ABSTRACT

Title of Document: ALTERNATE CONFORMATIONS REGULATE RIBOSOMAL RECODING IN A POSITIVE-SENSE RNA VIRUS

Michelle Kuhlmann
Doctor of Philosophy, 2016

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

Positive-sense RNA viruses are important animal, plant, insect and bacteria pathogens and constitute the largest group of RNA viruses. Due to the relatively small size of their genomes, these viruses have evolved a variety of non-canonical translation mechanisms to optimize coding capacity expanding their proteome diversity. One such strategy is codon redefinition or recoding. First described in viruses, recoding is a programmed translation event in which codon alterations are context dependent. Recoding takes place in a subset of messenger RNA (mRNAs) with some products reflecting new, and some reflecting standard, meanings. The ratio between the two is both critical and highly regulated. While a variety of recoding mechanisms have been documented, (ribosome shunting, stop-carry on, termination-reinitiation, and translational bypassing), the two most extensively employed by RNA viruses are Programmed Ribosomal Frameshifting (PRF) and Programmed Ribosomal Readthrough (PRT). While both PRT and PRF subvert normal decoding for expression of C-terminal extension products, the former involves an alteration of reading frame, and the latter requires decoding of a non-sense

codon. Both processes occur at a low but defined frequency, and both require Recoding Stimulatory Elements (RSE) for regulation and optimum functionality. These stimulatory signals can be embedded in the RNA in the form of sequence or secondary structure, or trans-acting factors outside the mRNA such as proteins or micro RNAs (miRNA). Despite 40+ years of study, the precise mechanisms by which viral RSE mediate ribosome recoding for the synthesis of their proteins, or how the ratio of these products is maintained, is poorly defined. This study reveals that in addition to a long distance RNA:RNA interaction, three alternate conformations and a phylogenetically conserved pseudoknot regulate PRT in the carmovirus *Turnip crinkle virus* (TCV).

ALTERNATE CONFORMATIONS REGULATE RIBOSOMAL RECODING IN A POSITIVE-SENSE RNA VIRUS

By

Michelle Kuhlmann

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Anne E. Simon, Chair
Professor James N. Culver
Professor Jeffrey J. DeStefano
Professor Jonathan D. Dinman
Assistant Professor George A. Belov

Dedication

I would like to dedicate this work to my husband, Dietrich. Thank you for all your support, love and encouragement over the last 30^{+} years of marriage.

Acknowledgements

I would like to thank my advisor, Dr. Anne Simon. You really were a perfect fit for me. Thank you also to my committee members, I sincerely appreciate your time and energy.

So many people to thank. I would like to include the many people in the department I have had the pleasure of interacting with. Dr. Edgar Moctezuma, Dr. Patty Shields, Dr. John Buchner, Dr. Michelle Brooks, Ms. Gwen Warman, Ms. LaJeune Stover, Ms. Dorothea O'Toole, Ms. Simone Lord-Attivor, Ms. Nalini Chelliah, Mr. Jeffrey Dunton and Mr. Ryan Muldoon (Radiation safety). I appreciate all your kindnesses over these (many) years.

A special thanks to my lab mates and my students - you have been my family and support system for seven years. I hope you know how very much you have meant to me. My Le, Maitreyi, Tina, Kalyani, Feng, Amulya, Phillip, Vera, Teresa, Jared, you taught me, mentored me, put up with me, laughed with me . . . . . you made my graduate school experience so much better.

Last but certainly not least, I would like to thank my family. To my husband Dietrich, my daughter Taylor, my son Nick and my parents: you have always supported and encouraged me, and I love you very much.

Table of Contents

List of Figures

List of Abbreviations

1M7	1-methyl-7-nitroisatoic anhydride
AMV	*Alfalfa mosaic virus*
AnFBV	*Angelonia flower break virus*
BEV	Baboon endogenous retrovirus
BWYV	*Beet western yellows virus*
BYDV	*Barley yellow dwarf virus*
CarMV	*Carnation mottle virus*
CbMV	*Calibrachoa mottle virus*
CCFV	*Cardamine chlorotic fleck virus*
CIRV	*Carnation Italian ringspot virus*
CITE	Cap-independent translational enhancer
CoPMV	*Cowpea mottle virus*
CP	Coat protein
CPB	Coat protein binding form
CTFV	*Colorado tick fever virus*
Cryo-EM	Cryo-electron microscopy
DIAV	*Duck infectious anemia virus*
DMSO	Dimethyl sulfoxide
eEF1A	Eukaryotic elongation factor 1A
eIF4E	Eukaryotic initiation factor 4E
FLV	*Feline leukemia virus*

gRNA	Genomic RNA
HCRSV	*Hibiscus chlorotic ringspot virus*
HCV	*Hepatitis C virus*
HIV	*Human immunodeficiency virus*
HnRSV	*Honeysuckle ringspot virus*
hpi	Hours post-inoculation
IBV	*Avian infectious bronchitis virus*
IRES	Internal ribosome entry site
ISS	I-shaped structure
JINRV	*Japanese iris necrotic spot virus*
miRNA	Micro RNA
MDEV	*Mus dunni endogenous virus*
MLV	*Murine leukemia virus*
MMTV	*Mouse mammary tumor virus*
MNSV	*Melon necrotic spot virus*
mRNA	Messenger RNA
NLVCV	*Nootka lupine vein-clearing virus*
NMD	Nonsense-mediated decay
NMIA	N-methylisatoic anhydride
NMR	Nuclear magnetic resonance
ORF	Open reading frame
OH	Hydroxyl
OHEV	*Odocoileus hemionus endogenous virus*

PEMV	*Pea enation mosaic virus*
PERV	*Porcine endogenous retrovirus*
PFBV	*Pelargonium flower break virus*
PMV	*Panicum mosaic virus*
PRF	Programmed ribosomal frameshifting
PRT	Programmed readthrough
PSNV	*Pea stem necrosis virus*
PTE	*Panicum mosaic virus*-like translational enhancer
RNA	Ribonucleic acid
RCNMV	*Red clover necrotic mosaic virus*
RdRp	RNA-dependent RNA polymerase
RfRV	*Rhinolophus ferrumequinum retrovirus*
RSE	Recoding stimulatory element
RSV	*Rous sarcoma virus*
sgRNA	Subgenomic RNA
SHAPE	Selective 2'-hydroxyl Acylation analyzed by Primer Extension
SLA	Stem loop A
r-protein	Ribosomal protein
rRNA	Ribosomal RNA
SCV	*Saguaro cactus virus*
SARS-CoV	*Sudden acute respiratory syndrome coronavirus*
SAXS	Small-angle X-ray scattering
STNV-D	*Satellite tobacco necrosis virus-D*

SYMMV	*Soybean yellow mottle mosaic virus*
TBSV	*Tomato bushy stunt virus*
TED	Translation enhancer domain
TMV	*Tobacco mosaic virus*
TNV-D	*Tobacco necrosis virus-D*
tRNA	Transfer RNA
TSS	tRNA shaped structure
UTR	Untranslated region
WGE	Wheat germ extract
WT	Wild type

Chapter 1: Introduction

Canonical translation

Protein synthesis is an integral part of The Central Dogma in molecular biology that involves the flow or transfer of genetic information from DNA sequence to messenger ribonucleic acid (mRNA) during transcription and mRNA to protein during translation. mRNAs are essentially blueprints containing nucleotide triplets that correspond to one of twenty canonical amino acids or one of three stop codons, which are decoded by the ribosome, protein factors and transfer RNAs (tRNA). The eukaryotic ribosome (80S), a complex of RNAs and ribosomal proteins (r-protein), has two constituents: the large subunit (60S), and the small subunit (40S). The fully assembled ribosome and subunits are named for their sedimentation coefficient and are expressed in Svedberg units. The 60S contains the catalytically active site involved in the formation of peptide bonds during elongation, and consists of up to 47 r-proteins and 3 RNA molecules 5S, 5.8S, and 25S (between 25S and 28S in plants)[1,2]. The 40S subunit contains the decoding site where base pairing between the mRNA codon tRNA anticodon is monitored, and contains the 18S ribosomal RNA (rRNA) and up to 33 r-proteins[3].

Protein synthesis can be divided into three stages: initiation, elongation and termination. The initial reading frame is defined during initiation, when the mRNA binds the small subunit, and the initiator tRNA binds the start codon. Elongation ensues with the next aminoacyl tRNA binding its cognate codon on the mRNA and continues until a stop codon is encountered and recognized by a protein release factor. During protein biosynthesis the tRNA traverses three binding sites: the aminoacyl (A-site) where the

incoming aminoacylated tRNA (aa-tRNA) binds, the peptidyl (P-site) harboring the tRNA attached to the growing peptide chain, and the exit (E-site) where the deacylated tRNA resides before leaving the ribosome. This 5' to 3' directional progression of codons through the ribosome is known as translocation. The aa-tRNA is delivered to the A-site by eukaryotic elongation factor 1A (eEF1A) in its GTP-bound form which forces the tRNA to assume a deformed or 'bent' conformation (A/T) for codon sampling[4]. Recognition causes 18S rRNA residues A1824 and A1825 on helix 44 (h44) to "flip out" allowing a minor groove interaction with the codon:anticodon helix formed between the cognate tRNA and mRNA (Fig. 1.1.)[4]. A progressive tightening in the decoding center occurs upon recognition of the cognate codon inducing a closed conformation [5]. This in turn stabilizes the tertiary interaction stimulating GTP hydrolysis for subsequent dissociation of the eEF1a-GDP complex from the ribosome, and facilitates the accommodation and reconfiguration of the tRNA from an A/T to an A/A state. The structural plasticity of the tRNA allows it to function like a molecular spring adopting low to high energy conformations[6,7]. Studies on the kinetics of tRNA discernment show incorporation of a near-cognate codon requires a tenfold increase in GTP expenditure, and for non-cognate codons no GTP hydrolysis was observed[8]. It has been suggested that although a closed conformation is induced by both cognate and near-cognate tRNAs, the closed conformation forces the near-cognate tRNAs to adopt Watson-Crick geometry that may explain the unfavorable GTP consumption[9]. Thus a high rate of fidelity for codon recognition by tRNAs, (an error frequency on the order of 10^{-3} to 10^{-4})[10,11], is achieved through both the steric complementarity of the codon-anticodon helix, and the

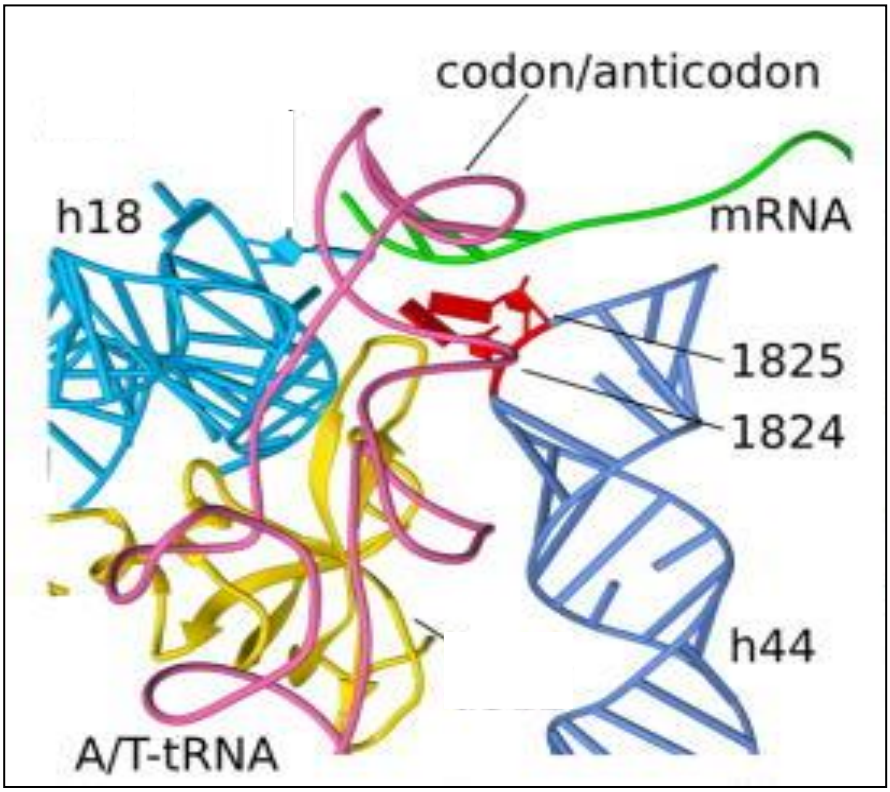


Fig. 1. 1 Model for codon recognition in eukaryotes . The GTP-bound aa-tRNA is delivered to the A-site (A/T tRNA pink) sampling the mRNA (green). Recognition of the cognate codon:anti-codon complex causes residues 1824 and 1825 (red) to "flip out" of h44 (blue) on the 18S rRNA. Image adapted from[4].

recognition of base pair geometry by the 18S adenosines. Canonical termination of translation is the last stage in protein biosynthesis and occurs when a stop codon (UAA-ochre, UGA-opal and UAG-amber) enters the aminoacyl-tRNA acceptor site (A-site) of the small ribosomal subunit where it is decoded by eukaryotic Release Factor 1 (eRF1). The accuracy of stop codon recognition is an order of magnitude higher than for tRNA selection[12]. Unlike prokaryotic Release Factors RF1 and RF2, which recognize UAA/UAG and UAA/UGA respectively, eRF1 recognizes all three stop codons[13]. The efficiency of termination is modulated by a ternary complex consisting of eRF1, eRF3 (which stimulates GTPase activity) and GTP[14]. eRF1 contains three functional domains: the N-terminal domain responsible for stop codon recognition; the M domain (with a conserved GGQ motif) essential for ribosome binding and release activity; and the C-terminal domain which engages eRF3[15-19]. These eRF1 domains are said to structurally and functionally mimic tRNA molecules, where domains M, C and N of eRF1 are analogous to the acceptor stem, T stem and anti-codon stems of the tRNA respectively [15,20,21]. Subsequent to the ternary complex binding the stop codon, GTP hydrolysis of the ester bond in the peptidyl-tRNA stimulates the dissociation of eRF3 inducing a conformational change in eRF1 which now becomes fully accommodated in the A-site[22]. This shift positions the GGQ-motif of the M domain in the peptidyl transferase center for peptide release[17]. Experimental procedures including cross-linking[23,24], Nuclear Magnetic Resonance (NMR)[25], Small-angle X-ray scattering[21] and Cryo-electron microscopy (cryo-EM)[22] have identified the specificity of stop codon recognition can be largely attributed to three conserved motifs within the N domain of human eRF1: NIKS (residues 61-64), YxCxxxF (residues 125-131) and GTS (residues 31-33).

Sequence alignments identified an Asn-Ile-Lys-Ser, or NIKS motif, located in the loop between α2 and α3 helices of the N-domain, a position analogous to the anticodon loop at the end of the double helix of a tRNA[26]. This highly conserved motif crosslinks within the ribosome to all three stop codons, specifically residue K63 links to the invariant first position uridine nucleotide[24]. Cryo-EM revealed that the side chains of N61 and K63 are within hydrogen bonding distance from the carbonyl groups of the uracil[22]. Substitution of lysine for glutamine (uncharged) or glutamate (negative charge) had a negative impact on release factor activity while a substitution of an arginine did not, suggesting the need for a positive charge for U1 recognition at position 63[27]. An additional requirement for efficient termination is the C4 hydroxylation of the K63 side chain. An almost two fold increase in the release of peptides from stalled pre-termination complexes occurred with partial hydroxylation of K63 in an in vitro system, and similarly, inhibition of hydroxylation caused an increase in ribosomal readthrough in a dual-luciferase reporter construct *in vivo*[28]. Steric hindrance with eRF1 would preclude purines in the +1 position of the stop codon, while the extensive hydrogen bonding requirements of the NIKS motif excludes cytidine[22]. Thus, the NIKS motif imposes a universal requirement for the +1 stop codon position due to the interaction between the Watson-Crick edge of the uridine and amino acid side chains.

In the YxCxxxF motif hydrogen bonding between invariant residuesTyr125 and Glu55 assists in the maintenance of spatial orientation in eRF1 and positions the glutamate side chain to hydrogen bond with the N6 atoms of the two stop codon adenosines at the +2 and +3 positions[29]. As these interactions are only possible with purines; pyrimidines would be excluded. For tryptophan, the O6 atoms of the

consecutively stacked guanosines would not satisfy hydrogen bonding requirements thus eliminating recognition of UGG by eRF1. Substitution of Alanine for Tyrosine at position 125 results in complete loss of termination activity at all three stop codons, however, substitution of Phenylalanine did not suggesting the need for an aromatic ring at position 125[27]. In a manner similar to a tRNA binding the A-site, binding of eRF1 flips the 18S nucleotide A1825[22]. Unlike sense codon recognition, A1824 does not flip out, and remains tucked into h44 (Fig. 2). eRF1 stacks on the +2 base, upon which the +3 base then stacks, and both are decoded as a single unit. This unique conformation allows for compaction of the mRNA pulling the +4 nucleotide into the A-site where it stacks against base G626 on h18 of the 18S effectively making it a tetra-nucleotide stop signal (Fig. 1.2)[22,30]. The main chain of Cys127 stabilizes the interaction, forming two hydrogen bonds with A1825 (Fig. 1.2)[27,31]. These data suggest the YxCxxxF motif contributes to the recognition of adenosines in the second and third positions in all three stop codons.

The third structural motif within the N-domain of eRF1, the highly conserved Gly-Thr-Ser loop or GTS motif, is located towards the sugar edge of the +3 stop codon base and adopts two different conformations that are interconnected with the YxCxxxF motif[25]. With a UAG stop codon, Thr32 faces the +3 base and can hydrogen bond with the N2 atom of the guanosine. A guanosine in the +2 position causes movement of the YxCxxxF motif towards the stacked pair causing a conformational shift in the GTS motif where Thr32 to faces away from the stop codon. Therefore, guanosine can occur at either the +2 or +3 positions[22].

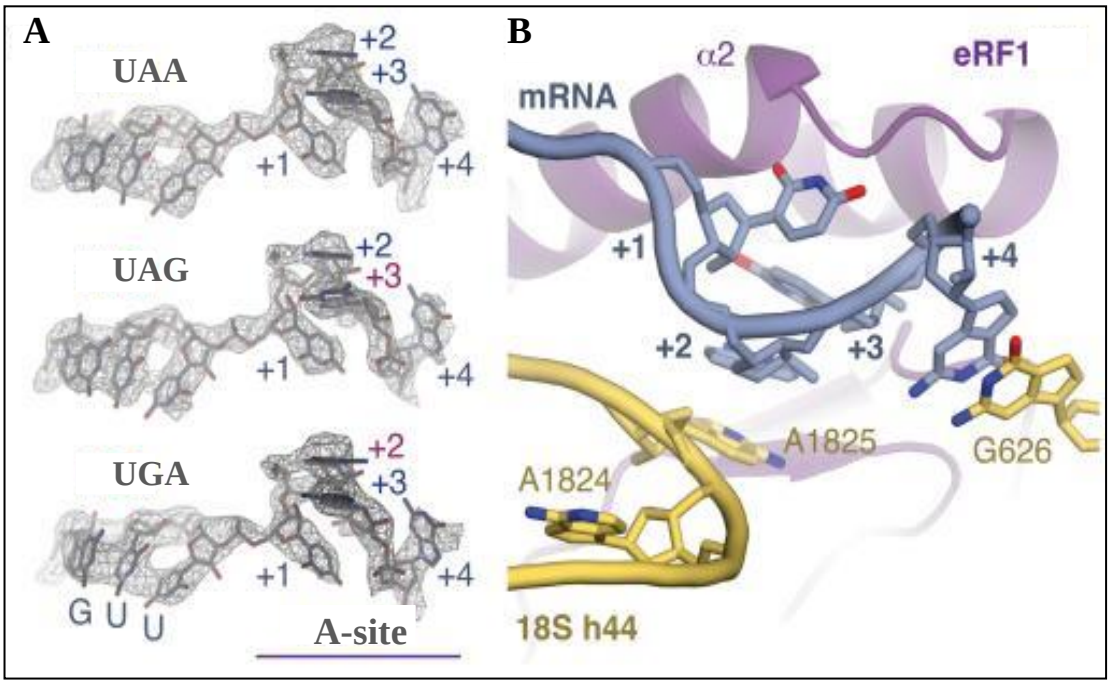


Fig. 1. 2 Model for eukaryotic stop codon recognition in the decoding center. (A) EM map densities of the UAA UAG and UGA stop codons reveal identical conformations with the +2 and +3 bases stacking and the +4 base being pulled into the A-site (purple bar) due to mRNA compaction. A Val^{GUU} codon occupies the P-site. (B) Recognition of the termination signal bases (+1 - +4 slate) by eRF1 (purple) in the A-site. Bases +2 and +3 stack on a "flipped" out A1825 from 18S h44 (yellow), and base +4 stacks on G626. Adapted from[22].

While there is structural similarity between a tRNA and eRF1 the NIKS motif alone, in a position analogous to a tRNA anti-codon loop is not sufficient for stop codon selectivity. The A-site decoding complex is a discontinuous recognition platform where stop codon geometry is recognized by the side chains of conserved amino acids and RNA motifs. This intricate canonical system of translation has evolved to be highly accurate, maintaining correct amino acid incorporation, stop codon recognition, and maintenance of reading frame.

Non-canonical translation

In 1968, Francis Crick published his paper on the origin of the genetic code and proposed two theories as to why the genetic code is Universal: The Stereochemical Theory, and The Frozen Accident Theory[32]. The former posits that the physico-chemical linkage between anti-codons and amino acids mandated the evolution of codon assignments. The latter arose from the assumption that all life evolved from a universal ancestor whose random assignment of codon/amino acid is universal, unevolvable, and unchangeable (frozen). However, in 1971 bacteriophage Qβ was found to read through its stop codon to produce a C-terminally extended protein, suggesting diversification of the genetic code[33]. There are currently more than twenty natural variant genetic codes, broadly characterized by two mechanisms: codon reassignment, and codon redefinition[34].

Codon redefinition

First described in viruses, codon redefinition is a programmed or regulated translation event in which codon alterations are context dependent. Whereas codon reassignment describes the difference between genetic codes throughout a transcriptome, codon redefinition, or recoding, takes place in a subset of mRNAs with protein products assuming altered function. RNA viruses have evolved to utilize a variety of recoding mechanisms to capitalize on their relatively small genomes thereby increasing the diversity of their proteomes. While a number of recoding mechanisms have been elucidated, (e.g. selenocysteine incorporation, translational bypassing, trans-translation and stop-go), the two most extensively employed by RNA viruses are Programmed Ribosomal Frameshifting (PRF) and Programmed Ribosomal Readthrough (PRT). While both PRF and PRT alter canonical translation for expression of C-terminal extension products, the former involves a change in reading frame, and the latter allows for decoding of a stop codon and insertion of an amino acid for production of a subset of protein products[35,36]. These processes occur at low frequency and require a Recoding Stimulatory Element (RSE) for full functionality.

-1 Programmed ribosomal frameshifting

Frameshifting was discovered in 1985 when Tyler Jacks and Harold Varmus found that frameshifting was required in *Rous Sarcoma virus* (RSV) for expression of the Gag-Pol polyprotein[37]. While frameshifting most often adjusts the reading frame in either the +1 or -1 direction, the latter is more common. In all known cases, -1 PRF requires a heptameric "slippery site" on the mRNA consisting of two homopolymeric triplets, N-

NNW-WWH [where the dash denotes reading frame, and N, W, H, follow IUPAC conventions where N is any nucleotides, W is a weak base (AAA or UUU) and H is not guanine][38]. The slippery sequence may be part of a larger motif, as mutations of nucleotides immediately upstream of the heptamer can reduce efficiency, suggesting involvement of the ribosomal E-site in frameshifting[39,40]. Two additional elements that modulate the efficiency of frameshifting are a downstream RSE, and a short spacer sequence of less than 12 nucleotides located between the slippery site and the RSE. The spacer sequence provides distancing so that as the ribosome pauses at the stimulatory structure the aminoacyl- and peptidyl-tRNAs are positioned over the slippery site, where they form a small stable helix with the mRNA[37,41]. The non-wobble bases subsequently establish new pairing in the -1 frame[42].

Productive PRF is common in viral decoding of positive-sense RNA viruses to generate additional proteoforms with different functional properties. In eukaryotes, PRF is often associated with expression of the viral replicase. Thus, in many RNA viruses PRF modulates the expression of the RNA-dependent-RNA-polymerase (RdRp), and for retroviruses frameshifting between the gag-pol overlap expresses the reverse transcriptase (Gag-Pol polyprotein)[43,44]. Some plant viruses, including the luteovirus *Barley yellow dwarf virus* (BYDV), the dianthovirus *Red clover necrotic mosaic virus* (RCNMV), and the umbravirus *Pea enation mosaic virus* (PEMV) have an additional requirement for a long distance interaction between the RSE and a 3' element for efficient frameshifting[45,46,110]. Recently it has been found that frameshifting during translation of cellular mRNAs does not necessarily lead to the synthesis of a functional protein product, but rather that "abortive" frameshifting can be utilized for the down-regulation of gene

expression[47,48]. For example, in the HIV co-receptor CCR5 mRNA, a micro RNA (miRNA) stimulates a -1 PRF event directing the elongating ribosome into a termination codon activating the nonsense-mediated (NMD) mRNA pathway[49].

Viral frameshift pseudoknots

Although the slippery site alone is sufficient for low levels of frameshifting, downstream RSE are required for levels needed for biological function[50]. Secondary structures that stimulate PRF can be broadly separated into two categories: pseudoknotted motifs and all other elements[51,52]. The folding motif known as a pseudoknot was discovered in the 3´end of *Turnip yellow mosaic virus* in 1982[53]. Pseudoknotted structures have since been identified in virtually all types of RNA molecules including tRNAs, viral genomic RNAs, rRNAs, catalytic and self-splicing RNAs, mRNAs, and ribonucleoprotein complexes such as telomerase[54-58]. An RNA pseudoknot in its most basic form is base-pairing between single-stranded nucleotides within a secondary structure loop (bulge loop, interior loop or bifurcation loop) with nucleotides outside the loop. Hairpin (H-type) pseudoknots result from Watson-Crick base pairing between a single-stranded segment within a hairpin, and a sequence outside the hairpin forming at least two helical stems (S1 and S2) and two non-equivalent loops (L1 and L2) that cross the grooves of the helices. Some H-type pseudoknots contain a third loop (L3) that is often a single nucleotide in the continuous strand and can intercalate between, or extrude from, the helices[59-61]. Coaxial stacking between S1 and S2 forms a semi-continuous helix, which can be a stabilizing force for the structure[62]. Additional variation in pseudoknot topology comes from the helix-helix junction

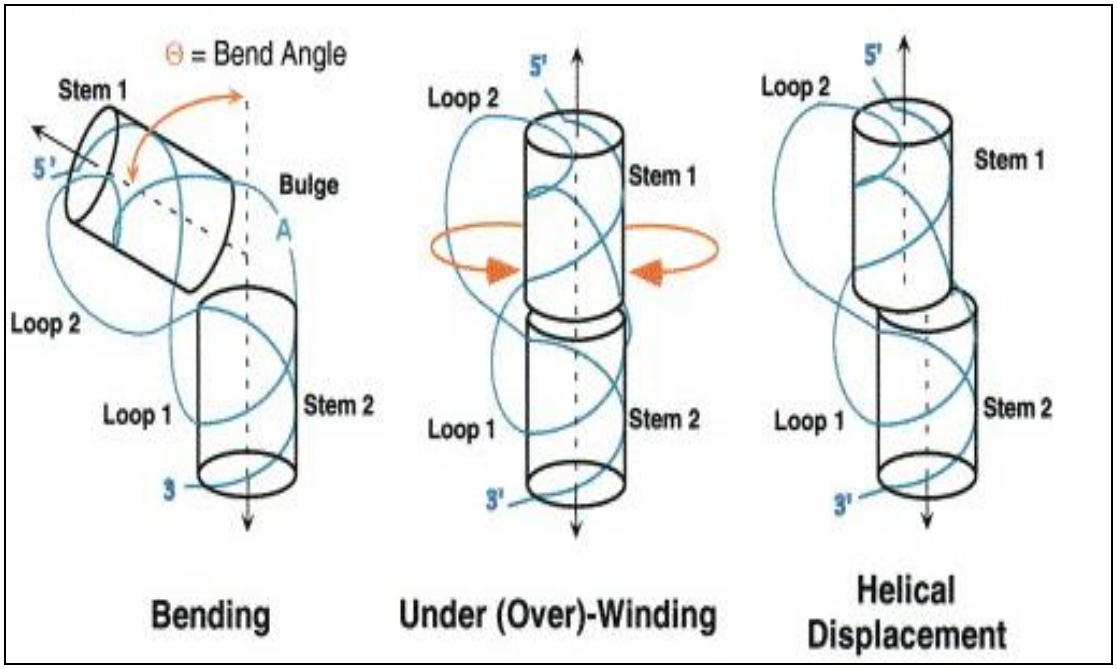


Fig. 1. 3 Cylinder models of the helical junction in an H-type pseudoknot. Structural diversity in pseudoknot conformation at the helical junction is shown. An intervening nucleotide on the continuous strand can induce bending (left), stems can coaxially stack (center) and the interhelical angle can be displaced (right). Adapted from[63].

including the interhelical angle between S1 and S2 stems, the degree of under- or over-rotation at the junction (relative to A-form helical geometry), and intervening nucleotides on the continuous strands (Fig. 1.3)[63]. *Severe acute respiratory syndrome* coronavirus (SARS-CoV) has an unusual three-stemmed pseudoknot configuration in its RSE[64]. RNA dimerization occurs through an intermolecular kissing-loop interaction between palindromic sequences in stem 3 which modulates synthesis of subgenomic and full-length RNA[65]. While a number of recoding pseudoknots exist, the H-type topology is the most common utilized by frameshifting pseudoknots. The pseudoknot provides an energetic barrier to the elongating ribosome by positioning it over the slippery site providing a 10- to 30-fold increase in frameshifting[66]. Cryo-EM at low resolution (~16 Å) imaged mammalian 80S ribosomes paused over the coronavirus *Avian infectious bronchitis virus* (IBV) frameshift signal[67]. As expected, the pseudoknot was located at the entrance to the mRNA channel making direct contact with ribosomal helicase elements. Eukaryotic Elongation Factor 2, the translocase, was trapped in the A-site, and the P-site tRNA was strongly bent in the 3´ direction, most likely pulled between the forces of translocation and the pseudoknot plug. A number of models have proposed that this tension causes the P-site tRNA to unpair and slip in the -1 direction where it subsequently re-pairs allowing translation to resume in the new reading frame. A recent study however, using single molecules and Optical Tweezers, found that the efficiency of -1 PRF was not due to resistance to unfolding; rather efficiency was correlated with an increased tendency to form alternate structural folds[68]. This conclusion was validated with SHAPE structure probing of the HIV frameshift element which demonstrated the importance of alternate helical conformations for -1PRT efficiency[69]. In addition to the

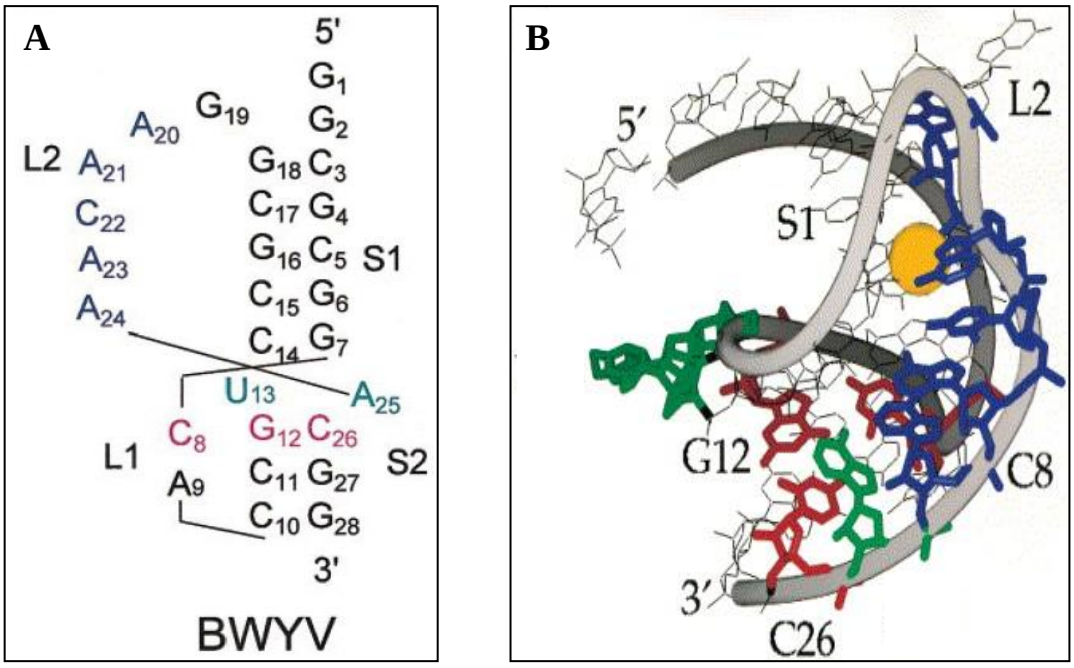


Fig. 1. 4 The BWYV pseudoknot. (A) Secondary structure showing the C8· G12- C26 base-triplet interaction (red), loop 2 bases that contact stem 1 (blue) and the U13-A25 predicted pairing (green). (B) The pseudoknot as determined by X-ray crystallography with nucleotides colored as in (A); the Na^+ ion (yellow) bound between loop L2 and stem S1. Adapted from[63].

stabilizing effects of bound cations on the tertiary stability of the pseudoknot, protonation (pH dependent) can significantly stabilize the molecule, with an increase in stability tied to an increase in frameshifting efficiency[70]. The crystal structure of the H-type pseudoknot of Beet western yellows virus (BWYV) shows a Na^+ bound between loop 2 and stem 1; substitution of K^+ or NH_4^+ showed no preference for monovalent ions (Fig. 1.4). The BWYV pseudoknot has a base-triple $C8^+ \cdot$ G12$\cdot$ C26 (C8 in loop 1 and G12-C26 base pair in stem 2) where C8 is protonated at N3 to form an H-bond to the O6 of G12. A deprotonated C8 can form two H-bonds with G12-C26, while a protonated C8 can form three H-bonds stabilizing the triple interaction modulating the loop closing energetics for pseudoknot folding.

Early studies broadly sub-divided viral RSE pseudoknotted motifs into three structural classes: (1) *Mouse mammary tumor virus* (MMTV) gag-pro like, (2) IBV 1a-1b junction like, and (3) *Beet western yellows virus* (BWYV) P1-P2 junction like[63]. The first stimulatory RNA element investigated in detail was the gag-pro -1 PRF of the Betaretrovirus MMTV (Fig. 1.5)[59,71,72]. The retroviral S1 stem is relatively short with five base-pairs and an eight nucleotide L2 loop; the S2 stem has six base-pairs and is crossed by a two nucleotide L2 loop. An unpaired adenosine is wedged between the two helical stems acting like a hinge inducing a strong bend (~70° angle) of S1 towards S2. Mutations in the pseudoknot that maintained the bent (U13C) conformation had no effect on frameshifting; however, a deletion of the adenosine (U13C + ΔA14) caused the two stems to coaxially stack. Frameshifting was subsequently reduced by 60% suggesting a correlation between the angular conformation and frameshifting efficiency[73,74]. Unlike the retrovirus gag-pol pseudoknots the IBV-like pseudoknots, which include those

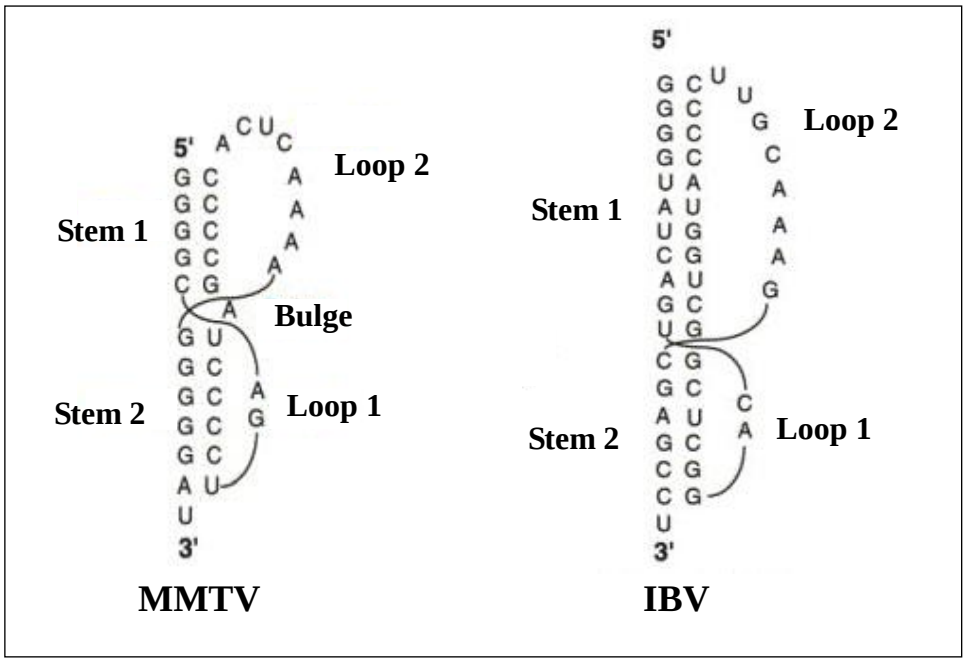


Fig. 1. 5. Secondary structure of H-type pseudoknots of MMTV and IBV. MMTV-type pseudoknot (left) characterized by a shorter stem S1 and loop L2, with a bulged nucleotide at the helix-helix junction. IBV type pseudoknot (right) with a longer stem S1 and loop S2 with no inter-helical bulge. Adapted from[63].

derived from coronaviruses, arteriviruses, and toroviruses, are characterized by both a long S1 stem of 11-12 base-pairs and a long L2 loop (Fig. 1.5)[43]. Similar to MMTV the IBV-type pseudoknots are associated with a six base pair stem S2, and two nucleotide loop L1. When placed downstream of the slippery sequence U-UUA-AAC these pseudoknots are highly efficient frameshift enhancers with frameshifting efficiencies reaching 40% and above[63]. Mutational analysis of the S1 stem found a minimum requirement of 11 base pairs for efficient frameshifting *in vitro*, and this requirement was not dependent upon the stability of the stem[75]. Additionally, a base substitution of G-U to G-C at the helical junction strongly stimulated frameshifting suggesting the junction is coaxially stacked, unlike the bent conformation of MMTV. Reducing the S1 stem of IBV to 6 bp renders the pseudoknot inactive, however, the addition of an unpaired adenosine at the helical junction and another at the 3' terminal position of loop L2 (mimicking MMTV) restores frameshifting efficiency[76].

Programmed ribosomal readthrough

As detailed above, ribosomes terminate translation with remarkable fidelity when one of three stop codons is encountered. With PRT the ribosome continues elongation by decoding the stop codon using aminoacylated suppressor tRNAs that have anticodons that can base pair with stop codons. These suppressor tRNAs are naturally occurring in the cellular environment and are primarily used in canonical translation for recognition of their cognate sense codons[77]. As such, additional signals embedded within the mRNA are required to stimulate readthrough by placing the stop codon in an unfavorable termination context. mRNA intrinsic signals can include the stop codon itself, the

surrounding nucleotide context, RNA secondary structure, tertiary contacts and long distance RNA:RNA interactions. As discussed earlier, flipped out 18S bases A1824 and A1825 allow codon recognition by cognate tRNAs, and a tucked-in A1825 with a flipped out A1824 facilitates recognition of stop codons by eRF1. For PRT, eukaryotic initiation factor 3 (eIF3) in complex with eRF1 specifically recognizes mismatches in the wobble position between the near-cognate tRNA and a stop codon in an unfavorable termination context[78,79]. Interference by eIF3 would result in premature ejection of the eRF1-eRF3-GTP complex. Since there are no cognate tRNAs for stop codons, there would be no competition for binding by near-cognate tRNAs. In addition to the codon-anticodon affinity of the suppressor tRNA, the efficiency of PRT can be up- or down-regulated depending upon the extent of base modifications within or 3' of the anticodon loop. For example, post-transcriptional isomerization of the uridine in a GUA anti-codon to pseudouridine (GΨA) is required for UAG suppression in the *Tobacco mosaic virus* (TMV) context (Fig. 1.6)[80].

Nucleotide context in both the 5' and 3' directions has been strongly implicated in readthrough efficiency for a number of viruses. The strongest 5' stimulatory signal involves two adjacent adenines. Two studies of animal and plant viruses analyzed nucleotides surrounding the stop codon including the ultimate and penultimate positions. Beier and Grim found of 53 sequences, 74% had adenine in the -1 position, 70% in the -2 position, and 55% in both positions[77]. Similarly Harrell et al. found of 91 sequences 71% had adenine in the -1 position, 76% in the -2 position and 55% in both positions[81]. Subsequent research specifically addressing the 5' major determinants likewise found that the majority of sequences involving two adjacent adenines immediately upstream of the

stop codon had the highest percentage of readthrough[82]. These authors propose the stacking potential of the tandem adenines induce structural modifications in the ribosomal P-site, which is transmitted to the A-site through the B2a eukaryotic bridge potentially interfering with recognition of the stop codon by the termination complex. The 3' nucleotide context has a much more pronounced effect on PRT. The $tRNA^{Tyr}$ GΨA anticodon sequence fails to suppress the TMV UAG stop codon without a specific six base nucleotide sequence immediately downstream in tRNA depleted wheat germ extract (WGE)[83]. Indeed, mutations of individual bases at each position either abolished or significantly reduced readthrough.

The hexanucleotide consensus sequence, CAR-YYA (R=purine, Y=pyrimidine) is a powerful PRT stimulator that can function independently to confer readthrough in heterologous RNA in the absence of an RSE[84]. Members of the genus *Tobamovirus* use this sequence for expression of replicase associated proteins, while Benyviruses, Pomoviruses and Tymoviruses have coat protein (CP) extensions[77]. While the mechanism of action is not currently known, base pairing between the consensus sequence and 18S rRNA may destabilize secondary structure inhibiting the binding of release factors (Fig. 1.7)[85]. A less stringent trinucleotide consensus sequence, CGG following a UGA stop codon, is found in members of the Pecluvirus, Furovirus, Pomovirus, Tobravirus, Coltivirus and Alphavirus genera. Single base alterations in this sequence for the Tobravirus *Tobacco rattle virus* had little effect upon translation in WGE, however, the effect was pronounced when two or three bases were mutated simultaneously[86]. Results for the Coltivirus *Colorado tick fever virus* (CTFV) were similar with a deletion of the entire CGG sequence reducing readthrough to 5% of wild

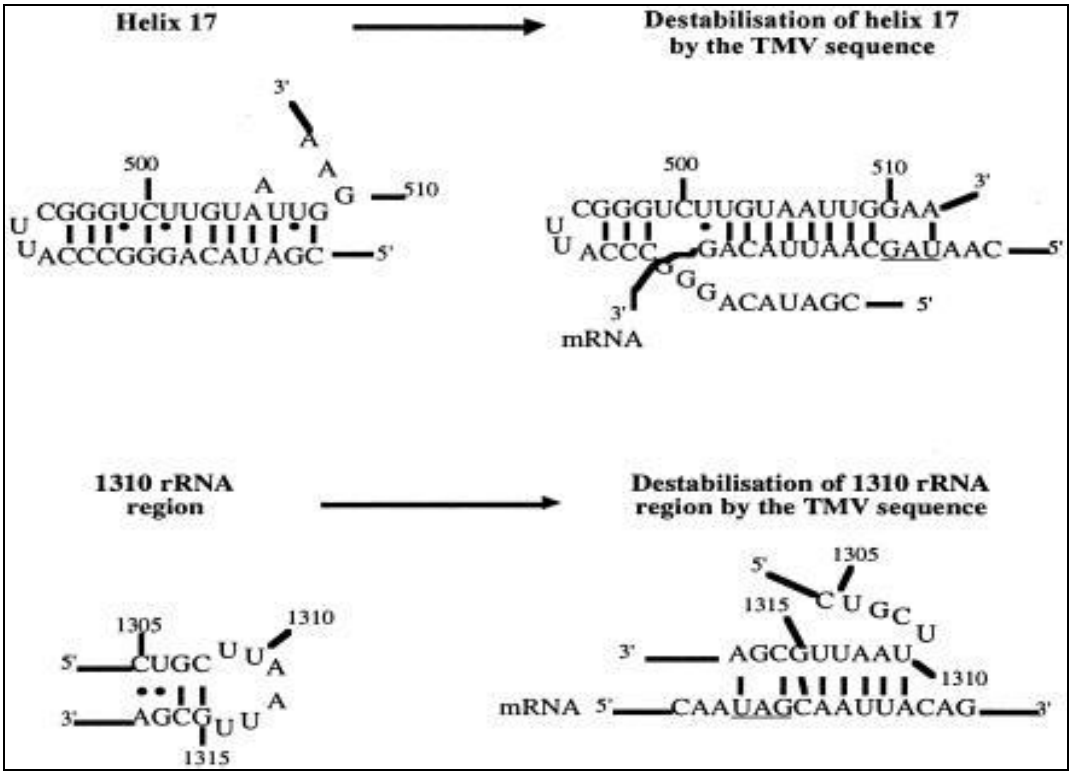


Fig. 1. 6 Base pairing between the TMV readthrough motif and the 18S rRNA. Possible pairing with h17 (top) or the 1310 region (bottom) by TMV mRNA could destabilize 18S rRNA stimulating readthrough. Unconventional (G-U) base pairing denoted by (*). Figure from[85].

type (WT)[87]. CTFV readthrough is also both UGA specific and dependent upon the stem region of a stable SL structure located 8 nucleotides downstream. Interestingly, unlike -1 PRF, there was no ribosomal pausing associated with translation of the CTFV recoding signal suggesting differing mechanisms for frameshifting and readthrough. Early work on Alphaviruses reported a single cytidine residue 3'-adjacent to the stop codon was the major determinant for efficient readthrough of the UGA stop codon for expression of the viral RdRp in rabbit reticulocyte lysates[88]. More recently however, it has been determined that a large SL structure located 8-12 bases downstream of the UGA stop codon increases readthrough by up to tenfold in Alphaviruses, and is likely present in Pecluvirus, Furovirus, Pomovirus and Tobravirus genera as well[89].

Members of the family *Leuteoviridae*, small positive-sense RNA plant viruses, have a CP-extension readthrough polypeptide product required for aphid transmission. PRT stimulatory elements for these viruses involve a long distance RNA:RNA interaction between a Cytidine-rich repeat (**CC**NNNN **CC**NNNN **CC**NNNN, N=any nucleotide) twenty bases 3' of the stop codon, and a distal sequence 700 to 750 nt downstream[90]. Similarly, members of *Tombusviridae* also require a long distance RNA:RNA interaction between two elements for readthrough of the amber stop codon for RdRp expression (with the exception of Dianthovirus which has a frameshift)[84]. *Tobacco necrosis virus-D* (TNV-D), the type member of the genus Betancecrovirus, has an extended RSE SL structure just 3' of the stop codon with a bulge loop (PRTE, proximal readthrough element) that participates in a long-range interaction with a sequence in the 3' untranslated region (UTR) of its genome (DRTE, distal readthrough element) (Fig. 1.8)[91].

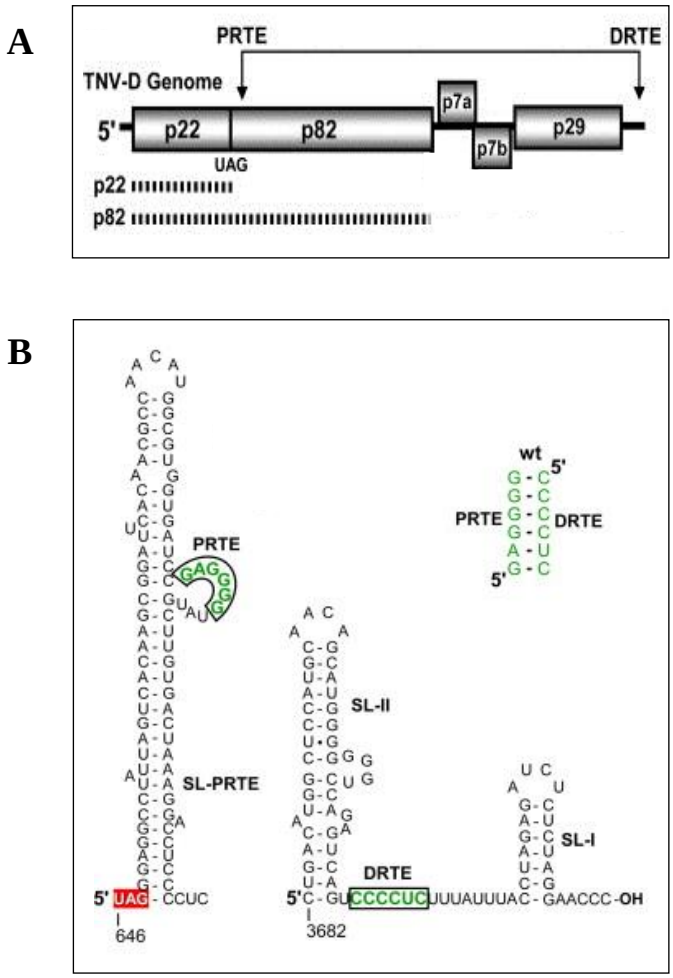


Fig. 1. 7 Requirement for a long distance interaction for readthrough in TNV-D.

A) Schematic of the TNV-D genome with encoded proteins boxed. Hashed lines delineate p22, a replicase associated protein and the readthrough product p88, the RdRp. (B) Mfold-predicted secondary structure of the SL-PRTE and 3' end of TNV-D. Complimentary sequences between the PRTE and DRTE (green) and the stop codon (red) are shown. Adapted from[91].

These elements are stop codon dependent, and are readthrough compatible only with UAG and UGA, but not UAA. For the tombusvirid *Carnation Italian ringspot virus* (CIRV), the addition of the BYDV heptanucleotide slippery sequence GGGUUU upstream of the stop codon allowed frameshifting for expression of RdRp both *in vitro* in WGE and *in vivo* in protoplasts (albeit at a much reduced level)[91]. This suggests RSE have similar functions in both -1 PRF and PRT.

Alternate conformations

In addition to its role in the genetic transfer of information from DNA to proteins, RNA is able to regulate a diverse number of cellular processes[92]. Many of these processes involve switching between conformations with discrete functions as exemplified by riboswitches, which act as on/off switches when bound by small molecule metabolites. The binding of the ligand to the riboswitch sensor induces a change in the secondary structure regulating transcription/translation downstream in the mRNA. Indeed, it is this structural plasticity in the RNA landscape which allows it to "sense" and respond to cellular environments with varying conditions.

While bacterial riboswitches are highly characterized with new classes being discovered each year, the knowledge of similar regulatory mechanisms for viral genomes is in an emergent state. For pathogenic viruses, such regulatory secondary structures can represent promising therapeutic targets. This is well exemplified in Hepatitis C virus (HCV), a small (9.7 kb) positive-sense infectious RNA hepacivirus which persists in the liver leading to an increased risk of cirrhosis and hepatocellular carcinoma[93]. There is

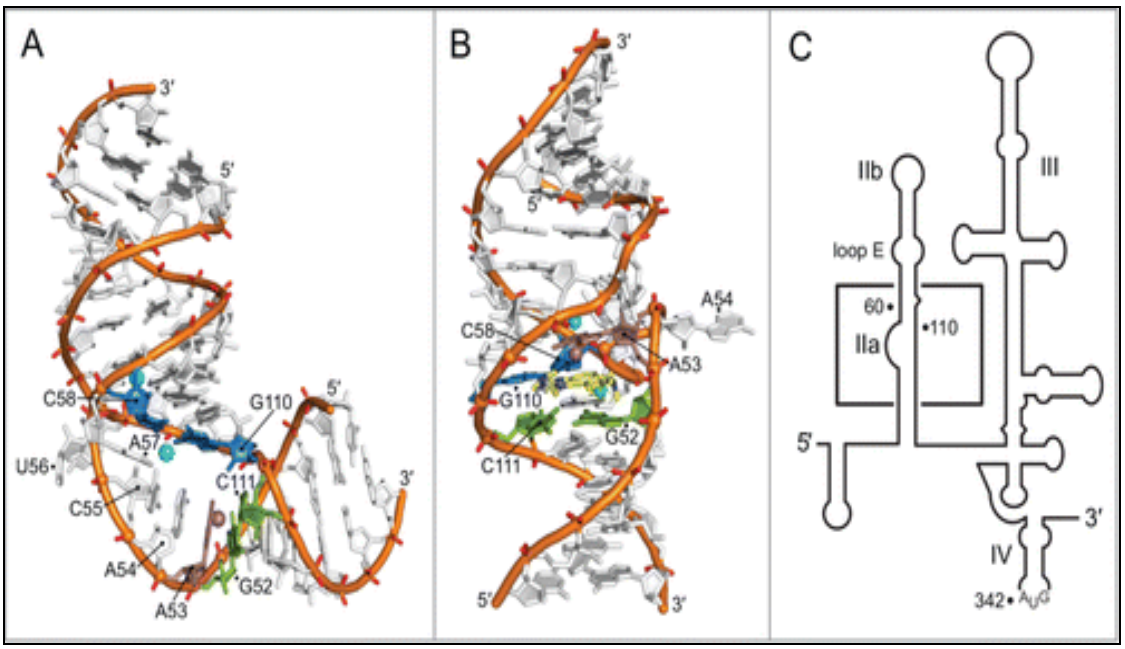


Fig. 1. 8 Structure of the HCV IRES subdomain IIa. (A) Crystal structure of the ligand-free structure exhibiting a bent conformation. Closing base pairs highlighted (top-blue, bottom-green) with a stabilizing Mg^{+} ion shown near base A53 (mauve). (B) Crystal structure of the ligand-captured structure adopting an elongated conformation, ligand shown (yellow). (C) Subdomain IIa outlined within the secondary structure of the HCV IRES element in the 5' UTR. Figure from[98].

currently no known vaccine against HCV making it an ideal candidate for anti-viral therapeutics. The 5' UTR is uncapped with translation initiation being driven by a highly structured Internal Ribosome Entry Site (IRES) which is composed of several domains[94,95]. The highly conserved domain II element contains an asymmetrical bulge on the lower stem, domain IIa, which is a target for viral benzimidazole translation inhibitors[96]. Domain IIa forms a deep pocket stabilized by magnesium ions which assumes a curved topology with a 90° bend (Fig. 1.5)[97]. Ligand binding induces a conformational switch to a more extended, linear conformation which inhibits translation elongation[99]. It is important to note this switch does not involve an alteration of base pairing. The domain IIa ligand pocket also selectively recognizes guanine suggesting a guanosine within a viral RNA sequence or rRNA could likewise trigger a similar 5' conformational switch thereby initiating a transition between viral processes.

Positive-sense RNA viruses regulate diverse "life" processes using their linear sequence, secondary structure and tertiary interactions. Many of these processes, such as the 5' to 3' procession of translation versus the 3' to 5' procession of minus strand synthesis, are mutually exclusive yet rely on the same template genome. Conformational switches between required functional motifs represent an additional layer in gene regulation for the optimal usage of small genomes. Many plant viruses contain relatively small 5' UTRs necessitating many regulatory elements to be located within the coding region and 3' UTRs[100]. Arguably the most widely characterized viral RNA switch occurs in the 3' terminus of *Alfalfa mosaic virus* (AMV) where the CP mediates the transition between replication and translation via the conformation changes of a 3' tRNA shaped element (TSS)(Fig. 1.9)[101,102]. AMV is an Ilarvirus, with a positive-sense RNA genome

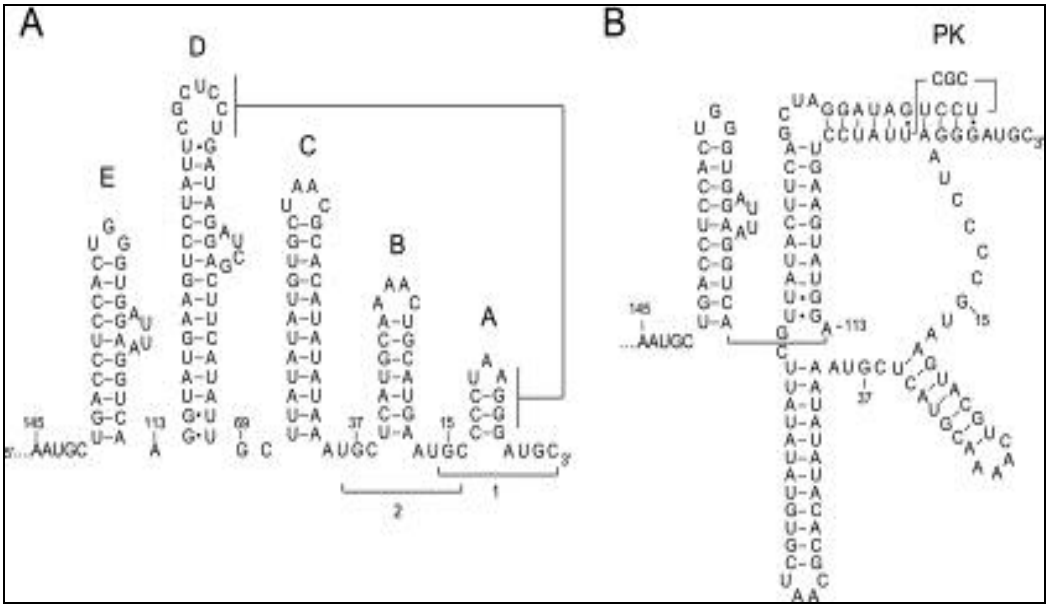


Fig. 1. 9 Conformational switch in the 3' terminal region of AMV. (A) CPB conformer with the two CP binding sites indicated by brackets; base pairing between loop D and stem A that promotes the conformational switch to the TL conformation. (B) Secondary structure of the TL conformer with pseudoknot shown. Figure from[104]

which requires the CP for infection. Preliminary reports detailed two conformations: a tRNA-like conformation required for replication and a CP binding form (CPB) necessary for efficient translation. The original findings were challenged by a study reporting replication was stimulated by the addition of CP in a dose-dependent manner, suggesting the CPB modulated replication rather than translation[103,104]. Most recently however, a systematic study shifting the population distribution of TL and CPB was carried out with three outcomes supporting the original findings. 1) Recognition (adenylation) of the 3' UTR by a tRNA-specific enzyme (ATP:tRNA nucleotidyl transferase, a CCA-adding enzyme) was significantly enhanced with mutations stabilizing the TL conformation; 2) Recognition of the 3' UTR by the viral replicase (minus strand synthesis) was inhibited by shifting the population towards the CPB conformation and 3) recognition of the 3' UTR by both the CCA-adding enzyme and the replicase was inhibited with the addition of CP[105]. These data combined suggest that the conformation for minus-strand synthesis in AMV is the TL, translation is the CPB, and that the genome must undergo a conformational shift to transition between the processes of replication and translation.

A number of cis-acting RNA elements mediate the viral life cycle of CIRV through coordinated interactions between RNA elements. CIRV has a positive-sense RNA genome of approximately 4.7 kb and encodes five functional proteins[106]. The 5' proximal open reading frame (ORF) encodes a replicase associated protein (p36), and its readthrough-polypeptide (p95) encodes the viral RdRp. Both proteins are directly translated from the viral genome and both are required for viral replication. As is the case for all tombusvirids, a long distance RNA:RNA interaction between an asymmetrical bulge sequence 3' proximal to the p36 stop codon (PRTE) and a hairpin

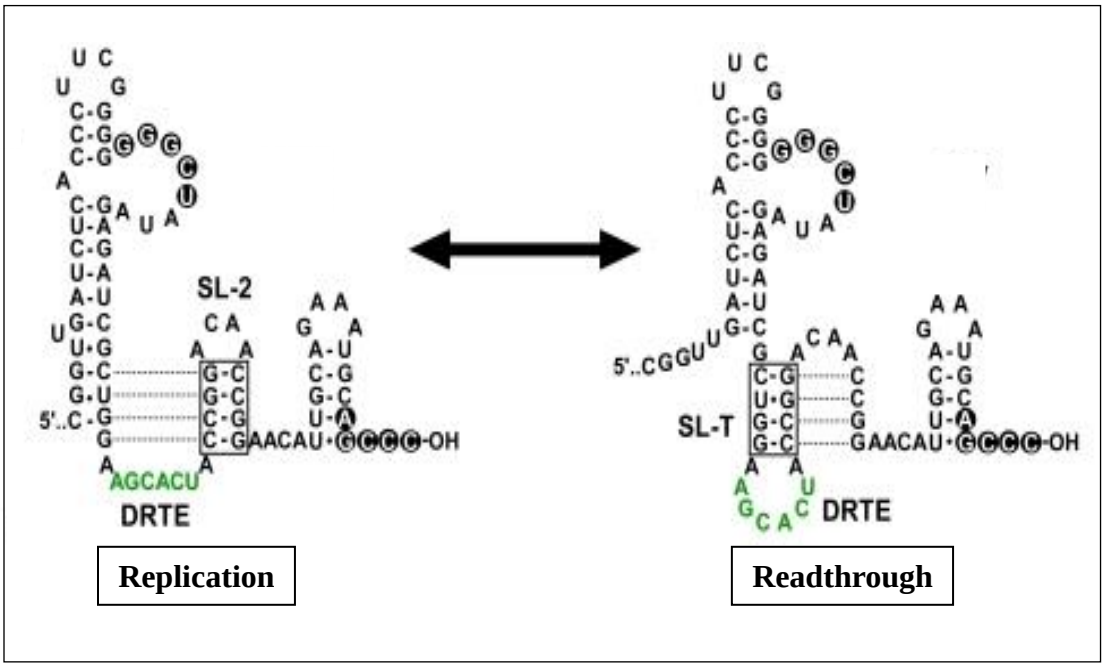


Fig. 1. 10 Alternate conformations driving switch between replication and translation in the 3' UTR of CIRV. Schematic of the secondary structure of two possible conformations surrounding DRTE (green) sequence is shown. The replication competent conformation (left) with formation of the SL-2 element (stem boxed) is shown, along with dotted lines denoting alternate pairing, and a 3' terminal pseudoknot in black circles. (B) The readthrough permissive conformation (right) is shown. Figure from [84].

terminal loop sequence within the 3' UTR (SL-T), is required for readthrough and expression of RdRp[84]. A separate hairpin terminal loop located within the 3' UTR, SL-2, is required for replication. Formation of either SL-T or SL-2 is mutually exclusive. Mutations preferentially stabilizing SL-2 down-regulate expression of RdRp, while stabilization of SL-T up-regulate RdRp expression strongly suggesting that the mechanism for the transition between replication and translation lies in the inter-conversion between these two alternate conformations (Fig. 1.10)[84].

Regulatory mechanisms for the maintenance of recoding protein ratios

The precise ratio of termination to extension products in recoding, including both -1 PRF and PRT, is critical for efficient viral accumulation. For example, mutations that alter the gag-pol ratio of *Human Immunodeficiency Virus Type 1* (HIV-1) reduce infectivity of viral progeny and destabilize RNA dimers[107]. While it is understood that the frequency of recoding is regulated by both *cis*- and *trans*-acting elements, very little mechanistic information is available for how the critical protein ratio is maintained.

A cis-acting element upstream of the SARS-CoV frameshift signal attenuates recoding efficiency by up to 50% *in vivo* and can function in heterologous RNA[108]. This stable secondary structure impedes ribosome processivity causing a fraction of ribosomes to dissociate from the mRNA template thereby reducing the number of ribosomes that actually encounter the frameshift signal[109]. This potential to form a stable element upstream of the recoding site is phylogenetically conserved in all coronaviruses surveyed suggesting an evolutionarily conserved mechanism for modulation of -1 PRF. The Umbravirus *Pea enation mosaic virus* (PEMV) utilizes a similar mechanism for the

modulation of -1 PRF. While deletion of the upstream element (SLA) had no obvious effect of frameshifting in WGE, in combination with the deletion of the RSE, the residual frameshifting efficiency of the slippery sequence was enhanced by 72%[110]. This suggests that SLA downregulates frameshifting in the absence of RSE formation. In the gammaretrovirus *Murine Leukemia virus* (MLV) a protonation-dependent switch in the RSE regulates PRT[111]. At physiological pH, readthrough of the gag stop codon for expression of the gag-pol fusion protein (reverse transcriptase) is~6%; a decrease in pH produces a corresponding decrease in the readthrough product. Base substitutions that eliminated protonation as observed by NMR likewise decreased readthrough levels. These findings correlate well with the pH-dependent stability of the frameshifting pseudoknot in BWYV as mentioned above suggesting a mechanism in which recoding levels for the virus are dependent upon the cellular environment within the host.

Turnip crinkle virus, a model system

TCV is the type member of the genus Carmovirus in the family *Tombusviridae* and has been used extensively as a model system for the identification and characterization of cis-acting elements involved in replication and translation. TCV has a positive-sense RNA genome of 4053 bases that is not capped or polyadenylated, with five overlapping open reading frames encoding proteins involved in replication, movement, packaging and host silencing[112,113]. The N-terminal protein p28 is a replicase associated protein with its readthrough product p88 encoding the viral RdRp. Both are expressed from the genomic RNA (gRNA). Two movement proteins, p8 and p9, are expressed from subgenomic RNA 1, and the CP (p38) is expressed from subgenomic RNA 2 (Fig. 1.11). It has been

discovered recently that p8, p9 and CP can all be expressed via internal initiation from the gRNA (Simon, unpublished), suggesting expression of these proteins from subgenomic RNAs has a temporal requirement.

Structural elements within the 3' UTR of carmoviruses

The highly structured 3' UTRs of carmoviruses contain a number of conserved elements at the terminal end (Fig. 1.12). The TCV 253 base 3' UTR is a highly interactive, complex network of secondary structures and tertiary interactions that regulate much of the virus life cycle (Fig. 1.12)[114,115]. These interconnecting elements include a tRNA-shaped structure (TSS), consisting of H4a, H4b, H5, Ψ_2 and Ψ_3, that forms a stable scaffold for the binding of ribosomes and functions as a Cap-Independent Translational Enhancer (CITE)[116,117]. Mutations in the TSS that negatively impact translation have parallel negative effects on ribosome binding. An element upstream of the TSS, H4, plays a multi-purpose role within the 3' UTR. Deletion of H4 reduced translation of reporter constructs *in vivo* by 85%, while mutation of a critical individual residue in the terminal loop reduced *in vitro* transcription by 70% demonstrating a role in both translation and replication[114,117]. Binding of the RdRp to various size 3' fragments had significant impacts on the structure of both the TSS and H4 causing an extensive conformational shift (Fig. 1.12)[118]. This disruption of translational elements was suggested to abrogate ribosome binding to the TSS, which could potentiate the transition between translation and replication. The 3' terminal hairpin (Pr), engages in a long distance kissing-loop interaction with a bulge loop on the RSE. This interaction is

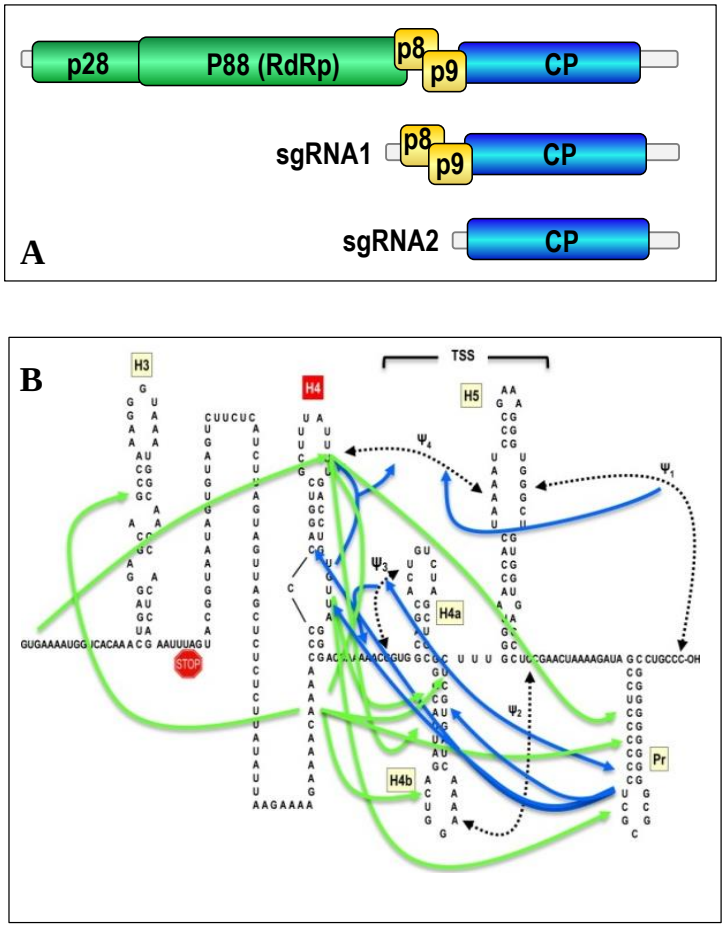


Fig. 1. 11 TCV genome and 3' UTR interactions. (A) Schematic depiction of the genomic RNA and two sub-genomic RNAs. Replicase proteins (green), movement proteins (yellow) and CP (blue) are denoted. (B) Secondary and tertiary interactions in the highly interactive 3' UTR. The CP termination codon (red stop sign), and TSS (bracket), and pseudoknots (dotted lines) are shown. Interactions between bases and/or elements identified through mutational analysis are designated by blue lines. Interactions identified through a second-site mutational analysis that arose from primary site mutations designated by green lines. Figure B from[115].

required for readthrough of the p28 stop codon and expression of the RdRp[84]. This interaction is phylogenetically conserved within *Tombusviridae* with the exception of dianthoviruses, which use the interaction to facilitate frameshifting. The positioning of the 3' readthrough element varies within the carmoviruses. Of the 15 viruses in the genus, 11 use the 3' Pr loop, with the remaining 3 using the terminal loop of H5 for the interaction. Mutations in the TCV Pr loop reduced *in vitro* transcription and RdRp binding to a 3' terminal fragment suggesting that in addition to its role in recoding, the Pr is also important for replication in TCV[119]. The positioning of a recoding enhancer element within the 3' UTR suggests a mechanism to separate the incompatible processes of replication and translation. Prior to RdRp synthesis the 3' end is effectively sequestered by long distance interaction with the RSE maintaining a translation competent genomic conformation. Conversely, readthrough of the amber stop codon would dissociate the Pr from the RSE down-regulating translation[84].

Two additional carmoviruses are known to have 3' CITEs: *Saguaro cactus virus* (SCV) and *Melon necrotic spot virus* (MNSV). SCV has a *Panicum mosaic virus* (PMV) type Translational Enhancer (PTE - this was the first such element characterized) that spans the junction between the CP and 3' UTR (Fig. 1.13)[120,121]. Seven of the 15 carmoviruses have, or are predicted to have PTE type 3' CITES. All PTE elements contain a terminal bifurcation, as well as an asymmetrical guanylate-rich bulge that forms a pseudoknot with sequences between the upper two hairpins. The PTE 5’ terminal hairpin engages in a long distance RNA:RNA interaction with a hairpin in the 5' coding region of the gRNA, and the 5' UTR of sgRNA2[121]. MNSV has an I-shaped (ISS) CITE in its 3' UTR which is known to interact eIF4E, and plants resistant to MNSV encode an

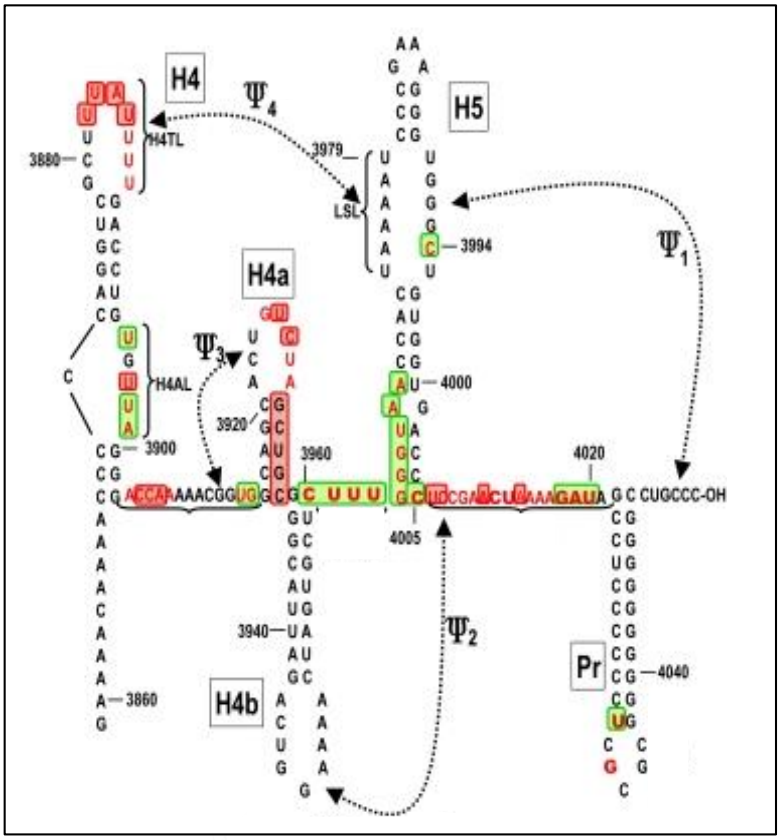


Fig. 1. 12 RdRp interaction with the TCV 3' end causes a widespread conformational shift. Location of hairpins, pseudoknots (dotted lines) and subregions (brackets) are shown. Red bases are susceptible to in-line cleavage in the presence of RdRp, and residues boxed in red show enhanced cleavage. Conversely, residues boxed in green show reduced cleavage. Figure from[118].

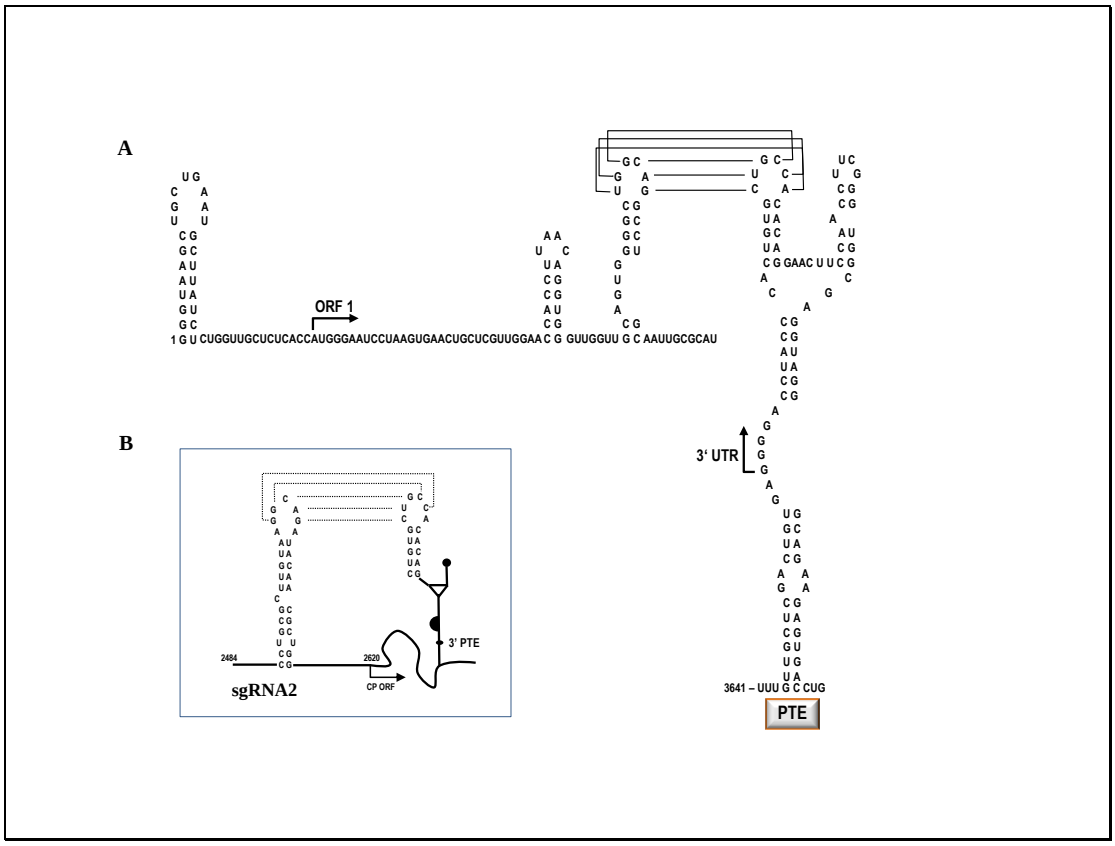


Fig. 1. 13 A kissing-loop interaction between the PTE and 5' ends of SCV gRNA and sgRNA2 enhances translation. (A) The RNA:RNA interaction with 6 base complementarity between the 5' end and the PTE. Start codon for ORF 1 and beginning of the 3' UTR are noted with arrows. (B) The 5 base RNA:RNA interaction between an element within the 5' UTR of sgRNA2 and the PTE is shown. The beginning of the CP coding region is noted with an arrow. Figure adapted[121].

eIF4E variant. While a defined pairing partner within the 5' region has not been identified, the 5' UTR is required for efficient translation suggesting an interaction.

Project rationale

Ribosomal recoding in TCV, Chapter 2

The study of viral recoding is relevant for a number of reasons. Viruses are the simplest, smallest form of life existing on the borderline between the biological and non-biological worlds. Because they complete their life cycles inside of cells, much can be revealed about complex biological processes within those very cells. Indeed, both frameshifting and readthrough were discovered in viruses, and are now known to regulate a number of cellular processes including mammalian tumorigenesis, as exemplified by the proangiogenic activity of VEGF-A, and the antiangiogenic properties of its readthrough product VEGF-Ax [122]. On the other side of the coin, ten percent of inherited diseases are caused by premature termination codon mutations that lead to the production of truncated non-functional proteins and mRNA degradation[123]. Using small, well characterized viruses as model systems for the study of recoding can provide valuable information on more complex systems that can potentially stimulate or attenuate readthrough. Lastly, viruses are a major cause of disease for both animals and crops. For humans, viruses such as Alphaviruses and Flaviviruses cause widespread morbidity and mortality. As vaccines are lacking for many of these, therapeutics aimed at critical RNA structural elements are attractive, but the prerequisite is the preliminary identification of functional targets.

The goal of this study is to use a small model plant virus to investigate mechanisms of viral recoding *in vivo* and *in vitro* using the entire genomic RNA specifically focusing on the structural dynamics involved in transitioning between canonical (termination) and non-canonical (readthrough) life cycle processes. To this end, structure probing using Selective 2'-hydroxyl Acylation analyzed by Primer Extension (SHAPE) in combination with mutational analyses was utilized for the identification of RSE in carmoviruses.

Generation of an infectious clone for *Calibrachoa mottle virus*, Chapter 3

CbMV was identified as a member of the genus *Carmovirus*, within the family *Tombusviridae*. While the CbMV genome had been sequenced and characterized, no infectious clone had been generated. Full-length infectious clones are powerful tools for the study of gene functions, gene expression, virus-host interactions, population dynamics and processes related to the viral life cycle. The ease with which these clones can be manipulated by site-directed mutagenesis or insertion of reporter genes for functional analysis has significantly impacted the field of molecular biology over the last three decades.

Toward that goal I generated an infectious clone of CbMV. Using SHAPE structure probing, phylogenetic comparisons and in silico analyses I compare and contrast structural elements and motifs between CbMV and members of the genus Carmovirus.

Chapter 2: Regulation of Translational Recoding in the Genus *Carmovirus*.

Introduction

Positive-sense RNA viruses employ non-canonical mechanisms for gene expression due to the relatively small size of their genomes. Two highly utilized methods, -1 programmed ribosomal frameshifting (-1 PRF) and programmed ribosomal readthrough (PRT), circumvent stop codons for expression of carboxy-terminal extension products. In -1 PRF, the elongating ribosome shifts back one residue at a seven residue adenylate- or uridylate-rich slippery sequence upstream of the stop codon, and then continues translating in the new reading frame generating a fusion polypeptide. In PRT, a suppressor tRNA decodes the stop codon and translation continues in the same reading frame. These processes are mainly used by viruses to express their RdRp or a CP extension product that is necessary for vector transmission and/or virion assembly[77]. Ribosome recoding leads to the tightly controlled synthesis of 5-10% fusion products, and maintaining this level is critical for efficient virus amplification[109,124,125].

While the slippery sequence is limited to frameshifting, both -1 PRF and PRT require RSE downstream of the stop codon for efficient recoding. Early studies identified a stable stem-loop in RSV and a pseudoknotted structure in IBV able to enhance frameshifting levels[54,126]. For -1 PRF, pseudoknotted structures are proposed to compromise the translocation process by forming a physical barrier that blocks the mRNA entrance channel, impeding mechanical unwinding of the mRNA by the translating ribosome[67]. Stability of the stem of simple RSE hairpins has also been

suggested to serve a similar function in promoting PRT and -1 PRF [87,89,110]. However, no consistent correlation has been found between frameshifting efficiency in reporter constructs and the mechanical force required to unfold RSE using optical tweezers [68,127]. Instead, the suggestion has been made that frameshifting efficiency correlates with conformational plasticity and the tendency of the pseudoknot to form alternate local folds. Protonation of the H-type pseudoknot RSE of MLV was suggested as important for formation of the readthrough permissive conformation[111]. Conversely, hairpins upstream of the recoding site have been reported that depress frameshifting levels in coronaviruses and the umbravirus *Pea enation mosaic* virus (PEMV)[108,110]. Strikingly, in the absence of the upstream hairpin, the PEMV RSE is no longer required to stimulate frameshifting from the slippery site, suggesting that upstream hairpins may play significant roles in -1PRF attenuation.

For this study, I assessed PRT within full length viral gRNA using TCV because of its small size, well-characterized genome and the simplicity of studying virus accumulation in single cells. TCV is a member of the *Carmovirus* genus within the *Tombusviridae* family, and has a 4053 nucleotide (nt) positive-sense RNA genome containing five overlapping open reading frames (ORFs) (Fig. 1A). The 5' proximal ORF (p28, a replicase associated protein) and the p88 RdRp readthrough product are translated from the gRNA, which lacks a 5' cap and 3' poly-A tail[112,113]. The TCV RSE, which is located just downstream of the p28 amber termination codon, contains two asymmetric bulges, one of which pairs with the terminal loop of a 3' end proximal hairpin[84]. Similar long-distance RNA:RNA interactions are required for efficient -1 PRF in PEMV, BYDV, and RCNMV, and for PRT in CIRV and TNV-D[45,46,84,110,128]. Since internal H-type

pseudoknots have not been reported in RSE from any of these plant viruses, proposals have been made that the long-distance RNA bridge forms an atypical pseudoknot that functionally replaces the internal pseudoknot that is common in animal virus RSE[2,46,129].

Here I report that, in addition to the long-distance interaction between the RSE and 3' sequences, the TCV RSE contains an internal H-type pseudoknot and both tertiary interactions are important for competent ribosome recoding. I also demonstrate that a short conserved RSE sequence just downstream from the recoding site is required for efficient PRT and this sequence is conserved in gamma retroviruses. Most significantly, I provide structural evidence that the lower stem of the TCV RSE adopts an alternate conformation that extends an upstream, phylogenetically conserved hairpin and this alternate structure is the predominant form the RSE *in vitro* and in plant cells. These data reveal that at least three RSE conformations exist in TCV, showcasing the dynamic nature of RSE.

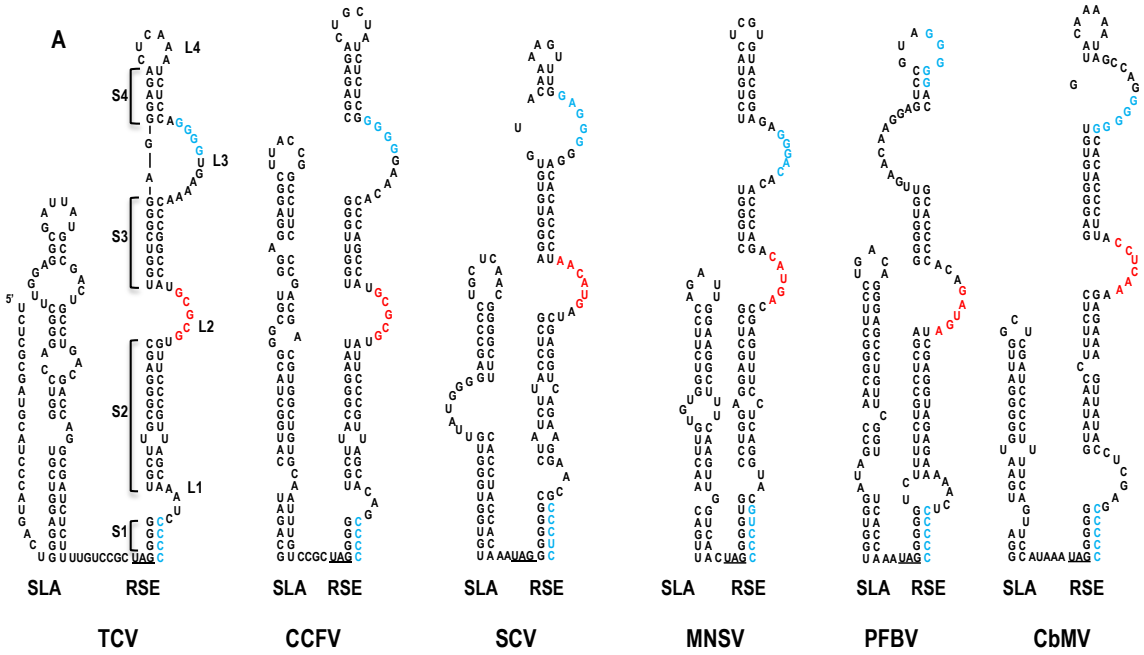
Results

The TCV RSE contains an internal pseudoknot

Computational and phylogenetic approaches indicate that members of the *Tombusviridae* contain two hairpins in the vicinity of the recoding site irrespective of whether they use frameshifting or readthrough to synthesize their RdRp (Fig. 2.1). The 5' hairpin, known as Stem-Loop A (SLA) is within 10 nt of the amber termination codon. The 3' hairpin, previously identified as an RSE, is an unbranched structure with two large asymmetric bulge loops (L2 and L3), with L2 engaged in the long-distance interaction with sequences at the 3' end[84]. The RSE is located just downstream of the UAG, with the

guanylate participating in the terminal base-pair at the base of the G:C-rich lower stem (S1).

Although RSE pseudoknots have not been reported in tombusvirids, in silico analyses of TCV and other carmovirus RSE revealed the potential for an internal pseudoknot between guanylate-rich residues in upper bulge loop L3 and cytidylates that are also predicted to be base-paired in the highly conserved S1 stem (Fig. 2.2). Bulge loop mutation m1, which disrupts the putative pseudoknot, reduced synthesis of p88 (and thus readthrough) by 97% in WGE, whereas stem S1 mutation m2, which would convert G:C to G:U pairs in S1 and the putative pseudoknot, did not affect readthrough (Fig. 2.2). When the gRNA contained both m1 and m2, which should maintain the pseudoknot and S1, synthesis of p88 increased to 127% of WT. Although the m1 mutation converted a proline to a serine in the RdRp, mutant and WT gRNAs were also evaluated for accumulation in Arabidopsis thaliana protoplasts (Fig 2.2). m1 reduced mRNA accumulation by 97% and m2 reduced gRNA accumulation by 34%. m1+m2 were compensatory, with accumulation increasing (over m1 levels) to 71% of WT. These results strongly suggest that, in addition to the phylogenetically conserved RSE structure that contains stem S1, the RSE must form an alternate structure without S1 and with an internal pseudoknot for efficient readthrough *in vitro.*


A
L4
S4
L3
S3
L2
S2
L1
S1
5'
SLA
RSE
TCV
CCFV
SCV
MNSV
PFBV
CbMV

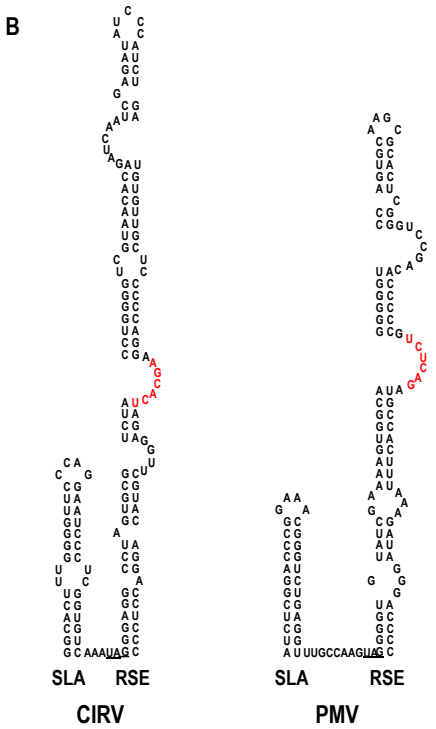

B
SLA
RSE
CIRV
PMV

Fig. 2. 1 Phylogenetic conservation of RSE and SLA structures. (A) Carmoviruses: CCFV, *Cardamine chlorotic fleck virus*; SCV, *Saguaro cactus virus*; MNSV, *Melon necrotic spot virus*; PFBV, *Pelargonium flower break virus*; CbMV, *Calibrachoa mottle virus*. Sequences in blue are predicted to form an internal pseudoknot in an alternate structure. Sequences in red are predicted to pair with complementary sequences at the 3' end of the genome. (B) CIRV, *Carnation Italian ringspot virus,* is a member of the genus Tombusvirus; PMV, *Panicum mosaic virus*, is a member of the genus Panicovirus.

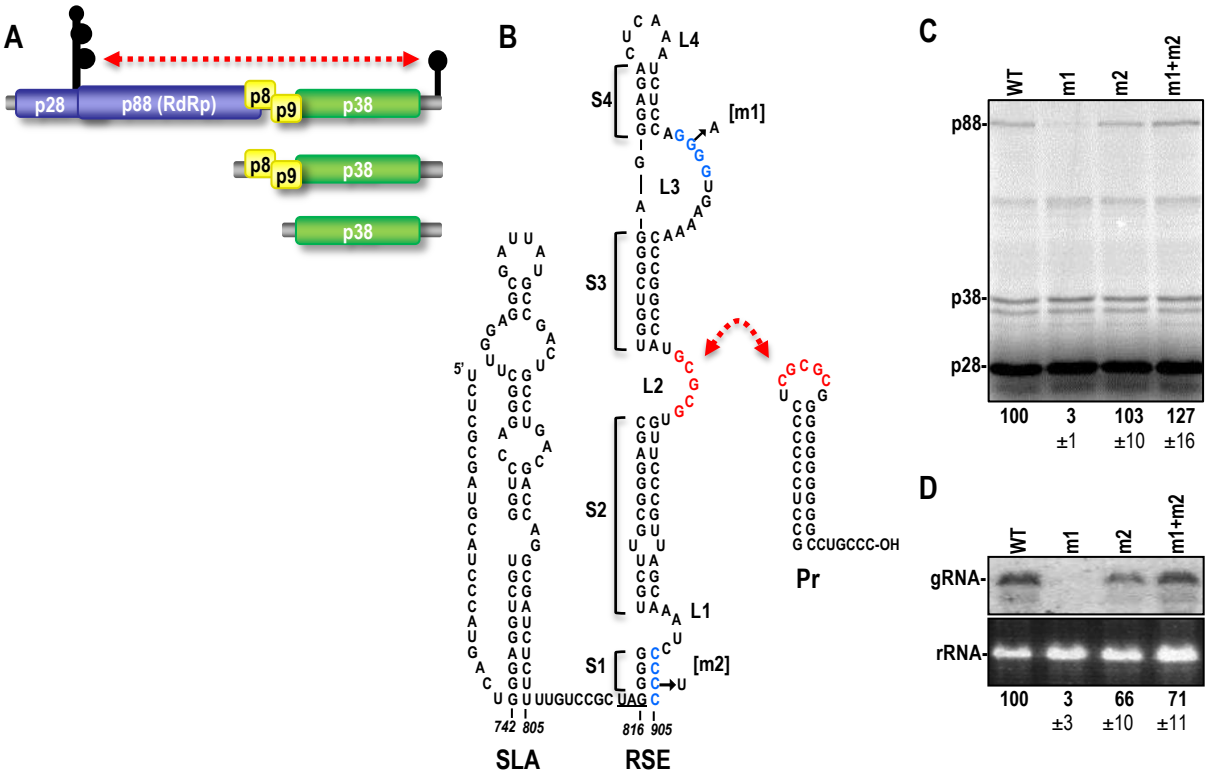

A
p28
p88 (RdRp)
p8
p9
p38
p8
p9
p38
p38
B
L4
S4
[m1]
L3
S3
L2
S2
L1
S1
[m2]
Pr
CCUGCCC-OH
5'
742 805
816 905
SLA
RSE
C
WT
m1
m2
m1+m2
p88-
p38-
p28-
100
3
±1
103
±10
127
±16
D
WT
m1
m2
m1+m2
gRNA-
rRNA-
100
3
±3
66
±10
71
±11

Fig. 2. 2 TCV RSE contains an internal pseudoknot.

(A) Genome organization of TCV. Genomic and subgenomic RNAs are shown. Long-distance RNA:RNA interaction connecting the RSE with the 3' terminal end is indicated by a dotted line. (B) Phylogenetically conserved structure of the TCV RSE and upstream hairpin SLA. Names of stems and loops in the RSE are shown. Sequences in L2 and Pr loop that engage in the long-distance interaction are in red. Putative pseudoknot residues are in blue. Location of m1 and m2 mutations generated to test for the pseudoknot are shown. (C) *In vitro* translation of full-length WT (wild type) and mutant gRNAs in WGE. Positions of p28 and readthrough product p88 are shown. P38 is the capsid protein and is translated by internal initiation. P38 was used as a loading control throughout this study. (D) RNA gel blot of WT and mutant gRNA accumulating at 40 hours post-inoculation in Arabidopsis protoplasts. Location of the gRNA and 26S rRNA is denoted.

Structure of the TCV RSE is important for function

To further analyze the importance of RSE sequence and structure, loop mutations and mutations designed to be compensatory in stems were generated throughout the hairpin (Fig. 2.3A). If stem alterations could be designed that were both compensatory in the RSE and silent in the RdRp, levels of mutant gRNAs were also assessed in protoplasts. Mutating the terminal loop (m3) or the base-pair just below bulge loop L3 (m6, m7) did not affect readthrough or gRNA accumulation *in vivo* (Fig. 2.3 B and C). Mutating the three base pairs just below L3 (m8, m9) reduced PRT to background levels, whereas the compensatory mutation (m10) restored p88 levels to WT. Altering the base pair in S4 that is adjacent to L3 (m4, m5) decreased gRNA accumulation *in vivo* but did not affect readthrough *in vitro*. Combining the two alterations was compensatory for gRNA accumulation, indicating that disrupting S4 might have consequences that were not discernable in the WGE assay. Altering the guanylate just below L2 to an adenylate (m11) disrupts pairing at this location and reduced PRT to 8% of WT and gRNA accumulation to 30% of WT. Maintaining pairing in this location (m12) had no effect on readthrough and only reduced gRNA accumulation by 26%. gRNA containing both mutations (m13) gave WT levels of readthrough and accumulation *in vivo*, suggesting that a base-pair in this position is important. Altering two base-pairs simultaneously in stem S2 (m15, m16) reduced readthrough and gRNA accumulation by 20-fold, whereas the compensatory mutations (m17) restored readthrough to 75% of WT and gRNA accumulation to WT levels. Stabilizing the S2 stem by converting the adjacent G:U pair and U U mismatch to a G:C and A:U (m14) enhanced readthrough to 275% of WT. This latter result correlates with published reports indicating that strengthening the stability of

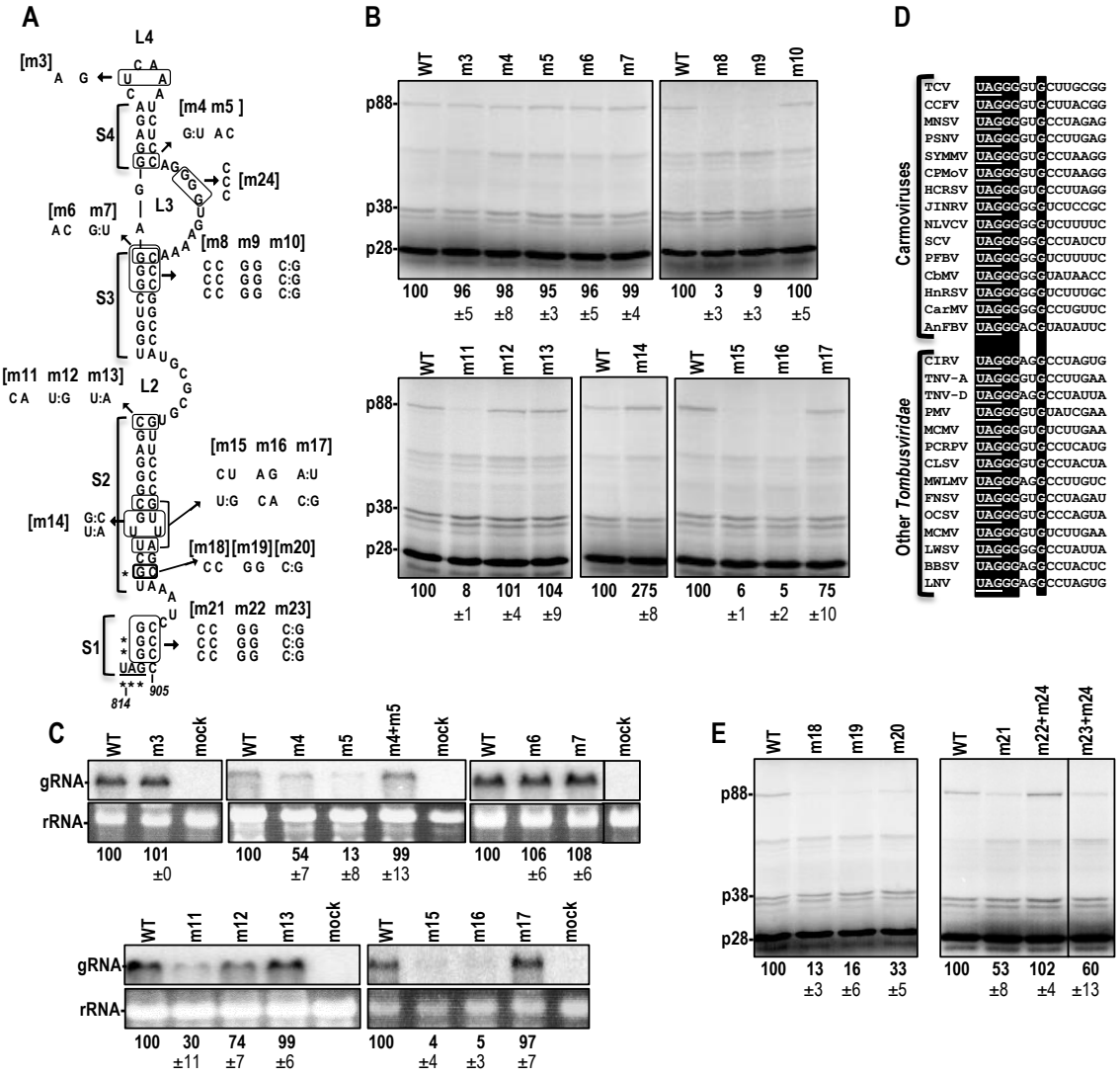

A
L4
[m3]
[m4 m5]
S4
[m24]
[m6 m7]
L3
[m8 m9 m10]
S3
[m11 m12 m13]
L2
[m15 m16 m17]
S2
[m14]
[m18][m19][m20]
[m21 m22 m23]
S1
814
905
B
WT m3 m4 m5 m6 m7 WT m8 m9 m10
p88
p38
p28
100 96 98 95 96 99 100 3 9 100
±5 ±8 ±3 ±5 ±4 ±3 ±3 ±5
WT m11 m12 m13 WT m14 WT m15 m16 m17
100 8 101 104 100 275 100 6 5 75
±1 ±4 ±9 ±8 ±1 ±2 ±10
D
Carmoviruses
TCV UAGGGGUGCUUGCGG
CCFV UAGGGGUGCUUACGG
MNSV UAGGGGUGCCUAGAG
PSNV UAGGGGUGCCUUGAG
SYMMV UAGGGGUGCCUAAGG
CPMoV UAGGGGUGCCUAAGG
HCRSV UAGGGGUGCCUUAGG
JINRV UAGGGGGGCUCUCGC
NLVCV UAGGGGGGCUCUUUC
SCV UAGGGGGGCCUAUCU
PFBV UAGGGGGGCUCUUUC
CbMV UAGGGGGGCUAUAACC
HnRSV UAGGGGGGCUCUUGC
CarMV UAGGGGGGCCUGUUC
AnFBV UAGGGACGUAUAUUC
Other Tombusviridae
CIRV UAGGGAGGCCUAGUG
TNV-A UAGGGGUGCCUUGAA
TNV-D UAGGGAGGCCUAUUA
PMV UAGGGGUGUAUCGAA
MCMV UAGGGGUGUCUUGAA
PCRPV UAGGGGUGCCUCAUG
CLSV UAGGGGUGCCUACUA
MWLMV UAGGGAGGCCUUGUC
FNSV UAGGGGUGCCUAGAU
OCSV UAGGGGUGCCCAGUA
MCMV UAGGGGUGUCUUGAA
LWSV UAGGGGGGCCUAUUA
BBSV UAGGGAGGCCUACUC
LNV UAGGGAGGCCUAGUG
C
WT m3 mock WT m4 m5 m4+m5 mock WT m6 m7 mock
gRNA
rRNA
100 101 100 54 13 99 100 106 108
±0 ±7 ±8 ±13 ±6 ±6
WT m11 m12 m13 mock WT m15 m16 m17 mock
gRNA
rRNA
100 30 74 99 100 4 5 97
±11 ±7 ±6 ±4 ±3 ±7
E
WT m18 m19 m20
p88
p38
p28
100 13 16 33
±3 ±6 ±5
WT m21 m22+m24 m23+m24
100 53 102 60
±8 ±4 ±13

Fig. 2. 3 Analysis of the TCV RSE. (A) Mutations generated in the RSE. Asterisks denote residues conserved at the readthrough site throughout the Tombusviridae (see D). When possible, base alterations were chosen for being silent in the encoded RdRp, and maintaining similar codon usage as the WT residue, which allowed for assaying the gRNA *in vivo* (see C). Names of the mutations are in brackets. (B) *In vitro* translation of full-length WT and mutant gRNAs in WGE. (C) Accumulation of selected gRNA mutants in protoplasts assayed using RNA gel blots. Lines in the upper right panel denote that a lane from the blot was discarded. (D) Sequence conservation adjacent to the amber codon in carmoviruses and members of *Tombusviridae*. (E) *In vitro* translation of full-length WT and mutant gRNAs with alterations in the conserved residues. Line denotes that a lane from the blot was discarded.

RSE lower stems positively impacts PRT and -1PRF [87,110]. These results establish the importance of RSE stems S2 and S3 for stimulating PRT.

Sequence adjacent to the readthrough site is conserved in tombusvirids and is important for readthrough

The readthrough sites in all 15 carmoviruses and in a random sampling of other tombusvirids (Fig. 2.3 D) contain a conserved sequence (UAG GGNNG; the termination codon is underlined). The guanylate in the +5 (3' terminal) position of the consensus sequence would occupy the central position of a codon, suggesting that this invariant residue (G821 in the TCV RSE) is not present because of a need to maintain a particular amino acid at this location in the RdRp. Alteration of either G821C (m18) or C986G (m19) reduced readthrough to 13% or 16% of WT, respectively (Fig. 2.3 A and E). The two alterations together (m20), which should reform the S2 stem, only improved readthrough to 33% of WT. This is in contrast with the two nearby base-pair alterations (m15 and m16) that decreased readthrough to 5% of WT, with the compensatory change (m17) improving readthrough to 75% of WT. This result suggests that both base pairing and the identity of G821 and/or C986 are important for readthrough.

To determine the importance of the guanylates adjacent to the termination codon, the three guanylates in positions +1, +2, and +3 were converted to cytidylates (m21) and the opposing cytidylates were converted to guanylates (m22). Because m22 mutations would disrupt the required pseudoknot, they were combined with corresponding changes in bulge loop L3 (m24) to maintain the tertiary interaction. m21 reduced readthrough by nearly 50% whereas m22 (+m24) had no detrimental effect (Fig. 2.3 E). These results indicate that the S1 stem can be disrupted without negatively impacting readthrough as long as mutations are on the 3' side and the required pseudoknot is not disrupted. Combining m21 with m22 (compensatory mutation m23), which along with m24 should

restore the S1 stem and maintain the pseudoknot, only slightly improved readthrough (to 60% of WT). The non-compensatory nature of the residues in stem S1 suggest that one or more of the conserved guanylates just downstream from the amber codon are required for efficient readthrough.

SHAPE structure probing of the RSE region *in vitro* and *in vivo*

SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) was used to structurally map the SLA, RSE, and 50 nt downstream of the RSE in full-length TCV gRNA prepared *in vitro* (Fig. 2.4). SHAPE detects flexible residues whose 2' OH can form covalent bonds with N-methylisatoic anhydride (NMIA), which then impedes reverse transcriptase-mediated primer extension. Residues with moderately-high to high reactivity with NMIA are colored red and residues with low to moderate reactivity are colored green. The 50 nt region downstream of the RSE was very flexible, with only one small hairpin predicted (Fig. 2.5). In the SLA/RSE region, SHAPE data supported the predicted structure of the SLA, but was only consistent with the upper portion of the RSE (Fig. 2.4 A, B and C). In the lower portion of the RSE, three residues on the 3' side of stem S2 that were predicted to be paired were strongly reactive with NMIA whereas their partner resides were not. This suggests that, despite the importance of the S2 stem for PRT *in vitro*, it was not present at discernable levels within the gRNA population used for both SHAPE and WGE. The strong reactivity of L3 pseudoknot guanylates, but not their paired cytidylates, also suggests that the PRT-required pseudoknot was not present in the *in vitro* synthesized gRNA. Bulge loop L2, which engages in the long-distance interaction with the Pr loop, contained a mixture of flexible and non-flexible residues.

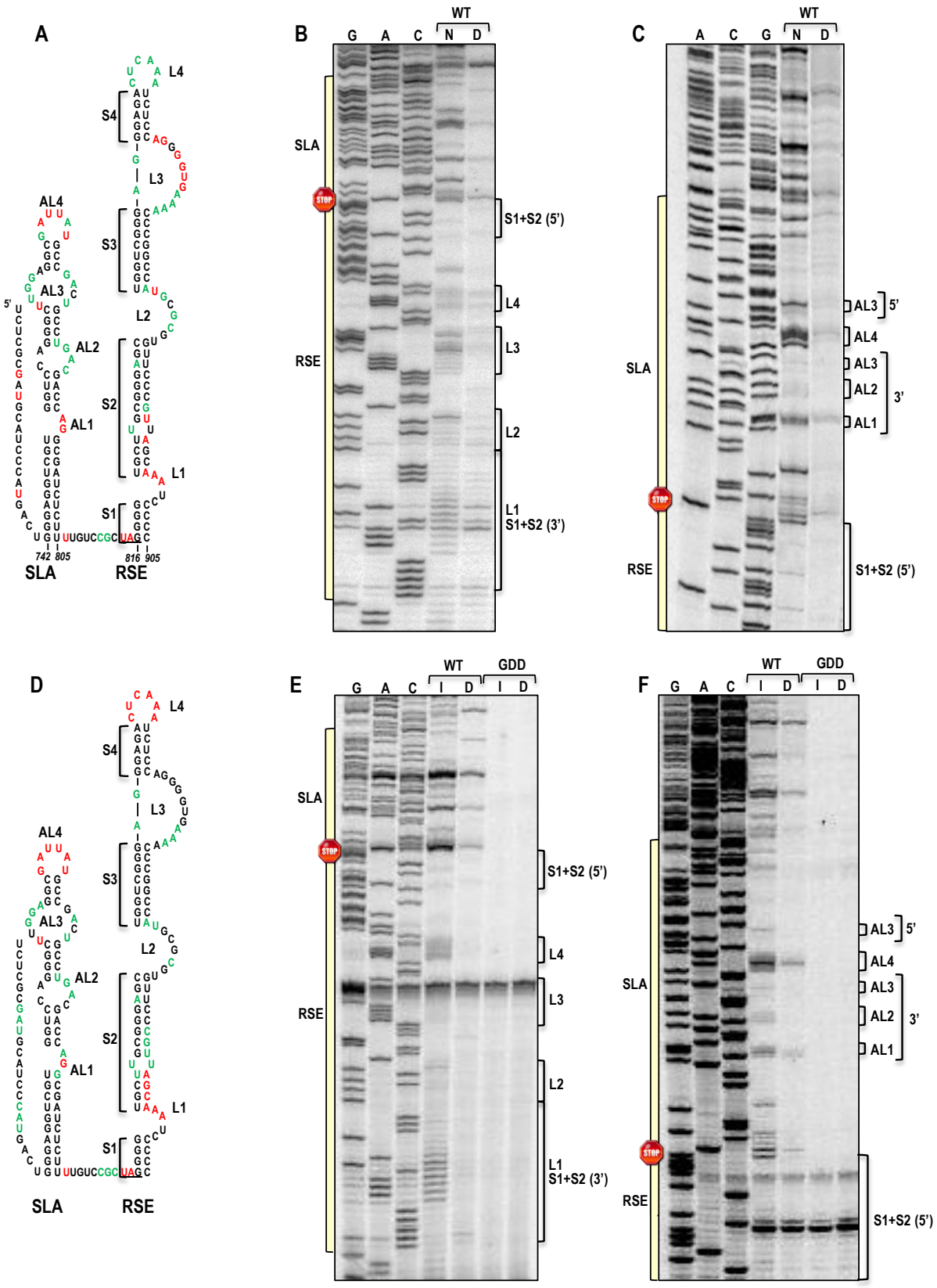

A
L4
S4
L3
AL4
S3
AL3
L2
AL2
S2
AL1
L1
S1
5'
742 805
816 905
SLA
RSE
B
WT
G A C N D
SLA
RSE
S1+S2 (5')
L4
L3
L2
L1
S1+S2 (3')
C
WT
A C G N D
SLA
RSE
AL3 5'
AL4
AL3
AL2 3'
AL1
S1+S2 (5')
D
E
WT GDD
G A C I D I D
SLA
RSE
S1+S2 (5')
L4
L3
L2
L1
S1+S2 (3')
F
WT GDD
G A C I D I D
SLA
RSE
AL3 5'
AL4
AL3
AL2 3'
AL1
S1+S2 (5')

Fig. 2. 4 SHAPE structure probing of the readthrough region in full-length gRNA *in vitro* and *in vivo*. (A) Location of NMIA-reactive residues in the phylogenetically-conserved structure of the RSE using SHAPE profile data generated from *in vitro*-synthesized gRNA (see B and C). Residues that are moderately-high to highly reactive are colored red and residues that are moderately-low to moderately reactive are colored green. (B and C) SHAPE mapping of the RSE and SLA regions in *in vitro*-synthesized gRNA. The first three lanes are nucleotide ladder lanes. N, NMIA-treated (modified); D, DMSO-treated (control). Locations of different regions in the RSE and SLA are shown to the right of the panels, with 5' and 3' denoting the side of the hairpin. The location of the amber stop codon is shown. (D) Location of 1-methyl-7-nitroisatoic anhydride (1M7)-reactive residues on the phylogenetically-conserved structure of the RSE using SHAPE profile data generated from gRNA accumulating in protoplasts (see E and F). Since 1M7 is membrane permeable, it was used in place of NMIA. (E and F) SHAPE mapping of the RSE and SLA regions in gRNA accumulating *in vivo* at 40 hpi. First three lanes are nucleotide ladder lanes. 1, 1M7-treated; D, DMSO-treated.

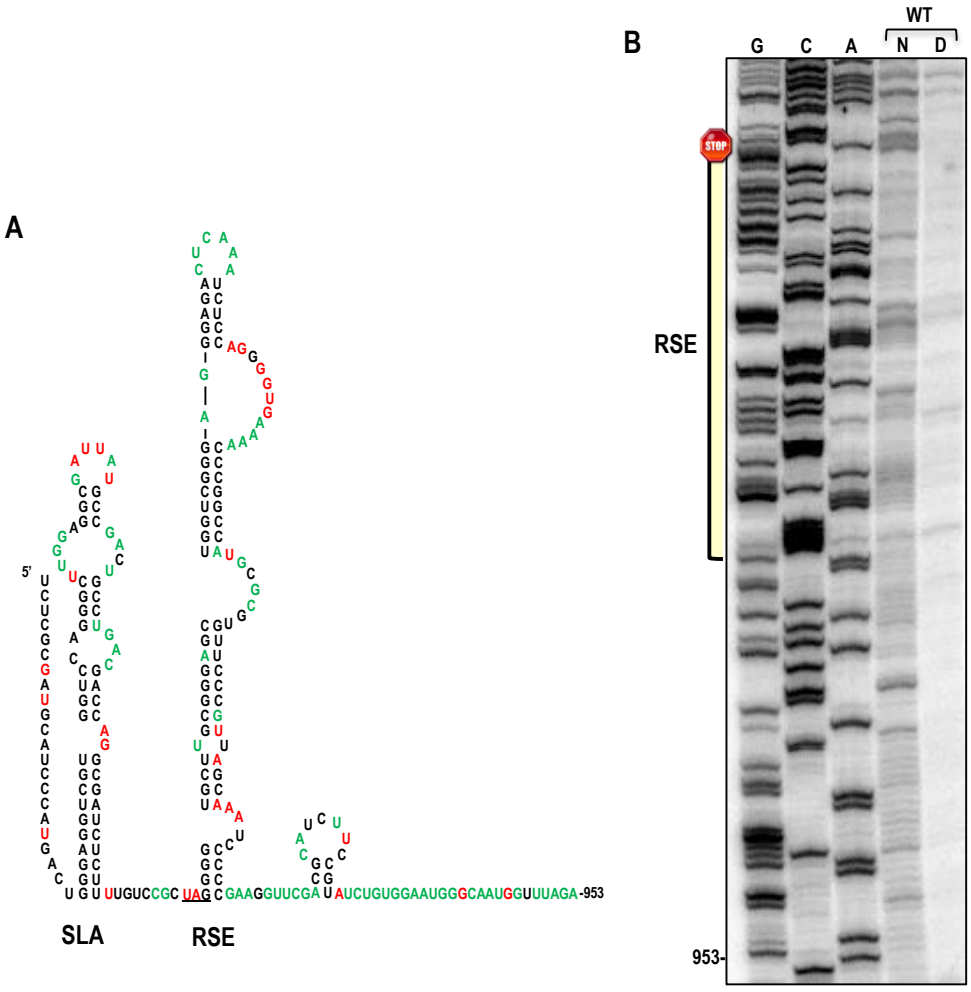

A
5'
UUGUCCGC UAG C GAAGGUUCGA U AUCUGUGGAAUGGGCAAUGGUUUAGA-953
SLA
RSE
B
WT
G
C
A
N
D
STOP
RSE
953-

Fig. 2. 5 Structure downstream of the TCV RSE. (A) SHAPE profile of the SLA, RSE and downstream region. Red and green residues denote strength of reactivity with NMIA (red is moderately-high to high reactivity, green is moderately-low to moderate reactivity). (B) SHAPE phosphorimage of the RSE and downstream region. First three lanes are nucleotide ladder lanes. N, NIMA-treated; D, DMSO-treated. Note that region downstream of the RSE is very reactive with NMIA denoting substantial flexibility. cytidylates, also suggests that the PRT-required pseudoknot was not present in the *in vitro* synthesized gRNA. Bulge loop L2, which engages in the long-distance interaction with the Pr loop, contained a mixture of flexible and non-flexible residues. These residues were identically flexible when gRNA was used that contained mutations in the Pr loop designed to disrupt the long-distance interaction (see Fig. S5). Thus this necessary tertiary interaction was also not detected in gRNA prepared *in vitro*.

These residues were identically flexible when gRNA was used that contained mutations in the Pr loop designed to disrupt the long-distance interaction. Thus this necessary tertiary interaction was also not detected in gRNA prepared *in vitro*.

Since RNA folds co-transcriptionally, the absence of both the phylogenetically-conserved stem S2 and the two required tertiary interactions could be due to misfolding of the *in vitro* synthesized, denatured and then renatured gRNA. Therefore, SHAPE was repeated using viral gRNA accumulating within infected protoplasts (*in vivo* SHAPE). gRNA that was capable (WT) or not capable (GDD) of amplification (due to mutations in the RdRp active site) was subjected to SHAPE at 40 hpi, since plant viruses are restricted to initially inoculated protoplasts and little additional virus accumulates after this point. Cells were treated with 1M7 (a membrane permeable electrophile that generates an *in vitro* SHAPE profile identical to NMIA [Fig. S3]), followed by cell lysis, total RNA extraction, and primer extension. As shown in Figure 3E and F, and on the secondary structure in Figure 3D, the *in vivo* SHAPE profile for the RSE region at this time point was very similar to the *in vitro* SHAPE profile, suggesting that the structure of the RSE region in the *in vitro* synthesized gRNA does not represent a kinetically trapped artificial structure. The *in vivo* SHAPE profile was not due to residual input RNA, as no SHAPE data was obtained using the GDD mutant (with the exception of the strong signal in all lanes at the position of the pseudoknot guanylates, which was likely associated with the radiolabeled oligonucleotide). The one major difference between the *in vivo* and *in vitro* SHAPE profiles was additional reactive residues on the 3' side of stem S2, strongly suggesting that at least some portion of the phylogenetically conserved S2 stem is not present in the majority of gRNA sampled *in vivo* and *in vitro*.

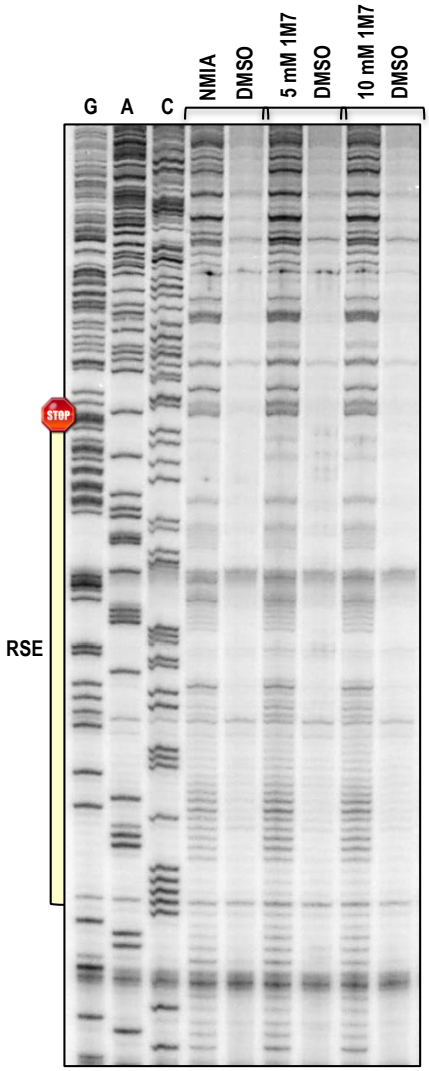


Fig. 2. 6 NMIA and 1M7 produce nearly identical SHAPE profiles in the RSE region using gRNA synthesized *in vitro*. (related to figure 2.4) First three lanes are nucleotide ladder lanes. The RSE region is denoted. Two different concentrations of 1M7 were used .

To investigate whether the S1 stem is present in *in vitro* synthesized gRNA, all four S1 3' side cytidylates were changed to guanylates (m32) and mutant gRNA subjected to *in vitro* SHAPE (Fig. 2.8A). The flexibilities of the altered residues on the 3' side of S1 (and adjacent sequences) were enhanced, however, no change in flexibility of 5' side guanylates was detected (Fig. 2.8B). In addition, the flexibility of L3 bulge loop guanylates that form a pseudoknot with these cytidylates was also unchanged. When the L3 guanylates were also converted to cytidylates (m32+m33) with the hope of driving formation of the pseudoknot, the mutated L3 residues were no longer flexible. However, the S1 3' guanylates in m32+m33 showed little change in flexibility over m32 alone, suggesting that the pseudoknot is still not forming in the majority of gRNA in the population and implying that the new cytidylate residues in L3 were likely pairing elsewhere. m32 reduced readthrough by 20-fold in WGE and m32+m33 regained WT levels of readthrough, indicating that the pseudoknot is still required and still capable of formation during translation of the gRNA (Fig. 2.7C). These results strongly suggest that the S1 stem and RSE pseudoknot are not present in the majority of the gRNA synthesized *in vitro*.

Sequences upstream of the SLA pair with RSE 5' side S1 and S2 residues

To determine if the middle portion of the S2 stem exists in gRNA synthesized *in vitro*, two additional mutant gRNAs were generated (m25, m26), each with seven altered residues in S2 that together are complementary (Fig. 2.9A, upper left). *In vitro*

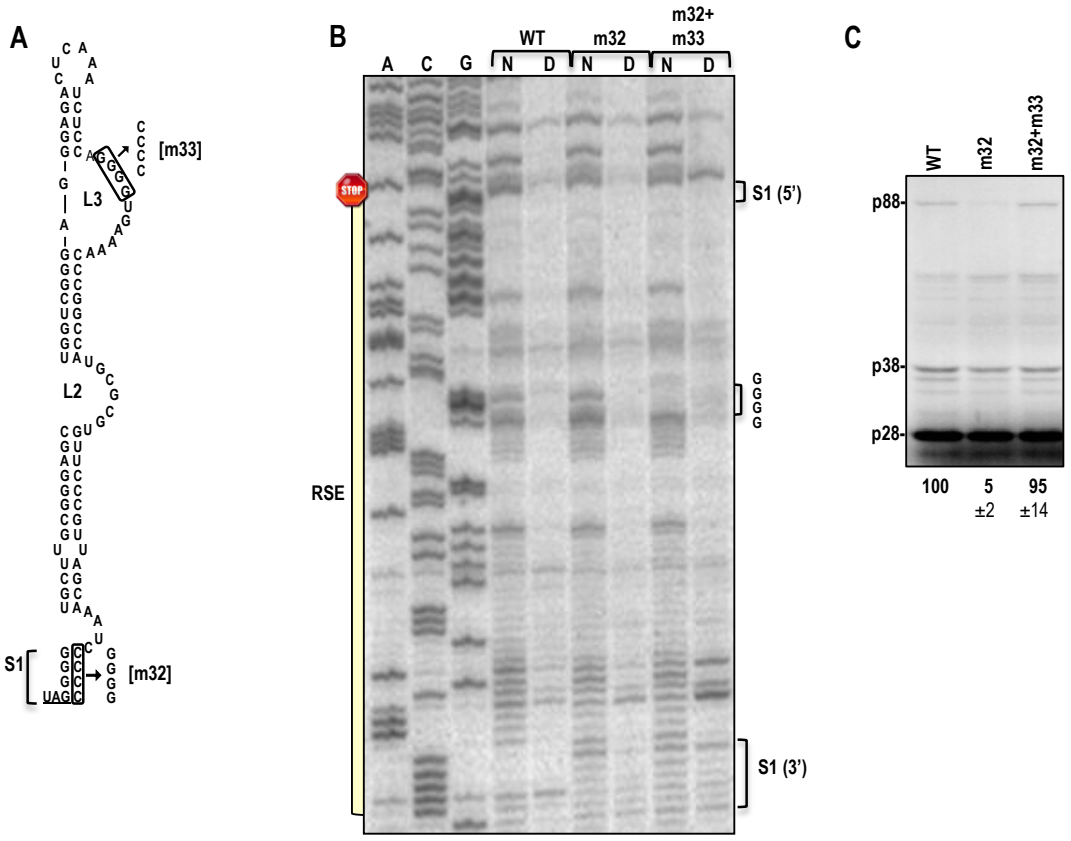


Fig. 2. 7 Compensatory mutations in the RSE pseudoknot do not result in detection of the pseudoknot in the *in vitro* synthesized gRNA. (A) Location of mutations. (B) First three lanes are nucleotide ladder lanes. N, NMIA-treated; D, DMSO treated. Specific locations in the RSE are denoted by brackets. Note that the m33 alteration reduced flexibility of local (mutated) residues, but did not substantially reduce flexibility of altered S1 partner residues. This indicates that the pseudoknot has not yet formed in this folded RNA. In addition, 5' S1 residue flexibility was unaltered, confirming that the S1 stem is also not present. (C) Effect of mutations on readthrough in WT and mutant gRNAs prepared *in vitro*.

translation of m25, m26, and m25+m26 gRNA generated only low (6 to 8% of WT) levels of p88 (Fig. 2.9B). Structure probing of m25, which contained mutations on the 5' side of S2, indicated that the mutated resides became flexible as well as a few adjacent residues (total of 12 nt) without any discernable changes in the SHAPE profile of residues located on the 3' side of S2 (Fig. 2.9C). m26, with mutations on the 3' side of S2, contained residues in the altered region that both gained and lost flexibility, with flexibility changes extending into the L2 bulge. No flexibility changes were observed on the opposing 5' side of S2. In m25+m26 gRNA, the mutated residues on the 3' side were no longer flexible suggesting that the majority of the S2 stem was now forming. Altogether, these results strongly suggest that the entire S2 stem, in addition to the S1 stem, does not form in the gRNA synthesized *in vitro*.

In m25+m26 gRNA, the upper portion of S2 and lower portion of L2 contained a large number of residues with enhanced flexibility. This result suggested that the structure of a portion of L2, which engages in the long-distance interaction with 3' sequences, is significantly altered when either the S2 stem forms or when the alternate pairing of S2 residues is disrupted. The SHAPE profile of m14 gRNA, which contains a strengthened S2 stem (see Fig. 2.3), displayed similar structural changes in this region as m25+m26 (Fig. 2.9). Since readthrough was enhanced in m14 gRNA (Fig. 2.3) and

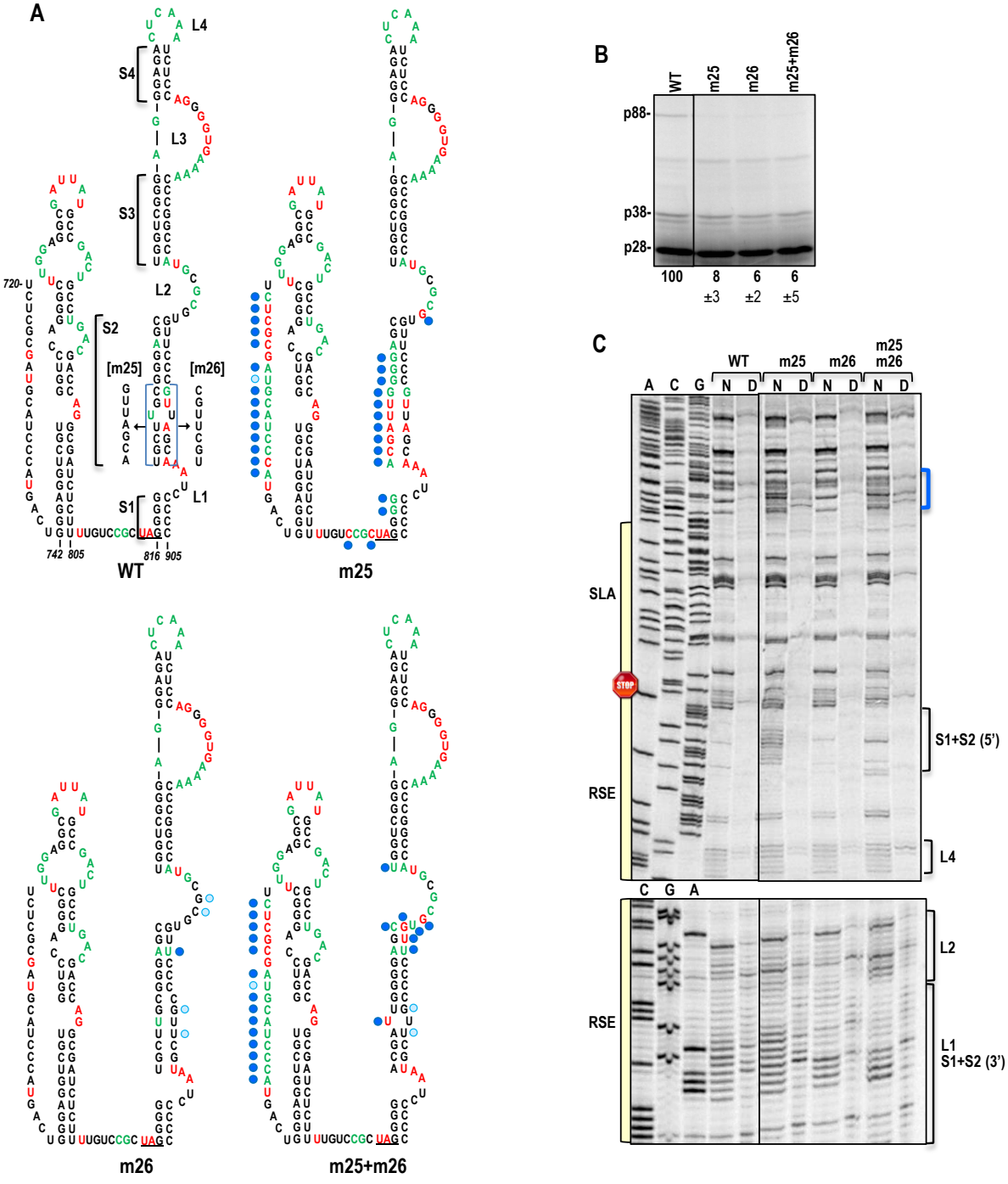

A
L4
S4
L3
S3
L2
S2
[m25]
[m26]
S1
L1
720-
742 805
816 905
WT
m25
m26
m25+m26
B
WT
m25
m26
m25+m26
p88-
p38-
p28-
100
8
±3
6
±2
6
±5
C
WT
m25
m26
m25
m26
A C G N D N D N D N D
SLA
RSE
STOP
S1+S2 (5')
L4
C G A
RSE
L2
L1
S1+S2 (3')

Fig. 2. 8 Alternate pairing partners for RSE 5' side stem S2 residues. (A) Upper left, location of mutations (m25 and m26) generated in full-length TCV gRNA. The WT SHAPE profile is shown. Color coding is as previously described. Upper right, effect of m25 mutations on NMIA-reactive residues in the RSE region. Dark blue and light blue circles designate residues that were more of less reactive than WT respectively. Lower left, effect of m26 mutations on reactive residues in the RSE region. Lower right, combination of m25 and m26 mutations that should be compensatory. (B) *In vitro* translation of full-length WT and mutant gRNAs in WGE. (C) SHAPE mapping of the RSE and SLA regions in WT and mutant gRNAs prepared *in vitro*. First three lanes are nucleotide ladder lanes. N, NMIA-treated; D, DMSO-treated. Blue bracket denotes region upstream of SLA that becomes flexible when 5' S2 residues are altered (m25). Note that the mutated residues on the 5' side of S2 are flexible in m25, whereas the partner residues on the 3' side of S2 retain WT-like flexibility.

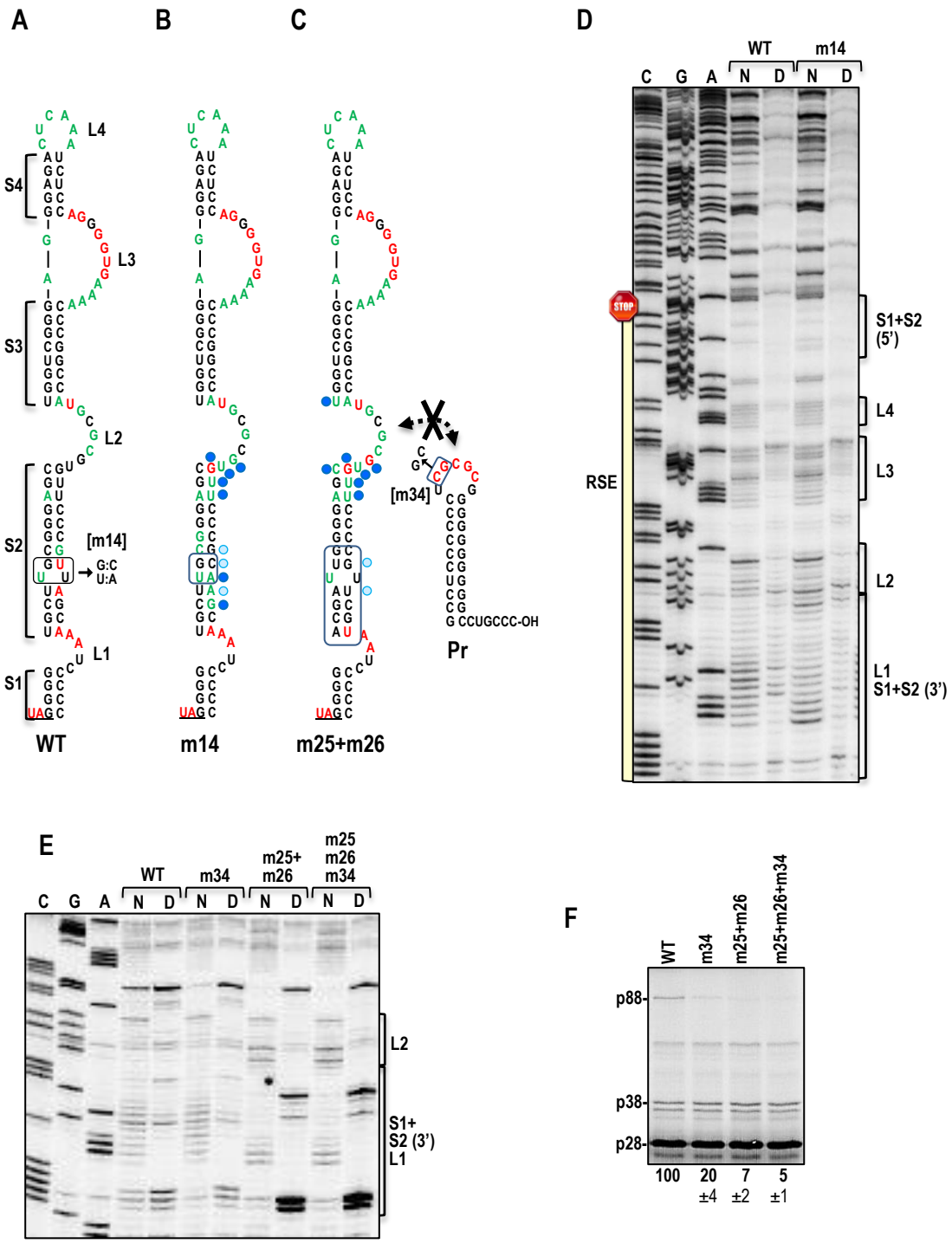

A
B
C
D
L4
S4
L3
S3
L2
S2
[m14]
G:C
U:A
L1
S1
WT
m14
m25+m26
[m34]
G CCUGCCC-OH
Pr
WT
m14
N
D
C
G
A
STOP
RSE
S1+S2
(5')
L4
L3
L2
L1
S1+S2 (3')
E
WT
m34
m25+
m26
m25
m26
m34
L2
S1+
S2 (3')
L1
F
WT
m34
m25+m26
m25+m26+m34
p88
p38
p28
100
20
±4
7
±2
5
±1

Fig. 2. 9 m14 and m25+m26 share similar structural alterations in upper 3' S2 and lower L2. (A) Location of mutations in the RSE that stabilize the stem of S2. (B) Effect of m14 mutations on the SHAPE profile of the RSE. Dark blue and light blue circles denote residues that were more or less reactive than WT respectively. (C) SHAPE profile of m25+m26 (compensatory mutations in the lower S2 stem) from Fig. 2.8, provided here for comparison purposes. The hairpin at the right is the 3' terminal Pr hairpin, whose apical loop is the pairing partner of L2. Residues complementary with L2 sequence are in red. The m34 mutations that disrupt the long-distance interaction are shown. (D) SHAPE mapping of the RSE region in WT and m14 gRNAs prepared *in vitro*. First three lanes are nucleotide ladder lanes. N, NMIA-treated; D, DMSO-treated. Note that the structural changes in upper 3' S2 and lower L2 are very similar to those found in m25+m26 (see E). (E) SHAPE mapping of the L2 region when the RSE contains m25+m26, and the long-distance interaction is either permitted (m25+m26) or not permitted (m25+m26+m34). This gel also shows that there is no difference in WT L2 sequence when the long-distance RNA:RNA interaction is disrupted. (F) *In vitro* translation of WT and m25+m26 gRNA that can and cannot (=M34) engage in the long-distance RNA:RNA interaction.

depressed in m25+m26 (Fig. 2.8B), the altered L2 structure was not by itself responsible for weak readthrough in m25+m26 gRNA. To determine if the S2/L2 structural changes affected the long-distance interaction between L2 and the 3' end, SHAPE was conducted on m25+m26 gRNA that also contained a two base alteration in the 3' end interacting sequence (m34). No difference in L2 flexibility was found in WT gRNA+m34 or m25+m26+m34, suggesting that the flexibility changes in L2 did not lead to detection of the long-distance interaction (Fig. 2.9E). In addition, m34 reduced gRNA readthrough by 2/3 but had no discernable effect on m25+m26. These results support the conclusion that the S2 stem forms in m25+m26 but not in WT gRNA synthesized *in vitro*, and that either the presence of the S2 stem or disruption of alternate pairing of S2 residues affects residue flexibility in and near the L2 bulge loop.

A region upstream of SLA is implicated in alternate pairing with RSE S2 residues

In the SHAPE profiles for m25 and m25+m26, a small region located just upstream of SLA (residues 720 to 735) showed enhanced flexibility (Fig. 2.8C, blue bracket). No discernable differences were detected either upstream or downstream of this region and no flexibility changes were discernable in this region in m26. When residues 720 to 735 were altered (m27) and the SHAPE profile of m27 compared with that of WT gRNA, the mutated region in m27 became more flexible (Fig. 2.10). In addition, 13 of 26 residues in S2 showed altered flexibility as did residues in the lower portion of L2. Since most S2 residues had increased flexibility, mutating residues 720-735 did not by itself result in formation of the phylogenetically conserved S2 stem. The presence of very similar flexibility changes in upper 3' S2 and lower L2 residues suggests that this flexibility

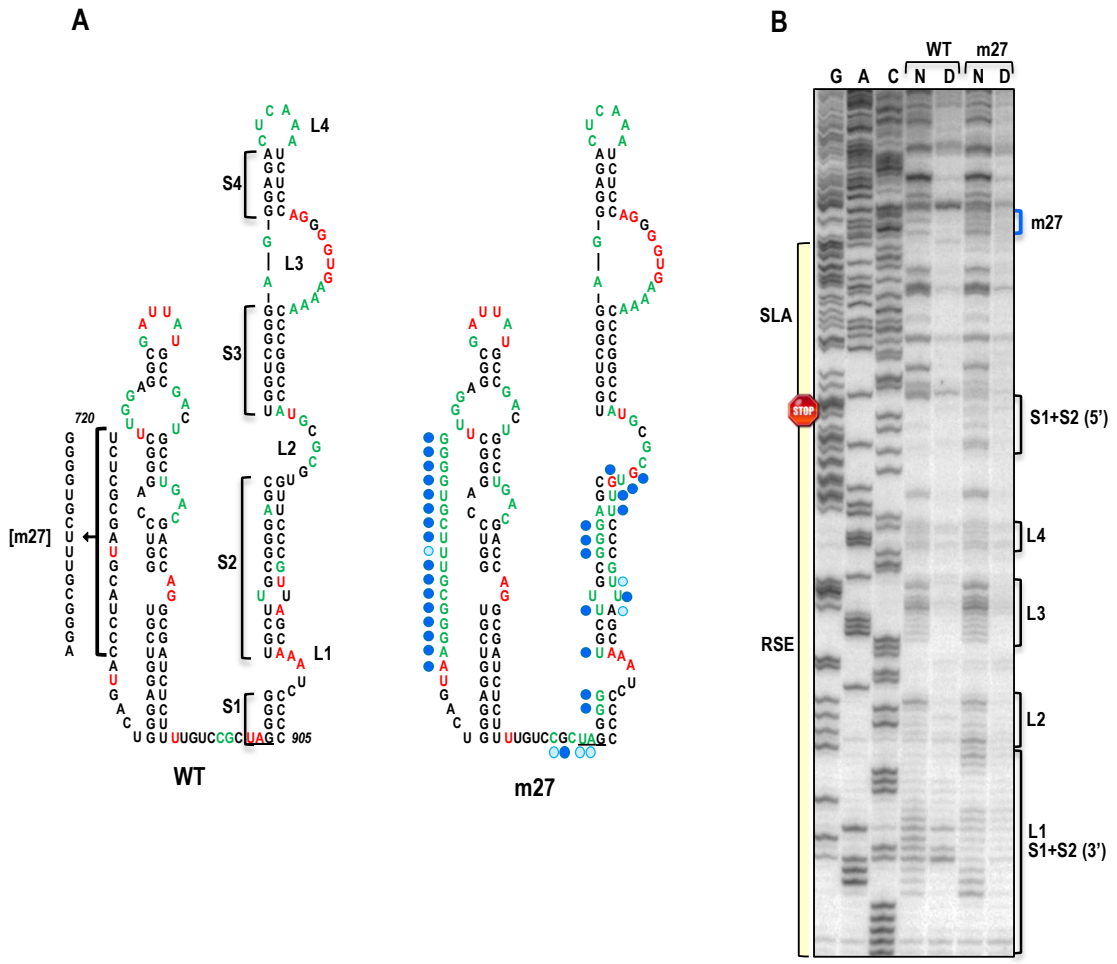


Fig. 2. 10 Altering the proposed pairing partners for 5' S1/S2 residues changes flexibility of S1/S2. (A) Left, location of m27 mutations generated in full-length TCV gRNA. The WT SHAPE profile is shown. Right, effect of m27 mutations on the SHAPE profile. Color coding is as previously described. (B) SHAPE mapping of the RSE and SLA regions in WT and mutant gRNAs prepared *in vitro*. First three lanes are nucleotide ladder lanes. N, NMIA-treated; D, DMSO-treated. Blue bracket denotes region upstream of SLA containing the m27 mutations. Note that the m27 mutations resulted in enhanced flexibility on both sides of the active RSE stem extending into the L2 region.

signature (also found for m14 and m25+m26 [Fig. 2.9]) is not directly associated with formation of the S2 stem, but rather may reflect loss of pairing between S2 residues and residues in positions 720-735.

Alternate structure of the TCV RSE

When the sequence between 720 and 735 was examined, clear base-pairing possibilities with residues on the 5' side of S1 and S2 were discernable, which extended the length of SLA (Fig. 2.11). This alternate conformation, labeled as the "basal" structure, agrees well with the SHAPE structure profile determined both *in vitro* (Fig. 2.11A) and *in vivo* (Fig. 2.11B). To determine the importance of SLA and the basal structure, three sets of silent mutations were generated on the 5' side of the presumptive elongated SLA: lower S1+S2 stem (m28), middle S5 stem (m29) and upper S6 stem (m30) (Fig. 2.12A). m28 gRNA caused several residues to gain flexibility on the opposing (3') side of the stem (Fig. 2.12B). These changes were accompanied by the signature structural changes in RSE upper S2 (3') and lower L2 regions (Fig. 2.12C) as was found when RSE stem S2 was strengthened (m14, m25+m26; Fig. 2.9) and/or when alternate pairing for S2 residues was disrupted (m27; Fig. 2.10). In addition, m28 mutations increased flexibility of residues in the S5 stem of the SLA, suggesting the presence of higher order structure within the extended hairpin (Fig. 2.12B). SHAPE structure probing of gRNA with m29 mutations in SLA S5 showed significant disruptions on both sides of the S5 stem (Fig. 2.12B). In addition, very similar structural changes were evident in the S1+S2 region of the extended SLA as were found for m28, accompanied by the signature changes in the RSE (Fig. 2.12B, C). These results support

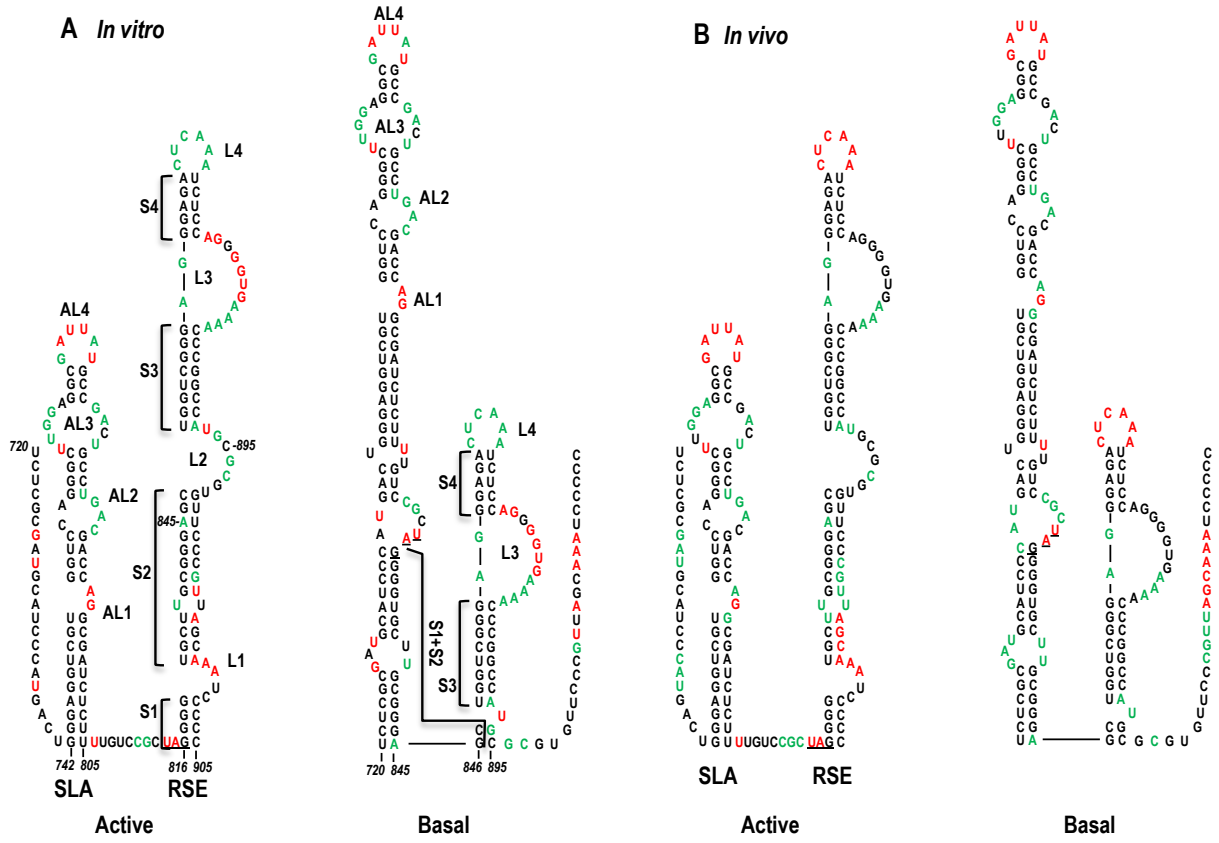


Fig. 2. 11 Proposed structure of the alternate, prominent RSE conformation

(A) WT SHAPE profile for gRNA synthesized *in vitro* is shown. Left, phylogenetically-conserved conformation is suggested to be the "active" form of the RSE, which also requires engagement in the long-distance interaction with the 3' end and formation of the pseudoknot for activity. Right, most prominent conformation of the RSE is proposed to be the "basal" form that must convert to the active form for readthrough.

(B) Left, phylogenetically-conserved conformation with the *in vivo* gRNA SHAPE profile. Right, proposed basal structure with the *in vivo* SHAPE profile.

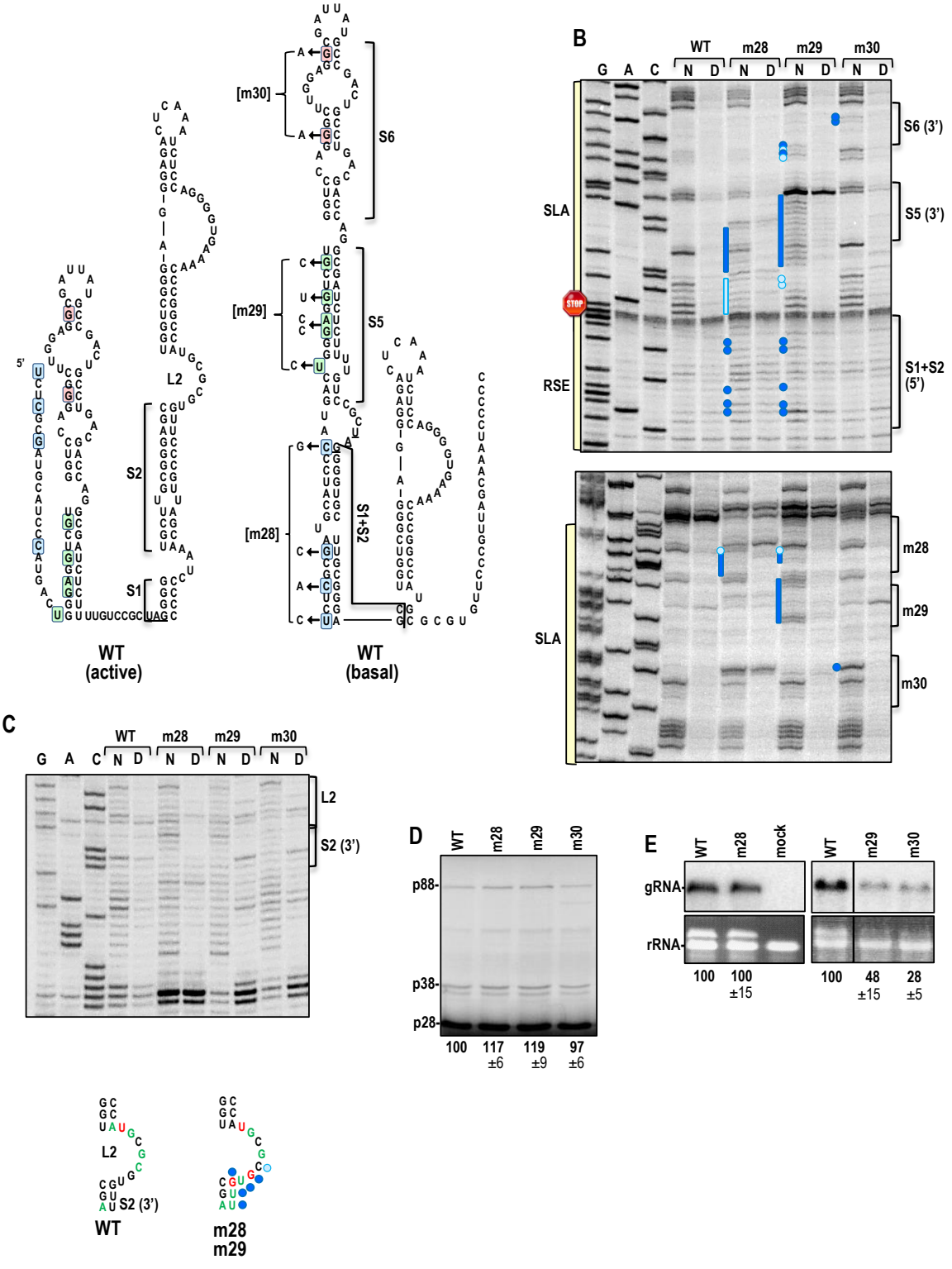

WT
(active)
WT
(basal)
[m30]
[m29]
[m28]
S6
S5
S1+S2
S2
S1
L2
B
WT
m28
m29
m30
G A C N D N D N D N D
SLA
RSE
STOP
S6 (3')
S5 (3')
S1+S2
(5')
m28
m29
m30
C
L2
S2 (3')
D
p88-
p38-
p28-
100 117 ±6 119 ±9 97 ±6
E
gRNA-
rRNA-
mock
100 100 ±15 100 48 ±15 28 ±5
WT
m28
m29

Fig. 2. 12 Extended SLA contains higher order structure and is important for gRNA accumulation *in vivo*. (A) Location of m28, m29, and m30 mutations in the basal and active structures of the RSE. Sets of mutations are color coded. (B) Effect of SLA mutations on the SHAPE profile of SLA and the RSE. Color coding has been previously described. Note that m28 and m29 cause nearly reciprocal structural changes in the adjacent stems. (C) Effect of SLA mutations on the structure of the RSE. Note that m28 and m29 cause alterations in residue flexibility in lower L2 and adjacent upper S2 (3') similar to m14, m25+m26, and m27 (Figs. 4, 5 and S5). (D) *In vitro* translation of WT and mutant full-length gRNAs in WGE. (E) RNA gel blot of WT and mutant gRNAs accumulating at 40 hpi. Line denotes that a lane was removed from the blot.

a structural connection between elements in the lower half of the basal SLA structure. In contrast, m30 mutations in the upper stem 6 of SLA enhanced the flexibility of several S6 stem residues without discernably affecting flexibility of the remainder of the SLA. m28 and m29 caused ~20% increases in translation of p88 in WGE, whereas m30 did not affect readthrough (Fig. 2.12D). This suggests that weakening the alternate pairing of RSE S1+S2 may slightly promote formation of the active form of the RSE. m29 and m30 reduced gRNA levels in protoplasts to 48% and 28% of WT, respectively (Fig. 2.12E). In contrast, m28 mutations did not affect the levels of gRNA, suggesting that the small enhancement in translation of p88 *in vitro* is not detrimental *in vivo*. These results support the importance of SLA for accumulation of gRNA *in vivo*, and suggest that the upper portion of SLA stem S5, which was only disrupted in m29, as well as apical stem S6 are important for function of SLA.

Multiple functions for the RSE: potential for ribosome recycling

Most eukaryotic cellular mRNAs have a modified nucleotide cap structure at the 5' end, 5' and 3' UTRs and a poly(A) tail at the 3' terminus. The 3' bound poly(A)-binding protein interacts with eukaryotic translation initiation factor 4G (eIF4G, a scaffold protein) and eukaryotic initiation factor 4E (eIF4E, the cap binding protein) forming a molecular bridge effectively circularizing the mRNA[130]. This circularization increases translation efficiency by recycling the terminating ribosomes on the same mRNA template[131]. Many plant viral RNAs often do not contain a 5' cap or poly(A) tail instead relying on 3' CITEs to enhance translation. These 3' elements can assist in the circularization of the gRNA by either direct RNA:RNA or protein mediated

interactions[117,121,132]. Given that the translation ratio of p28 to p88 is 95% to 5%, it is reasonable to hypothesize an element exists in proximity to the p28 stop codon that could potentially interact with the 5' end of the genome to facilitate the recycling of ribosomes (Fig. 2.13). Alternatively, a 3' element could potentially transfer ribosomes to either the readthrough region, or the 5' end. To test this hypothesis portions of the 5' end (the 5' UTR or 5' UTR + 60 bases), and the readthrough elements (SLA or RSE), were deleted. Additionally the 3' end was digested with either SpeI or HindIII effectively deleting 103 or 694 nt from the 3' end respectively. Using these six constructs, SHAPE structure probing was employed to survey the 5' end, recoding region, and 3' ends of the virus (Fig. 2.13). No differences were noted on any of the gels (Fig. 2.14). While this may suggest no possible role for the recoding elements SLA and RSE in ribosome recycling, two alternative options exist. A protein bridge may be required for the interaction to take place similar to the TCV ribosomal subunit bridge between the TSS and a 5' translation enhancer sequence. Alternatively, the conformation represented on the gels is a pre-translation snapshot. Translation factors or translating ribosomes may be required to shift the conformation.

Discussion

Conserved features of tombusvirid RSEs

I examined RSE regions in tombusvirids involved in either -1PRF or PRT. With few exceptions, all of these RSE are ca. 90 nt in length with a G:C-rich lower stem that incorporates the guanylate of the amber stop codon at the base. Tombusvirid RSE contain at least two large asymmetric bulge loops, with the lower loop engaged in an

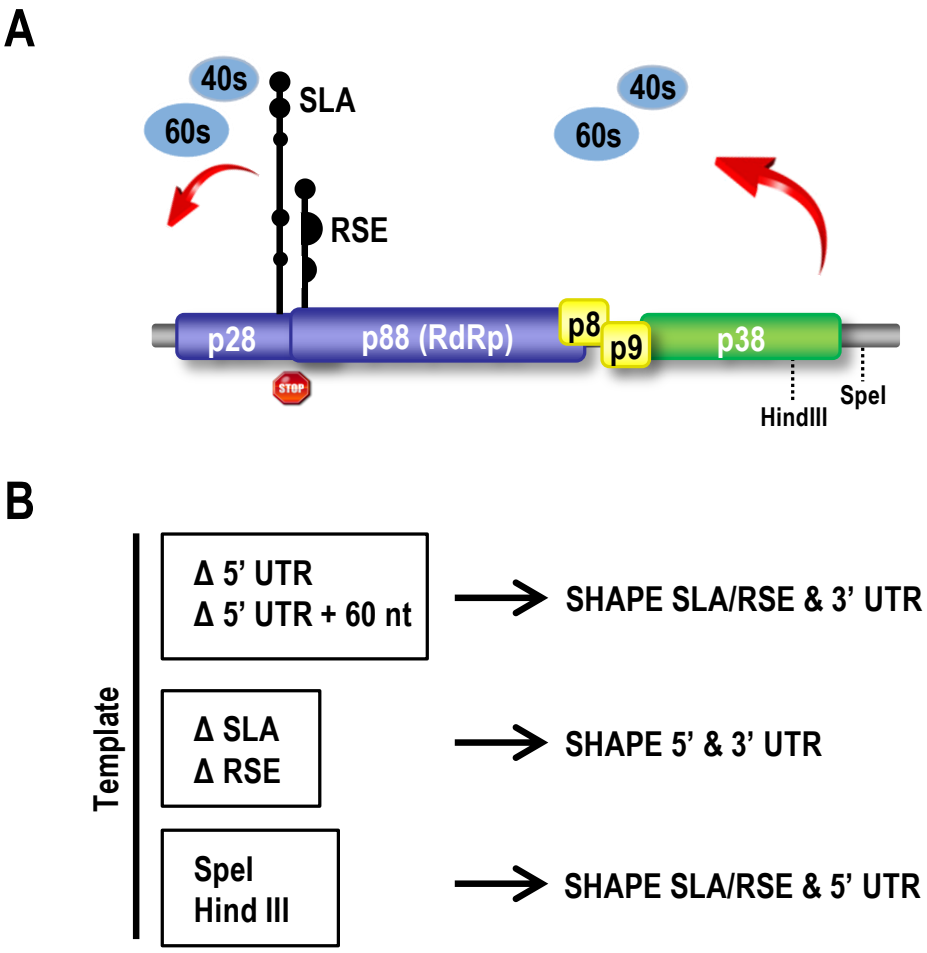


Fig. 2. 13 Potential role of SLA and RSE in ribosome recycling. (A) TCV genome schematic. SLA and RSE can potentially engage in an interaction with the 5' end for the recycling of ribosomes. The p28 stop codon is designated with a stop sign. 3' restriction sites HindIII and SpeI are designated with hashed lines. (B) Outline of mutation templates shown left, with SHAPE regions shown to the right.

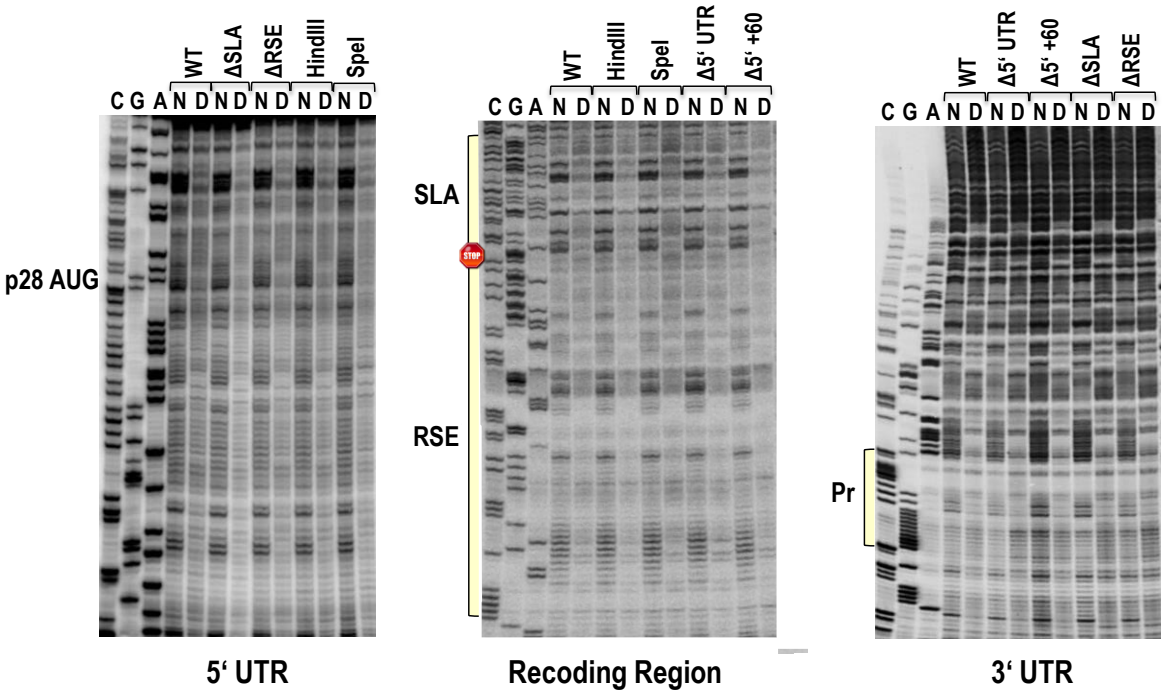


Fig. 2. 14 SHAPE gels of potential ribosome recycling sites. The three regions probed, from left to right are the 5' end, the recoding region and the 3' end. First three lanes are sequencing ladders. N, NIMA (modified RNA) and D, DMSO (control) are shown above. Six deletions were SHAPED: Δ5' UTR; Δ5' UTR + 60 bases; ΔSLA; ΔRSE; HindIII restriction site and SpeI restriction site (see Fig. 2.13).

important long-distance interaction with the 3' end[46,84,91]. This RNA bridge had been proposed to form an atypical pseudoknot in the RSE that functionally replaces the pseudoknot commonly found in animal virus RSE[84]. However, compensatory mutagenesis revealed that a phylogenetically conserved pseudoknot is also a critical feature of carmovirus RSE (Fig. 2.2), suggesting that the function of the long-distance interaction is not a simple substitution for a local H-type pseudoknot.

Conserved residues (UAGGGNNG) were also identified just downstream of the amber termination codon in tombusvirids that use PRT to synthesize their RdRp. The conserved guanylates adjacent to the UAG contributed to PRT efficiency, but not as simple participants in the S1 stem. In addition, the +5G was also important for readthrough (Fig. 2.3). Gamma retroviruses share the same conserved sequence downstream of their amber termination codon at the gag/pol junction (Fig. 2.13A). Alteration of +1G, +2G and +5G were reported to be highly detrimental for readthrough using dual-luciferase reporter constructs but not the intervening +3A and +4G[133]. +5G is located within the ribosome when the amber codon is in the A-site, suggesting that specific contacts may exist between this guanylate and ribosome components that support readthrough in these viruses.

The TCV RSE adopts basal and active structures

Tombusvirid RSE are structurally conserved, particularly in the lower stem regions (S1 and S2; Fig. 1B). However, SHAPE conducted both *in vivo* and *in vitro* did not support the presence of these stems in the TCV RSE. Rather, SHAPE profiles supported the predominance of an alternate structure, where residues on the 5' side of RSE S1 and

S2 pair with sequences located 100 nt upstream, extending the base of the upstream SLA hairpin. Hairpins in the same location as SLA are predicted to form upstream of RSE in all tombusvirids examined (Fig. S1). A hairpin in this location was also reported in umbraviruses and in coronaviruses where they serve as repressors of recoding[108,110]. In TCV, SLA also weakly repressed readthrough (Fig. 2.12D). Mutations disrupting the middle and upper portions of the extended TCV SLA also reduced gRNA accumulation *in vivo* (Fig. 2.12F). Although SLA in tombusvirids vary in sequence and structure, strong conservation of a hairpin just upstream of the recoding site suggests an important role for this hairpin in both readthrough and frameshifting.

While SHAPE only provides a structural snapshot of the majority of RNAs sampled, the results demonstrate that the structure with expanded SLA and truncated RSE (the basal structure) is the predominant form *in vitro* and at the 40 hpi time point *in vivo*. We propose that the basal structure is the readthrough inactive form of the RSE, which, when present, results in the ribosome terminating translation at the p28 amber codon. Although no structural evidence for the phylogenetically conserved active form of the TCV RSE currently exists, genetic evidence clearly demonstrates that RSE containing the S2 stem is critical for recoding *in vitro* (Fig. 2). The phylogenetically conserved structure that includes stem S1 is also not compatible with the required pseudoknot between S1 and L3 residues. This suggests that a series of conformational changes are needed for the recoding event. One possible scenario has ribosomes unwinding the lower stem of the extended SLA basal structure during translation, which causes structural changes in RSE L2 that lead to the establishment of the long-distance interaction with the 3'end, possibly helping to stabilizing the RSE active structure as was found for PEMV[110]. When the RSE

3' side S1 cytidylates form the pseudoknot, this might further stabilize the active structure leading to ribosome pausing and insertion of a tRNA that decodes the amber stop codon. Since mutating residues in positions 720-735 (the equivalent of ribosomes unwinding the lower stem of the extended SLA) does not by itself lead to formation of RSE stems S1 and S2, this implies a requirement to also release residues on the 3' side of S1 and S2 from their alternate pairing location in the basal structure, which is currently unknown.

Evidence for additional RSE conformations in unrelated viruses

Since viral RNAs are known to adopt alternate conformations of overlapping sequences to facilitate transitions between reversible processes, conformational switches modulating ribosome recoding might be widespread[134]. The presence of alternate structures has been suggested by SHAPE structure modeling of the RSE region in full-length HIV *in vitro*, which indicates that the previously proposed two-helix model was missing upstream and downstream sequences that together generate a more complex structure [69,135,136]. Ribosomes translating through this larger frameshifting domain were proposed to remodel portions of the extended structure into the conventionally accepted two-helix active structure. The RSE of MLV, in addition to having the same conserved residues as the TCV RSE (Fig. 2.13A), can also adopt an alternate hairpin conformation[137,138]. Mutations specifically affecting this alternate hairpin structure (but not the pseudoknotted structure) were highly detrimental to virus viability[137]. These data are consistent with MLV also having basal and active conformations of its RSE. Since

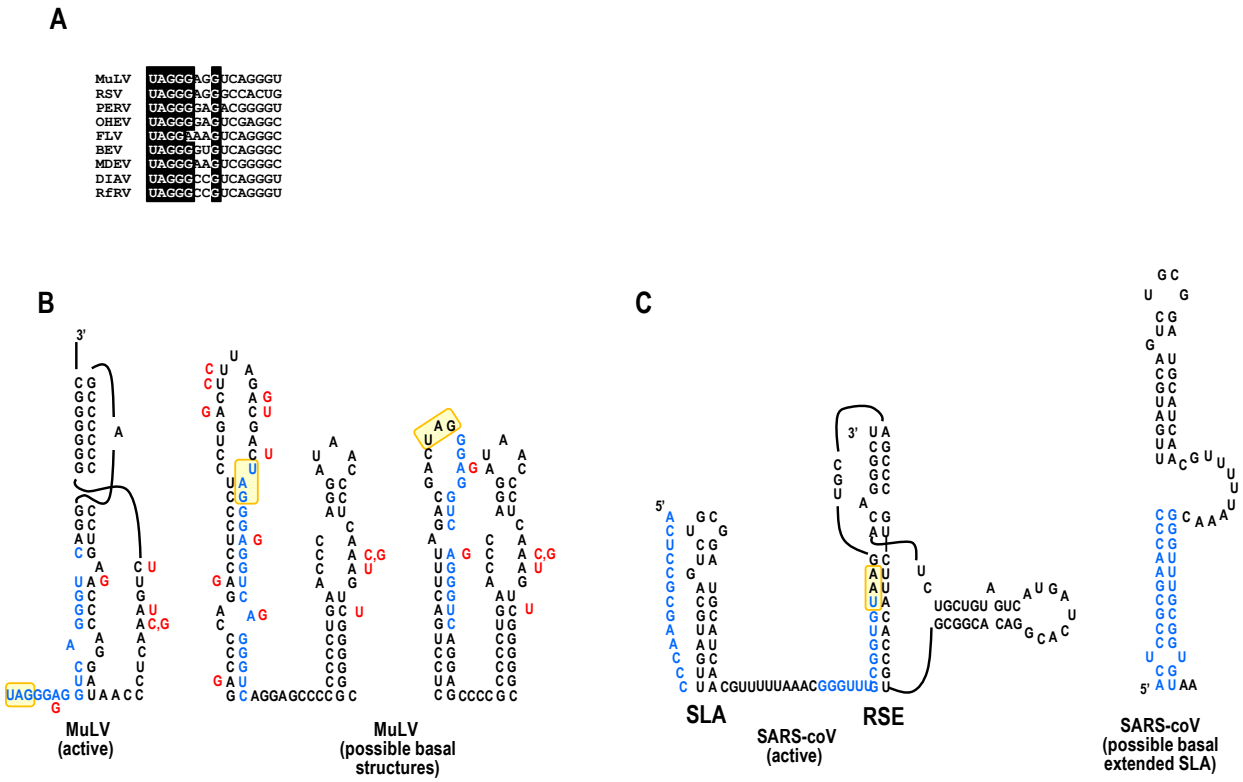

A
MuLV UAGGGAGGUCAGGGU
RSV UAGGGAGGSCCACUG
PERV UAGGGSAGACGGGGU
OHEV UAGGGSAGUCGAGGC
FLV UAGGGAAGUCAGGGC
BEV UAGGGGUGUCAGGGC
MDEV UAGGGAAGUCGGGGC
DIAV UAGGGCCGUCAGGGU
RfRV UAGGGCCGUCAGGGU
B
MuLV
(active)
MuLV
(possible basal
structures)
C
SLA
SARS-coV
(active)
RSE
SARS-coV
(possible basal
extended SLA)

Fig. 2. 15 Sequence conservation and conformational similarities between RSE of TCV and unrelated viruses.

(A) Conserved sequence at PRT site in tombusvirids is conserved in gamma retroviruses engaged in PRT of amber codons. Conserved sequence is highlighted. MLV, Murine leukemia virus; RSV, Rous sarcoma virus; PERV, Porcine endogenous retrovirus; OHEV, Odocoileus hemionus endogenous virus; FLV, Feline leukemia virus; BEV, Baboon endogenous retrovirus; MDEV, Mus dunni endogenous virus; DIAV, Duck infectious anemia virus; RfRV, Rhinolophus ferrumequinum retrovirus. (B) Proposed basal structure for MLV RSE using sequences upstream of RSE. Possible basal structure on right was previously proposed[137,138]. Structure in center is an alternate possibility involving additional 5' sequences not considered in the earlier study. Residues are colored blue to assist with orientation. All substitutions in 24 viruses examined (in red) maintain both active and proposed basal structures. Stop codon is boxed. (C) Proposed basal SLA structures for SARS-CoV. As with TCV, upstream SLA hairpin serves as a repressor of PRF (Su et al., 2005). Sequences in blue are capable of pairing.

the SARS-CoV SLA hairpin was also repressive for recoding [108], we examined the SARS-coV sequence in the SLA/RSE region to determine whether an alternate conformation might be possible that similarly extends the length of the SLA. As shown in Figure S6C, an alternate structure is predicted that extends the length of the SLA and contains a 10 bp stem. This SLA extension is formed by pairing sequence upstream of the SLA with sequence from the 5' side of the lower RSE stem, similar to TCV. Thus for TCV and possibly SARS-CoV, the SLA may serve as a scaffold for extension of its lower stem by base-pairing residues upstream of SLA with RSE residues, thereby disrupting the structure of the active RSE.

Note: Materials and methods can be found in Chapter 5

Chapter 3: Phylogenetic conservation of structural elements in *Calibrachoa mottle virus;* generation of an infectious clone

Introduction

Calibrachoa, a species of plants native to South America, is an economically important horticultural plant with small petunia-like flowers. CbMV was first isolated from Calibrachoa plants displaying leaf mottling, chlorotic blotch and interveinal yellowing symptoms[144]. CbMV is a positive-sense RNA virus of 3919 nucleotides with five overlapping open reading frames and is a member of the genus *Carmovirus*, family *Tombusviridae*. CbMV is transmitted mechanically, with no known vector. The genome is neither capped nor polyadenylated and the 5' and 3' UTRs have lengths of 34 and 234 nt respectively. The 5' ORF (p28) is a replicase associated protein, and its readthrough product (p87) is the RdRp (Fig. 3.1). ORF3 (p8) and ORF4 (p9) encode movement proteins expressed from sgRNA1, and ORF5 (p37, CP) encodes the capsid protein expressed from sgRNA2. A small additional ORF of 51 amino acids starting at nt 2962 and ending at nt 3117 has been predicted [144]. This putative 6kDa protein is nested within the CP in the same reading frame as the replicase and the p9 movement proteins. Two similar novel carmovirus proteins were reported nested within the CP of *Hibiscus Chlorotic Ringspot Virus* (HCRSV) as detected in WGE[145]. The 27 kDa protein was

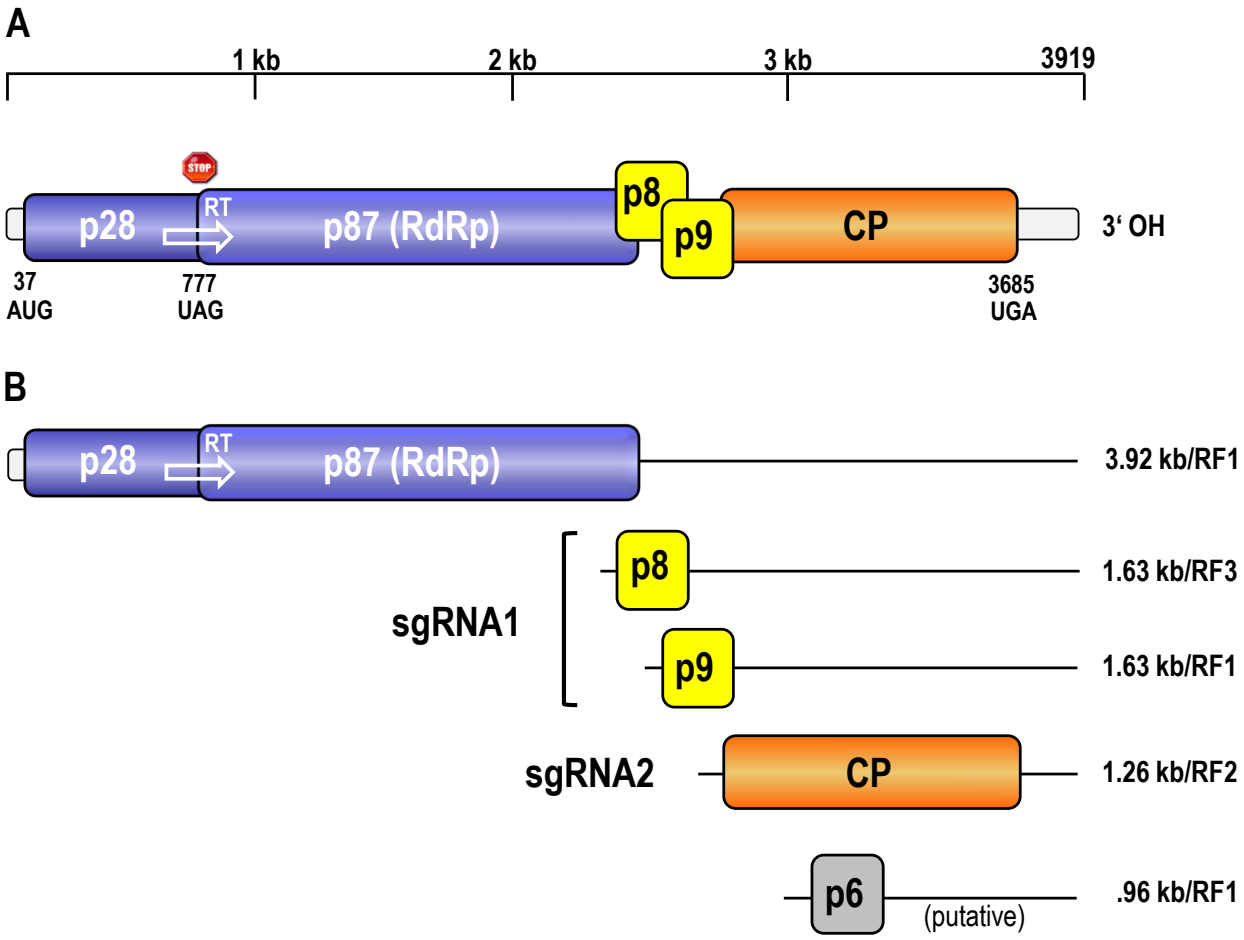


Fig. 3. 1 Genome organization of CbMV. (A) The five overlapping ORFs are shown, with the p28 readthrough site denoted with a white arrow. p28 is a replicase associated protein, with its readthrough product p87 expressing the RdRp. P8 and p9 are movement proteins, and p37 is the capsid. The ORF1 start codon, p28 readthrough stop codon and CP termination codon sites shown under genome. (B) Size of genomic and subgenomic RNAs are shown, immediately followed by the reading frame (RF 1-3). The putative p6 protein boxed in gray below CP.

shown to be a determinant of symptom severity, and the 25kDa protein is suggested to facilitate viral systemic movement[146].

I undertook this study to generate an infectious clone of CbMV and to characterize phylogenetically conserved structural elements within the genome, as no clone previously existed.

Results

Construction of the CbMV full-length infectious clone

CbMV infected tissue from *Nicotiana benthamiana* was obtained from the USDA (Genbank accession number GQ244431 – hereafter WT). Briefly, healthy *N. benthamiana* plants were mechanically inoculated with infected tissue. At 14 days post inoculation (dpi) symptomatic leaves were harvested and total RNA extracted. Reverse transcription was performed to produce full length cDNA that was used as template for PCR. Primers used for PCR were: forward, Kpn1 + T7 + CbMV sequence (see table for all primers); and reverse, CbMV sequence + SmaI + EcoRV (Fig. 3.2). Insert and vector (pUC19) were digested with KpnI and EcoRV and gel purified. Klenow was used to polish the ends. The vector was dephosphorylated, the insert phosphorylated and both ligated (1:3 molar ratio respectively) and transformed. Colony PCR was used to screen for the insert. Candidate plasmids were linearized and *in vitro* transcribed, and the RNA inoculated onto *N. benthamiana* plants.

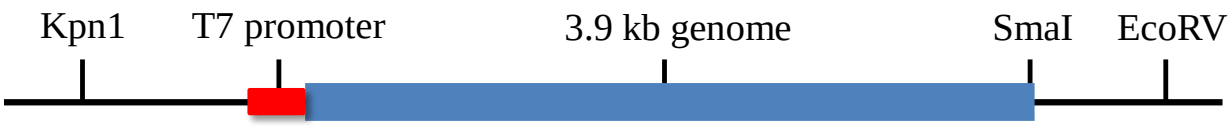


Fig. 3. 2 Construction of the CbMV infectious clone. Design of the CbMV insert with 5' and 3' restriction sites and T7 promoter location shown in relation to the CbMV genome.

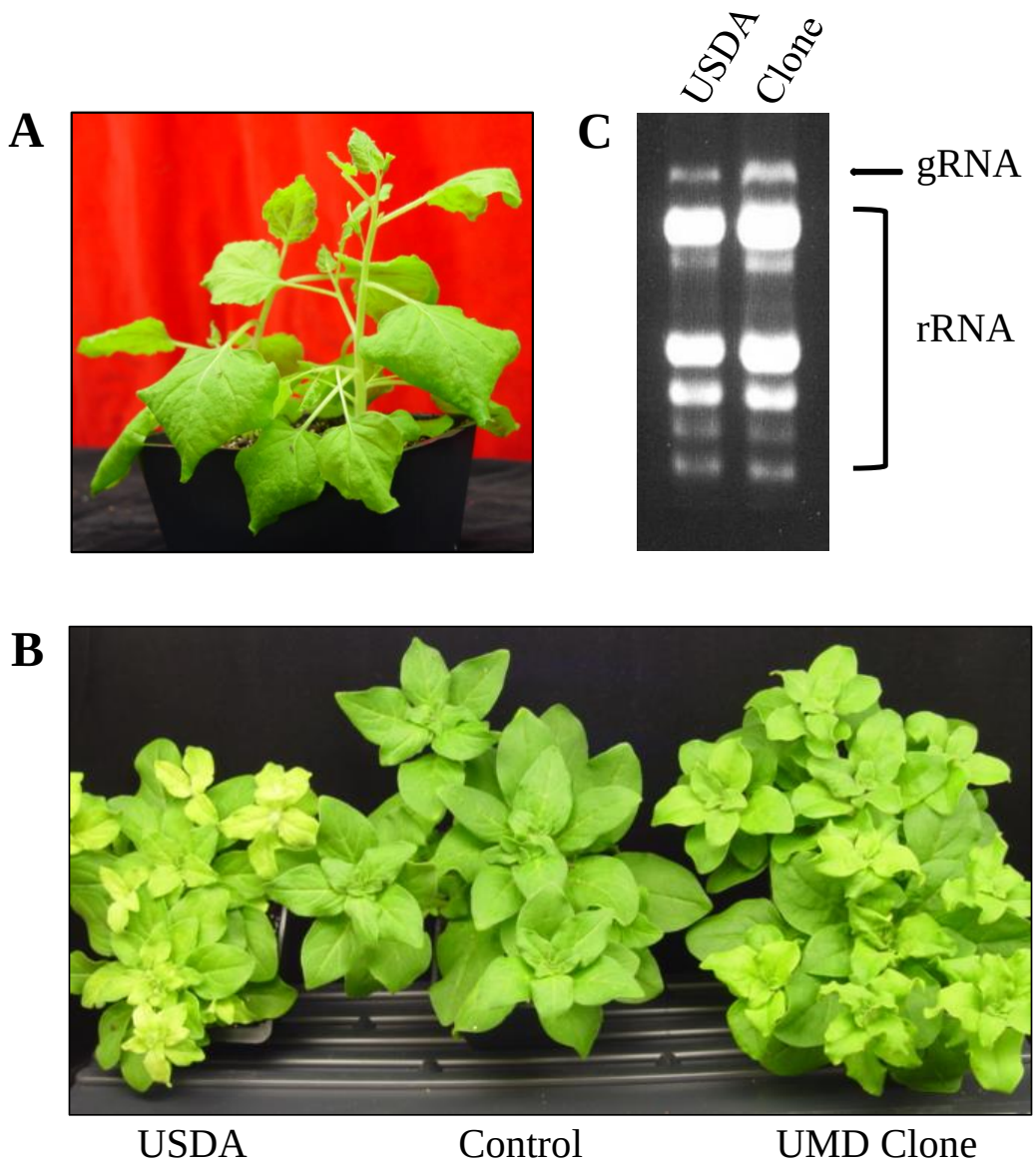


Fig. 3. 3 Identification of the CbMV infections clone. (A) Viral symptoms in *N. benthamiana* plants exhibiting mild chlorosis and severe epinasty. (B) Comparison of viral symptoms in wild petunia. Left, USDA WT isolate exhibiting chlorosis, stunting and leaf curling; center, control plant; and right, UMD CbMV clone displaying mild chlorosis and leaf curling. (C) Total RNA extracted from infected leaves of run on an ethidium stained agarose gel. Tissue from USDA or UMD clone inoculated plants, as well as viral and ribosomal bands noted.

Identification of the infectious clone

A single *N. benthamiana* plant began to exhibit symptoms (chlorosis, mottling and stunting) similar to that of WT approximately 10 dpi (Fig. 3.3A). The clone additionally exhibited pronounced leaf epinasty (leaf curling), a symptom not observed with WT. Symptoms in petunia (a species closely related to Calibrachoa) were less severe than that of WT, with less chlorosis and less stunting observed overall (Fig. 3.3B). Total RNA was extracted from symptomatic leaves and run on an agarose gel where a putative viral band was visible above the ribosomal bands, with the size of the band corresponding to the WT virus (Fig. 3.3C). The plasmid was sent for sequencing using both forward and reverse primers; contigs were assembled using CodonCode Aligner software (CodonCode Co., MA, USA), and alignments performed with ClustalW (Fig. 3.4)[147]. The UMD clone differed from WT in 11 positions, all of them silent. No differences were observed within the 5' UTR. P28 (p88) contained 9 changes upstream of the p28 stop codon: 3 purine:purine; 4 pyrimidine:pyrimidine; 1 purine:pyrimidine; and 1 pyrimidine:purine. Neither movement proteins contained changes, however, the CP had a single purine:purine mutation and the 3' UTR had a single pyrimidine:pyrimidine alteration.

***In Silico* analysis of CbMV 3' UTR secondary structure**

As discussed in Chapter 1, the TCV 3' UTR is comprised of five hairpins, and three pseudoknots. As compared to TCV, the type member of the genus Carmovirus, CbMV is predicted to have the ubiquitous 3' elements H5, Ψ_1 and Pr (Fig. 3.5). The CbMV Pr loop is predicted to pair with an asymmetrical internal bulge loop on the RSE to facilitate

CbMV UMD Clone Alignment with USDA Isolate GQ244431

5'UTR

```
     1 GGATAAACTTAGCTTTATCTCGGTGATTGAGCAATC 36
       |||||||||||||||||||||||||||||||||||
     1 GGATAAACTTAGCTTTATCTCGGTGATTGAGCAATC 36
```

p28 Replicase-associated protein
p377 RdRp

```
CbMV_p28/RdRp_GQ244431  ATGGGTATACTACAATTGGCAAAAGAACTAACCGTAGGAGGAGTGTGTTG 50
UMD _p28/RdRp           ATGGGTATACTACAATTGGCAAAAGAACTGACCGTAGGAGGAGTGTGTTG 50
                        ************************** ********************

CbMV_p28/RdRp_GQ244431  CACAACTCTACTATGCGTTGGGGTGGCAGCCCTGGAAGTACGCCTAGCTT 100
UMD _p28/RdRp           CACTACTCTACTATGCGTTGGGGTGGCAGCCCTGGAAGTACGCCTAGCTT 100
                        *** ********************************************

CbMV_p28/RdRp_GQ244431  TAGGAGCATACGAGTTTAACAAGCAATGTATCAGCAACGTTAGAGGTTTT 150
UMD _p28/RdRp           TAGGAGCATACGAGTTTAACAAGCAATGTATCAGCAACGTTAGAGGTTTC 150
                        ************************************************

CbMV_p28/RdRp_GQ244431  ATAGAATCAAGGGGACAGTCACACCCTTCAGAACCCAAGTATGCTAAAGC 200
UMD _p28/RdRp           ATAGAATCAAGGGGACAGTCACACCCTTCAGAACCCAAGTATGCTAAAGC 200
                        ************************************************

CbMV_p28/RdRp_GQ244431  TGAAGTGTCACCTTTCCAAGCTGAATTGGATGACGATCTGGAAGATGACG 250
UMD _p28/RdRp           TGAAGTGTCACCTTTCCAAGCTGAATTGGATGACGATCTGGAAGATGACG 250
                        ************************************************

CbMV_p28/RdRp__GQ244431 CAGAAATGCGAAACTATCTGGAAAAACATGAAGACAAGGAAAAGGATGAC 300
UMD _p28/RdRp           CAGAAATGCGAAACTATCTAGAAAAACATGAAGACAAGGAAAAGGATGAC 300
                        ****************** ****************************

CbMV_p28/RdRp_GQ244431  GAGGGCAAAGTGGTGGTGAAAAATGTGCGAACAACCTTACCTAGAAATAG 350
UMD _p28/RdRp           GAGGGCAAAGTGGTGGTGAAAAACGTGCGAACAACCTTACCTAGAAATAG 350
                        ********************** *************************

CbMV_p28/RdRp_GQ244431  GCATACCAAAGGAAAATTCTTGAAAAGGTTAGTCGCCGACACTAAAAACC 400
UMD _p28/RdRp           GCATACCAAAGGAAAATTTTTGAAGAGGTTAGTCGCCGACACTAAAAATC 400
                        ****************** ***** ********************** *

CbMV_p28/RdRp_GQ244431  ACTTTGGCGGAACCCCCACCCCCACTGATGCTAATAGACTAGCTGTGATG 450
UMD _p28/RdRp           ACTTTGGCGGAACCCCCACCCCCACTGATGCGAATAGACTAGCTGTGATG 450
                        ******************************** ******************

CbMV_p28/RdRp_GQ244431  AAATACATGGTGGGAAGGTGTCGAGAACATCATATGGTAGACTTGCACAT 500
UMD _p28/RdRp           AAATACATGGTGGGAAGGTGTCGAGAACATCATATGGTAGACTTGCACAT 500
                        ************************************************

CbMV_p28/RdRp_GQ244431  TCGGCAGGTCACTGAACTGGCAAAAGCAGCAGTGTTCACTCCAGACATTC 550
UMD _p28/RdRp           TCGGCAGGTCACTGAACTGGCAAAAGCAGCAGTGTTCACTCCAGACATTC 550
                        ************************************************

CbMV_p28/RdRp_GQ244431  TAGAAGTGCAAAGTGTGCAATTGCTCAACTCTTATCCAGCCTACCGGCGG 600
UMD _p28/RdRp           TAGAAGTGCAAAGTGTGCAATTGCTCAACTCTTATCCAGCCTACCGGCGG 600
                        ************************************************


CbMV_p28/RdRp_GQ244431  CGGTGCGCTCTCCACAAAGCGCATCAAGTGCAGGTGTGGAAAGAGTTGCT 650
UMD _p28/RdRp           CGGTGCGCTCTCCACAAAGCGCATCAAGTGCAGGTGTGGAAAGAGTTGCT 650
```

```
                          **************************************************
CbMV_p28/RdRp_GQ244431    GACTAATTGCTTCCATAAGGATGCCTGGGAGTATATTTGGTTCCGGATGA 700
UMD _p28/RdRp             GACTAATTGCTTCCATAAGGATGCCTGGGAGTATATTTGGTTCCGGATGA 700
                          **************************************************

CbMV_p28/RdRp_GQ244431    ATGGGGGATTGGCTCGATCCCCTTTTCAGTTCCATAAATAGGGGGGTATA 750
UMD _p28/RdRp             ATGGGGGATTGGCTCGATCCCCTTTTCAGTTCCATAAATAGGGGGGTATA 750
                          **************************************************

CbMV_RdRp__GQ244431       ACCTTTCTCGAGGGTGTGTGTACAAAAATAGCCAGGGGGGCACACCCTTA 800
UMD _RdRp                 ACCTTTCTCGAGGGTGTGTGTACAAAAATAGCCAGGGGGGCACACCCTTA 800
                          **************************************************

CbMV_RdRp__GQ244431       CCTCAAAGAGAAAGTTATACCTCGAGCCCCCAAACTTAGGAAACTTTACC 850
UMD _RdRp                 CCTCAAAGAGAAAGTTATACCTCGAGCCCCCAAACTTAGGAAACTTTACC 850
                          **************************************************

CbMV_RdRp__GQ244431       TCCAACAATCCTGCACCTCCGGTTTGCAGTACCGGGTGCATAATAACAGC 900
UMD _RdRp                 TCCAACAATCCTGCACCTCCGGTTTGCAGTACCGGGTGCATAATAACAGC 900
                          **************************************************

CbMV_RdRp__GQ244431       ATCGCCAATTTACGCAGAGGATTACTGGAGCGGGTTTTCTACGTGGAAAA 950
UMD _RdRp                 ATCGCCAATTTACGCAGAGGATTACTGGAGCGGGTTTTCTACGTGGAAAA 950
                          **************************************************

CbMV_RdRp__GQ244431       TAAAGATACAAAAAAACTGCAGACTTGTCCAGAGCCAGAAGCCGGTATTT 1000
UMD _RdRp                 TAAAGATACAAAAAAACTGCAGACTTGTCCAGAGCCAGAAGCCGGTATTT 1000
                          **************************************************

CbMV_RdRp__GQ244431       TTAAAGAGCTTAAAACGATCAGACAGCAATTTGTAAGGTTATGCGGTCAT 1050
UMD _RdRp                 TTAAAGAGCTTAAAACGATCAGACAGCAATTTGTAAGGTTATGCGGTCAT 1050
                          **************************************************

CbMV_RdRp__GQ244431       CATACCCGGATCTCCACAGAGCAATTTGTGGATTGTTATCAGGGCAGGAA 1100
UMD _RdRp                 CATACCCGGATCTCCACAGAGCAATTTGTGGATTGTTATCAGGGCAGGAA 1100
                          **************************************************

CbMV_RdRp__GQ244431       ACGCACAATCTACCAAAAAGCTGCAGACTCATTGAGCGAAATAGCCATAG 1150
UMD _RdRp                 ACGCACAATCTACCAAAAAGCTGCAGACTCATTGAGCGAAATAGCCATAG 1150
                          **************************************************

CbMV_RdRp__GQ244431       ACCGGTCTGACTCACGTCTAAAGACTTTTGTGAAGGCAGAGAAGTTTTGT 1200
UMD _RdRp                 ACCGGTCTGACTCACGTCTAAAGACTTTTGTGAAGGCAGAGAAGTTTTGT 1200
                          **************************************************

CbMV_RdRp__GQ244431       ATTGATCTCAAACCAGATCCGGCCCCCAGAGTGATCCAACCTAGAATGCC 1250
UMD _RdRp                 ATTGATCTCAAACCAGATCCGGCCCCCAGAGTGATCCAACCTAGAATGCC 1250
                          **************************************************

CbMV_RdRp__GQ244431       CAGATATAACGTAGAGCTGGGAAGATACCTAAAGAAGGTGGAACACTCTG 1300
UMD _RdRp                 CAGATATAACGTAGAGCTGGGAAGATACCTAAAGAAGGTGGAACACTCTG 1300
                          **************************************************

CbMV_RdRp__GQ244431       TTATCGCGCCTTAGATAAAATTTGGGGTGGCAGGACGGTGATGAAAGGG 1350
UMD _RdRp                 CTTATCGCGCCTTAGATAAAATTTGGGGTGGCAGGACGGTGATGAAAGGG 1350
                          **************************************************

CbMV_RdRp__GQ244431       TATACTGTCGAAGAAGTTGGTATGATAATCAGTGATGCCTGGGACCAATT 1400
UMD _RdRp                 TATACTGTCGAAGAAGTTGGTATGATAATCAGTGATGCCTGGGACCAATT 1400
                          **************************************************

CbMV_RdRp__GQ244431       CCACATGCCAGTTGCTGTTGGATTTGACATGAGTCGATTCGACCAACACG 1450
UMD _RdRp                 CCACATGCCAGTTGCTGTTGGATTTGACATGAGTCGATTCGACCAACACG 1450
                          **************************************************



CbMV_RdRp__GQ244431       TAAGTGTGCCAGCATTACAGTTCGAACACACCTGCTACATGTCTCTATTT 1500
UMD _RdRp                 TAAGTGTGCCAGCATTACAGTTCGAACACACCTGCTACATGTCTCTATTT 1500
                          **************************************************
```

```
CbMV_RdRp__GQ244431     CCAGGAGACAGGCACCTGCAACAACTACTGTCATGGCAATTACGAAATTA 1550
UMD _RdRp               CCAGGAGACAGGCACCTGCAACAACTACTGTCATGGCAATTACGAAATTA 1550
                        **************************************************

CbMV_RdRp__GQ244431     TGGAGTCGGGGTAGCAAGCAATGGCATCCTTAGGTACAAAGTAGATGGCA 1600
UMD _RdRp               TGGAGTCGGGGTAGCAAGCAATGGCATCCTTAGGTACAAAGTAGATGGCA 1600
                        **************************************************

CbMV_RdRp__GQ244431     AGAGAATGAGTGGTGACATGAATACAGCTTTGGGTAATTGCATCTTGGCA 1650
UMD _RdRp               AGAGAATGAGTGGTGACATGAATACAGCTTTGGGTAATTGCATCTTGGCA 1650
                        **************************************************

CbMV_RdRp__GQ244431     TGTTTGATAACCAAACATTTGTTTCCGGGTAATTACAGGCTCATCAACAA 1700
UMD _RdRp               TGTTTGATAACCAAACATTTGTTTCCGGGTAATTACAGGCTCATCAACAA 1700
                        **************************************************

CbMV_RdRp__GQ244431     TGGTGACGATTGTGTTCTCATAACAGAGCGCAACAACCTCCCCGAGATTA 1750
UMD _RdRp               TGGTGACGATTGTGTTCTCATAACAGAGCGCAACAACCTCCCCGAGATTA 1750
                        **************************************************

CbMV_RdRp__GQ244431     CTTCCAAGCTGGAGGCAGGGTGGAGGAGATTTGGATTCACCTGCATTTCT 1800
UMD _RdRp               CTTCCAAGCTGGAGGCAGGGTGGAGGAGATTTGGATTCACCTGCATTTCT 1800
                        **************************************************

CbMV_RdRp__GQ244431     GAGAAGCCTGTTTTTGTCAAAGAGGAAATAGAATTTTGCCAAATGCAACC 1850
UMD _RdRp               GAGAAGCCTGTTTTTGTCAAAGAGGAAATAGAATTTTGCCAAATGCAACC 1850
                        **************************************************

CbMV_RdRp__GQ244431     TGTCTATGATGGTCAAACTTATGTCATGGTCCGTAAGCCCTATATCTCTA 1900
UMD _RdRp               TGTCTATGATGGTCAAACTTATGTCATGGTCCGTAAGCCCTATATCTCTA 1900
                        **************************************************

CbMV_RdRp__GQ244431     TGTCTAAGGATGCTTTTAGTTTAACCCCGTGGCCAAACGAAAAAGCATGT 1950
UMD _RdRp               TGTCTAAGGATGCTTTTAGTTTAACCCCGTGGCCAAACGAAAAAGCATGT 1950
                        **************************************************

CbMV_RdRp__GQ244431     CAGCAGTGGTTGGGTGCAGTAGGTATGTGTGGAGAGAGACTGTGCGGAAA 2000
UMD _RdRp               CAGCAGTGGTTGGGTGCAGTAGGTATGTGTGGAGAGAGACTGTGCGGAAA 2000
                        **************************************************

CbMV_RdRp__GQ244431     GATACCAATCTTGCAGTCTTATTATCAAGCATATTGCAGAGCTAACAAAG 2050
UMD _RdRp               GATACCAATCTTGCAGTCTTATTATCAAGCATATTGCAGAGCTAACAAAG 2050
                        **************************************************

CbMV_RdRp__GQ244431     GGCGAAAGCCAAAAATGGATGCGCACGAGAAGGGGGGTATGTACATGCTA 2100
UMD _RdRp               GGCGAAAGCCAAAAATGGATGCGCACGAGAAGGGGGGTATGTACATGCTA 2100
                        **************************************************

CbMV_RdRp__GQ244431     GCTATGAATTCGAAGCGGGCTTTTGGGGAGATTTCCCAGGATTGTAGGTT 2150
UMD _RdRp               GCTATGAATTCGAAGCGGGCTTTTGGGGAGATTTCCCAGGATTGTAGGTT 2150
                        **************************************************

CbMV_RdRp__GQ244431     TTCATTCTATAAAGCTTTTGGTATTACACCTGACCAACAGGTTGCCATCG 2200
UMD _RdRp               TTCATTCTATAAAGCTTTTGGTATTACACCTGACCAACAGGTTGCCATCG 2200
                        **************************************************

CbMV_RdRp__GQ244431     AAAGCTTGATAGATTGCCATGAAATCTCCACGGAACCCGGCCCGCCTGAA 2250
UMD _RdRp               AAAGCTTGATAGATTGCCATGAAATCTCCACGGAACCCGGCCCGCCTGAA 2250
                        **************************************************

CbMV_RdRp__GQ244431     TCATGGAGCGCGAACATCCCACTATTAATAAAGCAATAA 2289
UMD _RdRp               TCATGGAGCGCGAACATCCCACTATTAATAAAGCAATAA 2289
                        ****************************************
```

p28 Replicase-associated Protein
p377 RdRp

```
p28/RdRp_GQ244431       MGILQLAKELTVGGVCCTTLLCVGVAALEVRLALGAYEFNKQCISNVRGFIESRGQSHPS 60
UMD p28/RdRp            MGILQLAKELTVGGVCCTTLLCVGVAALEVRLALGAYEFNKQCISNVRGFIESRGQSHPS 60
                        ************************************************************

p28/RdRp _GQ244431      EPKYAKAEVSPFQAELDDDLEDDAEMRNYLEKHEDKEKDDEGKVVVKNVRTTLPRNRHTK 120
UMD _p28/RdRp           EPKYAKAEVSPFQAELDDDLEDDAEMRNYLEKHEDKEKDDEGKVVVKNVRTTLPRNRHTK 120
                        ************************************************************

p28/RdRp _GQ244431      GKFLKRLVADTKNHFGGTPTPTDANRLAVMKYMVGRCREHHMVDLHIRQVTELAKAAVFT 180
UMD _p28/RdRp           GKFLKRLVADTKNHFGGTPTPTDANRLAVMKYMVGRCREHHMVDLHIRQVTELAKAAVFT 180
                        ************************************************************

p28/RdRp _GQ244431      PDILEVQSVQLLNSYPAYRRRCALHKAHQVQVWKELLTNCFHKDAWEYIWFRMNGGLARS 240
UMD _p28/RdRp           PDILEVQSVQLLNSYPAYRRRCALHKAHQVQVWKELLTNCFHKDAWEYIWFRMNGGLARS 240
                        ************************************************************

p28/RdRp_GQ244431       PFQFHKSTOPGGITFLEGVCTKIARGAHPYLKEKVIPRAPKLRKLYLQQSCTSGLQYRVH 300
UMD _p28/RdRp           PFQFHKSTOPGGITFLEGVCTKIARGAHPYLKEKVIPRAPKLRKLYLQQSCTSGLQYRVH 300
                        ************************************************************

CbMV_RdRp_GQ244431      NNSIANLRRGLLERVFYVENKDTKKLQTCPEPEAGIFKELKTIRQQFVRLCGHHTRISTE 360
UMD _RdRp               NNSIANLRRGLLERVFYVENKDTKKLQTCPEPEAGIFKELKTIRQQFVRLCGHHTRISTE 360
                        ************************************************************

CbMV_RdRp_GQ244431      QFVDCYQGRKRTIYQKAADSLSEIAIDRSDSRLKTFVKAEKFCIDLKPDPAPRVIQPRMP 420
UMD _RdRp               QFVDCYQGRKRTIYQKAADSLSEIAIDRSDSRLKTFVKAEKFCIDLKPDPAPRVIQPRMP 420
                        ************************************************************

CbMV_RdRp_GQ244431      RYNVELGRYLKKVEHSAYRALDKIWGGRTVMKGYTVEEVGMIISDAWDQFHMPVAVGFDM 480
UMD _RdRp               RYNVELGRYLKKVEHSAYRALDKIWGGRTVMKGYTVEEVGMIISDAWDQFHMPVAVGFDM 480
                        ************************************************************

CbMV_RdRp_GQ244431      SRFDQHVSVPALQFEHTCYMSLFPGDRHLQQLLSWQLRNYGVGVASNGILRYKVDGKRMS 540
UMD _RdRp               SRFDQHVSVPALQFEHTCYMSLFPGDRHLQQLLSWQLRNYGVGVASNGILRYKVDGKRMS 540
                        ************************************************************

CbMV_RdRp_GQ244431      GDMNTALGNCILACLITKHLFPGNYRLINNGDDCVLITERNNLPEITSKLEAGWRRFGFT 600
UMD _RdRp               GDMNTALGNCILACLITKHLFPGNYRLINNGDDCVLITERNNLPEITSKLEAGWRRFGFT 600
                        ************************************************************

CbMV_RdRp_GQ244431      CISEKPVFVKEEIEFCQMQPVYDGQTYVMVRKPYISMSKDAFSLTPWPNEKACQQWLGAV 660
UMD _RdRp               CISEKPVFVKEEIEFCQMQPVYDGQTYVMVRKPYISMSKDAFSLTPWPNEKACQQWLGAV 660
                        ************************************************************

CbMV_RdRp_GQ244431      GMCGERLCGKIPILQSYYQAYCRANKGRKPKMDAHEKGGMYMLAMNSKRAFGEISQDCRF 720
UMD _RdRp               GMCGERLCGKIPILQSYYQAYCRANKGRKPKMDAHEKGGMYMLAMNSKRAFGEISQDCRF 720
                        ************************************************************

CbMV_RdRp_GQ244431      SFYKAFGITPDQQVAIESLIDCHEISTEPGPPESWSANIPLLIKQSTOP 769
UMD _RdRp               SFYKAFGITPDQQVAIESLIDCHEISTEPGPPESWSANIPLLIKQSTOP 769
                        ************************************************
```

p37 Movement protein

```
CbMV_p37_GQ244431       ATGGAGCGCGAACATCCCACTATTAATAAAGCAATAAGCACCAAGGAAAAGAGTAAACAG 60
UMD _p8                 ATGGAGCGCGAACATCCCACTATTAATAAAGCAATAAGCACCAAGGAAAAGAGTAAACAG 60
                        ************************************************************

CbMV_p37_GQ244431       CTGAACAGTTCAAAGGATAAAAATAAACTGAGTGGCAAGTTAACTGCAGCTAAAGCTGTA 120
UND _p8                 CTGAACAGTTCAAAGGATAAAAATAAACTGAGTGGCAAGTTAACTGCAGCTAAAGCTGTA 120
                        ************************************************************
```

```
CbMV_p37_GQ244431       GCTAATGAACAAGCACGTGGTAGTGTTTACGGGGGTAGTTTCACTAATGTTGCTAGGGAG 180
UMD _p8                 GCTAATGAACAAGCACGTGGTAGTGTTTACGGGGGTAGTTTCACTAATGTTGCTAGGGAG 180
                        ************************************************************

CbMV_p37_GQ244431       ATTAAGATGGAGATCCATTTCCATTTTTGA 210
UMD _p8                 ATTAAGATGGAGATCCATTTCCATTTTTGA 210
                        ****************************
```

p9 Movement Protein

```
CbMV_p9_GQ244431        ATGAACAAGCACGTGGTAGTGTTTACGGGGGTAGTTTCACTAATGTTGCTAGGGAGATTA 60
UMD _p9                 ATGAACAAGCACGTGGTAGTGTTTACGGGGGTAGTTTCACTAATGTTGCTAGGGAGATTA 60
                        ************************************************************

CbMV_p9_GQ244431        AGATGGAGATCCATTTCCATTTTTGACTTCTTTCCCCCAGTGAATATCCCTAACATAAAC 120
UMD _p9                 AGATGGAGATCCATTTCCATTTTTGACTTCTTTCCCCCAGTGAATATCCCTAACATAAAC 120
                        ************************************************************

CbMV_p9_GQ244431        CAATCGATAGCACTTGCACTCTGTGGATTAATCCTAAACTGTATAGGAAAGGCAGAACCC 180
UMD _p9                 CAATCGATAGCACTTGCACTCTGTGGATTAATCCTAAACTGTATAGGAAAGGCAGAACCC 180
                        ************************************************************

CbMV_p9_GQ244431        TCATACAGCTATTATTCTACACACGACAGCAGTAAAACCCAATACATCAAGATCAATACA 240
UMD _p9                 TCATACAGCTATTATTCTACACACGACAGCAGTAAAACCCAATACATCAAGATCAATACA 240
                        ************************************************************

CbMV_p9_GQ244431        CCTGATGGATAA 252
UMD _p9                 CCTGATGGATAA 252
                        ************
```

p37 Coat protein

```
CbMV_p37_GQ244431       ATGGATAACTACAAGAACTCACCTGTCATCACTACCTTAGCTAACAAGGGTGTTCCATGG 60
UMD _p37                TGGATAACTACAAGAACTCACCTGTCATCACTACCTTAGCTAACAAGGGTGTTCCATGGG 60
                        ************************************************************

CbMV_p37_GQ244431       GCGATCAAGTTTAAAACTAAAACTTGGCAGGCGTTAACGCCAAACCAAAAGAAGCTTGCT 120
UMD _p37                GCGATCAAGTTTAAAACTAAAACTTGGCAGGCGTTAACGCCAAACCAAAAGAAGCTTGCT 120
                        ************************************************************

CbMV_p37_GQ244431       CGTGAGGCGTTAGGCATGAACCTAACTGCCACCGTGATAATACCCAAGAGAGCTCGAGGT 180
UMD _p37                CGTGAGGCGTTAGGCATGAACCTAACTGCCACCGTGATAATACCCAAGAGAGCTCGAGGT 180
                        ************************************************************

CbMV_p37_GQ244431       AGCCCAGCTGTAGCCAAGCCGAATAGGCTTGGCCCTGGTACGGCTGGCAAAACGTCTACC 240
UMD _p37                AGCCCAGCTGTAGCCAAGCCGAATAGGCTTGGCCCTGGTACGGCTGGCAAAACGTCTACC 240
                        ************************************************************

CbMV_p37_GQ244431       TGCACAGGATCAGAACTACTCCTAACTTTGCCCAAGCAGACTGGTTACACGCCACTTACT 300
UMD _p37                TGCACAGGATCAGAACTACTCCTAACTTTGCCCAAGCAGACTGGTTACACGCCACTTACT 300
                        ************************************************************

CbMV_p37_GQ244431       AAATCATGGATATTGAATCCAGGTCAATACAGCCCCTTCCGCCGAGCGTCATTAATGTCC 360
UMD _p37                AAATCATGGATATTGAATCCAGGTCAATACAGCCCCTTCCGCCGAGCGTCATTAATGTCC 360
                        ************************************************************

CbMV_p37_GQ244431       AGCATGTACAACAAGTATATGTTTACAAACATAAAGGTGCGGTGGACCACTACAGCCTCT 420
UMD _p37                AGCATGTACAACAAGTATATGTTTACAAACATAAAGGTGCGGTGGACCACTACAGCCTCT 420
                        ************************************************************

CbMV_p37_GQ244431       TTCGAATCATCAGGAAGAATTGTCCTGGCTTACAACAGTGATAGTTCAGACCCGGTGCCT 480
UMD_p37                 TTCGAATCATCAGGAAGAATTGTCCTGGCTTACAACAGTGATAGTTCAGACCCGGTGCCT 480
                        ************************************************************

CbMV_p37_GQ244431       ACTAAAGCTGGAATAGTGGAGTTTAAAACAAGGGCGGAGAACGTGGTGACCACAAGCTTT 540
UMD_p37                 ACTAAAGCTGGAATAGTGGAGTTTAAAACAAGGGCGGAGAACGTGGTGACCACAAGCTTT 540
                        ************************************************************
```

```
CbMV_p37_GQ244431       ATATTGGACATTCCAGGAGATGGAAAATATAGATACTGCAGAGACTCGACAAGTAACGAC 600
UMD _p37                ATATTGGACATTCCAGGAGATGGAAAATATAGATACTGCAGGGACTCGACAAGTAACGAC 600
                        ************************************** ******************

CbMV_p37_GQ244431       CCCAAACTAGTGGACTTCGGAAGACTTATCGTAATGCACTATGGGGCCGCGGAATCAGAC 660
UMD _p37                CCCAAACTAGTGGACTTCGGAAGACTTATCGTAATGCACTATGGGGCCGCGGAATCAGAC 660
                        ************************************************************

CbMV_p37_GQ244431       TCAGCTTACTTGGGAGAGGTTTTAGTTGACTACACCGTCGTGTTTTCTGAACCAATTCCT 720
UMD _p37                TCAGCTTACTTGGGAGAGGTTTTAGTTGACTACACCGTCGTGTTTTCTGAACCAATTCCT 720
                        ************************************************************

CbMV_p37_GQ244431       ACCGGGTCGATCACGCAACAAGGCGAACAGCTCGTCTCAGATGGTCCAGGATATGCCTTT 780
UMD _p37                ACCGGGTCGATCACGCAACAAGGCGAACAGCTCGTCTCAGATGGTCCAGGATATGCCTTT 780
                        ************************************************************

CbMV_p37_GQ244431       GTCACTGTACAACCGACTACTTTTCGCCTCACGATATACGGAGAGGGTAAATGGCTGGTA 840
UMD _p37                GTCACTGTACAACCGACTACTTTTCGCCTCACGATATACGGAGAGGGTAAATGGCTGGTA 840
                        ************************************************************

CbMV_p37_GQ244431       GTTTGGCAGTCGAGTACTGCCACTCCCGATGTAAATATCAAGGGAGACGGTGCCAAGGCA 900
UMD _p37                GTTTGGCAGTCGAGTACTGCCACTCCCGATGTAAATATCAAGGGAGACGGTGCCAAGGCA 900
                        ************************************************************

CbMV_p37_GQ244431       CACATCACATCATCTGCTGACAGTAAGACGGTTATAGCTGTGGTTACAGCTGAATTAGAA 960
UMD _p37                CACATCACATCATCTGCTGACAGTAAGACGGTTATAGCTGTGGTTACAGCTGAATTAGAA 960
                        ************************************************************

CbMV_p37_GQ244431       GGGGCTTACTTGGAAAGTACTACTTTAGCTGCTGTATCAGGGCTTAAGTGGTACGTTTCT 1020
UMD _p37                GGGGCTTACTTGGAAAGTACTACTTTAGCTGCTGTATCAGGGCTTAAGTGGTACGTTTCT 1020
                        ************************************************************

CbMV_p37_GQ244431       CGGTTATGA 1029
UMD _p37                CGGTTATGA 1029
                        *********
```

p37 translation

```
CbMV_p37_GQ244431       MDNYKNSPVITTLANKGVPWAIKFKTKTWQALTPNQKKLAREALGMNLTATVIIPKRARG 60
UMD _p37                MDNYKNSPVITTLANKGVPWAIKFKTKTWQALTPNQKKLAREALGMNLTATVIIPKRARG 60
                        ************************************************************

CbMV_p37_GQ244431       SPAVAKPNRLGPGTAGKTSTCTGSELLLTLPKQTGYTPLTKSWILNPGQYSPFRRASLMS 120
UMD _p37                SPAVAKPNRLGPGTAGKTSTCTGSELLLTLPKQTGYTPLTKSWILNPGQYSPFRRASLMS 120
                        ************************************************************

CbMV_p37_GQ244431       SMYNKYMFTNIKVRWTTTASFESSGRIVLAYNSDSSDPVPTKAGIVEFKTRAENVVTTSF 180
UMD _p37                SMYNKYMFTNIKVRWTTTASFESSGRIVLAYNSDSSDPVPTKAGIVEFKTRAENVVTTSF 180
                        ************************************************************

CbMV_p37_GQ244431       ILDIPGDGKYRYCRDSTSNDPKLVDFGRLIVMHYGAAESDSAYLGEVLVDYTVVFSEPIP 240
UMD _p37                ILDIPGDGKYRYCRDSTSNDPKLVDFGRLIVMHYGAAESDSAYLGEVLVDYTVVFSEPIP 240
                        ************************************************************

CbMV_p37_GQ244431       TGSITQQGEQLVSDGPGYAFVTVQPTTFRLTIYGEGKWLVVWQSSTATPDVNIKGDGAKA 300
UMD _p37                TGSITQQGEQLVSDGPGYAFVTVQPTTFRLTIYGEGKWLVVWQSSTATPDVNIKGDGAKA 300
                        ************************************************************

CbMV_p37_GQ244431       HITSSADSKTVIAVVTAELEGAYLESTTLAAVSGLKWYVSRLSTOP 346
UMD _p37                HITSSADSKTVIAVVTAELEGAYLESTTLAAVSGLKWYVSRLSTOP 346
                        *********************************************
```

3′ UTR

```
CbMV_3'UTR_GQ244431     ACACTAAGAGTTACTGCCACCAAAAGAGGATAATCAGACCACTTGTAGTT 50
UMD _3'UTR              ACACTAAGAGTTACTGCCACCAAAAGAGGATAATCAGACCACTTGTAGTT 50
```

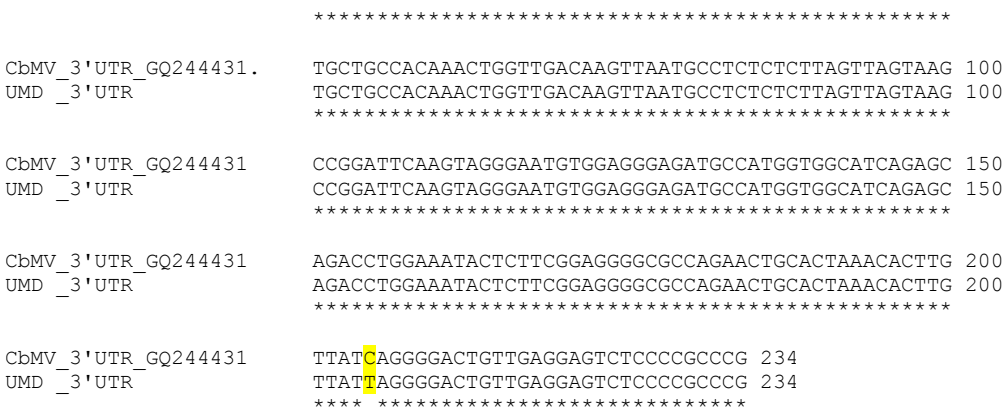


Fig. 3. 4 Sequence alignment of the WT and UMD clone CbMV sequence. The 5' UTR, p28, p87, p8, p9, p37 and 5' UTR alignments are shown. Yellow highlight denotes a sequence alteration. Protein translations are provided for ORFs that contained nt differences. Stop codons are displayed in red.

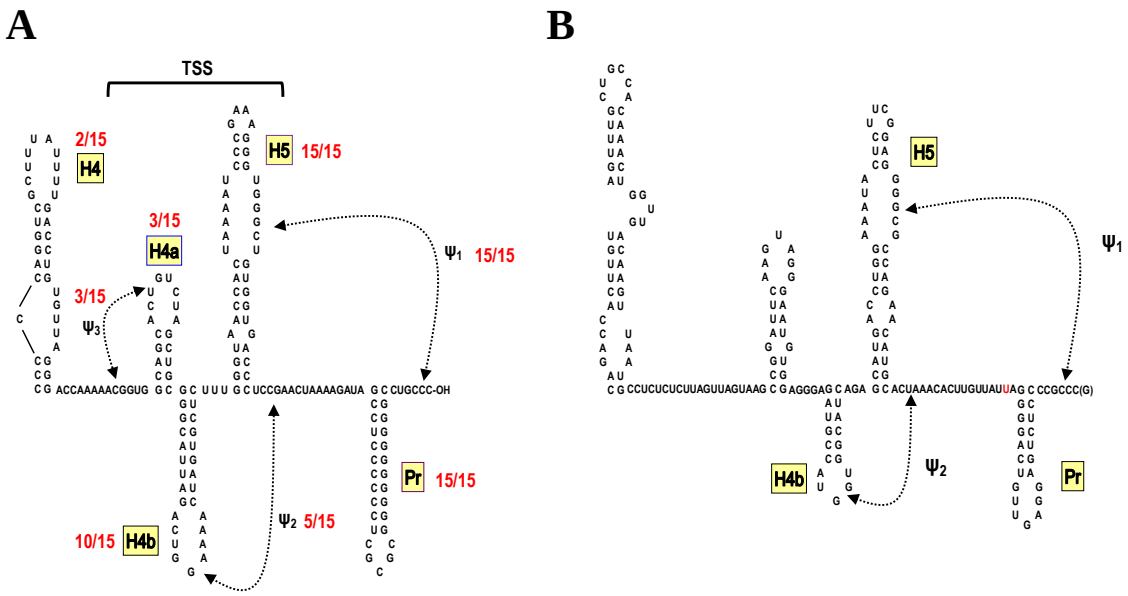


Fig. 3. 5 Conserved elements within the 3' UTRs of Carmoviruses. (A) Structural elements within the 3' UTR of carmovirus type member TCV, with conservation of elements noted in red. Three elements are 100% conserved within the genus: Ψ_1, Pr and H5. The TCV TSS (Ψ_2, Ψ_3, H4a, H4b and H5) is demarcated. Structure solution of the TCV 3' UTR was solved using SHAPE analysis and in-line probing. (B) Predicted structure of elements within the 3' UTR of CbMV using phylogenetic analyses and Mfold[149] computational modeling.

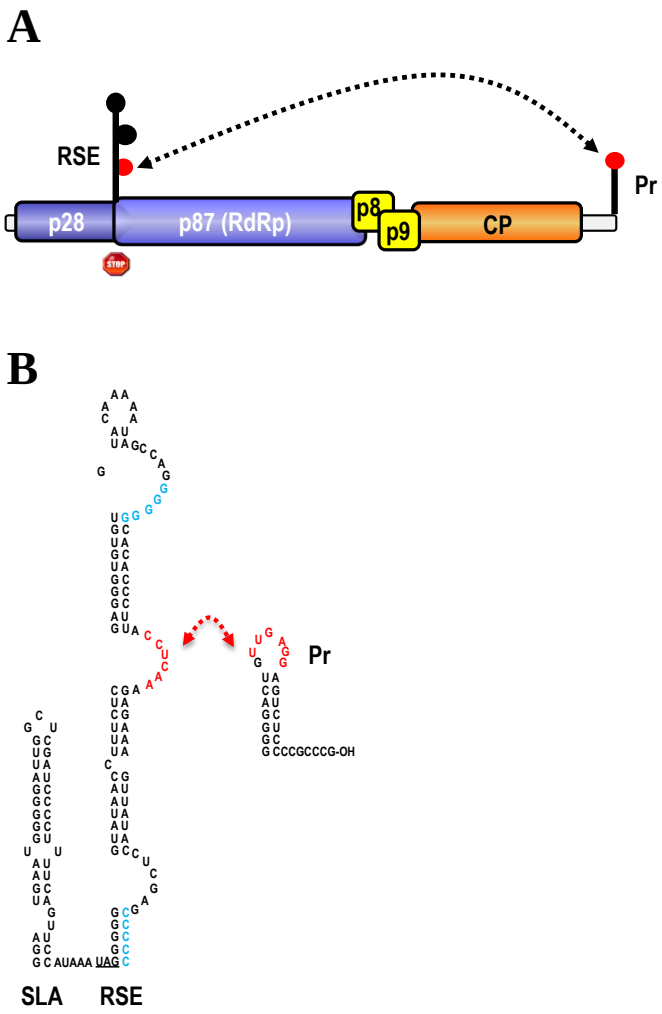


Fig. 3. 6 A long distance kissing loop interaction facilitates readthrough of the p28 stop codon. (A) Schematic of the CbMV genome showing five overlapping open reading frames. The putative long distance kissing-loop interaction between sequences on the RSE and Pr are denoted with a dotted arrow. P28 stop codon shown with red stop sign. (B) Secondary structure and tertiary interactions between the RSE and Pr. The conserved carmovirus pseudoknot residues shown in blue, with long distance interacting residues shown in red. The p28 stop codon is underlined, and the conserved upstream element SLA is shown.

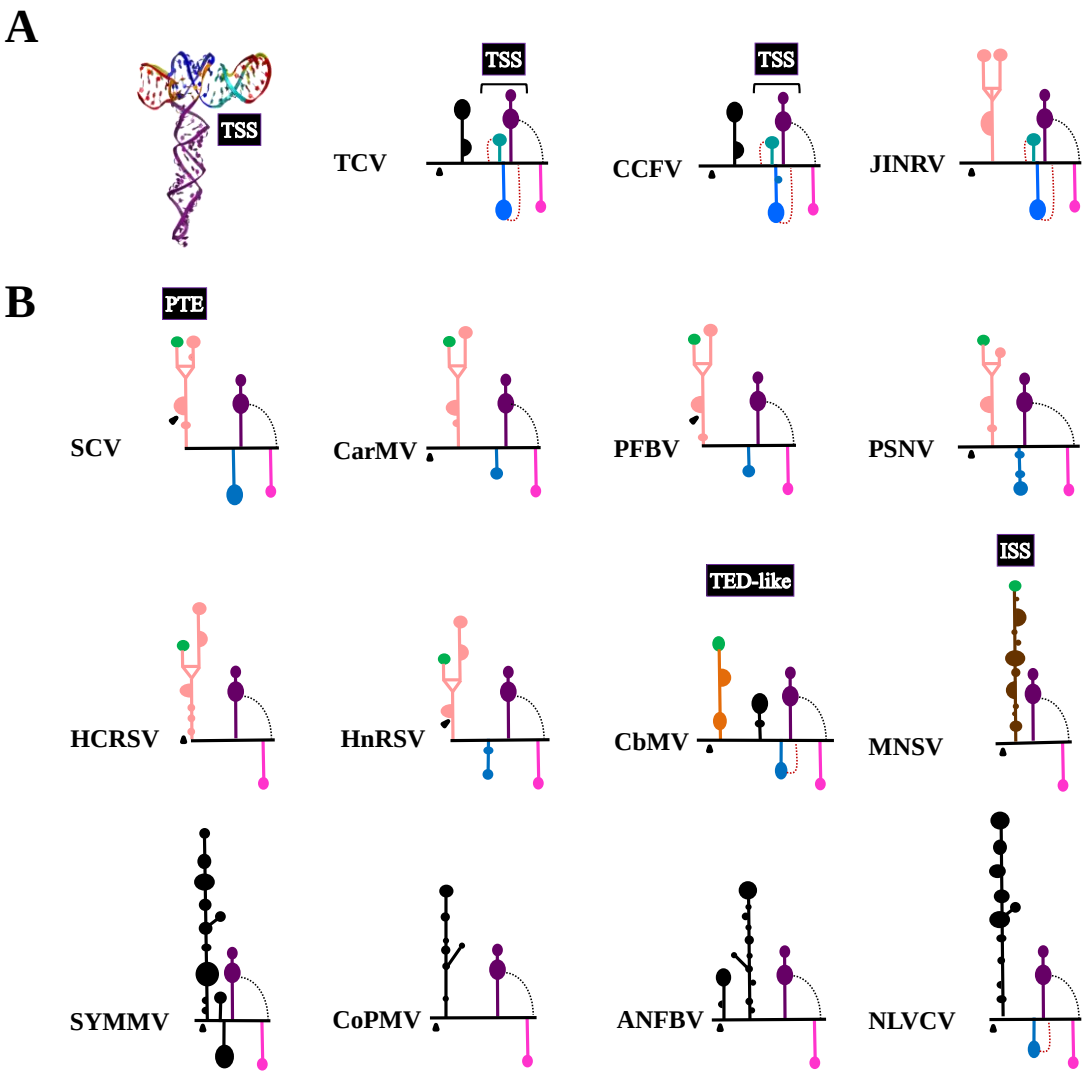
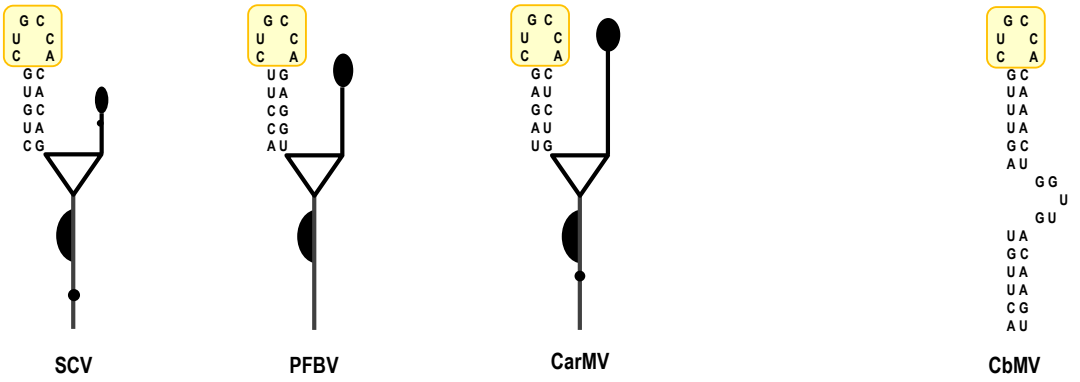

A
TSS
TCV
TSS
CCFV
TSS
JINRV
B
PTE
SCV
CarMV
PFBV
PSNV
TED-like
ISS
HCRSV
HnRSV
CbMV
MNSV
SYMMV
CoPMV
ANFBV
NLVCV

Fig. 3. 7 Carmovirus 3' UTR secondary and tertiary structures. (A) 3-D structural model of the TCV TSS 3' CITE is shown to the left. Right: color coded schematics of the 3 carmoviruses that are known or predicted to contain the five elements that comprise the TSS. Black hairpins represent H4, turquoise H4a, blue H4b, wine H5, pink Pr, red dotted lines Ψ_2 and Ψ_3, and black dotted lines Ψ_1. Salmon structures are PTE-like CITEs. Black triangles denote start site of the 3' UTR. (B) Color coded models of the remaining carmoviruses using TCV elements as a reference. The rust structure is a TED-like 3' CITE and the brown an I Shaped Structure (ISS) 3' CITE. Green terminal loops are known or predicted to engage in long distance RNA:RNA kissing-loop interactions. No data is available for putative 3' CITES for *Soybean yellow mottle mosaic virus* (SYMMV), *Cowpea mottle virus* (CoPMV), *Angelonia flower break virus* (ANFB) or *Nootka lupine vein-clearing virus* (NLVCV). While *Japanese iris necrotic ring virus* (JINRV) may contain elements which could putatively form a TSS and PTE, it is not known if either is functional. Virus names not previously mentioned: PSNV *Pea stem necrosis virus*, HCRSV *Hibiscus chlorotic ringspot virus*, HnRSV *Honeysuckle ringspot virus*. All viral structures are computationally and phylogenetically predicted with the exception of TCV and SCV.

readthrough, as this long distance (~3.5 kb) interaction is required for recoding in all tombusvirids (Fig. 3.6). Downstream of the Pr is the 3' terminal tail, which is present in all carmoviruses and ranges in length from 4 to 9nt, usually terminating with three cytidylates. Two exceptions to this rule exist: *Angelonia flower-break virus* (AnFBV) and CbMV which end in CAA and CCG respectively. CbMV also putatively contains H4b and Ψ_2, but not H4a or Ψ_3 which along with H5 comprise the TSS in TCV (Fig. 3.7). Upstream of H4b is a predicted TED-like 3' CITE (Translation Enhancer Domain first identified in *Satellite tobacco necrosis virus D* (STNV-D) which, in the latter virus, binds eIF4F with high affinity[148]. The terminal loop sequence of this CbMV TED-like hairpin has the same identity as the 3' PTE upstream terminal loop sequence in SCV, which is known to engage in a long distance RNA:RNA interaction with a 5' genomic (gH3) subgenomic RNA2 (sgH1) sequences to facilitate translation of a reporter construct *in vitro* (Fig. 1.14). Two additional carmoviruses, PFBV and *Carnation mottle virus* (CarMV) are also predicted to contain the same 3' sequence on the upstream branch of the PTE (Fig. 3.8). A primer extension inhibition assay (toeprint) was used to identify the location of the CbMV sgRNA2 start site (Fig. 3.9), and putative 5´interacting sequences were subsequently identified in both the genomic and subgenomic RNAs (Fig. 3.8) . A number of positive-sense RNA viruses including Dengue are known to engage in long distance RNA:RNA interactions between the 5´and 3´ends effectively circularizing the genome facilitating the 5´bound RdRp to reposition to the 3´end during replication.

A



B

```
gRNA:
SCV   GGGUAAGCUGCUGAaugCUUAUCUCUGGUUGCUCUCACCaugGGAAUCCUAAGUGAACUGCUCGUUGGAACCACCUUAACAGGUGGGUUAGUUGCAGUGGGGCUGGCAGGCCUG
PFBV  GGGAUACCUUACACUCGGUAUCUGGUUCCAACaugCUACGAUUCGGAUCUCAGUUGGUGGUGGGUUCCACCUUAAUCGGUGGUCUCACUGCUGUCGGGCUGGCAGGCCUG
CbMV  GGAUAAACUUAGCUUUAUCUCGGUGAUUGAGCAAUCaugGGUAUACUACAAUUGGCAAAAGAACUAACCGUAGGAGGAGUGUGUUGCACAACUCUACUAUGCGUUGGGGUGGCAGCCCUG


sgRNA2:
SCV  GGUAACGCUAGAACAAAUCAUCGUACUUAGCUUUUGUGGUCUCUUAAUUUCCUGCGCUUGUAAGGCAGAUACAACGCUGGUUUGGCACUAUUCCGACAACAAUUCAAAGAAUCAACACAUCACAAUCUCCACACCUAaug
PFBV GAUAAACCUCCAGCACAUAUUGAAUCUUACUUUGUUGUCCCUGAUUCUAAGUUCUUUCAUCUUGGCAGAGAGAUAACGCACAAUCACACUCAAACGACAACAGCAAAGCGCAAUACAUACGGAUCUCAACUGGGCAAUAACaug
CbMV AUAAACCAAUCGAUAGCACUUGCACUCUGUGGAUUAAUCCUAAACUGUAUAGGAAAGGCAGAACCCUCAUACAGCUAUUAUUCUACACACGACAGCAGUAAAACCCAAUACAUCAAGAUCAAUACACCUGaug
```

Fig. 3. 8 The CbMV TED-like 3' element shares terminal loop sequence identity with 3 other Carmoviruses and can potentially pair with 5' sequences. (A) Left: three carmoviruses contain identical upstream terminal loop sequences (CUGCCA) on their PTEs (boxed in yellow). Right: the CbMV 3' UTR TED-like element shares loop sequence identity with SCV, PFBV and CarMV. (B) Putative interacting sequences located in the 5' gRNA coding regions and 5' UTRs of sgRNA2 in red, with associated start codons in blue. No potential 5' interacting sequences for CarMV were identified.

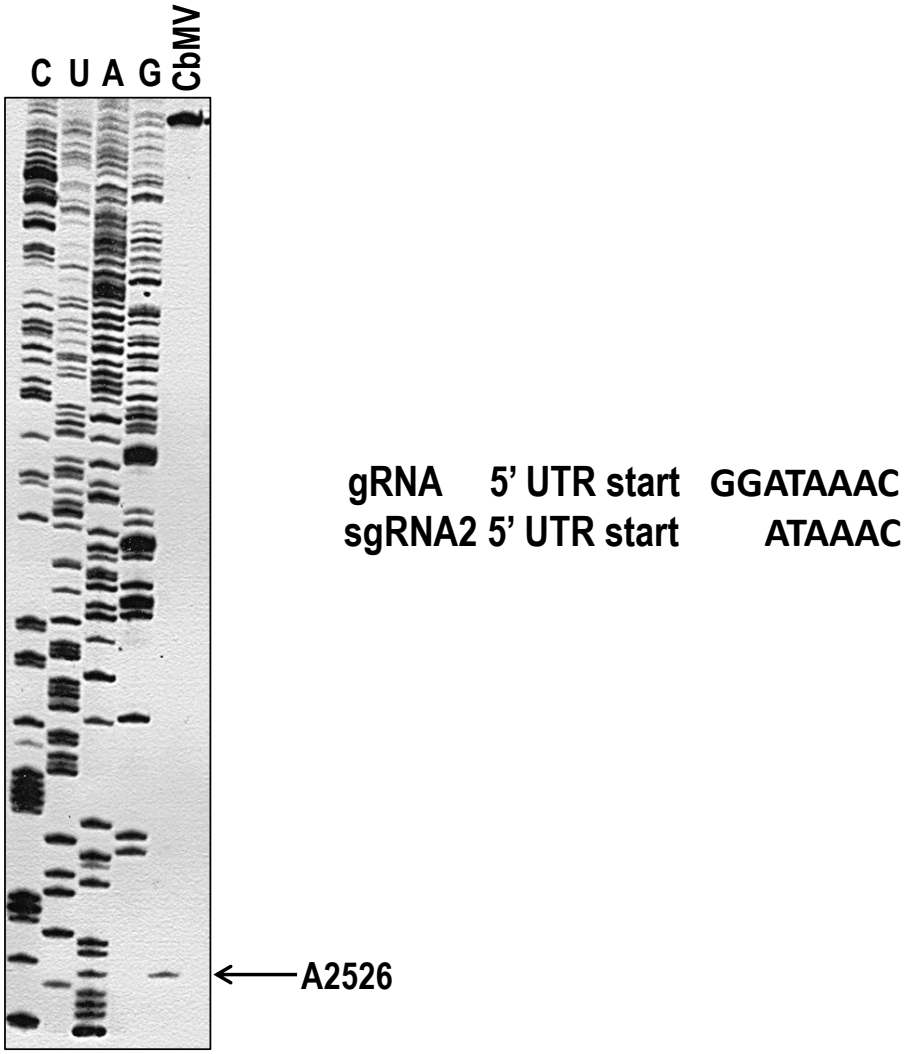


Fig. 3. 9 The CbMV sgRNA2 start site. Toeprint assay noting the nt A2526 start site for the sgRNA2 (CP) 5' UTR. The first four lanes are sequencing ladders. Identity between the genomic and subgenomic start sites is shown to the right. The start site for sgRNA1 is undetermined.

Internally located structural elements

While many RNA structures and sequences that modulate replication are located at the termini of viral genomes, some *cis*-acting RNA elements involved in replication may also be located internally within coding regions[150,151]. The RdRp of *Tomato bushy stunt virus* (TBSV), a Tombusvirus, interacts with its auxiliary replication protein p33 binding as a dimer to an Internal Replication Element (IRE) facilitating membrane targeting and replicase complex assembly[152]. Most members of the *Tombusviridae* family (with the exception of *Panicovirus*, *Machlomovirus* and *Avenavirus* genera) are proposed to contain IRE within the coding region of the RdRp downstream of the readthrough stop codon[153]. An additional exception are the dianthoviruses, whose IRE are positioned within the movement protein coding region of gRNA2[154]. Members of the genus *Carmovirus* are predicted to form IRE with a conserved central core containing a replicase protein binding site (Fig. 3.10). SHAPE structural analysis of TCV and CbMV IRE region validated computational models revealing long, unbranched hairpins with several symmetrical and asymmetrical internal loops (Fig. 3.11). Both elements bear a marked resemblance to the IRE of TBSV suggesting carmoviruses and tombusviruses may share a similar replication strategy.

A CbMV 5´terminal hairpin can functionally replace SCV gH1

CbMV contains 5´and 3´structural elements with identical terminal sequences as the SCV gH3-PTE/sgH1-PTE functional interaction (Fig. 1.14) (Fig. 3.8). Both CbMV and SCV harbor putative 5´terminal hairpins with similar stem sequence but no loop sequence identity. The SCV hairpin however contains an out-of-frame AUG which could possibly

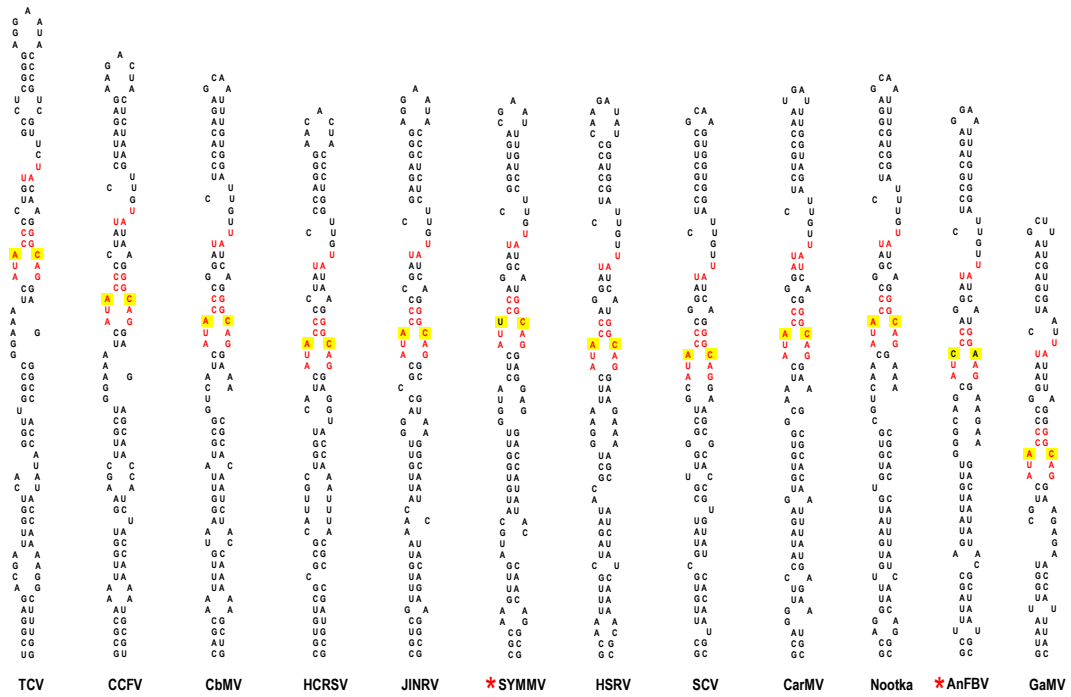


Fig. 3. 10 Phylogenetic comparison of Carmovirus IRE elements. Predicted secondary structure of IRE elements with conserved residues shown in red. The known (TCV)[153] or predicted site for replicase protein binding highlighted in yellow. A red asterisk by SYMMV and AnFBV denote exceptions to the binding residues where UC and CA respectively replace the conserved AC.

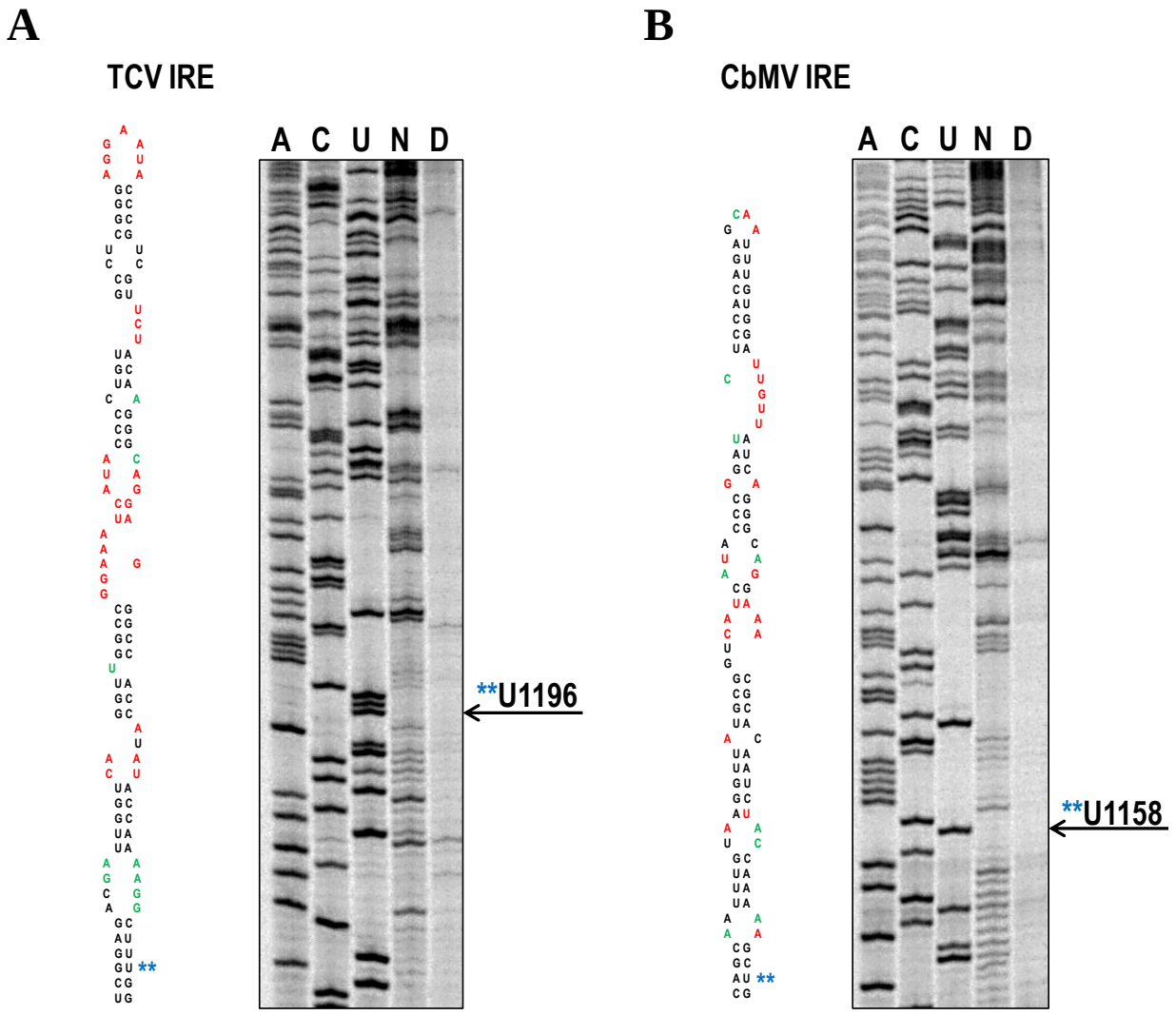


Fig. 3. 11 SHAPE structure probing of the TCV and CbMV IRE regions using full-length gRNA. (A) Secondary structure of the TCV IRE showing the location of NMIA-reactive residues generated by SHAPE. Flexibility color coding as stated previously. The first three lanes are sequencing ladders. N, NMIA treated (modified) and D, DMSO treated (control). 3' end site noted with blue asterisks for orientation. (B) Color coded secondary structure of CbMV IRE region and SHAPE gel.

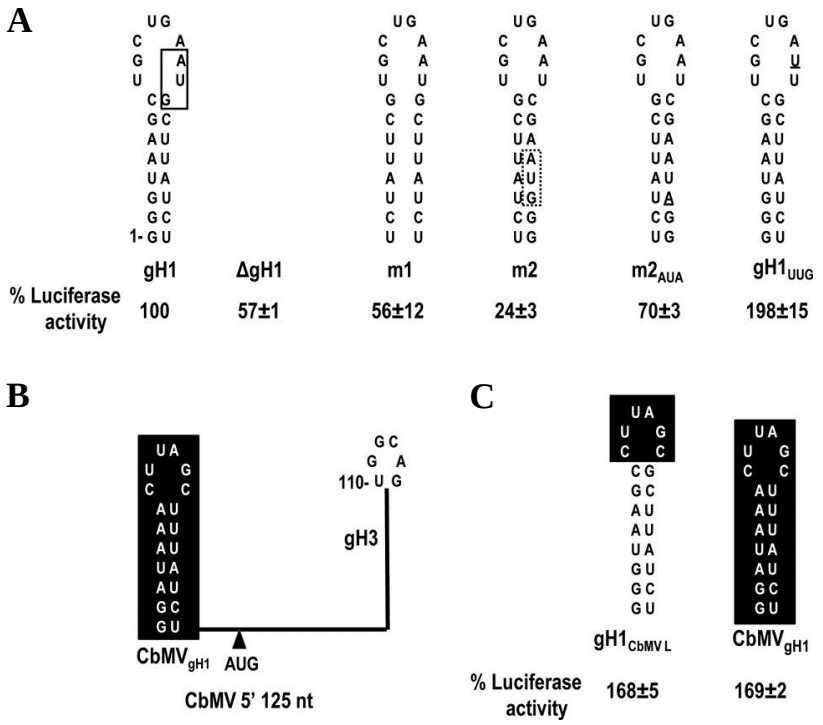


Fig. 3. 12 The CbMV 5´terminal hairpin can functionally replace SCV gH1. (A) A natural AUG located upstream and out-of-frame with the initiator AUG is boxed in a solid line in SCV gH1. An out of frame AUG that was introduced into the stem in construct m2 when the 3´ side sequence was replaced with the 5´ side sequence is boxed with a dotted line. This AUG was altered to AUA in construct $m2_{AUA}$. The naturally occurring AUG was mutated to UUG in construct $gH1_{UUG}$. (B) Schematic representation of the CbMV 5´ 125 nt showing predicted structures gH1 and gH3. CbMV contains a potential TED-like structure with a terminal loop capable of engaging in a kissing-loop interaction with gH3 in a manner very similar to the SCV RNA:RNA interaction. C. Hairpins that were substituted for SCV gH1 in the reporter construct and do not contain an AUG.

compete with the ORF 1 start codon. Deleting SCV gH1 reduced translation of a reporter construct *in vitro* to 57% of WT, as did a construct where pairing of the stem was abrogated (Fig. 3.12). Switching the stem eliminated the first AUG but introduced a second AUG and translation was further reduced to 24% of WT. Mutation of the second AUG to AUA increased translation to 70% of WT suggesting that an upstream out-of-frame AUG is interfering with translation of the reporter construct. A subsequent mutation of the naturally occurring out-of-frame gH1 from AUG to UUG increased translation to 198% of WT suggesting this upstream start codon was indeed negatively impacting translation. Another possibility however, is that alteration of the gH1 loop sequence affected a function that is independent of the start codon. To assess this possibility the apical loop or entire gH1 sequence from SCV was replaced with gH1 sequences from CbMV, which does not contain an AUG. Both the CbMV loop and gH1sequences increased translation to 168% and 169% respectively. These results suggest that CbMV gH1 can functionally replace SCV gH1 in the translation assays, and enhanced translation of constructs containing the CbMV loop and gH1 sequences is due to the absence of an upstream AUG. These results further suggest the translation initiation complex binds to gH1 and scans to the first AUG in good context. The function/mechanism of the 5´terminal hairpin as a translational enhancer is as of yet undetermined.

Discussion

Generation of the CbMV infectious clone has provided important phylogenetic data regarding the genus Carmovirus. From an evolutionary perspective, many elements and

motifs that are conserved at both the genus and family level are predictive of functional (vs. template) RNA.

While the three terminal elements in carmoviruses (Pr, Ψ_1 and H5) are ubiquitous, the genus seems to have diverged upstream of H5 (Fig. 3.7). For TCV, CCFV (and potentially JINRV), five centrally located 3' UTR elements (H5, H4b, H4a, Ψ2 and Ψ3) fold into a TSS; a 3' CITE capable of binding ribosomes and forming a protein bridge with the 5' end of the virus. Seven of the carmoviruses harbor a PTE as their 3' CITEs. Although CbMV has a putative TED-like 3' CITE, it shares a six-base terminal loop nucleotide identity with three other members of the genus and has the potential to form a long distance RNA:RNA interaction with the 5' region of both the gRNA and sgRNA (Fig. 3.8).

Investigation into the central region of the genomes has identified two elements whose structure is conserved, but there is little sequence identity: the RSE and the IRE. Three regions of sequence identity do exist however. All members of the *Tombusviridae* family have a G:C rich S1 on the RSE, and all members of the genus *Carmovirus* have G:C pairing partners in the RSE pseudoknot (Fig. 2.1). SHAPE structure probing of the CbMV and TCV IREs has again revealed a similar structure but dissimilar sequence, the exception being conservation of thirteen core residues within the putative protein binding region (Fig. 3.11).

Very little sequence or structural conservation has been found within the 5' regions of carmoviruses. CbMV, PFBV and SCV do share the 5' gRNA sequence UGGCAG, and 5' sgRNA sequence GGCAG, both of which can pair with the 3' CITE terminal loop sequence CUGCCA (Fig. 3.8). Additionally, the CbMV 5' terminal hairpin can

functionally replace the SCV 5' terminal hairpin as a translational enhancer in reporter constructs in vitro.

Cis-acting regulatory elements in viral RNAs serve important functions often mediating RNA:protein/RNA:RNA interactions. Functional constraints often lead to evolutionary conservation of secondary structure as is evidenced at the genus and family level in this study. The generation of the CbMV infectious clone has provided additional layers of data for the interrogation of regulatory mechanisms in these small virus model systems.

Note: Materials and methods can be found in Chapter 5.

List of oligonucleotides used for the construction and sequencing of the CbMV infectious clone are located in Appendix A

Chapter 4: Discussion and Perspectives

With the initial deciphering of the genetic code arose the assumption that the decoding of nucleic acids was universal and unchangeable (Crick Frozen Accident Hypothesis). Ironically, exceptions to the standard 'rules' were observed shortly thereafter and new discoveries continue to be made with alterations predicted to exist in all organisms. There are now more than 20 variant genetic codes (mitochondrial for example) and two non-universal proteinogenic amino acids: selenocystein and pyrrolysine. mRNAs have evolved to encode linear or structural signals which alter the meaning of specific codons under certain circumstances, and this is termed 'recoding'. Two well-studied recoding events utilized to extend the C-terminus of a polypeptide are -1PRF and PRT.

Eukaryotic translation has evolved machinery with sophisticated checks and balances to ensure the highest rates of fidelity during protein synthesis. Quality control errors can lead to the production of truncated, aberrant or misfolded proteins. The precision of this system is regulated at RNA and protein levels, and is fine tuned for the most economical energy output. Thus, to refer to PRT as 'leaky scanning' is misleading. 'Leaky scanning' is perhaps better exemplified by my 5' SCV work (Fig. 3.12) where the ribosome, in scanning from the 5' terminus, sometimes starts at the upstream out-of-frame AUG. This would most likely occur only during the pioneer round of translation (quality control), and not in the course of steady-state cycles[155]. It is interesting from an evolutionary perspective, that the AUG still remains in gH1, suggesting that it may serve some lifecycle function in SCV.

For the 40+ years of investigating recoding, much more is understood about the mechanisms of -1PRF than for PRT. Dissecting what is known about canonical translation provides a good foundation for understanding the mechanics of PRT. A brief summary:

- Recognition of cognate codons by natural cellular tRNAs requires steric complementarity of the codon-anti-codon helix as well as discernment of base pair geometry by 18S adenosines on h44.
- Recognition of stop codons (for which there are no cognate tRNAs) involves eRF1 N-terminal domain motifs (NIKS, YxCxxxF and GTS) whose residues specifically recognize uridylates in the first position, and purines in the second and third position. The nature of the amino acid side chains recognize purines in U<u>AG</u>, U<u>GA</u> and U<u>AA</u>, while specifically discriminating against the tryptophan sense codon purines in U<u>GG</u>.

Thus, the intricate specifications involved in sense and non-sense codon recognition preclude stop codon "leaks" except when the translation machinery encounters programmed signals embedded within the mRNA that specifically stimulate readthrough. These signals include but are not limited to:

- The stop codon identity and surrounding nucleotide context.
- Downstream RSE
- Long distance RNA:RNA interactions

Presumably nucleotide context contributes to readthrough by 18s rRNA interference, although this has not been definitively proven, nor has an explanation as to how the percentage of readthrough is regulated with this mechanism. These intrinsic signals are

known to place the stop codon in a poor termination context prompting eIF3 to disable decoding of the third stop codon (wobble) position by interfering with eRF1. Less understood are the mechanisms of action for RSE and long distance interactions. Given that all three stop codons assume the same conformation (Fig. 1.2A, the first residue lies orthogonal to the second and third which are stacked) it is conceivable that these secondary structures and tertiary interactions work to distort the stop codon geometry precluding recognition by eRF1 motifs. This may involve interaction with the helicase as it seeks to unwind the mRNA.

The stoichiometric protein ratios are regulated through upstream RNA structures and physiological conditions within the cell. For TCV, regulation of p28 expression is achieved through adoption of the basal structure (SLA), with readthrough regulated through the active (RSE) structure (Fig. 2.12). Mutations of SLA increased readthrough but decreased viral accumulation *in vivo*. As alteration of RdRp levels are known to be deleterious in several positive-sense RNA viruses, this would suggest a role for SLA in attenuation/regulation of readthrough. Uniquely, the TCV RSE engages in two tertiary interactions: the internal H-type pseudoknot and the long distance RNA:RNA interaction with a 3' terminal hairpin loop. Both of these interactions are required for efficient readthrough levels. It is possible they serve to enhance the structural plasticity of the recoding region and, in concert with SLA, stabilize recoding ratios. The question remains whether activation of these tertiary interactions is pH dependent. In other words, is there a molecular switch involved between the basal and active conformations, or is it a matter of populations based on physiological conditions?

A clue to the mechanism may lie in the pairing of the RSE lower stem. All tombusvirids have a cnformation with guanylates on the upstream side and cytidylates on the downstream side of the RSE S1(Fig. 2.1). SHAPE structure probing combined with a mutational analysis revealed the guanylates are paired upstream on SLA. The cytidylates are also not flexible, which may indicate base stacking, however, it is more likely they are paired elsewhere.

Interestingly identification of the putative p6 protein nested within the coat protein of CbMV prompted a search for a similar ORF in TCV. A potential 22 kDa candidate has tentatively been identified and is currently being investigated. The start codon begins at base 3164 (TCV CP 2743-3799), with the stop codon being 3' co-terminal with the CP. A protein band of this approximate size can be seen on both WGE gels, and Western blots. As stated in Chapter 3, HCRSV has two functional proteins (25 and 27 kDa) nested within the CP. Additionally, published WGE gels from related tombusviruses (data not shown) show numerous unexplained protein bands, which arguably may be an artifact of WGE. This serves to reinforce the importance of phylogenetic analyses in the investigation of functional RNA elements and conserved proteins.

Chapter 5: Materials and Methods

Generation of Plasmid Constructs in TCV

Mutant clones were generated using PCR-based site-directed mutagenesis from plasmid pTCV66, which contains the full-length TCV genome downstream of a T7 RNA polymerase promoter. Overlapping oligonucleotides (Integrated DNA Technologies) were used to introduce desired mutations using Phusion high-fidelity polymerase (New England Biolabs) according to a one-step mutagenesis protocol [139]. PCR products were subjected to DpnI digestion for 2 hours at 37°C prior to transformation into DH5α competent cells. Mutations were confirmed by sequencing in the region (Eurofins Genomics).

In vitro *translation*

SmaI-digested TCV plasmids were subjected to *in vitro* transcription for 2 h at 37°C using T7 RNA polymerase. Uncapped RNA transcripts (1 pmole) were translated in 10 µl of wheat germ extract (WGE, Promega) according to the manufacturer's instructions supplemented with 100 mM potassium acetate and ^{35}S-methionine. The reaction was incubated at 25°C for 2 h and resolved on a 10% SDS-PAGE gel. The dried gel was exposed to a phosphorimager screen and scanned by a FLA-5100 fluorescent image analyzer (Fujifilm). Band intensity was quantified using Quantity One software (Bio-Rad). Experiments were performed in triplicate using independently transcribed RNAs.

Protoplast transfection and RNA gel blots

Protoplasts generated from seed-derived *Arabidopsis thaliana* callus tissue (ecotype Col-0) were transfected with *in vitro* transcribed full-length TCV gRNA as described previously [140]. Briefly, 20 µg of uncapped gRNA was transfected into 5 x 10^6 cells, and incubated in the dark for 40 h. Total RNA was extracted, subjected to electrophoresis,

transferred to a nitrocellulose membrane and probed using [γ-^{32}P] ATP-labeled oligonucleotides complementary to positions 3931-3953, 3869-3883, and 4035-4054 in the 3' UTR. The membrane was exposed to a phosphorimager screen as described above. Experiments were performed in triplicate using independently transcribed RNAs.

SHAPE structure probing

Structure probing was performed using SHAPE as previously described[115]. Briefly, for *in vitro* SHAPE, 6 pmoles of *in vitro* transcribed, full-length TCV gRNA was denatured for 5 minutes at 65°C, snap cooled on ice and then incubated in folding buffer at 37° for 20 min. The folded RNA was treated with either N-methylisatoic anhydride (NMIA, 15 mM final concentration) for base modification, or an equal volume of dimethyl sulfoxide (DMSO) as a negative control. The reactions were incubated at 37°C for 35 minutes, ethanol precipitated and resuspended in 8 µl 0.5x Tris-EDTA (TE) buffer. Primer extension was carried out using [γ-^{32}P] ATP-labeled oligonucleotides and SuperScript III reverse transcriptase (Invitrogen) as previously described[141]. Primers complimentary to TCV positions 940-964, 875-894, 839-862 and 995-1018 were used for structure probing of the RSE, SLA and upstream and downstream regions, respectively. Reaction products and dideoxy sequencing ladders (Roche) were resolved on an 8% denaturing polyacrylamide gel, exposed overnight to a phosphorimager screen and imaged using the FLA-5100 fluorescent analyzer (Fujifilm). For *in vivo* SHAPE, protoplasts prepared from *Arabidopsis thaliana* seed callus cultures (5 x 10^6 cells) were transfected with 20 µg of uncapped full-length WT TCV gRNA or an equal amount of a non-replicating (GDD→GAA) polymerase active site mutant. After 40 h incubation at room temp in the

dark, 1-methyl-7-nitroisatoic anhydride (1M7; final concentration of 5 mM) was added to the protoplasts for base modification or an equal volume of DMSO was added as a negative control, and cultures gently shaken for 5 min at room temperature. IM7 was used in place of NMIA as NMIA was reported to be inefficient at modifying RNA in a cellular environment[142,143]. Unmodified WT TCV was used to generate dideoxy sequencing ladders. Cells were collected by centrifugation, total RNA was extracted and the RNA resuspended in H_2O. Primer extension, structure probing, resolution and visualization were performed as described above.

cDNA reaction (RT-PCR) for the CbMV clone

5 µg of total RNA extracted from symptomatic N. benthamiana leaf tissue was used for the cDNA reaction performed with SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Briefly, 2.5 pmol RNA, 2.5 mM CbMV R oligo (Appendix A), 1.5 mM dNTP in a 13.0 µl reaction was heated to 65° for 5 minutes. 4.0 µl 5x First strand buffer,1.0 µl 5mM DTT, 1.0 µl RNase inhibitor (NEB Biolabs) and 1.0 µl SuperScript III was added for a 20.0 µl reaction. The reaction was incubated in a thermocycler (LabNet Multigene) for: 37° 12 minutes; 42° 12 minutes; 47° 12 minutes; 51° 12 minutes; 55° 12 minutes; 85° 1 minute.

Amplification of the CbMV cDNA by PCR

A PCR reaction using with Platinum Taq (Invitrogen) was used according to the manufacturer's protocol. Briefly, 8.0 µl CbMV cDNA (above), 20 µl 10x PCR buffer (-$MgCl_2$), 0.2 mM dNTP, 1.5 mM $MgCl_2$, 0.2 mM CbMV F (Appendix A), 0.2 mM CbMV R (Appendix A), 0.8 µl Platinum Taq and H_2O in a volume of 200 µl were added to a

PCR tube. Thermocycler (LabNet Multigene) conditions were set at 94° for 2 minutes, followed by 25 cycles of 94° 30 seconds, 50° 30 seconds, 72° 4 minutes with a final extension of 72° for 10 minutes.

Restriction enzyme digestion.

The PCR product and pUC19 vector were digested with KpnI-HF and EcoRV-HF (NEB) according to the manufacturer's protocol, and gel purified using QUAEX II gel purification kit (QIAGEN).

Ligation

Digested insert was treated with Klenow (DNA Polymerase Large Fragment NEB) according to the manufacturer's protocol, and heat inactivated for 20 minutes at 75°. The insert was phosphorylated with T4 polynucleotide kinase (NEB), and the vector dephosphorylated with Antarctic phosphatase (NEB) according to the manufacturer's protocol. Insert and vector were ligated using a 3:1 molar ratio respectively using the Quick Ligation Kit (NEB) according to the manufacturer's protocol.

Transformation

5.0 µl ligation reaction was added to 50 µl DH5α competent cells and incubated on ice for 30 minutes. The cells were heat shocked for 30 seconds and placed on ice for 2 minutes. 1 ml Luria Broth (LB) was added, and put on the shaker (225 RPM) at 37° for 1 hour. Transformed cells were subsequently plated on LB plates with Ampicillin, and incubated at 37° overnight.

Colony PCR

Candidate colonies were screened using GoTaq DNA polymerase kit according to the manufacturer's protocol. 20.0 µl was aliquoted per PCR tube, and inoculated with small amount of the colony. PCR conditions: 94° for 5 minutes followed by 30 cycles of 94° for 30 seconds, 55° for 30 seconds, 72° for 2 minutes, with a final extension of 72° for 10 minutes. PCR product was run on a 1% agarose gel.

Appendix A

Oligonucleotides used in the construction of the CbMV infectious clone

Name	Position	Sequence	Polarity
CbMV F3	1-23	CGGGGTACCTAATACGACTCACTA TAGGATAAACTTAGCTTTATCTCGG	+
CbMV R3	3901-3919	GAGGAGTCTCCCCGCCCGGGGATA TCAGGC	-

Oligonucleotides used in the sequencing of the CbMV infectious clone

CbMV 678F	pUC19	GCTCGATCCCCTTTTCAGTT	+
CbMV 973R	254-973	AGGATTACTGGAGCGGGTTT	-
CbMV 1352F	652-671	TAAAATTTGGGGTGGCAGGA	+
CbMV 1677R	956-977	TGCAATTACCCAAAGCTGTATT	-
CbMV 2026F	1326-1345	TGTGCGGAAAGATACCAATC	+
CbMV 2302R	1585-1602	ATGTTCGCGCTCCATGAT	-
CbMV 2709F	2009-2028	TCCATGGGCGATCAAGTTTA	+
CbMV 2966R	2278-2296	GTCAATACAGCCCCTTCCG	-
CbMV 3335F	2635-2657	TTTTAGTTGACTACACCGTCGTG	+
CbMV 3569R	2849-2869	CAAGGCACACATCACATCATC	-
CbMV 3576F	2876-2902	CAGTAAGACGGTTATAGCTGTGGTTAC	+

Underlined: restriction sites. Red lettering: T7 promoter

Appendix B

Copyright release was obtained to reuse the following figures:

Elsevier grants blanket permission to reuse figures in a thesis or dissertation. The following figures in Chapter 1 are from Elesvier publications: Fig. 1.1; Fig. 1.3; Fig. 1.4; Fib. 1.5; Fig. 1.7; Fig. 1.13.

https://www.elsevier.com/about/company-information/policies/copyright/permissions

Permission Guidelines

For further guidelines about obtaining permission, please review our Frequently Asked Questions below:

When is permission required? +

When is permission not required? +
From whom do I need permission? +
How do I obtain permission to use photographs or illustrations? +
Do I need to obtain permission to use material posted on a website? +
What rights does Elsevier require when requesting permission? +
How do I obtain permission from another publisher? +
What is Rightslink? +
What should I do if I am not able to locate the copyright owner? +
What is Elsevier's policy on using patient photographs? +
Can I obtain permission from a Reproduction Rights Organization (RRO)? +
Is Elsevier an STM signatory publisher? +
Do I need to request permission to re-use work from another STM publisher? +
Do I need to request permission to text mine Elsevier content? +
Can I post my article on ResearchGate without violating copyright? +
Can I post on ArXiv? +
Can I include/use my article in my thesis/dissertation? +

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.

Which uses of a work does Elsevier view as a form of 'prior publication'? +

Fig. 1.2 Model for Eukaryotic Stop Codon Recognition in the Decoding Center.

Fig. 1.6 Potential Base Pairing Between the TMV readthrough motif and the 18s

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

May 02, 2016

This Agreement between Micki Kuhlmann ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3860910269642
License date	May 02, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	EMBO Reports
Licensed Content Title	Impact of the six nucleotides downstream of the stop codon on translation termination
Licensed Content Author	Olivier Namy,Isabelle Hatin,Jean-Pierre Rousset
Licensed Content Date	Sep 1, 2001
Pages	7
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 3
Will you be translating?	No
Title of your thesis / dissertation	Alternative Conformations moculate translational recoding in a positive-sense RNA virus
Expected completion date	May 2016
Expected size (number of pages)	200
Requestor Location	Micki Kuhlmann 702 warrington ave SE WASHINGTON, DC 20003 United States

Fig. 1.8 Structure of the HCV IRES subdomain IIa

Fig. 1.9 Conformational Switch in the 3' Terminal Region of AMV

Fig. 1.10 Alternative Conformations Drive the Conformational Switch

Content License

The following policy applies to all of PLOS journals, unless otherwise noted.

PLOS applies the Creative Commons Attribution (CC BY) license to works we publish. This license was developed to facilitate open access – namely, free immediate access to, and unrestricted reuse of, original works of all types.

Under this license, authors agree to make articles legally available for reuse, without permission or fees, for virtually any purpose. Anyone may copy, distribute or reuse these articles, as long as the author and original source are properly cited.

Using PLOS Content

No permission is required from the authors or the publishers to reuse or repurpose PLOS content provided the original article is cited. In most cases, appropriate attribution can be provided by simply citing the original article.

Example citation:

Kaltenbach LS et al. (2007) Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. *PLOS Genet* 3(5): e82. doi:10.1371/journal.pgen.0030082.

If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared.

For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

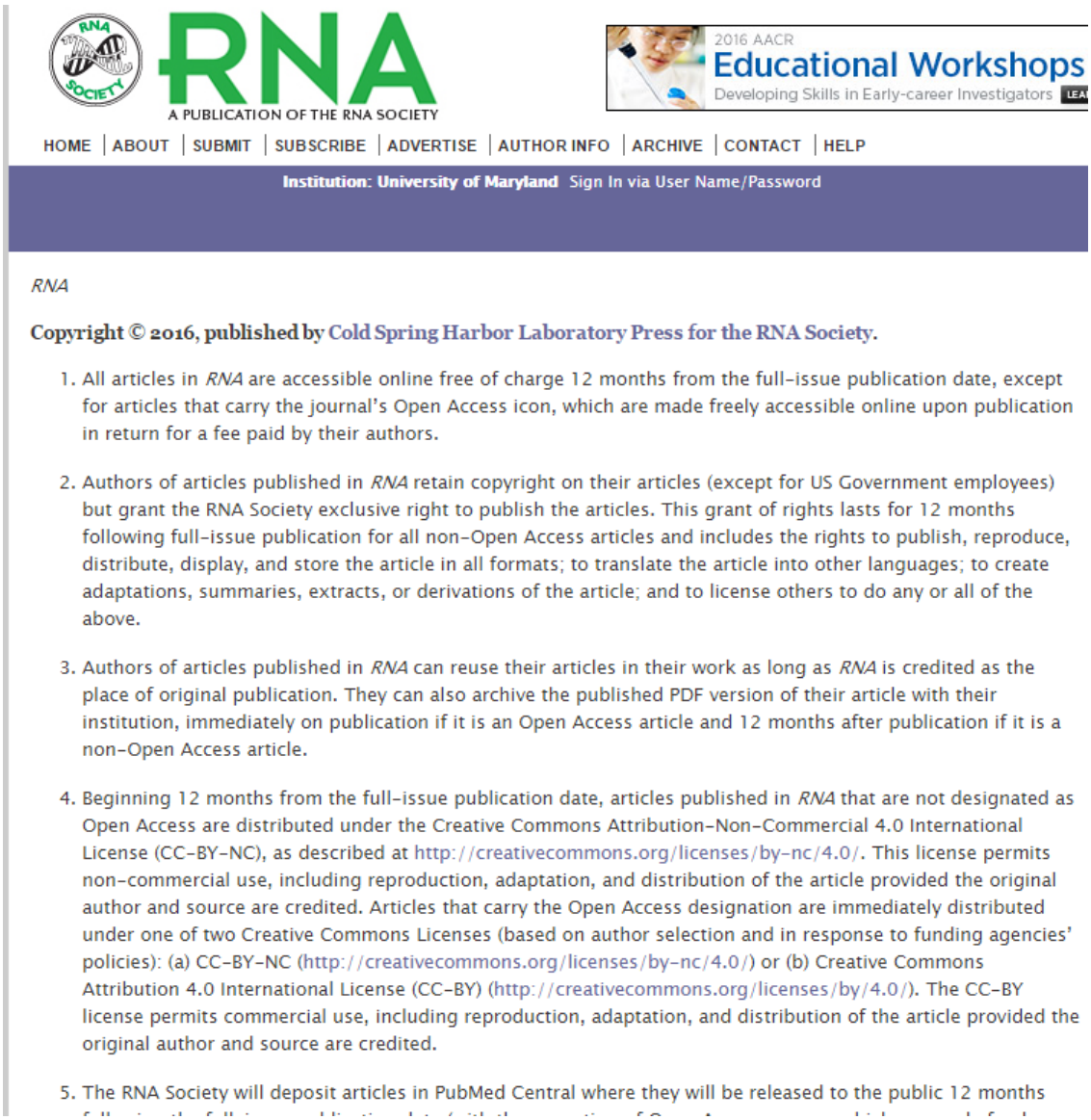
Figures, Tables, and Images

Figures, tables, and images are published under the Creative Commons Attribution (CC

BY) license.

Figure 1.11 TCV Genome and 3' UTR Interactions

Fig. 1.12 RdRp interaction with the TCV 3' End Causes Widespread . . . .



RNA

A PUBLICATION OF THE RNA SOCIETY

RNA

References

1 Hepburn, A. G. & Ingle, J. Molecular integrity of plant ribosomal ribonucleic Acid. Plant Physiol 57, 410-414 (1976).

2 Cornish, P. V., Hennig, M. & Giedroc, D. P. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. Proc Natl Acad Sci U S A 102, 12694-12699, (2005).

3 Wilson, D. N. & Doudna Cate, J. H. The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol 4, (2012).

4 Budkevich, T. V. et al. Regulation of the mammalian elongation cycle by subunit rolling: a eukaryotic-specific ribosome rearrangement. Cell 158, 121-131, (2014).

5 Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111, 721-732 (2002).

6 Yarus, M., Valle, M. & Frank, J. A twisted tRNA intermediate sets the threshold for decoding. RNA 9, 384-385 (2003).

7 Frank, J. et al. The role of tRNA as a molecular spring in decoding, accommodation, and peptidyl transfer. FEBS Lett 579, 959-962, (2005).

8 Thompson, R. C. & Stone, P. J. Proofreading of the codon-anticodon interaction on ribosomes. Proc Natl Acad Sci U S A 74, 198-202 (1977).

9 Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. Nature 484, 256-259, (2012).

10 Kurland, C. G. & Ehrenberg, M. Optimization of translation accuracy. Prog Nucleic Acid Res Mol Biol 31, 191-219 (1984).

11 Parker, J. Errors and alternatives in reading the universal genetic code. Microbiol Rev 53, 273-298 (1989).

12 Jørgensen, F., Adamski, F. M., Tate, W. P. & Kurland, C. G. Release factor-dependent false stops are infrequent in Escherichia coli. J Mol Biol 230, 41-50, (1993).

13 Kisselev, L., Ehrenberg, M. & Frolova, L. Termination of translation: interplay of mRNA, rRNAs and release factors? EMBO J 22, 175-182, (2003).

14 Taylor, D. et al. Cryo-EM structure of the mammalian eukaryotic release factor eRF1-eRF3-associated termination complex. Proc Natl Acad Sci U S A 109, 18413-18418, (2012).

15 Song, H. et al. The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. Cell 100, 311-321, (2000).

16 Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G. & Stansfield, I. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. RNA 6, 1236-1247 (2000).

17 Frolova, L. Y. et al. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. RNA 5, 1014-1020 (1999).

18 Ito, K., Ebihara, K. & Nakamura, Y. The stretch of C-terminal acidic amino acids of translational release factor eRF1 is a primary binding site for eRF3 of fission yeast. RNA 4, 958-972 (1998).

19 Seit-Nebi, A., Frolova, L. & Kisselev, L. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. EMBO Rep 3, 881-886, (2002).

20 Ito, K., Ebihara, K., Uno, M. & Nakamura, Y. Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. Proc Natl Acad Sci U S A 93, 5443-5448 (1996).

21 Cheng, Z. et al. Structural insights into eRF3 and stop codon recognition by eRF1. Genes Dev 23, 1106-1118, (2009).

22 Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. Nature 524, 493-496, (2015).

23 Bulygin, K. N. et al. Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome. RNA 16, 1902-1914, (2010).

24 Chavatte, L., Seit-Nebi, A., Dubovaya, V. & Favre, A. The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. EMBO J 21, 5302-5311 (2002).

25 Wong, L. E., Li, Y., Pillay, S., Frolova, L. & Pervushin, K. Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. Nucleic Acids Res 40, 5751-5765, (2012).

26 Frolova, L., Seit-Nebi, A. & Kisselev, L. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. RNA 8, 129-136 (2002).

27 Kryuchkova, P. et al. Two-step model of stop codon recognition by eukaryotic release factor eRF1. Nucleic Acids Res 41, 4573-4586, (2013).

28 Feng, T. et al. Optimal translational termination requires C4 lysyl hydroxylation of eRF1. Mol Cell 53, 645-654, (2014).

29 Kolosov, P. et al. Invariant amino acids essential for decoding function of polypeptide release factor eRF1. Nucleic Acids Res 33, 6418-6425, (2005).

30 Brown, C. M., Stockwell, P. A., Trotman, C. N. & Tate, W. P. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. Nucleic Acids Res 18, 6339-6345 (1990).

31 Matheisl, S., Berninghausen, O., Becker, T. & Beckmann, R. Structure of a human translation termination complex. Nucleic Acids Res 43, 8615-8626, (2015).

32 Crick, F. H. The origin of the genetic code. J Mol Biol 38, 367-379 (1968).

33 Weiner, A. M. & Weber, K. Natural read-through at the UGA termination signal of Q-beta coat protein cistron. Nat New Biol 234, 206-209 (1971).

34 Knight, R. D., Freeland, S. J. & Landweber, L. F. Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2, 49-58, (2001).

35 Dinman, J. D., Icho, T. & Wickner, R. B. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. Proc Natl Acad Sci U S A 88, 174-178 (1991).

36 Wills, N. M., Gesteland, R. F. & Atkins, J. F. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. Proc Natl Acad Sci U S A 88, 6991-6995 (1991).

37 Jacks, T. & Varmus, H. E. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230, 1237-1242 (1985).

38 Harger, J. W., Meskauskas, A. & Dinman, J. D. An "integrated model" of programmed ribosomal frameshifting. Trends Biochem Sci 27, 448-454 (2002).

39 Bekaert, M. & Rousset, J. P. An extended signal involved in eukaryotic -1 frameshifting operates through modification of the E site tRNA. Mol Cell 17, 61-68, (2005).

40 Léger, M., Dulude, D., Steinberg, S. V. & Brakier-Gingras, L. The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. Nucleic Acids Res 35, 5581-5592, (2007).

41 Lopinski, J. D., Dinman, J. D. & Bruenn, J. A. Kinetics of ribosomal pausing during programmed -1 translational frameshifting. Mol Cell Biol 20, 1095-1103 (2000).

42 Dinman, J. D. Control of gene expression by translational recoding. Adv Protein Chem Struct Biol 86, 129-149, (2012).

43 Brierley, I. Ribosomal frameshifting viral RNAs. J Gen Virol 76 (Pt 8), 1885-1892, (1995).

44 Dreher, T. W. & Miller, W. A. Translational control in positive strand RNA plant viruses. Virology 344, 185-197, (2006).

45 Barry, J. K. & Miller, W. A. 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, (2002).

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

47 Baranov, P. V., Atkins, J. F. & Yordanova, M. M. Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. Nat Rev Genet 16, 517-529, (2015).

48 Plant, E. P., Wang, P., Jacobs, J. L. & Dinman, J. D. A programmed -1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. Nucleic Acids Res 32, 784-790, (2004).

49 Belew, A. T. et al. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. Nature 512, 265-269, (2014).

50 Dulude, D., Baril, M. & Brakier-Gingras, L. Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. Nucleic Acids Res 30, 5094-5102 (2002).

51 Kollmus, H., Hentze, M. W. & Hauser, H. Regulated ribosomal frameshifting by an RNA-protein interaction. RNA 2, 316-323 (1996).

52 Yu, C. H., Noteborn, M. H., Pleij, C. W. & Olsthoorn, R. C. Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. Nucleic Acids Res 39, 8952-8959, (2011).

53 Rietveld, K., Van Poelgeest, R., Pleij, C. W., Van Boom, J. H. & Bosch, L. The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. Nucleic Acids Res 10, 1929-1946 (1982).

54 Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55, 447-458 (1988).

55 McPheeters, D. S., Stormo, G. D. & Gold, L. Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J Mol Biol 201, 517-535 (1988).

56 Powers, T. & Noller, H. F. A functional pseudoknot in 16S ribosomal RNA. EMBO J 10, 2203-2214 (1991).

57 Haas, E. S., Brown, J. W., Pitulle, C. & Pace, N. R. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. Proc Natl Acad Sci U S A 91, 2527-2531 (1994).

58 Gilley, D. & Blackburn, E. H. The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. Proc Natl Acad Sci U S A 96, 6621-6625 (1999).

59 Shen, L. X. & Tinoco, I. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. J Mol Biol 247, 963-978 (1995).

60 Egli, M., Minasov, G., Su, L. & Rich, A. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. Proc Natl Acad Sci U S A 99, 4302-4307, (2002).

61 Pleij, C. W., Rietveld, K. & Bosch, L. A new principle of RNA folding based on pseudoknotting. Nucleic Acids Res 13, 1717-1731 (1985).

62 Walter, A. E. & Turner, D. H. Sequence dependence of stability for coaxial stacking of RNA helixes with Watson-Crick base paired interfaces. Biochemistry 33, 12715-12719 (1994).

63 Giedroc, D. P., Theimer, C. A. & Nixon, P. L. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. J Mol Biol 298, 167-185, (2000).

64 Plant, E. P. et al. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. PLoS Biol 3, e172, (2005).

65 Ishimaru, D. et al. RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. Nucleic Acids Res 41, 2594-2608, (2013).

66 Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. Virus Res 139, 193-208, (2009).

67 Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature 441, 244-247, (2006).

68 Ritchie, D. B., Foster, D. A. & Woodside, M. T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. Proc Natl Acad Sci U S A 109, 16167-16172, (2012).

69 Low, J. T. et al. Structure and dynamics of the HIV-1 frameshift element RNA. Biochemistry 53, 4282-4291, (2014).

70 Nixon, P. L. & Giedroc, D. P. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. J Mol Biol 296, 659-671, (2000).

71 Chamorro, M., Parkin, N. & Varmus, H. E. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc Natl Acad Sci U S A 89, 713-717 (1992).

72 Jacks, T., Townsley, K., Varmus, H. E. & Majors, J. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc Natl Acad Sci U S A 84, 4298-4302 (1987).

73 Chen, X. et al. Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. EMBO J 14, 842-852 (1995).

74 Chen, X. et al. A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. J Mol Biol 260, 479-483 (1996).

75 Napthine, S., Liphardt, J., Bloys, A., Routledge, S. & Brierley, I. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. J Mol Biol 288, 305-320, (1999).

76 Liphardt, J., Napthine, S., Kontos, H. & Brierley, I. Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. J Mol Biol 288, 321-335, (1999).

77 Beier, H. & Grimm, M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res 29, 4767-4782 (2001).

78 Beznosková, P. et al. Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. PLoS Genet 9, e1003962, (2013).

79 Beznosková, P., Wagner, S., Jansen, M. E., von der Haar, T. & Valášek, L. S. Translation initiation factor eIF3 promotes programmed stop codon readthrough. Nucleic Acids Res 43, 5099-5111, (2015).

80 Zerfass, K. & Beier, H. Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. Nucleic Acids Res 20, 5911-5918 (1992).

81 Harrell, L., Melcher, U. & Atkins, J. F. Predominance of six different hexanucleotide recoding signals 3' of read-through stop codons. Nucleic Acids Res 30, 2011-2017 (2002).

82 Tork, S., Hatin, I., Rousset, J. P. & Fabret, C. The major 5' determinant in stop codon read-through involves two adjacent adenines. Nucleic Acids Res 32, 415-421, (2004).

83 Pelham, H. R. Leaky UAG termination codon in tobacco mosaic virus RNA. Nature 272, 469-471 (1978).

84 Cimino, P. A., Nicholson, B. L., Wu, B., Xu, W. & White, K. A. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. PLoS Pathogens, (2011).

85 Namy, O., Hatin, I. & Rousset, J. P. Impact of the six nucleotides downstream of the stop codon on translation termination. EMBO Rep 2, 787-793, (2001).

86 Urban, C., Zerfass, K., Fingerhut, C. & Beier, H. UGA suppression by tRNACmCATrp occurs in diverse virus RNAs due to a limited influence of the codon context. Nucleic Acids Res 24, 3424-3430 (1996).

87 Napthine, S., Yek, C., Powell, M. L., Brown, T. D. & Brierley, I. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. RNA 18, 241-252, (2012).

88 Li, G. & Rice, C. M. The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. J Virol 67, 5062-5067 (1993).

89 Firth, A. E., Wills, N. M., Gesteland, R. F. & Atkins, J. F. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. Nucleic Acids Res 39, 6679-6691, (2011).

90 Brown, C. M., Dinesh-Kumar, S. P. & Miller, W. A. Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. J Virol 70, 5884-5892 (1996).

91 Newburn, L. R., Nicholson, B. L., Yosefi, M., Cimino, P. A. & White, K. A. Translational readthrough in Tobacco necrosis virus-D. Virology 450-451, 258-265, (2014).

92 Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157, 77-94, (2014).

93 Kiyosawa, K. et al. Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 12, 671-675 (1990).

94 Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. Internal ribosome entry site within hepatitis C virus RNA. J Virol 66, 1476-1483 (1992).

95 Wang, C., Sarnow, P. & Siddiqui, A. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. J Virol 67, 3338-3344 (1993).

96 Seth, P. P. et al. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. J Med Chem 48, 7099-7102, (2005).

97 Dibrov, S. M., Johnston-Cox, H., Weng, Y. H. & Hermann, T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. Angew Chem Int Ed Engl 46, 226-229, (2007).

98 Boerneke, M. A. & Hermann, T. Ligand-responsive RNA mechanical switches. RNA Biol 12, 780-786, (2015).

99 Pestova, T. V., de Breyne, S., Pisarev, A. V., Abaeva, I. S. & Hellen, C. U. eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. EMBO J 27, 1060-1072,(2008).

100 Simon, A. E. 3'UTRs of carmoviruses. Virus Res 206, 27-36, (2015).

101 Olsthoorn, R. C., Mertens, S., Brederode, F. T. & Bol, J. F. A conformational switch at the 3' end of a plant virus RNA regulates viral replication. EMBO J 18, 4856-4864, (1999).

102 Krab, I. M., Caldwell, C., Gallie, D. R. & Bol, J. F. Coat protein enhances translational efficiency of Alfalfa mosaic virus RNAs and interacts with the eIF4G component of initiation factor eIF4F. J Gen Virol 86, 1841-1849, (2005).

103 Guogas, L. M., Laforest, S. M. & Gehrke, L. Coat protein activation of alfalfa mosaic virus replication is concentration dependent. J Virol 79, 5752-5761, (2005).

104 Petrillo, J. E., Rocheleau, G., Kelley-Clarke, B. & Gehrke, L. Evaluation of the conformational switch model for alfalfa mosaic virus RNA replication. J Virol 79, 5743-5751, (2005).

105 Chen, S. C. & Olsthoorn, R. C. *In vitro* and *in vivo* studies of the RNA conformational switch in Alfalfa mosaic virus. J Virol 84, 1423-1429, (2010).

106 Rubino, L., Burgyan, J. & Russo, M. Molecular cloning and complete nucleotide sequence of carnation Italian ringspot tombusvirus genomic and defective interfering RNAs. Arch Virol 140, 2027-2039 (1995).

107 Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. J Virol 75, 1834-1841, (2001).

108 Su, M. C., Chang, C. T., Chu, C. H., Tsai, C. H. & Chang, K. Y. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. Nucleic Acids Res 33, 4265-4275, (2005).

109 Plant, E. P., Rakauskaite, R., Taylor, D. R. & Dinman, J. D. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. J Virol 84, 4330-4340, (2010).

110 Gao, F. & Simon, A. E. Multiple Cis-acting elements modulate programmed -1 ribosomal frameshifting in Pea enation mosaic virus. Nucleic Acids Res,(2015).

111 Houck-Loomis, B. et al. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. Nature 480, 561-564, (2011).

112 Carrington, J. C., Heaton, L. A., Zuidema, D., Hillman, B. I. & Morris, T. J. The genome structure of turnip crinkle virus. Virology 170, 219-226 (1989).

113 Hacker, D. L., Petty, I. T., Wei, N. & Morris, T. J. Turnip crinkle virus genes required for RNA replication and virus movement. Virology 186, 1-8 (1992).

114 Yuan, X., Shi, K. & Simon, A. E. A local, interactive network of 3' RNA elements supports translation and replication of Turnip crinkle virus. J Virol 86, 4065-4081, (2012).

115 Chattopadhyay, M. et al. Requirement for Host RNA-Silencing Components and the Virus-Silencing Suppressor when Second-Site Mutations Compensate for Structural Defects in the 3' Untranslated Region. J Virol 89, 11603-11618, (2015).

116 McCormack, J. C. et al. Structural domains within the 3' untranslated region of Turnip crinkle virus. J Virol 82, 8706-8720, doi:JVI.00416-08 (2008).

117 Stupina, V. A. et al. The 3' proximal translational enhancer of Turnip crinkle virus binds to 60S ribosomal subunits. RNA 14, 2379-2393, (2008).

118 Yuan, X., Shi, K., Meskauskas, A. & Simon, A. E. The 3' end of Turnip crinkle virus contains a highly interactive structure including a translational enhancer that is disrupted by binding to the RNA-dependent RNA polymerase. RNA 15, 1849-1864, (2009).

119 Yuan, X., Shi, K., Young, M. Y. & Simon, A. E. The terminal loop of a 3' proximal hairpin plays a critical role in replication and the structure of the 3' region of Turnip crinkle virus. Virology 402, 271-280, (2010).

120 Batten, J. S., Desvoyes, B., Yamamura, Y. & Scholthof, K. B. A translational enhancer element on the 3'-proximal end of the Panicum mosaic virus genome. FEBS Lett 580, 2591-2597, (2006).

121 Chattopadhyay, M., Shi, K., Yuan, X. & Simon, A. E. 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, (2011).

122 Eswarappa, S. M. et al. Programmed translational readthrough generates antiangiogenic VEGF-Ax. Cell 157, 1605-1618,(2014).

123 Bidou, L., Allamand, V., Rousset, J. P. & Namy, O. Sense from nonsense: therapies for premature stop codon diseases. Trends Mol Med 18, 679-688 (2012).

124 Csibra, E., Brierley, I. & Irigoyen, N. Modulation of stop codon read-through efficiency and its effect on the replication of murine leukemia virus. J Virol 88, (2014).

125 Dulude, D., Berchiche, Y. A., Gendron, K., Brakier-Gingras, L. & Heveker, N. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. Virology 345, 127-136, (2006).

126 Brierley, I., Digard, P. & Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57, 537-547 (1989).

127 Green, L., Kim, C. H., Bustamante, C. & Tinoco, I. Characterization of the mechanical unfolding of RNA pseudoknots. J Mol Biol 375, 511-528, (2008).

128 Paul, C. P., Barry, J. K., Dinesh-Kumar, S. P., Brault, V. & Miller, W. A. A sequence required for -1 ribosomal frameshifting located four kilobases downstream of the frameshift site. J Mol Biol 310, 987-999, (2001).

129 Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. J Gen Virol 93, 1385-1409, (2012).

130 Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem 68, 913-963, (1999).

131 Thivierge, K. et al. Plant virus RNAs. Coordinated recruitment of conserved host functions by (+) ssRNA viruses during early infection events. Plant Physiol 138, 1822-1827, doi:10.1104/pp.105.064105 (2005).

132 Gao, F., Kasprzak, W. K., Szarko, C., Shapiro, B. A. & Simon, A. E. The 3' untranslated region of Pea Enation Mosaic Virus contains two T-shaped, ribosome-binding, cap-independent translation enhancers. J Virol 88, 11696-11712, (2014).

133 Feng, Y. X., Yuan, H., Rein, A. & Levin, J. G. Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the gag termination codon followed by an RNA pseudoknot. J Virol 66, 5127-5132 (1992).

134 Simon, A. E. & Gehrke, L. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. Biochim Biophys Acta 1789, 571-583, (2009).

135 Gaudin, C. et al. Structure of the RNA signal essential for translational frameshifting in HIV-1. J Mol Biol 349, 1024-1035, (2005).

136 Staple, D. W. & Butcher, S. E. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. J Mol Biol 349, 1011-1023, (2005).

137 Jones, D. S. et al. The effect of specific mutations at and around the gag-pol gene junction of Moloney murine leukaemia virus. Nucleic Acids Res 17, 5933-5945 (1989).

138 Alam, S. L., Wills, N. M., Ingram, J. A., Atkins, J. F. & Gesteland, R. F. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. J Mol Biol 288, 837-852, (1999).

139 Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8, 91, (2008).

140 McCormack, J. C. & Simon, A. E. Callus cultures of Arabidopsis. Curr Protoc Microbiol Chapter 16, Unit16D.11, (2006).

141 Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 1, 1610-1616, 2006).

142 Spitale, R. C. et al. RNA SHAPE analysis in living cells. Nat Chem Biol 9, 18-20, (2013).

143 McGinnis, J. L. & Weeks, K. M. Ribosome RNA assembly intermediates visualized in living cells. Biochemistry 53, 3237-3247, (2014).

144 Gulati-Sakhuja, A. & Liu, H. Y. Complete nucleotide sequence and genome organization of Calibrachoa mottle virus (CbMV)--a new species in the genus Carmovirus of the family Tombusviridae. Virus Res 147, 216-223, (2010).

145 Huang, M. et al. Complete nucleotide sequence and genome organization of hibiscus chlorotic ringspot virus, a new member of the genus Carmovirus: evidence for the presence and expression of two novel open reading frames. J Virol 74, 3149-3155 (2000).

146 Zhou, T., Fan, Z. F., Li, H. F. & Wong, S. M. Hibiscus chlorotic ringspot virus p27 and its isoforms affect symptom expression and potentiate virus movement in kenaf (Hibiscus cannabinus L.). Mol Plant Microbe Interact 19, 948-957, (2006).

147 Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948, (2007).

148 Gazo, B. M., Murphy, P., Gatchel, J. R. & Browning, K. S. A novel interaction of Cap-binding protein complexes eukaryotic initiation factor (eIF) 4F and eIF(iso)4F with a region in the 3'-untranslated region of satellite tobacco necrosis virus. J Biol Chem 279, 13584-13592, (2004).

149 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31, 3406-3415 (2003).

150 Liu, Y., Wimmer, E. & Paul, A. V. Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochim Biophys Acta 1789, 495-517, (2009).

151 Pathak, K. B., Pogany, J. & Nagy, P. D. Non-template functions of the viral RNA in plant RNA virus replication. Curr Opin Virol 1, 332-338, (2011).

152 Pogany, J., White, K. A. & Nagy, P. D. Specific binding of tombusvirus replication protein p33 to an internal replication element in the viral RNA is essential for replication. J Virol 79, 4859-4869, (2005).

153 Nicholson, B. L., Lee, P. K. & White, K. A. Internal RNA Replication Elements are Prevalent in Tombusviridae. Front Microbiol 3, 279, (2012).

154 Tatsuta, M., Mizumoto, H., Kaido, M., Mise, K. & Okuno, T. The red clover necrotic mosaic virus RNA2 trans-activator is also a cis-acting RNA2 replication element. J Virol 79, 978-986, (2005).

155 Maquat, L. E., Tarn, W. Y. & Isken, O. The pioneer round of translation: features and functions. Cell 142, 368-374, (2010).

Principia Mathematica

"This most beautiful system of the sun, planets, and comets could only proceed from the counsel and dominion of an intelligent and powerful Being. This Being governs all things, not as the soul of the world, but as Lord over all."

Sir Isaac Newton