

Recombination between satellite RNAs of turnip crinkle virus

Pamela J. Cascone¹, Clifford D. Carpenter²,
Xiao Hua Li² and Anne E. Simon^{1,2}

¹Program in Molecular and Cellular Biology and ²Department of Plant Pathology, University of Massachusetts, Amherst, MA 01003, USA

Communicated by A. van Kammen

Turnip crinkle virus (TCV) is associated with satellite (sat) RNAs (sat-RNA D, sat-RNA F), defective interfering (DI) RNAs (DI RNA G, DI1 RNA), and one RNA with properties of both sat-RNAs and DI RNAs (sat-RNA C). When plants were inoculated with TCV, sat-RNA D and *in vitro* sat-RNA C transcripts containing non-viable mutations in the 5' domain, recombinant sat-RNAs were recovered. These recombinants were composed of sat-RNA D at the 5' end and sat-RNA C sequences at the 3' end. Analysis of 20 independent recombination junctions revealed that unequal crossing-over had occurred *in planta* in a region of sequence similarity between the two sat-RNAs which resulted in the duplication of 3–16 nucleotides. Thirty percent of the sat-RNA recombinants also had one to three additional nucleotides inserted at the crossover junctions which did not correspond to either sat-RNA C or sat-RNA D sequence. The right side of the recombination junctions always began with one of three consecutive nucleotides of sat-RNA C. Based on the similarity between this sequence of sat-RNA C, the right side junction of DI RNA G and the 5' end of TCV, as well as the sequence similarity between right side junctions of DI1 RNA and sat-RNA C and the 5' end of the sat-RNAs, a replicase-driven copy choice mechanism is proposed. The replicase, while replicating viral or sub-viral minus strands, can dissociate from the template along with the nascent plus strand and reinitiate synthesis at one of two internal replicase recognition sequences on the same or different template thus generating recombinant sat-RNAs or DI RNAs.

Key words: defective interfering RNAs/plant virus/RNA recombination/satellite RNAs/viral replication

Introduction

Recombination among RNA genomes has been described for a number of viruses including picornaviruses (Cooper *et al.*, 1974; King *et al.*, 1982) and coronaviruses (Lai *et al.*, 1985). RNA recombination can be either homologous, where the two parental genomes are related and the crossover event occurs at the same location in both RNAs thereby preserving any open reading frames, or non-homologous, where these restrictions do not apply (for a review, see King, 1988). Homologous recombination is very precise; analysis of 46 crossover sites in picornavirus recombinants revealed no insertions, deletions or mismatched bases (King, 1988). Furthermore, homologous recombination is not sequence-

or structure-specific; 17 different crossover events between foot and mouth disease virus (FMDV) subtypes occurred randomly throughout the genome (King *et al.*, 1985), and 13 independent poliovirus recombinants revealed neither striking sequence specificity nor a preference for recombination occurring in regions of high sequence similarity (Kirkegaard and Baltimore, 1986). An enzymatic cutting and re-ligation model and a replicase-driven 'copy choice' model have been proposed to explain both types of recombination. According to the copy choice model, the replicase complexed with the nascent RNA strand is released from the template being copied, and synthesis is reinitiated at a new site on a second viral template (Cooper *et al.*, 1974).

In plant RNA viruses, recombination has only been conclusively demonstrated for brome mosaic virus (Bujarski and Kaesberg, 1986), although it is also thought to occur in tobnaviruses (Robinson *et al.*, 1987; Angenent *et al.*, 1989). Recombination in brome mosaic virus, a tripartite, positive (+) sense virus, occurs between similar sequences in the 3' region of each viral RNA, and can result in the duplication of sequences at the crossover junction. In this paper, we present evidence for RNA recombination between two sub-viral RNAs associated with turnip crinkle virus (TCV).

Turnip crinkle is a single-stranded, positive (+) sense virus which has been classified as a member of the carmoviruses (Morris and Carrington, 1988). Unlike other animal or plant viruses examined to date, TCV supports the replication of a variety of linear small RNAs: (i) satellite (sat) RNAs (Altenbach and Howell, 1981; Simon and Howell, 1986) which are small, single-stranded molecules, do not function as mRNAs, and require a helper virus for replication and propagation; (ii) defective interfering (DI) RNAs (Li *et al.*, 1989), derived almost exclusively from helper virus sequence; (iii) chimeric RNAs (Simon and Howell, 1986), composed of sat-RNA sequence at the 5' end and viral sequence at the 3' end (Figure 1). TCV isolate TCV-M supports two avirulent sat-RNAs (sat-RNA D, 194 bases; sat-RNA F, 231 bases) which differ mainly by a 36 base insert located near the 3' end of sat-RNA F, and one chimeric RNA (sat-RNA C, 356 bases), which produces severe symptoms when inoculated along with helper virus on a variety of host plants (Li and Simon, 1990). The 5' 190 bases of sat-RNA C share 88% similarity with full-length sat-RNA D while the 3' 166 bases share 94% similarity with two regions at the 3' end of TCV (Simon and Howell, 1986). DI RNA G (346 bases), normally associated with the TCV-B isolate, is a mosaic molecule composed of 21 nucleotides of unknown origin at the 5' end as well as 5' and 3' viral segments and a centrally repeated block of viral sequence (Li *et al.*, 1989). DI1 RNA (383 bases) was generated *de novo* following inoculation of plants with material derived from *in vitro* synthesized viral transcripts (TCV-B isolate), and contains the exact 5' and 3' ends of TCV as well as an internal viral sequence (Li *et al.*, 1989). Both monomeric

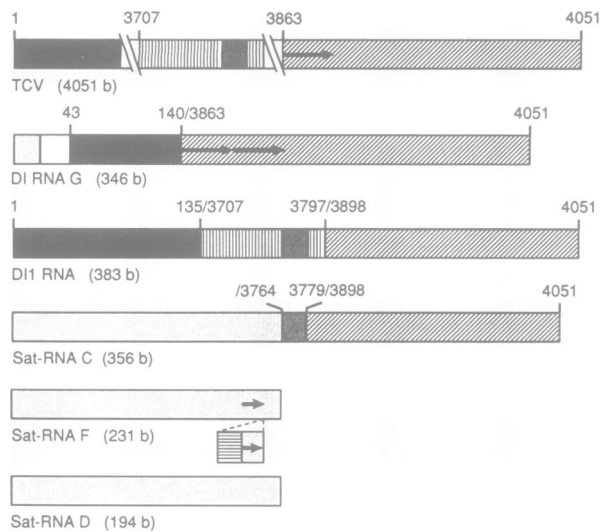


Fig. 1. Schematic representation of sequence similarities between TCV and associated RNAs. TCV isolate TCV-M (formerly TCV WT) contains sat-RNAs C, D and F (Simon and Howell, 1986). Isolate TCVB (formerly TCV-T) is associated with sat-RNA D and DI RNA G. DI1 RNA was generated *de novo* after inoculating turnip with cloned viral material (Li *et al.*, 1989). Similar sequences are shaded alike. Arrows of similar size represent repeated sequences. Numbering in the DI RNAs and sat-RNA C refer to the positions of TCV-related sequence. The sequence of the viral genomic RNA has only been determined for TCV-B (Carrington *et al.*, 1989) and it is not known what differences, if any, exist between the two isolates.

and multimeric forms of the sat-RNAs and DI RNAs are normally found in infected tissue.

In this paper, we report the second definitive example of recombination in a plant virus system; the *in planta* generation of recombinants between sat-RNA D and the similar sequences in the 5' domain of sat-RNA C. Crossover events occur in a region which is related, but not identical, in the two RNAs. Based on the location of 20 independent recombination sites, and the previously determined sequences at the junction sites in the DI RNAs and sat-RNA C, we have formulated a model which accounts for the production of sat-RNA recombinants. Our model involves a copy choice mechanism coupled with recognition by the replicase of specific initiation signals on minus (-) strand RNA templates. Such a mechanism also explains the origin of the chimeric sat-RNA C and the formation of DI RNAs.

Results

Recombinant molecules are formed *in vivo* between sat-RNA D and sat-RNA C sequences

A series of mutations have been constructed in the virulent sat-RNA C in order to identify sequences involved in sat-RNA processes such as replication and pathogenicity (Simon *et al.*, 1988; Carpenter, Cascone and Simon, unpublished). All mutations were incorporated into pPM2-2-47M+, a plasmid containing a full-length, monomeric copy of sat-RNA C downstream from a modified *Escherichia coli* RNA polymerase promoter (Ahluquist and Janda, 1984). Transcripts of sat-RNA C synthesized *in vitro* are infectious when inoculated with helper inoculum (HVI) on turnip cv. Just Right (Simon and Howell, 1987). The helper virus inoculum contains both genomic TCV (from isolate TCV-M) and sat-RNA D, an avirulent sat-RNA which has been

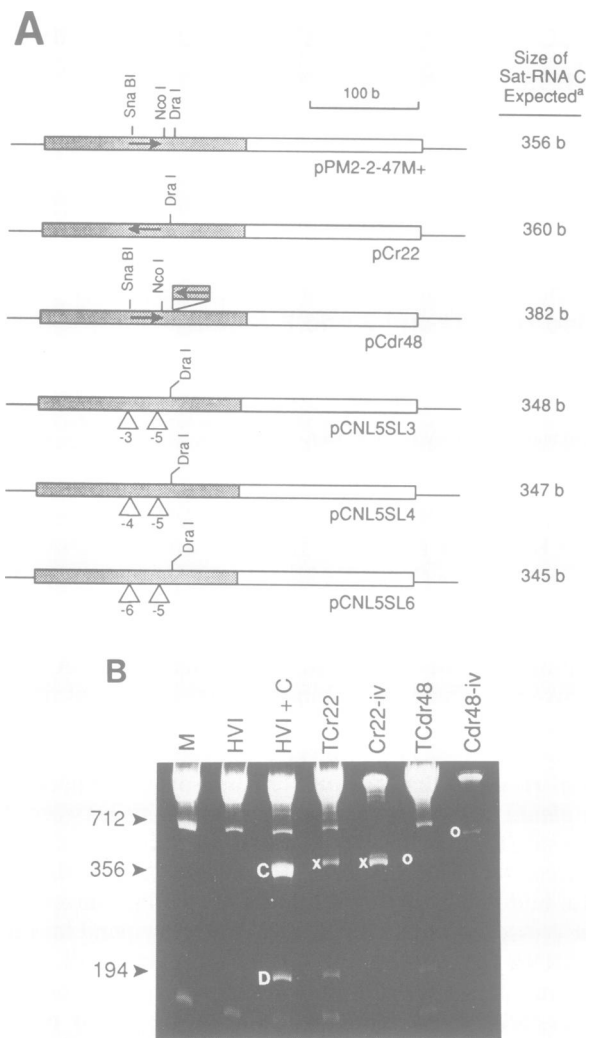


Fig. 2. (A) Mutations and modifications constructed in sat-RNA cDNA. Shaded domain is sat-RNA D related sequence. Triangles represent deletions and the number of deleted bases is indicated. Arrows denote the sequence from base 79 to 100 of sat-RNA C. In pCr22, this sequence is inverted. In pCdr48, a copy of the sequence is inserted in reverse orientation in the *Dra*I site. (B) Gel analysis of total leaf RNA isolated from mock (buffer) infected turnip (M), or turnip infected with helper virus (HVI), HVI plus transcripts of wild-type sat-RNA C synthesized from plasmid pPM2-2-47M+ (HVI + C), HVI plus transcripts from pCr22 (TCr22), or HVI plus transcripts from pCdr48 (TCdr48). The gel is composed of 5% acrylamide and 50% urea. HVI is a total RNA preparation from turnip infected with TCV genomic RNA from TCV-M, and sat-RNA D. Lanes labeled pCr22-iv and pCdr48-iv are samples of the *in vitro* synthesized transcripts. Sat-RNA C and sat-RNA D accumulating *in vivo* are labeled C or D, respectively. 'X' and 'O' denote the transcript or sat-RNA accumulating *in vivo* derived from pCr22 or pCdr48, respectively. Note that the *in vitro* transcript of pCdr48 migrates more slowly than the *in vivo* sat-RNA. ^aIn plants infected with transcripts derived from the following constructs.

difficult to eliminate completely from TCV inocula (Simon and Howell, 1987). Two weeks post-inoculation with *in vitro* transcripts, putative sat-RNA C is recovered from plants and sequenced to determine if the RNA accumulating *in vivo* is identical to the inoculated RNA.

Sat-RNA C is able to accommodate a variety of insertions and deletions in the middle two-thirds of the molecule (Simon *et al.*, 1988; Carpenter, Cascone and Simon, unpublished).

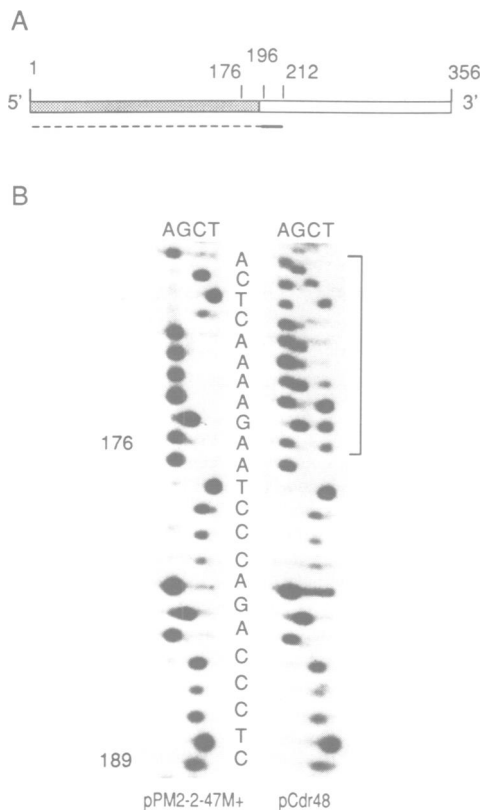


Fig. 3. (A) Schematic representation of sat-RNA C. The short, thick bar represents the oligonucleotide used for the reverse transcriptase-mediated sequencing of the RNA. The extension product is indicated by a dashed line. The region of similarity between sat-RNAs C and D is shaded. (B) Dideoxynucleotide sequence of RNA accumulating in plants infected with transcripts derived from pPM2-2-47M+ (wild-type sat-RNA C) or pCdr48. The sequence heterogeneity in the sat-RNA from pCdr48 is bracketed. Sequence heterogeneity of the *in vivo* accumulating RNA from each construct described in Figure 2 began at sat-RNA C nucleotide 176.

However, a number of mutations can be constructed in sat-RNA C cDNA which do not result in the accumulation of RNA of the expected size in plants (Figure 2). Two to three weeks post-inoculation with transcripts derived from these constructs, total RNA was extracted from leaves and analyzed by gel electrophoresis. An RNA which migrated slightly more slowly than the wild-type sat-RNA C was present in each plant regardless of the size of the inoculated transcript (Figure 2B). Attempts were made to sequence this RNA species from individual plants using primer extension, dideoxynucleotide chain termination methods (Carpenter and Simon, 1990). Although the oligonucleotide used for the sequencing reaction (complementary to nucleotides 196–212 of sat-RNA C) hybridized to the new RNA species and the oligonucleotide sequence was extended with reverse transcriptase, the sequence of the RNA species could not be determined due to heterogeneity of the RNA (Figure 3). The sequence heterogeneity began precisely at nucleotide 176, 20 bases upstream from the oligonucleotide primer used in the sequencing reaction.

Because of the inability to sequence this RNA species directly, cDNA clones were generated from two to three individual plants infected with transcripts from each construct described in Figure 2. Sequence analysis of the cDNA clones revealed that the RNA species accumulating in each of these

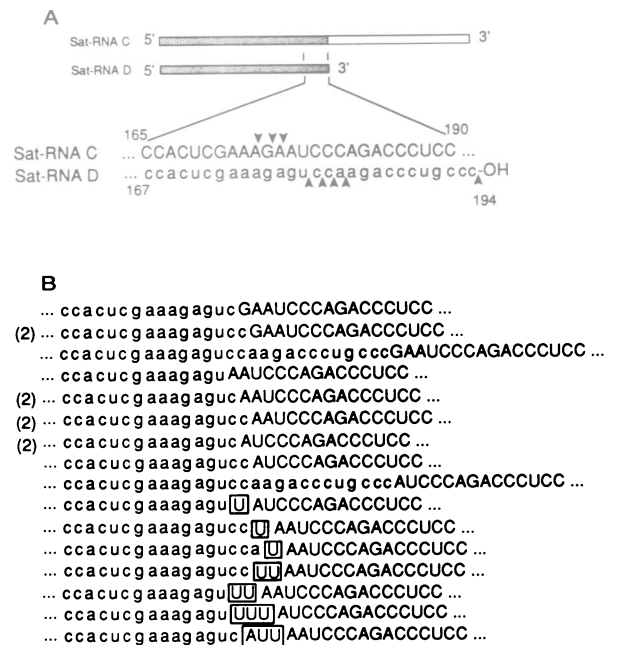


Fig. 4. (A) Sequence of the region involved in recombination between sat-RNA D and sat-RNA C. Arrows denote different recombination crossover points. The region of sequence similarity between the two sat-RNAs is shaded. (B) Sequences surrounding the junctions of 20 independent recombinants between sat-RNA D and sat-RNA C. All recombinant sat-RNAs contained sat-RNA D sequence in the 5' region and sat-RNA C sequence in the 3' region. Numbers in parentheses indicate the number of independent isolations of a particular recombinant. Boxed nucleotides correspond to non-template encoded bases at the recombination junctions.

plants contained sat-RNA D sequence at the 5' end joined to sat-RNA C sequence at the 3' end. The exact location where the homology to sat-RNA D ended and the sat-RNA C specific sequence began (i.e. the site of recombination) could be determined by virtue of the imprecise nature of the event (Figure 4A). The right side of the recombination junction point was always one of three consecutive nucleotides in sat-RNA C (bases 175–177), indicating a preference for a specific sequence in the recombination event. Eighteen out of 20 independent recombination events occurred between nucleotides 180 and 183 of sat-RNA D and bases 175–177 in sat-RNA C. The inexact crossover events resulted in the duplication of three to six nucleotides in the sat-RNA recombinants. Two cDNA clones contained the complete sequence of sat-RNA D joined at the 3' end to base 175 or 177 of sat-RNA C (Figure 4B). Curiously, between one and three additional nucleotides had been inserted at the crossover junction in seven of the recombinant sat-RNAs. These additional nucleotides were all uridine residues with the exception of one insert of AUU. cDNA clones derived from recombinant sat-RNAs with different crossover points and/or additional nucleotides could be isolated from individual plants, thereby accounting for the original observation of heterogeneity in the RNA.

Similarity among the right side junction sequences of the recombinant RNAs and subviral RNAs formed from discontinuous TCV sequence

The generation of DI RNAs also involves the joining of discontinuous RNA fragments, and may result from the same or a similar mechanism to that which produces recombinant



Fig. 5. (A) Comparison of sequence between the 5' end of TCV (Carrington *et al.*, 1989) and the junction points in the recombinant (D/C) sat-RNAs and between TCV bases 140 and 3863 in DI RNA G (Li *et al.*, 1989). (B) Comparison of sequences between the 5' ends of DI RNA G and sat-RNA F (Simon *et al.*, 1988), and the junctions of discontinuous sequences in sat-RNA C (Simon and Howell, 1986) and DI1 RNA (Li *et al.*, 1989). Arrowheads indicate junction points. The positions of the junctions in the various species are shown in Figure 1. Numbers in parentheses denote the first (1) or second (2) junction in the RNAs. Only the junction 140/3863 is shown for DI RNA G due to the unknown origin of the first 21 nucleotides at the 5' end. The 5' end sequence of DI RNA G was determined by cloning the joining region between monomeric units of DI RNA G dimers as described in Materials and methods. This sequence differs slightly from the previously reported sequence determined by primer-extension sequencing on a total population of DI RNA G molecules (Li *et al.*, 1989). Since sat-RNAs C, D and F have similar 5' end sequences, only one is shown. Shading of nucleotides is meant to aid in viewing sequence similarities and does not imply a requirement of any specific bases in the replicase recognition signals.

sat-RNAs. Therefore, we examined the crossover points of the TCV DI RNAs and wild-type sat-RNA C (which has a 3' domain consisting of two regions of TCV) for sequence similarity with the sat-RNA recombinant junction. The sequence at the right side of one DI RNA junction, where TCV base 140 is joined to base 3863 in DI RNA G, is very similar to the right side of the recombinant sat-RNA junction, sharing 10 of 13 nucleotides (Figure 5A). The similarity between the right side of the recombinant sat-RNA junction also extends four bases to the left of the DI RNA G junction. This consensus sequence will be referred to as motif 1.

Sequence similarity did not exist between other DI RNA or sat-RNA C junctions and motif 1. However, when all other junction sequences in the discontinuous subviral RNAs were compared with each other, a second consensus sequence of ~20 nucleotides was found, located on the right sides of the junctions (motif 2; Figure 5B). Nearly all of these junctions contain the trinucleotide AAA followed closely by the dinucleotide GG. Furthermore, the 20 nucleotides on the right sides of the junctions are rich in purines (75–80%) when compared with 20 nucleotides on the left sides of the junctions (45–55% purines). More evidence for the non-random nature of the right side junction sequences is that two independent events, the formation of sat-RNA C and DI1 RNA, resulted from the targeting of the same TCV sequence (base 3898). There are no sequences in common among the left sides of the junctions.

The junctions of the DI RNAs, recombinant sat-RNAs and sat-RNA C also share a second feature: the presence of short repeated nucleotides within 20 bases on both sides of nearly all crossover points. The pentanucleotide ACCCC is found on both sides of the junction of DI RNA G, where TCV bases 140 and 3863 are joined. The second junction of sat-RNA C, where TCV base 3779 is joined to base 3898, has the sequence CCAA on both sides. The first junction of DI1 RNA contains the pentanucleotide AGGUA on the left and right sides, while the recombinant sat-RNAs have the hexanucleotide AUCCCA on both sides of the crossover point. DI1 RNA has the sequence AGCUAU at the left side of the second junction; the original sequence in this position

of the TCV transcript used for the *de novo* generation of DI1 RNA is AGCACU which also is found at the right side of the junction.

Sequence at the right side of the recombinant sat-RNAs, DI RNAs and sat-RNA C junctions is similar to the sequence at the 5' ends of TCV genomic or sat-RNAs

Since replicase-driven copy choice models have been proposed as the mechanism for recombination between viral RNA molecules, we looked for sequence similarity between motif 1 and motif 2 and the 5' and 3' ends of TCV and its associated subviral RNAs. Our rationale was that the replicating enzyme may recognize specific viral 3' end sequences in order to initiate replication of the infectious (+) strands, and may recognize the complement of the 5' end sequence to replicate newly synthesized (–) strands.

Extensive similarity was found between motif 1 and a sequence located 11 nucleotides from the 5' end of TCV (Figure 5A). The TCV 5' proximal sequence shares 14 of 18 nucleotides with the right side recombinant sat-RNA junction and 12 of 16 nucleotides with the same sequence surrounding the junction of DI RNA G. Motif 2 resembles 15 nucleotides at the very 5' end of the sat-RNAs (Figure 5B). The 5' 20 nucleotides of the sat-RNAs are rich in purines (75%) and contain the trinucleotide AAA followed closely by the dinucleotide GG, characteristic of the junctions shown in Figure 5B.

We had previously noted that the partial sequence obtained for the 5' end of DI RNA G was similar to the 5' end of the sat-RNAs and not related to TCV (Li *et al.*, 1989). In order to complete the sequence of the 5' end of DI RNA G, cDNAs of DI RNA G dimers were cloned. The sequence corresponding to the 5' end of DI RNA G was obtained and is presented in Figure 5B. The first 10 nucleotides of DI RNA G and sat-RNA F are identical. As in the junction sequences of motif 2, a trinucleotide AAA is followed closely by the dinucleotide GG. The 5' 20 nucleotides of DI RNA G are also rich in purines (80%). Two of the five cDNA clones sequenced had additional nucleotides inserted at the

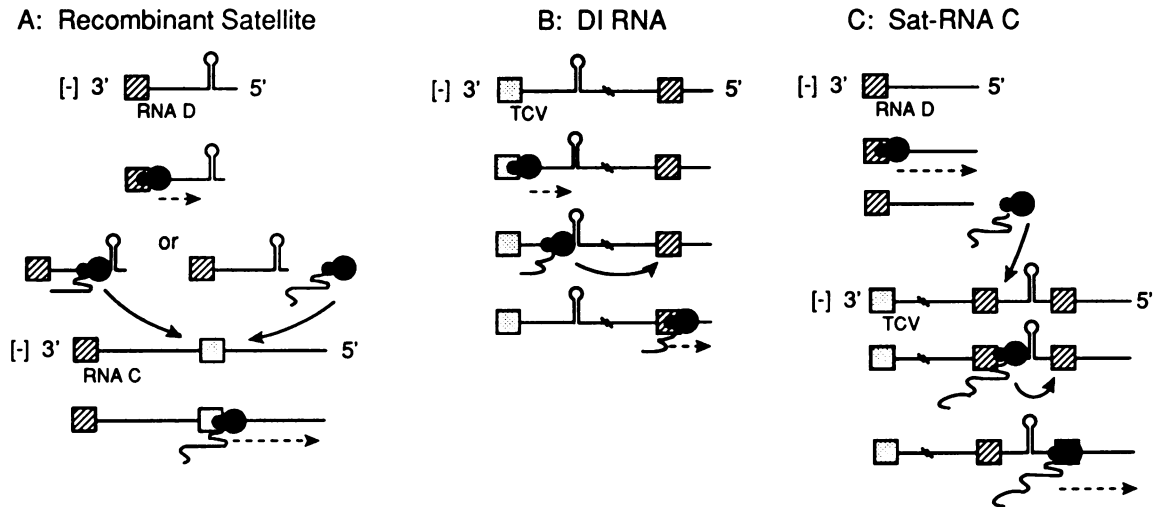


Fig. 6. Model for the generation of (A) recombinant sat-RNAs, (B) DI RNAs and (C) sat-RNA C. The model suggests that during replication of [−] strand templates, the replicase (represented by two filled-in circles) dissociates from the template due to a yet undetermined constraint (represented by a hairpin structure) or after completion of the daughter strand, and then reinitiates synthesis at one of two recognition signals (motifs 1 and 2, represented by shaded boxes) without first releasing the nascent [+] strand. The two recognition signals are shaded differently.

junction between the monomeric units (GG or GA). The relevance of this observation will be discussed below.

Discussion

Model for the generation of DI RNAs, recombinant sat-RNAs and sat-RNA C

Two different mechanisms have been postulated to explain the phenomena of RNA recombination and DI RNA formation in animal virus systems. The first model involves the breakage and joining of postreplication RNAs using a mechanism which might resemble the *trans*-splicing reaction thought to occur in *Caenorhabditis elegans* (Krause and Hirsh, 1987), trypanosomes (Murphy *et al.*, 1986) and *Chlamydomonas reinhardtii* chloroplasts (Choquet *et al.*, 1988). *Trans*-splicing, as described in these systems, is similar to *cis*-splicing in its requirement for specific splicing signals. The second model suggests that recombinant and DI RNA formation occur during replication of the RNAs, when the replicase, while copying the original template, detaches along with the nascent daughter strand, and re-establishes synthesis downstream on the same template (forming DI RNAs) or on a second template (forming recombinants). Using temperature- and chemical-sensitive poliovirus parental RNAs, Kirkegaard and Baltimore (1986) demonstrated that template switching of the replicase during (+) strand synthesis was responsible for the generation of poliovirus recombinants.

Based on the lack of specific splicing signals at the junction points as well as the sequence similarity between the 5' end of TCV, DI RNA G and the sat-RNAs, and the right sides of all junction sequences, we believe our data best support a replication driven mechanism. Our model suggests that there are two minus (−) strand recognition signals where replication can be initiated or reinitiated (Figure 6). One signal is found at the 5' end of TCV and right side of two junctions (motif 1) while the other is found at the 5' end of the sat-RNAs and DI RNA G as well as four junctions (motif 2; see Figure 5). As the replicase copies the (−) strand, it can detach from the template either prematurely or when it reaches the natural end-point.

Premature termination may not involve a specific sequence since there are no sequence similarities at the left sides of the junctions, but rather may be due to RNA secondary or tertiary structures or obstruction by proteins. Before the replicase releases the nascent (+) strand, it reinitiates synthesis at a recognition signal on the same or a different (−) strand template. Recombinant sat-RNAs are formed if the replicase dissociates from sat-RNA D after completing or nearly completing (+) strand synthesis and reinitiates synthesis at a replicase recognition signal (motif 1) on the (−) strand of sat-RNA C (Figure 6A). DI RNAs are formed if the replicase, while synthesizing new TCV (+) strands, dissociates from the template and reinitiates synthesis on the same or a second molecule of TCV (−) strand at either a motif 1 or motif 2 sequence (Figure 6B). This model also accounts for the original formation of sat-RNA C (Figure 6C). When replication of sat-RNA D (−) strand was completed, the replicase along with the nascent daughter strand reinitiated synthesis at the motif 2 sequence at TCV base 3764. After synthesis of an additional 15 nucleotides, the replicase released from the TCV template and reinitiated synthesis at a second motif 2 sequence at TCV base 3898, thus generating a sat-RNA–DI RNA hybrid molecule. The presence of short stretches of homologous nucleotides on both sides of most junctions may be important in aiding the replicase–nascent strand complex to position itself correctly at an internal replicase recognition signal. The occurrence of short homologous sequences on both sides of six out of seven junctions is not likely to be random. However, caution should be used in interpreting this observation since several other repeated sequences of five or six nucleotides can be found scattered throughout sat-RNA C and sat-RNA D.

Our model does not explain the insertion of 36 nucleotides into sat-RNA D at base 186 which resulted in the formation of sat-RNA F (Simon and Howell, 1986; Figure 1). This insert consists of a 20 nucleotide segment of unknown origin, which shares no similarity with either motif 1 or motif 2, followed by a 16 nucleotide repeat of the sequence directly upstream from the insert. Because of the unusual composition of the inserted sequence, a mechanism separate from the one described in Figure 6 may have been responsible for the

formation of sat-RNA F. Our model also does not explain the origin of the 5' 21 nucleotides of DI RNA G. The first 10 nucleotides of DI RNA G and sat-RNA F are identical. However, the following 11 nucleotides of DI RNA G share little similarity with either TCV genomic or subviral RNAs (Li *et al.*, 1989). The sequence surrounding TCV base 43, where the homology between DI RNA G and TCV begins, does not resemble either motif 1 or motif 2. It therefore seems likely that additional uncharacterized mechanisms may contribute to the formation of TCV associated subviral RNAs.

The replicase-driven copy choice model described in Figure 6 can account for the normal accumulation of linear sat- and DI RNA multimers in infected plants. After completing synthesis of a full-length sat- or DI RNA (+) strand, the replicase, before releasing the newly synthesized strand, could initiate a second round of replication either using the same template or a second homologous template thus generating tandemly copied multimeric species. Circular sat-RNAs (virusoids) are thought to replicate by a 'rolling circle' mechanism whereby higher order multimers undergo a self-cleavage reaction (Prody *et al.*, 1986; Forster and Symons, 1987). However, circular forms of linear sat-RNAs, such as the TCV sat-RNAs or cucumber mosaic virus sat-RNA, have not been found (Linthorst and Kaper, 1984; A.E.Simon, unpublished) nor has self-cleavage been found to occur (A.E.Simon, unpublished).

Seven out of 20 recombinant sat-RNAs had additional nucleotides inserted at the crossover junction. The origin of these additional nucleotides is not clear. In four of these seven recombinant sat-RNAs, the nucleotides preceding (sat-RNA D sequence) and following (sat-RNA C sequence) the additional nucleotide(s) are different, thereby arguing against a replicase stuttering-type mechanism (Schwer *et al.*, 1987; Thomas *et al.*, 1988). One explanation is that the TCV replicating enzyme, while detached from a template with the nascent daughter strand, is able to polymerize nucleotides without a template. It is interesting to note that natural dimers of sat-RNA C (Simon and Howell, 1986) and DI RNA G (this report) can also have one or two additional residues at the junction between monomers. Since we believe that formation of multimers and sat-RNA recombinants are different manifestations of the same process, the ability to add additional nucleotides to the junction may be an intrinsic property of the replicase.

Our model for the formation of recombinant sat-RNAs and DI RNAs, in which the replicase detaches from the (-) strand template and reinitiates synthesis at a specific signal on the same or a different molecule, does not explain the formation of homologous recombinants or DI RNAs in most other systems. DIs and/or viral recombinants of vesicular stomatitis virus (Yang *et al.*, 1983; Meier *et al.*, 1984), poliovirus (Kirkegaard and Baltimore, 1986; Kuge *et al.*, 1986) and mouse hepatitis virus (Baric *et al.*, 1987; Makino *et al.*, 1988) have no sequence homology between different joining sites, but may be generated in regions of high secondary structure (Baric *et al.*, 1987; Wilson *et al.*, 1988). However, Re *et al.* (1985) have proposed a similar copy choice/replicase recognition signal model for the generation of Sendai virus DI RNAs during replication of the infectious (-) strand template based on their finding of pyrimidine rich, putative promoter signals on only one side of all junctions. These signals are similar to a sequence located 5-16

nucleotides from the 3' end of Sendai virus. As noted in our results, the right side junction sequences of the TCV sat-RNAs and recombinants are purine rich. However, since our model predicts that recombinants and DIs are formed while the (-) strand is being copied, the proposed signal recognized by the replication enzyme is actually the pyrimidine rich complement. It is interesting to note that direct evidence for poliovirus recombinants (Kirkegaard and Baltimore, 1986) and indirect evidence in the Sendai virus and TCV systems suggest that polymerase template switching occurs during synthesis of only one RNA strand. Template availability or structural differences between the two strands may contribute to this limitation.

In this report, we have described the second definitive example of recombination in a plant RNA virus system, between two sat-RNAs associated with TCV. Recombination in this system has several characteristics not previously detected. First, crossover points between sat-RNA D and the related sequence in the 5' half of sat-RNA C have only been found at two sites: 14-16 nucleotides from the 3' end of sat-RNA D, or the very 3' end of sat-RNA D, and nucleotides 175-177 of sat-RNA C. Secondly, seven out of 20 recombinant events resulted in the addition of non-template encoded residues. Since homologous recombination has been defined as a precise yet random exchange between related viral genomes, recombination between TCV sat-RNAs can be considered to be non-homologous. Non-homologous recombination has only been reported in two systems: Sindbis virus, where cellular tRNA sequences can be found joined to Sindbis DI RNAs (Monroe and Schlesinger, 1983), and brome mosaic virus, where duplicated sequences have been found at junctions (Bujarski and Kaesberg, 1986).

Over the years, we have constructed a large number of mutations in sat-RNA C which have no apparent deleterious effects on infectivity (Simon *et al.*, 1988; Carpenter, Cascone and Simon, unpublished). In these cases, recombination would not be detected due to the abundance of infectious, unrecombined sat-RNA. Although it is not known why some viruses are able to generate DI RNAs and/or recombine RNAs, this process allows the 'correction' of deleterious mistakes in sat-RNA C by recombination with the similar region in sat-RNA D. TCV is then able to associate with an infectious, near-normal sat-RNA C. Since recombinant sat-RNAs are almost always recovered in infections where sat-RNA C contains a non-viable mutation in the 5' domain, association with a normal or near-normal sat-RNA C may confer a selective advantage on TCV in virus replication, packaging or spread.

Materials and methods

Virus isolates and plant inoculations

Complete descriptions of the TCV isolates and methods of inoculations have been published (Li *et al.*, 1989). TCV-M and TCV-B were previously designated TCV WT and TCV-T, respectively. Sat-RNA transcripts were synthesized *in vitro* from linearized template using *E. coli* RNA polymerase as previously described (Simon and Howell, 1987). Transcription was initiated at a promoter directly upstream from the sat-RNA cDNA sequence. The resultant transcripts contain the correct 5' end of sat-RNA C and five additional nucleotides at the 3' end. *In vitro* synthesized RNA (~125 ng/plant) was combined with inoculation buffer (Li *et al.*, 1989) and 5 µl of helper virus inoculum (HVI) and then mechanically inoculated onto the leaves of 2-week-old turnip cv. Just Right (Burpee). HVI was prepared by isolating total RNA from turnip previously infected with genomic

TCV (TCV-M isolate) and sat-RNA D (Simon and Howell, 1987). Two to three weeks post-inoculation, total leaf RNA was extracted using a LiCl precipitation procedure (Simon and Howell, 1986), then subjected to electrophoresis on 5% polyacrylamide, 50% urea gels.

Sat-RNA C mutagenesis

To create pCr22 and pCdr48, pPM2-2-47M+ (Simon and Howell, 1987) was digested with *Sna*BI and *Nco*I, and the purified 22 base fragment treated with *E. coli* polymerase large fragment (Klenow, BRL) and deoxyribonucleotides according to the manufacturer's suggested conditions to create blunt ends. The resultant 26 base fragment was ligated into the following positions in pPM2-2-47M+: the original location in reverse orientation to create pCr22; the *Dra*I site in reverse orientation to construct pCdr48. To create pCNL5SL3, 4 and 6, pCNL5 (which contains a deletion of sat-RNA C bases 96–100) was digested with *Sna*BI, ethanol precipitated, then treated with the slow form of *Bal*31 (IBI) for various times according to the manufacturer's suggested conditions. The extent of *Bal*31 digestion was monitored by *Sma*I restriction analysis of an aliquot of each sample. The remainder was digested with *Pst*I and the fragment containing leftward deletions purified and re-ligated into *Sna*BI-*Pst*I-digested pCNL5 to regenerate the intact plasmids. Dideoxynucleotide chain termination sequencing (Sequenase, USB) was used to identify plasmids with three (pCNL5SL3), four (pCNL5SL4) or six (pCNL5SL6) base deletions.

cDNA cloning

30 µg of total leaf RNA was subjected to electrophoresis on 5% polyacrylamide, 50% urea gels. The species to be cloned was excised from the ethidium bromide stained gel, electroeluted into dialysis bags and ethanol precipitated. The purified RNA was resuspended along with an oligonucleotide complementary to sat-RNA C bases 196–212 or 175–199 (for cloning sat-RNA recombinants) or complementary to DI RNA G bases 90–101 (for cloning DI RNA G dimers). The hybridization buffer contained 10 mM PIPES, pH 6.8, and 0.4 M NaCl in a total volume of 12.5 µl. The oligonucleotide was annealed to the RNA by heating at 10° above the T_m of the oligonucleotide for 8 min followed by slow cooling to 10° below the T_m . Following ethanol precipitation, first strand cDNA synthesis was carried out by resuspending the samples in a 12.5 µl MMLV reverse transcriptase reaction according to the manufacturer's suggested conditions (BRL). After incubating for 30 min at 37°C in the presence of 100 U MMLV reverse transcriptase, an additional 100 U of enzyme was added and incubation continued for an additional 30 min. The second strand was synthesized by adding 59 µl H₂O, 20 µl second strand buffer (0.5 M KCl, 25 mM DTT, 25 mM MgCl₂, 125 mM Tris-HCl, pH 8.3), 7.5 µl of 2.5 mM deoxyribonucleotides, 2 U of *E. coli* DNA polymerase I (BRL) and 1 U RNase H (BRL). After a 2 h incubation at 16°C, samples were phenol-chloroform extracted and ethanol precipitated with 5 µg carrier tRNA. cDNA was then treated with 2 U Klenow in a 40 µl reaction for 15 min at 23°C according to the manufacturer's suggested conditions. Following phenol-chloroform extraction and ethanol precipitation, the cDNA was resuspended in 10 µl H₂O and a 2.5 µl aliquot was ligated overnight at 16°C to 100 ng of pUC19 which had been previously digested with *Sma*I and treated with calf intestinal phosphatase (CIP, Boehringer Mannheim).

RNA sequencing

Dideoxynucleotide chain termination methods using oligonucleotide primers were performed as previously described (Carpenter and Simon, 1990). To confirm that the 3' end sequence of the recombinant sat-RNAs originated from sat-RNA C, most RNAs were sequenced directly using an oligonucleotide complementary to nucleotides 339–356. The sequence was clearly sat-RNA C in all cases and showed no heterogeneity until the recombination site described.

Acknowledgements

We thank Drs Andrea Branch and Hugh Robertson for helpful discussions and critical reading of the manuscript. This work was supported by the NSF (DBM 8704124 and DBM 8803853), University of Massachusetts Faculty Research Award and the Massachusetts Agricultural Experiment Station.

References

Ahlquist, P. and Janda, M. (1984) *Mol. Cell Biol.*, **12**, 2876–2882.
 Altenbach, S.B. and Howell, S.H. (1981) *Virology*, **112**, 25–33.
 Angenent, G.C., Posthumus, E., Brederode, F.T. and Bol, J.F. (1989) *Virology*, **171**, 271–274.

Baric, R.S., Shieh, C.K., Stohlman, S.A. and Lai, M.M.C. (1987) *Virology*, **156**, 342–354.
 Biebricher, C.K., Eigen, M. and Luce, R. (1981) *J. Mol. Biol.*, **148**, 391–410.
 Biebricher, C.K., Eigen, M. and Gardiner, W.C. (1983) *Biochemistry*, **22**, 2544–2559.
 Bujarski, J.J. and Kaesberg, P. (1986) *Nature*, **321**, 528–531.
 Carpenter, C.D. and Simon, A.E. (1990) *Biotechniques*, **8**, 26–27.
 Carrington, J.C., Heaton, L.A., Zuidema, D., Hillman, B.I. and Morris, T.J. (1989) *Virology*, **170**, 219–226.
 Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kuck, U., Bennoun, P. and Rochaix, J.D. (1988) *Cell*, **52**, 903–913.
 Cooper, P.D., Steiner-Pryor, A., Scotti, P.D. and Delong, D. (1974) *J. Gen. Virol.*, **23**, 41–49.
 Forster, A.C. and Symons, R.H. (1987) *Cell*, **49**, 211–220.
 King, A.M.Q. (1988) In Domingo, E., Holland, J.J. and Ahlquist, P. (eds), *RNA Genetics. Retroviruses, Viroids and RNA Recombination*. CRC Press, Boca Raton, FL, Vol. II, pp. 149–165.
 King, A.M.Q., McCahon, D., Slade, W.R. and Newman, J.W.I. (1982) *Cell*, **29**, 921–928.
 King, A.M.Q., McCahon, D., Saunders, K., Newman, J.W.I. and Slade, W.R. (1985) *Virus Res.*, **3**, 373–384.
 Kirkegaard, K. and Baltimore, D. (1986) *Cell*, **47**, 433–443.
 Krause, M. and Hirsh, D. (1987) *Cell*, **49**, 753–761.
 Kuge, S., Saito, I. and Nomoto, A. (1986) *J. Mol. Biol.*, **192**, 473–487.
 Lai, M.M.C., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L. and Stohlman, S.A. (1985) *J. Virol.*, **56**, 449–456.
 Li, X.H. and Simon, A.E. (1990) *Phytopathology*, **80**, 238–242.
 Li, X.H., Heaton, L.A., Morris, T.J. and Simon, A.E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9173–9177.
 Linthorst, H.J.M. and Kaper, J.M. (1984) *Virology*, **139**, 317–329.
 Makino, S., Shieh, C.D., Soe, L.H., Baker, S.C. and Lai, M.M.C. (1988) *Virology*, **166**, 550–560.
 Meier, E., Harmison, G.G., Keene, J.D. and Schubert, M. (1984) *J. Virol.*, **51**, 515–521.
 Monroe, S.S. and Schlesinger, S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3279–3283.
 Morris, T.J. and Carrington, J.C. (1988) In Koenig, R. (ed.), *The Plant Viruses*. Plenum Publishing Corp., Vol. III, pp. 73–112.
 Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell*, **47**, 517–525.
 Perrault, J. (1981) *Curr. Top. Microbiol. Immunol.*, **93**, 151–207.
 Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R. and Bruening, G. (1986) *Science*, **231**, 1577–1580.
 Re, G.G., Morgan, E.M. and Kingsbury, D.W. (1985) *Virology*, **146**, 27–37.
 Robinson, D.J., Hamilton, W.D.O., Harrison, B.D. and Baulcombe, D.C. (1987) *J. Gen. Virol.*, **68**, 2551–2561.
 Schwer, B., Visca, P., Vos, J.C. and Stunnenberg, H.G. (1987) *Cell*, **50**, 163–169.
 Simon, A.E. and Howell, S.H. (1986) *EMBO J.*, **5**, 3423–3428.
 Simon, A.E. and Howell, S.H. (1987) *Virology*, **156**, 146–152.
 Simon, A.E., Engel, H., Johnson, R.P. and Howell, S.H. (1988) *EMBO J.*, **7**, 2645–2651.
 Thomas, S.M., Lamb, R.A. and Paterson, R.G. (1988) *Cell*, **54**, 891–902.
 Wilson, V., Taylor, P. and Desselberger, U. (1988) *Arch. Virol.*, **102**, 131–139.
 Yang, F. and Lazzarini, R.A. (1983) *J. Virol.*, **45**, 766–772.

Received on November 2, 1989; revised on March 6, 1990