Journal of General Virology (1997), **78**, 525–534. Printed in Great Britain

Genome organization and gene expression of saguaro cactus carmovirus

Ziming Weng and Zhongguo Xiong

Department of Plant Pathology, University of Arizona, Forbes 204, Tucson, AZ 85721, USA

The complete sequence of the single-stranded, ( )-sense RNA genome of saguaro cactus carmovirus (SCV) has been determined. The 3879 nucleotide genome contains five open reading frames (ORFs). The 5′-proximal ORF encodes a 26 kDa protein (p26) and terminates with an amber codon which is readthrough into an in-frame p57 ORF to generate an 86 kDa fusion protein (p86). Two small, centrally located ORFs encode a 6 kDa protein (p6) and a 9 kDa protein (p9), respectively. The 3′-proximal ORF encodes a 37 kDa (p37) capsid protein (CP). Analysis of the nucleotide and predicted amino acid sequences supports the classification of SCV in the genus *Carmovirus* in the family *Tombusviridae*. All predicted SCV proteins are expressed in an *in vitro* translation system. SCV p26 and the readthrough fusion protein p86 are synthesized from the genomic RNA while p6, p9 and p37 CP ORFs at the 3′ half of the genome are expressed from two subgenomic (sg) RNAs. The 5′ termini of both sg RNAs have been mapped. The large 1614 nucleotide sg RNA contains the p6 and p9 ORFs as the first and the second ORFs respectively from its 5′ end. It directs the synthesis of abundant p6 but a small amount of p9. While a synthetic transcript with the p9 ORF at the 5′ end is a more efficient messenger for p9, no corresponding sg RNA has been identified *in vivo*. The smaller 1396 nucleotide sg RNA contains only the p37 ORF and directs the synthesis of SCV CP.

Introduction

Saguaro cactus virus (SCV) is a small icosahedral virus, approximately 32 nm in diameter. It was first isolated during a cactus virus survey in Arizona (Milbrath & Nelson, 1972; Milbrath *et al.*, 1973). In nature, SCV only infects the giant saguaro cactus (*Carnegiea gigantea*), causing an asymptomatic, systemic infection. The virus can be mechanically transmitted to *Chenopodium amaranticolor* and *C. quinoa*, producing local lesions, and to *C. capitatum* producing systemic vein-clearing and mottle symptoms (Milbrath & Nelson, 1972). Nelson & Tremaine (1975) reported a single RNA component of approximately 4 kb and a major capsid protein (CP) of about 38·9 kDa. Cluster analysis of the amino acid composition of SCV CP showed that it was not closely related to ten other small icosahedral viruses including two carmoviruses, carnation mottle virus (CarMV) and turnip crinkle virus (TCV). SCV antiserum did not cross-react with these two, nor to several other small isometric viruses (Milbrath & Nelson, 1972; Nelson & Tremaine, 1975). However, SCV was classified as a member of the genus *Carmovirus* of the family *Tombusviridae* on the basis of virion morphology and sizes of the genomic RNA and CP (Murphy *et al.*, 1995).

Complete nucleotide sequences have been determined for five carmoviruses including the type member CarMV (Guilley *et al.*, 1985), TCV (Carrington *et al.*, 1989), melon necrotic spot virus (MNSV) (Riviere & Rochon, 1990), cardamine chlorotic fleck virus (CCFV) (Skotnicki *et al.*, 1993) and cowpea mottle virus (CPMoV) (You *et al.*, 1995). These viruses have similar genome organizations and a high degree of nucleotide and amino acid sequence similarities. *Carmovirus* genomes contain five major open reading frames (ORFs). Mutagenesis studies in TCV have demonstrated that all five ORFs are essential for infectivity (Hacker *et al.*, 1992).

In vitro translation studies of CarMV (Carrington & Morris, 1985, 1986) showed that the first two ORFs were expressed from the viral genomic RNA. The 3′-terminal CP ORF was expressed from a 1·5 kb subgenomic (sg) RNA. One of the two centrally located small ORFs, p7 ORF, was expressed from a 1·7 kb sg RNA. Although the remaining small ORF, p9 ORF, encodes an essential protein for carmovirus cell-to-cell movement (Hacker *et al.*, 1992), its protein product has not been identified either *in vitro* or *in vivo*. It has been speculated that

Author for correspondence: Zhongguo Xiong.
Fax +1 520 621 9290. e-mail zxiong@ag.arizona.edu
The GenBank accession number of the sequence reported in this paper is U72332.

the two small, centrally located ORFs might be expressed from the larger sg RNA, but this has yet to be demonstrated. Subgenomic RNAs of similar sizes were also reported in TCV (Carrington *et al*., 1989), MNSV (Riviere & Rochon, 1990) and CPMoV (Kim & Bozarth, 1992).

In this paper, we report SCV genome organization predicted from the complete nucleotide sequence and characterization of its two sg RNAs. In addition, SCV gene expression strategy determined by *in vitro* translation studies of SCV *in vitro* transcripts and fractionated virion RNA is presented. Analysis of the nucleotide sequence and the gene expression strategy supports the classification of SCV in the genus *Carmovirus* of the family *Tombusviridae*. Our data support the hypothesis that the two small, centrally located ORFs of carmoviruses are expressed from a single, bicistronic sg RNA. This is the first report that translation products of all predicted ORFs in *Carmovirus* are expressed *in vitro*.

Methods

■ **Virus purification and RNA extraction.** SCV was propagated in *C. capitatum*. Virus purification and subsequent viral RNA extraction were carried out as described previously (Weng & Xiong, 1995). Total RNA preparations of infected and healthy *C. capitatum* leaves were extracted using the TRIzol RNA extraction kit (GIBCO-BRL).

■ **cDNA synthesis and cloning.** Polyadenylation and subsequent cDNA cloning of SCV RNA into pBR322 were carried out as described previously (Xiong & Lommel, 1989). The resulting cDNA library was initially screened by colony hybridization using ^{32}P-labelled cDNA probes synthesized from viral RNA with random primers and Superscript II reverse transcriptase (RT) (GIBCO-BRL) (Sambrook *et al*., 1989). Several large overlapping clones were subsequently identified by reciprocal Southern hybridization using ^{32}P-labelled DNA probes nick-translated from selected clones. These overlapping clones were subcloned into pBS(+) (Stratagene) at the *Pst*I site.

■ **Nucleotide sequencing.** Clones in pBS(+) were linearized with *Eco*RI and *Kpn*I to produce a 5′ overhang and a 3′ overhang, respectively. Nested deletion clones were generated by digesting the linearized plasmid with ExoIII nuclease followed by S1 nuclease treatment (Xiong & Lommel, 1989). ExoIII deletion clones approximately 150–200 nucleotides apart were selected for sequencing by dideoxynucleotide chain termination method using T7 or T3 primer, [α-thio-^{35}S]dATP and Sequenase version 2.0 (USB).

■ **Computer analysis of sequence data.** Nucleotide and amino acid sequences were analysed using University of Wisconsin Genetics Computer Group software. SCV related sequences were initially identified using the FASTA program. Pair-wise sequence comparisons were subsequently performed by the GAP program using 25 randomizations. The significance of sequence alignments was assessed by calculating adjusted alignment score AS [$AS = (S_o - S_r)/\alpha$, where S_o is the observed score, S_r is the average score of 25 randomizations and α is the standard deviation of S_r] (Koonin, 1991; Tavazza *et al*., 1994).

■ **Northern hybridization.** Northern hybridization of viral RNA and total plant RNA was performed essentially as described by Xiong *et al*. (1993). A ^{32}P-labelled riboprobe complementary to the (+)-sense viral RNA was prepared from a full-length SCV cDNA clone, pSCV15, by *in vitro* transcription with T3 RNA polymerase. After hybridization, membranes were rinsed and then washed for 15 min at 65 °C in 1 × SSC and 0·1% SDS, followed by a 15 min wash at 65 °C in 0·5 × SSC and 0·1% SDS and a 15 min wash at 65 °C in 0·1 × SSC and 0·1% SDS.

■ **Primer extension analysis.** The 5′ termini of two SCV sg RNAs were determined by primer extension using two primers, SCV2663 (5′ AGCTGCACCAGTAGCAGC 3′) complementary to nucleotides 2663–2680 and SCV2401 (5′ TACCCCCATTAGTAGTGC 3′) complementary to nucleotides 2401–2418 of the SCV genome (see Fig. 2). RNA was denatured at 80 °C for 3 min and annealed to the primers by 45 min incubation at 53 °C. Primer extension was carried out by Superscript II RT in the presence of 1·0 μM [α-thio-^{35}S]dATP and 1·5 μM each of dGTP, dCTP and dTTP at 45 °C for 10 min. The extension reaction was completed by further incubation in the presence of 0·1 mM dNTP at 45 °C for 45 min. To identify the precise 5′ termini of the sg RNAs, clone pSCV8 (see Fig. 2) was sequenced using the same primers.

■ **Virion RNA fractionation.** Virion-associated RNA was fractionated by linear-log sucrose density gradient centrifugation (Brakke & Van Pelt, 1970). Viral RNA was denatured by a 3 min incubation in RNA dissociation buffer (0·2 M Tris, 2 mM EDTA, 2% SDS, pH 9·0) at 65 °C and quickly chilled in ice water before loading on the gradients. The gradients were centrifuged at 39000 r.p.m. for 5 h at 4 °C in a Beckman SW41 rotor. Fractions (0·7 ml) were collected with an ISCO gradient fractionator and concentrated by ethanol precipitation. The RNA pellet was resuspended in a small volume of sterile water and used for *in vitro* translation studies.

■ **Synthesis of full-length and 3′-coterminal transcripts.** A full-length infectious SCV cDNA clone (pSCV15) was generated by PCR using two primers. One primer contains T7 RNA polymerase promoter sequence plus the 18 5′-terminal nucleotides. A second primer contains 18 nucleotides complementary to the SCV 3′ terminus plus an *Afl*II restriction site. Digestion with *Afl*II followed by *in vitro* transcription produced full-length infectious SCV RNA transcripts identical to wild-type viral RNA (unpublished data). A series of 3′-coterminal SCV cDNA clones were constructed using three ExoIII deletion subclones generated for sequencing. The first nucleotide of pSCV83, pSCV84 and pSCV85 corresponded to 2220, 2348 and 2463 of the final compiled SCV sequence, respectively. All of these clones were derived from clone pSCV8, which did not possess the 3′-terminal 285 nucleotides of the SCV genome (see Fig. 2). A 1238 nucleotide *Bam*HI–*Afl*II fragment at the 3′ terminus of the full-length cDNA clone was transferred to each of these clones to generate the final 3′-coterminal clones. Uncapped full-length SCV RNA and 3′-coterminal transcripts were synthesized with T7 RNA polymerase as described previously (Xiong & Lommel, 1991).

■ ***In vitro* translation.** Purified viral RNA and synthetic RNA transcripts were translated in a rabbit reticulocyte *in vitro* translation system (Reticulocyte Type 1, Boehringer Mannheim). One μg of viral RNA or synthetic transcripts was translated in 25 μl of the reticulocyte lysate. After incubation at 30 °C for 1·5 h, translation was terminated with 2 vols of Laemmli dissociation buffer (Laemmli, 1970). ^{35}S-labelled *in vitro* translation products were resolved by electrophoresis in either 10% or 18·2% SDS–polyacrylamide gel and visualized by fluorography (Xiong *et al*., 1993).

Results

Analysis of SCV capsid protein and viral RNAs

A single 37 kDa CP was observed when purified virions were fractionated by SDS-PAGE (data not shown). The estimated size of the CP is similar to that reported earlier (Nelson & Tremaine, 1975).

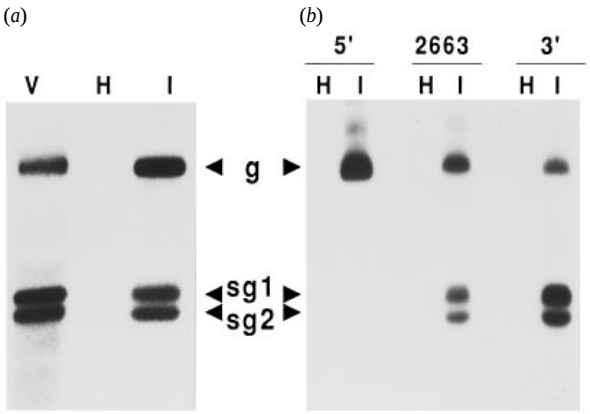

Fig. 1. Northern hybridization analysis of SCV RNA. (*a*) Hybridization of virion RNA (V), total RNA extracted from healthy (H) and SCV-infected *C. capitatum* leaves (I) with a ^{32}P-labelled riboprobe specific for (+)-sense SCV RNA. The riboprobe was synthesized with T3 RNA polymerase from a full-length infectious clone, pSCV15. (*b*) Hybridization of ^{32}P-labelled oligonucleotide probes to total RNA purified from healthy (H) and SCV-infected (I) *C. capitatum*. Oligonucleotides complementary to nucleotides 1–18 (5′), to nucleotides 2663–2680 (2663) and to nucleotides 3862–3879 (3′) were labelled by T4 kinase in the presence of [γ-^{32}P]ATP. Positions of the genomic RNA (g), the large sg RNA (sg1) and the small sg RNA (sg2) are indicated by arrowheads.

Purified SCV virion RNA resolved into three distinct RNA species of approximately 4 kb, 1·6 kb and 1·4 kb in 1·5% agarose gels (Fig. 1). A full-length riboprobe complementary to the (+)-sense viral RNA hybridized to all three RNA species, indicating that all of them were SCV-specific (Fig. 1*a*). To determine whether these RNAs were also present in infected tissues, total RNA preparations were extracted from SCV-infected and mock-inoculated *C. capitatum*. The same riboprobe also hybridized with the three RNAs of identical sizes from SCV-infected but not from mock-inoculated *C. capitatum*. Taken together, these data indicate that SCV produces three (+)-sense viral RNAs during infection and encapsidates them into virions.

Among the three RNA species, the 4 kb RNA was previously identified as SCV genomic RNA (Nelson & Tremaine, 1975). The two smaller RNAs were further analysed by Northern blot hybridization with ^{32}P-labelled oligonucleotide primers: a 5′ primer and a 3′ primer complementary respectively to the first 18 nucleotides and the last 18 nucleotides, and primer SCV2663 complementary to nucleotides 2663–2680 of the SCV genome (Fig. 1*b*). Both the 3′ primer and primer SCV2663 hybridized to the genomic RNA and the two smaller RNAs. The 5′ primer hybridized only with the genomic RNA. These data indicate that the two smaller RNAs are 3′-coterminal with the genomic RNA and are probably SCV sg RNAs. The 1·6 kb and the 1·4 kb sg RNAs are referred to as sg1 and sg2, respectively.

cDNA cloning and sequencing

Most of the SCV-specific clones contained cDNA inserts of about 0·8–1·8 kb in length. Several large overlapping clones were identified by reciprocal hybridization (Fig. 2*b*). These overlapping cDNA clones together represented most of the SCV genome and were sequenced. Except for approximately 300 nucleotides at the 3′ terminus and 4 nucleotides at the 5′ terminus, the SCV cDNA sequence was determined from these overlapping clones. Each clone was sequenced from both directions.

Determination of terminal sequences

The cloning and sequence determination of the 5′ and 3′ termini of SCV RNA genome were reported previously (Weng & Xiong, 1995). The 5′ 200 nucleotides were sequenced from 12 independent clones. Except for the first 4 nucleotides, the 5′-terminal sequence was also confirmed by direct sequencing of the SCV genomic RNA. The 3′-terminal 1200 nucleotides were cloned by anchored cloning (Weng & Xiong, 1995). The 371 nucleotides at the 3′ end of SCV genome were determined by sequencing three independent 3′-terminal clones. The first 10 nucleotides, 5′ GGGTAAGCTG 3′, and the last 7 nucleotides of the SCV genome, 5′ CCCGCCC 3′, are identical to those of CarMV (Guilley *et al.*, 1985).

Complete nucleotide sequence and genome organization

Sequences generated from various cDNA and terminal clones were compiled and assembled with the GCG GEL ASSEMBLE program. The complete sequence of the SCV genome containing 3879 nucleotides has been deposited in the
GenBank database (accession number U72332) and is not duplicated here. SCV genome organization predicted from the sequence is summarized in Fig. 2(*a*). Five major ORFs were identified. The 5′-proximal ORF starts at nucleotide 40 and extends to nucleotide 753 encoding a polypeptide of M_r 26371 (p26). If the p26 amber termination codon is readthrough, the ORF extends to nucleotide 2307 to produce a readthrough protein of M_r 85700 (p86). Alternatively, an in-frame AUG codon downstream from the p26 terminator may potentially initiate the translation of a polypeptide of M_r 57190 (p57), representing the readthrough portion of the p86 protein.

One of the two centrally located and overlapping small ORFs begins at nucleotide 2292 and stops at nucleotide 2471 to produce an M_r 6379 protein (p6). The other begins at nucleotide 2377 and terminates at nucleotide 2631 to produce an M_r 9336 protein (p9). The 3′-proximal ORF encoding a CP of M_r 37289 (p37) begins at nucleotide 2621 and extends to nucleotide 3655. The SCV genome contains 39 non-coding nucleotides at the 5′ end and 224 non-coding nucleotides at the 3′ end.

In addition to the five ORFs described above, two small ORFs with sizes similar to or larger than p6 are also predicted from the sequence. One starts at nucleotide 1460 and ends at

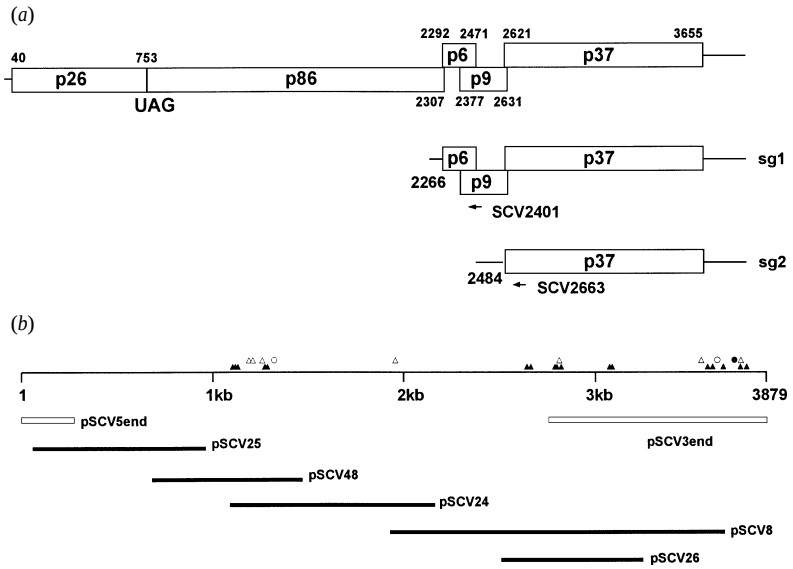

Fig. 2. (*a*) Schematic of SCV genome organization and gene expression strategy. Each open rectangle represents an ORF. ORFs are labelled according to the size of the predicted protein each encodes. Nucleotide positions of initiation and termination codons defining each ORF are labelled either above or below the rectangles. Overlapping rectangles represent overlapping ORFs. The first ORF is in-frame with the second ORF and terminated with an amber termination codon, UAG. Lengths of two sg RNAs (sg1, sg2) and ORFs contained within each sg RNA are shown below the genome representation. Numbers at the front of each subgenomic RNA indicate the nucleotide position corresponding to its 5′ terminus. Locations of two oligonucleotide primers used to map the 5′ termini of the sg RNAs are marked by two short arrows. (*b*) Overlapping cDNA clones used in sequencing analysis. Solid lines represent major cDNA clones used for sequencing determination. Open rectangles represent terminal regions of SCV genome whose sequences were determined earlier (Weng & Xiong, 1995). Positions of nucleotide substitutions observed in the SCV genome are indicated above the linear representation of the SCV genome. ▲, C to U transition; △, A to G transition; ○, A to U transversion; ●, G to U transversion.

nucleotide 1621, capable of encoding a protein of M_r 5700. Another begins at nucleotide 2634 and terminates at nucleotide 2861, capable of encoding a protein of M_r 8200. However, no significant sequence identity was found between these two putative proteins and any proteins deposited in the GenBank databases.

A high frequency of nucleotide substitutions was observed in regions of the SCV genome where overlapping cDNA clones were sequenced (Fig. 2*b*). Most of these substitutions were C to U or A to G transitions; there were only two A to U transversions and one G to U transversion. None of these substitutions changed the amino acid sequence of the predicted ORFs. As not all the sequence has been determined from more than one clone, polymorphisms in the SCV population could be more extensive.

Identification of SCV sg RNA 5′ termini

Two 3′-coterminal sg RNAs of 1·6 and 1·4 kb were produced during SCV infection (Fig. 1). The 5′ termini of these two sg RNAs were mapped by primer extension with two primers (Fig. 2). Primer SCV2663 is complementary to a region 42 nucleotides downstream of the CP AUG initiation codon. Primer SCV2401 is complementary to a region 24 nucleotides downstream of the p9 AUG initiation codon. A cDNA clone, pSCV8, was sequenced using the same primers. The primer extension and sequencing products are shown in Fig. 3. Two extension products of equal intensity were obtained with primer SCV2663. They corresponded to two guanosine residues at nucleotides 2484 and 2485 and represented the 5′ terminus of sg2. A single extension product synthesized from primer SCV2401 co-migrated with guanosine 2266, representing the 5′ terminus of sg1. Both the 1614 nucleotide sg1 and 1396 nucleotide sg2 begin with 5′ nucleotides GGU. Positions of sg1 and sg2 relative to the SCV genome are illustrated in Fig. 2. The 5′ terminus of sg1 is located 26 nucleotides upstream of the p6 AUG initiation codon and 111 nucleotides upstream of the p9 AUG initiation codon. The sg1 RNA thus contains the overlapping p6 and p9 ORFs as well as the p37 CP ORF. The 5′ terminus of sg2 is located 137 nucleotides upstream of the CP initiation codon and therefore encodes only the CP.

The double extension bands observed with primer SCV2663 may represent two species of sg2 initiating at two adjacent guanosine residues. Alternatively, they may be an artifact caused by non-uniform primer lengths or by the reverse transcription reaction. Carrington & Morris (1986) observed similar double bands in mapping CarMV sg RNAs. A cap- and template-independent activity of reverse transcriptase on some completed transcripts was thought to be responsible for their appearance.

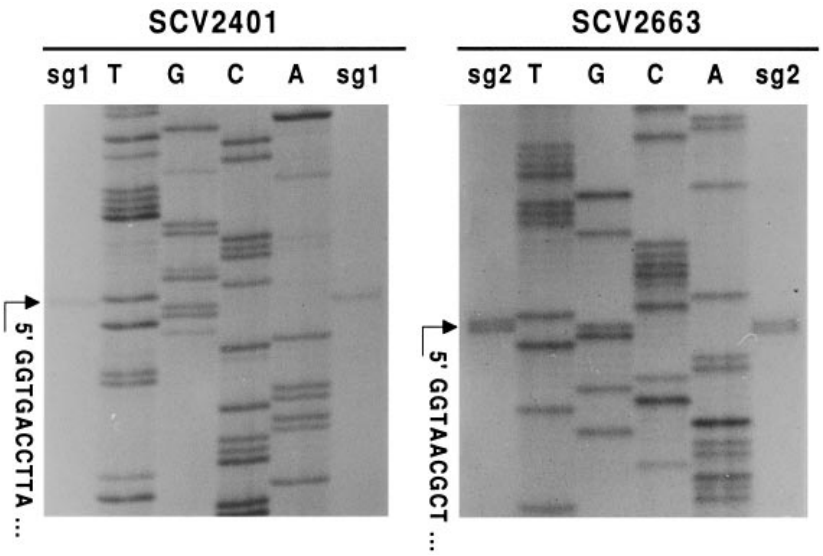

Fig. 3. Primer extension analysis of SCV sg RNAs. Primer SCV2401 complementary to nucleotides 2401 to 2418 and primer SCV 2663 complementary to nucleotides 2663–2680 were extended on SCV RNA by Superscript II reverse transcriptase. ^{35}S-labelled extension products of the large sg RNA (sg1) and the smaller sg RNA (sg2) were resolved in a 6% denaturing polyacrylamide gel along both sides of sequencing products of clone pSCV8 primed by the same primers. Nucleotides at the 5′ terminus of each sg RNA are shown as DNA. Note the double bands in primer SCV2663 extension products.

In vitro translation products from SCV RNA and full-length infectious transcripts

To determine if all the predicted SCV ORFs are expressed, virion RNA and the infectious full-length transcripts were translated in a rabbit reticulocyte *in vitro* translation system. Translation products programmed by these RNA templates were analysed by SDS-PAGE in 10% and 18·2% gels (Fig. 4). In 10% PAGE, proteins corresponding to p26, p57 and p37 ORFs were observed among translation products of virion RNA (Fig. 4*b*, lane 2). The most abundant proteins were p26 and p37 CP. In addition, a small amount of a protein of M_r 86000 was present in the translation products. This protein is likely the result of translation readthrough of p26 ORF into p57 ORF. The amount of p86 is estimated to be 4% of p26, consistent with the infrequent terminator readthrough event. Electrophoresis in 18·2% SDS-PAGE revealed that two small polypeptides, with M_rs of approximately 6000 and 9000, corresponding to p6 and p9 ORFs, were translated from virion RNA (Fig. 4*c*, lane 2). The p6 protein was similar in intensity to those of p26 and p37, but p9 was translated at a much reduced level. These results indicate that all major ORFs predicted in the SCV genome are translated from SCV virion RNA *in vitro*. Since virion RNA represents all the virus-specific RNA produced in the infected tissue, these proteins are also likely to be expressed during infection.

In contrast to the virion RNA translation profile, only two major proteins, p26 and the readthrough protein p86, were translated from the full-length transcripts (Fig. 4*b*, *c*, lanes 1). Both of these proteins were encoded by the first two ORFs at the 5′ end of the SCV genome. The ratio of p86 to p26 is estimated to be about 13%, higher than the ratio of p86/p26 translated from virion RNA. The lower ratio of p86/p26 translated from virion RNA could possibly be caused by partial degradation of virion RNA, resulting in incomplete translation of p86.

Degradation of virion RNA may also contribute to a faint protein band with an estimated M_r similar to the predicted p57 ORF product (Fig. 4*b*, lane 2). The AUG codon of the p57 ORF may have been exposed as the first initiation codon in the degrade genomic RNA. The same protein was not detected among the translation products of the full-length transcripts which were more homogeneous and not significantly degraded.

SCV 3′-terminal genes were expressed from sg RNAs

Total virion RNA contains genomic RNA as well as sg RNAs (Fig. 1) while the *in vitro* transcript represents only SCV genomic RNA. The difference in translation profiles of these two RNA templates (Fig. 4*b*) suggests that p6, p9 and p37 are expressed from sg RNAs encapsidated within virions. To test this hypothesis, total virion RNA was size-fractionated by sucrose density gradient centrifugation. Seven fractions were collected but only fractions F3-F6 contained detectable amounts of RNA (Fig. 4*a*). Sg RNAs are most abundant in F4 and gradually decreased to an undetectable level in F6. The genomic RNA was not detected in F3 and gradually increased in proportion to the sg RNAs in subsequent fractions. Proteins programmed by RNA in fractions F3-F6 are shown in Fig. 4(*c*). Proteins p37, p9 and p6, corresponding to the three 3′-terminal ORFs, were most abundant in the F3 translation products and decreased to undetectable levels in F6. The amount of p6, p9 and p37 correlates with that of sg RNAs in these fractions, further suggesting that these proteins are expressed from SCV sg RNAs. F6 contained mostly genomic RNA but a small amount of p37 CP was produced along with p26 and p86. The appearance of a small amount of CP in F6 may have been caused by contaminating sg RNA.

Minor protein products were also observed among the translation products of fractionated viral RNA, but were absent in the translation profiles of full-length transcripts or

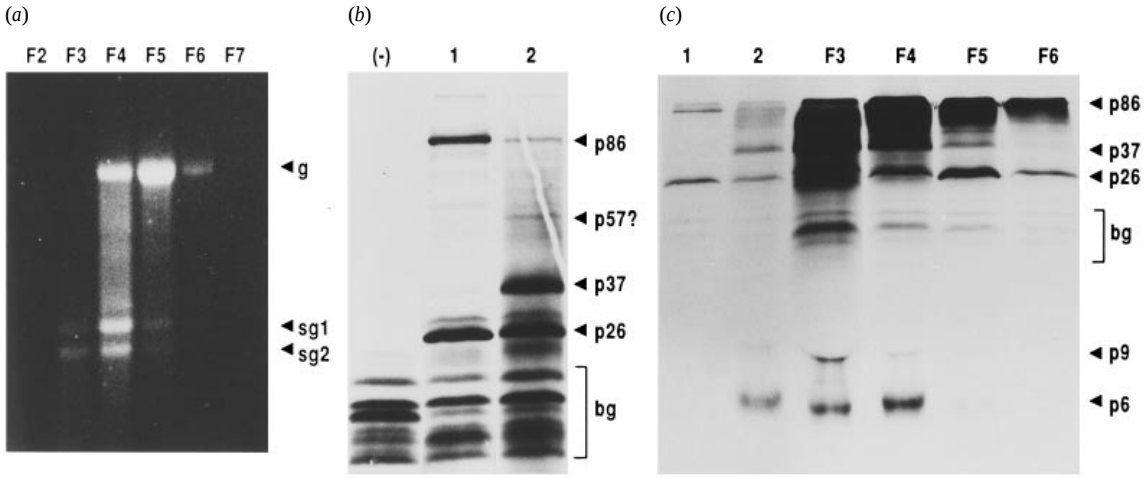

Fig. 4. Analysis of SCV genome expression by *in vitro* translation in rabbit reticulocyte lysates. (*a*) SCV virion RNA fractionated by linear-log sucrose density gradient centrifugation. Fractions of 0·7 ml were collected. Fractions 2 (F2) to 7 (F7) were electrophoresed in an agarose gel and stained with ethidium bromide. Positions of genomic RNA (g) and sg RNAs (sg1 and sg2) are marked to the right. (*b*) *In vitro* translation products of full-length infectious transcripts (lane 1) and unfractionated SCV virion RNA (lane 2) electrophoresed in a 10% SDS-polyacrylamide gel. A negative control with no exogenous RNA template (-) is shown to indicate background translation products (bg) from the reticulocyte lysate. (*c*) *In vitro* translation products of full-length infectious transcripts (lane 1), unfractionated SCV virion RNA (lane 2) and RNA fractions 3 (F3) to 6 (F6) separated in an 18·2% SDS-polyacrylamide gel. ^{35}S-labelled *in vitro* translation products in (*b*) and (*c*) were visualized by fluorography. Protein products corresponding to identified SCV ORFs are labelled to the right of each fluorograph.

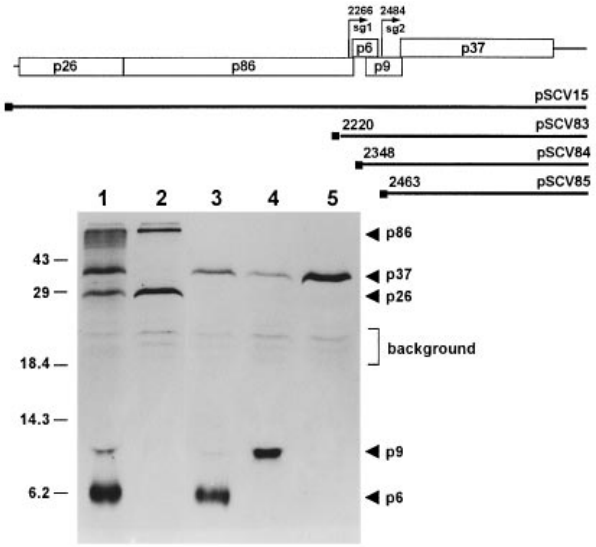

Fig. 5. Analysis of *in vitro* translation products of SCV 3′-coterminal deletion transcripts. Relationship of the four *in vitro* transcripts to SCV genomic RNA is indicated in the schematic. Transcription initiation sites of two SCV sg RNAs are marked by arrows above the genome schematic. ORFs were represented by open rectangles. Full-length infectious transcripts were transcribed from clone pSCV15 (lane 2). Transcripts derived from pSCV83 (lane 3), pSCV84 (lane 4) and pSCV85 (lane 5) represent portions of SCV genome from nucleotide 2220, 2348 and 2463 to the 3′ terminus, respectively. These transcripts along with SCV virion RNA (lane 1) were translated in a rabbit reticulocyte lysate system. The *in vitro* translation products were resolved in an 18·2% polyacrylamide gel and visualized by fluorography. Positions of the prestained low molecular mass protein standards are indicated to the left. Major *in vitro* translation products are labelled to the right.

unfractionated viral RNA. Degradation of viral RNA during fractionation and the subsequent exposure of internal, non-initiating AUG codons at the 5′ end of the degraded viral RNA may have contributed to the appearance of these minor bands.

p6 and p9 are expressed from a single sg RNA

To conclusively demonstrate that p6, p9 and p37 are translated from sg RNAs and to identify which sg RNA is responsible for their expression, 3′-coterminal transcripts were synthesized to approximate SCV sg RNAs (Fig. 5). pSCV83 is nearly identical to sg1 but has 46 extra nucleotides at the 5′ end. pSCV85 is nearly identical to sg2 but has 21 extra nucleotides at the 5′ end. pSCV84 does not correspond to any sg RNA; however, its 5′ terminus is 29 nucleotides upstream of the initiation AUG codon of the p9 ORF.

Translation products directed by these three 3′-coterminal transcripts as well as the full-length transcripts are shown in Fig. 5. No detectable p6, p9 and p37 were observed in the translation products of the full-length pSCV15 transcripts, suggesting that they are not translated from the genomic RNA. pSCV85 transcripts programmed the translation of only the p37 CP. The major protein translated from pSCV84 transcripts was p9. pSCV83 transcript programmed the synthesis of a large amount of p6 in addition to a small amount of p9. A small amount of p37 CP was translated from both pSCV83 and pSCV84 transcripts. These results demonstrate that p37 CP are expressed from sg2, and to a lesser degree from sg1 and that both p6 and p9 were expressed from sg1.

Discussion

Comparison of SCV nucleotide sequence with related viruses

The complete 3879 nucleotides of the SCV genome have been determined from cDNA clones obtained by traditional cloning methods and by novel techniques (Weng & Xiong, 1995). The latter techniques allow precise identification of the terminal nucleotides of the SCV genome. The SCV sequence significantly aligns to the published sequences of five carmoviruses (Table 1). Among carmoviruses, CarMV is most closely related to SCV with a nucleotide identity of 53·8%. The close relationship between SCV and CarMV is also supported by the 10 identical nucleotides at the 5′ end and 7 identical nucleotides at the 3′ end of the two genomes. No other carmoviruses share these many identical terminal nucleotides. Furthermore, 31 nucleotides at the transcription initiation site of CP sg RNA are identical between these two viruses. In addition to carmoviruses, several other small isometric viruses including members of the genera *Dianthovirus*, *Tombusvirus*, *Necrovirus* and *Machlomovirus* have a high degree of relatedness at the nucleotide level (AS > 7) (Table 1).

Comparison of genome organization in carmoviruses

The predicted genome organization of SCV is similar to that previously reported for CarMV (Guilley *et al.*, 1985), TCV (Carrington *et al.*, 1989), CCFV (Skotnicki *et al.*, 1993), MNSV (Riviere & Rochon, 1990) and CPMoV (You *et al.*, 1995) (Fig. 6). However, there are several distinctions. In CarMV, a double-readthrough event was predicted to extend the translation of p87 into the in-frame p9 ORF, resulting in a fusion protein of M_r 98000 (Guilley *et al.*, 1985). A protein of M_r approximately 98000 was indeed observed in an *in vitro* translation reaction supplemented with suppressor tRNA (Harbison *et al.*, 1985). This double-readthrough event is not predicted in the SCV genome. The p86 ORF is in-frame with p9 in SCV, but several additional terminator codons between the p86 amber terminator and the p9 initiation codon preclude the second readthrough. *In vitro* translation data (Figs 4 and 5) also do not support a double-readthrough event. In MNSV, two small, central ORFs, p7A and p7B, are separated by an in-frame amber terminal codon which might be readthrough to produce a p14 fusion protein. The two central small ORFs in SCV are overlapping and out of frame with each other. CPMoV has the most unique genome organization among the

Table 1. Percentage nucleotide identity, amino acid identity and similarity (in parentheses) between SCV and related viruses

Additional abbreviations and sources for sequences used: AMCV, artichoke mottle crinkle virus (Tavazza *et al.*, 1994); BYDV, barley yellow dwarf virus (Miller *et al.*, 1988); CNV, cucumber necrosis virus (CNV) (Rochon & Tremaine, 1989); CRSV, cymbidium ringspot virus (Dalmay *et al.*, 1993); MCMV, maize chlorotic mottle virus (Nutter *et al.*, 1989); RCNMV, red clover necrotic mosaic virus (Xiong & Lommel, 1989); SBMV, southern bean mosaic virus (Wu *et al.*, 1987); TBSV, tomato bushy stunt virus (Hearne *et al.*, 1990); TNV-A and -D, tobacco necrosis virus strain A (Meulewaeter *et al.*, 1990) and strain D (Coutts *et al.*, 1991).

Virus	Nucleotide identity (%)	Amino acid identity (similarity)				
		p26	p86	p6	p9	p37 CP
CarMV	53·8	39·6 (57·4)	55·2 (70·1)	58·3 (71·7)	60·0 (78·8)	41·8 (59·4)
TCV	47·8	30·2 (54·3)	45·6 (65·8)	45·0 (58·3)	38·8 (52·6)	36·5 (56·5)
CCFV	47·4	27·3 (51·5)	44·0 (65·3)	53·3 (61·7)	45·8 (66·3)	36·4 (54·0)
CPMoV	45·4	27·7 (51·6)	43·3 (62·5)	36·2 (63·8)	34·6 (56·4)*	32·8 (50·9)
MNSV	45·1	25·3 (51·1)*	43·5 (64·1)	38·6 (56·1)*	27·9 (50·8)*	29·5 (52·7)
MCMV	42·3	34·4 (56·1)*	43·8 (63·4)	–	–	–
TNV-A	45·3	31·1 (53·7)*	39·7 (59·7)	–	–	28·3 (48·6)*
CRSV	41·6	24·0 (45·9)*	36·8 (54·2)	–	–	32·2 (54·6)
TNV-D	–	30·5 (56·8)*	38·4 (61·4)	–	–	–
RCNMV	40·9	32·0 (55·5)*	36·2 (58·1)	–	–	30·2 (52·1)
TBSV	–	26·5 (51·6)*	37·0 (59·7)	–	–	32·3 (52·7)
CNV	–	26·5 (50·4)*	37·2 (59·0)	–	–	32·7 (55·5)
BYDV	–	23·1 (55·2)*	33·8 (58·8)	–	–	–
AMCV	–	–	–	–	–	33·1 (52·9)
SBMV	–	–	–	–	–	28·5 (49·6)*

* Adjusted alignment scores (AS) < 7·0.
– Indicates no significant sequence alignment.

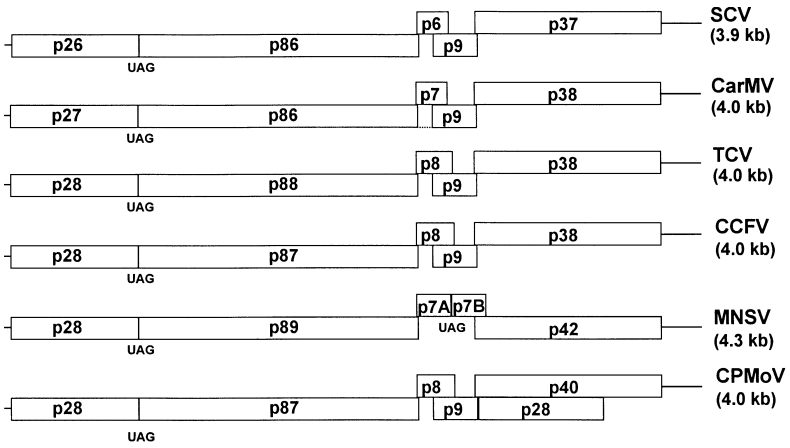

Fig. 6. Comparison of *Carmovirus* genome organizations. Open rectangles represent predicted ORFs in each viral genome. The in-frame amber termination codon UAG of the first ORF which may be readthrough to generate a larger fusion protein is shown. Potential readthrough of CarMV p86 UAG terminator into the in-flame p9 ORF to produce a p98 protein is indicated by a dashed line connecting the p86 and p9 ORFs. The UAG terminator between the in-frame p7A and p7B ORFs of MNSV may also potentially be readthrough to produce a p14 fusion protein.

carmoviruses (You *et al.*, 1995). A p28 ORF nested within the 3′ CP ORF is predicted. A small ORF capable of encoding a protein of M_r 8600 is present within the SCV CP ORF, but the putative protein has no homology with CPMoV p28.

Comparison of the amino acid sequences of SCV-encoded proteins

(i) Proteins involved in virus replication. Previous studies with TCV have shown that p28 and its terminator readthrough fusion protein p88 are required for TCV replication (Hacker *et al.*, 1992; White *et al.*, 1995). The SCV genome encodes two homologous proteins: p26 translated from the first ORF and the p86 readthrough protein (Fig. 2). A GDD motif conserved in all RNA-dependent RNA polymerases of (+)-stand RNA viruses (Kramer & Argos, 1984) is present in p86. Both p26 and p86 show a high degree of relatedness with the homologous proteins in other carmoviruses (Table 1). Pair-wise amino acid sequence alignments reveal that p86 is highly conserved in carmoviruses with the similarity ranging from 64% to 70% and the identity ranging from 44% to 55% (Table 1). In contrast, p26 shows lower but significant similarity with its homologous proteins (51–57% similarity and 25–40% identity).

The polymerases of eight other small RNA viruses also showed significant similarity to SCV p86 with adjusted AS values > 7. These viruses include members of the genera *Dianthovirus*, *Luteovirus*, *Machlomovirus*, *Necrovirus* and *Tombusvirus* (Table 1). The homologous protein of maize chlorotic mottle machlomovirus (MCMV) (Nutter *et al.*, 1989) is nearly 48% identical and 66% similar to SCV p86. All these viruses are small icosahedral viruses with a single-stranded, (+)-sense RNA genome. Significant identity in the replication-related proteins suggests that they may have a common evolutionary ancestry (Koonin & Dolja, 1993).

(ii) Movement proteins. Two small overlapping ORFs in the middle of the SCV genome encode p6 and p9. Mutational analysis of the TCV genome has indicated that both p8 and p9, homologous to p6 and p9 of SCV, are involved in virus cell-to-cell movement (Hacker *et al.*, 1992). Amino acid sequence comparisons show that these two proteins are conserved within the genus *Carmovirus*. No significant identity was found with any proteins from other viruses. Sequence identities of p6 and p9 vary considerably within the carmoviruses. SCV p6 and p9 are 58% and 60% identical to CarMV homologous proteins, but are only 39% and 28% identical to their counterparts in MNSV. The variability in the carmovirus movement proteins may reflect the diverse host ranges of each carmovirus.

(iii) Capsid protein. Amino acid sequence comparisons indicate that p37 CP is less conserved than either the movement proteins or the p86 polymerase (Table 1). SCV CP is only 41·8% identical to CarMV CP while the polymerase and movement proteins are nearly 60% identical to the CarMV homologues. There is also considerable variability in the sequence identity. MNSV CP has only 30% identity with SCV CP.

As with p86, SCV CP shows significant similarity with several other small isometric viruses including members of the genera *Necrovirus*, *Tombusvirus*, *Dianthovirus* and *Sobemovirus* (Table 1). Several viruses have significant similarity with SCV in CP, but not in p86. Conversely, several viruses have significant similarity with SCV in p86, but not in CP (Table 1). This observation may reflect evolution of RNA viruses by recombining modules of polymerases and CPs.

SCV gene expression strategy

The SCV gene expression strategy has been elucidated by *in vitro* translation studies. This is the first report where products of all the predicted carmovirus ORFs are expressed *in vitro*. Our experiments demonstrate that SCV 5′-proximal p26 ORF is expressed from the genomic RNA. Suppression of the p26 amber terminator allows ribosomes to translate the downstream, in-frame p57 ORF to produce a p86 fusion protein (Figs 4 and 5). The readthrough expression of the SCV p57 ORF observed in this study is consistent with CarMV and TCV gene expression (Carrington & Morris, 1985).

SCV transcribes two 3′-coterminal sg RNAs for expression of the 3′-terminal ORFs. Substantial amounts of both sg RNAs are packaged into SCV virions (Fig. 1). The 5′ termini of the SCV sg RNAs were mapped to nucleotide 2266 and 2484, respectively. Analysis of the nucleotide sequence from −50 to +50 of the two transcription initiation sites did not reveal extensive sequence identity or similar secondary structures. The only conserved feature is nucleotides CUGGU with the transcription starting at the third nucleotide. It is not clear what role this pentanucleotide may play in sg RNA transcription. SCV and CarMV nucleotide sequence alignments revealed 31 identical nucleotides spanning from −25 to +6 of the CP sg RNA (sg2) initiation site (data not shown). This and the extensive sequence identity between SCV and CarMV replication-related proteins suggest a similar if not identical mechanism of CP sg RNA transcription.

The p6, p9 and CP ORFs at the 3′ half of the SCV genome are expressed from the two sg RNAs. Two overlapping ORFs, p6 and p9, are expressed from the 1·6 kb sg1. Protein p6 encoded by the first ORF in sg1 is expressed at a significantly higher level than the p9 protein encoded by the second ORF (Figs 4 and 5). The 3′-proximal p37 CP ORF is expressed through the 1·4 kb sg2. Although *in vitro* translation studies suggest that the p9 ORF is more efficiently expressed from an sg RNA containing the p9 ORF as the first ORF, no such sg RNA has been identified either in SCV or any other carmovirus infections. Thus, we conclude that sg1 serves as a bicistronic messenger RNA for the gene expression of both p6 and p9. It is not clear how the p9 ORF, the second ORF on sg1, is expressed. A ribosomal leaky scanning mechanism has been proposed for the expression of overlapping internal ORFs in some plant viruses including barley stripe mosaic (Zhou & Jackson, 1996) and cucumber necrosis virus (Johnston & Rochon, 1996). Several factors have been suggested to promote ribosomes bypassing the first AUG initiation codon. They include short 5′ noncoding leader sequences, absence of considerable secondary structure downstream of the 5′-proximal AUG codon, and close proximity of the second AUG codon to the first initiating codon. A similar ribosome leaky scanning could explain the expression of p9 since the leader sequence of SCV sg1 contains only 26 nucleotides and the p9 AUG initiation codon is only 85 nucleotides downstream of the p6 AUG codon. Alternatively, internal ribosome initiations reported in picornaviruses and hepatitis C virus (Borman *et al*., 1995) may also lead to the expression of p9. Further studies are need to address the mechanism of p9 gene expression.

We thank Drs Steven A. Lommel, Christina K. Kennedy and Merritt R. Nelson for their critical review of this manuscript.

References

Borman, A. M., Bailly, J. L., Girard, M. & Kean, K. M. (1995). Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation *in vitro*. *Nucleic Acids Research* **23**, 3656–3663.

Brakke, M. K. & Van Pelt, N. (1970). Linear-log gradients for estimating sedimentation coefficients of plant viruses and nucleic acids. *Analytical Biochemistry* **38**, 56–64.

Carrington, J. C. & Morris, T. J. (1985). Characterization of the cell-free translation products of carnation mottle virus genomic and sg RNAs. *Virology* **144**, 1–10.

Carrington, J. C. & Morris, T. J. (1986). High resolution mapping of carnation mottle virus-associated RNAs. *Virology* **150**, 196–206.

Carrington, J. C., Heaton, L., Zuidema, D., Hillman, B. I. & Morris T. J. (1989). The genome structure of turnip crinkle virus. *Virology* **170**, 219–226.

Coutts, R. H. A., Ridgen, J. E., Slabas, A. R., Lomonosoff, G. P. & Wise, P. J. (1991). The complete nucleotide sequence of tobacco necrosis virus D. *Journal of General Virology* **72**, 1521–1529.

Dalmay, T., Rubino, L., Burgyan, J., Kollar, A. & Russo, M. (1993). Functional analysis of cymbidium ringspot virus genome. *Virology* **194**, 697–704.

Guilley, H., Carrington, J. C., Balazs, E., Jonard, G., Rochards, K. & Morris, T. J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Research* **13**, 6663–6677.

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.

Harbison, S. A., Davies, J. W. & Wilson, T. M. A. (1985). Expression of high molecular weight polypeptides by carnation mottle virus RNA. *Journal of General Virology* **66**, 2597–2604.

Hearne, P. Q., Knorr, D. A., Hillman, B. I. & Morris, T. J. (1990). The complete genome structure and synthesis of infectious RNA from clones of tomato bushy stunt virus. *Virology* **177**, 141–151.

Johnston, J. C. & Rochon, D. M. (1996). Both codon context and leader length contribute to efficient expression of two overlapping open reading frames of a cucumber necrosis virus bifunctional subgenomic mRNA. *Virology* **221**, 232–239.

Kim, J. W. & Bozarth, R. F. (1992). Mapping and sequence analysis of the capsid protein gene of cowpea mottle virus. *Intervirology* **33**, 135–147.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerase of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197–2206.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemical and Molecular Biology* **28**, 375–430.

Kramer, G. & Argos, P. (1984). Primary structural comparison of RNA-dependent polymerase from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7282.

Laemmli, U. K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Meulewaeter, F., Seurinck, J. & Van Emmelo, J. (1990). Genome structure of tobacco necrosis virus strain A. *Virology* **177**, 699–709.

Milbrath, G. M. & Nelson, M. R. (1972). Isolation and characterization of a virus from saguaro cactus. *Phytopathology* **62**, 739–742.

Milbrath, G. M., Nelson, M. R. & Wheeler, R. E. (1973). The distribution and electron microscopy of viruses of cacti in southern Arizona. *Phytopathology* **63**, 1133–1139.

Miller, W. A., Waterhouse, P. M. & Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Research* **16**, 6097–6111.

Murphy, F. A. , Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors) (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*. Vienna & New York: Springer-Verlag.

Nelson, M. R. & Tremaine, J. H. (1975). Physiochemical and serological properties of virus from saguaro cactus. *Virology* **65**: 309–319.

Nutter, R. C., Scheets, K., Panganiban, L. C. & Lommel, S. A. (1989). The complete nucleotide sequence of maize chlorotic mottle virus genome. *Nucleic Acids Research* **17**, 3163–3177.

Riviere, C. J. & Rochon, D. M. (1990). Nucleotide sequence and genomic organization of melon necrotic spot Virus. *Journal of General Virology* **71**, 1887–1896.

Rochon, D. M. & Tremaine, J. H. (1989). Complete nucleotide sequence of the cucumber necrosis virus genome. *Virology* **169**, 251–259.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning*: *A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Skotnicki, M. L., Mackenzie, A. M., Torronen, M. & Gibbs, A. J. (1993). The genomic sequence of cardamine chlorotic fleck carmovirus. *Journal of General Virology* **74**, 1933–1937.

Tavazza, M., Lucioli, A., Calogero, A., Pay, A. & Tavazza, R. (1994). Nucleotide sequence, genomic organization and synthesis of infectious transcripts from a full-length clone of artichoke mottle crinkle virus. *Journal of General Virology* **75**, 1515–1524.

Weng, Z. & Xiong, Z. (1995). A method for accurate determination of terminal sequences of viral genomic RNA. *Genome Research* **5**, 202–207.

White, K. A., Skuzeski, J. M., Li, W. Z., Wei, N. & Morris, T. J. (1995). Immunodetection, expression strategy and complementation of turnip crinkle virus p28 and p88 replication components. *Virology* **211**, 525–534.

Wu, S., Rinehart, C. A. & Kaesberg, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161**, 73–80.

Xiong, Z. & Lommel, S. A. (1989). The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1. *Virology* **171**, 543–554.

Xiong, Z. & Lommel, S. A. (1991). Red clover necrotic mosaic virus infectious transcripts synthesized *in vitro*. *Virology* **182**, 388–392.

Xiong, Z., Kim, K. H. Giesman Cookmeyer, D. & Lommel, S. A. (1993). The roles of the red clover necrotic mosaic virus capsid and cell-to-cell movement proteins in systemic infection. *Virology* **192**, 27–32.

You, X. J., Kim, J. W., Sturat, G. W. & R. F. Bozarth. (1995). The nucleotide sequence of cowpea mottle virus and its assignment to the genus *Carmovirus*. *Journal of General Virology* **76**, 2841–2845.

Zhou, H. & Jackson, A. O. (1996). Expression of the barley stripe mosaic virus RNA beta 'triple gene block'. *Virology* **216**, 367–379.

Received 27 September 1996; Accepted 8 November 1996