doi:10.1016/S0022-2836(02)01366-9 *J. Mol. Biol.* (2003) **326**, 35–48

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

A Multifunctional Turnip Crinkle Virus Replication Enhancer Revealed by *in vivo* Functional SELEX

Guohua Zhang and Anne E. Simon*

Department of Cell Biology and Molecular Genetics University of Maryland College Park College Park, MD 20742, USA

*Corresponding author

The motif1-hairpin (M1H), located on (−)-strands of *Turnip Crinkle Virus* (TCV)-associated satellite RNA C (satC), is a replication enhancer and recombination hotspot. Results of *in vivo* genetic selection (SELEX: systematic evolution of ligands by exponential enrichment), where 28 bases of the M1H were randomized and then subjected to selection in plants, revealed that most winners contained one to three short motifs, many of which in their (−)-sense orientation are found in TCV and satC (−)-strand promoter elements. Ability to replicate in protoplasts correlated with fitness to accumulate in plants with one significant exception. Winner UC, containing only a seven-base replacement sequence, was the second most fit winner, yet replicated no better than a 28-base random replacement sequence. Fitness of satC containing different M1H replacement sequences could be due to enhanced satC replication or enhanced ability to affect TCV movement, since satC interferes with TCV virion accumulation, which is correlated with enhanced movement to younger tissue. Cells inoculated with TCV and UC accumulated fewer virions when compared to other winners that replicated better in protoplasts but were less fit in plants. UC, and other first and second round winners, contained structures that were on average 33% more stable in their (+)-strand orientation, and most formed hairpins with a A-rich sequence at the base. These results suggest that M1H replacement sequences contribute to the fitness of satC by either containing (−)-strand elements that enhance satRNA replication and/or a (+)-strand hairpin flanked with single-stranded sequence that enhances TCV movement.

Keywords: RNA virus replication; RNA recombination; viral movement enhancer; satellite RNA; RNA-dependent RNA polymerase

Introduction

Replication of (+)-sense RNA viruses is a two-step process that begins with the synthesis of an RNA strand complementary to the input strand and finishes when the complementary strand is copied into an identical version of the input strand. A virus-encoded enzyme, the RNA-dependent RNA polymerases (RdRp), contains the active site for nucleotide polymerization and may be associated with additional virus and/or host-encoded subunits whose functions remain largely unknown.[1] Promoters for RdRp vary widely in both sequence and structure, even for multiple promoters of a single RdRp.[2,3] Since promoters are recognition sites for the RdRp or associated factors, it is currently unclear how specificity of the recognition is achieved.

Core promoters for full length complementary strand synthesis of (+)-strand RNA viruses have been identified for a number of RdRp and usually comprise one or more hairpins located within untranslated sequence near the 3′ end, with the terminal nucleotides usually, but not always, single-stranded.[4–7] Some viral RNAs can support a low level of replication in the absence of the 5′[8] or 3′[9] untranslated region, suggesting that internal sequences can functionally replace the normally utilized promoter.[8] Recently, a number of internal sequences in viral RNAs have been identified that contribute either directly or indirectly to viral RNA replication. An internal region that enhances

Abbreviations used: TCV, *Turnip Crinkle Virus*; SELEX, systematic evolution of ligands by exponential enrichment; M1H, motif1 hairpin; satC, satellite RNA C; RdRp, RNA-dependent RNA polymerase; wt, wild-type; CP, coat protein; sg, subgenomic.

E-mail address of the corresponding author: anne_simon@umail.umd.edu

the replication of a defective interfering RNA associated with *Tomato Bushy Stunt Virus* is also an enhancer in the context of the genomic RNA.[10] Internal regions within the coat protein (CP) coding sequence of *Tobacco Etch Virus*,[11] the RNA3 intergenic region of *Brome mosaic virus* (BMV)[12] and the 2C coding sequence of poliovirus[13,14] also contain important replication elements. Functions for the latter two internal elements have recently been determined. The internal poliovirus hairpin, known as the CRE, is the template for uridylylation of the protein replication primer, VPg,[14] while the BMV internal sequence is required to produce an active replicase complex[15] and for template sequestration by the BMV 1a protein.[16]

We have been examining *Turnip Crinkle Virus* (TCV) and its satellite (sat) RNAs for sequences involved in RNA replication.[17] TCV (4054 bases), a member of the genus *Carmovirus* in the family *Tombusviridae*, is a (+)-sense RNA virus with five open reading frames (ORFs) that encode two subunits of the viral RdRp (p28 and the readthrough product p88, which contains the polymerase active site motif GDD); two proteins involved in virus movement (p8 and p9); and the CP that forms a 180 subunit, $T = 3$ icosahedral capsid[18–20] (Figure 1(A)). The movement proteins and the CP are translated from two subgenomic (sg) RNAs of 1.7 and 1.45 kb, respectively,[21] which are synthesized from internal promoters on the TCV (−)-strand.[22,23]

The association of TCV with several small satRNAs that are also replicated by the TCV RdRp provides a large number of promoter elements that can be assayed for functionality and compared for sequence and structural similarities[2,22,24–28] (Figure 1). SatC, a satRNA composed of two regions of TCV at the 3′ end and the nearly complete sequence of satD at the 5′ end (Figure 1(A)), contains a number of elements important for transcription of (+) and (−)-strands. SatC (+)-strands contain a 3′ hairpin flanked by six single-stranded 3′ terminal bases that are conserved among all TCV-associated RNAs (Figure 1(B)). SatC (−)-strands contain two small linear elements, the 3′PE and 5′PE, which can function as independent promoters in a cell-free (*in vitro*) system containing partially purified TCV RdRp[2] (Figure 1(D)). Either the 3′PE or 5′PE is required for complementary strand synthesis *in vitro*[2] but only the 5′PE is required *in vivo*.[27,28] Also required for replication *in vivo* is the (−)-strand 3′ terminal *Carmovirus* consensus sequence (CCS: $C_{2\text{-}3}A/U_{3\text{-}7}$), which is present at the transcription initiation sites of all TCV-associated RNAs[27] (Figure 1(C)).

A 30-base interior hairpin, the motif1-hairpin (M1H; Figure 1(D)), is also required for efficient replication of satC monomers (but not dimers) *in vivo*.[29,30] The M1H was originally identified as a hotspot for RNA recombination between satD and satC during (+)-strand synthesis *in vivo*.[31] Evidence that the TCV RdRp is recruited by the M1H to the (−)-strand satC acceptor template during recombination-mediated template switching comes from the composition of the (−)-sense M1H, which includes motifs found in the 5′PE and TCV 3′CCS promoter elements[30] (Figure 1(D)); the ability of the (−)-sense M1H to direct 3′ terminal extension from primers base-paired 3′ of the hairpin *in vitro*[32,33]; the ability of the p88 polymerase subunit expressed in *Escherichia coli* to extend internally base-paired 3′ ends, a reaction that requires the M1H;[34] and competition experiments between wt and mutant M1H using template containing an internally base-paired 3′ end.[33] These results suggest a model where the M1H is able to promote recombination between satD and satC by directing the RdRp (still complexed with newly synthesized satD (+)-strand) to re-initiate synthesis 3′ of the M1H on satC (−)-strands. In addition, this model suggested that the M1H might be an enhancer of satC replication by promoting the interaction of the RdRp with satC (−)-strand templates. In agreement with this model, the M1H was found to function as a replication enhancer *in vitro*, stimulating complementary strand synthesis from a (−)-strand promoter (the 3′PE) by tenfold in its (−)-sense orientation and sevenfold in its (+)-strand orientation.[29] In addition, the M1H could convert to an independent promoter *in vitro* when placed downstream of a CA-rich sequence.[29] However, the M1H in either orientation was only weakly stimulative of transcription from the hairpin promoter at the 3′ end of satC (+)-strands,[29] supporting earlier suggestions that the M1H functions mainly in its (−)-sense orientation during (+)-strand synthesis to aid in attracting the RdRp to satC (−)-strand templates.

Extensive mutagenesis of the satC M1H has been conducted and alteration of bases in the upper stem, lower stem, interior asymmetric loop and terminal loop were all found to reduce the ability of the hairpin to promote both recombination with satD *in vivo* and the analogous self-primed 3′ terminal extension *in vitro*.[32,33] While these results indicated a role for both sequence and structure of the hairpin in recombination and 3′ terminal extension, most individual mutations had little discernable effect on satC accumulation in plants.[31,35] Thus, to clarify the role of the M1H in enhancement of satC replication, we initiated an analysis of the hairpin by *in vivo* functional selection also known as *in vivo* SELEX (systematic evolution of ligands by exponential amplification). Unlike traditional *in vitro* SELEX, i.e. SELEX that uses binding as an assay or searches for a desired catalytic activity,[36] function-based *in vivo* SELEX searches for functional sequences that impart fitness to the molecule, regardless of the mechanism of action.[26–28,37–39] Our results indicate that the M1H can be replaced by a diverse group of sequences that, in the (−)-sense orientation, contain multiple consecutive cytidylates and motifs found in TCV promoter and enhancer elements as well as novel motifs. Surprisingly, the vast majority

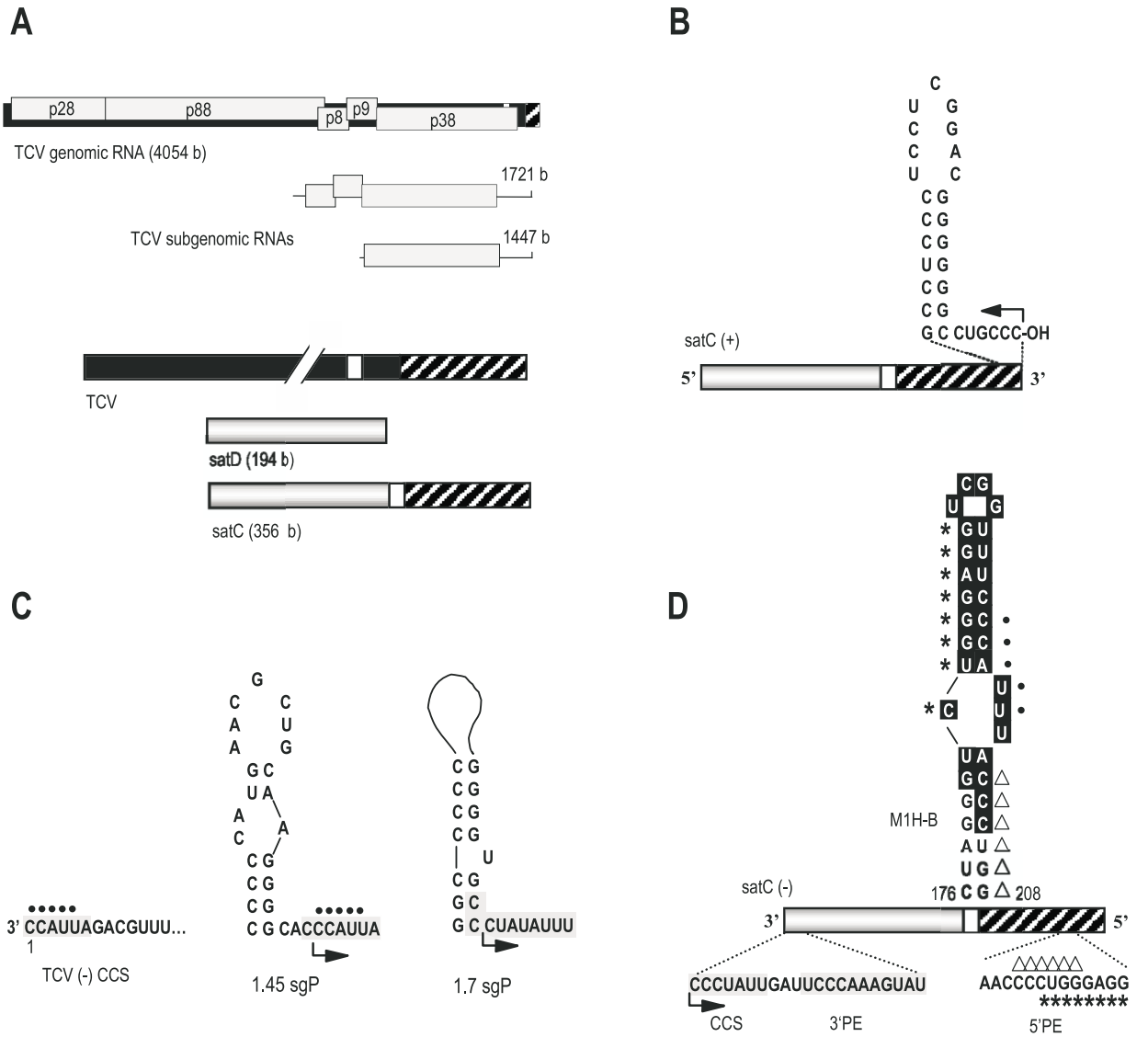

Figure 1. TCV-associated RNAs and *cis*-acting sequences involved in replication. (A) RNAs associated with TCV. Upper panel shows the TCV genomic RNA and two sgRNAs. p28 and p88 are the viral encoded subunits of the RdRp; p8 and p9 are movement-related proteins and p38 is the CP. The 1.7 kb sgRNA is the mRNA for p8 and p9 and the 1.4 kb sgRNA is the mRNA for the CP. Genomic RNA shading is related to the bottom panel, which shows the relationship between the recombinant satC and its parental RNAs, satD and TCV. Similar sequences are shaded alike. (B) Hairpin promoter on satC (+)-strands required for complementary (−)-strand synthesis. Initiation is at the 3′ terminal nucleotide (arrow). (C) TCV (−)-strand promoter elements. The TCV CCS (boxed), is located at the 3′ end of TCV (−)-strands; the 1.45 sgRNA promoter (1.45 sgP) is a minimal version of the sgP that is equally active as the wt promoter, which contains a much more extensive hairpin stem.[22] Note that both sgP have a CCS ($C_{2\text{-}3}A/U_{3\text{-}7}$; boxed) at the transcription start sites (arrow). Filled circles above certain bases denote sequence identity with bases in the M1H on satC (−)-strands shown in (D). (D) SatC (−)-strand elements required for (+)-strand synthesis. The satC CCS and 3′PE are boxed. The 3′PE and 5′PE can serve as independent promoters *in vitro*. The M1H sequence shown (M1H-B) has a single change at the base of the stem (U to C at position 176) that created a *Bam* HI site in the corresponding cDNA required for the generation of satC with randomized M1H sequences. Sequence homology between motifs in the M1H and various TCV promoter elements are denoted by similar symbols. M1H boxed sequence was subjected to randomization for the *in vivo* functional SELEX.

of 28-base M1H replacement sequences fold into stem-loops that are more stable in the (+)-sense orientation and retain A-rich sequences on either side of the (+)-strand hairpin. This (+)-strand structure may contribute to the fitness of the satRNA in plants by reducing the accumulation of TCV virions, which promotes virus long-distance movement in the host.

Results

Functional selection of M1H replacement sequences

In the presence of satC, TCV is able to move more rapidly to younger tissue (F. Zhang, Q. Kong and A.E.S., unpublished results), which is likely

Table 1. M1H replacement sequences in first round SELEX winners

Plant	Name	Sequence
1	1-1	CCAGCAUAAAAUCUACGUCCAUACGAAU
	1-2	UCCAAACUAUUAAGUUACGCAGUGACCU
2	2-1	CAAAGGGUGAGUUUUCCUAGGUCAGGAA
	2-2	CCGUUAGUACCACACCAGGAAGACCCCG
4	4-1	CAACCCCAUCAGCAAUCGUAGUGAGGCC
5	5-1	ACCGGCGAAGGUAGUUCCAUCACCAGAA
6	6-1	AGCCGCAAUAUAAAUGCCACACGGAACC
7	7-1	CCCAGGAUGAACAUCAACCCCGGCACCA
8	8-1	GGACCCCUCACCUACGAUGAACCCCUCC
9	9-1	CCCCAAUAGCAAUGUAGAUCCUAAGCCA
10	10-1	CAUUGUCUGAAACCAGAACCUAUCGCAA
11	11-1	ACCCAGGGGACAUAACCCCUCAUCCGCA
12	12-1	CAACCCCGAAGAGGUCCCCAAAACCCGG
13	13-1	UUCAACCUUAGUGAACCCAUAAAAGCG
14	14-1	AAAAUACUACCUCUCUAACACUCCCCCA
15	15-1	UACGAUUCACCACAUGACUCGACCCGAU
16	16-1	CUCAGAGGUAUCUUACAGGCCCCGAGUC
17	17-1	CACACGACCAAACACAGUUCGCGCCAGA
18	18-1	CCACCUUUAGUACCGGCCCGUCGAAAAU
	18-3	ACGGUAGCACAUCGUUCGGGAUUUUACA
	18-5	ACACGCCCUGCUCACAUCUAGUGUUGCC
	18-6	CUAGUUACAUCGCCCGGCCAGGCUUCCG
19	19-1	CUGACACCCAUUAUUCGUGCACAACCCC
	19-2	AGUAUAGCUGUUAGUAGCUCAACCGCC
20	20-1	CCCACCCGCGGCCAACGCAGGACCUCGA

responsible for the enhanced virulence of TCV when associated with satC.[40] Since satC containing either a deletion of the M1H, or random sequences replacing the hairpin, accumulates poorly in protoplasts,[29,30] there is a strong selection for replacement sequences that either improve its ability to accumulate in plants and/or enhance TCV movement.

A total of 28 bases of the M1H (Figure 1(D), boxed) were randomized using a PCR strategy. Two PCR products were generated using a full-length satC cDNA clone: a 5′ fragment that included an upstream T7 RNA polymerase promoter and a 3′ fragment containing randomized bases replacing positions 181 to 208 of the M1H. After proper ligation of the two fragments, satC transcripts containing the randomized sequence in place of the M1H were synthesized using T7 RNA polymerase and then inoculated onto 19 turnip seedlings. Included in the inoculation were transcripts of TCV genomic RNA, which provides the template for translation of the RdRp and CP required to replicate and package the satRNA. At 21-days post-inoculation, total RNA was extracted from uninoculated leaves and full-length satC cDNA clones generated by reverse transcriptase (RT) PCR. Two to 18 full length cDNA clones were sequenced per plant and were designated as winners of the first round (Table 1). In all, 15 plants contained a single species of satC that was unique for each plant, while three plants contained mixtures of satC with two different M1H replacement sequences (unique for each plant) and one plant contained satC with four different sequences. Thus, a total of 25 satC with different M1H replacement sequences were recovered from the 19 infected plants.

To subject the first round winners to further competition, transcripts from all 25 satC first round winners were combined, and equal amounts inoculated onto six additional plants along with TCV genomic RNA. Sequencing the 32 satC clones generated from RNA isolated at 21-days post-inoculation revealed only a single species (clone 1-2) accumulating in all plants (Table 2). The second round screen was repeated with nine new plants by omitting 1-2 transcripts from the mixture and inoculating with transcripts of the remaining 24 first round winners. Five of the 24 satC species were recovered from at least two of these plants at 21-days post-inoculation (Table 2). In addition, a new satC species (clone UC) was recovered that contained a seven-base M1H replacement sequence. This sequence, "UCAGGAA" (since the sequence of the M1H is thought to be important in its (−)-sense orientation, all M1H replacement sequences are shown in (−)-sense orientated 3′ to 5′ unless otherwise stated) could have originated from several possible first round winners by a deletion of 21 bases from either clone 2-1 (the 28-base replacement sequence terminates with UCAGGAA); 2-2 (interior sequence contains CCA GGAA) or 7-1 (contains the sequence CCAGGAU)

Table 2. Second round 28-base SELEX

		# Recovered[a]	
Name	Sequence of second round winners[b]	Six plants	Nine plants[c] (total)
1-2	UCCAAACUAUUAAGUUACGCAGUGACCU	32	
2-2	CCGUUAGUACCACACCAGGAAGACCCCG	0	4-0-1-4-0-2-0-0-8 (19)
UC[d]	UCAGGAA	0	2-0-0-6-1-0-0-5-0 (14)
4-1	CAACCCCAUCAGCAAUCGUAGUGAGGCC	0	0-0-0-0-6-2-0-1-0 (9)
11-1	ACCCAGGGGACAUAACCCCUCAUCCGCA	0	0-0-0-0-0-2-5-0-0 (7)
6-1	AGCCGCAAUAUAAAUGCCACACGGAACC	0	0-3-0-0-0-3-1-0-0 (7)
5-1	ACCGGCGAAGGUAGUUCCAUCACCAGAA	0	0-2-0-1-0-0-1-0-1 (5)

[a] Only sequences recovered in two or more plants are presented.
[b] (−)-strand sequences are shown in a 3′ to 5′ orientation.
[c] Rows correspond to the nine different plants and columns indicate the number of times a particular sequence was recovered for that particular plant.
[d] UC was recovered only in second round plants. "CCAGGAA" was also recovered in one plant.

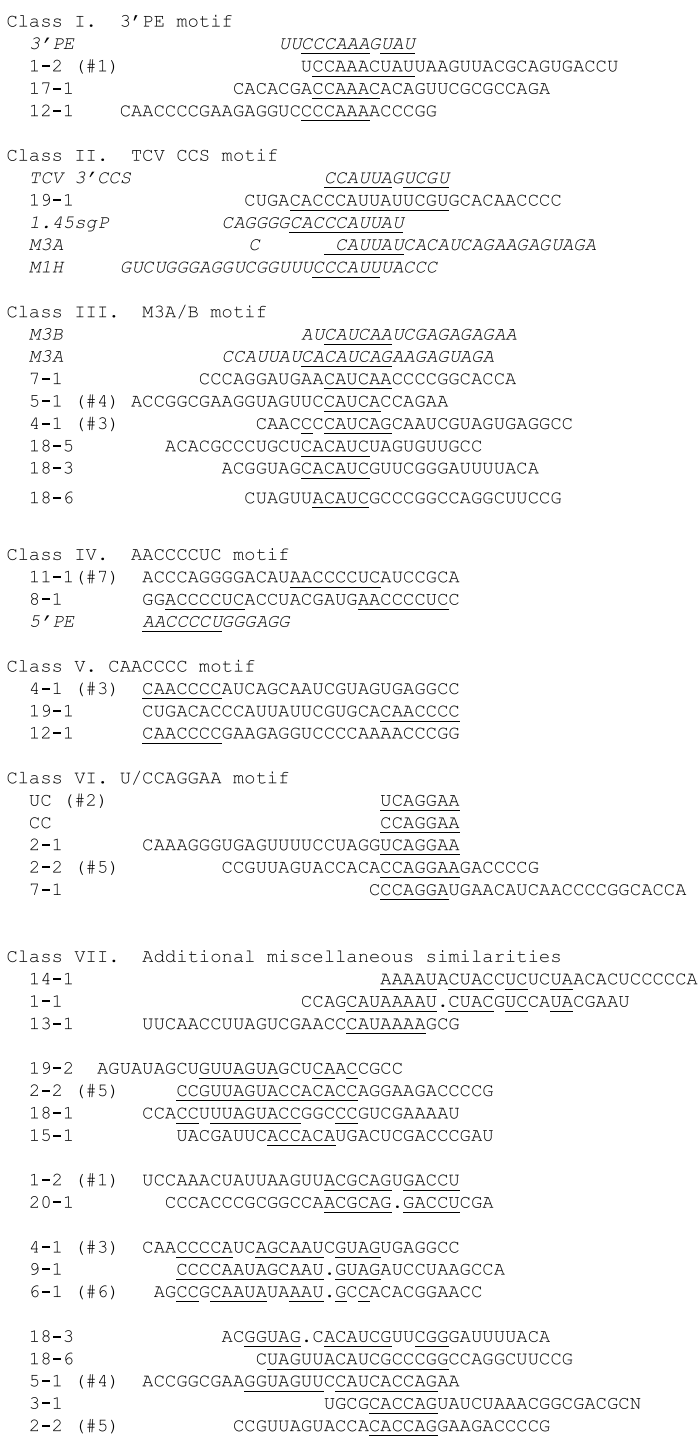

Figure 2. Motifs found in many first and second round M1H replacement sequences are also found in TCV and satC (−)-strand promoter and recombination elements. Full-length M1H replacement sequences from first and second round winners are shown (in (−)-sense oriented 3′ to 5′) and motifs with similarity to promoter elements are underlined (classes I through IV). Promoter elements (and their abbreviations) are shown in italics and are described in the legend to Figure 1. Other sequences in common among the replacement sequences are shown underlined in classes V to VII. All second round winner names are followed by numbers in parentheses that denote their relative fitness in direct competition assays in plants (see Table 3).

(see Figure 2, class VI). SatC containing a similar seven-base M1H replacement sequence, CCA GGAA, was also was recovered from one plant.

To ascertain if UC was a functional satRNA and not an artifact of the cloning process, full length satC containing the UCAGGAA M1H replacement sequence was generated and transcripts inoculated onto three plants along with TCV genomic RNA. At 21-days post-inoculation, total RNA was extracted from plants and analyzed by ethidium bromide-stained gel electrophoresis (satC accumulates to the level of 5 S ribosomal RNA and is easily visible in stained gels). SatRNA migrating slightly faster than wt satC was detected in all plants and sequencing revealed that the seven-base replacement sequence was stable (data not shown).

M1H replacement sequences contain motifs found in TCV promoter elements as well as novel motifs

The base composition of the recovered sequences was not random. Cytidylate and adenylate residues comprised 35% and 29% of the M1H

Table 3. Competition for fitness in plants

Co-inoculation	Number recovered (three to four plants)
wt satC	57
1-2	0
wt satC	30
UC	0
1-2	28
UC	9
UC	30
4-1	16
5-1	10
UC	23
6-1	6
2-2	0
UC	7
11-1	1
4-1	31
5-1	2
4-1	6
2-2	0
5-1	32
2-2	3
5-1	7
6-1	0
2-2	27
6-1	4
11-1	1
11-1	28
6-1	8

replacement sequences, respectively, while uridylates and guanylates each made up 18% of the remaining residues. Furthermore, there were 2.6-fold more CCA triplets and 3.6-fold more CCC triplets in the replacement sequences than expected for a purely random sequence, while no UGG triplets were recovered in any sequence.

While the overall sequences of the first round and second round winners differed from each other and from the wt M1H sequence, winners could be grouped in classes according to common sequence motifs (Figure 2). Some of the short motifs (generally from six to ten bases) that were shared among M1H winners comprised portions of previously defined (−)-promoter elements or *in vivo* RNA recombination hotspots. For example, 1-2, the top second round winner, contained the sequence CCAAACUAU, which is very similar to the 3′PE (UCCCAAAGUAU; identical sequence is underlined) (Figure 2, class I). The sequence CACCCAUUAUUCGU in first round winner 19-1 is identical to sequence at the transcription start site of the 1.45 kb sgRNA promoter located on TCV (−)-strands (CACcCAUUAU; transcription start site is in lower case) and similar to the 3′CCS of TCV (−)-strands (3′OH-CCAUUAGUCGU…) (Figure 2, class II). A portion of the underlined sequence above from clone 19-1 (CCAUUAU) is also in one of the two imperfect repeats (M3A) that together comprise the TCV (−)-strand recombination hotspot element known as the motif3-hairpin (M3H).[33,41] Previous results indicated that one of the two M3H repeats was required for TCV replication in protoplasts.[41] Sequence similarities between other short sequences in the M3H repeats and first and second round winners were also found (Figure 2, class III).

The motif AACCCCUC was repeated twice in second round winner 11-1 and once in first round winner 8-1 (Figure 2, class IV). Seven of eight contiguous bases are also part of the 5′PE (AAC CCCU). The similar motif CAACCCC, not previously found in any TCV promoter element, was recovered in second round winner 4-1 and first round winners 19-1 and 12-1 (Figure 2, class V). As described above, the second round winner UC, with the seven-base replacement sequence UCA GGAA, contains sequence found in three other first and second round winners (Figure 2, class VI). This motif was not previously found in any known TCV promoter element. In addition to the motifs just described, additional regions of sequence similarity could be found between various first and second round winners. For example, second round winner 4-1 contained the sequence CCCCAUCAGCAAUCGUAG, which is very similar to the 9-1 sequence CCCCAAU AGCAAU·GUAG (Figure 2, class VII). These results indicate that the majority of the M1H replacement sequences resemble the wt M1H by being composed of short motifs, some of which can be found in (−)-strand TCV promoters and recombination hotspots.

Fitness to accumulate in plants of most satC with M1H replacement sequences correlates with ability to replicate in protoplasts

To determine which of the replacement sequences found in the second round winners confers the most fitness for satC to accumulate in plants, equal amounts of transcripts of two or three of the winners were combined and inoculated with TCV genomic RNA onto three to four plants. At 21-days post-inoculation, total RNA was extracted and approximately equal numbers of clones generated from all plants were sequenced and the results combined and shown in Table 3. In direct competition between 1-2 and wt satC, only wt satC was recovered. UC, with its seven-base M1H replacement sequence, was surprisingly fit, accumulating better than 4-1, 5-1, 6-1 and 11-1 in direct competition. Only 1-2 out-competed UC, with 76% of clones recovered from plants having the 1-2 replacement sequence. Based on the results from Table 3, the order of fitness in plants of second round SELEX winners is: 1-2, UC, 4-1, 5-1, 2-2, 11-1 and 6-1.

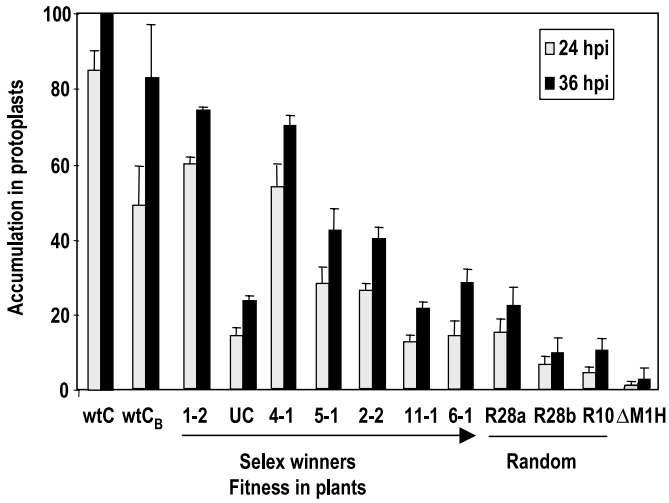

Figure 3. Accumulation of satC SELEX winners in protoplasts. Arabidopsis protoplasts (5×10^6) were inoculated with 20 μg of TCV genomic RNA transcripts and 2 μg of transcripts of wt satC (wtC), or satC with a U to C alteration at position 176 (wtC$_B$) that was the parental construct of the SELEX winners, or SELEX second round winners 1-2, UC, 4-1, 5-1, 2-2, 11-1 and 6-1 (fitness order in plants shown by a directional arrow). Control satC constructs R28a and R28b have randomly selected 28-base M1H replacement sequences, R10 has a ten-base randomly selected M1H replacement sequence and ΔM1H contains a 28-base deletion of the M1H. RNA was extracted at 24 or 36 hours post-inoculation and subjected to RNA gel blot analysis using an oligonucleotide probe that hybridized to TCV genomic RNA and satC. Values shown were calculated from at least three independent experiments.

The ability of satC, containing different M1H replacement sequences, to accumulate in plants is likely a combination of one or more of the following attributes: (i) ability to replicate; (ii) stability; and (iii) ability to direct the movement of TCV towards veins and thus promote more rapid systemic infection. To determine the effect of different M1H replacement sequences on accumulation in protoplasts, protoplasts prepared from TCV host *Arabidopsis thaliana* were co-inoculated with transcripts of each second round SELEX winner and TCV genomic RNA. Accumulation levels were compared to both wt satC and satC containing a single base change in the base of the M1H (satC$_B$) that was necessary to simplify the original construction of the SELEX constructs. As shown in Figure 3, this single base change in construct satC$_B$ resulted in an 18% reduction in accumulation compared with wt satC at 36-hours post-inoculation. The most fit SELEX winner, 1-2, accumulated to 89% of satC$_B$ levels while the least fit clones (11-1 and 6-1) did not accumulate significantly different from satC containing a non-selected 28-base replacement sequence (R28a). Clones 4-1, 5-1, and 2-2, which were intermediate in fitness in plant competition assays, accumulated in protoplasts in correlation with their fitness levels. The one exception, UC, accumulated more poorly in protoplasts than 4-1, 5-1, 2-2, 6-1 and R28a, while being more competitive when co-inoculated in plants (Table 3). SatC containing randomly selected 28-base or ten-base replacement sequences (R28a, R28b and R10) replicated better than satRNA containing a deletion of the M1H (ΔM1H), which is possibly a function of satRNA size differences. We have previously demonstrated that accumulation of satC and another TCV subviral RNA *in vivo* is substantially affected by reductions in size.[42–44]

To test for stability differences among the second round winners, protoplasts were inoculated with full-length transcripts of wt satC, satC$_B$, 1-2, UC, and R28a in the absence of TCV genomic RNA. RNA was extracted between zero and six hours post-inoculation and satRNA levels detected by RNA gel blot hybridization. All satC tested were equally stable (data not shown), agreeing with previous findings that wt satC and satC with a deletion of the M1H did not differ in their stability.[33]

These results suggest that most SELEX winners were selected due to enhanced replication of satC in infected cells, with the selected sequences functionally replacing the replication enhancer attribute of the M1H. However, the poor accumulation of UC in protoplasts suggests that factors other than replication enhancement influenced the selection of M1H replacement sequences in plants.

Enhanced fitness of winner UC correlates with inhibition of virion accumulation

SatC enhances the movement of TCV towards younger leaves in *Arabidopsis*, resulting in more virulent symptoms and the death of the plants within 14 to 21 days post-inoculation (F. Zhang, Q. Kong and A.E.S., unpublished results). This rapid movement of TCV correlates with a marked reduction in virion accumulation in the presence of satC in co-inoculated protoplasts (F. Zhang, Q. Kong and A.E.S., unpublished results). TCV requires CP, but not virion formation, for movement[23] and thus a reduction in virions, without a concomitant reduction in CP, could allow TCV to move as a ribonucleoprotein complex in a manner that accelerates movement towards veins.

To determine if the seven-base M1H replacement sequence in UC enhances fitness by interfering with virion formation, *Arabidopsis* protoplasts were inoculated with TCV in the absence or

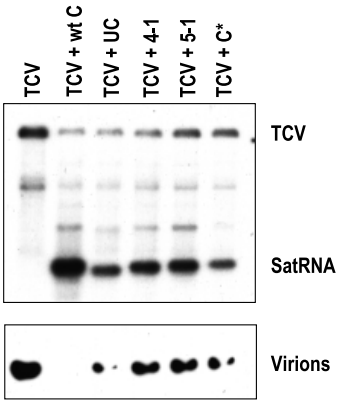

Figure 4. Effect of co-inoculation of TCV with satC on virion accumulation in protoplasts. Arabidopsis protoplasts (5×10^6) were inoculated with 20 μg of TCV transcripts alone or with 2 μg of wt satC (wtC), or second round 28-base SELEX winners UC, 4-1, and 5-1, or satC* (satC with the 3′ terminal 100 bases of TCV). RNA and virions were extracted at 40 hours post-inoculation. Upper panel is a representative RNA gel blot of total RNA probed with an oligonucleotide complementary to both TCV and satC. Virions (lower panel) were visualized by chemiluminescence following treatment with anti-TCV CP antibody.

presence of wt satC, UC, 4-1, 5-1 and satC*. SatC* is a variant of satC that has the wt M1H but contains the 3′ 100 bases of TCV in place of its own 3′ 100 bases, resulting in reduced accumulation in protoplasts to levels comparable with those of UC. As described above and shown in Figure 4, upper panel, UC replicated more poorly in protoplasts than 4-1 and 5-1, even though UC was more fit

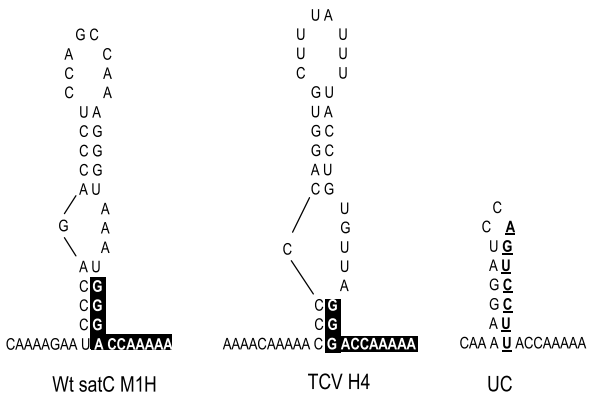

Figure 5. Plus-strand structures of the wt M1H, the M1H replacement sequence in SELEX winner UC, and TCV H4. TCV H4 is a hairpin similar in both structure and locational proximity to the 3′ end of the corresponding RNAs. The replacement sequence in UC is underlined. The second recombination event that produced satC occurred in the lower stem of TCV H4. Boxed sequence in satC M1H is thus derived from the boxed sequence in H4. Note that H4 forms a structure similar to the M1H in the upstream sequence that is unrelated to satC, including an A-rich region at the 5′ base of the stem.

Table 4. Stability of local hairpins formed by first and second round SELEX winner replacement sequences

			Randomized[a]	
Name[b]	ΔG (+)-strand	ΔG (−)-strand	ΔG (+)-strand	ΔG (−)-strand
wt	−15.9	−17.8		
1-1	−1.16	−2.1	−7.6	−6.1
1-2 (1)	−10.9	−7.3	−6.9	−6.2
2-1	−9.5	−14.6	−4.0	−7.3
2.2 (5)	−14.6	−9.9	−12.0	−9.2
4-1 (3)	−9.1	−8.5	−8.1	−8.0
5-1 (4)	−13.8	−7.9	−7.8	−8.8
6-1 (7)	−10.6	−9.1	−7.9	−6.3
7-1	−14.8	−8.5	−10.5	−7.4
8-1	−15.7	−9.3	−9.6	−6.7
9-1	−9.7	−2.8	−6.9	−5.9
10-1	−7.1	−6.4	−10.7	−6.9
11-1 (6)	−14.3	−12.4	−11.0	−7.3
12-1	−16.1	−7.7	−12.3	−6.8
13-1	−12.8	−8.9	−6.9	−6.7
14-1	−14.2	−4.2	−7.3	−3.6
15-1	−9.1	−7.4	−7.3	−5.3
16-1	−11.7	−12.3	−6.7	−8.5
17-1	−11.7	−4.5	−9.8	−5.9
18-1	−9.1	−8.2	−8.1	−6.8
18-3	−3.5	−11.3	−4.3	−6.5
18-5	−8.0	−8.3	−9.0	−8.2
18-6	−9.4	−7.6	−8.5	−9.5
19-1	−11.9	−5.2	−9.1	−5.1
19-2	−10.7	−10.3	−8.1	−7.7
20-1	−13.5	−11.1	−13.3	−10.4
Average[c]	−10.9 ± 3.5	−8.2 ± 2.9	−8.5 ± 2.2	−7.0 ± 1.5
UC (2)	−4.4	−3.7	−0.8	−1.6

Structures were ascertained using mFOLD.[45]

[a] Minus-strand sequences were shuffled using a randomization program from Arizona Research Labs. The ΔG values shown are the average of five independent randomizations. The complements of the (−)-strand randomized sequences were also folded to give the averaged (+)-strand randomized values.

[b] Names followed by parentheses are second round winners and numbers in parentheses denote order of fitness from Table 3.

[c] Averages were calculated using all 28-base replacement sequences and thus excludes UC, which contains a seven-base replacement sequence.

than 4-1 and 5-1 in direct competition assays (Table 3). Levels of virions isolated from equal numbers of infected protoplasts varied markedly depending on the co-inoculated satRNA. No virions were detected in Western blots when TCV was co-inoculated with wt satC. Levels of virions associated with UC co-inoculation were reduced by 84%, similar to the 73% reduction by the comparably replicating satC*. This result suggests that UC is equally capable of virion reduction as this variant satC containing a wt M1H. Virion accumulation in the presence of 4-1 and 5-1 was only reduced by 48% and 45%, respectively, compared with levels in the absence of satRNA, even though these satRNA accumulated ~twofold better than UC and satC*. These results suggest that the seven-base replacement sequence in UC contributes to the fitness of the satRNA by interfering with virion accumulation, thus enhancing TCV movement.

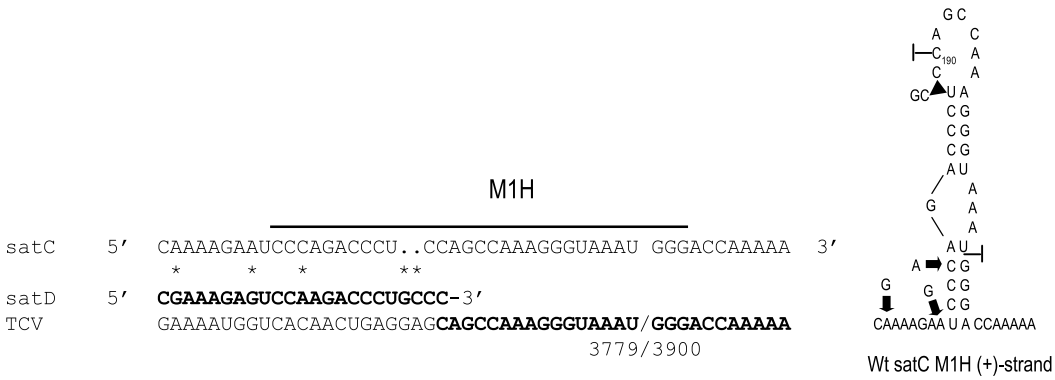

Figure 6. Evolution of the M1H from the original parental sequences. The M1H is comprised of three non-contiguous sequences: the 3′ end of satD and two segments from TCV. Sequences in bold are the original parental sequences and asterisks denote position differences with wt satC. These differences are also shown at the right on the (+)-strand structure of the M1H. Lines at or between bases in the structure denote the crossover sites. The cytidylate at position 190 could be derived from either satD or TCV.

Structures of the 28-base M1H replacement sequences

We previously determined that the structure of the M1H was essential for recombination *in vivo*.[31,33] The M1H is predicted to form similar structures on both (+) and (−)-strands according to computer predictions using the mFOLD program[45] and biochemical structural determination[41,46] (Figures 1(D) and 5). To ascertain if the M1H replacement sequences of the SELEX winners might also form hairpins in (+) and/or (−)-strands, all 25 replacement sequences (both (+) and (−)-orientation) recovered in the first round (along with 11 upstream and eight downstream bases) were subjected to computer analysis using mFOLD. In addition, all (−)-sense sequences were randomized using the Shuffle program from Arizona Research Labs and the randomized sequences along with their complements also examined for possible hairpins. Unexpectedly, as shown in Table 4, (+)-strand structures of the replacement sequences were, on average, 33% more stable than their (−)-strand complements. Furthermore, 80% of the structures were more stable in their (+)-sense orientation than in their (−)-sense orientation, including all second round winners. The hairpin formed by the UC replacement sequence was also 19% more stable in its (+)-sense orientation (−4.4 kcal/mol compared with −3.7 kcal/mol). These results suggest that a hairpin in the (+)-strand might be more significant to satC fitness than a corresponding hairpin in the (−)-strand. In addition, 67% of the (+)-strand hairpins, including that of UC, formed such that the A-rich sequences at the base maintained their single-stranded structure, similar to the wt (+)-strand M1H (Figure 5).

Discussion

DNA enhancers are defined as position-independent sequences that are functional in both orientations, which reduce, but do not abolish transcription when removed.[47] This definition also describes the M1H recombination hotspot/enhancer of satC. The M1H is active as an enhancer when located distal to its original position or when reversed in orientation, and eliminating the M1H reduces, but does not abolish, RNA replication.[29] As described in the Introduction, analysis of the M1H over the past 12 years indicates that the M1H functions in recombination and as an enhancer in its (−)-sense orientation.[29,31,33,48]

The M1H was originally derived from three non-contiguous sequences: the 3′ end of satD, 16-bases from the TCV CP coding sequence, and the remainder from the TCV 3′ untranslated region (Figure 6).[49] Based on analysis of the parental sequences, five changes have occurred since the original recombination events that formed satC, all in the satD-derived sequence: the first two changes (moving 5′ to 3′ in the (+)-strand) increased the number of adenylates in the A-rich sequence at the 5′ base of the hairpin; the next change strengthened the hairpin stem and the final two-base deletion strengthened the hairpin stem and also resulted in an eight-base sequence identity (in the (−)-sense orientation) with the 5′PE. The (−)-sense M1H contains two additional motifs found in elements located on satC (−)-strands known to be important for (+)-strand synthesis (Figure 1(D)). In addition, the 3′ end of one M1H motif, 3′CCCUGG, is precisely the junction of the second recombination event that led to the formation of satC. The identical motif is also found in the 5′PE of satC, which can serve as an independent promoter *in vitro*.[28] Since the consecutive cytidylates in this 5′PE motif are absolutely required for promoter activity,[28] it is likely that the RdRp was attracted to this identical motif in TCV to re-initiate synthesis during the crossover event that formed satC. Thus, the TCV sequences responsible for formation of satC by RdRp-mediated recombination are likely those responsible for the replication enhancer properties of the M1H.

The discovery that the (−)-strand M1H contained motifs found in other (−)-strand promoter elements in TCV and TCV-associated RNAs prompted initiation of the SELEX study, to determine if satC fitness in plants was related to M1H replacement sequences containing similar short motifs. The M1H replacement sequences in (−)-strands of the first round winners of the 28-base SELEX were highly variable but not random. Short three to four consecutive cytidylates were present in the replacement sequences at three to fivefold more than expected for a purely random sequence. Short stretches of consecutive cytidylates are also present in all TCV promoter elements, including the wt (−)-strand M1H (Figure 1). As with the wt M1H, many of the (−)-strand 28-base replacement sequences were composed of several short six to ten-base motifs found in (−)-strand TCV promoter elements (Figure 2), the most common being the CCS motif found in the 3′PE (Figure 2, class I). The CCS motifs located at the 3′ ends of satC and TCV (−)-strands also were represented in several SELEX winners (Figure 2, class II). Since the most fit of the SELEX winners (1-2, 4-1, 5-1 and 2-2) replicated between 1.5 and 3.7-fold better than the average of two non-selected satC with 28-base replacement sequences (R28a and R28b; Figure 3), and all contained at least one of the identified motifs, it is likely that these motifs contributed to the enhanced replication of the SELEX winners.

Since the reverse complement of the M1H is also able to enhance replication of satC *in vivo* and transcription *in vitro*,[29] we examined the reverse complement sequence for similar short motifs. This sequence, 3′CCAGGGUAAAUGGGAAACCGAC CUCCCAGA*CCCUA*AG5′, (presented in 3′ to 5′ orientation to better compare with the presentation of sequences in Table 1 and Figure 2), also contains several stretches of contiguous cytidylates (in italics) and the motif CCCUA (underlined), which is the 3′CCS of satC. While the (+)-sense orientation of the M1H is functional as a replication enhancer, previous *in vitro* results indicated that the M1H in either orientation provided less than twofold enhancement when combined with the hairpin promoter at the 3′ end of satC (+)-strands (compared with nearly tenfold enhancement using a (−)-sense promoter). Therefore, the M1H's replication enhancer role is likely in its (−)-sense orientation, to enhance (+)-strand synthesis.

The replication enhancer property of the M1H is thus consistent with the presence of different short promoter-like motifs, which can be located throughout the sequence accounting for the variability of the SELEX results. These results also suggest that a linear comparison of sequence may not reveal similarities that may exist between promoter elements for other viral RdRp. Other studies using an *in vivo* functional SELEX approach to analyze a protein binding sequence have also recovered a large number of different functional sequences. Six classes of sequences containing short six to ten-base motifs that could serve as splicing enhancers by binding to a family of proteins called SR, were selected at high frequency from a random population of RNA sequences *in vivo*.[38] While some recovered sequences resembled natural purine-rich SR-binding sequences, other recovered sequences were novel and non-purine rich. These recovered sequences were significantly more degenerate than sequences recovered using *in vitro* SELEX and selecting only for SR protein binding.[50]

One reason for the recovery of diverse sequences following *in vivo* SELEX is that, unlike *in vitro* SELEX, where selection is for a single attribute (e.g. binding to a particular protein), *in vivo* SELEX does not presume a unique function for the selected sequence. Thus, it is possible that the selected sequences might serve more than a single function, which could account for variability in the recovered population. While we initiated *in vivo* functional SELEX of the M1H with the presumption that sequences recovered would substitute for the missing replication enhancer, a second *in vivo* function for the M1H region was suggested by the recovery of clone UC. UC, with a seven-base replacement sequence did not replicate better in protoplasts than randomly selected clone R28a, which contains a 28-base replacement sequence (Figure 3). UC did replicate 1.9-fold better than satC with a ten-base randomly selected replacement sequence and eightfold better than satC with a deletion of the M1H, suggesting that the UC replacement sequence may be weakly enhancing replication. However, UC replicated more poorly than clones 4-1, 5-1, 2-2 and 6-1 while being more fit than these satRNAs in direct competition assays in plants. An explanation for the unusual fitness is that UC, as with wt satRNA, is able to strongly interfere with virion accumulation (Figure 4), which correlates with more rapid helper virus trafficking to younger tissue (F. Zhang, Q. Kong, and A.E.S., unpublished results). When assaying equal numbers of infected protoplasts, second round winners 4-1 and 5-1 accumulated 46% better than UC, yet 2.3-fold fewer virions were isolated from UC-infected cells. Since the only difference between the satRNAs was the M1H replacement sequence, the M1H, possibly in conjunction with surrounding sequences, or its complement comprises the movement enhancer that increases fitness. These results suggest that the M1H region is multifunctional, serving as both an enhancer of replication and an enhancer of viral movement.

Previous results indicated that the structure of the M1H in (−)-strands was important for recombination *in vivo*[31] and the analogous primer extension reaction *in vitro*.[33,51] Computer structural modeling of the M1H replacement sequences here, however, suggest a role for a (+)-strand hairpin in basic satRNA fitness. The seven-base UC replacement sequence, both in full-length satRNA and a portion consisting of 11 upstream and eight downstream bases, is predicted to form the stem of a

short hairpin (Figure 5) that, like the wt M1H, contains A-rich ((+)-sense) or U-rich ((−)-sense) sequences at the base of the stem. The (+)-sense UC hairpin is nearly 20% more stable than the complementary hairpin, suggesting that a (+)-strand hairpin in this location might be important for fitness (Table 4). Hairpin stability predictions for all first round SELEX replacement sequences revealed that 80% of the clones contain sequences with (+)-strand structures more stable than their (−)-strand complements. Furthermore, all second round SELEX winners contain hairpins that are between 7% and 74% more stable in their (+)-strand orientation.

Based on the evidence that selection of satC to accumulate in plants favors (+)-strand hairpins flanked by single-stranded A-rich sequences, and that (+)-strands accumulate to levels >100-fold higher than (−)-strands in infected cells (A.E.S., unpublished), it is likely that the (+)-strand hairpin flanked by A-rich single-stranded sequence comprises the movement enhancer. Since the composition of the stems and loops of the hairpins varied considerably in size, sequence, and stability, it is possible that the function of the hairpin is to keep the A-rich sequences single-stranded. How this structure and flanking sequence might influence virion accumulation is not known. A specific packaging sequence that folds into a hairpin near the 3′ end of the CP coding region has been identified as the assembly initiation site for the TCV capsid.[52] However, there are no obvious sequence interactions between the M1H region and this packaging signal. Interestingly, TCV genomic RNA contains a (+)-strand hairpin (H4), also flanked by A-rich sequences, which is similar to the (+)-strand M1H in both structure and proximity to 3′ end of the RNA (Figure 5). The 3′ base of TCV H4 is the 3′ recombination junction site of satC, and thus this sequence, and the A-rich downstream region, comprises the 3′ base and flanking sequences of the (+)-strand M1H (Figure 5, boxed). The A-rich sequence at the 5′ side of H4 is unrelated to the A-rich sequence at the 5′ side of the M1H. Hairpin 4 and its flanking sequences are highly conserved in the related *Carmovirus*, *Cardamine Chlorotic Fleck Virus*, differing only by a single change in the loop and a C:G to G:C reversal in the stem. We are currently determining if the U-rich sequence in the H4 loop interacts with the A-rich sequence at the base of the hairpin and if this interaction is disrupted by the presence of similar A-rich sequences in satC.

In conclusion, our results indicate that a single region of satC has evolved to both enhance replication and interfere with virion formation. While multifunctional proteins are well known to exist for viruses,[53] multifunctional RNA sequences are only beginning to be defined. The 250 base intergenic region in RNA 3 of *Brome Mosaic Virus* is the subgenomic promoter in (−)-strands and on (+)-strands is involved in membrane sequestration mediated by the 1a protein.[16] The (−)-strand complement of the internal ribosome entry site (IRES) of hepatitis C virus is involved in mRNA transcription, viral replication and binding to the hepatitis C RdRp.[54] A second multifunction sequence in satC also exists. The 5′PE on satC (−)-strands is important for complementary strand synthesis in the absence of the 3′PE *in vitro*[2] and comprises part of a stem on a (+)-strand hairpin that is critical for virus replication (J. Zhang, R. Stuntz and A.E.S., unpublished). The low probability of a series of recombination/alteration events occurring necessary to produce the multifunctional replication and movement enhancer of satC is reflected in our 12 year inability to detect any other recombinant satRNA capable of amplification to detectable levels in TCV-infected cells even though recombination events between satD and TCV, similar to those that created satC, occur at high frequency in infected cells.[41,55]

Materials and Methods

In vivo genetic selection using 28-base M1H replacement sequences

In vivo genetic selection was performed as described.[26,27] To generate a population of full-length satC containing randomized bases in positions 181 to 208, two separate PCR were carried out using pT7C +, a full-length cDNA clone of wt satC. To generate the satC 5′ fragment, primers T7C5′ (5′-GTAATACGACTCA CTATAGGGATAACTAAGGG-3′), which contains a T7 polymerase promoter at the 5′ end, and SEL5′ (5′-GACT GGATCCTTTTGAGTGGGAAACAG-3′) were used. For generating the 3′ fragment containing the M1H randomized bases, SEL3′ (5′-GACTGGATCCNNNNNNNNNN NNNNNNNNNNNNNNNNNNACCAAAAACGGCG GCAGCACC-3′) and oligo7 (5′-GGGCAGGCCCCCCGT CCGA-3′) were used. These two PCR were designed to generate a new *Bam* HI site in the satC cDNA by creating a U to C alteration at position 176 of satC. This new site was required to link the two fragments to avoid cloning and thus reduction of the complexity of the population. Both the 5′ and 3′ PCR fragments were digested with *Bam* HI, purified through an agarose gel and then ligated. SatC transcripts were directly synthesized from the ligated product using T7 RNA polymerase and 5 μg used to inoculate each of 19 turnip seedlings along with 2 μg/plant of TCV transcripts. Total RNA was extracted from uninoculated leaves at 21 days post-inoculation. SatC accumulating in plants was amplified by RT-PCR using primers T7C5′ and oligo7, cloned into the *Sma* I site of pUC19, and then sequenced. For the second round selection, equal amounts of transcripts from all cloned first round winners were combined and used to inoculate six turnip seedlings. Total RNA was extracted at 21 days post-inoculation and satC cloned as described above.

Fitness of second round 28-base SELEX winners to accumulate in plants

To compare fitness of the second round winners to accumulate in turnip, transcripts of SELEX winners were synthesized by T7 RNA polymerase and equal amounts of transcripts from two or three winners

combined and inoculated onto three turnip seedlings (0.2 μg/plant) together with TCV transcripts (2 μg/plant). Total RNA was extracted at 21 days post-inoculation and satC cloned and sequenced as described above.

Protoplast inoculations and RNA gel-blot analysis

Protoplasts were prepared from callus cultures of *A. thaliana* ecotype Col-0, as described.[56] Protoplasts (5×10^6) were inoculated with 20 μg of TCV genomic RNA transcripts and 2 μg of transcripts of various satC constructs using polyethelene glycol as described.[56] Equal amounts of total RNA isolated from protoplasts at various times post-inoculation were used for RNA gel-blot analysis of TCV and satC.[22] Oligo13 (5′-AAAGAGCACTAGTTTTCCAG-3′), which is complementary to positions 3950–3970 of TCV genomic RNA and positions 250–269 of satC, was labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and used as the probe. SatC with non-selected random sequences replacing the M1H contained the following replacement sequences: R28a, 3′GUAUAGAACUUACAUCCUUCCCUAUUAU5′; R28b, 3′AAAUAGCACGGAGACCAUCCCCCAAUAA5′; R10, 3′AUGUUAUUAG5′.

Stability of satC with different 28-base M1H replacement sequences

Sat-RNA stability in protoplasts was determined as described.[29] Briefly, protoplasts were inoculated with satC in the absence of TCV and levels at one to six hours post-inoculation determined by RNA gel blots after treatment of cells with RNase A to remove any unabsorbed satC transcripts. The probe was [α-^{32}P]UTP-labeled (−)-strands of satC transcribed by T7 RNA polymerase from *Dra* I-linearized pT7C(−), which contains full-length satC cDNA downstream of a T7 RNA promoter.[57]

Virion isolation and Western blots

Virus particles were isolated from infected protoplasts and analyzed as described.[52] At 40 hours post-inoculation with TCV and various satRNA, protoplasts (2.5×10^6) were collected and resuspended in 200 μl of 0.2 M sodium acetate (pH 5.2). Sterile glass beads (30 μl, 0.1 mm diameter, Biospec Products, Inc.) in the same buffer were added, and the mixture was vortexed. The aqueous phase was recovered after brief microcentrifugation, and the solid phase was re-extracted twice. The aqueous phases were combined and incubated on ice for 60 minutes followed by centrifugation to remove cell debris. Virions were precipitated by addition of 0.25 vol. of 40% polyethlene glycol (M_r 8000)/1 M NaCl, incubated overnight on ice, and then collected by microcentrifugation. Pelleted virions were dissolved in 15 μl of 0.01 M sodium acetate (pH 5.5), and analyzed by electrophoresis through 1% (w/v) agarose gels in 50 mM Tris–base/38 mM glycine (pH 8.3). Following electrophoresis, gels were soaked in 50 mM sodium hydroxide for 20 minutes and then transferred to 0.2 M sodium acetate (pH 5.5), for 20 minutes. Disrupted virions were blotted to NitroPlus membranes (Micron Separations Inc.) and probed with a polyclonal antiserum raised against TCV coat protein. Chemiluminescent staining was performed with the Western Lighting Chemiluminescence Reagent kit (Perkin Elmer Life Sciences).

Acknowledgements

This work was supported by grants from the US Public Health Service (GM61515-01) and the National Science Foundation (MCB-0096274 and MCB-0086952) to A.E.S.

References

1. 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.
2. Guan, H., Song, C. & Simon, A. E. (1997). RNA promoters located on (−)-strands of a subviral RNA associated with turnip crinkle virus. *RNA*, **3**, 1401–1412.
3. Koev, G. & Miller, W. A. (2000). A positive-strand RNA virus with three very different subgenomic RNA promoters. *J. Virol.* **74**, 5988–5996.
4. Blight, K. J. & Rice, C. M. (1997). Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA. *J. Virol.* **71**, 7345–7352.
5. Dreher, T. W. (1999). Functions of the 3′-untranslated regions of positive strand RNA viral genomes. *Annu. Rev. Phytopathol.* **37**, 151–174.
6. Kao, C. C., Singh, P. & Ecker, D. J. (2001). *De novo* initiation of viral RNA-dependent RNA synthesis. *Virology*, **287**, 251–260.
7. Koev, G., Liu, S. J., Beckett, R. & Miller, W. A. (2002). The 3′-terminal structure required for replication of Barley yellow dwarf virus RNA contains an embedded 3′ end. *Virology*, **292**, 114–126.
8. Wu, B., Vanti, W. B. & White, K. A. (2001). An RNA domain within the 5′ untranslated region of the Tomato bushy stunt virus genome modulates viral RNA replication. *J. Mol. Biol.* **305**, 741–756.
9. Todd, S., Towner, J. S., Brown, D. M. & Semler, B. L. (1997). Replication-competent picornaviruses with complete genomic RNA 3′ non-coding region deletions. *J. Virol.* **71**, 8868–8874.
10. Ray, D. & White, K. A. (1999). Enhancer-like properties of an RNA element that modulates tombusvirus RNA accumulation. *Virology*, **256**, 162–171.
11. Haldeman-Cahill, R., Daros, J-A. & Carrington, J. C. (1998). Secondary structures in the capsid protein coding sequence and 3′ nontranslated region involved in amplification of the tobacco etch virus genome. *J. Virol.* **72**, 4072–4079.
12. Sullivan, M. L. & Ahlquist, P. (1999). A brome mosaic virus intergenic RNA3 replication signal functions with viral replication protein 1a to dramatically stabilize RNA *in vivo*. *J. Virol.* **73**, 2622–2632.
13. Paul, A. V., Rieder, E., Kim, D. W., van Boom, J. H. & Winner, E. (2000). Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the *in vitro* uridylylation of VPg. *J. Virol.* **74**, 10359–10370.
14. Rieder, E., Paul, A. V., Kim, D. W., van Boom, J. H. & Winner, E. (2000). Genetic and biochemical studies of poliovirus *cis*-acting replication element cre in

relation to VPg uridylylation. *J. Virol.* **74**, 10371–10380.
15. Quadt, R., Ishikawa, M., Janda, M. & Ahlquist, P. (1995). Formation of brome mosaic virus RNA-dependent RNA polymerase in yeast requires coexpression of viral proteins and viral RNA. *Proc. Natl Acad. Sci. USA*, **92**, 4892–4896.
16. Schwartz, M., Chen, J. B., Janda, M., Sullivan, M., den Boon, J. & Ahlquist, P. (2002). A positive-strand RNA virus replication complex parallels form and function of retrovirus capsids. *Mol. Cell*, **9**, 505–514.
17. Simon, A. E. (1999). Replication, recombination, and symptom-modulation properties of the satellite RNAs of turnip crinkle virus. *Cur. Top. Microbiol. Immunol.* **239**, 19–36.
18. Carrington, J. C., Heaton, L. A., Zuidema, D., Hillman, B. I. & Morris, T. J. (1989). The genome structure of *Turnip crinkle virus*. *Virology*, **170**, 219–226.
19. Hacker, D. L., Petty, I. T. D., Wei, N. & Morris, T. J. (1992). Turnip crinkle virus genes required for RNA replication and virus movement. *Virology*, **186**, 1–8.
20. Oh, J.-W., Kong, Q., Song, C., Carpenter, C. D. & Simon, A. E. (1995). Open reading frames of turnip crinkle virus involved in satellite symptom expression and incompatibility with *Arabidopsis thaliana* ecotype Dijon. *Mol. Plant-Microbe Interact.* **8**, 979–987.
21. Li, W. Z., Qu, F. & Morris, T. J. (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.
22. Wang, J. & Simon, A. E. (1997). Analysis of the two subgenomic RNA promoters for turnip crinkle virus *in vivo* and *in vitro*. *Virology*, **232**, 174–186.
23. Wang, J. & Simon, A. E. (1999). Symptom attenuation by a satellite RNA *in vivo* is dependent on reduced levels of virus coat protein. *Virology*, **259**, 234–245.
24. Song, C. & Simon, A. E. (1995). Requirement of a 3′-terminal stem-loop in *in vitro* transcription by an RNA-dependent RNA polymerase. *J. Mol. Biol*, **254**, 6–14.
25. Stupina, V. & Simon, A. E. (1997). Analysis *in vivo* of turnip crinkle virus satellite RNA C variants with mutations in the 3′ terminal minus strand promoter. *Virology*, **238**, 470–477.
26. Carpenter, C. D. & Simon, A. E. (1998). Analysis of sequences and putative structures required for viral satellite RNA accumulation by *in vivo* genetic selection. *Nucl. Acids Res.* **26**, 2426–2432.
27. Guan, H., Carpenter, C. D. & Simon, A. E. (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.
28. Guan, H., Carpenter, C. D. & Simon, A. E. (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.
29. Nagy, P. D., Pogany, J. & Simon, A. E. (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.
30. Nagy, P. D., Pogany, J. & Simon, A. E. (2001). *In vivo* and *in vitro* characterization of an RNA replication enhancer in a satellite RNA associated with Turnip crinkle virus: comparison of sequences and structures stimulating primer-dependent and primer-independent RNA synthesis. *Virology*, **288**, 315–324.
31. Cascone, P. J., Haydar, T. & Simon, A. E. (1993). Sequences and structures required for RNA recombination between virus-associated RNAs. *Science*, **260**, 801–805.
32. Nagy, P. D. & Simon, A. E. (1998). *in vitro* characterization of late steps of RNA recombination in turnip crinkle virus I: role of the motif1-hairpin structure. *Virology*, **249**, 379–392.
33. Nagy, P. D., Zhang, C., Simon, A. E. & Dissecting, R. N. A. (1998). recombination *in vitro*: role of RNA sequences and the viral replicase. *EMBO J.* **17**, 2392–2403.
34. Rajendran, D. S., Pogany, J. & Nagy, P. D. (2002). Comparison of Turnip crinkle virus RNA-dependent RNA polymerase preparations expressed in *Escherichia coli* of derived from infected plants. *J. Virol.* **76**, 1707–1717.
35. Cascone, P. J. (1992). RNA recombination in the Turnip crinkle virus system: an analysis of sequences and structures required for RNA recombination. PhD Thesis, University of Massachusetts.
36. Wilson, D. S. & Szostak, J. W. (1999). *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647.
37. Tian, H. H., Kole, R. & Strong, R. N. A. (2001). splicing enhancers identified by a modified method of cycled selection interact with SR protein. *J. Biol. Chem.* **276**, 33833–33839.
38. Schaal, T. D. & Maniatis, T. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* **19**, 1705–1719.
39. Lund, M. & Kjems, J. (2002). Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end. *RNA*, **8**, 166–179.
40. Li, X. H. & Simon, A. E. (1990). Symptom intensification on cruciferous hosts by the virulent sat-RNA of turnip crinkle virus. *Phytopathology*, **80**, 238–242.
41. Carpenter, C. D., Oh, J.-W., Zhang, C. & Simon, A. E. (1995). Involvement of a stem-loop structure in the location of junction sites in viral RNA recombination. *J. Mol. Biol.* **245**, 608–622.
42. Carpenter, C. D., Cascone, P. J. & Simon, A. E. (1991). Formation of multimers of linear satellite RNAs. *Virology*, **183**, 586–594.
43. Li, X. H. & Simon, A. E. (1991). *In vivo* accumulation of a turnip crinkle virus DI RNA is affected by alterations in size and sequence. *J. Virol.* **65**, 4582–4590.
44. Zhang, C. & Simon, A. E. (1994). Effect of template size on replication of defective interfering RNAs. *J. Virol.* **68**, 8466–8469.
45. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940.
46. Song, C. (1995). *In vitro* study of RNA replication using isolated viral RNA-dependent RNA polymerase. PhD Thesis, University of Massachusetts.
47. Hertel, K. J., Lynch, K. W. & Maniatis, T. (1997). Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* **9**, 350–357.
48. Cascone, P. J., Carpenter, C. D., Li, X. H. & Simon, A. E. (1990). Recombination between satellite RNAs of turnip crinkle virus. *EMBO J.* **9**, 1709–1715.
49. 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.

50. Tacke, R. & Manley, J. L. (1999). Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* **11**, 358–362.
51. Song, C. & Simon, A. E. (1995). Synthesis of novel products *in vitro* by an RNA-dependent RNA polymerase. *J. Virol.* **69**, 4020–4028.
52. Qu, F. & Morris, T. J. (1997). Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size. *J. Virol.* **71**, 1428–1435.
53. Callaway, A., Giesman-Cookmeyer, D., Gillock, E. T., Sit, T. L. & Lommel, S. A. (2001). The multifunctional capsid proteins of plant RNA viruses. *Annu. Rev. Phytopathol.* **39**, 419–460.
54. Kashiwagi, T., Hara, K., Kohara, M., Iwahashi, J., Hamada, N., Honda-Yoshino, H. & Toyoda, T. (2002). Promoter/origin structure of the complementary strand of Hepatitis C virus genome. *J. Biol. Chem.* **277**, 28700–28705.
55. Zhang, C., Cascone, P. J. & Simon, A. E. (1991). Recombination between satellite and genomic RNAs of turnip crinkle virus. *Virology*, **184**, 791–794.
56. Kong, Q., Wang, J. & Simon, A. E. (1997). Satellite RNA-mediated resistance to turnip crinkle virus in Arabidopsis involves a reduction in virus movement. *Plant Cell*, **9**, 2051–2063.
57. Song, C. & Simon, A. E. (1994). RNA-dependent RNA polymerase from plants infected with turnip crinkle virus can transcribe (+)- and (−)-strands of virus-associated RNAs. *Proc. Natl Acad Sci. USA*, **91**, 8792–8796.

Edited by N.-H. Chua

(Received 11 September 2002; received in revised form 20 November 2002; accepted 22 November 2002)