J. Mol. Biol. (1995) **245**, 608–622

JMB

Involvement of a Stem-loop Structure in the Location of Junction Sites in Viral RNA Recombination

Clifford D. Carpenter[1], Jong-Won Oh[2], Chunxia Zhang[1] and Anne E. Simon[1]*

[1]*Department of Biochemistry and Molecular Biology Program in Molecular and Cellular Biology and*

[2]*Department of Microbiology University of Massachusetts Amherst, MA 01003, U.S.A.*

**Corresponding author*

Recombination between RNAs associated with turnip crinkle virus is thought to occur during plus-strand synthesis at motifs resembling the 5'-ends of genomic, subgenomic and satellite RNAs. Common structural regions encompassing the motifs have been found for major crossover sites on two different minus-strand templates, with junctions preferentially located in a single-stranded region at the 3' base of a hairpin. Base changes, deletions and compensatory alteration constructed in and around the hairpin in the region of the turnip crinkle virus genomic RNA involved in recombination support the importance of the hairpin for normal crossover site selection. This region of the genomic RNA is also important for replication of the viral genomic RNA in plants and protoplasts, suggesting a common link between sequences required for recombination and viral replication.

Keywords: RNA structure; RNA replication; RNA replicase; RNA virus; satellite RNA

Introduction

RNA replication by RNA-dependent RNA polymerases (RdRps) leads to nucleotide misincorporations and RNA recombination; such errors, which result in heterogeneous populations of viral RNA molecules, have played an important role in the evolution of RNA viruses (Holland *et al*., 1982; Strauss & Strauss, 1988; Rao & Hall, 1993; Duarte *et al*., 1994). RNA recombination, the joining of normally discontinuous RNA segments of mainly viral origin (Jarvis & Kirkegaard, 1991; Lai, 1992; Simon & Bujarski, 1994) is thought to occur when processive transcription mediated by the viral RdRp is interrupted (possibly by structural constraints) allowing the enzyme to switch to a second template or an alternative position on the same template where polymerization of the nascent strand continues (Kirkegaard & Baltimore, 1986; King, 1988; Nagy & Bujarski, 1993).

RNA recombination is defined as either homologous or non-homologous (Lai, 1992). Homologous recombination occurs when sequences are exchanged between two closely related virus strains or within regions of homology in otherwise unrelated segments. For coronaviruses, homologous recombination occurs between genomic RNAs (g-RNAs) of different strains, between subviral defective RNAs and the g-RNA, and between subgenomic and g-RNAs (Makino *et al*., 1986; Koetzner *et al*., 1992; van der Most *et al*., 1992). Non-homologous recombination occurs between RNA molecules with little or no homology at the crossover site. Defective interfering (DI) RNAs associated nearly ubiquitously with animal viruses and a growing number of plant viruses, are also thought to be generated by non-homologous recombination (Roux *et al*., 1991).

Crossover site selection in RNA recombination has been studied mainly by sequencing through the junctions of natural recombinants. Crossover sites, however, can be influenced by selective pressure for molecules that are replication competent (Banner *et al*., 1990; Lai, 1992). For example, crossover sites were found to be clustered in coronavirus recombinants when screening was based on the presence of selectable markers, which eliminated recombinants that were unable to replicate (Banner *et al*., 1990). In contrast, crossover sites seem to be randomly distributed when a polymerase chain reaction (PCR) strategy was used to amplify and clone recombinant molecules (Banner & Lai, 1991). The secondary structure of the template appears to

Abbreviations used: RdRp, RNA-dependent RNA polymerase; g-RNA, genomic RNA; DI, defective interfering; BMV, brome mosaic virus; TCV, turnip crinkle virus; sat-RNA, satellite RNA; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR; ORF, open reading frame; CMCT, carbodiimide; DEPC, diethylpyrocarbonate; DMS, dimethylsulfate; WT, wild-type.

0022–2836/95/050608–14 $08.00/0

influence the generation of internal deletions in the genomic RNA of flock house virus and tomato bushy stunt virus (Li & Ball, 1993; White & Morris, 1994), by bringing the deletion sites into proximity. Homologous recombination in picornaviruses (Kuge *et al*., 1986; Romanova *et al*., 1986; Tolskaya *et al*., 1987) and non-homologous recombination in brome mosaic virus (BMV; Bujarski & Dzianott, 1991; Nagy & Bujarski, 1993) occur more frequently in, or can be artificially targeted to, single-stranded sequences that precede heteroduplexes containing the two parental templates. With the exception of junctions in recombinant species associated with turnip crinkle virus (TCV; Cascone *et al*., 1990; Zhang *et al*., 1991; Carpenter & Simon, 1994), and possibly junctions in Sendai virus DI RNAs (Re *et al*., 1985), sequence similarity at crossover sites has not been reported.

TCV is a single-stranded RNA virus with a monopartite genome that infects a broad range of dicotyledonous plants. TCV naturally provides replication and encapsidation functions for a number of dispensable, non-coding, subviral RNAs that are either products of recombination or templates for further recombination. These include: satellite (sat-) RNAs, such as sat-RNA D (194 bases), which share no sequence similarity with the genomic RNA beyond seven 3'-terminal bases (Simon & Howell, 1986); DI RNAs, whose sequence is nearly completely derived from the genomic RNA (Li *et al*., 1989); chimeric species such as sat-RNA C (356 bases), containing full-length or nearly full-length sat-RNA D at the 5'-end and one or more segments from the 3'-end region of the genomic RNA at the 3'-end (Cascone *et al*., 1990; Zhang *et al*., 1991). Recombinant RNAs derived *de novo* from sat-RNA D and sat-RNA C, or sat-RNA D and the g-RNA have been described (Cascone *et al*., 1990, 1993; Zhang *et al*., 1991).

Analyses of junction sequences of TCV recombinant molecules revealed no evidence for any heteroduplexes at crossover sites. Rather, one of three, dissimilar ~20 nt motifs was nearly always located downstream of the junctions of TCV-associated recombinant RNAs (Cascone *et al*., 1990; Zhang *et al*., 1991; Carpenter & Simon, 1994). The similarity between the motifs and sequences at the 5'-ends of the g-, sat- and 1.45 kb subgenomic RNAs suggested a model whereby recombination occurs during plus (+)-strand synthesis and the RdRp recognizes the motifs, or structures containing the motifs, as sites for internal initiation (Cascone *et al*., 1990; and see Figure 1A). Mutagenesis of sequence within and surrounding the motif in sat-RNA C that is the target for recombination with sat-RNA D (motif I) revealed that a hairpin encompassing part of the motif sequence was required for recombination, with the junction sites located in a single-stranded region at the base of the hairpin (Cascone *et al*., 1993).

To determine if a second hot-spot for recombination shares similar morphological features with the motif I site in sat-RNA C, we examined recombination between sat-RNA D and the 3'-non-coding region of g-RNA. We provide evidence that a stem-loop structure encompassing a sequence similar to the 5'-ends of g-RNA and the 1.45 kb subgenomic RNA is important for normal crossover site selection. Furthermore, this region involved in recombination is required for accumulation of the viral genomic RNA in whole plants and isolated leaf cells, suggesting a link between sequences required for recombination and replication.

Results and Discussion

Regions involved in recombination share similar structural features

We previously determined by RNA blot hybridizations and sequence analysis that turnip plants infected with g-RNA and sat-RNA D accumulate recombinant molecules composed of full-length to nearly full-length sat-RNA D at the 5'-end joined to different lengths of g-RNA at the 3'-end (Zhang *et al*., 1991). The majority of recombinants had crossover sites between positions 3800 and 3825 in g-RNA, just downstream from the termination codon of the coat protein open reading frame and about 250 nt from the 3'-terminus. This ''hot spot'' for recombination in g-RNA contains two imperfect 24 nt tandem repeats (motifs IIIA and IIIB) that contain sequences similar to the 5'-terminus of g-RNA and the 1.45 kb subgenomic RNA (Figure 1B). The similarity consists of the consensus sequence CCAUUAU/C, followed by a region of variable spacing and then a purine-rich sequence. Preliminary mapping of the transcription start site for the 1.7 kb subgenomic RNA also indicates that a similar sequence is located just downstream of the putative start site: CACUAU followed by 20 nucleotides and then a 12-base perfect homology with the purine-rich sequence near the transcription start site of the 1.45 kb subgenomic RNA (M. Farley & A. E. S., unpublished results). Since subgenomic RNAs of RNA viruses are known to be synthesized by internal initiation on (–)-strand templates producing 3' co-terminal (+)-strand molecules (Miller *et al*., 1985; Watanabe & Okada, 1986; Levis *et al*., 1990), sequence similarity between motif IIIA and B and the 5'-ends of the subgenomic RNAs suggest that the motif IIIA/B sequences on the (–)-strand of g-RNA, or structures containing these sequences, may function as internal start sites where transcription can be re-initiated by the RdRp following switching of template strands.

We conducted *in vitro* secondary structure analysis of a 697 nt (–)-strand g-RNA segment containing the motif IIIA/B region using chemical and enzymatic probes that distinguish between single-stranded and helical regions. The results, presented in Figure 2A, indicate the presence of a hairpin and 3' single-stranded region, features predicted by computer-aided structural analysis using an algorithm that predicts optimal and suboptimal structures (Zuker *et al*., 1991; data not shown).

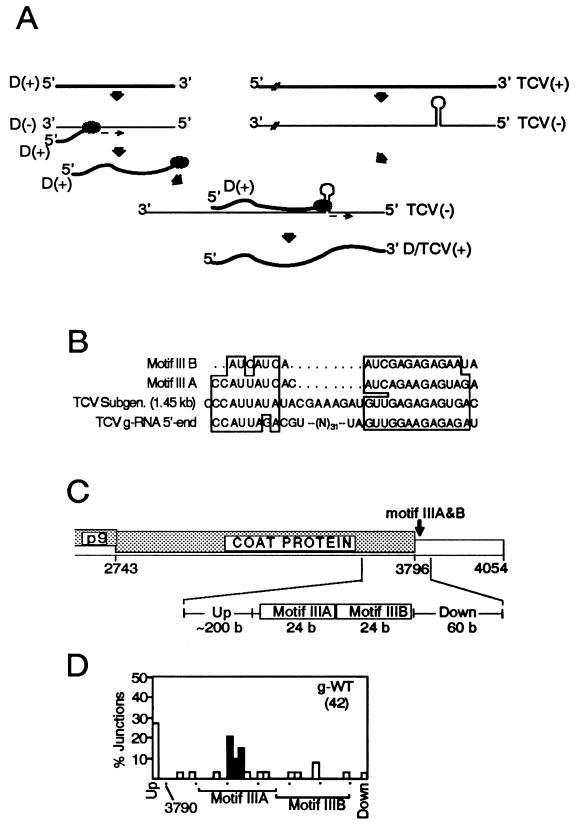

Figure 1. Recombination between sat-RNA D and TCV g-RNA. A, Model for recombination. Based on the finding that sequences at the right sides of the junctions in the (+)-strands of recombinant molecules resemble sequences at the 5'-ends of TCV associated RNAs (genomic, subgenomic and sat-RNAs), recombination is thought to occur during (+)-strand synthesis. The TCV replicase (filled oval), after synthesizing a (+)-strand of sat-RNA D, detaches from the template and before releasing the nascent strand, recognizes a sequence and/or structure near the 5'-end of TCV (–)-strand genomic RNA (represented as a hairpin) and reinitiates synthesis from the 3'-end of the sat-RNA D nascent strand on the TCV genomic RNA template producing a recombinant molecule. Thick lines denote (+)-strands and thin lines denote (–)-strands. B, Sequence similarity between two 24nt imperfect repeats (motif IIIA and motif IIIB), where most crossover sites in the TCV g-RNA are located, and sequence at the 5' termini of TCV g-RNA and 1.45kb subgenomic RNA (+)-strands (Carrington *et al*., 1987). Sequence from the corresponding (–)-strands is presented (in 3' to 5' orientation), since (–)-strands are thought to be the template for recombination and subgenomic RNA production. Boxed sequences are meant to aid in viewing sequence similarities. C, Location of motifs IIIA and IIIB in the untranslated 3'-region of the TCV g-RNA. Groupings of junction sites for the junction profiles are indicated. D, Junction profiles indicating the location of junction sites in the g-RNA for recombinants amplified from 2 plants previously inoculated with sat-RNA D and wild-type g-RNA (g-WT). The total number of recombinants sequenced (not including identical sibling clones) is noted in parentheses. Junctions found upstream (Up) or downstream (Down) of the motif III region are grouped together. Within motifs IIIA and IIIB (and 11 bases upstream), junctions found for every 2 bases are grouped together. Filled bars denote junctions at the major recombination site in the region.

Previous analysis of the motif I recombination region of sat-RNA C indicated that a stable stem-loop (hairpin B) was required for recombination, with crossover sites located at the 3'-base of the hairpin (Cascone *et al*., 1993). Computer-aided structure analysis also predicted a second stem-loop (hairpin A) 3' of hairpin B in the (–)-strand. Since we were interested in determining whether structural similarities exist between the motif I region of sat-RNA C and the motif III region of g-RNA, full-length sat-RNA C (–)-strands were subjected to chemical and enzymatic probing (Figure 2B). *In vitro* structural determination of the motif I region supported the existence of hairpin B but indicated that 3' of hairpin B is an extensive single-stranded region. The (–)-strands of both recombination sites, therefore, have similar structural features *in vitro*: a single-stranded region adjacent to the 3'-side of a stem-loop structure.

Positions of junctions in TCV g-RNA and sat-RNA D

To generate a more complete profile of junction sites in recombinant molecules generated from sat-RNA D and wild-type g-RNA, plants were inoculated with sat-RNA D and wild-type g-RNA and total RNA was isolated from uninoculated leaves three weeks later. Recombinant species containing portions of sat-RNA D and g-RNA were amplified by reverse transcriptase (RT) PCR using g-RNA- and sat-RNA D-specific primers, and the resultant fragments were cloned and sequenced. Only unique recombinants were scored from each plant; independence was based on the crossover points in sat-RNA D and g-RNA, and the presence or absence of non-template bases at the junctions, the occurrence of which has been described (Cascone *et al*., 1990, 1993; Carpenter *et al*., 1991*a*). The elimination of potential sibling clones controls for the possibility of biased amplification during RT-PCR. However, it should be noted that nearly all recombinant clones were unique and sibling clones were generally associated with the most common junction positions in sat-RNA D and g-RNA.

The junction sites in wild-type (and mutant) g-RNA are presented in histogram form that will be referred to as the junction profile (Figure 1C and D). Junctions located 11 nt or more upstream of motif IIIA were grouped together and labeled ''upstream''. Upstream junctions were favored to occur between 50 and 75 nt upstream of motif IIIA, where there is a previously noted motif II site (motif II sequences are very similar to the 5'-ends of TCV sat-RNAs; Cascone *et al*., 1990). From 11 bases upstream of motif IIIA to the end of motif IIIB, junctions that occurred at every two nucleotide positions were grouped together. Junctions downstream from motif IIIB were randomly distributed and were grouped together. Ambiguities in junction positions due to identical bases at both sides of the junctions were arbitrarily assigned to the g-RNA sequence.

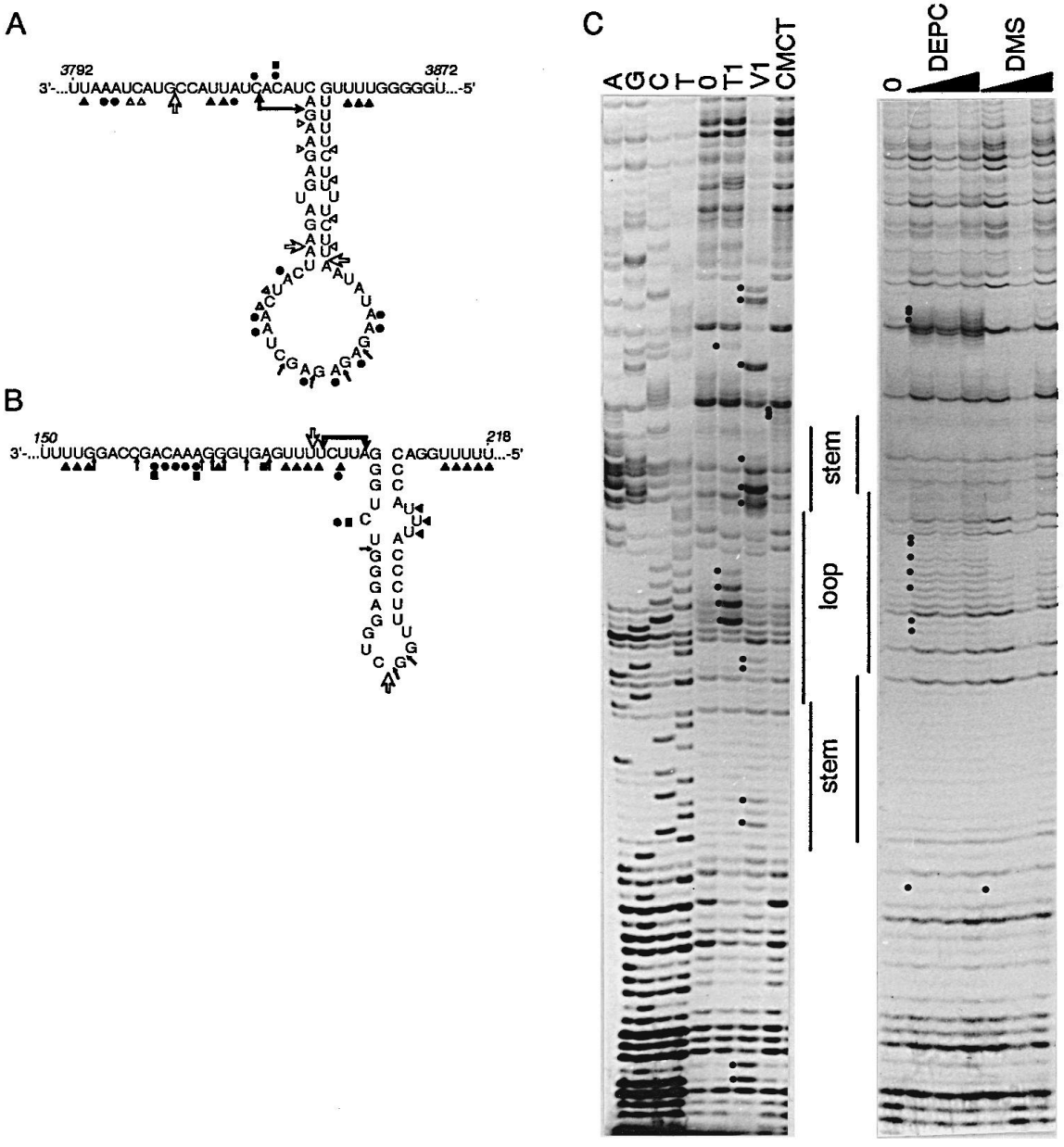

Figure 2. Secondary structure of the RNA near preferred recombination sites as determined by chemical and enzymatic probes. A, The motif III region of g-RNA. B, The motif I region of sat-RNA C. Single-stranded probes were DMS (filled boxes), DEPC (filled circles), CMCT (filled triangles), and RNase T_1 (small arrows); the double-stranded probe was RNase V_1 (open triangle). The (–)-strand sequence is shown. Open arrowheads encompass the motif sequences. Double arrowheads enclose the most prevalent recombination sites within the region shown. Junctions within this region in the g-RNA are denoted by filled bars in the junction profiles. For motif I of sat-RNA C, 198 of 200 recombinants had junctions within the prevalent site (Cascone *et al*., 1993). C, Examples of gels used to determine the reactive nucleotides shown in A and B. For the right side gel, 3 levels of DEPC and DMS were used to better identify the reactive bases. Bases needed to be clearly reactive on several gels for designations in A and B. A, G, C and T denote the dideoxynucleotide used in the reverse transcription sequencing reactions to generate a sequencing ladder and 0 denotes reverse-transcriptase only reaction. Location of the putative stem and loop regions are shown. Reactive bases clearly visible on these gels are indicated.

The junction profile for wild-type g-RNA (g-WT) is presented in Figure 1D. Of the junctions, 42% were within a six-base region (positions 3809 to 3815) in the middle of motif IIIA. The location of this preferred crossover region is at the 3' base of the putative hairpin (see Figure 2A), similar to the location of crossover sites at the base of the hairpin in the motif I region of sat-RNA C (see Figure 2B; Cascone *et al*., 1993). Another 33% of the junctions were located upstream of the motifs, while 2% were found downstream. There were only two duplicate clones in the 44 recombinants sequenced, indicating

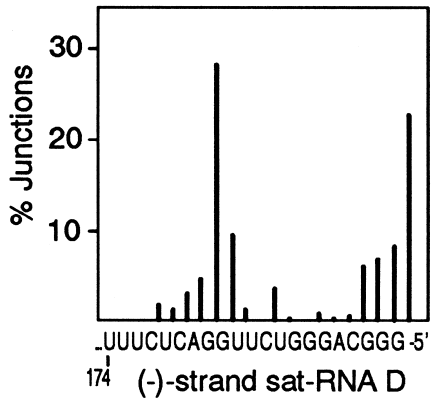

Figure 3. Junction sites in sat-RNA D: 430 randomly selected recombinants were analyzed. The sequence shown is from the 5'-end of the (–)-strand of sat-RNA D. No junction was found 3' of these positions.

a large number of independent recombination events. This suggests that recombinant molecules, once generated, may not be further amplified or are not amplified to the extent that selection takes place. In addition, we have been unable to generate infectious constructs for full-length versions of even the most often duplicated recombinants (data not shown). These results are consistent with prior findings that indicate only subviral RNAs of specific size and sequence are able to be amplified in plants (Li & Simon, 1991; A. E. S., unpublished results) or protoplasts (Zhang & Simon, 1994).

The junction sites for sat-RNA D are presented in Figure 3. There was a clear bimodal distribution of sat-RNA D junctions with over 50% at the 3'-terminus of the (+)-strand (i.e. recombinants contained the full-length sequence of sat-RNA D) or 13 nt upstream of the (+)-strand 3'-end. No junctions were detected more than 18 nt upstream from the 3'-end. If our model is correct, then the RdRp would switch from transcribing sat-RNA D to transcribing g-RNA. The location of most of the junctions within a few bases of the 3'-end of the (+)-strand or 12 or 13 bases upstream may arise from premature transcription termination by the RdRp due to RNA-bound proteins or sequences that result in pausing and then release of the RdRp-nascent strand complex. Alternatively, post-transcriptional cleavage of the newly synthesized 3'-end sequence can be postulated, analogous to the endonucleolytic cleavage induced by factors associated with DNA-dependent RNA polymerases (Surratt *et al*., 1991; Borukhov *et al*., 1993; Mote *et al*., 1994); this evolutionarily conserved mechanism permits DNA-dependent RNA polymerases to continue RNA synthesis past template sequences that are recalcitrant to processive transcription and might be a mechanism for releasing ''stuck'' TCV RdRp from the sat-RNA D template.

Effect of point mutations in the motif III region of g-RNA on the location of crossover sites in g-RNA

Point mutations in the stem and loop of the sat-RNA C motif I hairpin eliminated detectable recombination in infected plants (Cascone *et al*., 1993). To analyze how the sequence and structure of the motif III region in g-RNA impacts on recombination with sat-RNA D, we generated a series of base alterations and deletions in the region and tested each construct for the ability to recombine with sat-RNA D in infected plants. Initially, TCV g-RNA containing 11 individual point mutations in the hairpin and upstream single-stranded region were assayed for recombination with sat-RNA D (Figure 4A). Between 39 and 52 independent recombinants from at least two plants were analyzed for each mutation.

The four mutations in the single-stranded region 3' of the stem-loop (g-U3800A, g-U3805A, g-U3806A and g-U3808A; all terminology refers to the (–)-strand of g-RNA) did not substantially alter the distribution of junctions when compared with g-WT (Figure 4B). g-U3847G, a single base alteration in the loop, also gave a junction profile similar to wild-type. g-G3816A, which contained a GU to AU stem alteration, and g-A3837C, which contained a mutation in the favored crossover region in motif IIIB, appeared to generate fewer junctions within motif IIIB, while maintaining the same preferred junction site in motif IIIA. A higher percentage of upstream junction sites was found for three, structurally clustered g-RNA mutants (g-U3827G, g-U3850G and g-U3830A). An increase in upstream recombination sites may indicate structural alterations in the motif III region that lead to a decreased preference for crossing over in the region when compared with upstream sites.

Mutant g-U3834C generated a highly unusual junction profile (Figure 4B). Of the 47 recombinants analyzed, 31 had position 3811 as the g-RNA junction. This was the only mutation analyzed for this report that resulted in many sibling recombinant clones. Of the 47 clones sequenced from two plants, 31 had the crossover site at position 3811 and, of these, 27 were identical, with position 181 of sat-RNA D being the upstream junction. This specific recombinant (sat-RNA D 181/g-RNA 3811) was found in nearly every population of recombinant molecules analyzed. Efforts to generate biologically active clones of the sat-RNA D 181/g-RNA 3811 recombinant containing the U3834C mutation were unsuccessful (data not shown), suggesting that this species was not amplified preferentially. It is unlikely that a PCR preference for this recombinant, due to a single additional base alteration, was responsible for the high number of sibling clones. Rather, it is likely that the high number of junctions at position 3811 reflects preferred crossing over at this position due to the U3834C mutation.

Results of the single base alterations in the hairpin and upstream single-stranded region indicate that,

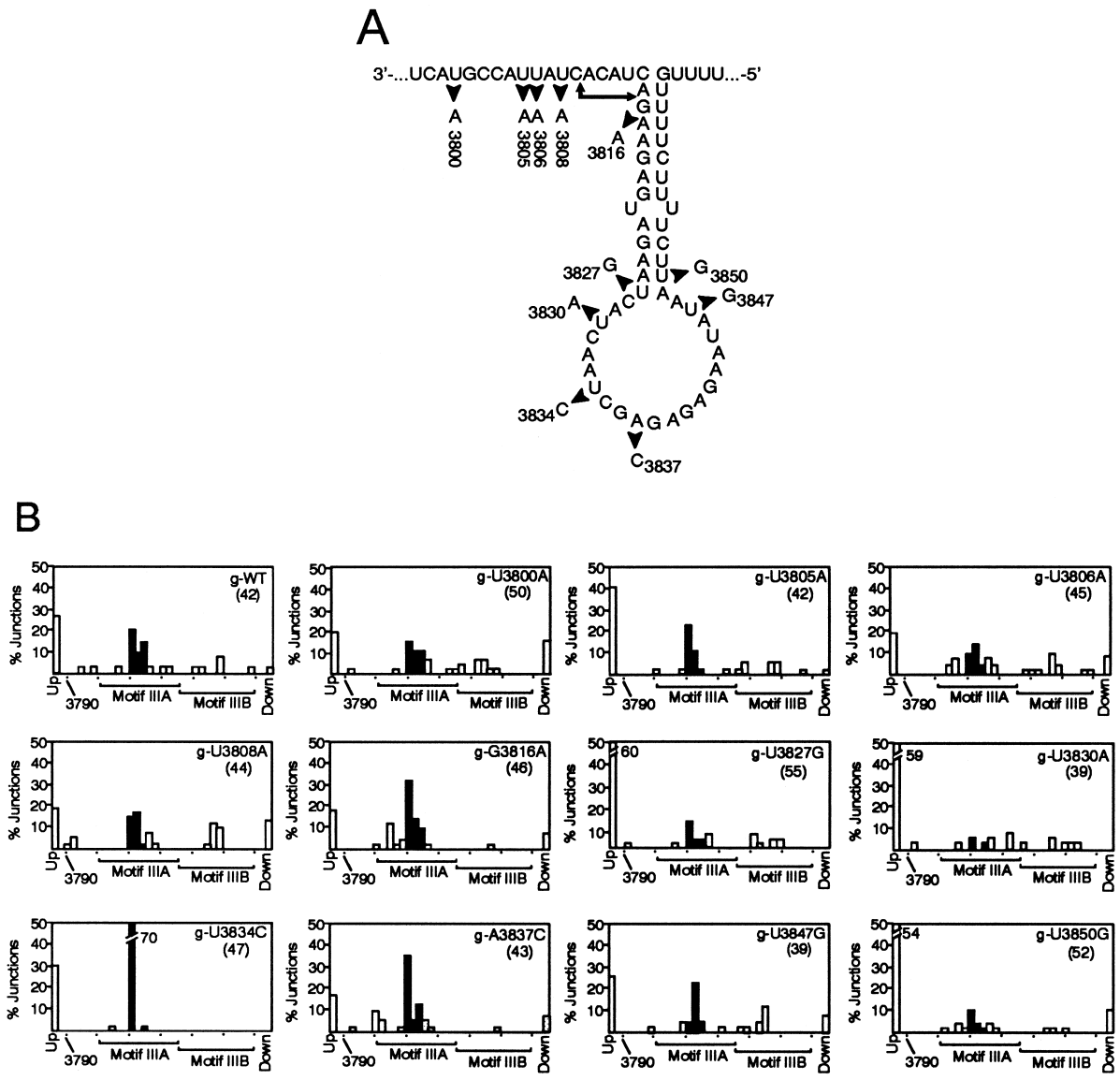

Figure 4. Effect of single base mutations in the motif III region of TCV genomic RNA on the junction profiles of recombinants. A, Location and composition of mutations in the putative structure. The (–)-strand sequence is shown. Numbering is from the corresponding (+)-strand. Major crossover region using wild-type g-RNA is between the joined arrowheads. B, Junction profiles indicating the location of junction sites in the g-RNA for recombinants amplified from 2 plants previously inoculated with sat-RNA D and mutant g-RNAs shown in A. The total number of recombinants sequenced (not including identical sibling clones) is noted in parentheses. Junctions were grouped together as described for Figure 1D. Filled bars denote recombinants within the double arrowhead region shown in A. Only non-identical recombinants were scored with the exception of g-U3834C (see the text).

unlike similar alterations in the motif I hairpin of sat-RNA C, single base changes in the motif IIIA/B region do not eliminate recombination. However, point mutations in the motif IIIA/B region can affect the distribution of junction sites although the functional significance of these base changes is not known.

Effect of loop deletions on the location of crossover sites in g-RNA

The g-U3847G mutation created a unique *Afl*II site in g-RNA cDNA, allowing for the generation of a stepwise deletion series in motif IIIB. Since motif IIIB comprises the loop of the putative hairpin, these deletions should indicate the effect of deleting loop sequence on recombination in the region. g-RNA containing this series of deletions was introduced into plants in a manner that differed from other constructs analyzed for this report. Transcripts containing various deletions were first inoculated separately onto turnip plants. After 18 days, total RNA was extracted, combined with transcripts of sat-RNA D and fresh plants inoculated. After three weeks, total RNA was extracted and recombinants analyzed as described. This sequential inoculation

method had an unexpected effect on crossover sites using g-WT. Although the junction site positions within motifs IIIA and IIIB were similar whether assayed by sequential inoculation or co-inoculation of g-RNA and sat-RNA D transcripts (compare g-WT junction profiles in Figure 5B and Figure 6B), the percentage of junction sites upstream and downstream were affected. The six constructs analyzed following sequential inoculation averaged 2% upstream junctions compared with an average of 31% upstream junctions for the 20 constructs analyzed following co-inoculation. It is not known why the method of inoculation had an effect on the generation of recombinants with crossover sites upstream and downstream of motifs IIIA and IIIB.

The junctions of all sequentially inoculated constructs were compared with g-WT subjected to the identical inoculation conditions (Figure 5B). Deletions of 17 bases or less were not predicted by computer structural analysis to disrupt the hairpin (data not shown) and did not substantially alter the junction site profile when compared with g-WT; the majority of recombination junctions were still found at the base of the stem. However, to maintain a hairpin in the region with a deletion of 21 bases, several nucleotides at the base of the stem would need to be recruited to form a new loop. The 21-base deletion affected the junction profile in two ways: (1) the peak was shifted to a new location ten nucleotides downstream from the wild-type position; (2) the number of junctions downstream of the

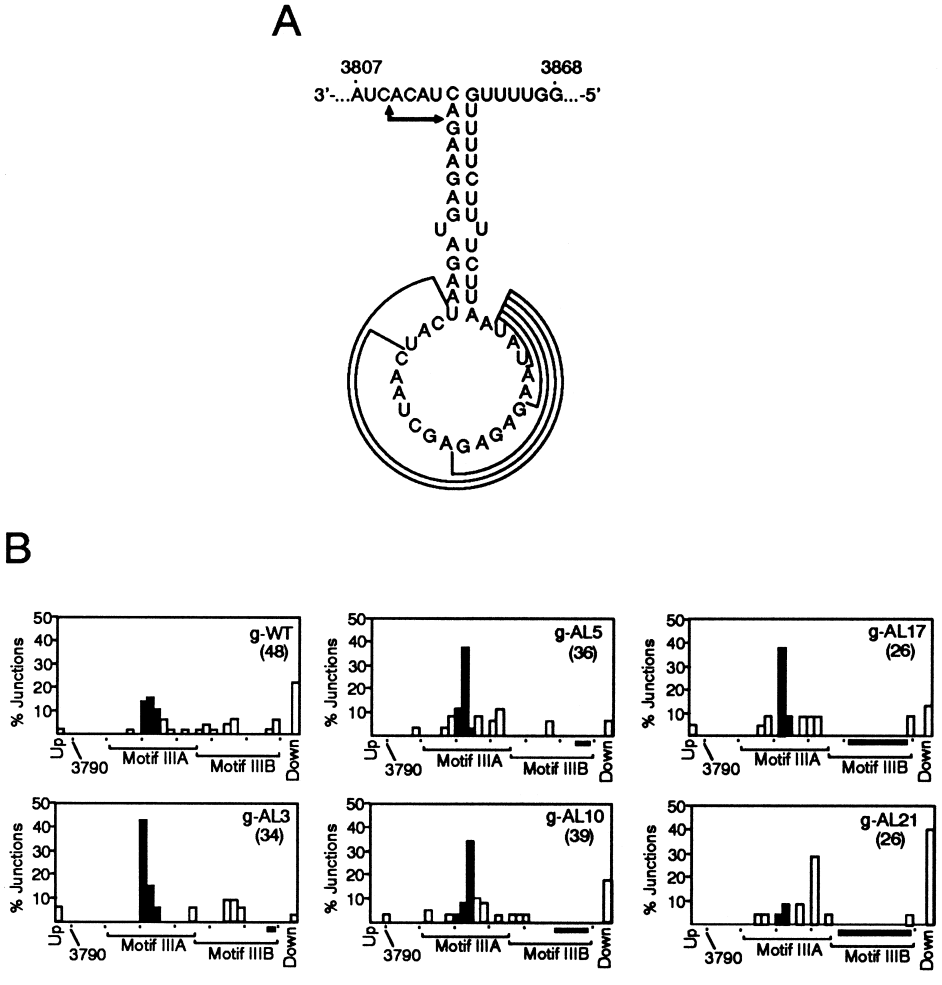

Figure 5. Effect of deletions in the putative loop region (motif IIIB) on the junction profiles of recombinants. A, Deletions between 3 and 21 bases (denoted by lines) were generated in the putative loop region. Major crossover region in recombinants produced using wild-type g-RNA (g-WT) is between the joined arrowheads. B, Junction profiles of recombinants amplified from 2 plants previously inoculated with sat-RNA D and the deletion mutants shown in A. Numbers in the names of the g-RNA constructs (e.g. g-AL17) denote the number of residues deleted. Horizontal bars beneath the junction profiles indicate the location and extent of the deletions. The number of non-sibling recombinants that were sequenced to produce the junction profiles is in parentheses. Junctions were grouped together as described for Figure 1D. Filled vertical bars denote junctions at the major recombination site in the region, located between the arrowheads in A.

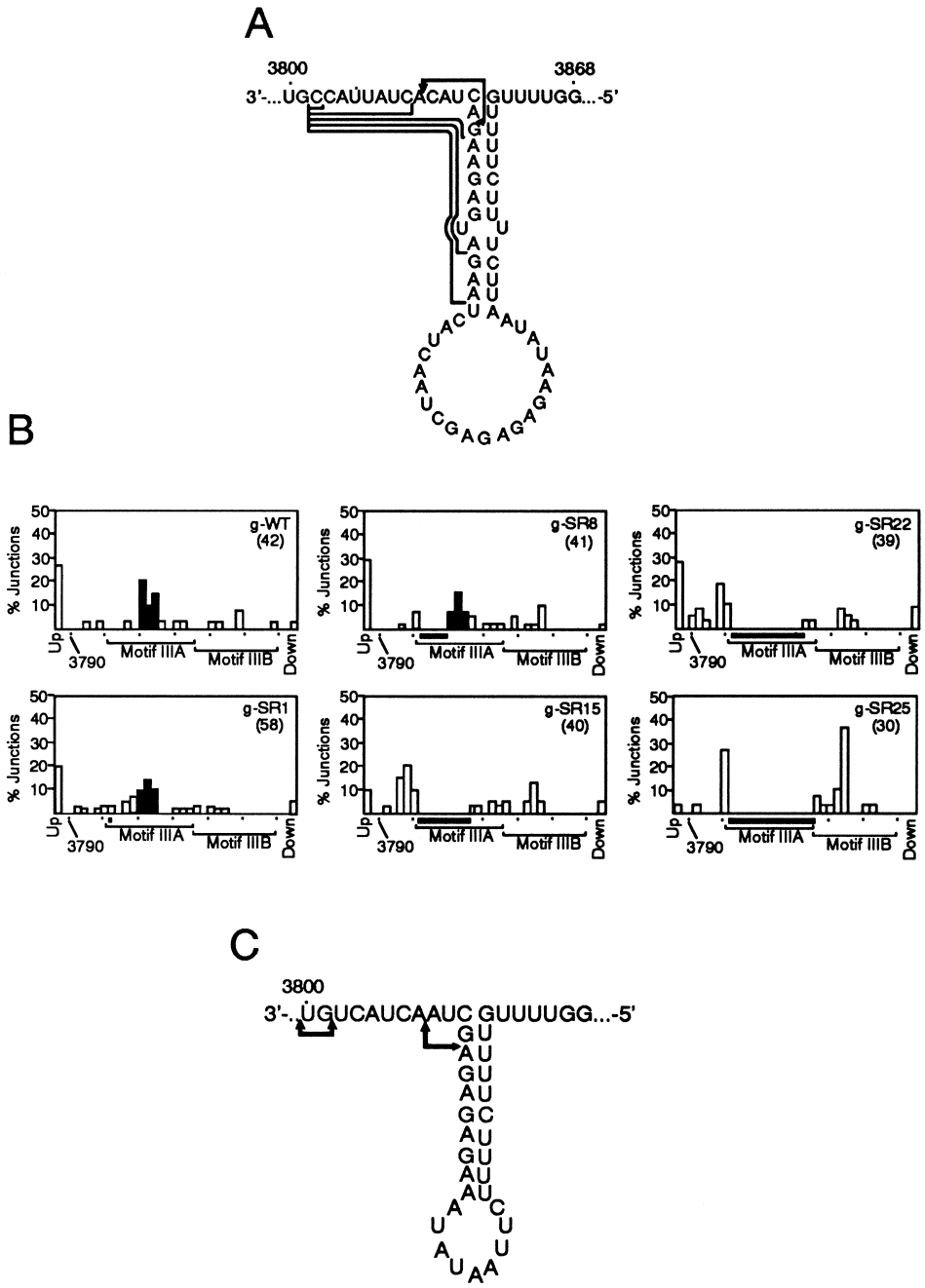

Figure 6. Effect of deletions in the 3'-side stem sequence and upstream single-stranded region (motif IIIA) on junction profiles of recombinants. A, Deletions of 1 through 25 bases are denoted by lines in the putative structure. Major crossover region in recombinants synthesized using wild-type g-RNA is between the joined arrowheads. B, Junction profiles of recombinants amplified from 2 plants previously inoculated with sat-RNA D and the deletion mutants shown in A. Numbers in the names of the g-RNA constructs (e.g., g-SR1) denote the number of residues deleted. Horizontal bars beneath the junction profiles indicate the location and extent of the deletions. The number of non-sibling recombinants that were sequenced to produce the junction profile is in parentheses. Junctions were grouped together as described for Figure 1D. Filled vertical bars denote junctions at the major recombination site in the region, located between the arrowheads in A. C, Putative structure predicted to form by computer secondary structure algorithm for g-SR25 (25-base deletion). New major crossover sites are encompassed by the joined arrowheads.

motifs increased, from an average of 7.6% for g-WT and g-RNA containing smaller deletions to 38% for the 21-base deletion. These results indicate that deletions in the loop region affect the distribution of recombination junctions only when the deletions are predicted to disrupt the bottom of the wild-type stem.

Effect of deleting the 3' single-stranded region on the location of crossover sites in g-RNA

Deletion of base 3802 generated a unique *Sna*BI site that was then used to construct a series of deletions extending through the putative single-

stranded region and into the stem of the hairpin (Figure 6A). Recombinations generated in plants inoculated with g-RNA containing deletions of one or eight bases (g-SR1 and g-SR8, respectively) did not contain junctions that differed substantially from wild-type, indicating that sequence 3' of the preferred junction site is not required for normal crossover site selection (Figure 6B). Deletion of 15 or 22 bases eliminated the preferred crossover site at the base of the stem and shortened the hairpin. The majority of recombinants using these constructs had crossover sites just 3' of the deletion, indicating that the composition of the sequence at the preferred crossover site is less important than the position of the sequence in relationship to a 5' sequence or structure. Recombinants generated using g-SR25, which contained a nearly complete deletion of the motif IIIA sequence, had a shift in crossover sites with the majority now in the centre of motif IIIB. One explanation for the dramatic shift in recombination junctions using g-SR25 is that motifs IIIA and IIIB are imperfect repeats allowing the motif IIIB sequence to replace the deleted motif IIIA sequence in the stem. This creates a new hairpin, and the majority of junctions are located at the 3' base of the new stem (Figure 6C).

Effect of deletions 5' of motifs IIIA and IIIB on location of junction sites in g-RNA

To determine if a hairpin containing the motif IIIA or IIIB sequence is important for recombination or whether the motif sequences alone are responsible for the normal junction profile, deletions in the uracil-rich portion of the putative stem were generated to disrupt the hairpin without altering the motifs. Short deletions within the uracil-rich region were not predicted to eliminate the hairpin due to partial replacement of the deleted bases with the four uracil residues that are not base-paired and are located just 5' of the wild-type stem. Short deletions, however, should affect the length of the stem. Deletion of the complete 5'-side of the stem is predicted by the computer secondary structure algorithm to eliminate the hairpin, which would affect the junction profile if the hairpin is involved in crossover site selection. Deletions of six, eight and 12 bases were generated in the uracil-rich region using the *Afl*II site at position 3847 in g-RNA construct g-U3847G (Figure 7A). Deletion of six and eight bases (g-AR6 and g-AR8, respectively), while eliminating the preferred recombination site at position 3815 found for the parental U3847G construct, still maintained a similar distribution of junction sites within the upstream, motif IIIA, motif IIIB and downstream groupings (Figure 7B). Curiously, all the recombinants sequenced from one of the two plants inoculated with the eight-base deletion construct had between three and 12 additional uracil residues at the position of the deletion, possibly the result of replicase stuttering on the four adjoining uracil residues (Carpenter *et al*., 1991*b*). The distribution of junction sites within the major groupings for recombinants with additional uracil residues did not differ, however, from recombinants without additional bases (data not shown).

Deletion of 12 bases (g-AR12), predicted to substantially disrupt the hairpin without altering the motif III sequences, had a dramatic effect on the location of junction sites. Of the 43 recombinants sequenced, all but three contained crossover sites 5' of the deletion (Figure 7B). The new favored position for crossover sites corresponds to the base of a hairpin predicted by both computer algorithm and chemical and enzymatic determination to exist just 5' of the motif IIIA/B hairpin (Figure 7A). Only two recombinants had junctions within motifs IIIA and IIIB; both of these recombinants also contained 21 additional uracil residues at the position of the deletion (no other recombinant generated using g-AR12 contained additional uracil residues). The ''reversion'' of the 12-base deletion by addition of 21 uracil residues would reform the (–)-strand stem, but not the (+)-strand stem due to the formation of four G-U base-pairs in the (–) orientation that become C-A base-pairs in the (+)-strand orientation. The finding that normal junctions are restored with the additional uracil residues, together with the results below that demonstrate the importance of the hairpin in crossover site selection in the region, support our model that suggests recombination is occurring during (+)-strand synthesis. Results from these deletions suggest that the motif sequences alone are not responsible for the normal wild-type junction profile since disruption of the hairpin without altering the motif sequences results in a significant alteration in the location of crossover sites.

Effect of compensatory mutations on the location of junctions sites in g-RNA

Single base changes in the stem of hairpin B in sat-RNA C eliminated recombination while compensatory changes that reformed the hairpin restored normal crossing over (Cascone *et al*., 1993). To conduct similar alterations in the motif IIIA/B hairpin of g-RNA, the four residues at the base of the putative stem were replaced with 3'-GUUU on the 3'-side and/or AAAC-5' on the 5'-side (Figure 8A); g-RNA containing only the 3'-GUUU (g-GUUU) alteration is predicted by computer algorithm to have a hairpin shortened by these four bases. If recombination is favored to occur in a single-stranded region at the base of a hairpin, then the favored crossover site should be shifted a few bases 5' of the wild-type position. The junction profile for g-GUUU was consistent with this prediction; 32% of the junction sites were located in an eight-base region just 5' of the preferred wild-type position (Figure 8B). The CAAA-5' alteration on the 5'-side of the hairpin was predicted by computer structural modeling to form a substantially altered hairpin with an 18-base stem and two bulge loops, beginning with

base 3814 and terminating with base 3869. Plants inoculated with g-CAAA had an unusual junction profile, with 81% of junction sites now located in the region upstream of the motifs.

g-RNA containing both alterations (g-CAAA + GUUU) should possess a stem similar to the wild-type stem. The junction profile for g-CAAA + GUUU was similar to wild-type g-RNA with the majority of junctions in the single-stranded sequence at the base of the stem (Figure 8B). However, g-RNA containing the compensatory mutations did not produce any junctions in the motif IIIB (loop) region out of 44 recombinants analyzed, unlike g-WT, which had 14% of the crossover sites in motif IIIB. The stem containing the compensatory changes differed from the wild-type stem by the conversion of a G-U base-pair to an A-U base-pair, resulting in a slightly more thermodynamically stable stem. One of the point mutants generated for this study (g-G3816A) had the identical G-U to A-U base-pair alteration; the junction profile for this mutant also had fewer crossovers in the loop region, only 2% (1 out of 46; see Figure 4).

Results of the compensatory mutation study, combined with results of the deletion studies described above, strongly suggest that normal crossover site selection in the motif IIIA / B region of the g-RNA is influenced by the presence of a stable

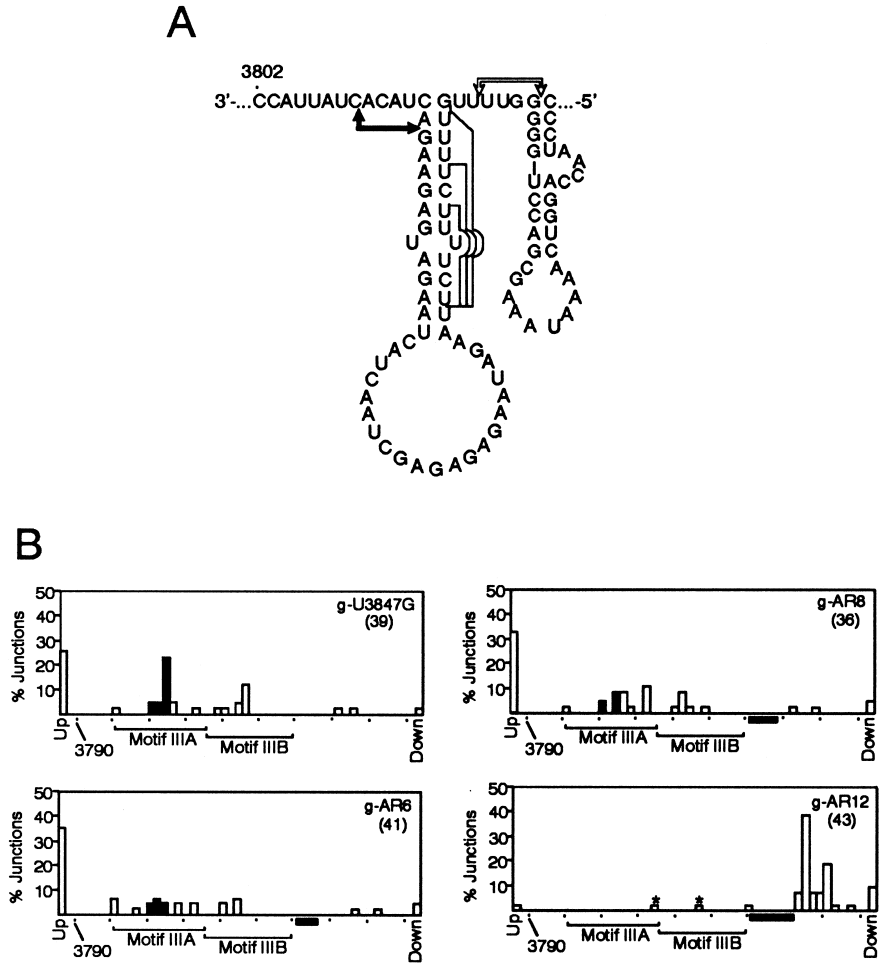

Figure 7. Effect of deletions in the uracil-rich region downstream from the motifs on junction profiles of recombinants. A, Deletions of 6, 8 and 12 bases are denoted by lines in the putative structure. Major crossover region in recombinants synthesized using wild-type g-RNA is between the joined, black arrowheads. The 5' hairpin was predicted to occur in the wild-type g-RNA by computer algorithm and chemical and enzymatic modification studies. Open, joined arrowheads denote new major crossover site that occurs using g-RNA with a 12-base deletion (g-AR12), which is predicted to eliminate the 3' hairpin. B, Junction profiles of recombinants amplified from 2 plants previously inoculated with sat-RNA D and the deletion mutants shown in A. g-U3847 is the parental construct for these deletions. The numbers in the names of the g-RNA constructs (e.g. g-AR12) denote the number of residues deleted. Horizontal bars beneath the junction profiles indicate the location and extent of the deletions. The number of non-sibling recombinants that were sequenced to produce the junction profile is in parentheses. Junctions were grouped together as described for Figure 1D with the exception that additional downstream 2-base groupings are shown to better convey the altered junction profile of g-AR12. Filled vertical bars denote junctions at the major recombination site in the region, located between the joined, black arrowheads in A. The asterisks (*) in the g-AR12 junction profile indicate that these recombinants contained 21 uracil residues immediately 3' of the deletion that were not in the parental template.

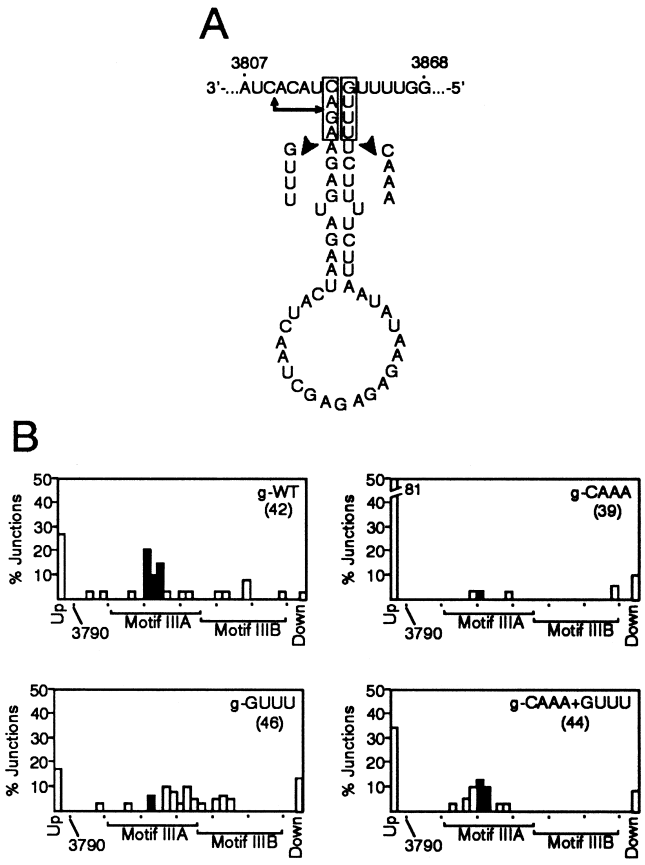

Figure 8. Effect of compensatory stem alterations on the junction profiles. A, Location and composition of the base alterations. Major crossover region in recombinants synthesized using wild-type g-RNA is between the joined arrowheads. B, Junction profiles of recombinants amplified from 2 plants previously inoculated with sat-RNA D and the mutants shown in A. The names of the g-RNA constructs (e.g. g-GUUU) denote the new altered residues. The number of non-sibling recombinants that were sequenced to produce the junction profile is in parentheses. Junctions were grouped together as described for Figure 1D. Filled vertical bars denote junctions at the major recombination site in the region, located between the arrowheads in A.

hairpin and that crossing over preferentially occurs at the 3' base of the hairpin ((–)-strand orientation). This is strikingly similar to recombination in the motif I region of sat-RNA C, where the junctions are also located at the 3'-base of a stable hairpin (Cascone *et al*., 1993). However, as is the case with sat-RNA C, the connection between the sequence motifs and recombination remains unclear. Deletion of the uracil-rich sequence 5' of motif IIIA/B resulted in a substantial alteration of the junction profile (see profile for g-AR12, Figure 7B), suggesting that the presence of the motifs without the structure encompassing the motifs is insufficient to target crossing over to the normally preferred site (positions 3809 to 3815). However, junctions were strongly favoured to occur at the base of a hairpin now located just 5' of the motifs following the AR12 deletion. It is possible that favored sites for reinitiation of transcription following template switching by the viral replicase requires a sequence motif similar to sequences at the 5'-ends of the viral and subviral RNAs, and a nearby stem-loop. The size and/or stability of the stem-loop may be important, since the large hairpin predicted to form for the g-CAAA mutant did not target crossing over to the base of the stem, but rather 81% of the junctions were found in the upstream region. We are continuing to investigate the role of the motif sequences and hairpin size and stability in recombination between TCV-associated RNAs.

The motif III region is important for normal accumulation of g-RNA in plants and protoplasts

One question concerning recombination among TCV-associated RNAs is why sequences have been maintained in the non-coding region of g-RNA that lead to recombination with sat-RNA D resulting in the production of molecules of seemingly limited value. One possibility is that this region of the TCV g-RNA is involved in processes other than recombination. During these experiments, we observed that the amount of g-RNA accumulating in plants that contained the larger deletions was inversely proportional to the size of the deletions (data not shown). To pursue this observation, g-RNA containing deletions in the motif IIIA/B region were inoculated onto three to four plants and the accumulation of g-RNA two weeks post-inoculation determined by RNA gel blot analysis (Figure 9A and B). Deletion of 15 bases in motif IIIA or 17 bases in motif IIIB, resulted in a reduction in g-RNA accumulation to 75% or 50% of wild-type, respectively. Deletion of 26 bases or greater in the motif IIIA region reduced the accumulation of g-RNA to undetectable levels while deletion of 25 and 29 bases in the motif IIIB region reduced the accumulation of g-RNA to less than 10% of wild-type. Deletion of 12 bases in the uracil-rich sequence downstream of the motifs reduced the level of g-RNA accumulating in plants by 88%, while deletion of an additional 8 bases reduced the amount of g-RNA to below the level of detection. These results suggest that the region involved in recombination with sat-RNA D is necessary for accumulation of normal levels of g-RNA in plants and that selection for maintaining these sequences may have occurred.

Since accumulation of g-RNA in whole plants is a combination of RNA stability, replication, packaging and movement, we inoculated several deletion constructs onto turnip leaf protoplasts (single plant cells), which eliminate the need for packaging and movement in the accumulation of RNA molecules in order to focus on whether these deletions affect replication/stability of g-RNA. As shown in Figure 9C, the accumulation of g-RNA with deletions of 21 and 24 bases in motif IIIB in protoplasts was reduced to about 10% of wild-type levels indicating

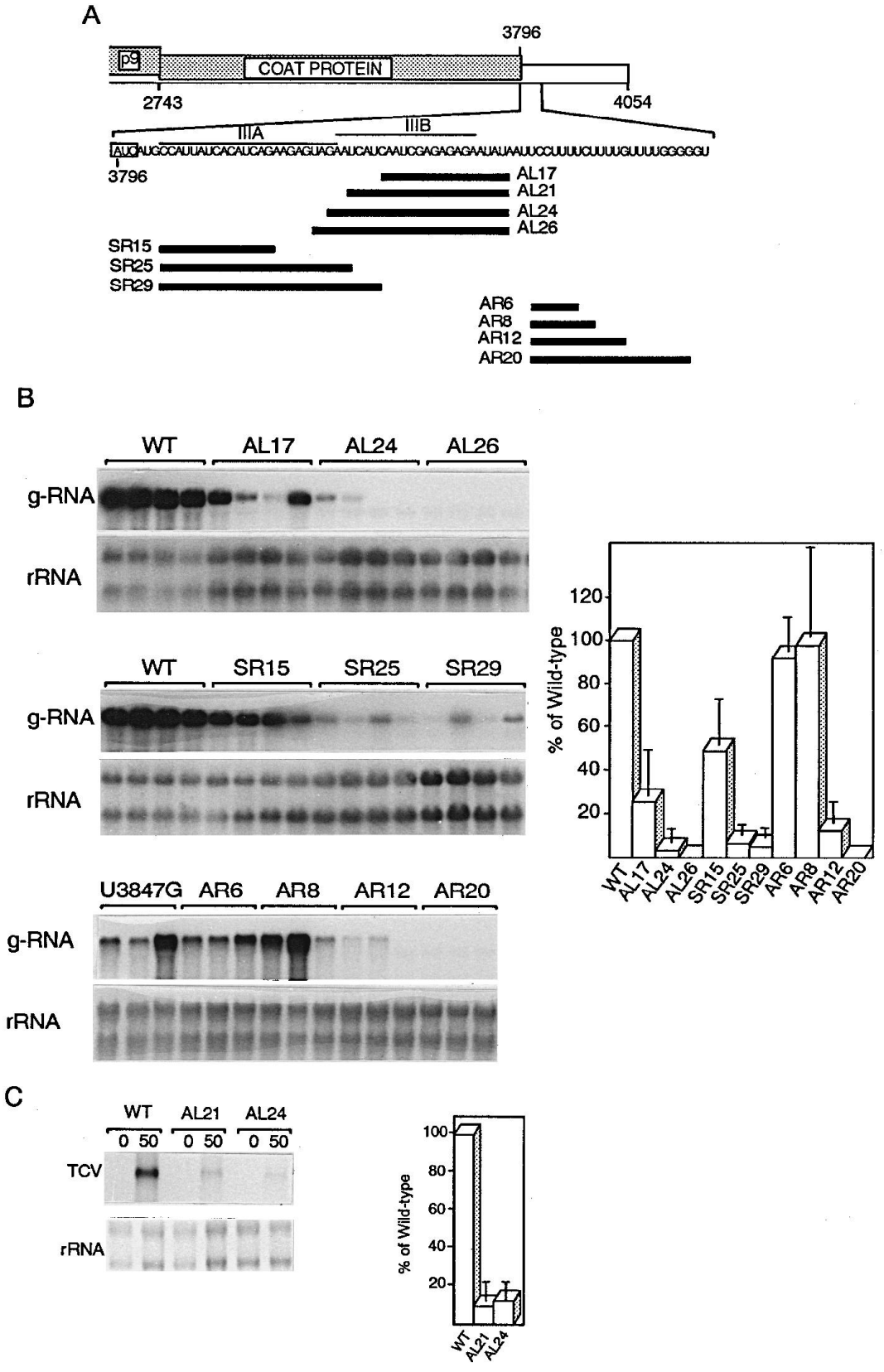

Figure 9. Accumulation of g-RNA containing deletions in the motif III region in plants and protoplasts. A, Location of the deletions (filled bars) in the transcripts used in these experiments. B, RNA gel blot analysis of g-RNA levels in uninoculated turnip leaves. Left, Plants (3 or 4) were inoculated with the transcripts designated above the lanes. After hybridization with the g-RNA-specific probe, blots were stripped and rehybridized with a rRNA-specific probe. Right, Levels of g-RNA were normalized and compared with levels of wild-type or U3847G, the parental transcript for the AR series. C, RNA gel blot analysis of g-RNA levels in turnip protoplasts. RNA was isolated at 0 or 50 h post-inoculation and hybridized with g-RNA and then rRNA probes (left). Only 1 of 2 experiments is shown. Right, Levels of g-RNA were normalized to rRNA levels and compared with wild-type.

that this region is important for either stability or replication of the RNA. Since large (>1000 base) deletions of the coat protein and movement protein open reading frames did not affect replication of g-RNA in protoplasts (and by analogy, stability; Hacker *et al*., 1992), it is likely that replication of g-RNA is affected by the 12 to 26 base deletions and not the stability of g-RNA. One possibility is that this region is a site of recognition by the TCV RdRp, involved in both initiation of replication from the 3'-end of the (+) or (–)-strand, and re-initiation of RNA synthesis following a crossing over event. We are currently analyzing whether specific sequences / structures or the spacing between two elements in the motif III region is required for viral replication and whether production of (+) or (–)-strands is affected.

Materials and Methods

Production of *in vitro* transcripts for TCV g-RNA and sat-RNA D

The cloning and sequencing of a full-length cDNA of TCV g-RNA from the laboratory isolate designated TCV-M (Li *et al*., 1989) will be described elsewhere. The full-length cDNA was inserted in both orientations into the *Sma*I site of pT7E19(+) (Petty, 1988), which had been modified to juxtapose the bacteriophage T7 RNA polymerase promoter and the *Sma*I site, generating pT7TCV(+) and pT7TCV(–). The CCCG residues between the T7 promoter and the exact 5'-end of the g-RNA sequence in pT7TCV(+) (containing the TCV cDNA oriented with the 5'-end of the corresponding g-RNA sequence adjacent to the T7 promoter) were deleted using a gapped duplex procedure (Morinaga *et al*., 1984) and an oligonucleotide with sequence 5'-CGACTCACTATAGGTAATCTGCAA-3' producing plasmid pT7TCVms. *In vitro* transcription from *Sma*I-digested pT7TCVms using T7 RNA polymerase (Li & Simon, 1991) produced transcripts containing the exact 5'- and 3'-ends of the natural genomic RNA. A full-length sat-RNA D cDNA clone was generated by ligating the 31 bp *Sma*I-*Sna*BI 5'-end and the 89 bp *Nco*I-*Sma*I 3'-end fragment from a sat-RNA D partial dimer clone (with precise junction sequences between monomeric units; Carpenter *et al*., 1991*b*) to the 74 bp *Sna*BI-*Nco*I interior region of sat-RNA D contained in cDNA clone pD22 (Simon & Howell, 1986). The full-length sat-RNA D cDNA was then cloned into the *Sma*I site of pPM2 (Simon *et al*., 1988) forming pPM2D; the *Sma*I site is just downstream from an *Escherichia coli* RNA polymerase promoter. Digestion of pPM2D with *Eco*RI and *in vitro* transcription using *E. coli* RNA polymerase (Carpenter *et al*., 1991a) resulted in the synthesis of transcripts with exact 5'-ends and seven plasmid-derived bases at the 3'-end. During the course of these experiments, the full-length sat-RNA D cDNA was inserted into the modified pT7E19(+) (Petty, 1988) forming pT7D. pT7D, when digested with *Bam*HI and subjected to transcription *in vitro* using T7 RNA polymerase, also produced transcripts containing the exact 5'-end and seven additional bases at the 3'-end. Transcripts generated from either pPM2D or pT7D were amplified efficiently when co-inoculated with TCV g-RNA transcripts onto turnip plants (data not shown).

Generation of mutations in TCV g-RNA

Point mutations in the g-RNA cDNA sequence in pT7TCVms were obtained using the method of Kunkel (1985). The oligonucleotide primers used to create the *Afl*II site at position 3841 (the altered base is underlined) and the *Sna*BI site at position 3801 (the position of the deleted base is underlined) were 5'-GCTCTCTCTTATCTTAAGAAAA-GAAAAC-3' and 5'-CAGAATTTAGTAC_GTAATAGTG-TAGTCT-3', respectively. To generate g-GUUU, g-CAAA, and g-GUUU + CAAA, oligonucleotides 5'-GTACGGTAA-TAGTGTACAAATCTCATCTTAGTAGTTAG-3' and 5'-CT-TATATTAAGAAAAGATTTGAAAACCCCCAGGTC-3', respectively, were used either separately or combined. Random point mutations were introduced into the motif III region of TCV g-RNA (bases 3799 to 3858) using a series of overlapping 40-mer oligonucleotides (3789 to 3838, 3819 to 3858 and 3839 to 3878) each of which was synthesized with mixed bases in the central 20 positions at a ratio of 95% wild-type to 1.33% each of the other bases. Plasmids containing mutated g-RNA sequences were detected by DNA sequencing (Sequenase, US Biochemicals). Unidirectional *Bal*31 deletions were generated using the TCV clones containing the *Afl*II and *Sna*BI sites. After linearization at the aforementioned sites, the DNA was treated with the slow form of *Bal*31 (IBI) for various periods according to the supplier's suggested procedures. After the reactions were terminated, ends were made flush with DNA polymerase I large fragment (Biolabs). The *Sna*BI and *Afl*II fragments were digested with *Xba*I and then ligated to the coresponding, untreated *Sna*BI-*Xba*I or *Afl*II-*Xba*I fragments.

Detection of sat-RNA D/g-RNA recombinants

For all constructs, except for the deletion series presented in Figure 5A, 3 μg of full-length TCV g-RNA transcripts were combined with 3 μg of sat-RNA D transcripts and inoculated onto each of three turnip cultivar 'Just Right' plants. After three weeks, sat-RNA D / g-RNA recombinant molecules were amplified from total RNA prepared from uninoculated leaves (Simon *et al*., 1988) by RT-PCR using primers homologous to positions 104 to 118 of sat-RNA D and complementary to positions 3943 to 3962 of TCV g-RNA as described (Carpenter & Simon, 1994). PCR products were not detected in control reactions using the TCV- and sat-RNA D-specific primers and approximately 1 μg of genomic and sat-RNA D transcripts (data not shown). For the deletion series in Figure 5A, 3 μg of TCV g-RNA transcripts were inoculated onto turnip plants without sat-RNA D transcripts. At 18 days postinoculation, total leaf RNA was isolated and 5 μg from a single plant was combined with 3 μg of sat-RNA D transcripts for inoculation on each of three new plants. At 20 days postinoculation, recombinants were sequenced as described above.

RNA structure probing

Minus-strand RNA transcripts of sat-RNA C, containing the exact 5'-end and five plasmid-derived bases at the 3'-end, were synthesized from *Bam*HI-digested pT7CAM(–), which contains a full-length cDNA of sat-RNA C (Cascone *et al*., 1993) downstream from a bacteriophage T7 promoter. Minus-strand transcripts of TCV g-RNA, containing 697 bases from the coat protein ORF through the 3' terminus (of the corresponding (+)-strand), were synthesized from *Hin*dIII-digested pT7TCV(–). Transcripts contained no additional vector sequences. From 2 to 3 μg of gel-purified transcripts in the appropriate chemical modification or enzyme digestion buffers were heated at 56°C for five minutes and then

slowly cooled to 25°C. Carbodiimide (CMCT; Aldrich Chemical Co.) modification was performed as described by Moazed *et al*. (1986), diethylpyrocarbonate (DEPC; Sigma) and dimethylsulfate (DMS; Sigma) modifications were performed as described (Peattie & Gilbert, 1980). Enzymatic cleavages were performed essentially as described (Shelness & Williams, 1985), incubating transcripts in 0.1 to 0.2 unit of RNase T_1 (Sigma) or 0.25 to 0.5 unit of RNase CV1 (Pharmacia) for two minutes at 37°C. The chemically modified or ribonuclease-cleaved transcripts were extracted with phenol, precipitated in ethanol and then annealed to 10 pmol of primer by incubation at 90°C and slow-cooling to 25°C. The sat-RNA C-specific primer was homologous to positions 57 to 78 and the TCV-specific primer was homologous to positions 3773 to 3789. Reverse transcription was performed using 0.25 unit of AMV reverse transcriptase (USB) in 3.8 μM dCTP, dGTP and dTTP, and 5 μCi of [α-^{35}S]dATP, in buffer supplied by the manufacturer. After five minutes at 45°C, 0.5 μmol of all four dNTPs were added and incubation continued for an additional five minutes followed by separation of cDNA on a urea / 7% (w / v) polyacrylamide sequencing gel. Dideoxy-sequencing of untreated, primer-extended transcripts was performed using an RNA sequencing kit from US Biochemicals.

Protoplast preparation and inoculation

One to two inch leaves from young turnip plants were sterilized in 5% (v / v) bleach, 0.5 (v / v) Tween-20 (Aldrich) for three minutes followed by three rinses with sterile water. Following removal of the lower epidermis and midribs, leaves were cut into small pieces and digested with 1% (w / v) cellulase (10,500 units / g; CalBiochem.) and 0.05% (w / v) macerase (3500 units / g, CalBiochem.) in 0.6 M mannitol for two hours at 25°C with gentle shaking. Protoplasts were filtered through a 0.53 μm nylon mesh and centrifuged for three minutes at 10 ***g***. The protoplasts pellet was washed four or five times in 0.6 M mannitol followed by centrifugation. Protoplasts (3×10^6) were inoculated with approximately 27 μg of TCV g-RNA transcripts in 1.2 ml of inoculation solution (3 mM $CaCl_2$, 40% PEG 1500; Boehringer Mannheim). After 15 seconds at 25°C, the inoculation mixture was combined with 12 ml of 0.6 M mannitol and 1 mM CaCl. The diluted inoculation mixture was incubated on ice for 15 minutes and then centrifuged at 34 ***g*** for five minutes. Pelleted protoplasts were washed three times in 0.6 M mannitol and 1 mM $CaCl_2$ and then resuspended in 6 ml of wash solution supplemented with 6 μg / ml nystatin and 20 μg / ml carbenicillin. Following incubation for 50 hours at 25°C under fluorescent lights (1500 lux), protoplasts were centrifuged at 10 g, combined with 0.2 ml RNA extraction buffer (50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 100 mM NaCl, 1% (w / v) SDS), extracted twice with phenol and the RNA precipitated in ethanol.

RNA gel blots

RNA gel blots were prepared and hybridized as described (Carpenter *et al*., 1994) except that for hybridizations to TCV g-RNA, a ^{32}P-labeled oligonucleotide complementary to bases 3943 to 3962 was used, and hybridizations were performed in buffer containing 25% (v / v) formamide at 32°C and the final wash temperature was 25°C.

Acknowledgements

We are grateful to M. Farley for technical assistance. This work was supported by grants MCB-9419303 and MCB-9315948 from the National Science Foundation to A.E.S.

References

Banner, L. R. & Lai, M. M. C. (1991). Random nature of coronavirus RNA recombination in the absence of selection pressure. *Virology*, **185**, 441–445.

Banner, L. R., Keck, J. G. & Lai, M. M. C. (1990). A clustering of RNA recombination sites adjacent to a hypervariable region of the peplomer gene of murine coronavirus. *Virology*, **175**, 548–555.

Borukhov, S., Sagitov, V. & Goldfarb, A. (1993). Transcript cleavage factors from *E. coli*, *Cell*, **72**, 459–466.

Bujarski, J. J. & Dzianott, A. M. (1991). Generation and analysis of nonhomologous RNA-RNA recombinants in brome mosaic virus: sequence complementarities at crossover sites. *J. Virol.* **65**, 4153–4159.

Carpenter, C. D. & Simon, A. E. (1994). Recombination between plus and minus strands of turnip crinkle virus. *Virology*, **201**, 419–423.

Carpenter, C. D., Cascone, P. J. & Simon, A. E. (1991*a*). Mutations in a satellite RNA of turnip crinkle virus result in addition of poly(U) in vivo. *Virology*, **183**, 595–601.

Carpenter, C. D., Cascone, P. J. & Simon, A. E. (1991*b*). Formation of multimers of linear satellite RNAs. *Virology*, **183**, 586–594.

Carpenter, C. D., Kreps, J. A. & Simon, A. E. (1994). Genes encoding glycine-rich *Arabidopsis thaliana* proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm. *Plant Physiol.* **104**, 1015–1025.

Carrington, J. C., Morris, T. J., Stockley, P. G. & Harrison, S. C. (1987). Structure and assembly of turnip crinkle virus. IV. Analysis of the coat protein gene and implications of the subunit primary structure. *J. Mol. Biol.* **194**, 265–276.

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.

Cascone, P. J., Haydar, T. & Simon, A. E. (1993). Sequences and structures required for recombination between virus-associated RNAs. *Science*, **290**, 801–805.

Duarte, E. A., Novella, I. S., Ledesma, S., Clarke, D. K., Moya, A., Elena, S. F., Domingo, E. & Holland, J. J. (1994). Subclonal components of consensus fitness in an RNA virus clone. *J. Virol.* **68**, 4295–4301.

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.

Holland, J., Spindler, F., Horodyski, F., Drabau, E., Nichol, S. & Van de Pol, S. (1982). Rapid evolution of RNA genomes. *Science*, **215**, 1577–1585.

Jarvis, T. C. & Kirkegaard, K. (1991). The polymerase in its labyrinth: mechanisms and implications of RNA recombination. *Trends Genet.* **7**, 186–191.

King, A. M. Q. (1988). Preferred sites of recombination in poliovirus RNA: an analysis of 40 intertypic cross-over sequences. *Nucl. Acids Res.* **29**, 921–928.

Kirkegaard, K. & Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell*, **47**, 433–443.

Koetzer, C. A., Parker, M. M., Ricard, C. S., Sturman, L. S. & Masters, P. S. (1992). Repair and mutagenesis of the

genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. *J. Virol.* **66**, 1841–1848.

Kuge, S., Saito, I. & Nomoto, A. (1986). Primary structure of poliovirus defective-interfering particle genomes and possible generation mechanisms of the particles. *J. Mol. Biol.* **192**, 473–487.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Nat. Acad. Sci., U.S.A.,* **82**, 488–492.

Lai, M. M. C. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**, 61–79.

Levis, R., Schlesinger, S. & Huang, H. V. (1990). Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. *J. Virol.* **64**, 1726–1733.

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

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.

Li, Y. & Ball, L. A. (1993). Nonhomologous RNA recombination during negative-strand synthesis of flock house virus RNA. *J. Virol.* **67**, 3854–3860.

Makino, S., Keck, J. G., Stohlman, S. A. & Lai, M. M. C. (1986). High frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**, 729–737.

Miller, W. A., Dreher, T. W. & Hall, T. C. (1985). Synthesis of brome mosaic virus subgenomic RNA in vitro by internal initiation on (–) sense genomic RNA. *Nature (London),* **313**, 68–70.

Moazed, D., Stern, S. & Noller, H. F. (1986). Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* **187**, 399–416.

Morinaga, Y., Franceschini, T., Inouye, S. & Inouye, M. (1984). Improvement of oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA. *Bio/Technology,* **2**, 636–639.

Mote, J., Jr, Ghanouni, P. & Reines, D. (1994). A DNA minor groove-binding ligand both potentiates and arrests transcription by RNA polymerase II. *J. Mol. Biol.* **236**, 725–737.

Nagy, P. D. & Bujarski, J. J. (1993). Targeting the site of RNA-RNA recombination in brome mosaic virus with antisense sequences. *Proc. Nat. Acad. Sci., U.S.A.,* **90**, 6390–6394.

Peattie, D. A. & Gilbert, W. (1980). Chemical probes for higher-order structure in RNA. *Proc. Nat. Acad. Sci., U.S.A.,* **77**, 4679–4682.

Petty, I. T. D. (1988). A plasmid vector for cloning directly at the transcription initiation site of a bacteriophage T7 promoter. *Nucl. Acids Res.* **16**, 8733.

Rao, A. L. N. & Hall, T. C. (1993). Recombination and polymerase error facilitate restoration of infectivity in brome mosaic virus. *J. Virol.* **67**, 969–979.

Re, G., Morgan, E. M. & Kingsbury, D. W. (1985). Nucleotide sequences responsible for generation of internally deleted Sendai virus defective interfering genomes. *Virology,* **146**, 27–37.

Romanova, L. I., Blinov, V. M., Tolskaya, E. A., Viktorova, E. G., Kolesnikova, M. S., Guseva, E. A. & Agol, V. I. (1986). The primary structure of crossover regions of intertypical poliovirus recombinants: a model of recombination between RNA genomes. *Virology,* **155**, 202–213.

Roux, L., Simon, A. E. & Holland, J. J. (1991). Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. *Advan. Virus Res.* **40**, 181–211.

Shelness, G. S. & Williams, D. (1985). Secondary structure analysis of apolipoprotein II mRNA using enzymatic probes and reverse transcriptase. *J. Biol. Chem.* **260**, 8637–8646.

Simon, A. E. & Bujarski, J. J. (1994). RNA-RNA recombination and evolution in infected plants. *Annu. Rev. Phytopath.* **32**, 337–362.

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.

Simon, A. E., Engel, H., Johnson, R. P. & Howell, S. H. (1988). Identification of regions affecting virulence, RNA processing and infectivity in the virulent satellite of turnip crinkle virus. *EMBO J.* **7**, 2645–2651.

Strauss, J. H. & Strauss, E. G. (1988). Evolution of RNA viruses. *Annu. Rev. Microbiol.* **42**, 657–683.

Surratt, C. K., Milan, S. C. & Chamberlin, M. J. (1991). Spontaneous cleavage of RNA in ternary complexes of Escherichia coli RNA polymerase and its significance for the mechanism of transcription. *Proc. Nat. Acad. Sci., U.S.A.,* **88**, 7983–7987.

Tolskaya, E. A., Romanova, L. I., Blinov, V. M., Viktorova, E. G., Sinyakov, A. N., Kolesnikova, M. S. & Agol, V. I. (1987). Studies on the recombination between RNA genomes of poliovirus: The primary structure and nonrandom distribution of crossover regions in the genomes of intertypic poliovirus recombinants. *Virology,* **161**, 54–61.

van der Most, R. G., Heijnen, L., Spaan, W. J. M. & de Groot, R. J. (1992). Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucl. Acids Res.* **20**, 3375–3381.

Watanabe, Y. & Okada, Y. (1986). In vitro viral RNA synthesis by a subcellular fraction of TMV-infected tobacco protoplasts. *Virology,* **149**, 64–73.

White, K. A. & Morris, T. J. (1994). Nonhomologous RNA recombination in tombusviruses: generation and evolution of defective interfering RNAs by stepwise deletions. *J. Virol.* **68**, 14–25.

Zhang, C. & Simon, A. E. (1994). Effect of template size on accumulation of defective interfering RNAs in protoplasts. *J. Virol.* In the press.

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

Zuker, M., Jaeger, J. A. & Turner, D. H. (1991). A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucl. Acids Res.* **19**, 2707–2714.

Edited by J. Karn

(Received 11 July 1994; accepted 25 October 1994)