RNA structure-mediated regulation of cap-independent translation by plant viruses

Jelena Jeremic Kraft
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd
Part of the Biochemistry, Biophysics, and Structural Biology Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/10074
RNA structure-mediated regulation of cap-independent translation by plant viruses

by

Jelena Jeremic Kraft

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
W. Allen Miller, Major Professor
   Alan Myers
   Cathy Miller
   Gustavo Macintosh
   Mark Hargrove

Iowa State University
Ames, Iowa
2011

Copyright © Jelena Jeremic Kraft, 2011. All rights reserved.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHAPTER 1. GENERAL INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>CHAPTER 2. CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS</strong></td>
<td>34</td>
</tr>
<tr>
<td>OF THE BARLEY YELLOW DWARF VIRUS CAP-INDEPENDENT TRANSLATION ELEMENT</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER 3. CATION-INDUCED STRUCTURE OF CAP-INDEPENDENT</strong></td>
<td>50</td>
</tr>
<tr>
<td>TRANSLATION ELEMENTS IN THE 3’ UNTRANSLATED REGION OF PLANT RNA VIRUSES</td>
<td></td>
</tr>
<tr>
<td>FACILITATES INTERACTION WITH EUKARYOTIC TRANSLATION INITIATION FACTOR 4F</td>
<td></td>
</tr>
<tr>
<td>(eIF4F)</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER 4. STRUCTURAL PLASTICITY OF BARLEY YELLOW DWARF</strong></td>
<td>82</td>
</tr>
<tr>
<td>VIRUS-LIKE CAP-INDEPENDENT TRANSLATION ELEMENTS IN FOUR GENERA OF PLANT</td>
<td></td>
</tr>
<tr>
<td>VIRAL RNAS</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER 5. GENERAL CONCLUSIONS</strong></td>
<td>119</td>
</tr>
<tr>
<td><strong>APPENDIX 1. CRYSTALLIZATION OF RELATED VIRAL</strong></td>
<td>125</td>
</tr>
<tr>
<td>CAP-INDEPENDENT TRANSLATION ELEMENTS</td>
<td></td>
</tr>
<tr>
<td><strong>APPENDIX 2. INVESTIGATION OF HELICAL STACKING IN THE BTE</strong></td>
<td>139</td>
</tr>
<tr>
<td>USING “RNA LEGO” SYSTEM</td>
<td></td>
</tr>
<tr>
<td><strong>APPENDIX 3. CONTRIBUTION TO BOOK CHAPTER PUBLISHED</strong></td>
<td>147</td>
</tr>
<tr>
<td>IN RECENT ADVANCES IN PLANT VIROLOGY</td>
<td></td>
</tr>
</tbody>
</table>

**ACKNOWLEDGEMENTS**                                                     | 193 |
ABSTRACT

Efficient eukaryotic translation is initiated upon recognition of the mRNA 5’ cap structure by initiation factors and binding of the small ribosomal subunit to the 5’ end of the mRNA. However, many viral mRNAs have alternatives to the cap and poly(A) tail that ensure efficient viral translation at the expense of host protein synthesis. Specifically, plant viral RNAs in the Luteovirus genus, including that of Barley yellow dwarf virus (BYDV) harbor a cap-independent translation element (CITE) located in the 3’ untranslated region (UTR) of viral genome. The BYDV-like translation element (BTE) is defined by the presence of a conserved 17 nt sequence (17 nt CS) GAUCCUUGGAAACAGG that adopts a stem-loop with paired underlined bases and a loop that can base pair to a complementary bases in the 5’UTR. The BTE promotes translation initiation at the 5’-proximal AUG on viral RNAs by recruiting translation machinery and delivering it to the 5’ end via a kissing stem-loop.

The key to understanding how the BTE outcompetes the host mRNA for protein synthesis machinery is to determine the structure of the BTE at atomic resolution. Thus, we conducted phylogenetic and mutagenesis analysis of BYDV and BTEs from diverse viruses to identify the most structurally homogeneous sequences before subjecting them to extensive crystallization screening. In this process, we discovered that Groundnut rosette virus (GRV) in genus Umbravirus also harbors a functional BTE in its genome despite violating 17 nucleotide (nt) consensus. We also demonstrated that BTE sequences are capable of crystallization under a variety of conditions suggesting that they all adopt a compact fold. One of the screened BTEs, the 87c BTE RNA gave crystals that diffracted initially to 30 Å.
Upon crystallization condition optimization, we obtained crystals that diffract below 5Å, with a complete data set collected to 6.9 Å. This crystal form indexes with an Rmerge of 0.094 in the monoclinic space group C2 with unit-cell parameters a=316.6 Å, b=54.2 Å, c=114.5 Å, α = γ =90°, β =105.1°.

Despite screening thousands of conditions and dozens of sequence variants, I have been unable to obtain BTE crystals that diffract at less than 4.5Å. Thus, I switched to an entirely different class of structure, the Pea enation mosaic virus RNA 2-like translation enhancer (PTE). The PTE binds and requires smaller translation initiation factor eIF4E for which crystal structure is known. Although the PTE alone did not crystallize, a 1:1 mixture of eIF4E with PTE gave large crystals that diffracted at 2.3Å resolution. Data analysis revealed that PTE-eIF4E RNA crystals belong to the same space group and have near identical unit cell parameters as eIF4E crystals suggesting that eIF4E simply crystallized free of the RNA despite being presence of the RNA in the mother liquor.

All these elements utilize unique magnesium dependent folds to bind the rate-limiting eIF4F complex, but the binding site for most CITE has not been mapped out yet. Thus, we used chemical and enzymatic probing to map the eIF4G-binding site on the BTE. Footprinting experiments revealed that eIF4G alone is capable of protecting most of the BYDV BTE structure from chemical and enzymatic modification. This BTE protection pattern is even more exaggerated in the presence of heterodimeric eIF4F complex suggesting binding to a very large area of the BTE with the exception of loop three. This loop known to base pair to the 5’ UTR remained solvent accessible even at high (1µM) eIF4G/4F concentrations. This data fit the proposed model where the eIF4F complex binds directly to
the BTE and once recruited, it is delivered to the 5'UTR via a long distance kissing interaction
CHAPTER 1: GENERAL INTRODUCTION

Central dogma of biology:

\[
\text{DNA \ Transcription \ RNA \ Translation \ Protein}
\]

This overly simplified scheme of protein synthesis involving making RNA from DNA template and subsequent incorporation of amino acids into the polypeptide chain was first coined by Francis Crick (Crick, 1958), but the foundations for it was laid years before in the independent works of Torbjörn Caspersson and Jean Brachet who showed that nucleic acids and RNA in particular is directly involved in protein synthesis (Thieffry and Burian, 1996). Years later, the site of protein synthesis on endoplasmic reticulum was discovered first in animal cells (Palade et al., 1954) and then in plants (Hodge et al., 1957). Following the discovery of endoplasmic reticulum and DNA structure (Watson and Crick, 1954) the major scientific goal was to understand how the information encoded in DNA is expressed.

The crucial discovery of the adaptor molecule tRNA, first proposed by Francis Crick and later confirmed by Zamecnik and Hoagland, helped explain an important step in protein synthesis. Using a cell-free extract and \(^{14}C\)-labeled amino acids, they showed that amino acids interact with tRNA before protein synthesis (Hoagland et al., 1958). Although discovery of tRNA answered how amino acids are delivered to the ribosome, it did not answer how information encoded in DNA in the nucleus was expressed via the ribosome in cytoplasm. Brenner, Jacob and Meselson (Brenner et al., 1961) found that missing link to be mRNA, a molecule synthesized directly from the DNA template of genes and used to provide the information for protein synthesis. However, the puzzle of how mRNA specifies particular amino acids remained to be solved. The discovery of polynucleotide phosphorylase (Grunberg-Manago and Ochoa, 1955) made it possible to synthesize
polynucleotides of a single kind and defined length. These polynucleotides were used in the seminal work of Niremberg and Matthai who showed that poly U codes for polyphenylalanine (Nirenberg and Matthai, 1961). Next, heteropolymers with random and defined sequence composition allowed determination of the codon composition for all 20 amino acids (Kohrana, 1977, Nirenberg, 1977).

The significance of the above-mentioned results in understanding the translation process cannot be overemphasized. Protein synthesis is a highly complex and an elaborate process with the ultimate goal to translate message into protein. Translation can be divided into three highly regulated stages: initiation, chain elongation and chain termination. The research focus of this thesis is at the initiation step of translation.

**Canonical translation initiation**

Translation initiation is the first and the most heavily regulated step of protein synthesis (Holcik et al., 2005; Sonenberg & Hinnebusch, 2009). The process begins with the formation of 43S preinitiation complex composed of initiator tRNA, 40S small ribosomal subunit and several initiation factors. Once the 43S complex is correctly assembled it is recruited to the mRNA by the eukaryotic initiation factor 4F (eIF4F) scaffold (Gingras et al, 1999; Marintchev et al., 2009). Plant eIF4F is composed of large scaffold protein, eIF4G, and a much smaller cap-binding protein subunit, eIF4E (Browning, 2004). Plants also contain an isoform of eIF4F, eIFiso4F, composed of eIFiso4E and eIFiso4G (Browning, 1996). In accordance with the scanning model (Kozak, 1989; Berthelot et al., 2004), the 43S complex moves along mRNA in the 5’ to 3’ direction in search of the AUG codon to yield 48S initiation complex. After start codon recognition, the protein factors surrounding the 40S
subunit are displaced by additional protein factors to allow 60S large ribosomal subunit joining and subsequent protein synthesis ensues (Jackson et al., 2010).

The efficiency of translation initiation is dictated by the sequence and structural elements of the mRNA (Kozak, 2002). Cellular mRNAs contain a 5’ m7G(5’)ppp(5’)N cap structure that stabilizes mRNA and assists in 43S complex recruitment to the mRNA via eIF4E binding (Haghighat, 1997). In contrast to the stabilizing effect of the 5’ cap structure, 5’ untranslated regions (UTRs) may contain stable secondary structures (stem loops) and upstream AUGs in suboptimal AUG contexts that negatively regulate translation initiation (Kozak, 2002, Rakotondrafara et al., 2006). Similar to the 5’ UTR sequences, the 3’ UTR also contains signals that impact mRNA stability and translatability. The 3’ poly(A) tail and its binding partner poly(A)-binding protein (PABP) interact with the 5’-located protein factors to circularize mRNA and stimulate translation efficiency. Message circularization through protein factors (Fig. 1) is thought to stimulate ribosome recycling by allowing ribosome to reinitiate another round of translation as well as to prevent translation of truncated mRNAs (Wells et al., 1998).

Protein synthesis is regulated primarily at initiation step (Gebauer and Hentze, 2004, Sonenberg & Hinnebusch, 2009). The reason for this tight regulation of translation initiation includes bypassing the need to go through upstream steps of regulation such as transcriptional control and providing additional level of control. Also, translation initiation regulation allows for fast response to external stimuli, reversibility via reversible modification of translation factors and flexibility to focus on single gene or whole classes of mRNAs via modulation of general factors. Here, I will focus on the most important regulators of protein synthesis in plants.
RNA specific regulation. RNA itself can function as regulator of translation by adopting specific regions where structure, sequence or both are important for translatability of the message. Stable upstream hairpin loops, alternative initiation codon selection in the optimal sequence context, short upstream open reading frames (uORFs) are just few examples of sequences that directly shape the outcome of translation (Kozak et al., 2002, Franceschetti et al., 2001). For example, sucrose-mediated translational repression of Arabidopsis ATB2/AtbZIP12 transcription factor is driven by a highly structured 5’ UTR and in particular conserved upstream ORF2 (Wiese et al., 2004). A possible mechanism of translational repression involves direct interaction of uORF2 with a component of the translational machinery that slows down translating ribosomes (Wiese et al., 2004). This ORF is also found in 5’ UTR of basic leucine zipper domains in Arabidopsis and other plants, suggesting that their translation is also regulated in a similar way. In plants, another metabolite driven translational repression system was also characterized for polyamine triggered translational repression of an S-adenosylmethionine decarboxylase gene driven by a specific small uORF Franceschetti et al., 2001, Hanfrey et al., 2002).

Another mechanism of translational regulation in plants involves hybridization of micro RNAs (miRNAs) to sequences to mRNAs to suppress translation. Compared to animals, miRNA complementarity to its target must be almost perfect, with no or very few base mismatches. Specific examples include miRNAs 156 and 157 that target the 3’ UTR of the Arabidopsis SBP box gene SPL3 to prevent very early flowering (Gandikota et al., 2007) and miR160 target regulation of Arabidopsis ARF17 mRNA that prevents developmental abnormalities (Mallory et al., 2005). The high degree of sequence complementarity between Arabidopsis miRNA172 and AP2 mRNA leads to translational repression to allow proper
flower development (Chen et al., 2004). These miRNAs isolated from Arabidopsis were also found in other plants, suggesting evolutionarily conserved roles for plant miRNAs (Axtell and Bartel, 2005).

**RNA binding proteins.** Protein factors bound to modifications present at both ends of the mRNA act together to ensure the integrity of the message and offer additional levels of regulation through changes in phosphorylation state of the factors (Gallie et al., 2004). In plants, the least abundant initiation factor, eIF4E (Dunkan et al., 1987), lacks phosphorylated serine 209 present in mammalian 4E that was shown to bind the cap structure with high affinity (Joshi, et al., 1995). Moreover, plants lack eIF4E-binding protein (eIF4E-BP) homolog that in its hypophosphorylated state binds eIF4E to suppress translation (Gingras et al., 1996), suggesting that plants employ different regulatory mechanisms. Indeed, it was shown that phosphorylation of recombinant plant eIF4E at serine 207 by mammalian CK2 in vitro leads to approximately 2-fold reduction in its cap binding ability (Khan and Goss, 2004).

As stated earlier, plants utilize two eIF4F complexes for translation of mRNA presumably to allow for differential selection of mRNAs. More complex, structured mRNA sequences are primarily recognized by eIF4F, while more abundant eIFiso4F interacts with unstructured mRNAs (Gallie, 2007). Furthermore, the N-terminal region of eIFiso4G was found to enhance the interaction of hypophosphorylated PABP for poly(A) stretch (Gallie et al., 1997; Hanh et al., 2000). In its phosphorylated state, eIF4B preferentially stimulates PABP cooperative binding to poly(A) indicating that modulation of PABP phosphorylation state serves to modulate the nature of its interaction with the initiation complex (Hanh et al., 2000).
As a part of the ternary complex, eIF2 first binds GTP, followed by the initiator Met-tRNA. Following the assembly of the pre-initiation complex and the start codon recognition eIF2B, catalyses hydrolysis of eIF2-bound GTP to GDP. Mammalian eIF2 is a trimer composed of α, β and γ subunits. Many mammalian regulatory pathways activated under stress conditions, often including virus infection, converge on the phosphorylation of the α subunit’s serine 51 (Gebauer and Hentze, 2004). Phosphorylation of serine 51 results in inability of eIF2 to exchange GTP for GDP that leads to inhibition of translation. Regulation of translation initiation via eIF2α phosphorylation in plants is more questionable. Although plants lack a homologue of eIF2B, all of its subunits were identified in genome of Arabidopsis indicating the possibility that plants may utilize eIF2B for GTP hydrolysis (Gallie, 2007). Furthermore, serine 51 and other residues are phosphorylated in wheat eIF2α and well as phosphorylation of eIF2β (Gallie, et al.,1997), but it remains to be determined if this phosphorylation functions in regulation of translation in plants.
Fig. 1. mRNA circularization mediated by translation initiation factors and their roles in initiation of protein synthesis.
RNA structure and role of ions in folding

Following the discovery of DNA structure and catalytic RNAs (Watson and Crick, 1954, Cech et al., 1981, Forster et al., 1990) the spotlight turned to RNA and understanding how RNA folds and utilizes unique three-dimensional shapes to accomplish multiple tasks. RNA structure is hierarchical with three levels of organization including primary sequence, secondary and tertiary structure.

Efforts to accurately model tertiary RNA structures are still in their infancy mostly because RNA architecture is less dependent on the canonical Watson-Crick (WC) type interactions and instead more dependent on non-WC type interactions that form specific motifs (Westhof and Fritsch, 2000). These motifs can occur between paired helical regions such as coaxial stacking of helices observed in the structure of tRNA^{Phe}, hammerhead ribozyme and PTE class of cap-independent translation elements (Quigley et al., 1976, Pley et al., 1994, Wang et al., 2011) where one helical stack continues onto another and is often stabilized by metal-ion binding near the site of the stack.

Alternatively, RNA motifs can occur between paired helical and unpaired regions, as is the case with tetraloops. The most prevalent GNRA (where N=any base, R=purine) tetraloop motif in biological RNAs owes its stability to a combination of hydrogen bonding and base stacking (Correll et al., 2003) and has been observed in crystal structures of P4-P6 domain of group I intron where it specifically interacts with tetraloop receptor (Cate et al., 1996). This tetraloop-tetraloop receptor interaction often pulls ribose-phosphate backbone close to itself resulting in tightly packing of RNA helices in the ribose zipper motif observed in the crystal structure of hammerhead ribozyme (Pley et al., 1994).

Unlike the sequence unspecific interaction of RNA strands in the ribose zipper, motifs
formed as result of interactions between unpaired segments require sequence specificity. In loop-loop interaction of Barley yellow dwarf RNA that specificity is imparted by five base pairs between complementary hairpins while Tomato bushy stunt virus (TBSV) RNA required 9 bases of complemenarity (Rakotondrafara et al., 2006; Fabian et al., 2004). Another recurring motif observed between unpaired RNA segments is the pseudoknot formed through interaction of loop bases with complementary bases outside of the loop. This type of the interaction is the primary driving force in folding of several viral RNAs (Wang et al., 2009; 2011; Zuo et al., 2010). Similarly to coaxial stacking interactions, pseudoknots are stabilized in the presence of mono or divalent cations. (Wyatt et al., 1990).

Numerous crystal structures provide insight into cation binding to RNA, especially for magnesium ions. There are two main modes of Mg binding to RNA structure. One mode of binding involves coordination of partially desolvated Mg\(^{2+}\) ions to sequence specific regions and the other involves interaction with anionic phosphate oxygen atoms in non-specific manner (Misra et al., 2002). The former mode is by far more prevalent and its role is to shield the negative phosphate charges in regions of sequence that gets progressively more compact as RNA folds (Woodson et al., 2005; Koculi et al., 2007). Without ions, the repulsive forces of compacted RNA structure would exceed the energetically favorable interactions resulting in disruption of RNA three-dimensional fold. Recent study by Woodson group showed that the shape and size of the cations are both important determinants of RNA folding. In general, the smaller the atomic radius and the greater the charge density of the metal ion the better charge neutralization leading to more compact RNA (Koculi et al., 2007) as result of closer interaction with RNA. High charge density cations are not only seen to neutralize phosphate charges but also to mediate attraction
between separate helical domains that further stabilizes the folded RNA of Azoarcus
ribozyme (Rangan and Woodson, 2003). Thus, in general terms, RNA fold can be described
as collection of preformed WC paired helices and RNA motifs stabilized by the non-WC
interactions and cation binding.

**Alternative translation initiation mediated by viral RNA structure**

Viruses depend heavily on the host cell translation apparatus for survival. Despite
lacking canonical modifications like 5’ cap and/or poly(A) tail, viral mRNAs must out
compete the host cell mRNAs for the ribosomes and protein factors. They accomplish this
by encoding sequences and complex structures in their UTRs that allow the virus to escape
the cellular antiviral responses and to steal the initiation factors away from the host mRNAs.
(Bushell et al., 2002)

**IRESes.** The 5’ terminal region of animal and some plant viral mRNAs contain an
internal ribosomal entry site (IRES) that recruits the 40S subunit to the proximity of the start
codon to promote cap–independent translation (Doudna and Sarnow, 2007). There is little
similarity in sequence, structure, and translation factor requirements amongst different
classes of IRESes. Plant viral RNAs harbor structurally simpler, approximately 100 base-
long cap-independent translation elements found at the both ends of the viral genome.
Regardless of their particular structure or mechanism of action, all IRESes have evolved
functions that compensate for the missing components of translation initiation machinery to
allow synthesis of essential proteins, both viral and cellular.

For example, tobacco etch potyvirus (TEV) naturally lacks a 5’ cap structure and yet
it is translated efficiently. The TEV 5’ leader sequence contains two cap-independent
regulatory elements (CIREs) that are both needed for optimal translation initiation (Niepel and Gallie, 1999). The TEV 5’ UTR promoted translation of a second ORF when placed between two ORFs in a dicistronic reporter construct, suggesting that they function as an IRES. Potato virus Y (PVY) is another potyvirus that contains an IRES element that directs efficient translation of a second ORF from the intercistronic region (Levis and Astier-Manifacier, 1993).

Other examples of plant IRESes include the 5’ leader sequences of both genomic RNAs (gRNA) of Blackcurrant reversion virus (BRV, family Comoviridae) that mediate translation by recruiting the 40S ribosomal subunit directly to 5’ element (Karetnikov and Letho, 2007, 2008). Presence of UC-rich stretches complementary to 18S rRNA and formation of closed loop structure via long-distance base-pairing between the UTRs all promote translation of BRV RNA.

CITEs. The 3’ UTRs of many plant viruses contain cap-independent translation elements (CITEs) that confer efficient translation initiation in the absence of cap and/or poly(A) tail. These 3’ CITEs differ from IRESes which are located upstream of the start codon. So far, seven different classes of 3’ cap-independent translation elements have been characterized (Miller et al., 2007) and they share little sequence or secondary structural similarity (Fig. 2).

- TED. Satellite tobacco necrosis virus (STNV) harbors the first-discovered 3’ CITE, termed translation enhancer domain (TED) (Meulewaeter et al., 1998). TED forms a long hairpin-like structure interrupted by several internal bulges. It resides in the proximal end of the 3’ UTR but it allows efficient initiation from the 5’ AUG probably via long distance base pairing to a 5’ complementary sequence in the 5’ UTR (Meulewaeter et al.,
The proposed mechanism of TED-mediated translation initiation involves direct binding of eIF4F/iso4F to TED followed by interaction with 5’ end bound 43S pre-initiation complex either through a protein bridge or by direct RNA-RNA interaction (Gazo et al., 2004).

**BTE.** Among the most characterized 3’ CITEs is that of Barley yellow dwarf virus. The BYDV BTE and almost all BYDV-like translation elements (BTEs) found in related viruses contain a highly conserved 17 nt tract GGAUCCUGGGAAACAGG (17 nt CS) which forms a stable stem-loop (SL-I). One of the first two conserved G’s in this sequence is highly modified by chemical reagent during structure probing experiments (Wang et al., 2010). Another defining characteristic of this group of elements is presence of a stable stem-loop (SL-III in the BYDV BTE) containing at least six base pairs and topped off by a loop capable of base pairing to complementary bases in the 5’UTR (Guo et al., 2001). Other shared characteristics include a long basal helix (S-IV) from which two to five helices radiate, depending on the virus, and a variable numbers of nonconserved, single-stranded bases around junction that link together radiating helices from the central hub.

**PTE.** A completely different group of cap-independent Panicum mosaic virus-like translation elements (PTEs), from Panicum mosaic (PMV) (Batten et al., 2006), Pea enation mosaic virus-2 (PEMV2) (Wang et al., 2009) and other viruses was recently discovered. The PTE consists of a long bulged basal helix from which two stem-loops radiate. Two major G and C rich domains are predicted or established to form a magnesium-dependent pseudoknot essential for PTE function. Unlike the BTE, PTEs bind to the eIF4E subunit of eIF4F complex with high affinity that makes them unique
among uncapped mRNAs (Wang et al., 2009).

- **YTE.** The Y-shaped translation element R3.5 of Tomato bushy stunt virus (TBSV, family *Tombusviridae*) is composed of three helical regions, of which stem-loop B forms long-distance base-pairing with 9 complementary bases of the T-shaped element located in the 5’ UTR (Fabian et al., 2004). This long-distance communication essential for efficient translation in vivo seems to tolerate more base changes than the similar interaction in the BTE, as long as base pairing is maintained.

- **TSS.** 3’ UTR of another member of *Tombusviridae*, Turnip crinkle virus (TCV) consisting of 5 stem loops, 4 pseudoknots and multiple unpaired single stranded linker regions harbor tRNA-shaped structure (TSS) essential for translation (Zuo et al., 2010). It is proposed that TSS acts by recruiting and binding directly the 60S ribosomal subunit (Stupina et al., 2008). Once recruited, the 60S subunit is then delivered to 5’end where the 40S subunit binds. Interestingly, the viral RNA dependent RNA polymerase (RdRp) was also found to bind near the TSS, suggesting that it acts to prevent further ribosomal binding to allow for the transcription of the negative strand (Yuan et al., 2009).

- **ISS.** The I-shaped structure found in the 3’ UTRs of Melon necrotic spot virus (MNSV) (Truniger et al., 2008) and Maize necrotic streak virus (MNeSV) (Scheets and Redinbaugh, 2006, Nicholson et al. 2010) forms a long extensively paired stem loop structure intersected with few internal loops. To allow efficient ISS-mediated translation initiation, the MNeSV ISS must bind to both viral 5’UTR and eIF4F in a similar mechanism as proposed for BYDV BTE. However, unlike the BYDV BTE that specifically interacts with eIF4G subunit, ISS binds to eIF4F via eIF4E subunit. Furthermore, mutations in the 3’ CITE of MNSV allowed it to break the eIF4E-mediated
resistance suggesting that specific binding to eIF4E is also required for MNSV-mediated translational enhancement and is general behavior of this class of 3' CITEs. To date, no crystal structure of any of plant viral cap-independent translation element elements has been reported and structure-based mechanistic details of translation initiation remain to be elucidated.
Fig. 2. Classes of 3’ cap-independent translation elements classified according to structure.
Barley yellow dwarf virus (BYDV)

Barley yellow dwarf virus (genus *Luteovirus*, family Luteoviridae) is a positive sense RNA virus. Its 5.6 kilobase genome encodes six open reading frames (ORFs) and makes three 3’ co-terminal subgenomic RNAs (sgRNAs) in infected cells (Fig. 3) (Kelly et al., 1994; Miller et al., 1997) (Fig. 3). Only two proximal ORFs are translated from genomic RNA. ORFs 1 and 2 that encode the RdRp complex are produced as a fusion protein via ribosomal frameshifting (Barry et al., 2002). The remaining downstream ORFs are expressed from sgRNA1. The movement protein encoded by ORF 4 is translated via leaky scanning and an aphid transmission functioning protein encoded by ORF 5 is produced by read through of the ORF 3 stop codon. ORF 6 has no known function and it is translated from sgRNA 2 (but only in vitro). sgRNA2 acts as *trans*-regulator of gRNA and sgRNA1 translation (Wang et al., 1997, Shen et al, 2006).

We have long determined that in the absence of a cap and poly(A) tail, BYDV utilizes its 3’ cap-independent translation element to mediate translation initiation at 5’end of viral RNA. In order to stimulate cap-independent translation from the 3’UTR, the stem-loop three (SLIII) of the BTE must base pair directly to a 5’-complementary loop to deliver recruited initiation factors (Fig. 2). However, when the BTE is moved to the 5’ UTR in artificial constructs, the long-distance base pairing is unnecessary (Rakotondrafara et al., 2006; Guo et al., 2000). This in vitro-defined 105 nt long BTE sequence is not sufficient to drive efficient translation initiation *in vivo*. In a more competitive environment of oat protoplast an extension of BTE’s basal helix IV is necessary for full cap-independent activity and probably acts by increasing the stability of the BTE structure (Shen et al., 2004).
The BTE requires and binds eIF4G with moderately high affinity (Kd=177 nM) and the heterodimeric eIF4F complex with even higher affinity (Kd= 37 nM) (Treder et al., 2008). The eIF4G/4F binding correlates well with ability of these factors to recover translation of a reporter mRNA that has been inhibited by addition of excess BTE or cap-analog *in trans* (Treder et al., 2008). In contrast, eIF4E alone does not bind the BTE (Kd=>>2000 nM), suggesting that if it does interact with the BTE it does so through mediation of eIF4G. These data strongly support the notion that the 3’ BTE binds the eIF4G subunit of eIF4F, and once recruited, the eIF4F complex is delivered to 5’UTR where translation initiates.
Figure 2. **BYDV genome organization and expression signals.** Open reading frames are indicated by the white boxes. Bold lines stand for genomic and subgenomic RNAs. The BYDV cap-independent translation element (BTE), located in the 3' UTR, promotes translation initiation at the 5’-proximal AUG on viral RNAs. The 3’ BTE base pairs with the 5’ UTR (dashed arrow) of gRNA or sgRNA1 via kissing stem-loops (red) to allow translation to occur. Black boxes indicate frameshift element on genomic RNA needed for RdRp translation; gray boxes are readthrough signals for translation of RTD protein required for aphid transmission.
Role of translation initiation factors in plant virus infections

The study of recessive resistance genes against plant viruses revealed that the key modulators in the plant virus infections are translation initiation factors and in particular the eIF4F scaffold. About one-half of all known plant virus resistance genes are recessive and so far all sequenced recessive genes encode alleles of eIF4G or eIF4E or their isoforms (Truniger & Aranda (2009) Adv Virus Res 75, 119-159). Examples of naturally occurring resistance genes that are alleles of eIF4E or its isoform eIFiso4E include ubiquitous Potyvirus resistance genes (Nicaise, et al., 2003, Robaglia et al., 2006, Ruffel et al., 2005). They function by incorporating mutations near or at the cap-binding pocket and are hypothesized or known to disrupt the binding of a virus-encoded protein VPg present at the 5’ end of the viral genome.

Understanding the role of eIF4G/eIF4E in RNA virus infection is particularly important for the viruses studied here because it has been shown that all of them function by binding to the eIF4F (Treder et al, 2008, Wang et al., 2008; 2010). For example, a single transversion in the genome-linked viral protein (VPg) gene of Rice yellow mottle sobemovirus led to resistance breaking (Hebrard et al., 2006). Furthermore, carmovirus and bymovirus 3’ UTRs were found to be important in overcoming the plant eIF4E-mediated resistance (Diaz et al., 2004, Kanyuka et al., 2005, Truniger et al., 2008) but little is known how they accomplish this. Thus, detailed mechanistic studies in this dissertation of viral elements that interact with eIF4G or eIF4E are likely to provide answers to understanding plant virus virulence.
**Dissertation summary**

One aim of this research is aimed at determining the high-resolution structure of a powerful stimulator of cap-independent translation, the Barley yellow dwarf virus (BYDV) translation element (BTE) by X-ray crystallography. This will further our understanding of alternative translation initiation mechanisms shared by many medically important viruses. The BTE naturally resides in the 3’ untranslated region (UTR) of the viral genome yet confers efficient translation initiation at the 5’ end. To accomplish this, the BTE must interact with components of the translation initiation machinery in order to deliver them to the 5’ UTR via kissing stem loops. Previously, we reported that a 105 nt long BTE element forms secondary structure consisting of three stem loops connected by single-stranded bases to a fourth helical region (stem-IV) that connects the BTE to the viral genome (Guo, et al. 2000). This stem-IV can be replaced by a GC-rich clamp sequence present at the ends of the BTE and still confer cap-independent translation. Thus, optimal RNA folding and not the sequence of stem-IV is essential for translation. Deletions of the proximal ends of stem-loops (SL) I, II or III, was deleterious for BTE function suggesting disrupted sequence and/or structure disrupted BTE’s function.

The research presented in Chapter 2 continues to attempt interrogation of the BTE structure by X-ray crystallography in a manuscript titled “Crystallization and preliminary X-ray diffraction analysis of Barley yellow dwarf virus cap-independent translation element” which was published in *Acta Cryst. F*. Most surveyed BYDV BTE sequences formed crystals under a variety of conditions suggesting that they are able to fold and maintain the folded state during crystallization process. One of the optimized sequences gave crystals that diffracted to 4.5 Å with a complete data set collected to 6.9 Å resolution. Furthermore, in
this study I pioneered a PCR scheme to generate the DNA template used for transcription of large quantities of RNA needed for crystallization that allowed for the rapid purification of multiple constructs. In this scheme, the upstream primer contained either the common class III T7 RNA polymerase promoter for guanosine-initiating transcripts or a class II T7 promoter for adenosine-initiating transcripts (Coleman et al., 2004). The class II T7 promoter gives superior 5’-end homogeneity and produced yields comparable to those obtained with the class III T7 promoter. Using a downstream primer containing 2’-O-methoxy modifications at its 5’-end prevented 3’-end heterogeneity of the RNA transcript. This approach allowed the rapid production of clean, homogeneous and acrylamide-free RNA samples suitable for structural studies as demonstrated by the 5 Å resolution diffracting crystals.

**Chapter 3** focuses on determining how the BTE RNA folds and utilizes its unique 3D shape to interact with translation machinery in a manuscript entitled “Cation-induced structure of cap-independent translation elements in the 3’ untranslated region of plant RNA viruses facilitates interaction with eukaryotic translation initiation factor 4F (eIF4F)” and written for submission to the *Journal of Molecular Biology*. We show that the BTE structure formation depends on cations and that it is driven by charge neutralization. Footprinting experiments revealed that most of the BYDV BTE is protected from chemical and enzymatic modification by eIF4G except for the loop known to base pair to the 5’ UTR. In contrast, the nonfunctional BTE BF mutant containing a four base insertion in the 17 nt consensus sequence did not show a strong protection pattern and was more evenly modified.

These data strongly support our model wherein the BTE simultaneously binds eIF4F and base pairs to the 5’ UTR via loop 3, in order for eIF4F to recruit the ribosome to the 5’
UTR. Similar behavior is reiterated among BTEs from different viruses suggesting a shared mechanism of action.

**Chapter 4** contains my contribution as second author to a paper entitled “Structural plasticity of Barley yellow dwarf virus-like cap-independent translation elements in four genera of plant viral RNAs” that was published in *Virology*. Here, I describe identification of functionally weaker Groundnut rosette virus (GRV) BTE in genus *Umbravirus*. The GRV BTE has the same overall secondary structure as that one from BYDV, but the 17 nt sequence in GRV BTE violates consensus that results in inability of this sequence to compete efficiently with other BTEs for translation initiation factors in standard wheat germ extract translations.

**Appendix 1** details purification, crystallization and preliminary X-ray data analysis of Panicum mosaic virus (PMV) RNA in complex with translation initiation factor eIF4E. The PTE-eIF4E RNA crystals belong to the same space group and have identical unit cell dimensions as eIF4E crystals suggesting that eIF4E simply crystallized free of the RNA that was present in the mother liquor.

**Appendix 2** describes implementation of the “RNA LEGO” system to study helical stacking in the BTE. By mutating two of the loops in the BTE to be complementary to each other, I established that stem loop-I (SL-I) is probably capable of kissing complementary SL-III sequences and is therefore coaxially stacked on SL-III.

**Appendix 3** contains my contribution as second author to a book chapter entitled “Roles of cis-acting elements in translation of viral RNAs” and published in *Recent Advances in Plant Virology*. 
REFERENCES


Guo, L., Allen, E., and W. A. Miller. 2000. Structure and function of a cap-independent translation element that functions in either the 3′ or the 5′ untranslated region. RNA 6, 1808–1820.


Hebrard, E., Pinel-Galzi, A., Bersoult, A., Siré, C., and D. Fargette. 2006. Emergence of a resistance-breaking isolate of Rice yellow mottle virus during serial inoculations is due to a single substitution in the genome-linked viral protein VPg. J. Gen. Virol. 87, 1369–1373.


binding protein specifies its binding to poly(A) RNA and its interaction with
eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B. J. Biol. Chem. 275, 17452-
17462.

Levis, C. and S. Astier-Manifacier. 1993. The 5’ untranslated region of PVY RNA, even
located in an internal position, enables initiation of translation. Virus Genes 7, 367–
379.

Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development

Marintchev, A., Edmonds, K. A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C.,

sequences of satellite tobacco necrosis virus RNA promoting translation in tobacco,
Plant J. 14, 169–176.


translation elements in the 3’ untranslated regions of plant viral RNAs. Biochem. Soc.
Trans. 35, 1629–1633.


stem-loop interactions mediate 5’ scanning-dependent translation by a viral 3’-cap-independent translation element. RNA 12, 1893-1906.


Treder, K., Kneller, E., L., P., Allen, E., M., Zhaohui, W., Browning, K., S., and W. A.


CHAPTER 2. CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF THE BARLEY YELLOW DWARF VIRUS CAP-INDEPENDENT TRANSLATION ELEMENT

A paper published in
Acta Crystallographica Section F (2011) 67, 561–564

Jelena J. Kraft, Julie A. Hoy and W. Allen Miller

Author contributions: JJK performed all experiments. JJK, JAH, and WAM designed and analyzed the data and wrote the papers.

ABSTRACT

Barley yellow dwarf virus (BYDV) RNA lacks a 5’ m7GTP cap, yet it is translated efficiently because it contains a 105-base BYDV-like cap-independent translation element (BTE) in the 3’ untranslated region (UTR). To understand how the BTE outcompetes the host mRNA for protein-synthesis machinery, its three-dimensional structure is being determined at high resolution. The purification using transcription from DNA containing 2’-O-methyl nucleotides and preliminary crystallographic analyses of the BTE RNA are presented here. After varying the BTE sequence and crystallization-condition optimization, crystals were obtained that diffracted to below 5 Å resolution, with a complete data set being collected to 6.9 Å resolution. This crystal form indexes with an Rmerge of 0.094 in the
monoclinic space group C2, with unit-cell parameters \( a = 316.6, b = 54.2, c = 114.5 \text{ Å}, \ a = \beta = 90, \gamma = 105.1^\circ. \)

**Synopsis.** The sequence design, purification, crystallization and preliminary X-ray diffraction analysis of the novel cap-independent translation RNA element from the 3’ untranslated region of the barley yellow dwarf virus genome is reported.

**Keywords.** Barley yellow dwarf virus-like translation element; translation initiation; 30 translation enhancers; RNA.

1. Introduction

Translation initiation is the first and the most regulated step of protein synthesis (Sonenberg & Hinnebusch, 2009; Jackson et al., 2010). The 40S ribosomal subunit complexed with tRNA and initiation factors is recruited to the mRNA by the eukaryotic initiation factor 4F (eIF4F) complex (Gingras et al., 1999; Marintchev et al., 2009). Plant eIF4F consists of the large scaffolding protein eIF4G and the cap-binding protein eIF4E (Browning, 2004). Unlike host mRNAs, many viral RNAs lack the 5’ m7GTP cap that is normally required for eIF4F to bind the mRNA and yet these viral RNAs are translated efficiently (Doudna & Sarnow, 2007). The 3’ UTRs of many plant viral RNAs contain cap-independent translation elements (CITEs) that facilitate efficient translation initiation at the 5’ end of the RNA (Kneller et al., 2006; Miller et al., 2007).

Among the best characterized and most efficient 3’ CITEs is that of barley yellow dwarf virus (BYDV) and related viruses (Guo et al., 2001). BYDV-like translation elements (BTEs) display diverse secondary structures (Wang et al., 2010), but they share the following
features: (i) a long basal helix (S-IV in the BYDV BTE) from which a variable number of helices radiate, (ii) a 17-nucleotide consensus sequence (17 nt CS), GGAUCCUGGGAAACAGG, that forms a stem-loop topped by a GNRNA pentaloop, (iii) a stable stem-loop (SL-III in the BYDV BTE) topped by a loop that base-pairs to the 5’ UTR (Guo et al., 2001) and (iv) variable numbers of essential but nonconserved non-Watson–Crick paired bases around the junction of the helices (Wang et al., 2010). To facilitate translation, the BTE binds and requires eIF4G but not eIF4E (Treder et al., 2008). To understand how the BTE interacts with the translational machinery, we are determining its structure. Here, we describe the purification, crystallization and preliminary X-ray diffraction analysis of BYDV BTE crystals in our attempt to determine the first three-dimensional structure of a 3’ CITE at near-atomic resolution.

2. Materials and methods

2.1. T7 RNA polymerase purification

T7 RNA polymerase (with a His_{6} tag) was purified by modifying the procedure of Grodberg & Dunn (1988). Briefly, 3 BL21/pT7-911 colonies (Ichetovkin et al., 1997) were suspended in 1 mL of 10mMTris, pH 7.5, 10mM MgSO_{4}. 100 µL of this suspension was added to 500 mL of M9TB expression medium (1% tryptone (v/v), 100 mM NaCl, 40 mM NH_{4}Cl, 44 mM KH_{2}PO_{4}, 100mM Na_{2}HPO_{4}, 0.1mM CaCl_{2}, 1mM MgSO_{4}, 0.2% Glycerol per 500 mL). The culture was grown overnight at 303 K until A_{600} reached 0.5-0.6.

To induce T7 RNA polymerase expression, 0.5 mL of 1M IPTG was added to the bacterial culture and cells were harvested 3 hours later. The pelleted cells were resuspended in 25 mL of lysis buffer (50mM Tris pH8, 100mM NaCl, 5% Glycerol (v/v), 5mM β-
mercaptoethanol, 1mM imidazole) supplemented with lysozyme and 1 protease inhibitor cocktail minitablet (Roche) per pellet. Each pellet was lysed by sonication and centrifuged (15,500 rpm, 30 minutes, 277 K). The supernatant was applied to a Ni-NTA column (Qiagen) equilibrated with 15 mL of lysis buffer. After washing with 15 mL of lysis buffer supplemented with 10mM imidazole, bound protein was eluted with lysis buffer supplemented with 500mM imidazole. Eluted T7 RNA polymerase was dialyzed into storage buffer (20mM KHPO$_4$ pH 7.5, 100mM NaCl, 50% Glycerol (v/v), 1mM DTT, 0.1mM EDTA, 0.2% NaN$_3$).

2.2. RNA purification

Templates to transcribe the BYDV BTE sequence variants used in this study (Fig. 1a) were prepared by PCR using the upstream oligonucleotide containing either class III T7 promoter (TAATACGACTCActATAG) for the BTE RNA transcripts that start with GTP or a T7 class II promoter (Coleman et al., 2004) (TAATACGAGCTCActATTA) for the non-GTP initiating transcripts. The downstream oligonucleotide contained two 2'-O-methyl RNA bases in place of the deoxyribonucleotides at the 5' end to prevent addition of nontemplated 3' nucleotides by the polymerase (Fig. 1b) (Kao et al., 1999). DNA was amplified in a 1 mL PCR reaction [100 ng plasmid DNA p5’UTR-LUC-TE869 (Guo et al., 2000), 0.3 mM dNTPs each, 2.5 mM MgCl$_2$, 0.4 µM primers, 1X High Fidelity PCR buffer and 10 U Platinum Taq DNA polymerase (Invitrogen)] for 30 cycles of 1 min at 367 K, 40 sec at 329 K, and 90 sec at 343 K, and a final extension for 6 min at 343 K.

The gel purified PCR product (0.7 µM) was incubated in 1X transcription buffer (30 mM Tris HCl, pH 8.1, 2 mM spermidine, 0.01% Triton X-100 and 10mM DTT), 25 mM
MgCl₂, 5 mM each NTP, and 50 µg/mL T7 RNA polymerase for 2 hours at 310 K. Following transcription, RNA was phenol-chloroform extracted and precipitated using ½ volume of 7.5 M CH₃COONH₄ and 2 volumes of ethanol. Dried product of a 1 mL transcription reaction was resuspended in 500 µL of nuclease-free water (Ambion) and passed through a Micro Bio-Spin 30 column (Bio-Rad) equilibrated with nuclease-free water. The integrity of eluted RNA sample was verified by 10% denaturing polyacrylamide gel electrophoresis (Fig. 1c).

2.3. Crystallization and data collection

Prior to crystal screening, BTE RNA was heated in water to 348 K followed by slow cooling to 298 K over 3 hours to drive RNA folding to its lowest energy native conformation. The RNA solution was subjected to crystal screening using Natrix and Nucleic Acid Mini screens (NAM) from Hampton Research. 1 µL of 0.3 mM BTE RNA was mixed with an equal volume of reservoir solution and equilibrated against 500 µL of well solution using hanging drop vapor-diffusion at 298 K.

Crystal screening and initial imaging were performed at the Iowa State University X-ray Crystallography Facility. BTE RNA crystals were soaked for 30 seconds in a solution containing 25%(w/v) of glycerol in Natrix 26 solution and flash cooled in liquid nitrogen before data collection. Initial diffraction images were collected from a single crystal plate at the Advanced Light Source (beamline 4.2.2) at 100 K and λ=0.979 Å on a CCD detector. Images were recorded with one second exposure at a crystal-to-detector distance of 200 mm using an oscillation range of 1°. According to analysis with Phaser (McCoy et al., 2007), the
most probable number of molecules in the asymmetric unit is 6, with the Matthews coefficient of 2.55 Å³ Da⁻¹ and solvent content of 51.77%. Diffraction data were processed using XDS (Kabsch, 2010).

3. Results and discussion

3.1. RNA construct design and preparation

To ensure the complete homogeneity of BTE RNA constructs while allowing rapid purification of multiple constructs, we developed a PCR scheme to generate the template for transcription. The upstream primer contained either the common class III T7 RNA polymerase promoter for guanosine-initiating transcripts or a class II T7 promoter for adenosine-initiating transcripts. The class II T7 promoter gives superior 5’ end homogeneity (Coleman et al., 2004) and produced yields comparable to those obtained with the class III T7 promoter. Using a downstream primer containing 2’-methoxy modifications at its 5’ end prevented 3’ end heterogeneity (Kao et al., 1999). This approach allowed fast production of clean, homogeneous, and acrylamide-free RNA samples suitable for structural studies as evidenced by the 5Å diffracting crystals.

Crystalline material obtained from wild type BTE105 RNA did not diffract X-rays. Therefore, we tested variant sequences designed to enhance crystallization, while taking care to maintain BTE functionality. The functional construct 87c BTE RNA (Guo et al. 2000), in which the terminal 24 bases of basal helix were truncated and a GC-rich "clamp" was appended to the ends (Fig. 1a), gave many crystals. Natrix solution 26 (0.05 M sodium cacodylate pH 6.5, 0.2 M potassium chloride, 0.1 M magnesium acetate and 10 % PEG 8,000) yielded plates nucleating from a common center ("RNA flowers") at 298 K two weeks
after set-up (Fig 2a). These BTE crystals diffracted to >25 Å and were subjected to additive screening (Hampton Research) for optimization. Single crystal plates, formed in 4% 1,3-propanediol and 0.025% dichloromethane, diffracted to <10 Å 3-5 days after screen set up (Fig. 2b). Finer additive screening with 1,3-propanediol and dichloromethane where percentages varied in 1% increments from 3% to 10% for 1,3-propanediol and 0.01% increments from 0.005 to 0.065% for dichloromethane yielded large RNA plates (Fig. 2c) that diffracted to <5Å resolution (Fig. 3).

3.2. Crystallization screening and diffraction experiments

Crystal screening of BTE RNA sequences using commercial sparse matrix screens revealed preference for positively charged ions, especially magnesium and potassium, lower percentages of PEG as precipitant, and a wide pH range. It is no surprise that magnesium was a critical ion in all conditions that generated BTE RNA crystals, because magnesium often facilitates ion-dependent RNA folding by neutralizing negatively charged phosphates (Woodson, 2005). The best diffracting 87c BTE RNA crystals formed large thin plates (>0.4mm in length and 0.01mm in diameter) in the presence of small organic molecules within 3 days that diffracted consistently to below 5 Å.

The large 87c BTE crystal plates were fragile and difficult to cryoprotect, resulting in radiation damage that weakened diffraction below 7 Å resolution. Despite the crystal damage, we processed data using the XDS program package with statistics summarized in Table 1. To generate better diffracting crystals, we will generate more crystallizable BTE RNAs by slight variation of the BTE sequence. The promise of this approach is shown by the significant improvement we obtained by altering BTE105 to 87c BTE. Also, native
PAGE analysis revealed that 8 mM magnesium chloride relieves conformational heterogeneity of 87c BTE and makes the BTE adopt a more compact conformation as indicated by its faster mobility (Fig 1d, lane 2). Hence, to ensure structural homogeneity of RNA, we are modifying the folding step by dialyzing purified RNA into buffer containing 8 mM magnesium chloride and 10 mM HEPES KOH pH 7.5 prior to dilution with reservoir solution.

In addition to optimizing the BTE crystal growth, we are pursuing different conditions of dehydration and crystal annealing to extend the diffraction limit of crystals we have generated so far. Preliminary cryo-screening using CryoPro kit (Hampton Research) indicated that 10-30% (v/v) MPD causes no crystal damage, even after prolonged 60 min incubation.

For phasing, we will soak RNA crystals with heavy atom lanthanide series compounds, which bind RNA at magnesium sites (Cate et al., 1996; Feig et al., 1998), and crystallize similarly to native RNA. Because there is no certainty that the metal will bind at the desired sites, we may position G-U pairs in helices to form a cation binding pocket lined with negatively charged phosphate groups (Keel et al., 2007). The BTEs already contain several G•U pairs surrounded by Watson-Crick pairs that could be mutated to fit the consensus phasing sequence (Keel et al., 2007; Edwards et al., 2009) while maintaining the structure and function of the BTE prior to heavy atom screening.

This research may provide fundamental insight on mechanisms of translation factor - and ultimately ribosome - recruitment by eukaryotic mRNAs. We are encouraged by initial incremental success in crystallizing the native RNA and therefore plan to pursue phasing and higher resolution data collection for structure determination.
ACKNOWLEDGEMENTS

This work was supported by NIH grant R01 GM067104 with an American Recovery and Reinvestment Act supplement. We thank Jay Nix at the Advanced Light Source of Lawrence Berkeley National Laboratory for assistance with data collection, Jennifer Doudna for advice, Mark Hargrove for laboratory support, and Jens Kaiser for assistance with data processing.
REFERENCES


FIGURE LEGENDS

**Figure 1.** (a) Secondary structure of construct 87c BTE used in crystallization trials is shown with deleted wild type bases in gray and base substitutions in red. The 17 nt conserved sequence (17 nt CS) is highlighted in italics. Bold bases participate in the long-distance base pairing to the 5' UTR. (b) Production of homogeneous T7 RNA polymerase transcripts by PCR, fusing a primer with modified 5'-terminal bases, followed by transcription with T7 polymerase. (c) 10% denaturing PAGE of purified BTE transcripts. Lane M, Low Range ssRNA ladder (New England Biolabs); lane 1, BTE105 RNA; lane 2, 87c BTE. (d) Folding of 87c BTE assessed by 10% native PAGE containing 10mM HEPES KOH pH 7.5 and either 0 mM (lane 1) or 8 mM (lane 2) magnesium chloride.

**Figure 2.** Typical crystals of 87c BTE obtained using Natrix 26 (0.2 M potassium chloride, 0.1 M magnesium acetate tetrahydrate, 0.05 M sodium cacodylate trihydrate pH 6.5, 10 % w/v polyethylene glycol 8,000) and (a) no additive (dimensions of 0.54 x 0.08 x 0.01 mm); (b) 6% 1,3-propandiol (dimensions of 0.2 x 0.02 x less than 0.01 mm in diameter); (c) 0.045% dichloromethane (dimensions of 0.4 x 0.06 x 0.01 mm).

**Figure 3.** X-ray diffraction pattern for the 87c BTE crystals showing diffraction extending to approximately 6.9 Å.

**Table 1** Data-collection statistics.

Values in parentheses are for the highest resolution shell.
Fig. 1. A diagram showing the T7 promoter, PCR, RNA synthesis, and 2'-O-methyl RNA bases. (a) and (b) are sequences of DNA and RNA, respectively. (c) and (d) are gel images of the experiment results.
Fig. 2.
Fig. 3.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>MBC 4.2.2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.979</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td>a = 316.6, b = 54.2, c = 114.5, α = γ = 90, β = 105.1</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>10409</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>3157</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>21.69–6.86 (7.27–6.86)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.5 (87.3)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.3 (3.0)</td>
</tr>
<tr>
<td>( R_{\text{merge}} )†</td>
<td>0.094 (0.54)</td>
</tr>
<tr>
<td>Average ( I/\sigma(I) )</td>
<td>8.4 (1.9)</td>
</tr>
</tbody>
</table>

\[ R_{\text{merge}} = \sum_{hkl} \sum_{i} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_{i} I_i(hkl). \]

† \( R_{\text{merge}} \) is the merging R-factor.
CHAPTER 3: CATION-INDUCED STRUCTURE OF CAP-INDEPENDENT TRANSLATION ELEMENTS IN THE 3’ UNTRANSLATED REGION OF PLANT RNA VIRUSES FACILITATES INTERACTION WITH EUKARYOTIC TRANSLATION INITIATION FACTOR 4F (eIF4F)

A paper prepared for submission to the Journal of Molecular Biology

Jelena J. Kraft, Krzysztof Treder, Mariko Peterson and W. Allen Miller

Author contributions: JJK performed all the experiments, KT purified all the recombinant protein and assisted with determination of the Hill coefficient of the magnesium-RNA association, MP assisted with RNA purification. JJK planned experiments, analyzed the data and wrote the paper.

ABSTRACT

The 3’ untranslated regions (UTR) of many plant viruses contain heterogeneous cap-independent translation elements (CITEs) that drive translation initiation at the 5’ end of the mRNA in the absence of a 5’ cap. CITEs differ in primary sequence, structure and the component of the eIF4F complex with which they interact. The Barley yellow dwarf virus-like CITE (BTE) binds and requires the eIF4G subunit of translation initiation factor eIF4F to stimulate translation. To understand this interaction, we characterized the dynamic structural properties of the BYDV BTE and mapped the eIF4G binding sites. Here we show that the BTE folding involves cooperative uptake of magnesium ion and it is driven primarily by
charge neutralization. Footprinting experiments revealed that eIF4G alone is capable of protecting most of the BYDV BTE structure from chemical and enzymatic modification. This BTE protection pattern is more evident in the presence of the heterodimeric eIF4F complex, supporting previous observations that eIF4E enhances eIF4G binding to the BTE. However, the BTE’s overall fold remained the same, suggesting that the BTE forms a stable platform for protein factor binding prior to protein recognition. Furthermore, the loop involved in long distance base pairing to the 5’ UTR remained solvent accessible even at high (1µM) eIF4G/4F concentrations. These data fit the proposed model in which the eIF4F complex binds directly to the BTE and, is delivered to the 5’UTR via a long distance kissing interaction, where it then recruits the 40S subunit to the 5’ end. These observations provide new insight into translation factor-mRNA interactions that effect efficient translation of mRNA.

**Keywords.** Barley yellow dwarf virus, translation initiation, translation enhancer, RNA structure.

**Abbreviations used.** CITE, cap-independent translation element; BTE, Barely yellow dwarf virus-like cap-independent translation element; BYDV, Barley yellow dwarf virus; eIF, eukaryotic initiation factor; 17 CS, the highly conserved 17-nucleotide sequence; IRES, internal ribosome entry site; RCNMV, Red clover necrotic mosaic virus; RSDaV, Rose spring dwarf-associated virus; TNV-D, Tobacco necrosis virus D; WGE, wheat germ extract;

**INTRODUCTION**

The efficiency of translation initiation is dictated by the sequence and structural elements on the mRNA (Kozak, 2002, Dreher and Miller 2006). Cellular mRNAs contain a 5’ m7GpppN cap structure that is co-transcriptionally added to the nascent mRNA. This
structure stabilizes mRNA and assists in recruiting the 43S ribosomal subunit complex to the mRNA via binding eIF4F (Haghighat, 1997, Pestova and Kolupaeva, 2002). Like the 5’ untranslated region (UTR), the 3’ UTR contains sequences that enhance mRNA translatability and stability. A 3’ poly(A) tail and its binding partner, poly(A)-binding protein (PABP), interact with the 5’ located protein factors to circularize mRNA and to keep eIF4G in close proximity and available to the translation machinery (Jackson et al., 2010). Message circularization through protein factors is thought to stimulate ribosome recycling by allowing the ribosome to reinitiate another round of translation as well as to prevent translation of truncated mRNAs (Wells et al., 1998).

Unlike cellular mRNAs, the RNA genomes of many plant viruses lack a 5’ cap. Instead, they contain cap-independent translation elements (CITEs) that differ from mammalian internal ribosome entry sites (IRESes) in that they are located in the 3’ untranslated region (UTR) and translation requires scanning from the 5’ end. These elements are classified into about seven distinct classes with little conservation of the sequence or secondary structure (Miller et al., 2007). Among these is well characterized CITE of Barley yellow dwarf virus (BYDV). The BYDV and other BYDV-like translation elements (BTEs) show remarkable diversity of secondary structure with few conserved features. All BTEs contain a highly conserved 17-nucleotide tract (17 CS) GGAUCCUGGGAAACAGG and a stable stem loop structure (loop 3 in BYDV) capable of base paring to complementary sequences in viral 5’ UTR. Other similarities include a number of nonconserved single-stranded bases around the BTE helical junction that links together radiating helices from the central hub, and all BTEs contain a basal helix that connects the BTE to the viral genome (Guo et al., 2000, Wang et al., 2010).
All CITEs function by recruiting and binding to a component of translation initiation machinery. The BYDV BTE binds and requires the translation initiation factor eIF4F (Treder et al. 2008 RNA), which in plants is composed of the large eIF4G scaffolding protein and the smaller cap-binding eIF4E (Gallie, et al., 2007). More precisely, full length eIF4G or a fragment of eIF4G lacking the eIF4E-binding site stimulated translation of uncapped reporter construct containing the 3’BTE. eIF4E additionally enhanced both the translation stimulatory activity of full-length eIF4G activity and the affinity of eIF4G for the BTE (Treder et al., 2008). However, the precise mechanism of eIF4F binding to the BYDV BTE remains to be determined. Furthermore, it is not known with what protein factors the BTEs in other viral genera interact to bring about efficient translation initiation.

Since the BTE function is inevitably tied to its unique three-dimensional structure, we sought to understand how the BTE folds and utilizes its structural features to recognize and bind to initiation factors and direct them to the 5’ end where initiation ensues. We also explored the diversity of other BTE sequences and secondary structures in Luteovirus, Necrovirus and Dianthovirus genera to see if they adopt and utilize structural features to recognize initiation factors in a similar way. Using structure probing, native polyacrylamide gel electrophoresis and footprinting methods we show that the BTE folding involves cooperative uptake of magnesium ion and it is driven primarily by charge neutralization. Point mutations that disrupt compact native state also impaired the BTE’s ability to support cap-independent translation. Footprinting experiments revealed that eIF4G alone is capable of protecting most of the BTE structure from chemical and enzymatic modification. This BTE protection pattern is more exaggerated in the presence of heterodimeric eIF4F complex confirming previous observations that eIF4E enhances eIF4G binding to the BTE. However,
the BTE’s overall fold remained the same suggesting that the BTE forms a stable platform for protein factor binding prior to protein recognition. Furthermore, the loop involved in long distance communication remained solvent accessible even at high (1µM) eIF4G/4F concentrations. These data fit the proposed model in which the eIF4F complex binds directly to the BTE and once recruited, is delivered to the 5’UTR via a long distance kissing interaction. The behavior is reiterated in the BTEs from two other viral genera suggesting shared mechanism of action of this class of CITEs.

RESULTS

Cation dependence of BTE folding. To understand how BTE folds, we monitored the mobility of the minimal functional BYDV 105 BTE RNA (Guo et al., 2000, Raktondrafara et al., 2006) on native polyacrylamide gel in the presence of increasing Mg$^{2+}$ concentrations as a direct measure of BTE’s ability to adopt a single compact species. In the absence of Mg$^{2+}$, 105 BTE adopts multiple non-native conformations (nn) manifested as a dispersed band of retarded mobility (TBM gels, lane 1, Fig. 1 (c)) suggesting absence of one specific three-dimensional conformation. Upon addition of 2.5 mM Mg$^{2+}$, the number of stable alternative conformations decreased significantly with complete folding to native state (N) observed in the presence of 10 mM Mg$^{2+}$ (TBM gel, lane 1, Fig 1 (c)). To determine if this cation-induced BTE folding specifically requires magnesium, we monitored BTE’s mobility on native PAGE in the presence of monovalent cations. Incomplete BTE folding was observed at lower K$^+$ concentrations, but at high 500 mM K$^+$ concentrations the BTE adopts single native conformation (TBK gel, lane 1, Fig. 1 (d)) indicative of folding driven by charge neutralization. The requirement for significantly higher monovalent cation
concentrations compared to divalent ions for RNA folding was observed for the *Tetrahymena*
group I ribozyme (Heilman-Miller et al., 2001) and the Hepatitis C virus IRES (Kieft et al.,
1999) and stems from higher charge density of Mg$^{2+}$ leading to better neutralization of the
RNA phosphate charge (Woodson, 2005).

**BTE helical junction mutants destabilize the compact state.** A major question
corns folding at the complex 4-way junction (4wj) in the BTE that has many non-Watson-
Crick paired bases. If cap-independent translation activity of the BTE were dependent on
formation of compact native structure, then point mutations that destabilize the compact fold
would destabilize BTE’s function. To address this, we perturbed the BTE helical junction by
mutating each unpaired base singly into each of the three other possible bases, taking care not
to introduce a start codon in the middle of the BTE. These 4wj mutants were added in trans
to a wheat germ reaction containing luciferase reporter BLucB mRNA which has the BYDV
genomic RNA 5’ and 3’ UTRs. We then tested for ability of mutant BTEs to inhibit
translation, presumably by binding and sequestering rate-limiting eIF4F. At 100-fold excess
over BLucB mRNA, only 4wj point mutants A4868C and U4900C were able to reduce
translation of BLucB mRNA to 20% suggesting that these point mutations did not alter the
BTE’s function. This 20% level of trans-inhibition was observed for the wild type 105 BTE
and for comparison purposes, was defined as 100% inhibition. Our results were normalized
to the 105 BTE and presented in figure 1 (a) and (b) as a fraction of this inhibition. Thus, the
functional mutants A4868C and U4900C that inhibited translation to 20% of the original
level are presented as 100% inhibitor activity.

To correlate BTE function with the ability of various 4wj mutants to adopt native
compact fold (N) we compared their mobilities on native gel in the presence of increasing Mg$^{2+}$ concentrations. All analyzed RNA species displayed structural heterogeneity at the low Mg$^{2+}$ concentrations that was relieved completely in presence of 10 mM Mg$^{2+}$ (TBM gels, Fig. 1 (c)). However, the level of compaction of nonfunctional BTE mutants did not match that of the 105 BTE and functional 4wj A4868C and U4900C mutants at lower 0-2.5 mM Mg$^{2+}$ concentrations (TBM gels, Fig. 1 (c)). One possible explanation is that addition of Mg$^{2+}$ leads to global change in shape as a result of formation of secondary and some tertiary contacts, but nonfunctional mutations probably disrupted local packing of the RNA backbone resulting in their inability to adopt a compact fold. To determine whether the differential mobilities of functional and non-functional BTEs are the result of disrupted native state, the 105 BTE and selected functional and nonfunctional RNAs were electrophoresed under denaturing conditions and in the presence of chelating agent. All analyzed RNAs ran as a single sharp band with the same mobility under denaturing conditions (dTBE gel, Fig. 1 (e)) indicating that the retarded mobilities of nonfunctional mutants are due to the disruption of the BTE native fold which, in turn, impaired the ability of the BTE to support cap-independent translation initiation. Like the 105 BTE, 4wj mutants were able to adopt a single native fold at 0.5 mM K+ conditions, confirming that charge-neutralization drives the BTE folding to native state (TBK gel, Fig. 1(d)).

**Cooperativity of the BTE folding.** To complement the native gel analysis and further investigate the role of Mg$^{2+}$ in the BTE folding, we probed the BYDV BTE structure in solution using Selective 2’-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) technology (Wilkinson et al., 2006). The SHAPE chemistry interrogates conformation of
each base independently of its sequence and if the nucleotide is unpaired or conformationally unconstrained it will react with SHAPE reagent to form a bulky 2’-O-adduct that will prevent primer extension by reverse transcriptase (Wilkinson et al., 2006, Mortimer and Weeks, 2007). Specifically, we used fast-acting benzoyl cyanide reagent (BzCN) (Mortimer and Weeks, 2008) to monitor the BTE folding in the presence of increasing Mg\(^{2+}\) concentration. If the BTE folding involves cooperative uptake of Mg\(^{2+}\) then we could expect widely separated regions of sequence to become modified at similar Mg\(^{2+}\) concentrations. Indeed, quantification of the observed BzCN-induced modifications at two remote exposed regions of the BYDV BTE sequence, universally conserved nucleotide G4837 and nucleotide C4899 (filled and open triangle, respectively Fig. 2(a)), showed an increase in modification at increasing magnesium concentrations. Calculated Hill coefficients for the modification at nucleotide G4837 and C4899 were 1.72 and 1.88 respectively (filled and open triangle, Fig. 2(b)). This cooperative nature of magnesium uptake was also confirmed using enzymatic probing with nuclease V1, which preferentially cleaves helical or stacked base regions (Lowman and Draper, 1986). Increase in V1 cleavage at nucleotides A4832 and A4868 was plotted as a function of Mg\(^{2+}\) concentration and calculated Hill coefficients were 1.6 and 1.92 respectively, indicating positive cooperative binding of Mg\(^{2+}\) to the BTE (filled and open square, respectively Fig. 2(b)). A slight difference in Hill coefficients calculated for BzCN-mediated modification versus RNase V1cleavages most likely reflects different chemistries of two reagents used in structure probing experiments. Most of the secondary structural elements were formed at 2.5 mM with complete folding observed at the 10 mM magnesium ion concentration. Higher Mg\(^{2+}\) concentrations of up to 40 mM showed no additional structural changes. Functional 4wj U4900C mutant showed near identical Hill coefficient of
1.7 at conserved G4837 compared to 105 BTE RNA indicating that mutation did not alter structure formation in any significant way (Fig. 2(d)). In contrast, disruption of the BTE helical junction in non-functional 4wj U4900G mutant resulted in lowered Hill coefficient by 20% compared to wild-type BTE (Fig. 2(d)). Further more, this non-functional 4wj mutant RNA required higher magnesium concentration for formation of most secondary structural elements and the Mg$^{2+}$ concentration for half-maximal modification (half-folding) was calculated to be 5.5 mM compared to 3.3 mM for functional U4900C RNA (Fig. 2(c)). Similar trend was observed for Azoarcus ribozyme, where point mutation in tetraloop motif destabilized the compact state leading to increase requirement for Mg$^{2+}$ for folding (Chauhan and Woodson, 2008).

**RNA elements in three viral genera also fold cooperatively.** To see if cooperative uptake of Mg$^{2+}$ upon folding is general behavior of this class of CITEs, we probed the BTE structure of Rose spring dwarf-associated luteovirus (RSDaV), Tobacco necrosis Necrovirus-D (TNV-D) and Red clover necrotic mosaic dianthovirus (RCNMV) with BzCN in the presence and absence of Mg$^{2+}$. Like the BYDV BTE, all probed BTEs underwent a Mg$^{2+}$ dependent conformational change at similar 2 mM Mg$^{2+}$ concentrations except RSDaV BTE (Fig. 3(a)). This RNA appears to adopt most of the secondary and some tertiary structure at low 0.5 mM Mg$^{2+}$. However, global compaction of structure was observed only at high 20 mM Mg$^{2+}$ indicated by the decrease in BzCN mediated modification at increased Mg$^{2+}$ concentrations (Fig. 3 (a) and (b)). This is not surprising since RSDaV BTE sequence does not match the 17 nt CS perfectly and it contains a large number of unpaired bases that likely require more Mg$^{2+}$ for complete folding. Also, the RSDaV BTE is a less efficient translation
stimulator than the BYDV or TNV BTEs (Wang et al., 2010). Uptake of Mg\(^{2+}\) upon TNV-D and RCNMV BTE folding is strongly cooperative with the Hill coefficients for the modification at universally conserved G residue in TNV-D G3591 (square) of 2.85 and RCNMV G3608 (circle) of 3.6 suggesting presence of multiple Mg\(^{2+}\) binding sites (Fig. 3(c)). Hence, all BTEs with perfect 17nt CS adopt stable compact native fold at physiological Mg\(^{2+}\) concentrations in the absence of protein factors.

**The BTE interaction with the initiation factor eIF4G.** We also investigated the physical interaction between the BTE and initiation factors to determine the sites on the BTE bound by protein factor and if this binding promotes further folding of the BTE. We found that whole eIF4G and smaller N-terminally truncated forms of eIF4G that are known to facilitate BTE-mediated translation (p86 and p100) showed similar BTE protection patterns from modification by BzCN or nucleases. Most nucleotides in the BTE were protected from chemical modification with the exception of loop 3 (lane 10, Fig. 4(a)). Loop 3 was expected to remain solvent-accessible because we showed previously it must base pair to the 5’ UTR for translation (Rakotondrafara et al., 2006). Full-length eIF4G, but not fully functional p100, caused a broad reduction in band intensity in the footprinting assays that we cannot explain. Incubation of the BTE and full eIF4G resulted in shifted band with retarded mobility on native gel compared to free RNA indicative of complex formation (data not shown). No other degradation products were observed on gel, thus ruling out RNase activity as a possible cause of broad reduction in band intensity. Addition of eIF4E enhanced protection in the presence of the only two forms that contain the eIF4E binding site (wt and p100) suggesting that eIF4E binding to eIF4G increases binding affinity of eIF4G to the BTE.
(Fig. 4(b)). This agrees with our previous observation that eIF4E enhances eIF4G stimulation of BTE-mediated translation by 20-30% in (Treder et al. RNA, 2008). As expected, negative controls (lanes 4-7, Fig. 4 (a)) containing no protein, eIF4E (4E), ovalbumin (O) or p70, all known not to bind the BTE, showed no changes in modification pattern. Functional relevance of this interaction is supported further by significantly reduced protection of nonfunctional mutant RNA, BTE BF, from modification by eIF4G or deletion mutants (BTE BF, Fig. 4(a)). Furthermore, loop 3 involved in long distance communication was partially protected from chemical modification in non-functional BF. This data strongly supports our model in which the BTE simultaneously binds eIF4F and base pairs to the 5’ UTR via loop 3, in order for eIF4F to recruit the ribosome to the 5’ UTR.

Next, we focused on analyzing the interaction between other luteovirus, necrovirus and dianthovirus BTEs and eIF4F. To ensure that all tested BTEs sequences are functional cap-independent translation elements we assayed their ability to trans-inhibit translation of reporter BlucB mRNA in the presence and absence of cap structure. A 200-fold excess of RSDaV<sub>5173-5302</sub>, TNV-D<sub>3573-3664</sub> and RCNMV<sub>3590-3740</sub> (numbering is according to position occupied in the viral genome) reduced translation of BLucB mRNA to 25% (Fig 5(a)). Capped versions of the same elements did not significantly improved trans–inhibitory activity of tested BTEs confirming that these viral sequences function as cap-independent translation elements.

To visualize this RNA-protein interaction at nucleotide level, we probed these BTEs with BzCN reagent in the presence and absence of eIF4G. In all BTEs, non-functional p70 and eIF4E had no effect on modification pattern of any of the nucleotides (Fig. 5 (c)). However, the addition of p100 induced protection at similar regions as the BYDV BTE.
Like in the BYDV BTE, the addition of eIF4E further exaggerated p100 induced protection. These results highlighted the shared BTE characteristics such as direct binding to eIF4G subunit of eIF4F complex, eIF4G binding enhancement by eIF4E and solvent accessibility of loop involved in long-distance communication.

**DISCUSSION**

Viral mRNAs usually adopt folded shapes to recruit specific host factors and facilitate efficient translation of viral mRNA at the expense of host protein synthesis. Here, we explored the diversity of sequences and secondary structures of CITEs found in 3’UTR of *Luteovirus, Necrovirus and Dianthovirus* genera of plant viruses to see how they arrange themselves prior to and after the interaction with translation machinery to stimulate cap-independent translation initiation. Using a combination of functional and structural assays, we show that all BTEs adopt a stable cation-dependent fold in the absence of the initiation factors. Mutations at the BTE helical junction not only decrease the cooperativity of base pairing with respect to Mg$^{2+}$ concentration, but they also impair ability of this RNA to bind initiation factors (Fig. 1 (a) and (b)). Thus the stability of the native state is directly related to ability to recruit and bind to components of the initiation machinery.

At very low (0.1 mM) and up to 2.5 mM Mg$^{2+}$ concentrations, only wild-type 105 BTE and two functional 4wj A4868C and U4900C RNAs predominantly adopt a compact native fold, while point mutants in the helical junction that greatly reduced BTE function populated non-native conformations that migrated with retarded mobility on native gel (Fig. 1 (c)). In functional point mutant U4900C, replacement of uridine with cytidine converts a weak guanine-uracil (G-U) wobble pair into a stable Watson-Crick guanine-cytosine (G-C) base
pair that probably stabilizes base of helix 3. The effect of replacing adenine with cytosine in
the other functional mutant A4868C cannot be explained by canonical Watson-Crick type of
interactions. One possibility is that the replacement of adenine with cytosine results in the
formation of a wobble type of cytidine-adenine (C-A) base pairing observed in high-
resolution crystal structure of octameric RNA (Jang et al., 1998). Like in the other functional
mutant, such C-A pairing would stabilize the base of helix 2.

At elevated (10 mM) Mg$^{2+}$ concentrations, electrophoretic migration rates of functional
and non-functional 4wj mutants were equal (Fig. 1 (c)), suggesting that even non-functional
4wj mutants collapsed to compact near-native fold. Such behavior was noted for the bi5core
RNA of a group I intron (Buchmueller et al., 2000) and the Azoarcus ribozyme (Rangan et
al., 2003). In the case of bi5core RNA it was proposed that Mg$^{2+}$ mediated compact but non-
native states are biologically relevant intermediates, but that bi5core RNA achieves its
complete native fold only in presence of protein (Buchmueller et al., 2000). This is not the
case with non-functional 4wj mutants because they were not able to bind and sequester
initiation factors from reporter mRNA in noncompetitive condition of wheat germ lysate.
We conclude that non-native compact BTE states are not biologically relevant states and
probably are gel-matrix arrested destabilized near-native intermediates as observed for
Azoarcus ribozyme (Rangan et al., 2003).

There are two possibilities for how Mg$^{2+}$ might stabilize the compact native fold of the
BTE. One possibility is the specific coordination of Mg$^{2+}$ ions to the RNA structure as
observed in crystal structures of the Tetrahymena group I intron (Cate et al., 1997; Juneau et
al. 2001) and the other involves Mg$^{2+}$ diffuse binding to neutralize RNA phosphates as
observed for HCV IRES and tRNA$^{Phe}$ (Kieft et al., 1999; Misra et al., 2002). To differentiate
between these possibilities, we ran 105 BTE and selected functional and non-functional 4wj mutants on native gel supplemented with half-millimolar K$^+$ concentrations (TBK gel, Fig. 1(d)). All analyzed mutants migrated as single RNA species indicating that the BTE RNA folding is driven by neutralization of the negative phosphate charges of RNA backbone. Requirement for 50-fold greater K$^+$ ion concentration compared to Mg$^{2+}$ to complete the BTE folding is likely a consequence of it’s larger size and lower charge density compared to Mg$^{2+}$, leading to less effective neutralization of the phosphate charges. If the BTE fold required specific coordination of Mg$^{2+}$, than K$^+$ would not be able to replace Mg$^{2+}$ in our folding native gel studies. However, in the absence of high-resolution structure of the BTE we cannot rule out possibility of specific coordination of Mg$^{2+}$ to the BTE structure.

To investigate the ion-induced folding in terms of secondary structure formation we probed the structure of functional and non-functional BYDV BTEs as well as BTEs from other viral genera by chemical modification with benzoyl cyanide (SHAPE) reagent and by partial digestion with RNase V1 as a function of Mg$^{2+}$ concentration. We found that most secondary structural elements are formed at physiological 2.5 mM Mg$^{2+}$ concentrations, however structural rearrangements in the BTE persisted until 10 mM Mg$^{2+}$ particularly for RSDaV BTE. The higher cation requirement for RSDaV BTE folding is probably the result of the inherent flexibility introduced by the large number of unpaired bases. Furthermore, BzCN-mediated base modification showed that base-pairing forms less cooperatively in the non-functional 4wj mutants compared to functional RNAs, suggesting that the destabilization of the BTE helical junction most likely introduces an alternative folding pathway. Similar folding traps were observed for Azoarcus ribozyme where the disruption of tetraloop-tetraloop receptor interaction led to the formation of slower moving non-native species on
native gel electrophoresis and required higher Mg$^{2+}$ concentrations to form base pairs compared to wild type ribozyme (Chauhan et al., 2008). Most exposed regions of the BTE sequence displayed an increase in the modification as a function of Mg$^{2+}$ concentration and quantification of these folding changes for two distant regions of the sequence nearly overlaps, suggesting that the BTE folds in a single step stabilized by cooperative binding of Mg$^{2+}$ in diffuse manner (Fig. 6).

Observed BTE folding correlated strongly with the ability to bind to initiation factor eIF4F and stimulate translation initiation. Mutants with defective helical junction and fold were not able to compete with reporter mRNA for the initiation machinery (Fig. 1). Mapped protein binding regions indicate that except for the long-distance communication loop, most of the nucleotides in BTE were protected from chemical modification by whole eIF4G. This loop (loop 3 in BYDV BTE) was expected to remain solvent-accessible because we showed previously it must base pair to the 5’ UTR for translation (Rakotondrafara et al., 2006). The predicted kissing loops in the other tested BTEs, loop 3 in TNV-D and RSDaV and loop 2 in RCNMV BTE were also solvent accessible in the presence of eIF4F, suggesting shared mechanism of translation machinery recruitment and delivery.

It is interesting to compare the BTE with completely unrelated cap-independent element of Hepatitis C virus (HCV) IRES. Like the BTE, the HCV IRES prefolds into a distinct cation-dependent fold prior to recognition by the translation initiation machinery and mutations that disrupt IRES folding also disrupted its function (Kieft et al., 1999). Like the BTE’s long-distance communication loop, structure probing of HCV IRES highlighted the loop with the embedded start codon that remained solvent accessible so that it can readily interact with translation machinery. Thus, it appears that these unrelated CITEs evolved
diverse structures in a similar way capable to accomplish the same task of mobilizing and activating the translation machinery. We propose that BTE class of CITEs preorganize in specific shapes that allow simultaneous eIF4F binding and base pairing to the 5’ UTR via loop 3, in order for eIF4F to recruit the ribosome to the 5’ UTR.

**MATERIALS AND METHODS**

**Plasmids.** Preparation of the individual BTE SHAPE cassettes used for structure probing was described in Wang et al. 2010, except for the BTE BF SHAPE. This construct containing GATC insertion in the middle of the BYDV BTE was prepared by BamHI cutting, Klenow-filling and re-ligation at the BamHI site of the BTE SHAPE construct. BLucB is the reporter plasmid containing the firefly luciferase gene flanked by the BYDV genomic 5’ and 3’ UTRs (Guo et al. 2000). For trans-inhibition studies, short BTE templates containing an upstream T7 RNA polymerase promoter sequence and BYDV 4814-4918 nt, TNV-D 3573-3664 nt, RSDaV 5173-5302 or RCNMV 3590-3740 genome segments were prepared by a PCR amplification.

**RNA synthesis and purification.** Smal-linearized BTE SHAPE cassettes and PCR-generated BTE templates were transcribed using Megashortscript kit (Ambion). Smal linearized BLucB was transcribed using MEGAscript (Ambion), while capped RNAs were transcribed from PCR generated short BTE templates using T7 Message mMachine (Ambion). All transcripts were purified by phenol/chloroform extraction and ethanol precipitation. RNA concentrations were determined spectrophotometrically and integrity was verified by 0.8% agarose gel electrophoresis.
Trans-inhibition of translation. Nonsaturating amounts (4 nM) of uncapped BLucB transcript pre-mixed with 400 nM or 800 nM of the designated viral BTEs were translated in wheat germ extracts (Promega) for 1 hour at 22 °C as described elsewhere (Rakotondrafara et al., 2006; Treder et al., 2008, Wang et al., 2010). The luciferase activity was estimated by the addition of 2 µL of the translation reaction to 40 µL of the Luciferase Assay Reagent (Promega) followed by GloMaxTM20/20 luminometer measurements.

Polyacrylamide gel electrophoresis. 30 pmols of the indicated BTE 4-way junction RNAs were heated at 92° C for 2 min and immediately cooled on ice. To these solutions, triple dye loading buffer (National Diagnostics) was added to final 1X concentration and RNA was analyzed separately by electrophoresis on 10% non-denaturing polyacrylamide gel (29:1 acrylamide to bisacrylamide ratio) containing 1X TBM buffer (89 mM Tris, 89 mM boric acid and 0 to 10 mM magnesium chloride) or 10% denaturing polyacrylamide gel (19:1 acrylamide to bisacrylamide ratio) containing 1X TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) at 4 °C. RNA was electrophoresed for 4 h at 15 W and then visualized by UV transillumination following ethidium bromide staining.

Recombinant protein expression and purification. His-tagged eIF4E in pET23d vector was introduced into E. coli (BL21 cells) and expression was induced, at OD600nm = 0.8, with 100 mM IPTG. Four hours after induction, cells were harvested from 1 L of culture by centrifugation at 10,000 x g for 10 min. The cells were frozen at -80°C for at least 1 h and sonicated 12 times for 30 sec each with 2 min cooling on ice in binding buffer (25 mM
HEPES-KOH at pH 7.6, 100 mM KCl, 2 mM MgCl₂, 10% glycerol plus 0.1 mM phenylmethyl-sulphonyl fluoride, 0.1% Soybean trypsin inhibitor, and 1 tablet/10 mL of Complete protease inhibitor cocktail, EDTA-free [Roche]). The homogenate from 1 L of cells was centrifuged at 38,000 x g for 20 min at 4°C and supernatant was applied to 1 mL of Ni-NTA Superflow Cartridge (Qiagen). The cartridge was washed with 10 vol of binding buffer plus 10 mM imidazole and then with 10 vol of binding buffer plus 20 mM imidazole. The his-tagged proteins were eluted with 250 mM imidazol in the same buffer.

Recombinant (wild-type) wheat eIF4F was expressed from dicistronic constructs in a pET3D vector encoding eIF4G and eIF4E from wheat, and purified as described (Mayberry et al. 2007). The dicistronic plasmids were introduced into E.coli (BL21 cells) and induced with 0.1 M IPTG. Four hours post-induction, cells were harvested by centrifugation and sonicated prior to purification. The lysates were loaded onto a phosphocellulose column, followed by a m⁷GTP sepharose affinity column, and lastly, the protein was concentrated on a second phosphocellulose column. The proteins were dialyzed against N’-100 (25 mM Hepes-KOH at pH 7.6, 100 mM KCl, 1 mM MgCl₂, 1 mM DTT) to remove excess m⁷GTP, and concentrated on Microcon YM-10 (Amicon) with three changes of N9-100. Recombinant scaffold proteins were expressed from pET3d harboring wheat eIF4G (Mayberry et al. 2007) and eIFiso4G (van Heerden and Browning 1994) and purified on a phosphocellulose column and centrifuged through Microcon YM-100 (eIF4G) or Microcon YM-50 (eIFiso4G). Expression and purification of p86 and p70 were performed as for eIF4G, followed by an additional step on 1 mL of Ni-NTA Superflow Cartridge (Qiagen) as described for his-tagged eIF4E. The purity of all proteins was verified by SDS-PAGE and
Coomassie Brilliant Blue staining and concentration determined by Bradford assay (BioRad Protein Assay).

**RNA structure probing and footprinting.** Chemical and enzymatic RNA structure probing was performed as described (Wang et al., 2009; 2011). Briefly, 500 ng of refolded RNA alone or preincubated with indicated proteins was treated with 10% (v/v) of benzoyl cyanide (Sigma-Aldrich) and incubated for 30 sec at 22 °C. Enzymatic structure probing was performed the same way except that 2 µg of yeast tRNA was premixed with the RNA or RNA-protein complex prior to addition of RNase V1 (0.03 units) (Ambion). The reactions containing RNA alone or RNA-protein complex were stopped by the addition of 0.6 M sodium acetate followed by phenol-chloroform extraction and ethanol precipitation. Reactions were resolved on an 8% denaturing polyacrylamide gel and dried following primer extension. Dried gels were exposed to a storage phosphor screen and the band intensity was quantified using Phoretix 1D software.

**ACKNOWLEDGEMENTS**

This work was supported by NIH grant R01 GM067104 with an American Recovery and Reinvestment Act supplement. We thank Dr. Zhaohui Wang for construction of the BTE SHAPE cassettes and advice on RNA structure probing methods.
REFERENCES


FIGURE LEGENDS

Figure 1. Correlation of BTE mutant activity and effect of magnesium chloride on mutant BTE folding. (a) Relative trans-inhibition translation levels of 4wj BTE mutants mapped onto the BTE secondary structure (b) Trans-inhibition of translation by BTE mutants. 4 nM BLucB mRNA was incubated in the presence of 400 nM 4wj mutant RNAs in a wheat germ extract (WGE) and luciferase activity was measured after 1-hour incubation. Trans-inhibition by 105 BTE was set as 100% and luciferase expression in the presence of each of the 4wj mutants was normalized to 105 BTE with standard error shown. Experiment is performed in triplicate and repeated in three independent experiments. (c) Native gels (TBM) containing Tris-borate buffer and indicated amounts of MgCl2. (d) Native gel (TBK) containing Tris-borate buffer and 500mM KCl. (e) Denaturing gel (dTBE) containing Tris-borate buffer and 2mM EDTA. For details on amount of RNA loaded see Methods.

Figure 2. Cation-driven BTE RNA folding. (a) The BYDV BTE modification and cleavage structure probing patterns generated using benzoyl cyanide (SHAPE) reagent or RNase V1 in the presence of increasing [Mg2+] are shown. (b) Folding induced increase in modification and V1 cleavage at four widely separated residues G4837 (filled triangle), C4899 (open triangle), A4832 (filled square), and A4868 (open square) in the BTE were plotted as a function of log [Mg2+]. The data were fit to a Hill equation as follows: fraction folded = [L]n/Kd + [L]n, where [L] is free ligand concentration, Kd is equilibrium dissociation constant and n is the Hill coefficient. (c) Mg2+ dependent folding of functional A4900C and non-functional A4900G 4wj mutant assessed by SHAPE probing. (d) Observed increase in modification at conserved highly modified nucleotide G4837 (open and closed circle) as a
function of $[\text{Mg}^{2+}]$ was analyzed and plotted as described in (b). A and G are dideoxy sequencing lanes. Unmodified RNA (- lane) shows background RNA hydrolysis.

**Figure 3.** Cooperativity of BTE RNA folding in three viral genera. (a) Treatment of RSDaV, TNV-D, and RCNMV BTEs with benzoyl cyanide in the presence of increasing $\text{Mg}^{2+}$ concentrations. (b) BTE secondary structures with mapped changes in the modification pattern as a function of $\text{Mg}^{2+}$ concentration. (c) BzCN induced modification was quantified for conserved highly-modified G nucleotide; RSDaV’s G5191 (triangle), TNV-D’s G3591 (square) and RCNMV’s G3608 (circle). BzCN induced modification for TNV-D’s G3591 (square) and RCNMV’s G3608 (circle) nucleotide were plotted as function of log $[\text{Mg}^{2+}]$ as described above. A and G are dideoxy sequencing lanes. Unmodified RNA (- lane) shows background RNA hydrolysis.

**Figure 4.** Analysis of the eIF4G-BYDV BTE interaction. (a) Benzoyl cyanide footprinting of the BTE and non-functional BTE BF RNAs in the presence of non-binding proteins eIF4E (4E), ovalbumin (O), and p70 truncation of eIF4G (70), or BTE binding proteins N-terminally truncated forms p86 (86), and p100 (100) or full eIF4G (4G) with or without added eIF4E. (b) The BTE secondary structure showing protection pattern in the presence of p100 and eIF4E. A and G are dideoxy sequencing lanes. RNA modified in the absence of protein factors (+ lane) and unmodified RNA (- lane) shows background RNA hydrolysis.
**Figure 5.** Footprinting analysis of BTEs from 3 viral genomes. (a) Relative translation levels of uncapped 4 nM BlucB mRNA in the presence of uncapped or capped 0.8 µM BTE RNAs from 3 genera, BYDV and RSDaV (Luteovirus), TNV-D (Necrovirus) and RCNMV (Dianthovirus). Luciferase activities obtained from the indicated RNAs are normalized to uncapped BLucB (defined as 100%) and shown as RLU (relative light units). (b) Benzoyl cyanide mediated footprinting of the TNV-D, RSDaV and RCNMV BTEs in the presence of non-binding proteins eIF4E (4E), or p70 truncation of eIF4G (70), or BTE binding protein N-terminally truncated form of eIF4G with retained the eIF4E binding site, p100 (100) with or without added eIF4E. All proteins were added to final concentration of 1 µM. (c) The TNV-D, RSDaV and RCNMV BTE secondary structures showing protection pattern in the presence of p100 and eIF4E.

**Figure 6.** A model of the Mg$^{2+}$-dependent folding of the BYDV BTE and eIF4F (eIF4G+eIF4E) binding based on native PAGE, structure probing and footprinting experiments.
Fig. 1.
Fig. 2.
Fig. 3.
Fig. 4.

- 100 + eIF4E induced protection of hypermodifed bases
- 100 + eIF4E mediated increase in reactivity with the benzoyl cyanide
Fig. 5

(a) mRNA: BLucB

(b) TNV-D

(c) 100 + eIF4E induced protection of indicated bases

RCNMV

RSDaV

SL-I

SL-II

SL-III

SL-IV

SL-V

SL-VI
Fig. 6
CHAPTER 4: STRUCTURAL PLASTICITY OF BARLEY YELLOW DWARF VIRUS-LIKE CAP-INDEPENDENT TRANSLATION ELEMENTS IN FOUR GENERA OF PLANT VIRAL RNAs


Zhaohui Wang, Jelena J. Kraft, Alice Y. Hui and W. Allen Miller

Author contributions: ZW performed the experiments for figures 1-5, JJK for figure 1 (D-F), figure 3 C and figure 4 (SHAPE analysis of GRV BTE). JJK and AYH performed in vivo translation experiments. ZW, JJK, AYH and WAM analyzed the data and wrote the paper.

ABSTRACT

The 3′ untranslated regions (UTRs) of many plant viral RNAs contain cap-independent translation elements (3′ CITEs). Among the 3′ CITEs, the Barley yellow dwarf virus (BYDV)-like translation elements (BTEs) form a structurally variable and widely distributed group. Viruses in three genera were known to harbor 3′ BTEs, defined by the presence of a 17-nt consensus sequence. To understand BTE function, knowledge of phylogenetically conserved structure is essential, yet the secondary structure has been determined only for the BYDV BTE. Here we show that Rose spring dwarf-associated luteovirus, and two viruses in a fourth genus, Umbravirus, contain functional BTEs, despite deviating in the 17 nt consensus sequence. Structure probing by selective 2′-hydroxyl acylation and primer extension (SHAPE) revealed conserved and highly variable structures in BTEs in all four
genera. We conclude that BTEs tolerate striking evolutionary plasticity in structure, while retaining the ability to stimulate cap-independent translation.

**INTRODUCTION**

Upon entry into the host cell, the genomic RNA of a positive-sense RNA virus must first be translated to generate the viral proteins necessary for RNA replication. The very small number of initially infecting RNAs must compete with actively translating host mRNAs for the host translation machinery. Thus, plant viral RNAs have evolved a variety of mechanisms to usurp host factors for their own translation. An increasing number of translation enhancer elements have been identified in the untranslated regions (UTRs) of plant positive-sense RNA viruses (Dreher and Miller, 2006; Kneller et al., 2006; Miller et al., in press).

All nonviral eukaryotic mRNAs have a 5′-cap structure, and all plant and most animal mRNAs have a 3′-poly (A) tail. These modifications are essential for recruitment of translation initiation factors and the ribosome to the mRNA and for mRNA stability (Hentze et al., 2007; Jackson et al., 2010). In contrast to canonical mRNAs, mRNAs of many positive-sense RNA viruses lack a 5′ cap structure and instead contain a cap-independent translation element (CITE), that allows for efficient translation initiation. Among the CITEs located in the 5’ UTRs of animal and some plant viruses, are internal ribosome entry sites (IRESes) which recruit the 40S ribosomal subunit directly to the mRNA (Filbin and Kieft, 2009; Jan, 2006; Niepel and Gallie, 1999; Zeenko and Gallie, 2005; Karetnikov and Lehto, 2007). In contrast, many uncapped plant viral RNAs harbor a CITE in the 3’ UTR of the genome (Kneller et al., 2006; Fabian and White, 2004; Scheets and Redinbaugh, 2006;
Stupina et al., 2008; Wang et al., 2009). The 3′-CITEs identified so far fall into about seven or eight distinct classes based on their sequence and secondary structures (Miller et al., 2007). These elements show no obvious sequence or structural similarity to each other, except that most harbor a stem-loop in which the loop sequence is known or predicted to base pair to the 5′-UTR (Guo et al., 2001; Fabian and White, 2004; Miller and White, 2006; Karetnikov and Lehto, 2008).

Barley yellow dwarf virus (BYDV) and viruses in four genera (below) harbor a BYDV-like CITE (BTE) in their 3′ UTRs. Phylogenetic comparisons, mutagenesis, and structural probing of the BYDV BTE revealed that the BTEs contain a 17 nt conserved sequence, (17 nt CS) GAUCCUCGGGAAACAGG that forms a stem-loop (SL-I) in which the underlined bases are paired (Guo et al., 2000; Mizumoto et al., 2003; Shen and Miller, 2004; Meulewaeter et al., 2004). The BYDV BTE adopts a cruciform secondary structure with three stem-loops radiating from the central hub which is connected to the rest of the viral genome by a fourth basal helix (Guo et al., 2000). The loop of one of the stem-loops (SL-III in the BYDV BTE) must base pair to the 5′ UTR in a kissing stem-loop interaction to facilitate cap-independent translation (Guo et al., 2001; Rakotondrafara et al., 2006). This interaction may ensure localization of initiation factor eIF4G, which binds the BTE (Treder et al., 2008), to the 5′ end where the 40S ribosomal subunit is recruited.

To understand how the BTE structure brings about its function, phylogenetic comparisons with other viral 3′ CITEs reveal which sequences and structures are essential and which are not. Indeed, 3′ CITEs resembling, but differing significantly from, the BYDV BTE have been demonstrated in other viruses, including Tobacco necrosis viruses A and D (TNV-A, TNV-D) (Meulewaeter et al., 2004; Shen and Miller, 2004) and Red clover necrotic
mosaic virus (RCNMV) RNA 1 (Mizumoto et al., 2003). These were identified as BTEs, because they harbor a sequence that is identical, or nearly-identical, to the 17 nt CS. The BTEs from TNVs A and D (which, despite their similar names, are quite distinct viruses), are predicted to lack a structural homolog to SL-II of the BYDV BTE (Meulewaeter et al., 2004; Shen and Miller, 2004). In contrast, the BTE of RCNMV is predicted to have five stem-loops radiating from the central core, in addition to the basal helix (Mizumoto et al., 2003; Sarawaneeyaruk et al., 2009). However, no direct structural analyses have been performed on any of the BTEs except that of BYDV (Guo et al., 2000).

To understand how the BTEs function, it is necessary to determine the essential sequence and structural features they share. Sequences matching or resembling the 17 nt CS are present in all members of the Luteo-, Diantho- and Necrovirus genera (Mizumoto et al., 2003; Shen and Miller, 2004; Wang et al., 1997) and, we report here, in two viruses of genus Umbravirus (Table 1). Here we determine the functions and structures of 3’ UTR elements which contain sequences resembling, but not perfectly matching, the 17 nt BTE consensus, or which contain secondary structures that differ from those characterized to date. We report: (i) the first known BTEs in genus Umbravirus, (ii) a functional BTE in the Rose spring dwarf-associated luteovirus (RSDaV) 3’ UTR which differs substantially from other known BTEs by having large tracts of unpaired bases, (iii) less functional BTEs and non-functional BTE-like sequences that differ from consensus in the 17 nt CS, (iv) chemical determination of actual secondary structures of BTEs from four viral genera, and (v) conserved structural features common to all BTEs. These results reveal that the BTE structure shows substantial plasticity, both in phylogenetic terms, and possibly physically, compared to other known cap-independent translation elements or IRESes.
RESULTS

A functional BTE in umbravirus 3' UTRs. The tract spanning nts 3742–3758 in the 3' UTR of the tobacco bushy top umbravirus (TBTV) genome (GenBank AF402620) matches the 17 nt CS consensus perfectly (Table 1). Moreover, the predicted secondary structure of the putative TBTV BTE resembles that of BYDV except that the first four bases of the conserved sequence may be weakly base paired to extend SL-I, and there is a small stem-loop expected to branch from the end of SL-III (Fig. 1A).

To determine whether the predicted TBTV BTE confers cap-independent translation, a reporter construct was prepared which consists of the full-length 5' and 3' untranslated regions of TBTV genome flanking the firefly luciferase ORF (construct TUlucTU, Fig 1B). The mRNA was produced by in vitro transcription driven by a T7 RNA polymerase promoter immediately upstream of the TBTV 5' UTR. The resulting TUlucTU transcripts were translated in wheat germ extract in amounts (4 nM) well below the saturating levels, so the levels of luciferase activity were proportional to the translation efficiency of the mRNA. Because the viral cap-independent translation element is defined as a sequence sufficient to functionally replace a 5' cap, translation of both capped and uncapped TUlucTU RNAs was measured. Uncapped TUlucTU translated 57% as efficiently as the capped version (Fig 1C). Deletion of the 3' UTR, or a four base insertion (GAUC) in the natural BamH I site in the 17nt CS (construct TUlucTUBF), reduced luciferase expression of uncapped RNA to 20% of the capped version. The GAUC insertion is known to knock out cap-independent translation activity of the BYDV and TNV BTEs (Guo et al., 2000; Shen and Miller, 2004). Notice that the capped transcripts lacking the viral 3' UTR (TUluc) or containing the GAUC insertion (TUlucTUBF) translated at levels similar to that of uncapped transcript harboring the intact
TBTV 3’ UTR (Fig. 1C). Moreover, when these constructs were tested in protoplasts, TULucTU translated about 30-fold more efficiently than TULucTUBF. However, translation of both constructs was low in protoplasts, probably due to the very short 10 nt 5’ UTR which may be removed by 5’–3’ cellular exonucleases in the absence of a 5’ cap. Overall, these results resemble those observed previously with BYDV (Guo et al., 2000; Wang et al., 1997, 1999) and TNV BTEs (Meulewaeter et al., 2004; Shen and Miller, 2004), indicating that TBTV harbors a 3’ BTE.

To map the minimal length of TBTV BTE necessary for cap-independent translation, deletions were made in the 3’ UTR of TULucTU and corresponding RNAs were tested in wheat germ extract for their ability to support translation. Translation levels with 3’ UTRs consisting of TBTV nts 3712–3872 or nt 3724–3838 were greater than 80% of the levels obtained from full-length viral 3’ UTR (TULucTU). In contrast, uncapped reporter mRNA containing only TBTV nts 3731–3828 (Fig. 1A) in the 3’ UTR translated less than 20% as efficiently as with the full 3’ UTR. Thus we define the element spanning nts 3724 to 3838 as the TBTV BTE.

We next tested the necessity of predicted structures of the TBTV BTE that are absent in all other BTEs (SL-IIIb) or absent in TNV BTEs (SL-II) (Fig. 1A). Deletion of SL-II (dm1), replacement of the G in the junction between SL-I and SL-III with AUAU (sm1) or deletion of SL-IIIB (dm2), reduced translation to approximately 20% of that obtained with uncapped, wild type TULucTU (Fig. 1A). Replacement of the GUAAC sequence in loop III, which is predicted to kiss with a loop in the 5’ UTR, with the sequence GUGGA (sm2) retained just over one-third of wild type translation activity. However, structural probing data (below) showed that the sequences at the end of SL-III may not exist in the predicted structures. In
summary, the TBTV BTE is sensitive to changes, even in regions that are not obviously conserved in other BTEs.

A potential BTE in the 3’ UTR of Groundnut rosette umbravirus (GRV) contains the sequence 3639-GGAUCCCGGAAACUGG-3655 (GenBank NC_003603), which differs from the 17nt CS at two positions (bold, italic). This would be expected to weaken or disrupt the SL-I base pairing in the consensus (underlined). However, Mfold predicts a four base pair SL-I, as in the canonical BTE, but with the two strands of the helix shifted, giving a GAAA tetraloop (Fig. 1D). Also, there is no structural equivalent to SL-II, instead there is a tract of 17 unpaired bases upstream of SL-III (Fig. 1D). The 17 nt CS tracts of both umbraviruses are located near the middle of the 645 nt (TBTV) or 535 nt (GRV) 3’ UTR (Fig. 1). This differs from the BYDV and TNV BTEs which are located at the 5’ end of the 3’ UTR, immediately down-stream of the stop codon of the upstream ORF.

We were unable to obtain clones of the GRV sequences because the virus is a serious pathogen limited to Africa. Thus, to test the GRV sequence for BTE function, the 12 nt GRV 5’ UTR and the 120 nt predicted GRV BTE sequence were synthesized, and subcloned into the TUlucTU clone in place of the TBTV 5’ UTR and BTE, respectively, giving rise to construct GlucG (Fig 1E). Uncapped GlucG mRNA translated slightly more efficiently than uncapped TUlucTU mRNA and 70% as efficiently as capped GlucG mRNA (Fig. 1F). Thus, functional replacement of the TBTV BTE with the GRV sequence indicates that the GRV sequence also functions as a cap-independent translation element. To further analyze the significance of the deviations from the 17 nt CS, we mutated the GRV 17 nt CS to match the BYDV consensus sequence. The uncapped transcript translated slightly more efficiently than wild type and nearly as well as the capped version (GlucGs, Fig. 1F). Thus, while the
perfect consensus may provide slightly more efficient cap-independent translation than the natural GRV 17 nt CS, it appears that the BTE may tolerate the base differences in the wild type 17 nt CS of GRV with little effect on function, at least in uncompetitive conditions in wheat germ extract.

The RSDaV genome contains a functional BTE element in the 3’ UTR. The genome organization and sequence of recently-discovered Rose spring dwarf-associated virus (RSDaV) are very similar to those of luteoviruses including BYDV (Salem et al., 2008), indicating that RSDaV belongs in genus Luteovirus. The 3’ UTR of RSDaV also harbors a 17 nt CS (bases 5190–5206) that differs from the BTE consensus at one base (a G–U transversion at position 10, Table 1). Downstream of this sequence is a predicted stem-loop capable of forming a kissing-loop interaction with the 5’ UTR, like the BYDV BTE. As in the case of BYDV, this stem-loop contains five predicted loop bases (UUGUC in RSDaV, UGUCA in BYDV) complementary to loop sequences in the predicted 5’ end of subgenomic RNA1 and the 5’ end of genomic RNA (Salem et al., 2008). However, the predicted RSDaV BTE secondary structure differs strikingly from other known and predicted (below) BTEs (Fig. 2A). All known BTEs are predicted to contain stem-loops that radiate from a central hub, while the putative RSDaV BTE is predicted to have large tracks of unpaired or weakly paired sequence between predicted stem-loops SL-I, SL-II and SL-III and these putative stem-loops don't all project from a central hub. Many suboptimal structures with very similar minimum free energies were predicted by Mfold (Salem et al., 2008), indicating a relatively unstructured RNA.

To test whether the predicted RSDaV BTE confers cap-independent translation, it was
tested by the same type of construct that was used for TBTV: an mRNA encoding the firefly luciferase gene flanked by the complete 5’ and 3’ UTRs of RSDaV genomic RNA (construct RSlucRS, Fig. 2B). The uncapped transcript yielded 50% as much luciferase as the capped transcript (Fig. 2C). The ratio of luciferase obtained from uncapped to capped transcripts dropped to 15–17% when reporter RNAs lacking the viral 3’ UTR, or containing the GAUC insertion in the BamH I site of the 17 nt CS were translated. Importantly, the capped versions of constructs with the mutant or deleted 3’ UTR translated about as efficiently as uncapped mRNA containing wild type UTRs, as was observed for other BTEs (e.g. Fig. 1C). Thus, we conclude that the RSDaV 3’ UTR contains a functional 3’ BTE.

**Structure solution probing of BTE elements.** Although the secondary structures of many BTEs have been predicted, only that of BYDV has been determined directly. Because the predicted structures are so diverse, we set out to directly probe the secondary structures of diverse BTEs by Selective 2’-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) (Mortimer and Weeks, 2007; Wang et al., 2009; Wilkinson et al., 2006). SHAPE chemistry reveals the position of unpaired or otherwise conformationally unconstrained nucleotides, whose 2’-hydroxyl group is able to form a 2’-O-ester product with SHAPE reagents N-methylisatoic anhydride (NMIA) or 1-methyl-7-nitroisatoic anhydride (1M7) (Mortimer and Weeks, 2007; Wilkinson et al., 2006). The modified sites block reverse transcriptase so they can be identified by primer extension followed by polyacrylamide gel electrophoresis. Based on the previous solution structure probing of Pea enation mosaic virus (PEMV) RNA2 (Wang et al., 2009), we designed a cassette which contains the T7 promoter sequence, and EcoR I and Hpa I sites upstream of PEMV RNA2 bases 3924–3981 (which are not part of a
translation enhancer), which are followed by a Sma I site (Fig. 3A). The PEMV RNA2 sequence between the Hpa I and Sma I sites forms a stem-loop that serves as an excellent primer binding site (Wang et al., 2009). To investigate the BTE structures, we took advantage of this sequence as a universal primer binding site, by placing the BTE of interest between the T7 promoter and the Hpa I site (Fig. 3A). Selected BTEs, were amplified and inserted between the EcoR I and Hpa I sites, and the resulting constructs were linearized with Sma I to allow transcription of RNAs for structure probing.

To ensure that each BTE folded into a functional conformation in the context of the structure probing cassette, we tested the efficiency with which each BTE inhibited translation in trans, comparing the BTE RNA lacking the extraneous sequences to the version containing the stem-loop from the expression cassette, i.e. the same RNA that will be used in structure probing assays. Previously it was established that the trans-inhibition efficiencies of three unrelated CITEs: the STNV TED, BYDV BTE and PEMV RNA2 PTE, correlated well with their translation enhancing activity in cis (Gazo et al., 2004; Guo et al., 2000; Wang et al., 2009). This is because they function by binding translation initiation factor complex eIF4F, and trans-inhibition is likely due to competition by the trans-inhibiting CITE with the mRNA for eIF4F. BTEs representing all four genera, BYDV and RSDaV (Luteovirus), TBTV (Umbravirus), TNV-D (Necrovirus), and RCNMV (Dianthovirus), were added to a translation reaction containing the BYDV reporter mRNA BlucB. Except for the RCNMV BTE, all BTEs inhibited translation by 75% to 85% when added in 100-fold excess over the mRNA (Fig 4B). The RCNMV BTE reduced BlucB translation by 60–75%. Importantly, the negative control, BTEBF, reduced translation of BlucB by less than 20% when present at the same concentrations as the trans-inhibiting RNAs. The presence of the 3′ stem-loop in the
structure probing cassette (Sma I-linearized transcripts) had no deleterious effect on trans-Inhibition (Fig. 3B). In some cases presence of the primer binding site enhanced trans-Inhibition (compare RCNMV transcripts from Sma I- vs. Hpa I-linearized templates, Fig. 3B).

In contrast to the above BTEs, the GRV BTE only slightly inhibited BlucB translation in trans (Fig. 3C). We speculated that because its 17 nt CS deviates most from consensus, the BTE of GRV may not compete well with the BYDV BTE in BlucB. To test this, we observed ability of the GRV BTE to trans-Inhibit mRNA containing the GRV BTE in cis, i.e. GlucG. Indeed, the 100-fold excess of GRV BTE (with or without the SHAPE primer sequence from PEMV RNA 2) trans-Inhibited translation of GlucG by 50%, and a 200-fold excess inhibited even more (Fig. 3C). We conclude that the RNAs used for structure probing are highly likely to form the conformations in which they function to stimulate translation in cis, and that the GRV BTE does not compete as well for translation machinery as the other BTEs.

Results of primer extension following exposure of the BTE transcripts to SHAPE reagents reveal that the predicted loops are modified more heavily than flanking paired bases (Fig. 4). The SHAPE reactivities for each nucleotide in the BTE are superimposed on the best fitting RNA secondary structure in Fig. 5. SHAPE probing revealed the presence of SL-I topped by the GNRNA (N = any base, R = purine) pentaloop sequence in all tested BTEs. The first four bases of the 17 nt CS, GGAU (GGAC in RCNMV RNA1) showed some intriguing common features. One of the first two G's of the 17 nt CS was highly modified by SHAPE. In BYDV and TNV-D, the first G of the 17 nt CS is highly modified, while the adjacent GAU sequence has potential to base pair to the AUC sequence (underlined, Fig. 5) adjacent to the opposite strand of stem-IV. In the RCNMV BTE, the second G of the 17 nt
CS is modified and the flanking G and AC may base pair to a GUC in the same relative position as the AUC in the above two BTEs. Thus there appears to be natural covariation to allow the C of the GGAC in RCNMV RNA1 to base pair in a similar fashion to stem-IV as the U in the GGAU sequence of the other BTEs. The TBTV and RSDaV BTEs also have potential base pairing between the GAU and an AUC at the end of stem-IV, but the SHAPE probing of the TBTV BTE favors different interactions, because in both cases the AUC is moderately accessible to SHAPE reagent, while the pairing most parsimonious with the data is between the AU doublet at positions 3 and 4 of the CS and a UG on the opposite side of stem-I (TBTV, Fig. 5). The only BTE lacking a highly modified G at the beginning of the 17 nt CS is the weaker, non-consensus GRV BTE in which G's at that position were modified only moderately.

Another characteristic structural element predicted by Mfold and verified by SHAPE in all BTEs is the highly stable SL-III. Although it is not always the third stem-loop downstream of SL-I, with the exception of the RCNMV BTE, we define SL-III as the GC-rich stem-loop containing a loop capable of base pairing to the 5′ UTR, as is the case for BYDV. The terminal loop of SL-III in BYDV and TNV-D BTEs was shown experimentally to base pair to a loop region in the 5′ UTR of the genomic RNA (Guo et al., 2001; Miller and White, 2006; Shen and Miller, 2004). What we call SL-III in the RSDaV BTE is predicted to base pair to a loop in the genomic 5′ UTR (Salem et al., 2008). In contrast, the TBTV BTE interaction with the 5′ UTR remains ambiguous because the genomic 5′ UTR has only 10 nucleotides. Moreover, SHAPE analysis is consistent with a different conformation at the distal end of SL-III than predicted by Mfold as in Fig. 1. The end branches into two very short stem-loops instead of one three base pair stem-loop protruding from a large
asymmetrical bulge (compare TBTV Fig. 5 with Fig. 1A).

The SHAPE results also differ from the three most stable structures of the RSDaV BTE predicted using Mfold (Salem et al., 2008) (Fig. 2A). The predicted stem-loop II does not exist, instead these bases participate in a long, bulged stem-loop that we call SL-II (Fig. 5). We were unable to identify a structure that reconciled all of the structure probing data, hence some unmodified nucleotides are shown as single-stranded and some modified ones are base paired. SHAPE probing revealed that SL-III is in a different position than predicted and, in fact, is located near SL-I, which more closely resembles its position in the other BTEs. However RSDaV SL-III is flanked by a very large number of unpaired bases. The basal helix of the RSDaV BTE also resembles the other BTEs more closely than predicted because SHAPE reveals that it connects to the viral genome at the end of the helix, rather than at the side of a stem-loop as predicted (compare Fig. 5 with Fig. 2A).

At 150nt long, the RCNMV RNA1 BTE (known as 3′TE-DR1, Mizumoto et al., 2003) is the longest and most structurally complicated of the BTEs. The SHAPE data are in good agreement with the structure predicted by Okuno's group (Mizumoto et al., 2003; Sarawaneeyaruk et al., 2009) (Fig. 5). However, the SL-II loop is larger than predicted and predicted loop V was poorly modified by 1M7 indicating that this region is inaccessible, base paired or constrained in some other way. The UU bases in the SL-III loop were not reactive to 1M7, perhaps because they are constrained in a highly stable UNCG tetraloop flanked by a C–G base pair (Molinaro and Tinoco, 1995). The loops of SL-II, SL-IV and SL-V each have potential to base pair to predicted single stranded tracts in the 5′ UTR (boxed Fig. 5), but Sarawaneeyaruk et al. (2009) showed that no single loop in the RCNMV BTE is required to base pair to the 5′ UTR
for cap-independent translation. However the authors did not quite rule out the possibility that any of the three loops may be sufficient for base pairing to the 5′ UTR.

DISCUSSION

We define BTEs as RNA sequence elements, located naturally in the 3′ UTR, that confer efficient cap-independent translation, and that harbor a sequence resembling the 17 nt CS. We showed here that BTEs exist in viruses in four genera representing portions of two virus families. One genus, Umbravirus, has not been assigned to a family. To our knowledge, this is the first report of BTEs in the Umbravirus genus. Interestingly, a different umbravirus, PEMV RNA 2, contains a completely different CITE (the PTE) (Wang et al., 2009), and a fourth umbravirus, Carrot mottle mimic virus, seems to have neither a BTE nor a PTE in its 3′ UTR. This lack of relationship between type of CITE, and classification of the virus, extends to the other genera and their families. Of the three genera in the Luteoviridae, only genus Luteovirus harbors a BTE. The CITEs of the other Luteoviridae genera, if any, are unknown. Of the eight genera in the Tombusviridae, only two, Necrovirus and Dianthovirus harbor BTEs. The other Tombusviridae genera contain a variety of diverse, and apparently unrelated CITES (Meulewaeter et al., 1998; Gazo et al., 2004; Fabian and White, 2006; Scheets and Redinbaugh, 2006; Truniger et al., 2008; Stupina et al., 2008; Miller et al., 2007), including one genus, Panicovirus, which harbors the same type of CITE (PTE) as the umbravirus PEMV RNA2 (Jeffrey et al., 2006; Wang et al., 2009).

Based on the results presented here and elsewhere (Guo et al., 2000; Mizumoto et al., 2003; Sarawaneeyaruk et al., 2009), the consensus of known functional 17nt conserved sequences is: GgAuCCuGGgAAACaGG (bases in lower case can be altered) and it
must be able to form a four base pair helix. The consensus of the 17 nt CS of all known and predicted BTEs is ggAuCCuGggaaACaGG (Table 1). While it is not known whether all the non-consensus 17 nt conserved sequences are functional in their natural context, it is likely that the 17 nt CS tolerates different types of deviations, but not a large number in the same sequence.

The situation is more ambiguous with regard to secondary structure. All functional and structure-probed BTEs have the following features. (i) A long, bulged, basal connecting helix from which a number of helices radiate at the distal end. (ii) The 17 nt CS which begins at the distal end of the basal helix. (iii) Three of the 5′-proximal four bases of the 17 nt CS have potential to base pair to a complementary sequence on the opposite side of the basal helix, but if this base pairing exists it is not essential, and in TBTV these bases may instead pair to bases downstream of SL-I (Fig. 5). (iv) The first or second guanidylate of the 17 nt CS is modified by SHAPE reagent (dots, Fig. 4) indicating it is exposed and unlikely to be base paired. The lesser modification of the G in the 17 nt CS of the GRV BTE may explain the weak activity of the GRV BTE as measured in the trans-inhibition assay (Fig. 3C). (v) Bases 5–17 of the 17 nt CS form a four base pair helix connected by a five base loop that fits the consensus GNRNA pentaloop motif (or in rare cases a GNRA tetraloop). The GNRNA pentaloop is highly stable, resembles a GNRA tetraloop, and is known to be a protein binding site in bacteriophage lambda RNA (Legault et al., 1998). (vi) A stable stem-loop we call SL-III has at least six uninterrupted base pairs, of which at least four are G–C or C–G pairs. Its loop is predicted to base pair to the 5′ UTR, except for RCNMV in which the kissing stem-loop, if any, may be one of three different loops. (vii) Unpaired bases link together many of the helices radiating from the central hub, suggesting a relatively floppy structure or structure
determined by non-Watson–Crick interactions. (viii) The orientation of SL-III relative to the hub is expected to be variable or flexible as it is linked to the hub by unpaired, SHAPE-modifiable bases.

We speculate that the basal helix functions as a platform for proper folding of the rest of the BTE. In all of the most stable structures predicted by Mfold analysis of the entire BYDV genome, the BTE folds correctly and protrudes distinctly (data not shown). We propose that this makes the BTE accessible to translational machinery and also aids in allowing the kissing stem-loop III access to the 5′ end of the genome to which it base pairs. One suggested model for the BTE mechanism was that the GGAUC at the 5′ end of the 17 nt CS may base pair to the 3′ end of the 18S rRNA, analogous to the Shine–Dalgarno interaction in bacteria (Wang et al., 1997). However, the presence of natural deviations from consensus, such as the GGACC in RCNMV, and the lack of SHAPE accessibility of all but one highly modified G in this sequence render that model unlikely.

Previous data indicate that the BTE binds the eIF4G subunit of eIF4F, leading to our model that the BTE delivers eIF4F to the 5′ UTR by long-distance base pairing and that this, in turn, recruits the 40S ribosomal subunit to the 5′ end (Treder et al., 2008). The 40S subunit then binds the 5′ end of the RNA and scans to the start codon by the same mechanism as on normal capped mRNAs. (Rakotondrafara et al., 2006).

One candidate for a protein binding site in the BTE is the 17 nt CS. In all BTEs (except for the weak GRV BTE, which contains a GNRA tetraloop), the 17 nt CS contains a consensus GNRNA pentaloop motif. High resolution NMR structural analysis of bacteriophage lambda box b RNA revealed a similar stem-loop containing the GNRNA loop. In this loop, the G and A are paired and all bases except the protruding fourth base are
stacked in a helix, folding as in a GNRA tetraloop (Legault et al., 1998). A GNRA or GNRNA loop is sufficient for binding of the lambda N protein, while the 4th protruding base of the GNRNA loop is required for subsequent binding of the E. coli transcription elongation factor NusA. We speculate that the 17 nt CS forms a similar structure which is bound specifically by eIF4F (directly by the eIF4G subunit and indirectly by the eIF4E subunit) and other as-yet unidentified BTE- binding proteins (Treder et al., 2008). The data presented here will guide future research into the RNA structural requirements for high affinity binding to these factors.

**MATERIALS AND METHODS**

**Plasmids.** A full-length clone of the Tobacco bushy top virus (TBTV) genome was assembled from cDNA segments as described (Mo et al., 2003). TUlucTU is a firefly luciferase (luc2, Promega) reporter construct with firefly luciferase gene flanked by the 5′ and 3′ UTRs of TBTV (Fig. 1). The 10 nt 5′ UTR was fused directly between the T7 promoter and the luciferase start codon by PCR using primer T7-TBTV5′Luc2 5′ and Luc2-r1. The PCR product was digested by EcoR I and Xba I and inserted into EcoR I and Xba I-digested pKF19K-2 vector (TaKaRa). The 3′ UTR of TBTV was amplified using primer TBTV 3′ UTRf and TBTV 3′ UTRr, digested with Pst I and Xba I and inserted into Pst I and Xba I-digested pKF19K-2 vector containing the TBTV 5′ UTR and luc2 as described above. The resulting construct was designated as pTUlucTU. pTUlucTUBF is a mutant vector generated from pTUlucTU with a duplication of the GATC sequence in the 17 CS of the BTE. It was generated by digesting with BamH I and then filling in the sticky ends with DNA polymerase Pfx (Invitrogen) followed by blunt-end re-ligation.
pGlucG was derived by inserting the PCR-amplified 12 nt 5′ UTR of GRV into EcoRI/ScaI-digested pTU
tucTU, in place of the TBTV 5′ UTR. The resulting intermediary
construct was cut with XbaI/XhoI in the 3′ UTR to replace the TBTV BTE sequence with the
PCR fragment containing the GRV BTE sequence to give rise to the finished pGlucG
construct (Fig. 1E).

pRS
lucRS is a firefly reporter construct with the RSDaV 5′ and 3′ UTRs. The RSDaV
5′ UTR was amplified by overlapping PCR using primer RSDaV 5′ UTRf/RSDaV OL1,
RSDaV OL1/luc72–53r to amplify two segments and then fused together using primer
RSDaV 5′ UTRf/ luc72–53r. The resulting PCR product was digested with Xba I and Hind
III and then ligated into Xba I and Hind III-digested pGEM®-luc (Promega). The 3′ UTR of
RSDaV was amplified using primers RSDaV 3′ UTRf/RSDaV OL3, RSDaV OL4/RSDaV 3′
UTRr from two cDNA segments and then fused to generate the full RSDaV 3′ UTR by PCR
using RSDaV 3′ UTRf/RSDaV 3′ UTRr. The PCR product was digested with Stu I and Sac I
and then ligated into Stu I and Sac I-digested pGEM®-luc vector with RSDaV 5′ UTR as
described above. The re-
sulting construct was designated pRS
lucRS.

To generate RNAs from plasmids for trans-inhibition assays and solution structure
probing, a universal cassette was generated and inserted into the PUC19 vector. The cassette
contains the T7 promoter sequence, EcoR I and Hpa I sites, a stem-loop segment from
PEMV RNA2 (nt 3924–3981) and a Sma I site sequence (Fig. 3). Individual BTE segments
(Table 2) were amplified and inserted between the EcoR I and Hpa I sites.

**RNA preparation.** To generate templates for transcription of mRNAs, plasmids
pTU
tucTU, pGlucG, pBlucB, and pRS
lucRS were linearized with Hind III, Hind III, Sma I
and Sac I, respectively. Capped and uncapped RNAs were synthesized by using mMessage mMachine® and MEGAscript® (Ambion), respectively, according to the manufacturer's instructions. Small RNA segments for trans-inhibition assays and solution structure probing were transcribed from Hpa I or Sma I-linearized plasmids using MEGAscriptTM (Ambion) except for the GLucG derived segments. The GLucG BTE with or without the PEMV primer sequence were transcribed from Kpn I-linearized plasmid to include the SHAPE cassette for structure probing or from PCR-amplified GRV BTE segment designed to exclude the cassette for trans-inhibition studies. For trans-inhibition and structure probing, RNA integrity was estimated by 0.8% agarose gel electrophoresis. Final concentration and purity was determined by spectrophotometry.

**In vitro translation.** In vitro translation reactions were performed in wheat germ extract (Promega) as described (Wang et al., 2009). Nonsaturating amounts of RNAs (0.2 pmol) were translated in wheat germ extract in a total volume of 25 µl with amino acids mixture, 93 mM potassium acetate and 2.1 mM MgCl2 according to the manufacturer's instructions. In trans-inhibition experiments, the mRNA and competitor RNA were mixed prior to addition to the translation reaction. After 1 h incubation at room temperature, 2µl of the translation reaction product was mixed into 50 µl of Luciferase assay reagent (Promega), and measured immediately on a GloMaxTM20/20 Luminometer (Promega).

**In vivo translation.** Uncapped TUlucTU or TUlucTUBF RNAs were co-electroporated into oat protoplasts with capped mRNA encoding Renilla luciferase as an internal control. Protoplasts were prepared and assays performed as described previously
(Rakotondrafara et al., 2006). Four hours following the electroporation, the luciferase activities were measured with Dual Luciferase Reporter Assay System™ of Promega and Firefly relative light units (RLUs) were normalized against the Renilla relative light units. RLU measured in the absence of added luciferase mRNA were subtracted from the values obtained with TUlucTU or TUlucTUBF mRNAs. All samples were tested in triplicate.

**Structure probing.** Plasmids with individual BTE segments in the PEMV RNA2-derived universal cassette were linearized with Sma I to be used as template for generation of RNA transcript using MEGAshortscript™ Kit (Ambion). Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) (Wilkinson et al., 2006) was applied to probe selected BTE elements following the procedure described previously (Wang et al., 2009; Wilkinson et al., 2006). Briefly, 500 ng of RNA was heat denatured and renatured in SHAPE buffer (50 mM HEPES–KOH, pH 7.2, 100 mM KCl and 8 mM MgCl2) at room temperature. 1- methyl-7-nitroisatoic anhydride (1M7) in 50 mM in anhydrous DMSO (Mortimer and Weeks, 2007) was mixed into the renatured RNA aliquot at a 1/10 (v/v) ratio. After 2 min at room temperature, the RNA was mixed with four-fold excess tRNA and precipitated in 3 volumes of ethanol and 1/10 volume 3 M sodium acetate. Control RNA was treated with same amount of DMSO without 1M7. The primer (GATCTTTTTGGCGAGACATC (Wang et al., 2009), shown here in Fig. 3, was 5′ end-labeled with γ-[32P] ATP and used for the extension reaction. Primer extension, gel electrophoresis and visualization by phosphorimager were performed as described previously (Wang et al., 2009; Wilkinson et al., 2006). RNA secondary structures were deduced from solution probing data and the best fitting Mfold prediction (Zuker, 2003).
ACKNOWLEDGMENTS

The authors thank Xiaohan Mo (Yunnan Academy of Tobacco Science, Kunming, PR China) for generously providing the TBTV cDNA clones, and Bryce W. Falk (University of California, Davis) for generously providing RSDaV cDNA clones. We thank Min Zeng for contributing to construction of pRSlucRS and preliminary translation assays. This work was supported by National Institutes of Health grant R01 GM067104 with supplemental funds from the American Recovery and Reinvestment Act of 2009. This journal paper of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA, Project No. 3608 was also supported by Hatch Act and State of Iowa funds.
REFERENCES


Guo, L., Allen, E., Miller, W.A., 2000. Structure and function of a cap-independent translation element that functions in either the 3′ or the 5′ untranslated region. RNA 6, 1808–1820.


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


Meulewaeter, F., van Lipzig, R., Gulyaev, A.P., Pleij, C.W., Van Damme, D., Cornelissen,


via the eIF4G subunit to initiate translation. RNA 14, 134–147.


**FIGURE LEGENDS**

**Figure 1.** Umbravirus 3’ UTRs contain a BTE. (A) Representative predicted (Mfold) secondary structure of the BTE in the 3’ UTR of TBTV genomic RNA, and relative translation activities of selected mutants. 17 nt CS is in bold italics. The shaded bases were deleted in mutants dm1 and dm2 as indicated. In mutants sm1 and sm2, bases in dashed boxes were replaced by bases in solid boxes (dashed arrows). The luciferase activity generated by translation of uncapped mutant RNAs, as a percentage of wild type (100%) is indicated. (B) Genome organization of TBTV RNA and maps of translation reporter constructs containing TBTV UTRs. TUlucTU contains both complete UTRs of TBTV flanking the firefly luciferase coding sequence (fLUC). TUluc has only the TBTV 5’ UTR. TUlucTUBF differs from TUlucTU by only a 4 nt insertion (underlined) in the BamH I site. Bases are numbered according to their position in the TBTV genome. (C) Relative translation activities of capped and uncapped reporter mRNAs in wheat germ extract. Luciferase activities obtained from the indicated RNAs are normalized to uncapped TUlucTU (defined as 100%) and shown as RLU (relative light units). Error bars indicate standard error. (D) Predicted secondary structure of the GRV BTE. The 17 nt CS (italics) is in the dashed box, with bases that deviate from consensus indicated in bold italics. Mutant sequence (CS) which is identical to the consensus 17 nt CS is indicated at right by dashed arrow. (E) Map of reporter RNA GlucG. GlucG contains the GRV 5’ UTR and a chimeric 3’ UTR consisting of the TBTV 3’ UTR with the putative GRV BTE in place of the TBTV BTE. Positions of bases from each virus are indicated. (F) Relative translation activities of capped and uncapped reporter mRNAs in wheat germ extract.
**Figure 2.** RSDaV 3′ UTR contains a functional BTE. (A) A representative predicted secondary structure of the RSDaV BTE (Salem et al., 2008). 17 nt CS is shown in bold italic. (B) Genome organization of RSDaV RNA and maps of translation reporter constructs. RSlucRS has both UTRs of RSDaV flanking the luciferase coding sequence (fLUC). RSluc has only the RSDaV 5′ UTR. RSlucRSBF differs from RSlucRS only by a 4 nt GAUC insertion (underlined) in the BamH I site. Bases are numbered according to their position in the RSDaV genome. (C) Relative translation activities of capped and uncapped reporter mRNAs in wheat germ extract. Luciferase activities are normalized to uncapped RSlucRS (defined as 100%). Error bars indicate standard error.

**Figure 3.** Cassette for structure probing of BTEs, and test of function by trans-inhibition of translation. (A) RNA cassette probing and reverse transcription. The cDNA of the RNA element of interest (bold gray line) was inserted between EcoR I and Hpa I sites of the structure probing cassette upstream of nts 3924–3981 from PEMV RNA2 which form the stem-loop between the Hpa I and Sma I sites. Sma I-linearized plasmids were used as templates to transcribe the RNA elements for trans-inhibition assays and structure probing. The black arrow along the sequence indicates sequence to which the 32P-labeled oligomer was annealed for primer extension. Large gray arrow indicates start site of transcription by T7 RNA polymerase. (B) Relative translation levels of uncapped 4 nM BlucB (reporter mRNA with BYDV UTRs flanking firefly luciferase coding sequence) in wheat germ extract containing 400 nM of the indicated viral BTE RNAs from the cassette in panel A (sequences are shown in Table 2). DNA was linearized with Sma I (S) or Hpa I (H) prior to transcription, as indicated. (C) Relative translation levels of 4 nM BlucB (black bars) or GlucG (gray bars)
in the presence of BTE RNAs lacking (BTE) or containing (BTE SHAPE) the PEMV sequence used for the SHAPE primer.

**Figure 4.** SHAPE analysis of BTEs from six viral genomes. Primer extension products from RNA treated (+) or not treated (−) with 5 mM 1M7 (See Materials and methods for details). The sequencing ladders showing positions of U and C residues in the BTEs (lanes U, C) were generated by reverse transcribing unmodified RNA in the presence of dideoxyATP (U lane) or dideoxyGTP (C lane) using the same 5′-labeled primer that was used for SHAPE. Numbers to the left of each gel indicate genomic positions of indicated bases. Position of the hypermodified G residue at the beginning of the 17 nt CS is indicated by the dot to the right of each gel. Regions corresponding to selected loops (Fig. 5) are indicated by L-[loop number].

**Figure 5.** Superposition of SHAPE reactivity of each nucleotide on the best fitting secondary structures of BTEs. Bases are color coded based on the intensity of the bands in Fig. 4 which reflects the level of modification by 1M7. Nucleotides are numbered according to their positions in the viral genome. 17 nt CS is in bold italics. The three base sequence (AUC or GUC) complementary to the 5′ end of the 17 nt CS (see text) is underlined. Boxed bases are known (BYDV and TNV-D) or predicted (other BTEs) to base pair to the 5′ UTR.

**Table 1.** 17 nt conserved sequence in all known or predicted BTEs. Dot: identical to BYDV-PAV, underlined: paired bases in SL-I of BYDV BTE. GenBank® accession
numbers of the viral genome sequence and references to publications that demonstrate a functional BTE are shown.

Table 2. Sequences of the selected BTE elements for structure probing. The cDNAs were amplified and inserted between EcoR I and Hpa I sites of the universal SHAPE vector (Fig. 3A). The 17 nt CS is underlined. Extra sequence introduced on the 3′ end of BYDV-PAV BTE for extension of SL-III was indicated in italic.
Fig. 1.
Fig. 2.
Fig. 3.
Fig. 5.
<table>
<thead>
<tr>
<th>Virus</th>
<th>17 nt conserved sequence</th>
<th>Accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Genus Luteovirus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barley yellow dwarf virus-PAV (BYDV-PAV)</td>
<td>GGAOFCCGSQGAAACACG</td>
<td>AF235167</td>
<td>(Wang et al., 1997)</td>
</tr>
<tr>
<td>Barley yellow dwarf virus-PAS (BYDV-PAS)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>AF218798</td>
<td></td>
</tr>
<tr>
<td>Soybean dwarf virus (SBDV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>L24049</td>
<td></td>
</tr>
<tr>
<td>Bean leafroll virus (BLRV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_003369</td>
<td></td>
</tr>
<tr>
<td>Rose spring dwarf-associated virus (RSDaV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>EU024678</td>
<td></td>
</tr>
<tr>
<td><em>Genus Dianthovirus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red clover necrotic mosaic virus (RCNMV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_003756</td>
<td>(Mizumoto et al., 2003)</td>
</tr>
<tr>
<td>Carnation ringspot virus (CRSV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>L18870</td>
<td></td>
</tr>
<tr>
<td>Sweet clover necrotic mosaic virus (SCNMV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_003806</td>
<td></td>
</tr>
<tr>
<td><em>Genus Necrovirus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olive latent-1 (OLV-1)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_001721</td>
<td>(Meulewaeter et al., 2004)</td>
</tr>
<tr>
<td>Olive mild mosaic virus (OMMV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_006939</td>
<td></td>
</tr>
<tr>
<td>Tobacco necrosis A (TNV-A)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_001777</td>
<td></td>
</tr>
<tr>
<td>Tobacco necrosis D (TNV-D)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>U62545</td>
<td>(Shen and Miller, 2004)</td>
</tr>
<tr>
<td>Leek white stripe virus (LWSV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_001822</td>
<td></td>
</tr>
<tr>
<td>Black beet scorch virus (BBSV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>AF452884</td>
<td></td>
</tr>
<tr>
<td><em>Genus Umbravirus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobacco bushy top virus (TBTv)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>AF402620</td>
<td></td>
</tr>
<tr>
<td>Groundnut rosette virus (GRV)</td>
<td>A. . . . . . . . . . . . . . . . .</td>
<td>NC_003603</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.**
<table>
<thead>
<tr>
<th>BTE origin</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYDV-PAV</td>
<td>CGUGAAGCAAGACCAUACCACUGUAGCAGACAAAUUCAGAUCUCCUGGGAACAGGCUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
<tr>
<td>TNV-D</td>
<td>CGCUAUCGUUCGUAGCUUUCGUAAUCGUUCGUAGCGUACAGCGGUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
<tr>
<td>TIBTV</td>
<td>CGUGAAGCAAGACCAUACCACUGUAGCAGACAAAUUCAGAUCUCCUGGGAACAGGCUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
<tr>
<td>RSDaV</td>
<td>CGUGAAGCAAGACCAUACCACUGUAGCAGACAAAUUCAGAUCUCCUGGGAACAGGCUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
<tr>
<td>RCNMV RNA1</td>
<td>CGUGAAGCAAGACCAUACCACUGUAGCAGACAAAUUCAGAUCUCCUGGGAACAGGCUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
<tr>
<td>GRV</td>
<td>CGUGAAGCAAGACCAUACCACUGUAGCAGACAAAUUCAGAUCUCCUGGGAACAGGCUAUCUGUGCUUACAUAGCUUCGUGUAGGCUGUCAAAACCCUACGCUUGGUGAUAGCCAUACUGCUUGUGACUGUCAAGGGGUAGGCAGGAGGCA AGUGCAUGACUGUGUAGGCAUAGUGUCUGUGACAGGUCAGUGUCAACCUCCUGGCUUAAUCCCGCGACCCGGGAGCCCGCGCGCCGGCGCGGCACCCGGCATCGAGAGA</td>
</tr>
</tbody>
</table>

Table 2.
CHAPTER 5: GENERAL CONCLUSIONS

BYDV is an economically important virus that impacts a large number of cereal crops throughout the world (Miller et al., 1997). In Sub-Saharan Africa peanuts, another major crop and food staple, have been ravaged by the occurrence of Groundnut rosette disease caused by the combination of Groundnut rosette assistor polerovirus and Groundnut rosette umbravirus (GRV), a virus demonstrated to contain novel functional BTE (Wang et al., 2010). Another Umbravirus used in structural studies, Pea enation mosaic virus (PEMV), infects legumes and lentils both in US and other parts of the world causing up to 100% crop loss (Makkouk et al., 1999; 2003). Like GRV, PEMV is a complex disease that requires cooperativity between a polerovirus and an umbravirus: two positive sense genomic RNA viruses that belong to different genera in order to cause disease.

In addition to their economical importance, viruses described here use an array of unconventional tools to express encoded genes that could provide insight into translation mechanisms used by related plant viruses. This in turn could help in the evolution of antiviral agents against viral translation mechanisms. Little is known about how viruses studied here are able to recruit translation machinery in the absence of canonical modifications in their genome. Even less is known about how they are able to break plant-mediated resistance. The study of recessive resistance genes against plant viruses revealed that the key modulators in the plant virus infections are translation initiation factors and in particular the eIF4F scaffold. About one-half of all known plant virus resistance genes are recessive and, so far all sequenced recessive genes encode alleles of eIF4E or eIF4G (Hébrard et al., 2010). The success of plant breeding for improved resistance depends on how much we know about virus gene expression and replication.
The research presented here focuses on understanding the nature of folded CITE RNA and the mechanism of its interaction with protein synthesis machinery. CITEs differ within and among different classes of elements in both structure, primary sequence and in the component(s) of the eIF4F complex with which they interact (Miller et al., 2007, Treder et al, 2008, Wang et al., 2009; 2010; 2011). What is similar is that all CITEs bind the rate-limiting eIF4F complex with similar affinities suggesting a similar underlying mechanism of action.

The BTE folding involves the cooperative uptake of magnesium and it is driven primarily by charge neutralization to form a stable platform for protein factor binding as suggested by structure probing and crystallization experiments. The BTE retained the same overall fold in the presence of protein suggesting that BTEs adopt their specific 3D fold independently of protein. Similarly, an unrelated group of CITEs, the PTEs, also adopt a magnesium-dependent compact fold prior to protein recognition (Wang et al., 2011). Both classes of unrelated CITE harbor a highly SHAPE-modifiable G residue in the conserved regions of the sequence. It is predicted for the PTE, that this highly reactive G residue positions itself directly in the cap-binding pocket in place of the normal cap structure (Wang et al., 2011). Thus it is possible that the role of a specific protruding G in the BTE is also to interact with a translation initiation factor. However, unlike the PTE, the BTE does not bind eIF4E directly, so the role of the G would not be to insert in the cap-binding pocket to bind eIF4E.

Future research in this area remains to determine the 3D structure of 3’CITE at a high resolution. This can be accomplished in several ways. First, existing BTE RNA crystals can be soaked with heavy atom lanthanide series compounds to obtain phase information and solve the BTE structure below 5Å. The low resolution X-ray analysis will not show the BTE
nucleotide sequence but the overall RNA backbone and the bends in the backbone will be visible. This in turn will allow testing of a new hypotheses and the generation of improved RNA crystallography targets. Also, the study of BTE folding in the presence of cations (Chapter 3) revealed that 8 mM magnesium chloride completely relieves conformational heterogeneity in the BTE leading it to adopt a more compact fold. Thus, to ensure structural homogeneity of RNA prior to crystallization, modification of the folding step is necessary to include dialysis of purified RNA into buffer containing 8 mM magnesium chloride and 10mM HEPES KOH pH 7.5 prior to dilution with reservoir solution.

To complement the crystallography investigations, small-angle X-ray scattering (SAXS) can be employed to observe BTE RNA folding under a variety of conditions and determine its overall size and shape in solution (Hammond et al., 2009). If the SAXS data reveal that the BTE folds similarly to structures already solved by X-ray crystallography such as tRNA, then we could employ a molecular replacement technique to solve the phase problem by making use of phases from a known structure (Rosmann, 1990).

Finally, a closer look at 3’ CITE interaction with initiation factors at high resolution will provide more biologically relevant structure and needs to be investigated in detail. In appendix I, PMV RNA interaction with eIF4E at high resolution was attempted, but eIF4E crystallized free of RNA. To improve the likelihood of obtaining PMV-eIF4E co-crystals further crystal screening of this complex assembled at different RNA to protein ratios using many different crystallization conditions needs to be pursued.

The prerequisite for attempting the BTE-eIF4G co-crystallization is mapping the minimal BTE-binding domain in eIF4G. The Miller lab predicts that there are several potential RNA binding domains located in eIF4G and precise mapping of the minimal RNA
binding region is underway. The smallest eIF4G fragment (35 kDa) purified so far encompasses only the essential middle region of eIF4G and it was found to bind BLucB mRNA in preliminary BTE inhibited wheat germ translations (K. Treder, unpublished). This future research will give direct insight into the specifics of the BTE-translation factor interaction. Understanding of the BTE mediated cap-independent translation will contribute not only to the understanding of translational mechanisms used by related viruses but to understanding of translation initiation in general.
REFERENCES


dwarf virus-like cap-independent translation elements in four genera of plant viral RNAs. Virology 402, 177-186.


APPENDIX 1. CRYSTALLIZATION OF RELATED VIRAL CAP-INDEPENDENT TRANSLATION ELEMENTS

A general strategy to improve the chances of obtaining diffraction quality crystals is to examine and include sequences from related viruses in crystallization trials since the choice of sequence is the single most important variable in the crystallization process (Holbrook et al., 2001). Thus, I focused on crystallizing RNA translation enhancer sequences from the related Necrovirus, Umbravirus and Panicovirus genera of plant viruses alone or in complex with protein. I reasoned that additional sequence variation and inclusion of protein that binds translation enhancer specifically in my crystallization trials could stabilize the RNA structure and increase the diffraction limit of the RNA crystals.

Towards this goal, I purified and included in crystallization trials, structurally simpler cap-independent translation sequences from Tobacco necrosis virus D (TNV-D, genus *Necrovirus*, family *Tombusviridae*) (Shen et al., 2004; 2007) and Pea enation mosaic virus-2 (PEMV2, genus *Umbravirus*, no assigned family) (Wang et al., 2009) recently characterized in our lab. Like Barley yellow dwarf virus (BYDV) BTE, TNV-D BTE harbors a 17 base consensus sequence (GGAUCCUGGGAAACAGG) and a stem-loop capable of base pairing to complementary bases in the 5’ UTR (Fig. 1). Although TNV-D BTE lacks stem loop 2 found in other BTEs, it confers efficient cap-independent translation both in vitro and in vivo (Shen et al., 2004). In contrast to the BTEs, a completely different group of cap-independent translation elements termed PTEs from PEMV2, Panicum mosaic (PMV) and other viruses was discovered recently (Batten et al., 2006; Wang et al., 2009). These PTEs contain a pseudoknot structure formed through base pairing between bases in a G bulged domain and
bases at a helical branch point (Fig. 1). Unlike the BTEs that bind to the large eIF4G scaffold (Treder et al., 2008), PTEs bind to the smaller eIF4E subunit of eIF4F complex with high affinity (Kd = 50 nM) (Wang et al., 2009) that makes them unique among uncapped mRNAs. Compared to the BYDV BTE, both TNV-D BTE and PEMV2 PTE adopt more compact secondary structure where three helices emerge from a single point in the structure with fewer unpaired bases (Fig. 1). Thus, the minimal functional 85 base TNV-D BTE RNA (viral bases 3576 to 3661) and 102 base PEMV 2 PTE (viral bases 3815-3917) were purified as described (Kraft et al., 2011) and subjected to crystallization screening using RNA biased crystal screens (Natrix, Nucleic Acid Mini Screen, and Crystal Screen) from Hampton Research. However, these RNAs produced either clear drop or precipitate but no crystalline material even after extended incubation of up to 4 months at 4 and 22 °C (Fig. 1).

Subsequent dynamic light scattering analysis on these RNA samples generated messy raw data that were impossible to analyze indicating that they underwent significant aggregation that impeded crystal formation.

To improve the odds of obtaining diffraction quality PTE crystals, I focused on the more compact PTE sequence from Panicum mosaic virus (genus Panicovirus, family Tombusviridae) (Batten et al., 2006) in crystallization trials alone or in complex with N-terminally truncated eIF4E (ΔN-eIF4E) for which the X-ray structure has already been solved (Monzingo et al., 2007). Hence, the first step towards determining the PMV PTE structure was to map the minimal functional element that is still able to confer cap-independent translation. For this purpose, progressive truncations in the PTE were made (Fig. 2A) and tested in our standard trans-inhibition assay for their ability to inhibit translation of a luciferase reporter P2 lucP2 mRNA with full-length PEMV RNA 2 UTRs (Wang et al.,
2009) (Fig. 2B). Most tested PMV PTE deletion mutants were functional and the minimal functional PMV PTE element is only 86 nt long. This construct spanning bases 4117-4197 of the viral genome was chosen for further testing by native polyacrylamide gel electrophoresis (PAGE) and dynamic light scattering. The polydispersity index of PTE\textsubscript{4117-4197} sequence measured by dynamic light scattering was 41.4%, indicating presence of oligomeric states. These alternative oligomeric forms were visible on native PAGE but were significantly relieved by heating and snap cooling of the RNA and the addition of 5mM magnesium ions. After following this RNA annealing protocol, PTE\textsubscript{4117-4197} was subjected to crystal screening using commercial screens. 1 ml 0.3 mM BTE RNA was mixed with 1 ml of reservoir solution and equilibrated against 500 ml well solution using hanging-drop vapor diffusion at 4 and 18 °C. Natrix conditions 18 (0.01 M Magnesium acetate, 0.05 M Sodium cacodylate pH 6.5, 1.3 M Lithium sulfate) and 24 (0.2 M Ammonium acetate, 0.01 M Calcium chloride, 0.05 M Sodium cacodylate pH 6.5, 10% Polyethylene glycol 4,000) yielded microcrystalline material at 18 °C 2 months after screen set up, but no individual crystals formed.

To improve crystallizability of PMV PTE sequence, I focused on including eIF4E in crystallization trials. The interaction between PMV RNA and ΔN-eIF4E has not been demonstrated, but interaction between full-length eIF4E and PMV PTE has been demonstrated (Wang et al., 2011), and likely involves the cap-binding pocket (Wang, et al., 2009) which is intact in ΔN-eIF4E. Given the high degree of conservation of secondary structure and tertiary contacts in both macromolecules, ΔN-eIF4E-PMV PTE interaction is likely to occur. Thus, I employed capillary electrophoresis mobility shift assay (CEMSA) to analyze PMV-eIF4E interactions. This method allows for fast and efficient separation of bound species using minimal amounts of reagents. For this purpose, equimolar amounts of
PMV PTE and eIF4E were mixed and incubated at 22° C for 15 minutes to allow complex formation. CE separation was achieved on a Beckman P/ACE-MDQ system with a UV/VIS detector in N-CHO - 50µm polyvinyl alcohol (PVA) coated capillaries (Beckman Coulter) with a cooled sample tray. Before runs, the capillary was rinsed for 1 min with 0.1 M HCl, 10 min with nuclease free water and 10 min with run buffer: 50 mM Tris-HCl pH 8.2. Samples were injected for 5 s at low pressure and electrophoresed at 20 kV in run buffer at 22°C. Compared to free PMV RNA that elutes in the reverse polarity mode about eight minutes after the start of electrophoretic separation, PMV-eIF4E complex was retained longer and eluted around 10 minutes indicating that ΔN-eIF4E binds to PMV PTE (Fig. 3).

Thus, purified ΔN eIF4E (Monzingo et al., 2007) at 4 mg/ml concentration was incubated with 0.15 mM PMV PTE (molar ratio of 1:1) at 22°C for 15 min before crystal tray set up. Single, large crystals appeared in mother liquor containing the 0.1 M Tris hydrochloride pH 8.5 and 2.0 M Ammonium sulfate at 18 ° C (Fig. 3B). These crystals were flash cooled in liquid nitrogen and prescreened on a home X-ray source. All of the screened crystals diffracted to below 3 angstroms indicating that adding protein helped convert poor quality into diffracting quality crystals (Fig. 4). These crystals were then sent to the Advanced Light Source (beamline 4.2.2.) for further screening and data collection.

The collected data was indexed, integrated, scaled and merged using d*TREK (Pflugrath, 1999) and findings are summarized in table 1. PTE-eIF4E co-crystals diffracted to atomic resolution of 2 Å and the completeness of the particular data set was 100%. Interestingly, the crystals belong to the same hexagonal space group as eIF4E protein crystals alone. This seems to indicate that the PTE RNA did not bind to the protein despite being present in solution at the same molar ratio as protein. However, since around half of the
protein crystal is made of liquid it is possible that RNA replaced water in protein crystal. To assess weather or not this occurred, the crystals were washed extensively with reservoir solution before they were redissolved in water and subjected to PAGE analysis. Unfortunately, only a single band corresponding to a protein of about 20 kDa was visible on the bis-Tris SDS-PAGE gel (Invitrogen) and no RNA band was visible when resuspended crystal solution was run on native PAGE gel. It appears that RNA was excluded from the crystal lattice and that the crystals contained only protein.

I believe that the key to growing well diffracting RNA crystals suitable for structural determination is through incorporation of a protein-binding partner. Also, the PTE-eIF4E co-crystal provides more biologically relevant structural information because the PTE RNA requires eIF4E to function. Hence, more solution conditions at varying ratios of RNA to protein need to be pursued to identify those that stimulate RNA-protein co-crystal growth.
REFERENCES


Shen R., and W. A. Miller. 2007. Structures required for poly(A) tail-independent translation overlap with, but are distinct from, cap-independent translation and RNA replication signals at the 3' end of Tobacco necrosis virus RNA. Virology 358, 448-458.

eIF4F via the eIF4G subunit to initiate translation. RNA 14, 134–147.

FIGURE LEGENDS:

Figure 1. Secondary structures of Tobacco necrosis virus D (TNV-D) BTE and Pea enation mosaic virus-2 (PEMV2) PTE with respective images of crystallization activity generated with Natrix condition 9 (0.1 M Potassium chloride, 0.01 M Magnesium chloride, 0.05 M MES pH 6.0, 10% Polyethylene glycol 400) and Nucleic acid mini screen condition 5 (10% MPD, 40mM Na Cacodylate pH 6.0, 12 mM Spermine tetra-HCl, 80 mM Potassium chloride, 20 mM Magnesium chloride)

Figure 2. Functional analysis of PMV PTE mutants by trans-inhibition assay. A. Examples of deletion (blue) and point mutants (red) designed for crystal growing and mapping PTE core element. B. Translation levels of reporter P2LucP2 mRNA containing the full 5’ and 3’ PEMVRNA2 UTRs in WGE supplemented with excess of inhibitor PMV PTEs. Positive control is PEMV2 PTE and a negative control PTEm2 contains two point mutations that disrupts functional pseudoknot (Wang et al., 2009). *PMV4117-4197 is the minimal functional PTE used in crystallography trials.

Figure 3. Capillary electrophoresis mobility shift assay (CEMSA) in 50 mM Tris-HCl pH 8.2 at 22 °C in reverse polarity mode showing eIF4E binding PTE. Free flow is left to right with PTE RNA alone shown in blue and shifted eIF4E-PTE complex in magenta.

Figure 4. Structural analysis of PMV PTE. X-ray diffraction pattern of crystals obtained from equimolar ΔN-eIF4E and PTE in 0.1 M Tris HCl pH 8.5 and 2.0 M NH4SO4 at 18 °C. Resolution: 2.3Å. Inset: Examples of crystals.
Table 1 Data-collection statistics for the PTE-eIF4E co-crystals.

Values in parentheses are for the highest resolution shell.
Fig. 1
Fig. 2
Fig. 4
<table>
<thead>
<tr>
<th><strong>Beamline</strong></th>
<th>MBC 4.2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Spacegroup</strong></td>
<td>P61</td>
</tr>
<tr>
<td><strong>Unit cell parameters (Å, °)</strong></td>
<td>a=b=66.10, c=181.51</td>
</tr>
<tr>
<td></td>
<td>α=β=90, γ=120</td>
</tr>
<tr>
<td><strong>Resolution range</strong></td>
<td>35.56 - 2.16 (2.24 - 2.16)</td>
</tr>
<tr>
<td><strong>Measured reflections</strong></td>
<td>142567</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>24008</td>
</tr>
<tr>
<td><strong>Average redundancy</strong></td>
<td>5.94 (4.19)</td>
</tr>
<tr>
<td><strong>% completeness</strong></td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td><strong>R_{merge}</strong></td>
<td>0.347 (0.951)</td>
</tr>
<tr>
<td><strong>&lt;I/sigI&gt;</strong></td>
<td>2.9 (0.7)</td>
</tr>
</tbody>
</table>

\[ R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl)| - \langle I(hkl) \rangle}{\sum_{hkl} \sum_i I_i(hkl)} \]

Table 1.
APPENDIX 2. INVESTIGATION OF HELICAL STACKING IN THE BTE USING THE “RNA LEGO” SYSTEM

Although we have determined the helices in the BTE (Guo et al., 2000; Wang et al., 2010), the tertiary structure is unknown. Specific helices in the BTE probably stack coaxially, which is an important driving force in the tertiary folding of RNA (Walter et al., 1994; Pyle et al., 1995). The preference for particular BTE helical branching pattern over others is likely the consequence of BTE’s particular function. Thus, to complement and aid the crystallographic investigations of the BTE, I applied the “RNA LEGO” system (Horiya et al. 2002; 2003) to study coaxial stacking in the BTE.

In the “RNA LEGO” system, an RNA element containing two hairpin loops capable of forming loop-loop “kissing” interaction is used as a basic building block for formation of larger RNA assemblies in the presence of magnesium. This kissing interaction is formed when 5 to 7 bases in one loop base pair to complementary bases in another loop and is aided by base stacking of neighboring bases. The resulting kissing-loop interaction is stable in the presence of magnesium and at low temperatures (4°C). If the helices in the RNA of interest are stacked, only end-to-end kissing stem-loops will form, resulting in linear chains of RNA manifested as a ladder of high mobility bands on polyacrylamide gel (linear assemblies). If the kissing helices are not stacked, “head-to-tail” base pairing forms between BTE molecules, forming circular complexes (circular assemblies) giving just a few bands, of limited size (Fig. 1a). These two alternative interactions migrate very differently on standard non-denaturing gels.

The wild type BTE sequence off of which subsequent kissing mutant constructs were made was modified to remove 24 terminal non-essential bases and the GC rich sequence was
append to molecules 5’ and 3’ ends to give raise to more stable and compact construct 87c BTE (Fig. 2a). In order to engineer kissing-loop modules capable of polymerization, I mutated individually stem loop-I (SLI) loop (construct SLI km) and SLII loop (construct SLII km) in the 87c BTE to UGACA (Fig. 2a), making it complementary to loop III which is already known to kiss to the same 5 loop bases located in the 5’ UTR (Raktondrafara, 2006). Some of the possible BTE helical stacking interactions are outlined in figure 1b.

I hypothesized that if SLI or SLII is stacked rigidly on SLIII, only linear arrangements could form manifested as a ladder of slower moving bands (Fig. 1a) and if there is flexibility between mutated loops and SLIII then I expect to see only single species on native PAGE.

As controls, I synthesized RNAs of known interaction used by Horiya et al., 2003, which are based on the dimerization initiation site (DIS) of the human immunodeficiency virus (HIV). The 56-nucleotide (nt) V-shaped +DIS construct contains two extensively base paired hairpin loops connected by a short 2 nt linker region (Fig. 2b). Both hairpins in +DIS are topped of by a loop with embedded 6 nt GC rich self-complementary region capable of kissing module formation (Fig. 2b). Negative control (construct –DIS) contains two point mutations in the middle of the GC self-complementary region that precludes kissing complex formation (Fig. 2b).

To distinguish between different helical stacking possibilities, I used a simple native gel electrophoresis to analyze BTE kissing mutants (Horiya et al., 2003). To eliminate formation of any alternative secondary structures, all RNA samples were heated (95°C, 2 min) and quickly cooled on ice. The RNA was then allowed to fold in 1X PN buffer (Horiya et al., 2003) containing 10 mM sodium phosphate pH 7.0 and 50mM sodium chloride for 30 minutes at 30°C. To RNA solution, a triple dye-loading buffer (National Diagnostics) was
added to a final concentration of 1X and samples were run on nondenaturing 10% polyacrylamide gel (3h, 200V). After the electrophoresis RNAs were stained with ethidium bromide and visualized by irradiation with a UV transilluminator.

The results presented in figure 2c show that SLI km is forming slower moving RNA assemblies similar to control +DIS RNA in presence of 0.1mM magnesium chloride (TBM gel) suggesting that SLI is probably kissing complementary SLIII sequences and is therefore coaxially stacked on SLIII. As expected, both +DIS and SLI km dissociated into monomeric forms in the presence of 2mM ethylenediaminetetraacetic acid (EDTA, TBE gels) at increased temperatures. These RNAs also showed smearing pattern in the TBE gels indicating disengagement of kissing loop interactions.

In contrast, SLII km migrated as a single species in the presence and absence of magnesium and at all temperatures. Although equal increasing amounts of each RNA were loaded (30 and 40 pmoles per lane), SLII km stained less effectively then either control DIS RNA or 87cBTE suggesting that this sequence is less structured than the other RNAs. However, even control 87c BTE sequence that does not contain kissing module formed dimeric and larger products on native gels. These larger 87c BTE products persisted (although decreased in abundance) even at high temperatures and in the presence of chelator suggesting that they are a product of extensive base pairing rather than kissing interactions. Probably, this is a result of suboptimal folding conditions introduced by 1X PN buffer lacking magnesium and containing low concentration of monovalent cation that favors kissing interaction and is insufficient to stabilize 87c BTE native fold. Even though the absence of the three-dimensional structure makes further data interpretation difficult, occurrence of larger SLI km RNA assemblies and not SLII km indicates that helices in the
SLI km are more constrained if not held in particular arrangement where SL-I continuously stacks with SLIII.
REFERENCES


FIGURE LEGENDS:

Figure 1. a. (Taken from Horiya et al. 2003) Polymerization of coaxially stacked helices (linear arrangements) and helices that are in bent orientations (circular arrangements). b. Examples of the possible BTE helical stacking arrangements tested using RNA LEGO approach.

Figure 2. Folding of BTE RNA kissing mutants. a. Secondary structure of the 87c BTE RNA and kissing mutants showing base substitutions in red, deletions in gray and wild type bases in black. b. Secondary structure of control 56-mer RNA capable of formation of kissing loop interaction (+DIS) and double mutant with disrupted kissing interaction (-DIS) (taken and modified from Horiya et al. 2003). c. 10% native polyacrylamide gel electrophoresis (PAGE) analysis of BTE RNA kissing mutants in the presence of 0.1 mM MgCl₂ (TBM gels) or 2mM EDTA (TBE gels) run at indicated temperatures.
Fig. 1.
Fig. 2.
APPENDIX 3. CONTRIBUTION TO BOOK CHAPTER PUBLISHED IN RECENT ADVANCES IN PLANT VIROLOGY


W. Allen Miller, Jelena J. Kraft, Zhaohui Wang, and Qiuling Fan.

This book chapter surveys the most important mechanisms employed by viruses to express encoded genes. It is co-written with Dr. W. Allen Miller as the first author and Drs. Zhaohui Wang and Qiuling Fan as third and fourth author respectively.

JJK described ribosome shunting mechanism, cap-independent translation via IRES and 3’ CITES (excluding the PTEs) and drew figure 1.1. This entire manuscript was reviewed and edited by Dr. W. Allen Miller.

ABSTRACT

Cis-acting signals regulate translation of viral RNAs to produce viral proteins at the appropriate levels and timing to maximize virus replication. Here we describe the cis-acting sequences that achieve this translational control via processes such as cap-dependent translation, leaky scanning to initiate translation at more than one start codon, ribosomal shunting, cap-independent translation initiation controlled from the 5’ and/or 3’ untranslated region, poly(A) tail-independent translation initiation, stop codon readthrough, and ribosomal
frameshifting. The secondary structures and, in some cases, three-dimensional structures, of the RNA sequences that control these events are described. We also discuss the mechanisms of the translation events facilitated by the cis-acting signals, and how they mesh with the overall replication strategies of the diverse viruses that employ these mechanisms.

INTRODUCTION

All viruses are parasites of their host's translational machinery. Translation is the first event that takes place during, or immediately after, disassembly of a positive strand RNA virion upon entering the cell, because it is required for synthesis of the viral replication proteins. The vast majority of plant viruses have positive sense RNA genomes, hence for most viruses the initial step to gain a foothold in the plant cell is translation initiation. Once viral replication proteins accumulate, the positive strand RNA genome must switch its role from messenger RNA to replicase template. These events are incompatible on the same molecule so the cis-acting signals that control translation and replication are highly regulated. Late in infection, viral coat protein is translated at extremely high levels, requiring additional translational control.

This chapter focuses on cis-acting signals that control translation initiation and recoding (noncanonical reading of the genetic code). The translation initiation elements are located primarily in the untranslated regions (UTRs), while those that control recoding events such as ribosomal frameshifting and in-frame readthrough of stop codons are primarily in the coding regions near the site of the recoding event.
**Translation initiation.** Initiation is the rate-limiting, and thus most regulated, step of translation. All viruses contain sequence elements that regulate translation to some extent, often by mechanisms unknown in cellular mRNAs. Normal cellular mRNAs contain a 5' m$^7$GpppN cap and a 3' poly(A) tail. To recruit mRNA to the ribosome, the 5' cap is bound by the heterodimeric eukaryotic initiation factor (eIF) 4F, which consists of cap-binding protein eIF4E and scaffolding protein eIF4G. Poly(A) binding protein binds simultaneously to the poly(A) tail and eIF4G. These interactions can circularize the mRNA (Hentze, 1997; Wells et al., 1998b) and are thought to attract eIF3 which is bound to the 43S ribosomal preinitiation complex (Pestova et al., 2007). The ribosome then scans from the 5' end to the start codon where the 60S subunit joins, and elongation ensues. eIF4B enhances many of these interactions (Cheng and Gallie, 2006), and eIF4A participates as an RNA helicase (Svitkin et al., 2001). Many other initiation factors participate that are outside the scope of this chapter.

In order to outcompete host mRNAs for the translational machinery and perhaps to avoid host defense responses that act via translation, viruses have evolved a plethora of cis-acting elements that facilitate viral translation. Here, we discuss some examples of such exquisite and sometimes bizarre mechanisms of translational control.

**Leaky Scanning.** Many plant viruses express overlapping genes, or generate proteins with different N-termini, by inducing the ribosome to initiate translation at the first and second AUG codons on the mRNA. Ribosomes scanning from the 5' end of the mRNA may skip an AUG in a poor "Kozak context" and initiate at the second AUG if it is in a better context (Kozak, 1989), or initiate infrequently at a CUG or AUA codon in an optimal start
codon context upstream of the first AUG at which initiation also takes place. In plants, the optimal context of the start codon (underlined) is \((A/G)NNA\underline{UGGC}\) (Joshi et al., 1997; Lukaszewicz et al., 2000). Secondary structure can influence initiation codon efficiency as well, by inducing the scanning ribosome to pause on the initiation codon (Dinesh-Kumar and Miller, 1993; Kozak, 1990). Leaky scanning initiation at two AUGs in the same reading frame leads to two versions of the same protein that differ in the length of the N terminus. An example is the p95/105 polyprotein of Cowpea mosaic virus, which leads to translation of the movement protein and an N-terminally extended version of the movement protein (Holness et al., 1989). In other cases, the two AUGs are in different reading frames, leading to synthesis of unrelated proteins from overlapping ORFs. Examples include long overlaps of replication genes in Tymo- and Poleroviruses, the triple gene block ORFs translated from a single subgenomic RNA of Potex- and Hordeiviruses, and movement or silencing suppressor proteins entirely embedded in other ORFs of Tombus- and Luteoviruses (Dreher and Miller, 2006). Genes in many other plant viruses are translated via leaky scanning. For a more expansive review on leaky scanning see Dreher and Miller (2006).

**Ribosome shunting.** Ribosome shunting involves ribosomal circumnavigation around the structured regions in the 5’ UTR to allow translation of downstream ORFs (Ryabova et al., 2002). In this pathway, first discovered in Cauliflower mosaic virus (CaMV) (Futterer et al., 1993), the 40S ribosomal subunit binds to the 5’ end of the mRNA, scans in the 3’ direction until structured regions are encountered. Then the ribosomal subunit dissociates from the RNA, skips over to a nearby landing pad and resumes scanning until the start codon is reached (Fig. 1). Translation by shunting of the first long ORF (ORF VII) in
the CaMV 35S RNA requires a 5’cap, short upstream ORF (sORF) A, and a stable branched structural element (Ryabova and Hohn, 2000). Ribosomes enter the 5’ end of the 35S RNA in the normal scanning mode and begin translation of the 12 nucleotide long (nt) sORF A, just upstream of the stable structure, until the termination site is reached (Pooggin et al., 2000; Ryabova and Hohn, 2000). A portion of terminating ribosomes are able to shunt over the 480 nt branched element and resume scanning until they encounter the first AUG and re-initiate translation of downstream ORF VII. Single stranded regions that surround the stable structure, as well as the base of the helix, most likely facilitate the launching and landing of the ribosome as it skips over the 480 nt element in search of the next start codon (Dominguez et al., 1998; Pooggin et al., 2006).

Similar to CaMV, the first long ORF (ORF I) of rice tungro bacilliform virus (RTBV) is preceded by a dozen short ORF and AUG start codons (Pooggin et al., 2006). In keeping with the conventional scanning model of translation initiation, only the 5’ proximal ORF is translated, while downstream ORF I would remain silent. To circumvent the strict rules of eukaryotic translation, the RTBV encodes short ORFs and secondary structures that act synergistically to promote expression of ORF I. Specifically, translation initiates at the first AUG of 21nt sORF 1 and terminates just few bases away from the large Y shaped secondary structure (Fig. 1). This secondary structure bridges the distant regions of the RTBV leader and positions the long ORF I in close proximity to the sORF I termination site. This allows terminating ribosomes to skip over the structured bases containing small ORFs and to initiate translation at the non-AUG start codon of ORF I. Efficient recognition of the unusual start codon is the result of the stalled shunted ribosomes (Pooggin et al., 2006).
The arrangement of *cis*-acting elements and secondary structure described for CaMV and RTBV is well conserved suggesting that all other plant pararetroviruses utilize a similar mechanism of translation initiation. Furthermore, the shunting process was subsequently discovered to be used by human viruses such as adenovirus and even by cellular mRNAs (Yueh and Schneider, 1996), indicating that this might be a widespread mechanism of translation utilized by mRNAs with complex leader sequences.

**Cap-dependent translation enhancers in the 5' UTR.** Many viral RNAs with capped 5' ends have very efficient 5' leaders. The prototype is the 68 nt Ω (omega) sequence in the 5' untranslated region (UTR) of Tobacco mosaic virus (TMV) which promotes highly efficient translation initiation in vitro and in vivo (Gallie et al., 1988; Gallie et al., 1987). The Ω sequence interacts with the host heat shock protein 101 (HSP101) (Wells et al., 1998a) to recruit eIF4F; the result being similar to that of the eIF4E-5’cap and PABP-poly(A) complexes in recruiting the initiation complex (Gallie, 2002). Ω is characterized by the absence of G residues, which eliminates G-C and G-U base pairs. The resulting lack of strong secondary structure probably facilitates efficient ribosome scanning.

Ω is one of the most efficient mRNA leaders known and is used widely for high level expression of transgenes in transgenic crops. It is perhaps surprising that the 5' leader of TMV genomic RNA is more efficient than the leader on the viral coat protein (CP) subgenomic mRNA, because enormous amounts of coat protein are synthesized in infected cells, while the genomic RNA serves as message for the replication proteins (p126 and p183, the RNA-dependent RNA polymerase) which are needed at far lower amounts than the CP. Disassembly of the rod-shaped TMV virion is thought to occur co-translationally as the
ribosome strips the CP as it advances on the viral RNA (Shaw et al., 1986). Perhaps a very powerful translation element is needed to give the ribosome a toehold on the 5’ end of the otherwise encapsidated RNA.

The leader of Potato virus X potexvirus (PVX) RNA which stimulates translation contains an unstructured 41 nt alpha domain which lacks G residues, followed by domain beta (Smirnyagina et al., 1991; Zelenina et al., 1992). It was proposed that an essential CCACC motif in the alpha domain base pairs with the 3’ end of 18S ribosomal RNA to facilitate translation in a prokaryotic-like fashion (Tomashevskaya et al., 1993). The 5' UTR of Alfalfa mosaic virus coat protein RNA is another powerful leader (Jobling and Gehrke, 1987). Probably owing to the fact that it is only 36 nt long and unstructured (Gehrke et al., 1983), it has reduced dependence on eIF4 initiation factors (Browning et al., 1988).

In a yeast system that permits replication of Brome mosaic virus (BMV) RNAs, a 31 nt tract in the 5' UTR of BMV RNA2 was shown to confer dependence on the yeast helicase and translation factor, Ded1 (Noueiry et al., 2000). This selective dependence of BMV RNA2, out of the four BMV mRNAs, on a specific host factor may allow selective regulation of expression of the RdRp, which is encoded by BMV RNA2.

**Poly(A) tail-independent translation**

*TMV pseudoknot-rich domain.* The RNA genomes of viruses in the Bromoviridae, Tombusviridae and Luteoviridae families, and the *Tobamovirus, Furovirus, Tobravirus, Pecluvirus, Pomovirus, Benyvirus, Sobemovirus, Umbravirus, Tymovirus* and *Hordeiviruses* genera lack a poly(A) tail. Other sequences in the 5’ UTR or 3’ UTR can replace the function of a poly(A) tail in translation initiation and RNA stability (Kneller et al., 2006).
Toward the 5' end of the 3' UTR of TMV lies a pseudoknot-rich domain that functionally substitutes for a poly(A) tail (Gallie and Walbot, 1990; Leathers et al., 1993), perhaps by binding a common factor that also binds to the TMV 5' UTR (Tanguay and Gallie, 1996).

**tRNA-like structures.** The 3' termini of viral RNAs in the Bromoviridae family and the Tobamovirus and Tymovirus genera can form tRNA-like structures (TLS) (Fig. 2). TLSes act as telomeres by serving as a template for the host CTP:ATP nucleotidyl transferase which adds CCA, in a nontemplated fashion, to the 3' ends of TLSes as it does on host tRNAs (Rao et al., 1989). Similarly, TLSes can be aminoacylated with a cognate amino acid. The TLS is the initiation site for synthesis of negative strand RNA (Bujarski et al., 1985; Dreher, 2009; Dreher et al., 1984), the origin of assembly for BMV (Choi et al., 2002) and is required for translation initiation on TYMV RNA (Matsuda and Dreher, 2004). Using only in vitro assays, Barends et al. (2003) claimed that the valine which is attached to the TLS at the 3' end of the Turnip yellow mosaic virus (TYMV) genome, is used to cap-Independently initiate protein synthesis at the 5' end of the genome by what they call a "Trojan Horse" model of translation initiation. In contrast, the same lab found that TLSes from another Tymovirus, as well as TLSes of BMV and TMV did not incorporate their cognate amino acids into viral protein (Barends et al., 2004; Rudinger-Thirion et al., 2006). Moreover, a series of in vitro and in vivo experiments by the Dreher lab provided compelling evidence that the Trojan Horse model is incorrect, and that initiation of TYMV RNA translation relies on conventional cap-dependent mechanism via leaky scanning (Matsuda and Dreher, 2007). Dreher's lab previously showed that the TLS is a translation enhancer that (i) stimulates translation synergistically with an upstream pseudoknot domain in the 3' UTR, (ii) requires binding by translation elongation factor eEF1A, and (iii) is most efficient
when aminoacylatable (Matsuda et al., 2004). Thus, the TLS may replace the need for a poly(A) tail, but it stimulates translation by a different mechanism. The 3’ UTR of BMV RNA contains a tRNA-like structure which was also shown to stimulate translation (Zeenko, et al., 2002; Barends, et al., 2004).

**Alfalfa mosaic virus.** The 3’ termini of AMV and ilarvirus genomic RNAs fold into a series of six stem-loops linked by an AUGC repeat motif (Fig. 2B). AMV coat protein (CP) binds this structure to stimulate translation (Neeleman et al., 2004). X-ray crystallography revealed that binding of the CP compresses the RNA into a novel, compact structure in which the U and C of each AUGC repeat base pair to the A and G of the downstream repeat in a conformation that requires sharp kinks in the phosphodiester backbone (Guogas et al., 2004). For efficient translation, this structure can be replaced with a poly(A) tail (Neeleman et al., 2001). The CP also binds the eIF4G and eIFiso4G subunits of eIF4F and eIFiso4F, respectively (Krab et al., 2005). Thus, CP may circularize the RNA by simultaneously binding 5’ cap-bound eIF4F and the 3’ terminus of AMV RNA, similar to the way PABP simultaneously binds eIF4F and the poly(A) tail (Krab et al., 2005).

This AMV 3’ terminus can form an alternative conformation with a pseudoknot (Fig. 2B), that has some properties of a TLS (Olsthoorn et al., 1999). It can be substrate for the CCA-adding enzyme, but it cannot be charged with an amino acid. It was proposed that the alternative conformation acts as a switch from a CP binding site which facilitates translation, to a replicase binding site in the TLS conformation (Olsthoorn et al., 1999). Switching is important because translation and replication on the same RNA molecule are mutually exclusive. However, Petrillo et al. (2005) showed that CP bound to the 3’ end equally well under conditions predicted to form the pseudoknot. This, along with other, mostly negative,
evidence led the authors to propose a more general "3' organization model" in which different viruses evolved different ways of organizing the 3' end. The observations of the two labs may not be incompatible. It is possible that, rather than formation of the pseudoknot preventing CP binding, instead localization of the viral RNA into the tonoplast, where AMV RNA synthesis takes place, by the replication proteins may somehow cause the CP to dissociate (Bol, 2003). Thus, dissociation of the CP may allow the pseudoknot to form rather than the other way around. Because of substantial similarity in sequence and structure of the 3' ends of ilarviruses and AMV, and ability of Tobacco streak ilarvirus CP to stimulate translation of AMV RNA (Reusken et al., 1995), the above mechanisms likely apply to the Ilarviruses as well.

**Cap-independent translation**

*Internal ribosome entry sites (IRESes).* Certain viral and cellular mRNAs utilize internal ribosome entry sites (IRESes) elements for direct recruitment of translation machinery to or near the start codon to efficiently express genes (Doudna and Sarnow, 2007; Jang, 2006). An assay used frequently to assess IRES activity of an RNA sequence involves placement of the potential IRES element between two reporter ORFs in a capped dicistronic construct. Translation of the first ORF proceeds via canonical cap-dependent mechanism, while the downstream ORF is translated only if the intercistronic sequence is capable of recruiting translation machinery directly to the start codon. Even in cases when the translation of the first ORF is abolished due to the absence of the cap and/or presence of a stable stem loop structure, translation of the downstream ORF is unchanged if the intercistronic sequence is an IRES (Poyry et al., 2001).
IRESes vary in their dependence on translation factors. Almost all IRESes function in the absence of eIF4E, picornavirus IRESes require eIF4G, Hepatitis C virus does not require any eIF4 factors, and the intergenic region IRES of dicistroviruses requires no translation initiation factors (Costantino et al., 2008; Pestova et al., 2004). Unlike these animal virus IRESes which are often 250 to 650 bases long and highly structured, plant viral RNAs harbor structurally simpler cap-independent translation elements (CITEs), usually about 100 bases long, that can be found at the either end of the viral genome. Many of them do not confer internal ribosome entry, and thus are not IRESes.

**Plant virus IRESes.** Members of the large Potyviridae family including tobacco etch virus (TEV) contain a viral genome-linked protein (VPg) covalently attached to the 5' end instead of a cap structure, yet they are translated efficiently. Deletions studies revealed that TEV 5’ leader sequence contains two cap-independent regulatory elements (CIREs) that are needed for optimal translation initiation (Niepel and Gallie, 1999). The TEV leader sequence as well as each CIRE promoted translation of the second ORF when placed in a dicistronic reporter construct suggesting that they function as an IRES. Under competitive conditions when the levels of eIF4F were reduced, the TEV IRES provided a translational advantage over non-viral mRNAs due to its ability to bind and sequester the eIF4G subunit of eIF4F (Gallie, 2001). Under conditions when levels of eIF4G are plentiful, this competitive advantage is lost. The 5’ leader of another potyvirus, the Potato virus Y (PVY), also contains an IRES that directs efficient translation of second ORF from the intercistronic region (Levis and Astier-Manifacier, 1993). IRES mapping studies showed that the last 55 bases of the 185 nt 5’UTR of PVY RNA are sufficient for full IRES activity in plant cells (Yang et al.).
Similar to the potyviral RNA, 5’ leader sequences of both genomic RNAs (gRNA) of Blackcurrant reversion virus (BRV, family Comoviridae) contain IRES elements that mediate translation by recruiting 40S ribosomal subunit directly to 5’ element (Karetnikov and Lehto, 2007). The translation of BRV RNA is stimulated both by presence of UC-rich tracts complementary to 18S rRNA and by formation of a closed loop structure via long-distance base-pairing between the 5’ and 3’ UTRs that may promote ribosomal recycling (Karetnikov et al., 2006). This IRES is novel in that it appears to be split, having essential components in the 5’ and 3’ UTRs (Karetnikov and Lehto, 2008).

Unlike the above-mentioned IRESes that stimulate cap-independent translation from the viral genomic RNA, synthesis of the CP and movement protein (MP) of TMV strain U1 occurs via internal ribosomal entry from the mono and dicistronic subgenomic RNA (sgRNA) respectively (Skulachev et al., 1999). Crucifer infecting tobamovirus (crTMV) contains a 148 nt IRES upstream of the CP coding region. This IRES forms a stable hairpin structure flanked by GAAA repeats that contribute the most towards IRES activity (Dorokhov et al., 2002). The first 25 nucleotides of CP IRES overlap with the MP coding region that is also expressed via an IRES. The 75 bases long MP IRES was also functional when placed in the TMV construct that lacks its own IRES indicating strong IRES activity (Dorokhov et al., 2002).

An IRES was discovered unexpectedly in the middle of ORF 1 of Potato leafroll polerovirus (PLRV). ORF1 encodes a protease, the genome-linked protein, and the N-terminal portion of the RdRp which requires translation of ORF2 by frameshifting to be expressed. The IRES facilitates translation of a small overlapping reading frame, encoding replication-associated protein 1 (Rap1) (Jaag et al., 2003), which is essential for virus
replication. A sequence, GGAGAGAGAGG, located downstream of the start codon was found to be required for IRES activity (Jaag et al., 2003). This is unusual in its location downstream of the start codon, but it resembles the CrTMV IRES in that it is a polypurine tract.

**Cap-Independent translation elements in the 3’UTR.** Unlike IRES elements that stimulate translation from the 5’UTR, many plant viruses contain sequences in 3’UTR that confer efficient translation initiation at the 5’ end. To date, several structurally unrelated classes of CITE have been identified in plant viral 3’UTRs (Miller et al., 2007). There seems to be little conservation of sequence or structure, between classes of these elements. Where known, the different elements all seem to act by binding in different ways to eIF4F.

A long, bulged hairpin-like translation enhancer domain (TED) of satellite tobacco necrosis virus (STNV) was the first discovered 3’ CITE (Danthinne et al., 1993; Timmer et al., 1993). TED resides in the 5’ end of the 600 nt 3’ UTR but it allows efficient initiation from the 5’ AUG. The terminal loop of the TED is complementary to a loop in the 5’ UTR, but experiments designed to disrupt and restore this base pairing did not support the existence of this long-distance base pairing (Meulewaeter et al., 1998). The proposed mechanism of TED mediated translation initiation involves direct binding of eIF4F/iso4F to TED followed by interaction with 5’ end-bound 43S ribosomal pre-initiation complex either through a protein bridge or RNA-RNA interaction (Gazo et al., 2004).

CITEs that are superficially similar to the TED in secondary structure are present in the 3’ UTRs of Melon necrotic spot virus (MNSV) (Truniger et al., 2008) and Maize necrotic streak virus (MNeSV) (Scheets and Redinbaugh, 2006) (both in the Tombusviridae). All three elements are predicted to form a bulged stem-loop of about 100 nt (Fig. 3). However,
there are no apparent conserved bulges, loops or sequences among these structures, and the structural and sequence requirements have not been determined for the MNSV and MNeSV elements. Genetic evidence revealed that the MNSV element has a striking dependence on eIF4E. In melon, recessive resistance to MNSV mapped to point mutations in eIF4E (Nieto et al., 2006). Mutations in the 3' CITE of MNSV allow it to break this resistance (Truniger et al., 2008). Using in vitro translation extracts dependent on added eIF4E, the 3' CITE from the MNSV strain that could not infect the resistant plants facilitated cap-independent translation only in the presence of the eIF4E from the susceptible plant and not the resistant plant. In contrast, the 3' CITE from the resistance-breaking MNSV strain permitted cap-independent translation in the presence of eIF4E from both the susceptible and resistant plants (Truniger et al., 2008). Thus it is likely that the MNSV translation element interacts with and requires eIF4E.

Barley yellow dwarf virus-like (BYDV) cap-independent translation elements (BTEs) form one of the largest classes of 3' CITEs. Members of this class are found in genus *Luteovirus* (but not in the other Luteoviridae genera), in the Diantho- and Necrovirus genera of the Tombusviridae, and in two viruses in genus *Umbravirus* (ZW, submitted). The identifying features of this class include a 17 nt consensus sequence, GGAUCCUGGGAAACAGG, and a stem-loop sequence capable of “kissing” the complementary bases in the 5'UTR (Miller et al., 2007). The BYDV BTE adopts a cruciform secondary structure with 3 helices held together by a fourth stem that connects the BTE to the rest of the viral genome (Guo et al., 2000) (Fig. 3). In other BTEs, the number of helices radiating from the connecting stem range from two (necroviruses) to five (dianthoviruses). To allow expression of encoded genes, the BTE must first recruit the eIF4F
complex through direct binding to eIF4G and then deliver it to the 5’ UTR via long distance base pairing involving the loop 3 in the BTE and a 5’ complementary stem loop (Rakotondrafara et al., 2006; Treder et al., 2008). Presumably, once translation initiation factors are delivered to the 5’UTR, assembled 43S preinitiation complex enters at the 5’ end and begins scanning until the start codon is encountered and elongation takes place (Rakotondrafara et al., 2006; Treder et al., 2008).

Behavior of BTEs isn't always as straightforward as described for BYDV. Sequences designed to restore disrupted 5’-3’ base pairing of the TNV-A (Meulewaeter et al., 2004) and TNV-D (Shen and Miller, 2004), and Red clover necrotic mosaic virus RNA 1 BTE (called 3’TE-DR1) (Sarawaneeyaruk et al., 2009) did not restore translation. Thus suggests that both the primary sequence and complementarity is important or the mutant RNAs misfolded in unpredictable ways. An important caution on assay systems was revealed by Sarawaneeyaruk et al. (2009). They found the sequence and structure of the 5’ UTR was crucial for cap-independent translation mediated by the 3’TE-DR1 in tobacco (BY-2) protoplasts but not in cowpea protoplasts. This was in part due to the 5’ UTR being necessary for stability in BY-2 but not cowpea cells, but also due to unexplained differences in translation, perhaps different levels of eIF4F in the two cell cultures, or different affinities of translation factors of the different plant species for the DR1-3’TE. Regardless of the cause, these observations provide a warning about interpretation of results from a single assay system.

An approximately 100 nt element called the PTE was discovered in the 3’ UTRs of Panicum mosaic panicovirus (PMV) (Tombusviridae) (Batten et al., 2006), and Pea enation mosaic umbravirus (PEMV) RNA2 (Wang et al., 2009). We also predict that PTEs are present in other panicoviruses such as Cocksfoot mild mosaic virus (Ziegler et al., 2009) and
in the carmoviruses (Z. Wang unpublished observation). Structure probing by SHAPE (selective 2’-hydroxyl acylation analyzed by primer extension) (Wilkinson et al., 2006), along with site-directed mutagenesis followed by translation assays, indicate that the PTEs form a structure with a bulged helix branched into two stem loops on the top (Wang et al., 2009). A C–rich bulge between the stem loops interacts with a G-rich bulge in the helix to form an RNA pseudoknot (Fig. 3). The base pairing, as well as the sequences of bases in the pseudoknot are crucial for cap-independent translation activity, but the lengths and sequences of the stem loops can vary and are not phylogenetically conserved ((Wang et al., 2009).

The PTE binds and requires the eIF4E subunit of eIF4F, and it binds only weakly to eIF4G. Thus, the PTE is a different class of CITE from the BTE or the STNV TED. The wild type and mutant forms of the PEMV-2 PTE bind eIF4E with affinities that correlate with their function (Wang et al., 2009). Mutations to key cap-interacting amino acids in the cap-binding pocket of eIF4E greatly reduced the PTE-eIF4E binding affinity, suggesting the PTE may interact with the cap-binding pocket (Wang et al., 2009). This is remarkable, as it is the first natural RNA known to bind eIF4E with high affinity in the absence of a m7G modification.

Tomato bushy stunt virus (TBSV) and other tombusviruses contain a Y-shaped RNA domain (R3.5) in the 3’ UTR of its genomic RNA (Fabian and White, 2004, 2006). Base pairing between a loop (SL-B) in this RNA and a loop in the 5’ UTR (SL3 in the T-shaped domain) is necessary for full cap independent translation activity (Fabian and White, 2006) (Fig. 3). The sequence and the secondary structure in R3.5 are both important for its activity. The Y-shaped 3’ CITE of TBSV confers cap-independent translation in tobacco protoplasts, but not in wheat germ translation extracts (Fabian and White, 2004). In contrast, a very
similar Y-shaped domain from Carnation Italian ringspot tombusvirus (CIRV) confers substantial cap-independent translation in the wheat germ extract (Nicholson and White, 2008). The authors attribute this difference to misfolding of TBSV RNA in vitro, and a requirement for in vivo conditions or factors to properly fold the TBSV BTE.

Yet another 3' CITE, called a T-shaped structure (TSS), exists in Turnip crinkle carmovirus (TCV), another member of the Tombusviridae. This element is part of a complex region of multiple bulged stem-loops containing four pseudoknots in the 3' end of the TCV genome. Exhaustive mutagenesis distinguished domains required for translation from those required for replication (McCormack et al., 2008). Solution structural analysis by in-line probing combined with computer modeling revealed a candidate T-shaped 3D structure that vaguely resembles a tRNA (Fig. 3) (McCormack et al., 2008). Unlike all the 3' CITEs discussed above, no base pairing between the TSS and the 5' UTR is predicted. Using a yeast ribosome binding assay, the TSS appears to function by binding directly to the 60S ribosomal subunit (Stupina et al., 2008). The authors propose that the 40S subunit binds the 5' UTR, and upon reaching the start codon the 60S subunit is delivered from the 3' end by the TSS. How the 40S subunit is recruited to this uncapped RNA is unknown.

As mentioned above, BRV RNAs contain 5' sequences that function as an IRES (Karetnikov et al., 2006; Karetnikov and Lehto, 2007) it also contains novel 3' CITE that are essential for the function of the 5' IRES in plant cells (Karetnikov and Lehto, 2008). In order for 3' CITE of BRV RNA2 to stimulate translation from 3'UTR it must base-pair with complementary sequences in the 5'UTR. Specifically, 9 bases of a predicted 5' stem loop were found to “kiss” the complementary stem loop 1 bases in the BRV CITE (Karetnikov and Lehto, 2008). Presence of a poly(A) tail in reporter mRNAs stimulated translation
several fold suggesting major role in CITE mediated translation. This element appears to be present in both RNAs of all Nepoviruses. This is noteworthy in that it is the only 3' CITE not in a Tombusviridae-like virus. Unlike the other 3' CITE-containing viruses, Nepoviruses are picornavirus-like in sequence of polymerase gene, polyprotein gene expression strategy and they contain a 5' VPg and a 3' poly(A) tail. The structure of the BRV 3' CITE and mechanism by which it stimulates translation awaits future research.

The observations to date lead to interesting questions. For example why is base pairing to the 5' UTR required in some cases but not others? With the exception of TCV, it appears that the vast majority of the 3' CITEs have a loop sequence of usually 5-6 bases predicted to base pair to a sequence in the 5' UTR, even in umbraviruses where the 5' UTR is only 10 nt long. In some cases (BYDV, TBSV, BRV) this base pairing is necessary, but in other cases (STNV, RCNMV RNA1, PEMV RNA2) the predicted long-distance base pairing was not necessary. We speculate that the complementarity is necessary because it is so universally conserved, but that the in vitro and in vivo assay conditions are such that translation of some RNAs can occur even without the base pairing. This seems to be especially true for RNAs with short 5' UTRs, such as umbraviruses or STNV, which would be predicted to have a reduced requirement for eIF4F (Browning et al., 1988; Pestova and Kolupaeva, 2002).

Another interesting question is the reason for location of the 3' CITEs in the 3' UTR. We propose that this location allows the replicase, which initiates at the 3' end, to turn off translation initiation far upstream by disrupting the 3' CITE (Barry and Miller, 2002; Miller and White, 2006). This would ensure there are no ribosomes on the RNA to be replicated. Ribosomes have been shown to block the replicase and prevent negative strand synthesis on
poliovirus RNA (Barton et al., 1999; Gamarnik and Andino, 1998). Indeed Yuan et al. (Yuan et al., 2009) have shown that binding of the TCV RdRp to the 3’ end of TCV RNA dramatically alters the structure of the YSS, which would be expected to disrupt translation enhancing activity.

**Recoding.** Recoding events such as ribosomal frameshifting and stop codon readthrough facilitate translation of many plant virus genomes. Usually less than 10% of the translating ribosomes undergo the recoding event, while most ribosomes are unaffected by the recoding signals and do not change reading frame or they stop at the first stop codon. This results in translation of only small amounts of the protein that depends on the recoding event, and large quantities of the amino-terminal portion of that protein.

**Stop codon readthrough.** Plant viruses in the Tombusviridae family, and the Tobravirus, Tobamovirus, Furovirus, Pomovirus and Tymovirus genera employ stop codon readthrough for translation of the catalytic domain of the RdRp (Beier and Grimm, 2001; Pelham, 1978). Viruses in the Luteoviridae family (Brown et al., 1996), and the Benyavirus, Pomovirus, and Furovirus genera generate a C-terminal extension of the coat protein (CP) by stop codon readthrough (Beier and Grimm, 2001). This carboxy-terminal extension to the CP allows virus transmission by aphid (Brault et al., 2005) or fungal (Adams et al., 2001) vectors.

Using a construct that required in-frame readthrough of the stop codon for expression of a reporter gene, Skuzeski et al. (1991), showed that the sequence CARYYA (R = purine, Y = pyrimidine) immediately following the UAG stop codon is necessary and sufficient for readthrough of the first (p126) ORF stop codon to translate the RdRp of TMV. This
CARYYA consensus sequence is also adjacent to leaky stop codons in the Benya-, Pomo- and Tymovirus genera (Beier and Grimm, 2001). A second readthrough motif, UGACGG, where UGA is the leaky stop codon, is present in the Tobra-, Peclu-, Furo and Pomovirus genera, and in the Alphavirus genus of animal viruses (Beier and Grimm, 2001). The cytidine residue adjacent to the UGA stop codon is the most important stimulator of readthrough (Li and Rice, 1993). The cis-acting signal for readthrough of the Luteoviridae CP is much more complex. It requires a tract of eight or more CCNNNNN repeats, beginning about 16 nt downstream of the leaky CP ORF stop codon, along with an ill-defined downstream element located about 720 nt downstream of the stop codon (Brown et al., 1996).

Certain suppressor tRNAs are preferentially incorporated at the leaky stop codons of TMV and Tobacco rattle virus (Zerfass and Beier, 1992a, b). However, the mechanisms by which the cis-acting signals cause the ribosome to insert an aminoacyl tRNA instead of the release factor at a stop codon are not understood for any of the many eukaryotic viruses that employ leaky termination to control viral protein synthesis. See detailed review by (Beier and Grimm, 2001).

**Ribosomal frameshifting.** Frameshifting allows translation of overlapping reading frames by an entirely different process from readthrough. Plant viruses in the Luteoviridae family (Barry and Miller, 2002; Paul et al., 2001; Prufer et al., 1992), and the Dianthovirus (Kim and Lommel, 1998) and Sobemovirus (Lucchesi et al., 2000; Makelainen and Makinen, 2005; Makinen et al., 1995; Meier and Truve, 2007; Tamm et al., 2009) genera, as well as Nidovirales, astroviruses and retroviruses in the animal kingdom (Brierley and Dos Ramos, 2006), use a minus one (-1) frameshift to translate the catalytic domain of the RdRp. Until
recently, it was thought that in most sobemoviruses the ribosome frameshifts out of the RdRp ORF, while in others it shifts into the RdRp ORF. Recent resequencing of sobemoviral genomes revealed that, like all of the above frameshifting viruses, the ribosome shifts into the RdRp ORF in all sobemoviruses (Meier and Truve, 2007).

Minus one frameshifting is brought about by two adjacent features in the viral RNA. The first is a shifty heptanucleotide at the actual site where the ribosome slips back one base relative to the mRNA. This heptanucleotide usually fits the consensus N NNX XXZ, where N is any base, X is A or U, and Z is any base except G and gaps indicate spaces between the codons in the unshifted (zero) reading frame. The second feature begins five to six bases downstream of the shifty site and consists of a highly structured region of RNA, usually a stable pseudoknot, but sometimes a simple bulged stem-loop. In the Enamovirus and Polerovirus genera of the Luteoviridae, this frameshift stimulator is a compact, well-characterized pseudoknot of only 28-30 nt (Giedroc and Cornish, 2009; Kim et al., 1999) (Fig. 4). In contrast, viruses in genus Luteovirus (Luteoviridae) and Dianthovirus (Tombusviridae) and probably Umbravirus have a long (80-90 nt) stem loop with large single-stranded bulges (Miller and Giedroc, 2010). In BYDV (Luteovirus) RNA, one of these bulges base pairs to a stem-loop located four kilobases downstream in the 3’ UTR (Fig. 4), and this interaction is required for frameshifting (Barry and Miller, 2002). Such long-distance interactions have been predicted for the Dianthovirus and Umbravirus frameshift enhancing structures (Miller and Giedroc, 2010). In the sobemoviruses, a simple stem-loop downstream of the shifty site suffices for % frameshifting in vitro (Tamm et al., 2009) (Fig. 4).
The mechanism of -1 frameshifting is understood much better than in-frame readthrough. When the 80S ribosome translating the first ORF (reading frame zero) reaches the shifty site, both tRNAs along with the ribosome are thought to simultaneously slip one base upstream relative to the mRNA. Translation resumes in the new (-1) reading frame (Plant et al., 2003). The shifty site permits the tRNA anticodons to pair at at least two of the three bases of the codon after the shift. The unpaired bases, if any, are in the wobble (third) position of the codon. The slippage is induced by the downstream structured region in the mRNA which is thought to create tension between the moving ribosome and the mRNA as it transiently "jams" the ribosome (Namy et al., 2006). This tension is relieved by slippage of the ribosome and tRNAs, which results in frameshift (Plant et al., 2003). It is important to bear in mind that most of the time, the ribosome successfully melts out the RNA structure without frameshifting.

Translation of the RdRp gene of viruses in the Closteroviridae family requires a net +1 change in reading frame (Karasev, 2000). This would be the first known frameshift on any viral RNA that is not -1. No clear cis-acting signal has been identified, but it is proposed that the ribosome pauses at a rare codon or a stop codon, causing it to pause and occasionally change reading frames without terminating (Karasev, 2000). This is the mechanism for +1 frameshifting by Ty1 retrotransposon in yeast (Belcourt and Farabaugh, 1990). However, the possibility that the ribosome instead undergoes a -2 frameshift or a +4 or longer hop on closteroviral RNA has not been ruled out. The RNA of Potato virus M (Carlavirus) induces a ribosome stop-start at the sequence AUGA in which the ribosome appears to slip back one base at the UGA stop codon and reinitiate at the AUG sequence without releasing the nascent peptide, thus undergoing a frameshift (Gramstat et al., 1994). Finally, -1 frameshifting may
be required for expression of a recently discovered small ORF in the Potyviridae, called PIPO (Chung et al., 2008). A conserved sequence ($G_{1-2}A_{5-6}$) at the beginning of the PIPO ORF may be a new kind of frameshift signal. No conserved secondary structure downstream of this site has been identified, despite much bioinformatic analysis. For a more comprehensive description of RNA structures that induce frameshifting by plant viral RNAs see the review by Miller and Giedroc (2010).

**Parting thoughts.**

This review only touches on selected cis-acting control signals. It is clear that many cis-acting signals that control translation have been identified, but many of the mechanisms by which they act remain unclear. Thus a major objective of future research is to identify the cellular and viral factors that interact with these cis-acting signals, as well as subcellular localization of the RNA and associated proteins. Even with only limited understanding of mechanisms, the known cis-acting signals and their effects on viral RNA translation make clear that viral RNA sequence and structure efficiently control viral gene expression at the level of protein synthesis. Often the same sequence performs more than one function, such as coding for a protein as well as regulating its translation, or enhancing translation initiation and serving as an origin of replication. This reveals the extreme economy of the use of RNA sequence space by RNA viruses.

**ACKNOWLEDGEMENTS**

The authors apologize to those authors whose work was not presented due to space restrictions. This work was funded by NIH grant number R01-GM067104 supplemented by
funds from the American Recovery and Reinvestment Act, a grant from Dow AgroSciences, and by Hatch funds from the Iowa State University Agricultural Experiment Station.
REFERENCES


viruses. Virology 344, 185–197.


positive-stranded RNA virus. Genes Dev. 12, 2293–2304.


sequences of satel-lite tobacco necrosis virus RNA promoting translation in tobacco


stem–loop interactions mediate 5′ scanning-dependent translation by a viral 3′-cap-independent translation element. RNA 12, 1893–1906.


of cocksfoot mottle virus RNA is indispensable for programmed −1 ribosomal frameshifting. Virus Res. 146, 73–80.


FIGURE LEGENDS

**Fig. 1.1** *Cis*-acting structures that facilitate ribosomal shunting. After translation of the small ORF (sORF), the ribosome launches from the sequence spanning the sORF stop codon through the nuclotides at the base of the helix (dashed line at left), then lands on the RNA at the other end of the secondary structure (dashed line at right), resuming scanning to the large ORF which is translated to synthesize functional viral protein. Secondary structures of CaMV and RTBV elements that are skipped by the shunting ribosome are shown. See Poogin et al. (Poogin et al., 2006) for details.

**Fig. 1.2.** 3' ends of TYMV and AMV RNAs. A. Pseudoknotted tRNA-like structure at the 3' end of TYMV genomic and subgenomic RNAs. TLSes of other viruses resemble this structure in the overall base pairing scheme. B. Secondary structure and repeated sequence motif at the 3' end of AMV RNAs. Square brackets indicate novel base pairing between the stem-loops in the highly compact structure formed upon CP binding. Dashed lines indicate pseudoknot base pairing proposed to form a tRNA-like structure required for minus strand synthesis.

**Fig. 1.3.** 3' CITES of selected viruses. TED indicates the general type of structure of the translation enhancer domain of STNV and predicted for the MNSV and MNeSV CITES. PTE is the CITE in panicoviruses and PEMV RNA2. Conserved bases are shown including those that structural probing indicates form a pseudoknot (bases connected by dashed lines). BTE, BYDV-like 3' CITE in many viruses (see text). Consensus 17 base sequence is shown. Dashed line indicates a region that can consist of zero to three stem-loops radiating from the
central hub. R3.5 is the Y-shaped 3' CITE (YSS) found in TBSV and other viruses in genus *Tombusvirus*. In the above four structures, the loop in gray is known (BTE, R3.5) or predicted (TED, PTE) to base pair to the 5' UTR. A 3D model of the ribosome-binding tRNA-shaped structure(TSS) that confers cap-independent translation on TCV RNA was determined using NMR and small angle X-ray scattering (Zuo et al., 2010). The authors thank Yun-Xing Wang for providing this image.

**Fig. 1.4.** Frameshift signals. Shifty heptanucleotide at which frameshifting takes place is in italics. Viruses in genera *Luteovirus, Dianthovirus,* and *Umbravirus* are known (BYDV) or predicted to form the long-distance base pairing between the bulged stem-loop adjacent to the frameshift site and the distant stem-loop over 4 kb downstream. Beet western yellows virus (BWYV) is an example of the small pseudoknot formed by viruses in genera *Polerovirus* and *Enamovirus*. Base triples, including one with a positively charged cytidine residue (C+) are shown. Cocksfoot mottle virus (CfMV) is an example of the sobemovirus frameshift element.
Fig. 1.2

A

TYMV TLS

B

AMV 3’ end
Fig. 1.3
Fig. 1.4
ACKNOWLEDGMENTS

First, I would like to thank my academic advisor and major professor Dr. W. Allen Miller for his guidance and support for all the years of my undergraduate and doctoral studies. Without his help, this dissertation would not have been possible. I am deeply grateful for the long discussions that helped me sort out the technical details of my work and for holding me to a high research standard and teaching me how to do research.

Also, I would like to thank my committee members, Drs. Cathy Miller, Gustavo MacIntosh, Mark Hargrove, Alan Myers and Marit Nilsen-Hamilton for meeting with me on such a short notice and for their constructive critiques throughout my graduate career. I am deeply indebted to Dr. Marit Nilsen-Hamilton and Dr. Mark Hargrove for their laboratory support of my X-ray crystallography work. I owe a special thanks to Dr. Nilsen Hamilton’s graduate student Dr. Allison Pappas for her help with capillary electrophoresis.

I thank Dr. Julie Hoy for her practical training, supervision, and advice on X-ray crystallography experiments. Julie has supported me not only by providing technical support, but also academically and emotionally to finish this thesis.

I am grateful to my current and past colleagues and friends; Krzysztof Treder, Mariko Peterson, Elizabeth Krueger, Jimena Carrillo-Tripp, Alice Hui, Randy Beckett, Jackie Jackson, Zhaohui Wang, Aurélie Rakotondrafara, Drew Dust, and many other rotational, summer and visiting students. This inspiring group has been a constant source of friendship, good advice and collaboration.

I thank my small Serbian community in Ames for being my support group and an extension of my family. Their support and care helped me adjust to a new country and stay focused on my graduate study.
Lastly, I would like to thank my family for their unconditional love and support during good and difficult times throughout this work. Without their love and understanding I would not have gotten this far. I am thankful for my parents who supported me in all my pursuits and thought me to set goals and work hard to achieve them. Also, I am deeply indebted to them for taking care of my son while I was finishing my experiments and writing thesis. This dissertation is dedicated to my family and particularly to my son, husband, mother, and father whose constant love and support, especially during the final stages of my Ph.D. is greatly appreciated. Od srca vam hvala na svemu!