

## Short Internal Sequences Involved in Replication and Virion Accumulation in a Subviral RNA of *Turnip Crinkle Virus*

Xiaoping Sun,<sup>†</sup> Guohua Zhang,<sup>†</sup> and Anne E. Simon\*

Department of Cell Biology and Molecular Genetics, University of Maryland College Park,  
College Park, Maryland

Received 20 May 2004/Accepted 18 September 2004

***cis*-acting sequences and structural elements in untranslated regions of viral genomes allow viral RNA-dependent RNA polymerases to correctly initiate and transcribe asymmetric levels of plus and minus strands during replication of plus-sense RNA viruses. Such elements include promoters, enhancers, and transcriptional repressors that may require interactions with distal RNA sequences for function. We previously determined that a non-sequence-specific hairpin (M1H) in the interior of a subviral RNA (satC) associated with *Turnip crinkle virus* is required for fitness and that its function might be to bridge flanking sequences (X. Sun and A. E. Simon, *J. Virol.* 77:7880–7889, 2003). To establish the importance of the flanking sequences in replication and satC-specific virion repression, segments on both sides of M1H were randomized and subjected to *in vivo* functional selection (*in vivo* SELEX). Analyses of winning (functional) sequences revealed three different conserved elements within the segments that could be specifically assigned roles in replication, virion repression, or both. One of these elements was also implicated in the molecular switch that releases the 3' end from its interaction with the repressor hairpin H5, which is possibly involved in controlling the level of minus-strand synthesis.**

Replication of positive-strand RNA viruses is a multistep process mediated by the virally encoded RNA-dependent RNA polymerase (RdRp). First, the invading plus-strand genomic RNA is released from the capsid, recruited by ribosomes, and then translated to produce components of the RdRp. Next, in a poorly understood process that may require the clearance of ribosomes (1), the genomic plus-strand RNA switches from a translation template to a replication template. RdRp and possibly other viral or host factors (16) form an initiation complex at the promoter and initiate *de novo* or primer-directed transcription of the complementary minus strand. The minus-strand intermediate is then used as a template for synthesis of plus strands. The relative levels of the two strands are often highly asymmetric, with ratios of up to 1,000 plus strands for every minus strand produced (2). This asymmetric accumulation of complementary strands suggests the existence of a second switch, one that represses minus-strand synthesis, thereby constraining the RdRp to synthesize plus strands. The switch from minus-strand to plus-strand synthesis is likely mediated by *cis*-acting elements that allow or deny access of the RdRp to the plus-strand promoter or 3'-terminal sequences.

Diverse RNA secondary or tertiary structures such as hairpins, pseudoknots, and tRNA-like structures exist in viral 3' untranslated regions that can play important roles in virus replication (4). Some viruses regulate switches that occur during viral RNA replication by changing the conformation of 3'-proximal structures, which may be mediated by one or more unstable base pairs occurring between complementary short

sequences located within and outside hairpins (13, 23, 25, 45). For example, *Barley yellow dwarf virus* is proposed to repress minus-strand RNA production by altering the conformation of its 3' end to a “pocket” structure in which the initiation site for transcription is embedded in a stem and unavailable to the RdRp (13). A similar molecular switch in the coronavirus *Mouse hepatitis virus* genome involves sequences within a stem-loop and pseudoknot in the 3' untranslated region (6). In addition to *cis*-acting sequences, *trans*-acting cellular factors or virally encoded proteins may also affect the balance between alternative structural conformations (23).

Sequences and structural elements involved in efficient viral RNA synthesis are not limited to the 3'-proximal regions of viral RNAs. The 5' ends of a growing number of viruses have been implicated in acting in minus-strand synthesis by communicating with the 3' end via RNA-RNA (12, 42) or RNA-protein (5, 10, 11) interactions, causing cyclization of the genome. RNA elements located in variable positions on both strands that function as transcriptional enhancers or repressors may also be involved in 3' structural conformation changes and/or sequestration of 3'-terminal bases (25, 28–30, 39, 45).

*Turnip crinkle virus* (TCV) (family *Tombusviridae*, genus *Carmovirus*) has a 4,054-base single-stranded RNA genome that encodes five proteins (9) (Fig. 1A). p28 and p88 (a translational readthrough product of p28 that contains the polymerase active-site consensus sequence GDD) are translated from the genomic RNA. Purified p88 expressed in *Escherichia coli* can correctly transcribe TCV-specific plus- and minus-strand templates into complementary strands (27). p8 and p9, which are required for cell-to-cell virus movement, are translated from the 1.7-kb subgenomic RNA and are dispensable for replication (9, 17). The coat protein (CP), which is translated from the 1.45-kb subgenomic RNA, packages TCV-associated RNAs into a 180-subunit, T = 3 icosahedral virion. Whereas

\* Corresponding author. Mailing address: Department of Cell Biology and Molecular Genetics, Microbiology Building, University of Maryland College Park, College Park, MD 20742. Phone: (301) 405-8975. Fax: (301) 805-1318. E-mail: Anne\_Simon@umail.umd.edu.

<sup>†</sup> X.S. and G.Z. contributed equally to this work.

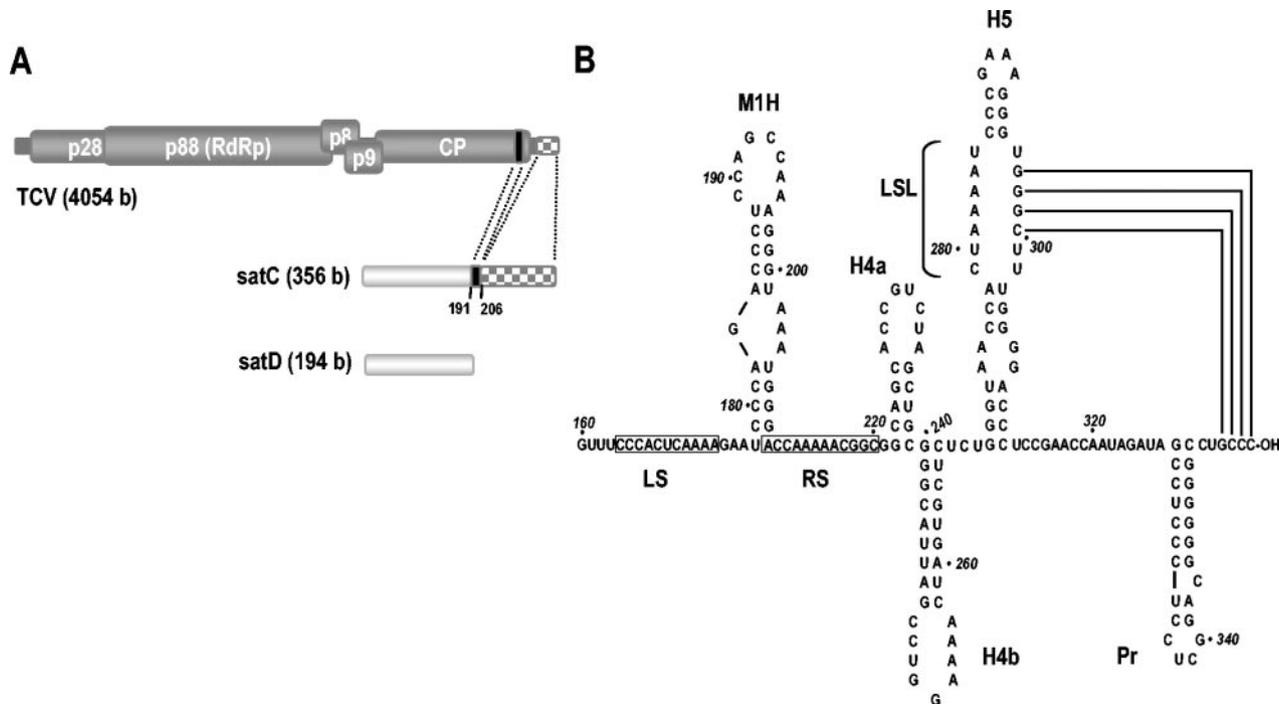


FIG. 1. Relationship among TCV-associated RNAs. (A) satC is a chimeric RNA composed of satD and TCV sequences. (B) Structure of the 3' region of satC. The structure was determined by computer structural predictions (47) and phylogenetic similarities and was partially confirmed by solution structural probing for position 170 to the 3' end (45). M1H is a replication enhancer in its minus-sense orientation and is required as a hairpin on plus strands for satC repression of virion accumulation (21, 37, 44). H4a and H4b have unknown functions and are structurally conserved among related carmoviruses (45). H5 is required for replication and is a repressor of minus-strand transcription in vitro (45). An interaction between the 3' side of the H5 LSL and the 3'-terminal GCCC (45) is shown. Pr is the core promoter for transcription of minus strands (35). Boxed sequences were subjected to in vivo genetic selection (in vivo SELEX). Numbering is from the 5' end.

the CP is dispensable for replication of TCV genomic and sub-genomic RNAs, it positively or negatively influences the replication of different TCV-associated subviral RNAs (14, 15).

TCV is naturally associated with several dependent subviral RNAs, including three satellite RNAs (satRNAs) and one defective interfering RNA (18, 32–34). None of these subviral RNAs are templates for translation, and all are dependent on TCV-encoded proteins for replication and encapsidation. The sequence relationship between TCV and two of its satRNAs, satC and satD, is shown in Fig. 1A. satC (356 bases) is composed of nearly full-length satD (194 bases) at its 5' end and two discontinuous segments from the TCV genome at its 3' end. Due to its sequence relationship with TCV and its small size, satC is an excellent template for examining *cis*-acting sequence and structural requirements for virus replication (32).

As shown in Fig. 1B, the TCV-derived region of satC, combined with a portion derived from satD, folds into a series of stem-loop structures as predicted by Mfold version 3.1 (47) and partially verified (for positions 170 to 356) by solution structure analysis (45). The 3'-proximal hairpin (Pr) is a conserved feature among carmoviruses (45) and can function independently as a core promoter for minus-strand synthesis (35). Pr is flanked on its 3' side by CCUGCCC-OH, which is conserved at the 3' termini of all TCV-associated plus-strand RNAs. The hairpin just upstream of Pr, hairpin 5 (H5), is also structurally and spatially conserved among carmoviruses. H5 was recently identified as a repressor of minus-strand synthesis in vitro (45) and is required for satC (46) and TCV (19)

replication in vivo. A major distinguishing feature of H5 is a large symmetrical internal loop (LSL) whose sequence is highly conserved among most carmoviruses (45, 46). Repression of minus-strand synthesis mediated by H5 was attributed to base pairing between the 3'-side LSL GGGC (positions 297 to 300) and the 3'-terminal GCCC-OH (positions 353 to 356) (Fig. 1B). Deletion of the three cytidylates at the 3' terminus of satC resulted in a substantial increase in transcription of both full-length and aberrantly initiated products. Whereas the H5 structure is not apparent in solution structure analysis of wild-type (wt) satC, a significant rearrangement of the H5 structure and 3' flanking sequences occurs after deletion of the three terminal cytidylates, which reveals the phylogenetically conserved structure (45). These findings led to the proposal that the interaction between 3'-terminal bases and H5 sequesters the 3' terminus from the RdRp.

The 30-base motif 1 hairpin (M1H) consists of sequences derived from satD and TCV. M1H is a hot spot for recombination between satC and satD in vivo (3, 20) and a replication enhancer in its minus-sense orientation in vivo and in vitro (21, 22). In addition, a plus-strand hairpin in this location is important for satC fitness, possibly functioning to bridge flanking sequences (37, 44). By presumptively juxtaposing these flanking sequences, satC is able to interfere efficiently with TCV virion formation, resulting in accumulation of additional free CP to better suppress RNA silencing (26, 37, 38, 43), a potent antiviral defense system in plants (24, 40). In addition to a role

in suppressing virion accumulation, sequences flanking M1H have been implicated in the replication of satC (22, 37).

In the present study, we determined the importance of specific sequences flanking M1H for satC fitness and established their role in replication and virion repression by sequence randomization followed by in vivo functional selection (in vivo SELEX). Analyses of winning (functional) satRNAs revealed three different conserved elements within the regions that could be specifically assigned to roles in replication, virion repression, or both. One of these elements was implicated in the molecular switch that releases the 3' end from H5 in satC and possibly controls the level of minus-strand synthesis.

## MATERIALS AND METHODS

**In vivo SELEX.** In vivo functional selection was performed as previously described (37). To produce full-length satC containing random bases replacing the LS and RS regions, two fragments were generated by separate PCRs with pC+ (pUC19 containing full-length satC cDNA) as a template. The 5' fragment was produced by using primers T7C5' (5'-GTAATACGACTACTATAGGG ATAATAAGGG-3'), which contains a T7 polymerase promoter at its 5' end, and SELF (5'-AGTCGGATCCNNNNNNNNNAAACAGCCAGGUUUU CACGC-3'); the BamHI site in the primer sequence [underlined] results in an adenylate-to-guanylate transition at position 176, and N represents equal ratios of the four nucleotides). The 3' fragment was generated by using primers SEL3F (5'-AGTCGGATCCAGACCCTCCAGCCAAAGGGTAAATGGGNNNN NNNNNNNGGCAGCACCGTCTAGCTGCG-3') and oligo 7 (5'-GGGCAG GCCCCCGTCCGA-3'), which is complementary to satC 3'-terminal sequence. satC containing the BamHI site is referred to as satC<sub>B</sub> (37). The 5' and 3' PCR fragments were treated with BamHI, purified through 1.5% agarose gels, and then ligated to produce full-length cDNA. satC<sub>B</sub> transcripts containing randomized sequences were directly synthesized from the ligation products by using T7 RNA polymerase. For the first round of selection, 5 µg of these satC<sub>B</sub> transcripts and 2 µg of TCV transcripts were coinoculated onto each of 30 turnip seedlings. Total RNA was extracted from uninoculated leaves at 21 days postinoculation (dpi). Full-length satC was amplified by reverse transcription-PCR (RT-PCR) with primers T7C5' and oligo 7 and then cloned into the SmaI site of pUC19 and sequenced. For the second round of selection, equal amounts of total RNAs extracted from all first-round plants were pooled. Approximately 5 µg of the pooled RNAs was inoculated onto each of six new turnip seedlings. Total RNA was extracted from uninoculated leaves at 21 dpi, and the recovered satC was cloned and sequenced as described above.

**Fitness of SELEX winners in turnip plants.** Fitness competitions between winners were performed as previously described (7, 8). Equal amounts of winning RNA transcripts (0.2 µg each/plant) were combined and inoculated onto a single leaf of three turnip seedlings together with TCV genomic RNA transcripts (2 µg/plant). Total RNA was extracted at 21 dpi, and satC was cloned and sequenced as described above.

**Construction of satC mutants.** Mutants LS1, LS2, RS1, RS2, LS3, and LS1RS1 were produced by ligation of two PCR fragments amplified with plasmid pC+ as a template. The 5' fragments of LS1, LS2, and LS3 were amplified by using primers T7C5' and 5'-AGTCGGATCCTCAATGGCTGGGAAACAG CCAGGUUUUACACGC-3', 5'-AGTCGGATCCTTTTGCTCGGAAACAGCC AGGUUUUACACGC-3', or 5'-GACTGGATCCTTTTGAGTCTCAAACAGC CAGTTTC-3', respectively (the BamHI site is underlined), while the 3' fragments were produced by PCR with primers oligo 7 and 5'-AGTCGGATCCCA GACCCTCCAGCCAAAGGGTAAATGGGTATGTGAAATGAGGCAGCA CCGTCTAGCTGCG-3'. For RS1 and RS2, the 5' fragments were generated by PCR with primers T7C5' and 5'-AGTCGGATCCTCAATGCTCGGAAACAG CCAGGUUUUACACGC-3'; the 3' fragments were obtained by using primers oligo 7 and 5'-AGTCGGATCCCAAGACCCTCCAGCCAAAGGGTAAATGG GACCAAAAAATGAGGCAGCACCGTCTAGCTGCG-3' or 5'-AGTCGGA TCCAGACCCTCCAGCCAAAGGGTAAATGGGTATGTGAACGGCGG CAGCACCGTCTAGCTGCG-3'. The 5' fragment for construction of LS1 and the 3' fragment for construction of RS1 were used to produce LS1RS1. 5' and 3' PCR fragments were digested with BamHI, gel purified, ligated together, and inserted into the SmaI site of pUC19. All clones were confirmed by sequencing. Transcripts for protoplast inoculation and in vitro transcription with TCV RdRp were synthesized in vitro by digestion of plasmids with SmaI followed by tran-

scription with T7 polymerase. This process generates RNAs with wt 3' and 5' ends.

**Protoplast inoculation.** Protoplasts ( $5 \times 10^6$ ) prepared from callus cultures of *Arabidopsis thaliana* ecotype Col-0 were inoculated with 20 µg of TCV genomic RNA transcripts and 2 µg of satC transcripts as described previously (7). Total RNA and virions were extracted from protoplasts at 40 h postinoculation (hpi) as previously described (41).

**Northern and Western blots.** Total RNA was denatured with formamide and separated by nondenaturing agarose gel electrophoresis as previously described (41). RNA was hybridized with a [ $\gamma$ -<sup>32</sup>P]ATP-labeled probe complementary to positions 3950 to 3970 of TCV genomic RNA and positions 250 to 269 of satC. Minus-strand RNAs were detected with a [ $\alpha$ -<sup>32</sup>P]UTP-labeled probe complementary to the 5'-terminal 151 bases of satC minus strands and the 5'-terminal 155 bases of TCV minus strands with nine single-base mismatches.

Western blotting was performed with polyclonal antibody against TCV CP and detection by chemiluminescence (Pierce Biotech), as previously described (41).

**In vitro transcription with purified recombinant TCV p88.** The plasmid expressing TCV p88 was a generous gift of P. D. Nagy (University of Kentucky). The expression of p88 in *E. coli* and purification of p88 were carried out as previously described (27). In vitro RdRp assays were performed as previously described (45). Briefly, 1 µg of purified RNA template was added to a 25-µl reaction mixture containing 50 mM Tris-HCl (pH 8.2); 100 mM potassium glutamate; 10 mM MgCl<sub>2</sub>; 10 mM dithiothreitol; 1 mM (each) ATP, CTP, and GTP; 0.01 mM UTP; 10 µCi of [ $\alpha$ -<sup>32</sup>P]UTP (Amersham); and 2 µg of p88. After 90 min of incubation at 20°C, 1 µg of yeast tRNA was added and the mixture was subjected to phenol-chloroform extraction and ammonium acetate-isopropanol precipitation. Radiolabeled products were analyzed by denaturing 8 M urea-5% polyacrylamide gel electrophoresis followed by autoradiography.

**RNA solution structure probing with reverse transcriptase.** Solution structure probing was performed as previously described (45). Briefly, purified satC transcripts (11 µg) synthesized by using T7 RNA polymerase and pT7C or pT7CA3C as a template (45) were mixed with 110 µg of yeast tRNA and 675 µl of modification buffer (70 mM HEPES [pH 7.5], 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 100 mM KCl). The mixture was heated to 90°C, slowly cooled to 35°C, and incubated at 25°C for 20 min. Fifty-microliter samples of the RNA were added to an equal volume of modification buffer containing either no additional reagents (control) or one of the following: 10% (vol/vol) diethylpyrocarbonate (Sigma), 1% (vol/vol) dimethyl sulfate (DMS) (Sigma), 0.05 U of RNase T<sub>1</sub> (Ambion), or 0.03 U of RNase V<sub>1</sub> (Ambion). Primer extension reactions were performed with 1 pmol of an oligonucleotide radiolabeled with [ $\alpha$ -<sup>35</sup>S]dATP that was complementary to positions 272 to 293.

**RNA solution structure probing with 3'-end-labeled transcripts.** RNA structure probing was performed by using protocols and reagents obtained from Ambion. Briefly, satC transcripts (50 pmol) were labeled at the 3' end with [<sup>32</sup>P]pCp (Amersham) and T4 RNA ligase as recommended by the manufacturer. Gel-purified transcripts were added to a mixture containing RNA structure buffer, yeast tRNA, and either water or RNase T<sub>1</sub> (0.01 to 0.001 U/µl), RNase A (0.01 to 0.001 µg/ml), or RNase V<sub>1</sub> (0.001 to 0.0001 U/µl) and incubated at 22°C for 15 min. Samples were precipitated and subjected to electrophoresis through 10% polyacrylamide-7 M urea gels, followed by autoradiography. Alkaline hydrolysis ladders were obtained by treatment of 3'-end-labeled RNA with alkaline hydrolysis buffer at 95°C for 5 min. RNase T<sub>1</sub> ladders were obtained by RNase T<sub>1</sub> (0.1 to 0.01 U/µl) digestion of heat-denatured 3'-end-labeled RNA at 22°C for 15 min.

## RESULTS AND DISCUSSION

**In vivo genetic selection of the satC LS and RS regions.** Our previous study on the sequence and structural requirements of the satC M1H replication enhancer suggested that flanking sequences contribute to both replication and interference with TCV virion accumulation (37). Understanding the underlying mechanisms responsible for these properties requires first identifying specific functional sequences within these regions. To identify such elements, in vivo SELEX was performed, which allows for selection of sequences important for fitness regardless of function. SELEX was performed by replacing the 11-base region on the 5' side of M1H (designated LS) and the 12-base region on the M1H 3' side (designated RS) with random sequences (Fig. 1B). This construction required an ade-

TABLE 1. Results of in vivo genetic selection

	Clone	Winning sequences <sup>a</sup>		Plants <sup>b</sup>						Total	
		LS	RS	1	2	3	4	5	6		
Round 1	S1-1	CCUACAAAACC	GUGGACCGUGGCgg	2						2	
	S1-2	CCUACAAAACC	GAGCUCGUCGGCgg	1						1	
	S1-3	UCCCCUGUAC	CGACCUACGCGGgg	1						1	
	S1-4	UCCUGAUUUCU	ACUACCGGAGGUgg	1						1	
	S1-5	CCCCAUUAAA	GAUUAACCGGUCgg	2						2	
	S1-6	UGCCCUUCCCC	AGGAUCCUGAUCgg	1						1	
	S1-7	UCCUAAAACCC	UUCUUAGCGGUCgg		1					1	
	S1-8	GACCAAAAGAC	AAGGACACGGCagg		1					1	
	S1-9	UUGCCCAUUC	AGGAAUAACGGCgg		1					1	
	S1-10	UUCUAAAACCG	GUUUCGCACCGUgg		1					1	
	S1-11	UCUAACACUUC	GAAGUCCCGGCagg		1					1	
	S1-12	UCCUAAAACCG	UUCUUAGCGGUCgg		2					2	
	S1-13	CCCACUCCGAC	ACCAAAAACGGCgg			1				1	
	S1-14	UCCCAACCUCU	CUUCCGAACGUGgg			1				1	
	S1-15	ACCAUAAUACA	UAGUCCAACGGGgg			1				1	
	S1-16	UCCUACAUACU	UCUAGAAGCGGCgg			1				1	
	S1-17	UUUCCAAAACA	GCCUGGCAGGUUgg			1				1	
	S1-18	CCUAAGCCUAU	GACUAGGAGGGUgg				4			4	
	S1-19	UCCUGACCCU	CCUAUCCUGGCgg				1			1	
	S1-20	ACCCAUAUAUC	GACAGCAUCGGCgg				1			1	
	S1-21	UCGAAGAAGAC	CGAGAACAAAGUgg					1		1	
	S1-22	AUCUCAAGCAC	UAGUGUACCGGGgg					3		3	
	S1-23	UCCCCGAACCU	CGUUCGGCGGUgg					1		1	
	S1-24	CCCACCCAGAC	CUGUCUCUGCGGgg						5	5	
	S1-25	CCUGCGAGAAC	GUCUGGUGGUUAUgg						1	1	
				Total	8	7	5	6	5	6	37
Round 2	S2-1	CCCACACAAUA	CAUCGAUUCGGCgg	3	2	2	5	2	6	20	
	S2-2	UCCCCAACAU	CUCUAAGCGGUUgg					2		2	
	S2-3	UCCACUCAAAA	CAUCGAUUCGGCgg		1					1	
	wtC	CCCACUCAAAA	ACCAAAAACGGCgg	3	4	5	2	2	1	17	
				Total	6	7	7	7	6	7	40

<sup>a</sup> Sequences are shown in 5'- to -3' orientation. wt satC (wtC) LS and RS sequences are shown in the bottom row and were recovered in the second round. Bases in lowercase were not selected but may be part of the recovered elements (see text). The C-rich LS elements (CCCA and CCUA/G) are in red. The G-rich RS elements (U/CGGCGG, UCGGCAGG, and GC/UGG) are in blue. Additional A-rich elements (CAAAA and ACCAAAA) are in green.

<sup>b</sup> Number of clones found in individual plants at 21 dpi.

nylate-to-guanylate transition at position 176 to create a new restriction site. This parental satRNA, satC<sub>B</sub>, replicates to 82% of wt satC levels in protoplasts and inhibits virion accumulation as efficiently as wt satC (37).

satC<sub>B</sub> transcripts containing randomized LS and RS sequences were coinoculated onto 30 turnip seedlings along with TCYV genomic RNA transcripts. All inoculated plants displayed satC symptoms (dark green, highly crinkled, and stunted leaves) within 2 weeks, and total RNA was extracted from uninoculated leaves at 21 dpi. Agarose gel electrophoresis of the extracted RNA indicated that all plants contained visible satC species as detected by ethidium bromide staining (data

not shown). Full-length satC was recovered by RT-PCR, and 37 clones from six randomly selected plants were sequenced. Twenty-five clones were unique and were designated first-round winners (Table 1) (all sequences in this report are presented in their plus-sense, 5'-to-3' orientation unless otherwise noted). No second-site alterations beyond the selected regions were observed.

The selected LS sequences were enriched for cytidylates (40%) and adenylylates (32%) (Table 1). Only 18 guanylates were present in 275 possible positions (7%). While no first-round winners contained wt LS, the LS of S1-13 differed from wt at only 3 of 11 positions, suggesting that most bases in this

region of wt satC are sequence specific (Table 1). Six of 25 first-round-selected LSs contained the wt sequence CCCA, with clone S1-24 containing two copies of this sequence. Six other winners contained CCUA (repeated twice in S1-18). Four winners had the related sequence CCUG. Based on the LS sequences of the first-round winners, the most common recovered elements can be summarized as CCCA and CCUA/G. Two winners also contained the wt element CAAAA located in the mid- to 3' region of the LS, similar to its location in wt satC. Lastly, two winners contained the wt element ACCA AAA, one in the LS and one in the RS.

Whereas the first-round winning RS sequences contained nucleotide ratios that were more evenly distributed (29% guanylates, 25% cytidylates, 23% adenylates, and 23% uridylates), guanylates were located mainly in the 3' portion of the RS (Table 1). One of the first-round winners (S1-13) contained the wt RS, indicating that the RS is also sequence specific in wt satC. Eight of the selected RSs contained the wt sequence CGGC, and three contained the related sequence UGGC. Eight of these 11 related sequences were located at the 3' edge of the RS, suggesting that additional nonselected 3' bases (e.g., guanylates G221 and G222) may form part of the element. This possibility is supported by the location in clone S1-23 of an internal CGGC sequence upstream of two guanylates. Three of the 11 related sequences were preceded by AGG (for two winners, this included G221 and G222). A second sequence, GC/UGG, was found in 12 RSs from first-round winners (not including occurrences as part of the major U/CGGCCGG recovered sequence). GC/UGG was located mainly near, but not necessarily adjacent to, the RS 3' end. The results of the first-round RS SELEX therefore suggest a requirement for one of three major G-rich elements (U/CGGCCGG, U/CGGC AGG, and GC/UGG). Interestingly, two first-round winners derived from the same plant (S1-1 and S1-2) had identical LSs but very different RSs. Since a well-studied satC recombination hot spot is located just downstream of the LS at the base of M1H (3), RNA recombination may have occurred to generate these satC species. The recovered RSs and LSs of two other winners (S1-7 and S1-12) differed by only a single base. Since these clones were isolated from the same plant, one sequence likely evolved from the other by a single base alteration.

Additional competition between satC species accumulating in first-round plants was carried out by inoculating six seedlings with a mixture of equal amounts of total RNA extracted from all 30 first-round plants. At 21 dpi, total RNA was isolated from each second-round plant and clones were generated by RT-PCR. Only four different satC species were recovered in the 40 clones sequenced, one of which was identical to wt satC. The majority of the recovered clones from each of the six plants contained a new sequence (S2-1) that differed from wt satC by 2 bases in the LS and by 9 of 13 bases in the RS (Table 1). S2-1 contained both of the most prevalent elements identified in the first round, CCCA at the 5' terminus of the LS and CGGCCGG at the 3' terminus of the RS (includes nonselected G221 and G222). wt satC, the next most prevalent species, may have been generated by recombination between S1-13 and S2-1, the former of which contains a wt RS and the latter of which contains a nearly wt LS, followed by sequence evolution. S2-2 and S2-3 were each found in only one of the six plants. S2-2 contained the major CCCA element in the LS and the

TABLE 2. Competition of SELEX winners for fitness in plants

Co-inoculation	No. recovered at 21 dpi
wt satC .....	24
S2-1 .....	0
S2-1 .....	14
S2-2 .....	6
S2-2 .....	20
S2-3 .....	1
S2-3 .....	16
S1-9 .....	2
S1-9 .....	12
S1-17 .....	10

prevalent RS element, GCGG. S2-3 did not contain either of the major LS elements. The presence of identical RS sequences in S2-3 and S2-1 was likely due to RNA recombination. In addition, the adenylate-to-guanylate transition at the base of the M1H that is necessary for construction of the SELEX cDNAs had reverted in S2-3 and in satC with wt LS and RS.

The higher recovery of S2-1 compared to wt satC in second-round plants did not necessarily reflect enhanced fitness of S2-1, since generation of wt satC would first require multiple events. To determine the relative fitnesses of second-round winners and also selected first-round winners, pairwise competitions for fitness were conducted. Two first-round winners were selected for inclusion in this experiment for the following reasons: S1-9 contained the prevalent CGGCCGG (includes G221 and G222) in its RS and CCCA in the LS yet was not recovered in the second round, and the LS of S1-17 did not contain either of the major LS C-rich elements but did contain CAAAA and the prevalent RS element U/CGGCCAGG.

Equal amounts of satC transcripts were combined and inoculated together with TCV onto three turnip seedlings. At 21 dpi, a total of 18 to 24 clones were sequenced from the three plants used for each pairwise competition (Table 2). Nearly equivalent numbers of clones were recovered in coinoculations with first-round winners S1-9 and S1-17 (12 and 10 clones, respectively). S2-3 (the least prevalent second-round winner) was considerably more fit than S1-9 (16 of 18 clones). This suggests that even low-level second-round winners were more fit than some first-round winners. In addition, this result indicates that the simple presence of the wt elements CGGCCGG and CCCA in S1-9 was insufficient to out-compete the second-round winner. S2-2 was substantially more fit than S2-3 and was recovered in 20 of 21 clones. The most prevalent winner in the second round, S2-1, out-competed S2-2 (14 of 20 clones) but was not recovered in direct competition with wt satC. These results indicate the following fitness order: wt satC, S2-1, S2-2, S2-3, S1-9, and S1-17. Since S2-1 and S2-3 contained identical RS sequences, the fitness levels must reflect differences in their LS sequences. The LS of S2-3 differed from wt satC at only one position, strongly suggesting the importance of the wt CCCA for fitness compared with the S2-3 UCCA.

**Winning sequences revealed by in vivo SELEX contribute to satC replication and repression of virion accumulation.** The

fitness of *satC* to accumulate depends on replication competence and the ability to restrict TCV virion accumulation, since high levels of free coat protein are better able to suppress the host's innate RNA silencing defense system. This is advantageous for TCV, since it allows the virus to spread more rapidly within a plant (37, 43; F. Zhang and A. E. Simon, unpublished data). To determine how well second-round winners (S2-1, S2-2, and S2-3) and selected first-round winners (S1-9 and S1-17) replicate and whether they interfere with virion accumulation, protoplasts were inoculated with transcripts prepared from these clones, parental construct *satC<sub>B</sub>*, and *satC<sub>B</sub>* with nonselected ("random") sequences in the LS and RS regions (*satC<sub>B</sub>*-Rd) (Fig. 2A). RNA gel blot analysis of total RNA extracted from protoplasts at 40 hpi showed that *satC<sub>B</sub>*-Rd accumulated to only 2% of *satC<sub>B</sub>* levels (Fig. 2B). First-round winners S1-9 and S1-17 accumulated 13-fold more efficiently than *satC<sub>B</sub>*-Rd, reaching levels that were 26% of *satC<sub>B</sub>* levels. This demonstrates that even after one round of selection, *satC* replication was significantly improved. The most fit second-round winner, S2-1, replicated to levels similar to those of wt *satC* (i.e., ~20% higher than *satC<sub>B</sub>*), and the second most fit winner (S2-2) replicated to levels similar to those of *satC<sub>B</sub>*. The least fit second-round winner accumulated to 73% of *satC<sub>B</sub>* levels. These results indicate that recovered sequences contribute to fitness of the winners by effectively increasing the ability of the *satRNA* to replicate. In addition, since there was a substantial difference in fitness between S2-1, S2-2, and wt *satC* based on direct competition assays in plants (Table 2), distinctions between these three *satC* species likely involve parameters other than simple replication efficiency.

To ascertain whether the first- and second-round winners interfered with virion accumulation, virions were extracted from infected protoplasts and analyzed by Western blotting with antibody against TCV CP (Fig. 2C). Compared with virion levels present during infection with TCV alone, *satC<sub>B</sub>*-Rd reduced the accumulation of virions by 58%, while *satC<sub>B</sub>* reduced the levels by 92%. This suggests that sequences outside the LS-RS region in *satC<sub>B</sub>*-Rd also contribute to reducing the levels of virions. All SELEX winners examined interfered with virion accumulation more effectively than *satC<sub>B</sub>*-Rd, ranging from an 85% reduction for the most fit second-round winner, S2-1, to undetectable levels for S1-9. Although the winners accumulated to different levels, which could affect the extent of virion interference both directly and by differentially influencing the level of TCV genomic RNA accumulation (and hence the level of CP), we previously determined that the *satRNA* that was most efficient at reducing virion levels had a deletion of the M1H enhancer and accumulated to only 3% of the wt level (37). These findings therefore suggest that while recovered sequences in the LS and/or RS regions contribute to fitness of *satRNAs* by inhibition of virion accumulation, other parameters are more important in determining *satC* fitness.

**CCCA and CGGCGG enhance replication of *satC*.** To determine whether the LS CCCA and RS CGGCGG elements in wt *satC* and SELEX winners contribute to replication and/or virion repression, the elements were introduced individually into the LS and RS of *satC<sub>B</sub>*-Rd by replacing corresponding random bases as follows: CCCA was placed at the 5' edge of the *satC<sub>B</sub>*-Rd LS next to U163, generating construct LS1, and

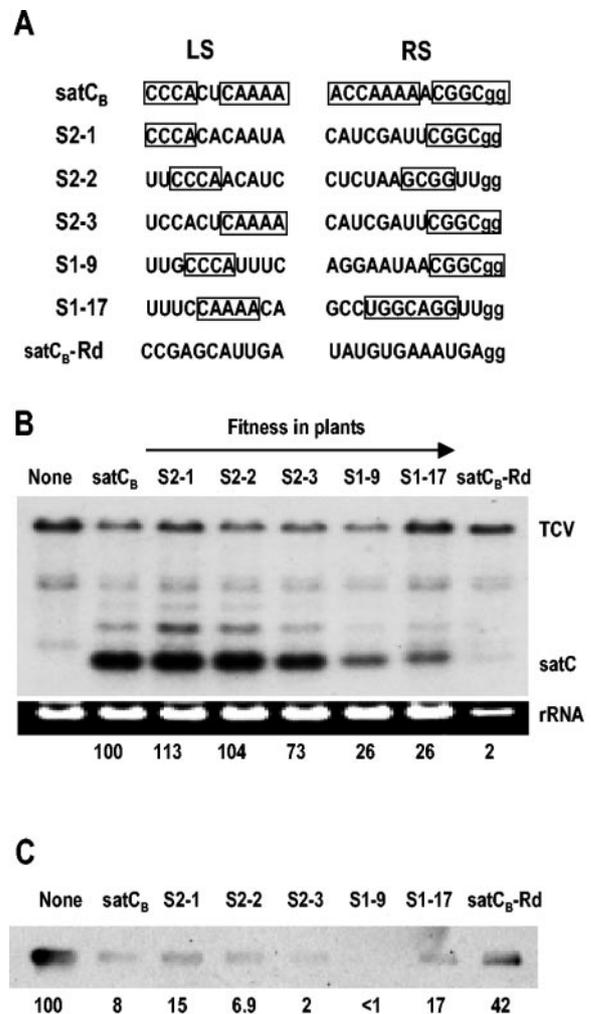


FIG. 2. Replication and virion repression of second-round and selected first-round SELEX winners. (A) Sequences in the LS and RS regions in the parental *satRNA* (*satC<sub>B</sub>*), second-round winners (S2-1, S2-2, and S2-3), first-round winners (S1-9 and S1-17) and *satC<sub>B</sub>* containing randomly selected LS and RS sequences (*satC<sub>B</sub>*-Rd). Elements identified in the text and in Table 1 are boxed. Bases in lowercase were outside the selected sequences but likely comprise part of some identified elements. (B) RNA gel blot of total RNA extracted 40 h after inoculation of protoplasts with TCV and the *satRNAs* listed above each lane. The blot was probed with an oligonucleotide specific for both TCV and *satC*. Positions of TCV and *satC* are shown. 26S rRNA from the ethidium-stained gel is shown and was used as a loading control. Values below each lane are normalized levels of *satC*. The order of fitness of the *satRNAs* (indicated by an arrow above the lanes) is from Table 2. (C) Virion levels in infected protoplasts. Virions were detected by chemiluminescence following treatment with anti-TCV CP antibody. Values given below each lane represent relative levels of virions.

CGGC was placed at the 3' edge of the RS, upstream of G221 and G222, generating construct RS1 (Fig. 3A).

Transcripts of individual constructs were inoculated onto protoplasts with TCV genomic RNA. *satC<sub>B</sub>*-Rd plus strands averaged 6% of *satC<sub>B</sub>* levels in two independent experiments, while accumulation of minus-strand *satC<sub>B</sub>*-Rd was much higher, reaching 27% of *satC<sub>B</sub>* levels. Plus strands of LS1 and RS1 accumulated to 3.6- and 2.8-fold-higher levels than *satC<sub>B</sub>*-Rd plus strands, respectively, reaching levels that were 22 and 18%

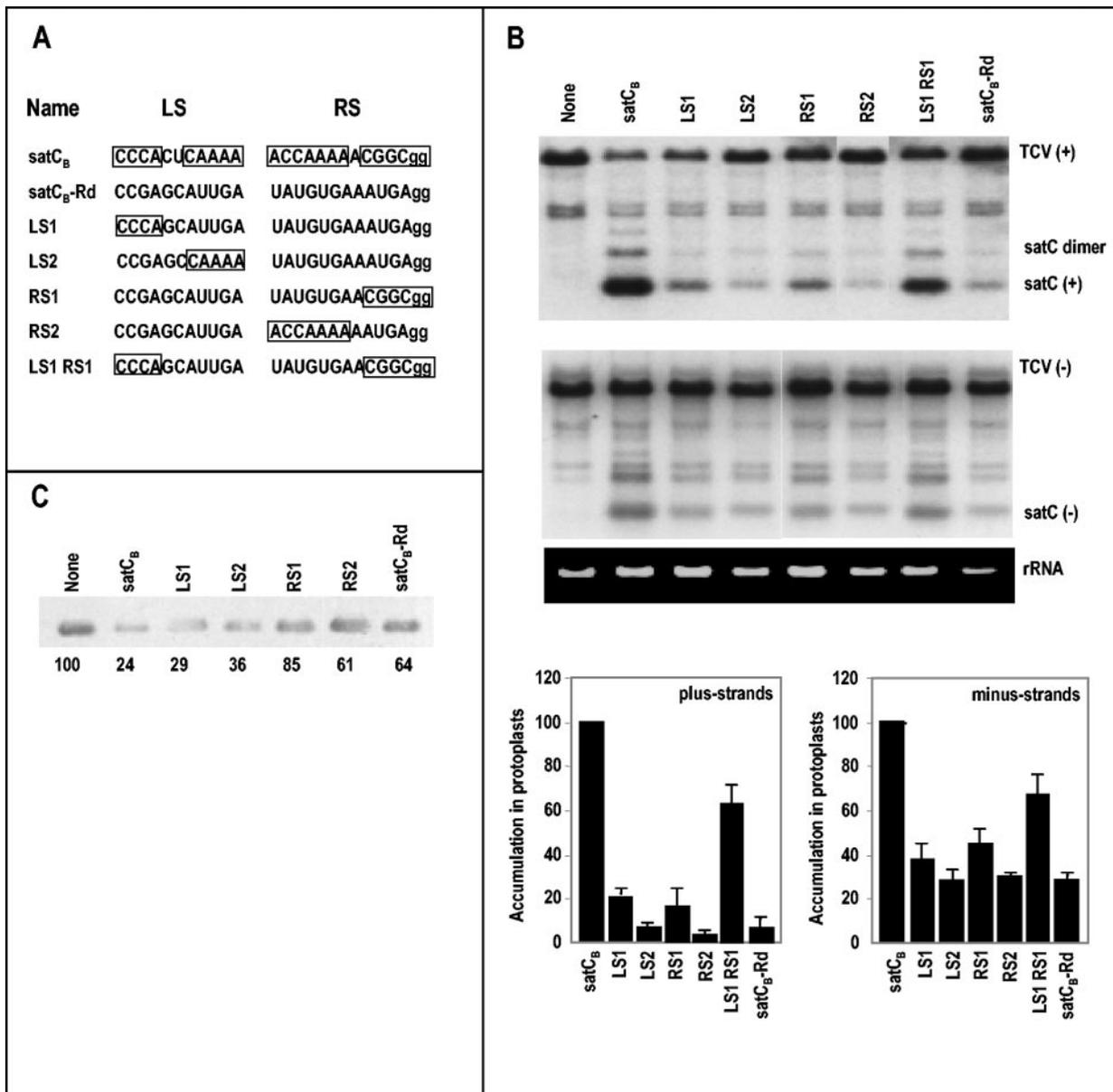


FIG. 3. Effect on replication and virion repression of individual elements identified by *in vivo* SELEX. (A) Elements (boxed) were inserted into the LS or RS region of satC<sub>B</sub>-Rd, replacing corresponding randomly selected LS and/or RS sequences. Names of the constructs are given. satC<sub>B</sub>, parental construct containing wt satC LS and RS sequences. Bases in lowercase were outside the selected sequences but likely comprise part of the identified elements. (B) RNA gel blot of accumulating TCV and satC plus strands (upper panel) and minus strands (middle panel) at 40 hpi of protoplasts. satC constructs included in the inoculum are given above each lane. The blot was probed with an oligonucleotide specific for both TCV and satC. Positions of TCV and satC are shown. The strong band migrating slightly faster than TCV minus strands is found in preparations from uninfected protoplasts and comigrates with 26S rRNA (data not shown). 26S rRNA from the ethidium-stained gel is shown and was used as a loading control. Normalized levels of accumulating satRNAs in replicate experiments are presented below the gel blots. (C) Virion levels in infected protoplasts at 40 hpi. Virions were detected by chemiluminescence following treatment with anti-TCV CP antibody. Values given below each lane represent relative levels of virions. Error bars indicate standard deviations.

of that for satC<sub>B</sub> (Fig. 3B). Minus strands of LS1 and RS1 were only slightly affected by the addition of the elements, accumulating to levels only 1.4- or 1.6-fold greater than that for satC<sub>B</sub>-Rd. Interestingly, mutations that disrupt H5 structure also have a greater influence on plus-strand levels than on minus-strand levels (J. Zhang and A. E. Simon, submitted for publication).

A new construct combining CCCA and CGGCGG (construct LS1RS1) was generated to determine whether together

these sequences have an additive or a synergistic effect on satRNA replication. At 40 hpi of protoplasts, LS1RS1 accumulated to 63% of satC<sub>B</sub> plus-strand levels and to 67% of satC<sub>B</sub> minus-strand levels (Fig. 3B). This synergistic effect on plus-strand accumulation suggests participation of the sequences in a related replicative function. However, these results also suggest that additional factors in the region are required to achieve wt levels of replication.

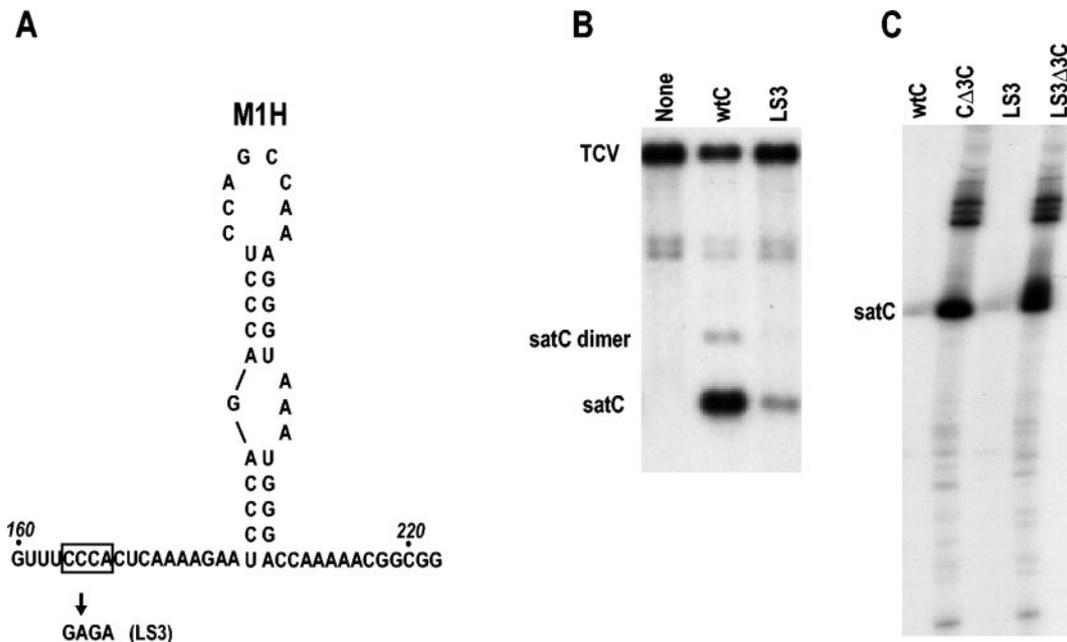


FIG. 4. Effect of mutations in the LS CCCA on accumulation in vivo and transcription in vitro. (A) Alterations in the CCCA sequence in construct LS3. (B) Effect of mutations on accumulation of satC in protoplasts at 40 hpi. The Northern blot was probed with an oligonucleotide complementary to both TCV and satC sequence. (C) Effect of mutations on transcription in vitro with TCV p88 purified from *E. coli*. Templates that also contain a deletion of the 3'-terminal three cytidylates ( $\Delta 3C$ ) are indicated.

The roles of CAAAA and ACCAAAA elements in replication were also assayed by replacing corresponding random bases in satC<sub>B</sub>-Rd with these sequences in their wt positions (generating constructs LS2 and RS2, respectively [Fig. 3A]). In contrast to the case for LS1 and RS1, plus and minus strands of LS2 and RS2 did not accumulate to levels greater than those for satC<sub>B</sub>-Rd (Fig. 3B). This result was surprising since we had previously shown the best winner of a previous in vivo SELEX of M1H had a duplication of CAAAA in the RS region that substantially contributed to replication and virion repression of the winner (37). Combined with our present results, this suggests that CAAAA cannot enhance the replication of satC independent of other elements in the LS and/or RS but may enhance replication when combined with other sequences not present in satC<sub>B</sub>-Rd.

Transcripts containing mutations in the CGGCGG element were very poor templates for in vitro transcription assays programmed with purified TCV p88 prepared in *E. coli* (45). To determine the effect on in vitro transcription of satC containing mutations in the CCCA element, satC that altered the CCCA element to GAGA was prepared (construct LS3 [Fig. 4A]). As shown in Fig. 4B, accumulation of LS3 in protoplasts was substantially reduced compared with that of wt satC, supporting an important replicative role for the CCCA element. However, LS3 and wt satC had equivalent template activities in vitro (Fig. 4C). Deletion of the 3'-terminal three cytidylates (LS3 $\Delta 3C$ ) relieved transcriptional repression similarly to wt satC with the equivalent deletion (C $\Delta 3C$ ). Since the products of the TCV RdRp-mediated in vitro transcription reactions are double stranded (36), while minus strands synthesized in vivo are likely single stranded, the in vitro assay examines the initiation event for only a single round of transcription. Thus, it is possible that the CCCA element functions downstream of ini-

tiation of minus-strand synthesis in an event of replication that cannot be evaluated by using this in vitro assay.

**CCCA and CAAAA affect virion accumulation.** To determine whether the most prevalent recovered elements affect virion accumulation, virions were isolated from protoplasts infected with TCV alone or accompanied by satC<sub>B</sub>, LS1, LS2, RS1, or RS2 and analyzed by Western blotting with antibody specific to TCV CP (Fig. 3C). In this experiment, satC<sub>B</sub> reduced virion levels to 24% of those of TCV alone, while virion levels were 64% in the presence of satC<sub>B</sub>-Rd. RS1 and RS2 did not further reduce virion levels compared with satC<sub>B</sub>-Rd, indicating that ACCAAAA and CGGCGG are not independently involved in virion repression. However, LS1 and LS2 reduced virion levels nearly as effectively as satC<sub>B</sub>. These results confirm our previous finding that duplication of CAAAA in the region enhanced virion repression (37). In addition, these results indicate that CCCA is multifunctional, enhancing satRNA replication and the satRNA's ability to repress virions.

How the CCCA and CAAAA in the LS region interfere with virion accumulation is not known. It is also not known why ACCAAAA in the RS region, which contains an embedded CAAAA element, had no effect on virion repression. Since satC<sub>B</sub>-Rd was able to repress virion accumulation by 36% in the absence of selected LS and RS sequences, it seems likely that sequences external to these regions influence virion assembly and thus that the position of the CAAAA element within the LS may be important.

**Structural probing of the CGGCGG and CCCA elements.** We previously observed that repression of minus-strand synthesis in vitro was mediated by the interaction between 3'-terminal bases and the 3' side of the H5 LSL and could be alleviated by deletion of the 3'-terminal three cytidylates (45). Mutations that converted the RS CGGCGG element to CC

GGGG strongly suppressed transcription *in vitro*, suggesting that this sequence functioned as a “derepressor” involved in freeing the 3′ end from the H5 LSL to promote initiation of minus-strand synthesis (45). We proposed that releasing the 3′ end from its interaction with the H5 LSL requires alternative base pairing between the 3′ end (CCUGCCC-OH) and the derepressor (CGGCGG) (45). This alternative base pairing between the 3′ end and the derepressor is present in the lowest-free-energy structure as predicted by Mfold version 3.1 (47). If this alternative base pairing is correct, then recovered RS that did not contain the wt derepressor should contain an element that still permits base pairing with the 3′ end. Eleven of the recovered RSs in first- and second-round winners contained the wt or near-wt element U/CGGCGG, and 16 winners contained the element GC/UGG or U/CGGCAGG. The latter two sequences can form four or five base pairs, respectively, with 3′-terminal sequences (CCUGCCC-OH), supporting the existence of this interaction. Of the remaining winners (Table 1), two clones (S1-15 and S1-22) contained the similarly positioned sequence CGGG, which is capable of a 4-base interaction with 3′ sequences (underlined, CCUGCCC-OH), and one clone (S1-5) contained the sequence GGUC, which could potentially form three base pairs.

To obtain further support for an interaction between the CGGCGG and 3′-terminal bases when the 3′ end is not interacting with the H5 LSL, we analyzed the structure of the CGGCGG element in the absence of the latter interaction by using construct H5RL, which contains a 3-base alteration in the H5 LSL (Fig. 5A). When H5RL was used as a template in the *in vitro* transcription assay programmed with TCV RdRp, products were essentially identical to those obtained with CΔ3C, indicating a similar disruption of the 3′ end-LSL interaction in H5RL (45).

Previous use of reverse transcriptase-mediated extension of primers to assay residues susceptible to single-strand-specific and double-strand-specific reagents revealed that the CGGCGG element was single stranded in wt satC (45). For the present study, the structures of wt satC and H5RL were probed by using 3′-end-labeled transcripts and partial digestion with RNase T<sub>1</sub> (specific for single-stranded guanylates), RNase A (specific for single-stranded pyrimidines) and RNase V<sub>1</sub> (cleaves at double-stranded and stacked residues). wt satC and H5RL differed substantially in the upper portion and 3′ side of H5 between positions downstream of U280 (Fig. 5C) and upstream of G315 (Fig. 5B). This result supports our previous finding that the phylogenetically conserved H5 structure (i.e., the structure presented in Fig. 5A) is not present in wt satC transcripts (45), since the H5RL mutations should not affect that structure to such a significant extent.

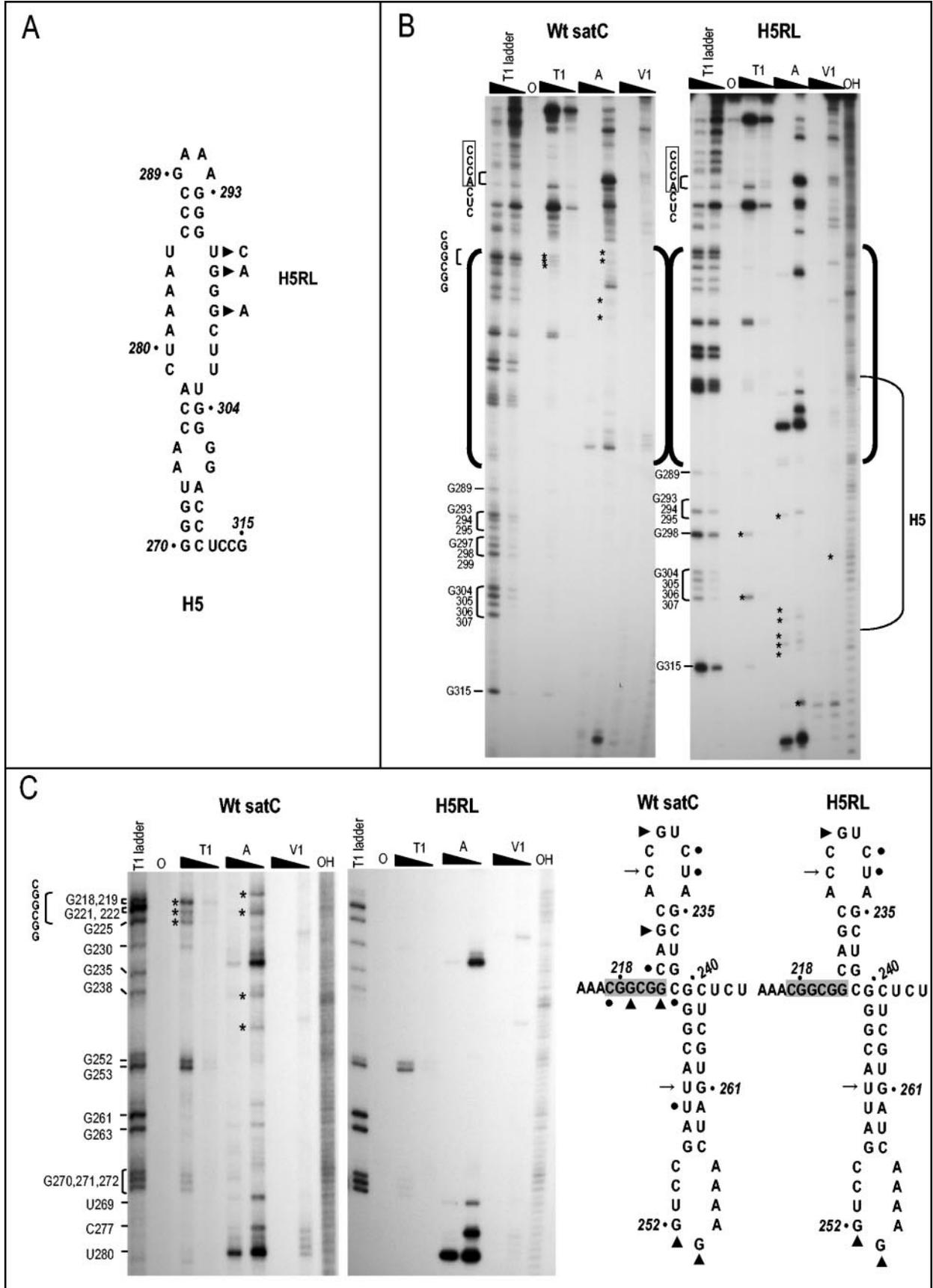
Enzyme susceptibility was very similar between wt satC and H5RL upstream of position U280, with several significant and reproducible exceptions (Fig. 5C). The CGGCGG element (C217 through G222) was single stranded in wt satC, as previously shown (45). However, none of these residues were susceptible to single-strand-specific enzymes in H5RL, indicating that these residues are likely involved in a base-paired configuration in the satC mutant. This result is consistent with our model that the 3′ end, when freed from the H5 LSL, is available for pairing with the CGGCGG element. Additional differences included C239 at the 3′ base of H4a and single residues in H4a (G225) and H4b (U246); all were single stranded only in wt satC (Fig. 5C). At present, we do not have an explanation for these differences.

We previously determined that transcripts of CΔ3C had a substantially altered structural rearrangement of H5 and 3′ flanking sequences, with reactive residues now supporting the phylogenetically conserved structure of H5 (45). The only other discernible differences between wt satC and CΔ3C within the region assayed (positions 170 to 330) were at positions 254 and 255 in the loop of H4b, which had slightly enhanced susceptibility to DMS in CΔ3C. The structural rearrangement in the H5 region was proposed to reflect a possible conformational switch between repressible and transcriptionally active templates mediated by the release of the 3′ end from the H5 LSL (45). Interestingly, the residues susceptible to single-strand- and double-strand-specific reagents in H5 were very different between H5RL and CΔ3C, indicating that simple release of the LSL-3′ end interaction can lead to different structures in the H5 region depending on whether the LSL contains alterations.

As described above, 20 of 29 first- and second-round winners contained CCCA or CUA/G in the 5′ region of their selected LS sequence. The nature of these recovered bases suggested that this element might contribute to satC replication through a base-pairing interaction with a sequence such as 5′ UGGG, which would account for differences between the recovered sequences by allowing G · U pairings. Examination of the nine winning LS sequences that did not contain CCCA or CUA/G revealed that seven contained a sequence in a similar 5′ side location that could also base pair with 5′ UGGG (UCUA, UCUA, UCCA ([two sequences]), UUCG, CUCA, and CCCG).

Full-length satC contains five UGGG sequences: one is at position 106, one is at the base of M1H, and three are within H5. Of the elements within H5, two are in stem regions in the phylogenetically conserved structure, and one is in the LSL (Fig. 1B). The LSL UGGG partially overlaps the sequence known to interact with the 3′-terminal GCCC-OH (GGGC) (overlapped bases are underlined) (45). Since the combined

FIG. 5. Structure probing of wt satC and H5RL. (A) Mutations constructed in the H5 LSL to generate H5RL. Numbering is from the 5′ end. The phylogenetically conserved structure of H5 is presented. (B) Solution structure probing of wt satC and H5RL with 3′-end terminally labeled transcripts. Identities of residues in the RNase T<sub>1</sub> ladder are given. Positions of the LS CCCA element (and downstream pyrimidines) and the RS CGGCGG element are shown. The region encompassing H5 is indicated by a single bracket. The region encompassed by the thick double bracket is the region shown in more detail in panel C. Triangles above adjacent lanes indicate that the transcripts were digested with high and low concentrations of enzymes. T1, RNase T<sub>1</sub>; A, RNase A; V1, RNase V<sub>1</sub>; O, no enzyme digestion. Asterisks indicate residues susceptible only in wt satC or H5RL. (C) Left panel, extended electrophoresis of samples shown in panel B. Locations of the CGGCGG element and several prominent RNase A-cleaved residues are given. Asterisks indicate residues susceptible only in wt satC or H5RL. Right panel, locations of reactive residues in the structures of wt satC and H5RL. ●, RNase A; [rtrif], RNase T<sub>1</sub>; →, RNase V<sub>1</sub>. The RS CGGCGG element is shaded.



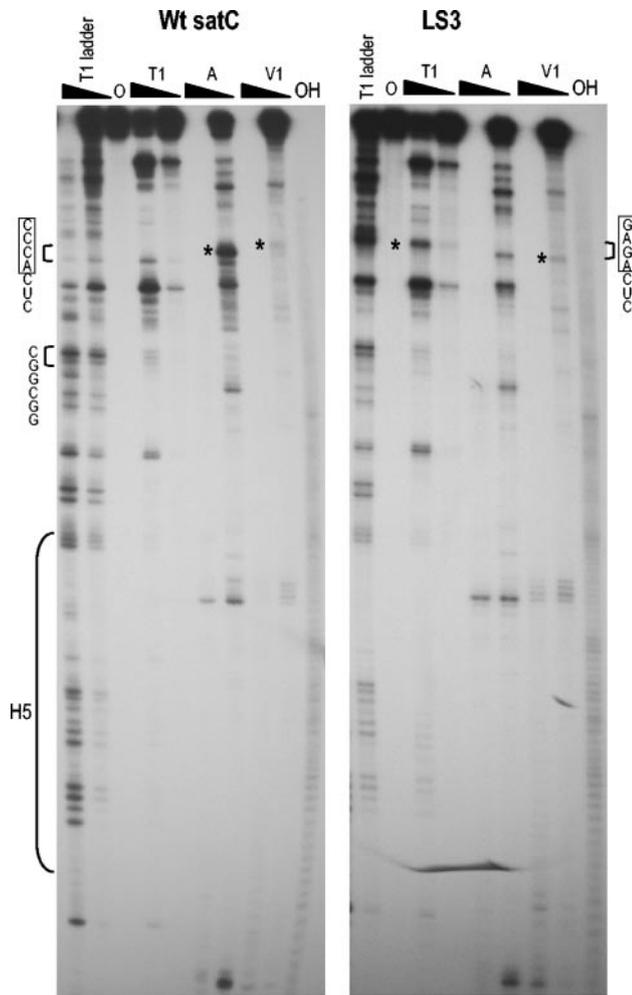


FIG. 6. Structure probing of wt satC and LS3. The LS3 mutant is described in Fig. 4A. Solution structure probing was conducted with 3'-end terminally labeled transcripts. Positions of the LS CCCA element (and downstream pyrimidines) and RS CGGCGG element are shown. The region encompassing H5 is indicated by a single bracket. Triangles above adjacent lanes indicate that the transcripts were digested with high and low concentrations of enzymes. T1, RNase T<sub>1</sub>; A, RNase A; V1, RNase V<sub>1</sub>; O, no enzyme digestion. Asterisks indicate residues susceptible only in wt satC or LS3.

presence of CCCA and CGGCGG was synergistic (Fig. 3), one possible model is that CCCA interacts with the H5 LSL in the absence of the LSL-3' end interaction.

To aid in determining a function for the CCCA element, we assessed its structure in wt satC. Since the CCCA element is located approximately 200 bases from the 3' end of satC, it was difficult to precisely place the element in gels that probed the structure of transcripts labeled at the 3' end. A strong RNase A signal appeared to correspond to pyrimidines just downstream from the CCCA element (CCCACUC), while at least one of the cytidylates in the element appeared to correspond with an RNase V<sub>1</sub> signal (Fig. 5B). To confirm this tentative assignment and to explore structural changes that might occur in the absence of the element, we compared the susceptibilities of residues in the region between wt satC and LS3, which contained the CCC→GAG alteration shown in Fig. 4A. Struc-

tural probing revealed that all residues between approximately position 110 and the 3' end of H5 were identical in LS3 and wt satC, with the exception of residues in the region tentatively assigned to the CCCA element and downstream pyrimidines (Fig. 6). In LS3, the strong RNase A signal and upstream adjacent RNase V<sub>1</sub> signal were absent, while new RNase T<sub>1</sub> and adjacent downstream RNase V<sub>1</sub> sites were present. These results indicate that the CCCA element is at least partially double stranded in wt satC and that alteration of the element does not result in gross structural changes in H5 or the remainder of the region probed.

Although mutations in the CCCA element did not detectably affect the H5 region, it was still possible that an interaction between these two regions might be evident in CΔ3C, since the H5 region in CΔ3C is structurally rearranged compared with that in wt satC (45). To test for this possibility, wt satC and CΔ3C were subjected to solution structural probing in the region encompassing the LS and RS sequences. Because of the distance from the 3' end, we performed the structural probing by using primer extension to assay for reactive bases. In both templates, the adenylate at position 167 in the LS CCCA element (CCCA) and the cytidylate at position 168 just downstream from the element (CCCAC) were susceptible to DMS, which methylates the N1 and N3 positions of unpaired adenosines and cytidines, respectively (Fig. 7). C168 most likely

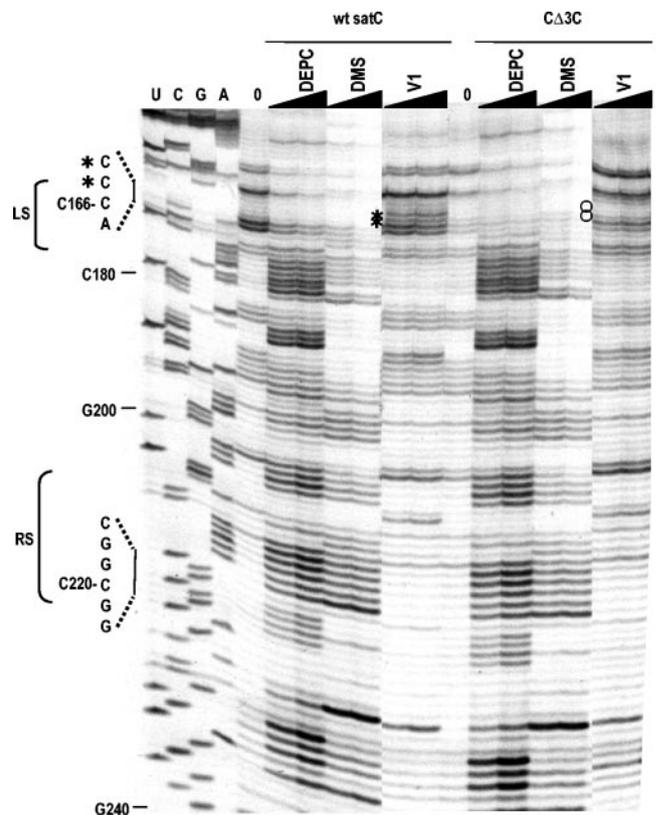


FIG. 7. Solution structure probing of the region encompassing the LS and RS regions of satC and CΔ3C, using extension of primers by reverse transcriptase. The locations of the CCCA and CGGCGG elements are shown. Asterisks indicate the locations of two RNase V<sub>1</sub>-sensitive bases in the CCCA sequence in wt satC. These bases are not sensitive to RNase V<sub>1</sub> in CΔ3C (circles).

corresponds to part of the strong RNase A signal in the prior structural probing (Fig. 6, left panel). The cytidylate at position 166 (CCCA) had a strong stop in the treatment-free lane and was therefore of unknown conformation. The cytidylates at positions 164 and 165 (CCCA) were susceptible to RNase V<sub>1</sub> cleavage in wt satC, indicating that these bases were most likely the RNase V<sub>1</sub>-reactive residues in the region of this sequence in Fig. 6. When CΔ3C was subjected to identical treatment, C164 and C165 lost their sensitivity to RNase V<sub>1</sub> treatment, indicating that a structural alteration that involved at least these bases had occurred. No further differences between the two templates were discernible upstream of this region until the 5' end, with the exception of an adenylate at position 4 (data not shown). These results therefore support a connection between the LS CCCA and the structural rearrangement that occurs in the H5 region of CΔ3C. However, the role of the CCCA element in replication remains elusive.

**Conclusions.** Our previous evidence that the internal CG GCGG element functions in replication was based on a loss of transcriptional activity in an in vitro RdRp assay and the ability of the element to base pair with 3'-terminal bases in Mfold computer structure predictions (45). These findings led to a model in which the 3' end, initially base paired with the H5 LSL and unavailable for initiation of complementary strand synthesis, requires the CGGCGG element to help facilitate access of the RdRp to the 3' end for de novo initiation. Our present results support this model by finding that bases recovered from the SELEX analysis maintain the possible pairing with the 3' end and that disruption of the LSL-3' end interaction leads to a structural alteration in the CGGCGG element, with residues no longer reactive to single-strand-specific enzymes.

This requirement for an additional element to allow access of RdRp to the 3' end is similar to a report of factors necessary for initiation of bacteriophage Qβ minus-strand synthesis (31). The 3'-terminal five bases of Qβ are also involved in a long-distance base pairing that does not permit efficient access to the polymerase. A bacterial protein (Hfq) or a series of mutations including alterations to the 3' end and interacting sequence are required for efficient replication and are thought to destabilize the secondary structure in the region and allow access of the polymerase to the 3' end. Interestingly, a number of structural and sequence similarities exist between Qβ and TCV genomic RNAs, including three guanylates at the 5' terminus, similar 3'-terminal sequences (TCV, CCUGCCC-OH; Qβ, CCUCCC), and a stable hairpin located just upstream from the 3' sequence that contains stacked pyrimidines on one side and stacked purines (mainly guanylates) on the other side (31, 45). The presence of alternative structural conformations of some viral 3' ends may be one mechanism to limit minus-strand synthesis and permit greater synthesis of progeny plus strands.

#### ACKNOWLEDGMENTS

Funding was provided by grants from the Public Health Service (GM61515-01) and the National Science Foundation (MCB-0086952) to A.E.S.

#### REFERENCES

- Barton, D. J., B. J. Morasco, and J. B. Fanegan. 1999. Translating ribosomes inhibit poliovirus negative-strand RNA synthesis. *J. Virol.* **73**:10104–10112.
- Buck, K. W. 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv. Virus Res.* **47**:159–251.
- Cascone, P. J., T. Haydar, and A. E. Simon. 1993. Sequences and structures required for RNA recombination between virus-associated RNAs. *Science* **260**:801–805.
- Dreher, T. W. 1999. Functions of the 3'-untranslated regions of positive strand RNA viral genomes. *Annu. Rev. Phytopathol.* **37**:151–174.
- Frolov, I., R. Hardy, and C. M. Rice. 2001. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA* **7**:1638–1651.
- Goebel, S. J., B. Hsue, T. F. Dombrowski, and P. S. Masters. 2004. Characterization of the RNA components of a putative molecular switch in the 3'-untranslated region of the murine coronavirus genome. *J. Virol.* **78**:669–682.
- Guan, H., C. D. Carpenter, and A. E. Simon. 2000. Analysis of cis-acting sequences involved in plus-strand synthesis of a TCV-associated satellite RNA identifies a new carmovirus replication element. *Virology* **268**:345–354.
- Guan, H., C. D. Carpenter, and A. E. Simon. 2000. Requirement of a 5'-proximal linear sequence on minus strands for plus-strand synthesis of a satellite RNA associated with TCV. *Virology* **268**:355–363.
- Hacker, D. L., I. T. D. Petty, N. Wei, T. J. Morris. 1992. Turnip crinkle virus genes required for RNA replication and virus movement. *Virology* **186**:1–8.
- Herold, J., and R. Andino. 2001. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol. Cell* **7**:581–591.
- Isken, O., C. W. Grassmann, R. T. Sarisky, M. Kann, S. Zhang, F. Grosse, P. N. Kao, and S.-E. Behrens. 2003. Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus. *EMBO J.* **22**:5655–5665.
- Khromykh, A. A., H. Meka, K. J. Guyatt, and E. G. Westaway. 2001. Essential role of cyclization sequences in flavivirus RNA replication. *J. Virol.* **75**:6719–6728.
- Koev, G., S. Liu, R. Beckett, and W. A. Miller. 2002. The 3'-terminal structure required for replication of barley yellow dwarf virus RNA contains an embedded 3' end. *Virology* **292**:114–126.
- Kong, Q., J.-W. Oh, C. D. Carpenter, and A. E. Simon. 1997. The coat protein of Turnip crinkle virus is involved in subviral RNA-mediated symptom modulation and accumulation. *Virology* **238**:478–485.
- Kong, Q., J.-W. Oh, and A. E. Simon. 1995. Symptom attenuation by a normally virulent satellite RNA of Turnip crinkle virus is associated with the coat protein open reading frame. *Plant Cell* **7**:1625–1634.
- Lai, M. M. C. 1998. Cellular factors in the transcription and replication of viral RNA genomes: a parallel to DNA-dependent RNA transcription. *Virology* **244**:1–12.
- Li, W.-Z., F. Qu, and T. J. Morris. 1998. Cell-to-cell movement of Turnip crinkle virus is controlled by two small open reading frames that function in trans. *Virology* **244**:405–416.
- Li, X. H., L. A. Heaton, T. J. Morris, and A. E. Simon. 1989. Turnip crinkle virus defective interfering RNAs intensify viral symptoms and are generated de novo. *Proc. Natl. Acad. Sci. USA* **86**:9173–9177.
- McCormack, J. C., and A. E. Simon. 2004. Biased hypermutagenesis associated with mutations in an untranslated hairpin of an RNA virus. *J. Virol.* **78**:7813–7817.
- Nagy, P. D., C. Zhang, and A. E. Simon. 1998. Dissecting RNA recombination in vitro: role of RNA sequences and the viral replicase. *EMBO J.* **17**:2392–2403.
- Nagy, P. D., J. Pogany, and A. E. Simon. 1999. RNA elements required for RNA recombination function as replication enhancers in vitro and in vivo in a plus strand RNA virus. *EMBO J.* **18**:5653–5665.
- Nagy, P. D., J. Pogany, and A. E. Simon. 2001. In vivo and in vitro characterization of an RNA replication enhancer in a satellite RNA associated with Turnip crinkle virus. *Virology* **288**:315–324.
- Olsthoorn, R. C., S. Mertens, F. T. Brederode, and J. F. Bol. 1999. A conformational switch at the 3' end of a plant virus RNA regulates viral replication. *EMBO J.* **18**:4856–4864.
- Plasterek, R. H. 2002. RNA silencing: the genome's immune system. *Science* **296**:1263–1265.
- Pogany, J., M. R. Fabian, K. A. White, and P. D. Nagy. 2003. A replication silencer element in a plus-strand RNA virus. *EMBO J.* **22**:5602–5611.
- Qu, F., T. Ren, and T. J. Morris. 2003. The coat protein of Turnip crinkle virus suppresses posttranscriptional gene silencing at an early initiation step. *J. Virol.* **77**:511–522.
- Rajendran, K. S., J. Pogany, and P. D. Nagy. 2002. Comparison of Turnip crinkle virus RNA-dependent polymerase preparations expressed in *Escherichia coli* or derived from infected plants. *J. Virol.* **76**:1707–1717.
- Ranjith-Kumar, C. T., X. Zhang, and C. C. Kao. 2003. Enhancer-like activity of a bromo mosaic virus RNA promoter. *J. Virol.* **77**:1830–1839.
- Ray, D., and K. A. White. 1999. Enhancer-like properties of an RNA element that modulates tombusvirus RNA accumulation. *Virology* **256**:162–171.
- Ray, D., and K. A. White. 2003. An internally located RNA hairpin enhances replication of Tomato bushy stunt virus RNAs. *J. Virol.* **77**:245–257.
- Schuppli, D., J. Georgijevic, and H. Weber. 2000. Synergism of mutations in bacteriophage Qβ RNA affecting host factor dependence of Qβ replicase. *J. Mol. Biol.* **295**:149–154.

32. **Simon, A. E.** 1999. Replication, recombination and symptom-modulation properties of the satellite RNAs of Turnip crinkle virus. *Curr. Top. Microbiol. Immunol.* **239**:19–36.
33. **Simon, A. E., M. J. Roossinck, and Z. Havelda.** 2004. Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu. Rev. Phytopathol.* **42**:415–447.
34. **Simon, A. E., and S. H. Howell.** 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.
35. **Song, C., and A. E. Simon.** 1995. Requirement of a 3'-terminal stem loop in in vitro transcription by an RNA-dependent RNA polymerase. *J. Mol. Biol.* **254**:6–14.
36. **Song, C., and A. E. Simon.** 1995. Synthesis of novel products in vitro by an RNA-dependent RNA polymerase. *J. Virol.* **69**:4020–4028.
37. **Sun, X., and A. E. Simon.** 2003. Fitness of a *Turnip crinkle virus* satellite RNA correlates with a sequence-nonspecific hairpin and flanking sequences that enhance replication and repress the accumulation of virions. *J. Virol.* **77**:7880–7889.
38. **Thomas, C. L., V. Leh, C. Lederer, and A. J. Maule.** 2003. Turnip crinkle virus coat protein mediates suppression of RNA silencing in *Nicotiana benthamiana*. *Virology* **306**:33–41.
39. **van Rossum, C. M. A., C. B. E. M. Reusken, F. T. Brederode, and J. F. Bol.** 1997. The 3' untranslated region of alfalfa mosaic virus RNA3 contains a core promoter for minus-strand synthesis and an enhancer element. *J. Gen. Virol.* **78**:3045–3049.
40. **Voinnet, O., Y. M. Pinto, and D. C. Baulcombe.** 1999. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proc. Natl. Acad. Sci. USA* **96**:14147–14152.
41. **Wang, J., and A. E. Simon.** 1999. Symptom attenuation by a satellite RNA in vivo is dependent on reduced levels of virus coat protein. *Virology* **259**:234–245.
42. **You, S., B. Falgout, L. Markoff, and R. Padmanabhan.** 2001. In vitro RNA synthesis from exogenous dengue viral RNA templates requires long-range interactions between 5'- and 3'-terminal regions that influence RNA structure. *J. Biol. Chem.* **276**:15581–15591.
43. **Zhang, F., and A. E. Simon.** 2003. Enhanced viral pathogenesis associated with a virulent mutant virus or a virulent satellite RNA correlates with reduced virion accumulation and abundance of free coat protein. *Virology* **312**:8–13.
44. **Zhang, G., and A. E. Simon.** 2003. A multifunctional Turnip crinkle virus replication enhancer revealed by in vivo functional SELEX. *J. Mol. Biol.* **326**:35–48.
45. **Zhang, G., J. Zhang, and A. E. Simon.** 2004. Repression and derepression of minus-strand synthesis in a plus-strand RNA virus replicon. *J. Virol.* **78**:7619–7633.
46. **Zhang, J., R. M. Stuntz, and A. E. Simon.** 2004. Analysis of a viral replication repressor: sequence requirements for a large symmetrical interior loop. *Virology* **326**:90–102.
47. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.