figure 1, (iii)**) (Ahlquist, 2006; Newburn and White, 2015; Nagy, 2016).
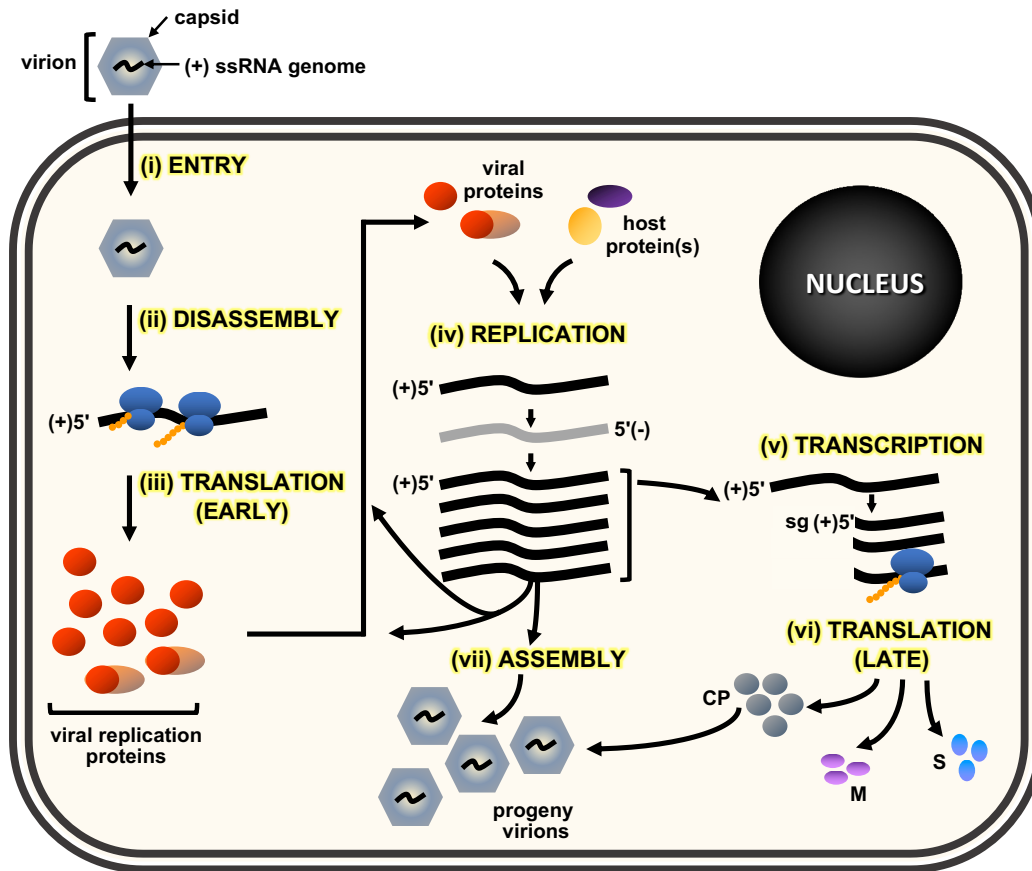
Figure 1: Simplified plus-strand RNA plant virus infectious cycle. Steps in the infectious cycle are labeled as **(i) – (vii)**. **(i)** Entry of the virion. **(ii)** Disassembly of the virion and release of the viral plus-strand (+) RNA genome into the cytoplasm. **(iii)** Translation of early proteins, i.e. viral replication proteins, from the genomic RNA. **(iv)** Replication of the viral genome mediated by viral and host proteins via a minus-strand RNA intermediate [(grey, (-)] and accumulation of progeny genomes. **(v)** Transcription of subgenomic mRNAs (sg) from the genome. **(vi)** Expression of late stage, accessory viral proteins, including the capsid protein (CP), movement protein (M), and host immune system suppressor protein (S). **(vii)** Packaging of the viral RNA genomes by CP into progeny virions. Not to scale. See text for details. Adapted from White (2011).

Many plus-strand RNA plant viruses express their RdRps via translational recoding mechanisms, such as translational readthrough or translational frameshifting, which are discussed in subsequent sections (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021). After RdRp protein synthesis, genome replication commences with the formation of virus replication centers that occur in association with select cellular membranes (e.g. endoplasmic reticulum or mitochondria) (**Figure 1, (iv)**) (Newburn and White, 2015; Zhang et al., 2019). Genome replication is catalyzed by the viral RdRp, the core of the replicase complex, and occurs through the synthesis of a full-length minus-strand RNA copy of the genome that is then used as a template to synthesize multiple copies of the RNA genome, termed progeny genomes (Ahlquist, 2006; Newburn and White, 2015; Venkataraman et al., 2018). Progeny genomes then serve as templates for further genome replication, viral protein translation, or, in some cases, transcription of smaller viral messages, called subgenomic (sg) mRNAs (**Figure 1, (v)**) (Ahlquist, 2006; Newburn and White, 2015). In plus-strand RNA plant viruses, the sg mRNAs that are transcribed encode capsid proteins or other accessory viral proteins involved in cell-to-cell and systemic movement in plant hosts or viral defence proteins that deactivate the plant antiviral immune system (Ahlquist, 2006; Newburn and White, 2015). The translation and accumulation of capsid protein then allows for packaging of progeny viral RNA genomes into progeny virions (**Figure 1 (vi)**) (Ahlquist, 2006; Newburn and White, 2015; Ye et al., 2021).

1.2 Gene expression strategies of plus-strand RNA plant viruses

One of the key steps in the infectious cycle of plus-strand RNA viruses is the initial translation of the RdRp from the invading viral genome by the cellular translation machinery (Ahlquist, 2006; Newburn and White, 2015). Some viral plus-strand RNA genomes harbor a 5'-cap and a 3'-poly(A) tail at their termini, and thus are translated in a manner similar to eukaryotic cellular mRNAs (Jaafar and Kieft, 2019; Geng et al., 2021). Canonical eukaryotic translation

involves the recognition of the 5'-cap structure in cellular mRNAs by eukaryotic initiation factor 4E (eIF4E). Cap-bound eIF4E is complexed with eIF4G and helicases eIF4A/B (collectively known as eIF4F), and along with the poly(A) tail bound by poly(A) binding protein lead to recruitment of the 43S preinitiation ribosome subunit complex to the 5'-end of the message (Aitken and Lorsch, 2012). The preinitiation complex scans in the 5'-to-3' direction in search of a start codon in a favorable Kozak context (i.e. G/AXX**AUG**G), where it pauses, and the 60S subunit joins to form the 80S initiation complex and translation commences (Kozak, 1987; Dever et al., 2018). Translation then continues until a stop codon is encountered by translating ribosomes, which directs translation termination mediated by eukaryotic release factors (Hellen, 2018).

A large proportion of plus-strand RNA plant virus genomes lack either a 5'-cap or a 3'-poly(A) tail, or both (Jaafar and Kieft, 2019; Geng et al., 2021). Thus, to recruit eukaryotic translation machinery these viruses must rely on cap-independent and/or poly(A) tail-independent translation initiation strategies (Jaafar and Kieft, 2019; Geng et al., 2021). In addition to lacking typical eukaryotic mRNA end modifications, the majority of plus-strand RNA plant viruses are polycistronic, that is, they encode more than one open reading frame (ORF) within their genomes (Wolf et al., 2018; Koonin et al., 2020). In this scenario, downstream ORFs in the viral RNA genome cannot be easily accessed by 5'-end-dependent eukaryotic 43S ribosomal subunits, thus they are not translated when situated in such genomic contexts (Aitken and Lorsch, 2012). To overcome this challenge, viruses have developed alternative gene expression and coding strategies, which include internal ribosome entry sites (IRESes), 3' cap-independent translation enhancers (CITEs), translational recoding (frameshifting and readthrough), leaky scanning, polyprotein ORFs, segmented genomes, and subgenomic mRNAs (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021). Each of these alternative gene expression strategies are described in more detail below.

1.2.1 Internal ribosome entry sites (IRESes)

IRESs are RNA elements that can mediate translation initiation in a cap- and 5'-end-independent manner by recruiting ribosomes close to a start codon (**Figure 2A**) (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021). IRESs are found among both plant and animal viruses and can vary greatly in size and RNA structure complexity, with those of plant viruses being smaller and less complex (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021). For example, encephalomyocarditis picornavirus contains an elaborate ~450 nt-long multidomain IRES at the 5'-end of its plus-strand RNA genome that directs translation of its encoded polyprotein ORF by recruiting different translation initiation factors (Martínez-Salas et al., 2015). In comparison, a plant virus turnip crinkle virus (family Tombusviridae) possesses a 116 nt-long, largely unstructured IRES that allows for limited early expression of its 3'-proximally encoded capsid protein directly from its polycistronic genome (May et al., 2017). In the case of insect-infecting dicistroviruses, which have genomes that encode two separate ORFs, each ORF has a different IRES upstream of it that allows for defined control of protein production (Wilson et al., 2000).

1.2.2 3' cap-independent translation enhancers (3'CITEs)

3'CITEs are RNA secondary structures located in the 3'UTRs of plus-strand RNA genomes that mediate cap-independent translation initiation at the 5'-end of these genomes (Truniger et al., 2017; Jaafar and Kieft, 2019) (**Figure 2B**). Unlike IRESs, 3'CITEs are cap-independent, but 5'-end dependent (Guo et al., 2001; Rakotondrafara et al., 2006; Nicholson and White, 2008). 3'CITEs are uniquely found among plant-infecting plus-strand RNA viruses lacking 5'-caps and 3'-poy(A) tails, typified by the large plus-strand RNA plant virus family Tombusviridae (collectively called tombusvirids), in which all 18 genera contain 3'CITEs (Truniger et al., 2017; Jaafar and Kieft, 2019). 3'CITEs recruit the translational machinery by
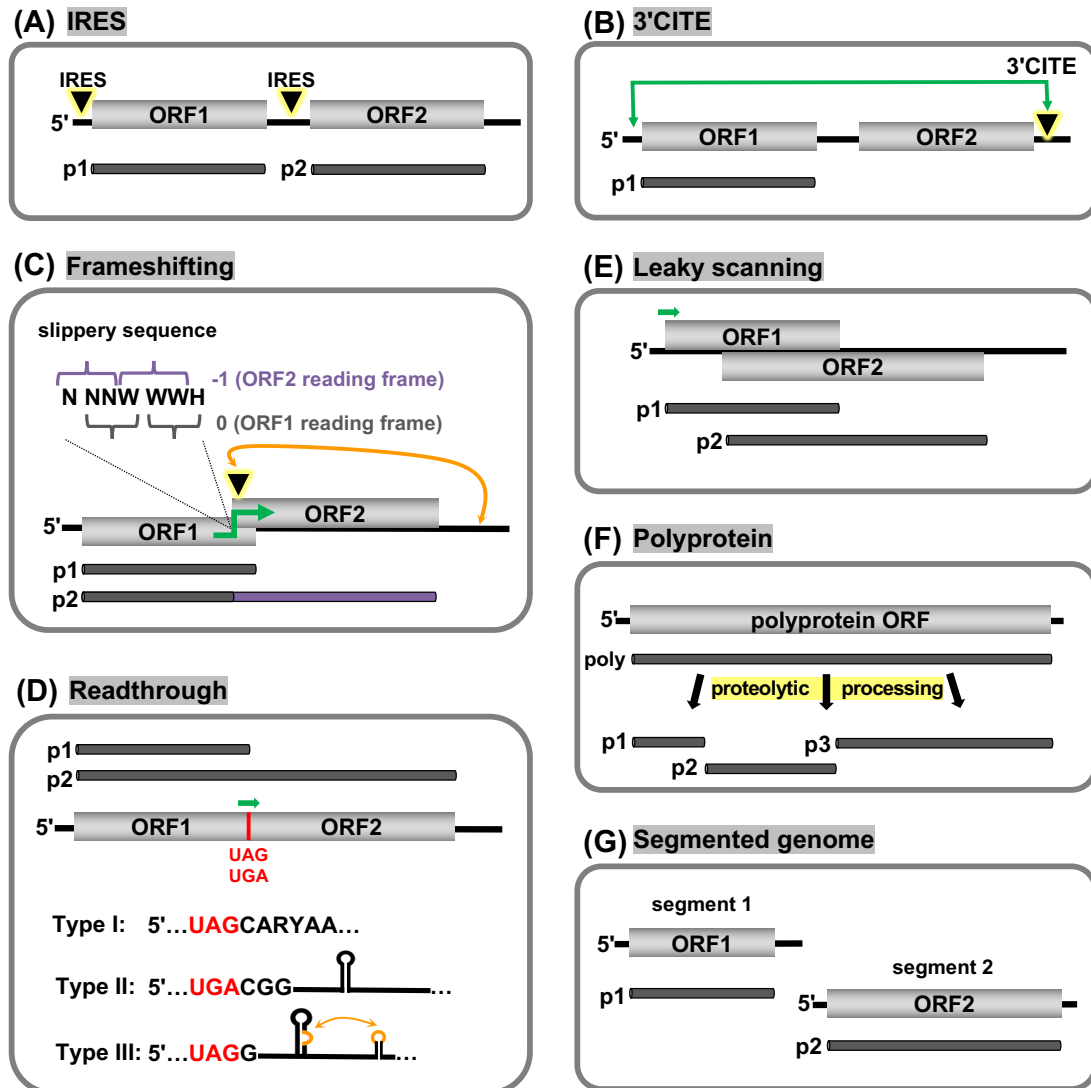
Figure 2: Alternative gene expression strategies of plus-strand RNA viruses. Graphical representations of hypothetical plus-strand viral genomes are shown with grey boxes representing encoded open reading frames (ORFs) and dark grey bars depicting corresponding protein products. **(A)** Internal ribosome entry site (IRES), an RNA element located near the 5'-end and/or internally, that recruits ribosomes for translation in a cap-independent manner. **(B)** 3' cap-independent translation enhancer (3'CITE) RNA structure in the 3'UTR that recruits translation initiation factors or ribosomes and can relocate them to the 5'-end of the genome for translation initiation via an RNA-RNA interaction (green double headed arrow). **(C)** Programmed ribosome frameshifting, where a translating ribosome shifts its reading frame due to a slippery sequence and a downstream RNA secondary structure (black triangle) to produce a C-terminally extended fusion protein (p2); sometimes requiring a long-distance RNA-RNA interaction (orange double-headed arrow). The conserved slippery sequence is shown and reading frames are indicated. **(D)** Programmed stop codon readthrough, where a translating ribosome decodes a stop codon as a sense codon and continues translation to produce a C-terminally extended protein (p2). Three types of readthrough elements are shown. **(E)** Leaky scanning, where a proportion of ribosomes bypass an upstream ORF's start codon and initiate translation at a downstream initiation codon. **(F)** Polyprotein synthesis, where one long polypeptide is translated that is proteolytically cleaved into individual viral proteins. **(G)** Segmented genome, where a viral genome is composed of multiple segments, each encoding a different ORF(s). Not to scale. See text for additional details.

interacting directly with eukaryotic initiation factors or ribosomal subunits, which are then

repositioned to the 5'-end of the genome where 43S scanning and subsequent translation

initiation takes place (Truniger et al., 2017; Jaafar and Kieft, 2019). Repositioning of the 3'CITE-

bound initiation factors to the 5'-end of the viral genome is typically mediated by a 5'UTR-3'CITE

RNA-RNA base-pairing interaction (Figure 2B) (Guo et al., 2001; Nicholson and White, 2008;

Chattopadhyay et al., 2011; Gao et al., 2012; Blanco-Pérez et al., 2016; Ilyas et al., 2021).

Tombusvirids contain different classes of 3'CITEs, and the first to be discovered was the

barley yellow dwarf virus translational enhancer, or BTE (Wang et al., 1997; Truniger et al.,

2017). BTEs are characterized by a long basal stem with two, three, or five helices radiating

from the top, giving them a cloverleaf-like shape (Wang et al., 1997; Mizumoto et al., 2003;

Shen and Miller, 2004; Ilyas et al., 2021). BTEs also include a conserved 17 nt consensus

sequence that is essential for eIF4G binding (Treder et al., 2008; Kraft et al., 2013; Simon and

Miller, 2013). As with most 3'CITEs, the BTE-bound eIF4G/4E complex is repositioned to the

5'UTR through a long-distance RNA-RNA base-pairing interaction (Guo et al., 2001; Treder et

al., 2008; Kraft et al., 2013; Sharma et al., 2015; Ilyas et al., 2021). In contrast, for certain

3'CITE-containing viruses, no apparent 5'-to-3' base-pairing interactions exist, leaving open the

question of how the 3'CITE-bound translation initiation factors promote translation at the 5'-end

of these viral genomes (Sarawaneeyaruk et al., 2009; Chkuaseli et al., 2015).

1.2.3 Translational recoding mechanisms

In order to initiate an infection, plus-strand RNA viruses must translate their RdRps

shortly after entering a host cell (Ahlquist, 2006; Wolf et al., 2018). In many cases, the RdRp is

encoded as the second ORF in the viral genome and is expressed through either programmed

ribosomal frameshifting or programmed stop codon readthrough (Firth and Birerley, 2012; Geng

et al., 2021; Penn and Mukhopadhyay, 2022). Programmed ribosomal frameshifting (FS) occurs

when a small proportion of elongating ribosomes shift their original reading frame either in the

-1, -2 or +1 direction upon encountering specific nucleotide sequences and a higher-order RNA

structure (Firth and Brierley, 2012; Geng et al., 2021; Penn and Mukhopadhyay, 2022). The

ribosomes then continue translation elongation in the new reading frame to produce a fusion

protein consisting of an N-terminal pre-frameshift portion in the original reading frame and a C-

terminal post-frameshift extension in a new reading frame (**Figure 2C**). Programmed stop codon

readthrough (RT), on the other hand, occurs when a proportion of terminating ribosomes do not

terminate at the first ORF's stop codon due to a specific RNA sequence or higher-order RNA

structure (Firth and Brierley, 2012; Geng et al., 2021). Instead, the stop codon is decoded as a

sense codon by a near cognate tRNA and translation proceeds in the first ORF's original

reading frame (**Figure 2D**). The resulting readthrough protein consists of the pre-readthrough

protein and a C-terminal readthrough extension.

Accordingly, both RT and FS result in proteins that have extended C-termini.

Importantly, these extended products have distinct functions compared to their pre-FS or pre-RT

counterparts and are expressed at defined, lower frequencies (Firth and Brierley, 2012; Geng et

al., 2021; Penn and Mukhopadhyay, 2022). Utilizing FS and RT mechanisms thus allows viruses

to fine-tune the amounts of RdRp synthesized during infections. Certain plus-strand RNA plant

viruses also use the RT strategy to produce C-terminally extended capsid proteins with unique

functions (Firth and Brierley, 2012; Geng et al., 2021). Additional information on the FS and RT

processes is presented in the sections below.

1.2.3.1 Frameshifting

The majority of RNA viruses that rely on FS use the -1 FS strategy to express their

RdRps (Firth and Brierley, 2012; Geng et al., 2021; Penn and Mukhopadhyay, 2022). The

efficiency of this process depends on two RNA elements: (i) a consensus slippery sequence,

where ribosomes change their frame, and (ii) a downstream, frameshift stimulating higher-order

RNA structure (Firth and Brierley, 2012) (**Figure 2C**). The slippery sequence follows the

consensus 5'-N<u>NNWW</u>WWH, where the original reading frame is underlined, NNN can be any

three identical nucleotides, WWW is either UUU or AAA, and H is either C, U, or A (Firth and

Brierley, 2012). The frameshift stimulatory higher-order RNA structure can differ greatly among

different RNA viruses; in poleroviruses it is a localized pseudoknot structure (Kim et al, 2000;

Cornish et al., 2005), while in tombusvirids an extended SL structure must base-pair with distal

genomic sequences via a long-distance RNA-RNA interaction (Barry and Miller, 2002; Tajima et

al., 2011; Gao and Simon, 2016). In all cases, the higher-order RNA structure is thought to stall

the elongating ribosome at the slippery site and promote ribosome slippage by resisting the

ribosome's RNA unwinding activity (Qu et al., 2011; Firth and Brierley, 2012; Tholstrup et al.,

2012; Allan et al., 2023). In addition to RNA stimulatory elements, viral and host proteins can

also play a role in promoting frameshifting, as demonstrated in animal plus-strand RNA viruses

(Allan et al., 2023). For example, viral protein 2A, encoded by encephalomyocarditis

picornavirus, binds to a small SL structure downstream of its polyprotein FS slippery site to

promote ribosome slippage and allows for 2A concentration-dependent temporal regulation of

frameshifting (Napthine et al., 2017). In the case of porcine reproductive and respiratory

syndrome arterivirus, a viral and a host protein form a complex with a short linear RNA

sequence downstream of its slippery site that promotes -1 and -2 frameshifting (Li et al., 2014;

Napthine et al., 2016; Patel et al., 2020).

1.2.3.2 Readthrough

Similar to frameshifting, programmed translational readthrough is a common mechanism

used by plus-strand RNA viruses to express RdRp genes, although some viruses express C-

terminally extended capsid proteins by this process (Firth and Brierley, 2012). RT efficiency is

influenced by the identity of the stop codon present, its flanking nucleotides, and in many cases, higher-order RNA structures (Firth and Brierley, 2012). RT sites in RNA viruses are dominated by UGA and UAG stop codons, which are naturally more permissive to readthrough than UAA (Beier and Grimm, 2001; Firth and Brierley, 2012). Based on the stop codon identity and the nature of the downstream RT stimulatory RNA elements, RT signals have been divided into three broad categories: type I, type II, and type III (**Figure 2D**). Type I is characterized by a 6-nt long conserved sequence, 5'-CARYAA (R represents A or G, Y represents C or U), located immediately downstream from a UAG stop codon (Skuzeski et al., 1991). This conserved element is believed to function at the primary sequence level to promote RT through a yet unknown mechanism (Skuzeski et al., 1991; Beier and Grimm, 2001; Firth and Brierley, 2012). Viruses harboring type II RT signals typically contain a UGA stop codon followed by a CGG motif and a downstream RNA SL structure, as exemplified by animal-infecting alphaviruses and plant-infecting furoviruses (Firth et al., 2011). Lastly, the type III readthrough element consists of a UGA stop codon, followed by a G and a SL structure immediately downstream, as well as a long-distance RNA-RNA interaction between the SL and a complementary 3'-distal sequence (Firth and Brierley, 2012). Type III RT is common among plant-infecting tombusvirids and poleroviruses, with the former group producing their RdRps and the latter generating C-terminally extended capsid proteins (Brown et al., 1996; Cimino et al., 2011; Newburn et al., 2014; Kuhlmann et al., 2016; Xu et al., 2018). How the readthrough stimulatory RNA elements of type II and type III categories promote stop codon decoding by ribosomes is unknown (Brown et al., 1996; Cimino et al., 2011; Firth et al., 2011; Newburn et al., 2014; Kuhlmann et al., 2016; Xu et al., 2018). However, it has been proposed that the RT stimulatory elements may physically stall ribosomes near the stop codon, similar to FS stimulatory elements, and make specific contacts with ribosomes that alter their activity, or interfere with the ability of eukaryotic release factors to bind to the stop codon and direct termination (Firth and Brierley, 2012). Additionally, similar to FS, protein factors can also regulate RT efficiency. In Moloney murine leukemia

retrovirus, the viral reverse transcriptase precursor protein interacts with eukaryotic release factor 1 (eRF1), which prevents eRF1 from binding to the stop codon and provides increased opportunity for non-cognate tRNA binding (Orlova et al., 2003).

1.2.4 Leaky scanning

Many plus-strand RNA viruses encode ORFs that overlap with each other, but in different reading frames (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021) (**Figure 2E**). To express a downstream overlapping ORF, leaky scanning is often involved where a proportion of scanning ribosomes bypass the first ORF's initiation codon and initiate translation at the downstream ORF's start codon (Firth and Brierley, 2012; Jaafar and Kieft, 2019; Geng et al., 2021). This bypass can occur either because the start codon of the upstream ORF is in a weak Kozak context or the start codon of the upstream ORF is very close to the 5'-end of the message (Kozak, 1987; 1991; 2002). This process can even involve three ORFs, as in plant potexviruses that encode three proteins in one of their sg mRNAs (Fujimoto et al., 2022). All three proteins are translated by virtue of the start codon for the first ORF being very close to the 5'-end of the message and the second ORF having a suboptimal Kozak consensus (Fujimoto et al., 2022).

1.2.5 Polyprotein ORFs and segmented genomes

Some plus-strand RNA viruses, such as animal-infecting flaviviruses and plant-infecting potyviruses, have evolved to encode a single large polyprotein ORF (**Figure 2F**) (Simmonds et al., 2017; Pasin et al., 2022). In this strategy, these viruses no longer face the challenge associated with having polycistronic genomes, i.e., inaccessibility of downstream ORFs to the eukaryotic translation machinery. Instead, after synthesis of the lone ORF, the large polyprotein

undergoes proteolytic processing, usually carried out by viral and/or host proteases, to yield individual functional viral proteins (**Figure 2F**) (Chambers et al., 1993; Pasin et al., 2022).

Other plus-strand RNA viruses deal with the expression of ORFs by having segmented genomes, where ORFs are distributed on different viral RNAs (**Figure 2G**) (Choi et al., 2002; Choi and Rao, 2003; Basnayake et al., 2006). Making a genome segmented, with each RNA segment encoding a protein, aids in ribosome access, but it also presents a challenge in that all segments are required simultaneously for successful infection. To achieve this, the different genome segments are either packaged within the same virion, as is the case for the bi-segmented tombusvirid red clover necrotic mosaic virus (Basnayake et al., 2006), or packaged separately in their own capsids, like the tri-segmented plant virus brome mosaic virus (Choi et al., 2002; Choi and Rao, 2003). In the latter case, each particle containing one of the genome segments must synchronously infect a cell, and this task is aided by the very high levels that the virus accumulates to during infections (Choi et al., 2002; Choi and Rao, 2003).

1.2.6 Subgenomic mRNAs

Many plus-strand RNA viruses rely on the transcription of subgenomic (sg) mRNAs during infections as a means to express ORFs encoded 3'-proximally in their viral genomes (Miller and Koev, 2000; White, 2002; Jiwan and White, 2011; Sztuba-Solińska, et al., 2011). As alluded to previously, in polycistronic plus-strand RNA genomes, ORFs encoded downstream of one or more ORFs are not easily accessed by 5'-end-loading and scanning 43S ribosomal subunits (Aitken and Lorsch, 2012). Sg mRNAs represent shorter, viral RdRp-transcribed messages that are 3'-coterminal with their cognate viral genomes (**Figure 3A**) (Miller and Koev, 2000; White, 2002; Sztuba-Solińska, et al., 2011). With this gene expression strategy, ORFs that are 3'-proximal in a genomic context are relocated near the 5'-end of sg mRNAs, making the ORFs readily accessible to the eukaryotic translation initiation machinery (**Figure 3A**). Additionally, by
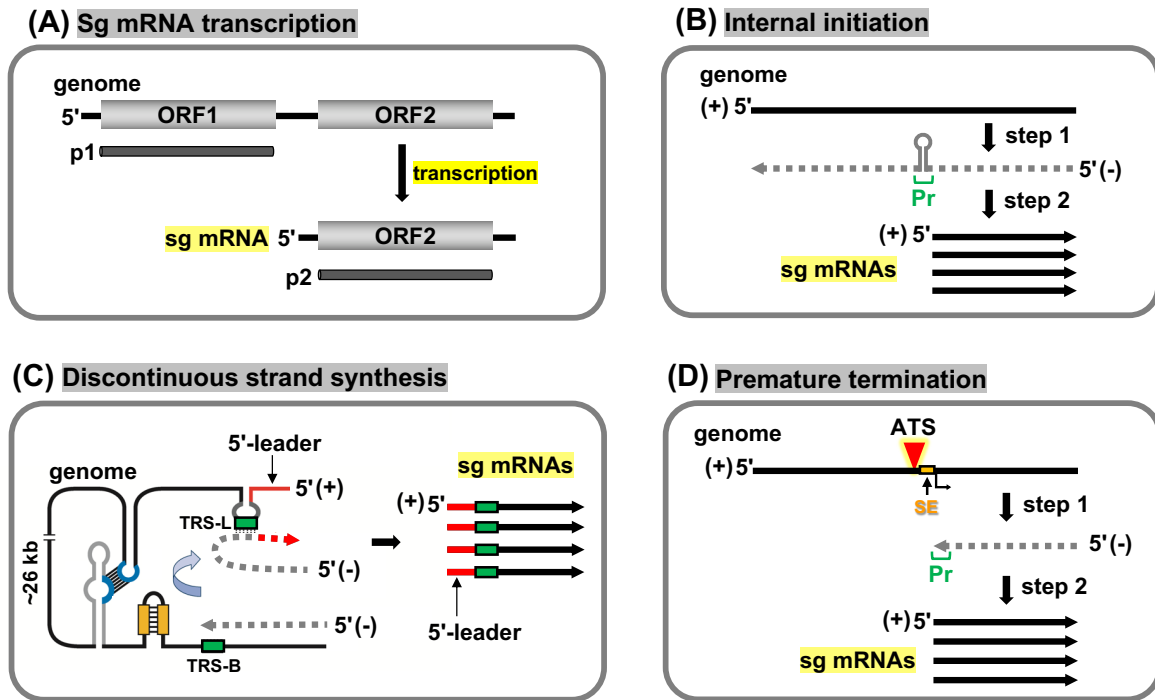
Figure 3: Subgenomic (sg) mRNA transcription mechanisms of plus-strand RNA viruses. (A) Sg mRNA transcription strategy used by plus-strand RNA viruses to express a downstream ORF (p2). **(B)** Internal initiation mechanism of sg mRNA transcription where a full-length minus-strand genome templates sg mRNA transcription using an internal promoter (pr). **(C)** Discontinuous synthesis mechanism of sg mRNA transcription where a minus-strand template for sg RNA transcription is synthesized by "jumping" (blue curved arrow) of the RdRp between TRS-B and RTS-L sequences. The discontinuous template is then used to transcribe a sg mRNA containing both 5'-leader and 3'-portion of the genome. Adapted from Sola et al. (2015). **(D)** Premature termination mechanism of sg mRNA transcription in which the RdRp generates a truncated minus-strand sg RNA due to encountering an attenuation structure (ATS) that forms a physical barrier in the plus-strand genome. Spacer element (SE, orange), between the initiation site and ATS, influences the accuracy of the RdRp's termination. The truncated minus-strand sg RNA generated harbors a promoter sequence (pr, green) at its 3'-end that is used by the RdRp to transcribe multiple 3'-coterminal plus-strand sg mRNAs. Not to scale. See text for details.

regulating the sg mRNA transcription process, the quantity and timing of the sg mRNA-encoded viral proteins can be controlled (Miller and Koev, 2000; White, 2002; Sztuba-Solińska, et al., 2011). Currently, there are three distinct mechanistic models for how sg mRNAs are transcribed during infections (**Figure 3B, C, D**): (i) internal initiation, (ii) discontinuous synthesis, and (iii) premature termination (White, 2002; Jiwan and White, 2011; Sztuba-Solińska et al., 2011).

1.2.6.1 The internal initiation mechanism of sg mRNA transcription

Both plant and animal plus-strand RNA viruses are known to use the internal initiation mechanism to transcribe their sg mRNAs (Miller et al., 1985; Wielgosz et al., 2001; Qiu et al., 2011). During the initial step of genome replication, a full-length minus-strand copy of the genome is generated by the viral RdRp (**Figure 3B, step 1**). This minus-strand genome is then used as a template to produce multiple plus-strand copies of the genome or, alternatively, is used for sg mRNA transcription via internal initiation (**Figure 3B, step 2**) (Miller et al., 1985; Sztuba-Solińska et al., 2011). The latter event is possible because the full-length minus-strand genome carries an internal promoter within its sequence that directs initiation of sg mRNA transcription (Haasnoot et al., 2000; Haasnoot et al., 2002; Sivakumaran et al., 2004; Qiu et al., 2011) (**Figure 3B, red**). Thus, in addition to being able to initiate viral RNA synthesis at the 3'-terminus of the genomic minus-strand (to synthesize the genome), the RdRp is also able to initiate in the middle of the minus-strand RNA sequence to transcribe a sg mRNA (Li and Stollar, 2004; Grdzelishvili et al., 2005; Li and Stollar, 2007). As with all transcriptional mechanisms, the minus-strand template is reused multiple times to produce many copies of sg mRNA (**Figure 3B, A**) (Miller et al., 1985; Haasnoot et al., 2000; Haasnoot et al., 2002; Sivakumar et al., 2004; Qiu et al., 2011).

1.2.6.2 The discontinuous synthesis mechanism of sg mRNA transcription

The discontinuous synthesis mechanism is unique to animal plus-strand RNA viruses belonging to the order Nidovirales, which includes coronaviruses (Sztuba-Solińska et al., 2011; Sola et al., 2015; Allan et al., 2023). These viruses transcribe several nested 3'-coterminal sg mRNAs during infections that allow for the expression of multiple 3'-proximally encoded structural proteins (Woo et al., 2023; Brinton et al., 2021). The distinguishing characteristic of these sg mRNAs is that, in addition to being 3'-coterminal with the genome, their 5'-ends also contain a short sequence identical to the 5'-end of the viral genome, termed the 5'-leader sequence (**Figure 3C, red**) (Sztuba-Solińska et al., 2011; Sola et al., 2015; Malone et al., 2022; Allan et al., 2023). The 5'-leader sequence is acquired through discontinuous transcription during minus-strand synthesis of the genome, where the viral RdRp initiates transcription at the 3'-end of the genome, but instead of generating a full-length minus-strand copy of the genome, it terminates after copying an internal transcription regulatory sequence, termed TRS-B (**Figure 3C**) (Pasternak et al., 2001; Zúñiga et al., 2004). In transmissible gastroenteritis coronavirus (TGEV), this termination step is mediated by an RNA structure located upstream from the termination site (**Figure 3C, yellow**) (Moreno et al., 2008; Mateos-Gómez, et al., 2011). After termination, the RdRp, together with the nascent minus-strand containing the TRS-B sequence at its 3'-end, re-primes on a complementary TRS-L element near the 5'-end of the genome, and synthesis resumes to copy the leader sequence (**Figure 3C, curved arrow**) (Van Marle et al., 1999; Sola et al., 2015; Allan et al., 2023). The net result of this RdRp "jumping" is that terminal regions of the genome are joined, while a large internal segment is omitted (Marle et al., 1999; Zúñiga et al., 2004; Sola et al., 2005). In the case of TGEV, a long-distance RNA-RNA interaction spanning 26 kb helps to position the TRS-B close to the TRS-L, which facilitates the RdRp jumping step (**Figure 3C, blue**) (Mateos-Gómez, et al., 2013). The "spliced" minus-strand sg mRNA template generated is then used repeatedly to transcribe plus-strand counterparts

through RdRp initiation at the 3'-end (**Figure 3C**) (Sola et al., 2015; Malone et al., 2022; Allan et al., 2023). For viruses with multiple 3'-encoded ORFs in their genomes, there is a TRS-B located just upstream from each ORF so that a corresponding sg mRNA is transcribed for each encoded protein (Zúñiga et al., 2004; Sola et al., 2015).

1.2.6.3 The premature termination mechanism of sg mRNA transcription

The premature termination (PT) mechanism resembles the discontinuous synthesis mechanism but excludes the template switching step. This process has been documented in both plant and animal plus-strand RNA viruses, including members of the large plant virus family Tombusviridae (i.e. tombusvirids) (White, 2002; Jiwan and White, 2011; Sztuba-Solińska et al., 2011). The PT mechanism consists of two major steps: minus-strand RNA template synthesis and plus-strand sg mRNA transcription (**Figure 3D**) (White, 2002). During step 1, the viral RdRp initiates minus-strand synthesis at the 3'-end of a plus-strand viral genome; however, instead of synthesizing a full-length genomic minus-strand, the RdRp terminates prematurely when it encounters a higher-order RNA structure, termed the attenuation structure (ATS) (**Figure 3D, ATS**) (White, 2002; Jiwan and White, 2011). This termination leads to the production of a truncated minus-strand RNA that is the same size as the plus-strand sg mRNA, and a promoter (pr) sequence is present at its 3'-end (**Figure 3D, step 1, pr**). During step 2, this promoter sequence is recognized by the RdRp, and plus-strand sg mRNAs are synthesized, which in turn allow for translation of 3'-encoded viral proteins (**Figure 3D, A**).

Formation of an active RNA attenuation structure is a rate limiting step in the process because it determines the efficiency of RdRp termination during step 1 (White, 2002; Jiwan and White, 2011). An AU-rich sequence located immediately downstream of the initiation site in the genome facilitates RdRp stalling and/or dissociation, and also corresponds to the transcriptional promoter at the 3'-end of the minus-strand sg RNA template (White, 2002; Jiwan and White,

16

2011). Among viruses that utilize PT, RNA attenuation structures vary in sequence and range in

structural complexity (White, 2002; Jiwan and White, 2011). In all cases, thermodynamic

stability of the RNA attenuation structure in the viral genome plays an important role in its ability

to physically block the progression of the RdRp (Lin et al., 2007; Wang et al., 2008; Jiwan and

White, 2011). Attenuation structures always include an RNA helix positioned just upstream from

the transcriptional initiation site, which forms by either local, long-distance, or *in trans* base-

pairing interactions (Sit et al., 1998; Choi and White, 2002; Lindenbach et al., 2002; Lin and

White, 2004; Xu and White, 2008; Xu and White, 2009; Wu et al., 2010; Jiwan et al., 2011;

Blanco-Pérez and Hernández, 2016). It is this helix that the leading edge of the RdRp

encounters during minus-strand synthesis, contributing to RdRp stalling and dissociation (White,

2002; Jiwan and White, 2011). The distance between this inhibitory helix and the sg mRNA

transcription initiation site, termed the spacer element, also plays an important role during step 1

of PT (**Figure 3D, SE, orange**) (Lin et al., 2007; Wang et al., 2008). Spacer lengths vary

between viruses, with the typical length ranging between 2-4 nucleotides (White, 2002; Jiwan

and White, 2011). Notably, altering this length negatively impacts the RdRp's ability to terminate

accurately after copying up to and including the transcription initiation site (Lin and White, 2004;

Lin et al., 2007; Wu and White, 2007; Wang et al., 2008; Jiwan et al., 2011). Proper termination

during minus-strand sg mRNA synthesis generates the functional 3'-terminal promoter sequence

(pr), which is then used repeatedly by the viral RdRp to transcribe multiple sg mRNAs (Choi and

White, 2002; Lin et al., 2007; Wu et al., 2010; Jiwan et al., 2011). It has been shown that the

viral RdRps using a PT mechanism initiate very poorly if the sg mRNA promoter sequence is

located internally in the minus-strand RNA template (Panavas et al., 2003; Lin et al., 2007).

Thus, the PT mechanism provides a way to generate a minus-strand template that contains the

transcriptional promoter sequence at the 3'-terminus, allowing for efficient RdRp initiation.

1.3 Carnation Italian ringspot virus (CIRV)

CIRV is a plus-strand RNA plant virus in the genus Tombusvirus within the family Tombusviridae (Sit and Lommel, 2015). Tombusviruses are distinct because they regulate multiple critical steps of their infectious cycle through different higher-order RNA structures and dynamic long-range RNA-RNA interactions in their genomes (Newburn and White, 2015; Chkuaseli and White, 2018) (**Figure 4**). Many aspects of tombusvirus infections have been studied using CIRV, as well as another genus member, tomato bushy stunt virus (TBSV), and both of these viruses have served as important model systems (White and Nagy, 2004; Yamamura and Scholthof, 2005; Sit and Lommel, 2015). CIRV forms non-enveloped 30 nm icosahedral particles composed of 180 subunits of viral capsid protein (Hollings et al., 1970; Olson et al., 1983; Rubino et al., 1995). Packaged inside CIRV virions is a single plus-strand RNA genome that is ~4.8 kb in length and encodes five functional ORFs (**Figure 4A**) (Rubino et al., 1995; White and Nagy, 2004). The 5'-proximal ORF directs translation of the p36 auxiliary replication protein, and readthrough of its stop codon generates p95, the RdRp (Cimino et al., 2011). Both of these proteins are translated directly from the genome and are essential for viral genome replication and sg mRNA transcription (Scholthof et al. 1995a; Oster et al., 1998; Rubino et al., 2001; Pantaleo et al., 2003; 2004; Hwang et al., 2008; Gunawardene et al., 2015). The next protein encoded in the genome is the capsid protein (p41), the sole structural protein in CIRV virions (Rubino et al., 1995; Olson et al., 1983; White and Nagy, 2004). The 3'-proximal overlapping ORFs are in different reading frames and encode p22 and p19, which are involved in mediating cell-to-cell movement and suppressing antiviral RNA silencing, respectively (Rubino et al., 1995; Scholthof et al., 1995b; Chu et al., 1999; Vargason et al., 2003; Ye et al., 2003; Lozsa et al., 2008; Várallyay et al., 2010).

(A) Long-range RNA-RNA base-pairing interactions in the CIRV genome
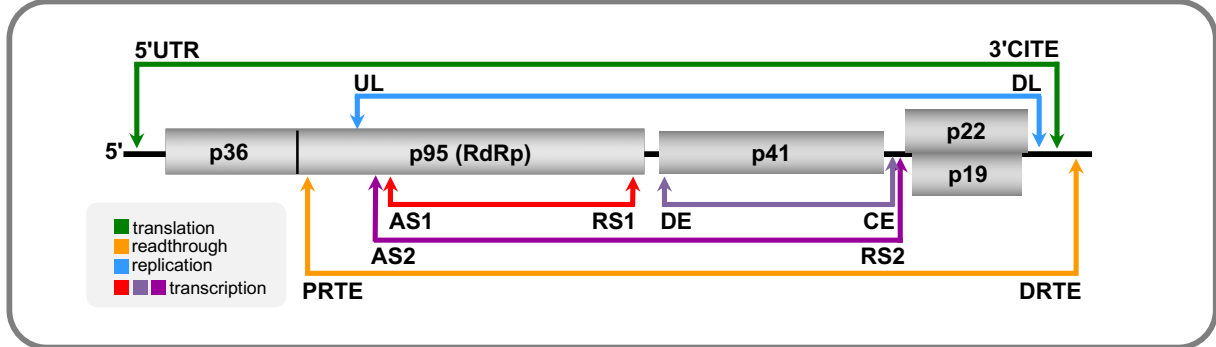
(B) Long-range interactions regulating translation in CIRV
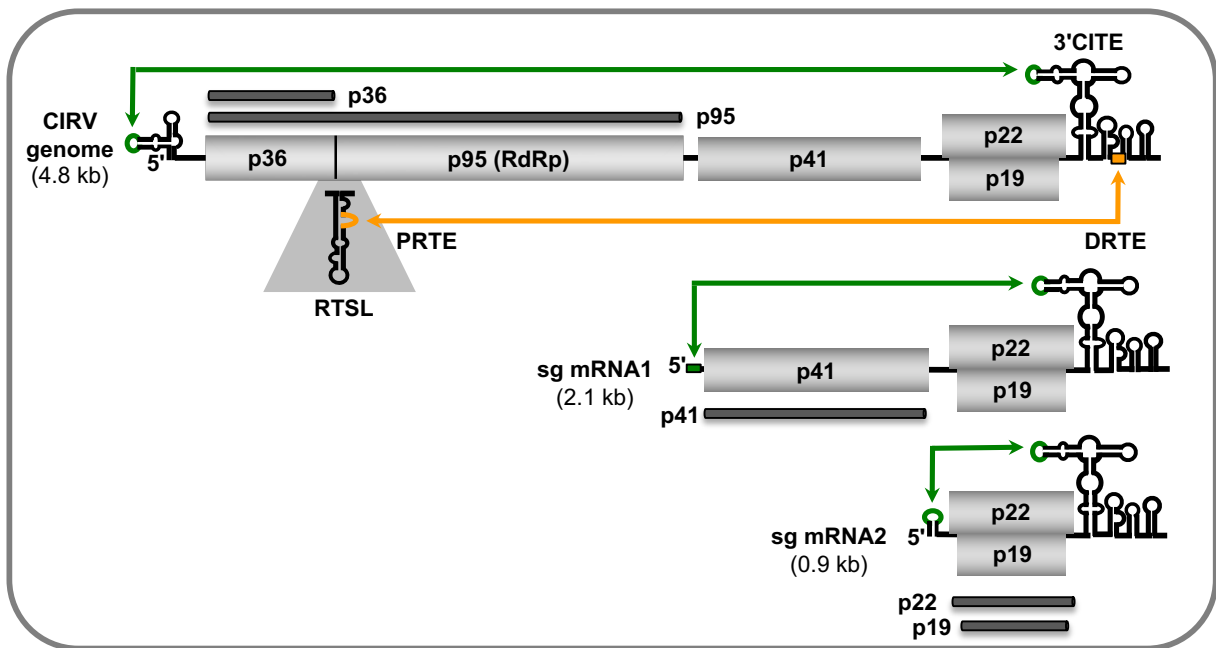
(C) A long-range interaction regulating genome replication in CIRV
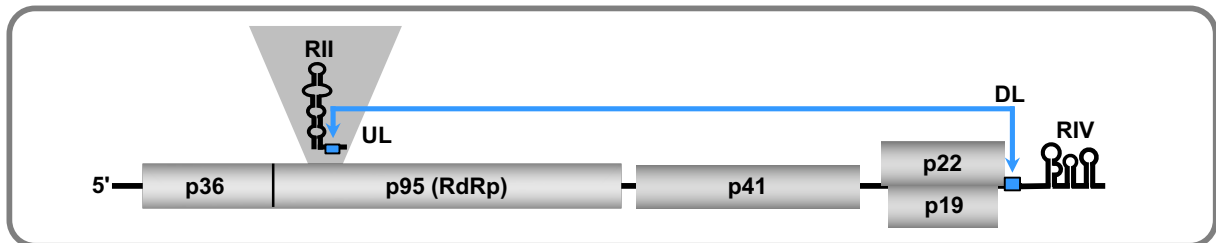
Figure 4: Genome organization, regulatory RNA elements, and long-range interactions in carnation Italian ringspot virus (CIRV). (A) CIRV genome with ORFs depicted as grey boxes. Colour-coded arrows indicate different long-range RNA-RNA interactions that regulate specific processes in the infectious cycle (see legend). **(B)** Long-range interactions regulating cap-independent translation in the CIRV genome and its sg mRNAs, as well as programmed translational readthrough generating p95. **(C)** Long-range interactions regulating genome replication. See text for details. Adapted from Chkuaseli and White (2018).

1.3.1 CIRV gene expression

CIRV's plus-strand RNA genome is polycistronic and lacks both a 5'-cap and a 3'-poly(A) tail; thus, it relies on an unconventional translational strategy for expression of its proteins (Rubino et al., 1995; White and Nagy, 2004). The CIRV genome contains a Y-shaped 3′CITE in its 3'UTR (**Figure 4B**) (Wu and White, 1999; Fabian and White, 2004; 2006; Nicholson and White, 2008) that binds to eukaryotic translation initiation complex eIF4F (Nicholson et al., 2010; Nicholson et al., 2013). Through an intragenomic long-distance RNA-RNA interaction with a complementary sequence in the 5'UTR of the genome, the 3'CITE relocates the 3′CITE-bound eIF4F to the 5'-end, where it promotes ribosome recruitment and translation of p36 and p95 (Nicholson et al., 2010; Nicholson et al., 2013) (**Figure 4B, green**).

The p95 RdRp is synthesized via a translational RT mechanism in which the stop codon for p36 is decoded as a sense codon (**Figure 4B; Figure 2D**) (Rubino et al., 1995; Nicholson and White, 2008; Cimino et al., 2011; Firth and Brierly, 2012). Efficiency of this decoding event in CIRV is dependent on the formation of a large RNA stem-loop structure (RTSL) located just 3' to the stop site, which contains a proximal readthrough element (PRTE) that, for RT to occur, must base-pair with a complementary distal readthrough element (DRTE) in the 3'UTR of the viral genome (**Figure 4B, top, orange**) (Cimino et al., 2011). The RT process is also dependent on the formation of a second long-range intragenomic interaction, involving the base-pairing partner sequences, upstream linker (UL) and downstream linker (DL) (**Figure 4A, blue**) (Cimino et al., 2011). The UL/DL interaction has a dual function because it also plays an essential role in viral genome replication; as will be discussed below (Wu et al., 2009) (**Figure 4C**). Notably, the DRTE in the 3'UTR of the genome is located within an important replication regulatory element called region IV (RIV) (**Figure 4B, top**) (Fabian et al., 2003; Pogany et al., 2003; Li et al., 2010; Sasvari et al., 2011; Pathak et al., 2012). When the PRTE is bound to the DRTE, it not only

activates RT, it also simultaneously inhibits minus-strand RNA synthesis of the genome, which would interfere with translation of p95 from the genome (Cimino et al., 2011).

The more 3'-proximally encoded CIRV proteins, p41, p22, and p19, are translationally silent in the context of the full-length viral genome (White and Nagy, 2004). Instead, they are expressed from two smaller viral messages, sg mRNA1 (~2.1 kb) and sg mRNA2 (~0.9 kb), which are transcribed by the p95 RdRp during viral infections (**Figure 4B**) (Rubino et al., 1995; Qiu and Scholthof, 2001; White and Nagy, 2004). Because the sg mRNAs are 3'-coterminal with the genome, they also contain and are dependent on the Y-shaped 3'CITE for their translation (Nicholson and White, 2008) (**Figure 4B**). The 5'UTRs of both sg mRNAs contain sequences that are complementary to the 3'CITE, thus allowing for repositioning of 3'CITE-bound eIF4F to their 5'-ends and efficient translation of their encoded proteins (**Figure 4B, green**). The p41 capsid protein is translated from the larger sg mRNA1, whereas p22 and p19 are both expressed from the smaller sg mRNA2, with p19 being translated via a leaky scanning mechanism (**Figure 4B; Figure 2E**) (Johnston and Rochon, 1996; White and Nagy, 2004).

1.3.2 CIRV genome replication

CIRV genome replication takes place on mitochondrial membranes, within small membrane invaginations called spherules (Rubino et al., 2001; Pantaleo et al., 2003; 2004; Hwang et al., 2008). Viral genome replication is a multistep process that involves both viral (p36 and p95) and cellular proteins, as well as regulatory RNA elements within the genome (White and Nagy, 2004; Nagy, 2016; Gunawardene et al., 2017). One of the key steps in the replication process is the assembly of viral replicase complexes that are responsible for synthesizing viral RNAs (Nagy, 2016). Replicase assembly is dependent upon two RNA elements within the viral genome termed RII-SL, an internally located RNA replication structure in the p95 ORF, and RIV in the 3'UTR (**Figure 4C**) (Panaviene et al., 2005; Monkewich et al., 2005; Pogany et al., 2005).

RII mediates direct binding to p36 and p95 replication proteins, while RIV recruits cellular

protein factors that help to reconfigure the 3'-end into an RdRp-accessible conformation

(Monkewich et al., 2005; Pogany et al., 2005; Pogany et al., 2003; Li et al., 2009; 2010; Sasvari

et al., 2011; Pathak et al., 2012). RII and RIV, which are separated by ~3000 nucleotides, are

united by the UL/DL interaction that forms a discontinuous RII-RV RNA platform that is

necessary for the RdRp to initiate minus-strand RNA synthesis from the 3'-end of the plus-

strand genome (**Figure 4C**) (Wu et al., 2009; Pathak et al., 2012). The full-length genomic

minus-strand generated serves as a template to synthesize multiple copies of plus-strand

progeny genomes. Additional *cis*-acting RNA elements (e.g., replication promoters at the 3'-

termini of plus- and minus-strand genomic RNAs) are also involved in facilitating and regulating

the replication process (Wu and White, 1998; Ray and White, 1999; Wu et al., 2001; Panavas et

al., 2002; Panavas and Nagy, 2003; Panavas et al., 2003; Ray and White, 2003; Ray et al.,

2003; Ray et al., 2004; Fabian and White, 2006; Gunawardene et al., 2021).

1.3.3 CIRV sg mRNA transcription

Two sg mRNAs are transcribed during CIRV infections, both of which are generated by a

premature termination (PT) mechanism (**Figure 5A; Figure 3D**) (Rubino et al., 1995; Choi and

White, 2002; White, 2002; Lin and White, 2004; Wu and White, 2007; Jiwan and White, 2011;

Gunawardene et al., 2015;). Step 1 of the PT process requires formation of an RNA attenuation

structure in the plus-strand genome, just 5' to the sg mRNA initiation site, that acts as a physical

barrier for the RdRp (**Figure 3D**) (White, 2002; Jiwan and White, 2011; Wang et al., 2008). In

CIRV, attenuation structures form just upstream from the transcription initiation sites for sg

mRNA1 and sg mRNA2, and are generated by different sets of long-range intragenomic RNA-

RNA base-pairing interactions (Zhang et al. 1999; Choi et la., 2001; Choi and White, 2002; Lin

and White, 2004; Lin et al., 2007). The attenuation structure for sg mRNA2 transcription

(A) Long-range RNA-RNA interactions regulating sg mRNA transcription in CIRV
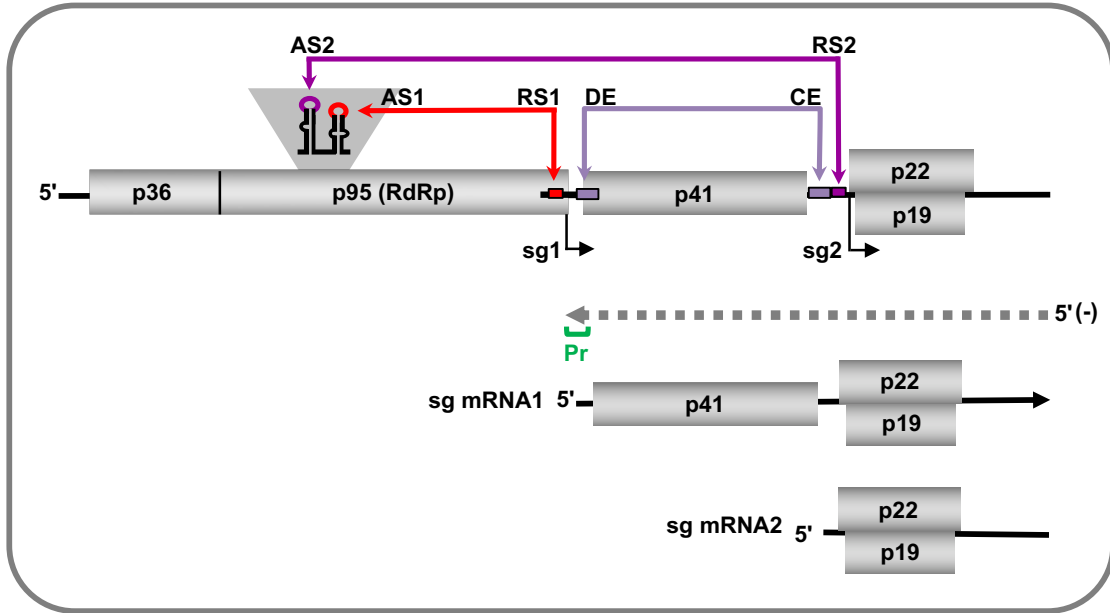
(B) Cartoon of long-range RNA-RNA interactions forming in the CIRV genome
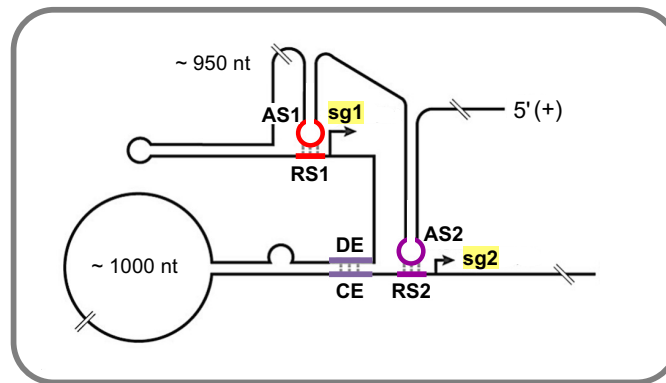
Figure 5: CIRV subgenomic mRNA transcription via a premature termination mechanism. (A) Long-distance RNA-RNA interactions involved in the formation of sg mRNA1 (red) and sg mRNA2 attenuation structures (purple and mauve). The truncated intermediate minus-strand template for sg mRNA1 is shown (grey dashed line), which contains a 3'-terminal promoter (green). Sg mRNA2 is also transcribed via a premature termination mechanism (not shown). Adapted from Chkuaseli and White (2018). **(B)** Sg mRNA1 and sg mRNA2 long-distance RNA-RNA interactions in the context of the folded CIRV genome. Adapted from Nicholson and White (2014). Not to scale. See text for details.

requires two long-range interactions: i) between activator sequence 2 (AS2) and receptor sequence 2 (RS2) spanning ~2000 nucleotides (**Figure 5A, dark purple**) (Lin and White, 2004), and ii) distal element (DE) and core element (CE) across ~1100 nucleotides (**Figure 5A, light purple**) (Zhang et al., 1999; Choi et al., 2001). In contrast, the attenuation structure for sg mRNA1 transcription involves a single long-range interaction between AS1 and RS1 sequences, separated by ~1000 nucleotides (**Figure 5A, red**) (Choi and White, 2002). The AS2/RS2, DE/CE, and AS1/RS1 interactions all function within the plus-strand genome, forming unique RNA attenuation structures that direct production of the minus-strand sg RNA2 and sg RNA1 intermediate templates (Choi et al., 2001; Choi and White, 2002; Lin and White, 2004). These minus-strand RNA intermediates carry promoter sequences at their 3'-ends that the RdRp uses to generate coding-sensed sg mRNAs during step 2 of the PT mechanism (**Figure 3D**). Accordingly, sg mRNA transcription in CIRV is a complex process involving large-scale folding of the viral RNA genome into global conformations that allow for distal sequences to interact intramolecularly and form functional RNA attenuation structures (**Figure 5B**). **Chapter 2 of this dissertation further investigates the extensive network of intragenomic interactions in CIRV and reports the identification of additional long-distance RNA-RNA interactions that are critical for activating sg mRNA1 transcription.**

1.4 The pea enation mosaic virus (PEMV) complex

The PEMV complex is a combination of two distinct but co-dependent RNA plant viruses, PEMV2 and PEMV1 (German and de Zoeten, 1975). Although each encodes its own RdRp and is capable of independent genome replication and sg mRNA transcription, these viruses are always found together in their infected natural hosts (Demler and de Zoeten, 1991; Demler et al., 1993; 1994; Syller, 2003; Taliansky and Robinson, 2003). Like CIRV, PEMV2 belongs to the Tombusviridae family (i.e. it is a tombusvirid), but it is a member of a different genus,

Umbravirus (Sit and Lommel, 2015). PEMV1, on the other hand, is in the genus Enamovirus, belonging to the family Solemoviridae (Sõmera et al., 2021). Thus, the complex consists of two unrelated viruses that have established a mutually beneficial relationship.

PEMV2, like all umbraviruses, lacks a gene for a capsid protein (Syller, 2003; Taliansky and Robinson, 2003) and instead relies on its assistor virus PEMV1 for capsid protein to form particles and for aphid transmission (Demler et al., 1993; 1994; 1997). PEMV2 in return provides PEMV1 with proteins required for cell-to-cell and systemic movement within infected plants (Ryabov et al., 1998; 1999a; 1999b; 2001a; 2001b; Nurkiyanova et al., 2001; Taliansky et al., 2003). Consequently, these two viruses from different families, have become co-dependent on each other for survival in nature. Virions isolated from PEMV1 and PEMV2 co-infections are non-enveloped icosahedral particles with diameters of 25 and 28 nm (German and de Zoeten, 1975). The 25 nm particles contain PEMV2 genomes, while the 28 nm ones encase the larger PEMV1 genomes (Doumayrou et al., 2016). In addition to the major capsid protein, these particles also contain a few copies of minor capsid protein, which is necessary for their aphid-mediated transmission from plant to plant (Demler et al., 1997). This minor CP protein is a C-terminally extended version of the CP and is generated by translational readthrough, earning it the term CP-readthrough domain, or CP-RTD (Demler et al., 1997; Liu et al., 2009). Below, molecular aspects of each virus in the PEMV complex are described individually.

1.4.1 Umbravirus PEMV2

The PEMV2 genome is a 4.3 kb-long plus-strand RNA that encodes four ORFs (**Figure 6A**) (Demler et al., 1993; Sit and Lommel, 2015). ORFs 1 and 2 yield p33 and its frameshift product p94 (the RdRp), which are both involved in viral RNA synthesis (Demler et al., 1993; Gao and Simon, 2016). However, unlike for tombusviruses such as CIRV, essentially nothing is known about the regulatory RNA elements in PEMV2 that control genome replication or sg

(A) PEMV2 genome and long-range RNA-RNA interactions regulating translation
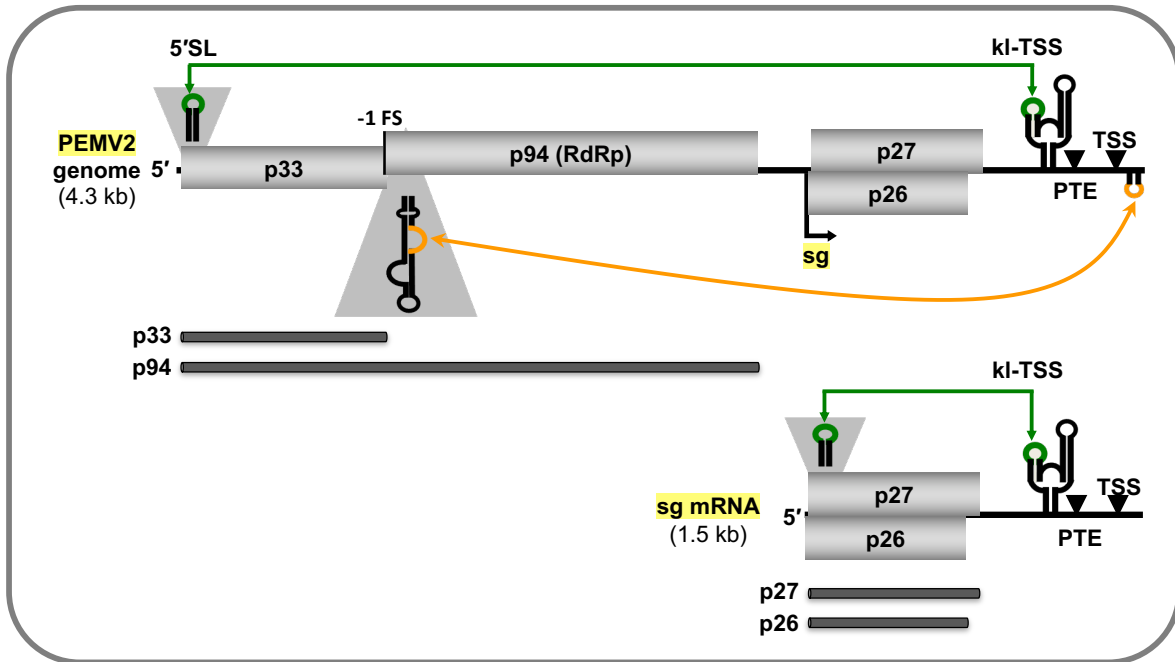
(B) PEMV1 genome and viral proteins
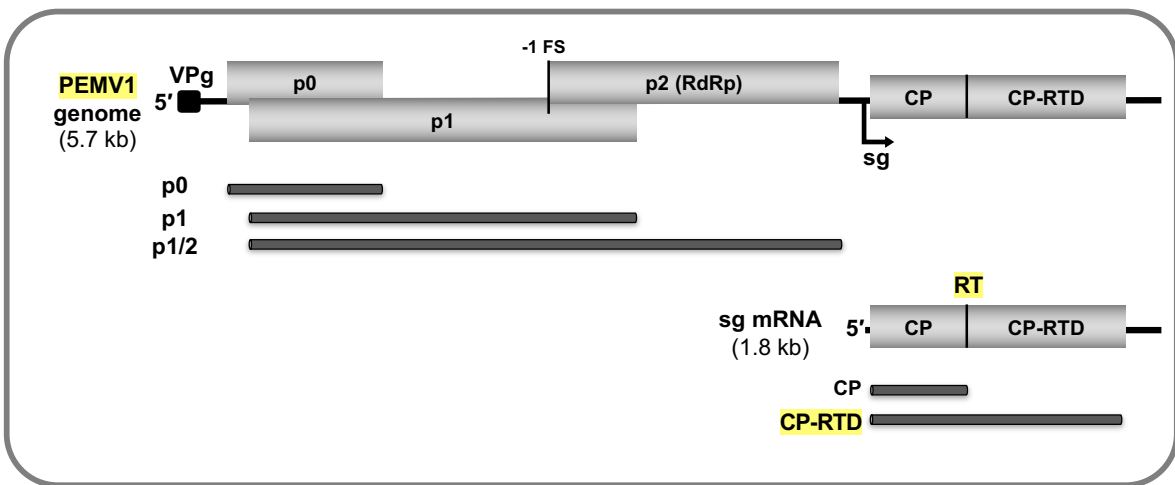
Figure 6: PEMV2 umbravirus and PEMV1 enamovirus that form the PEMV viral complex. (A) PEMV2 genome, sg mRNA, and encoded proteins. Long-range Interactions involved in promoting translation initiation in the genome and sg mRNA are indicated by the green arrows. The long-range interaction mediating -1 frameshifting is depicted by the orange arrow. The 3'-CITEs in the 3'UTR are indicated. **(B)** PEMV1 genome, sg mRNA, and encoded proteins. Not to scale. See text for details.

mRNA transcription. The 3'-proximal ORFs, which overlap in different reading frames, code for p26 and p27 and are expressed from a sole sg mRNA (Gao and Simon, 2017) (**Figure 6A, bottom**). P26 mediates formation of viral ribonucleoprotein particles, protects viral RNAs from nonsense mediated decay, and allows systemic spread of the infection (Ryabov et al., 1999a; Ryabov et al., 2001; Taliansky et al., 2003; Kim et al., 2007a; Kim et al., 2007b; May et al., 2020), whereas p27 is a cell-to-cell movement protein (Ryabov et al., 1998; Ryabov et al., 1999b; Nurkiyanova et al., 2001; Ryabov et al., 2001). As mentioned above, PEMV2 does not encode a capsid protein gene.

The PEMV2 genome, like that of CIRV and all other tombusvirids, is uncapped and non-polyadenylated (Demler et al., 1993). To allow for efficient translation initiation, PEMV2 has three different 3'CITEs in its 3'UTR; a kissing-loop T-shaped structure (kl-TSS), a panicum mosaic virus-like translational enhancer (PTE), and a 3' T-shaped structure (3'TSS) (Wang et al., 2009; Gao et al., 2012; 2013; 2014; Du et al., 2017) (**Figure 6A**). The kl-TSS directly binds ribosomal subunits and base-pairs with a complementary sequence in the 5'-end of the genome to allow initiation at the p33 start codon (Gao et al., 2012; 2013) (**Figure 6A, green**). The PTE recruits eIF4E and is proposed to enhance the ribosome-binding ability of kl-TSS (Wang et al., 2009; 2011; Gao et al., 2012; 2013). The 3'TSS also binds ribosomal subunits, but was found to only be necessary for efficient translation initiation from the sg mRNA (Gao et al., 2014; Gao and Simon, 2017; Du et al., 2017).

The p94 RdRp is translated directly from the genome via programmed -1 frameshifting near the end of the p33 ORF (Gao and Simon, 2016). Efficiency of this process depends on a slippery sequence at the frameshift site upstream from the p33 stop codon, formation of an extended SL structure downstream from the p33 stop codon, and an RNA-RNA interaction between the frameshift SL and the 3'UTR (Gao and Simon, 2016) (**Figure 6A, orange**), similar to the RT stimulatory structure of CIRV (**Figure 4B, orange**) (Cimino et al., 2011). The overlapping p26 and p27 ORFs are expressed from the sg mRNA that is transcribed by p94

during infections (Gao and Simon, 2017) (**Figure 6A, bottom**). All three of the 3'CITEs present

in the 3'UTR direct translation initiation from the sg mRNA, and a base-pairing interaction

between kl-TSS and a complementary sequence near the 5'-end of the sg mRNA is necessary

for this process (**Figure 6a, bottom, green**) (Gao and Simon, 2017).

While it is known that PEMV2 transcribes a sg mRNA, the details of this process are

currently not known. The 5'-terminus of the sg mRNA was mapped to position 2772 nt in the

genome, making it 1.5 kb in length (Gao and Simon, 2017). Analysis of the sequence around

the sg mRNA initiation site revealed a promoter-like sequence immediately downstream of the

transcription start site and a predicted stem-loop RNA structure just upstream (Gao and Simon,

2017). It has been clearly established that several genera in the family Tombusviridae utilize a

PT mechanism for sg mRNA transcription (Jiwan and White, 2011), therefore PEMV2's

membership in this family suggests that it too likely uses a PT mechanism. **Chapter 3 of this**

dissertation explores the sg mRNA transcriptional strategy utilized by PEMV2 and

reports that it employs a PT mechanism in which the RNA attenuation structure is formed

through dimerization of its viral RNA genome.

1.4.2 Enamovirus PEMV1

PEMV1, the assistor enamovirus of PEMV2, possesses a 5.7 kb-long, plus-strand,

polycistronic RNA genome that codes for five ORFs (**Figure 6B, top**) (Demler and de Zoeten,

1991). Three 5'-proximally encoded nonstructural proteins p0, p1 and p2, are all translated from

the genome (Demler and de Zoeten, 1991). P0 is the suppressor of antiviral RNA silencing

(Fusaro et al., 2012), while the p1 polyprotein, based on conserved amino acid motifs, contains

a helicase, a protease, and a VPg (viral protein genome-linked) (Demler and de Zoeten, 1991;

Li et al, 2007; Delfosse et al., 2021). P2 is generated by a -1 frameshift in the p1 ORF and is

the RdRp responsible for genome and sg mRNA synthesis (Demler and de Zoeten, 1991;

Osman et al., 2006a). Structural proteins, CP and CP-RTD, are both translated from a 1.8 kb-long sg mRNA, and these proteins are utilized by both PEMV1 and PEMV2 for genome packaging (Liu et al., 2009) (**Figure 6B, bottom**). CP-RTD is an important minor capsid protein that is incorporated, in low numbers, into virus particles, and as mentioned previously, is required for aphid transmission of PEMV1 and PEMV2 virions between host plants (Demler et al., 1997).

Details of PEMV1 genome replication and sg mRNA transcription regulation have not been investigated. PEMV1 VPg has been suggested to play a role in genome replication (Skaf et al., 2000) based on the comparable VPgs in related poleroviruses that serve as protein primers (likely via an -OH group on a tyrosine) for initiation of viral RNA synthesis (Osman et al., 2006b). For sg mRNA transcription, based on other plus-strand RNA plant viruses, either a PT or internal initiation mechanism could be utilized (Sztuba-Solińska et al., 2011; Delfosse et al., 2021; Sõmera et al., 2021; Miller and Lozier, 2022).

PEMV1 genome, like that of CIRV and PEMV2, does not have 5'-cap or 3'-poly(A) tail structures. Instead, a 3 kDa VPg, derived from proteolytic processing of p1, is covalently linked to its 5'-terminus, and a SL structure is predicted to form at its 3'-terminus (**Figure 6B, top**) (Demler and de Zoeten, 1991; Wobus et al., 1998; Skaf et al., 2000). What promotes initiation of translation in this virus is not known (Sõmera et al., 2021), however, based on studies on related poleroviruses (also members of the family Solemoviridae), a 3'CITE (Miras et al., 2022) or the VPg (Hébrard et al., 2010; Reinbold et al., 2013; Miller and Lozier, 2022) may be involved. It should be noted that, although free VPgs (i.e. not genome linked) of several poleroviruses have been shown to directly interact with eukaryotic translation initiation factors, it has not been clearly demonstrated that 5'-genome-tethered VPgs function to facilitate viral translation initiation (Jiang and Laliberté, 2011; Hébrard et al., 2010; Reinbold et al., 2013).

Three proteins are translated from the PEMV1 genome. P1 is proposed to be expressed via a leaky scanning mechanism, due to the weak Kozak context of the start codon for P0

(Kozak, 1987; Demler and de Zoeten, 1991; Delfossee et al., 2021). P2 is expressed as a C-terminal extension of p1 via programmed -1 frameshifting (Demler and de Zoeten, 1991; Nixon et al., 2002a; 2002b) that requires a conserved RNA pseudoknot structure positioned just downstream from a slippery site (Kim et al., 2000; Nixon et al., 2002a; 2002b; Cornish et al., 2005) (**Figure 6b, top**).

CP and CP-RTD are both translated from a 1.8 kb-long sg mRNA, with the latter protein produced via programmed translation readthrough (**Figure 6b, bottom**) (Demler and de Zoeten, 1991; Liu et al., 2009). Readthrough stimulating RNA elements have not been investigated in PEMV1, however, in the related polerovirus potato leafroll virus (PLRV) this process was found to involve complementary sequences, one located just downstream from the CP stop codon and another sequence positioned ~700 nt downstream and present within the CP-RTD coding region (Xu et al., 2018). While efforts to demonstrate a functional interaction between the two sequences in PLRV were unsuccessful, the conservation of similarly positioned complementary sequences in polerovirus and enamovirus sg mRNAs implied that a long-range interaction was required for RT production of CP-RTD (Xu et al., 2018). **Chapter 4 of this dissertation examines programmed translational readthrough in PEMV1 and reports that long-distance RNA-RNA interactions forming an elaborate RNA structure are needed for the expression of CP-RTD from its sg mRNA.**

1.5 Research objectives

Plus-strand RNA plant virus genomes encode essential viral proteins that ensure successful takeover of their host cells. These proteins are produced by employing different gene expression strategies, including transcriptional and translational approaches. **The goal of this dissertation is to examine the mechanisms of sg mRNA transcription in two plant viruses within the family Tombusviridae, CIRV and PEMV2, as well as the mode of activation of programmed translational readthrough in the sg mRNA of PEMV1, the obligate partner of PEMV2 in infections.** Collectively, this work has provided significant novel insights into the molecular mechanisms of how sg mRNAs are transcribed and how sg mRNAs are translated.

The above stated sg mRNA-related goals were achieved through three studies which are presented in separate chapters in the form of two publications (Ch 2 and 4) and a submitted manuscript (Ch 3). The specific objectives of each chapter are provided below:

(1) To characterize a novel RNA interaction required for transcription of sg mRNA1 in CIRV
(Chapter 2 – published article**)**

(2) To determine the mechanism of PEMV2 sg mRNA transcription
(Chapter 3 – submitted manuscript**)**

(3) To define the readthrough-promoting RNA structure present in the sg mRNA of PEMV1
(Chapter 4 – published article**)**

1.6 References

Ahlquist, P. (2006). Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Reviews Microbiology*, *4*(5), 371–382. https://doi.org/10.1038/nrmicro1389

Aitken, C. E., & Lorsch, J. R. (2012). A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, *19*(6), 568–576. https://doi.org/10.1038/nsmb.2303

Allan, M. F., Brivanlou, A., & Rouskin, S. (2023). RNA levers and switches controlling viral gene expression. *Trends in Biochemical Sciences*, *48*(4), 391–406. https://doi.org/10.1016/j.tibs.2022.12.002

Barry, J. K., & Miller, W. A. (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. *Proceedings of the National Academy of Sciences*, *99*(17), 11133–11138. https://doi.org/10.1073/pnas.162223099

Basnayake, V. R., Sit, T. L., & Lommel, S. A. (2006). The genomic RNA packaging scheme of Red clover necrotic mosaic virus. *Virology*, *345*(2), 532–539. https://doi.org/10.1016/j.virol.2005.10.017

Beier, H. & Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Research*, *29*(23), 4767–4782. https://doi.org/10.1093/nar/29.23.4767

Blanco-Pérez, M., & Hernández, C. (2016). Evidence supporting a premature termination mechanism for subgenomic RNA transcription in Pelargonium line pattern virus: Identification of a critical long-range RNA–RNA interaction and functional variants through mutagenesis. *Journal of General Virology*, *97*(6), 1469–1480. https://doi.org/10.1099/jgv.0.000459

Blanco-Pérez, M., Pérez-Cañamás, M., Ruiz, L., & Hernández, C. (2016). Efficient Translation of Pelargonium line pattern virus RNAs Relies on a TED-Like 3´-Translational Enhancer that Communicates with the Corresponding 5´-Region through a Long-Distance RNA-RNA Interaction. *PLOS ONE*, *11*(4), e0152593. https://doi.org/10.1371/journal.pone.0152593

Brinton, M. A., Gulyaeva, A. A., Balasuriya, U. B. R., Dunowska, M., Faaberg, K. S., Goldberg, T., Leung, F. C. C., Nauwynck, H. J., Snijder, E. J., Stadejek, T., & Gorbalenya, A. E. (2021). ICTV Virus Taxonomy Profile: Arteriviridae 2021. *Journal of General Virology*, *102*(8). https://doi.org/10.1099/jgv.0.001632

Brown, C. M., Dinesh-Kumar, S. P., & Miller, W. A. (1996). Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. *Journal of Virology*, *70*(9), 5884–5892. https://doi.org/10.1128/jvi.70.9.5884-5892.1996

Chambers, T. J., Nestorowicz, A., Amberg, S. M., & Rice, C. M. (1993). Mutagenesis of the yellow fever virus NS2B protein: Effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *Journal of Virology*, *67*(11), 6797–6807. https://doi.org/10.1128/jvi.67.11.6797-6807.1993

Chattopadhyay, M., Shi, K., Yuan, X., & Simon, A. E. (2011). Long-distance kissing loop interactions between a 3′ proximal Y-shaped structure and apical loops of 5′ hairpins enhance translation of Saguaro cactus virus. *Virology*, *417*(1), 113–125. https://doi.org/10.1016/j.virol.2011.05.007

Chkuaseli, T., Newburn, L. R., Bakhshinyan, D., & White, K. A. (2015). Protein expression strategies in Tobacco necrosis virus-D. *Virology*, *486*, 54–62. https://doi.org/10.1016/j.virol.2015.08.032

Chkuaseli, T., & White, K. A. (2018). Intragenomic Long-Distance RNA–RNA Interactions in Plus-Strand RNA Plant Viruses. *Frontiers in Microbiology*, 9, 529. https://doi.org/10.3389/fmicb.2018.00529

Choi, I.-R., Ostrovsky, M., Zhang, G., & White, K. A. (2001). Regulatory Activity of Distal and Core RNA Elements in Tombusvirus Subgenomic mRNA2 Transcription. *Journal of Biological Chemistry*, *276*(45), 41761–41768. https://doi.org/10.1074/jbc.M106727200

Choi, I.-R., & White, K. A. (2002). An RNA Activator of Subgenomic mRNA1 Transcription in Tomato Bushy Stunt Virus. *Journal of Biological Chemistry*, *277*(5), 3760–3766. https://doi.org/10.1074/jbc.M109067200

Choi, Y. G., Dreher, T. W., & Rao, A. L. N. (2002). tRNA elements mediate the assembly of an icosahedral RNA virus. *Proceedings of the National Academy of Sciences*, *99*(2), 655–660. https://doi.org/10.1073/pnas.022618199

Choi, Y. G., & Rao, A. L. N. (2003). Packaging of Brome Mosaic Virus RNA3 Is Mediated through a Bipartite Signal. *Journal of Virology*, *77*(18), 9750–9757. https://doi.org/10.1128/JVI.77.18.9750-9757.2003

Chu, M., Park, J.-W., & Scholthof, H. B. (1999). Separate Regions on the Tomato Bushy Stunt Virus p22 Protein Mediate Cell-to-Cell Movement versus Elicitation of Effective Resistance Responses. *Molecular Plant-Microbe Interactions®*, *12*(4), 285–292. https://doi.org/10.1094/MPMI.1999.12.4.285

Cimino, P. A., Nicholson, B. L., Wu, B., Xu, W., & White, K. A. (2011). Multifaceted Regulation of Translational Readthrough by RNA Replication Elements in a Tombusvirus. *PLoS Pathogens*, *7*(12), e1002423. https://doi.org/10.1371/journal.ppat.1002423

Cornish, P. V., Hennig, M., & Giedroc, D. P. (2005). A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated –1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences*, *102*(36), 12694–12699. https://doi.org/10.1073/pnas.0506166102

Delfosse, V. C., Barrios Barón, M. P., & Distéfano, A. J. (2021). What we know about poleroviruses: Advances in understanding the functions of polerovirus proteins. *Plant Pathology*, *70*(5), 1047–1061. https://doi.org/10.1111/ppa.13368

Demler, S. A., Borkhsenious, O. N., Rucker, D. G., & de Zoeten, G. A. (1994). Assessment of the Autonomy of Replicative and Structural Functions Encoded by the Luteo-phase of Pea Enation Mosaic Virus. *Journal of General Virology*, *75*(5), 997–1007. https://doi.org/10.1099/0022-1317-75-5-997

Demler, S. A., & de Zoeten, G. A. (1991). The Nucleotide Sequence and Luteovirus-like Nature of RNA 1 of an Aphid Non-transmissible Strain of Pea Enation Mosaic Virus. *Journal of General Virology*, *72*(8), 1819–1834. https://doi.org/10.1099/0022-1317-72-8-1819

Demler, S. A., Rucker, D. G., & de Zoeten, G. A. (1993). The chimeric nature of the genome of pea enation mosaic virus: The independent replication of RNA 2. *Journal of General Virology*, *74*(1), 1–14. https://doi.org/10.1099/0022-1317-74-1-1

Demler, S. A., Rucker-Feeney, D. G., Skaf, J. S., & de Zoeten, G. A. (1997). Expression and suppression of circulative aphid transmission in pea enation mosaic virus. *Journal of General Virology*, *78*(3), 511–523. https://doi.org/10.1099/0022-1317-78-3-511

Dever, T. E., Dinman, J. D., & Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *10*(8), a032649. https://doi.org/10.1101/cshperspect.a032649

Doumayrou, J., Sheber, M., Bonning, B., & Miller, W. (2016). Role of Pea Enation Mosaic Virus Coat Protein in the Host Plant and Aphid Vector. *Viruses*, *8*(11), 312. https://doi.org/10.3390/v8110312

Du, Z., Alekhina, O. M., Vassilenko, K. S., & Simon, A. E. (2017). Concerted action of two 3′ cap-independent translation enhancers increases the competitive strength of translated viral genomes. *Nucleic Acids Research*, *45*(16), 9558–9572. https://doi.org/10.1093/nar/gkx643

Fabian, M. R., Na, H., Ray, D., & White, K. A. (2003). 3′-Terminal RNA secondary structures are important for accumulation of tomato bushy stunt virus DI RNAs. *Virology*, *313*(2), 567–580. https://doi.org/10.1016/S0042-6822(03)00349-0

Fabian, M. R., & White, K. A. (2004). 5′-3′ RNA-RNA Interaction Facilitates Cap- and Poly(A) Tail-independent Translation of Tomato Bushy Stunt Virus mRNA. *Journal of Biological Chemistry*, *279*(28), 28862–28872. https://doi.org/10.1074/jbc.M401272200

Fabian, M. R., & White, K. A. (2006). Analysis of a 3′-translation enhancer in a tombusvirus: A dynamic model for RNA–RNA interactions of mRNA termini. *RNA*, *12*(7), 1304–1314. https://doi.org/10.1261/rna.69506

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *Journal of General Virology*, *93*(7), 1385–1409. https://doi.org/10.1099/vir.0.042499-0

Firth, A. E., Wills, N. M., Gesteland, R. F., & Atkins, J. F. (2011). Stimulation of stop codon readthrough: Frequent presence of an extended 3′ RNA structural element. *Nucleic Acids Research*, *39*(15), 6679–6691. https://doi.org/10.1093/nar/gkr224

Fujimoto, Y., Keima, T., Hashimoto, M., Hagiwara-Komoda, Y., Hosoe, N., Nishida, S., Nijo, T., Oshima, K., Verchot, J., Namba, S., & Yamaji, Y. (2022). Short 5′ Untranslated Region Enables Optimal Translation of Plant Virus Tricistronic RNA via Leaky Scanning. *Journal of Virology*, *96*(7), e02144-21. https://doi.org/10.1128/jvi.02144-21

Fusaro, A. F., Correa, R. L., Nakasugi, K., Jackson, C., Kawchuk, L., Vaslin, M. F. S., & Waterhouse, P. M. (2012). The Enamovirus P0 protein is a silencing suppressor which inhibits local and systemic RNA silencing through AGO1 degradation. *Virology*, *426*(2), 178–187. https://doi.org/10.1016/j.virol.2012.01.026

Gao, F., Gulay, S. P., Kasprzak, W., Dinman, J. D., Shapiro, B. A., & Simon, A. E. (2013). The Kissing-Loop T-Shaped Structure Translational Enhancer of Pea Enation Mosaic Virus Can Bind Simultaneously to Ribosomes and a 5′ Proximal Hairpin. *Journal of Virology*, *87*(22), 11987–12002. https://doi.org/10.1128/JVI.02005-13

Gao, F., Kasprzak, W. K., Szarko, C., Shapiro, B. A., & Simon, A. E. (2014). The 3′ Untranslated Region of Pea Enation Mosaic Virus Contains Two T-Shaped, Ribosome-Binding, Cap-Independent Translation Enhancers. *Journal of Virology*, *88*(20), 11696–11712. https://doi.org/10.1128/JVI.01433-14

Gao, F., Kasprzak, W., Stupina, V. A., Shapiro, B. A., & Simon, A. E. (2012). A Ribosome-Binding, 3′ Translational Enhancer Has a T-Shaped Structure and Engages in a Long-Distance RNA-RNA Interaction. *Journal of Virology*, *86*(18), 9828–9842. https://doi.org/10.1128/JVI.00677-12

Gao, F., & Simon, A. E. (2016). Multiple Cis-acting elements modulate programmed -1 ribosomal frameshifting in Pea enation mosaic virus. *Nucleic Acids Research*, *44*(2), 878–895. https://doi.org/10.1093/nar/gkv1241

Gao, F., & Simon, A. E. (2017). Differential use of 3'CITEs by the subgenomic RNA of Pea enation mosaic virus 2. *Virology*, *510*, 194–204. https://doi.org/10.1016/j.virol.2017.07.021

Geng, G., Wang, D., Liu, Z., Wang, Y., Zhu, M., Cao, X., Yu, C., & Yuan, X. (2021). Translation of Plant RNA Viruses. *Viruses*, *13*(12), 2499. https://doi.org/10.3390/v13122499

German, T. L., & de Zoeten, G. A. (1975). Purification and properties of the replicative forms and replicative intermediates of pea enation mosaic virus. *Virology*, *66*(1), 172–184. https://doi.org/10.1016/0042-6822(75)90188-9

Grdzelishvili, V. Z., Garcia-Ruiz, H., Watanabe, T., & Ahlquist, P. (2005). Mutual Interference between Genomic RNA Replication and Subgenomic mRNA Transcription in Brome Mosaic Virus. *Journal of Virology*, *79*(3), 1438–1451. https://doi.org/10.1128/JVI.79.3.1438-1451.2005

Gunawardene, C. D., Donaldson, L. W., & White, K. A. (2017). Tombusvirus polymerase: Structure and function. *Virus Research*, *234*, 74–86. https://doi.org/10.1016/j.virusres.2017.01.012

Gunawardene, C. D., Im, J. S. H., & White, K. A. (2021). RNA Structure Protects the 5′ End of an Uncapped Tombusvirus RNA Genome from Xrn Digestion. *Journal of Virology*, *95*(20), e01034-21. https://doi.org/10.1128/JVI.01034-21

Gunawardene, C. D., Jaluba, K., & White, K. A. (2015). Conserved Motifs in a Tombusvirus Polymerase Modulate Genome Replication, Subgenomic Transcription, and Amplification of Defective Interfering RNAs. *Journal of Virology*, *89*(6), 3236–3246. https://doi.org/10.1128/JVI.03378-14

Guo, L., Allen, E. M., & Miller, W. A. (2001). Base-Pairing between Untranslated Regions Facilitates Translation of Uncapped, Nonpolyadenylated Viral RNA. *Molecular Cell*, *7*(5), 1103–1109. https://doi.org/10.1016/S1097-2765(01)00252-0

Haasnoot, P. C. J., Brederode, F. Th., Olsthoorn, R. C. L., & Bol, J. F. (2000). A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. *RNA*, *6*(5), 708–716. https://doi.org/10.1017/S1355838200992471

Haasnoot, P. C. J., Olsthoorn, R. C. L., & Bol, J. F. (2002). The Brome mosaic virus subgenomic promoter hairpin is structurally similar to the iron-responsive element and functionally equivalent to the minus-strand core promoter stem-loop C. *RNA*, *8*(1), 110–122. https://doi.org/10.1017/S1355838202012074

Hébrard, E., Poulicard, N., Gérard, C., Traoré, O., Wu, H.-C., Albar, L., Fargette, D., Bessin, Y., & Vignols, F. (2010). Direct Interaction Between the *Rice yellow mottle virus* (RYMV) VPg and the Central Domain of the Rice eIF(iso)4G1 Factor Correlates with Rice Susceptibility and RYMV Virulence. *Molecular Plant-Microbe Interactions®*, *23*(11), 1506–1513. https://doi.org/10.1094/MPMI-03-10-0073

Hellen, C. U. T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *10*(10), a032656. https://doi.org/10.1101/cshperspect.a032656

Hollings, M., Stone, O. M., & Bouttell, G. C. (1970). Carnation Italian ringspot virus. *Annals of Applied Biology*, *65*(2), 299–309. https://doi.org/10.1111/j.1744-7348.1970.tb04591.x

Hwang, Y. T., McCartney, A. W., Gidda, S. K., & Mullen, R. T. (2008). Localization of the Carnation Italian ringspot virus replication protein p36 to the mitochondrial outer membrane is mediated by an internal targeting signal and the TOM complex. *BMC Cell Biology*, *9*(1), 54. https://doi.org/10.1186/1471-2121-9-54

Ilyas, M., Du, Z., & Simon, A. E. (2021). *Opium Poppy Mosaic Virus* Has an Xrn-Resistant, Translated Subgenomic RNA and a BTE 3′ CITE. *Journal of Virology*, *95*(9), e02109-20. https://doi.org/10.1128/JVI.02109-20

Jaafar, Z. A., & Kieft, J. S. (2019). Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology*, *17*(2), 110–123. https://doi.org/10.1038/s41579-018-0117-x

Jiang, J., & Laliberté, J.-F. (2011). The genome-linked protein VPg of plant viruses—A protein with many partners. *Current Opinion in Virology*, *1*(5), 347–354. https://doi.org/10.1016/j.coviro.2011.09.010

Jiwan, S. D., & White, K. A. (2011). Subgenomic mRNA transcription in Tombusviridae. *RNA Biology*, *8*(2), 287–294. https://doi.org/10.4161/rna.8.2.15195

Jiwan, S. D., Wu, B., & White, K. A. (2011). Subgenomic mRNA transcription in tobacco necrosis virus. *Virology*, *418*(1), 1–11. https://doi.org/10.1016/j.virol.2011.07.005

Johnston, J. C., & Rochon, D. M. (1996). Both Codon Context and Leader Length Contribute to Efficient Expression of Two Overlapping Open Reading Frames of a Cucumber Necrosis Virus Bifunctional Subgenomic mRNA. *Virology*, *221*(1), 232–239. https://doi.org/10.1006/viro.1996.0370

Kim, S. H., MacFarlane, S., Kalinina, N. O., Rakitina, D. V., Ryabov, E. V., Gillespie, T., Haupt, S., Brown, J. W. S., & Taliansky, M. (2007a). Interaction of a plant virus-encoded protein with the major nucleolar protein fibrillarin is required for systemic virus infection. *Proceedings of the National Academy of Sciences*, *104*(26), 11115–11120. https://doi.org/10.1073/pnas.0704632104

Kim, S. H., Ryabov, E. V., Kalinina, N. O., Rakitina, D. V., Gillespie, T., MacFarlane, S., Haupt, S., Brown, J. W. S., & Taliansky, M. (2007b). Cajal bodies and the nucleolus are required for a plant virus systemic infection. *The EMBO Journal*, *26*(8), 2169–2179. https://doi.org/10.1038/sj.emboj.7601674

Kim, Y.-G., Maas, S., Wang, S. C., & Rich, A. (2000). Mutational study reveals that tertiary interactions are conserved in ribosomal frameshifting pseudoknots of two luteoviruses. *RNA*, *6*(8), 1157–1165. https://doi.org/10.1017/S1355838200000510

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Kuhn, J. H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews*, *84*(2), e00061-19. https://doi.org/10.1128/MMBR.00061-19

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, *15*(20), 8125–8148. https://doi.org/10.1093/nar/15.20.8125

Kozak, M. (1991.). A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expression*, *1*(2), 111–115.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, *299*(1–2), 1–34. https://doi.org/10.1016/S0378-1119(02)01056-9

Kraft, J. J., Treder, K., Peterson, M. S., & Miller, W. A. (2013). Cation-dependent folding of 3′ cap-independent translation elements facilitates interaction of a 17-nucleotide conserved sequence with eIF4G. *Nucleic Acids Research*, *41*(5), 3398–3413. https://doi.org/10.1093/nar/gkt026

Kuhlmann, M. M., Chattopadhyay, M., Stupina, V. A., Gao, F., & Simon, A. E. (2016). An RNA Element That Facilitates Programmed Ribosomal Readthrough in Turnip Crinkle Virus Adopts Multiple Conformations. *Journal of Virology*, *90*(19), 8575–8591. https://doi.org/10.1128/JVI.01129-16

Li, M.-L., & Stollar, V. (2004). Identification of the amino acid sequence in Sindbis virus nsP4 that binds to the promoter for the synthesis of the subgenomic RNA. *Proceedings of the*

National Academy of Sciences, *101*(25), 9429–9434.
https://doi.org/10.1073/pnas.0400995101

Li, M.-L., & Stollar, V. (2007). Distinct Sites on the Sindbis Virus RNA-Dependent RNA
Polymerase for Binding to the Promoters for the Synthesis of Genomic and Subgenomic
RNA. *Journal of Virology*, *81*(8), 4371–4373. https://doi.org/10.1128/JVI.02672-06

Li, X., Halpin, C., & Ryan, M. D. (2007). A novel cleavage site within the potato leafroll virus P1
polyprotein. *Journal of General Virology*, *88*(5), 1620–1623.
https://doi.org/10.1099/vir.0.82627-0

Li, Y., Treffers, E. E., Napthine, S., Tas, A., Zhu, L., Sun, Z., Bell, S., Mark, B. L., Van Veelen, P.
A., Van Hemert, M. J., Firth, A. E., Brierley, I., Snijder, E. J., & Fang, Y. (2014).
Transactivation of programmed ribosomal frameshifting by a viral protein. *Proceedings of
the National Academy of Sciences*, *111*(21). https://doi.org/10.1073/pnas.1321930111

Li, Z., Pogany, J., Panavas, T., Xu, K., Esposito, A. M., Kinzy, T. G., & Nagy, P. D. (2009).
Translation elongation factor 1A is a component of the tombusvirus replicase complex and
affects the stability of the p33 replication co-factor. Virology, 385(1), 245–260.
https://doi.org/10.1016/j.virol.2008.11.041

Li, Z., Pogany, J., Tupman, S., Esposito, A. M., Kinzy, T. G., & Nagy, P. D. (2010). Translation
Elongation Factor 1A Facilitates the Assembly of the Tombusvirus Replicase and Stimulates
Minus-Strand Synthesis. *PLoS Pathogens*, *6*(11), e1001175.
https://doi.org/10.1371/journal.ppat.1001175

Lin, H.-X., & White, K. A. (2004). A complex network of RNA–RNA interactions controls
subgenomic mRNA transcription in a tombusvirus. *The EMBO Journal*, *23*(16), 3365–3374.
https://doi.org/10.1038/sj.emboj.7600336

Lin, H.-X., Xu, W., & White, K. A. (2007). A Multicomponent RNA-Based Control System
Regulates Subgenomic mRNA Transcription in a Tombusvirus. *Journal of Virology*, *81*(5),
2429–2439. https://doi.org/10.1128/JVI.01969-06

Lindenbach, B. D., Sgro, J.-Y., & Ahlquist, P. (2002). Long-Distance Base Pairing in Flock House
Virus RNA1 Regulates Subgenomic RNA3 Synthesis and RNA2 Replication. *Journal of
Virology*, *76*(8), 3905–3919. https://doi.org/10.1128/JVI.76.8.3905-3919.2002

Liu, S., Sivakumar, S., Wang, Z., Bonning, B. C., & Allen Miller, W. (2009). The readthrough
domain of pea enation mosaic virus coat protein is not essential for virus stability in the
hemolymph of the pea aphid. Archives of Virology, 154(3), 469–479.
https://doi.org/10.1007/s00705-009-0327-7

Lozsa, R., Csorba, T., Lakatos, L., & Burgyan, J. (2008). Inhibition of 3' modification of small
RNAs in virus-infected plants require spatial and temporal co-expression of small RNAs and
viral silencing-suppressor proteins. *Nucleic Acids Research*, *36*(12), 4099–4107.
https://doi.org/10.1093/nar/gkn365

Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, *23*(1), 21–39. https://doi.org/10.1038/s41580-021-00432-z

Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G., & Diaz-Toledano, R. (2015). Picornavirus IRES elements: RNA structure and host protein interactions. *Virus Research*, *206*, 62–73. https://doi.org/10.1016/j.virusres.2015.01.012

Mateos-Gomez, P. A., Morales, L., Zuñiga, S., Enjuanes, L., & Sola, I. (2013). Long-Distance RNA-RNA Interactions in the Coronavirus Genome Form High-Order Structures Promoting Discontinuous RNA Synthesis during Transcription. *Journal of Virology*, *87*(1), 177–186. https://doi.org/10.1128/JVI.01782-12

Mateos-Gómez, P. A., Zuñiga, S., Palacio, L., Enjuanes, L., & Sola, I. (2011). Gene N Proximal and Distal RNA Motifs Regulate Coronavirus Nucleocapsid mRNA Transcription. *Journal of Virology*, *85*(17), 8968–8980. https://doi.org/10.1128/JVI.00869-11

May, J., Johnson, P., Saleem, H., & Simon, A. E. (2017). A Sequence-Independent, Unstructured Internal Ribosome Entry Site Is Responsible for Internal Expression of the Coat Protein of Turnip Crinkle Virus. *Journal of Virology*, *91*(8), e02421-16. https://doi.org/10.1128/JVI.02421-16

May, J. P., Johnson, P. Z., Ilyas, M., Gao, F., & Simon, A. E. (2020). The Multifunctional Long-Distance Movement Protein of *Pea Enation Mosaic Virus 2* Protects Viral and Host Transcripts from Nonsense-Mediated Decay. *MBio*, *11*(2), e00204-20. https://doi.org/10.1128/mBio.00204-20

Miller, W. A., Dreher, T. W., & Hall, T. C. (1985). Synthesis of brome mosaic virus subgenomic RNA in vitro by internal initiation on (−)-sense genomic RNA. *Nature*, *313*(5997), 68–70. https://doi.org/10.1038/313068a0

Miller, W. A., & Koev, G. (2000). Synthesis of Subgenomic RNAs by Positive-Strand RNA Viruses. *Virology*, *273*(1), 1–8. https://doi.org/10.1006/viro.2000.0421

Miller, W. A., & Lozier, Z. (2022). Yellow Dwarf Viruses of Cereals: Taxonomy and Molecular Mechanisms. *Annual Review of Phytopathology*, *60*(1), 121–141. https://doi.org/10.1146/annurev-phyto-121421-125135

Miras, M., Aranda, M. A., & Truniger, V. (2022). Different RNA Elements Control Viral Protein Synthesis in Polerovirus Isolates Evolved in Separate Geographical Regions. *International Journal of Molecular Sciences*, *23*(20), 12503. https://doi.org/10.3390/ijms232012503

Mizumoto, H., Tatsuta, M., Kaido, M., Mise, K., & Okuno, T. (2003). Cap-Independent Translational Enhancement by the 3' Untranslated Region of Red Clover Necrotic Mosaic Virus RNA. *Journal of Virology*, *77*(22), 12113-12121. https://doi.org/10.1128/JVI.77.22.12113-12121.2003

Monkewich, S., Lin, H.-X., Fabian, M. R., Xu, W., Na, H., Ray, D., Chernysheva, O. A., Nagy, P. D., & White, K. A. (2005). The p92 Polymerase Coding Region Contains an Internal RNA Element Required at an Early Step in Tombusvirus Genome Replication. *Journal of Virology*, *79*(8), 4848–4858. https://doi.org/10.1128/JVI.79.8.4848-4858.2005

Moreno, J. L., Zúñiga, S., Enjuanes, L., & Sola, I. (2008). Identification of a Coronavirus Transcription Enhancer. *Journal of Virology*, *82*(8), 3882–3893. https://doi.org/10.1128/JVI.02622-07

Nagy, P. D. (2016). Tombusvirus-Host Interactions: Co-Opted Evolutionarily Conserved Host Factors Take Center Court. *Annual Review of Virology*, *3*(1), 491–515. https://doi.org/10.1146/annurev-virology-110615-042312

Napthine, S., Ling, R., Finch, L. K., Jones, J. D., Bell, S., Brierley, I., & Firth, A. E. (2017). Protein-directed ribosomal frameshifting temporally regulates gene expression. Nature Communications, 8(1), 15582. https://doi.org/10.1038/ncomms15582

Napthine, S., Treffers, E. E., Bell, S., Goodfellow, I., Fang, Y., Firth, A. E., Snijder, E. J., & Brierley, I. (2016). A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Research*, *44*(12), 5491–5503. https://doi.org/10.1093/nar/gkw480

Newburn, L. R., Nicholson, B. L., Yosefi, M., Cimino, P. A., & White, K. A. (2014). Translational readthrough in Tobacco necrosis virus-D. *Virology*, *450–451*, 258–265. https://doi.org/10.1016/j.virol.2013.12.006

Newburn, L. R., & White, K. A. (2015). Cis-acting RNA elements in positive-strand RNA plant virus genomes. *Virology*, *479–480*, 434–443. https://doi.org/10.1016/j.virol.2015.02.032

Nicholson, B. L., & White, K. A. (2008). Context-influenced cap-independent translation of Tombusvirus mRNAs in vitro. *Virology*, *380*(2), 203–212. https://doi.org/10.1016/j.virol.2008.08.003

Nicholson, B. L., Wu, B., Chevtchenko, I., & White, K. A. (2010). Tombusvirus recruitment of host translational machinery via the 3′ UTR. *RNA*, *16*(7), 1402–1419. https://doi.org/10.1261/rna.2135210

Nicholson, B. L., Zaslaver, O., Mayberry, L. K., Browning, K. S., & White, K. A. (2013). Tombusvirus Y-Shaped Translational Enhancer Forms a Complex with eIF4F and Can Be Functionally Replaced by Heterologous Translational Enhancers. *Journal of Virology*, *87*(3), 1872–1883. https://doi.org/10.1128/JVI.02711-12

Nixon, P. L., Cornish, P. V., Suram, S. V., & Giedroc, D. P. (2002a). Thermodynamic Analysis of Conserved Loop−Stem Interactions in P1−P2 Frameshifting RNA Pseudoknots from Plant *Luteoviridae*. *Biochemistry*, *41*(34), 10665–10674. https://doi.org/10.1021/bi025843c

Nixon, P. L., Rangan, A., Kim, Y.-G., Rich, A., Hoffman, D. W., Hennig, M., & Giedroc, D. P. (2002b). Solution Structure of a Luteoviral P1–P2 Frameshifting mRNA Pseudoknot. *Journal of Molecular Biology*, *322*(3), 621–633. https://doi.org/10.1016/S0022-2836(02)00779-9

Nurkiyanova, K. M., Ryabov, E. V., Kalinina, N. O., Fan, Y., Andreev, I., Fitzgerald, A. G., Palukaitis, P., & Taliansky, M. (2001). Umbravirus-encoded movement protein induces tubule formation on the surface of protoplasts and binds RNA incompletely and non-cooperatively. *Journal of General Virology*, *82*(10), 2579–2588. https://doi.org/10.1099/0022-1317-82-10-2579

Olson, A. J., Bricogne, G., & Harrison, S. C. (1983). Structure of tomato bushy stunt virus IV. *Journal of Molecular Biology*, *171*(1), 61–93. https://doi.org/10.1016/S0022-2836(83)80314-3

Orlova, M., Yueh, A., Leung, J., & Goff, S. P. (2003). Reverse Transcriptase of Moloney Murine Leukemia Virus Binds to Eukaryotic Release Factor 1 to Modulate Suppression of Translational Termination. *Cell*, *115*(3), 319–331. https://doi.org/10.1016/S0092-8674(03)00805-5

Osman, T. A. M., Coutts, R. H. A., & Buck, K. W. (2006b). In Vitro Synthesis of Minus-Strand RNA by an Isolated *Cereal Yellow Dwarf Virus* RNA-Dependent RNA Polymerase Requires VPg and a Stem-Loop Structure at the 3′ End of the Virus RNA. *Journal of Virology*, *80*(21), 10743–10751. https://doi.org/10.1128/JVI.01050-06

Osman, T. A. M., Morris, J., Coutts, R. H. A., & Buck, K. W. (2006a). Synthesis of genomic and subgenomic RNAs by a membrane-bound RNA-dependent RNA polymerase isolated from oat plants infected with cereal yellow dwarf virus. *Archives of Virology*, *151*(11), 2229–2242. https://doi.org/10.1007/s00705-006-0789-9

Oster, S. K., Wu, B., & White, K. A. (1998). Uncoupled Expression of p33 and p92 Permits Amplification of Tomato Bushy Stunt Virus RNAs. *Journal of Virology*, *72*(7), 5845–5851. https://doi.org/10.1128/JVI.72.7.5845-5851.1998

Panavas, T., & Nagy, P. D. (2003). The RNA Replication Enhancer Element of Tombusviruses Contains Two Interchangeable Hairpins That Are Functional during Plus-Strand Synthesis. *Journal of Virology*, *77*(1), 258–269. https://doi.org/10.1128/JVI.77.1.258-269.2003

Panavas, T., Panaviene, Z., Pogany, J., & Nagy, P. D. (2003). Enhancement of RNA synthesis by promoter duplication in tombusviruses. *Virology*, *310*(1), 118–129. https://doi.org/10.1016/S0042-6822(03)00105-3

Panavas, T., Pogany, J., & Nagy, P. D. (2002). Analysis of Minimal Promoter Sequences for Plus-Strand Synthesis by the Cucumber necrosis virus RNA-Dependent RNA Polymerase. *Virology*, *296*(2), 263–274. https://doi.org/10.1006/viro.2002.1423

Panaviene, Z., Panavas, T., & Nagy, P. D. (2005). Role of an Internal and Two 3'-Terminal RNA Elements in Assembly of Tombusvirus Replicase. *Journal of Virology*, *79*(16), 10608–10618. https://doi.org/10.1128/JVI.79.16.10608-10618.2005

Pantaleo, V., Rubino, L., & Russo, M. (2003). Replication of *Carnation Italian Ringspot Virus* Defective Interfering RNA in *Saccharomyces cerevisiae*. *Journal of Virology*, *77*(3), 2116–2123. https://doi.org/10.1128/JVI.77.3.2116-2123.2003

Pantaleo, V., Rubino, L., & Russo, M. (2004). The p36 and p95 replicase proteins of Carnation Italian ringspot virus cooperate in stabilizing defective interfering RNA. *Journal of General Virology*, *85*(8), 2429–2433. https://doi.org/10.1099/vir.0.80063-0

Pasin, F., Daròs, J.-A., & Tzanetakis, I. E. (2022). Proteome expansion in the *Potyviridae* evolutionary radiation. *FEMS Microbiology Reviews*, *46*(4), fuac011. https://doi.org/10.1093/femsre/fuac011

Pasternak, A. O. (2001). Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *The EMBO Journal*, *20*(24), 7220–7228. https://doi.org/10.1093/emboj/20.24.7220

Patel, A., Treffers, E. E., Meier, M., Patel, T. R., Stetefeld, J., Snijder, E. J., & Mark, B. L. (2020). Molecular characterization of the RNA-protein complex directing −2/−1 programmed ribosomal frameshifting during arterivirus replicase expression. *Journal of Biological Chemistry*, *295*(52), 17904–17921. https://doi.org/10.1074/jbc.RA120.016105

Pathak, K. B., Pogany, J., Xu, K., White, K. A., & Nagy, P. D. (2012). Defining the Roles of *cis* - Acting RNA Elements in Tombusvirus Replicase Assembly *In Vitro*. *Journal of Virology*, *86*(1), 156–171. https://doi.org/10.1128/JVI.00404-11

Penn, W. D., & Mukhopadhyay, S. (2022). Abracadabra, One Becomes Two: The Importance of Context in Viral −1 Programmed Ribosomal Frameshifting. *MBio*, *13*(4), e02468-21. https://doi.org/10.1128/mbio.02468-21

Pogany, J., Fabian, M.R., White, K.A., Nagy, P.D. (2003). A replication silencer element in a plus-strand RNA virus. *The EMBO Journal*, *22*(20), 5602–5611. https://doi.org/10.1093/emboj/cdg523

Pogany, J., White, K. A., & Nagy, P. D. (2005). Specific Binding of Tombusvirus Replication Protein p33 to an Internal Replication Element in the Viral RNA Is Essential for Replication. *Journal of Virology*, *79*(8), 4859–4869. https://doi.org/10.1128/JVI.79.8.4859-4869.2005

Qiu, W., & Scholthof, H. B. (2001). Effects of inactivation of the coat protein and movement genes of Tomato bushy stunt virus on early accumulation of genomic and subgenomic RNAs. *Journal of General Virology*, *82*(12), 3107–3114. https://doi.org/10.1099/0022-1317-82-12-3107

Qiu, Y., Cai, D., Qi, N., Wang, Z., Zhou, X., Zhang, J., & Hu, Y. (2011). Internal Initiation Is Responsible for Synthesis of Wuhan Nodavirus Subgenomic RNA. *Journal of Virology*, *85*(9), 4440–4451. https://doi.org/10.1128/JVI.02410-10

Qu, X., Wen, J.-D., Lancaster, L., Noller, H. F., Bustamante, C., & Tinoco, I. (2011). The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature*, *475*(7354), 118–121. https://doi.org/10.1038/nature10126

Rakotondrafara, A. M., Polacek, C., Harris, E., & Miller, W. A. (2006). Oscillating kissing stem–loop interactions mediate 5′ scanning-dependent translation by a viral 3′-cap-independent translation element. *RNA*, *12*(10), 1893–1906. https://doi.org/10.1261/rna.115606

Ray, D., Na, H., & White, K. A. (2004). Structural Properties of a Multifunctional T-Shaped RNA Domain That Mediate Efficient Tomato Bushy Stunt Virus RNA Replication. *Journal of Virology*, *78*(19), 10490–10500. https://doi.org/10.1128/JVI.78.19.10490-10500.2004

Ray, D., & White, K. A. (1999). Enhancer-like Properties of an RNA Element That Modulates Tombusvirus RNA Accumulation. *Virology*, *256*(1), 162–171. https://doi.org/10.1006/viro.1999.9630

Ray, D., & White, K. A. (2003). An Internally Located RNA Hairpin Enhances Replication of *Tomato Bushy Stunt Virus* RNAs. *Journal of Virology*, *77*(1), 245–257. https://doi.org/10.1128/JVI.77.1.245-257.2003

Ray, D., Wu, B., & White, K. A. (2003). A second functional RNA domain in the 5′ UTR of the *Tomato bushy stunt virus* genome: Intra- and interdomain interactions mediate viral RNA replication. *RNA*, *9*(10), 1232–1245. https://doi.org/10.1261/rna.5630203

Reinbold, C., Lacombe, S., Ziegler-Graff, V., Scheidecker, D., Wiss, L., Beuve, M., Caranta, C., & Brault, V. (2013). Closely Related Poleroviruses Depend on Distinct Translation Initiation Factors to Infect *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions®*, *26*(2), 257–265. https://doi.org/10.1094/MPMI-07-12-0174-R

Rubino, L., Burgyan, J., & Russo, M. (1995). Molecular cloning and complete nucleotide sequence of carnation Italian ringspot tombusvirus genomic and defective interfering RNAs. *Archives of Virology*, *140*(11), 2027–2039. https://doi.org/10.1007/BF01322690

Rubino, L., Weber-Lotfi, F., Dietrich, A., Stussi-Garaud, C., & Russo, M. (2001). The open reading frame 1-encoded ('36K') protein of Carnation Italian ringspot virus localizes to mitochondria. *Journal of General Virology*, *82*(1), 29–34. https://doi.org/10.1099/0022-1317-82-1-29

Ryabov, E. V., Fraser, G., Mayo, M. A., Barker, H., & Taliansky, M. (2001a). Umbravirus Gene Expression Helps Potato leafroll virus to Invade Mesophyll Tissues and to be Transmitted Mechanically between Plants. *Virology*, *286*(2), 363–372. https://doi.org/10.1006/viro.2001.0982

Ryabov, E. V., Oparka, K. J., Santa Cruz, S., Robinson, D. J., & Taliansky, M. E. (1998). Intracellular Location of Two Groundnut Rosette Umbravirus Proteins Delivered by PVX and TMV Vectors. *Virology*, *242*(2), 303–313. https://doi.org/10.1006/viro.1997.9025

Ryabov, E. V., Roberts, I. M., Palukaitis, P., & Taliansky, M. (1999a). Host-Specific Cell-to-Cell and Long-Distance Movements of Cucumber Mosaic Virus Are Facilitated by the Movement Protein of Groundnut Rosette Virus. *Virology*, *260*(1), 98–108. https://doi.org/10.1006/viro.1999.9806

Ryabov, E. V., Robinson, D. J., & Taliansky, M. (2001b). Umbravirus-Encoded Proteins both Stabilize Heterologous Viral RNA and Mediate Its Systemic Movement in Some Plant Species. *Virology*, *288*(2), 391–400. https://doi.org/10.1006/viro.2001.1078

Ryabov, E. V., Robinson, D. J., & Taliansky, M. E. (1999b). A plant virus-encoded protein facilitates long-distance movement of heterologous viral RNA. *Proceedings of the National Academy of Sciences*, *96*(4), 1212–1217. https://doi.org/10.1073/pnas.96.4.1212

Sarawaneeyaruk, S., Iwakawa, H., Mizumoto, H., Murakami, H., Kaido, M., Mise, K., & Okuno, T. (2009). Host-dependent roles of the viral 5′ untranslated region (UTR) in RNA stabilization and cap-independent translational enhancement mediated by the 3′ UTR of Red clover necrotic mosaic virus RNA1. *Virology*, *391*(1), 107–118. https://doi.org/10.1016/j.virol.2009.05.037

Sasvari, Z., Izotova, L., Kinzy, T. G., & Nagy, P. D. (2011). Synergistic Roles of Eukaryotic Translation Elongation Factors 1Bγ and 1A in Stimulation of Tombusvirus Minus-Strand Synthesis. *PLoS Pathogens*, *7*(12), e1002438. https://doi.org/10.1371/journal.ppat.1002438

Scholthof, K.-B. G., Scholthof, H.B., & Jackson, A.O. (1995a). The Tomato Bushy Stunt Virus Replicase Proteins Are Coordinately Expressed and Membrane Associated. *Virology, 208* (1), 365–369. https://doi.org/10.1006/viro.1995.1162

Scholthof, H. B., Scholthof, K.-B. G., Kikkert, M., & Jackson, A. O. (1995b). Tomato Bushy Stunt Virus Spread Is Regulated by Two Nested Genes That Function in Cell-to-Cell Movement and Host-Dependent Systemic Invasion. *Virology*, *213*(2), 425–438. https://doi.org/10.1006/viro.1995.0015

Sharma, S. D., Kraft, J. J., Miller, W. A., & Goss, D. J. (2015). Recruitment of the 40S Ribosome Subunit to the 3′-Untranslated Region (UTR) of a Viral mRNA, via the eIF4 Complex, Facilitates Cap-independent Translation. *Journal of Biological Chemistry*, *290*(18), 11268–11281. https://doi.org/10.1074/jbc.M115.645002

Shen, R., & Miller, W. A. (2004). The 3' Untranslated Region of Tobacco Necrosis Virus RNA Contains a Barley Yellow Dwarf Virus-Like Cap-Independent Translation Element. *Journal of Virology*, *78*(9), 4655–64. https://doi.org/10.1128/JVI.78.9.4655-4664.2004

Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Muerhoff, S., Pletnev, A., Rico-Hesse, R., Smith, D. B., Stapleton, J. T., & ICTV Report Consortium. (2017). ICTV Virus Taxonomy Profile: Flaviviridae. *Journal of General Virology*, *98*(1), 2–3. https://doi.org/10.1099/jgv.0.000672

Simon, A. E., & Miller, W. A. (2013). 3′ Cap-Independent Translation Enhancers of Plant Viruses. *Annual Review of Microbiology*, *67*(1), 21–42. https://doi.org/10.1146/annurev-micro-092412-155609

Sit, T. L., & Lommel, S. A. (2015). Tombusviridae. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed., pp. 1–9). Wiley. https://doi.org/10.1002/9780470015902.a0000756.pub3

Sit, T. L., Vaewhongs, A. A., & Lommel, S. A. (1998). RNA-Mediated Trans-Activation of Transcription from a Viral RNA. *Science, New Series*, *281*(5378), 829–832. https://doi.org/10.1126/science.281.5378.829

Sivakumaran, K., Choi, S.-K., Hema, M., & Kao, C. C. (2004). Requirements for Brome Mosaic Virus Subgenomic RNA Synthesis In Vivo and Replicase-Core Promoter Interactions In Vitro. *Journal of Virology*, *78*(12), 6091–6101. https://doi.org/10.1128/JVI.78.12.6091-6101.2004

Skaf, J. S., Schultz, M. H., Hirata, H., & de Zoeten, G. A. (2000). Mutational evidence that the VPg is involved in the replication and not the movement of Pea enation mosaic virus-1. *Journal of General Virology*, *81*(4), 1103–1109. https://doi.org/10.1099/0022-1317-81-4-1103

Skuzeski, J. M., Nichols, L. M., Gesteland, R. F., & Atkins, J. F. (1991). The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *Journal of Molecular Biology*, *218*(2), 365–373. https://doi.org/10.1016/0022-2836(91)90718-L

Sola, I., Almazán, F., Zúñiga, S., & Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology*, *2*(1), 265–288. https://doi.org/10.1146/annurev-virology-100114-055218

Sola, I., Moreno, J. L., Zúñiga, S., Alonso, S., & Enjuanes, L. (2005). Role of Nucleotides Immediately Flanking the Transcription-Regulating Sequence Core in Coronavirus Subgenomic mRNA Synthesis. *Journal of Virology*, *79*(4), 2506–2516. https://doi.org/10.1128/JVI.79.4.2506-2516.2005

Sõmera, M., Fargette, D., Hébrard, E., Sarmiento, C., & ICTV Report Consortium. (2021). ICTV Virus Taxonomy Profile: Solemoviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. *Journal of General Virology*, *102*(12). https://doi.org/10.1099/jgv.0.001707

Syller, J. (2003). Molecular and biological features of umbraviruses, the unusual plant viruses lacking genetic information for a capsid protein. *Physiological and Molecular Plant Pathology*, *63*(1), 35–46. https://doi.org/10.1016/j.pmpp.2003.08.004

Sztuba-Solińska, J., Stollar, V., & Bujarski, J. J. (2011). Subgenomic messenger RNAs: Mastering regulation of (+)-strand RNA virus life cycle. *Virology*, *412*(2), 245–255. https://doi.org/10.1016/j.virol.2011.02.007

Tajima, Y., Iwakawa, H., 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*(1), 169–178. https://doi.org/10.1016/j.virol.2011.05.012

Taliansky, M. E., & Robinson, D. J. (2003). Molecular biology of umbraviruses: Phantom warriors. *Journal of General Virology*, *84*(8), 1951–1960. https://doi.org/10.1099/vir.0.19219-0

Taliansky, M., Roberts, I. M., Kalinina, N., Ryabov, E. V., Raj, S. K., Robinson, D. J., & Oparka, K. J. (2003). An Umbraviral Protein, Involved in Long-Distance RNA Movement, Binds Viral RNA and Forms Unique, Protective Ribonucleoprotein Complexes. *Journal of Virology*, *77*(5), 3031–3040. https://doi.org/10.1128/JVI.77.5.3031-3040.2003

Tholstrup, J., Oddershede, L. B., & Sørensen, M. A. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Research*, *40*(1), 303–313. https://doi.org/10.1093/nar/gkr686

Treder, K., Pettit Kneller, E. L., Allen, E. M., Wang, Z., Browning, K. S., & Miller, W. A. (2008). The 3′ cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the

eIF4G subunit to initiate translation. *RNA*, *14*(1), 134–147. https://doi.org/10.1261/rna.777308

Truniger, V., Miras, M., & Aranda, M. A. (2017). Structural and Functional Diversity of Plant Virus 3′-Cap-Independent Translation Enhancers (3′-CITEs). *Frontiers in Plant Science*, *8*, 2047. https://doi.org/10.3389/fpls.2017.02047

Van Marle, G., Dobbe, J. C., Gultyaev, A. P., Luytjes, W., Spaan, W. J. M., & Snijder, E. J. (1999). Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proceedings of the National Academy of Sciences*, *96*(21), 12056–12061. https://doi.org/10.1073/pnas.96.21.12056

Várallyay, É., Válóczi, A., Ágyi, Á., Burgyán, J., & Havelda, Z. (2010). Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *The EMBO Journal*, *29*(20), 3507–3519. https://doi.org/10.1038/emboj.2010.215

Vargason, J. M., Szittya, G., Burgyán, J., & Hall, T. M. T. (2003). Size Selective Recognition of siRNA by an RNA Silencing Suppressor. *Cell*, *115*(7), 799–811. https://doi.org/10.1016/S0092-8674(03)00984-X

Venkataraman, S., Prasad, B., & Selvarajan, R. (2018). RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*, *10*(2), 76. https://doi.org/10.3390/v10020076

Wang, S., Browning, K. S., & Miller, W. A. (1997). A viral sequence in the 3'-untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. *EMBO J.*, 16(13), 4107–4116. https://doi.org/10.1093/emboj/16.13.4107

Wang, S., Mortazavi, L., & White, K. A. (2008). Higher-Order RNA Structural Requirements and Small-Molecule Induction of Tombusvirus Subgenomic mRNA Transcription. *Journal of Virology*, *82*(8), 3864–3871. https://doi.org/10.1128/JVI.02416-07

Wang, Z., Parisien, M., Scheets, K., & Miller, W. A. (2011). The Cap-Binding Translation Initiation Factor, eIF4E, Binds a Pseudoknot in a Viral Cap-Independent Translation Element. *Structure*, *19*(6), 868–880. https://doi.org/10.1016/j.str.2011.03.013

Wang, Z., Treder, K., & Miller, W. A. (2009). Structure of a Viral Cap-independent Translation Element That Functions via High Affinity Binding to the eIF4E Subunit of eIF4F. *Journal of Biological Chemistry*, *284*(21), 14189–14202. https://doi.org/10.1074/jbc.M808841200

White, K. A. (2002). The Premature Termination Model: A Possible Third Mechanism for Subgenomic mRNA Transcription in (+)-Strand RNA Viruses. *Virology*, *304*(2), 147–154. https://doi.org/10.1006/viro.2002.1732

White, K.A. (2011). The Springer Index of Viruses. Tidona, C., & Darai, G. (Eds.). Springer New York. https://doi.org/10.1007/978-0-387-95919-1

White, K. A., & Nagy, P. D. (2004). Advances in the Molecular Biology of Tombusviruses: Gene Expression, Genome Replication, and Recombination. *Progress in Nucleic Acid Research and Molecular Biology*, *78*, 187–226. https://doi.org/10.1016/S0079-6603(04)78005-8

Wielgosz, M. M., Raju, R., & Huang, H. V. (2001). Sequence Requirements for Sindbis Virus Subgenomic mRNA Promoter Function in Cultured Cells. *Journal of Virology*, *75*(8), 3509–3519. https://doi.org/10.1128/JVI.75.8.3509-3519.2001

Wilson, J. E., Powell, M. J., Hoover, S. E., & Sarnow, P. (2000). Naturally Occurring Dicistronic Cricket Paralysis Virus RNA Is Regulated by Two Internal Ribosome Entry Sites. *Molecular and Cellular Biology*, *20*(14), 4990–4999. https://doi.org/10.1128/MCB.20.14.4990-4999.2000

Wobus, C. E., Skaf, J. S., Schultz, M. H., & De Zoeten, G. A. (1998). Sequencing, genomic localization and initial characterization of the VPg of pea enation mosaic enamovirus. *Journal of General Virology*, *79*(8), 2023–2025. https://doi.org/10.1099/0022-1317-79-8-2023

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., & Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. *MBio*, *9*(6), e02329-18. https://doi.org/10.1128/mBio.02329-18

Woo, P.C.Y., de Groot, R.J., Haagmans, B., Lau, S.K.P., Neuman, B.W., Perlman, S., Sola, I., van der Hoek, L., Wong, A.C.P. and Yeh, S-H. (2023) ICTV Virus Taxonomy Profile: Coronaviridae. *Journal of General Virology*, (in press).

Wu, B., Oliveri, S., Mandic, J., & White, K. A. (2010). Evidence for a Premature Termination Mechanism of Subgenomic mRNA Transcription in a Carmovirus. *Journal of Virology*, *84*(15), 7904–7907. https://doi.org/10.1128/JVI.00742-10

Wu, B., Pogany, J., Na, H., Nicholson, B. L., Nagy, P. D., & White, K. A. (2009). A Discontinuous RNA Platform Mediates RNA Virus Replication: Building an Integrated Model for RNA–based Regulation of Viral Processes. *PLoS Pathogens*, *5*(3), e1000323. https://doi.org/10.1371/journal.ppat.1000323

Wu, B., Vanti, W. B., & White, K. A. (2001). An RNA domain within the 5′ untranslated region of the tomato bushy stunt virus genome modulates viral RNA replication. *Journal of Molecular Biology*, *305*(4), 741–756. https://doi.org/10.1006/jmbi.2000.4298

Wu, B., & White, K. A. (1998). Formation and Amplification of a Novel Tombusvirus Defective RNA Which Lacks the 5′ Nontranslated Region of the Viral Genome. *Journal of Virology*, *72*(12), 9897–9905. https://doi.org/10.1128/JVI.72.12.9897-9905.1998

Wu, B., & White, K. A. (1999). A Primary Determinant of Cap-Independent Translation Is Located in the 3'-Proximal Region of the Tomato Bushy Stunt Virus Genome. *Journal of Virology*, *73* (11), 8982–8988. https://doi.org/10.1128/JVI.73.11.8982-8988.1999

Wu, B., & White, K. A. (2007). Uncoupling RNA virus replication from transcription via the polymerase: Functional and evolutionary insights. *The EMBO Journal*, *26*(24), 5120–5130. https://doi.org/10.1038/sj.emboj.7601931

Xu, W., & White, K. A. (2008). Subgenomic mRNA transcription in an aureusvirus: Down-regulation of transcription and evolution of regulatory RNA elements. *Virology*, *371*(2), 430–438. https://doi.org/10.1016/j.virol.2007.09.035

Xu, W., & White, K. A. (2009). RNA-Based Regulation of Transcription and Translation of Aureusvirus Subgenomic mRNA1. *Journal of Virology*, *83*(19), 10096–10105. https://doi.org/10.1128/JVI.00376-09

Xu, Y., Ju, H.-J., DeBlasio, S., Carino, E. J., Johnson, R., MacCoss, M. J., Heck, M., Miller, W. A., & Gray, S. M. (2018). A Stem-Loop Structure in *Potato Leafroll Virus* Open Reading Frame 5 (ORF5) Is Essential for Readthrough Translation of the Coat Protein ORF Stop Codon 700 Bases Upstream. *Journal of Virology*, *92*(11), e01544-17. https://doi.org/10.1128/JVI.01544-17

Yamamura, Y., & Scholthof, H. B. (2005). Tomato bushy stunt virus: A resilient model system to study virus-plant interactions. *Molecular Plant Pathology*, *6*(5), 491–502. https://doi.org/10.1111/j.1364-3703.2005.00301.x

Ye, K., Malinina, L., & Patel, D.J. (2003). Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature, 426*, 874–878. https://doi.org/10.1038/nature02213

Ye, L., Ambi, U. B., Olguin-Nava, M., Gribling-Burrer, A.-S., Ahmad, S., Bohn, P., Weber, M. M., & Smyth, R. P. (2021). RNA Structures and Their Role in Selective Genome Packaging. *Viruses*, *13*(9), 1788. https://doi.org/10.3390/v13091788

Zhang, G., Slowinski, V., & White, K. A. (1999). Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA*, *5*(4), 550–561. https://doi.org/10.1017/S1355838299982080

Zhang, Z., He, G., Filipowicz, N. A., Randall, G., Belov, G. A., Kopek, B. G., & Wang, X. (2019). Host Lipids in Positive-Strand RNA Virus Genome Replication. *Frontiers in Microbiology*, *10*, 286. https://doi.org/10.3389/fmicb.2019.00286

Zúñiga, S., Sola, I., Alonso, S., & Enjuanes, L. (2004). Sequence Motifs Involved in the Regulation of Discontinuous Coronavirus Subgenomic RNA Synthesis. *Journal of Virology*, *78*(2), 980–994. https://doi.org/10.1128/JVI.78.2.980-994.2004

CHAPTER 2

ACTIVATION OF VIRAL TRANSCRIPTION BY STEPWISE LARGESCALE FOLDING OF AN RNA VIRUS GENOME

CIRV transcribes its CP-encoding sg mRNA1 via the premature termination mechanism. A previous study identified that sg mRNA1 attenuation structure is formed by a long-distance RNA-RNA base-pairing interaction between complementary sequences AS1 and RS1 (Choi and White, 2002). This chapter provides further analysis of RNA elements involved in promoting sg mRNA1 transcription. The results demonstrate that activation of sg mRNA1 transcription is more complex than previously appreciated and involves additional long-range interactions for assembly of a functional attenuation structure.

This chapter is presented as a peer-reviewed journal article - "Activation of viral transcription by stepwise largescale folding of an RNA virus genome" by Tamari Chkuaseli and K. Andrew White, published in *Nucleic Acids Research* (Chkuaseli and White, 2020). I conceptualized and designed the experiments for the study together with Dr. K. Andrew White. I performed all experiments and all data analyses and wrote the first draft of the manuscript.

Activation of viral transcription by stepwise largescale folding of an RNA virus genome

Tamari Chkuaseli and K. Andrew White [ID]*

Department of Biology, York University, Toronto, Ontario M3J 1P3, Canada

ABSTRACT

The genomes of RNA viruses contain regulatory elements of varying complexity. Many plus-strand RNA viruses employ largescale intra-genomic RNA-RNA interactions as a means to control viral processes. Here, we describe an elaborate RNA structure formed by multiple distant regions in a tombusvirus genome that activates transcription of a viral subgenomic mRNA. The initial step in assembly of this intramolecular RNA complex involves the folding of a large viral RNA domain, which generates a discontinuous binding pocket. Next, a distally-located protracted stem-loop RNA structure docks, via base-pairing, into the binding site and acts as a linchpin that stabilizes the RNA complex and activates transcription. A multi-step RNA folding pathway is proposed in which rate-limiting steps contribute to a delay in transcription of the capsid protein-encoding viral subgenomic mRNA. This study provides an exceptional example of the complexity of genome-scale viral regulation and offers new insights into the assembly schemes utilized by large intra-genomic RNA structures.

INTRODUCTION

Positive-strand RNA viruses comprise a large group of agriculturally and medically important pathogens that infect a wide range of hosts. The successful takeover of their hosts requires multiple steps that involves precise regulation and careful coordination. A critical component of this control is the modulation of different viral processes by RNA sequences and structures located within viral genomes (1–4). In some cases, the RNA-based regulation is mediated by large functional RNA folds, some of which span the entire length of a viral genome (5). Accordingly, overall RNA genome architecture and dynamics can contribute significantly to the orchestration of different phases that occur during viral infections (6,7). Notably, this large-scale form of riboregulation is employed by many significant plant and animal messenger-sensed RNA viruses, including luteoviruses (8,9), carmoviruses (10,11), umbraviruses (12,13), flaviviruses (14–17), hepacivirus (18–24) and coronaviruses (25–27).

Tombusviruses (family Tombusviridae) are important model plus-strand RNA viruses (28). Studies performed on members of this genus have resulted in pioneering discoveries (29–31) and led to significant progress in the identification of pro- and antiviral host factors (32–34). Tombusviruses have also been invaluable for investigating how global viral RNA genome structure actively controls essential viral processes (5,6). Their 4.8 kb-long coding-sensed ssRNA genomes contain a vast network of intra-genomic, base pair-mediated, long-distance RNA–RNA interactions (LDRIs) that play different critical roles during the viral reproductive cycle (7). In particular, two tombusviruses, the prototype of the genus, Tomato bushy stunt virus (TBSV), and the closely-related Carnation Italian ringspot virus (CIRV) (Figure 1A), have been instrumental in deducing the structure and function of this complex LDRI network (35).

Tombusvirus RNA genomes are not 5′-capped or 3′-polyadenylated, thus they rely on an unconventional mode of translation, which has been studied extensively in CIRV (36–39). An RNA structure in CIRV's 3′-untranslated region (3′UTR), termed the 3′-cap independent translation enhancer (3′CITE), binds to eukaryotic translation initiation factor 4F (eIF4F). The eIF4F-bound 3′CITE then simultaneously base-pairs with the 5′UTR via an LDRI, which positions eIF4F near the 5′-end of the genome, where it mediates ribosome recruitment (38) (Figure 1A). This results in translation of the auxiliary RNA replication protein, p36. Production of the p95 RNA-dependent RNA polymerase (RdRp) requires translational readthrough of the p36 stop codon. This recoding event involves an extended RNA stem-loop (SL) structure, termed the readthrough SL (RTSL), located immediately 3′ to the p36 termination codon, UAG (Figure 1A, green asterisk). The RTSL is not able to direct readthrough on its own and, to function, requires the formation of an LDRI between a bulged sequence in RTSL (the proximal readthrough element, PRTE) and a complementary sequence (the distal readthrough element, DRTE) in the 3′UTR of the genome (39) (Figure 1A). This LDRI not only promotes readthrough, it also concomitantly inhibits genomic minus-

*To whom correspondence should be addressed. Tel: +1 416 736 2100 (Ext 40890 or 70352); Fax: +1 416 736 5698; Email: kawhite@yorku.ca
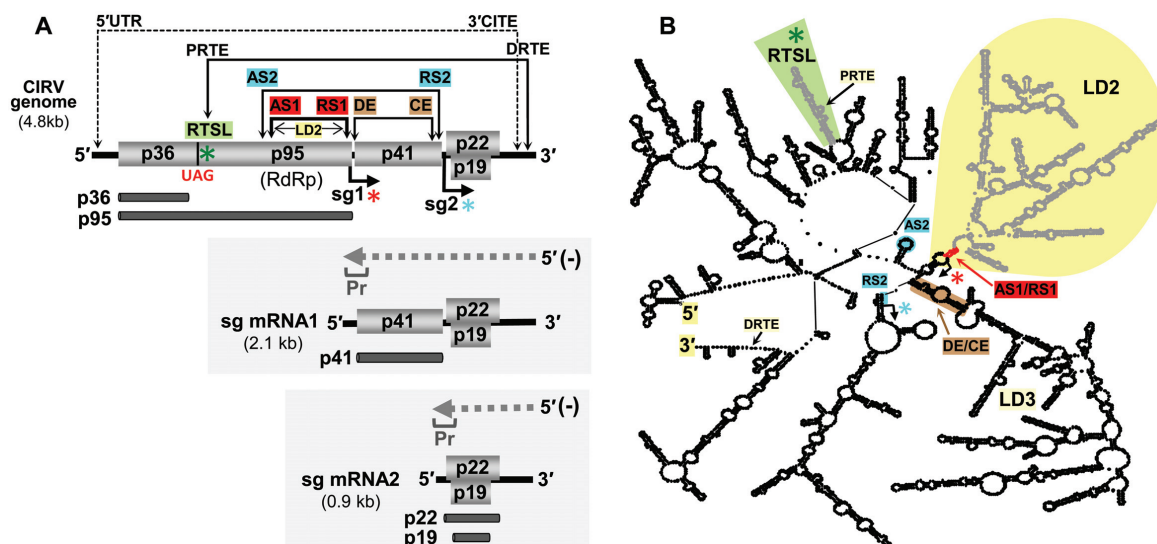
Figure 1. Carnation Italian ringspot virus (CIRV) genome and subgenomic mRNAs. (**A**) Schematic representation of the CIRV genome (top) and sg mRNAs (below in shaded boxes) with encoded proteins indicated. The small black arrows below the genome indicate the initiation sites for sg mRNA1 (sg1, red asterisk) and 2 (sg2, turquoise asterisks) in the genome. Proteins translated from the genome and sg mRNAs are depicted as dark gray bars immediately below encoded proteins. The positions of the sequences that form long-distance RNA-RNA interactions (LDRIs) involved in translation (5′UTR/3′CITE), readthrough (PRTE/DRTE), sg mRNA2 transcription (DE/CE, brown and AS2/RS2, turquoise) and sg mRNA1 transcription (AS1/RS1, red) are indicated by double-headed arrows. The minus-strand intermediates in transcription are shown above each sg mRNA as dashed grey arrows, with promoter sequences (Pr) indicated at their 3′-ends. (**B**) Simplified model of the RNA secondary structure for the TBSV genome (48) with relevant large RNA domains (LD2 and LD3), RNA elements (RTSL, green shading, PRTE and DRTE), and LDRIs indicated. The LDRI color-coding scheme corresponds to that depicted in panel (A).

strand RNA synthesis, which would interfere with translation. Thus, the RTSL, via an LDRI, functions as a dual regulator that coordinates translational recoding and genome replication.

LDRIs are also involved in controlling the production of tombusvirus subgenomic (sg) mRNAs, which are small virus genome-derived mRNAs that are transcribed by the viral RdRp during infections (40,41). Structurally, sg mRNAs are 3′-coterminal with the viral genome, while their 5′-ends map to internal regions. Consequently, they encode 3′-proximal ORFs that are translationally silent within the context of the full-length genome. By modulating sg mRNA transcription, the virus is able to control the amount and timing of viral protein production during infections. Tombusviruses transcribe two sg mRNAs (Figure 1A). The smaller sg mRNA2 is transcribed earlier during infections, and mediates translation of both the p19 suppressor of gene silencing and the p22 cell-to-cell movement protein. The larger sg mRNA1 is transcribed later in infections, and encodes the capsid protein (CP) (42,43).

Tombusviruses (40), nodaviruses (family Nodaviridae) (44) and toroviruses (family Tobaniviridae) (45) transcribe their sg mRNAs using a premature termination mechanism (46). In this process, the viral RdRp terminates transcription prematurely while synthesizing a minus-strand from a full-length plus-strand viral RNA genome. The stalling of the RdRp occurs when it encounters an RNA element within the genome called an attenuation structure. This termination event leads to the production of a 3′-truncated minus-sense RNA species that possess a promoter sequence

at its 3′-end (Figure 1A, Pr). The promoter is then recognized by the viral RdRp, which transcribes the coding-sense sg mRNAs from the truncated intermediate.

The attenuation structures that block the progression of RdRps are helical RNA structures that are located ∼2–5 nt upstream from where the copying RdRp stalls. In some viruses, the inhibitory stem is formed by LDRIs (40,44). Production of tombusvirus sg mRNA2 involves two sets of LDRIs. One occurs between activator sequence 2 (AS2) and receptor sequence 2 (RS2), spanning ∼2100 nucleotides (47), and the other involves distal element (DE) and core element (CE), traversing ∼1100 nucleotides (42) (Figure 1A, turquoise and brown). When viewed in the context of the RNA secondary structure model for the TBSV genome (48), the DE/CE interaction corresponds to the closing stem of a sizable RNA domain, termed large domain 3 (LD3), which, along with formation of the adjacent LD2, acts to unite the AS2 and RS2 sequences (Figure 1B).

Efficient sg mRNA1 transcription requires an LDRI between AS1 and RS1, which spans ∼1000 nucleotides (Figure 1A, red) and forms a helix just three nucleotides upstream of the sg mRNA1 initiation site (Supplementary Figure S1) (49). The 7 nt long AS1 sequence is the terminal loop of an RNA hairpin, designated AS1-SL, that facilitates its accessibility (Supplementary Figure S1) (49). The AS1/RS1 interaction has been verified experimentally to (i) pair and operate in the plus-strand of the genome, (ii) occur *in cis* and (iii) promote production of sg mRNA1 minus-strand intermediates (49). The AS1/RS1 interaction is also predicted to form the closing helix of LD2 (Figure 1B), thus

accurate folding of LD2 is proposed to be important for formation of the AS1/RS1 LDRI (48,49).

In this study, we show that the attenuation structure for sg mRNA1 is far more complex than previously appreciated, with the AS1/RS1 interaction being a component of a group of critical LDRIs. Unexpectedly, the active RNA structure includes the recoding RNA element, RTSL, as well as specific subsections of LD2. Formation of a functional attenuation structure requires multiple LDRIs within LD2 that generate a discontinuous binding site for RTSL. The docking of RTSL into this binding pocket acts as a linchpin that stabilizes an active conformation of the RNA complex. Functional and structural aspects of these LDRIs are discussed and a likely path for assembly of this intragenomic RNA attenuation structure is presented.

MATERIALS AND METHODS

Plasmid DNA construction

Nucleotide substitutions were introduced into a cloned cDNA copy of the full-length wt CIRV genome (50) through standard PCR-based site-directed mutagenesis. Each of the mutated CIRV clones was sequenced over the entire inserted PCR fragment containing the modification to confirm that only the desired change was present.

Preparation of infectious CIRV genomic RNA

SmaI-linearized wild type (wt) and mutant full-length CIRV genome cDNAs were used as templates for *in vitro* transcription reactions using a T7 Flashscribe transcription kit (CellScript) to synthesize uncapped genomic CIRV RNAs, as described previously (51).

In vitro translation assay

In vitro-generated viral genomic RNAs (0.5 pmol) were assessed for translation and readthrough using a wheat germ extract (wge) *in vitro* translation system (Promega) and proteins were monitored by incorporation of [^{35}S]-methionine, as described previously (36,39). Translation products were separated by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis, detected using a Typhoon FLA 9500 variable mode imager (GE Healthcare), and quantified using QuantityOne software (Bio-Rad). Each wge translation experiment was performed three times independently and averages and standard errors of the mean (SEM) were calculated. Readthrough levels were calculated as a ratio of the amount of p95 readthrough product relative to that of its corresponding p36 pre-readthrough product, with the ratio for the wt genome set as 100% (36,39).

Protoplast transfection and viral RNA analysis

Production of genomic and subgenomic CIRV RNAs were assessed after protoplast infections, as described previously (51). Protoplasts were prepared from the cotyledons of 6-day old cucumber plants. For each viral RNA genome tested, ∼500 000 protoplasts were transfected with 5 μg of CIRV transcript using polyethylene glycol and CaCl$_2$

(51). Transfected protoplasts were incubated under constant fluorescent light at 22°C for 22 h. Total nucleic acids were extracted and separated by agarose gel electrophoresis and plant 25S ribosomal RNA bands were monitored as controls to ensure even loading. Total nucleic acids were then transferred to a nylon membrane and plus-strand viral RNA accumulation levels were assessed using a [γ -^{32}P]-labeled oligonucleotide probe complementary to the 3′-end of the CIRV genome and subgenomic mRNAs (coordinates 4739–4760). Northern blots were imaged using a Typhoon FLA 9500 and RNA bands were quantified using the QuantityOne software. Relative sg mRNA1 levels were calculated as the ratio of sg mRNA1 levels to their cognate genome levels, with the wt ratio set to 100%. Each set of protoplast transfections was carried out three times independently and averages and SEM values were calculated.

Minus-strand viral RNA accumulation was analyzed as described previously (52). Briefly, total nucleic acids isolated from protoplast infections were denatured with dimethyl sulfoxide and glyoxal and separated by agarose gel electrophoresis in 10 mM sodium phosphate buffer (pH 7.0). Northern blotting, imaging, and data analysis was performed as described earlier, except that [α-^{32}P]-UTP-labeled riboprobe, corresponding to the 3′-end of CIRV cDNA (coordinates 4381–4760), was used for detection.

Electrophoretic mobility shift assay (EMSA)

DNA fragments of RTSL and LD2 and their derivatives were generated by standard PCR that incorporated a T7 promoter upstream of the 5′-ends of the RNA-encoding region. Individual, or mixtures of, *in vitro*-transcribed RNA fragments (10 pmol each in 3.6 μl of water) were heated at 94°C for 3 min, then combined with 0.4 μl RNA binding buffer (final concentration: 5 mM HEPES pH7.8, 6 mM MgCl$_2$, 100 mM KCl. 3.8% glycerol) (39,53,54). The tubes were placed at 37°C for 30 min and snap-cooled on ice for 2 min. An equal volume of sterile 20% glycerol was added to each sample and the entire contents were separated by non-denaturing 4% (or 8%) polyacrylamide gel electrophoresis in a running buffer containing 45 mM Tris, pH 8.3, 43 mM boric acid and 1 mM MgCl$_2$ (54). Gels were then stained with 1 mg/mL ethidium bromide (39), imaged using Typhoon FLA 9500 scanner, and RNA bands were quantified using the QuantityOne software. Relative binding efficiencies were determined by quantifying the amount of shifted LD2 or LD2-core by comparing their levels in LD2-only or LD2-core-only lanes with their corresponding unbound levels in mixtures with RTSL. Thus, relative binding efficiency is presented as a percentage of shifted LD2 or LD2-core. Each EMSA experiment was conducted three times independently, with averages and SEMs provided.

In-line probing of RNA secondary structure

In vitro-generated RNA transcripts of wt LD2-core and wt RTSL were purified using two cycles of the crush-soak RNA purification method (55). Purified transcripts were then dephosphorylated using calf-intestinal phosphatase (NEB) and 5′-end labeled using [γ-^{32}P]-ATP and T4 polynucleotide kinase (NEB). End-labeled transcripts were recovered by G-50 column chromatography and used for in-line

reactions that were carried out at 25°C for 40 hours in 1x in-line reaction buffer (50 mM Tris–HCl pH 8.3, 100 mM KCl, 20 mM MgCl$_2$) (56,57). Reactions contained labeled fragments individually (∼1 pmol) or as a mixture with their unlabeled partner fragment (30 pmol) (56,57). Labeled fragments were also used to generate untreated controls, as well as size ladders generated by alkaline hydrolysis or RNase T1 digestion. All samples were separated in 10% denaturing polyacrylamide gels (58) and imaged and quantified as described in the previous sections. In-line probing was performed twice, with consistent results. Reactivities were used to generate an in-line-guided secondary structure model for LD2-core, RTSL or a complex of both as described in Supplementary Figures S10–S12. RNA secondary structures presented were generated using *RNA2Drawer* software (59).

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)

SHAPE analysis of the LD2 region of the CIRV genome was performed using 1-methyl-7-nitroisatoic anhydride (1M7), as described previously (39). Four primers were used to map the secondary structure of the LD2 (primer coordinates in the CIRV genome: 1871–1895, 2169–2191, 2444–2464 and 2800–2824). Following fluorescent capillary sequencing, the raw data was analyzed using the *ShapeFinder* software (60) to generate relative reactivities at single nucleotide resolution. SHAPE reactions were performed twice for each of the four primers and average reactivities were used. The reactivity data was normalized against the average of the ten highest reactivity values, as described previously (39). The *RNAStructure* web server was used to combine SHAPE reactivity data with thermodynamic prediction to generate a secondary structure model of LD2 as described in Supplementary Figure S2 (61). RNA secondary structures presented were generated using *RNA2Drawer* software (59).

RESULTS

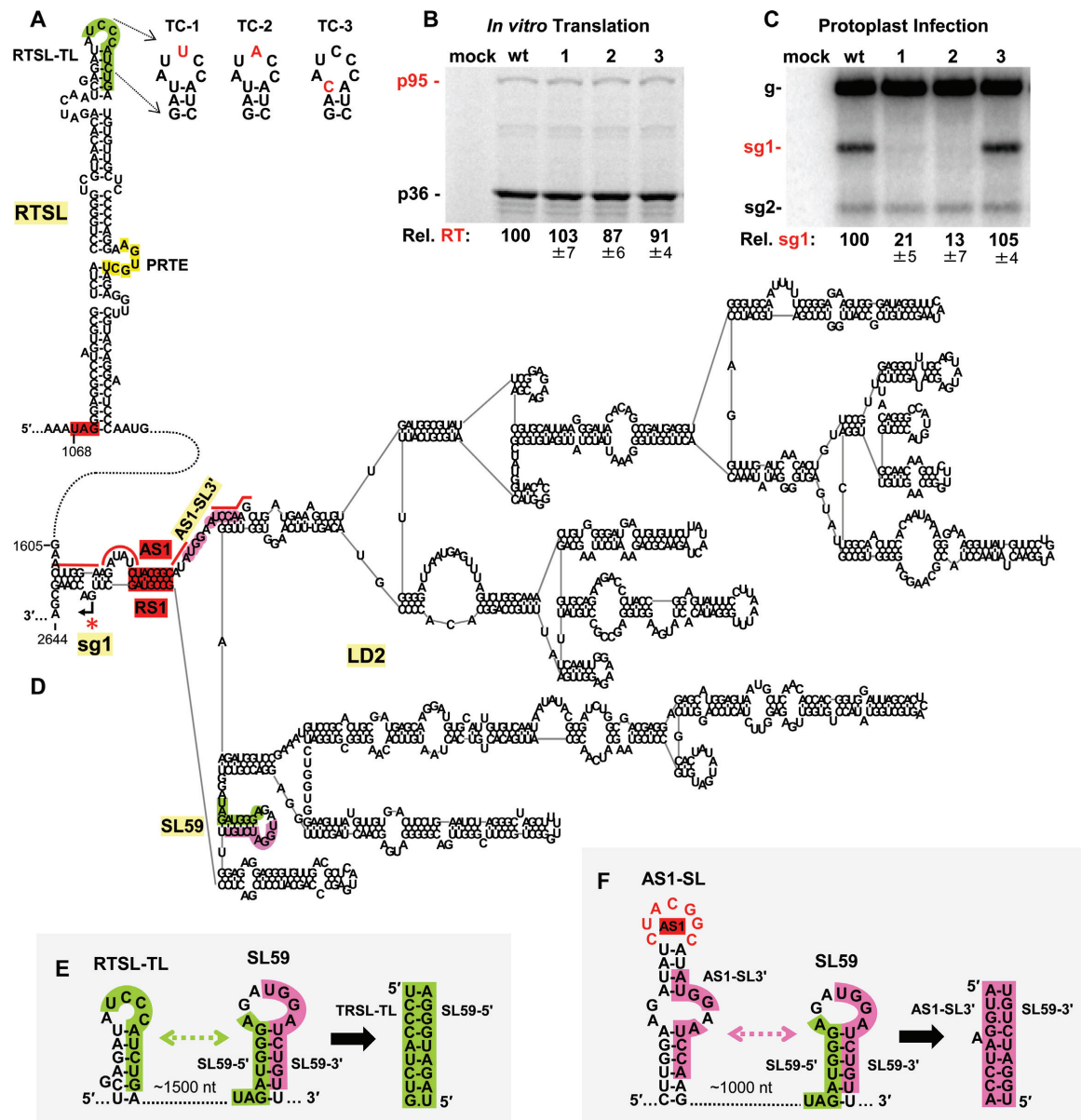### The RTSL terminal loop (RTSL-TL) regulates sg mRNA1 accumulation via an LDRI

Translational readthrough for the CIRV genome requires a long-distance RNA–RNA interaction (LDRI) between RTSL and the 3′UTR, involving the PRTE and DRTE partner sequences, respectively (Figure 1A, B) (39). To investigate the possible involvement of other regions of the RTSL in the readthrough process, silent nucleotide substitutions were introduced into its terminal loop (mutants TC-1 and TC-2) and closing base pair (TC-3) (Figure 2A). *In vitro* translation of CIRV genomes containing these modifications showed that readthrough production of p95 was similar to wt, or moderately affected (∼103% to ∼87%) (Figure 2B). However, northern blot analysis of protoplasts transfected with the same mutant viral genomes revealed an unanticipated role for the terminal loop of RTSL (RTSL-TL) in facilitating the accumulation of sg mRNA1. In these infections, sg mRNA1 levels were quantified relative to the corresponding levels of their cognate genomes, with that for wt set at 100%. Both terminal loop substitutions resulted in a ∼fivefold decrease in relative sg mRNA1 accumulation,

whereas alteration in the loop's closing base pair yielded wt levels (Figure 2C). Notably, the negative effects of the RTSL-TL mutants were specific for sg mRNA1, as typical levels of sg mRNA2 were maintained. Also, because the modifications introduced were not present in sg mRNA1, altered RNA stability was ruled out as a cause for the observed decreases. Instead, the results indicated a role for RTSL-TL in regulating the transcriptional efficiency of sg mRNA1.

Modulation of sg mRNA1 transcription by RTSL-TL could occur by it interacting with a protein factor or a complementary RNA sequence in the CIRV genome. As tombusviruses are known for controlling important viral processes via intra-genomic RNA-RNA interactions, the latter possibility was deemed more probable (36,39,41,47,49,62). To regulate sg mRNA1 transcription, RTSL-TL would likely have to interact with a sequence located near the initiation site for sg mRNA1 transcription. In tombusviruses, this initiation site is positioned just downstream from the transcription-promoting AS1/RS1 interaction (Supplementary Figure S1), which forms the closing stem of the RNA domain LD2, as shown for TBSV (Figure 1B) (48,49). Corresponding structure probing analysis (63) of the CIRV genome predicted a comparable AS1/RS1-containing LD2 (Figure 2D and Supplementary Figure S2A) that structurally mimicked that in TBSV (Supplementary Figure S2B). Also, the CIRV AS1/RS1 LDRI was shown, as demonstrated previously for TBSV (49), to be necessary for sg mRNA1 transcription (Supplementary Figure S3).

CIRV's LD2 was examined for a potential base-pairing partner for RTSL-TL, and a candidate 9 nt long segment was identified 70 nt upstream from the transcription initiation site for sg mRNA1. This sequence was present within a predicted RNA hairpin structure, SL59, located within the 3′-end of the p95 ORF, some ∼1500 nt away from RTSL-TL (Figure 2D, green). Its partner sequence, RTSL-TL, was present in the terminal loop of RTSL and extended into the adjoining 3′-stem region (Figure 2A, green); thus, for the RTSL/SL59 interaction to occur, the helical region of RTSL-TL would need to unpair. Similarly, to associate with RTSL-TL, the complementary partner sequence in SL59, comprising the 5′-half of this hairpin (herein termed SL59–5′, green), would have to unpair from its 3′-half (SL59–3′, pink) (Figure 2E). A potential base pairing partner for the displaced SL59–3′ was also identified that mapped to the 3′-half of AS1-SL, termed AS1-SL3′ (pink) (Figure 2F). Consequently, the binding of RTSL-TL to SL59–5′ (green interaction) could be accompanied by an intra-LD2 interaction (pink) (Supplementary Figure S4A), both of which (in addition to the AS1-RS1 interaction) were supported by comparative sequence analysis showing maintenance of the base pairing, despite sequence variations (Supplementary Figure S4B).

The binding of RTSL-TL to SL59–5′ was investigated functionally by introducing compensatory nucleotide substitutions into the candidate partner sequences and assessing the effects on sg mRNA1 accumulation following transfection of mutant viral RNA genomes into protoplasts (Figure 3A–C). Whenever possible, translationally neutral or conservative substitutions were used. Disruption of base

Figure 2. RTSL regulates sg mRNA1 transcription. (**A**) RNA secondary structure of RTSL (39) with the p36 stop codon highlighted in red, the PRTE depicted in yellow, and a sequence of interest (RTSL-TL) in the terminal loop colored in green. Nucleotide substitutions introduced into RTSL in genomic mutants are shown in red. All substitutions were made at wobble positions and did not alter the p95 amino acid sequence. (**B**) *In vitro* translation analysis in wheat germ extract of wt and mutant CIRV RNA genomes shown in panel (A). Samples, labeled at the top, were separated in a 12% SDS polyacrylamide gel and the positions of CIRV proteins are indicated on the left. Average p95 readthrough levels (Rel. RT) for each mutant relative to that of the wt (set as 100%) are shown below the gel, along with standard errors from three independent experiments. (**C**) Northern blot analysis of total nucleic acids isolated from plant protoplasts transfected with wt and mutant CIRV RNA genomes shown in panel (A). Respective viral infections, labeled at the top, were probed for plus-strand viral RNAs. The positions of the viral genome (g), sg mRNA1 (sg1) and sg mRNA2 (sg2), are shown on the left. Average sg mRNA1 accumulation levels, relative to that of the wt (set as 100%), are shown below the blot with standard errors determined from three independent experiments. (**D**) RNA secondary structure of the LD2 domain in CIRV deduced from structure probing analysis (Supplementary Figure S2A). The ∼400 nucleotides between RTSL and LD2 are shown as a connecting dashed line. Relevant features of LD2 include the AS1/RS1 LDRI in red, and SL59, composed of sequences that are complementary to RTSL-TL (green) and AS1-SL3′ (pink). The sg mRNA1 transcription initiation site is denoted by an arrow and a red asterisk. Sequences corresponding to the two halves of the stem in AS1-SL are indicated by red lines. (**E**) Sequences and structures involved in forming the RTSL-TL/SL59–5′ (green) interaction. (**F**) Sequences and structures involved in forming the AS1-SL3′/SL59–3′ (pink) interaction.

Figure 3. Functional analysis of candidate LDRIs in regulating sg mRNA1 transcription. (**A**, **D**) Compensatory nucleotide substitutions introduced in RTSL-TL/SL59–5′ (green) and AS1-SL3′/SL59–3′ (pink) LDRIs in mutant genomes are shown as red and white nucleotides, respectively. Any corresponding changes in the p95 amino acid sequence are shown under the mutant sequences. (**B, E**) Northern blot analysis of plus-strand CIRV RNAs isolated from protoplasts transfected with wt and mutant CIRV RNA genomes depicted in panels (A) and (D), respectively. Identities of the viral genomes tested are indicated above the blots and positions of positive-sense genome and sg mRNAs are shown on the left. Average sg mRNA1 accumulation levels relative to that of the wt are provided below the blots with standard errors obtained from three independent experiments. (**C, F**) Northern blot analysis of minus-strand CIRV RNAs isolated from protoplast infections. Identities of the viral genomes tested are indicated above the blots and positions of the minus-sense genome and sg mRNAs are shown on the left. Average minus-strand sg mRNA1 accumulation levels relative to that of the wt are provided below the blots with standard errors obtained from three independent experiments. (**G**) Simplified RNA secondary structure depiction of CIRV's LD2. Regulatory RNA elements important for sg mRNA1 transcription are highlighted: AS1/RS1 (red), AS1-SL3′ (pink), S38 (orange), S56 (blue), SL59–5′ (green) and SL59–3′ (pink). A core regulatory region at the base of LD2 is defined by a black dashed line. The arrow and the red asterisk show the sg mRNA1 transcription initiation site.

pairing potential in mutants TC-6 and TC-7 diminished sg mRNA1 plus- and minus-strand levels below ∼10% of wt, while regenerating pairing capacity with alternate nucleotides in mutant TC-8 restored levels up to ∼50–62% of wt (Figure 3B, C). This correlation between base pairing stability and sg mRNA1 accumulation is consistent with a role for the interaction in mediating transcription of sg

mRNA1. Notably, the low levels of accumulation of the intermediate minus-strand sg mRNA1 templates in TC-6 and TC-7 indicated that disrupting the RTSL-TL/SL59–5′ interaction hindered proper formation of the RdRp attenuation structure for sg mRNA1 (Figure 3C). Similar results were observed when comparable mutational analysis was performed to assess the proposed AS1-SL3′/SL59–3′ inter-

action (Figure 3D-F). Thus, in addition to AS1/RS1, two other LDRIs, RTSL-TL/SL59–5′ and AS1-SL3′/SL59–3′, are critical for generating an effective RdRp attenuation structure. Since formation of the intra-LD2 interaction (pink) would free SL59–5′ for binding to RTSL-TL (green) (Supplementary Figure S4A), the probable order of these interactions would be the former followed by the latter.

Additional LD2 substructures are necessary for sg mRNA1 transcription

The organisation of LD2 includes two subdomains, LD2-sub1 and LD2-sub2, which have closing stems (S38, orange, and S56, blue, respectively) that are proximal to the sequences involved in the RTSL-TL/SL59–5′ and AS1-SL3′/SL59–3′ interactions (Figure 3G). These closing stems, which are maintained in the genus (Supplementary Figure S5), could therefore influence formation of the latter-mentioned interactions. To address this question, compensatory mutational analysis was performed on S38 and S56, which yielded results supporting the importance of their helical stability (Supplementary Figure S6). Accordingly, the closing stems of both LD2 subdomains also contribute to the assembly of an effective RdRp attenuation structure. This allowed for approximate delineation of a core region of functional importance at the base of LD2 (Figure 3G, black dashed line).

RTSL-TL and SL59–5′ interact physically by intra-genomic association

Having obtained *in vivo* genetic evidence for the RTSL-TL/SL59–5′ interaction, we next sought physical support for this pairing event. To achieve this, fragments of RTSL (106 nt) and LD2 (1040 nt) (Figure 4A) that contained the same compensatory mutations in RTSL-TL and SL59–5′ that were tested earlier (Figure 3A) were used in RNA-RNA electrophoretic mobility shift assays (EMSAs) (Figure 4B). Incubation of wt fragments of RTSL and LD2 led to the formation of an RNA–RNA complex, observed as an upward shift of the LD2 fragment (Figure 4B, compare lane 4 with 6). Combinations of fragments in which the RTSL-TL/SL59–5′ interaction was destabilized diminished shifting, while restoration of pairing regenerated the shift (Figure 4B, lanes 7 & 8 and lane 9, respectively). Thus, formation of an RTSL/LD2 complex is dependent on the RTSL-TL/SL59–5′ (green) interaction.

The demonstration of an *in trans* interaction *in vitro* raised the possibility that the same could be true during viral infections. To address this prospect, virus genome mutants TC-6 and TC-7 (Figure 3A), each of which was unable to form the RTSL-TL/SL59–5′ interaction *in cis*, but could potentially form it between each other *in trans*, were co-transfected into protoplasts. Levels of sg mRNA1 in the co-transfection (TC-6+TC-7) were similar to those for the individual transfections, i.e. ∼10%, and well below the ∼59% observed for the compensatory mutant TC-8 (Figure 4C, lanes 3–6). As a further test, a small noncoding CIRV genome-derived RNA replicon, DI8, containing a wt RTSL was utilized (Figure 4D, lower section). DI8 replicates only in the presence of the CIRV genome, which provides RdRp

for its reproduction (39). Thus, DI8 amplification is limited to cells occupied by both the replicon and the CIRV genome. Co-transfection of DI8 and TC-6 (containing a mutated RTSL-TL and a wt SL59–5′ compatible with the wt RTSL in DI8) resulted in high levels of accumulation of both viral RNAs (Figure 4C, lane 9). However, despite robust co-accumulation, no increase in sg mRNA1 levels was observed (Figure 4C, compare lane 9 with lane 3). Collectively, these results indicate that during CIRV infections, the RTSL-TL/SL59–5′ interaction occurs as an intra-genomic event.

RTSL-TL/SL59-5′ complex formation requires both AS1/RS1 and AS1-SL3′/SL59-3′ interactions

Three key interactions are involved in efficient formation of the RdRp attenuation structure, RTSL-TL/SL59–5′ (green), AS1-SL3′/SL59–3′ (pink) and AS1/RS1 (red). To investigate the order in which these binding events occur, additional RNA–RNA EMSAs were performed (Figure 5). When the AS1/RS1 interaction corresponding to the closing stem of LD2 (Figure 5A, red) was assessed via compensatory mutations (Figure 5B), the results indicated its requirement for RTSL-TL binding to SL59–5′ (Figure 5C, lanes 7 to 10). Interestingly, disruption of AS1/RS1 in the LD2 fragment led to a slight decrease in its mobility, suggesting a more open conformation, consistent with AS1/RS1′s role in stabilizing the basal region of this large RNA domain (Figure 5C, compare lanes 4 and 5 with lanes 3 and 6). A dependence on the AS1-SL3′/SL59–3′ (pink) interaction was also observed, as RTSL/LD2 complex formation was inhibited by the CU mismatch in mutant f10 (Figure 5D, E lane 9). In contrast, the AG mismatch in f9 allowed for complex formation (Figure 5D, E lane 8). In the CIRV genome, this modification led to strong inhibition of sg mRNA1 levels in protoplast transfections (Figure 3D–F). The differing results observed for the AG mismatch in the EMSA is likely the consequence of this common non-canonical base pair being less destabilizing under the higher salt conditions of the assay. Notwithstanding, inhibition of complex formation with the CU mismatch and its recovery with the AU pair indicates that the AS1-SL3′/SL59–3′ (pink) interaction is indeed required for RTSL-TL binding to SL59–5′ (Figure 5D, 5E). These results, when considered along with the requisite for partner sequence accessibility and proximity, support the following sequential order for the formation for the three critical interactions. The AS1/RS1 (red) interaction would occur first and position AS1-SL3′ proximal to SL59–3′. Next, formation of the AS1-SL3′/SL59–3′ (pink) interaction would concurrently liberate SL59–5′ (green). Lastly, pairing of SL59–5′ with RTSL-TL (green) would complete assembly of the attenuation structure.

Defining the structural requirements for the RTSL-TL/SL59-5′ interaction

Infections with CIRV genome mutants revealed that efficient activation of sg mRNA1 transcription required the basal region of LD2, bounded by AS1/RS1, S38 and S56 (Figure 3G, black dotted line). To determine if more distal
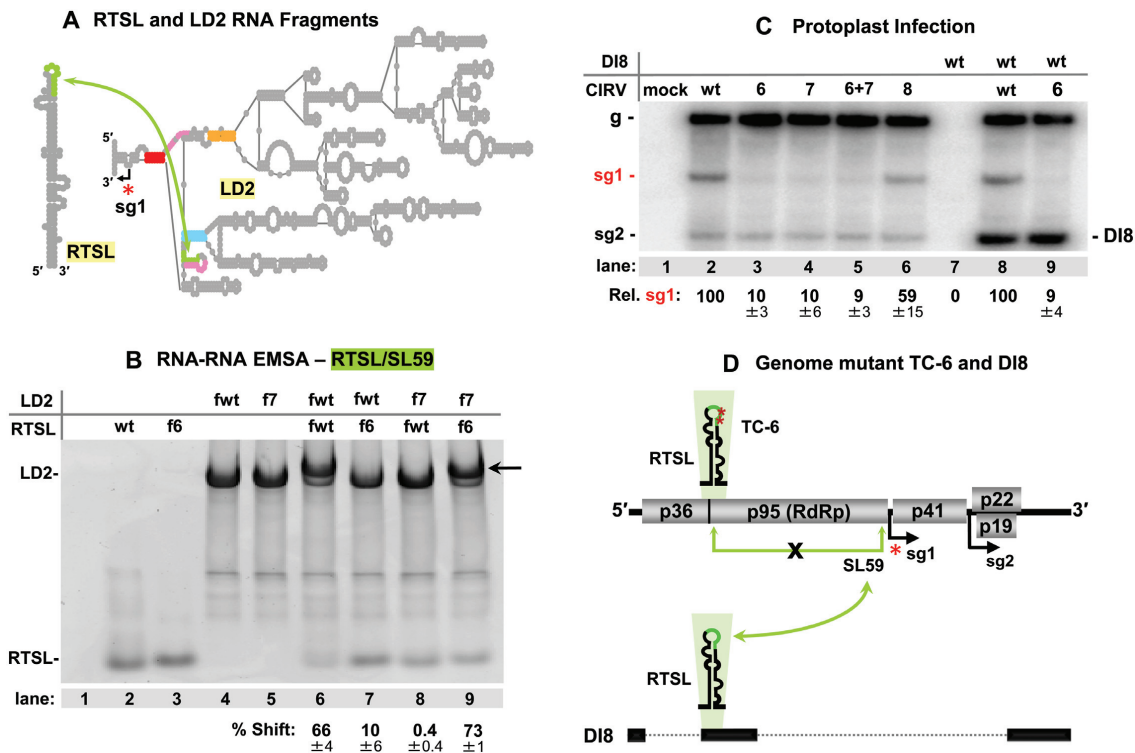
Figure 4. RTSL-TL and SL59–5′ sequences interact physically and via an intragenomic interaction. (**A**) Simplified depiction of CIRV RTSL (106 nt) and LD2 (1040 nt) RNA fragments tested in RNA–RNA EMSAs. (**B**) Ethidium bromide-stained native 4% polyacrylamide gel of EMSA with the RNA fragments containing mutations shown in Figure 3A. The contents of each lane are indicated above the gel, with the fragment type shown to the far left. Lane 1 represents a mock lane containing only RNA binding buffer and glycerol. The black arrow on the right side of the gel points to the position where the RTSL/LD2 complexes migrate. The percentages with standard errors of shifted LD2 RNAs are provided below and were obtained from three independent EMSA experiments. (**C**) Northern blot analysis of protoplast co-transfection and single transfections. Identities of the viral genomic and DI RNAs tested are indicated above the gel. Modifications in the RNAs are shown in Figure 3A. Positions of positive-sense genome, sg mRNAs, and DI8 are indicated on the left and right of the blot. Average sg mRNA1 accumulation levels relative to that of the wt are provided below the blot with standard errors obtained from three independent experiments. (**D**) Schematic depiction of the co-transfection involving mutant TC-6 CIRV genome (top) and wt DI8 (bottom, horizontal black bars correspond to regions of the viral genome present in DI8). TC-6 modifications (two burgundy asterisks) in RTSL prevents formation of an intragenomic RTSL-TL/SL59–5′ LDRI (green double-headed arrow with a black **X**). The DI8 RNA contains wt RTSL in its sequence that, when co-inoculated with TC-6, can potentially base pair *in trans* with the wt SL59–5′ sequence in TC-6 (the curved green arrow connects DI8's wt RTSL and TC-6's wt SL59).

sequences or structures in LD2-sub1 or LD2-sub2 were required for RTSL-TL binding, a 188 nt-long RNA fragment containing only the core region of LD2 was constructed (Figure 6A, LD2-core). In LD2-core, the subdomain sequences beyond S38 and S56 were replaced with ultra-stable UNCG-type tetraloops. When LD2-core and RTSL fragments containing compensatory mutations in the RTSL-TL/SL59–5′ interaction (Figure 3A) were tested by EMSA, the results were equivalent to those observed with the complete LD2 fragment (compare Figure 6B, lanes 6–9, with Figure 4B, lanes 6–9). LD2-core thus accurately recapitulated the binding activity of the full-length LD2, implying that all determinants for efficient RTSL binding are present in this smaller fragment. RTSL binding to LD2-core also exhibited equivalent binding activities compared to full-length LD2 in terms of dependence on the AS1/RS1 (Supplementary Figure S7B, lanes 7–10, with Figure 5C, lanes 7–10) and AS1-SL3′/SL59–3′ interactions (Supplementary Figure S7C, lanes 7–10, with Figure 5E, lanes 7–10). Thus,

LD2-core behaves comparably to full length LD2. Additionally, the potential involvement of SL60 in RTSL binding was assessed by deleting it from LD2-core and the results indicated no role for this substructure in complex formation (Supplementary Figure S8).

The portion of RTSL required for binding to LD2-core was also sought by generating fragments with increasingly larger truncations of its lower region (Figure 6C) The EMSA results revealed that the bottom half of RTSL, including the PRTE, was dispensable for binding (Figure 6D). Therefore, the portion of RTSL essential for translational readthrough (i.e. the PRTE) is not required for formation of the RTSL-TL/SL59–5′ interaction.

In-line probing analysis of LD2-core and RTSL reveals a second key binding site

With both genetic and physical evidence supporting the formation and function of the RTSL/LD2 interaction, we next
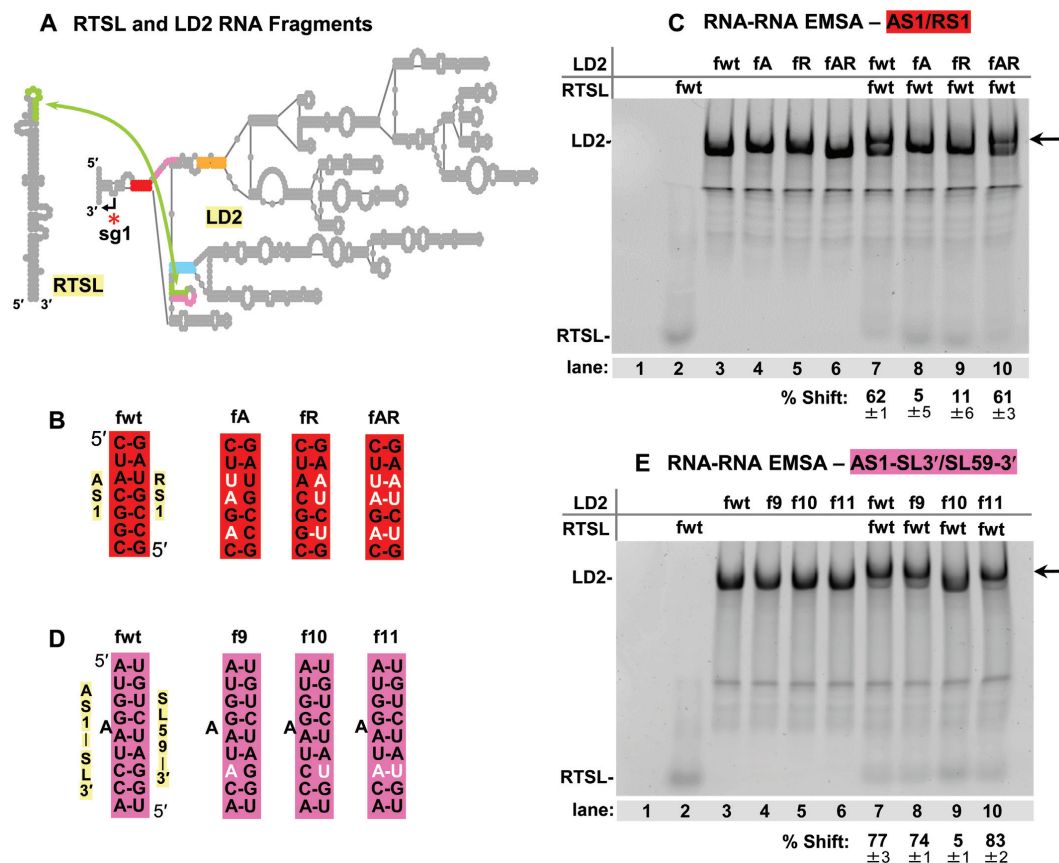
Figure 5. Formation of the RTSL/LD2 complex is dependent on both the AS1/RS1 and AS1-SL3′/SL59–3′ interactions. (**A**) Simplified depiction of CIRV RTSL (106 nt) and LD2 (1040 nt) RNA fragments tested in RNA-RNA EMSAs. (**B, D**) Compensatory nucleotide substitutions (in white) that were introduced, respectively, into the AS1/RS1 (red) and AS1-SL3′/SL59–3′ (pink) complementary partner sequences. (**C, E**) Ethidium bromide-stained native 4% polyacrylamide gels of EMSAs testing the RNA fragments containing mutations shown in panels (B) and (D), respectively. The contents of each lane are indicated above the gels, with the fragment type shown to the far left. Lane 1 represents a mock lane containing just the RNA binding buffer and glycerol. The black arrows on the right side of the gels point to the position where the RTSL/LD2 complexes migrate. The percentages with standard errors of shifted LD2 RNAs are provided below and were obtained from three independent EMSA experiments.

sought to gain additional insights into the nature of this RNA complex through solution structure probing analysis. To this end, in-line probing was used to assess the RNA structure of RTSL and LD2-core, both individually and in complex. Under the assay conditions, residues that are flexible, and thus likely single-stranded, undergo spontaneous hydrolysis (56). Information gained from the analysis is then used to build structural models consistent with the chemical reactivity data.

LD2-core was assessed first (Figure 7A). In its free state, the structural status of SL59–5′ (green), its adjacent partner sequence SL59–3′ (pink), and the alternate partner of the latter, AS1-SL3′ (pink), were of particular interest. The reactivity data (Figure 7A, lane 4) suggested that unbound LD2-core likely exists as a conformational mixture that includes SL59 (Figure 7Bi) and the mutually-exclusive AS1-SL3′/SL59–3′ (pink) interaction (Figure 7Bii). Probing results with free LD2-core that were consistent with the formation of SL59 included (i) high reactivity in the 5′-portion of AS1-SL3′ (pink, coordinates 25 to 30), indicating that

a proportion of this sequence does not pair with SL59–3′ (Figure 7A, lower black bar and 7Bi, brown-shaded triangles) and (ii) high reactivity in the loop residues in SL59, which would be reactive in the context of SL59 (Figure 7A, upper black bar and 7Bi, brown-shaded triangles). Further evidence for SL59′s functional relevance and structural existence was provided, respectively, by comparative sequence analysis supporting its conservation (Supplementary Figure S9) and RNA structure modelling, guided by the in-line reactivity data, that predicted its presence in the optimally-folded LD2-core (Supplementary Figure S10A). Conversely, the moderate reactivity of residues in SL59–5′ (green, 106–110), which indicated an unpaired state in a proportion of the structural population, was consistent with an alternative non-SL59-containing structure (Figure 7A, white bar and 7Bii, brown-shaded triangles); a concept bolstered by the prerequisite for the SL59–5′-freeing AS1-SL3′/SL59–3′ (pink) interaction for complex formation (Figure 5E and Supplementary Figure S7C). Collectively, these data suggest that, when unbound, the core re-
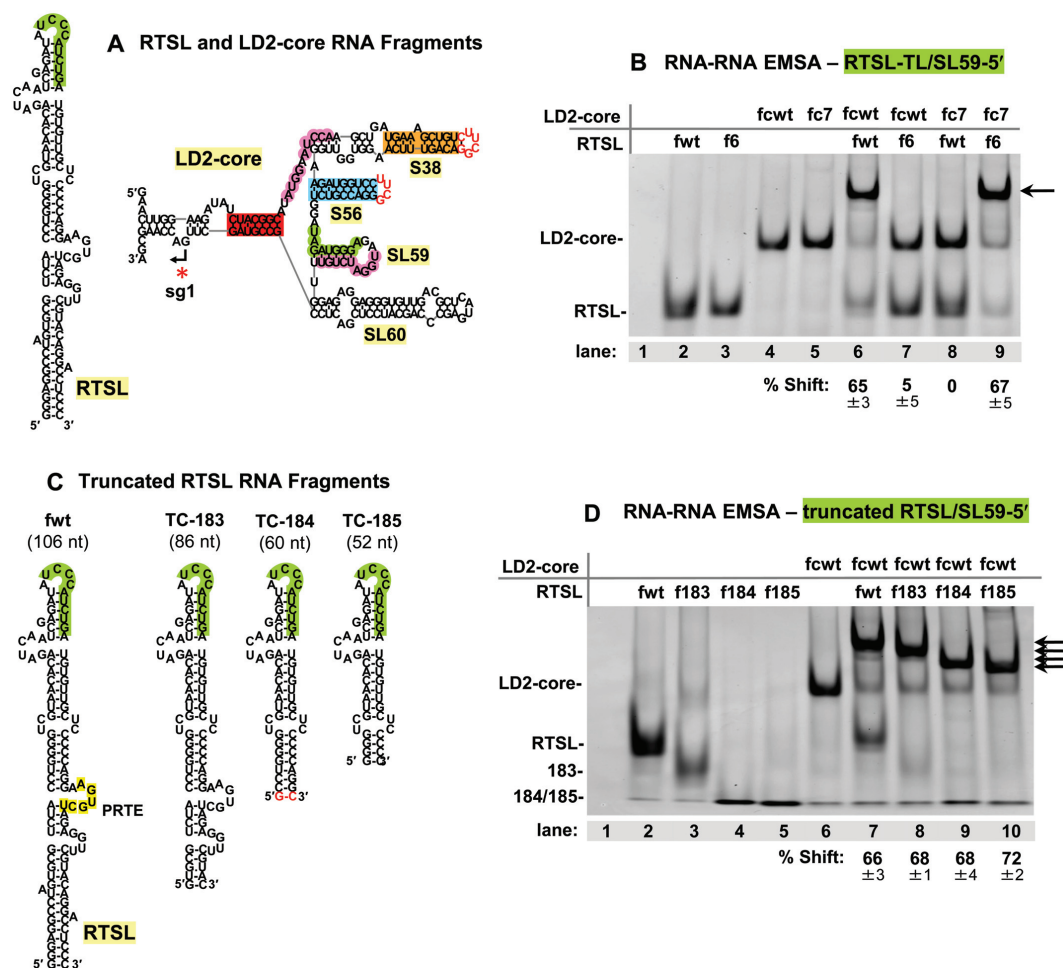
Figure 6. Structural requirements for RTSL/LD2 complex formation. (**A**, **C**) Secondary structures of CIRV RTSL (106 nt) and LD2-core (188 nt) RNA fragments tested in RNA-RNA EMSAs. The red nucleotides in the LD2-core secondary structure represent the added UNCG-type tetraloops that replaced LD2 sub1 and sub2 beyond S38 and S56, respectively. The G–C pair shown in red in RTSL mutant TC-184 was added to allow for its transcription from PCR templates using T7 RNA polymerase. (**B**, **D**) Ethidium bromide-stained 8% native polyacrylamide gels of EMSAs testing the RNA fragments containing modifications shown in Figure 3A and panel C, respectively. The contents of each lane are indicated above the gels, with the fragment type shown to the far left. Lane 1 represents a mock lane containing only RNA binding buffer and glycerol. The black arrows on the right side of the gels point to the positions where the RTSL/LD2 complexes migrate. The percentages with standard errors of shifted LD2-core RNAs are provided below the gels and were obtained from three independent EMSA experiments.

gion of LD2 is comprised of a mixture that includes the two structural conformations presented (Figure 7B), however other configurations are also plausible (Supplementary Figure S10B).

Probing results for LD2-core when in complex with RTSL revealed a notable reduction in reactivity of SL59–5′ (green), consistent with it base-pairing with RTSL-TL (Figure 7A, compare lanes 4 and 5). Correlative results were observed when RTSL was probed individually (Figure 7C, lane 4, 7D, and Supplementary Figure S11) or in complex, the latter of which showed a corresponding reduction in reactivity of RTSL-TL (green) in the bound state (Figure 7C, compare lanes 4 and 5). The probing results also revealed a potential second inter-fragment interaction involving two 5 nt-long complementary sequences (*i.e.* corresponding re-

duced reactivities in the bound states) (Figure 7A, C, purple) located between S38 and S56 in LD2-core and in a bulged region of RTSL (Figure 7B and D, purple, respectively). Thus, in addition to the RTSL-TL/SL59–5′ (green) interaction, a second interaction between RTSL and LD2 (purple) could also be functionally relevant, as structurally modeled (Figure 7E and Supplementary Figure S12).

The potential second interaction was initially assessed in protoplast infections with CIRV genomes containing compensatory mutations in the partner sequences. The results indicated that base pairing of these sequences was required for both sg mRNA1 plus- and minus-strand synthesis (Supplementary Figure S13A–C). EMSA analysis of the same mutations in the context of the LD2-core and RTSL fragments indicated that complex formation was depen-
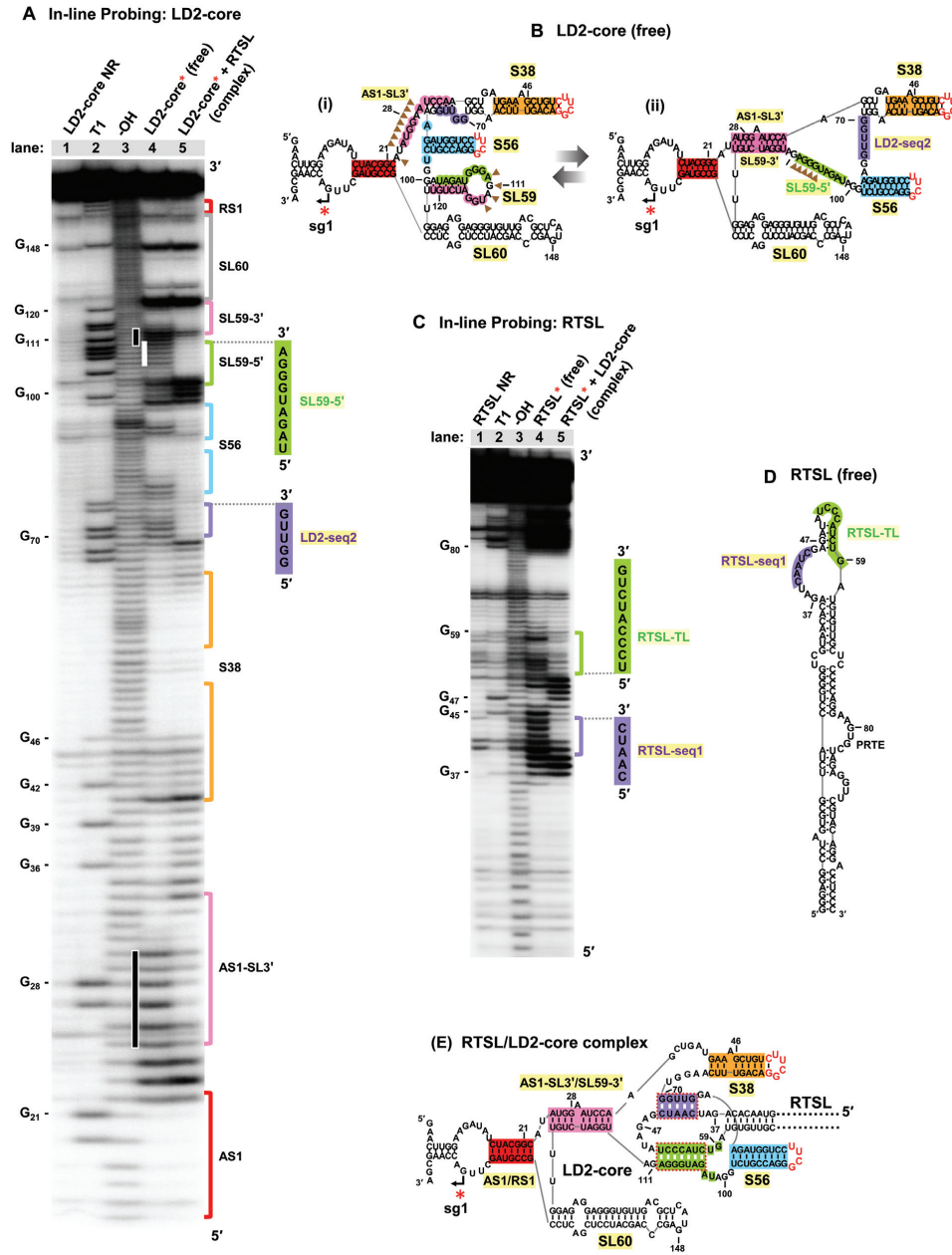
Figure 7. In-line structural probing analysis of CIRV LD2-core and RTSL RNAs. (**A**, **C**) Sequencing gel analysis following in-line probing of radiolabeled LD2-core (LD2-core*) and RTSL (RTSL*) RNA fragments, respectively. Lane 1 contains untreated LD2-core* or RTSL* RNA samples (NR, no reaction). Lane 2 contains the RNase T1-digested LD2-core* or RTSL* RNA samples to generate G ladders. Lane 3 contains LD2-core* or RTSL* RNA samples that were subjected to alkaline hydrolysis reaction (–OH) to generate cleavages at every nucleotide position. Lane 4 contains in-line reactions from free LD2-core* or free RTSL* RNA fragments (free). Lane 5 shows in-line reactions when LD2-core* or RTSL* was incubated with unlabeled RTSL and LD2-core, respectively, to generate a complex. Nucleotide positions of selected G residues are indicated on the left. Different regulatory sequences are color coded and labeled on the right side of the gels. Black bars on the left of lane 4 in panel (A) indicate SL59 and AS1-SL3′ sequences that show high cleavage levels in free LD2-core. The white bar on the left of lane 4 shows moderate cleavage levels for SL59–5′ in free LD2-core. (**B**) Two alternative RNA secondary structure conformations for free LD2-core. The structure on the left (i) was deduced as the optimal structure by in-line-guided folding of LD2-core, as described in Supplementary Figure S10A. Areas of notable reactivity are indicated by brown arrowheads (which correspond to vertical black bars in panel A). The structure on the right (ii) was generated with folding constraints that maintained nucleotides 106–110 as unpaired (brown arrowheads, which correspond to the vertical white bar in panel A). (**D**) RNA secondary structure of free RTSL was deduced as the optimal structure by in-line-guided folding of RTSL, as described in Supplementary Figure S11A. RTSL-TL and RTSL-seq1 are shown in green and purple, respectively. (**E**) RNA secondary structure of the RTSL/LD2-core complex, deduced by in-line probing results from analysis of the RNA complex (Supplementary Figure S12).

dent upon complementarity of the sequences in RTSL and LD2, termed RTSL-seq1 and LD2-seq2, respectively (Supplementary Figure S13D). Additionally, RTSL-seq1/LD2-seq2 pairing was found to be well conserved among the members of the genus Tombusvirus (Supplementary Figure S13E). These findings support a critical role for the RTSL-seq1/LD2-seq2 (purple) interaction in mediating formation of an effective attenuation structure for sg mRNA1 transcription (Figure 7E).

DISCUSSION

Global architecture of viral RNA genomes can contribute significantly to the regulation of critical viral functions. Accordingly, there is considerable interest in understanding how these genome-level RNA structures assemble and function. Our investigation of a tombusvirus led to the discovery of a novel intra-genomic RNA complex that activates sg mRNA1 transcription. Notably, this RNA-based attenuation structure is comparatively complex and provides new perspectives into this higher-order level of viral riboregulation.

The initial goal of this study was to investigate the possible role of the apical region of RTSL in the readthrough process, thus its observed involvement in sg mRNA1 transcription was unexpected. This additional function hinted at possible regulatory cross-talk between readthrough and transcription activities. Indeed, the presence of transcriptional regulatory sequences in the RdRp coding region of the viral genome would require the suppression of readthrough to allow for unimpeded transcription. Though an appealing possibility, *in vitro* translation analysis showed either no effect or minor decreases in readthough when the RTSL-TL/SL59–5′ interaction was disrupted (Figure 2B), whereas a notable increase (i.e. derepression) would have been expected if it was involved in coordinating the two processes. Nonetheless, the new transcriptional function uncovered adds to its previously known roles in promoting readthrough and inhibiting minus-strand RNA synthesis and classifies RTSL as a unique multifunctional RNA element controlling three distinct viral processes.

The complexity of the RdRp attenuation signal formed between RTSL and LD2 provided a unique opportunity to explore the assembly of this functional RNA complex. The comparatively smaller and localized components involved in the interaction, RTSL, AS1-SL and SL59, are anticipated to fold independently and relatively rapidly after their emergence during progeny viral RNA genome synthesis (Figure 8A). In contrast, formation of larger and more complex structures, such as subdomain-1 and -2 of LD2, would likely require additional time (Figure 8B). A role for these subdomains in assembly of the functional complex is supported by the observed importance of their closing stems for mediating efficient transcription (Supplementary Figure S6). Notably, the establishment of these subdomains unites AS1 and RS1 (red) to within ~80 nt, thereby markedly reducing their ~1000 nt distance of separation in the linear genome (compare Figure 8A with B). This colocalization would in turn facilitate base pairing of AS1 and RS1 (Figure 8B) and complete formation of LD2 (Figure 8C).

The AS1/RS1 (red) interaction also mediates formation of the core region of LD2 that ultimately forms the binding pocket for RTSL. Probing data suggests that this core region likely exists as a conformational ensemble that includes incompatible and compatible forms, with respect to RTSL binding (Figure 8D and E, respectively). The presence of SL59 precludes formation of the essential AS1-SL3′/SL59–3′ (pink) interaction (Figure 8D), while formation of the latter is needed to free SL59–5′ (green) and LD2-seq2 (purple) to allow their binding with partner sequences in RTSL (Figure 8E). SL59 thus represents an integral but transient component in the folding process. Functionally, the formation of SL59 could prevent its critical halves from interacting with non-cognate complementary sequences that would interfere with correct folding of the binding pocket. In this capacity, SL59 would provide a safe, temporary, storage form for its component sequences until their requirement for binding pocket formation, initiated by the AS1/RS1 interaction.

Formation of the RTSL binding pocket requires both global folding of the large RNA domain LD2 and detailed conformational arrangements within its basal core region. Key features of the resulting docking site includes two discontinuous sequences (LD2-seq2, purple and SL59–5′, green) that map to either side of S56, the closing stem of subdomain-2 (Figure 8F). The docking of RTSL, via bipartite binding of RTSL-seq1 and RTSL-TL with these sites, acts as a linchpin in the formation and stabilization of the higher-order RNA complex capable of blocking progression of the viral RdRp (Figure 8G). This final docking step could confer its effect by bolstering the AS1/RS1 (red) interaction by either direct or allosteric means. In the latter case, RTSL pairings could stabilize the adjacent AS1-SL3′/SL59–3′ (pink) interaction, which, in turn, could structurally support the juxtaposed AS1/RS1 (red) helix (Figure 8G). Alternatively, direct, presumably noncanonical, interactions between RTSL and AS1/RS1 could function to stabilize the latter. A third possibility is that an additional part(s) of the RNA complex, in addition to the AS1/RS1 helix, contacts the RdRp and contributes to the stalling activity. Future, higher-resolution structural analysis will be required to investigate further the precise mode of RNA-based inhibition of the RdRp.

The formation and stability of the RNA attenuation structure is highly cooperative, as verified by the strong inhibitory effects of nucleotide mismatches in any of its component interactions (Figure 3 and Supplementary Figures S3, S6, and S13). Although our analyses indicate that these interactions can occur spontaneously *in vitro* (Figure 4 and Figure 5), viral or host proteins could also assist in folding of the attenuation RNA complex during infections (e.g. RNA chaperones (64)). Assembly of the RNA complex follows a multistep folding pathway involving the spatial unification of numerous distant regions of the genome (Figure 8). In this folding scheme, two steps in particular are likely to be rate limiting, and thus determinants of the timing of active complex formation leading to sg mRNA1 transcription. The first is the generation of LD2, including formation of the binding pocket, which would depend on overall domain folding and subsequent refinement of the docking site. A second restrictive step would be the docking event,

61
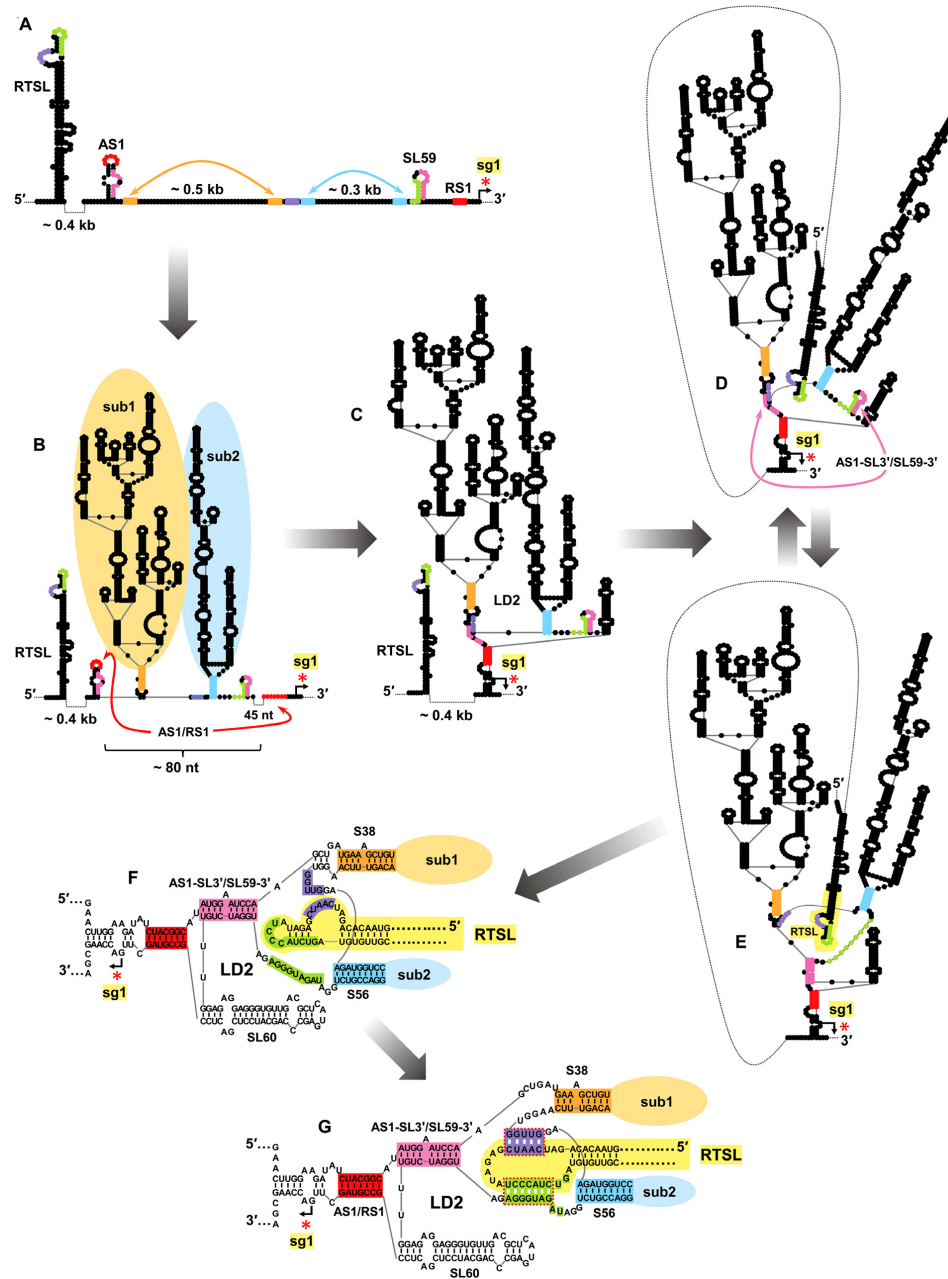
Figure 8. Proposed RNA genome folding pathway leading to activation of sg mRNA1 transcription in CIRV. Note, this is a highly simplified folding pathway based on conjectured temporally-distinct transitions dictated by differences in stability, complexity and the relative spatial positions of the sub-components. The relative timescales for transitions are not known and the structures shown represent approximations of probable intermediates within ensemble populations. (**A**) Schematic depiction of partially folded sections of the CIRV genome including RTSL, AS1-SL and SL59. AS1 and RS1 are separated by ∼1000 nucleotides in the linear sequence. Orange and blue double-headed arrows point to the complementary sequences involved in formation of S38 of sub1 and S56 of sub2, respectively. Small secondary structures within sub1 and sub2 are not shown, but are anticipated to form on a similar timescale as RTSL, AS1 and SL59. (**B**) RNA secondary structure after folding of LD2 sub1 (orange) and sub2 (blue), which are closed by stems S38 (orange) and S56 (blue), respectively. In this conformation AS1 and RS1 are brought within ∼80 nt from one another, which facilitates their base pairing (red double headed arrow). (**C**) Formation of the AS1/RS1 interaction completes folding of LD2 and its basal core region. (**D**) SL59, when present in the basal core region, prevents the docking of RTSL into the binding pocket by sequestering its partner sequences (refer also to Figure 7B(i)). Base pairing between AS1-SL3′ and SL59–3′ (pink double-headed arrow) leads to a conformational change in the basal core region. (**E**) Formation of the AS1-SL3′/SL59–3′ interaction (pink) frees the key binding pocket sequences (purple and green) and results in a functional binding site for RTSL. (**F**) Detailed pre-docking depiction of RTSL and the functional binding pocket in LD2. (**G**) Detailed post-docking depiction of RTSL and LD2 resulting in formation of an active attenuation structure.

the rate of which would be determined by the facility with which RTSL stochastically and productively encounters the binding pocket. Together, these steps could function to delay the production of CP from sg mRNA1 to later in the infection, when packaging is required. Indeed, time course experiments of viral RNA accumulation with tombusviruses show that, compared with that for sg mRNA2, sg mRNA1 transcription is delayed (42,43).

Intra-genomic LDRIs are also used to control sg mRNA transcription in other genera of the family *Tombusviridae*, including aureusviruses (65) and pelarspoviruses (66), while dianthoviruses utilize an inter-genomic interaction (67). However, in these cases the attenuation RNA structures are less complex than that described here; though, based on our unexpected results, further analyses may be warranted. This mode of sg mRNA regulation via LDRIs also extends beyond plant viruses to include plus-strand RNA viruses that infect insects and mammals. The sole sg mRNA produced by the insect-infecting Flock house virus (family *Nodaviridae*) is produced via a premature termination mechanism that utilizes an RNA-based attenuation structure composed of a three-helix junction formed by distant sequences (44). In contrast, coronaviruses use an alternate discontinuous transcription mechanism for sg mRNA production, where 3′ and then 5′ segments of the viral genome are copied discontinuously during minus-strand synthesis (68). In Transmissible gastroenteritis coronavirus, the discontinuous step for the production of the mRNA encoding the nucleocapsid protein is facilitated by an LDRI in the genome that unites the regions where the viral RdRp dissociates and reinitiates (69). Other plus-strand RNA viruses that infect humans and animals also depend on LDRIs for regulating viral processes, most notably flaviviruses (e.g. dengue virus (14,16,17) and zika virus (15)), hepaciviruses (e.g. hepatitis C virus (18,70)) and aphthoviruses (e.g. foot-and-mouth disease virus (71,72)). Moreover, other categories of RNA virus such are retroviruses (e.g. HIV (73)) and negative-strand RNA viruses (e.g. influenza virus (74)) also rely on LDRIs.

Understanding the molecular mechanisms of large-scale RNA circuits and their structural and functional integration is key to determining how RNA viruses regulate their infectious cycles. Deciphering such LDRI networks, however, has remained a challenge because many reside in coding regions and have multiple functions, as illustrated herein. In this study, we uncovered a new function for the folding of a large viral RNA domain in creating a distinctive binding pocket, and showed that subsequent docking of a distal RNA structure into this binding site acts as a linchpin that stabilizes an RNA complex required for viral transcription. We also proposed a plausible multistep pathway for the formation of the active intra-genomic RNA complex, an area of LDRI research that remains largely unexplored. These novel findings reinforce the importance and often overlooked underlying role of global RNA structure in viruses. Indeed, in many instances viral RNA genomes should be viewed as large complex RNA switches, and tombusviruses, with no fewer than eight functional LDRIs, serve as valuable prototypes for understanding this intriguing category of RNA-mediated regulation.

REFERENCES

1. Liu,Z.Y. and Qin,C.F. (2020) Structure and function of cis-acting RNA elements of flavivirus. *Rev. Med. Virol.*, **30**, e2092.
2. Madhugiri,R., Fricke,M., Marz,M. and Ziebuhr,J. (2016) Coronavirus cis-acting RNA elements. *Adv. Virus Res.*, **96**, 127–163.
3. Liu,Y., Wimmer,E. and Paul,A.V. (2009) Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim. Biophys. Acta.*, **1789**, 495–517.
4. Newburn,L.R. and White,K.A. (2015) Cis-acting RNA elements in positive-strand RNA plant virus genomes. *Virology*, **479–480**, 434–443.
5. Miller,W.A. and White,K.A. (2006) Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annu. Rev. Phytopathol.*, **44**, 447–467.
6. Nicholson,B.L. and White,K.A. (2014) Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nature Rev. Microbiol.*, **12**, 493–504.
7. Chkuaseli,T. and White,K.A. (2018) Intragenomic long-distance RNA-RNA interactions in plus-strand RNA plant viruses. *Front Microbiol.*, **9**, 529.
8. Guo,L., Allen,E.M. and Miller,W.A. (2001) Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. *Mol. Cell*, **7**, 1103–1109.
9. Barry,J.K. and Miller,W.A. (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.
10. Kuhlmann,M.M., Chattopadhyay,M., Stupina,V.A., Gao,F. and Simon,A.E. (2016) An RNA element that facilitates programmed ribosomal readthrough in turnip crinkle virus adopts multiple conformations. *J. Virol.*, **90**, 8575–8591.
11. Chattopadhyay,M., Shi,K., Yuan,X. and Simon,A.E. (2011) Long-distance kissing loop interactions between a 3′ proximal Y-shaped structure and apical loops of 5′ hairpins enhance translation of Saguaro cactus virus. *Virology*, **417**, 113–125.
12. Gao,F. and Simon,A.E. (2016) Multiple cis-acting elements modulate programmed -1 ribosomal frameshifting in Pea enation mosaic virus. *Nucleic Acids Res.*, **44**, 878–895.
13. Du,Z., Alekhina,O.M., Vassilenko,K.S. and Simon,A.E. (2017) Concerted action of two 3′ cap-independent translation enhancers increases the competitive strength of translated viral genomes. *Nucleic Acids Res.*, **45**, 9558–9572.
14. Sanford,T.J., Mears,H.V., Fajardo,T., Locker,N. and Sweeney,T.R. (2019) Circularization of flavivirus genomic RNA inhibits de novo translation initiation. *Nucleic Acids Res.*, **47**, 9789–9802.
15. Li,P., Wei,Y., Mei,M., Tang,L., Sun,L., Huang,W., Zhou,J., Zou,C., Zhang,S., Qin,C.F. *et al.* (2018) Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe*, **24**, 875–886.
16. de Borba,L., Villordo,S.M., Iglesias,N.G., Filomatori,C.V., Gebhard,L.G. and Gamarnik,A.V. (2015) Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J. Virol.*, **89**, 3430–3437.

17. Filomatori,C.V., Lodeiro,M.F., Alvarez,D.E., Samsa,M.M., Pietrasanta,L. and Gamarnik,A.V. (2006) A 5′ RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.*, **20**, 2238–2249.

18. Romero-López,C. and Berzal-Herranz,A. (2020) The role of the RNA-RNA interactome in the hepatitis C virus life cycle. *Int. J. Mol. Sci.*, **21**, 1479.

19. Rance,E., Tanner,J.E. and Alfieri,C. (2018) Genomic-scale interaction involving complementary sequences in the hepatitis C virus 5′UTR domain IIa and the RNA-dependent RNA polymerase coding region promotes efficient virus replication. *Viruses*, **11**, E17.

20. Romero-López,C., Barroso-Deljesus,A., García-Sacristán,A., Briones,C. and Berzal-Herranz,A. (2014) End-to-end crosstalk within the hepatitis C virus genome mediates the conformational switch of the 3′X-tail region. *Nucleic Acids Res.*, **42**, 567–582.

21. Romero-López,C. and Berzal-Herranz,A. (2012) The functional RNA domain 5BSL3.2 within the NS5B coding sequence influences hepatitis C virus IRES-mediated translation. *Cell. Mol. Life Sci.*, **69**, 103–113.

22. Romero-López,C. and Berzal-Herranz,A. (2009) A long-range RNA–RNA interaction between the 5′ and 3′ ends of the HCV genome. *RNA*, **15**, 1740–1752.

23. Tuplin,A., Struthers,M., Simmonds,P. and Evans,D.J. (2012) A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. *Nucleic Acids Res.*, **40**, 6908–6921.

24. Shetty,S., Stefanovic,S. and Mihailescu,M.R. (2013) Hepatitis C virus RNA: molecular switches mediated by long-range RNA–RNA interactions? *Nucleic Acids Res.*, **41**, 2526–2540.

25. Moreno,J.L., Zúñiga,S., Enjuanes,L. and Sola,I. (2008) Identification of a coronavirus transcription enhancer. *J. Virol.*, **82**, 3882–3893.

26. Mateos-Gómez,P.A., Zuñiga,S., Palacio,L., Enjuanes,L. and Sola,I. (2011) Gene N proximal and distal RNA motifs regulate coronavirus nucleocapsid mRNA transcription. *J. Virol.*, **85**, 8968–8980.

27. Mateos-Gomez,P.A., Morales,L., Zuñiga,S., Enjuanes,L. and Sola,I. (2013) Long distance RNA–RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J. Virol.*, **87**, 177–186.

28. White,K.A. and Nagy,P.D. (2004) Advances in the molecular biology of tombusviruses: gene expression, genome replication and recombination. *Prog. Nucleic Acids Res. Mol. Biol.*, **78**, 187–226.

29. Harrison,S.C., Olson,A.J., Schutt,C.E., Winkler,F.K. and Bricogne,G. (1978) Tomato bushy stunt virus at 2.9 A resolution. *Nature*, **276**, 368–373.

30. Hillman,B.I., Carrington,J.C. and Morris,T.J. (1987) A defective interfering RNA that contains a mosaic of a plant virus genome. *Cell*, **51**, 427–433.

31. Ye,K., Malinina,L. and Patel,D.J. (2003) Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature*, **426**, 874–878.

32. Jaag,H.M., Pogany,J. and Nagy,P.D. (2010) A host Ca^{2+}/Mn^{2+} ion pump is a factor in the emergence of viral RNA recombinants. *Cell Host Microbe*, **7**, 74–81.

33. Chuang,C., Prasanth,K.R. and Nagy,P.D. (2017) The glycolytic pyruvate kinase is recruited directly into the viral replicase complex to generate ATP for RNA synthesis. *Cell Host Microbe*, **22**, 639–652.

34. Nagy,P.D. (2016) Tombusvirus-host interactions: co-opted evolutionarily conserved host factors take center court. *Annu. Rev. Virol.*, **3**, 491–515.

35. Nicholson,B.L. and White,K.A. (2015) Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.*, **12**, 66–74.

36. Nicholson,B.L. and White,K.A. (2008) Context-influenced cap-independent translation of Tombusvirus mRNAs in vitro. *Virology*, **380**, 203–212.

37. Nicholson,B.L., Zaslaver,O., Mayberry,L.K., Browning,K.S. and White,K.A. (2013) Tombusvirus Y-shaped translational enhancer forms a complex with eIF4F and can be functionally replaced by heterologous translational enhancers. *J. Virol.*, **87**, 1872–1883.

38. Nicholson,B.L., Wu,B., Chevtchenko,I. and White,K.A. (2010) Tombusvirus recruitment of translational machinery via the 3′UTR. *RNA*, **16**, 1402–1419.

39. Cimino,P.A., Nicholson,B.L., Wu,B., Xu,W. and White,K.A. (2011) Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. *PLoS Pathog.*, **12**, e1002423.

40. Jiwan,S.D. and White,K.A. (2011) Subgenomic mRNA transcription in Tombusviridae. *RNA Biol.*, **8**, 287–294.

41. Wu,B. and White,K.A. (2007) Uncoupling RNA virus replication from transcription via the polymerase: functional and evolutionary insights. *EMBO J.*, **26**, 5120–5130.

42. Zhang,G., Slowinski,V. and White,K.A. (1999) Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA*, **5**, 550–561.

43. Qiu,W. and Scholthof,H.B. (2001) Effects of inactivation of the coat protein and movement genes of Tomato bushy stunt virus on early accumulation of genomic and subgenomic RNAs. *J. Gen. Virol.*, **82**, 3107–3114.

44. Lindenbach,B.D., Sgro,J.Y. and Ahlquist,P. (2002) Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. *J. Virol.*, **76**, 3905–3919.

45. van Vliet,A.L., Smits,S.L., Rottier,P.J. and de Groot,R.J. (2002) Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J.*, **21**, 6571–6580.

46. White,K.A. (2002) The premature termination model: a possible third mechanism for subgenomic mRNA transcription in (+)-strand RNA viruses. *Virology*, **304**, 147–154.

47. Lin,H.X. and White,K.A. (2004) A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. *EMBO J.*, **23**, 3365–3374.

48. Wu,B., Grigull,J., Ore,M.O., Morin,S. and White,K.A. (2013) Global organization of a positive strand RNA virus genome. *PLoS Pathog.*, **9**, e1003363.

49. Choi,I.R. and White,K.A. (2002) An RNA activator of subgenomic mRNA1 transcription in Tomato bushy stunt virus. *J. Biol. Chem.*, **277**, 3760–3766.

50. Rubino,L., Burgyan,J. and Russo,M. (1995) Molecular cloning and complete nucleotide sequence of carnation Italian ringspot tombusvirus genomic and defective interfering RNAs. *Arch. Virol.*, **140**, 2027–2203.

51. White,K.A. and Morris,T.J. (1994) Nonhomologous RNA recombination in tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions. *J. Virol.*, **68**, 14–24.

52. Choi,I.R., Ostrovsky,M., Zhang,G. and White,K.A. (2001) Regulatory activity of distal and core RNA elements in tombusvirus subgenomic mRNA2 transcription. *J. Biol. Chem.*, **276**, 41761–41768.

53. Fabian,M.R. and White,K.A. (2006) Analysis of a 3′-translation enhancer in a tombusvirus: a dynamic model for RNA-RNA interactions of mRNA termini. *RNA*, **12**, 1304–1314.

54. Lafuente,E., Ramos,R. and Martínez-Salas,E. (2002) Long-range RNA-RNA interactions between distal regions of the hepatitis C virus internal ribosome entry site element. *J. Gen. Virol.*, **83**, 1113–1121.

55. Petrov,A., Wu,T., Viani Puglisi,E. and Puglisi,J.D. (2013) RNA purification by preparative polyacrylamide gel electrophoresis in Laboratory methods in enzymology: RNA. *Methods Enzymol.*, **530**, 315–330.

56. Nahvi,A. and Green,R. (2013) Structural analysis of RNA backbone using in-line probing in laboratory methods in enzymology: RNA. *Methods Enzymol.*, **530**, 381–397.

57. Regulski,E.E. and Breaker,R.R. (2008) In-line probing analysis of riboswitches in Post transcriptional gene regulation. *Methods Mol. Biol.*, **419**, 53–67.

58. Slatko,B.E. and Albright,L.M. (1992) Denaturing gel electrophoresis for sequencing. *Curr. Protoc. Mol. Biol*, **16**, 7.6.1–7.6.13.

59. Johnson,P.Z., Kasprzak,W.P., Shapiro,B.A. and Simon,A.E. (2019) RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA Biol.*, **16**, 1667–1671.

60. Vasa,S.M., Guex,N., Wilkinson,K.A., Weeks,K.M. and Giddings,M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.

61. Bellaousov,S., Reuter,J.S., Seetin,M.G. and Mathews,D.H. (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, **41**, W471–W474.

62. Wu,B., Pogany,J., Na,H., Nicholson,B.L., Nagy,P.D. and White,K.A. (2009) A discontinuous RNA platform mediates RNA virus

replication: building an integrated model for RNA-based regulation of viral processes. *PLoS Pathog.*, **5**, e1000323.

63. Boerneke,M.A., Ehrhardt,J.E. and Weeks,K.M. (2019) Physical and functional analysis of viral RNA genomes by SHAPE. *Annu. Rev. Virol.*, **6**, 93–117.

64. Stork,J., Kovalev,N., Sasvari,Z. and Nagy,P.D. (2011) RNA chaperone activity of the tombusviral p33 replication protein facilitates initiation of RNA synthesis by the viral RdRp in vitro. *Virology*, **409**, 338–347.

65. Xu,W. and White,K.A. (2008) Subgenomic mRNA transcription in an aureusvirus: down-regulation of transcription and evolution of regulatory RNA elements. *Virology*, **371**, 430–438.

66. Blanco-Pérez,M. and Hernández,C. (2016) Evidence supporting a premature termination mechanism for subgenomic RNA transcription in Pelargonium line pattern virus: identification of a critical long-range RNA-RNA interaction and functional variants through mutagenesis. *J. Gen. Virol.*, **97**, 1469–1480.

67. Sit,T.L., Vaewhongs,A.A. and Lommel,S.A. (1998) RNA-mediated trans-activation of transcription from a viral RNA. *Science*, **281**, 829–832.

68. Sola,I., Almazán,F., Zúñiga,S. and Enjuanes,S. (2015) Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol.*, **2**, 265–288.

69. Mateos-Gomez,P.A., Morales,L., Zuñiga,S., Enjuanes,L. and Sola,I. (2013) Long-distance RNA-RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J. Virol.*, **87**, 177–186.

70. Mauger,D.M., Golden,M., Yamane,D., Williford,S., Lemon,S.M., Martin,D.P. and Weeks,K.M. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3692–3697.

71. Serrano,P., Pulido,M.R., Sáiz,M. and 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.

72. Diaz-Toledano,R., Lozano,G. and Martinez-Salas,E. (2017) In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. *Nucleic Acids Res.*, **45**, 1416–1432.

73. Ooms,M., Abbink,T.E., Pham,C. and Berkhout,B. (2007) Circularization of the HIV-1 RNA genome. *Nucleic Acids Res.*, **35**, 5253–5261.

74. Tomescu,A.I., Robb,N.C., Hengrung,N., Fodor,E. and Kapanidis,A.N. (2014) Single-molecule FRET reveals a corkscrew RNA structure for the polymerase- bound influenza virus promoter. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3335–E3342.

SUPPLEMENTARY FIGURES (Figures S1 – S13)

Activation of Viral Transcription by Stepwise Largescale Folding
of an RNA Virus Genome

Tamari Chkuaseli and K. Andrew White

Department of Biology, York University, Toronto, Ontario, Canada M3J 1P3
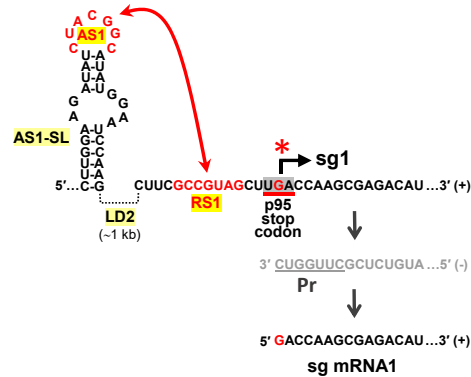
Figure S1. **Detailed depiction of the AS1 and RS1 RNA elements in CIRV that activate sg mRNA1 transcription.** AS1 and RS1 are shown in red, with AS1 positioned in the terminal loop of AS1-SL. The p95 stop codon (red underline) overlaps with the initiating nucleotide for sg mRNA1 transcription (red G). The truncated minus-strand RNA intermediate synthesized during sg mRNA1 transcription is shown in grey nucleotides, with the promoter (Pr) sequence underlined. The minus-strand intermediate is generated when the RdRp encounters the AS1/RS1 stem in the genome, which causes it to terminate synthesis. The 3′-truncated minus-strand generated then serves as a template for transcription of sg mRNA1.
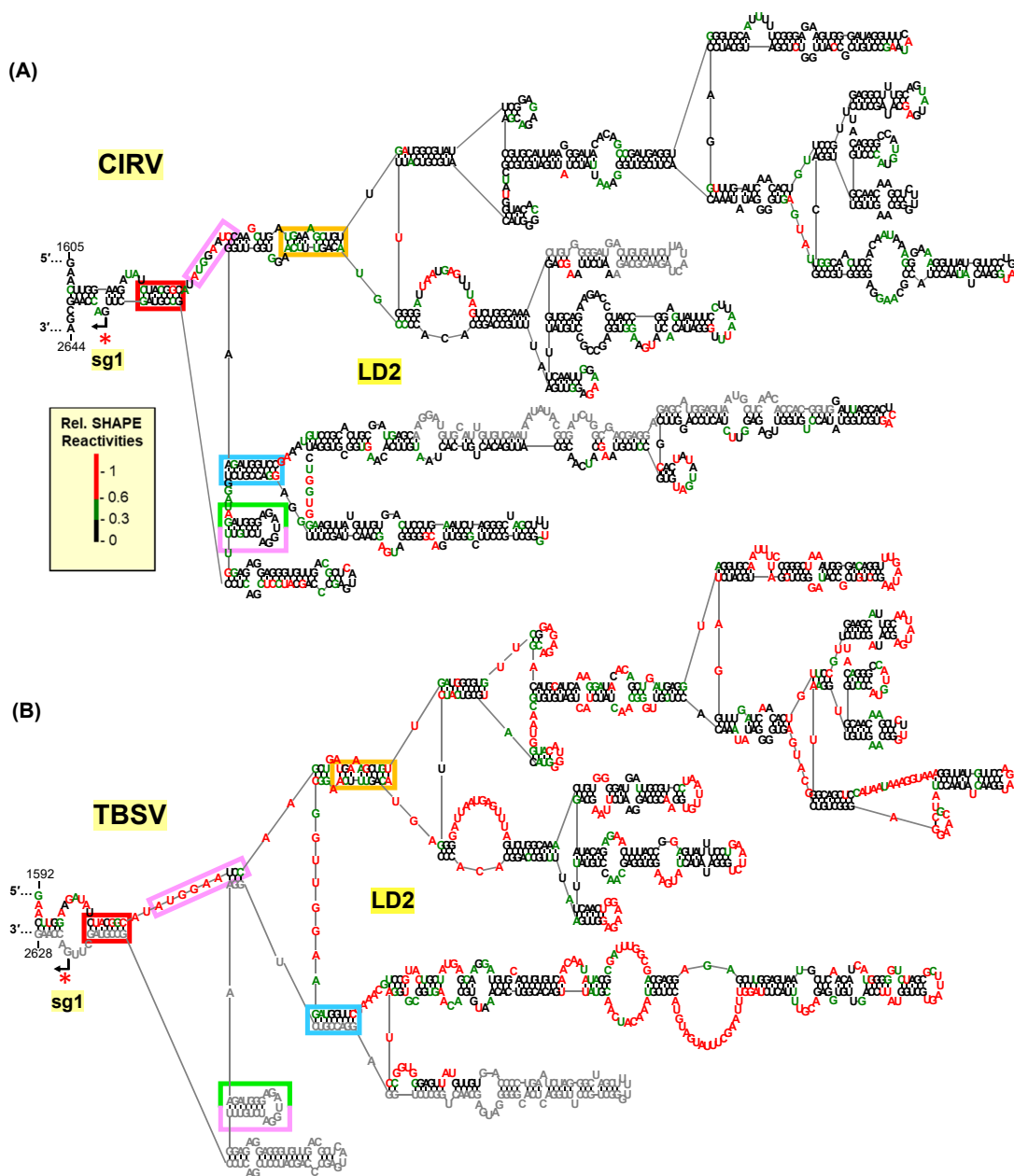
Figure S2. RNA secondary structure of LD2 in the CIRV genome deduced through selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). **(A)** SHAPE analysis was performed on the full-length CIRV RNA genome. The averages of normalized SHAPE reactivities from two independent SHAPE experiments were used as folding constraints in *RNAstructure* (slope value = 1.8 kcal/mol and intercept value = -0.6 kcal/mol) to deduce an RNA secondary structure for LD2. Nucleotides coloured in red were highly reactive, in green were moderately reactive, and in black were weakly reactive or unreactive. Nucleotides in grey correspond to regions for which no SHAPE data was obtained. **(B)** For comparison, SHAPE-guided secondary structure of LD2 in TBSV (adapted from 48). Key corresponding structural features are colour-coded (see text for details).
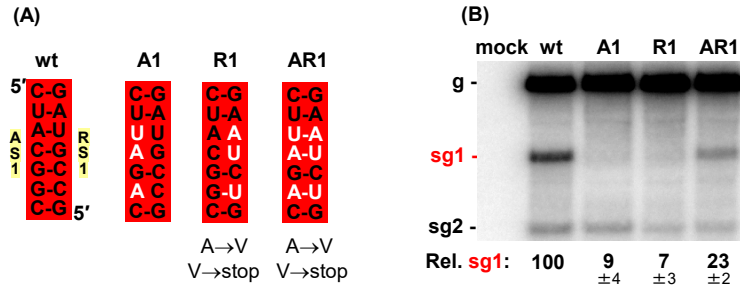
Figure S3. **The AS1/RS1 LDRI regulates sg mRNA1 transcription in CIRV. (A)** Compensatory mutations introduced in AS1 and RS1 in CIRV are shown in white. Amino acid changes in p95 are indicated under the mutant sequences. **(B)** Northern blot analysis of plus-strand CIRV RNAs extracted from protoplasts transfected with wt and mutant CIRV genomic RNAs shown in panel (A). Identities of the tested genomes are indicated above the blot, and positions of CIRV genome (g) and subgenomic mRNAs (sg1 and sg2) are shown on the left. Average plus-strand sg mRNA1 accumulation levels relative to that of the wt are provided below the blot with standard errors obtained from three independent experiments.

Figure S4

Figure S4. Structural depiction and comparative sequence analysis of the AS1/RS1, AS1-SL3'/SL59-3', and RTSL-TL/SL59-5' interactions. (A) CIRV secondary structure showing formation of the intra-LD2 AS1-SL3'/SL59-3' (pink) interaction. **(B)** Comparative sequence analysis of AS1/RS1 (red), AS1-SL3'/SL59-3' (pink) and RTSL-TL/SL59-5' (green) interactions between the members of the Tombusvirus genus, and Zeavirus genus (*i.e.* MNeSV, the most closely related genus to tombusviruses). Nucleotide substitutions that maintain base pairing are in white, while those that do not preserve pairing are in red. The asterisks below correspond to the nucleotides that are 100% conserved. The AS1-SL3'/SL59-3' (pink) interaction for CBLV does not conform to the pairing scheme observed in the other viruses, and when the CBLV genome was analyzed by *mFold* an alternative base pairing scheme was predicted (boxed sequences at bottom of table). Tombusviruses: Carnation Italian ringspot virus (CIRV, NC_003500.3), Lisianthus necrosis virus (LNV, DQ011234.1), Pear latent virus (PLV, AY100482.1), Tomato bushy stunt virus cherry isolate (TBSV-Ch, M21958.1), TBSV pepper isolate (TBSV-P, U80935.1), Artichoke mottled crinkle virus (AMCV, NC_001339.1), TBSV nipplefruit isolate (TBSV-Nf, AY579432), TBSV statice isolate (TBSV-St, AJ249740.1), Grapevine Algerian latent virus (GALV, NC_011535.1), Pelargonium necrotic spot virus (PNSV, NC_005285.1), Cucumber necrosis virus (CuNV, NC_001469.1), Cymbidium ringspot virus (CymRSV, NC_003532.1), Cucumber Bulgarian latent virus (CBLV, NC_004725.1), Lettuce necrotic stunt virus isolate L2 (LNCV-L2, JN700748.1), Moroccan pepper virus isolate PM75 (MPV-PM75, NC_020073.2), Eggplant mottled crinkle virus (EMCV, NC_023339.1), CIRV isolate CZ (KP888563.1), Pelargonium leaf curl virus isolate T46 (PLCV-T46, NC_030452.1). Zeavirus: Maize necrotic streak virus (MNeSV, NC_007729.1).

Figure S5. Comparative sequence analysis of S38 and S56 in the Tombusvirus and Zeavirus genera. Nucleotide substitutions that maintain base pairing are depicted in white, while those that do not preserve pairing are in red. The asterisks below correspond to the nucleotides that are 100% conserved among the analysed viral sequences. Tombusviruses: Carnation Italian ringspot virus (CIRV, NC_003500.3), Lisianthus necrosis virus (LNV, DQ011234.1), Pear latent virus (PLV, AY100482.1), Tomato bushy stunt virus cherry isolate (TBSV-Ch, M21958.1), TBSV pepper isolate (TBSV-P, U80935.1), Artichoke mottled crinkle virus (AMCV, NC_001339.1), TBSV nipplefruit isolate (TBSV-Nf, AY579432), TBSV statice isolate (TBSV-St, AJ249740.1), Grapevine Algerian latent virus (GALV, NC_011535.1), Pelargonium necrotic spot virus (PNSV, NC_005285.1), Cucumber necrosis virus (CuNV, NC_001469.1), Cymbidium ringspot virus (CymRSV, NC_003532.1), Cucumber Bulgarian latent virus (CBLV, NC_004725.1), Lettuce necrotic stunt virus isolate L2 (LNCV-L2, JN700748.1), Moroccan pepper virus isolate PM75 (MPV-PM75, NC_020073.2), Eggplant mottled crinkle virus (EMCV, NC_023339.1), CIRV isolate CZ (KP888563.1), Pelargonium leaf curl virus isolate T46 (PLCV-T46, NC_030452.1). Zeavirus: Maize necrotic streak virus (MNeSV, NC_007729.1).

72

Figure S6. Functional analysis of S38 and S56. (A) Secondary structure of RTSL-TL and LD2 regions in the CIRV genome. Key structures are colour-coded. **(B)**, **(E)** Compensatory substitutions that were introduced in the full-length CIRV genome to functionally assess S38 (orange) and S56 (blue), respectively. Amino acid changes in the p95 ORF are indicated under each mutant. **(C)**, **(F)** Northern blot analysis of plus-strand RNAs isolated from protoplasts transfected with CIRV wt and mutant genomic RNAs shown in (B) and (E), respectively. Identities of the tested samples are indicated above the blots and positions of positive-sense genome and sg mRNAs are shown on the left. Average sg mRNA1 accumulation levels relative to that of the wt are provided below the blots with standard errors obtained from three independent experiments. **(D)**, **(G)** Northern blot analysis of minus-strand CIRV RNAs isolated from protoplast infections. Average minus-strand sg mRNA1 accumulation levels relative to that of the wt are provided below the blots with standard errors obtained from three independent experiments. For mutant 12 in panel D and mutant 16 in panel G, relative sg RNA minus-strand levels were higher than their plus-strand counterparts. This is likely due to inaccurate RdRp termination (caused by a mutated attenuation structure) during (-)sgRNA synthesis, which results in a promoter (Pr, see supplemental Fig. S1) with either missing or added 3′-terminal nucleotides that inhibits its use as transcriptional promoter.

73
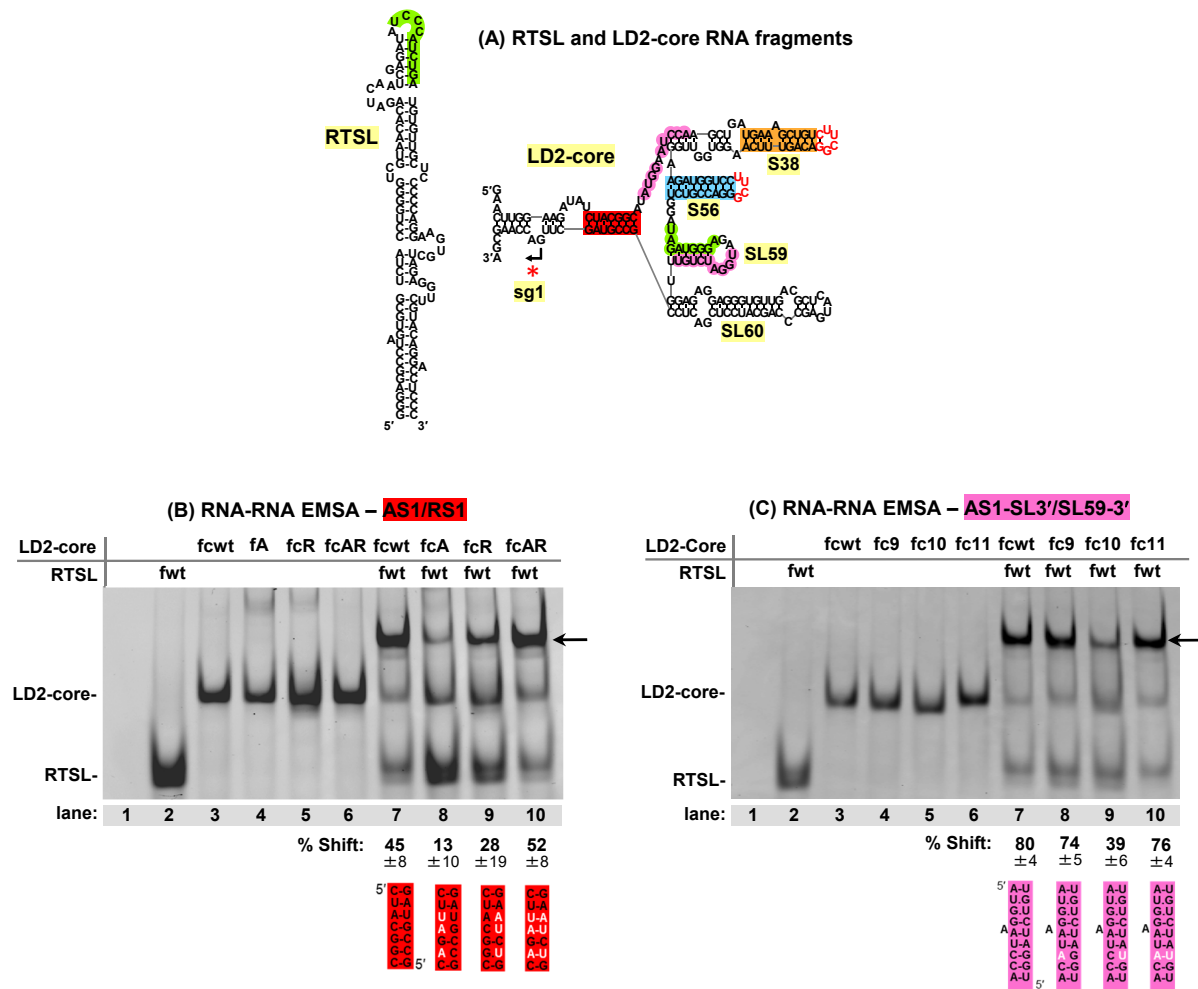
Figure S7. RTSL/LD2-core complex formation relies on both the AS1/RS1 and AS1-SL3'/SL59-3' interactions. (A) Secondary structures of RTSL (106 nt) and LD2-core (188 nt) RNA fragments tested in RNA-RNA EMSAs. RTSL-TL (green), AS1/RS1 (red), AS1-SL3' (pink), S38 (orange), S56 (blue), SL59 (green and pink), and LS60 are indicated. Sg mRNA1 initiation site is depicted with a small black arrow and a red asterisk. **(B)**, **(C)** RNA-RNA EMSA results for RTSL and LD2-core RNA fragments shown in panel (A) with the modifications indicated below the gels. RNAs were separated in native 8% polyacrylamide gels, which were stained with ethidium bromide. The contents of each lane are indicated above the gels with the fragment type shown to the far left. Lane 1 contains only RNA binding buffer and glycerol. The black arrows on the right side of the images point to where the RTSL/LD2-core complexes migrate. The percentages and standard errors of shifted LD2-core RNAs compared to the corresponding non-shifted LD2-core RNAs are displayed below the gels and were obtained from three independent EMSA experiments.
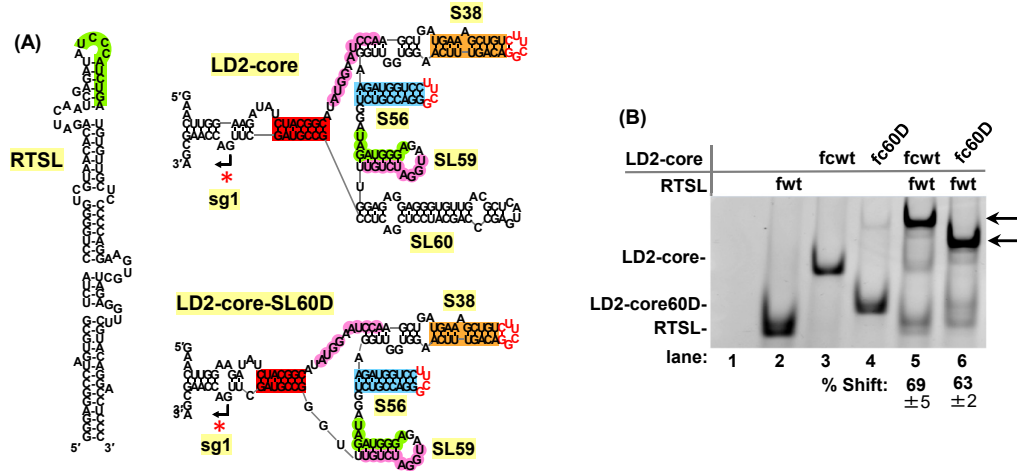
Figure S8. SL60 is not required for RTSL/LD2-core complex formation. (A) RNA secondary structures of RTSL (106 nt), LD2-core (188 nt), and LD2-core-SL60D (145 nt, with SL60 deleted) fragments tested by RNA-RNA EMSA. **(B)** RNA-RNA EMSA results for the RNA fragments shown in (A). The contents of each lane are indicated above the native 8% polyacrylamide gel stained with ethidium bromide, with the fragment type shown on the far left. Lane 1 represents a mock lane containing only RNA binding buffer and glycerol. The black arrows on the right side of the image point to where the RTSL/LD2-core and RTSL/LD2-core-SL60D complexes migrate. The percentages and standard errors of shifted LD2-core RNAs compared to the corresponding non-shifted LD2-core RNAs are displayed below the gels and were obtained from three independent EMSA experiments.

Figure S9. Comparative sequence analysis of SL59 in the Tombusvirus and Zeavirus genera. Boxed nucleotides represent complementary sequences that form the stem of SL59. The green and pink nucleotides depict SL59-5' and AS1-SL3' sequences that are complementary to RTSL-TL and AS1-SL3' sequences, respectively. Nucleotide substitutions that maintain base pairing are in white, while those that do not preserve pairing are in red. The asterisks below correspond to the nucleotides that are 100% conserved. Tombusviruses: Carnation Italian ringspot virus (CIRV, NC_003500.3), Lisianthus necrosis virus (LNV, DQ011234.1), Pear latent virus (PLV, AY100482.1), Tomato bushy stunt virus cherry isolate (TBSV-Ch, M21958.1), TBSV pepper isolate (TBSV-P, U80935.1), Artichoke mottled crinkle virus (AMCV, NC_001339.1), TBSV nipplefruit isolate (TBSV-Nf, AY579432), TBSV statice isolate (TBSV-St, AJ249740.1), Grapevine Algerian latent virus (GALV, NC_011535.1), Pelargonium necrotic spot virus (PNSV, NC_005285.1), Cucumber necrosis virus (CuNV, NC_001469.1), Cymbidium ringspot virus (CymRSV, NC_003532.1), Cucumber Bulgarian latent virus (CBLV, NC_004725.1), Lettuce necrotic stunt virus isolate L2 (LNCV-L2, JN700748.1), Moroccan pepper virus isolate PM75 (MPV-PM75, NC_020073.2), Eggplant mottled crinkle virus (EMCV, NC_023339.1), CIRV isolate CZ (KP888563.1), Pelargonium leaf curl virus isolate T46 (PLCV-T46, NC_030452.1). Zeavirus: Maize necrotic streak virus (MNeSV, NC_007729.1).

76

Figure S10. **RNA secondary structure for the free CIRV LD2-core deduced from in-line probing. (A)** LD2-core secondary structure with nucleotides coloured according to their relative in-line cleavage values, with red being frequently cleaved, green, cleaved at intermediate frequency, and black, infrequently cleaved. Nucleotides in grey represent the areas were relative in-line values could not be determined. The optimal secondary structure was obtained by inputting the relative in-line probing values into the *RNAstructure* web server to generate an in-line-guided structural model for LD2-core. The quantification of the in-line probing data was obtained by determining the intensity of each band in the lane containing free LD2-core RNA (Figure 7A, lane 4) using the QuantityOne (Biorad) software, with background level subtracted. The value for each band was normalized relative to the average value for top ten highest values. The normalized values were then used as folding constraints in *RNAstructure* web server with slope value set to 1.8 kcal/mol and intercept value set to -0.6 kcal/mol. **(B)** Suboptimal secondary structure predictions of LD2-core, within 10% of the optimal structure, using the in-line constraints as described above. **(C)** A table showing nucleotide positions and their corresponding normalized in-line reactivity values. Values for nucleotides 1-5, 91-96, as well as, 137-188 could not be quantified. The relative in-line values are colour-coded according to the scale shown in panel (A).

(C) LD2-core (free)

Nucleotide Position	Relative Self-cleavage Levels
1	
⋮	
6	0.040040646
7	0.03515452
8	0.045114822
9	0.058427861
10	0.392942868
11	7.609050702
12	0.316103884
13	1.322595886
14	0.057581272
15	0.0978398
16	0.030831967
17	0.078571861
18	0.015884511
19	0.026641631
20	0.028344097
21	0.067648945
22	0.608071655
23	1.256700119
24	1.28850328
25	0.952240121
26	1.250727204
27	1.282791756
28	0.755331877
29	0.39196196
30	0.601232828
31	0.120207651
32	0.052005275
33	0.06856054
34	0.144956913
35	0.163907352
36	0.069895001
37	0.044261757
38	0.068727743
39	0.059825054
40	0.131817132
41	0.500915033
42	0.119450034
43	0.122159585
44	0.135530505
45	0.265524221
46	0.126033583
47	0.023026281
48	0.018163751
49	0.008363368
50	0.007977678
51	0.008150851
52	0.011071323
53	0.034372167
54	0.027644828
55	0.024532527
56	0.008492562
57	0.010522237
58	0.021023791
59	0.017607648
60	0.01775595
61	0.018715794
62	0.028279951
63	0.052278867
64	0.072441533
65	0.112018445
66	0.090650356
67	0.114595522
68	0.21537867
69	0.29762994
70	0.473021657
71	0.486586091
72	0.458656859
73	0.320926827
74	0.278096687
75	0.233592283
76	0.276041193
77	0.331253323
78	0.345518189
79	0.151356442
80	0.144029298
81	0.073911609
82	0.060491121
83	0.065710792
84	0.068131652
85	0.061910728
86	0.066709512
87	0.104327843
88	0.1715945
89	0.36423537
90	0.398896246
⋮	
97	0.109499439
98	0.239826449
99	0.695350753
100	0.495289192
101	0.508191516
102	0.726523823
103	0.463966552
104	0.259090554
105	0.285351337
106	0.337929429
107	0.350208813
108	0.418854725
109	0.496571309
110	0.51071755
111	0.756571111
112	0.851467171
113	0.866534537
114	0.481567292
115	0.339172218
116	0.235600272
117	0.18348433
118	0.200666561
119	0.300627662
120	0.544912081
121	0.910791629
122	1.80236667
123	2.496116778
124	0.516867829
125	0.17459244
126	0.183243381
127	0.376783691
128	0.254370437
129	0.100060882
130	0.051306745
131	0.036506825
132	0.118011955
133	0.171852158
134	0.304033532
135	0.938825274
136	1.726774857
⋮	
188	

(A) Rel. In-line Self-cleavage

RTSL

△G = -73.1 kcal/mol

(B)

△G = -66.5 kcal/mol

(C) RTSL (free)

Nucleotide Position	Relative Self-cleavage Levels
1	
:	
12	0.032174798
13	0.048571818
14	0.092572552
15	0.265122695
16	0.242744463
17	0.204024133
18	0.24169773
19	0.220472326
20	0.063513145
21	0.078613681
22	0.052566625
23	0.142214006
24	0.093279343
25	0.081848904
26	0.182420835
27	0.070370122
28	0.085911526
29	0.067887161
30	0.073947621
31	0.047083458
32	0.041017752
33	0.039472058
34	0.036411826
35	0.074207204
36	0.22650011
37	0.530208673
38	1.833890429
39	1.998503264
40	1.607942131
41	0.967800335
42	0.770818043
43	0.938028716
44	0.103485769
45	1.035845176
46	0.362990164
47	0.212275249
48	0.28265575
49	0.098620577
50	0.397987724
51	0.489659982
52	0.44700461
53	0.503803801
54	0.540372313
55	0.343621719
56	0.218472353
57	0.252114337
58	1.071830807
59	0.253703607
60	0.237210128
61	0.180344905
62	0.14576307
63	0.140598662
64	0.14876417
65	0.139807313
66	0.094979323
67	0.149709286
68	0.192952415
69	0.11782957
70	0.082312072
71	0.055672344
72	0.04215052
73	0.041954368
74	0.059783675
75	0.067494974
76	0.081995281
77	0.168993896
:	
106	

Figure S11. **RNA secondary structure for the free CIRV RTSL determined from in-line probing analysis. (A)** RTSL secondary structure with nucleotides coloured according to their relative in-line cleavage values, with red being frequently cleaved, green, cleaved at intermediate frequency, and black, infrequently cleaved. Nucleotides in grey represent the areas were relative in-line values could not be determined. The secondary structure was determined by inputting the relative in-line probing values into the *RNAstructure* web server to generate an in-line-guided structural model for RTSL. To obtain the relative in-line probing data the intensity of each band in the lane containing free RTSL RNA (Figure 7C, lane 4) was quantified using the QuantityOne (Biorad) software, with background level subtracted. The value for each band was normalized relative to the average value for top ten highest values. The normalized values were then used as folding constraints in *RNAstructure* web server with slope value set to 1.8 kcal/mol and intercept value set to -0.6 kcal/mol. **(B)** A suboptimal secondary structure of RTSL, within 10% of the optimal structure, predicted through using in-line constraints as described above. **(C)** A table showing nucleotide positions and their corresponding normalized in-line reactivity values. Values for nucleotides 1-11 and 78-106 could not be quantified. The relative in-line values are colour-coded according to the scale shown in panel (A).
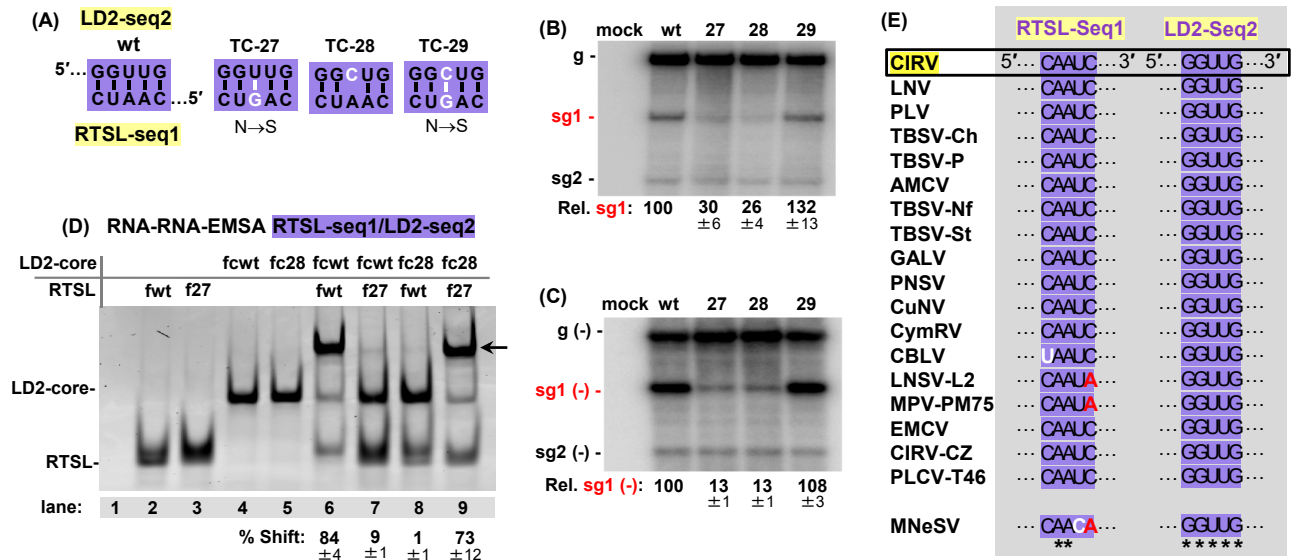
(A)

Rel. In-line Self-cleavage

- 1
- 0.6
- 0.3
- 0
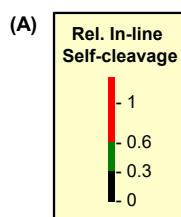
Figure S12. **Secondary structure analysis of the RTSL/LD2-core complex**. **(A)** Secondary structure of the RTSL/LD2-core complex with nucleotides coloured according to their relative in-line cleavage values, with red being frequently cleaved, green, cleaved at intermediate frequency, and black, infrequently cleaved. Nucleotides in grey represent the areas were relative in-line values could not be determined. The relative in-line probing data was obtained by quantifying the intensity of each band in the lane containing complexed RTSL and LD2-core RNAs (Figure 7C, lane 5; Figure 7A, lane 5) using the QuantityOne (Biorad) software, with background level subtracted. The values were normalized in the same manner as described in captions for Figure S9 and S10. The normalized self-cleavage values for RTSL and LD2-core were then mapped onto the RTSL/LD2-core complex, with the *trans*-interacting sequences in the RTSL/LD2-core complex deduced from the relative reactivities in the two RNAs. **(B), (C)** Tables showing nucleotide positions and their corresponding normalized in-line self-cleavage values of RTSL and LD2-core, respectively when complexed. Values for nucleotides 1-11 and 78-106 of RTSL and 1-5, 91-96, as well as, 137-188 of LD2-core could not be quantified. The relative in-line values are colour-coded according to the scale shown in panel (A).

(B) RTSL (complexed)

Nucleotide Position	Relative Self-cleavage Levels
1	
⋮	
12	0.023553
13	0.028482
14	0.093097
15	0.286619
16	0.242127
17	0.216130
18	0.239524
19	0.199685
20	0.043011
21	0.061902
22	0.051894
23	0.156932
24	0.089603
25	0.082613
26	0.198506
27	0.058691
28	0.106066
29	0.066709
30	0.073593
31	0.036710
32	0.027699
33	0.017375
34	0.018316
35	0.054870
36	0.147941
37	0.900962
38	1.974990
39	0.338849
40	0.329944
41	0.205065
42	0.196099
43	0.165729
44	0.241558
45	0.252262
46	0.399798
47	0.586928
48	9.851030
49	1.872102
50	1.245696
51	0.608206
52	0.251464
53	0.204472
54	0.364220
55	0.300678
56	0.178039
57	0.177791
58	0.250911
59	0.232509
60	0.145921
61	0.126760
62	0.128564
63	0.141068
64	0.153013
65	0.144609
66	0.107221
67	0.170729
68	0.167898
69	0.096529
70	0.050942
71	0.024704
72	0.014345
73	0.015775
74	0.033291
75	0.045561
76	0.072410
77	0.195089
⋮	
106	

(C) LD2-core (complexed)

Nucleotide Position	Relative Self-cleavage Levels
1	
⋮	
6	0.001473
7	0.00388
8	0.035427
9	0.076936
10	0.246933
11	0.650302
12	0.711707
13	0.931633
14	0.213634
15	0.198475
16	0.005693
17	0.054576
18	0.004939
19	0.009842
20	0.007937
21	0.03121
22	0.629676
23	1.890612
24	0.77078
25	0.291076
26	0.335198
27	0.132877
28	0.141725
29	0.183578
30	0.243749
31	0.058285
32	0.037203
33	0.120904
34	0.444934
35	0.360546
36	0.221611
37	0.163055
38	0.194632
39	0.108563
40	0.221114
41	0.804529
42	0.179884
43	0.199504
44	0.16693
45	0.305328
46	0.126038
47	0.00978
48	0.004443
49	0.002372
50	0.001134
51	2.79E-06
52	0.002467
53	0.017115
54	0.011677
55	0.008754
56	0.000249
57	0.000159
58	0.005881
59	0.002068
60	0.000952
61	0.004309
62	0.022678
63	0.06389
64	0.126041
65	0.205804
66	0.141211
67	0.170532
68	0.308481
69	0.632720
70	0.123865
71	0.076743
72	0.068669
73	0.080093
74	0.08131
75	0.106665
76	0.103481
77	0.091952
78	0.076336
79	0.052306
80	0.057136
81	0.043733
82	0.038206
83	0.040138
84	0.044485
85	0.042235
86	0.042943
87	0.072407
88	0.125225
89	0.168186
90	0.128331
⋮	
97	0.058693
98	0.189289
99	0.850742
100	1.15013
101	1.055292
102	1.578908
103	0.912252
104	0.145481
105	0.087883
106	0.081512
107	0.08773
108	0.09309
109	0.116757
110	0.157805
111	0.224963
112	0.409056
113	0.493022
114	0.269072
115	0.178872
116	0.089739
117	0.083439
118	0.109507
119	0.170251
120	0.341581
121	0.723714
122	2.08539
123	2.309699
124	0.443910
125	0.128304
126	0.146346
127	0.391248
128	0.264826
129	0.067765
130	0.002121
131	0.009574
132	0.038115
133	0.098375
134	0.273933
135	0.968608
136	1.852198
⋮	
188	

Figure S13. Functional analysis of the proposed second RTSL/LD2 interaction involving RTSL-seq1 and LD2-seq2. (A) Compensatory mutations introduced into the full-length CIRV genome in the proposed RTSL-seq1/LD2-seq2 LDRI (purple) are highlighted in white. Amino acid changes in the p95 ORF are indicated under each mutant. **(B)** Northern blot analysis of plus-strand RNAs isolated from protoplasts transfected with wt and mutant CIRV genomic RNAs shown in panel (A). Identities of the tested samples are indicated above the blot with the positions of positive-sense genome and sg mRNAs shown on the left. Average sg mRNA1 accumulation levels relative to that of the wt are provided below the blot with standard errors obtained from three independent experiments. **(C)** Northern blot analysis of minus-strand CIRV RNAs isolated from protoplasts transfected with wt and mutant CIRV genomic RNAs shown in (A). Identities of the tested samples are indicated above the blot and the positions of the minus-sense genome and sg mRNAs are shown on the left. Average minus-strand sg mRNA1 accumulation levels relative to that of the wt are provided below the blot with standard errors obtained from three independent experiments. **(D)** RNA-RNA EMSA results for the RTSL (106 nt) and LD2-core (188 nt) RNA fragments containing substitutions shown in (A). The contents of each lane are indicated above the native 8% polyacrylamide gel stained with ethidium bromide, with the fragment type shown on the far left. Lane 1 represents a mock lane containing only RNA binding buffer and glycerol. The black arrow on the right side of the image points to the position at which RTSL/LD2-core complex migrates. The percentages and standard errors of shifted LD2-core RNAs compared to the corresponding not-shifted LD2 RNAs are displayed below the gels obtained from three independent EMSA experiments. **(E)** Comparative sequence analysis of RTSL-seq1 and LD2-seq2 in the members of the Tombusvirus and Zeavirus genera. Red nucleotides represent substitutions that disrupt RTSL-seq1/LD2-seq2 interaction and the white nucleotide represents the substitution that maintains the interaction. Asterisks depict nucleotides that are 100% conserved among the compared viral sequences. See legend for Figure S4 for full names and accession numbers of the viral species.

CHAPTER 3

DIMERIZATION OF AN UMBRAVIRUS RNA GENOME ACTIVATES SUBGENOMIC mRNA TRANSCRIPTION

Umbravirus PEMV2 transcribes a sg mRNA that encodes two viral proteins necessary for cell-to-cell and systemic movement of the infection (Gao and Simon, 2017). However, the mechanism of sg mRNA transcription and RNA elements involved in its regulation have not been explored. This chapter provides a detailed analysis on PEMV2 sg mRNA transcription regulation which demonstrates that PEMV2 transcribes its sg mRNA via a premature termination mechanism and formation of a functional attenuation structure requires viral genome dimerization.

This chapter is presented as a submitted manuscript - "Dimerization of an umbravirus RNA genome activates subgenomic mRNA transcription", by Tamari Chkuaseli and K. Andrew White, accepted for publication by *Nucleic Acids Research* on June 1st, 2023. I conceptualized and designed the experiments for the study together with Dr. K. Andrew White. I performed all experiments and data analyses and wrote the first draft of the manuscript.

Dimerization of an Umbravirus RNA Genome Activates
Subgenomic mRNA Transcription

Tamari Chkuaseli and K. Andrew White*

Department of Biology, York University, Toronto, Ontario, Canada M3J 1P3

Running title: viral RNA genome dimerization activates transcription

Key words: RNA dimerization, RNA virus, RNA structure, RNA-RNA interaction, transcription, subgenomic RNA, plant virus, HIV, Tombusviridae

*Corresponding author:
K. Andrew White
Department of Biology
York University
4700 Keele St.
Toronto, Ontario
Canada M3J 1P3
Ph: (416) 736-2100 (ext. 40890 or 70352)
Fax: (416) 736-5698
email: kawhite@yorku.ca

ABSTRACT

Many eukaryotic RNA viruses transcribe subgenomic (sg) mRNAs during infections to control expression of a subset of viral genes. Such transcriptional events are commonly regulated by local or long-range intragenomic interactions that form higher-order RNA structures within these viral genomes. In contrast, here we report that an umbravirus activates sg mRNA transcription via base pair-mediated dimerization of its plus-strand RNA genome. Compelling in vivo and *in vitro* evidence demonstrate that this viral genome dimerizes via a kissing-loop interaction involving an RNA stem-loop structure located just upstream from its transcriptional initiation site. Both specific and non-specific features of the palindromic kissing-loop complex were found to contribute to transcriptional activation. Structural and mechanistic aspects of the process in umbraviruses are discussed and compared with genome dimerization events in other RNA viruses. Notably, probable dimer-promoting RNA stem-loop structures were also identified in a diverse group of umbra-like viruses, suggesting broader utilization of this unconventional transcriptional strategy.

INTRODUCTION

The majority of eukaryote-infecting viruses possess plus-strand RNA genomes (1). These single-stranded, coding-sensed genomes serve as mRNAs for viral protein translation upon entering host cells, however they can differ from eukaryotic mRNAs in multiple aspects, one of which is being polycistronic. Due to this coding strategy, the 5'-proximal open reading frame (ORF) is efficiently translated by eukaryotic translation machinery, while downstream ORFs are inaccessible and translationally silent. Despite this challenge, plus-strand RNA viruses successfully exploit eukaryotic translation machinery by employing alternative gene expression strategies, one of which is to transcribe subgenomic (sg) mRNAs (2-7).

Sg mRNAs are shorter viral messages that include 3'-portions of a viral genome and are transcribed during infections by the virally encoded RNA-dependent RNA polymerase (RdRp). This results in repositioning of downstream ORFs in polycistronic genomes to the 5'-end of shorter sg mRNAs, making them accessible for translation. This transcriptional strategy also provides a mechanism to temporally and quantitatively regulate viral protein production. Sg mRNAs are transcribed by different mammalian viruses, such as members of Coronaviridae (8), Togaviridae (9), and Nodaviridae (10) families, and is a common gene expression approach among plus-strand RNA viruses of plants, including Bromoviridae (11), Solemoviridae (12), and Tombusviridae (5,13) families.

Members of the family Tombusviridae (collectively termed tombusvirids) are classified into 18 different genera of plant viruses, all of which produce sg mRNAs during infections (13). Umbravirus is a unique genus in this family as all of its members lack a gene for a capsid protein (14,15). Umbraviruses instead rely on coinfections with assistor viruses from the genus Polerovirus or Enamovirus (both in the family Solemoviridae) to obtain capsid proteins for packaging their genomes (14,15). Pea enation mosaic virus 2 (PEMV2) is an umbravirus that *trans*-encapsidates its genome using capsid proteins produced by its assistor enamovirus PEMV1 (16,17). PEMV1 in turn relies on PEMV2-encoded proteins for cell-to-cell and systemic

spread (18). Because of this co-dependence, PEMV1 and PEMV2 are always found together in natural infections, even though they each encode their own RdRp and are capable of independent genome replication (16,18).

The umbravirus PEMV2 contains a 4.3 kb-long, uncapped, non-polyadenylated, plus-strand RNA genome that encodes four ORFs (**Figure 1A**) (16). As it is devoid of a 5'-cap and a 3'-poly(A) tail, protein translation from the genome and sg mRNA is initiated through the activities of 3' cap-independent translation enhancers (3'CITEs) located in the 3'UTR (19-24). The 5'-proximal ORF is translated into p33 (16), and a -1 ribosomal frameshifting event near the 3'-end of this ORF directs translation of a fusion protein, p94, the RdRp (**Figure 1A**) (16,25). Downstream ORFs for p26 and p27 overlap in different reading frames (16). P26 mediates formation of viral ribonucleoprotein particles, protects viral RNAs from nonsense mediated decay (NMD), and mediates systemic movement of infections through phloem (26-31), while p27 is a cell-to-cell movement protein (32-35). Both of these ORFs are translated from a sg mRNA that is transcribed during PEMV2 infections (24,36).

The mechanism by which PEMV2 sg mRNA is transcribed is not known, however the majority of tombusvirids transcribe their sg mRNAs using a premature termination (PT) mechanism involving two major steps (3,5) (**Figure 1A**). In step 1, while synthesizing a minus-strand of the genome, the viral RdRp terminates prematurely when it encounters a higher-order RNA structure, termed attenuation structure, located in the plus-strand genome. The attenuation structure acts, at least in part, as a physical barrier that causes the RdRp to stall and terminate, releasing a sg mRNA-sized minus-strand RNA, the 3'-terminus of which contains a promoter sequence (**Figure 1A, pr**). In step 2, the promoter sequence is recognized by the viral RdRp, and a plus-strand sg mRNA is transcribed (**Figure 1A**). Success of step 1 depends on the formation of an active attenuation structure, and the majority of these in tombusvirids are generated via intragenomic RNA-RNA interactions that form local structures (37-39) or span long distances (40-46). However, one tombusvirid, red clover necrotic mosaic virus (RCNMV;

85
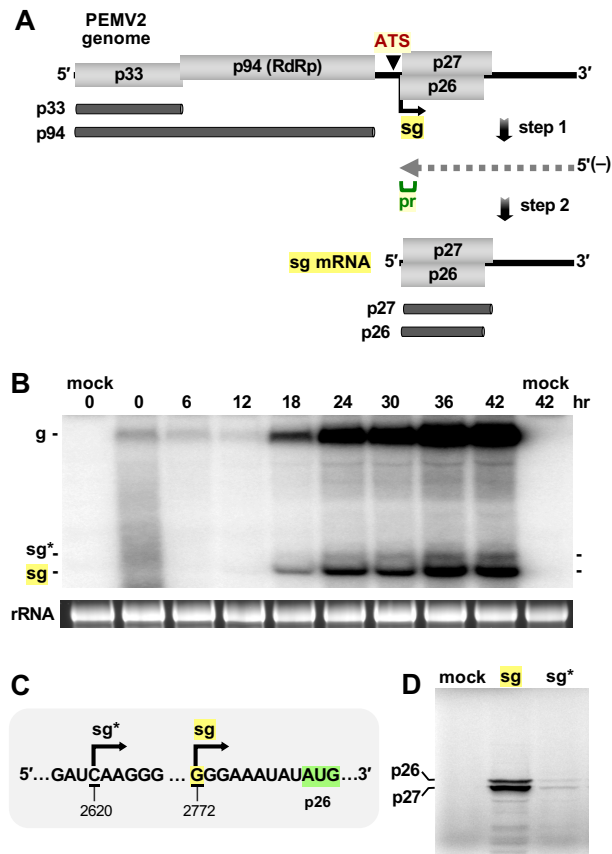
Figure 1: PEMV2 genome organization, viral RNA accumulation and p26/27 expression. (A) Schematic representation of the PEMV2 genome and subgenomic (sg) mRNA. Open reading frames encoding viral proteins are shown as grey boxes. The position of the sg mRNA transcription start site is indicated with a black arrow beneath the genome. Major steps in the premature termination mechanism of sg mRNA transcription are shown. The minus-strand intermediate is depicted as a dotted line (grey) with the promoter (pr, green) at the 3'-end. The relative position of the attenuation structure (ATS) is shown. Translation products of the genome and sg mRNA are depicted below as dark grey bars. **(B)** Northern blot analysis of PEMV2 viral RNA accumulation in infected cucumber protoplasts over the course of 42 hours. Total nucleic acids were extracted post-inoculation at the hours (hr) indicated. The identities of viral RNA bands are indicated to the left of the blot; genome (g), sg mRNA (sg), and novel subgenomic RNA (sg*). Ethidium bromide-stained rRNA are loading controls. **(C)** PEMV2 genomic coordinates for the 5'-termini of sg* and sg mRNA (underlined). The start codon for p26 is highlighted in green. **(D)** *In vitro* translation analysis of uncapped sg mRNA and sg* transcripts in wheat germ extract. The positions of ^{35}S-methionine-labeled viral proteins p26 and p27 are indicated on the left. Note that p26 migrates more slowly than p27 (24).

genus Dianthovirus), is different from other tombusvirids in that it contains a bi-segmented

genome consisting of RNA1 and RNA2 (47), and sg mRNA transcription from RNA1 is mediated

by formation of a *trans*-attenuation structure that is generated by base pair-mediated

heterodimerization of RNA1 with RNA2 (48).

In the present study we demonstrate that PEMV2 sg mRNA is transcribed via a PT

mechanism and, unexpectedly, utilizes a *trans*-attenuation structure formed through genome

homodimerization involving a palindrome-mediated kissing-loop interaction. RNA secondary

structures and sequences involved in PEMV2 genome dimerization and sg mRNA transcription

activation were investigated in detail and compared with existing and proposed examples of

viral RNA genome dimerization.

RESULTS

Accumulation profile of PEMV2 RNAs

PEMV2 infections were monitored over a 42-hour period to determine the timing and

relative levels of accumulation of viral RNAs. Total nucleic acids were extracted from protoplasts

transfected with uncapped *in vitro* transcripts of the viral genome at six-hour intervals and

assessed by northern blotting (**Figure 1B**). Residual PEMV2 transcripts from the transfection

were present between 0 and 12 hours post inoculation, while from hour 18 on, both genome and

sg mRNA levels increased (**Figure 1B**). Unexpectedly, another, previously unreported, less

abundant subviral RNA, designated as sg*, was also observed (**Figure 1B**). The 5'-terminus of

sg* was mapped by 5'RACE to nucleotide C_{2620} (PEMV2 genome coordinate), making it 152

nucleotides longer than the primary sg mRNA (**Figure 1C**). *In vitro* translation analysis in wheat

germ extract showed that, compared to the primary sg mRNA, the sg* message produced much

lower levels of p26 and p27 and did not generate any additional products (**Figure 1D**). This

finding argues against a role for sg* in markedly boosting p26 and 27 production or serving as a

message for a unique protein; the latter being the case for another umbravirus, opium poppy mosaic virus (61). Additionally, the nucleotides at the 5'-terminus of sg* (5'-CAA) were not consistent with the promoter sequence present at the 5'-end of typical umbravirus sg mRNAs (5'-GGG) that are transcribed by the viral RdRp (24) (**Figure 1C**), leaving the origin of sg* an open question. The mode of production of sg* and its possible function will be addressed in future investigations, while the current study examines the mechanism by which the primary sg mRNA is transcribed.

Uncoupling the steps in sg mRNA transcription supports a PT mechanism

Based on PEMV2 being a tombusvirid, we anticipated that its primary sg mRNA would be produced by a premature termination (PT) mechanism (3,5). This two-step mechanism involves viral RdRp-mediated synthesis of a minus-strand sg mRNA that is then used as a template for transcription of a plus-strand sg mRNA (**Figure 2A**). Previous studies have shown that minus-strand synthesis can be uncoupled from subsequent plus-strand transcription by introducing mutations in the sg mRNA promoter (37-39,42-45). To investigate this possibility for PEMV2, the first nucleotide of the sg mRNA promoter sequence was substituted in genome mutant m1 (**Figure 2B**) and its effect examined in protoplast infections. Relative sg mRNA accumulation was calculated as the ratio of a sg mRNA's level to its cognate genome's level, with the wt ratio set to 100%. The modification in m1 reduced plus-strand sg mRNA accumulation to ~27% of wt (**Figure 2C, left panel**), while corresponding minus-strands were ~74% of wt (**Figure 2D, left panel**). More compelling uncoupling was observed when the fourth nucleotide in the promoter was substituted in m2 (**Figure 2C, D, right panels**). These observations show that minus-strand sg RNA synthesis (step 1) can occur independently of plus-strand sg mRNA production (step 2) (**Figure 2A**), a finding that supports transcription of the PEMV2 sg mRNA via a PT mechanism (3,5).
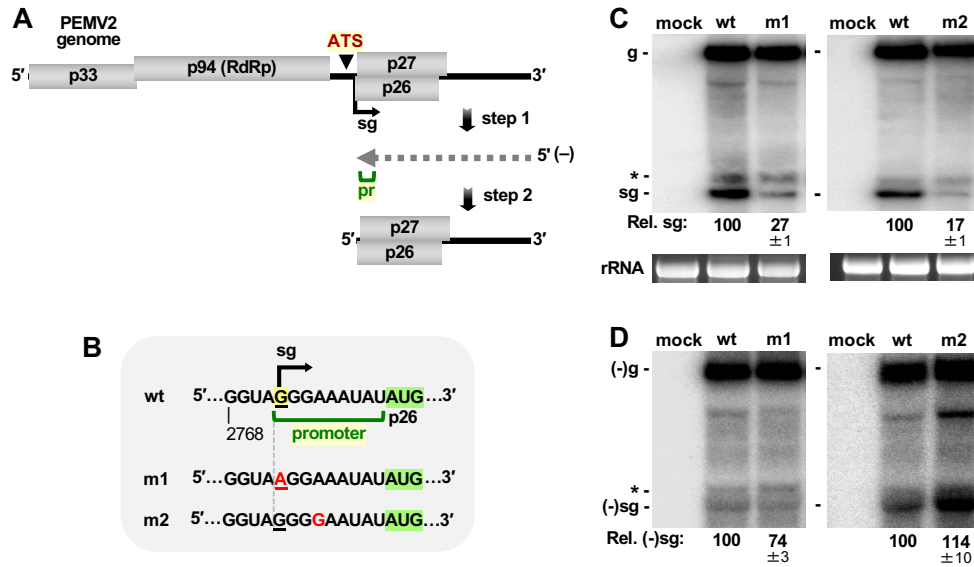
Figure 2: Assessing sg mRNA transcription in PEMV2. (A) Major steps in the premature termination mechanism. In step 1, the viral RdRp stalls at an RNA attenuation structure (ATS) generating a sg-sized minus-strand intermediate (dotted line). In step 2, the promoter (pr, green) at the 3'-end of the minus-strand is used by the RdRp to transcribe the sg mRNA. (B) The wt sg mRNA promoter sequence is shown along with promoter mutants, where nucleotide substitutions are indicated in red and sg mRNA initiating nucleotide is underlined. (C) Northern blot analysis of viral RNA accumulation 40 hours post-transfection of protoplasts with wt and promoter mutant PEMV2 genomic transcripts. The identities of the viral genomes in the infections are labeled at the top, and the blots were probed to detect plus-strand viral RNAs. The positions of viral RNAs are shown on the left, with an asterisk (*) representing sg*. Relative sg mRNA levels were calculated as the ratio of sg mRNA levels to their cognate genome levels, with the wt ratio set to 100%. Average sg mRNA accumulation levels, relative to that of the wt are shown below the blot along with standard errors determined from three independent experiments. (D) Northern blot analysis of minus-strand viral RNAs from the same protoplast infections as shown in panel C.

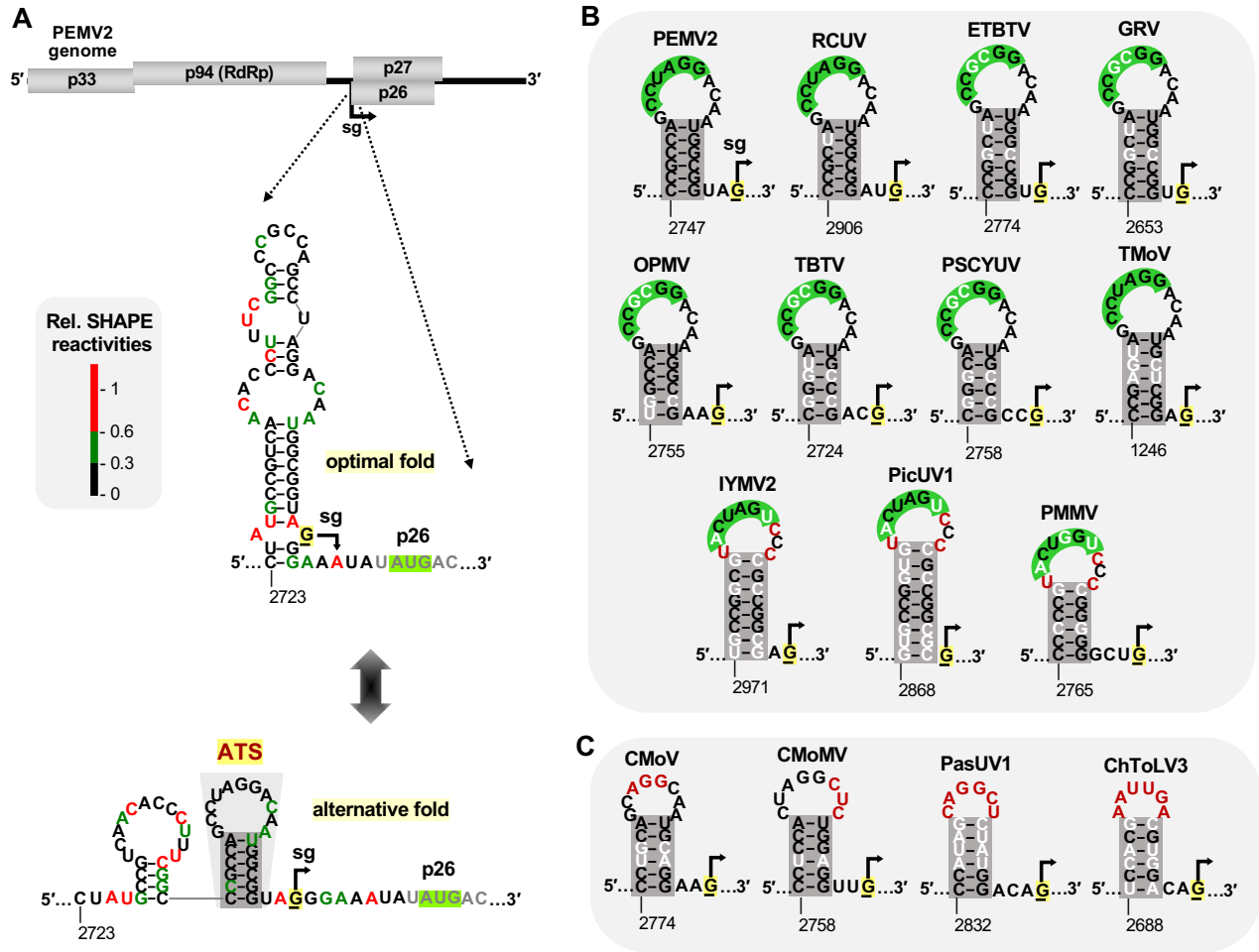## Structural analysis of potential attenuation structures

In the PT mechanism, sg mRNA transcription is mediated by an RNA attenuation structure usually positioned about two to three nucleotides upstream from the sg mRNA start site (3,5). In search of a possible attenuation structure in PEMV2, SHAPE RNA secondary structure analysis (52,53) was carried out on full-length wt PEMV2 genome transcripts. SHAPE

reactivity data (**Supplementary Table S1**), where high reactivities are indicative of flexible, single-stranded nucleotides, were used as a folding constraint with folding software to create RNA secondary structure models for the region upstream of the sg mRNA transcription start site (56-58). The results yielded two different mutually-exclusive RNA conformations, each of which contained an RNA stem-loop structure just upstream from the sg mRNA transcription initiation site (**Figure 3A**), thus both represented potential candidates for the attenuation structure. The most thermodynamically stable, optimal fold, predicted a more extended stem-loop structure (**Figure 3A, middle**), while a less stable alternative fold included two smaller stem-loop structures (**Figure 3A, bottom**).

In the alternative fold, the larger of the two stem-loops, located two nucleotides upstream from the initiation site (**Figure 3A, bottom**), was previously proposed to be important for PEMV2 sg mRNA transcription (24). Comparative structural analysis of this stem-loop with corresponding stem-loops in other umbraviruses revealed mono- and co-variation of base pairs (white nucleotides) that maintained bonding in the stem (**Figure 3B, C**), while no conservation was identified for the extended stem-loop in the optimal fold. Interestingly, in the conserved stem-loops, the loop sequences in 11 of 15 contained 6-nt-long palindromes (green), including coupled substitutions (white nucleotides) that maintained the palindromic characteristic of the sequences (**Figure 3B**). Four umbraviruses, however, did not contain a 6-nt palindrome and instead had variable loop sequences (**Figure 3C**). Overall, the high degree of covariation in the stems of these stem-loops, along with their positions just upstream from sg mRNA initiation sites, implicate them as strong candidates for attenuation structures involved in PT-mediated sg mRNA transcription. Accordingly, the stem-loop in the alternative fold in PEMV2 was prognostically termed ATS (<u>at</u>tenuation <u>s</u>tructure) (**Figure 3A, bottom).**

Figure 3: RNA secondary structure models for the PEMV2 attenuation structure. (A) SHAPE-guided prediction of optimal and alternative RNA folds of potential attenuation structures. The optimal structure is shown directly below the genome with colour-coded SHAPE reactivities for individual nucleotides (no SHAPE data for nucleotides in grey). An alternative fold is shown below the optimal fold, with a double-headed arrow depicting interconversion between the two mutually exclusive conformations. The experimentally-confirmed attenuation structure, a 6-bp-long stem-loop, is labelled as ATS in the alternative fold. (B) and (C) Comparative RNA secondary structure and sequence analysis of corresponding ATS stem-loops in umbraviruses. RNA secondary structures were generated using *RNAStructure* webserver (56). Mono- and co-variations that maintain base-pairing in stems (dark grey) or a palindromic sequence (green) are depicted as white nucleotides, while substitutions that either disrupt the stem or the palindrome are depicted in maroon. PEMV2 (NC_003853.1), red clover umbravirus (RCUV, MG596237.1), Ethiopian tobacco bushy top virus (ETBTV, NC_024808.1), groundnut rosette virus (GRV, NC_003603.1), opium poppy mosaic virus (OPMV, NC_027710.2), tobacco bushy top virus (TBTV, NC_004366.1), paedria scandens chlorosis yellow umbravirus (PSCYUV, OP053684.1), patrinia mild mottle virus (PMMV, MH922775.1), ixeridium yellow mottle virus 2 (IYMV2, NC_034243.1), picris umbravirus 1 (PicUV1, OL472232.1), tobacco mottle virus (TMV, AY007231.1), carrot mottle virus (CMoV, NC_011515.1), carrot mottle mimic virus (CMoMV, NC_001726.1), pastinaca umbravirus 1 (PasUV1, OL472236.1), Changjian tombus-like virus 3 (ChToLV3, NC_033104.1).

91

The stem of the ATS is important for sg mRNA transcription

To investigate if the ATS was involved in sg mRNA transcription during PEMV2 infections, the role of its stem was assessed by compensatory mutational analysis (**Figure 4A**). Disruption of base pairing in the stem (mutants m3 and m4) had specific and severe adverse effects on sg mRNA accumulation in protoplast infections and did not allow for either plus- or minus-strand sg mRNA accumulation to detectable levels (**Figure 4B, C**). Restoration of the stem with transposed base pairs in compensatory mutant m5 rescued sg mRNA levels for both strands to ~125% of wt (**Figure 4B, C**), demonstrating a clear dependence of transcription on the formation of the stem of the ATS. Moreover, the reliance of minus-strand intermediate accumulation on ATS formation is consistent with the ATS promoting premature RdRp termination during minus-strand RNA synthesis (**Figure 4C**). The results also argue against the optimal stem-loop predicted by SHAPE (**Figure 3A**) being important for transcription, because nucleotide substitutions in m5, which restored sg mRNA levels past that of wt, simultaneously disrupted its stem (**Figure 4D**).
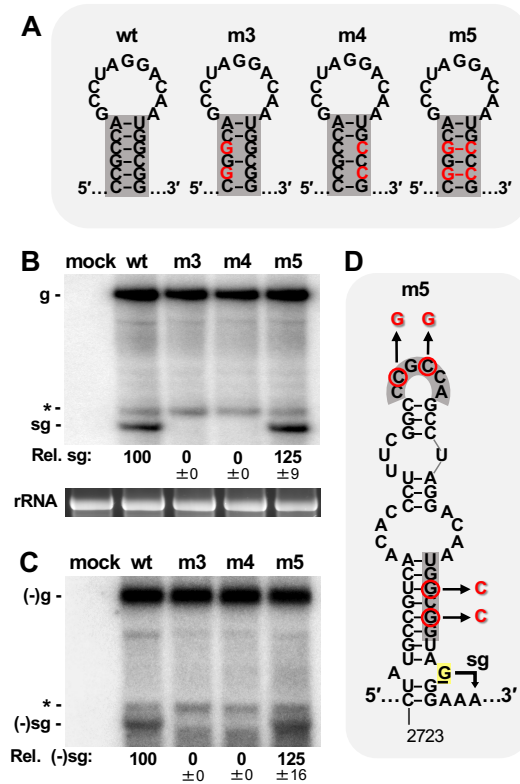
Figure 4: Functional assessment of the ATS stem. (A) Wt and mutant genomes with disruptions and then restoration of the stem of the ATS. Nucleotide substitutions are indicated in red. **(B)** Northern blot analysis of plus-strand viral RNAs extracted from protoplasts transfected with wt and mutant PEMV2 genome transcripts shown in panel A. **(C)** Northern blot analysis of minus-strand viral RNAs corresponding to the protoplast infections in panel B. **(D)** Optimal fold conformation of initiation site region (see Figure 3) with m5 nucleotide substitutions (red) mapped onto the structure.

PEMV2 genome dimerization via the ATS activates sg mRNA transcription

Having established the importance of the ATS's stem for sg mRNA transcription, we next

explored the potential role of its loop. The thermodynamic stability of the stem-loop forming the

ATS is predicted to be relatively weak (ΔG = -9.0 kcal/mol), therefore, on its own, it presumably

would not be an effective attenuation structure. One approach tombusvirids use to bolster the

stability of their attenuation structures is to have initiation site-proximal sequences/structures

base pair with upstream RNA sequences/structures (42-46). For the ATS, this could be

accomplished if the loop sequence paired with sequences more upstream in the PEMV2

genome, however exploration of this possibility did not identify any promising upstream partner

candidates. Consequently, we considered the possibility that the conserved 6-nt-long

palindrome in the loop instead allowed for pairing between two PEMV2 genomes (**Figure 5A**),

via a kissing-loop interaction (**Figure 5B, left**), to form a genomic dimer. This interaction could

generate a more stable *trans*-paired ATS capable of stalling the RdRp during minus-strand

synthesis (**Figure 5A, step 1**). Additionally, the possibility existed for transition of the kissing-

loop interaction to a more stable extended duplex by further *trans*-pairing between stem

sequences (**Figure 5B, right**).

To investigate the proposed involvement of genome dimerization, compensatory

mutational analysis assessing the ATS-mediated *trans*-interaction was carried out by targeting

the palindromic sequence. The strategy employed was to first independently substitute each of

the two central nucleotides of the palindrome to disrupt self-complementarity (**Figure 6A, top,

m8 and m9**), and then combine the two substitutions to restore self-complementarity, with

different central nucleotides (**Figure 6A, top, m10**). The net result being that the two central

nucleotides were converted from self-complementary UA in the wt PEMV2 genome to self-

94

complementary GC in m10. Substitutions in non-self-complementary m8 and m9 greatly

diminished sg mRNA accumulation in protoplast infections, while restoration of dimerization
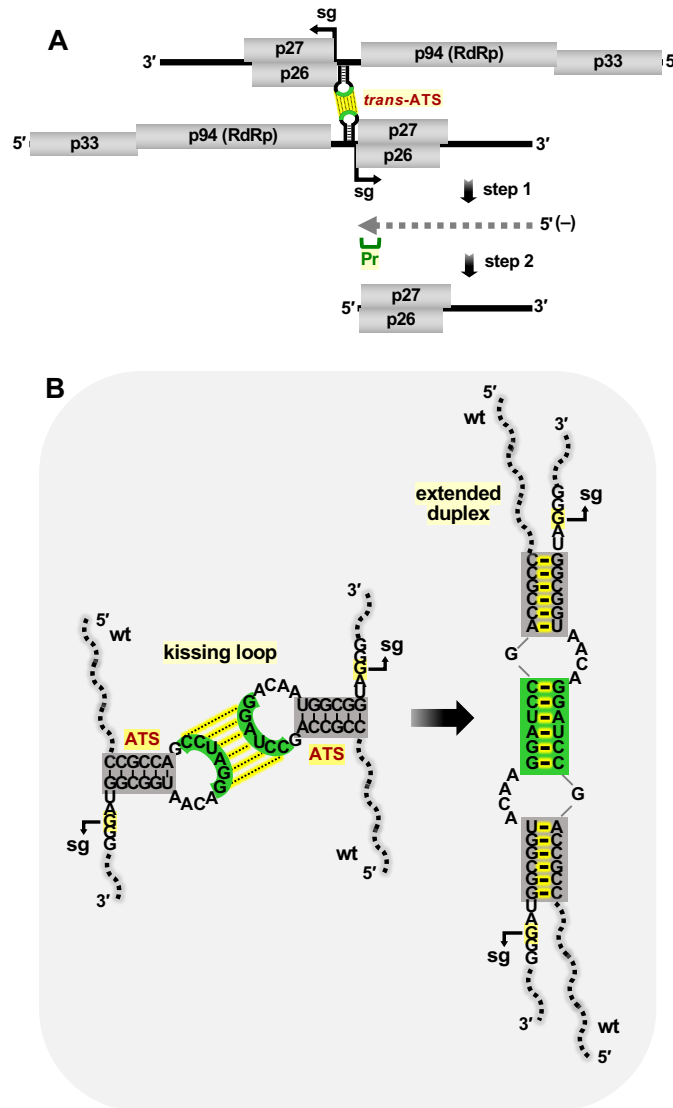
Figure 5: ATS-mediated dimerization of PEMV2 genomes. (A) Two PEMV2 genomes dimerizing via the palindromic sequences in their ATSs. The RNA structure formed by the paired ATSs could form a trans-conformation capable of directing premature termination during genomic minus-strand synthesis. For simplicity, sg mRNA production is shown for only one of the genomes in the dimer. **(B)** Possible ATS-mediated base-pairing interactions. Pairing of palindromic sequences would form a kissing-loop interaction that could potentially transition to an extended duplex involving ATS stem sequences. Base pairs that occur between the two genomes are highlighted in yellow.
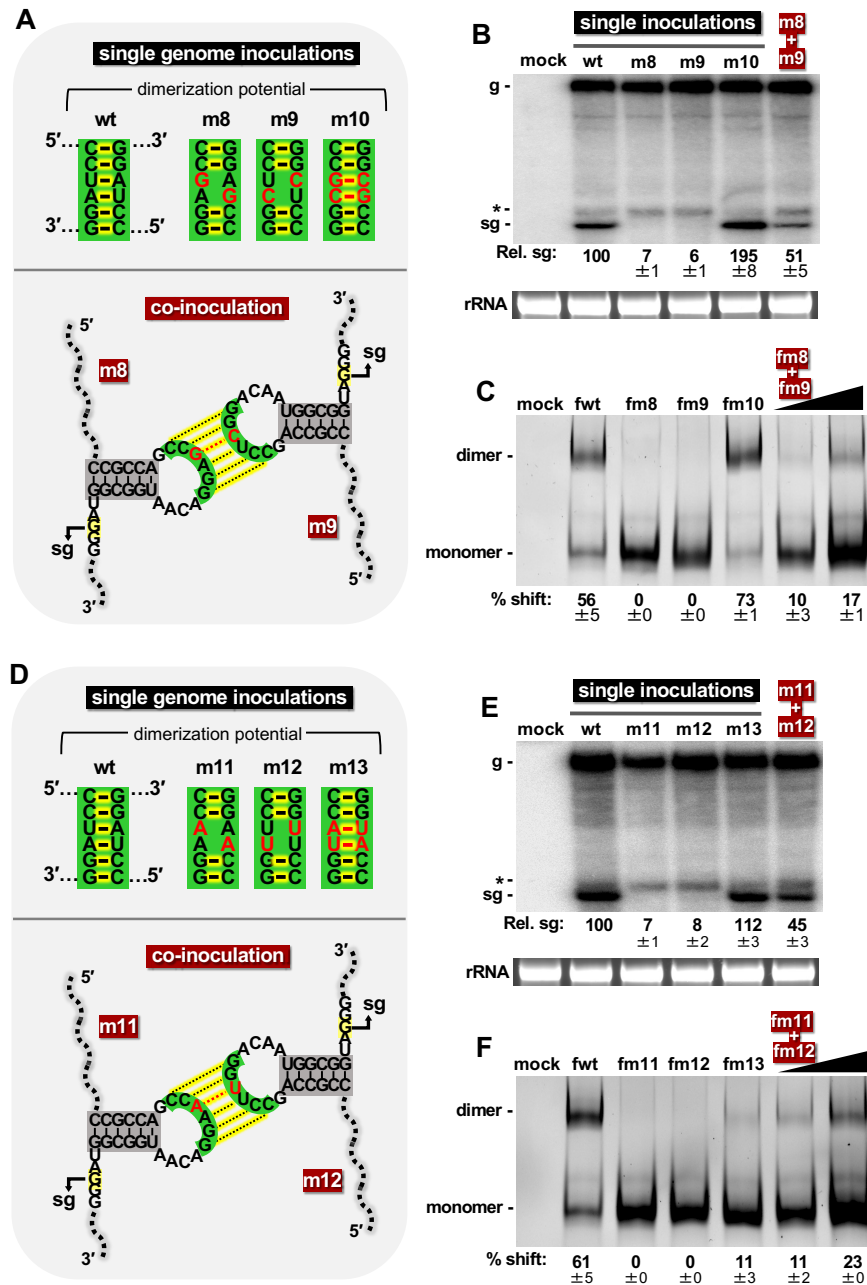
Figure 6: Analysis of PEMV2 genome dimerization via the ATS palindrome. (A) Wt and mutant genomes with substitutions (red) in the 6-nt-long palindromic sequence that disrupt and then restore self-complementarity (top). Shown below, is the *trans*-complementarity of the palindromes present in mutants m8 and m9. **(B)** Northern blot analysis of plus-strand viral RNAs from protoplast transfections with single genomes or co-transfections with two different genomes (m8+m9). **(C)** RNA-RNA EMSA analysis of RNA fragments (denoted by an f) corresponding to the 224-nt-long intergenic region in the PEMV2 genome that contains the ATS. The RNA fragments contained the same substitutions in the palindrome as in their genomic counterparts (panel A, top). The fragments tested (5 pmol each) are listed above each lane, and the second lane in the fm8+fm9 co-incubation contained double the fragment concentration (10 pmol each). The positions of monomers and dimers are indicated on the left, and % shift values [dimer/dimer+monomer] (±SE), calculated from three independent experiments, are shown below each lane. **(D), (E) and (F)** As described for A, B and C, respectively, but with a different set of palindromic mutants.

96

potential with stronger GC pairs in m10 revived sg mRNA levels past that of wt (**Figure 6B**). The requirement for maintenance of a self-complementary palindromic sequence is consistent with genome dimerization via the ATS being necessary for sg mRNA transcription.

To provide more definitive evidence for the involvement of an intermolecular base-pairing interaction between two genomes during infections, equimolar amounts of the single-nucleotide mutants m8 and m9 were co-transfected into protoplasts. Neither of these mutants was capable of homodimerizing, presumably due to disruption of the palindrome and self-complementarity (**Figure 6A, top**). However, in coinfections, m8 and m9 could potentially heterodimerize with loop sequences that were *trans*-complementary but not palindromic (**Figure 6A, bottom**). Indeed, northern blot analysis of the co-inoculation of m8 and m9 showed an over 7-fold increase in sg mRNA levels, from ~6-7% in single inoculations to ~51% in the co-inoculation (**Figure 6B**). This result provides compelling evidence supporting the importance of ATS loop complementarity and supports genome dimerization as the mechanism activating sg mRNA transcription.

Next, to demonstrate that PEMV2's ATS can physically interact to form dimers, RNA-RNA EMSAs were carried out. Wt and mutant fragments, based on corresponding wt and mutant genomic mutants (**Figure 6A**) and spanning the 224-nt-long intergenic region between the p94 and p26 ORFs (**Figure 1A**) were transcribed *in vitro* and equimolar amounts were incubated either separately (fwt, fm8, fm9, or fm10) or in combination (fm8 and fm9). Non-denaturing PAGE analysis showed that individual wt or mutant fm10 fragments containing intact palindromes dimerized, while individual fm8 or fm9 with disrupted palindromes did not (**Figure 6C**). However, co-incubation of fm8 and fm9 led to detectable dimerization, which was absent when they were tested individually (**Figure 6C**). These EMSA results correlate well with those from in vivo protoplast infections (**Figure 6B**) and provide physical evidence for a palindrome-mediated ATS interaction based on complementarity.

97

Comparable findings to those described above were also obtained using a different set of mutants (m11, m12 and m13) containing alternative substitutions (i.e. AU) in the central pair of the palindrome (**Figure 6D, E, F**). Collectively, these in vivo and *in vitro* observations provide compelling evidence that PEMV2 genome dimerization activates PT-mediated sg mRNA transcription via base pairing of the palindromic sequence in the ATS.

An extended duplex is not required for activation of sg mRNA transcription

Having established the importance of the palindromic sequence for dimerization, we then investigated the possible requirement for a transition from a kissing-loop interaction to an extended duplex structure (**Figure 5B**). To explore this prospect, different pairs of genomic mutants were created in which the ATS stems of the individuals in the pairs were either complementary to each other or not complementary to each other. Toward this objective, a new set of mutant PEMV2 genomes were generated that harbored a modified ATS stem in combination with a wt loop or with loop mutations present in m8, m9, or m10 (**Figure 6A**) to produce m23, m27, m28 and m29, respectively (**Figure 7A, top four rows**). In terms of self-complementarity, m23 and m29 contained compatible loops, m27 and m28 loops were incompatible, and all stems were compatible for self-pairing (**Figure 7A, top four rows**). Then, to assess if extended pairing that included stem sequences was required, m27 and m28 were tested in select combinations with the previously described m8 and m9 palindrome mutants containing wt stems (**Figure 6A**). For co-inoculations of m9+m27 or m8+m28, the ATS stems of the individual genomes in the pairs were not compatible with each other, but their loops were, while in m27+m28, both the ATS loops and stems in the two genomes were compatible with each other (**Figure 7A, last three rows, and Figure 7B**). Also, note that the ATS loops of all genomes tested in pairs (i.e. co-inoculations) were not self-complementary (**Figure 7A, last three rows**), thus loop pairing could only occur between the two different genomes in co-inoculations.

98
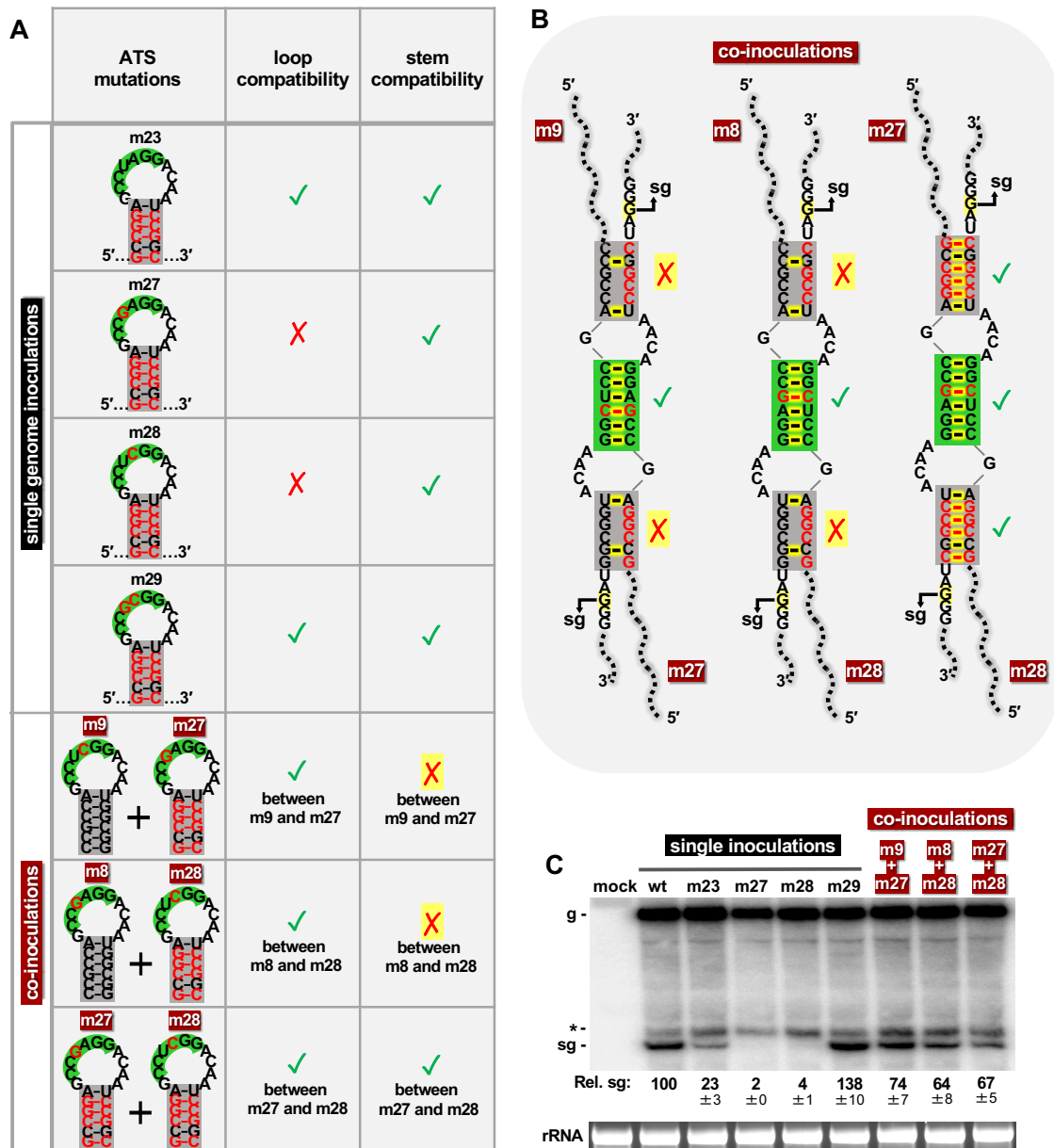
Figure 7: Assessing the requirement for an extended duplex. **(A)** Different ATS mutants showing their corresponding loop and stem compatibilities (i.e. complementarity). The top four rows correspond to self-compatibility of individual genomes, whereas the last three rows relate to compatibility between two different genomes. **(B)** Depictions of stem and loop compatibilities in the three different co-inoculations. **(C)** Northern blot analysis of plus-strand viral RNA accumulation in protoplasts transfected with wt and mutant PEMV2 genome transcripts depicted in panel A.

The functionality of the initial set of mutants were tested first (**Figure 7A, top four rows**). Single inoculations with m23, m27, m28, or m29 exhibited anticipated effects on sg mRNA accumulation (**Figure 7C**) that were similar to those for wt, m8, m9, and m10, respectively (**Figure 6B**). Next, co-transfections of m9+m27 or m8+m28, with *trans*-incompatible stems (due to a total of eight mismatches), were compared with co-transfection of m27+m28 with *trans*-compatible stems (**Figure 7B**). Both m9+m27 and m8+m28 accumulated sg mRNA at ~74% and ~64% that of wt, which were levels comparable to the ~67% observed for the m27+m28 co-inoculation (**Figure 7C**). These data indicate that intergenomic stem base-pairing leading to extended duplex formation is not required for the activation of PEMV2 sg mRNA transcription.

Further functional analysis of ATS stem and loop sequences

Recognizing the central role played by the palindromic sequence, we next sought to examine the importance of other nucleotides in the loop. The 5'-G and ACAA-3' loop sequences flanking the palindrome in PEMV2 are conserved in 8 of the 11 palindrome-containing umbraviruses (**Figure 3B, top two rows**), while the remaining three in the group are flanked by 5'-U and CCC-3' (**Figure 3B, bottom row**). Notably, the palindrome in the latter group begins with 5'-A and ends with U-3', whereas those in the former group have palindromic termini of 5'-C and G-3' (**Figure 3B**). These correlations and conservations suggested that the flanking residues play important functional roles in combination with their cognate palindromes. To address this possibility, different single nucleotide substitutions were introduced into certain nucleotides in the palindrome-flanking sequences in PEMV2. The substitutions in m15 and m17 extended the palindrome, corresponding substitutions in m14 and m16 did not, and the conserved C was converted to A in m18 (**Figure 8A**). All modifications elicited severe negative effects on sg mRNA accumulation in infections (**Figure 8B**). EMSA analysis of corresponding RNA fragments showed varying effects, with fm14 and fm17 dimerizing at levels higher (~82%)
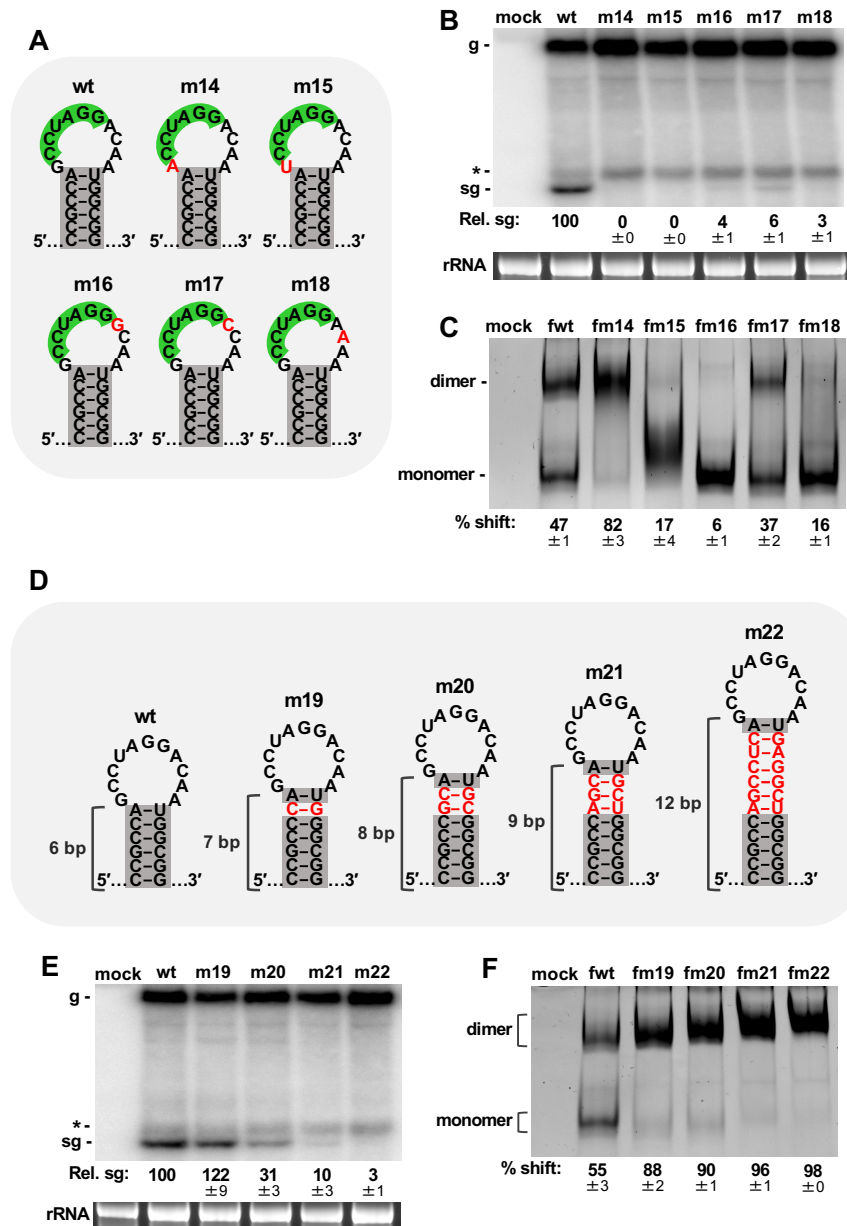
Figure 8: Testing the importance of ATS non-palindromic loop sequences and stem length. (A) Wt and mutant PEMV2 genomes with substitutions (red) in non-palindromic loop sequence in the ATS. **(B)** Northern blot showing effects of mutations in panel A on plus-strand PEMV2 genome and sg mRNA accumulation in protoplasts. **(C)** RNA-RNA EMSA of fragments corresponding to the intergenic region of mutants shown in panel A. **(D)** Wt and mutants with extended (red) stems. **(E)** Northern blot showing effects of mutations in panel D on plus-strand PEMV2 genome and sg mRNA accumulation in protoplasts. **(C)** RNA-RNA EMSA of fragments corresponding to the intergenic region of mutants shown in panel D.

or similar (~37%), respectively, to wt (~47%), while fm15, fm16, and fm18 displayed far lower than wt levels of dimerization, ~17%, ~16%, ~6%, respectively (**Figure 8C**). Also, extension of the palindrome with canonical base pairs (m15 and m17) did not lead to enhancement of dimerization over that of wt (**Figure 8A, C**). Collectively, the EMSA results suggest that the defects observed in infections could be related to either an inability to dimerize efficiently (fm15, fm16, and fm18) or formation of dimers that were not capable of functioning as attenuation signals (fm14 and fm17). Regardless, the flanking sequences clearly play essential, but yet to be defined, roles in mediating efficient sg mRNA transcription.

The stem of the ATS was also examined further. The length of ATS stems in umbraviruses varies, but if a 2-nt single-stranded spacer (that is predicted in most, 8 of 15) is assumed for all, then the stem length would fall into a narrower range of 6 to 7 bps (**Figure 3B, C**). To assess its importance, the ATS stem length in PEMV2 was increased by one, two, three, or six bps, while maintaining the AU base pair closing the loop (**Figure 8D**). The 7-bp stem in m19 was tolerated well in protoplast infections and yielded ~122% of wt sg mRNA levels, while lengthening the stem further in the other mutants had increasingly detrimental effects (**Figure 8E**). EMSA analysis with corresponding RNA fragments showed that dimerization capabilities increased with increasing stem length (**Figure 8F**). These results indicate that, although longer ATS stems allow for increased dimerization *in vitro*, corresponding dimerizations in infections are either impaired or the dimers formed are not capable of acting as attenuation signals. Collectively, the data indicate that both the palindrome-flanking nucleotides and the length of the ATS stem are critical for formation of an active attenuation structure capable of promoting PEMV2 sg mRNA transcription.

DISCUSSION

Infectious cycles of RNA viruses consist of multiple critical steps, each of which can be regulated by different RNA elements that form functional higher-order RNA structures within viral genomes locally, across long distances, or *in trans* via genome dimerization (62-68). Here we examined the role of an RNA structure in mediating transcription of the primary sg mRNA of PEMV2. Our results revealed a critical role for genome dimerization in this process that relies on specific RNA sequences and structural components. Below we examine this mechanism in more detail, compare the PEMV2 dimerization event with those in other viruses, and explore the prevalence of this transcriptional mechanism in related viruses.

Mechanistic aspects of PEMV2 sg mRNA transcription

Results from our analyses indicate that transcription of PEMV2's sg mRNA occurs via a premature termination mechanism (**Figure 2 and 4**). However, unlike for other non-segmented tombusvirids, which utilize intragenomic RNA interactions (37-46), activation of PEMV2 transcription requires genome dimerization. Assembly of PEMV2's functional *trans*-ATS relies on the formation of individual ATSs (**Figure 4**) that mediate genome dimerization via 6 nt-long palindromic sequences in their loops (**Figure 6**). Several lines of experimental evidence support the existence and functional relevance of PEMV2 genome dimerization for transcription: (i) comparative sequence analysis showed that self-complementary palindromes in umbravirus ATSs are preserved via covariation (**Figure 3B**); (ii) viral RNA dimerization *in vitro* correlated with transcriptional activity in infections (**Figure 6A, C, D, F**) (iii) compensatory mutational analysis confirmed that transcription was dependent on self-complementarity (**Figure 6A, B, D, E**, **single inoculations**); and (iv) co-transfection of two different genomes, each defective for self-dimerization, activated sg mRNA transcription via complementarity-mediated heterodimerization (**Figure 6A, B, D, E, co-inoculations**). Collectively, these results provide

compelling experimental support for PEMV2 genome dimerization activating sg mRNA transcription.

The functional interaction formed between two ATSs depends on complementary loop sequences predicted to form a kissing-loop structure (**Figure 5B, left**). This interaction likely involves coaxial stacking that stabilizes the complex and allows the ATS stems to be more effective RdRp barriers. Although the potential exists for this interaction to expand into an extended duplex (**Figure 5B, right**), our data showed that this transition is not required for activation of sg mRNA transcription (**Figure 7**). Consequently, it is the kissing-loop configuration of the *trans*-ATS that triggers viral RdRp stalling/termination during minus-strand synthesis (**Figure 5A**). Additional features of the ATS, loop sequences flanking the palindrome and stem length, also contribute to its activity (**Figure 8**). The neighbouring loop residues could function by (i) assisting in palindrome presentation for dimerization, (ii) stabilizing the kissing loop via coaxial stacking and/or noncanonical interactions, and/or (iii) acting as specificity determinants needed for RdRp stalling/termination. ATSs with longer stem lengths inhibited sg mRNA transcription in infections, but maintained *in vitro* dimerization activity (**Figure 8D-F**). If these genomic mutants also dimerize during infections, the observed lack of sg mRNA transcription suggests that the distance and/or orientation of the kissing loop relative to the RdRp is important for stalling/termination. The significance of these features suggests that, in addition to the ATS stem serving as a physical RNA barrier for the RdRp, other sub-elements in the *trans*-ATS structure perform essential roles. For example, the RdRp may need to precisely contact the kissing loop region of the *trans*-ATS for successful termination. In tombusviruses, the C-terminus of the RdRp plays a specific and essential role in recognizing a highly complex ATS that directs transcription of the larger of two viral sg mRNAs (46,69).

The adoption of a genomic *trans*-activation mechanism for transcription provides a clever means of controlling the timing of sg mRNA production (48). As a bimolecular interaction, dimer formation is concentration dependent. This means that transcriptional activity would be

delayed until the accumulation of sufficient levels of PEMV2 genome to allow for productive

contacts. It is not known whether only one genome in the dimer engages in sg mRNA

transcription, as shown in **Figure 5A**. Alternatively, both genomes could be transcriptionally

active, either simultaneously or sequentially. In either case, the ultimate result is the production

of sg mRNAs encoding p26 and p27; viral proteins involved in long-distance and cell-to-cell

movement and inhibiting NMD (26,27,31-33). It is possible that early expression of these viral

proteins interferes with initial steps of the infection, such as genome translation or replication.

Alternatively, or additionally, production of these proteins later in infections may be optimal for

their activities. Regardless, the maintenance of this expression strategy indicates that the virus

benefits from this method of temporal gene regulation.

Comparison of PEMV2 with other viral RNA genomes that dimerize

The best-known examples of RNA genome dimerization come from studies on

retroviruses, with human immunodeficiency virus 1 (HIV-1) being the best characterized (70,71).

Similar to the ATS in PEMV2, HIV-1 RNA genomes dimerize via a 6-nt-long palindromic

sequence in a stem-loop structure termed the dimerization initiation site (DIS) (72-75). As we

have suggested likely occurs in the *trans*-ATS, the HIV-1 kissing-loop dimer is stabilized via

coaxial stacking interactions (76,77) and, for both interactions, residues flanking the

palindromes in their loops are important for dimerization (**Figure 8A, C**) (78). Importantly, there

is compelling experimental evidence for dimerization during infections for both PEMV2 and HIV-

1, through the analysis of compensatory mutations in the DIS palindrome (74,75) and the ATS

palindrome (**Figure 6**), respectively. The HIV-1 kissing-loop dimer is able to transition into an

extended duplex conformation *in vitro* under certain conditions (76,79-82), however the

formation or importance of this lengthened interaction during infections is unclear (70).

Conversely, we found that extended duplex formation of the *trans*-ATS is not required for its

activity (**Figure 7**), defining the kissing-loop structure as the relevant complex. Lastly, although HIV-1 and PEMV2 share the feature of genome dimerization, these events lead to distinct functions, with the DIS interaction facilitating co-packaging (75,83,84) and the *trans*-ATS promoting sg mRNA transcription.

Genome dimerization has also been proposed for certain non-segmented plus-strand RNA viruses (85,86). The hepatitis C virus (HCV) genome contains a palindromic stem-loop in its 3'UTR, termed the dimer linkage sequence (DLS) that directs kissing-loop dimerization of viral RNA fragments *in vitro* (87,88). Mutation of the DLS showed that it is important for optimal genome replication in infections (89,90). HCV dimerization has been suggested to mediate a switch between translation competent and replication competent conformations of the 3'UTR, but this is based largely on *in vitro* studies (91-95). Other plus-strand RNA viruses, namely severe acute respiratory syndrome coronavirus (SARS-CoV) and SARS-CoV-2, have also been proposed to dimerize with palindromic sequences located, respectively, in a frameshift-promoting structure (96) or the s2m element in the 3'UTR (97). However, for both HCV and these coronaviruses, compelling evidence for functionally relevant dimerization during infections is lacking. Thus, with respect to non-segmented plus-strand RNA viruses, PEMV2 represents the only example that has been experimentally validated to functionally dimerize during infections.

In many ways PEMV2 genome dimerization is related to dimerization in RCNMV. Both are members of the family Tombusviridae and belong to the same subclade based on RdRp similarity (13,98). Also, for these two viruses, dimerization activates sg mRNA transcription and mediates the expression of certain genes later in infections. However, they diverge with respect to the number and function of the proteins expressed from their respective sg mRNAs, and also in the type of dimer formed. PEMV2's sg mRNA encodes p26 and 27 that are involved in viral movement and inhibiting NMD (26,27,31-33), while RCNMV's sg mRNA encodes only the viral capsid protein (99). Another difference is that the bi-segmented RCNMV genome forms a

heterodimer between its two genome segments (48), whereas PEMV2 forms a homodimer.

RCNMV dimerization also mediates co-packaging of both genome segments (100,101),

whereas the packaging of two copies of the PEMV2 genome is unlikely due to limiting capsid

capacity (18,102,103). Another difference is that, unlike PEMV2's kissing-loop dimerization, the

RCNMV dimer involves pairing between the loop of a stem-loop in RNA2 and a linear (i.e. non-

loop) sequence in RNA1 (48,104). Also, in RCNMV, the RdRp barrier is formed by the *trans*-

paired RNA1 and RNA2 sequences, while in PEMV2 *cis*-formed ATS stems function as barriers.

Thus, there are significant differences in the individual structures involved in dimer formation, as

well as in the dimer interface generated. Nonetheless, both interactions generate effective

ATSs, though they likely operate differently because the *trans*-interaction in RCNMV can be

functionally replaced by a stable local hairpin in RNA1 (104), whereas the presence of

comparable stable local hairpins in PEMV2 was ineffective (**Figure 8D-F**).

Prevalence and emergence of dimerization-activated sg mRNA transcription

It was surprising to find that sg mRNA transcription for PEMV2 required genome

dimerization, as transcription in all other non-segmented tombusvirids that have been examined

involves ATS formation via intragenomic RNA interactions (37-46). Interestingly, four

umbraviruses did not contain palindromes in their ATS loops (**Figure 3C**), suggesting that they

do not use dimerization for transcriptional activation. As alluded to earlier, the ATSs of some

other tombusvirids are generated when transcription initiation site-proximal

sequences/structures base pair with nearby upstream sequences (37-39). For three of the four

non-palindromic umbravirus ATSs, we were able to identify nearby upstream stem-loops that

had loop sequences complementary to those in their corresponding ATSs (**Figure 9A**). Thus, it

is possible that for these viruses, *cis*-ATSs are generated by kissing-loops formed

intragenomically. Additionally, the pseudo-palindromic nature of the ATS loop sequence in
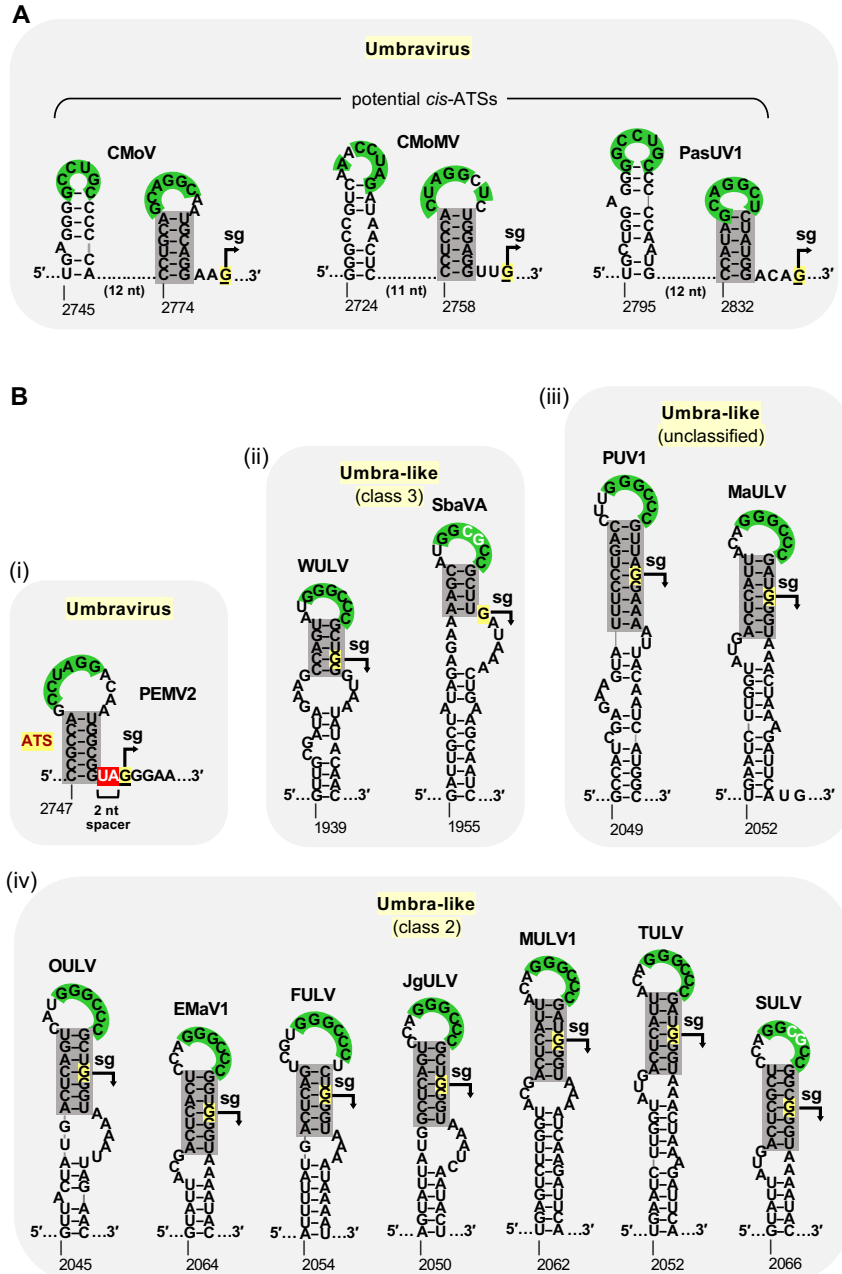
Figure 9: Alternative ATSs in some umbraviruses and in umbra-like viruses. (A) Potential umbravirus intragenomic RNA-RNA interactions between non-palindromic ATS loops (green) and complementary loop sequences (green) of small stem-loop structures located 11 to 12 nt upstream in CMoV, CMoMV and PasUV1. (B) Comparison of PEMV2 ATS (i) with those predicted in umbra-like viruses (ii) – (iv). Palindromes (green) are indicated, and covariations maintaining a palindromic sequence within the umbra-like groups are shown in white. Wheat umbra-like virus (WULV, OK573479), strawberry-associated virus A (SbaVA, MK211274), parsley umbravirus 1 (PUV1, OM419177.1), maize-associated umbra-like virus 1 (MaULV, OK018180.1), opuntia umbra-like virus (OULV, MT909563), Ethiopia maize-associated virus 1 (EMaV1, MN715238), fig umbra-like virus (FULV, MW480893), johnsongrass umbra-like virus 1 (JgULV1, OM937760), maize umbra-like virus 1 (MULV1, OM937759), teosinte-associated umbra-like virus (TULV, OK018180), sugarcane umbra-like virus (SULV, MN868593). RNA secondary structures were generated using RNAStructure webserver (56).

carrot mottle virus (CMoV) (**Figure 9A**) suggests that it could also dimerize utilizing non-canonical central AG/GA pairs. Moreover, a single base change in the central nucleotides, from AG to either CG or AU, would render the sequence fully palindromic, and this could provide an evolutionary pathway for it to transition from a *cis*-ATS to a *trans*-ATS. Indeed, it is logical to believe that some type of *cis*-interaction would need to precede the development of a corresponding *trans*-interaction, therefore the above scenario represents a plausible path for the emergence of *trans*-ATSs.

In addition to umbraviruses, there exists a group of smaller viruses that possess RdRps that are closely related to those in umbraviruses, and collectively these are termed umbra-like viruses (105-113). Umbra-like viruses have been organized into different classes (class 1, 2 or 3), with those in class 2 and 3 producing sg mRNAs (106). Comparative structural analysis in the region spanning sg mRNA initiation sites in these two classes (and two unclassified viruses) showed that, like the ATSs in umbraviruses such as PEMV2, they possessed ATS-like structures with a 6 nt-long palindromic sequence (including covariations) in their loops (**Figure 9B**). This suggests that these viruses also utilize genome dimerization for activation of sg mRNA transcription. However, the position of the umbra-like virus ATSs relative to their sg mRNA initiation sites were different, in that their initiation sites (except for strawberry-associated virus A, SbaVA) were located within the 3'-half of their ATS stems (**Figure 9B, compare i with ii - iv**). For the PEMV2 ATS and other umbraviruses, the sequence between the initiation site and the base of the stem (**Figure 9B(i)**) is thought to provide a ~2-nt spacer sequence that permits the RdRp to copy up to the initiating nucleotide as it encounters the base of the stem, which serves as a physical barrier (114). A similar scenario would not be possible with the umbra-like virus ATSs, because RdRp copying up to the initiating nucleotide would disrupt the stem of the ATS. Instead, similar to RCNMV (48,104), the intergenomically-paired helix, formed by the palindromic sequence and located about 3 nt from the start sites, likely functions as the RdRp barrier in these umbra-like viruses (**Figure 9B, ii – iv**). Thus, although dimer formation via

109

kissing loops would likely be similar in the two groups, precisely how these distinct *trans*-ATSs

function is predicted to be different, and will be the focus of future studies.

MATERIALS AND METHODS

Plasmid Construction

All PEMV2 constructs were derived from a cDNA copy of the wild-type (wt) full-length

PEMV2 genome (16) that was kindly provided by W. Allen Miller (Iowa State University). Desired

nucleotide substitutions were introduced into viral clones through site-directed mutagenesis and

confirmed by sequencing.

Viral RNA Preparation *in vitro*

PEMV2 cDNA constructs contained a T7 promoter directly upstream from the viral 5'-end

and a unique SmaI restriction enzyme site at the 3'-end of the genomic sequence. SmaI-

linearized PEMV2 cDNA templates were transcribed *in vitro* using T7-FlashScribe Transcription

Kit (CellScript) according to the manufacturer's instructions, to generate uncapped infectious

viral transcripts with authentic viral 5'- and 3'-ends.

Protoplast Transfections and Viral RNA Analyses

Cucumber protoplast transfections were carried out to monitor effects of different

nucleotide substitutions on PEMV2 genomic and sg mRNA accumulation in infections (49).

Approximately 500,000 protoplasts isolated from 6-day old cucumber leaves were transfected

with 5 μg of uncapped PEMV2 *in vitro*-transcribed genomic RNAs in the presence of

polyethylene glycol and $CaCl_2$. After a 40-hour incubation period under constant light at 22°C

total nucleic acids were extracted. For the time-course experiment, nucleic acid extractions were

carried out at 0, 6, 12, 18, 24, 30, 36, and 42 hours post inoculation. The extracted nucleic acids

were separated by agarose gel electrophoresis and ribosomal RNA bands were monitored by

ethidium bromide staining to ensure even loading. Subsequently, the nucleic acids were transferred to nylon membrane (Hybond-N+, Amersham) and plus-strand viral RNAs were detected by northern blotting using three 5'-[γ -^{32}P]-labeled DNA oligonucleotide probes complementary to the 3'-end of the PEMV2 genome (genome coordinates: 4232 – 4253, 4053 – 4079, 3937 – 3984). Northern blots were imaged using a Typhoon FLA 9500 (GE Healthcare) variable mode imager and viral RNA bands were quantified using the *Quantity One* software (Bio-Rad). Three biological replicates of protoplast inoculations were performed and average values with standard errors (SE) were calculated. Relative sg mRNA levels were determined by calculating the ratio of a sg mRNA to its cognate genome level, with the wt sg/genome ratio set to 100%.

For minus-strand viral RNA accumulation analysis, equal aliquots of total nucleic acids isolated from transfected cucumber protoplasts were treated with dimethyl sulfoxide and glyoxal to denature viral minus-strand RNA/plus-strand RNA hybrids and then separated by agarose gel electrophoresis in 10 mM sodium phosphate buffer (pH 7.0) (41,46). Northern blotting, imaging and data analysis were performed as described for plus-strand viral RNAs, however, a riboprobe was used for detection. A 3'-terminal segment of PEMV2 cDNA was amplified through PCR (with the forward primer containing a T7 promoter) and this PCR product was used as a template for *in vitro* transcription reaction (as described above) in the presence of [α-^{32}P]-UTP to generate an internally-labeled riboprobe (genome coordinates: 3864 – 4253) that was complementary to the minus-strand PEMV2 genome.

5'-rapid amplification of cDNA ends (5'-RACE)

The 5'-end of sg* RNA was mapped using the 5'-RACE approach, as described previously (50). Total nucleic acids were extracted from protoplasts transfected with wt PEMV2 genome transcripts. An adapter DNA oligonucleotide (5'-

GGCCTCTCACCACCAAAGAAGCAGTCAAGCCGTACCGAATTC) was ligated to 5'-ends of the

extracted total nucleic acids using T4 RNA ligase I (NEB), according to the manufacturer's

instruction. After ligation, first-strand cDNA synthesis was performed using a DNA primer

complementary to the PEMV2 genome (coordinates 2874 – 2895) and SuperScript IV reverse

transcriptase (Invitrogen), followed by two rounds of PCR amplification. The same reverse

primer used in the reverse transcription reaction was used for PCR, along with a forward DNA

primer (5'-GCAGTCAAGCCGTACCGAATTC) nested within the adapter oligonucleotide. PCR

products were purified by gel extraction, cloned into a pUC19 vector, and transformed into DH5-

α E. coli cells. Plasmid DNA was extracted from five different colonies and sequenced to identify

the very 5'-nucleotide of PEMV2 sg* RNA. All five sequenced plasmids mapped sg* 5'-

nucleotide to nucleotide C_{2620}.

In vitro Translation Assay

To examine the coding capacity of sg*, uncapped transcripts of wt sg mRNA and sg*

were generated *in vitro* using T7 polymerase. Transcripts, 0.5 pmol each of sg (24) and sg*,

were incubated in wheat germ extract (wge; Promega) containing [^{35}S]-Methionine at 25°C for 1

hr. Concentration of KOAc was increased to 133 mM to optimize translation efficiency (17,51).

The translated viral proteins were separated by 12% SDS-PAGE and detected through

phosphorimaging with a Typhoon FLA 9500 Variable Mode Imager (GE Healthcare).

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)

SHAPE was performed on full-length PEMV2 genome transcripts to model the RNA

secondary structures surrounding the sg mRNA transcription initiation site (46,52-54). 1-methyl-

7-nitroisatoic anhydride electrophile was used to modify flexible nucleotides in PEMV2 genome.

An oligonucleotide primer, fluorescently labeled at its 5'-end and complementary to a region

downstream from the sg mRNA transcription start site (coordinates: 2856–2890) was used for

primer extension reactions following electrophile treatment. The products of primer extension

were subjected to fluorescent capillary electrophoresis and the raw data obtained were

analyzed using *ShapeFinder* software (55) to generate relative reactivities for each nucleotide.

These reactivity values were normalized against the ten highest reactivities in the pool. The

SHAPE experiment was performed twice, with consistent results, and averaged values of the

two repeats were used for secondary structure prediction. The *RNAStructure* web server was

used (56) to combine SHAPE reactivity data (slope = 1.8 kcal/mol; intercept = -0.6 kcal/mol)

(57) with minimal free energy prediction and generate the secondary structure models for the

RNA region upstream from PEMV2 sg mRNA start site. RNA secondary structure models were

drawn using *RNA2Drawer* software (58).

RNA-RNA Electrophoretic Mobility Shift Assay (EMSA)

RNA-RNA EMSA was carried out as described previously (46). Briefly, wt and mutant

DNA templates containing a T7 promoter and corresponding to the PEMV2 intergenic region

(coordinates: 2556 - 2780) were generated through PCR amplification using corresponding

genomic clones as templates. The PCR products were transcribed *in vitro* (as described above)

to generate 224 nt-long viral RNA fragments. Five pmol of each transcript was heat-denatured

at 94°C for 3 minutes and then incubated in RNA binding buffer (5 mM HEPES pH7.8, 6 mM

$MgCl_2$, 100 mM KCl, 3.8% Glycerol) (59,60) at 37°C for 30 minutes. Samples were separated by

8% nondenaturing polyacrylamide gel electrophoresis in TBM buffer (45 mM Tris pH 8.3, 43 mM

boric acid, 2 mM $MgCl_2$) (60). The gels were stained with ethidium bromide and imaged using

Typhoon FLA 9500 variable mode imager (GE Healthcare). RNA bands representing monomers

and dimers were quantified for each RNA sample using *Quantity One* software (BioRad). The

%-shift values were calculated by determining the percentages of dimeric molecules from

pooled dimeric and monomeric band intensity values i.e., dimer/(dimer+monomer). Three trials

of EMSA experiments were conducted and average %-shift values and their corresponding SEs were calculated.

REFERENCES

1. Wolf,Y.I., Kazlauskas,D., Iranzo,J., Lucía-Sanz,A., Kuhn,J.H., Krupovic,M., Dolja,V.V. and Koonin,E.V. (2018) Origins and Evolution of the Global RNA Virome. *mBio*, **9**, e02329-18.

2. Miller,W.A. and Koev,G. (2000) Synthesis of Subgenomic RNAs by Positive-Strand RNA Viruses. *Virology*, **273**, 1–8.

3. White,K.A. (2002) The Premature Termination Model: A Possible Third Mechanism for Subgenomic mRNA Transcription in (+)-Strand RNA Viruses. *Virology*, **304**, 147–154.

4. Miller,W.A. and White,K.A. (2006) Long-Distance RNA-RNA Interactions in Plant Virus Gene Expression and Replication. *Annu. Rev. Phytopathol.*, **44**, 447–467.

5. Jiwan,S.D. and White,K.A. (2011) Subgenomic mRNA transcription in Tombusviridae. *RNA Biol.*, **8**, 287–294.

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

7. Firth,A.E. and Brierley,I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.*, **93**, 1385–1409.

8. Woo,P.C.Y., de Groot,R.J., Haagmans,B., Lau,S.K.P., Neuman,B.W., Perlman,S., Sola,I., van der Hoek,L., Wong,A.C.P. and Yeh,S-H. (2023) ICTV Virus Taxonomy Profile: Coronaviridae. *J. Gen. Virol.*, (in press).

9. Chen,R., Mukhopadhyay,S., Merits,A., Bolling,B., Nasar,F., Coffey,L.L., Powers,A., Weaver,S.C., and ICTV Report Consortium (2018) ICTV Virus Taxonomy Profile: Togaviridae. *J. Gen. Virol.*, **99**, 761–762.

10. Sahul Hameed,A.S., Ninawe,A.S., Nakai,T., Chi,S.C., Johnson,K.L., and ICTV Report Consortium (2019) ICTV Virus Taxonomy Profile: Nodaviridae. *J. Gen. Virol.*, **100**, 3–4.

11. Bujarski,J., Gallitelli,D., García-Arenal,F., Pallás,V., Palukaitis,P., Reddy,M.K., Wang,A., and ICTV Report Consortium (2019) ICTV Virus Taxonomy Profile: Bromoviridae. *J. Gen. Virol.*, **100**, 1206–1207.

12. Sõmera,M., Fargette,D., Hébrard,E., Sarmiento,C., and ICTV Report Consortium (2021) ICTV Virus Taxonomy Profile: Solemoviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. *J. Gen. Virol.*, **102**.

13. Sit,T.L. and Lommel,S.A. (2015) Tombusviridae. In John Wiley & Sons, Ltd (ed), *eLS*. Wiley, pp. 1–9.

14. Syller,J. (2003) Molecular and biological features of umbraviruses, the unusual plant viruses lacking genetic information for a capsid protein. *Physiol. Mol. Plant Pathol.*, **63**, 35–46.

15. Taliansky,M.E. and Robinson,D.J. (2003) Molecular biology of umbraviruses: phantom warriors. *J. Gen. Virol.*, **84**, 1951–1960.

16. Demler,S.A., Rucker,D.G. and de Zoeten,G.A. (1993) The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *J. Gen. Virol.*, **74**, 1–14.

17. Chkuaseli,T. and White,K.A. (2022) Complex and simple translational readthrough signals in pea enation mosaic virus 1 and potato leafroll virus, respectively. *PLoS Pathog.*, **18**, e1010888.

18. Demler,S.A., Borkhsenious,O.N., Rucker,D.G. and de Zoeten,G.A. (1994) Assessment of the Autonomy of Replicative and Structural Functions Encoded by the Luteo-phase of Pea Enation Mosaic Virus. *J. Gen. Virol.*, **75**, 997–1007.

19. Wang,Z., Treder,K. and Miller,W.A. (2009) Structure of a Viral Cap-independent Translation Element That Functions via High Affinity Binding to the eIF4E Subunit of eIF4F. *J. Biol. Chem.*, **284**, 14189–14202.

20. Gao,F., Kasprzak,W., Stupina,V.A., Shapiro,B.A. and Simon,A.E. (2012) A Ribosome-Binding, 3′ Translational Enhancer Has a T-Shaped Structure and Engages in a Long-Distance RNA-RNA Interaction. *J. Virol.*, **86**, 9828–9842.

21. Gao,F., Gulay,S.P., Kasprzak,W., Dinman,J.D., Shapiro,B.A. and Simon,A.E. (2013) The Kissing-Loop T-Shaped Structure Translational Enhancer of Pea Enation Mosaic Virus Can Bind Simultaneously to Ribosomes and a 5′ Proximal Hairpin. *J. Virol.*, **87**, 11987–12002.

22. Gao,F., Kasprzak,W.K., Szarko,C., Shapiro,B.A. and Simon,A.E. (2014) The 3′ Untranslated Region of Pea Enation Mosaic Virus Contains Two T-Shaped, Ribosome-Binding, Cap-Independent Translation Enhancers. *J. Virol.*, **88**, 11696–11712.

23. Du,Z., Alekhina,O.M., Vassilenko,K.S. and Simon,A.E. (2017) Concerted action of two 3′ cap-independent translation enhancers increases the competitive strength of translated viral genomes. *Nucleic Acids Res.*, **45**, 9558–9572.

24. Gao,F. and Simon,A.E. (2017) Differential use of 3'CITEs by the subgenomic RNA of Pea enation mosaic virus 2. *Virology*, **510**, 194–204.

25. Gao,F. and Simon,A.E. (2016) Multiple Cis-acting elements modulate programmed -1 ribosomal frameshifting in Pea enation mosaic virus. *Nucleic Acids Res.*, **44**, 878–895.

26. Ryabov,E.V., Robinson,D.J. and Taliansky,M.E. (1999a) A plant virus-encoded protein facilitates long-distance movement of heterologous viral RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 1212–1217.

27. Ryabov,E.V., Robinson,D.J. and Taliansky,M. (2001) Umbravirus-Encoded Proteins both Stabilize Heterologous Viral RNA and Mediate Its Systemic Movement in Some Plant Species. *Virology*, **288**, 391–400.

28. Taliansky,M., Roberts,I.M., Kalinina,N., Ryabov,E.V., Raj,S.K., Robinson,D.J. and Oparka,K.J. (2003) An Umbraviral Protein, Involved in Long-Distance RNA Movement, Binds Viral RNA and Forms Unique, Protective Ribonucleoprotein Complexes. *J. Virol.*, **77**, 3031–3040.

29. Kim,S.H., MacFarlane,S., Kalinina,N.O., Rakitina,D.V., Ryabov,E.V., Gillespie,T., Haupt,S., Brown,J.W.S. and Taliansky,M. (2007) Interaction of a plant virus-encoded protein with the major nucleolar protein fibrillarin is required for systemic virus infection. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 11115–11120.

30. Kim,S.H., Ryabov,E.V., Kalinina,N.O., Rakitina,D.V., Gillespie,T., MacFarlane,S., Haupt,S., Brown,J.W.S. and Taliansky,M. (2007) Cajal bodies and the nucleolus are required for a plant virus systemic infection. *EMBO J*, **26**, 2169–2179.

31. May,J.P., Johnson,P.Z., Ilyas,M., Gao,F. and Simon,A.E. (2020) The Multifunctional Long-Distance Movement Protein of *Pea Enation Mosaic Virus 2* Protects Viral and Host Transcripts from Nonsense-Mediated Decay. *mBio*, **11**, e00204-20.

32. Ryabov,E.V., Oparka,K.J., Santa Cruz,S., Robinson,D.J. and Taliansky,M.E. (1998) Intracellular Location of Two Groundnut Rosette Umbravirus Proteins Delivered by PVX and TMV Vectors. *Virology*, **242**, 303–313.

33. Ryabov,E.V., Roberts,I.M., Palukaitis,P. and Taliansky,M. (1999b) Host-Specific Cell-to-Cell and Long-Distance Movements of Cucumber Mosaic Virus Are Facilitated by the Movement Protein of Groundnut Rosette Virus. *Virology*, **260**, 98–108.

34. Nurkiyanova,K.M., Ryabov,E.V., Kalinina,N.O., Fan,Y., Andreev,I., Fitzgerald,A.G., Palukaitis,P. and Taliansky,M. (2001) Umbravirus-encoded movement protein induces tubule formation on the surface of protoplasts and binds RNA incompletely and non-cooperatively. *J. Gen. Virol.*, **82**, 2579–2588.

35. Ryabov,E.V., Fraser,G., Mayo,M.A., Barker,H. and Taliansky,M. (2001) Umbravirus Gene Expression Helps Potato leafroll virus to Invade Mesophyll Tissues and to be Transmitted Mechanically between Plants. *Virology*, **286**, 363–372.

36. Gao,F., Alekhina,O.M., Vassilenko,K.S. and Simon,A.E. (2018) Unusual dicistronic expression from closely spaced initiation codons in an umbravirus subgenomic RNA. *Nucleic Acids Res.*, **46**, 11726–11742.

37. Xu,W. and White,K.A. (2009) RNA-Based Regulation of Transcription and Translation of Aureusvirus Subgenomic mRNA1. *J. Virol.*, **83**, 10096–10105.

38. Wu,B., Oliveri,S., Mandic,J. and White,K.A. (2010) Evidence for a Premature Termination Mechanism of Subgenomic mRNA Transcription in a Carmovirus. *J. Virol.*, **84**, 7904–7907.

39. Jiwan,S.D., Wu,B. and White,K.A. (2011) Subgenomic mRNA transcription in tobacco necrosis virus. *Virology*, 418, 1–11.

40. Zhang,G., Slowinski,V. and White,K.A. (1999) Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA*, **5**, 550–561.

41. Choi,I.-R., Ostrovsky,M., Zhang,G. and White,K.A. (2001) Regulatory Activity of Distal and Core RNA Elements in Tombusvirus Subgenomic mRNA2 Transcription. *J. Biol. Chem.*, **276**, 41761–41768.

42. Choi,I.-R. and White,K.A. (2002) An RNA Activator of Subgenomic mRNA1 Transcription in Tomato Bushy Stunt Virus. *J. Biol. Chem.*, **277**, 3760–3766.

43. Lin,H.-X. and White,K.A. (2004) A complex network of RNA–RNA interactions controls subgenomic mRNA transcription in a tombusvirus. *EMBO J*, **23**, 3365–3374.

44. Xu,W. and White,K.A. (2008) Subgenomic mRNA transcription in an aureusvirus: Down-regulation of transcription and evolution of regulatory RNA elements. *Virology*, **371**, 430–438.

45. Blanco-Pérez,M. and Hernández,C. (2016) Evidence supporting a premature termination mechanism for subgenomic RNA transcription in Pelargonium line pattern virus: identification of a critical long-range RNA–RNA interaction and functional variants through mutagenesis. *J. Gen. Virol.*, **97**, 1469–1480.

46. Chkuaseli,T. and White,K.A. (2020) Activation of viral transcription by stepwise largescale folding of an RNA virus genome. *Nucleic Acids Res.*, **48**, 9285–9300.

47. Lommel,S.A., Weston-Fina,M., Xiong,Z. and Lomonossoff,G.P. (1988) The nucleotide sequence and gene organization of red clover necrotic mosaic virus RNA-2. *Nucleic Acids Res.,* **16**, 8587–8602.

48. Sit,T.L., Vaewhongs,A.A. and Lommel,S.A. (1998) RNA-Mediated Trans-Activation of Transcription from a Viral RNA. *Science,* **281**, 829–832.

49. White,K.A. and Morris,T.J. (1994) Nonhomologous RNA recombination in tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions. *J. Virol.*, **68**, 14–24.

50. Gunawardene,C.D., Newburn,L.R. and White,K.A. (2019) A 212-nt long RNA structure in the Tobacco necrosis virus-D RNA genome is resistant to Xrn degradation. *Nucleic Acids Res.*, **47**, 9329–9342.

51. Nicholson,B.L. and White,K.A. (2008) Context-influenced cap-independent translation of Tombusvirus mRNAs in vitro. *Virology*, **380**, 203–212.

52. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States. *PLoS Biol.*, **6**, e96.

53. Low,J.T. and Weeks,K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.

54. Cimino,P.A., Nicholson,B.L., Wu,B., Xu,W. and White,K.A. (2011) Multifaceted Regulation of Translational Readthrough by RNA Replication Elements in a Tombusvirus. *PLoS Pathog.*, **7**, e1002423.

55. Vasa,S.M., Guex,N., Wilkinson,K.A., Weeks,K.M. and Giddings,M.C. (2008) ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.

56. Bellaousov,S., Reuter,J.S., Seetin,M.G. and Mathews,D.H. (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, **41**, W471–W474.

57. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.

58. Johnson,P.Z., Kasprzak,W.K., Shapiro,B.A. and Simon,A.E. (2019) RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA Biol.*, **16**, 1667–1671.

59. Fabian,M.R. and White,K.A. (2006) Analysis of a 3′-translation enhancer in a tombusvirus: A dynamic model for RNA–RNA interactions of mRNA termini. *RNA*, **12**, 1304–1314.

60. Lafuente,E., Ramos,R. and Martínez-Salas,E. (2002) Long-range RNA–RNA interactions between distant regions of the hepatitis C virus internal ribosome entry site element. *J. Gen. Virol.*, **83**, 1113–1121.

61. Ilyas,M., Du,Z. and Simon,A.E. (2021) *Opium Poppy Mosaic Virus* Has an Xrn-Resistant, Translated Subgenomic RNA and a BTE 3′ CITE. *J. Virol.*, **95**, e02109-20.

62. Nicholson, B.L., and White, K.A. (2014) Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nature Rev. Microbiol*., **12**, 493–504.

63. Chkuaseli, T., and White, K.A. (2018) Intragenomic long-distance RNA-RNA interactions in plus-strand RNA plant viruses. *Front. Microbiol*., **9**, 529.

64. Jaafar,Z.A. and Kieft,J.S. (2019) Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol*, **17**, 110–123.

65. Newburn,L.R. and White,K.A. (2019) Trans-Acting RNA–RNA Interactions in Segmented RNA Viruses. *Viruses*, **11**, 751.

66. Geng,G., Wang,D., Liu,Z., Wang,Y., Zhu,M., Cao,X., Yu,C. and Yuan,X. (2021) Translation of Plant RNA Viruses. *Viruses*, **13**, 2499.

67. Ye,L., Ambi,U.B., Olguin-Nava,M., Gribling-Burrer,A.-S., Ahmad,S., Bohn,P., Weber,M.M. and Smyth,R.P. (2021) RNA Structures and Their Role in Selective Genome Packaging. *Viruses*, **13**, 1788.

68. Allan,M.F., Brivanlou,A. and Rouskin,S. (2023) RNA levers and switches controlling viral gene expression. *Trends Biochem. Sci.*, **48**, 391–406.

69. Wu,B. and White,K.A. (2007) Uncoupling RNA virus replication from transcription via the polymerase: functional and evolutionary insights. *EMBO J.*, **26**, 5120–5130.

70. Dubois,N., Marquet,R., Paillart,J.-C. and Bernacchi,S. (2018) Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.*, **9**, 527.

71. Hanson,H.M., Willkomm,N.A., Yang,H. and Mansky,L.M. (2022) Human Retrovirus Genomic RNA Packaging. *Viruses*, **14**, 1094.

72. Paillart,J.C., Marquet,R., Skripkin,E., Ehresmann,B. and Ehresmann,C. (1994) Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J. Biol. Chem.*, **269**, 27486–27493.

73. Clever,J.L., Wong,M.L. and Parslow,T.G. (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J. Virol.*, **70**, 5902–5908.

74. Moore,M.D., Fu,W., Nikolaitchik,O., Chen,J., Ptak,R.G. and Hu,W.-S. (2007) Dimer Initiation Signal of Human Immunodeficiency Virus Type 1: Its Role in Partner Selection during RNA Copackaging and Its Effects on Recombination. *J. Virol.*, **81**, 4002–4011.

75. Chen,J., Nikolaitchik,O., Singh,J., Wright,A., Bencsics,C.E., Coffin,J.M., Ni,N., Lockett,S., Pathak,V.K. and Hu,W.-S. (2009) High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 13535–13540.

76. Ennifar,E., Walter,P., Ehresmann,B., Ehresmann,C. and Dumas,P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.*, **8**, 1064–1068.

77. Ennifar,E. and Dumas,P. (2006) Polymorphism of Bulged-out Residues in HIV-1 RNA DIS Kissing Complex and Structure Comparison with Solution Studies. *J. Mol. Biol.*, **356**, 771–782.

78. Paillart,J.-C., Westhof,E., Ehresmann,C., Ehresmann,B. and Marquet,R. (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.*, **270**, 36–49.

79. Muriaux,D., De Rocquigny,H., Roques,B.-P. and Paoletti,J. (1996) NCp7 Activates HIV-1Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer. *J. Biol. Chem.*, **271**, 33686–33692.

80. Ennifar,E., Yusupov,M., Walter,P., Marquet,R., Ehresmann,B., Ehresmann,C. and Dumas,P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, **7**, 1439–1449.

81. Brigham,B.S., Kitzrow,J.P., Reyes,J.-P.C., Musier-Forsyth,K. and Munro,J.B. (2019) Intrinsic conformational dynamics of the HIV-1 genomic RNA 5′UTR. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 10372–10381.

82. Blakemore,R.J., Burnett,C., Swanson,C., Kharytonchyk,S., Telesnitsky,A. and Munro,J.B. (2021) Stability and conformation of the dimeric HIV-1 genomic RNA 5′UTR. *Biophys. J.*, **120**, 4874–4890.

83. Chen,J., Rahman,S.A., Nikolaitchik,O.A., Grunwald,D., Sardo,L., Burdick,R.C., Plisov,S., Liang,E., Tai,S., Pathak,V.K., *et al.* (2015) HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc. Natl. Acad. Sci. U.S.A.*, **113**.

84. Ferrer,M., Clerté,C., Chamontin,C., Basyuk,E., Lainé,S., Hottin,J., Bertrand,E., Margeat,E. and Mougel,M. (2016) Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Res.*, **44**, 7922–7934.

85. Bou-Nader,C. and Zhang,J. (2020) Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions. *Molecules*, **25**, 2881.

86. Romero-López,C., and Berzal-Herranz,A. (2020) The role of the RNA-RNA interactome in the hepatitis C virus life cycle. *Int. J. Mol. Sci.* **21**pii: E1479.

87. Cristofari,G., Ivanyi-Nagy,R., Gabus,C., Boulant,S., Lavergne,J-P., Penin,F., and Darlix, J.-L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic Acids Res.*, **32**, 2623–2631.

88. Ivanyi-Nagy,R., Kanevsky,I., Gabus,C., Lavergne,J.-P., Ficheux,D., Penin,F., Fosse,P. and Darlix,J.-L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res.*, **34**, 2618–2633.

89. Shetty,S., Kim,S., Shimakami,T., Lemon,S.M. and Mihailescu,M.-R. (2010) Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. *RNA*, **16**, 913–925.

90. Masante,C., Jaubert,C., Palau,W., Plissonneau,J., Besnard,L., Ventura,M. and Di Primo,C. (2015) Mutations of the SL2 dimerization sequence of the hepatitis C genome abrogate viral replication. *Cell. Mol. Life Sci.*, **72**, 3375–3385.

91. You,S., Stump,D.D., Branch,A.D. and Rice,C.M. (2004) A *cis* -Acting Replication Element in the Sequence Encoding the NS5B RNA-Dependent RNA Polymerase Is Required for Hepatitis C Virus RNA Replication. *J. Virol.*, **78**, 1352–1366.

92. Romero-López,C. and Berzal-Herranz,A. (2009) A long-range RNA–RNA interaction between the 5′ and 3′ ends of the HCV genome. *RNA*, **15**, 1740–1752.

93. Palau,W., Masante,C., Ventura,M. and 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.

94. Cantero-Camacho,Á. and Gallego,J. (2018) An unexpected RNA distal interaction mode found in an essential region of the hepatitis C virus genome. *Nucleic Acids Res.*, **46**, 4200–4212.

95. Castillo-Martínez,J., Fan,L., Szewczyk,M.P., Wang,Y.-X. and Gallego,J. (2022) The low-resolution structural models of hepatitis C virus RNA subdomain 5BSL3.2 and its distal complex with domain 3′X point to conserved regulatory mechanisms within the Flaviviridae family. *Nucleic Acids Res.*, **50**, 2287–2301.

96. Ishimaru,D., Plant,E.P., Sims,A.C., Yount,B.L., Roth,B.M., Eldho,N.V., Pérez-Alvarado,G.C., Armbruster,D.W., Baric,R.S., Dinman,J.D., *et al.* (2013) RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.*, **41**, 2594–2608.

97. Imperatore,J.A., Cunningham,C.L., Pellegrene,K.A., Brinson,R.G., Marino,J.P., Evanseck,J.D. and Mihailescu,M.R. (2022) Highly conserved s2m element of SARS-CoV-2 dimerizes via a kissing complex and interacts with host miRNA-1307-3p. *Nucleic Acids Res.*, **50**, 1017–1032.

98. Gunawardene,C.D., Donaldson,L.W. and White,K.A. (2017) Tombusvirus polymerase: Structure and function. *Virus Res.*, **234**, 74–86.

99. Xiong,Z. and Lommel,S.A. (1989) The Complete Nucleotide Sequence and Genome Organization of Red Clover Necrotic Mosaic Virus RNA-1. *Virology*, **171**, 543–554.

100. Basnayake,V.R., Sit,T.L. and Lommel,S.A. (2009) The Red clover necrotic mosaic virus origin of assembly is delimited to the RNA-2 trans-activator. *Virology*, **384**, 169–178.

101. Park,S.-H., Sit,T.L., Kim,K.-H. and Lommel,S.A. (2013) The red clover necrotic mosaic virus capsid protein N-terminal amino acids possess specific RNA binding activity and are required for stable virion assembly. *Virus Res.*, **176**, 107–118.

102. German,T.L. and de Zoeten,G.A. (1975) Purification and properties of the replicative forms and replicative intermediates of pea enation mosaic virus. *Virology*, **66**, 172–184.

103. Sivakumar,S., Wang,Z., Harrison,R.L., Liu,S., Miller,W.A. and Bonning,B.C. (2009) Baculovirus-expressed virus-like particles of Pea enation mosaic virus vary in size and encapsidate baculovirus mRNAs. *Virus Res.*, **139**, 54–63.

104. Im,J.S.H., Newburn,L.R., Kent,G. and White,K.A. (2021) Trans-Activator Binding Site Context in RCNMV Modulates Subgenomic mRNA Transcription. *Viruses*, **13**, 2252.

105. Quito-Avila,D.F., Alvarez,R.A., Ibarra,M.A. and Martin,R.R. (2015) Detection and partial genome sequence of a new umbra-like virus of papaya discovered in Ecuador. *Eur. J. Plan.t Pathol.*, **143**, 199–204.

106. Liu,J., Carino,E., Bera,S., Gao,F., May,J.P. and Simon,A.E. (2021) Structural Analysis and Whole Genome Mapping of a New Type of Plant Virus Subviral RNA: Umbravirus-Like Associated RNAs. *Viruses*, **13**, 646.

107. Redila,C.D., Prakash,V. and Nouri,S. (2021) Metagenomics Analysis of the Wheat Virome Identifies Novel Plant and Fungal-Associated Viral Sequences. *Viruses,* **13**, 2457.

108. Tahir,M.N., Bolus,S., Grinstead,S.C., McFarlane,S.A. and Mollov,D. (2021) A new virus of the family Tombusviridae infecting sugarcane. *Arch. Virol.*, **166**, 961–965.

109. Wang,X., Olmedo-Velarde,A., Larrea-Sarmiento,A., Simon,A.E., Kong,A., Borth,W., Suzuki,J.Y., Wall,M.M., Hu,J. and Melzer,M. (2021) Genome characterization of fig umbra-like virus. *Virus Genes*, **57**, 566–570.

110. Lappe,R.R., Elmore,M.G., Lozier,Z.R., Jander,G., Miller,W.A. and Whitham,S.A. (2022) Metagenomic identification of novel viruses of maize and teosinte in North America. *BMC Genomics*, **23**, 767.

111. Fox,A., Gibbs,A.J., Fowkes,A.R., Pufal,H., McGreig,S., Jones,R.A.C., Boonham,N. and Adams,I.P. (2022) Enhanced Apiaceous Potyvirus Phylogeny, Novel Viruses, and New Country and Host Records from Sequencing Apiaceae Samples. *Plants*, **11**, 1951.

112. Koloniuk,I., Přibylová,J., Čmejla,R., Valentová,L. and Fránová,J. (2022) Identification and Characterization of a Novel Umbra-like Virus, Strawberry Virus A, Infecting Strawberry Plants. *Plants*, **11**, 643.

113. Quito-Avila,D.F., Reyes-Proaño,E.G., Mendoza,A., Margaria,P., Menzel,W., Bera,S. and Simon,A.E. (2022) Two new umbravirus-like associated RNAs (ulaRNAs) discovered in maize and johnsongrass from Ecuador. *Arch. Virol.*, **167**, 2093–2098.

114. Wang,S., Mortazavi,L., White,K.A. (2008) Higher-order RNA structural requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. *J. Virol*. **82**, 3864–3871.

Dimerization of an Umbravirus RNA Genome Activates Subgenomic mRNA Transcription

Tamari Chkuaseli and K. Andrew White*

Department of Biology, York University, Toronto, Ontario, Canada M3J 1P3

Table S1: Normalized SHAPE reactivities for PEMV2 intergenic region shown in Figure 3A. The SHAPE analysis method used is describe in the materials and methods section. Nucleotide positions and values for each independent replicate, as well as, the averages are shown. Cells are color-coded according to the SHAPE reactivity scale where no fill represents low reactivity(0 - 0.3), green – moderate (0.3 – 0.6), red – high (>0.6). Grey fill indicates nucleotides with no SHAPE data.

Position (nt)	Relative SHAPE Reactivities		
	Replicate 1	Replicate 2	Average
1			
⋮			
2723	0.183722674	0.220083268	0.201902971
2724	0.267404421	0.310549085	0.288976753
2725	1.233786203	1.45393654	1.343861372
2726	0.681613517	0.804409024	0.74301127
2727	0.335719935	0.418583038	0.377151487
2728	0.144202253	0.167677227	0.15593974
2729	0.170174594	0.178626062	0.174400328
2730	0.115668136	0.100390519	0.108029327
2731	0.068810358	0.087189848	0.078000103
2732	0.143068028	0.17098999	0.157029009
2733	0.147693221	0.148943269	0.148318245
2734	0.426591704	0.499566309	0.463079006
2735	0.743394916	0.781465269	0.762430093
2736	0.203391748	0.130986902	0.167189325
2737	0.135543506	0.126590358	0.131066932
2738	0.16358067	0.220524079	0.192052374
2739	0.963548833	0.729634845	0.846591839
2740	0.329323597	0.285574896	0.307449246
2741	0.131577594	0.190819613	0.161198603
2742	1.075737125	1.262703592	1.169220359
2743	1.256871762	1.488222056	1.372546909
2744	0.369045951	0.358930892	0.363988422
2745	0.224480173	0.405472397	0.314976285
2746	0.004013214	0.011500058	0.007756636
2747	0.101926066	0.115388634	0.10865735
2748	0.341381551	0.362834631	0.352108091
2749	0.118660801	0.100473388	0.109567095
2750	0.255920076	0.324287127	0.290103601
2751	0.085726542	0.133811219	0.109768881
2752	0.133037991	0.147967034	0.140502512
2753	0.228143361	0.247513908	0.237828635
2754	0.072613586	0.092316291	0.082464938
2755	0.06241932	0.074834041	0.068626681
2756	0.065729075	0.087121575	0.076425325
2757	0.165267167	0.19652867	0.180897918
2758	0.133197808	0.148477386	0.140837597
2759	0.066014726	0.118767865	0.092391295
2760	0.127615348	0.127342846	0.127479097
2761	0.392506383	0.411233985	0.401870184
2762	0.040167612	0.057202274	0.048684943
2763	0.507357851	0.578527379	0.542942615
2764	0.435846469	0.541035976	0.488441223
2765	0.122727686	0.140850133	0.13178891
2766	0.004310986	0.021509906	0.012910446
2767	0.009371373	0.011285558	0.010328465
2768	0.046250207	0.074623807	0.060437007
2769	0.022186944	0.021630066	0.021908505
2770	0.071094776	0.073889025	0.0724919
2771	0.559971019	0.729174131	0.644572575
2772	0.118694372	0.119458605	0.119076489
2773	0.125835182	0.168425818	0.1471305
2774	0.398871074	0.312481346	0.35567621
2775	0.583030685	0.450135663	0.516583174
2776	0.222691606	0.201102162	0.211896884
2777	0.802790772	0.723314455	0.763052614
2778	0.079416626	0.094021281	0.086718954
2779	0.026864262	0.036429053	0.031646658
⋮			
4253			

CHAPTER 4

COMPLEX AND SIMPLE TRANSLATIONAL READTHROUGH SIGNALS IN PEA ENATION MOSAIC VIRUS 1 AND POTATO LEAFROLL VIRUS, RESPECTIVELY

PEMV1 and PLRV both produce C-terminally extended capsid proteins, termed CP-RTDs, from their sg mRNAs via programmed stop codon readthrough. Readthrough in both of these viruses was previously predicted to involve long-range RNA-RNA interactions, however experimental evidence for this occurrence and the nature of the active structures was not provided (Xu et al., 2018). The study presented in this chapter demonstrates that efficient readthrough production of CP-RTD requires formation of distinct RNA higher order structures in PLRV and PEMV1, with the former being simpler and formed by a single long-distance RNA-RNA interaction, and the latter being more complex and requiring three different long-distance RNA-RNA interactions.

This chapter is presented as a peer-reviewed journal article - "Complex and simple translational readthrough signals in pea enation mosaic virus 1 and potato leafroll virus, respectively" by Tamari Chkuaseli and K. Andrew White published in *PLOS Pathogens* (Chkuaseli and White, 2022). I conceptualized and designed the experiments for the study together with Dr. Andrew White. I performed all experiments and data analyses and wrote the first draft of the manuscript.
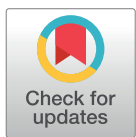
PLOS PATHOGENS

Complex and simple translational readthrough signals in pea enation mosaic virus 1 and potato leafroll virus, respectively

Tamari Chkuaseli, K. Andrew White ⓘ *

Department of Biology, York University, Toronto, Ontario, Canada

* kawhite@yorku.ca

Abstract

Different essential viral proteins are translated via programmed stop codon readthrough. Pea enation mosaic virus 1 (PEMV1) and potato leafroll virus (PLRV) are related positive-sense RNA plant viruses in the family Solemoviridae, and are type members of the Enamovirus and Polerovirus genera, respectively. Both use translational readthrough to express a C-terminally extended minor capsid protein (CP), termed CP-readthrough domain (CP-RTD), from a viral subgenomic mRNA that is transcribed during infections. Limited incorporation of CP-RTD subunits into virus particles is essential for aphid transmission, however the functional readthrough structures that mediate CP-RTD translation have not yet been defined. Through RNA solution structure probing, RNA secondary structure modeling, site-directed mutagenesis, and functional in vitro and in vivo analyses, we have investigated in detail the readthrough elements and complex structure involved in expression of CP-RTD in PEMV1, and assessed and deduced a comparatively simpler readthrough structure for PLRV. Collectively, this study has (i) generated the first higher-order RNA structural models for readthrough elements in an enamovirus and a polerovirus, (ii) revealed a stark contrast in the complexity of readthrough structures in these two related viruses, (iii) provided compelling experimental evidence for the strict requirement for long-distance RNA-RNA interactions in generating the active readthrough signals, (iv) uncovered what could be considered the most complex readthrough structure reported to date, that for PEMV1, and (v) proposed plausible assembly pathways for the formation of the elaborate PEMV1 and simple PLRV readthrough structures. These findings notably advance our understanding of this essential mode of gene expression in these agriculturally important plant viruses.

Author summary

RNA viruses use a variety of strategies to express their encoded proteins, and one mechanism used is translational readthrough of stop codons. Some viruses in the family Solemoviridae use this expression strategy to make a C-terminally extended minor capsid protein that, when incorporated into virus particles, mediates plant-to-plant transmission by aphids. In this study we investigated the RNA sequences and structures that facilitate

127

readthrough production of the minor capsid proteins of pea enation mosaic virus 1 (PEMV1; genus Enamovirus) and potato leafroll virus (PLRV; genus Polerovirus). A combination of structural and functional analyses allowed for the identification of RNA elements that contribute to the readthrough process. In both cases, the RNA structures that facilitate readthrough were formed via long-distance RNA-RNA interactions, with the active readthrough structure for PEMV1 being significantly more complex than that for PLRV. These results provide the first structural models for readthrough signals in these virus genera, and provide new insights into how these critical RNA structures assemble. This information significantly advances our understanding of this important gene expression strategy that is employed by these agriculturally important viruses.

Introduction

Programmed stop codon readthrough is an alternative protein expression strategy utilized by different viruses. The readthrough process involves the decoding of a stop codon as a sense codon by near-cognate tRNAs, which then allows ribosomes to continue translating in the original reading frame. The resulting C-terminally extended readthrough product is functionally distinct from its pre-readthrough protein, which expands the coding capacity of viral mRNAs. Positive sense RNA viruses represent the largest proportion of the viruses that employ readthrough [1–3]. Some of these viruses, like alphaviruses, alphacarmoviruses, tombusviruses, betanecroviruses, tobamoviruses, etc., [4–9] express their RNA-dependent RNA polymerase (RdRp) via readthrough. Other viruses, such as members of Benyvirus, Furovirus, Pomovirus, Luteovirus, Polerovirus and Enamovirus genera [1,10–14] translate a C-terminally extended minor capsid protein (CP) by means of readthrough.

Readthrough efficiency is promoted and fine-tuned by downstream regulatory RNA sequences and structures [1–3] that are either positioned 3′-proximally to readthrough sites [9] or involve both proximal and distal RNA elements that are united by RNA-RNA interactions [5–8]. Some distal readthrough elements (DRTEs) are separated from their complementary proximal readthrough elements (PRTEs) by shorter distances (~50 to 150 nt), as observed for RdRp production in alphaviruses and predicted for furoviruses, tobraviruses, pecluviruses, and pomoviruses [5]. However, other DRTEs involved in RdRp translation are positioned several thousand nucleotides downstream from their cognate PRTEs, as in viruses in the Tombusvirus, Betanecrovirus, and Alphacarmovirus genera [6–8].

Members of Luteovirus (Tombusviridae), Polerovirus (Solemoviridae) and Enamovirus (Solemoviridae) genera, all of which are related by similarities in their CPs and corresponding readthrough products, have been proposed to regulate CP readthrough from their subgenomic (sg) mRNAs via long-distance RNA-RNA interactions (spanning ~700 nt); based on observed complementarity between corresponding PRTEs and DRTEs [12,13]. Viruses belonging to these three genera are economically important pathogens causing major crop losses of pea, bean, potato and other food crops around the world [15–18]. These viruses require aphid vectors for plant-to-plant transmission and readthrough of the CP stop codon generates a C-terminally extended minor CP, referred to as CP-readthrough domain (CP-RTD), that plays a key role in aphid transmission [14,19–24]. CP-RTD also facilitates other viral events, such as systemic movement in infected plants, persistence of virions in aphid vectors, tissue tropism, and phloem loading [20,21,23,25–31].

To date, only limited information is available about the regulation of readthrough-mediated translation of CP-RTD from sg mRNAs in Luteovirus, Polerovirus and Enamovirus genera

128

[12,13]. The luteovirus barley yellow dwarf virus (BYDV) was the first virus shown to require RNA sequences both proximal (i.e. PRTE) and distal (i.e. DRTE) from its CP stop codon for efficient readthrough [12] and, subsequently, the polerovirus potato leaf roll virus (PLRV) was shown to have similar requirements [13]. Despite the existence of complementarity between the PRTEs and DRTEs in these and other related viruses, no experimental evidence confirming the functional importance of such long-distance interactions has been reported, nor have there been any studies investigating the structural nature of the functional readthrough-promoting RNA signals.

Pea enation mosaic virus (PEMV1) is an enamovirus with a 5.7 kb-long plus-strand RNA genome that contains a 5′ viral protein genome-linked (VPg) and no 3′ poly(A) tail (**Fig 1A**) [32]. The enamovirus protein coding scheme is very similar to that of poleroviruses, like PLRV, except that poleroviruses encode a few additional smaller proteins [18]. The PEMV1 genome codes for three 5′-proximally encoded non-structural proteins, p0, p1, and p1/2 (RdRp, expressed via frameshifting), all of which are translated from the genome (**Fig 1A**) [32]. Structural proteins, CP and its readthrough product CP-RTD, are expressed from a 1.8 kb-long sg mRNA that is transcribed during infections (**Fig 1B**). Similar to poleroviruses and luteoviruses, the sg mRNAs of enamoviruses were predicted to contain complementary readthrough-promoting PRTEs and DRTEs [13]. However, although the interactions proposed for PEMV1 were in the right general areas of the RTD (**Fig 1B and 1C, red circles**), the base-pairing partner sequences that were suggested were not correct [13], as revealed by results presented herein.

In this study we investigated, in detail, the CP readthrough signal in PEMV1's sg mRNA and determined that it adopts an elaborate RNA structure, which contrasts the simple readthrough structure that was deduced for PLRV. We also confirmed the functional requirement for long-distance RNA-RNA interactions between the PRTEs and DRTEs in both PEMV1 and PLRV. Lastly, we propose putative folding pathways for formation of the complex PEMV1 and simple PLRV readthrough structures.

Results

Secondary structure analysis of the PRTE and DRTE in PEMV1 sg mRNA

Prior to investigating the functional involvement of the PRTE and DRTE in regulating PEMV1 CP stop codon readthrough (**Fig 1B and 1C**), the local RNA secondary structures in these regions were analyzed via selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) [33]. SHAPE was conducted on in vitro synthesized transcripts of the full-length wild type (wt) PEMV1 sg mRNA and the reactivity data gathered (SHAPE reactivity correlates with flexibility of the corresponding nucleotide) were integrated into the *RNAStructure* folding program to predict the most probable secondary structure [34]. The results for the PRTE region revealed the presence of two small RNA stem-loop (SL) structures, termed SL1 and SL3, located downstream of the CP stop codon (**Fig 2A, left**). Interestingly, an alternative fold was also possible for the sequence between the stop codon and SL3, in which SL1 is replaced by a mutually-exclusive SL2 (**Fig 2A, right**). Notably, SL2 contains four cytidylate residues (red) in its terminal loop (**Fig 2A, right**), which are complementary to four guanylate residues (red) in the terminal loop of the SHAPE-predicted SL4 in the DRTE (**Fig 2B**). Consequently, the complementary terminal loops of SL2 and SL4 could potentially engage in a kissing-loop base-pairing interaction and nucleate contact between the PRTE and DRTE. Subsequent to this initial interaction, additional regions of complementarity, such as the identified orange-highlighted segments, would then be able to pair (**Fig 2A and 2B**).

The sequence in the PRTE between the CP stop codon and SL3 is highly conserved in enamoviruses. However, one genus member, citrus vein enation virus (CVEV), was found to
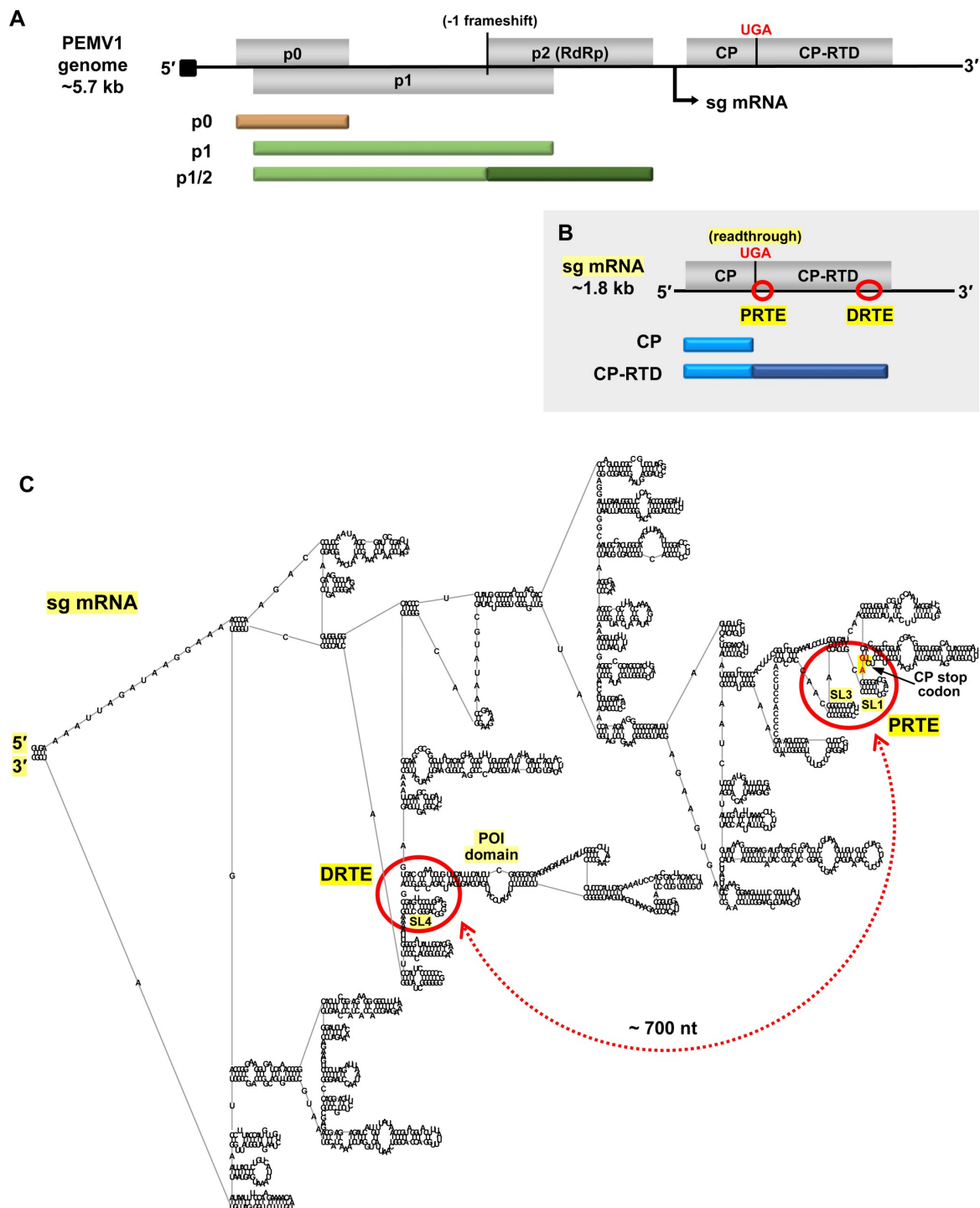
Fig 1. Organization of PEMV1 genome and subgenomic mRNA. (A) Linear representation of PEMV1 genome showing encoded ORFs (grey boxes) for p0, p1, p2, coat protein (CP) and coat protein-readthrough domain (CP-RTD). Proteins translated from the genome are shown beneath it as tan and green bars. P1/2 RdRp protein is expressed via programmed -1 frameshifting within the p1 ORF. Black arrow beneath the genome indicates the

transcription initiation site for the subgenomic (sg) mRNA. The black square at the 5′-end of the genome represents the VPg. **(B)** PEMV1 sg mRNA encoding CP and CP-RTD. Corresponding translation products are indicated below as blue bars. CP-RTD is expressed via programmed readthrough of the CP UGA stop codon. Relative positions of the proposed readthrough-regulating proximal readthrough element (PRTE) and distal readthrough element (DRTE) are shown as red circles. **(C)** RNA secondary structure model of full-length PEMV1 sg mRNA, as predicted by *RNAStructure* using default settings [34] and rendered using *RNA2Drawer* [57]. Labelled are the 5′ and 3′ ends, PRTE, DRTE, CP stop codon, SL1, SL3, SL4 and the pink-orange intervening (POI) domain. The red circles on the folded structure correspond to the regions circled on the linear sg mRNA in panel B. The proposed long-distance RNA-mediated interaction between PRTE and DRTE is indicated by a red double-headed arrow, and spans approximately 700 nt.

differ significantly in this region [35]. In CVEV, a SL2 equivalent with a tandem base pair covariation (boxed) in the center of its stem was predicted, however a corresponding SL1 could not be identified (**Fig 2C**). Additionally, CVEV has a comparable, but distinct, SL4 in its DRTE (described in a later section) that contains a complementary terminal loop sequence for CCCCA (red) in SL2 (**Fig 2C**). These comparative observations support the existence and proposed relevance of PEMV1's SL2 in mediating the initial union of PRTE and DRTE via a SL2/SL4 kissing-loop interaction.

Functional analysis of the CCCC/GGGG (red) interaction in PEMV1 sg mRNA

A wheat germ extract (wge) in vitro translation system was employed to assess modulation of PEMV1 CP readthrough by the identified RNA elements, starting with the red partner sequences (**Fig 3A**). To accurately assign the identity of translational products, the wt sg
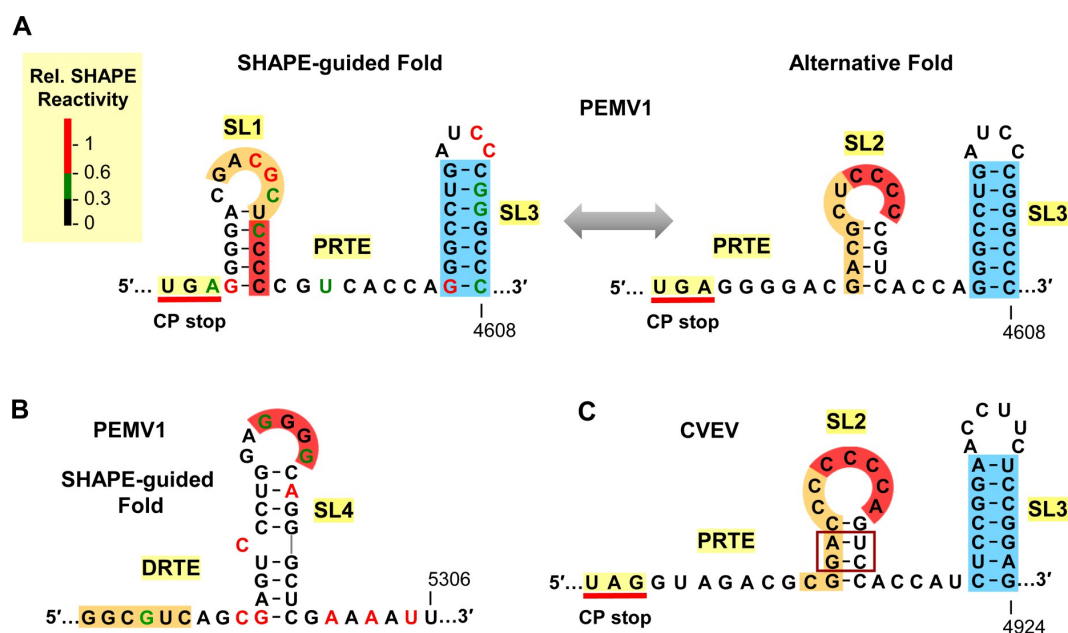
Fig 2. SHAPE-guided RNA secondary structures of PRTE and DRTE. (A) SHAPE-guided and alternative folds for the PRTE in PEMV1. Relative SHAPE reactivities of individual nucleotides are colour-coded (see key) in the SHAPE-guided fold (left). The CP stop codon, SL1, and SL3 are labeled. An alternative fold in which SL2 forms is shown on the right, with a double-headed arrow depicting the proposed interconversion between the two conformations. The stem of SL3 is highlighted in blue, while orange and red highlights denote sequences that have complementary partner segments in the DRTE, which is shown in panel B. **(B)** SHAPE-guided fold of DRTE in PEMV1. Nucleotide reactivities are colour-coded and segments complementary to red and orange RNA segments in the PRTE are indicated. **(C)** SL2 and SL3 PRTE equivalents predicted in citrus vein enation virus (CVEV, NC_021564), with corresponding orange, red, and blue segments highlighted.
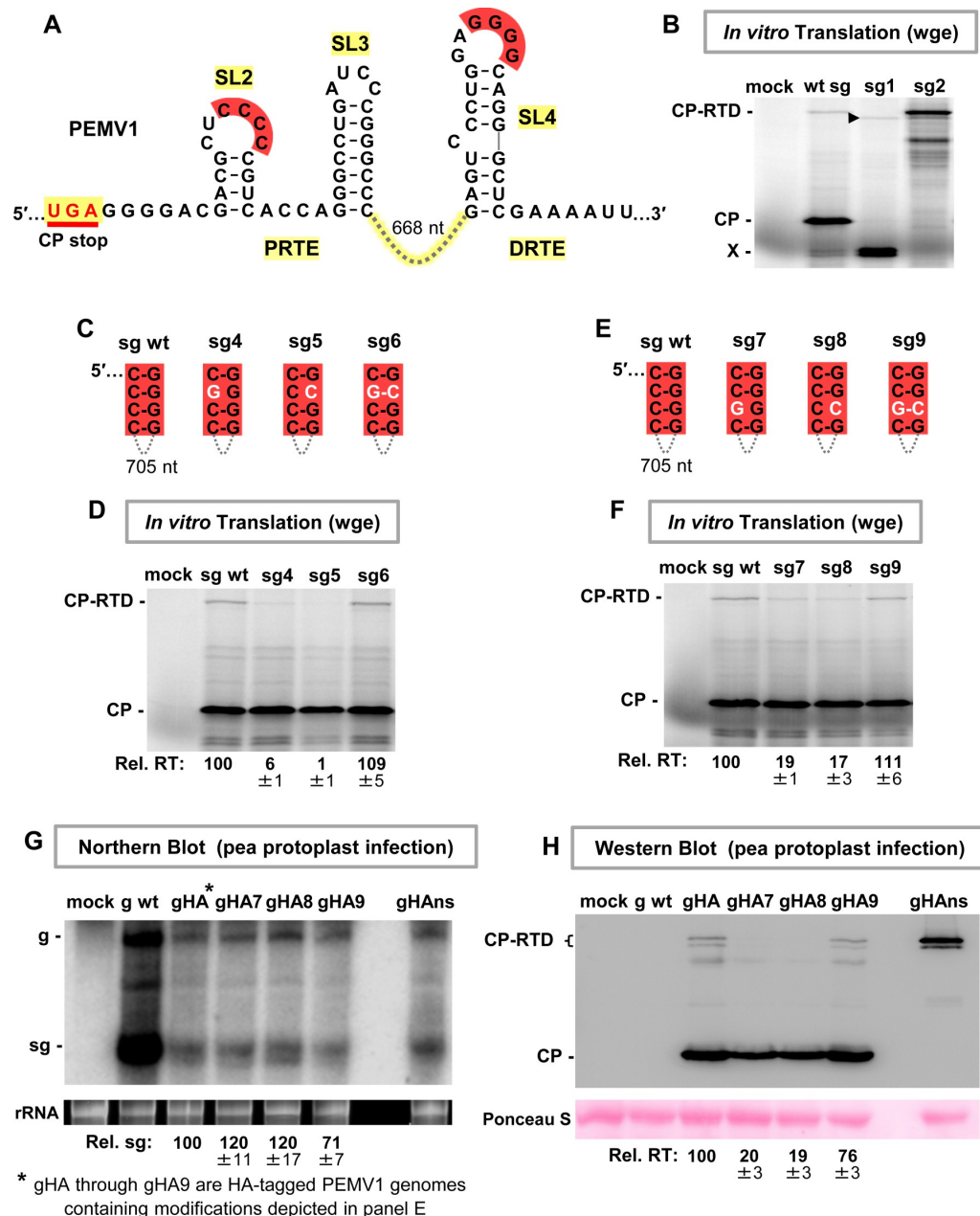
Fig 3. Assessing the red long-distance RNA-RNA interaction. (A) Secondary structures for the PRTE alternative fold and DRTE in PEMV1. The intervening 668 nucleotides between SL3 and SL4 are depicted by a connecting dashed line. **(B)** Wheat germ extract (wge) in vitro translation assay testing wt and mutant PEMV1 sg mRNAs. In mutants sg1 and sg2, the CP start codon and the CP stop codon, respectively, were inactivated (AUG → CAG and UGA → GGA). The sg mRNAs tested are indicted above each lane and the identities of the translated viral proteins are indicated on the left. The X-designated doublet band likely represents translation initiation at internal start codons in the CP ORF, and their probable readthrough products are indicated by the arrowhead. **(C)** and **(E)** Compensatory mutations introduced in the sg mRNA to test the red interaction. Nucleotide substitutions are shown in white. **(D)** and **(F)** In vitro translation analyses of the sg mRNAs shown in panels C and E, respectively. Average relative readthrough (Rel. RT) levels (±SE) calculated from three independent trials are shown below each lane. **(G)** Northern blot analysis of total nucleic acids isolated from pea protoplasts transfected with wt and HA-tagged mutant PEMV1 genomes. gHA, gHA7, gHA8, gHA9 and gHAns each contain a triple HA tag inserted 6 amino acids from the CP N-terminus. Tagged genomic mutants gHA7, gHA8 and gHA9

contain the same compensatory mutations as shown in panel E, and genomic mutant gHAns has the same CP stop codon knockout substitution as mutant sg2 in panel B. Substitutions in the DRTE in gHA8 and gHA9 lead to an arginine to serine amino acid change in CP-RTD. Positions of the genome (g) and sg mRNA (sg) are shown on the left side of the blot. Average sg levels (±SE) were calculated from three independent trials and are displayed below each lane. An ethidium bromide-stained rRNA loading control is shown below the Northern blot. (G) Western blot analysis of total proteins extracted from the same pea protoplast infections as in panel G. Identities of the detected viral proteins are indicated on the left and averaged Rel. RT levels (±SE) from three independent trials are shown under each lane. Ponceau S-stained loading control of the blot is shown below.
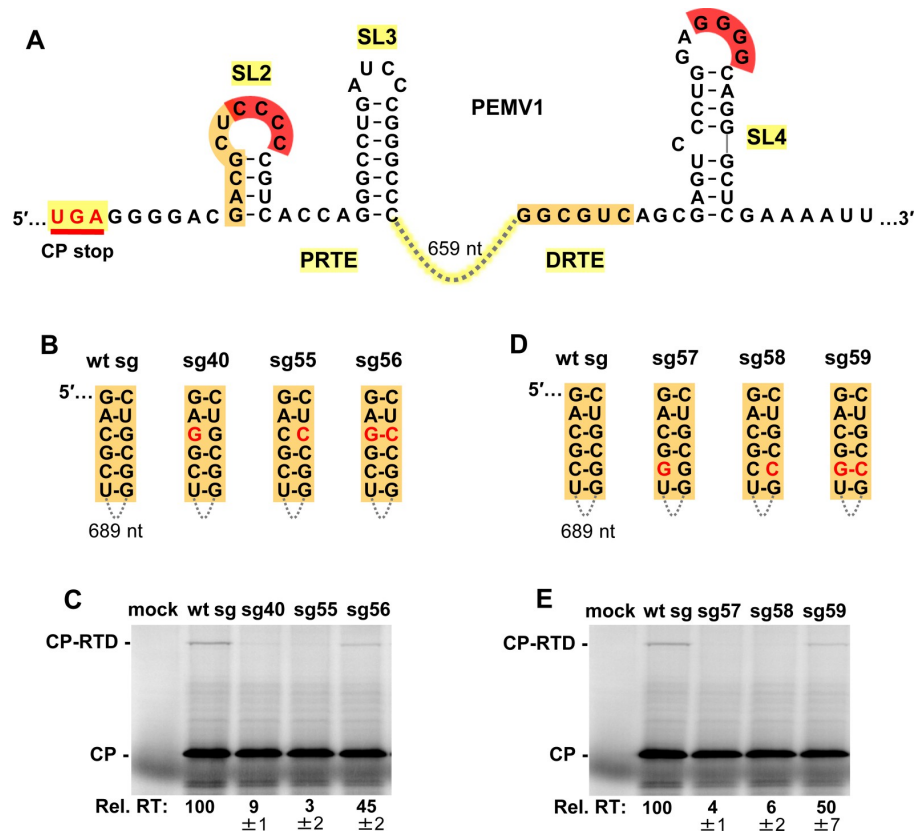
mRNA, a sg mRNA with the CP start codon inactivated by changing AUG to CAG (mutant sg1), and a sg mRNA with the UGA CP stop codon altered to glycine-coding GGA (mutant sg2) were tested. The results showed that both CP and CP-RTD were produced from wt sg mRNA, as confirmed by the absence of the former in the CP AUG knockout (sg1) and the increased levels of the latter in the CP UGA knockout (sg2) (Fig 3B). Two smaller minor products were also generated from wt sg mRNA (Fig 3B, denoted by X). These bands likely represent translational initiation at inframe downstream AUGs in the CP open reading frame (ORF), because in mutant sg1 (CP AUG knockout) their accumulation increased and a corresponding smaller readthrough product(s), denoted by an arrowhead, appeared (Fig 3B).
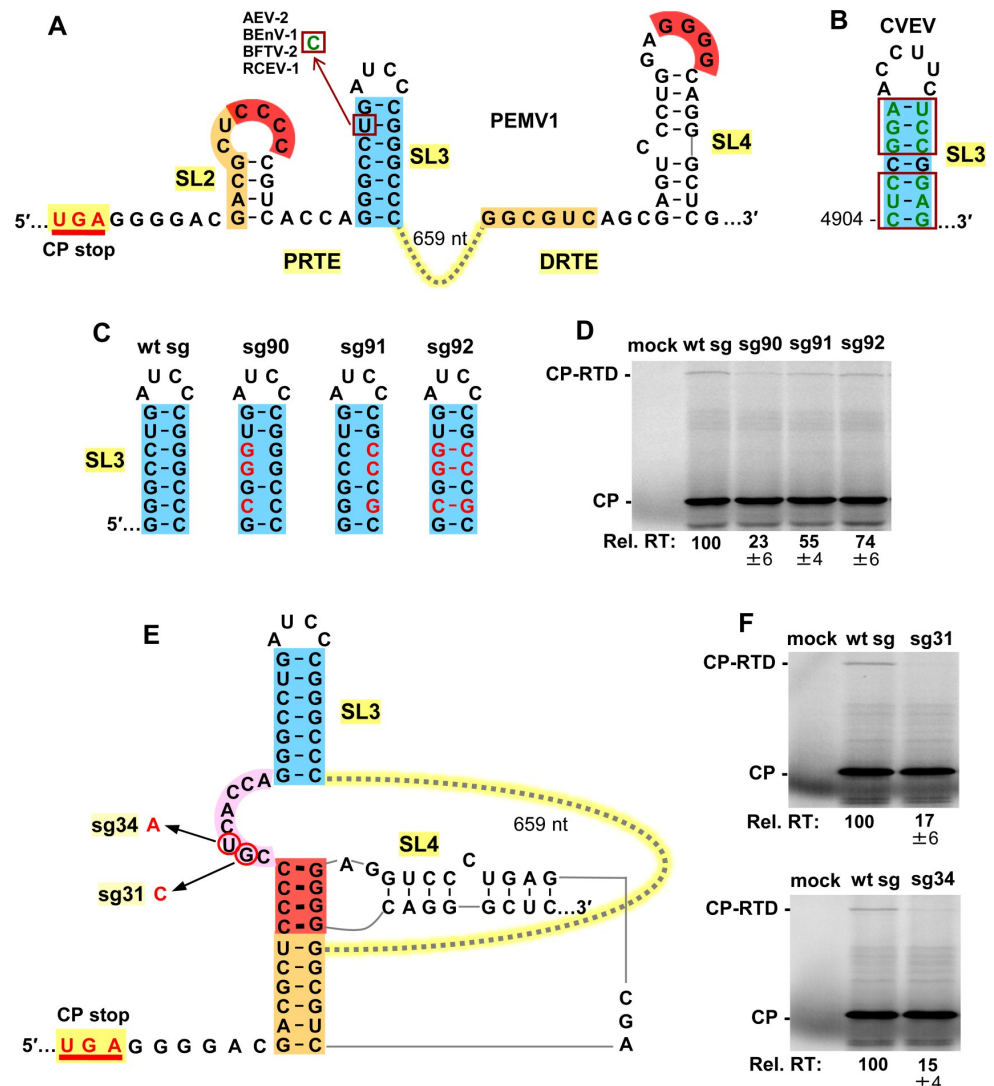
Having established that the wge system yielded readily detectable amounts of CP-RTD, the proposed red CCCC/GGGG interaction between SL2 and SL4 was investigated (Fig 3A). Sets of compensatory substitutions were introduced individually at two different nucleotide positions in the complementary red sequences (Fig 3C and 3E). In vitro translation analysis of corresponding wt and mutant sg mRNAs showed that the relative readthrough levels correlated with base-pairing capacity between the terminal loop sequences (Fig 3D and 3F). That is, when base pairing was disrupted, relative readthrough levels dropped below 20% that of wt (Fig 3D, mutants sg4 and sg5; Fig 3F, mutants sg7 and sg8), while restoration of base pairing in compensatory mutants sg6 and sg9 rescued relative readthrough to wt levels (Fig 3D and 3F).

To determine if the results obtained from in vitro translation assays reflected activity in corresponding in vivo viral infections, an N-terminal triple-HA tag was introduced into the CP ORF in the full-length PEMV1 genome (creating gHA), thus allowing for immunological detection of CP and CP-RTD. The same red sequence-targeting mutations in sg mRNA mutants sg7, sg8 and sg9 (Fig 3E) were then introduced into the gHA genomic context, creating gHA7, gHA8, and gHA9, and the tagged viral genomes were transfected into pea protoplasts. Infections also included gHAns as a control, which was a mutant genome in which the CP stop codon was converted to a glycine sense codon (UGA → GGA). Northern blot analysis revealed that HA-tagged genomes and sg mRNAs accumulated to lower levels than their wt counterpart (Fig 3G), likely due to the tag interfering with virus packaging and/or other intracellular viral processes. However, the accumulation levels of the sg mRNAs in the tagged virus infections were reasonably comparable, and examination of corresponding relative readthrough levels revealed results that were consistent with those from in vitro assays (compare Fig 3H and 3F). Combined, these in vitro and in vivo findings provide compelling evidence for the requirement of the red CCCC/GGGG interaction for optimal readthrough and validated use of the wge system for further analysis.

Additional RNA elements are required for efficient readthrough

Formation of the long-distance red CCCC/GGGG interaction would position the identified complementary orange sequences in the PRTE and DRTE in close proximity (Fig 4A). In vitro translation of compensatory mutants targeting two different base pairs in the orange partner sequences in the sg mRNA (Fig 4B and 4D) supported functional base-pairing (Fig 4C and

133

Fig 4. Assessing the orange long-distance RNA-RNA interaction. (A) RNA secondary structures of PRTE and DRTE in PEMV1. **(B)** and **(D)** Compensatory mutations introduced in the sg mRNA to test the orange interaction. Nucleotide substitutions are shown in red. **(C)** and **(E)** In vitro translation analyses of the wt and mutant sg mRNAs shown in panels B and D, respectively. Averaged Rel. RT levels (±SE) collected from three independent trials are shown below each lane.

4E). Notably, in both cases, only partial rescue of readthrough (~45–50% of wt) was observed for compensatory mutants sg56 and sg59 (**Fig 4C and 4E**). This lower level of rescue could be related to the substitutions in the orange sequence in the PRTE interfering with presentation of the red CCCC in the terminal loop of SL2, because the orange sequence forms the 5′ half of the stem in SL2 (**Fig 4A**). Regardless, the obtained results support an important role for the orange interaction in promoting readthrough efficiency.

The importance of SL3 (blue), localized within the PRTE region, was also assessed due to its proximity to the other functionally relevant PRTE sequences (i.e. orange and red) and its conservation among enamoviruses (**Fig 5A**). Regarding the latter point, four enamoviruses contain a U-to-C substitution in the stem of their SL3s that maintains pairing (**Fig 5A**), while CVEV contains a SL3 with multiple covariant base pairs (**Fig 5B, boxes**). Compensatory mutations in sg mRNAs were designed to simultaneously target three base pairs in the GC-rich stem of PEMV's SL3 (**Fig 5C**) and results from wge assays indicated that stability of the stem contributes to CP stop codon readthrough (**Fig 5D**), albeit to a lesser degree than the associated long-distance interactions.

134

Fig 5. Assessing the local SL3 (blue) in the PRTE. (A) RNA secondary structures of PRTE and DRTE in PEMV1. A nucleotide mono-variation (U to C) in the SL3 of four enamoviruses that maintains base pairing is shown (boxed). Alfalfa enamovirus-2 (AEV-2, KY985463.1), bean enamovirus-1 (BEnV-1, MZ361924), birdsfoot trefoil virus-2 (BFTV-2, NC_048296) and red clover enamovirus-1 (RCEV-1, MN412742). **(B)** SL3 of citrus vein enation virus (CVEV, NC_021564). Covariations in the SL3 stem that maintain pairing are boxed. **(C)** Compensatory mutations introduced into SL3, with substitutions depicted in red. **(D)** Results of in vitro translation reactions for the sg mRNAs shown in panel C. Averaged Rel. RT levels (±SE) calculated from three independent trials are shown below each lane. **(E)** Proposed RNA secondary structure when the red and orange long-distance interactions between the PRTE and DRTE occur. The linker sequence between red and blue helices is highlighted in pink, with corresponding substitutions in mutant sg mRNAs circled and indicated in red. **(F)** Results of in vitro translation reactions for mutant sg mRNAs shown in panel E. Averaged Rel. RT levels (±SE) calculated from three independent trials are shown below each lane.

Formation of the two long-distance RNA-RNA interactions (red and orange) between the PRTE and DRTE would lead to an RNA structure with a large intervening sequence (659 nt) (**Fig 5E**). In this structure, the red helix would likely coaxially stack on the orange helix below, with the red helix separated from SL3 (blue) by an 8 nt long intervening sequence (pink)

(**Fig 5E**). The location of this small linker sequence between two functionally important structures suggested that it too could be important for readthrough. Consequently, two separate single nucleotide substitutions were introduced into the intervening pink sequence (**Fig 5E**). Results from translational assays revealed that both substitutions had notable detrimental effects on relative readthrough levels (**Fig 5F**), confirming an important role for the pink linker sequence.

A third long-distance RNA-RNA interaction is required for readthrough

Like the orange and red sequences in the PRTE, we reasoned that the pink sequence (**Fig 6A, top left**) could also function by pairing with a complementary sequence. Potential base-pairing partner sequences for the PRTE's pink segment were initially sought close to the red and orange sequences in the DRTE. Although complementary sequences were identified nearby, none proved to be functionally relevant. A continued search ultimately identified a partially complementary 5 nt long sequence (pink) located some 170 nucleotides upstream from the orange segment in the DRTE (**Fig 6A, bottom right**). Compensatory mutagenesis of two different base pairs followed by translational analyses revealed a critical role for the pink PRTE-DRTE long-distance interaction in facilitating optimal CP readthrough (**Fig 6B–6E**). As with the orange interaction (**Fig 4**), the inability to recover full activity with restored pink pairing may be related to concurrent destabilization of the stem of SL2 and reduced presentation of the red CCCC (**Fig 6A**). Notably, although the 5 nt long pink sequence is located 170 nucleotides upstream from the orange and red in the DRTE, the intervening 170 nucleotides are predicted, in the context of the full-length wt sg mRNA (**Fig 1C**), to fold into a small RNA domain, herein termed the pink-orange intervening (POI) domain (**Fig 6A, grey shading**). Formation of the POI domain would colocalize the red, orange, and pink sub-elements of the DRTE (**Fig 6A, bottom right**), thereby facilitating their simultaneous interaction with their corresponding localized partner sequences in the PRTE. Collectively, the results show that optimal PEMV1 CP stop codon readthrough depends on three long-distance RNA-RNA interactions (red, orange, and pink) and a local stem-loop structure, SL3 (blue).

Role of DRTE's SL4 and a potential fourth long-distance interaction

Simultaneous base pairing between complementary red, orange, and pink sequences would collectively lead to the assembly of an extended quasi-continuous helix (**Fig 7A**), with the 170 nt long POI domain and a larger 482 nt long domain extending from the helical intersections. The junctions of the adjacent helices are likely stabilized via coaxial stacking, which for the blue-pink and pink-red helical joints could involve non-canonical base pairs forming above (AG, CA) and below (CC) the pink helix (**Fig 7A**). In this structure, SL4's stem could, as shown, remain paired while its loop interacts with its red partner sequence in the PRTE (**Fig 7A**). However, an alternative long-distance interaction (green) was noted in which the 5'-portion of SL4's stem could base-pair with a 6 nt complementary sequence immediately downstream from the CP stop codon (**Fig 7A, green**). Thus, the stem sequence in SL4 could first function locally in the DRTE to present the GGGG (red) sequence and subsequently participate in a fourth PRTE/DRTE (green) interaction. SL4's role in presenting GGGG (red) is strongly supported by comparative structural analysis among enamoviruses, which revealed covariation within the stem (alfalfa enamovirus-2, AEV-2; bird's-foot trefoil enamovirus, BFTV-2; and red clover enamovirus-1, RCEV-1) and alternative SL4 folds (Bean enamovirus-1, BEnV-1 and citrus vein enation virus, CVEV) (**Fig 7B**). Additionally, the analysis of PEMV1 sg mRNAs with compensatory mutations in the stem of SL4 (**Fig 7C, right**) confirmed the importance of pairing in its stem (**Fig 7D, boxed lanes**).
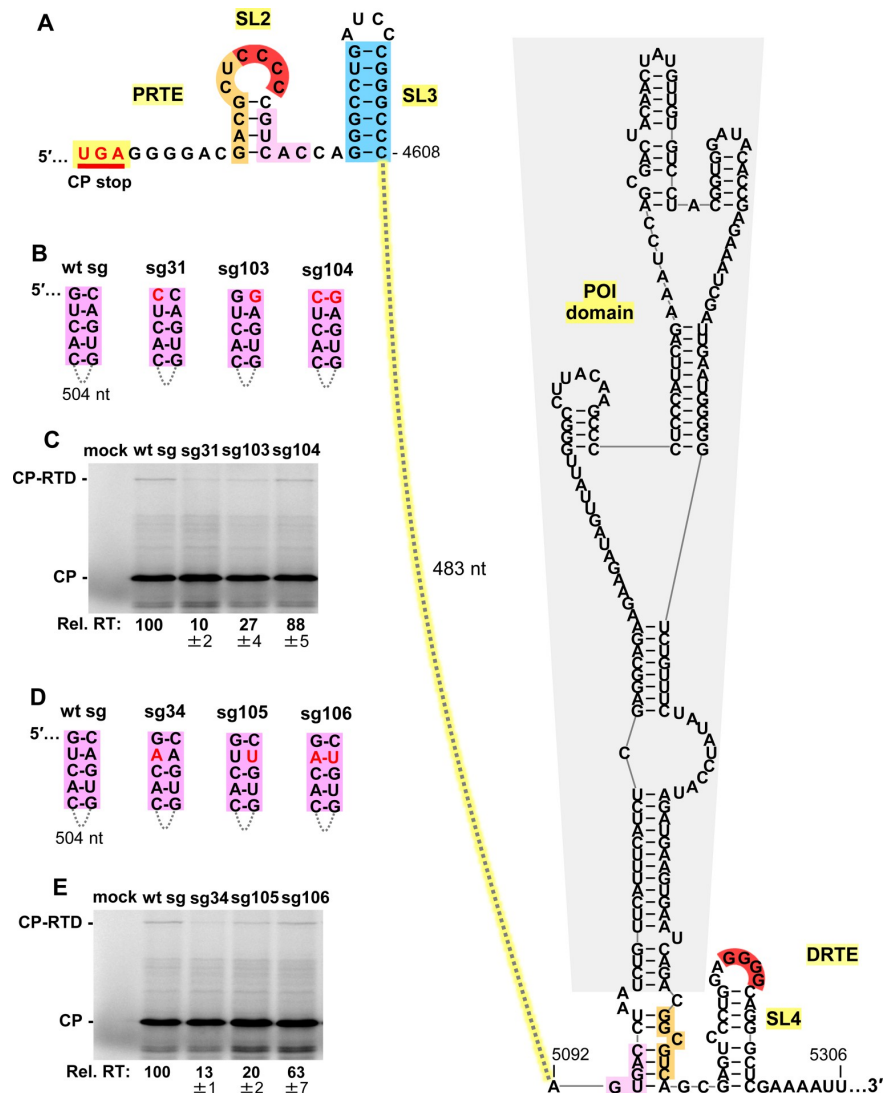
Fig 6. Assessing the pink long-distance RNA-RNA interaction. (A) RNA secondary structures of PRTE and DRTE in PEMV1, including the POI domain (grey shading). The complementary sequences highlighted in pink represent a third long-distance RNA-RNA interaction between PRTE and DRTE. **(B)** and **(D)** Compensatory mutations introduced in the sg mRNA context to test the pink interaction. Nucleotide substitutions are shown in red. **(C)** and **(E)** In vitro translation analyses of the wt and mutant sg mRNAs shown in panels B and D, respectively. Averaged Rel. RT levels (±SE) collected from three independent trials are shown below each lane.

Comparative structural analysis of the potential long-distance green interaction revealed that for most enamoviruses (except CVEV) the green sequence in the PRTE is strictly conserved, with substitutions in partner green sequences in their DRTEs that generally maintained complementarity (nucleotides in green), or generated non-canonical GA or AG pairs (nucleotides in red) (**Fig 7E**). CVEV's green sequence in its PRTE contains two substitutions (boxed) compared to that of the other enamoviruses (**Fig 7E**), and collectively maintains a potential green interaction that could include GA and AG pairs [36–38]. Accordingly, the structural

137
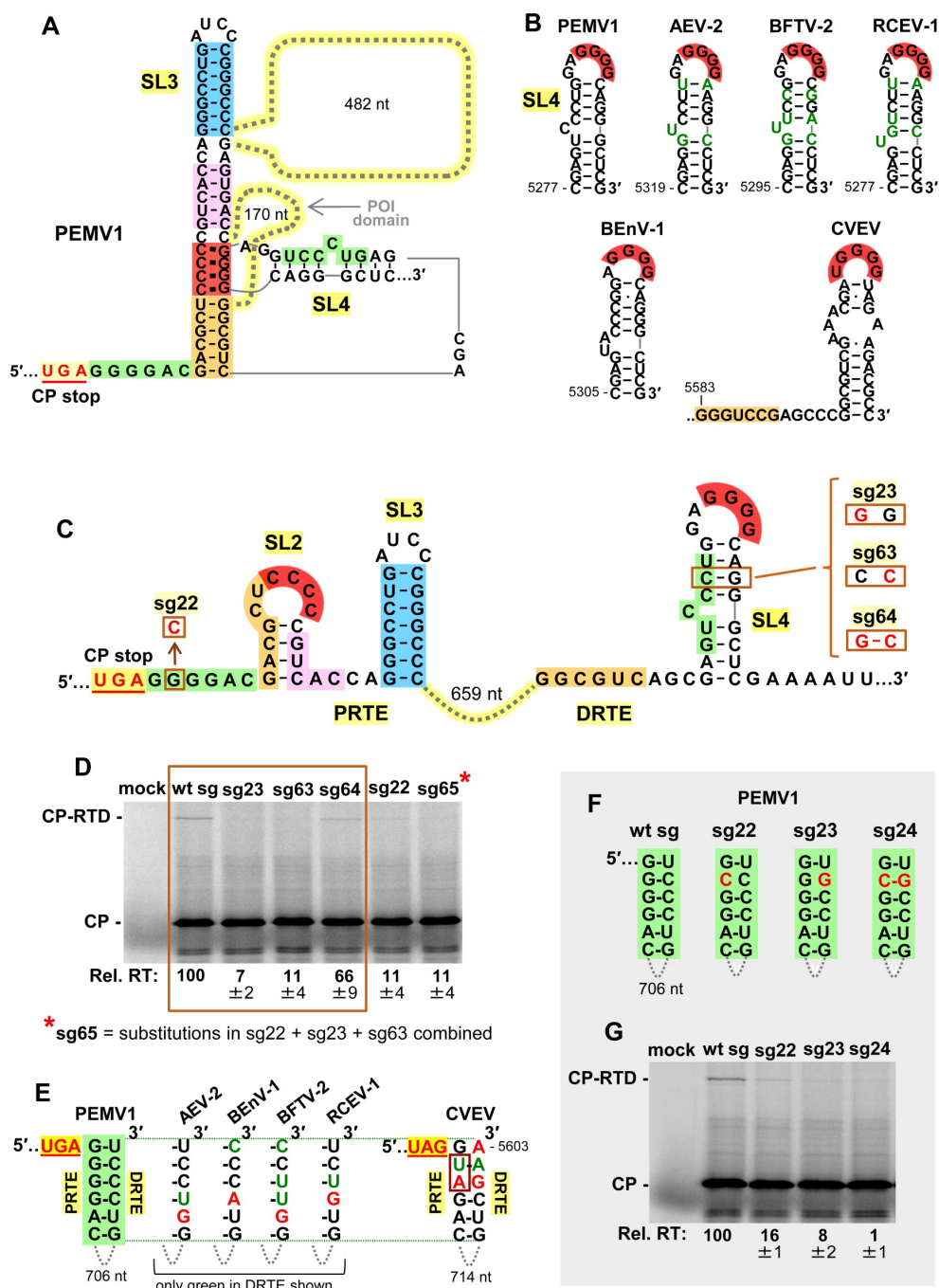
Fig 7. Assessing SL4 and a potential fourth PRTE/DRTE interaction. (A) RNA secondary structure model for the readthrough structure when red, orange, and pink complementary sequences are paired. **(B)** Conservation of SL4 in DRTEs among the members of the genus Enamovirus: PEMV1, AEV-2, BFTV-2, RCEV-1, BEnV-1, and CVEV. In the top row, mono- and co-variations that maintain base pairing of SL4 are shown in green. **(C)** RNA secondary structures of PRTE and DRTE in PEMV1, with substitutions targeting the green PRTE sequence and the stem of SL4 boxed and shown in red. Segments in the putative fourth PRTE/DRTE long-distance interaction are highlighted in green. **(D)** In vitro translation analyses of the wt and mutant sg mRNAs. The boxed area represents results from the SL4 stem compensatory mutants shown to the right in panel C.

138

comparisons suggest the possibility of a fourth functionally relevant long-distance green PRTE/DRTE interaction. Indeed, if sterically feasible, the green interaction would extend the quasi-continuous helix at the base and presumably further enhance the structure's stability (**Fig 7A**). To address this possibility, compensatory mutations were introduced into the green partner sequences in sg mRNAs and tested in translational assays (**Fig 7F and 7G**). Disruptive mutants (sg22 and sg23) notably decreased readthrough, while the restorative mutant (sg24) caused further reduction (**Fig 7G**). Moreover, combining the green interaction-restoring changes in mutant sg24 with an additional substitution (**Fig 7C, sg63**) that simultaneously restored pairing in the stem of SL4 (thereby generating sg65) did not lead to recovery of readthrough (**Fig 7D**). Therefore, nucleotide identity within the PRTE's green sequence is important, but its role may be independent of pairing with the DRTE's complementary green sequence. Accordingly, while our results corroborate an important role for the stem of SL4 in presenting the red GGGG partner sequence in the DRTE, they do not support, but also do not conclusively preclude, its involvement in a fourth long-distance green PRTE/DRTE interaction.

PLRV readthrough signal involves a long-distance RNA-RNA interaction

A previous study identified a PRTE and DRTE in PLRV that were both shown to be essential for efficient CP-RTD production [13]. Although these sequences exhibited notable complementarity, efforts to experimentally demonstrate a PRTE/DRTE pairing requirement for readthrough were unsuccessful [13]. Due to PLRV's close relationship to PEMV1, we sought to assess the necessity for such pairing and deduce the RNA structure formed. Initially, the secondary structure of wt PLRV sg mRNA was modeled using the *RNAStructure* folding program [34]. Interestingly, within the full-length sg mRNA fold, the previously identified PRTE and DRTE sequences (orange) were predicted to be paired to each other at the base of a large RNA domain (**Fig 8A**). In the prior attempt to generate informative sg mRNA compensatory mutants, several nucleotides were targeted simultaneously for substitution [13]. We reasoned that this approach likely hindered important local folding in one or both regions and/or the modified PRTE or DRTE inadvertently bound to non-cognate partner sequences elsewhere in the sg mRNA. We therefore designed our compensatory mutations as single nucleotide changes that would disrupt the bottom of the proposed structure while minimally altering the partner sequences. This strategy would both destabilize the overall structure and alter the functionally important distance between the UAG and the base of the readthrough-promoting structure (**Fig 8B and 8C**). Also, contrary to prior reports [13,39], we were able to detect synthesis of a PLRV CP-RTD product using wge assays, as confirmed by its level increasing upon knockout of the CP stop codon in sg mRNA mutant PLns (**Fig 8D**). Using the wge system to test wt and mutant PLRV sg mRNAs, we observed that both sets of compensatory mutants yielded results consistent with base pairing of the orange sequences in the PRTE and DRTE being required for optimal CP-RTD production (**Fig 8E and 8F**). These results demonstrate that the previously proposed long-distance interaction in PLRV [13] is indeed essential for optimal readthrough of its CP stop codon.
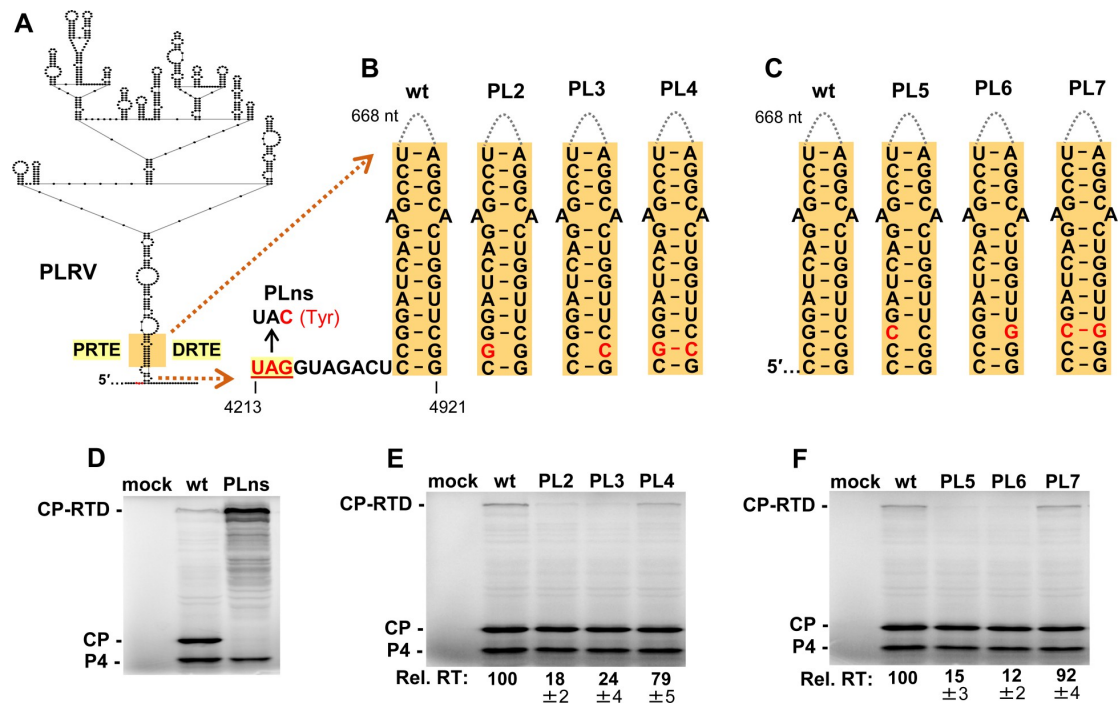
Fig 8. Assessing the PRTE/DRTE interaction in PLRV. (A) RNA secondary structure model for a central region of the PLRV sg mRNA based on *RNAStructure* [34] and rendered using *RNA2Drawer* [57]. In the structure, small black circles represent nucleotides, with those corresponding to the PLRV CP UAG stop codon shown in red. Highlighted in orange are the PRTE and DRTE sequences previously proposed to base-pair and regulate CP readthrough in PLRV [13]. **(B)** and **(C)** Compensatory mutations introduced in PLRV sg mRNA that target the orange PRTE/DRTE base-pairing interaction, and mutant PLns, in which the CP stop codon was inactivated. **(D)**, **(E)** and **(F)** In vitro translation analyses of the wt and mutant PLRV sg mRNAs shown in panels B and C. Averaged Rel. RT levels (±SE) collected from three independent trials are shown below each lane.

Discussion

Survival of PEMV1 and PLRV depends on aphid-mediated host-to-host transmission, which is conferred by their CP-RTD minor capsid proteins generated via programmed ribosome readthrough [40]. In this study we performed a detailed investigation of the regulation of CP-RTD production in PEMV1 and developed an elaborate multi-helix model for the read-through structure. In contrast, our assessment of the PLRV readthrough signal indicated a simple single-helix RNA structure. Below, different readthrough structures are discussed, the PEMV1 and PLRV readthrough structures are compared, and hypothetical models for the assembly of PEMV1 and PLRV readthrough signals are proposed.

Long-distance readthrough structures in other viruses

Programmed stop codon readthrough is commonly used by RNA viruses to produce their RdRps or minor CPs [1]. In some cases, readthrough stimulating signals are localized immediately downstream from corresponding stop codons. Murine leukemia retrovirus relies on a compact RNA pseudoknot structure situated 8-nt downstream from its gag stop codon for pol translation [41], while in tobacco mosaic virus a 6 nt-long linear sequence directly after the stop codon promotes readthrough production of its RdRp [9]. In other viruses, bipartite read-through signals, separated by intervening sequences, are employed. For example, alphaviruses
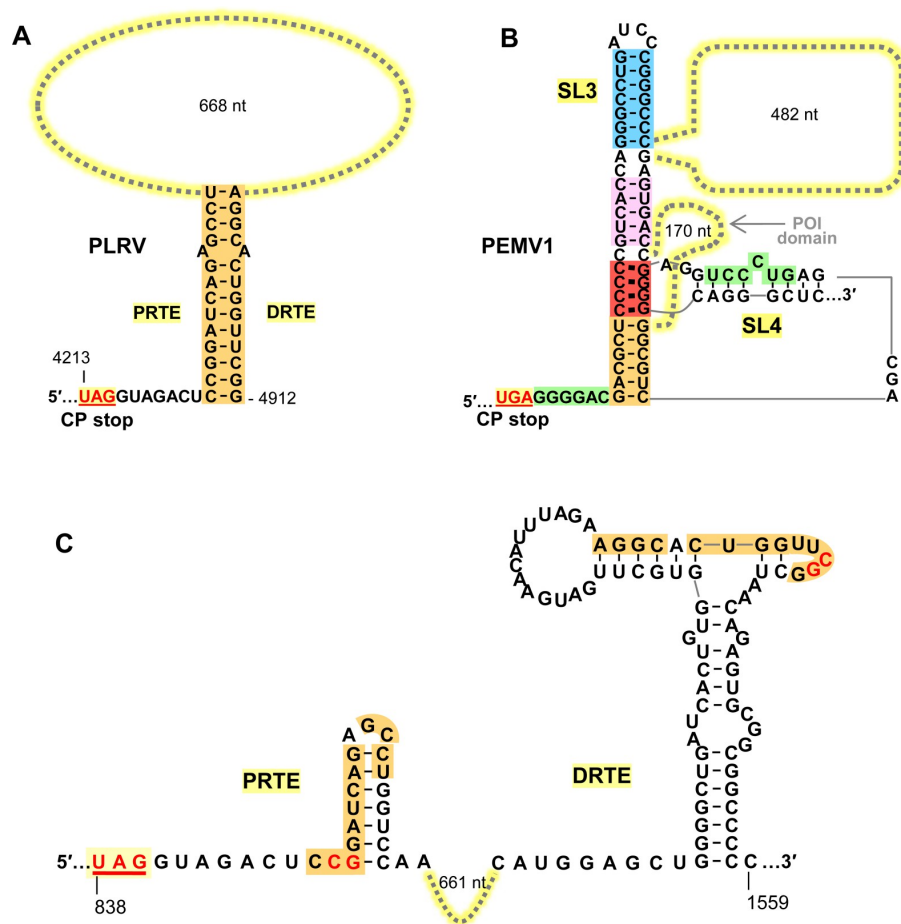
Fig 9. Comparison of PLRV and PEMV1 readthrough structures. (A) Predicted readthrough structure of PLRV showing the key orange PRTE/DRTE interaction. (B) Proposed readthrough structure for PEMV1, including local and long-distance interactions. (C) Predicted local RNA secondary structures for PRTE and DRTE [13] regions in PLRV. Nucleotides shown in red were targeted for compensatory mutational analysis in Fig 8.

https://doi.org/10.1371/journal.ppat.1010888.g009

utilize a simple helical readthrough structure, similar to that in PLRV (**Fig 9A**), for production of their RdRps [5]. However, although comparable with respect to their basic stem structures, the intervening sequences in alphaviruses are considerably shorter than that in PLRV (i.e. ~100–150 nt versus ~670 nt, respectively).

Arguably the best studied viruses employing long-distance interactions for readthrough are genera in the family Tombusviridae (Tombusvirus, Betanecrovirus and alphacarmovirus), all of which use long-distance RNA-RNA base pairing (spanning kilobases) to mediate read-through expression of their RdRps [6–8]. In contrast to readthrough in the sg mRNAs in PEMV1 and PLRV, readthrough in tombusvirids occurs in the full-length viral genomes, and with corresponding DRTEs located in their genomic 3′UTRs. This placement coincides with genomic replication elements, allowing for potential crosstalk between the two processes. For instance, the DRTE of the tombusvirus carnation Italian ringspot virus is integrated with a genome replication element in the genomic 3′UTR and, importantly, the functional structures of the DRTE and replication element are mutually-exclusive RNA conformations. This

141

overlapping arrangement acts as an RNA switch that dictates whether genomic minus-strand synthesis or translational readthrough proceeds, thereby coordinating these two opposing processes [6]. In contrast, the DRTEs for PEMV1 and PLRV are positioned centrally in the coding regions of their RTDs (**Fig 1B**). Thus, although possible, the remote locations of these DRTEs are likely not related to regulation of other sg mRNA processes. Instead, their positions are more likely the consequence of random but productive (for readthrough) initial long-distance interactions, which were maintained and further optimized.

The PRTEs of some tombusvirids can assume alternate structures or have flexible adjacent structures important for readthrough efficiency. The PRTE of the alphacarmovirus turnip crinkle virus can adopt two alternative structures, one which is nonfunctional and the other that is functional [8]. In tobacco necrosis betanecrovirus, a downstream PRTE-adjacent structure that influences readthrough efficiency has both active and inactive conformations [42]. Alternative RNA conformations such as these provide additional avenues for regulating readthrough, and illustrate the importance of considering local context and structural flexibility when investigating regulatory RNA elements. Indeed, as alluded to earlier (**Fig 2A**), alternative local conformations are also likely relevant in PEMV1's PRTE.

For both betanecroviruses and tombusviruses, in addition to their PRTE/DRTE interactions, efficient RdRp readthrough expression requires an extra long-distance RNA-RNA interaction, termed the upstream linker/downstream linker (UL/DL) interaction [6,43], which is also essential for viral genome replication [44]. Accordingly, these viruses employ two distinct long-distance interactions for readthrough, one involved in forming the readthrough structure (PRTE/DRTE) and another that serves an essential accessory role (UL/DL). Since the UL/DL interactions reside within the ~3 kb intervening sequence between the PRTE and DRTE partner sequences, it was proposed that they likely function to help unite the PRTE and DRTE [6,43]. In this regard, the possibility of intervening sequence assisting in the formation of the PRTE/DRTE interaction in PLRV is discussed in the next section.

The PEMV1 readthrough structure versus PLRV's

Enamovirus, Luteovirus, and Polerovirus genera are related based on amino acid conservation of their CP and CP-RTD [32,45]. Members of these genera are also predicted to contain bipartite readthrough regulatory signals separated by ~600 to ~800 nt [12,13]. Notably, they all have the same relative positioning of their PRTEs and DRTEs in the CP-RTD coding region [13]. This suggests that CP/CP-RTD coding and associated readthrough signal were adopted by an enamo/polerovirus common ancestor prior to its divergence into two distinct genera, while a recombination event introduced the 3′-proximal structural gene cassette into luteoviruses, which contain tombus-like polymerases [46–48]. Despite their distinct evolutionary histories, these genera have maintained commonalities in their strategies for mediating readthrough.

Of the three genera, poleroviruses and enamoviruses are most similar [16]. Yet a comparison of the prototype species, PLRV and PEMV1, revealed clear differences in their approach to inducing readthrough (**Fig 9A and 9B**). PLRV's CP ORF and those of all known poleroviruses terminate with an UAG stop codon, while PEMV1 and all known enamoviruses (except for CVEV) use UGA. Proteomic analysis of PLRV's CP-RTD revealed that the UAG is decoded ~89% of the time by tRNA^{Gln} [13]. The tRNA responsible for decoding PEMV1's UGA is currently unknown. Corresponding PRTEs and DRTEs in PLRV and PEMV do not share any noteworthy sequence identity (**Fig 9A and 9B**). Dissimilarity also extends to the predicted local RNA secondary structures at these two locations. For PLRV, prior solution structure probing and mutational analyses [13] determined that the orange DRTE sequence involved in forming the readthrough structure resides in a local stem-loop structure, with most of the

orange nucleotides paired (**Fig 9C, right**). The local structure in the PRTE region was not investigated [13], but thermodynamic predictions suggest that this segment likely includes a small RNA stem-loop that sequesters most of the PRTE's orange sequence (**Fig 9C, left**). Based on these predictions, the PRTE and DRTE regions do not adopt conformations that would effectively nucleate PLRV's orange interaction. This suggests that PLRV uses a different strategy for uniting these sequences, and secondary structure predictions of the full-length PLRV sg mRNA indicate that this could be accomplished through global folding, where PRTE and DRTE form the closing ends of a large RNA domain (**Fig 8A**). That is, the folding of subdomains within the large domain would act to bring the partner sequences together. In contrast, folding predictions for PEMV1 sg mRNA indicate that the PRTE and DRTE are located in different RNA domains (**Fig 1C**), thus a unification mechanism akin to that suggested for PLRV would be less likely. Accordingly, the differences in sequence and predicted RNA structures for PLRV and PEMV1 indicate that the former likely mediates formation of its readthrough structure primarily through the folding of an independent RNA domain, while the latter initiates readthrough structure formation by stochastic nucleation of key partner sequences (i.e. red) located in different RNA domains (see next section for details).

The proposed readthrough signal for PLRV, a contiguous helix, is relatively simple (**Fig 9A**). In comparison, the PEMV1 readthrough structure is considerably more complex, consisting of a quasi-contiguous helix stabilized by coaxial stacking at stem junctions and assembled via multiple long-distance interactions involving different regions (**Fig 9B**). Though these structures differ greatly, they are both able to direct production of the requisite amounts of CP-RTD. It is intriguing that two closely related viruses have found such radically different structural solutions for readthrough. These differences are presumably the consequence of repeated sequential sampling of distinct structural variants, resulting in maintenance of those that adequately addressed functional requirements. The net result being that these viruses have evolved via divergent pathways to give rise to secondary structures of vastly contrasting complexity. Considering these extreme examples, and the predicted variability of PRTE/DRTE interactions [13], we anticipate the existence of a range of readthrough structures with different levels of complexity within the expansive and diverse polerovirus and luteovirus genera [49].

An assembly model for PEMV1 readthrough structure

SHAPE data indicated that the default structure of the PRTE is comprised of SL1 and SL3 (**Fig 2A, left**). Importantly, although the orange sequence is predicted by SHAPE to be single stranded in the loop of SL1 (**Fig 2A**), its orange partner sequence in the DRTE is predicted to be paired (i.e. low SHAPE reactivity) and thus unavailable for pairing (**Fig 2B**). The latter interpretation is supported by the prediction that, in the context of the full-length sg mRNA, the orange sequence in the DRTE is paired with the DRTE's pink sequence (**Fig 6A, bottom right**). Accordingly, SL1 would be limited in its ability to nucleate the PRTE/DRTE interaction via an orange pairing interaction. In the alternative PRTE fold where SL2 forms and presents the red sequence in its loop, the pink and orange sequences are paired in its stem (**Fig 6A, top left**) and thus would not be available for long-distance base-pairing with partner sequences in the DRTE; which are also predicted to be paired and unavailable (**Fig 6A, bottom right**). Consequently, the predicted local structural contexts in the alternatively-folded PRTE and the DRTE would favor the red CCCC/GGGG kissing-loop interaction, and concurrently impede the orange and pink interactions (**Fig 6A**).

Based on our experimental results, we propose a theoretical model for the assembly of the PEMV1 readthrough structure (**Fig 10**). In the PRTE, the red CCCC sequence in SL1 is

144
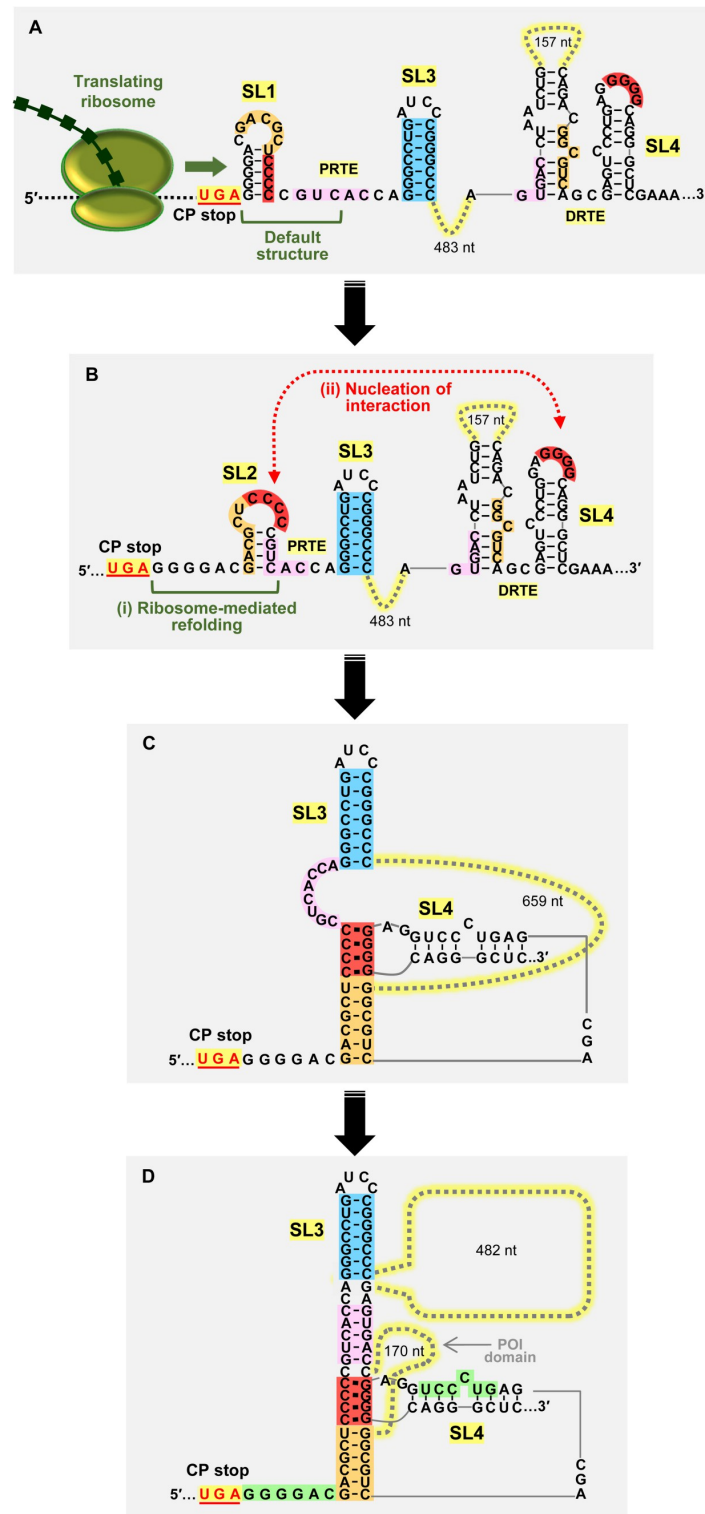
Fig 10. Model for assembly of the PEMV1 readthrough structure. (A) Default structures for PRTE and DRTE. The helicase activity of a terminating ribosome extends over SL1 and unfolds it. **(B)** (i) The sequence refolds into an alternative conformation that includes SL2. (ii) SL2 pairs with SL4 via a red sequence kissing-loop interaction and nucleates the assembly process. **(C)** Other key interactions then form, such as pairing of the orange partner sequences. **(D)** Addition of the pink interaction generates an extended helical structure, stabilized by coaxial stacking at stem junctions, that promotes efficient readthrough production of CP-RTD. A potential fourth green PRTE/DRTE interaction may also occur (not depicted), which would extend the quasi-continuous helix to the stop codon. See text for details.

initially paired in the stem of SL1 and is not available for pairing with its available red GGGG partner sequence in SL4 in the DRTE (**Fig 10A**). However, the unfolding of SL1 by the helicase activity of terminating ribosomes [50] would facilitate a SL1 to SL2 conversion (**Fig 10A**). Refolding of the PRTE sequence into the alternative CCCC-presenting SL2 (**Fig 10B, i**) would then allow for a red CCCC/GGGG kissing loop interaction with SL4 in the DRTE (**Fig 10B, ii**). In this model, SL1 acts as an attenuator of readthrough structure formation in the absence of CP translation and presumably contributes to the regulation of readthrough levels. Following the red-mediated nucleation of the interaction, additional secondary interactions, such as the orange (**Fig 10C**) or the pink would form in turn and lead to the assembly of an active readthrough structure (**Fig 10D**).

Not depicted in **Fig 10D** is the potential formation of an additional interaction involving the green partner sequences in the PRTE and DRTE. This pairing would extend the helical region at the base and could help to stabilize the structure via coaxial stacking with the orange helix (**Fig 10D**). However, this green interaction would need to be temporary and disengage during ribosome readthrough, so as to allow for the necessary spacer distance (~7–9 nt) between the stop codon and the base of the readthrough structure [1]. Either with or without the involvement of this latter interaction, the active RNA structure, postulated to be that depicted in **Fig 10D**, would then be able to efficiently trigger CP stop codon readthrough, presumably by increasing utilization of near cognate tRNAs or decreasing recruitment of release factors by an unknown mechanism [1]. Active translation of the RTD coding region would cause disruption of PRTE/DRTE interactions and their local RNA structures. Accordingly, for subsequent rounds of readthrough to occur, ribosome-mediated conversion of SL1 to SL2 would again be required to initiate assembly of an active readthrough structure (**Fig 10A and 10B**). It is also noted that the readthrough structure folding process described could also involve other protein factors, such as RNA chaperones and/or RNA helicases.

Conclusion

This study has provided the first higher-order RNA models for readthrough structures in the Enamovirus and Polerovirus genera. Compelling experimental evidence demonstrating the importance of long-distance RNA-RNA interactions in the formation of these structures was also presented. Compared to other readthrough structures, the proposed structure for PEMV1 is arguably the most elaborate readthrough signal reported to date, and its suggested folding pathway, as well as that for PLRV, provide new insights into readthrough structure assembly. Collectively, these findings significantly advance our understanding of the strategies used by viruses to mediate the production of essential readthrough proteins.

Materials and methods

cDNA preparation

Standard PCR-based site-directed mutagenesis was utilized for introducing nucleotide substitutions in different parts of full-length PEMV1 genome (gene bank: NC_003629.1) and

PEMV1 sg mRNA. Cloned cDNA of the full-length PEMV1 genome [32,45] (Kindly provided by W. Allen Miller, Iowa State University) was used to create genomic mutants, as well as wt PEMV1 sg mRNA and its mutant derivatives. All viral mutants utilized in this study were sequenced to confirm that only the intended modifications were present.

Full-length PEMV1 genome construct gHA, contained three tandem HA-tag sequences (UACCCAUACGAUGUUCCAGAUUACGCU) introduced at the N-terminal region of CP ORF (genome coordinates 4015–4095, immediately downstream from the first 6 codons of CP). gHA was then utilized as a backbone to insert PRTE-DRTE compensatory nucleotide substitutions, thereby creating gHA7, gHA8 and gHA9.

Mutants constructed to investigate CP-RTD production from the PLRV sg mRNA [51] were derived from PLRV genome cDNA (gene bank: KP090166.1) that was kindly provided by Michelle Heck (Cornell University).

Synthesis of viral RNAs in vitro

All of the PEMV1 genome and sg mRNA constructs investigated in this study contained a T7 promoter at the 5′-end of the viral sequence and a unique PstI restriction enzyme cut site at its 3′-end. PstI-linearized wt and mutant clones were treated with T4 DNA polymerase (NEB) to remove the 3′-overhang left after PstI cleavage and then were transcribed in vitro using Ampli-Cap-Max T7 High Yield Message Maker Kit (Cellscript) to create 5′-capped sg RNAs and Mes-sageMax T7 ARCA-Capped Message Transcription Kit (Cellscript) to create 5′-capped genomic RNAs, both with authentic viral 3′ ends.

The PLRV sg mRNA constructs utilized in this study contained a T7 promoter at the 5′-end of the viral sequence and a unique 3′-terminal ScaI restriction enzyme cut site. ScaI-linearized wt and mutant cDNAs were transcribed in vitro using AmpliCap-Max T7 High Yield Message Maker Kit (Cellscript) to create 5′-capped sg mRNAs with authentic viral 3′ ends.

In vitro translation assays

To test readthrough levels of CP-RTD, 0.5 pmol of 5′-capped transcripts of wt or mutant PEMV1 sg mRNAs (sub-saturating levels) were incubated in wheat germ extract (wge, Promega) in the presence of [^{35}S]-Methionine at 25˚C for 1 hr according to the manufacturer's instructions, except that the concentration of KOAc was increased to 133 mM for each reaction to optimize translation and readthrough efficiency. The viral proteins translated during the incubation were detected and quantified through 12% SDS-PAGE and phosphorimaging, respectively [52,53]. Imaging was carried out using Typhoon FLA 9500 Variable Mode Imager (GE Healthcare). QuantityOne software (BioRad) was used to quantify protein bands, from which ratios of the readthrough product CP-RTD and the pre-readthrough product CP were calculated for each tested mRNA. Percentages of the mutant ratios relative to the wt ratio were determined and used as relative readthrough levels (Rel. RT). Three independent repeats were carried out for each of the in vitro translation experiment and means with standard errors (SE) were calculated.

The same steps were followed as above for obtaining readthrough levels of CP-RTD from PLRV sg mRNAs in vitro, except that 0.4 pmol of 5′-capped transcripts (sub-saturating levels) was used per in vitro translation reaction.

Pea protoplast transfection

Pea protoplasts were isolated from 12-day old, fully expanded *Pisum sativum* leaves by first removing the lower epidermis and then incubating the remaining tissue in a cellulase mixture at 26˚C for 4 hours [54,55]. Two million protoplasts were transfected with 20 μg of 5′-capped

146

PEMV1 transcripts using polyethylene glycol (PEG 1450) and $CaCl_2$ and incubated at 22˚C for 40 hours under constant fluorescent light [55]. After the incubation, one half of the infection was used for total protein isolation and western blotting and the other half for total nucleic acid extraction and Northern blotting.

Western blotting

Total proteins were separated by 12% SDS-PAGE and transferred to membrane (Amersham Hybond P 0.45 PVDF). Ponceau S staining was carried out for visualizing total proteins and confirming equal loading and transfer prior to proceeding with blotting. HA-tagged CP and CP-RTD were detected by blotting with Anti-HA-peroxidase high affinity (3F10) rat monoclonal antibodies (Roche) at 1:2000 dilution. CP and CP-RTD bands were detected using ECL Select western blotting detection reagent (GE Healthcare) and captured through MicroChemi imager (DNR Bio-Imaging Systems). Detected viral protein bands were quantified using *QuantityOne* software. Three independent repeats of pea protoplast infections/western blotting were carried out and means with SE were calculated. Rel. RT levels were calculated as described for in vitro translation assays.

Northern blotting

Total nucleic acids were extracted from infected protoplasts, separated by agarose gel electrophoresis, and transferred to nylon membrane (Hybond-N$^+$, Amersham), after ensuring even loading via monitoring rRNA levels. Coding sense PEMV1 genome and sg mRNA were detected by blotting with nine oligonucleotide probes 5′-end-labeled with [γ -^{32}P] complementary to both the genome and sg mRNA (genome coordinates: 4004–4031, 4324–4359, 4401–4434, 4681–4708, 4749–4780, 4936–4968, 5122–5389, 5406–5439, 5679–5703). Northern blots, from three independent repeats, were captured using Typhoon FLA 9500 Variable Mode Imager and viral RNA bands were quantified using *QuantityOne* software. Sg mRNA levels of each mutant PEMV1 were calculated to generate average values with SE.

SHAPE RNA structure analysis

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) was performed and the data was used to model the RNA secondary structures of PRTE and DRTE regions in full-length PEMV1 sg mRNA, as described previously [33,6,53]. SHAPE was carried out using 1-methyl-7-nitroisatoic anhydride (1M7) that modifies flexible (*i.e.* single stranded) nucleotides. Two primers, fluorescently labeled at their 5′-ends, one complementary to a region downstream from the PRTE (genome coordinates– 4828–4857) and the other to a region downstream from DRTE (genome coordinates –5504–5533), were used for primer extension reactions following 1M7 treatment of wt PEMV1 sg mRNA. After fluorescent capillary electrophoresis of the products of primer extension, the raw data was analyzed using the *ShapeFinder* software [56] to generate relative reactivities for each nucleotide. These reactivity values were normalized against the ten highest reactivities in the pool. The SHAPE experiment was performed twice, with consistent results, and averaged values of the two repeats were used for secondary structure prediction. The *RNAStructure* web server was used [34] to combine SHAPE reactivity data (slope = 1.8 kcal/mol; intercept = -0.6 kcal/mol) with thermodynamic prediction to generate secondary structure models of PEMV1 PRTE and DRTE in the sg mRNA context. *RNA2Drawer* software was utilized to draw RNA secondary structure models depicted throughout the paper [57].

147

Acknowledgments

We thank members of our laboratory for reviewing the manuscript. We are also grateful to Allen Miller (Iowa State University) for providing the PEMV1 infectious clone and Michelle Heck (Cornell University) for providing the PLRV infectious clone.

Author Contributions

Conceptualization: Tamari Chkuaseli, K. Andrew White.

Formal analysis: Tamari Chkuaseli, K. Andrew White.

Funding acquisition: Tamari Chkuaseli, K. Andrew White.

Investigation: Tamari Chkuaseli, K. Andrew White.

Methodology: Tamari Chkuaseli, K. Andrew White.

Project administration: K. Andrew White.

Resources: K. Andrew White.

Supervision: K. Andrew White.

Writing – original draft: Tamari Chkuaseli.

Writing – review & editing: Tamari Chkuaseli, K. Andrew White.

References

1. Firth AE, Brierley I. Non-canonical translation in RNA viruses. J Gen Virol. 2012; 93: 1385–1409. https://doi.org/10.1099/vir.0.042499-0 PMID: 22535777

2. Newburn LR, White KA. Cis-acting RNA elements in positive-strand RNA plant virus genomes. Virology. 2015; 479–480:434–443. https://doi.org/10.1016/j.virol.2015.02.032 PMID: 25759098

3. Rodnina MV, Korniy N, Klimova M, Karki P, Peng B-Z, Senyushkina T, et al. Translational recoding: canonical translation mechanisms reinterpreted. Nucleic Acids Res. 2020; 48:1056–1067. https://doi.org/10.1093/nar/gkz783 PMID: 31511883

4. Li G, Rice CM. The signal for translational readthrough of a UGA codon is Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. J Virol. 1993; 67:5062–5067.

5. Firth AE, Wills NM, Gesteland RF, Atkins JF. Stimulation of stop codon readthrough: frequent presence of an extended 3′ RNA structural element. Nucleic Acids Res. 2011; 39:6679–6691. https://doi.org/10.1093/nar/gkr224 PMID: 21525127

6. Cimino PA, Nicholson BL, Wu B, Xu W, White KA. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. PLoS Pathog. 2011; 12:e1002423.

7. Newburn LR, Nicholson BL, Yosefi M, Cimino PA, White KA. Translational readthrough in Tobacco necrosis virus-D. Virology. 2014; 450–451:258–265. https://doi.org/10.1016/j.virol.2013.12.006 PMID: 24503089

8. Kuhlmann MM, Chattopadhyay M, Stupina VA, Gao F, Simon AE. An RNA element that facilitates programmed ribosomal readthrough in turnip crinkle virus adopts multiple conformations. J Virol. 2016; 90:8575–8591. https://doi.org/10.1128/JVI.01129-16 PMID: 27440887

9. Skuzeski JM, Nichols LM, Gesteland RF, Atkins JF. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. J Mol Biol. 1991; 218:365–373. https://doi.org/10.1016/0022-2836(91)90718-l PMID: 2010914

10. Reavy B, Arif M, Cowan GH, Torrance L. Association of sequences in the coat protein/readthrough domain of potato mop-top virus with transmission by Spongospora subterranean. J Gen Virol. 1998; 79:2343–2347.

11. Tamada T, Schmitt C, Saito M, Guilley H, Richards K, Jonard G. High resolution analysis of the readthrough domain of beet necrotic yellow vein virus readthrough protein: a KTERmotif is important for efficient transmission of the virus by Polymyxa betae. J Gen Virol. 1996; 77:1359–1367. https://doi.org/10.1099/0022-1317-77-7-1359 PMID: 8757975

148

12. Brown MC, Dinesh-Kumar SP, Miller WA. Local and Distant Sequences Are Required for Efficient Readthrough of the Barley Yellow Dwarf Virus PAV Coat Protein Gene Stop Codon. J Virol. 1996; 70:5884–5892. https://doi.org/10.1128/JVI.70.9.5884-5892.1996 PMID: 8709208

13. Xu Y, Ju H-J, DeBlasio S, Carino EJ, Johnson R, MacCoss MJ, et al. A stem-loop structure in potato leafroll virus open reading frame 5 (ORF5) is essential for readthrough translation of the coat protein ORF5 stop codon 700 bases upstream. J Virol. 2018; 92:e01544–17. https://doi.org/10.1128/JVI.01544-17 PMID: 29514911

14. Demler SA, Rucker-Feeney DG, Skaf JS, de Zoeten GA. Expression and suppression of circulative aphid transmission in pea enation mosaic virus. J Gen Virol. 1997; 78:511–523. https://doi.org/10.1099/0022-1317-78-3-511 PMID: 9049399

15. Mayo MA, Ziegler-Graff V. Molecular Biology of Luteoviruses. Adv Virus Res. 1996; 46:413–460. https://doi.org/10.1016/s0065-3527(08)60077-9 PMID: 8824705

16. Delfosse VC, Barrios Barón MP, Distéfano AJ. What we know about poleroviruses: Advances in understanding the functions of polerovirus proteins. Plant Pathol. 2021; 70:1047–1061.

17. de Zoeten GA, Skaf JS. Pea enation mosaic and the vagaries of a plant virus. Adv Virus Res. 2001; 57:323–350. https://doi.org/10.1016/s0065-3527(01)57007-4 PMID: 11680388

18. Sõmera M, Fargette D, Hébrard E, Sarmiento C, ICTV RC. ICTV Virus Taxonomy Profile: Solemoviridae 2021. J Gen Virol. 2021; 102:001707. https://doi.org/10.1099/jgv.0.001707 PMID: 34951396

19. Gray S, Gildow FE. Luteovirus-aphid interactions. Annu Rev Phytopathol. 2003; 41:539–566. https://doi.org/10.1146/annurev.phyto.41.012203.105815 PMID: 12730400

20. Brault V, van den Heuvel JFJM, Verbeek M, Ziegler-Graff V, Reutenauer A, Herrbach E, et al. Aphid transmission of beet western yellows luteovirus requires the minor capsid read-through protein P74. EMBO J. 1995; 14:650–659. https://doi.org/10.1002/j.1460-2075.1995.tb07043.x PMID: 7882968

21. Brault V, Mutterer J, Scheideker D, Simonis MT, Herrbach E, Richards K, et al. Effects of point mutations in the readthrough domain of the beet western yellows virus minor capsid protein on virus accumulation in planta and on transmission by Aphids. J Virol. 2000; 74:1140–1148. https://doi.org/10.1128/jvi.74.3.1140-1148.2000 PMID: 10627524

22. Filichkin SA, Lister RM, McGrath PF, Young MJ. In vivo expression and mutational analysis of the barley yellow dwarf virus readthrough gene. Virology. 1994; 205:290–299. https://doi.org/10.1006/viro.1994.1645 PMID: 7975225

23. Chay CA, Gunasinge UB, Dinesh-Kumar SP, Miiler WA, Gray SM. Aphid transmission and systemic plant infection determinants of barley yellow dwarf luteovirus-PAV are contained in the coat protein readthrough domain and 17-kDa protein, respectively. Virology. 1996; 219:57–65. https://doi.org/10.1006/viro.1996.0222 PMID: 8623554

24. Peter KA, Liang D, Palukaitis P, Gray SM. Small deletions in the potato leafroll virus readthrough protein affect particle morphology, aphid transmission, virus movement and accumulation. J Gen Virol. 2008; 89:2037–2045. https://doi.org/10.1099/vir.0.83625-0 PMID: 18632976

25. Brault V, Périgon S, Reinbold C, Erdinger M, Scheidecker D, Herrbach E, et al. The Polerovirus Minor Capsid Protein Determines Vector Specificity and Intestinal Tropism in the Aphid. J Virol. 2005; 79:9685–9693. https://doi.org/10.1128/JVI.79.15.9685-9693.2005 PMID: 16014930

26. Bruyère A, Brault V, Ziegler-Graff V, Simonis M-T, van den Heuvel JFJM, Richards K, et al. Effects of mutations in the beet western yellows virus readthrough protein on its expression and packaging and on virus accumulation, symptoms, and aphid transmission. Virology. 1997; 230:323–334. https://doi.org/10.1006/viro.1997.8476 PMID: 9143288

27. Mutterer JD, Stussi-Garaud C, Pichler P, Richards KE, Jonard G, Ziegler-Graff V. Role of the beet western yellows virus readthrough protein in virus movement in Nicotiana clevelandii. J Gen Virol. 1999; 80:2771–2778. https://doi.org/10.1099/0022-1317-80-10-2771 PMID: 10573174

28. van den Heuvel JFJM, Bruyère A, Hogenhout SA, Ziegler-Graff V, Brault V, van den Wilk F, et al. The N-terminal region of the luteovirus readthrough domain determines virus binding to buchnera GroEL and is essential for virus persistence in the aphid. J Virol. 1997; 71:7258–7265. https://doi.org/10.1128/JVI.71.10.7258-7265.1997 PMID: 9311800

29. Liu S, Sivakumar S, Wang Z, Bonning BC, Miller WA. The readthrough domain of pea enation mosaic virus coat protein is not essential for virus stability in the hemolymph of the pea aphid. Arch Virol. 2009; 154:469–479. https://doi.org/10.1007/s00705-009-0327-7 PMID: 19240978

30. Peter KA, Gildow F, Palukaitis P, Gray SM. The C terminus of the polerovirus P5 readthrough domain limits virus infection to the phloem. J Virol. 2009; 83:5419–5429. https://doi.org/10.1128/JVI.02312-08 PMID: 19297484

149

31. Xu Y, Da Silva WL, Qian Y, Gray SM. An aromatic amino acid and associated helix in the C-terminus of the potato leafroll virus minor capsid protein regulate systemic infection and symptom expression. PloS Pathog. 2018; 14:e1007451. https://doi.org/10.1371/journal.ppat.1007451 PMID: 30440046

32. Demler SA, de Zoeten GA. The nucleotide sequence and luteovirus-like nature of RNA1 of an aphid non-transmissible strain of pea enation mosaic virus. J Gen Virol. 1991; 72:1819–1834. https://doi.org/10.1099/0022-1317-72-8-1819 PMID: 1875194

33. Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. Methods. 2010; 52:150–158. https://doi.org/10.1016/j.ymeth.2010.06.007 PMID: 20554050

34. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013; 41:W471–W474. https://doi.org/10.1093/nar/gkt290 PMID: 23620284

35. Vives MC, Lelàzquez K, Pina JA, Moreno P, Guerri J, Navarro L. Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. Virology. 2013; 103:1077–1086. https://doi.org/10.1094/PHYTO-03-13-0068-R PMID: 23718835

36. SantaLucia J Jr, Kierzek R, Turner DH. Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. Biochemistry. 1990; 29:8813–9. https://doi.org/10.1021/bi00489a044 PMID: 2271557

37. Limmer S. Mismatch base pairs in RNA. Prog Nucleic Acid Res Mol Biol. 1997; 57:1–39. https://doi.org/10.1016/s0079-6603(08)60276-7 PMID: 9175429

38. Morgado CA, Svozil D, Turner DH, Šponer J. Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes. Phys Chem Chem Phys. 2012; 14:12580–91. https://doi.org/10.1039/c2cp40556c PMID: 22722325

39. Juszczuk M, Zagórski-Ostoja W, Hulanicka DM. Studies on the translation mechanism of subgenomic RNA of potato leafroll virus. Acta Biochim Pol. 1997; 44:69–78. PMID: 9241356

40. Heck M, Brault V. Targeted disruption of aphid transmission: a vision for the management of crop diseases caused by Luteoviridae members. Curr Opin Virol. 2018; 33:24–32. https://doi.org/10.1016/j.coviro.2018.07.007 PMID: 30031985

41. Wills NM, Gesteland RF, Atkins JF. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. Proc Natl Acad Sci USA. 1991; 88:6991–6995. https://doi.org/10.1073/pnas.88.16.6991 PMID: 1871115

42. Newburn LR, White KA. Atypical RNA elements modulate translational readthrough in tobacco necrosis virus D. J Virol. 2017; 91:e02443–16. https://doi.org/10.1128/JVI.02443-16 PMID: 28148800

43. Newburn LR, Wu B, White KA. Investigation of novel RNA elements in the 3′UTR of tobacco necrosis virus-D. Viruses. 2020; 12:856. https://doi.org/10.3390/v12080856 PMID: 32781505

44. Wu B, Pogany J, Na H, Nicholson BL, Nagy PD, White KA. A discontinuous RNA platform mediates RNA virus replication: building an integrated model for RNA–based regulation of viral processes. PLoS Pathog. 2009; 5: e1000323. https://doi.org/10.1371/journal.ppat.1000323 PMID: 19266082

45. Demler SA, Borkhsenious ON, Rucker DG, de Zoeten GA. Assessment of the autonomy of replicative and structural functions encoded by the luteo-phase of pea enation mosaic virus. J Gen Virol. 1994; 75:997–1007. https://doi.org/10.1099/0022-1317-75-5-997 PMID: 8176385

46. Miller WA, Liu S, Beckett R. Barley yellow dwarf virus: Luteoviridae or Tombusviridae? Mol Plant Pathol. 2002; 3:177–183. https://doi.org/10.1046/j.1364-3703.2002.00112.x PMID: 20569325

47. Gunawardene CD, Jaluba K, White KA. Conserved motifs in a tombusvirus polymerase modulate genome replication, subgenomic transcription, and amplification of defective interfering RNAs. J Virol. 2015; 89:3236–3246. https://doi.org/10.1128/JVI.03378-14 PMID: 25568204

48. Gunawardene CD, Donaldson LW, White KA. Tombusvirus polymerase: Structure and function. Virus Res. 2017; 234:74–86. https://doi.org/10.1016/j.virusres.2017.01.012 PMID: 28111194

49. LaTourette K, Holste NM, Garcia-Ruiz H. Polerovirus genomic variation. Virus Evol. 2021; 7:1–18.

50. Takyar S, Hickerson RP, Noller HF. mRNA helicase activity of the ribosome. Cell 2005; 120:49–58. https://doi.org/10.1016/j.cell.2004.11.042 PMID: 15652481

51. Miller JS, Mayo MA. The location of the 5' end of the potato leafroll luteovirus subgenomic coat protein mRNA. J Gen Virol. 1991; 72:2633–2638. https://doi.org/10.1099/0022-1317-72-11-2633 PMID: 1940859

52. Nicholson BL, White KA. Context-influenced cap-independent translation of Tombusvirus mRNAs in vitro. Virology. 2008; 380:203–212. https://doi.org/10.1016/j.virol.2008.08.003 PMID: 18775547

53. Chkuaseli T, White KA. Activation of viral transcription by stepwise largescale folding of an RNA virus genome. Nucleic Acids Res. 2020; 48:9285–9300. https://doi.org/10.1093/nar/gkaa675 PMID: 32785642

150

54. Demler SA, Rucker DG, de Zoeten GA. The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. J Gen Virol. 1993; 74:1–14. https://doi.org/10.1099/0022-1317-74-1-1 PMID: 8423443

55. White KA, Morris TJ. Nonhomologous RNA recombination in tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions. J Virol. 1994; 68: 14–24. https://doi.org/10.1128/JVI.68.1.14-24.1994 PMID: 8254723

56. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 2008; 14:1979–1990. https://doi.org/10.1261/rna.1166808 PMID: 18772246

57. Johnson PZ, Kasprzak WP, Shapiro BA, Simon AE. RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. RNA Biol. 2019; 16:1667–1671. https://doi.org/10.1080/15476286.2019.1659081 PMID: 31441369

151

CHAPTER 5
DISCUSSION AND FUTURE DIRECTIONS

5.1 Summary of key findings

Subgenomic mRNA production provides plus-strand RNA viruses with the capacity to express downstream ORFs that would otherwise be inaccessible for translation in the context of their polycistronic genomes. Additionally, by regulating sg mRNA transcription and translation processes, these viruses can precisely control the timing and quantity of sg mRNA-encoded viral proteins to maximize the success of their infections. This dissertation focused on determining RNA sequences and structures that regulate (i) sg mRNA transcription in the tombusvirids CIRV and PEMV2; and (ii) readthrough-mediated translation of a C-terminally extended capsid protein from the sg mRNA of PEMV1, PEMV2's assistor virus. The findings of this study demonstrate that in all three cases, higher-order RNA structures mediate activation of the above-mentioned viral processes. For CIRV and PEMV1, the regulatory RNA structures form via multiple long-distance intragenomic RNA-RNA interactions, whereas for PEMV2, an intergenomic kissing-loop RNA-RNA interaction between two PEMV2 genomes is involved. The major findings for each of the three viruses are summarized below:

1. Investigation of the attenuation structure for CIRV sg mRNA1 transcription revealed it to be complex, involving at least six critical long-distance RNA-RNA interactions that assemble sequentially to form an active higher-order RNA configuration. The assembly process involved two major steps: (i) formation of a binding pocket upstream from sg mRNA1 transcription start site requiring four key long-distance RNA-RNA interactions; and (ii) docking of the RTSL into the binding pocket via two additional long-distance RNA-RNA interactions. This RNA conformation generated represents the most complex

attenuation structure reported to date and its elaborate pathway of formation is proposed to help delay transcription and translation of the CP to later stages of infections.

2. PEMV2 was found to transcribe its sg mRNA via a premature termination mechanism similar to CIRV and other tombusvirids examined. However, in contrast to CIRV, PEMV2's attenuation structure forms in *trans* through a palindrome-mediated kissing-loop interaction between two PEMV2 genomes. Further extension of the genomic duplex into the stem regions was not required, implicating the loop-loop interaction as the key interface, with probable coaxial stacking providing further stabilization. The concentration dependence of this process could also provide a means to delay transcription and translation of the encoded proteins during infections. This represents the first report of RNA genome homodimerization activating sg mRNA transcription, a mechanism that likely extends to related umbra-like viruses.

3. Translational readthrough results in a C-terminally extended protein. In PEMV1, the readthrough protein, CP-RTD, is required for aphid transmission, and is translated from a sg mRNA during infections. Although a long-range interaction was previously implicated in the formation of the PEMV1 readthrough structure, the sequences involved, and the nature of the conformation generated were unknown. Structure-function analysis revealed that a complex higher-order RNA structure forms just downstream of the readthrough site, involving three distinct long-range RNA-RNA interactions. In contrast to the corresponding structure formed in related PLRV (i.e. a simple helix), the PEMV1 structure was composed of multiple discontinuous shorter helices, with coaxially stacking implicated in enhancing its stability. These diverging results indicate that these related viruses have evolved highly contrasting solutions for generating functional readthrough-promoting RNA structures.

5.2 Viral riboregulation by long-distance interactions and genome dimerization

CIRV, PEMV2, and PEMV1 all rely on RNA-based regulation to control gene expression from their sg mRNAs. For these viruses, the regulation involves long-range RNA-RNA interactions or genome dimerization. While dimerization is uncommon, long-range interactions are frequently involved in plus-strand RNA virus replication and gene expression (Nicholson and White, 2014). Why these viruses have adopted this largescale folding strategy is not always clear. The initial formation of such interactions is presumably related to inherent folding of RNA virus genomes into more compact, thermodynamically stable structures (Nicholson and White, 2015; Vicens and Kieft, 2022). This in turn provides opportunities for random sampling of different long-distance RNA-RNA contacts, with beneficial ones being maintained (Nicholson and White, 2014; Chkuaseli and White, 2018). The adoption of such interactions could provide several benefits during viral infections, including: (i) bringing together distal RNA regions that need to be in close proximity for their function, such as the function of the UL/DL interaction for generating a discontinuous RNA platform for genome replication; (ii) relocation of RNA-bound protein factors from one part of the genome to another, for example 3'CITE-5'UTR interactions; (iii) formation of specialized higher-order RNA structures that directly regulate viral processes, such as attenuation or RT/FS stimulatory structures; and (iv) allowing crosstalk and coordination between different viral processes that proceed in opposite directions or at the opposite ends of the viral RNA genome such as coordination of RT and genome replication (Nicholson and White, 2014; Chkuaseli and White, 2018).

Functional genome dimerization events, similar to that identified for PEMV2, are rare and have only been experimentally confirmed for one plus-strand RNA virus, red clover necrotic mosaic virus, and a retrovirus, human immunodeficiency virus 1 (HIV-1) (Sit et al., 1998; Dubois et al., 2018). Utilizing genome dimerization as a regulatory mechanism can provide several advantages: (i) allowing genome concentration-dependent temporal regulation of different viral processes; (ii) allowing co-packaging of different viral genomes into same particles to form

154

infectious virions; and (iii) promoting recombination events between genomes to increase viral diversity. Given these potential benefits, it is likely that other viruses also rely on this method of RNA-based regulation. Indeed, experimental validation of other viruses proposed to use this mechanism (e.g. coronaviruses, hepatitis C virus, West Nile virus, and umbra-like viruses) would help to further establish its utility and prevalence (Bou-Nader and Zhang, 2020; Romero-López and Berzal-Herranz, 2020; Romero-López et al., 2023; Chkuaseli and White, 2023).

5.3 Regulation of sg mRNA transcription

CIRV and PEMV2 both transcribe their sg mRNAs via the premature termination mechanism involving two steps (**Figure 1A, C).** In step 1, a minus-strand sg RNA is synthesized when the RdRp terminates transcription at an attenuation structure. Step 2 involves transcription of multiple copies of the plus-strand sg mRNA from the minus-strand template. Regulation of step 1 is governed by the formation of an active RNA attenuation structure.

CIRV and PEMV2 form distinct RNA attenuation structures that differ in several respects (**Figure 1B, D**) (Chkuaseli and White, 2020; 2023). The length of the spacer element is 3 nt in CIRV and 2 nt in PEMV2, suggesting different relative positioning of the leading edge of their RdRps upon contact with their respective attenuation structures. Immediately upstream from the spacer elements, comparable helices are present in both viruses (**Figure 1D, grey; 1B, red**), but although these helices are similar in length and form intragenomically, CIRV's involves a ~1000-nt long-distance RNA-RNA interaction between AS1 and RS1 (**Figure 1A, B, red**), while that for PEMV2 is formed by a local stem-loop (**Figure 1B**). In both cases these helices are surrounded by additional base-pairing that contributes to their overall stability. The attenuation structure for CIRV contains five stabilizing base-paired stems around the AS1/RS1 helix, each formed by an intragenomic long-distance RNA-RNA interaction (**Figure 1B, pink, orange, purple, green, and blue**). In contrast, stabilization in PEMV2 is mediated by the *trans*-base-
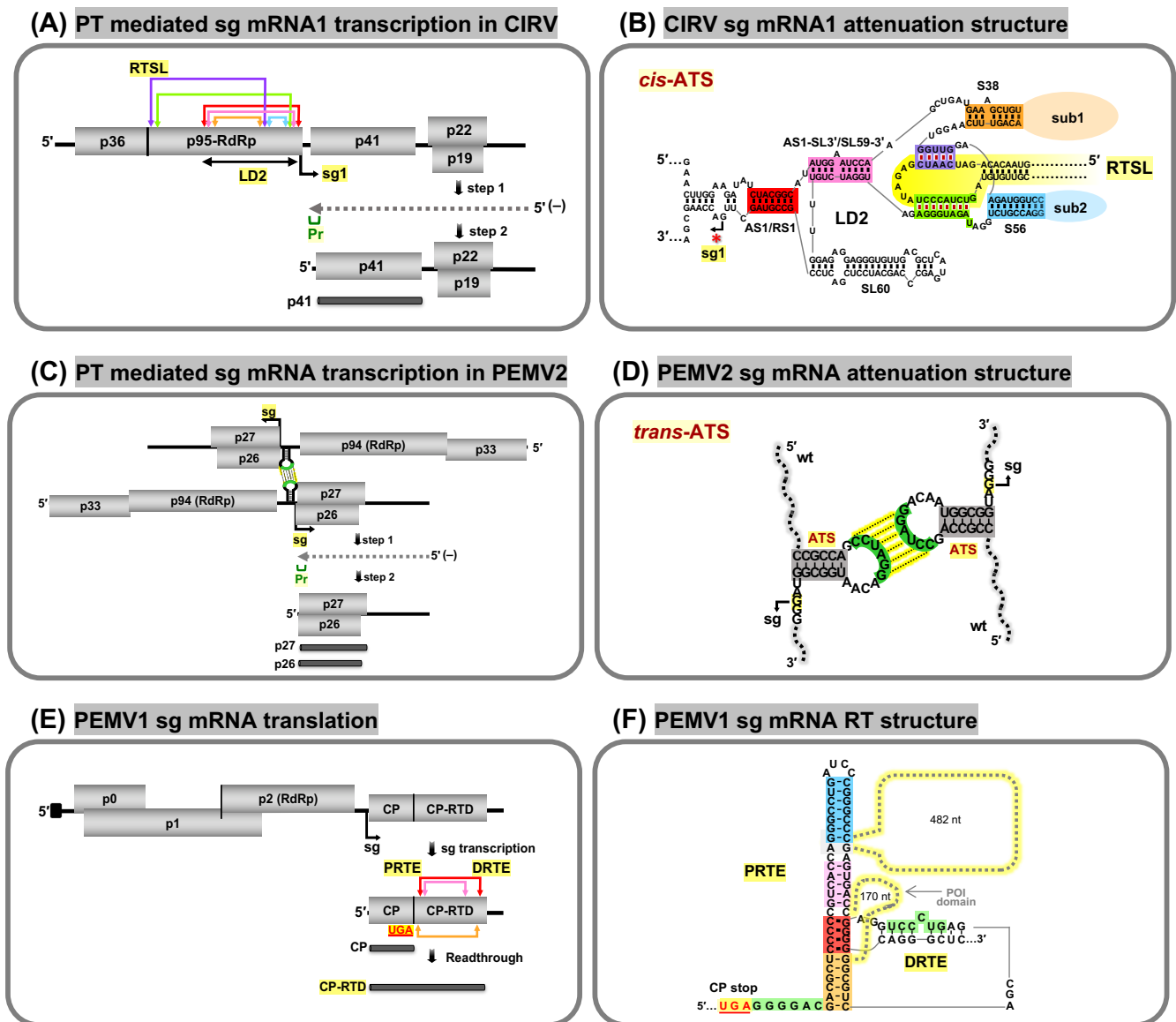
Figure 1: Sg mRNA transcription and translation regulation mechanisms of CIRV, PEMV2, and PEMV1. (A) Premature termination (PT) mechanism of sg mRNA1 transcription in CIRV. Long-distance *cis* RNA-RNA interactions involved in the formation of the attenuation structure are shown as differently colored double headed arrows. Positions of RTSL, LD2, and sg mRNA1 transcription start site are indicated. **(B)** The *cis* attenuation (ATS) structure assembled by the RNA-RNA interactions shown in (A). **(C)** PT mechanism of sg mRNA transcription in PEMV2. Two PEMV2 genomes are shown base pairing in *trans* via a palindrome-mediated kissing-loop interaction (green loops connected with dashed lines) that forms the active attenuation structure. **(D)** A more detailed view of the PEMV2 *trans*-ATS. **(E)** Readthrough-mediated translation of CP-RTD from PEMV1 sg mRNA. Long-distance RNA-RNA interactions between PRTE and DRTE RNA elements involved in the formation of the readthrough stimulatory RNA structure are indicated. **(F)** PEMV1 readthrough stimulatory structure formed by the RNA-RNA interactions shown in (E) and a local SL structure (blue).

156

paired palindromic loop sequences (**Figure 1D, green**). Despite these notable differences, both structures perform equally well in their assigned functions.

Overall, the CIRV attenuation structure could be considered more complex due to its multiple components. This higher structural complexity also correlates with differences in the folding pathways leading to their formation. For CIRV, the assembly pathway is seemingly more elaborate, requiring sequential folding of sub-components (**Figure 1B**). This requirement likely makes formation of the active attenuation structure a relatively low frequency event during CIRV infections, and could account for the observed delay in sg mRNA1 accumulation during infections (Zheng et al., 1999; Qiu and Scholthof, 2001). However, post-translational modification of the RdRp, or the involvement of host proteins, could also control the timing of transcription. For PEMV2, its attenuation structure is dependent first on the formation of the ATS stem loop and second on the concentration of PEMV2 genomic RNA. Thus, in this latter case, the timing of sg mRNA transcription activation could be influenced by the efficiency of one or both of these steps. Consequently, CIRV and PEMV2 have adopted different mechanistic solutions to delay the transcription of their sg mRNAs that encode viral proteins required at later stages of infections.

Although the attenuation structures of CIRV and PEMV2 are distinct in terms of overall structure and assembly pathways, they both represent thermodynamically stable higher-order RNA complexes that are comprised of canonical, non-canonical, and stacking interactions (**Figure 1B, D**). Formation of these structures could occur spontaneously, with their relative abundance in the population based on their thermodynamic stability. However, it is also possible that viral or host proteins aid in the formation of these attenuation structures. For instance, the pre-readthrough replication protein of tombusviruses, p36 in CIRV, has demonstrated RNA chaperone activity in vitro (Stork et al., 2011), and thus could potentially assist in formation of the attenuation structure. Also, several host RNA helicases are involved in tombusvirus genome replication, which could possibly assist in reorganizing the viral genome architecture to favor

attenuation structure formation (Kovalev et al., 2012; Kovalev and Nagy, 2014). Conversely, these RNA structure-adjusting proteins could also act to downregulate sg mRNA transcription, by unwinding or destabilizing attenuation structures.

Regardless of the involvement of host or viral proteins, it is clear that the stability of attenuation structures is critical for their activity. This stability allows them to act, at least in part, as physical barriers for the transcribing RdRp, with the downstream A/U-rich promoter region likely aiding dissociation of the RdRp from the template and nascent strand (White, 2002). During minus-strand synthesis, the leading edge of the RdRp is envisioned to contact the attenuation structure, causing it to stall over the A/U-rich sequence. Although the stalling is believed to be mediated primarily by nonspecific physical restriction (Wang et al., 2008), it is also possible that the attenuation structures have distinctive features that allows the RdRp to specifically recognize them (Chkuaseli and White, 2023). Such interactions could occur if the RdRp contains a specific recognition domain. Tombusviruses, such as CIRV, have RdRps in which the C-terminal motif is essential for recognition of their attenuation structures (Wu and White, 2007), and other further upstream RdRp motifs were also shown to specifically promote sg mRNA levels during infections (Gunawardene et al., 2015). Thus, at least for tombusviruses, features of the RdRp are important for sensing attenuation structures.

5.4 Regulation of stop codon readthrough

PEMV1, the assistor virus of PEMV2, translates a C-terminally extended capsid protein CP-RTD from its subgenomic mRNA via a programmed stop codon readthrough (RT) mechanism. Formation of the RT structure depends on three long-distance RNA-RNA interactions involving sequences downstream from the CP stop codon and partner sequences located within the CP-RTD coding region (**Figure 1E**) (Chkuaseli and White, 2022). These interactions lead to the assembly of an extended non-contiguous SL structure that stimulates ribosome readthrough and allows for production of the minor capsid protein, CP-RTD. CP-RTD

is required in low quantities for aphid-mediated plant-to-plant transmission of PEMV1 and PEMV2 virions (Demler et al., 1997). Because CP-RTD contains the CP as its N-terminal domain, it is incorporated into the virions. However, only a few copies of CP-RTD are needed per virion, thus the readthrough gene expression strategy acts to maintain a consistent high to low concentration stoichiometry between CP and CP-RTD, respectively.

PLRV is a close relative of PEMV1 and also produces its CP-RTD via CP stop codon readthrough. PLRV's RT structure is simpler than PEMV1's and is formed by the closing stem of a large RNA domain (Chkuaseli and White, 2022). This is similar to the formation of the AS1/RS1 interaction in CIRV's attenuation structure, which is also mediated by folding of a large RNA domain, LD2 (**Figure 1A, B**). In contrast, the assembly process for PEMV1's RT structure is predicted to be more complex, involving a ribosome-mediated conformational transition, nucleation of the interaction via the kissing-loops, and then additional long-range base-pairing interactions completing a non-contiguous helix stabilized by coaxial stacking (**Figure 1E, F**). The complexity of this multistep assembly pathway parallels that for the assembly of CIRV's elaborate attenuation structure (**Figure 1B**).

The question remains as to why PEMV1 or PLRV use long-distance RNA-RNA interactions to form their readthrough structures. While the requirement for a long-range interaction in the tombusvirid CIRV is related to coordinating readthrough production of the RdRp with genome replication (Cimino et al., 2011), there does not seem to be an obvious reason for employing this long-distance strategy in PEMV1 or PLRV. Instead, as previously alluded to, the global folding of these sg mRNAs may have provided the opportunity for testing and adopting these functional long-range interactions (Chkuaseli and White, 2022). Simply put, it worked, so it was maintained.

5.5 Future Directions

Data presented in this dissertation provided detailed information on RNA-based regulation of sg mRNA transcription and translation in tombusvirids CIRV and PEMV2, and a solemovirid PEMV1, however, additional future experiments could be performed to expand understanding of these and related processes. Examples of possible future studies are provided below.

5.5.1 Investigating the origin and function of PEMV2's sg*

A novel subgenomic RNA termed sg* was detected during PEMV2 protoplast infections that is 152 nt longer than the primary sg mRNA and does not appear to greatly contribute to the production of p26 and p27 (Chkuaseli and White, 2023). The fact that sg* is produced during infections suggests that it provides some fitness advantage to the virus. It is unlikely that it is produced via transcription by the viral RdRp, because it does not possess the typical sg mRNA promoter sequence. Instead, it could represent a degradation product of a cellular 5'-to-3' exoribonuclease, such as Xrn (Steckelberg et al., 2018; Ilyas et al., 2021). For example, following endoribonucleolytic digestion within the genomic RNA, a 5'-to-3' exoribonuclease could engage the 3'-half and hydrolyze it until reaching an RNA structure resistant to digestion. This seems plausible because a small sg-like RNA, that serves as a message for a novel viral protein, is produced by Xrn digestion in infections with another umbravirus, OPMV (Ilyas et al., 2021).

To determine whether sg* originates through Xrn digestion, in vitro Xrn digestion assays could be performed (Gunawardene et al., 2019). A degradation product, matching sg* in size (in gels) and in the 5'-end nucleotide identity (determine through 5'-RACE), would be expected if sg* is generated as a product of Xrn digestion. Further structural analysis (e.g. SHAPE) to define the Xrn-blocking structure and mutational analysis (protoplast infections and in vitro Xrn

digestion assays) to define the structural features important for function, would also be performed.

Additionally, although PEMV2's sg* does not appear to greatly contribute to p26 and p27 translation (based on in vitro translation) (Chkuaseli and White, 2023), it could have other functions during natural whole plant infections when present with its assistor virus, PEMV1. To understand potential functions of sg*, production of sg* could be inhibited in whole plant infections by mutagenesis inactivating the Xrn-blocking structure and effects of sg*'s absence on genome accumulation and symptomatology assessed.

5.5.2 In vivo evidence for PEMV2 genome dimerization

Although in vitro EMSA and in vivo compensatory mutagenesis clearly support functional dimerization of PEMV2 genomes, an additional piece of evidence to provide further support would be to perform RNA-RNA crosslinking during infections. Techniques such as SPLASH (sequencing of psoralen crosslinked, ligated, and selected hybrids) or COMRADES (cross-linking of matched RNAs and deep sequencing) could be used (Aw et al., 2016; Ziv et al., 2018; 2020). Both techniques rely on psoralen that selectively and reversibly crosslinks certain paired RNA bases in vivo. Briefly, protoplasts infected with PEMV2 would be treated with psoralen to allow crosslinking of base-paired RNAs and crosslinked viral RNAs would be isolated. The cross-linked viral RNAs would then be proximity ligated, followed by crosslink reversal, adapter ligation, reverse transcription, PCR, and deep sequencing to identify RNAs that were ligated together, and thus base-paired. The PEMV2 palindrome chimeras would be predicted to be among the crosslinked sequences. However, success of this detection would be dependent on the frequency with which PEMV2 genomes dimerize during infections and the presence of crosslinkable residues across the base-paired region.

5.5.3 The potential cis-acting ATSs of umbraviruses CMoV, CMoMV, and PasUV1

Comparative sequence analysis indicated that the umbraviruses carrot mottle virus (CMoV), carrot mottle mimic virus (CMoMV), and pastinaca umbravirus 1 (PasUV1), do not contain palindromic loop sequences in their attenuation structures (ATSs) (Chkuaseli and White, 2023). Instead, these viruses have additional SL structures, 11-12 nt upstream from their ATSs, with complementary loop sequences. Thus, *cis*-kissing loop interactions could be involved in the formation of active ATSs in these viruses. To explore this possibility, standard compensatory mutagenesis could be carried out between the complementary loop sequences for the three viruses and effects of these mutations analyzed in protoplast infections. Additionally, the loop sequence of CMoMV could be converted to a palindrome by changing one of the two central nucleotides to test if its *cis* attenuation structure can be converted to a *trans* attenuation structure. Information gathered from these experiments could provide a better evolutionary understanding of how the *trans* attenuation structure initially appeared in umbraviruses.

5.5.4 RT stimulatory structures of other poleroviruses and luteoviruses

PRTE and DRTE RNA sequences involved in CP-RTD readthrough production are predicted in similar locations in the sg mRNAs of other poleroviruses, enamoviruses, and luteoviruses (Xu et al., 2018). However, due to variability in the PRTE/DRTE primary sequences and differences in complementarity, it is predicted that RT stimulating structures of differing complexity are present in these viruses (LaTourette et al., 2021; Chkuaseli and White, 2022). RNA secondary structure predictions for some of these viruses suggest the existence of RT structures that are of intermediate complexity between those of PLRV and PEMV1 (TC, unpublished). Accordingly, diverse RT promoting structures of these viruses could be investigated (as in chapter 4) and compared with one another to gain additional insights into evolutionary relationships among these related groups of viruses.

5.6 Conclusion

The discoveries reported in this dissertation were uncovered through comprehensive structure-function characterization of novel types of RNA-based regulatory structures involved in gene expression of viruses in Tombusviridae and Solemoviridae families. The findings provide new mechanistic insights and further expand on our existing knowledge of RNA-based regulation in plus-strand RNA viruses.

5.7 References

Aw, J. G. A., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K.-L., Tapsin, S., Chan, Y.-S., Tan, C.-P., Sim, A. Y. L., Zhang, T., Susanto, T. T., Fu, Z., Nagarajan, N., & Wan, Y. (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell*, *62*(4), 603–617. https://doi.org/10.1016/j.molcel.2016.04.028

Bou-Nader, C., & Zhang, J. (2020). Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions. *Molecules*, *25*(12), 2881. https://doi.org/10.3390/molecules25122881

Chkuaseli, T., & White, K. A. (2018). Intragenomic Long-Distance RNA–RNA Interactions in Plus-Strand RNA Plant Viruses. *Frontiers in Microbiology*, *9*, 529. https://doi.org/10.3389/fmicb.2018.00529

Chkuaseli, T., & White, K. A. (2020). Activation of viral transcription by stepwise largescale folding of an RNA virus genome. *Nucleic Acids Research*, *48*(16), 9285–9300. https://doi.org/10.1093/nar/gkaa675

Chkuaseli, T., & White, K. A. (2022). Complex and simple translational readthrough signals in pea enation mosaic virus 1 and potato leafroll virus, respectively. *PLOS Pathogens*, *18*(9), e1010888. https://doi.org/10.1371/journal.ppat.1010888

Chkuaseli, T., & White, K. A. (2023). Dimerization of an umbravirus RNA genome activates subgenomic mRNA transcription. *Nucleic Acids Research*, *In press*.

Choi, I.-R., & White, K. A. (2002). An RNA Activator of Subgenomic mRNA1 Transcription in Tomato Bushy Stunt Virus. *Journal of Biological Chemistry*, *277*(5), 3760–3766. https://doi.org/10.1074/jbc.M109067200

Cimino, P. A., Nicholson, B. L., Wu, B., Xu, W., & White, K. A. (2011). Multifaceted Regulation of Translational Readthrough by RNA Replication Elements in a Tombusvirus. *PLoS Pathogens*, *7*(12), e1002423. https://doi.org/10.1371/journal.ppat.1002423

Demler, S. A., Rucker-Feeney, D. G., Skaf, J. S., & de Zoeten, G. A. (1997). Expression and suppression of circulative aphid transmission in pea enation mosaic virus. *Journal of General Virology*, *78*(3), 511–523. https://doi.org/10.1099/0022-1317-78-3-511

Dubois, N., Marquet, R., Paillart, J.-C., & Bernacchi, S. (2018). Retroviral RNA Dimerization: From Structure to Functions. *Frontiers in Microbiology*, *9*, 527. https://doi.org/10.3389/fmicb.2018.00527

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *Journal of General Virology*, *93*(7), 1385–1409. https://doi.org/10.1099/vir.0.042499-0

Firth, A. E., Wills, N. M., Gesteland, R. F., & Atkins, J. F. (2011). Stimulation of stop codon readthrough: Frequent presence of an extended 3′ RNA structural element. *Nucleic Acids Research*, *39*(15), 6679–6691. https://doi.org/10.1093/nar/gkr224

Gao, F., & Simon, A. E. (2017). Differential use of 3'CITEs by the subgenomic RNA of Pea enation mosaic virus 2. *Virology*, *510*, 194–204. https://doi.org/10.1016/j.virol.2017.07.021

Gunawardene, C. D., Jaluba, K., & White, K. A. (2015). Conserved Motifs in a Tombusvirus Polymerase Modulate Genome Replication, Subgenomic Transcription, and Amplification of Defective Interfering RNAs. *Journal of Virology*, *89*(6), 3236–3246. https://doi.org/10.1128/JVI.03378-14

Gunawardene, C. D., Newburn, L. R., & White, K. A. (2019). A 212-nt long RNA structure in the Tobacco necrosis virus-D RNA genome is resistant to Xrn degradation. *Nucleic Acids Research*, *47*(17), 9329–9342. https://doi.org/10.1093/nar/gkz668

Ilyas, M., Du, Z., & Simon, A. E. (2021). *Opium Poppy Mosaic Virus* Has an Xrn-Resistant, Translated Subgenomic RNA and a BTE 3′ CITE. *Journal of Virology*, *95*(9), e02109-20. https://doi.org/10.1128/JVI.02109-20

Kovalev, N., & Nagy, P. D. (2014). The Expanding Functions of Cellular Helicases: The Tombusvirus RNA Replication Enhancer Co-opts the Plant eIF4AIII-Like AtRH2 and the DDX5-Like AtRH5 DEAD-Box RNA Helicases to Promote Viral Asymmetric RNA Replication. *PLoS Pathogens*, *10*(4), e1004051. https://doi.org/10.1371/journal.ppat.1004051

Kovalev, N., Pogany, J., & Nagy, P. D. (2012). A Co-Opted DEAD-Box RNA Helicase Enhances Tombusvirus Plus-Strand Synthesis. *PLoS Pathogens*, *8*(2), e1002537. https://doi.org/10.1371/journal.ppat.1002537

Kuhlmann, M. M., Chattopadhyay, M., Stupina, V. A., Gao, F., & Simon, A. E. (2016). An RNA Element That Facilitates Programmed Ribosomal Readthrough in Turnip Crinkle Virus Adopts Multiple Conformations. *Journal of Virology*, *90*(19), 8575–8591. https://doi.org/10.1128/JVI.01129-16

LaTourrette, K., Holste, N. M., & Garcia-Ruiz, H. (2021). Polerovirus genomic variation. *Virus Evolution*, *7*(2), veab102. https://doi.org/10.1093/ve/veab102

Nagy, P. D. (2016). Tombusvirus-Host Interactions: Co-Opted Evolutionarily Conserved Host Factors Take Center Court. *Annual Review of Virology*, *3*(1), 491–515. https://doi.org/10.1146/annurev-virology-110615-042312

Newburn, L. R., Nicholson, B. L., Yosefi, M., Cimino, P. A., & White, K. A. (2014). Translational readthrough in Tobacco necrosis virus-D. *Virology*, *450–451*, 258–265. https://doi.org/10.1016/j.virol.2013.12.006

Newburn, L. R., Wu, B., & White, K. A. (2020). Investigation of Novel RNA Elements in the 3′UTR of Tobacco Necrosis Virus-D. *Viruses*, *12*(8), 856. https://doi.org/10.3390/v12080856

Nicholson, B. L., & White, K. A. (2014). Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nature Reviews Microbiology*, *12*(7), 493–504. https://doi.org/10.1038/nrmicro3288

Nicholson, B. L., & White, K. A. (2015). Exploring the architecture of viral RNA genomes. *Current Opinion in Virology*, *12*, 66–74. https://doi.org/10.1016/j.coviro.2015.03.018

Qiu, W., & Scholthof, H. B. (2001). Effects of inactivation of the coat protein and movement genes of Tomato bushy stunt virus on early accumulation of genomic and subgenomic RNAs. *Journal of General Virology*, *82*(12), 3107–3114. https://doi.org/10.1099/0022-1317-82-12-3107

Romero-López, C., & Berzal-Herranz, A. (2020). The Role of the RNA-RNA Interactome in the Hepatitis C Virus Life Cycle. *International Journal of Molecular Sciences*, *21*(4), 1479. https://doi.org/10.3390/ijms21041479

Romero-López, C., Roda-Herreros, M., Berzal-Herranz, B., Ramos-Lorente, S. E., & Berzal-Herranz, A. (2023). Inter- and Intramolecular RNA–RNA Interactions Modulate the Regulation of Translation Mediated by the 3′ UTR in West Nile Virus. *International Journal of Molecular Sciences*, *24*(6), 5337. https://doi.org/10.3390/ijms24065337

Sit, T. L., Vaewhongs, A. A., & Lommel, S. A. (1998). RNA-Mediated Trans-Activation of Transcription from a Viral RNA. *Science, New Series*, *281*(5378), 829–832. https://doi.org/10.1126/science.281.5378.829

Steckelberg, A.-L., Vicens, Q., & Kieft, J. S. (2018). Exoribonuclease-Resistant RNAs Exist within both Coding and Noncoding Subgenomic RNAs. *MBio*, *9*(6), e02461-18. https://doi.org/10.1128/mBio.02461-18

Stork, J., Kovalev, N., Sasvari, Z., & Nagy, P. D. (2011). RNA chaperone activity of the tombusviral p33 replication protein facilitates initiation of RNA synthesis by the viral RdRp in vitro. *Virology*, *409*(2), 338–347. https://doi.org/10.1016/j.virol.2010.10.015

Vicens, Q., & Kieft, J. S. (2022). Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences*, *119*(17), e2112677119. https://doi.org/10.1073/pnas.2112677119

Wang, S., Mortazavi, L., & White, K. A. (2008). Higher-Order RNA Structural Requirements and Small-Molecule Induction of Tombusvirus Subgenomic mRNA Transcription. *Journal of Virology*, *82*(8), 3864–3871. https://doi.org/10.1128/JVI.02416-07

White, K. A. (2002). The Premature Termination Model: A Possible Third Mechanism for Subgenomic mRNA Transcription in (+)-Strand RNA Viruses. *Virology*, *304*(2), 147–154. https://doi.org/10.1006/viro.2002.1732

Wu, B., & White, K. A. (2007). Uncoupling RNA virus replication from transcription via the polymerase: Functional and evolutionary insights. *The EMBO Journal*, *26*(24), 5120–5130. https://doi.org/10.1038/sj.emboj.7601931

Xu, Y., Ju, H.-J., DeBlasio, S., Carino, E. J., Johnson, R., MacCoss, M. J., Heck, M., Miller, W. A., & Gray, S. M. (2018). A Stem-Loop Structure in *Potato Leafroll Virus* Open Reading Frame 5 (ORF5) Is Essential for Readthrough Translation of the Coat Protein ORF Stop Codon 700 Bases Upstream. *Journal of Virology*, *92*(11), e01544-17. https://doi.org/10.1128/JVI.01544-17

Zhang, G., Slowinski, V., & White, K. A. (1999). Subgenomic mRNA regulation by a distal RNA element in a (+)-strand RNA virus. *RNA*, *5*(4), 550–561. https://doi.org/10.1017/S1355838299982080

Ziv, O., Gabryelska, M. M., Lun, A. T. L., Gebert, L. F. R., Sheu-Gruttadauria, J., Meredith, L. W., Liu, Z.-Y., Kwok, C. K., Qin, C.-F., MacRae, I. J., Goodfellow, I., Marioni, J. C., Kudla, G., & Miska, E. A. (2018). COMRADES determines in vivo RNA structures and interactions. *Nature Methods*, *15*(10), 785–788. https://doi.org/10.1038/s41592-018-0121-0

Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F., & Miska, E. A. (2020). The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Molecular Cell*, *80*(6), 1067-1077.e5. https://doi.org/10.1016/j.molcel.2020.11.004

APPENDICES

Appendix A – Additional Research Contributions

Chkuaseli, T., Newburn, L. R., Bakhshinyan, D., & White, K. A. (2015). Protein expression strategies in Tobacco necrosis virus-D. *Virology*, *486*, 54–62. https://doi.org/10.1016/j.virol.2015.08.032

Chkuaseli, T., & White, K. A. (2018). Intragenomic Long-Distance RNA–RNA Interactions in Plus-Strand RNA Plant Viruses. *Frontiers in Microbiology*, 9, 529. https://doi.org/10.3389/fmicb.2018.00529

Appendix B – Copyright Permissions

Figure 1: Copyright permission for CHAPTER 1, Figure 1.

Order Date	16-May-2023	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1355720-1	**Publisher**	Annual Reviews
ISSN	2327-0578	**Portion**	Cartoon

LICENSED CONTENT

Publication Title	Annual review of virology	**Rightsholder**	Annual Reviews, Inc.
Date	01/01/2014	**Publication Type**	e-Journal
Language	English	**URL**	http://www.annualreviews.org/journal/vir...
Country	United States of America		

REQUEST DETAILS

Portion Type	Cartoon	**Distribution**	Canada
Number of Cartoons	1	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the Disabled?**	No
Who Will Republish the Content?	Academic institution	**Minor Editing Privileges?**	Yes
Duration of Use	Life of current and all future editions	**Incidental Promotional Use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	CAD
Rights Requested	Main product		

Figure 2: Copyright permission for CHAPTER 1, Figure 3C.

Creative Commons License Deed

Attribution 4.0 International (CC BY 4.0)

Figure 3: Copyright permission for CHAPTER 1, Figures 4 and 5A.

Figure 4: Copyright permission for CHAPTER 1, Figure 5B.

Print This Page