RNA-based Regulation of Transcription and Translation of Aureusvirus Subgenomic mRNA1

Wei Xu and K. Andrew White*

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

Running title: Transcription and translation of CLSV sg mRNA1

Keywords: Cucumber leaf spot virus, Tombusvirus, Tombusviridae, RNA virus, plant virus, tRNA-like, RNA structure, U-turn, RNA-RNA, RNA replication

*Corresponding author: K. Andrew White
Department of Biology
York University
4700 Keele St.
Toronto, Ontario
Canada M3J 1P3
Ph: (416) 736-2100 (ext. 40890 or 70352)
Fax: (416) 736-5698
email: kawhite@yorku.ca
Cucumber leaf spot virus (CLSV) is an aureusvirus (family Tombusviridae) that has a positive-sense RNA genome encoding five proteins. During infections, CLSV transcribes two subgenomic (sg) mRNAs and the larger of the two, sg mRNA1, encodes coat protein. Here, the viral RNA sequences and structures that regulate transcription and translation of CLSV sg mRNA1 were investigated. A medium-range RNA-RNA interaction in the CLSV genome, spanning 148 nt, was found to be required for efficient transcription of sg mRNA1. Further analysis indicated that the structure formed by this interaction acts as an attenuation signal required for transcription of sg mRNA1 via a premature termination mechanism. Translation of coat protein from sg mRNA1 was determined to be facilitated by a 5'-terminal stem loop structure in the message that resembled a tRNA anticodon stem loop. Results from mutational analysis indicated that the 5'-terminal stem loop mediated efficient base pairing with a 3'-cap-independent translational enhancer at the 3' end of the message, leading to efficient translation of coat protein from sg mRNA1. Comparison of the regulatory RNA structures for sg mRNA1 of CLSV with those used by the closely related tombusviruses and certain cellular RNAs revealed interesting differences and similarities that provide evolutionary and mechanistic insights into RNA-based regulatory strategies.
INTRODUCTION

Positive-strand RNA viruses use a variety of mechanisms for expressing their genes. For viruses in the family *Tombusviridae*, subgenomic (sg) mRNA transcription is utilized to mediate the expression of viral proteins encoded 3'-proximally in their genomes (28, 50). These smaller, 3'-coterminal messages are efficient templates for translation and allow for expression of a subset of viral proteins. Interestingly, the plus-strand RNA genomes of the members of *Tombusviridae* are not 5'-capped or 3'-polyadenylated and several have been shown to contain 3'-cap independent translational enhancers (3'CITEs) in their 3'-untranslated regions (3'UTRs) (1, 6, 9, 10, 19, 30, 31, 33, 38, 39, 41, 43, 45, 46, 51). Some of these 3'CITEs are known to function by recruiting translational machinery to the message (13, 41, 44) and several have been shown to communicate with their cognate 5'UTRs via RNA-RNA interactions (9, 10, 14, 16, 17).

Additionally, as sg mRNAs are 3'-coterminal with their genomes, they also contain the 3'CITEs and can utilize them for efficient translation of their encoded proteins (18, 29).

*Cucumber leaf spot virus* (CLSV) is a member of the genus *Aureusvirus* (family *Tombusviridae*) (27) and it is very closely related to *Tomato bushy stunt virus* (TBSV; genus *Tombusvirus*, family *Tombusviridae*) (15). These viruses share a similar coding strategy in which RNA replication-related proteins are encoded 5'-proximally in the viral genome, followed by coat protein (CP) and, finally, overlapping cell-to-cell movement and gene silencing suppressor proteins that are coded in different reading frames (Fig. 1) (26, 27, 35, 37, 50). In both cases, the CP and the 3'-proximally encoded proteins are translated from two sg mRNAs that are transcribed during infections (Fig. 1) (15, 27). For tombusviruses, translation of viral proteins has been shown to be mediated by 3'CITEs (32, 38, 51) that interact with their 5'UTRs via RNA-RNA base pairing (9, 10, 32).
Transcription of TBSV sg mRNAs occurs via a premature termination (PT) mechanism during minus-strand synthesis (48) that involves long-distance RNA-RNA interactions involving activator sequences (ASs) and receptor sequences (RSs) (Fig. 1B) (5, 20). The RSs are located just 5' to the transcription initiation sites and they base pair with complementary upstream ASs to form structures that, along with sequences immediately 3' to the RSs, cause termination of minus-strand synthesis by viral RNA-dependent RNA polymerases (RdRps) (21, 47). The sg mRNA-sized minus strands generated then serve as templates for transcription of the sg mRNAs. Previous work showed that CLSV sg mRNA2 is transcribed by a premature termination mechanism (53). Interestingly, CLSV contains functional AS2 and RS2 elements for sg mRNA2 transcription that directly correspond to those in TBSV (Fig. 1) (53). Notably, both of these interactions span over 2000 nt and, accordingly, are termed long-range RNA-RNA interactions (Fig. 1). The AS1-RS1 interaction required for sg mRNA1 transcription in TBSV spans ~1000 nt and is also categorized as a long-range interaction (Fig. 1B). CLSV presumably contains AS1 and RS1 elements necessary for sg mRNA1 transcription; however these sequences have not yet been identified (Fig. 1A).

Here we have determined the functional AS1-RS1 interaction in CLSV and find that it differs significantly from that in TBSV. The distinct AS1 and RS1 elements were shown to mediate the transcription of a sg mRNA1 containing a 5'-terminal RNA stem-loop (5'SL) that structurally mimicked a tRNA anticodon SL and was found to be critical for efficient translation of CP. These findings are discussed in relation to (i) how distinct AS and RS elements arise, (ii) how tRNA mimicry can aid viral processes, and (iii) how viruses are able to program multiple functions into a small genomic sequence.
MATERIALS AND METHODS

Plasmid construction. All wt and mutant viral genomes used in this study were derived from an infectious clone of CLSV (CLSV-JR3) (35); a generous gift from D'Ann Rochon (Agriculture and Agri-food Canada, Summerland, British Columbia). Transcripts corresponding to CLSV sg mRNA1 were generated from construct sg1. All modifications in the CLSV genome or sg mRNA1 were introduced using PCR-based mutagenesis and standard cloning techniques. The modifications present in each mutant are presented in the respective figures. The PCR-derived regions of all mutants were sequenced completely to ensure that only the desired mutations were present.

In vitro transcription. Prior to in vitro transcription, plasmid constructs corresponding to CLSV viral genome and sg mRNA1 were linearized with a restriction enzyme to define the 3’ terminus of the message (49). Full-length genome or sg mRNA1 transcripts were generated by digestion with Smal, while 3’-truncated versions of sg mRNA1 were generated by digestion with BtgZI, which generated a message lacking the 3’-terminal 313 nt of the message. These DNA templates were used along with an AmpliScribe T7-Flash Transcription Kit (Epicentre Technologies) to synthesize viral RNA transcripts.

Protoplast transfection and RNA isolation. Protoplasts from 6- to 7-day-old cucumber cotyledons were prepared as previously described (49). Approximately 3 × 10^5 protoplasts were transfected with 3 μg of each viral RNA transcript using a PEG-mediated method (49). After a 22 h incubation in a growth chamber under fluorescent lighting at 25 °C, total nucleic acids were isolated from the protoplasts as previously described (49).

Northern blot analysis. For detection of plus-strand viral RNAs, the isolated nucleic acids were separated in a 1.5% agarose gel, followed by electroblotting to membrane and
hybridization with two $^{33}\text{P}$-5′-end-labeled DNA oligonucleotide probes complementary to segments 4304–4338 and 4151–4170 of the CLSV genome (53). For detection of minus-strand viral RNAs, glyoxal-treated samples were separated in a 1.0% agarose gel and then subjected to Northern blot analysis using $^{33}\text{P}$-labeled in vitro-generated RNA transcripts corresponding to CLSV sg mRNA2 (53). Relative levels of viral RNAs were quantified by radioanalytical scanning of blots with a PharosFx Plus Molecular Imager (Bio-Rad).

**5′-end mapping of sg mRNA1 by primer extension.** Analysis of the 5′-terminus of sg mRNA1 was performed by adding 1 pmol of $^{33}\text{P}$-5′-end-labeled primer (5′CTACTATGGACTTACCATCTGATAAGCAGCACCT; position 2431-2465) to the templates derived from one-tenth volume of the total nucleic acids extracted from infected protoplasts. The 13 µl mixture was incubated at 65 °C for 5 min and transferred to 37 °C for 2 min, followed by ice for 1 min. Extension of the annealed primer was achieved by adding 6 µl of buffer (5X SSIII FS buffer (Gibco-BRL): 0.1 M DTT: 10 mM dNTP; 4:1:1; v:v:v) and 1 µl (200 units) of superscript III reverse transcriptase (Gibco-BRL), followed by incubating at 52 °C for 10 min. The sample was then treated with 1 µl of 4 M NaOH at 95°C for 5 min, 29 µl of acid stop mix (1M Tris-HCl pH 3.5: stop dye: 4:25; v:v) was then added to stop the reaction. Finally samples were separated in a 6% polyacryamide (29:1) 7 M urea sequencing gel along with a corresponding sequencing ladder generated using the same primer. Results were visualized by radioanalytical scanning of the gel with a PharosFx Plus Molecular Imager (Bio-Rad).

**Computer analysis of RNA structure.** The lowest free-energy secondary structures for viral RNAs were predicted by the computer program mfold, version 3.2 (22, 54). Large RNA structures were drawn using RnaVis2 (7).
In vitro translation. Translation of RNA transcripts was carried out in nuclease-treated wheat germ extract (Promega) in the presence of \[^{35}\text{S}\]-Met. For each reaction, a subsaturating amount (0.5 pmol) of RNA template was used. The relative level of p41 expression from sg mRNA1 was quantified using a PharosFx Plus Molecular Imager (Bio-Rad) after separation by 12 % acrylamide sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Western blot analysis. Total proteins were extracted from infected protoplast by vortexing the cells in 20 μl extraction buffer (125 mM Tris-HCL, pH 6.8, 0.1% SDS and 20% glycerol \([\text{v/v}]\)), followed by centrifugation at 10,000 xg for 5 min to collect the supernatants. Aliquots were then heated at 95°C for 5 min and separated along with the protein ladder by 10 % acrylamide SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and CP (p41) was detected using a polyclonal antibody (1:2000 dilution). HRP-conjugated secondary antibody (Sigma, A0545) and Amersham ECL Plus western blotting detection reagents (GE Healthcare, RPN 2132) were used for further processing of the blot. Membrane was covered with Saran wrap and exposed to X-ray film for 30 min before developing the film.

RESULTS

Mapping the transcription initiation site for CLSV sg mRNA1. Since the transcriptional regulatory elements for sg mRNA2 in TBSV and CLSV were found to be highly comparable in both relative genomic positions and sequence identity (Fig. 1) (53), we anticipated an equivalent level of similarity for the elements controlling sg mRNA1. Indeed, a comparison of the sequences surrounding the known initiation site for sg mRNA1 of TBSV with the corresponding region in CLSV identified a contiguous 6 nt segment of identity (Fig. 2A, box). If CLSV sg mRNA1 initiated at the same relative position as in TBSV, its site would map to a
guanylate within this conserved region (Fig. 2A, asterisk). However, a previous study reported
that CLSV sg mRNA1 initiates at G2345, which is located 7 nt upstream from the conserved
sequence (Fig. 2A) (27). To distinguish between these two positions, we performed primer
extension analysis on sg mRNA1 using total RNA isolated from CLSV-transfected protoplasts.
The results indicated that the 5’-terminus of sg mRNA1 maps to the same upstream guanylate
(position 2345) that was implicated previously (27) (Fig. 2B).

To further confirm the results from the primer extension analysis, the identified initiation
nucleotide was substituted with cytidylate, uridylate or adenylate. Full-length infectious
transcripts of CLSV harboring these single nucleotide changes (G-sg1-C, G-sg1-U and G-sg1-A,
respectively) were transfected into plant protoplasts and viral RNA levels were subsequently
monitored by Northern blotting. All substitutions had substantial negative effects on sg mRNA1
accumulation levels, consistent with the proposed role for the guanylate as the transcriptional
initiation site (Fig. 2C). Primer extension analysis performed on viral RNAs from the mutant
infections did not yield any products corresponding to the mapped initiation site, confirming that
the band seen for wt CLSV was sg mRNA1-specific (Fig. 2B). Analysis of minus strands from
the infections identified sg mRNA1-sized viral RNAs for all (Fig. 2D). The presence of minus,
but not plus, strands of sg mRNA1 in the mutant infections is consistent with a PT mechanism in
which minus-strand intermediates are synthesized first (48). That is, in these mutants, the
promoter activity responsible for positive-strand sg mRNA1 production was inhibited, while the
attenuation activity generating minus-strands was maintained.

In the mutants above, the initiating guanylate that was substituted corresponds to the
central nucleotide of the last sense codon for p84, AGA (Fig. 2A). Consequently, the alterations
made in G-sg1-C, G-sg1-U and G-sg1-A changed the C-terminal wt Arg residue to Thr, Ile and
Lys, respectively. Although it is possible that the phenotypes observed were related to these changes in the RdRp (52), the similar levels of sg mRNA2 observed for the wt and mutant genomes suggest that this was not the case (Fig. 2C). As for the notably higher accumulation level seen for sg mRNA1 minus strands in the G-sg1-A mutant (Fig. 2D), this effect has also been observed in a corresponding TBSV mutant (5). For these two cases, adenylate likely acts to enhance termination of the RdRp during minus-strand synthesis, while its complement, uridylate, does not support efficient initiation of sg mRNA1 transcription on the intermediate template generated. Collectively, the results indicate that CLSV sg mRNA1 transcription occurs via a PT mechanism, but that CLSV uses a different relative initiation site than does TBSV.

Identification of a medium-range RNA-RNA interaction that facilitates transcription of CLSV sg mRNA1. Long-range genomic RNA-RNA interactions (i.e. AS-RS interactions) are required for transcription of TBSV sg mRNAs 1 and 2 (5, 20) and also for CLSV sg mRNA2 (53) (Fig. 1). Accordingly, we anticipated a similar long-range requirement for CLSV sg mRNA1. However, mfold analysis (22, 54) of the full-length CLSV RNA genome instead predicted a potential AS1-RS1 interaction between a 7 nt long segment located 3 nt just 5' to the sg mRNA1 start site and a complementary segment positioned only 148 nt upstream (Fig. 3A). This length of intervening sequence was almost seven times smaller than that in the corresponding AS1-RS1 interaction required for TBSV sg mRNA1 transcription (Fig. 1B). Consequently, this putative CLSV AS1-RS1 interaction was categorized as a medium-range interaction.

The functional relevance of the proposed interaction was supported by comparative sequence analysis among sequenced aureusviruses, which revealed two different positions in the stem where nucleotide covariation maintained the putative AS1-RS1 base pairing interaction.
Two sets of CLSV mutants were generated to test if the proposed interaction was important for sg mRNA1 transcription (Fig. 3A). All of the substitutions were made at degenerate codon positions in the p84 ORF so that the identities of the encoded amino acids would not be altered. Modifications that were predicted to decrease the stability of the AS1-RS1 interaction resulted in reduced relative levels of sg mRNA1 in transfected protoplasts, while those designed to restore the interaction by combining the disruptive changes lead to improved levels of accumulation (Fig. 3B, C). The genomic mutants harboring GU wobble pairs (LRS1-2 and LAS1-1) yielded notably higher levels of sg mRNA1 than their counterparts with CA mismatches (LAS1-2 and LRS1-1), consistent with the AS1-RS1 interaction functioning in the positive strand of the viral genome (Fig. 3). Also, the higher relative level of recovery of LARS1-1 compared to that for LARS1-2 can be explained by the replacement of an AU base pair with a stronger GC pair in the former, and the converse exchange in the latter (Fig. 3).

Together, these results support the proposed medium-range AS1-RS1 interaction in the CLSV genome as being functionally relevant for transcription of sg mRNA1. Notably, this CLSV pairing strategy is distinct from that used by TBSV.

**Activating a cryptic initiation site for CLSV sg mRNA1.** Although our results indicated that CLSV uses a sg mRNA1 initiation site that is positioned upstream relative to that used in TBSV, we wondered whether the TBSV-like start site could have been used by CLSV in the past (Fig. 2A). Indeed, phylogenetic analysis suggests that tombusviruses (e.g. TBSV) and aureusviruses (e.g. CLSV) have evolved from a common ancestor (27). If the TBSV-like initiation site was used previously, it would have had to possess both attenuator and promoter activities to allow for production of minus- and plus-strand sg mRNA1, respectively (21). Accordingly, it would have needed a cognate upstream AS1-RS1 structure to mediate efficient
RdRp attenuation. To test whether the downstream sequence possessed attenuator/promoter properties, a stable RNA hairpin followed by a CAA spacer was inserted immediately upstream of the TBSV-like start site in the CLSV genome (Fig. 4A). The introduced RNA hairpin would provide an appropriately positioned artificial version of the required AS1-RS1 element (47). Additionally, wt sg mRNA1 transcription was inactivated in this mutant, HX78-1, by a substitution in AS1 (i.e. the same modification present in LAS1-2; Fig. 3) (Fig. 4A). Unlike LAS1-2, which produced only trace amounts of sg mRNA1, HX78-1 yielded comparatively higher levels of a sg mRNA1-sized message in transfected protoplasts (Fig. 4B, left panel). To confirm that the sg mRNA1 being produced from HX78-1 originated from the TBSV-like start site, the guanylate at this position was mutated to a cytidylate, creating HX78-1C (Fig. 4A). HX78-1C did not produce detectable sg mRNA1-sized product (Fig. 4B, right panel). Moreover, primer extension analysis that mapped the sg mRNA transcribed from HX78-1 to the TBSV-like start site did not identify a corresponding product for HX78-1C (Fig. 4C). These functional studies show that the sequence downstream of the TBSV-like initiation site possesses intrinsic attenuator/promoter activities.

Translational properties of wt and TBSV-like sg mRNA1 of CLSV. Based on the results above, it is possible that CLSV could have previously used the TBSV-like initiation site to transcribe a sg mRNA1, termed sg1-T-like, which was 7 nt shorter at its 5' end (Fig. 5A). If so, what could have been the driving force leading to use of the upstream site? TBSV successfully uses the downstream site; therefore, we conjectured that this divergence could be related to differences in translational efficiency between the two messages.

Viruses in *Tobrusviridae* are either known or proposed to contain 3'CITEs; CLSV falls into the latter category. To test whether CLSV sg mRNA1 utilized a 3'CITE, uncapped full-
length and 3'-truncated in vitro transcripts were generated and tested in a wheat germ extract system. Translation of CP (p41) from the full-length wt transcript, sg1, was ~10-fold more efficient than from its 3'-truncated counterpart, sg1-d3' (Fig. 5B, left panel). Similar results were observed when uncapped full-length and 3'-truncated transcripts corresponding to the TBSV-like sg mRNA1, sg1-T-like and sg1-T-like-d3', respectively, were tested (Fig. 5B, right panel). Accordingly, both versions of CLSV sg mRNA1 were translationally active and dependent on 3'-proximal sequences for optimal efficiency. Next, to determine which of the full-length versions directed more efficient translation, the activities of sg1 and sg1-T-like were compared. Interestingly, wt sg1 yielded ~4-fold more CP than sg1-T-like, suggesting that the extra 5'-terminal sequence enhanced its translational efficiency (Fig. 5C). This notion was further bolstered after assessing the decay rates of the different mRNAs, which did not reveal any major differences (data not shown).

**Structure-function analysis of the 5' terminus of CLSV sg mRNA1.** Mfold analysis of the 5' end of wt sg mRNA1 predicted the formation of a 5'SL (Fig. 6A). Similar structures with the same loop sequence, but a shorter stem, were also predicted for two other aureusviruses (Fig. 6A). The additional GC base pair in the shorter stem would partially compensate, in terms of structural stability, for the lost central UA pair. Also, the covarying upper base pair and the concurrent deletion of the residues in the UA pair—both of which maintain the stem—support a functional role for the proposed 5'SL (Fig. 6A).

To assess the possible role of the proposed 5'SL structure in sg mRNA1 translation of CP, wt and mutant transcripts of sg mRNA1 were analyzed in wheat germ extract. In the mutants, the central AU base pair in the stem of the 5'SL was targeted for compensatory mutational analysis (Fig. 6A). Four different sets of compensatory mutants were generated and,
in all cases, the translational efficiency of CP correlated positively with the predicted stability of
the stem (Fig. 6B). Specifically, all mismatches decreased translation below wt levels, while
replacement of the wt UA with more stable GC or CG pairs enhanced translation above wt levels
(Fig. 6B). Furthermore, no notable differences in message decay were observed for these
different mutants (data not shown). These analyses support the concept that formation of the
5'SL facilitates translation of sg mRNA1.

Assessing the role of the loop sequence in 5'SL. Translation of viral mRNAs mediated
by 3'CITEs requires communication between the 3'CITEs and their cognate 5'UTRs (29). In
TBSV and other viruses, this is mediated by an RNA-RNA base pairing interaction between the
two terminal elements (9, 10, 14, 16, 17). Structural analysis of the 3' region of CLSV sg
mRNA1 revealed a Y-shaped domain (data not shown) that was similar to the Y-shaped domain
in TBSV, which is a verified 3'CITE (9, 10). The putative 3'CITE also contained a stem-loop
that was complementary to the 5'UTRs of the CLSV genome, sg mRNA2 (data not shown) and
sg mRNA1 (Fig. 7A). To determine if the predicted 5'SL-3'CITE interaction was functionally
relevant, a set of sg mRNA1 mutants with disruptive and restorative changes in the interaction
was generated (Fig. 7A). Results from translation (Fig. 7B) and stability (data not shown) assays
performed on these mutants indicated that 5'SL-3'CITE base pairing facilitates efficient CP
translation. Notably, the complementary GACCAUU sequence in the 5'SL of sg mRNA1 is also
in sg1-T-like, but it is located at the very 5'-end of this message (Fig. 5A). This presumably less
favorable context likely accounts for the lower translational efficiency observed for sg1-T-like
(Fig. 5C).

Another observation that was consistent with the proposed loop-mediated 5'SL-3'CITE
interaction in wt sg mRNA1 was the presence of a U-turn motif, YUNR (Y, pyrimidine; N, any
base; R, purine) (11), in the loop of the 5'SL. This motif participates in intra-loop interactions that act to project the Watson-Crick edges of the bases located 3' to the uridylate outward; thereby facilitating their probability of productively encountering pairing partners (11). To test the possible relevance of this motif to translational efficiency, the conserved uridylate was substituted with all three alternative bases, generating the mutants, sg1-UtoC, sg1-UtoA and sg1-UtoG. All three changes reduced the relative level of translation of CP, with sg1-UtoA and sg1-UtoG being affected to a greater degree (Fig. 7C). The comparatively smaller decrease seen for sg1-UtoC is consistent with results from SL and RNA target binding studies, which showed that a cytidylate substitution at this position in the motif retains a relative binding rate of 0.31 (compared to 1.0 for uridylate), while adenylate and guanylate substitutions display rates of 0.095 and 0.071, respectively (12). As no major changes in RNA stability were observed for these messages (data not shown), the combined data suggest that the YUNR motif contributes to efficient sg mRNA1 translation.

In vivo analysis of sg mRNA1 translation. The translational activity of several sg mRNA1 mutants were assessed in CLSV infections of protoplasts. To accomplish this, a number of the previously-tested modifications in the 5'UTR of sg mRNA1 were introduced into the corresponding region of the infectious clone of CLSV. Both viral RNA and CP levels were then assessed by Northern and Western blot analyses, respectively, at 22 hr post-transfection (Fig. 8). All mutants tested exhibited levels of sg mRNA1 that were consistently lower than that of wt CLSV (Fig. 1A). When corresponding levels of CP were determined, mutants G-6B and G-7B (producing sg mRNA1s with disruptions in the stem of 5'SL; Fig. 6) showed low but readily detectable CP accumulation (Fig. 8B). The compensatory mutants for these disruptions in the 3' half of the stem were not generated as the necessary additional substitutions in the 5' half of the
stem would have changed the UAA stop codon for p84 to sense codons. As for the other types of mutants tested, HX78-1 (producing a sg1-T-like message; Fig. 5A), G-5’UTR4 (producing a sg mRNA1 with a destabilized 5’-3’ interaction), and G-UtoA and G-UtoG (producing a sg mRNA1 with mutated U-turn motifs), all produced levels of sg mRNA1 similar or greater than that for G-6B (Fig. 8A), however, unlike for G-6B, no CP was detectible for any of the mutants (Fig. 8B). Collectively, these data indicate that the TBSV-like sg mRNA1 is translationally non-functional in vivo, as is the 5’-3’ destabilizing mutant and the U-turn mutants; suggesting critical roles for the corresponding RNA sequences and structures. Conversely, destabilization of the stem of the 5’sL did allow for CP production, but at substantially lower levels than wt. Overall, modifications to sg mRNA1 seem to be more detrimental to translation in vivo than in vitro, which may be due to the more competitive and hostile environment of cells.

DISCUSSION

RNA viruses utilize a variety of different regulatory mechanisms to control the expression of their genes, and cis-acting RNA elements invariably play prominent roles in such processes. As one would expect, viruses that are closely related often share regulatory strategies, however, each can possess unique features that provide novel insights into mechanism and/or evolution. Here we compare and contrast our results on the RNA elements that regulate gene expression of CLSV CP in terms of the novel features that control transcription of sg mRNA1 and their subsequent impact on the structure and function of this message at the translation level.

Transcription of sg mRNA1. Our results suggest that CLSV sg mRNA1 is produced by a PT mechanism; a conclusion that is consistent with our previous findings supporting this same mechanism for CLSV sg mRNA2 transcription (53). Interestingly, although CLSV and TBSV
use essentially the same set of corresponding long-range AS2-RS2 interactions to regulate sg mRNA2 transcription (53), they employ different AS1-RS1 elements to control sg mRNA1 transcription (Fig. 9). Thus, despite predicted common ancestry and considerable genetic similarity (27), these viruses have diverged with respect to their AS1-RS1 interactions.

What selective pressure(s) led CLSV to adopt a different AS1-RS1 pairing strategy?

CLSV and TBSV generate approximately the same relative amounts of sg mRNA1 during infections (53). Therefore, there does not seem to be an obvious advantage related to transcriptional efficiency. Alternatively, it is possible that the closer linear positioning of these regulatory elements in the CLSV genome reduces the involvement of genome structure in mediating the interaction. That is, a comparatively smaller portion of the CLSV genome, defined by the intervening sequence, would need to fold in a manner that allows the AS1-RS1 interaction to form. This reduced commitment would mean that more sequence would be available to engage in other RNA-based functions which could, in turn, provide a fitness advantage. Nonetheless, even at this smaller 148 nt scale (versus 996 nt in TBSV), the structure of the intervening sequence in CLSV still likely plays an active role in facilitating the AS1-RS1 interaction, as suggested by the two mfold-predicted intervening SLs that would help to juxtapose the two elements in space (Fig. 3).

Another central issue relates to how the current AS1-RS1 pairing in CLSV was acquired. Despite some divergence, the sg mRNA1 initiation site in TBSV and some of its surrounding sequence are correspondingly present in CLSV (i.e. AGACCA; Fig. 2A). Our ability to activate this cryptic initiation site in CLSV using an artificial SL-based AS1-RS1-like structure (Fig. 4) showed that the TBSV-like segment still maintains attenuator/promoter activities that can properly guide the CLSV RdRp in the transcriptional process. Also, since the heterologous AS1
and RS1 elements functioned well, there does not appear to be any strict structural requirements for a functional AS1-RS1 interaction in this virus. Based on these observations, it is conceivable that CLSV previously utilized TBSV-like regulatory RNA elements, including a different long-range AS1-RS1 interaction, for sg mRNA1 transcription. However, subsequently, some selection pressure would have directed their replacement with the current set of regulatory elements. In vitro, we showed that the existing wt CLSV mRNA1 is more efficient at mediating translation than the TBSV-like version of this message (sg1-T-like), while in vivo the latter is unable to produce detectable amounts of CP (Fig. 9A). Currently, it is unclear how the corresponding wt TBSV sg mRNA1, that contains the sequence complementary to its 3'CITE at the very 5' end of the message (Fig. 9B), is able to successfully direct CP translation in vivo. Differences in sequence and/or structure between the TBSV and CLSV messages are likely responsible for the disparity in translational activity.

If CLSV did use a TBSV-like sg mRNA1 in the past, the message would have had to contain additional changes that allowed it to translate functionally relevant levels of CP. Even if such conditions were met, the translational activity of the TBSV-like message may still have been suboptimal for CLSV. Under these circumstances, the transition to the upstream initiation site and its cognate medium-range AS1-RS1 interaction could have been driven by an advantage provided by comparatively higher levels of CP. Indeed, one or more of the important roles played by the CLSV CP including, viral genome protection, systemic spread, and vector transmission, could have benefited from such improved CP levels (27, 35).

**Translation of sg mRNA1.** Analysis of wt CLSV sg mRNA1 revealed the presence of a 5'SL that facilitates translation by base pairing with the 3'CITE (Fig. 7). This 5'UTR-3'CITE base pairing requirement is consistent with the cap-independent translational mechanism...
currently proposed for TBSV and other viruses that possess this class of 3'CITE (9, 10, 14, 29).

The model posits that the translational machinery is recruited by the 3'CITE and is able to access the 5'UTR via a 5'-3' RNA-RNA bridge (14, 29). For both TBSV and Barley yellow dwarf virus, ribosomes load at the very 5' ends of viral messages and then scan 5'-to-3' in search of start codons (10, 34). This linear scanning is proposed to disrupt the 5'-3' RNA-RNA bridge, which would subsequently need to reform before another ribosome could be loaded at the 5' end (10, 34). Accordingly, translation via this mechanism would require rapid reformation of the disrupted 5'-3' interaction in order to maintain a productive rate of translation.

For CLSV sg mRNA1, the 5'SL structure could mediate efficient formation of the 5'-3' RNA-RNA interaction by presenting the interacting nucleotides, which are primarily located in the loop, in single-stranded form (Fig. 7A). This notion is supported by in vitro (Fig. 6) and in vivo (Fig. 8) evidence for the importance of complementarity between partner residues in the stem of the 5'SL. Moreover, the YUNR motif in the loop also appears to contribute to the efficiency of translation in vitro (Fig. 7C) and is even more relevant in vivo (Fig. 8). YUNR motifs are found in tRNA anticodon loops and in bacterial regulatory RNAs that function via base-pairing (11). Interestingly, the loop and closing base pair of the sg mRNA1 5'SL are highly homologous to those in the tRNA-\textsuperscript{−}Phe anticodon SL from yeast (36); differing by only a single nucleotide out of nine total (Fig. 10A). Furthermore, the stem lengths in both structures are the same; 5 bps long. To our knowledge, this is the first report of anticodon SL structural mimicry in an aureusvirus. In contrast, it has long been known that tRNA mimicry exists in members of the genera Tymovirus, Tobamovirus and Bromovirus, which harbor such structures in their 5'UTRs, 3'UTRs and/or intergenic regions (8). These RNA elements have been shown to function in a variety of capacities, including translational enhancement (23), regulation of minus-
strand synthesis (24), recruitment of viral RNA into replication (3) and virus assembly (4).

Structures that mimic an entire tRNA are present at the 3'-termini of these viruses (8), while 5'- and internally-located tRNA mimics in bromoviruses correspond only to a single SL of tRNA, the TpsiC SL (2). This TpsiC-like SL structure is a necessary component of a larger bromovirus RNA recruitment signal (42). For viruses in the family Tombusviridae, the carmovirus Turnip crinkle virus was recently shown to contain a tRNA-like structure in its 3'UTR (25) that binds to 60S subunits and 80S ribosomes (41). This ~100 nt long RNA element, which can be modeled into a T-shaped structure that resembles folded tRNA, is proposed to recruit the ribosome or a subunit needed for viral protein translation; however, the mode of transfer to the 5'UTR is not currently known (25, 41). In comparison, the CLSV tRNA-like structure is a single anticodon-like SL that is much smaller (17 nt), it is located in the 5'UTR of the message, and it functions in 5'-3' communication. Thus, members of two different genera in the family Tombusviridae use different levels of tRNA mimicry to perform dissimilar functions in the process of cap-independent translation.

The 5'SL in CLSV sg mRNA1 also bears significant similarity to that in HokC mRNA from E. coli, which pairs with a regulatory antisense Sok-RNA (Fig. 10A) (11). Since this type of YUNR-containing non-viral RNA element and tRNA anticodons are known to facilitate rapid base pairing interactions with their partner RNAs in trans (11), it is reasonable to propose that the YUNR-containing 5'SL in sg mRNA1 functions analogously, but in cis. Indeed, the rapid reformation of the 5'SL-3'CITE interaction following its disruption by scanning ribosomes (10, 34) would mediate efficient 5'-reloading of 43S subunits on sg mRNA1. The proposed importance of the YUNR motif in this process is further underscored by its presence in predicted 5'-proximal SLs in the 5'UTRs of both the CLSV genome and sg mRNA2 (Fig. 10B). In either
case, only the N in the YUNR is changed, thus maintaining the motif consensus (Fig. 10B). Additionally, the active involvement of the YUNR motif in this type of cap-independent process likely extends to other viruses such as TBSV and Barley yellow dwarf virus that also contain the consensus motif in their interacting loops (29).

**An extreme case of overlapping regulatory RNA elements.** Close examination of the sequence that forms the 5'SL of sg mRNA1 reveals a high level of integration of different functional and structural features into a very small region of the CLSV genome. Based on a PT mechanism for transcription and what is known for TBSV (40, 48), the 5' end of this sequence will contain nucleotides that contribute to RdRp attenuation (to generate the minus-strand intermediate), while corresponding sequence in the intermediate template will act as a transcriptional promoter. The attenuator element and transcriptional promoter have not yet been mapped for CLSV sg mRNA1, therefore their relative positions have been estimated based on results from analysis of TBSV (21) (Fig. 10C) (21). The promoter and attenuator elements would also overlap with sense and stop codons of p84 and with the sequence that forms the 5' half of the stem of 5'SL (Fig. 10C). The YUNR motif is present in the loop along with sequence that forms the 5'-3' long-range interaction with the 3'CITE (Fig. 10C). Accordingly, at least six distinct functional/structural features are integrated into this 17 nt long RNA sequence, underscoring the inherent capacity of RNA viruses to encode diverse elements within a tiny genetic footprint.

**ACKNOWLEDGEMENTS:** We thank D'Ann Rochon for the infectious clone of CLSV and CP anti-serum. We are also grateful to members of our laboratory for reviewing the manuscript. This work was supported by NSERC and CFI.
REFERENCES


41. Stupina, V. A., A. Meskauskas, J. C. McCormack, Y. G. Yingling, B. A. Shapiro, J. D. 
   Dinman, and A. E. Simon. 2008. The 3' proximal translational enhancer of Turnip crinkle virus 
   binds to 60S ribosomal subunits. RNA 14:2379-2393.

   signal functions with viral replication protein 1a to dramatically stabilize RNA in vivo. J. Virol. 
   73:2622-2632.

43. Timmer, R. T., L. A. Benkowski, D. Schodin, S. R. Lax, A. M. Metz, J. M. Ravel, and 
   K. S. Browning. 1993. The 5' and 3' untranslated regions of satellite tobacco necrosis virus RNA 
   affect translational efficiency and dependence on a 5' cap structure. J. Biol. Chem. 268:9504-
   9510.

   2008. The 3' cap-independent translation element of Barley yellow dwarf virus binds eIF4F via 
   the eIF4G subunit to initiate translation. RNA 14:134-147.

   region mimics a 5' cap in facilitating translation of uncapped mRNA. EMBO J. 16:4107-4116.

46. Wang, S., and W. A. Miller. 1995. A sequence located 4.5 to 5 kilobases from the 5' end of 
   the barley yellow dwarf virus (PAV) genome strongly stimulates translation of uncapped mRNA. 
   J. Biol. Chem. 270:13446-13452.

   requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. J. 
   Virol. 82:3864-3871.


   tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions. J. 


   located in the 3'-proximal region of the tomato bushy stunt virus genome. J. Virol. 73:8982-
   8988.

52. Wu, B., and K. A. White. 2007. Uncoupling RNA virus replication from transcription via 
   the polymerase: functional and evolutionary insights. EMBO J. 26:5120-5130.


FIGURE LEGENDS

Figure 1. Schematic representation of CLSV and TBSV RNA genomes. (A) Linear version of the CLSV RNA genome showing its coding organization. The relative positions of interacting RNA elements, AS2 and RS2, involved in sg mRNA2 transcription are shown above the genome and are denoted by double-headed arrows. Initiation sites for sg mRNA transcription are labeled sg1 and sg2 and the corresponding sg mRNAs are represented by arrows below the genome. (B) Linear version of the TBSV RNA genome showing its coding organization. The relative positions of long-range interactions, AS1-RS1 and AS2-RS2, involved in sg mRNA1 and sg mRNA2 transcription, respectively, are shown above the genome. Initiation sites for sg mRNA transcription are labeled sg1 and sg2.

Figure 2. Analysis of the initiation site of sg mRNA1. (A) Sequence comparison between the TBSV sequence containing the sg mRNA1 initiation and the corresponding segment in CLSV. A 6 nt tract of identity between the two RNA genomes is boxed. Initiation sites for sg mRNA1 (sg1) in TBSV and CLSV are denoted by arrows and genome coordinates. The position in the CLSV genome that corresponds to the TBSV sg mRNA1 initiation site is identified by an asterisk. Stop codons for p92 (TBSV) and p84 (CLSV) are italicized and underlined. (B) Primer
extension analysis of wt and mutant CLSV genomes. Primer extension products were separated
in a denaturing high-resolution gel along with an appropriate sequencing ladder. The identities
of the infections analyzed are indicated above the lanes. The position of the primer extension
product for the wt infection is denoted by an arrowhead and corresponds to coordinate G_{2345} in
the sequencing ladder. (C), (D) Northern blot analysis of plus and minus stands of wt and
mutant CLSV genomes. Wt CLSV and mutant genomes used in the transfections are indicated
above the lanes. The mock lane is a negative control and corresponds to transfection with water
only. The positions of genomic (g) and subgenomic mRNAs (sg) are indicated to the left.
Detection of positive- and negative-strand viral RNAs are indicated by (+) and (−), respectively.
In this and all other experiments, uncapped transcripts of wt and mutant genomes were
transfected into cucumber protoplasts and viral RNAs were analyzed by Northern blotting
following a 22 hr incubation at 25 °C.

Figure 3. Compensatory mutational analysis of the proposed AS1-RS1 interaction in
CLSV. (A) Mfold-predicted secondary structure for CLSV genome sequence that includes the
predicted AS1 and RS1 elements for sg mRNA1. The AS1-RS1 interaction is labeled at the
bottom of the structure, as is the sg mRNA1 (sg1) initiation site. Base pairs that covary in the
AS1-RS1 interaction in other aureusviruses are shown and their positions indicated by dashed
lines (SNMV, Sesame necrotic mosaic virus; MWLMV, Maize white line mosaic virus;
JCSMV, Johnsongrass chlorotic stripe mosaic virus; PoLV, Pothos latent virus). Base pairs in
the AS1-RS1 interaction that were targeted for mutational analysis (boxed) are shown on either
side of the structure with substituted residues (bold and underlined) and the names of the
corresponding mutants. (B), (C) Northern blot analysis of wt and mutant genomes and graphical
representation of CLSV sg mRNA1 accumulation levels. The relative sg mRNA1 levels correspond to means with standard errors (SEs) from three independent experiments and represent the ratios of sg mRNA1 levels to their corresponding genomic RNA levels, all normalized to that for wt CLSV, set at 100.

Figure 4. Assessment of a cryptic sg mRNA1 initiation site. (A) Mutant HX78-1 has its wt AS1-RS1 interaction inactivated by a substitution in AS1 (circle) that inhibits wt sg mRNA1 transcription (denoted by the "X"). The stable RNA hairpin followed by a CAA spacer (boxed) was inserted just 3’ to the p84 stop codon (italicized and underlined) and 5’ to the TBSV-like initiation site (sg1-T-like). In HX78-1C, the cryptic initiation residue, guanylate, is substituted with a cytidylate, as shown. (B) Northern blot analysis of wt and mutant CLSV genomes. (C) Primer extension of sg mRNA1-like RNA from HX78-1 and HX78-1C transfections. The arrowhead denotes the position of the primer extension product for HX78-1 that corresponds to wt coordinate G2352.

Figure 5. In vitro translation of CLSV sg mRNA1. (A) 5’-terminal segments of wt sg mRNA1 (sg1) and TBSV-like sg mRNA1 (sg1-T-like) are shown below the CLSV genomic sequence along with CP start codon location (bold and underlined). (B) SDS-PAGE analysis of products generated from translation of 0.5 pmol of uncapped full-length sg mRNA1 (sg1) and 3’-truncated sg1 (sg1-d3’) (left panel) or uncapped full-length TBSV-like sg mRNA1 (sg1-T-like) and its 3’-truncated form (sg1-T-like-d3’) (right panel) after a 1 hour incubation at 25 °C in a 20 µl wheat germ extract reaction containing [35S]-Met. The viral RNA that was analyzed is labeled above the lane and the presence (+) or absence (−) of the 3’CITE in the message is indicated.
The mock lane is a negative control in which the translation reaction contained no added transcript, while the BMV lane is a positive control for translational competence of the system. The relative accumulation levels of CP (p41) were quantified by radioanalytical scanning of the gel and the relative translation level (±SEs) for each message is presented at the bottom of each lane. (C) SDS-PAGE analysis of translational products generated from sg1 and sg1-T-like, as described above.

Figure 6. Analysis of the 5′SL of sg mRNA1. (A) Mfold-predicted 5′-terminal stem loop structure in sg mRNA1. The corresponding stem present in two other aureusviruses is shown to the right (MWLMV, *Maize white line mosaic virus*; JCSMV, *Johnsongrass chlorotic stripe mosaic virus*). The start codon for coat protein (CP) is indicated and the pair of nucleotides that were targeted for compensatory mutational analysis is boxed. (B) SDS-PAGE analysis of products generated from translation of sg mRNA1 transcripts harboring substitutions in the stem of the 5′SL, as described in the legend to Fig. 5B. The nucleotide pairs above each lane represent the residues present in the 5′SL of the sg mRNA1 transcript tested.

Figure 7. Analysis of the loop sequence in the 5′SL. (A) Depiction of the 5′SL structure in sg mRNA1 and a SL in the 3′CITE. The complementary nucleotides in the two structures are connected by dotted lines. The two nucleotides that were targeted for compensatory mutational analysis are boxed and the changes made at these positions are shown in the large boxed inset. The conserved uridylate residue in the YUNR motif (thick grey line) is circled. (B) SDS-PAGE analysis of products generated from translation of sg mRNA1 transcripts harboring compensatory substitutions in the 5′SL and 3′CITE, as described in the legend to Fig. 5B. (C)
SDS-PAGE analysis of products generated from translation of sg mRNA1 transcripts harboring substitutions in the conserved uridylate in the YUNR motif.

Figure 8. In vivo analysis of sg mRNA1. (A) Northern blot analysis of wt and mutant CLSV RNAs from protoplast infections. Wt CLSV and mutant genomes used in the transfections are indicated above the lanes and the positions of genomic (g) and subgenomic mRNAs (sg) are indicated to the left. (B) Western blot analysis of CLSV CP accumulation in protoplast infections corresponding to those in (A). CLSV CP was detected using a polyclonal antiserum following separation of total proteins from infected protoplasts by SDS-PAGE and transfer to membrane. The asterisk denotes non-specific reactivity with a cellular protein and the positions of molecular mass markers (in kDa) are indicated to the left.

Figure 9. Schematic depiction of differences in transcriptional and translational regulatory elements for sg mRNA1 in CLSV and TBSV. (A) Schematic version of CLSV genome showing the medium-range AS1-RS1 elements (above) and the local 5′-terminal SL structure in sg mRNA1 (below). Sequences and predicted structures of the wt 5′SL and the 5′-end of the mutant sg1-T-like are shown in the box. The shaded residues are complementary to those in the 3′CITE. Relative levels of CP translation from the two messages in vitro and in vivo are shown to the left. (B) Schematic version of TBSV genome showing the long-range AS1-RS1 elements and the predicted local unstructured 5′-terminal sequence in sg mRNA1. The shaded residues are complementary to those in the TBSV 3′CITE.
Figure 10. Comparison of the 5’SL in CLSV sg mRNA1 with functional SLs in cellular and viral RNAs. (A) Secondary structure of the CLSV 5’SL compared with SLs in tRNA-Phe anticodon and HokC mRNA. The YUNR motif present in all three structures is delineated by the grey line and loop residues in the cellular RNA structures that are different from those in the CLSV SL are underlined. (B) Mfold-predicted SLs in the 5’UTRs of the CLSV genome and sg mRNA2. (C) Multifunctional nature of the sequence corresponding to the 5’SL in sg mRNA1. Different features in this structure are denoted as follows: “attenuation element”, thick diagonally-hatched line; “transcriptional promoter”, large dashed line; sense/stop codons, rectangles; YUNR motif, grey line; stem of SL, bracket; sequence that interacts with 3’CITE, dotted line inside the loop sequence. The relative positions of the attenuation element and transcriptional promoter are approximations that are based on those defined for TBSV (21). For convenience, the elements have also been labeled on the figure.