7-2018

SELEX and SHAPE Reveal that Sequence Motifs and an Extended Hairpin in the 5' Portion of Turnip Crinkle Virus Satellite RNA C Mediate Fitness in Plants

Charlie F. Bayne
Dickinson College

Max E. Widawski
Dickinson College

Feng Gao

Mohammed H. Masab
Dickinson College

Maitreyi Chattopadhyay

See next page for additional authors

Follow this and additional works at: https://scholar.dickinson.edu/faculty_publications

Part of the Plant Sciences Commons

Recommended Citation
Bayne, Charlie F.; Widawski, Max E.; Gao, Feng; Masab, Mohammed H.; Chattopadhyay, Maitreyi; Murawski, Allison M.; Sansevere, Robert M.; Lerner, Bryan D.; Castillo, Rinaldy J.; Griesman, Trevor; Fu, Jiantao; Hibben, Jennifer K.; García-Perez, Alma D.; Simon, Anne E.; and Kushner, David B., "SELEX and SHAPE Reveal that Sequence Motifs and an Extended Hairpin in the 5' Portion of Turnip Crinkle Virus Satellite RNA C Mediate Fitness in Plants" (2018). Dickinson College Faculty Publications. Paper 939.
https://scholar.dickinson.edu/faculty_publications/939

This article is brought to you for free and open access by Dickinson Scholar. It has been accepted for inclusion by an authorized administrator. For more information, please contact scholar@dickinson.edu.
SELEX and SHAPE reveal that sequence motifs and an extended hairpin in the 5’ portion of Turnip crinkle virus satellite RNA C mediate fitness in plants

Charlie F. Bayne\textsuperscript{a,1}, Max E. Widawski\textsuperscript{a,1}, Feng Gao\textsuperscript{b,1}, Mohammed H. Masab\textsuperscript{a}, Maitreyi Chattopadhyay\textsuperscript{b}, Allison M. Murawski\textsuperscript{a}, Robert M. Sansevere\textsuperscript{a}, Bryan D. Lerner\textsuperscript{a}, Rinaldys J. Castillo\textsuperscript{a}, Trevor Griesman\textsuperscript{a}, Jiantao Fu\textsuperscript{a}, Jennifer K. Hibben\textsuperscript{a}, Alma D. Garcia-Perez\textsuperscript{a}, Anne E. Simon\textsuperscript{b}, and David B. Kushner\textsuperscript{a,*}

\textsuperscript{a}Department of Biology, Dickinson College, Carlisle, PA 17013, USA

\textsuperscript{b}Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA

\textsuperscript{1}Authors equally contributed to this work

*Corresponding author

David B. Kushner, Department of Biology, Dickinson College, PO Box 1773, Carlisle, PA 17013, USA

Tel: +1 717-245-1328

Fax: +1 717-254-8138

E-mail: kushnerd@dickinson.edu
Abstract

Noncoding RNAs use their sequence and/or structure to mediate function(s). The 5’ portion (166 nt) of the 356-nt noncoding satellite RNA C (satC) of Turnip crinkle virus (TCV) was previously modeled to contain a central region with two stem-loops (H6 and H7) and a large connecting hairpin (H2). We now report that in vivo functional selection (SELEX) experiments assessing sequence/structure requirements in H2, H6, and H7 reveal that H6 loop sequence motifs were recovered at nonrandom rates and only some residues are proposed to base-pair with accessible complementary sequences within the 5’ central region. In vitro SHAPE of SELEX winners indicates that the central region is heavily base-paired, such that along with the lower stem and H2 region, one extensive hairpin exists composing the entire 5’ region. As these SELEX winners are highly fit, these characteristics facilitate satRNA amplification in association with TCV in plants.

Key Words: Turnip crinkle virus, satellite RNA, RNA evolution, SELEX, RNA structure
INTRODUCTION

There has recently been a concerted effort to use genome-wide approaches to identify the structure of RNAs present in cells. The development of the so-called RNA structurome has provided insights into structures of mRNAs and noncoding (nc)RNAs in prokaryotic and eukaryotic cells, as well as viral RNAs (Wan et al., 2011; Mortimer et al., 2014; Bevilacqua et al., 2016; Ignatova and Narberhaus, 2017). While genome-wide approaches have been useful in gaining global insights into RNA structure-function relationships, other molecular, biochemical, and biophysical approaches are needed to obtain more detailed analyses about important RNAs. One such approach has been to use in vitro SELEX (systematic evolution of ligands by exponential enrichment [Ellington and Szostak, 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990]) to examine requirements in RNA structure, by starting with a pool of random sequences that are subjected to selection to produce functional molecules. The value of SELEX in assessing sequence-structure relationships is that no a priori assumptions are made about sequence and/or structure requirements for function, as any such requirements can be identified during the selection process. Also, subjecting a large region of an RNA to SELEX allows for many possible sequence/structure permutations to exist in an initial pool. SELEX can also be applied in vivo, although pools are usually not saturating (e.g., randomizing a 30-nt region would require $4^{30}$ [1.2 x $10^{18}$] molecules to achieve saturation). However, the variety in the initial pool is typically sufficient to allow for selection of functional molecules and for further evolution of these initial molecules, followed by genetic and biochemical analyses of possibly functional features (Carpenter and Simon, 1998; van Bel et al., 2014).

As ncRNAs by definition do not encode proteins, determining their secondary structure (and any tertiary interactions) is needed before structure can be linked to possible functions.
RNA viruses and virus-associated coding and ncRNAs (known as satellite [sat] viruses and satRNAs) fold into one or more functional structures that are important for successful propagation of both the satellites and the helper virus, which share a mutualistic co-existence (Simon and Gehrke, 2009). A number of examples of satRNA structure-function exist; a few include the following. (i) Cucumber mosaic virus satD4 RNA is necrogenic in tomato and the necrogenic domain of satD4 forms a stem-loop; three nucleotide changes in this region that could alter the local structure caused loss of necrogenicity by this mutant satRNA (Rodriguez-Alvarado and Roossinck, 1997; Irian et al., 2007). (ii) The 3’-terminal region of Cymbidium ringspot virus sat-Cym RNA toggles between two different structures; mutational analysis has shown that both sets of these secondary structures are required for replication in protoplasts (Ashton et al., 2015). (iii) In regard to Bamboo mosaic virus, the conserved apical hairpin stem-loops in the 5’ UTR of satBaMV is important for replication of the satRNA and in the competition between satBaMV and BaMV gRNA for replication complexes (Yeh et al. 2004; Chen et al., 2012).

One satRNA that has been extensively studied in regard to sequence-structure-function relationships is the noncoding 356-nt satRNA C (satC) of Turnip crinkle virus (TCV). While most satRNAs do not share sequence similarity with the genomic (g)RNA of their helper virus, 166 nt at the 3’ end of satC are derived from two nearly-adjacent regions at the 3’ end of the TCV gRNA (Fig. 1A). Although sequences at the 3’ ends of TCV gRNA and satC are very similar, they feature distinct 3D structures related to their different functional requirements. For TCV, hairpins H4a, H4b, and H5, and pseudoknots Ψ2 and Ψ3, form a T-shaped structure (TSS; McCormack et al., 2008) that serves both as a ribosome-binding cap-independent translation enhancer and as a hub for extensive non-canonical RNA interactions in the region (Stupina et al., 2011; Yuan et al., 2012). The current model for the gRNA switching from a translated template
to a replication template posits that the RNA-dependent RNA polymerase (RdRp) binds to an A-rich region upstream of H4a/Ψ3, leading to disruption of H4a/Ψ3 that disassembles the TSS, resulting in a widespread conformational switch (Yuan et al., 2009; Yuan et al., 2012; Le et al., 2017). SatC, despite having similar versions of all these hairpins and pseudoknots, is not predicted to form a TSS and does not bind ribosomes due to 6 single-nt differences with the TSS of TCV (Guo et al., 2011). However, initiation of minus-strand synthesis by the RdRp requires that the 3’ end of satC toggles between pre-active and active structures (Zhang et al., 2006a; Zhang et al., 2006b).

In vivo SELEX has been used to select for satC sequences throughout the 3’ portion of satC that mediate satC fitness (the most fit satC are those that best enhance movement/accumulation of the TCV helper virus in plants). These hairpins include M1H, H4a, H4b, H5, Ψ2, and the 3’-terminal Pr hairpin (Carpenter and Simon, 1998; Guan et al., 2000a; Guan et al., 2000b; Sun and Simon, 2003; Zhang and Simon, 2003b; Zhang et al., 2004; Sun et al., 2005; Zhang and Simon, 2005; Zhang et al., 2006b). SELEX revealed that H4a and H4b, which are not part of a TSS in satC, can exist as either two stem-loops or a single larger hairpin, provided that Ψ2 is maintained (Guo et al., 2009). The 5’ portion of satC, which is derived from a second satRNA of TCV (satD), is more poorly characterized. Mfold (Zuker, 2003) and in vitro SHAPE (selective 2’-hydroxyl acylation analyzed by primer extension [Merino et al., 2005]) analyses have suggested that the 5’ 166 nt of satC form a separate RNA domain (Murawski et al., 2015). This 5’ portion was suggested to contain (i) a lower stem; (ii) an upper connecting hairpin termed H2; and (iii) a central region composed of two opposing stem-loops (H6 and H7; Fig. 1B). The H2 region (76 nt, positions 48-123) was extensively examined using in vivo SELEX by replacing it with either 76, 38, 19, or 0 random nt. The most fit satCs that were
recovered possessed either a 38 or 39 nt H2 region that folded into a stable hairpin, with one recovered satRNA having fitness similar to WT satC for propagation with the helper virus in plants (Murawski et al., 2015). Furthermore, consistent with prior studies examining the effect of deletions within the H2 region (Carpenter et al., 1991; Zhang, 2006), satC with reduced-length H2 accumulated mainly in dimeric form in protoplasts unlike WT satC, which mainly accumulates as monomers along with low levels of multimeric forms (Murawski et al., 2015).

For the current study, we sought to determine the sequence and/or structure requirements that facilitate satC fitness within the poorly characterized central region in the 5’ portion of satC. Data from multiple SELEX experiments combined with in vitro SHAPE analyses of SELEX winners suggest that satC is most fit when sequences in the central region maximize base-pairing and extend hairpin H2, which results in a single hairpin encompassing the entire 5’ portion of satC. However, in several instances, short WT sequence motifs were recovered that only partially contribute to the secondary structure of satC, suggesting that there is a preference, but not a requirement, for sequences that engage in either higher order structure or interaction with trans factors. In combination with the prior findings that a satC with a half-sized H2 region can replicate to WT levels (Murawski et al., 2015), this suggests that the 166-nt 5’ portion of satC, as with the 3’ portion, features structural plasticity that fulfills functions necessary for satRNA amplification and mutualistic association with its helper virus in plants.

RESULTS

Wild-type H6 loop sequence contributes to fitness of satC in plants

The previously designated central region of satC was predicted by mfold (Zuker, 2003) to fold into two stem-loops (H6 and H7), with the smaller H7 flanked by two stretches of unpaired
sequences (Murawski et al., 2015); this structure will be referred to as Structure $\alpha$ (Fig. 1B). However, RNA structure probing using in vitro SHAPE (Merino et al., 2005) only partially conformed to Structure $\alpha$ in the central region. For example, none of the hairpin H6 loop residues (positions 131-139) were flexible and only a minority of other unpaired residues (positions 23, 24, 29, 37, and 47) in the predicted secondary structure of this region showed moderate or high reactivity with NMIA (Murawski et al., 2015; Fig. 1B). One possibility for the lack of flexible H6 loop residues could be their participation in tertiary interactions in the central region and H2 hairpin, such as those depicted in Fig. 1B, and/or satC folds into a different structure in vitro that constrains the flexibility of these residues.

To gain a better understanding of the sequence and/or structural requirements within the satC central region, a series of in vivo SELEX and in vitro SHAPE experiments were performed. We first subjected H6 loop residues alone to in vivo SELEX. A pool of satC RNAs, each with random sequence replacing WT residues in positions 131-139, was mixed with TCV gRNA and inoculated onto 30 turnip seedlings. After 21 days, total RNA was harvested from young, emerging leaves and a portion of the full-length satCs within the population was cloned and sequenced. Total cellular RNA from all plants was pooled and used to inoculate 6 new turnip seedlings. RNA was harvested 21 days later from emerging leaves, followed by satC cloning. These steps were repeated for a total of 6 rounds.

Thirty-six satC containing 31 distinct, non-WT sequences in positions 131-139 were cloned in Round 1 (Fig. 2A). The recovered sequences could be roughly divided into four classes: (I) WT-like sequences containing between 5 and 8 WT satC residues (15 of 31); (II) non-class I sequences possessing the motif “AUAG” (3 of 31); (III) sequences with most or part of the motif “AUUAGA” (5 of 31); and (IV) unclassified sequences (8 of 31). Three of the
recovered class I clones (Alp, Blp, and Clp; the subscript “lp” denotes SELEX was of the H6 loop) had 8 of 9 WT residues, and 6 clones (Dlp, Elp, and Glp through Jlp) had 5 or 6 consecutive WT residues. When examining the 35 distinct clones from Rounds 1 and 2, the most common motifs recovered were WT residues in the 3’ half of the loop, in particular the overlapping sequences “AAGU” and “AGUAC” (each found in 8 of these 35 clones). When a computer program was used to generate 31 random 9-mer sequences, only a single “AAGU” and no “AGUAC” (or “GUAC”) were present (data not shown), strongly suggesting that at least a portion of the SELEX sequences recovered, even in Round 1, were non-random.

Only three different satCs were recovered in Round 4, which were also the only species cloned in Rounds 5 and 6. As shown in Fig. 2A, these SELEX “winners” were (i) WT satC (H6 loop: 5’-AUCAAGUAC-3’); (ii) class I sequence A lp (UUCAAGUAC; underlined residues are identical to WT), which also had second-site mutations at positions 39 and 201; and (iii) class II sequence W lp (AUAUAGAGC). To determine which satC winner was most fit, in planta competition experiments were conducted in which different combinations of the three satRNAs were inoculated in equal amounts onto turnip seedlings in the presence of the TCV helper virus, and satRNA cloned from emerging leaves 21 days later. As shown in Table 1, A lp was more fit than W lp when inoculated together or when included with WT satC (Competitions A and B, respectively). When A lp was directly compared with WT satC (Competition C), 41% of the clones (n=22) were A lp, suggesting that A lp and WT satC were similarly fit for accumulation in plants. Both A lp and W lp accumulated to levels similar to WT satC when separately inoculated with TCV into Arabidopsis thaliana protoplasts (Figs. 2D and 2E), indicating similar replication efficiencies for these satCs. Since this accumulation level was 5-fold higher than satC with a
random sequence in the H6 loop (ctrlip; see Supplemental Table 1 for the sequence), these results strongly suggest that specific sequences in the H6 loop contribute to satC fitness.

Two possible structures were predicted for A_{lp} and W_{lp}. One structure retained the WT two-hairpin central region (Structure α; Fig. 2B). However, unlike A_{lp} and WT satC, no >4 nt tertiary pairing opportunities are discernable for W_{lp} between its H6 loop sequence and putative unpaired residues in its 5’ region (or 3’ region). An alternative structure, termed Structure β, converted the satC central region into an extension of the H2 stem, eliminating H6 and H7 (Fig. 2C). In A_{lp} Structure β, U131 (the single H6 loop difference with WT satC) pairs with A39, one of the two second-site A_{lp} alterations (Fig. 2C, in green). W_{lp} Structure β was very similar to that of A_{lp}, and many Round 1 recovered sequences (23 of 31) were also capable of pairing across the stem, forming part of an elongated H2 structure (Fig. S1).

As A_{lp} and W_{lp} could be modeled to form Structure β, we revisited the in vitro SHAPE analysis of WT satC and found that the SHAPE data was also consistent with WT satC forming either Structure α or β (Figs. 1B and 1C). Although the modeled Structure βs for A_{lp} and W_{lp} were consistent with most of the WT satC SHAPE results (compare Figs. 1C and 2C), only 5 of 9 H6 loop residues in WT satC, A_{lp}, or W_{lp} were predicted to base-pair across the stem (for WT: AUCAAGUAC; underlined). Therefore, Structure β still did not account for why the H6 loop SELEX gave rise to so many WT-like sequences or sequences with specific motifs.

To gain a better understanding for why these specific H6 loop sequences were selected, we repeated the H6 loop SELEX but this time in combination with randomization of residues in positions 15 to 41. We reasoned that this could help determine if recovery of these non-random H6 loop sequences was predicated on WT sequences in the 5’ portion of the central region. As shown in Fig. 3A, none of the new selected H6 loop sequences contained more than 3
consecutive WT residues, nor did any contain either of the previously found motifs. All of the winning sequences in the final round (Round 5) were derived from one of two parental sequences. Within each of these two groups of cloned satC, most of the differences were due to second-site mutations (Fig. 3A). Direct competition in protoplasts between a member from each set (H_b and U_b) as well as an additional winner from the SELEX experiment described below (C_a) revealed that 19 of 24 were H_b (Table 1, Competition D). In direct competition between WT satC and either H_b or U_b, 2 of 12 clones were H_b whereas 0 of 12 clones were U_b (Table 1, Competitions E and F, respectively). In contrast, U_b accumulated to near WT levels in protoplasts, which was over 2-fold higher than accumulation of H_b (Fig. 3B); similar relative degrees of accumulation were observed for (-)-strand replication intermediates (Fig. S2), indicating that these SELEX winners are not blocked in progeny (+)-strand synthesis from their (-)-strand replication intermediate. These results indicate that despite its weaker accumulation in protoplasts, H_b was substantially more fit than U_b for accumulation in young leaves with TCV, suggesting that sequence differences in these regions may enhance replication and/or stability in addition to allowing for more robust movement in association with the helper virus in plants.

Similar to WT satC, two structures were predicted for competition winner H_b. In Structure α, the H_b-selected sequence in positions 15 to 41 was predicted to fold into a hairpin occupying a similar position as WT hairpin H7, with a 5-bp stem and 7-nt terminal loop (Fig. 3C, left). This H7 loop sequence could potentially pair with 6 consecutive residues in its selected H6 loop sequence (H7: 5’-GUGGUGA and H6: 3’-CCUACCACA). In Structure β, pairing of a portion of the two sequences extended the H2 stem (Fig. 3C, right). SHAPE structure probing was most consistent with full-length H_b adopting Structure β in vitro. In addition, only Structure β could be modeled for the less fit (but better able to replicate in protoplasts) U_b SELEX winner
(Fig. 3D). Altogether, results of the first two SELEX experiments suggest (i) recovery of WT-like sequences in the H6 loop was in part due to the ability of a portion of these sequences to form Watson-Crick base pairs with sequences present at least partially within positions 15-41 of satC; (ii) there does not seem to be an absolute requirement for specific WT sequences in the central region; and (iii) Structure α, where H6 loop sequence pairs with accessible complementary sequences within the 5’ central region, and/or Structure β, with an elongated H2 stem, may be important for satC fitness in plants.

To further investigate the possible structure of the 5’ portion of satC, an additional SELEX was performed where only positions 15-41 were randomized. We predicted that, like Hb (Fig. 3C), the fittest winners of this SELEX should contain sequences complementary to the WT H6 loop sequence in an accessible (single-stranded) location, and/or that these sequences should also at least partially pair in an alternative extended H2 stem structure. As shown in Fig. 4A, the clear winner of this SELEX was Ca, which was the only sequence recovered in the final round (Round 4). The selected sequence alone was not predicted to contain any secondary structure, but did contain 5’-GCUUGA, which was complementary to WT H6 loop sequence 3’-UGAACU (Fig. 4B, left). Full pairing between these same two 6-mer sequences was also modeled to extend the H2 stem in Structure β (Fig. 4B, right). Although Ca accumulated plus-strands (as well as [-]strand replication intermediates) to levels approaching WT satC in protoplasts (Figs. 4C and S2), Ca, even with the WT satC H6 loop sequence, was less fit than Hb and was uncompetitive against WT satC in plants (Table 1, Competitions D and G, respectively), suggesting that additional fitness may require more paired residues in the central region of Structure β.
To gain further insights into the structure of the satC central region, and in particular whether an H6-type hairpin (within Structure α) contributes to fitness, SELEX was performed with random sequences replacing the entire H6 hairpin (positions 125-145). Six rounds of selection were completed as described above, and results are presented in Fig. 5A. After Round 1, 29 satC were cloned in which 23 were distinct, non-WT sequences. Several Round 1 clones contained sequence similar to the WT H6 loop (5’-AUCAAGUAC-3’). For example, SH6 contained “AAGUAC”, CH6 had “AGUAC”, and 7 additional clones (FH6, GH6, HHH6, OHH6, QHH6, and RHH6) contained either “GUAC” or “AAGU”. In contrast, when a computer program generated 23 21-nt random sequences, only 1 “GUAC” and no “AAGU” resulted (data not shown). Despite the non-random recovery of WT satC H6 loop sequences, the SELEX sequence in SH6 within full-length satC was not predicted to form a hairpin either independently or together with adjacent satC sequences, and this clone (as with most Round 1 clones) was not recovered after Round 1. In fact, of all Round 1 clones, only GH6 was cloned subsequently, and then only through Round 3.

In contrast, Round 6 winner ZH6 (first cloned in Round 2 even though it would have been present in one of the 30 Round 1 plants) did not contain more than three consecutive WT satC H6 residues. Two additional clones, XH6 and YH6 (recovered in Round 2 only), differed minimally from ZH6 and thus all three were likely derived from a single parental satC sequence. This suggests that while acquisition of WT H6 loop sequences aids in fitness, other parameters, such as those found in ZH6, are more important. The selected sequence in ZH6 was capable of forming an H6-like hairpin in Structure α (Fig. 5C) and also was capable of folding into Structure β (Fig. 5D). Five residues in the ZH6 H6 loop could possibly pair with (predominantly) H7 loop sequences (Fig. 5C, pink sequences). SHAPE structure probing of ZH6 indicated that the
5’ portion of the H6 SELEXed sequence formed nearly perfect base-pairing with sequence on the opposing side of the stem (Fig. 5D). However, the remainder of the SELEXed sequence was unpaired and the resulting large internal loop was similar to Structure β of winning clone Hb (Fig. 3C, right).

To determine the fitness level of ZH6, equal amounts of ZH6 transcripts were co-inoculated with TCV and either WT satC or Alp onto turnip plants. After 21 days, only WT satC (n=14) or Alp (n=15) were recovered from these plants (Table 1, Competitions H and I, respectively), indicating that ZH6 was of lesser fitness. However, ZH6 accumulated to WT levels in protoplasts, which was 20-fold higher than a satC with randomly selected sequence replacing H6 (Fig. 5B; Supplemental Table 1). Overall, the results of the H6 SELEX supported selection for two (possibly connected) criteria within the 5’ central region: (i) satC with an extended H2 stem; and (ii) satC with accessible sequences (i.e., such as those in a terminal “H6” loop) that can pair with sequence in the opposing side of the stem.

Deletion of H2/H6 and replacement with random sequences

To continue testing the validity of a two-criteria selection, we performed SELEXes of satC that replaced 98 nt (including the 76-nt H2 hairpin, A124 (which links H2 and H6), and the 21-nt H6 sequence) with either (i) 21 random nt plus A124, or (ii) 60 random nt (these SELEXes were labeled as Δ98+21 and Δ98+60, respectively). Our reasoning was that outcomes might be further clarified by testing shorter satC, as prior work revealed a half-sized H2 can be at least as fit as WT satC (Murawski et al., 2015). In addition, if WT satC H6 loop sequence motifs continued to be selected, we could eliminate possible interaction with the H2 region as the reason for the presence of these specific sequence motifs.
The results of the Δ98+21 SELEX are shown in Fig. 6A. Twelve of 20 cloned satC from Round 1 (A_{Δ98+21}, B_{Δ98+21}, and I_{Δ98+21} through R_{Δ98+21}) contained unique SELEXed sequences. Round 1 clone A_{Δ98+21} contained sequence (5′-AUC_AGUAC-3′) nearly identical to the WT H6 loop (5′-AUCAAGUAC-3′). Round 5 clone C_{Δ98+21} contained eight consecutive WT H6 loop sequences (5′-UCAAGUAC-3′). This indicates that acquisition of WT H6 loop sequence is unrelated to any sequence present in H2. Starting with Round 4, the only clones isolated (B_{Δ98+21} through H_{Δ98+21}) were from a single parental satC, with B_{Δ98+21} cloned most frequently in Round 6 (16 of 20; Fig. 6A). When B_{Δ98+21} was subjected to competition in plants with the other three related Round 6 clones (D_{Δ98+21}, E_{Δ98+21}, and G_{Δ98+21}), similar amounts of only G_{Δ98+21} and E_{Δ98+21} were recovered (Table 1, Competition J). Direct competition between Z_{H6} and B_{Δ98+21} or G_{Δ98+21} resulted in recovery of only Z_{H6}, indicating that these truncated satC were of lesser fitness (Table 1, Competitions K and L, respectively). In protoplasts, none of the four Round 6 winners accumulated monomeric satC significantly better than ctrl_{Δ98+21}, a satC in which the 98-nt deletion was replaced with A124 and a random 21-nt sequence (Fig. 6B, top and lower left; Supplemental Table 1). However, protoplasts infected with either B_{Δ98+21}, E_{Δ98+21}, or G_{Δ98+21} accumulated substantial amounts of dimeric satC (Fig. 6B, top and lower right). Previous analysis of satC clones with truncations in the H2 region showed similar low accumulation of monomeric satC and higher accumulation of dimeric satC (Carpenter et al., 1991; Zhang, 2006; Murawski et al., 2015). Accumulation of (-)-strand replication intermediates in protoplasts mirrored levels seen for (+)-strands (Figs. 6B and S2). Curiously, D_{Δ98+21} accumulated low levels of both monomeric and dimeric forms of the satRNA.

*In vitro* SHAPE analysis of G_{Δ98+21} revealed that the selected sequence was predicted to pair with most residues in positions 28-43, forming a substantially zipped structure similar to
WT Structure $\beta$ (Fig. 6C, right). When compared with $B_{\Delta 98+21}$, the three single-nt differences (two in the selected sequence and one second-site alteration) were predicted to reduce the size of the terminal loop and alter pairing in the stem (Fig. 6C, left). Curiously, symptoms displayed by plants inoculated with TCV gRNA and $G_{\Delta 98+21}$ or $B_{\Delta 98+21}$ were quite distinct, with $G_{\Delta 98+21}$ producing extreme crinkling of the leaves compared with $B_{\Delta 98+21}$ (and all other infections in this report) (Fig. 6C, bottom). The reason for these unusually severe symptoms remains under investigation.

Eight different satC's were recovered in Round 1 of the $\Delta 98+60$ SELEX (Fig. 7A). One Round 1 clone, $F_{\Delta 98+60}$, was the progenitor for all winning clones found in Round 6. $F_{\Delta 98+60}$ contained the WT H6 loop sequence 5'-AUCAAG-3' (WT positions 131-136), and two clones derived from $F_{\Delta 98+60}$ ($O_{\Delta 98+60}$ and $P_{\Delta 98+60}$) contained a 3-nt deletion within this sequence while still possessing the same “AUCAAG” sequence due to an additional “AUC” just upstream. Competition in planta between $F_{\Delta 98+60}$ and the related clones found in Round 6 ($J_{\Delta 98+60}$, $K_{\Delta 98+60}$, $L_{\Delta 98+60}$, and $O_{\Delta 98+60}$) revealed that $F_{\Delta 98+60}$ and $K_{\Delta 98+60}$ were the most fit (Table 1, Competition M). Similar competitions with $Z_{H6}$ and WT satC did not lead to recovery of any $F_{\Delta 98+60}$ (Table 1, Competitions N and O, respectively). However, $F_{\Delta 98+60}$ was more competitive than $G_{\Delta 98+21}$ (Table 1, Competition P), which is consistent with prior results correlating H2 length and fitness (Murawski et al., 2015). All Round 6 $\Delta 98+60$ winners accumulated more monomeric satC in protoplasts compared with $\Delta 98+21$ winners, but still only reached about 20% of WT satC (Fig. 7B, top). Similar to the $\Delta 98+21$ winners, WT or greater levels of dimeric forms of satC were recovered for all five $\Delta 98+60$ winners (Fig. 7B, bottom). In vitro SHAPE structure probing of $F_{\Delta 98+60}$ revealed an elongated stem similar to Structure $\beta$ found for $G_{\Delta 98+21}$, with the $F_{\Delta 98+60}$ second-site mutation (A23C) increasing the stability of the lower stem (Fig. 7C). Although no
obvious hairpin could form that placed the “AUCAAG” sequence in a terminal loop, most of this sequence was predicted to be accessible within an internal loop.

**Evolution of satC lacking H6 sequence**

From the SELEX experiments described above, fitness of satC appeared to be correlated with extensive base-pairing throughout the 5’ region. In addition, the non-random recovery (usually in early rounds) of WT H6 loop sequence motifs by some clones in the absence of H2 and without concomitant hairpin generation suggests that some additional fitness of WT satC is conveyed by the presence of this sequence (possibly when in an accessible form within Structure α, as found in WT satC). It is probable that insufficient evolutionary time was available for the SELEX-winning satRNAs to present this sequence in an accessible hairpin loop within one structure, and in an extended hairpin in an alternative structure. We thus speculated that (i) satC with a deletion of the entire H6 sequence, without addition of any random sequence, would adopt only one structure with the short H7 hairpin flanked by unpaired residues; and (ii) evolution of this truncated satC should involve acquisition of many second-site alterations that provide new opportunities to base-pair the remaining sequence in the central region, if this provided additional fitness. To test this hypothesis, satC without the 21-nt H6 sequence (satC_{ΔH6}) was subjected to 8 rounds of *in planta* evolution. In Round 1, 15 of 22 clones contained the input satC_{ΔH6}, indicating that satC_{ΔH6} is still capable of amplification when H6 is absent (Fig. 8A). In addition, the remaining 7 recovered satC contained an assortment of second-site mutations, one of which (A34C) was also found in all clones beginning in Round 3. In Round 2, new clone L_{ΔH6} contained A42G and three adjacent second-site changes: C48G, C49U, and C50A, which together with upstream residue A47, resulted in acquisition of the WT satC H6 loop sequence
“AGUA”. \(L_{\Delta H6}\) likely was the progenitor for all subsequently cloned satC in Rounds 3-7, since nearly all recovered satC contained these three adjacent alterations at positions 48-50. \(M_{\Delta H6}\), the most prominent clone in Rounds 3-7, gained the A34C mutation and deleted the residue in position 42 (either the original A or the second-site G). \(N_{\Delta H6}\), which first appeared in Round 7 and was the predominant clone in Round 8 (9 of 13), had a deletion of one of five consecutive uridylates (arbitrarily designated as U114), an alteration that had also independently appeared in clones \(J_{\Delta H6}\) and \(K_{\Delta H6}\) in Round 2. In competition with WT satC, no \(N_{\Delta H6}\) were recovered (\(n=19\); Table 1, Competition Q). In addition, although on average \(N_{\Delta H6}\) accumulated 3-fold better than starting molecule satC\(_{\Delta H6}\), the level reached was 5-fold less than WT satC (Fig. 8B); similar relative levels of (-)-strand replication intermediates were observed (Fig. S2).

As described above, the predicted structure of satC\(_{\Delta H6}\) contained a mostly undisturbed H2 hairpin as well as the short H7 hairpin and, flanking H7, a total of 17 unpaired residues (Fig. 8C, left). A34C, which appears to have arisen in several independent clones, altered a residue in the H7 terminal loop without any obvious consequence to the structure or engagement in interactions elsewhere in the satRNA, and thus its contribution to fitness remains unclear. The four second-site changes in \(L_{\Delta H6}\) were predicted to extend the stem of hairpin H7 at the expense of the H2 lower stem, decreasing the number of unpaired bases in the secondary structure (Fig. 8C, middle). \(M_{\Delta H6}\) and \(N_{\Delta H6}\) deleted the residue at position 42, which shortened a bulge loop from four residues to three (Fig. 8C, right). \(N_{\Delta H6}\) replaced \(M_{\Delta H6}\) as the major cloned sequence after deleting U114 (described above), which also decreased the size of a bulge loop in H2 from four to three residues. In vitro SHAPE structure probing of \(N_{\Delta H6}\) was consistent with the predicted structure (Fig. 8D). Interestingly, \(O_{\Delta H6}\), which appeared in Round 8, contained all \(N_{\Delta H6}\) alterations and one additional change at position 20 (U20C). This alteration is predicted to
change a U:G pair to a C:G pair at the top of the \( N_{AH6} \) lower stem. Mfold estimates of the stability of the satC region (positions 1-166) suggest that these second-site changes are neutral with regard to the overall stability of the structure. However, all second-site changes reduced the number of unpaired residues, which may be of higher importance to satC fitness. Some second-site alterations found in Round 1 clones (e.g. A29G and A44U) are also predicted to extend the H7 hairpin (data not shown). However, these alterations were not present in subsequent rounds, suggesting that their contribution was dwarfed by the more robust alterations found in Round 2 progenitor clone \( L_{AH6} \).

**DISCUSSION**

Prior computational and structural analyses suggested that the central region in the satD-derived 5’ portion of satC contained the mfold-predicted hairpins H6 and H7. However, SHAPE revealed that the terminal loops of both putative hairpins contained mainly inflexible residues, suggesting that if these hairpins exist, then the loop sequences were likely involved in tertiary interactions (Murawski *et al.*, 2015). In this study, we sought to better understand the structure and importance of this 5’ central region, *i.e.*, whether putative hairpins H6 and H7 contribute to fitness and whether any particular sequences within these putative structures are important for satC accumulation. The *in vivo* SELEX experiments reported here revealed that sequences identical to all or a portion of the H6 loop sequence (5’-AUCAAGUAC-3’) were frequently selected, but only when the satRNA maintained WT residues in positions 15-41 (the H7 region). For example: (i) overlapping motifs “AAGU” and “AGUAC” were found in nearly half of the clones recovered during the first two rounds of SELEX of the H6 loop alone (Fig. 2A); (ii)
replacing the 98-nt H2/H6 region with 21 random nt (and A124) generated Round 1 clone A_{Δ98+21} with 5’-AUC_AGUAC-3’, and Round 1 clone C_{Δ98+21} with 5’-UCAAGUAC-3’ (Fig. 6A); (iii) replacing the 98-nt H2/H6 region with 60 random nt generated “AUCAAG” in F_{Δ98+60} (Figs. 7A and 7C); (iv) substituting randomized residues for the complete H6 sequence resulted in early-round recovery of “AAGUAC”, “AAGUA”, “AGUAC”, “AAGU”, or “GUAC” (Fig. 5A); and (v) “AGUA” was generated within the H7 region via three second-site mutations that arose during the passaging of satC with an H6 deletion and no randomized residues (Figs. 8A and 8C).

At first glance, the recovery of so many H6 loop-related sequences when positions 15-41 are WT suggests that the H6 loop motifs pair with position 15-41 residues located on the opposing side of the central region. However, there are only limited pairing partners for these recovered motifs within positions 15-41 (Fig. 1B). These results therefore suggest that sequences in the H6 loop contribute to fitness of satC for reasons other than purely Watson:Crick pairing of residues between the H6 and H7 regions.

The selected H6 loop motifs recovered were infrequently present as terminal loop sequences in putative hairpins similar to H6 (for an example, see Fig. 5C). This suggests that a hairpin similar to H6 is not important for fitness but rather that the H6 loop motifs may favor placement in an accessible location for maximum (e.g., WT satC) fitness. Examination of many of the recovered sequences that were unrelated to H6 revealed that a substantial number contained 5 to 6 consecutive residues that could pair with sequence in the H7 region, if H6 and H7 collapse into a zipped structure that extends hairpin H2 (Structure β; Fig. S1). This suggests that a second criteria for fitness is an extended H2 stem. Re-examination of the prior WT satC SHAPE data revealed results that were also consistent with Structure β (Murawski et al., 2015;
Fig. 1C). Additional support for an extended H2 structure contributing to fitness comes from examining the following SELEX winners: (i) winners H_b and U_b (SELEX of H6 loop and positions 15-41), which are predicted to fold into an extended H2 structure, with the H_b structure supported by SHAPE (Figs. 3C, right, and 3D); (ii) winner C_a (positions 15-41 SELEX), which recovered 6 consecutive residues that can pair with sequence in WT H6 (Fig. 4B, right); (iii) winner Z_{H6} (SELEX of complete H6), whose extended H2 structure was confirmed by SHAPE (Fig. 5D); and (iv) winners B_{A98+21} and G_{A98+21} (SELEX of 21 nt replacing H2/H6), which was predicted to adopt an H2-like structure that was also supported by SHAPE (Fig. 6C).

Altogether, our SELEX results suggest two fitness criteria for the central region of satC: (i) H6 loop motifs in an accessible location such as the terminal loop of a hairpin (as found in WT satC) and (ii) an extension of the H2 stem formed from zipping together H6 and H7 sequences. Of particular interest is winner F_{A98+60} from the Δ98+60 SELEX (Fig. 7A). F_{A98+60}, which was highly competitive among related SELEX winners (Table 1, Competition M) contains both an H6-type motif (AUCAAG), with the AA residues highly flexible as shown by SHAPE, and an overall highly-zipped 5’ region (akin to Structure β; Fig. 7C). Although there is accumulating evidence for fitness associated with Structure β, an open question is whether hairpins H6 and H7 (i.e., Structure α) still exist in WT satC, possibly as an alternative form.

Results from the current study are also consistent with previous findings that decreasing the length of the 5’ portion of satC both reduces fitness and enhances satC dimer accumulation in protoplasts (Murawski et al., 2015). In the current study, Competitions K, L, and N through Q (Table 1) included SELEX winners possessing different lengths. In each competition, the longer satC was dominant. Furthermore, WT-length SELEX winners A_{lp} and W_{lp} accumulated ratios of monomers to dimers that were similar to WT satC (Fig. 2E) whereas winners of the Δ98+21
SELEX, which were 76 nt shorter than WT satC, accumulated almost exclusively as dimeric satRNA (Fig. 6B). Interestingly, winners of the Δ98+60 SELEX, which typically were 38 nt shorter than WT satC, still accumulated a disproportional amount of dimers (Fig. 7B), whereas winners from the in planta evolution of satCΔH6 (21 nt shorter than WT) had ratios of monomers to dimers similar to WT satC (Fig. 8B). This raises the question of what accounts for an enhanced dimer to monomer ratio? *Cucumber mosaic virus* (CMV) Q-satRNA dimers are generated in the presence, but also in the absence, of helper virus (Choi *et al.*, 2012). When CMV is absent, Q-satRNA localizes to the nucleus where multimers accumulate containing a 7-nt motif at the junction between monomers. Since the CMV helper virus is absent, and these are nuclear events, host RNA polymerase II was proposed to mediate dimerization (Choi *et al.*, 2012). Recent work proposes that the bromodomain-containing host protein BRP1 transits Q-satRNA (as well as *Potato spindle tuber viroid*) into the nucleus (Chaturvedi *et al.*, 2014). As satC also accumulates in a dimer form in the absence of TCV helper virus following agroinfiltration (R. Guo and A.E.S., unpublished), it is possible that shorter satRNAs with a single extended 5’ hairpin (*i.e.*, not NΔH6 with its two hairpin conformation) may also preferentially localize to the nucleus and accumulate dimers while being poorly replicated by the helper virus RdRp in the cytoplasm.

**Concluding remarks**

The relationship between satC and TCV is mutualistic: satC is replicated by the helper virus RdRp and in return, TCV trafficking through the plant is more rapid (Zhang and Simon, 2003a). These SELEX experiments thus select for satRNAs that are both more efficient templates for accumulation and can enhance TCV systemic movement. Although our findings
clearly show that an extended hairpin in the 5’ portion of satC contributes to fitness, the most fit satC by competition experiments is not necessarily the satC that best accumulates in protoplasts. Thus a major question is how does an extended hairpin in the 5’ portion of satC, combined with acquisition of specific sequence motifs, contribute to assisting virus systemic movement? One possibility is if satC occupies both cytoplasmic and nuclear locations within a cell, analogous to the CMV satRNA. The extended hairpin H2 of satC could be a substrate for a nuclear dicer-like endonuclease, releasing a small RNA product that contributes to systemic movement; indeed, a number of satRNAs can be processed to form small RNAs with various functions in the infected host (Palukaitis, 2016; Wang and Smith, 2016). In addition, a shorter H2 could enhance satC transit into the nucleus increasing dimer accumulation, thereby reducing satC available for RdRp-mediated cytoplasmic replication, which would generate monomers. Work is currently ongoing to examine this and other explanations for the still enigmatic relationship between satellite and helper virus.
Acknowledgments

The authors thank Aaron Brumbaugh, Daud Byanyuma, Tracy Campbell, Brittany Coyle, Ridge Dershem, Breanna Goldner, Antonio Marrero, Colleen McCabe, Charles McElwee, Matthew Murray, Stephanie Neal, Thomas Nowlan, Rosemary Palmieri, Jennifer Rush, Jessica Snyder, Amelia Szabat, Christine Tallo, Arielle Verdi, Brett Weidman, Benjamin Weissman, Yuzhi Xi, Mesrob Yeterian, Allison Younkins, and Songhui Zhao, who along with C.F.B., M.H.M., B.D.L., J.F., and J.K.H. performed the first two rounds of the in vivo SELEXes during the laboratory portion of one of two offerings of BIOL 419 at Dickinson College. This research was supported by a grant from the National Science Foundation (MCB-0918624), and funds from the Research and Development Committee of Dickinson College, to D.B.K. and a grant from the National Science Foundation (MCB-1157906) to A.E.S.
Materials and Methods

**In vivo SELEX.** In vivo SELEX was performed essentially as previously described (Carpenter and Simon, 1998; Guo et al., 2009; Murawski et al., 2015). In this work, six templates were generated for in vitro transcription of satC RNA with distinct sequences randomized. Oligonucleotides described below were synthesized by Integrated DNA Technologies (Coralville, IA).

(i and ii) To generate the template for in vitro transcription of satC with 9 random nt replacing the putative H6 loop (positions 131-139), or 21 random nt replacing the putative H6 stem-loop (positions 125-145), two fragments were generated by separate PCRs with pC(+) (pUC19 containing full-length satC cDNA) as template. The 5’ fragment containing either 9 or 21 random nt was produced by using primers T7C5’ (5’-GTAATACGACTCATAAGGGATAACTAAGGGTTTCA-3’; T7 promoter sequence is italicized) and either reverse oligo 185-109loop-randomize (5’-GTCTGGGATTTTGGAGTGGAACACAGCCAGGTGTTCACGCTCCC[N9]GGGAACCTCCCGCCTTTAAAAAC-3’; HinfI site is underlined) or reverse oligo 185-109H6-randomize (5’-GTCTGGGATTTTGGAGTGGAACACAGCCAGGTGTTCAC[N21]TCCCGCTTTAAAAAC-3’; HinfI site is underlined). In both cases the 3’ fragment was generated by using forward primer 145-161 (5’- GTGAAAACCTGGCTGT-3’ and reverse primer oligo 7 (5’-GGGCAGGCCACCCCCGTCCGA-3’, which is complementary to 19 nt at the 3’ end of satC).

(iii) To generate the template for satC with H2 (positions 48-123) deleted and H6 replaced with 21 random nt (∆98+21), the 5’ fragment was generated using forward primer T7C5’, reverse primer 185-33randomize (5’-GTCTGGGATTTTGGAGTGGAACACAGCCAGGTGTTCAC[N21]TTGGTCTGCATTAGT
T-3’; HinfI site is underlined), and pΔH2 as template (pΔH2 is a satC clone that lacks H2; Murawski et al., 2015). The 3’ fragment was amplified from pC(+) using primers 145-161 and oligo 7.

(iv) To generate the template for satC with 60 random nt replacing the 98-nt region containing H2 and H6 (positions 48-145; Δ98+60), PCRs with pC(+) as template were performed. The 5’ fragment was produced by using primers T7C5’ and reverse primer 47-28 (5’-TGTTCTGCATTTAGTGCGTA-3’; MwoI site underlined). The 3’ fragment containing 60 random nt was generated by using forward primer 28-157randomize (5’-TACGCAACTAATGCAGAACA[N60]GTGAAAAACCTGG-3’; MwoI site underlined) and oligo 7.

(v) To generate the template for satC with randomization of the 27 nt composing the 5’ side of the central region (positions 15-41; Murawski et al., 2015), PCRs with pC(+) as template were performed. The 5’ fragment was produced by using forward primer T7+1-54randomize (5’-GTAATACGACTCATATAGGGATAACTAAAGGG[N27]AGAACACCCATGC-3’; T7 promoter sequence italicized) and reverse primer 114-93 (5’-AAAAACCCACCATGGTTAAACC-3’; NcoI site underlined). The 3’ fragment was produced by using primers 93-114 (5’-GGTTAACCATGGTGTTTTT-3’; NcoI site underlined) and oligo 7.

(vi) To generate the template for satC with randomization of both positions 15-41 and the putative H6 loop (positions 131-139), PCRs with pC(+) as template were performed. The 5’ fragment was produced by using forward primer T7+1-54randomize and reverse primer 114-93. The 3’ fragment was produced using forward primer 93-158randomize (5’-GGTTAACCATGGTGTTTTTAAAGGCGGGAGTTCCC[N9]GGGAGCGTGAAAACCTGGC-3’; NcoI site underlined) and oligo 7.
All PCR products were subjected to electrophoresis, and DNA fragments of expected mobility were purified using Wizard® SV Gel and PCR Clean-up System columns (Promega, Madison, WI), digested with Hinfl (for templates i, ii, and iii), MwoI (template iv), or NcoI (templates v and vi; all enzymes procured from New England Biolabs, Ipswich, MA, except where noted), phenol/chloroform extracted, and respective 5’ and 3’ fragments ligated together to produce full-length satC cDNA. Separate in vitro transcription reactions from these templates and synthesis of TCV genomic RNA (from pTCV66) were performed as described (Murawski et al., 2015). For the first round of selection, 2 µg of WT TCV gRNA transcripts and 5 µg of satC transcripts with specific randomized sequences were inoculated onto each of 30 turnip seedlings (Just Right Turnip Hybrid, Gurney’s, Greendale, IN or Wammock Farm Service, Sylvania, GA). Total RNA was extracted from uninoculated leaves after 21 days, pooled, and inoculated onto 6 new turnip seedlings for an additional 21-day infection. This procedure was repeated for a total of 4 rounds (SELEX v), 5 rounds (SELEX vi), or 6 rounds (SELEXes i-iv). After each round, satCs were cloned and sequenced (Clemson University Genomics Institute, Clemson, SC) as described (Murawski et al., 2015).

In vivo self-evolution. In the presence of TCV gRNA, satC with a deletion of H6 (satC_{AH6}) was passaged through plants for 8 rounds. In two steps, a plasmid was constructed to allow transcription of satC_{AH6} using T7 RNA polymerase. First, to delete H6, pC(+) (which contains satC but no T7 promoter) was used as template in PCR reactions with forward primer 93-170_{AH6} (5’-GGTTTAACCATGGTGGGTTTTTAAAGGCGGGAGTGAAAACCTGGCTGTTTCCCACTC-3’; NcoI site underlined) and reverse primer 301-283 (5’-AGCCCCACCCCTTTCCGGGATT-3’).
The 188-bp PCR product was digested with NcoI and BstEII and ligated into NcoI-BstEII digested pC(+) to make intermediate clone pC(+)<sub>ΔH6</sub>. Second, a T7 promoter was added 5’ of satC within pC(+)<sub>ΔH6</sub> via PCR with forward oligo BamHI-T7C5’ (5’-CTTCCCGGATCCGTAATACGACTCACTATAGGGATAACTAAGGTTTCA-3’; BamHI site underlined; T7 promoter sequence italicized) and reverse oligo 121-104 (5’-CGCCTTTAAAAACCCACC-3’). The 151-bp product was digested with BamHI and NcoI and ligated into BamHI-NcoI digested pC(+)<sub>ΔH6</sub> to make satC<sub>ΔH6</sub>, which was sequence-verified.

SatC<sub>ΔH6</sub> and pTCV66 each were linearized using SmaI and in vitro transcripts were synthesized using T7 RNA polymerase. Six plants were inoculated with transcripts; after 21 days, total RNA was harvested, pooled, and used to inoculate 3 new plants for a total of 8 rounds. SatC from each plant was cloned and sequenced (Murawski et al., 2015).

**In vivo competitions between various evolved and cloned satC.** Competition experiments between equal amounts of T7 polymerase-generated transcripts of SELEX winners, the in vivo satC<sub>ΔH6</sub> passaging winner, and/or WT satC were performed as previously described (Zhang and Simon, 2005; Guo et al., 2009; Murawski et al., 2015). WT satC transcription template was generated by PCR using pC(+) as template and primers T7C5’ and oligo 7. The PCR product was gel purified and phenol/chloroform extracted before use in transcription reactions with T7 RNA polymerase. For SELEX/in vivo passaging winners, T7 transcription reactions were performed after SmaI-linearizing plasmid (with cloned satC) templates. TCV gRNA was transcribed from SmaI-linearized pTCV66 using T7 RNA polymerase. For all competitions, controls in which individual plants were co-infected with TCV gRNA and RNA of each satC SELEX winner were performed simultaneously, to verify that each satC transcript was functional.
*in planta* (movement from inoculated to new leaf after 21 days) and able to be cloned from RNA extracted from the new leaf.

**Accumulation of viral RNAs in protoplasts.** Analysis of accumulation of viral RNAs in protoplasts was performed as previously described (Murawski *et al.*, 2015). TCV gRNA and satC transcripts were *in vitro* transcribed with T7 RNA polymerase using SmaI-linearized plasmids pT7TCVms and pT7C(+) (Song and Simon, 1994), or directly from PCR products. Protoplasts (5 x 10⁶) prepared from callus cultures of *A. thaliana* ecotype Col-0 were inoculated with 20 µg of TCV gRNA transcripts with or without 2 µg of satC RNA transcripts using polyethylene glycol-CaCl₂ as described (Zhang *et al.*, 2006b). Total RNA isolated from protoplasts at 40 hours post-inoculation was subjected to RNA gel blot analysis. A [γ-³²P]ATP-labeled, satC-specific oligonucleotide (oligo 7) was used to probe for satC (+)-strands. To obtain clones of satC with randomized sequences in H6, H6 and H7, or H6 and H2 for use as negative controls (Supplemental Table 1) in protoplast experiments, ligated satC cDNAs (described above) were directly cloned into the SmaI site of pUC19.

**In vitro SHAPE RNA structure probing.** SHAPE analysis was performed essentially as described previously (Murawski *et al.*, 2015). Six pmoles of *in vitro* transcribed WT and mutant (ZH6, GΔ98+21, FΔ98+60, Hb, and NλH6) satC transcripts each were heated for 5 min at 65°C, snap-cooled for 2 min on ice, then incubated for 20 min at 37°C in SHAPE Folding Buffer (80 mM Tris-HCl pH 8, 11 mM Mg(CH₃COO)₂, 160 mM NH₄Cl). Three pmoles of the folded RNA was combined with either N-methylisatoic anhydride (NMIA) or DMSO at a final concentration of 15 mM, incubated for 40 min (5 half-lives of NMIA) at 37°C, then ethanol precipitated.
was resuspended in 8 μl of 0.5x TE buffer. Primer extension reactions were performed using a [γ-32P]ATP-labeled oligonucleotide (complementary to positions 235-221) and Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA) as previously described (Wilkinson et al., 2006). Radioactively-labeled products of reverse transcription were resolved on 8% denaturing polyacrylamide gels and visualized using a PhosphorImager. RNA secondary structures were generated from structure probing results and the best-fitting mfold predictions (Zuker, 2003).
References


Simon, A.E., 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.


Figure legends

Fig. 1. TCV, satC, and satC structure. (A) Top, TCV genomic RNA (gRNA) is 4053 nt and encodes five proteins (p28, p88 [RdRp], p8, p9, and coat protein [CP]). Middle/bottom, satD and satC are two satellite RNAs that can be associated with TCV gRNA; the 5’ end of satC is derived from satD (grey shading), and the 3’ end of satC is derived from two nearly-adjacent regions at the 3’ end of TCV gRNA (striped region and black shading; Simon and Howell, 1986). (B and C) In vitro SHAPE structural analysis of the 5’ portion of satC (Murawski et al., 2015 and this report) reveals two possible structures. Red, green, and black nt denote strongly flexible, moderately/weakly flexible, and inflexible nt, respectively. (B) Structure α features two inflexible stem-loops termed H6 and H7; residues in the inflexible H6 loop could base-pair with other inflexible regions in the 5’ portion of satC (pink and blue dashed lines). (C) Structure β features H6 and H7 sequences integrated into an extended H2 hairpin.

Fig. 2. In vivo SELEX of the putative H6 loop. (A) Results from 6 rounds of SELEX with 9 random nt replacing the putative H6 loop (lp; satC positions 131-139). Samples were grouped into 4 categories based on either possessing a similar sequence motif (I-III) or due to lack of relatedness (IV). The number of times each sequence was cloned per round is indicated in the respective columns. Along with WT satC, only A lp and W lp (yellow boxes) were isolated after Round 6. Second-site mutations for A lp are defined at the bottom of the figure. (B and C) Predicted secondary structures α (B) and β (C) for non-WT SELEX winners A lp and W lp. Class I and II motifs are shown in blue and magenta, respectively; for A lp, the U39A second-site mutation and non-WT U131 are shown in green. Structure α (B) features stem-loops H6 and H7.
while structure $\beta$ (C) features an extended H2 hairpin. (D and E) Relative accumulation of (+) satC monomers in *A. thaliana* protoplasts. Data in (D) is from three independent experiments. Total RNA was extracted from protoplasts at 40 hours post-inoculation with transcripts of TCV gRNA and relevant satRNA, electrophoresed, and subjected to Northern blotting. Representative Northern blot lanes from the same gel are shown in (E). The sequence of ctrlp (and controls for other protoplast experiments in this paper) is listed in Supplemental Table 1.

Fig. 3. *In vivo* SELEX of positions 15-41 and the H6 loop (positions 131-139). (A) Results from 5 rounds of SELEX. The two groups of related samples (Hb-Ob and Qb-AAb) are boxed. Within the boxes, red residues indicate differences from Hb or Qb; blue residues in Hb and Qb (and WT) denote potential base-pairing. Second-site mutations are listed below the figure. (B) Relative accumulation of (+) satC monomers in protoplasts (see legend to Fig. 2D). (C) Structure $\alpha/\beta$ models for winner Hb; Structure $\beta$ best matches *in vitro* SHAPE analysis of Hb (red/green indicate flexible nt as noted in Fig. 1). Position of second-site mutation C130U is denoted with an arrow. Blue residues in Structure $\alpha$ can putatively pair. (D) Winner Ub can be modeled to form Structure $\beta$; position of second-site mutation G142A is denoted with an arrow.

Fig. 4. *In vivo* SELEX of positions 15-41. (A) Results from 4 rounds of SELEX. In magenta, UACU (located 5' of WT H7) can interact with sequence in the loop of WT H6 (Fig. 1B). The two groups of related samples (Ca-Fa and Ma-Na) are boxed; for the former group, the blue GCUUG sequence, along with A42, can base-pair with sequence in WT H6. For related clones Ca-Fa, the red C nt in Fa indicates a difference in the recovered sequence in the randomized region. Second-site mutations are listed below the figure. (B) Structure $\alpha/\beta$ models for winner
Fig. 5. *In vivo* SELEX of H6 (positions 125-145). (A) Results from 6 rounds of SELEX with 21 random nt replacing H6. Nt in blue indicate sequence also found in the WT H6 loop. SELEX winner $Z_{H6}$ is shown in a yellow box with related clones $X_{H6}$ and $Y_{H6}$; the green nt indicate differences between these related clones and the pink nt can base-pair with sequence in H7 as shown in (C) and, for 4 of these 5 nt, in (D). The second-site mutation for winner $Z_{H6}$ is noted at the bottom. (B) Relative accumulation of (+) satC monomers in protoplasts (see legend to Fig. 2D). (C) Structure α model for winner $Z_{H6}$. (D) Structure β model for winner $Z_{H6}$ based on *in vitro* SHAPE analysis (red/green indicate flexible nt as noted in Fig. 1 legend). In (C) and (D), the A29G second-site mutation in $Z_{H6}$ is noted by a carat.

Fig. 6. *In vivo* SELEX of satC with H2 and H6 (positions 48-145) replaced by A124 and 21 random nt. (A) Results from 6 rounds of SELEX in which A124 followed by 21 random nt replaced the H2 (positions 48-123) and H6 (positions 125-145) regions. Blue nt note sequence found also in the WT H6 loop. The SELEX winners (and related satC cloned prior to Round 6) are boxed in yellow; nt differences from $B_{98+21}$ are in red. Second-site mutations are listed at the bottom of the figure. (B) Northern blot (*top*) and relative accumulation of (+) satC monomers (*lower left*) and dimers (*lower right*) in protoplasts (see legend to Fig. 2D). (C) Predicted secondary structure for the 5’ region of winner $B_{98+21}$ and *in vitro* SHAPE-based model of $G_{98+21}$ (red/green indicate flexible nt as described in legend for Fig. 1). The location of the H2 deletion is noted in orange; the bracket shows the location of the SELEXed sequences (now part of the hairpin). Numbering is based on WT satC. Boxed residues denote differences.
between $B_{\Delta 98+21}$ and $G_{\Delta 98+21}$. For $B_{\Delta 98+21}$, the A44G second-site mutation is denoted with a carat. Photographs illustrate differential crinkling symptoms of turnip plants 21 days after co-inoculation of TCV gRNA and transcripts of the respective satC SELEX winner.

Fig. 7. *In vivo* SELEX of satC with H2 and H6 (positions 48-145) replaced by 60 random nt. (A) Results from 6 rounds of SELEX. Two groups of related sequences are boxed in yellow. In each box, red nt indicate differences from the first sequence listed. For $F_{\Delta 98+60}-P_{\Delta 98+60}$, blue nt denote a 6-mer motif found in the WT H6 loop. For $R_{\Delta 98+60}$, the magenta nt match Class II sequence from the H6-loop SELEX (Fig. 2A). Second-site mutations are shown below the figure. (B) Relative accumulation of (+) satC monomers (top) and dimers (bottom) in protoplasts (see legend for Fig. 2D). (C) Winner $F_{\Delta 98+60}$ was subjected to *in vitro* SHAPE (red/green nt are as described in legend for Fig. 1). Although 38 nt shorter than WT satC, $F_{\Delta 98+60}$ has an H2-like extended hairpin featuring a 6-mer found in the WT H6 loop (red box).

Fig. 8. Evolution of satC with H6 deleted. (A) Eight rounds of passaging satC$_{\Delta H6}$ through turnip plants. Six common second-site mutations are color-coded. (B) Northern blot (top) and relative accumulation of (+) satC monomers in protoplasts (bottom; see legend for Fig. 2D). (C) Structure models for $L_{\Delta H6}$, $N_{\Delta H6}$, and starting template satC$_{\Delta H6}$. For satC$_{\Delta H6}$, the locations of the six second-site mutations (at positions 34, 42, 48-50, and 114) found in $N_{\Delta H6}$ are noted in red. For $L_{\Delta H6}$ and $N_{\Delta H6}$, the second-site mutations are color-coded as in part (A); the AGUA motif at positions 47-50, which matches a 4-mer in the WT H6-loop, is underlined in green. (D) Model of $N_{\Delta H6}$ based on *in vitro* SHAPE analysis.