Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 322 (2004) 41–50

VIROLOGY

www.elsevier.com/locate/yviro

cis-acting elements at opposite ends of the *Citrus tristeza virus* genome differ in initiation and termination of subgenomic RNAs

María A. Ayllón,[1] Siddarame Gowda, Tatineni Satyanarayana, and William O. Dawson*

Department of Plant Pathology, University of Florida, Citrus Research and Education Center, Lake Alfred, FL 33850, USA

Received 24 November 2003; returned to author for revision 8 January 2004; accepted 12 January 2004

Abstract

Citrus tristeza virus (CTV), a member of the *Closteroviridae* with a plus-stranded genomic RNA of approximately 20 kb, produces 10 3′-coterminal subgenomic (sg) RNAs that serve as messenger (m)RNAs for its internal genes. In addition, a population of 5′-terminal sgRNAs of approximately 700 nts are highly abundant in infected cells. Previous analysis demonstrated that the controller elements (CE) are responsible for the 3′-terminal mRNAs and the small 5′-terminal sgRNAs differ in the number of additional sgRNAs produced. A feature of both types of CE is production of 5′- and 3′-terminal positive-stranded sgRNAs, but the 3′ CEs additionally produce a negative-stranded complement of the 3′-terminal mRNAs. Here, we found that the termination (for 5′-terminal sgRNAs) and initiation (for 3′-terminal sgRNAs) sites of the 5′ vs. the 3′ CEs occur at opposite ends of the respective minimal active CEs. The initiation site for the 3′ CE of the major coat protein gene, and probably those of the p20 and p23 genes, was outside (3′ in terms of the genomic RNA) the minimal unit, whereas the termination sites were located within the minimal CE, 30–50 nts upstream of the initiation site (referring to the positive-strand sequence). In contrast, the initiation site for the 5′ CE was in the 5′ region of the minimal unit, with the termination sites 20–35 nts downstream (referring to the positive-strand sequence). Furthermore, the CEs differ in initiation nucleotide and response to mutagenesis of that nucleotide. The 3′ CE initiates sgRNA synthesis from a uridylate, whereas the 5′ CE initiates from a cytidylate. We previously found that the 3′ CEs were unusually tolerant to mutagenesis of the initiation sites, with initiation proceeding from alternative sites. Mutagenesis of the initiation site of the 5′ CE prevented synthesis of either the 5′- or 3′-terminal sgRNAs. Thus, the *cis*-acting elements at opposite ends of the genome are remarkably different, perhaps having arisen from different origins and or with different functions in the life cycle of this virus.

Keywords: Controller elements; *Citrus tristeza virus*; sgRNAs

Introduction

Citrus tristeza virus (CTV), a member of the genus *Closterovirus* of the *Closteroviridae*, has a positive-sense single-stranded RNA genome of approximately 20 kb. The 5′ half of the genome encodes proteins involved in replication and the 3′ half of the genome encodes 10 genes not needed for replication in protoplasts (Karasev et al., 1995; Pappu et al., 1994; Satyanarayana et al., 1999). Based on amino acid sequences of the RNA-dependent RNA polymerase and other replication-associated proteins, closteroviruses have been placed in the alphavirus-like supergroup of positive-stranded RNA viruses (Koonin and Dolja, 1993). However, the large interdomain region between the methyltransferase- and helicase-like domains, the large genome size, along with multiple subgenomic (sg) messenger (m)RNAs appear to be more similar to some of the members of the animal *Nidovirales*, perhaps suggesting an intermediate position of the closteroviruses among those groups.

Closteroviruses produce an unusually large number of sgRNAs during replication. In addition to the 3′-coterminal mRNAs, two other types of sgRNAs are produced (Fig. 1A). One is a negative-stranded complement to the 3′-terminal mRNA. Although this sgRNA is produced in relatively large amounts, it is not known whether it serves a function in gene expression. It is possible that these minus strands are produced from the genomic template by termination to function as templates for transcription of the mRNAs, as has been proposed for other viruses (Sawicki

* Corresponding author. Citrus Research and Education Center, 700 Experiment Station Road, Lake Alfred, FL 33850. Fax: +1-863-956-4631.
E-mail address: wodtmv@C.REC.iFAS.ufl.edu (W.O. Dawson).
[1] Present address: Departamento de Biotecnología, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2004.01.005

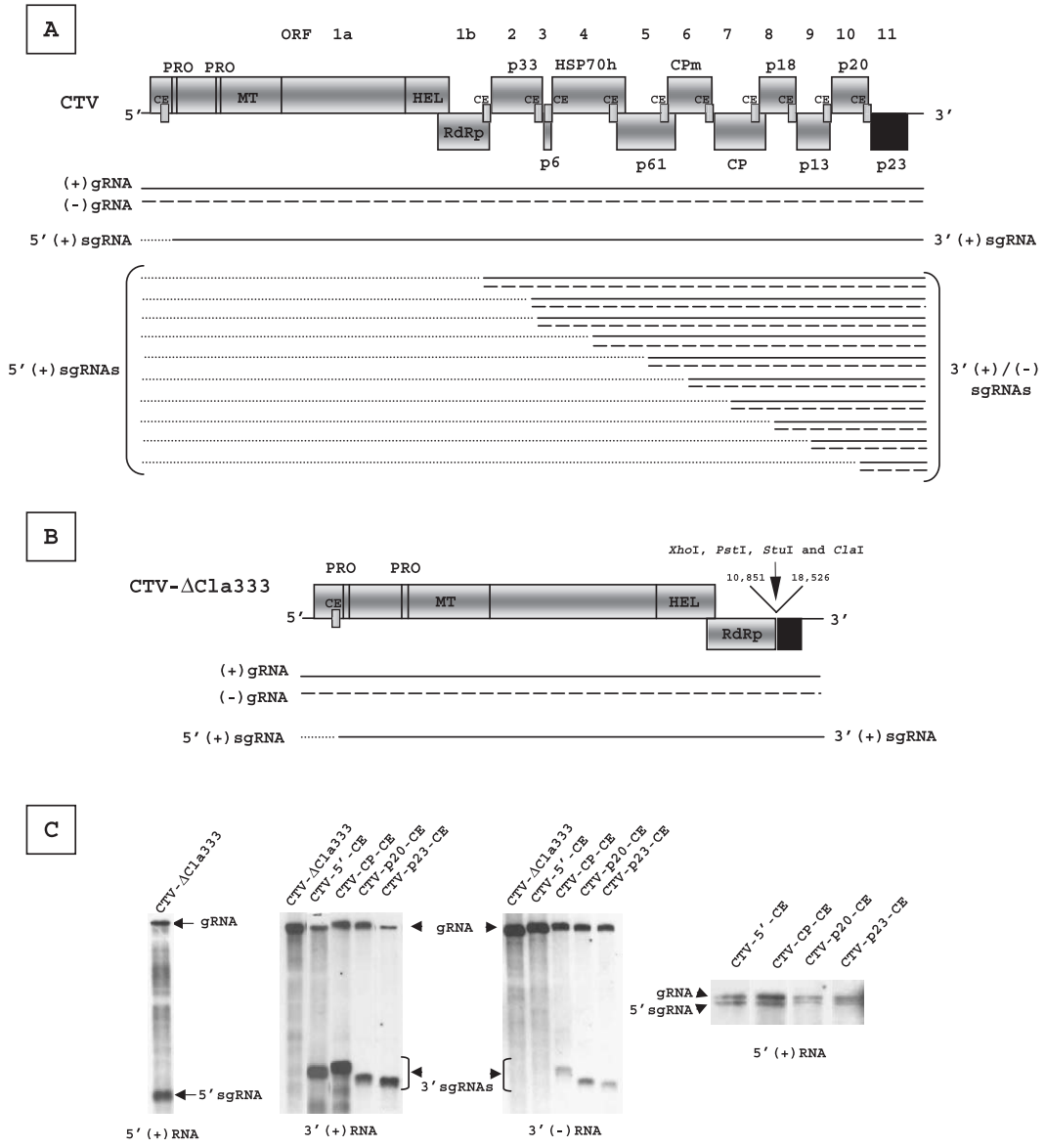

Fig. 1. (A) A schematic diagram of the CTV genome. Boxes represent the ORFs with the respective numbers and encoded products. The two papain-like proteases (PRO), methyltransferase (MT), helicase (HEL), and polymerase-like (RdRp) domains, HSP70 homologue (HSP70h), and major (CP) and minor (CPm) coat proteins are indicated. Small boxes along the diagram indicate the controller elements (CEs). The genomic (gRNA) and subgenomic RNAs (sgRNAs) produced are shown below. Positive and negative strands of the gRNA and the 3′-terminal sgRNAs are indicated by solid and dashed lines, respectively, and the 5′-terminal sgRNAs are indicated by dotted lines. (B) Schematic representation of the CTV-ΔCla333, the parental plasmid used for construction of replicons CTV-CP-CE, CTV-p20-CE, CTV-p23-CE, and CTV-5′-CE. The RNAs produced by CTVΔCla333, 5′-terminal sgRNAs (dotted line), 3′-terminal sgRNA and gRNA positive strands (solid lines) and negative-stranded gRNA (dashed line), the nucleotide numbers, and the restriction enzymes sites introduced at the end of ORF 1b are indicated. The 5′ CE is indicated by a small box. (C) Northern blot analysis of the total RNA isolated 4 dpi from *N. benthamiana* mesophyll protoplasts inoculated with transcripts of CTV-ΔCla333, CTV-5′-CE, CTV-CP-CE, CTV-p20-CE, or CTV-p23-CE, hybridized with a riboprobe specific to the plus-stranded 5′ terminus [5′ (+)RNA] or 3′ terminus [3′ (+)RNA] or to the negative-stranded complement of the genomic 3′ terminus [3′ (−)RNA]. Arrows denote gRNA, LMT, and 3′-terminal sgRNAs.

and Sawicki, 1990; van Marle et al., 1999; White, 2002). Alternatively, the 3′-coterminal mRNAs might be produced by internal promotion from the complement of the genomic RNA. In this case, the negative-stranded complements could be nonfunctional artifacts, copied from the mRNAs or produced by independent termination during minus-strand synthesis. The other sgRNA is a 5′-terminal positive-stranded sgRNA with a 3′ terminus corresponding to genomic sequences near the 3′ CE (Che et al., 2001; Gowda et al., 2001). Apparently, the 5′-terminal sgRNAs are produced by premature termination during genomic RNA synthesis. We have found no function for these RNAs. Cumulatively, we have identified at least 32 different CTV sgRNA species (Gowda et al., 2001, 2003; Hilf et

al., 1995; Ayllon et al., unpublished results). The production of these sgRNAs is regulated by *cis*-acting elements (Fig. 1A) within the genomic RNA or its complement which we refer to as controller elements (CE) because we do not know whether they function as promoters or terminators.

Not only does CTV have more sgRNA controller elements than its alphavirus-like relatives, the 3′ CEs of CTV appear to represent a distinct class of sgRNA controller elements, particularly in terms of initiation of plus-stranded sgRNA synthesis. The initiation site—the nucleotide on the minus strand corresponding to the first nucleotide of the 3′-terminal positive-stranded sgRNA—of the alphavirus supergroup tends to reside within the minimal promoter element and modification of the initiation site tends to prevent or greatly inhibit sgRNA synthesis in vivo (Grdzelishvili et al., 2000; Koev and Miller, 2000; van der Vossen et al., 1995). In contrast, the initiation site of the 3′ CE of the major coat protein (CP) is located outside of the minimal active unit, a region of 57 nts located upstream of the CP ORF (Gowda et al., 2001). In addition, there was considerable flexibility for alteration of the CTV 3′ CE initiation nucleotide or its context (Ayllón et al., 2003). The initiation sites of the 3′ CEs tend to be uridylates since all of the sgRNAs whose 5′ termini have been determined (7 of 10) are initiated with an adenylate (as was the 5′ end of the genome) (Ayllon et al., 2003, unpublished results; Karasev et al., 1995, 1997). Mutation of the initiation site usually resulted in continued sgRNA production, with initiation at an alternate nucleotide (Ayllón et al., 2003). In fact, the contexts of the initiation sites of different 3′ CEs appear to serve as a mechanism to regulate gene expression.

CTV appears to have yet a different type of controller element near the 5′ end of the genome (5′CE). The obvious products of the 5′ CE are copious amounts of a population of 5′-coterminal positive-stranded sgRNAs of approximately 700 nts whose function is unknown (Che et al., 2001; Mawassi et al., 1995). The 5′ CE was mapped to an approximately 60-nt region that was predicted to fold into two stem and loop (SL) structures (Gowda et al., 2003). Unexpectedly, we found that the 5′ CE also controlled the production of a 3′-terminal positive-stranded sgRNA, which was produced in nearly undetectable amounts from the 5′ CE in its native position near the 5′ end of the genome, but in large amounts from a duplicated copy of the 5′ CE near the 3′ end of the genome (Gowda et al., 2003). Yet, in contrast to the 3′ CEs, the 5′ CE appeared to function as a classic ‘internal promoter,’ producing the 3′-terminal positive-stranded sgRNA from the genomic minus strand, without a corresponding complementary minus-stranded sgRNA.

Here, we report experiments comparing these two types of sgRNA controller elements. These results show that the 5′ CE and 3′ CEs also differ substantially in initiation of 3′-terminal sgRNAs and termination of 5′-terminal sgRNAs.

Results

Conventions

To avoid confusion when referring to different sites on positive vs. negative RNA sequences and measured vs. inferred sites, we use the following terminology. The termination site is inferred as the position on the negative-stranded RNA corresponding to the mapped 3′ terminus of the 5′-terminal sgRNA. The initiation site is inferred as the position on the negative-stranded RNA corresponding to the determined 5′ terminus of the 3′-terminal positive-stranded sgRNAs. We show termination and initiation sites in the predicted secondary structures of the negative-sense sequence. For simplicity, we use numbering of the positive-sense sequence and we refer to the 5′-terminal nucleotide of the 3′-terminal positive-stranded sgRNA as the +1 nt.

Determination of the termination sites of the 3′ CEs

The initiation sites of several 3′ CEs have been determined by mapping the 5′ termini of the 3′-terminal sg mRNAs (Ayllón et al., 2003; Karasev et al., 1997). To examine termination that results in production of the 5′-terminal positive-stranded sgRNAs near the 3′ CEs of the CP, p20, and p23 genes (Fig. 1A), we inserted the respective CEs into CTV-ΔCla333 (Fig. 1B) so that only sgRNAs from the inserted controller element would be produced and inoculated *Nicotiana benthamiana* protoplasts with RNA transcripts of each construct. Total RNA samples extracted at 4 days post-inoculation (dpi) were analyzed by Northern blot hybridizations. In the absence of an inserted CE, CTV-ΔCla333 produces only genomic plus and minus strands along with the small 5′-terminal sg RNAs produced from the native 5′ CE (Fig. 1C). Each inserted 3′ CE (CTV-CP-CE, CTV-p20-CE, and CTV-p23-CE) produced three additional sgRNAs: the 3′-terminal positive-stranded sgRNA along with its complementary minus strand, plus the 5′-terminal positive-stranded sgRNA (Fig. 1C).

The termination sites near the 3′ CEs that result in production of the 5′-terminal sgRNAs were inferred by determining the 3′ termini of these RNAs. Total RNA extracted from the infected protoplasts was polyadenylated, amplified by RT-PCR, cloned, and sequenced. Twenty-nine cDNA clones from the CTV-CP-CE RNA were analyzed (Fig. 2A). The 3′ termini of the first 13 clones that were analyzed mapped (number of clones with an identical terminus are in parentheses) to nts 16155 (1), 16161 (10), and 16165 (2), which were located within 40–50 nts downstream of the initiation site. However, we realized that this area was located near an adenylate-rich region, which could have served as a template region for the oligo (dT) primer. To examine this supposition, we repeated the RT-PCR amplification, cloning, and sequencing using the same total RNA, but without the polyadenylation, and obtained similar products, thus making the previous result question-

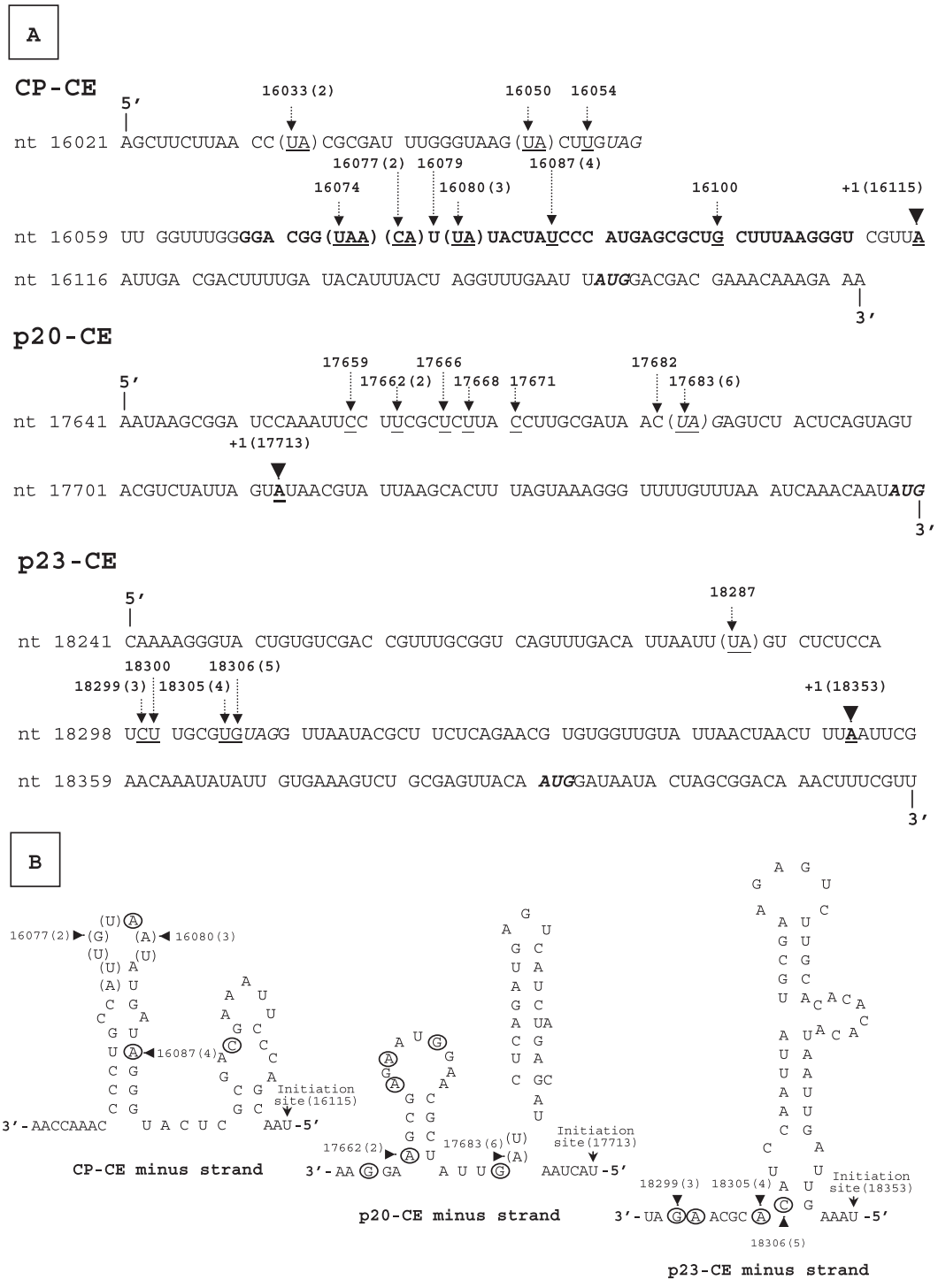

Fig. 2. Location of the 3′ termini of the 5′-terminal positive-stranded sgRNAs of CP, p20, and p23 CE′s. (A) Positive-stranded sequences of the 3′ CEs (5′ and 3′ ends are indicated in each sequence), with the upstream stop codons (in italics), downstream start codons (in italics and bold), and the +1 nts (underlined and bold). Termination sites are identified with arrows and nucleotide positions and number of times duplicated shown in parentheses, ambiguous termination sites are shown in parentheses and underlined, with only the first nucleotide identified with an arrow. The sequence of the minimal CP CE, which has been mapped, is presented in bold. (B) The predicted secondary structures of the negative-stranded sequences of the 3′ CEs (5′ and 3′ ends are indicated in each sequence) showing the location of the initiation site and the termination sites (circled nucleotides; ambiguous sites are in parentheses) with position numbers and number of times duplicated (in parentheses).

able. We then repeated the experiment with polyadenylated RNA using increased amounts of cDNA with a decreased annealing temperature or alternatively increasing the time of the amplification cycles and obtained an additional smaller amplified fragment. Sixteen clones from the smaller PCR product had heterogeneous 3′ termini at positions 16033–4

(2), 16050–1 (1), 16054 (1), 16074–6 (1), 16077–8 (2), 16079 (1), 16080–1 (3), 16087 (4), and 16100 (1) (Fig. 2A, CTV-CP-CE). Ambiguities resulted when the apparent terminus was adjacent to a 3′ adenylate residue in the CTV sequence because we could not distinguish whether the adenylate(s) resulted from the sequence or the poly (A) tail. In these cases, 2 (or 3) nts are identified as possible termini. Most of the inferred termination sites mapped to a region of 14 nts, situated 28–41 nts upstream of the initiation site of the CP CE. The locations of these termini relative to the predicted secondary structure of the 57-nt core CP CE and a larger 200-nt region were primarily throughout the first SL structure (Fig. 2B, CTV-CP-CE).

To map the termination sites at the p20 CE, we similarly sequenced 13 clones of the corresponding polyadenylated 5′-terminal sgRNAs. The 3′ termini corresponded to nts 17659 (1), 17662 (2), 17666 (1), 17668 (1), 17671 (1), 17682 (1), and 17683–4 (6), (Fig. 2A, CTV-p20-CE). The 3′ termini of more than half of the clones clustered near nts 17682–4, 30–31 nts upstream of the initiation site of the p20 CE. The minimal p20 CE has not been mapped. To predict a secondary structure of the p20 CE, we used MFOLD to analyze a region of 60–70 nts upstream of the initiation site and a larger region of 200 nts, which was predicted to fold into two SL structures with the 3′-most SL predicted to be more stable. Most of the inferred termination sites occurred upstream of the second SL (Fig. 2B, CTV-p20-CE) similar to that found for the CP CE.

Fourteen clones from the polyadenylated p23 5′-terminal sgRNA were sequenced, and the resulting 3′ termini were mapped to a region of 20 nts, at positions 18287–8 (1), 18299 (3), 18300 (1), 18305 (4), and 18306 (5) (Fig. 2A, CTV-p23-CE). Most of the 3′ termini (9 of 14 clones) mapped 47–48 nts upstream of the initiation site of the p23 CE. The minimal active unit of the p23 CE also has not been determined. The secondary structure of the 60- to 70-nt region upstream of the initiation site of this CE was predicted to fold into a single complex SL structure, unlike the predictions for either of the other 3′ CEs (Fig. 2B, CTV-p23-CE). This predicted SL was maintained when larger sequences were folded (data not shown). Yet, the inferred termination sites were all located upstream of that putative structure.

Overall, the termination sites of all three 3′ CEs were located 30–50 nts upstream of the initiation site and generally upstream of the 3′-most SL.

Determination of the termination site of the 5′ CE

We previously mapped the minimal 5′ CE responsible for production of the small 5′-terminal sgRNAs (Gowda et al., 2003). The 3′ termini of these RNAs were previously determined for the VT strain of CTV (Che et al., 2001); however, the minimal 5′ CE was determined only for the T36 strain (Gowda et al., 2003). Because these isolates differ in sequence in this region, it was necessary to map the 3′ termini of the T36 5′-terminal sgRNAs to correlate their 3′ termini to inferred termination sites associated with the T36 5′ CE. This determination was made for three different types of RNA samples: (1) 5′-terminal sgRNA from total RNA extracted from protoplasts inoculated with the replicon CTV-ΔCla333; (2) 5′-terminal sgRNA extracted with dsRNA from plants infected with wild-type CTV T36; and (3) the much longer 5′-terminal sgRNA from total RNA extracted from protoplasts infected with construct CTV-5′-CE, which has the 5′ CE duplicated in the 3′ region of CTV-ΔCla333 (Figs. 1B and C). The RNAs were polyadenylated at the 3′ termini, reverse transcribed, amplified using an oligo (dT) primer and an upstream CTV-specific primer, and the resulting cDNAs were cloned and sequenced. Sequences of 22 clones from the three different sources of the 5′-terminal sgRNAs gave eight different termination sites within a region of 18 nts. The termini of the sgRNAs produced from the 5′ CE in its native location mapped at positions 749 (1), 751 (1), 752 (3), 757 (1), 759 (1), 765 (8), and 766 (1) (Fig. 3A). The termini of the sgRNAs produced from the 5′ CE inserted into the 3′ end of the replicon mapped at positions 751 (3), 752 (1), 755–6 (1), and 765 (1). Most (17 of 22) of the 3′ termini were mapped to positions 751, 752, and 765, all within the 3′ region of the 5′ CE (Fig. 3A), about 20–35 nts downstream of the initiation site (determined in the next section), within the second predicted SL structure (Fig. 3B). This result contrasts with the 3′ CEs, where terminations occurred within the 5′ region of the minimal controller element and upstream of the initiation site.

Determination of the initiation site of the 5′ CE

To map the initiation site of the 3′-terminal sgRNA produced by the 5′ CE, this controller element was duplicated by insertion of its sequence into the 3′ region of CTV-ΔCla333 (Fig. 1B) to create construct CTV-5′-CE. This replicon accumulates large amounts of the 3′ positive-stranded sgRNA (Gowda et al., 2003 and Fig. 1C, CTV-5′-CE). The 5′ terminus of this 3′-terminal sgRNA was determined by a primer extension assay. Complementary DNA was synthesized using total RNA extracted from *N. benthamiana* protoplasts infected with CTV-5′-CE and the CTV specific primer C682. The same primer was used to generate the sequence ladder from the corresponding plasmid DNA. The primer extension product migrated as a double band, with the upper band coinciding with position 729 (guanylate) of the adjacent sequence ladder, and the lower band mapping between this guanylate and the uridylate at position 730 (Fig. 3C). This lower band tends to be produced only in primer extension experiments using total RNA isolated from protoplasts as templates, in contrast to RNA extracted from infected plants (Ayllón et al., 2003). Thus, we interpreted that the 5′-terminal nt of the 5′ CE sgRNA corresponded to the guanylate at position 729. This interpretation was supported by the mutagenesis experiments shown below.

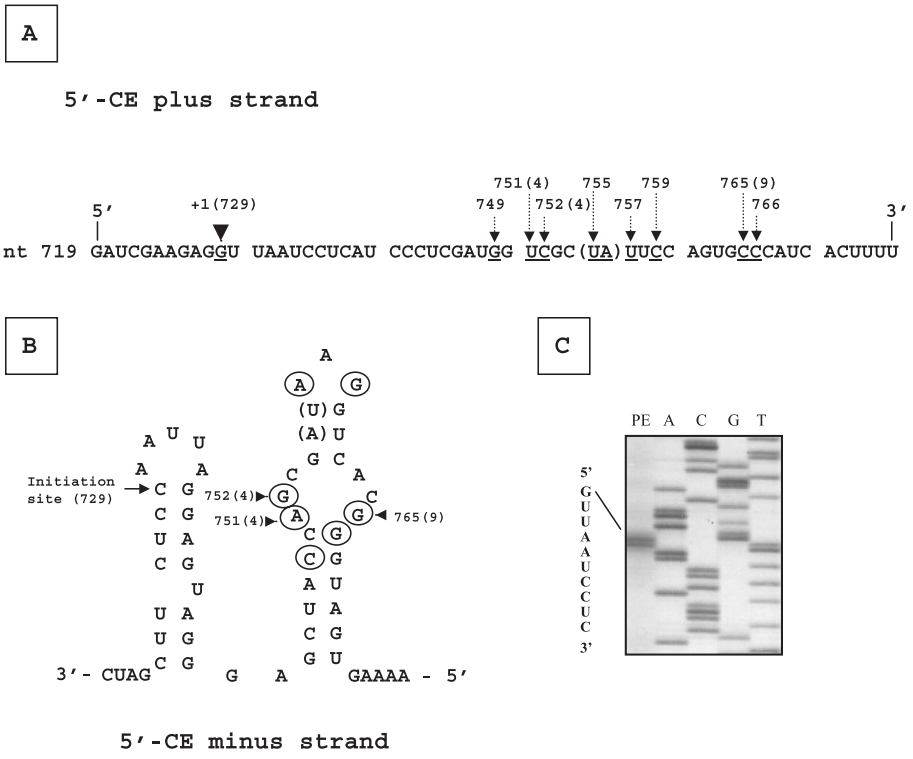

Fig. 3. Analysis of the 3′ termini of the 5′-terminal positive-stranded sgRNAs of the 5′ CE. (A) The sequence of the minimal active controller element of the 5′ CE (5′ and 3′ ends are indicated); the +1 nt (arrowhead) and the 3′-terminal nucleotides of the 5′-terminal sgRNAs (arrows) are underlined (ambiguous termini are in parentheses) with positions within the genome and the number of duplicate termini (in parentheses) indicated above the arrows. (B) The predicted secondary structures of the negative-stranded sequences of the 5′ CE showing the location of the initiation site and the termination sites (circled nucleotides; ambiguous termini are in parentheses) with position numbers, and number of times duplicated (in parentheses). (C) Mapping of the +1 nt of the 3′-terminal positive-stranded sgRNA of the 5′ CE. Analysis of the primer extension product (PE, left lane) in a 6% denaturing polyacrylamide gel beside the sequencing ladder (right four lanes) obtained from pCTV-5′-CE. Partial nucleotide sequence of the 5′-terminus of the 5′ CE sgRNA is presented at left.

A 3′-terminal sgRNA with a guanylate at the 5′ terminus contrasts with the results obtained for the seven 3′-terminal sgRNAs produced by the 3′ CEs in that all were mapped to terminal adenylates (Ayllón et al., 2003, unpublished results; Karasev et al., 1997).

Modification of the initiation site of the 3′-terminal sgRNA of the 5′ CE

Modification of the initiation site of the 3′-terminal sgRNAs of several members of the alphavirus-like supergroup results in inhibition or abolition of sgRNA synthesis. However, modification of the initiation sites of several CTV 3′ CEs had relatively little effect on sgRNA synthesis (Ayllón et al., 2003). To compare the importance of the initiation sites of the 5′ CE and the 3′ CEs on sgRNA production, we mutated the guanylate in position 729 (positive strand sequence) to A, C, or U to create the mutants 5′-CE + 1G/A, 5′-CE + 1G/C, and 5′-CE + 1G/U, respectively (Fig. 4A). In vitro-produced RNA transcripts obtained from the three mutants were used to inoculate *N. benthamiana* protoplasts, and the production of sgRNAs was analyzed by Northern blot hybridizations using riboprobes specific to the 3′ or 5′ end of the genomic RNA. The mutants replicated efficiently in protoplasts, but only CTV-5′-CE (wild-type) accumulated the 3′-terminal sgRNA (Fig. 4B) and the 5′-terminal sgRNAs (data not shown), demonstrating that modification of this nucleotide prevented accumulation of both sgRNAs. This result raised two questions: Do all mutations in this area prevent sgRNA synthesis? Did the structural changes caused by the mutation of the G at position 729 prevent sgRNA synthesis? The 5′ CE was predicted to fold into a secondary structure composed of two SL structures (Gowda et al., 2003). Nucleotide 729 would be within the 5′-most SL and the mutations of this nucleotide were predicted to modify the secondary structure predicted for the negative strand of the 5′ CE (Fig. 4C). To provide information concerning both questions, we modified the nucleotides at positions −2 (A to U) and +6 (U to A) with reference to nt 729; these were mutations on each side of the proposed initiation site. This mutant, which retained the +1G, was predicted to simulate the secondary structure predicted for the change of the +1G to A, C, or U (Fig. 4C). Mutant 5′-CE-2A/U + 6U/A accumulated both 3′-terminal and 5′-terminal sgRNAs at lower levels than that of the wild-type construct (Fig. 4B, data not shown), suggesting that the substitution of the +1G, rather than the predicted structural change, was responsible

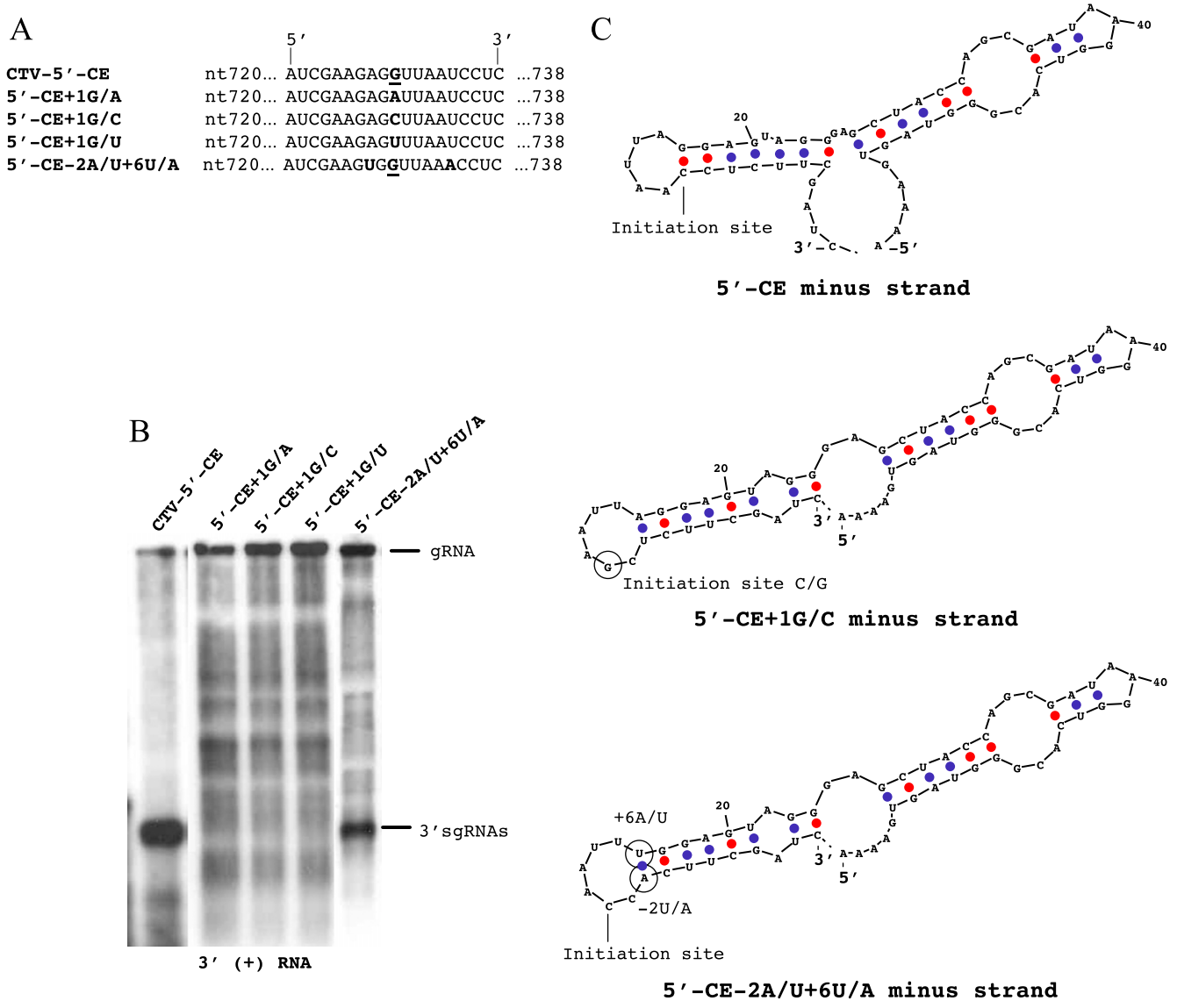

Fig. 4. Modification of the initiation site of the 5′-CE sgRNA. (A) Sequence shown around the +1 nt (bold and underlined) of the positive-stranded sequence of the wild-type construct and related mutants (5′ and 3′ ends are indicated). The mutated nucleotides are indicated in bold. (B) Northern blot analysis of the total RNA isolated 4 dpi from *N. benthamiana* mesophyll protoplasts inoculated with RNA transcripts from wild-type and mutants of CTV-5′-CE. The blots were hybridized with a digoxigenin-labeled riboprobe specific to the positive strand of the genomic 3′-terminal 900 nts. The positions of genomic (gRNA) and 3′-terminal positive-stranded sgRNAs are indicated. (C) Predicted secondary structures of the negative-stranded sequence of wild-type and mutants of CTV-5′-CE.

for the prevention of sgRNA production and that other nearby mutations did not prevent sgRNA production. Thus, these data support the mapping of the 5′ terminus of the 3′ sgRNA produced by the 5′ CE to guanylate 729. In contrast to the 3′ CEs, the modification of the initiation site of the 5′ CE completely prevented sgRNA accumulation.

Discussion

We showed previously that the 5′ CE and the 3′ CEs differed in production of minus-strand complements to the 3′-terminal sgRNAs (Gowda et al., 2003). The present work demonstrated other differences in termination and initiation of sgRNAs. Because the 5′-terminal sgRNAs have neither corresponding minus strands to suggest transcription from a sgRNA nor stoichiometrical amounts of corresponding 3′-terminal sgRNAs to suggest cleavage, it appears that they are produced by premature termination somehow related to the controller element functions (Che et al., 2001; Gowda et al., 2001, 2003). The termination sites of the different 3′ CE (CP, p20, and p23), in general, were similar. Termination was not precise, but occurred over a range of nucleotides about 30–50 nts upstream of the initiation site. A similar result was reported for the 5′-terminal sgRNA terminating at the p33 CE of the VT strain (Che et al., 2001). Termination by the CTV 3′ CEs appears to be similar to that of several other viruses in which this phenomenon has been examined.

Terminations of 5′-terminal sgRNAs of *Sindbis virus* (SIN) (Wielgosz and Huang, 1997), *Citrus leaf blotch virus* (Vives et al., 2002), and *Grapevine virus A* (GVA) (Galiakparov et al., 2003) occur upstream of the initiation site of the 3′ sgRNAs. The termination of the 5′ sgRNA at the controller elements of SIN and GVA is proportional to the amount of initiation of 3′ sgRNAs (Galiakparov et al., 2003; Wielgosz and Huang, 1997). Accumulation of CTV 5′-terminal sgRNA also appears to be proportional to the production of 3′ sgRNAs (Gowda et al., 2001, 2003), but also might be related to distance from the 5′ terminus (Che et al., 2001).

These results are consistent with the model of pause-and-release of the replicase complex at the sgRNA controller elements that has been proposed to explain the origin of the 5′ sgRNAs of other viruses (Galiakparov et al., 2003; German et al., 1992; Vives et al., 2002; Wielgosz and Huang, 1997). In this model, during the synthesis of the genomic RNA, the replicase complex first pauses and then releases the uncompleted genomic strand near the controller element, which is occupied by an obstructing replicase complex attempting to initiate sgRNA synthesis. The inability to uncouple the production of the 5′ and 3′ sgRNAs in CTV by mutation of the controller element or the initiation site (Ayllón et al., 2003; Gowda et al., 2001) suggests that termination and initiation of sgRNAs are connected events, supporting this model.

Termination at the 5′ CE also was heterologous, resulting in a population of 5′-terminal sgRNAs, as was reported previously for the VT strain of CTV (Che et al., 2001). However, the 5′ CE termination sites were downstream of the initiation site. Termination at this position could be consistent with a premature termination model similar to that described above if the replicase complex were to move downstream a few nucleotides after initiation and then pause, as has been shown for *Brome mosaic virus* (Adkins et al., 1998) and *Turnip crinkle virus* (Nagy et al., 1997). Alternatively, termination has been shown to occur during transcription by the *Escherichia coli* DNA-dependent RNA polymerase due to folding of the nascent transcript. The formation of a stable G–C rich hairpin followed by a run of U residues cause separation of the complex from the template at the poly (U) tract (Uptain and Chamberlin, 1997; Von Hippel and Yager, 1992; Wilson and von Hippel, 1995). The termination sites of the CTV 5′ CE are downstream of a G–C rich hairpin followed by a run of 4 U's, but the stall sites do not map to the poly (U) tract. Further experiments are required to determine whether there are commonalities among the termination mechanisms of the different types of controller elements.

The 5′ CE and the 3′ CEs also differed in initiation of 3′-terminal sgRNAs. Because the 3′-terminal sgRNA produced by the 5′ CE does not have a corresponding negative strand, it appears that this RNA is produced by internal promotion from the genomic-length minus strand. However, it is unclear whether the 3′-terminal sgRNAs produced by the 3′ CEs are produced in agreement with an internal promotion model or with a termination model in which sg minus strands are produced as a template for transcription of the plus-stranded sg mRNAs, as has been demonstrated for arteriviruses (van Marle et al., 1999). If the latter model is correct, the initiation site for the positive-stranded sgRNA would correspond to one of the nucleotides at the 3′ terminus of the sg minus strand. All of the 3′ CEs characterized initiate their 3′-terminal sgRNAs with an adenylate (Ayllón et al., 2003; Karasev et al., 1997), whereas the 5′ CE initiated its 3′-terminal sgRNA with a guanylate. In addition, the positions of the initiation sites relative to the *cis*-acting regions were different between the CP CE and 5′ CE. The CP CE initiation site was 3′ and outside of the minimal controller element, whereas the initiation site of the 5′ CE was located within the minimal active element, in fact at the opposite (5′) end. They also differed in flexibility of the initiation nucleotide and its context. In contrast with the considerable flexibility of modification of the +1 nt of the 3′ CEs (Ayllón et al., 2003), mutation of the initiation nucleotide of the 5′ CE completely abolished production of both 3′-terminal and 5′-terminal sgRNAs. The 5′ CE was more similar to that expected for members of the alphavirus-like supergroup (Grdzelishvili et al., 2000; Koev and Miller, 2000; van der Vossen et al., 1995).

Thus, we return to the basic question of how one virus replication complex is able to differentially interact with at least 11 different controller elements to produce regulated amounts of sgRNAs. Although we expect to find commonalities, these results suggest that the CTV replication complex is able to interact with different types of controller elements that initiate sgRNA synthesis differently. Another question is how these different controller elements evolved within this virus. The genomic sequences of several CTV isolates have an uneven distribution of sequence similarity, with the 3′ half of the genomes conserved among different isolates but increasing differences towards the 5′ end. This irregular distribution of sequence similarity has been interpreted as the result of a recombination event between CTV and some other closterovirus (Mawassi et al., 1996; Vives et al., 1999). It is possible that the different sgRNA controller elements of CTV have different origins. Yet, it should be noted that we have found no functions associated with either sgRNA produced by the 5′ CE. It is possible that this element has a role different from a 'controller element' in the virus life cycle. It has not escaped our notice that this element is located within the genomic RNA near the predicted transition of the different coat proteins defining the 'body' and 'tail' of the polar virions and might have a function in assembly.

Materials and methods

Plasmid constructions

The basis of all constructs is pCTV9R (Fig. 1A) of CTV T36 (Satyanarayana et al., 1999, 2003) with GenBank

accession no. AY170468. All mutants were constructed in pCTV-ΔCla333 (Fig. 1B), a replicon with a deletion between nt 10851 (end of ORF 1b) and nt 18525 (leaving the last 495 nts of the p23 and the 3′ noncoding region), with unique restriction sites (*Xho*I, *Pst*I, *Stu*I, and *Cla*I) inserted (Gowda et al., 2001). To generate constructs CTV-CP-CE, CTV-p20-CE, CTV-p23-CE, and CTV-5′-CE, DNA fragments of pCTV9 of nts 15895–16631, 17510–17943, 18163–18526, or 524–959 were amplified by PCR using a positive-sense primer containing an *Xho*I restriction site and ligated into pCTV-ΔCla333 between *Xho*I and *Stu*I restriction sites, respectively. The DNA fragments used to create the mutations in 5′ CE were generated by overlap extension PCR (Ho et al., 1989) using primers containing specific single nucleotide substitutions. All the nucleotide mutations introduced were confirmed by sequencing by the Interdisciplinary Center for Biotechnology Research DNA sequencing core facility of the University of Florida (Gainesville, FL). The secondary structures were predicted with the MFOLD program (Zuker et al., 1999).

Protoplast transfections and Northern blot hybridization

Isolation and infection of mesophyll protoplasts from *N. benthamiana* leaves were performed following the procedures described by Navas-Castillo et al. (1997) and Satyanarayana et al. (1999). SP6 RNA polymerase-produced (Epicentre Technologies, WI) capped (m^7G) RNA transcripts from *Not*I-linearized DNA constructs were used directly for protoplast inoculation (approximately 1×10^6) as described (Satyanarayana et al., 1999). Protoplasts were harvested at 4 dpi, total RNA was extracted (Navas-Castillo et al., 1997) and analyzed by Northern blot hybridization using 5′-terminal (600 nts) or 3′-terminal (900 nts) positive- or negative-stranded RNA specific digoxigenin-labeled riboprobes. The riboprobes were equalized using dsRNA as described (Satyanarayana et al., 2002).

Determination of the initiation site of the 5′ CE

Total RNA extracted from *N. benthamiana* mesophyll protoplasts infected with RNA transcripts of CTV-5′-CE was used as template to determine the 5′ terminus of the 3′-terminal sgRNA. A negative-strand primer, C682 (5′-TGAGTGGAGAAATCAGTTTCAG-3′), complementary to positions 874–853, was used for primer extension analysis as described by Ayllon et al. (2003). The products of primer extension were analyzed in a 6% denaturing polyacrylamide gel adjacent to a sequencing ladder obtained with the same primer on the respective fragment of pCTV-5′-CE.

Determination of the termination sites of the 5′ CE and the 3′ CEs

Total RNA extracted from *N. benthamiana* mesophyll protoplasts infected with RNA transcripts of constructs CTV-CP-CE, CTV-p20-CE, CTV-p23-CE, or CTV-5′-CE was used to determine the 3′ termini of the corresponding 5′-terminal sgRNAs. Total RNA extracted from *N. benthamiana* protoplasts inoculated with the replicon CTV-ΔCla (Satyanarayana et al., 1999) and double-stranded RNA extracted from CTV infected plants (Moreno et al., 1990) were used as template to determine the termini of the 5′ sgRNAs of approximately 700 nts. The RNAs were denatured at 90 °C, 3′-polyadenylated by yeast poly (A) polymerase (US Biochemicals), and reverse transcribed using an oligo (dT) [M111: 5′ GGTCTCGAG$(T)_{18}$ 3′] and avian myeloblastosis virus reverse transcriptase (U.S. Biochemicals). The amplifications were carried out using the primers M111 and C586 (5′-*GCAAT*CTCGAGAACGTGTAGTTTACGACAATCGGGTGAG-3′, positions 10726–10754) for 5′ CE sgRNA; C456 (5′-*TATT*CTCGAGGGCTACCTATCAGCTGACTT-3′, positions 15895–15916) for CP CE sgRNA; C466 (5′-*TTCT*CTCGAGACTTTCTACGCATCGTTATC-3′, positions 17510–17529) for p20 CE sgRNA; C469 (5′-*TTCT*CTCGAGGCGGTACTGTCGGATTTATC-3′, positions 18163–18182) for p23 CE sgRNA; and C585 (5′-*ATAC*CTCGAGCCGGGCCTCTTCTCGCTTCT-3′, positions 534–543) for LMT RNA (*Xho*I restriction site is underlined and non-template nucleotides are in italics). The first PCR for CP sequences was performed in 25 μl at one cycle of 5 min at 94 °C for denaturation, 30 cycles of 30 s at 94 °C, 30 s at 58 °C, 1 min at 72 °C for amplification, and one cycle of 5 min at 72 °C for final extension, using 1 μl of a total of 25 μl reverse transcription product as template. For the second PCR for CP sequences, we used 2 μl template and one 2-min cycle at 94 °C, 30 cycles of 30 s at 94 °C, 30 s at 58 °C, 2 min at 72 °C, and one cycle of 5 min at 72 °C, or alternatively, one cycle of 2 min at 94 °C, 30 cycles of 30 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C, and one cycle of 5 min at 72 °C. The amplified products were cloned into pGEMR-T Easy plasmid, according to the manufacturer's instructions (Promega) and sequenced as described above.

Acknowledgments

We thank John Cook, Cecile Robertson, and Judy Harber for technical assistance; Theo Dreher for helpful suggestions; and Scott Adkins, Jaime Cubero, and John Cook for critically reading the manuscript. This research was supported by an endowment from the J.R. and Addie Graves family and grants from the Florida Citrus Production Research Advisory Council, US-Israel BARD, and USDA/ARS cooperative agreement 58-6617-4-018, and was approved for publication as University of Florida Agricultural Experiment Station Journal Series Number R09974. M.A. Ayllon was the recipient of a postdoctoral fellowship from the Ministerio de Educaci□n y Ciencia (Spain).

References

Adkins, S., Stawicki, S.S., Faurote, G., Siegel, R.W., Kao, C.C., 1998. Mechanistic analysis of RNA synthesis by RNA-dependent RNA polymerase from two promoters reveals similarities to DNA-dependent RNA polymerase. RNA 4, 455–470.

Ayllón, M.A., Gowda, S., Satyanarayana, T., Karasev, A.V., Adkins, S., Mawassi, M., Guerri, J., Moreno, P., Dawson, W.O., 2003. Effects of modification of the transcription initiation site context on *Citrus tristeza* virus subgenomic RNA synthesis. J. Virol. 77, 9232–9243.

Che, X., Piestun, D., Mawassi, M., Satyanarayana, T., Gowda, S., Dawson, W.O., Bar-Joseph, M., 2001. 5′-coterminal subgenomic RNAs in citrus tristeza virus-infected cells. Virology 283, 374–381.

Galiakparov, N., Goszczynski, D.E., Che, X., Batuman, O., Bar-Joseph, M., Mawassi, M., 2003. Two classes of subgenomic RNA of *grapevine virus* A produced by internal controller elements. Virology 312, 434–448.

German, S., Candresse, T., Le Gall, O., Lanneau, M., Dunez, J., 1992. Analysis of the dsRNAs of apple chlorotic leaf spot virus. J. Gen. Virol. 73, 767–773.

Gowda, S., Satyanarayana, T., Ayllón, M.A., Albiach-Martí, M.R., Mawassi, M., Rabindran, S., Garnsey, S.M., Dawson, W.O., 2001. Characterization of the *cis*-acting elements controlling subgenomic mRNAs of *Citrus tristeza* virus: production of positive- and negative-stranded 3′-terminal and positive-stranded 5′-terminal RNAs. Virology 286, 134–151.

Gowda, S., Ayllón, M.A., Satyanarayana, T., Bar-Joseph, M., Dawson, W.O., 2003. Transcription strategy in a closterovirus: a novel 5′-proximal controller element of citrus tristeza virus produces 5′- and 3′-terminal subgenomic RNAs and differs from 3′ open reading frame controller elements. J. Virol. 77, 340–352.

Grdzelishvili, V.Z., Chapman, S.N., Dawson, W.O., Lewandowski, D.J., 2000. Mapping of the tobacco mosaic virus movement protein and coat protein subgenomic RNA promoters in vivo. Virology 275, 177–192.

Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D.J., Niblett, C.L., Garnsey, S.M., 1995. Characterization of citrus tristeza virus subgenomic RNAs in infected tissue. Virology 208, 576–582.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using polymerase chain reaction. Gene 77, 51–59.

Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K., Gumpf, D.J., Lee, R.F., Garnsey, S.M., Lewandowski, D.J., Dawson, W.O., 1995. Complete sequence of the citrus tristeza virus RNA genome. Virology 208, 511–520.

Karasev, A.V., Hilf, M.E., Garnsey, S.M., Dawson, W.O., 1997. Transcriptional strategy of closteroviruses: mapping the 5′termini of the citrus tristeza subgenomic RNAs. J. Virol. 71, 6233–6236.

Koev, G., Miller, W.A., 2000. A positive-strand RNA virus with three very different subgenomic RNA promoters. J. Virol. 74, 5988–5996.

Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375–430.

Mawassi, M., Mietkiewska, E., Hilf, M.E., Ashoulin, L., Karasev, A.V., Gafny, R., Lee, R.F., Garnsey, S.M., Dawson, W.O., Bar-Joseph, M., 1995. Multiple species of defective RNAs in plants infected with citrus tristeza virus. Virology 214, 264–268.

Mawassi, M., Mietkiewska, E., Goffman, R., Yang, G., Bar-Joseph, M., 1996. Unusual sequence relationship between two isolates of citrus tristeza virus. J. Gen. Virol. 77, 2359–2364.

Moreno, P., Guerri, J., Muñoz, N., 1990. Identification of Spanish strains of citrus tristeza virus (CTV) by analysis of double-stranded RNAs (dsRNA). Phytopathology 80, 477–482.

Nagy, P.D., Carpenter, C.D., Simon, A.E., 1997. A novel 3′-end repair mechanism in an RNA virus. Proc. Natl. Acad. Sci. U.S.A. 94, 1113–1118.

Navas-Castillo, J., Albiach-Martí, M.R., Gowda, S., Hilf, M.E., Garnsey, S.M., Dawson, W.O., 1997. Kinetics of accumulation of citrus tristeza virus RNAs. Virology 228, 92–97.

Pappu, H.R., Karasev, A.V., Anderson, E.J., Pappu, S.S., Hilf, M.E., Febres, V.J., Oakleaf, R.M.G., McCaffrey, M., Boyko, V., Gowda, S., Dolja, V.V., Koonin, E.V., Gumpf, D.J., Cline, K.C., Garnsey, S.M., Dawson, W.O., Lee, R.F., Niblett, C.L., 1994. Nucleotide sequence and organization of eight 3′open reading frames of the citrus tristeza closterovirus genome. Virology 199, 35–46.

Satyanarayana, T., Gowda, S., Boyko, V.P., Albiach-Martí, M.R., Mawassi, M., Navas-Castillo, J., Karasev, A.V., Dolja, V., Hilf, M.E., Lewandowski, D.J., Moreno, P., Bar-Joseph, M., Garnsey, S.M., Dawson, W.O., 1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. U.S.A. 96, 7433–7438.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Albiach-Martí, M.R., Rabindran, S., Dawson, W.O., 2002. The p23 protein of Citrus tristeza virus controls asymmetrical RNA accumulation. J. Virol. 76, 473–483.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Dawson, W.O., 2003. Frame-shift mutations in infectious cDNA clones of *Citrus tristeza virus*: a strategy to minimize toxicity of viral sequences to *Escherichia coli*. Virology 313, 481–491.

Sawicki, S.G., Sawicki, D.L., 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. 64, 1050–1056.

Uptain, S.M., Chamberlin, M.J., 1997. *Escherichia coli* RNA polymerase terminates transcription efficiently at rho-independent terminators on single-stranded DNA templates. Proc. Natl. Acad. Sci. U.S.A. 94, 13548–13553.

van der Vossen, E.A.G., Notenboom, T., Bol, J.F., 1995. Characterization of sequences controlling the synthesis of alfalfa mosaic virus subgenomic RNA in vivo. Virology 212, 663–672.

van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J.M., Snijder, E.J., 1999. Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. Proc. Natl. Acad. Sci. U.S.A. 96, 12056–12061.

Vives, M.C., Rubio, L., López, C., Navas-Castillo, J., Albiach-Martí, M.R., Dawson, W.O., Guerra, J., Flores, R., Moreno, P., 1999. The complete genome sequence of the major component of a mild citrus tristeza virus isolate. J. Gen. Virol. 80, 811–816.

Vives, M.C., Galipienso, L., Navarro, L., Moreno, P., Guerri, J., 2002. Characterization of two kinds of subgenomic RNAs produced by citrus leaf blotch virus. Virology 295, 328–336.

Von Hippel, P.H., Yager, T.D., 1992. The elongation-termination decision in transcription. Science 255, 809–812.

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

Wielgosz, K.S., Huang, H., 1997. Novel viral RNA species in Sindbis virus-infected cell. J. Virol. 71, 9108–9117.

Wilson, K.S., von Hippel, P.H., 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc. Natl. Acad. Sci. U.S.A. 92, 8793–8797.

Zuker, M., Mathews, D.H., Turner, D.H., 1999. Algorithms and thermodynamics for RNA secondary structure prediction. A practical guide. In: Barciszewski, J., Clark, B.F.C. (Eds.), RNA Biochemistry and Bio/Technology. Kluwer Academic Publishing, Dordrecht/Norwell, MA, pp. 11–43.