

***In vivo* analysis of the TSWV cap-snatching mechanism: single base complementarity and primer length requirements**

Daniël Duijsings, Richard Kormelink¹ and Rob Goldbach

Laboratory of Virology, Wageningen University, Binnenhaven 11, 6709 PD, Wageningen, The Netherlands

¹Corresponding author
e-mail: richard.kormelink@viro.dpw.wag-ur.nl

Requirements for capped leader sequences for use during transcription initiation by tomato spotted wilt virus (TSWV) were tested using mutant alfalfa mosaic virus (AMV) RNAs as specific cap donors in transgenic *Nicotiana tabacum* plants expressing the AMV replicase proteins. Using a series of AMV RNA3 mutants modified in either the 5'-non-translated region or in the subgenomic RNA4 leader, sequence analysis revealed that cleaved leader lengths could vary between 13 and 18 nucleotides. Cleavage occurred preferentially at an A residue, suggesting a requirement for a single base complementarity with the TSWV RNA template, which could be confirmed by analyses of host mRNAs used *in vivo* as cap donors.

Keywords: cap snatching/negative-strand RNA virus/transcription initiation/TSWV

Introduction

Segmented, negative-strand RNA viruses share the transcription initiation mechanism generally referred to as 'cap snatching'. During this process, a 7^mG-capped host mRNA is recruited by the viral transcriptase complex and subsequently cleaved by a virally encoded endonuclease. The resulting capped leader RNA is used to prime transcription on the viral genome, as described most extensively for influenza A virus (Caton and Robertson, 1980; Dhar *et al.*, 1980; Plotch *et al.*, 1981; Ulmanen *et al.*, 1981; Braam *et al.*, 1983).

However, knowledge of the requirements for sequence specificity, length and structure of a suitable donor RNA has remained rather limited. Commonly, cap donor RNAs are cleaved at a distance of ~15 nucleotides from the cap structure, though variation in length occurs between 10 and 20 nucleotides (Caton and Robertson, 1980; Dhar *et al.*, 1980; Bishop *et al.*, 1983; Patterson and Kolakofsky, 1984; Eshita *et al.*, 1985; Ihara *et al.*, 1985; Collett, 1986; Gerbaud *et al.*, 1987; Bouloy *et al.*, 1990; Simons and Pettersson, 1991; Grò *et al.*, 1992; Kormelink *et al.*, 1992a,b; Huiet *et al.*, 1993; Jin and Elliott, 1993a,b; Ramirez *et al.*, 1995; Shimizu *et al.*, 1996; van Poelwijk *et al.*, 1996). Exceptions have been reported for members of the Arenaviridae (Tacaribe virus) (Garcin and Kolakofsky, 1990; Raju *et al.*, 1990) and Nairovirus genus (Dugbe virus) (Jin and Elliott, 1993b), which use relatively short (1–4 and 5–16 nucleotides, respectively)

non-viral leader sequences. For many of these viruses, sequence analyses of their mRNAs have shown a nucleotide preference at the 3' end of the non-viral leader, assumed to reflect a sequence preference for cleavage by the viral endonuclease. For example, in the case of Dugbe virus, endonucleolytic cleavage has been proposed to take place exclusively after a C residue (Jin and Elliott, 1993b), whereas for Bunyamwera virus a strong preference for cleavage after a U residue has been proposed (80% of the mRNAs studied) (Jin and Elliott, 1993a). However, most mRNAs analysed in these cases were produced *in vivo*, hence the particular mRNAs that were used to provide these capped leader sequences (the cap donors) remained unknown. Therefore, it is still unknown whether cleavage of the cap donor has indeed taken place immediately after the assumed 3' end of the capped leader sequence or whether this cleavage has taken place farther downstream, e.g. one or two nucleotides, which would be complementary to the 3'-terminal residues of the viral template.

For several viruses, *in vitro* studies have provided information about leader length requirements and have suggested that base pair interactions can contribute to alignment of the capped leader RNA sequence with the viral template RNA (Shaw and Lamb, 1984; Honda *et al.*, 1986; Chung *et al.*, 1994; Hagen *et al.*, 1994, 1995; Garcin *et al.*, 1995). For Hantaan virus, an additional 'prime-and-realign' mechanism has been proposed to explain data obtained from transcription initiation studies (Garcin *et al.*, 1995). During this 'prime-and-realign' mechanism, transcription initiation starts with a capped leader RNA that base-pairs to the viral template RNA and becomes extended for a few nucleotides only. Subsequently, it is released from the viral RNA template and realigns backward by virtue of the terminal sequence repeats. Only then is progressive elongation of the nascent mRNA chain thought to take place. The reason for a 'prime-and-realign' mechanism is still unclear; however, it is proposed to account for the viral RNA polymerase initiating Hantaan (anti)genome and mRNA synthesis with GTP (Garcin *et al.*, 1995). Meanwhile, the occurrence of a 'prime-and-realign' mechanism has been proposed for several other negative-strand RNA viruses, e.g. Germiston virus and LaCrosse virus (Bouloy *et al.*, 1990; Dobie *et al.*, 1997), in order to explain the presence of repetitive sequences within the heterogeneous sequences at the 5' end of the viral mRNAs.

Tomato spotted wilt virus (TSWV) is the representative of the Tospovirus genus within the Bunyaviridae, a family of negative-strand RNA viruses with a tripartite genome. TSWV is the first plant virus for which the occurrence of cap snatching has been investigated (Kormelink *et al.*, 1992b; van Poelwijk *et al.*, 1996). These studies have shown the presence of non-viral leader sequences, 12–21 nucleotides in length, at the 5' ends of viral subgenomic

mRNAs. Recent findings have demonstrated that during a co-infection of *Nicotiana benthamiana* with TSWV and alfalfa mosaic virus (AMV), all capped (sub)genomic RNAs of the latter (positive-strand RNA) virus can act *in vivo* as cap donors for TSWV (Duijsings *et al.*, 1999), as likewise was shown to occur for the maize stripe tenuivirus during a co-infection with the barley stripe mosaic hordeivirus (Estabrook *et al.*, 1998). Cleavage of the AMV leaders preferentially took place at an A residue. The AMV capped leader sequences found at the 5' ends of TSWV N and NSs mRNAs varied in length, even when originating from the same AMV RNA molecule, except for RNA3, where obviously a single cleavage site was used. Apparently the capped leader sequence of AMV RNA3 matched strict requirements resulting in a unique cleavage site. However, alignment of the original AMV RNA3 sequence, and also those of RNA1, 2 and 4, with the chimeric AMV-TSWV mRNA sequences did not allow identification of the cleavage site as the position of this site is dependent on the possible need for complementarity between snatched leader and template RNA (Figure 1).

In order to test for leader length preference, cleavage specificity and a possible base pairing requirement, mutant AMV RNAs were analysed for their ability to act as cap donors in a manipulable *in vivo* system. To this end, transgenic (p12) tobacco plants expressing the AMV p1 and p2 replicase subunits were used, allowing *in vivo* replication of AMV RNA3 molecules specifically mutated in their leader sequences (Neeleman *et al.*, 1991, 1993; Taschner *et al.*, 1991). Here it is shown that mutant AMV RNAs provided in this way can be tested as cap donors during TSWV transcription initiation. These analyses are complemented, furthermore, by testing selected host mRNAs as cap donors. Altogether, the *in vivo* observations indicate that for successful transcription initiation, a single base complementarity between the snatched leader and the viral template RNA is required.

Results

Use of mutant AMV RNA3 leader sequences for transcription initiation by TSWV

From previous studies (Duijsings *et al.*, 1999), it was known that all four AMV RNAs are accepted as cap donors for TSWV transcription initiation and that cleavage of the AMV leaders took place preferentially at an A residue. These results were in support of a possible base pairing requirement between snatched leader and template RNA, but still left room for an alternative explanation (Figure 1) (Duijsings *et al.*, 1999). To collect further evidence that the preferential cleavage at A residues indeed reflects the requirement for base pairing, two sets of experiments were performed. First, it was investigated whether the length of the snatched leader RNA strictly co-varies along with the position of an available A residue. Secondly, it was tested whether leader sequences with a G instead of an A residue at the preferred cleavage site would be accepted as cap donors, but now by base pairing with the penultimate residue (C) of the TSWV template RNA.

To this end, specific AMV RNA3 leader mutants were presented to TSWV using transgenic (p12) tobacco plants

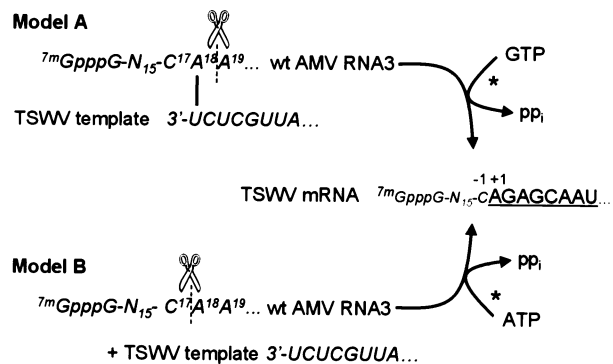


Fig. 1. Possible models for the cap-snatching mechanism. Model A: endonucleolytic cleavage occurs 3' of an A residue within the cap donor RNA (e.g. wild-type AMV RNA3), which subsequently base-pairs with the 3'-ultimate residue of the viral template RNA. The capped RNA primer is elongated according to the viral template, with a G (complementary to the 3'-penultimate residue of the template) being the first incorporated residue during elongation. Model B: the cap donor RNA is cleaved (non-)specifically at a certain distance from the cap structure and is elongated according to the viral template, with an A (complementary to the 3'-ultimate residue of the template) being the first incorporated residue during elongation.

expressing functional AMV p1 and p2 replicase subunits. As reported before (Neeleman *et al.*, 1991, 1993; Taschner *et al.*, 1991), such plants support the replication and systemic spread of AMV RNA3, even when provided as cauliflower mosaic virus (CaMV) 35S promoter-driven cDNA clones. To test the feasibility of this *in vivo* system, first a wild-type construct of AMV RNA3 was co-inoculated along with TSWV, and 7 days later chimeric AMV-TSWV mRNAs were rescued from systemically infected top leaves using RT-PCR (Figure 2A). This analysis demonstrated that AMV RNA3 molecules accumulating in p12 plants were indeed used as cap donors for TSWV mRNA synthesis (Figure 3A, lane WT). Sequence analyses of several independent clones (Table I) showed that the junction site found between the AMV RNA3 leader and the TSWV mRNA sequence completely matched the sequence data from earlier experiments during which *N.benthamiana* plants were co-inoculated with TSWV and AMV (Duijsings *et al.*, 1999).

As a next step, to investigate whether cleavage would specifically take place at an A residue, single point mutants (Table I) were made at nucleotide positions 17 or 18 and subsequently used in co-inoculation experiments with TSWV on p12 plants. It was anticipated that a change of residue C17 into an A would result in a -1 shift of the cleavage site, the resulting AMV leader sequence within the TSWV mRNA thus becoming one nucleotide shorter. Similarly, a change of A18 into a C would result in a +1 shift of the endonucleolytic cleavage site to still meet the supposed cleavage specificity for an A residue. In this case, the added AMV leader sequence of the TSWV mRNA would increase in size by one nucleotide, nucleotide A19 now being the residue at which cleavage would take place. Sequence data from independent RT-PCR clones (Figure 3A; Table I), collected from co-inoculations of these AMV RNA3 mutants and TSWV on p12 plants, indeed showed a -1 shift when C17 was changed into an A (mutant C17A), and a +1 shift where A18 was changed into a C (mutant A18C). The leader

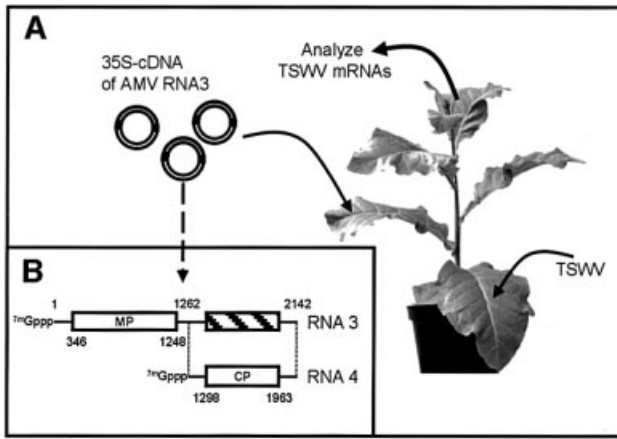


Fig. 2. *In vivo* amplification of (mutant) AMV RNA3 and rescue of chimeric TSWV mRNAs. (A) Transgenic *N. tabacum* p12 plants expressing the AMV replicase proteins were inoculated mechanically with a 35S promoter-driven plasmid containing the AMV RNA3 segment. Subsequently, the plants were inoculated with TSWV strain BR-01. Newly formed leaves showing symptoms of TSWV infection were harvested at 7 days post-infection and total RNA isolated from these leaves was analysed for the presence of AMV leader-containing TSWV mRNAs. (B) Schematic drawing of the AMV RNA3 segment. The genomic RNA3 segment (transcribed from a 35S-driven plasmid) is replicated by the p1 and p2 replicase proteins in the p12 plant. Subgenomic AMV RNA4 is subsequently transcribed from this RNA segment by the p1 and p2 replicase proteins.

length thus shifted along with the position of an available A residue, demonstrating that cleavage specificity (at an A) determines leader length and supporting the model in which a single base complementarity between leader and template RNA is required (Figure 1).

To demonstrate more conclusively that base pairing was involved, two point mutants were made in which G residues were introduced (mutants C17G and A18G; Table I), which would potentially lead to cleavage at G and base pairing with the penultimate nucleotide (C) of the TSWV template. Cloning and sequence analysis of progeny AMV RNA revealed that these mutants rapidly converted into mutant A18C. As a consequence, RT-PCR clones obtained from co-inoculation of these AMV RNA3 mutants with TSWV invariably showed an endonucleolytic cleavage pattern for the AMV leader sequence identical to that observed with mutant A18C (Table I).

Also, additional AMV RNA3 mutants tested, i.e. C17U and A18U, rapidly converted into mutant A18C (Table I), indicating restrictions in the mutability of the AMV RNA3 leader.

Mutant AMV RNA4 leaders confirm donor sequence and length requirements for priming TSWV transcription

As further mutagenesis of its leader led to genetic instability of AMV RNA3, we next presented modified AMV RNA4 leaders to TSWV-infected p12 plants to confirm the cleavage specificity at an A residue and to test for leader length preference and base pairing requirement. AMV RNA4 has been used as cap donor before in several *in vitro* studies with other negative-strand RNA viruses (Bouloy *et al.*, 1980; Plotch *et al.*, 1981; Ulmanen *et al.*,

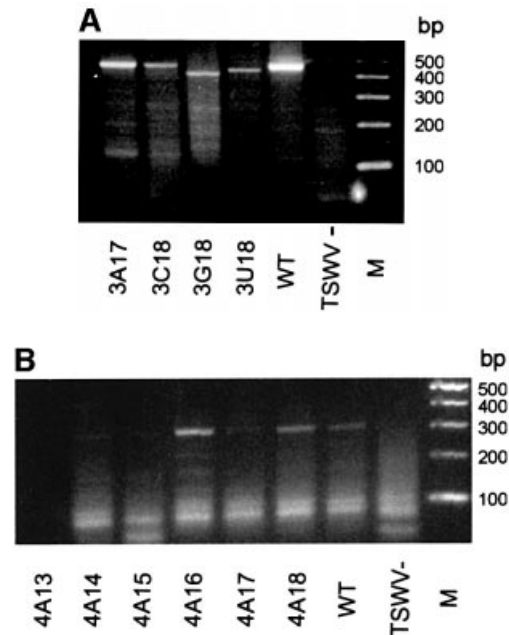


Fig. 3. RT-PCR products of TSWV mRNAs containing (mutant) AMV RNA leaders. (A) TSWV NSs mRNAs containing (mutant) AMV RNA3 leaders. Total RNA was isolated from infected *N. tabacum* p12 at 7 days post-infection and TSWV NSs mRNAs containing AMV RNA3-derived leaders were RT-PCR amplified using primer NSs1 (RT) and primers NSs2 and A3 (PCR). Products of ~420 bp represent TSWV NSs mRNAs. 3A17, AMV RNA3 mutant 3A17; 3C18, AMV RNA3 mutant 3C18; 3G18, AMV RNA3 mutant 3G18; 3U18, AMV RNA3 mutant 3U18; WT, wild-type AMV RNA3; TSWV-, uninfected, wild-type AMV RNA3-containing plants; M, 100 bp molecular marker. (B) TSWV N mRNAs containing (mutant) AMV RNA4 leaders. Total RNA was isolated from infected *N. tabacum* p12 at 7 days post-infection and TSWV N mRNAs containing AMV RNA4-derived leaders were amplified using primer N1 (RT) and primers N2 and A4 (PCR). Products of ~315 bp represent TSWV N mRNAs. 4A13, AMV RNA4 mutant A13; 4A14, AMV RNA4 mutant A14; 4A15, AMV RNA4 mutant A15; 4A16, AMV RNA4 mutant A16; 4A17, AMV RNA4 mutant A17; 4A18, AMV RNA4 mutant A18; WT, wild-type AMV RNA4; M, 100 bp molecular marker.

1981; Patterson *et al.*, 1984; Galarza *et al.*, 1996) and *in vivo* with TSWV (Duijsings *et al.*, 1999). AMV RNA4 is a subgenomic mRNA (Figure 2B), its leader residing internally in genomic RNA3 and not containing any *cis*-replication signals. Therefore, RNA4 leader mutants might be more stable than RNA3 leader mutants. The wild-type AMV RNA4 contains only two A residues at positions 13 and 14 within a U-rich context (Table II). Hence, only a minimal modification would be required to obtain a set of mutants with a single A residue at different positions within an oligo(U) stretch running from position 12 to 18 (Table II, mutants denoted as A12–A18). Additionally, a mutant was made lacking any A residue between positions 12 and 18 (mutant RNA4-noA). These mutants would allow us to test whether the length of the snatched leader sequence would co-vary precisely with the position of the A residue in this leader and provide information about leader length preference. First, the fitness and genetic stability of the AMV RNA4 mutants were tested by inoculation of the cDNA constructs harbouring these mutations on p12 plants and sequence analysis of progeny RNA4. All mutants, except RNA4-noA and A12, were

Table I. AMV RNA3 mutant leader sequences in TSWV mRNAs

| Mutant | Sequence | After p12 amplification | Chimeric TSMV mRNA | No. of clones |
|--------|---|-------------------------|--|---------------|
| WT | <i>7mGpppG-N₁₅-C¹⁷A¹⁸A¹⁹...</i> | WT | <i>7mGpppG-N₁₅-C¹⁷A¹⁸GAGCAAU...</i> | 10 |
| C17A | <i>7mGpppG-N₁₅-A¹⁷A¹⁸A¹⁹...</i> | C17A + A18C | <i>7mGpppG-N₁₅-A¹⁷GAGCAAU...</i> <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹GAGCAAU...</i> | 4 2 |
| C17G | <i>7mGpppG-N₁₅-G¹⁷A¹⁸A¹⁹...</i> | A18C | <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹GAGCAAU...</i> | 1 |
| C17U | <i>7mGpppG-N₁₅-U¹⁷A¹⁸A¹⁹...</i> | unstable ^a | ND | ND |
| A18C | <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹...</i> | A18C | <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹GAGCAAU...</i> | 7 |
| A18G | <i>7mGpppG-N₁₅-C¹⁷G¹⁸A¹⁹...</i> | A18C | <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹GAGCAAU...</i> | 8 |
| A18U | <i>7mGpppG-N₁₅-C¹⁷U¹⁸A¹⁹...</i> | A18C | <i>7mGpppG-N₁₅-C¹⁷C¹⁸A¹⁹GAGCAAU...</i> | 9 |

TSWV mRNAs primed by (mutant) AMV RNA3 leaders were amplified using nested RT-PCR. Individual clones from these amplifications were sequence analysed.

^aNot amplified in the p12 system; ND, not done.

Table II. AMV RNA4 mutant leader sequences in TSWV mRNAs

| Mutant | Sequence | After p12 amplification | Chimeric TSMV mRNA | No. of clones |
|--------|--|-------------------------|--|---------------|
| WT | <i>7mGpppG-N₁₁-A¹³A¹⁴UUUUC...</i> | WT | <i>7mGpppG-N₁₁-A¹³GAGCAAU...</i> <i>7mGpppG-N₁₁-A¹³A¹⁴GAGCAAU...</i> | 1 9 |
| A12 | <i>7mGpppG-N₁₀-A¹²UUUUUUC...</i> | A15, A16, A17, WT | ND | ND |
| A13 | <i>7mGpppG-N₁₀-UA¹³UUUUUUC...</i> | A13, A16 | <i>7mGpppG-N₁₁-A¹³GAGCAAU...</i> <i>7mGpppG-N₁₁-UUUA¹⁶GAGCAAU...</i> | 0 1 |
| A14 | <i>7mGpppG-N₁₀-UUA¹⁴UUUUC...</i> | A14, A16 | <i>7mGpppG-N₁₁-UA¹⁴GAGCAAU...</i> <i>7mGpppG-N₁₁-UUUA¹⁶GAGCAAU...</i> | 0 3 |
| A15 | <i>7mGpppG-N₁₀-UUUA¹⁵UUUC...</i> | A15, A16 | <i>7mGpppG-N₁₁-UA¹⁴GAGAGCAAU...</i> <i>7mGpppG-N₁₁-UUA¹⁵GAGCAAU...</i> | 1 2 |
| A16 | <i>7mGpppG-N₁₀-UUUUA¹⁶UUC...</i> | A16 | <i>7mGpppG-N₁₁-UUUA¹⁶GAGCAAU...</i> <i>7mGpppG-N₁₁-UUUUA¹⁶GCAAU...</i> | 3 5 |
| A17 | <i>7mGpppG-N₁₀-UUUUUA¹⁷UC...</i> | A17 | <i>7mGpppG-N₁₁-UUUUA¹⁷GAGCAAU...</i> | 2 |
| A18 | <i>7mGpppG-N₁₀-UUUUUUA¹⁸C...</i> | A18 | <i>7mGpppG-N₁₁-UUUUUA¹⁸GAGCAAU...</i> | 4 |
| noA | <i>7mGpppG-N₁₀-UUUUUUUC...</i> | unstable ^a | ND | ND |
| G15 | <i>7mGpppG-N₁₀-UUUG¹⁵UUUC...</i> | G15 | <i>7mGpppG-N₁₁-UUG¹⁵AGCAAU...</i> | 1 |

TSWV mRNAs primed by (mutant) AMV RNA4 leaders were amplified using nested RT-PCR. Individual clones from these amplifications were sequence analysed.

^aNot amplified in the p12 system; ND, not done.

relatively stable. Progeny RNA from mutant RNA4-noA could not be recovered by RT-PCR and no coat protein production was observed, indicating that this mutant was not viable (data not shown). RT-PCR and sequence analysis of several independent clones showed that mutant A12 was unstable, rapidly converting into wild-type RNA4 as well as into other mutant sequences (A15–A17). For most of the other mutants, minor amounts of another mutant sequence were found in addition to the expected mutant genotype after a single round of replication in p12 plants (Table II).

Having tested the fitness and stability of the AMV RNA4 mutants, these were provided as cap donors during a co-infection with TSWV on p12 plants. RT-PCR cloning of TSWV N mRNAs obtained from these co-infection experiments (Figure 3B) and subsequent sequence analyses of the 5'-capped leader sequences indicated that all of the RNA4 mutants with an A residue at different positions in the AMV RNA4 leader could serve as cap donors (Table II). Sequence data from several clones of TSWV N mRNAs, obtained from co-inoculation experiments with AMV RNA4 mutants, showed a shift of the snatched leader length along with the position of the A residue within the oligo(U) tract. A number of the AMV RNA4

mutants produced a polymorphous population of RNA4 molecules (Table II), the differently sized capped leaders found on the resulting TSWV mRNAs reflecting this variation. Despite this variation, the leader length always shifted along with the position of the A residue. Although the numbers were not statistically significant, the sequence results from the subset of different RNA4 molecules obtained after p12 amplification pointed towards a preference for molecules containing an A residue at position +16 (Table II), as was found when using host mRNAs as cap donors (van Poelwijk *et al.*, 1996).

The strict co-variance of leader length with the position of the A residue in both AMV RNA3 and RNA4 indicates that endonucleolytic cleavage of the leader takes place 3' of this A residue to allow base pairing with the first U residue of the TSWV template RNA. However, alternatively, the results could be explained by an endonuclease activity that cleaves cap donor RNAs 5' of the A residue (Figure 1B), with the A residue in the TSWV mRNA being the first nucleotide incorporated during capped primer elongation within the transcription process. This scenario seems less likely in view of the sequence data obtained using wild-type AMV RNA4 as a cap donor. The leader of this RNA contains two A residues (at positions

13 and 14), and sequence data of TSWV mRNAs containing leaders derived from RNA4 show that residue A13 is retained preferentially, which does not fit a cleavage mechanism meant to recruit a leader without a 3'-terminal A residue.

The results with the A13–A18 mutants indicated that the leader sequence of RNA4 could be modified to a certain extent without complete loss of *in vivo* RNA3 or RNA4 amplification in p12 plants. Therefore, to substantiate the requirement for a single base pairing between leader and TSWV template during the cap-snatching process, AMV RNA4 mutant G15 was made. This mutant lacked any A residue and contained a single G at position 15 within an oligo(U) stretch (Table II). Only a low RT-PCR signal was obtained for TSWV mRNAs primed with G15 leaders (Figure 4A), but the single clone obtained supported the base pairing model as the mRNA sequence lacked the 5'-terminal A residue of the TSWV template. This could only happen by cleavage of the mutant RNA4 leader 3' of residue G15 and base pairing to the penultimate C of the TSWV template (Figure 4B).

Recognition of host mRNAs during cap snatching.

The low recovery of TSWV mRNAs primed with the AMV RNA4 mutant G15 (Figure 4A) could be due to both decreased efficiency of priming on the penultimate residue and decreased replication of this mutant in p12 plants. To circumvent the second complication, we extended our *in vivo* studies to host mRNA leaders used by TSWV for transcription initiation. Two different host (*Nicotiana tabacum*) genes from the GenBank database were selected for which the transcription initiation sites were known unequivocally (Table III): the first gene (polyubiquitin; ubi.U4; DDBJ/EMBL/GenBank accession No. X77456) with a predictable cleavage site in view of the single A residue at position 17 and the other (*S*-adenosylmethionine decarboxylase; SAM-DC; DDBJ/EMBL/GenBank accession No. AF033100) potentially giving a more complicated picture, as its mRNA contained multiple possible cleavage sites between positions 12 and 21 (Table III). Using a specific RT-PCR primer for the leader of the polyubiquitin mRNA, chimeric TSWV mRNAs could indeed be obtained from total RNA of TSWV-infected tobacco plants (Figure 5). Sequence analysis of several clones revealed that the ubi.U4 mRNA was cleaved exclusively at residue 17 (Table III), confirming the results obtained with the AMV leaders. The SAM-DC mRNA also appeared to be cleaved at a single site, but now after a G residue at the preferred nucleotide position 16. All mRNAs containing the SAM-DC leader lacked the terminal A residue of the authentic TSWV sequence (Table III), and this result could only be explained by endonucleolytic cleavage downstream of G16 and subsequent base pairing to the penultimate C residue of the viral template. The use of the SAM-DC leader therefore indicates the plasticity of the cap-snatching mechanism of TSWV: position seems more important than nucleotide identity, as long as the 3'-terminal nucleotide of the snatched leader can base-pair with the ultimate or penultimate residue of the viral template.

From the co-infection studies with AMV (Duijsings *et al.*, 1999), the data obtained using mutant AMV RNA3 and RNA4 donors, and with the two host mRNAs, the

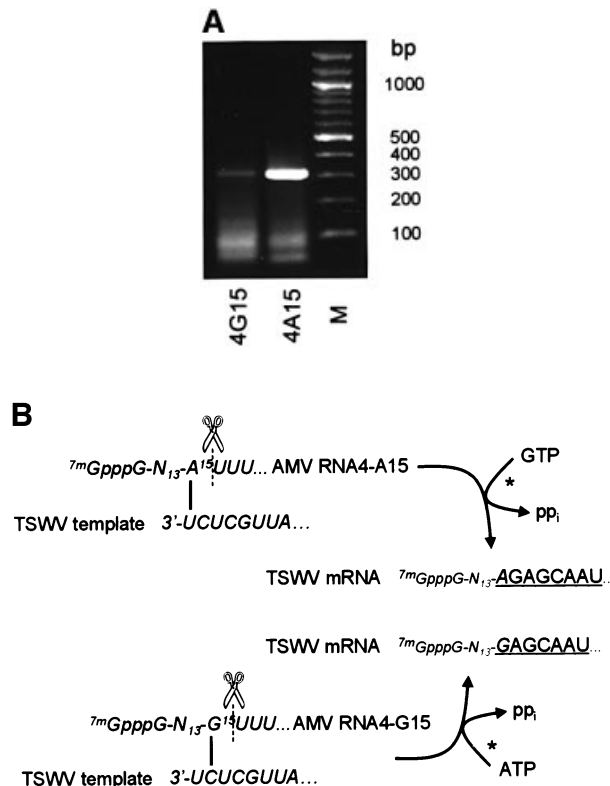


Fig. 4. Transcription initiation at the ultimate and penultimate base of the TSWV template RNA. (A) RT-PCR amplification of TSWV N mRNAs containing AMV RNA4 leaders mutated at nucleotide 15. Total RNA was isolated from infected *N. tabacum* p12 at 7 days post-infection and TSWV N mRNAs containing AMV RNA4-derived leaders were amplified using primer N1 (RT) and primers N2 and A4 (PCR). Products of ~315 bp represent TSWV N mRNAs. 4G15, AMV RNA4 mutant G15; 4A15, AMV RNA4 mutant A15; M, 100 bp molecular marker. (B) Model for transcription initiation on the penultimate residue. The cap donor RNA is cleaved by the viral endonuclease 3' of an A residue and base-pairs to the 3'-ultimate U residue of the viral template RNA. Elongation takes place according to the viral RNA, with a G residue being the first residue incorporated during elongation. Alternatively, if no A residue is available at an optimal distance from the cap structure, a G residue may be used in base pairing to the 3'-penultimate C residue of the viral template. Subsequently, elongation will take place, with an A residue being the first residue incorporated. The resulting viral mRNA will miss the ultimate A residue present at the 5' end of the viral genome.

conclusion can be drawn that for cap snatching by TSWV, capped leaders are cleaved preferentially behind an A residue, which, moreover, should occur preferentially at or close to position 16 from the cap. Knowing now that a single base complementarity is a prerequisite for accepting the leader of an mRNA as cap donor, previously obtained sequence data on cloned TSWV mRNAs containing host-derived leaders (van Poelwijk *et al.*, 1996) can be used to validate our conclusions. In 80% of the sequences obtained, the host-derived leader sequences fit a mechanism whereby cleavage occurred after an A residue at a distance of 13–21 nucleotides (average length 16 nucleotides). For the other 20%, the leaders fit cleavage after a G residue (at a distance of 14–22 nucleotides from the cap, average length 17 nucleotides), allowing priming by base pairing at the penultimate C residue of the TSWV template, a minor alternative also found with the AMV leaders.

Table III. Host leader sequences in TSWV mRNAs

| Host gene | Obtained sequence of TSWV mRNA | No. of clones |
|-----------|--|---------------|
| ubi.U4 | ^{7m} GpppAUCCUUUGAUUUCUCUA ¹⁷ UUCUC... ^{7m} Gppp-N ₁₆ -A ¹⁷ GAGCAAUU... | 4 |
| SAM-DC | ^{7m} GpppAUGGAGUCGAAAGGUG ¹⁶ GUAAAAAC... ^{7m} Gppp-N ₁₅ -G ¹⁶ AGCAAUU... | 11 |

TSWV mRNAs primed by host mRNA leaders were amplified using nested RT-PCR. Individual clones from these amplifications were sequence analysed.

Discussion

Cap snatching, as a general mechanism for transcription initiation among different segmented negative-strand RNA viruses, has been studied by both *in vivo* and *in vitro* methods. While some of the data obtained in these experiments suggested that a complementarity or base pairing between the donor RNA and the viral template might be required (Jin and Elliot, 1993a; Garcin *et al.*, 1995; Dobie *et al.*, 1997), other data disagreed with this view (Krug *et al.*, 1980; Hagen *et al.*, 1995). However, both the *in vivo* and *in vitro* methods used to study the cap-snatching process had some disadvantages. The *in vivo* methods were based mainly on analysis of viral mRNAs containing host-derived sequences obtained through 5'-RACE amplification, which implied that it was virtually impossible to determine the sequence of a given cap donor RNA before its use in transcription initiation. The influence of a specific sequence within the cap donor and the exact endonucleolytic cleavage site therefore remained unknown, nor could any alternative cleavage sites within the same cap donor be identified. The *in vitro* methods, on the other hand, allowed the possibility of supplying cap donors with precisely known features in the cap-snatching mechanism (Bouloy *et al.*, 1980; Ulmanen *et al.*, 1981; Patterson *et al.*, 1984; Galarza *et al.*, 1996), but the conditions under which it would take place might not have reflected the *in vivo* situation at all.

The approaches described in this study combine the advantage of natural, *in vivo* conditions with the use of known and even mutable leaders. Specific mutant AMV leaders can be generated easily and inoculated mechanically either as DNA constructs or as *in vitro* transcripts on transgenic ('p12') plants, to become amplified by the AMV p1 and p2 replicase proteins to high levels throughout the plant.

Following this *in vivo* approach, combined with the *in vivo* analyses of two selected host mRNAs as cap donors, we could demonstrate that a single base complementarity is required for a capped leader RNA to prime successfully on the viral template. This base pairing should occur preferentially at position +16 of the donor RNA, although all positions between nucleotides 13 and 18 can be used with different efficiencies (Figure 3B). Furthermore, this base pairing can occur not only with the 3'-ultimate A residue of the viral template (apparently the most optimal scenario), but also with the penultimate G and even the antepenultimate A residue, as was observed

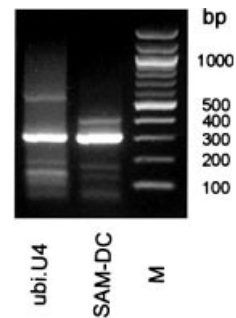


Fig. 5. RT-PCR detection of TSWV mRNAs initiated with specific host leaders. TSWV N mRNAs containing host leaders were amplified from total RNA, isolated from infected *N.tabacum* p12 at 7 days post-infection. TSWV N mRNAs containing host leaders were amplified using primer N1 (RT) and primers N2 and UbiU4-1 or SAM-DC-1 (PCR). Products of ~315 bp represent TSWV N mRNAs. ubi.U4, RT-PCR product of TSWV N mRNAs containing ubi.U4-derived leaders; SAM-DC, RT-PCR product of TSWV N mRNAs containing SAM-DC-derived leaders; M, 100 bp molecular marker.

for an AMV RNA4 A16 leader-primed mRNA, which lacked the first two nucleotides of the authentic TSWV sequence (Table II). When evaluating the sequences of host leader-primed TSWV mRNAs from an earlier study (van Poelwijk *et al.*, 1996), a preference for cleavage at position 16 could be observed with priming on the 3'-ultimate residue of the viral template, as well as a possible priming of the used leaders to the 3'-penultimate residue.

Evidence for realignment of the recruited capped RNA primer on the viral template, resulting in repeated insertions of the first few nucleotides of the viral genomic sequence between leader sequence and (authentic) viral RNA sequence, has hardly been monitored in our studies. Only on one occasion (Table II; AMV4 A15) was an extra AG dinucleotide insertion found. Such repeated sequences have been observed more frequently with some animal-infecting viruses (e.g. Germiston, Hantaan, Bunyamwera, Dugbe, influenza A and B viruses) (Shaw and Lamb, 1984; Bouloy *et al.*, 1990; Jin and Elliot, 1993a,b; Garcin *et al.*, 1995), as well as with the plant-infecting Tenuiviruses (Huiet *et al.*, 1993; Shimizu *et al.*, 1996; Estabrook *et al.*, 1998). The inserted sequences have been explained first for Hantaan virus as being the result of a 'prime-and-realign' mechanism (Garcin *et al.*, 1995). The low frequency of insertions of 5'-terminal viral sequences between leader sequence and viral sequence (van Poelwijk *et al.*, 1996; Duijsings *et al.*, 1999) suggests that initiation of transcription for TSWV occurs at the 3'-ultimate A residue, rather than at the antepenultimate A residue. Therefore, a 'prime-and-realign' mechanism seems not to be favoured as a means for initiation of transcription. Future experiments may reveal why repeats of the 5'-terminal nucleotides are rarely seen within the TSWV mRNAs.

Although specific nucleotide composition and distance of the base pairing residue from the 5' end within the leader play an important role in the efficiency of use as a cap donor, the effects of specific secondary and tertiary structures within the leader have not been investigated yet. If secondary and tertiary structures indeed occur within the 5' end of the leader during the cap-snatching process, these may influence the physical distance between cap structure

and possible cleavage sites, thereby altering the optimal site for endonucleolytic cleavage. It is likely, though, that the viral polymerase complex disrupts these secondary and tertiary structures within the 5' end of the leader. The viral nucleoprotein, a single-strand RNA-binding protein (Richmond *et al.*, 1998) that is part of the viral transcriptase complex, may play a role in this.

In summary, the combined analyses of mutated AMV RNAs and host mRNAs have led to improved insight into the requirements for the length and specific nucleotide composition of cap donors during TSWV transcription initiation. Moreover, it has resolved the base pairing requirement during cap snatching, which may apply to all segmented negative-strand RNA viruses.

Materials and methods

Host plants

Transgenic *N. tabacum* cv. Samsun NN plants expressing AMV replicase proteins P1 and P2 (referred to as p12 plants) were used for *in vivo* replication of wild-type and mutant AMV RNA3 and RNA4 from cloned cDNAs as described previously (Neeleman *et al.*, 1993).

Construction of plasmids

Plasmid pCa32T, which contains a cDNA copy of the wild-type AMV RNA3 flanked by the CaMV 35S promoter and nopaline synthase (nos) terminator (Neeleman *et al.*, 1991), was used as a source to create mutants of AMV RNA3 and RNA4. Mutant AMV RNA3 constructs that contained a point mutation at either position 17 or 18 of the AMV RNA3 sequence were made by amplifying pCa32T using primers α 35S (ctctccaagaatgaactcc, complementary to the 35S promoter) and A3/D17 (gtattaataccatttDaaatattccaattc, identical to nucleotides 1–32 of the AMV RNA3 sequence; D = A, G or T) or A3/B18 (gtattaataccatttcBaaatattccaattc; B = C, G or T) and the Expand Long template PCR system (Roche). Amplified PCR fragments were purified using the High Pure PCR purification kit (Roche), restriction enzyme digested with *DpnI* (to destroy input template DNA) and ligated using T4 DNA ligase (Promega). Individual clones were verified by sequence analysis. Point mutants of AMV RNA3 at nucleotide 17 are referred to as C17A (in which nucleotide 17 was changed from the wild-type C residue into an A residue), C17G and C17U. Point mutants of AMV RNA3 at nucleotide 18 are referred to as A18C (where nucleotide 18 was mutated into a C residue), A18G and A18U. Similarly, point mutants in the subgenomic promoter region of AMV RNA4 were derived from pCa32T using primers A4/rev (aaaataaaaacgcccattaccg, complementary to nucleotide positions 1250–1272 of the AMV RNA3 sequence), A4/A12–A4/A18 (ATtttttttccaataacttccatcatgag; TAttttttccaataacttccatcatgag; TTAAtttttccaataacttccatcatgag; TTTTAtttttccaataacttccatcatgag; TTTTAtttttccaataacttccatcatgag; TTTTAtttttccaataacttccatcatgag; TTTTAtttttccaataacttccatcatgag; TTTTAtttttccaataacttccatcatgag) and A4/G15 (TTTtGtttttccaataacttccatcatgag) (all identical to nucleotide positions 1273–1302 of the wild-type RNA3 sequence). Point mutants of AMV RNA4 are referred to as A12 (containing an A residue at nucleotide 12 of the wild-type RNA4 sequence and U residues at nucleotides 13 and 14), A13, A14, A15, A16, A17, A18 and G15 (Table II). Likewise, mutant RNA4-noA was made, containing a poly(U) tract between nucleotides 12 and 18 of the wild-type RNA4 sequence.

Inoculation of p12 plants

Transgenic *N. tabacum* cv. Samsun NN plants expressing AMV replicase proteins P1 and P2 (p12 plants) were grown under greenhouse conditions and mechanically inoculated with 35S-cDNA constructs and TSWV strain BR-01 as described previously (Neeleman *et al.*, 1991; Taschner *et al.*, 1991; Duijsings *et al.*, 1999).

Analyses of AMV-TSWV mRNA sequences

TSWV N mRNAs containing capped 5' nucleotide sequences derived from AMV RNA3 and RNA4 were detected and cloned into the pGEM-T vector (Promega) as described previously (Duijsings *et al.*, 1999). Briefly, total RNA was isolated from systemically infected leaf material as described by Gurr and McPherson (1992). First-strand cDNA was synthesized from this total RNA and a nested PCR amplification was

performed subsequently on the synthesized first-strand cDNA using a primer identical to the first 11 nucleotides of the AMV RNA leader sequences. The PCR products obtained were purified using the High Pure PCR purification kit (Roche) and cloned into pGEM-T (Promega) according to the manufacturer's procedures. Sequence analysis of the clones obtained was performed using the Sanger dideoxy method (Amersham-Pharmacia).

Analyses of host leader sequences in TSWV mRNAs

TSWV N mRNAs containing capped 5' leader sequences derived from different host (*N. tabacum*) genes were detected by nested RT-PCR and analysed by sequence determination as described above. In brief, amplification on first-strand cDNA material was performed with a nested TSWV primer in combination with a primer specific for the first 11 nucleotides of the 5' end of different host genes. Host genes chosen were a polyubiquitin gene, with corresponding primer UbiU4-1 (CCC~~GGATCC~~ATCC~~TTT~~GATT) and a SAM-DC gene with corresponding primer SAM-DC-1 (CCC~~GGATCC~~ATGGAGT~~CGAA~~).

Acknowledgements

The authors would like to thank Professor John Bol, Leiden University, for providing the p12 plants and plasmid pCa32T. This research was supported by The Netherlands Foundation for Chemical Sciences (C.W.) with financial aid from The Netherlands Organisation for Scientific Research (N.W.O.).

References

- Bishop, D.H.L., Gay, M.E. and Matsuoko, Y. (1983) Nonviral heterogeneous sequences are present at the 5' ends of one species of snowshoe hare bunyavirus S complementary RNA. *Nucleic Acids Res.*, **11**, 6409–6418.
- Bouloy, M., Plotch, S.J. and Krug, R.M. (1980) Both the 7-methyl and the 2'-O-methyl groups in the cap of mRNA strongly influence its ability to act as primer for influenza virus RNA transcription. *Proc. Natl Acad. Sci. USA*, **77**, 3952–3956.
- Bouloy, M., Pardigon, N., Vialat, P., Gerbaud, S. and Girard, M. (1990) Characterization of the 5' and 3' ends of viral messenger RNAs isolated from BHK21 cells infected with Germiston virus (Bunyavirus). *Virology*, **175**, 50–58.
- Braam, J., Ulmanen, I. and Krug, R.M. (1983) Molecular model of a eukaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell*, **34**, 609–618.
- Caton, A.J. and Robertson, J.S. (1980) Structure of the host-derived sequences present at the 5' ends of influenza virus mRNA. *Nucleic Acids Res.*, **8**, 2591–2603.
- Chung, T.D., Cianci, C., Hagen, M., Terry, B., Matthews, J.T., Krystal, M. and Colonno, R.J. (1994) Biochemical studies on capped RNA primers identify a class of oligonucleotide inhibitors of the influenza virus RNA polymerase. *Proc. Natl Acad. Sci. USA*, **91**, 2372–2376.
- Collett, M.S. (1986) Messenger RNA of the M RNA segment RNA of Rift Valley fever virus. *Virology*, **151**, 151–156.
- Dhar, R., Chanock, R.M. and Lai, C.J. (1980) Nonviral oligonucleotides at the 5' terminus of cytoplasmic influenza viral mRNA deduced from cloned complete genomic sequences. *Cell*, **21**, 495–500.
- Dobie, D.K., Blair, C.D., Chandler, L.J., Rayms-Keller, A., McGaw, M.M., Wasieloski, L.P. and Beaty, B.J. (1997) Analysis of La Crosse virus S mRNA 5' termini in infected mosquito cells and *Aedes triseriatus* mosquitoes. *J. Virol.*, **71**, 4395–4399.
- Duijsings, D., Kormelink, R. and Goldbach, R. (1999) Alfalfa mosaic virus RNAs serve as cap donors for tomato spotted wilt virus transcription during coinfection of *Nicotiana benthamiana*. *J. Virol.*, **73**, 5172–5175.
- Eshita, Y., Ericson, B., Romanowski, V. and Bishop, D.H.L. (1985) Analyses of the mRNA transcription processes of snowshoe hare bunyavirus S and M RNA species. *J. Virol.*, **55**, 681–689.
- Estabrook, E.A., Tsai, J. and Falk, B.W. (1998) *In vivo* transfer of barley stripe mosaic hordeivirus ribonucleotides to the 5' terminus of maize stripe tenuivirus RNAs. *Proc. Natl Acad. Sci. USA*, **95**, 8304–8309.
- Galarza, J.M., Peng, Q., Shi, L. and Summers, D.F. (1996) Influenza A virus RNA-dependent RNA polymerase: analysis of RNA synthesis *in vitro*. *J. Virol.*, **70**, 2360–2368.
- Garcin, D. and Kolakofsky, D. (1990) A novel mechanism for the

- initiation of Tacaribe arenavirus genome replication. *J. Virol.*, **64**, 6196–6203.
- Garcin,D., Lezzi,M., Dobbs,M., Elliott,R.M., Schmaljohn,C., Kang,C.Y. and Kolakofsky,D. (1995) The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis. *J. Virol.*, **69**, 5754–5762.
- Gerbaud,S., Vialat,P., Pardigon,N., Wychowski,C., Girard,M. and Bouloy,M. (1987) The S segment of the Germiston virus RNA genome can code for three proteins. *Virus Res.*, **8**, 1–13.
- Grò,M.C., Di Bonito,P., Accardi,L. and Giorgi,C. (1992) Analysis of 3' and 5' ends of N and NSs messenger RNAs of Toscana phlebovirus. *Virology*, **191**, 435–438.
- Gurr,S.J. and McPherson,M.J. (1992) Nucleic acids techniques. In Gurr,S.J., McPherson,M.J. and Bowles,D.J. (eds), *Molecular Plant Pathology: A Practical Approach*. Vol. 1. IRL Press, Oxford, UK, pp. 112–113.
- Hagen,M., Chung,T.D.Y., Butcher,J.A. and Krystal,M. (1994) Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J. Virol.*, **68**, 1509–1515.
- Hagen,M., Tiley,L., Chung,T.D.Y. and Krystal,M. (1995) The role of template–primer interactions in cleavage and initiation by the influenza virus polymerase. *J. Gen. Virol.*, **76**, 603–611.
- Honda,A., Mizumoto,K. and Ishihama,A. (1986) RNA polymerase of influenza virus: dinucleotide-primed initiation of transcription at specific positions on viral RNA. *J. Biol. Chem.*, **261**, 5987–5991.
- Huiet,L., Feldstein,P.A., Tsai,J.H. and Falk,B.W. (1993) The maize stripe virus major noncapsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5' termini. *Virology*, **197**, 808–812.
- Ihara,T., Matsuura,Y. and Bishop,D.L. (1985) Analyses of the mRNA transcription processes of Punta Toro phlebovirus (Bunyaviridae). *Virology*, **147**, 317–325.
- Jin,H. and Elliott,R.M. (1993a) Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus. *J. Virol.*, **67**, 1396–1404.
- Jin,H. and Elliott,R.M. (1993b) Non-viral sequences at the 5' ends of Dugbe nairovirus S mRNAs. *J. Gen. Virol.*, **74**, 2293–2297.
- Kormelink,R., de Haan,P., Peters,D. and Goldbach,R. (1992a) Viral RNA synthesis in tomato spotted wilt virus-infected *Nicotiana rustica* plants. *J. Gen. Virol.*, **73**, 687–693.
- Kormelink,R., van Poelwijk,F., Peters,D. and Goldbach,R. (1992b) Non-viral heterogeneous sequences at the 5' ends of tomato spotted wilt virus mRNAs. *J. Gen. Virol.*, **73**, 2125–2128.
- Krug,R.M., Broni,B.A., LaFiandra,A.J., Morgan,M.A. and Shatkin,A.J. (1980) Priming and inhibitory activities of RNAs for the influenza viral transcriptase do not require base pairing with the virion template RNA. *Proc. Natl Acad. Sci. USA*, **77**, 5874–5878.
- Neeleman,L., van der Kuyl,A.C. and Bol,J.F. (1991) Role of alfalfa mosaic virus coat protein gene in symptom formation. *Virology*, **181**, 687–693.
- Neeleman,L., van der Vossen,E.A. and Bol,J.F. (1993) Infection of tobacco with alfalfa mosaic virus cDNAs sheds light on the early function of the coat protein. *Virology*, **196**, 883–887.
- Patterson,J.L. and Kolakofsky,D. (1984) Characterization of La Crosse virus small-genome transcripts. *J. Virol.*, **49**, 680–685.
- Patterson,J.L., Holloway,B. and Kolakofsky,D. (1984) La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. *J. Virol.*, **52**, 215–222.
- Plotch,S.J., Bouloy,M., Ulmanen,I. and Krug,R.M. (1981) A unique cap (m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, **23**, 847–858.
- Raju,R., Raju,L., Hacker,D., Garcin,D., Compans,R. and Kolakofsky,D. (1990) Nontemplated bases at the 5' ends of Tacaribe virus mRNAs. *Virology*, **174**, 53–59.
- Ramirez,B.-C., Garcin,D., Calvert,L.A., Kolakofsky,D. and Haenni,A.-L. (1995) Capped nonviral sequences at the 5' end of the mRNAs of rice Hoja Blanca virus RNA4. *J. Virol.*, **69**, 1951–1954.
- Richmond,K.E., Chenault,K., Sherwood,J.L. and German,T.L. (1998) Characterization of the nucleic acid binding properties of tomato spotted wilt virus nucleocapsid protein. *Virology*, **248**, 6–11.
- Shaw,M.W. and Lamb,R.A. (1984) A specific sub-set of host-cell mRNAs prime influenza virus mRNA synthesis. *Virus Res.*, **1**, 455–467.
- Shimizu,T., Toriyama,S., Takahashi,M., Akutsu,K. and Yoneyama,K. (1996) Non-viral sequences at the 5' termini of mRNAs derived from virus-sense and virus-complementary sequences of the ambisense RNA segments of rice stripe tenuivirus. *J. Gen. Virol.*, **77**, 541–546.
- Simons,J.F. and Pettersson,R.F. (1991) Host-derived 5' ends and overlapping complementary 3' ends of the two mRNAs transcribed from the ambisense S segment of Uukuniemi virus. *J. Virol.*, **65**, 4741–4748.
- Taschner,P.E., van der Kuyl,A.C., Neeleman,L. and Bol,J.F. (1991) Replication of an incomplete alfalfa mosaic virus genome in plants transformed with viral replicase genes. *Virology*, **181**, 445–450.
- Ulmanen,I., Broni,B.A. and Krug,R.M. (1981) Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m⁷GpppNm) on RNAs and in initiating viral RNA transcription. *Proc. Natl Acad. Sci. USA*, **78**, 7355–7359.
- van Poelwijk,F., Kolkman,J. and Goldbach,R. (1996) Sequence analysis of the 5' ends of tomato spotted wilt virus N mRNAs. *Arch. Virol.*, **141**, 177–184.

Received January 15, 2001; revised March 6, 2001;
accepted March 23, 2001