VIROLOGY 226, 153–160 (1996)
ARTICLE NO. 0641

In Vivo Repair of 3′-End Deletions in a TCV Satellite RNA May Involve Two Abortive Synthesis and Priming Events

CLIFFORD D. CARPENTER and ANNE E. SIMON[1]

Department of Biochemistry and Molecular Biology and the Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, Massachusetts 01003

Received August 28, 1996; accepted October 3, 1996

RNA viruses that do not have the stabilizing features of poly(A) tails or amino acids covalently linked to their 3′ ends must develop other means for protecting or repairing their genomes from damage caused by cellular RNases. We previously found that deletions in the single-stranded tails of a satellite RNA (sat-RNA D) associated with turnip crinkle virus are repaired *in vivo* (C. D. Carpenter and A. E. Simon, 1996, *J. Virol.* 70, 478–486). We now extend this analysis to show that sat-RNA D transcripts with 3′-end deletions of 5 bases give rise to wild-type sat-RNA, while deletions of 6 to 11 bases result in sat-RNA with additional deletions to the −14 position joined to internal TCV genomic RNA (or other) sequence followed by replacement of the terminal $C_{1-2}UGC_{1-3}$ motif. In addition, we have determined that the selection of internal TCV sequence used in the repair of sat-RNA D 3′ ends is not random and generation of these short TCV segments likely involves primer-mediated synthesis of abortive products facilitated by base-pairing between internal regions of TCV genomic RNA and oligoribonucleotides generated by abortive cycling from the 3′ end of the TCV genome.

INTRODUCTION

For RNA viruses to be successful pathogens, they must be able to undergo their life cycle of unpackaging, translation, replication, and repackaging in a cytoplasmic environment where their RNA genomes are exposed to the action of cellular RNases. Host mRNAs are protected from the activity of 3′ to 5′ exonucleases by the presence of a multimeric complex of poly(A)-binding proteins regularly spaced on the mRNA poly(A) tail (Kuhn and Pieler, 1996; Wormington *et al.,* 1996; Ross, 1996). RNA viruses may have developed different strategies to cope with the presence of cytoplasmic nucleases such as sequestering replicating RNA in specific membrane-associated, virus-induced amorphous cytopathic structures or vesicles (Garnier *et al.,* 1986; Froshauer *et al.,* 1988; Wellink *et al.,* 1988; Chu and Westaway, 1992; Bienz *et al.,* 1992). Although the tRNA-like structure at the 3′ end of many plant viruses plays an important role in replication (Skuzeski *et al.,* 1996), the amino acid covalently linked to the end of the tRNA-like structure may also provide protection against 3′-end degradation. In addition, loss of the 3′-terminal CCA of the tRNA-like structure *in vivo* can be rapidly and efficiently repaired, possibly using a cellular nucleotidyltransferase, suggesting that the structure has a telomeric function (Rao *et al.,* 1989).

Viruses that do not contain tRNA-like structures may have evolved other means of repairing damage to the 3′ ends of their RNAs. For example, poliovirus encodes a polymerase that functions not only as the viral RNA-dependent RNA polymerase (RdRp), but also as a terminal adenylyl transferase, an activity that could be used to repair truncations of the poliovirus poly(A) tail (Neufeld *et al.,* 1994). Furthermore, deletion of the 3′-terminal CCC from the genomic RNA of cymbidium ringspot tombusvirus or up to eight bases from the 3′ end of the associated satellite (sat-) RNA was repaired *in vivo* by an unknown mechanism (Dalmay *et al.,* 1993; Dalmay and Rubino, 1995).

Repair of 3′-end deletions of RNAs associated with turnip crinkle virus (TCV) was recently discovered to occur *in vivo* (Carpenter and Simon, 1996a). TCV has a single-stranded RNA genome of 4054 bases with five open reading frames (including one readthrough of an amber termination codon) and is associated with several sat-RNAs that are noncoding and therefore completely dependent on products encoded by the genomic RNA for replication, packaging, and movement through plants (Simon and Howell, 1986; Fig. 1A). Since the sat-RNAs (194 to 356 bases) are substantially shorter than TCV (the use of "TCV" refers to the genomic RNA), they have been excellent templates for determining sequences and structures required for replication by the TCV RdRp (Song and Simon, 1994). Sat-RNA D (194 bases) shares little sequence similarity with TCV beyond a 7-nucleotide sequence at the 3′ end ($CCUGCC_{1-2}$-3′) that is also found at the 3′ ends of all TCV-associated RNAs (Simon and Howell, 1986; Li *et al.,* 1989).

Recombinants generated *de novo* between TCV and

[1] To whom correspondence and reprint requests should be addressed. Fax: (413) 545-4529; E-mail: simon@biochem.umass.edu.

0042-6822/96 $18.00

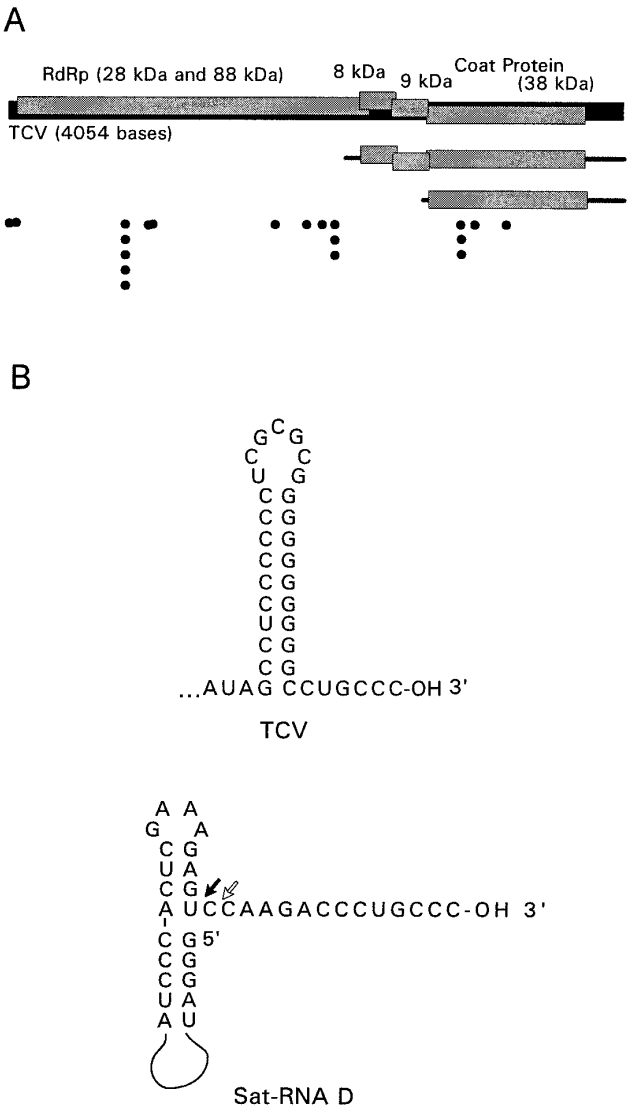

FIG. 1. TCV and its associated sat-RNA. (A) Map of the TCV genomic RNA. The two TCV subgenomic RNAs, used in the translation of downstream open reading frames, are shown below the full-length viral RNA. Filled circles denote the locations of homologous sequence in TCV found in the repaired 3′ ends of sat-RNA D (positions 3, 49, 768, 911, 922, 1703, 1899, 2022, 2081, 3001, 3063, and 3227). (B) Putative structures at the 3′ ends of TCV and sat-RNA D. The structure of sat-RNA D was determined by computer modeling using the MFOLD program from the GCG sequence analysis package (University of Wisconsin). The structure of TCV is based on computer modeling and enzymatic and chemical cleavages of the similar structure of sat-RNA C (Carpenter *et al.,* 1995). The filled arrow denotes the location of the 3′-end truncation (−14 bases) in sat-RNA D molecules with or without repaired 3′ ends. The open arrow denotes the major crossover site in recombinants generated *in vivo* between sat-RNA D and TCV (Carpenter *et al.,* 1995).

sat-RNA D can be detected in both plants (Zhang *et al.,* 1991; Carpenter *et al.,* 1995) and protoplasts (Carpenter and Simon, 1996b). Such recombination events occur frequently based on the recovery of recombinant molecules that were nearly all unique within a single infected leaf or preparation of infected protoplasts (Carpenter *et al.,* 1995; Carpenter and Simon, 1996a,b). Since the sat-RNA D/TCV recombinants generated in plants and protoplasts are apparently not biologically active molecules (Carpenter *et al.,* 1995; Carpenter and Simon, 1996a,b), the value of recombination leading to their production is not clear. However, we recently determined that transcripts of sat-RNA D containing deletions of up to 15 but not 18 bases of 3′-end sequence could be repaired by incorporation of sequence from internal regions of TCV joined to a 3′-terminal motif ($C_{1-2}UGC_{1-3}$-3′). Recent results analyzing repair of deletions of 3′-end sequence from a second TCV satellite RNA (sat-RNA C) *in vivo* and *in vitro* indicate that the origin of the $C_{1-2}UGC_{1-3}$ motif is the 3′ end of TCV, which is abortively synthesized in large quantities *in vitro* by the TCV RdRp (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted).

In this report, we extend our analysis of the 3′-end repair of sat-RNA D *in vivo* and have found that deletions of 5 bases are probably repaired differently from deletions of 6 to 13 bases. In addition, we have determined that the selection of internal TCV sequence used in the repair of sat-RNA D 3′ ends is not random and generation of these short TCV segments probably involves primer-mediated synthesis. We also propose that multiple rounds of abortive synthesis/primer-mediated synthesis may have led to the original formation of sat-RNA D.

MATERIALS AND METHODS

RNA synthesis and plant inoculations

cDNA clones of sat-RNA D (pT7D) and TCV (pTCV66) have been previously described (Song and Simon, 1994; Oh *et al.,* 1995). Both plasmids contain full-length cDNA of the respective genomic and subviral RNA downstream of a T7 RNA polymerase promoter. Transcripts generated from pTCV66 digested with *Sma*I contain exact 5′- and 3′-end sequence. Sat-RNA D transcripts prepared by digesting pT7D with *Bam*HI contain seven plasmid-derived bases (GGGAUC-3′) at the 3′ end. Six turnip plants were inoculated with 3 μg of TCV and 3 μg of sat-RNA D transcripts as previously described (Carpenter and Simon, 1996a). Control inoculations consisting of TCV transcripts only, were routinely performed to ensure that contaminating sat-RNA D was not in plants 2 weeks postinoculation. Total RNA was isolated from systemically infected leaves of plants 2 weeks postinoculation as previously described (Simon and Howell, 1986). RNAs were subjected to electrophoresis on 1% agarose gels and/or 5% polyacrylamide/8 *M* urea gels subsequently stained with ethidium bromide to assay for the presence of TCV and sat-RNA D.

Generation of deletions and mutations in the 3′ end of sat-RNA D

Generation of deletions in sat-RNA D has previously been described (Carpenter and Simon, 1996a). To gener-

ate specific mutations in sat-RNA D, pT7D was used as the template for polymerase chain reaction (PCR)-based mutagenesis. Primers containing altered bases spanned the 3′ ends of the cloned molecules and adjacent vector sequences that contained restriction enzyme sites. Upstream primers were selected that allowed restriction digestion of the PCR products to facilitate recloning into similarly digested parental clone vectors treated with shrimp alkaline phosphatase (Amersham). Mutations were confirmed by DNA sequence analysis.

Analysis of 3′-end sequences of sat-RNA D accumulating in plants

Sat-RNA D was gel purified from total plant RNA, and poly(A) tails were added to the 3′ ends using poly(A) polymerase as previously described (Carpenter and Simon, 1996a). cDNAs were synthesized using a primer containing a 19-base sequence joined to 17 T residues as previously described (Carpenter and Simon; 1996a). PCR was carried out using a primer homologous to positions 104 to 118 of sat-RNA D and the 19-base sequence described above. PCR products were cloned into the *Sma*I site of pUC19 and sequenced.

RESULTS AND DISCUSSION

Repair of sat-RNA D 3′ ends *in vivo*

In a previous study, we demonstrated that turnip plants inoculated with sat-RNA D transcripts containing deletions of one or all three of the terminal C residues (DΔ1, DΔ3; numbers refer to the number of bases deleted from the 3′ end of sat-RNA D) accumulated wild-type sat-RNA D in every plant, indicating that replacement of the C residues had occurred (Carpenter and Simon, 1996a). Sat-RNA D derived from DΔ6 or DΔ13 transcripts had additional deletions of 8 or 1 bases, respectively, resulting in a sat-RNA D core sequence missing the 3′-terminal 14 nucleotides (the deletion of an additional sat-RNA D nucleotide from transcripts of DΔ13 was mistakenly not noted in Carpenter and Simon, 1996a). The sat-RNA D core sequence was joined in most cases to a short sequence derived either from various regions of TCV or from a nonviral template followed by the motif $C_{1-2}UGC_{1-3}$-3′ (sat-RNA D containing this type of 3′-end sequence will be referred to as having a repaired 3′ end). Furthermore, clones of sat-RNA D with a 14-base deletion and no additional 3′-end sequence were generated from RNA isolated from many of the plants, including plants inoculated with wild-type sat-RNA D transcripts.

To determine if transcripts containing deletions of between 5 and 11 bases also resulted in truncations of the sat-RNA D sequence to the −14 position, six plants per construct were inoculated with transcripts of TCV and mutant sat-RNA D (all sat-RNA D transcripts also contained 6 plasmid-derived nucleotides at the 3′ end [GGGAUC-3′]) and RNA extracted 2 weeks later. All plants accumulated sat-RNA D-sized species (data not shown). The sat-RNA D-like species from two plants per construct was purified and cloned using a procedure that allows for the determination of the 3′ end of the RNA. As presented in Table 1, 12 of 13 sat-RNA D clones derived from wild-type transcripts contained full-length, wild-type sat-RNA D sequence, terminating in 2 or 3 C residues. The remaining clone contained a deletion of the 3′-terminal 14 bases. DΔ5 transcripts gave rise exclusively to clones containing full-length wild-type sequence in one plant; in the second plant, the majority of clones were of full-length wild-type sequence (with or without nontemplate U residues at the 3′ end) while the remaining

TABLE 1

New 3′-End Sequences in Sat-RNA D Generated *in Vivo* from Transcripts with 3′-End Deletions

T	Plant	3′-end sequence
wt		[GAGUCCAAGACCCUGCCCgggauc]
	1 and 2	GAGUCCAAGACCCUGCC(C) (12)
	1 and 2	GAGU (1)
DΔ5		[GAGUCCAAGACCCgggauc]
	1	GAGUCCAAGACC*CUGCC*(*C*) (12)
	2	GAGUCCAAGACC*CUGCCC*(*UU*) (5)
	2	GAGU (1)
	2	GAGUU (1)
	2	GAGUUAUC (1)
	2	GAGU*UAUCAAUUCUGCC* (1)
DΔ6		[GAGUCCAAGACCgggauc]
	1	GAGU*UUCUGAACUGC*(*CC*) (7)
	2	GAGU*UUAUUAUUCUGCC*(*C*) (3)
	2	GAGU*UUCUCUCUGCC* (2)
	2	GAGU (2)
DΔ9		[GAGUCCgggauc]
	1	GAGU*UUCUGAACUGC*(*CC*) (8)
	2	GAGU*UUAUUAUUCUGCC*(*C*) (2)
	2	GAGU*UUCUCUCUGCC* (2)
	2	GAGU (2)
DΔ11		[GAGUCCAgggauc]
	1	GAGUC*AUGUCUGCC*(*C*) (1)
	1	GAGUC*CUGUCUGCCC*(*UU*) (3)
	1	GAGU*UUUGUAAACUGCC* (1)
	2	GAGUC*AGAUGAUUGGCUGCC*(*C*) (6)
	2	GAGUC*AG* (1)
	2	GAGU*UUCACUGAACUGCCC* (1)

Note. The transcript sequence is enclosed in brackets. Only the sat-RNA D sequence from position 177 to the 3′ end is shown. Nucleotides in lowercase in the transcript sequence denote plasmid-derived bases. Sequence in italics denotes the repaired 3′ ends of the molecules. Sequence in bold denotes identity with a contiguous internal sequence of TCV. Sequence similar to the terminal motif found at the 3′ ends of all TCV-associated RNAs (CCUGCCC-3′) is underlined. Nucleotides in parentheses reflect clones that differed only in the number of C residues at the 3′ end or the presence of nontemplate U residues. Numbers in parentheses denote number of clones that were found with the sequence. All clones terminated in poly(A), which was used in the cloning procedure. T, transcript used in the inoculum; wt, wild-type transcript.

clones contained the 14-base terminal deletion alone, or the 14-base terminal deletion joined either to nontemplate U residues or to a repaired 3′ end. In contrast, no clones containing wild-type sequence were found in the previous study using transcripts of DΔ6 (1 additional deleted base; Carpenter and Simon, 1996a). To confirm this difference, sat-RNA D molecules were cloned from two additional plants inoculated with DΔ6 transcripts. As shown in Table 1, clones containing wild-type sat-RNA D sequence were not recovered from either plant, confirming our previous results. Furthermore, all clones had sat-RNA D sequence truncated by an additional 8 bases (14-base deletion total) either joined to no additional sequence or with repaired 3′ ends. Plants inoculated with DΔ9 and DΔ11 transcripts accumulated sat-RNA D molecules similar to those found for DΔ6. These results suggest that sat-RNA D with a 5-base deletion at the 3′ end is repaired differently from sat-RNA D deleted by 6 to 13 bases since only DΔ5 transcripts gave rise to wild-type sat-RNA D molecules. Furthermore, transcripts containing deletions of 6 to 13 bases were not used directly for 3′ end repair but required a further truncation, usually to the −14 position.

TABLE 2

3′-End Sequences in Sat-RNA D Generated *in Vivo* from Transcripts with 3′-End Deletions Using Helper Virus TCV_{A4050}

T	Plant	3′-end sequence
DΔ13		[GAGUCgggauc]
	1	GAGU*UACUCUACUGCC*(*C*) (6)
	1	GAGU*UACUCUACU* (1)
	1	GAGU*CUUCUCUGCC* (1)
	1	GAGU*UUUCGUACCAACUG* (1)
	1	GAGU*UGUCGUACCAACUGCU* (1)
	1	GAGU*UUUCUCUGCC* (1)
	2	GAGU*UCGUACCGACUGCCC* (1)
	2	GAGU*UUUCUCUGCC* (2)
	2	GAGU*UGUCGCAGCCACUG*CC (1)
	3	GAGU*AUUAUUCUGCCC* (1)
	3	GAGU*CCGCCCAACGACCUGCC(C)* (3)
	3	GAGU*CCGCCCAACGAACUGCCC* (1)
	3	GAGU*UAUUACUCUGCCC* (1)
TCV_{A4050}		[GGGGGGGGCCAGCCC]
	1 and 2	GGGGGGGGCCAGCCC (19)

Note. The underlined nucleotide is the base in sat-RNA D that corresponds to the position of the A residue mutation in the terminal motif of TCV_{A4050}. The TCV sequence is from position 4040 to the 3′ end. For other details, see footnote to Table 1.

Origin of the terminal motif used to repair deletions at the 3′ ends of sat-RNA D

We recently determined that the 3′-end motif of TCV (CCUGCCC-3′) was the source of the motif used to repair deletions in a second TCV sat-RNA, sat-RNA C, *in vivo.* This was shown by changing the terminal motif of TCV from CUGCCC-3′ to CAGCCC-3′ (TCV_{A4050}) and then demonstrating that inoculating turnip plants with sat-RNA C containing a 6-base, 3′-terminal deletion and TCV_{A4050} resulted in all sat-RNA C recovered terminating in CAGCCC-3′ (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted). To determine if the 3′ end of TCV was also the source of the $C_{1-2}UGC_{1-3}$ used to repair sat-RNA D transcripts with 3′-end deletions, turnip plants were inoculated with TCV_{A4050} and DΔ13. Sat-RNA D molecules accumulating in the plants at 2 weeks postinoculation were cloned. All plants inoculated with DΔ13 transcripts accumulated sat-RNA D-sized RNA (data not shown). Twenty of the 21 clones generated from sat-RNA D accumulating in the plants were deleted by an additional nucleotide (to −14) followed by 3′-repair sequences (Table 2). All sat-RNA D clones terminated in CUGCCC-3′ downstream from the various TCV- and non-TCV-related sequences.

Three alternative explanations exist for the presence of CUGCCC-3′, and not CAGCCC-3′, at the ends of the sat-RNA D accumulating with the helper virus TCV_{A4050}: (i) TCV_{A4050} reverted to wild-type in the plants inoculated with DΔ13 and the 3′ end of the revertant TCV supplied the 3′-terminal sequence; (ii) the sequence used for the repair of the 3′ ends of sat-RNA D is not the sequence from the 3′ end of TCV (i.e., the origin of the motifs used to repair sat-RNA C and sat-RNA D are different); (iii) the origin of the motif used in the repair of sat-RNA D and sat-RNA C is the same (the 3′ end of TCV), but due to the lack of substantial biological activity of sat-RNA D with an A at position 190, only sat-RNA D with a reversion to the wild-type U residue at position 190 are recovered.

To help distinguish among these possibilities, TCV was cloned from the same plants that were used to examine the sequence of sat-RNA D molecules derived from infection with TCV_{A4050} and DΔ13. All 19 clones contained TCV sequence terminating in CAGCCC-3′, indicating that TCV revertants were not accumulating to detectable levels in these plants (Table 2). To determine if sat-RNA D cannot tolerate nucleotides other than a U at position 190 (CUGCCC-3′), the U residue in wild-type sat-RNA D was modified to A, G, or C. Our reason for trying all three nucleotides at this position was that if an A residue was not tolerated, one of the other nucleotides might be, thereby allowing repetition of the experiment described above using TCV helper virus with the corresponding 3′-end alteration. Six turnip plants were inoculated with wild-type TCV and sat-RNA D transcripts with an A, G, or C residue at position 190. All plants accumulated sat-RNA D-sized RNA at 2 weeks postinoculation (data not shown), and all full-length clones generated from four plants terminated in CUGCCC-3′ (Table 3), indicating that simple reversion followed by selection had occurred. Several plants also contained sat-RNA D truncated to the −14 or −15

TABLE 3

Effect of Changing Sat-RNA D Position 190 from U to A, G, or C

Base 190	Plant	3′-end sequence
A		[GAGUCCAAGACCCAGCCCgggauc]
	1, 2, 3, and 4	GAGUCCAAGACCCUGCC(C) (5, 6, 5, 7)
	1	GAGU (1)
	2	GAG*CACUUCUGCC* (2)
	2	GAG*UUUUGACUGCC* (1)
	4	GAG*UUUUGACUGCCC* (1)
	4	GAG*UU* (2)
G		[GAGUCCAAGACCCGGCCCgggauc]
	1, 2, 3, and 4	GAGUCCAAGACCCUGCC(C) (5, 4, 4, 6)
	3	GAGUCCAAGACCC (1)
C		[GAGUCCAAGACCCCGCCCgggauc]
	1, 2, and 4	GAGUCCAAGAACCUGC(C*CU*) (3, 3, 4)
	1	GAGUCCAAGA*ACUGC* (1)
	1	GAGUCCAAG (1)
	1	GAGUCCAAGA. CCUGCC (1)
	2	GAGUCCAAGA*ACUGC* (1)
	3	GAGUCCAAGA. CCUGCC(C) (6)
	4	GAG*UU* (1)

Note. Position 190 in the transcript sequence is in bold. The base corresponding to position 190 in sat-RNA D accumulating in plants is underlined. Numbers in parentheses reflect the number of clones with the sequence derived from one or more plants. For other details, see footnote to Table 1.

position followed by typical 3′-end repair sequences. Interestingly, plants inoculated with sat-RNA D containing a C residue at position 190 accumulated sat-RNA D with previously unseen alterations. For example, two plants contained sat-RNA D with identical eight-base truncations followed by ACUGC at the 3′ terminus. In addition, several plants inoculated with the same construct contained sat-RNA D with second site mutations at position 187 consisting of a single base deletion or conversion from C to A. These results indicate that, unlike sat-RNA C and TCV, sat-RNA D does not tolerate any nucleotide other than U at position 190. In addition, simple reversion alone cannot account for the second site alterations found in several plants inoculated with transcripts containing a C at position 190.

To determine if the biological activity of sat-RNA D was also affected by alterations at other positions in the terminal CUGCCC-3′ motif, the following changes were made to sat-RNA D: C_{193} to G, C_{193} to A, and G_{191} to C. Plants inoculated with these transcripts and wild-type TCV accumulated sat-RNA D with wild-type sequence (29 of 45 clones), truncations to −14 and typical repaired 3′ ends (14 of 45 clones), or truncations of 12 or 14 nucleotides (2 of 45 clones; Table 4). The apparent strict requirement for the wild-type sequence at these positions in the terminal motif of sat-RNA D precludes determining conclusively the origin of the $C_{1-2}UGC_{1-3}$ at the 3′ ends of repaired sat-RNA D.

Primer-mediated synthesis for repair of sat-RNA 3′-end deletions

RdRp associated with RNA viruses are thought to initiate full-length genome replication in the absence of a primer *in vivo* although primer-directed synthesis has been shown to occur *in vitro* (McGeoch and Kitron, 1974; Flanegan and Baltimore, 1977; Van Dyke *et al.,* 1982; Honda *et al.,* 1986). Several viruses, including the plus-strand coronaviruses and minus-strand influenza virus, use a short RNA primer for transcription of subgenomic mRNAs (Bouloy *et al.,* 1978; Krug *et al.,* 1979; Lai *et al.,* 1983; Spaan *et al.,* 1983; Lai *et al.,* 1984). One possible mechanism for the addition of a fragment of TCV (or other sequence) and the TCV 3′ end motif to truncated sat-RNA D could begin with the synthesis of a primer complementary to the TCV 3′ six bases followed by primer-directed synthesis at internal TCV regions. If primer-mediated synthesis is involved in 3′-end repair, selection of internal TCV sequence might be aided by base-pairing between the primer and TCV just downstream from the sequence found in the sat-RNA D repaired 3′ ends.

The locations of TCV sequences found in sat-RNA D repaired 3′ ends to date (Carpenter and Simon, 1996a; this report) are shown in Fig. 1A (sequences were said to have a TCV origin if they contained at least 9/9 base identity). Although the internal TCV sequences used to repair sat-RNA D truncations are interspersed throughout TCV, specific sequences have been used to repair deleted sat-RNA D transcripts in multiple plants. The next six bases downstream from the internal regions used to repair sat-RNA D deletions were examined to determine if base-pairing with the complement of the TCV terminal motif (GACGGG-5′) could occur. As shown in Fig. 2A, 65% of the sequences just downstream from the TCV

TABLE 4

Effect of Mutations in the 3′-Terminal Motif on Sat-RNA D Sequence *in Vivo*

Mutation	Plant	3′ end sequence
C193 to G		[GAGUCCAAGACCCUGCGCgggauc]
	1 and 2	GAGUCCAAGACCCUGCCC(C) (2, 10)
	1	GA*UUUUCGACGCAUCUGCC*(*C*) (5)
	2	GAGU*UU* (1)
C193 to A		[GAGUCCAAGACCCUGCACgggauc]
	1	GAGUCCAAGACCCUGCCC(C) (10)
G191 to C		[GAGUCCAAGACCCUCCCCgggauc]
	1	GAGU*UAUUUGCAACUGCC*(*CU*) (8)
	1	GAGU*UAUUUGCAACUGCU* (1)
	2	GAGUCCAAGACCCUGCCCC(C) (7)
	2	GAGUCC (1)

Note. Mutation in the transcript is in bold. The base corresponding to the mutation in sat-RNA D accumulating in plants in underlined. Numbers in parentheses reflect the number of clones with the sequence derived from one or more plants. For other details, see footnote to Table 1.

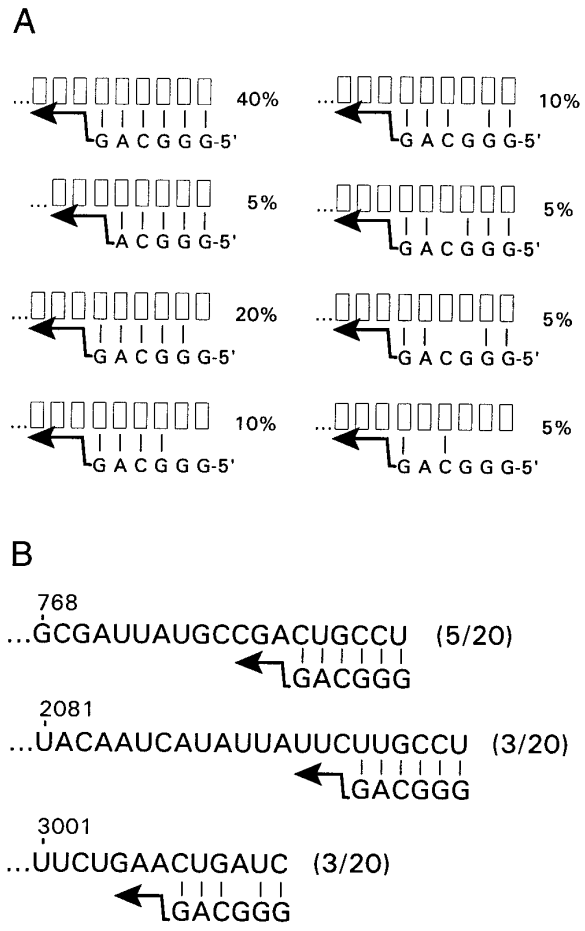

FIG. 2. Potential base-pairing between the putative primer synthesized from the 3′ end of TCV and sequences immediately downstream from internal regions of TCV used in repair of sat-RNA D 3′ ends. (A) The sequences downstream from 20 TCV sequences used for repair of sat-RNA D were examined for ability to base-pair with the primer (for the positions of these sequences, see legend to Fig. 1). Numbers denote percentage of downstream sequences able to base-pair as shown. (B) Potential priming downstream of the 3 most common sequences in TCV used in sat-RNA D 3′-end repair. Numbers in parentheses denote the number of repaired sat-RNA D with this TCV sequence over the total number of sat-RNA D repaired with TCV sequence.

internal regions can base-pair with five or six bases of contiguous sequence in the presumptive primer while an additional 25% of the sequences can base-pair with four contiguous or five of six noncontiguous bases of the presumptive primer. In addition, two of the three multiply selected internal sequences (positions 768 and 2081) had six of six base pairs between the downstream sequence and the complement of the TCV terminal motif while the remaining multiply selected sequence (position 3001) could form five of six base pairs (Fig. 2B). This result suggests that a six-base oligoribonucleotide synthesized from the 3′ end of TCV is used in primer-directed synthesis of internal regions of TCV. However, one of the internal TCV sequences could form only two of six downstream base pairs, suggesting that while base-pairing with the primer probably aids in synthesis of internal TCV sequence, it is not required. The ability of the TCV RdRp to participate in primer-directed synthesis using internal TCV sequence as a template is supported by our previous finding that longer than full-length products can be synthesized *in vitro* using TCV RdRp-active fractions from infected plants and sat-RNA C minus-strands as template (Song and Simon, 1995b); these products were generated by terminal elongation of the 3′ end of the template following base-pairing between the 3′ end and an internal region on the same template.

Model for the repair of deletions at the 3′ ends of the sat-RNAs

Our finding that the terminal motif used to repair sat-RNA C (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted) and probably sat-RNA D is derived from TCV and that primer-mediated synthesis of internal TCV regions is likely occurring allows for the refinement of the model proposed in Carpenter and Simon (1996a). It should be noted that the new model, shown in Fig. 3, describes the generation of viable sat-RNA D molecules, and that subsets of the steps may also be occurring but the molecules produced might not be good templates for further amplifi-

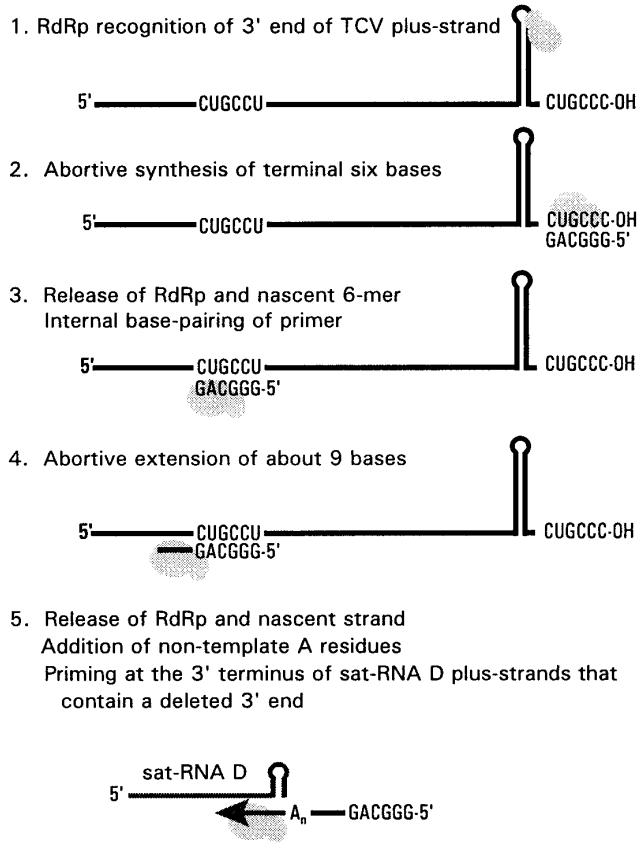

FIG. 3. Model for repair of sat-RNA D 3′ ends. Double oval denotes the TCV RdRp. Only synthesis of the minus strand is shown. See text for details.

cation in plants. In addition, although we have demonstrated that the motif used to repair truncations of sat-RNA C is derived from the 3′ end of TCV (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted), such a demonstration has not yet been possible for sat-RNA D due to the inviability of point mutations in the terminal motif of sat-RNA D (see above).

The initial step of the model involves the recognition by the RdRp of the stable hairpin near the 3′ end of the genomic RNA; the analogous hairpin in sat-RNA C and the 3′-terminal motif are known to be sufficient for promoting minus-strand synthesis *in vitro* (Song and Simon, 1995a). The complement of the single-stranded terminal motif could be synthesized by abortive cycling, which is known to occur for DNA-dependent RNA polymerases (Martin *et al.,* 1988). Abortive synthesis of short oligoribonucleotides complementary to the 3′ end of TCV may occur if some of the RdRp is inhibited from forming an elongation complex by the inability to efficiently unwind the very stable hairpin just upstream from the terminal motif. *In vitro,* high levels of 5- and 6-mer oligoribonucleotides are synthesized along with lower amounts of the 4-, 7-, and 8-mer species (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted). The RdRp, along with the oligoribonucleotide, could be released from the template after failure to establish an elongation complex, and then could reinitiate synthesis at an internal region of TCV (or other template) facilitated by base-pairing between the oligoribonucleotide primer and the internal sequence. The RdRp then synthesizes a short sequence from TCV (or other template) by a second abortive synthesis reaction (alternatively, addition of sufficient nontemplate bases to the oligoribonucleotide would obviate the apparent need for this step). Following release of the RdRp and the nascent RNA consisting now of the primer and internal TCV (or other) sequence, the RdRp can add nontemplate A residues to the nascent minus strand (resulting in nontemplate U residues on the plus strand), and then extend synthesis on the nascent RNA from position −14 of sat-RNA D. Extention by the RdRp on the nascent RNA from the −14 position of the sat-RNA D template should be possible without base-pairing between the nascent strand and the template since recent studies have shown that base-pairing is not required for primer extension using oligoribonucleotide primers and sat-RNA C containing 3′-end truncations (P. D. Nagy, C. D. Carpenter, and A. E. Simon, manuscript submitted). Repair of DΔ5 transcripts does not require steps 3 to 5 of the model, but may use the 3′ hydroxyl of the primer directly to elongate on DΔ5 transcripts with initiation beginning just past the plasmid-derived bases. It is possible that some portion of normal sat-RNA replication is also initiated with a primer derived from TCV, and therefore primer-mediated synthesis could also be responsible for the reversion to wild-type of mutations within the terminal motif of sat-RNA D.

Computer predictions of the secondary structure at the 3′ end of sat-RNA D suggest that the terminal 14 nucleotides are single-stranded, whereas nucleotides just upstream are base-paired (Fig. 1B). The finding of sat-RNA D with deletions of the 14 3′-terminal bases in nearly every plant could be due to deletion of the unprotected 3′ end by cellular 3′ to 5′ exonucleases, or initiation of minus-strand synthesis by the TCV RdRp at the −14 position at the base of the hairpin. Interestingly, the major crossover site in sat-RNA D in recombinants generated *in vivo* between sat-RNA D and TCV is at position −13 (Fig. 1B; Carpenter *et al.,* 1995), suggesting that the mechanism leading to recombination between sat-RNA D and TCV and repair of 3′ ends using internal TCV sequence and replacement of the terminal motif occur by different overall mechanisms.

This mechanism, employing elongation of primers for replication leading to functional molecules, is similar to that proposed for replication of the Mauriceville plasmid associated with some *Neurospora* strains (Lambowitz and Chiang, 1995). The Mauriceville plasmid, although composed of DNA that replicates through an RNA intermediate, has many similarities to plant RNA viruses including the presence of a tRNA-like structure at the 3′ end of the RNA intermediate. Reverse transcription of the RNA either begins one base from the 3′ end in the absence of a primer or can proceed by extension from the 3′ ends of random short DNA or RNA primers *in vitro* or *in vivo* at or near the 3′ end of the RNA template without the need for base-pairing between the primer and the RNA (Wang *et al.,* 1992; Kennell *et al.,* 1994). The presence of nontemplate U residues between the primer and template-derived sequence in some Mauriceville plasmid cDNAs is also similar to our findings for sat-RNA D repaired 3′ ends.

Primer-mediated synthesis and the origin of sat-RNAs

The mechanism used by the RdRp to repair the 3′ ends of sat-RNA D may have been used in the original formation of sat-RNA D. Attempts to align full-length sat-RNA D with TCV using the GAP program of the GCG package (University of Wisconsin) did not reveal any obvious similarities outside the 3′-terminal 7 nucleotides. However, if sat-RNA D originated from the 3′-terminal motif of TCV and interior short regions of TCV produced by multiple abortive synthesis reactions, then aligning short segments of sat-RNA D with both the plus and the minus strands of TCV might reveal additional similarities. The BESTFIT program of the GCG package was therefore used to align 25 base segments of sat-RNA D with the plus and minus strands of TCV. The sequence just upstream from the terminal motif of sat-RNA D and TCV (positions 179 to 188) is identical to positions 3538 to 3547 of plus-strand TCV (10/10). In addition, positions 81 to 91 of sat-RNA D is identical to positions 558 to 568 of

plus-strand TCV (11/11). No other region of sat-RNA D shares nonrandom similarity with plus- or minus-strand TCV although it is possible that mutations accumulating over years of propagation of both TCV and sat-RNA D have obscured TCV as the origin of other segments.

ACKNOWLEDGMENTS

We thank Dr. Peter D. Nagy for helpful discussions and critical review of the manuscript. The work was supported by National Science Foundation Grants MCB-9315948 and MCB-9419303 to A.E.S.

REFERENCES

Bienz, K., Pfister, T., and Troxler, M. (1992). Structural and functional characterization of the poliovirus replication complex. *J. Virol.* 66, 2740–2747.

Bouloy, M., Plotch, S. J., and Krug, R. M. (1978). Globin mRNAs are primers for the transcription of influenza viral RNA *in vitro. Proc. Natl. Acad. Sci. USA* 75, 4886–4890.

Carpenter, C. D., and Simon, A. E. (1996a). *In vivo* restoration of biologically active 3′ ends of virus-associated RNAs by nonhomologous RNA recombination and replacement of a terminal motif. *J. Virol.* 70, 478–486.

Carpenter, C. D., and Simon, A. E. (1996b). Changes in locations of crossover sites over time in *de novo*-generated RNA recombinations. *Virology* 223, 165–173.

Carpenter, C. D., Oh, J.-W., Zhang, C., and Simon, A. E. (1995). Involvement of a stem–loop structure in the location of junction sites in viral RNA recombination. *J. Mol. Biol.* 245, 608–622.

Chu, P. W. G., and Westaway, E. G. (1992). Molecular and ultrastructural analysis of heavy membrane fractions associated with the replication of Kunjin virus RNA. *Arch. Virol.* 125, 177–191.

Dalmay, T., and Rubino, L. (1995). Replication of cymbidium ringspot virus satellite RNA mutants. *Virology* 206, 1092–1098.

Dalmay, T., Russo, M., and Burgyán, J. (1993). Repair *in vivo* of altered 3′ terminus of cymbidium ringspot tombusvirus RNA. *Virology* 192, 551–555.

Flanegan, J. B., and Baltimore, D. (1977). Poliovirus-specific primer-dependent RNA polymerase able to copy poly(A). *Proc. Natl. Acad. Sci. USA* 74, 3677–3680.

Froshauer, S., Kartenbeck, J., and Helenius, A. (1988). Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J. Cell Biol.* 107, 2075–2086.

Garnier, M., Candresse, T., and Bove, J. M. (1986). Immunocytochemical localization of TYMV-coded structural and nonstructural proteins by the protein A–gold technique. *Virology* 151, 100–109.

Honda, A., Mizumoto, K., and Ishihama, A. (1986). RNA polymerase of influenza virus: Dinucleotide-primed initiation of transcription at specific positions on viral RNA. *J. Biol. Chem.* 261, 5987–5991.

Kennell, J. C., Wang, H., and Lambowitz, A. M. (1994). The Mauriceville plasmid of *Neurospora* spp. uses novel mechanisms for initiating reverse transcription *in vivo. Mol. Cell. Biol.* 14, 3094–3107.

Krug, R. M., Broni, B. B., and Bouloy, M. (1979). Are the 5′ ends of influenza viral mRNAs synthesized *in vitro* donated by host mRNAs? *Cell* 18, 329–334.

Kühn, U., and Pieler, T. (1996). *Xenopus* poly(A) binding protein: Functional domains in RNA binding and protein–protein interaction. *J. Mol. Biol.* 256, 20–30.

Lai, M. M. C., Baric, R. S., Brayton, P. R., and Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. USA* 81, 3626–3630.

Lai, M. M. C., Patton, C. D., Baric, R. S., and Stohlman, S. A. (1983). Presence of leader sequences in the mRNA of mouse hepatitis virus. *J. Virol.* 46, 1027–1033.

Lambowitz, A. M., and Chiang, C.-C. (1995). The Mauriceville and Varkud plasmids, primitive retroelements found in *Neurospora* mitochondria. *Can. J. Bot.* 73, S173–S179.

Li, X. H., Heaton, L., Morris, T. J., and Simon, A. E. (1989). Defective interfering RNAs of turnip crinkle virus intensify viral symptoms and are generated *de novo. Proc. Natl. Acad. Sci. USA* 86, 9173–9177.

Martin, C. D., Muller, K., and Coleman, J. E. (1988). Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* 27, 3966–3974.

McGeoch, D., and Kitron, N. (1974). Influenza virion RNA-dependent RNA polymerase, stimulation by guanosine and related compounds. *J. Virol.* 15, 686–695.

Neufeld, K. L., Galarza, J. M., Richards, O. C., Summers, D. F., and Ehrenfeld, E. (1994). Identification of terminal adenylyl transferase activity of the poliovirus polymerase $3D^{pol}$. *J. Virol.* 68, 5811–5818.

Oh, J.-W., Kong, Q., Song, C., Carpenter, C. D., and Simon, A. E. (1995). Open reading frames of turnip crinkle virus involved in satellite symptom expression and incompatibility with *Arabidopsis* ecotype Dijon. *Mol. Plant–Microbe Interact.* 8, 979–987.

Rao, A. L. N., Dreher, T. W., Marsh, L. E., and Hall, T. C. (1989). Telomeric function of the tRNA-like structure of brome mosaic virus RNA. *Proc. Natl. Acad. Sci. USA* 86, 5335–5339.

Ross, J. (1996). Control of messenger RNA stability in higher eukaryotes. *Trends Genet.* 12, 171–175.

Simon, A. E., and Howell, S. H. (1986). The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3′-end of the helper virus genome. *EMBO J.* 5, 3423–3428.

Skuzeski, J. M., Bozarth, C. S., and Dreher, T. W. (1996). The turnip yellow mosaic virus tRNA-like structure cannot be replaced by generic tRNA-like elements or by heterologous 3′ untranslated regions known to enhance mRNA expression and stability. *J. Virol.* 70, 2107–2115.

Song, C., and Simon, A. E. (1994). RNA-dependent RNA polymerase from plants infected with turnip crinkle virus can transcribe (+)- and (−)-strands of virus-associated RNAs. *Proc. Natl. Acad. Sci. USA* 91, 8792–8796.

Song, C., and Simon, A. E. (1995a). Requirement of a 3′-terminal stem-loop in *in vitro* transcription by an RNA-dependent RNA polymerase. *J. Mol. Biol.* 254, 6–14.

Song, C., and Simon, A. E. (1995b). Synthesis of novel products *in vitro* by an RNA-dependent RNA polymerase. *J. Virol.* 69, 4020–4028.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. M., and Siddell, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* 2, 1839–1844.

Van Dyke, T. A., Rickles, R. J., and Flanegan, J. B. (1982). Genome-length copies of poliovirion RNA are synthesized *in vitro* by the poliovirus RNA-dependent RNA polymerase. *J. Biol. Chem.* 257, 4610–4617.

Wang, H., Kennell, J. C., Kuiper, M. T. R., Sabourin, J. R., Saldanha, R., and Lambowitz, A. M. (1992). The Mauriceville plasmid of *Neurospora crassa:* Characterization of a novel reverse transcriptase that begins cDNA synthesis at the 3′ end of template RNA. *Mol. Cell. Biol.* 12, 5131–5144.

Wellink, J., Van Lent, J., and Goldbach, R. (1988). Detection of viral proteins in cytopathic structures in cowpea protoplasts infected with cowpea mosaic virus. *J. Gen. Virol.* 69, 751–755.

Wormington, M. W., Searfoss, A. M., and Hurney, C. A. (1996). Overexpression of poly(A) binding protein prevents maturation-specific deadenylation and translational inactivation in *Xenopus* oocytes. *EMBO J.* 15, 900–909.

Zhang, C., Cascone, P. J., and Simon, A. E. (1991). Recombination between satellite and genomic RNAs of turnip crinkle virus. *Virology* 184, 791–794.