The entire nucleotide sequence of foxtail mosaic virus RNA

J. B. Bancroft,¹* Michèle Rouleau,² R. Johnston,¹ L. Prins² and George A. Mackie²

Departments of ¹Plant Sciences and ²Biochemistry, University of Western Ontario, London, Ontario, Canada N6A 5B7

The nucleotide sequence of the RNA genome of foxtail mosaic virus (FMV), a member of the potexvirus family, is 6151 nucleotides long, exclusive of a poly(A) tail. The RNA contains five principal open reading frames (ORFs), designated from the 5' terminus as encoding proteins with Mr values of 152.3K (ORF1), 26.4K (ORF2) which overlaps an 11.3K (ORF3) product, 5.8K (ORF4) which overlaps a 28.8K readthrough protein (ORF5A) which leads into the coat protein cistron of 23.7K (ORF5). The sizes and composition of the proteins encoded by the ORFs are generally similar to those found in other potexviruses; the least similar is the coat protein which nonetheless retains apparently critical consensus regions. The 5' terminus of the previously reported 0.9 kb subgenomic (sg) RNA was determined by S1 nuclease mapping and shown to begin with the sequence GAAGA, 43 nucleotides upstream from the first nucleotide of the coat protein initiation codon. The positions of the 5' end of this sgRNA and of that deduced from the nucleotide sequence for a 1.9 kb sgRNA are entirely consistent with the previously published sizes of these sgRNAs.

Introduction

The potexviruses form a large group of flexuous filamentous plant viruses with lengths between 470 and 580 nm (Koenig & Lesemann, 1978) and positive-sense ssRNA genomes. The sequence of the genomic RNA (gRNA) of some members of the potexvirus group has been completed: white clover mosaic virus (WCIMV) (strain M, Forster et al., 1988; strain O, Beck et al., 1990), potato virus X (PVX) (Russian strain, Skryabin et al., 1988a; strain X3, Huisman et al., 1988; Andean strain, Orman et al., 1990), narcissus mosaic virus (NMV) (Zuidema et al., 1989), papaya mosaic virus (PMV) (Sit et al., 1989) and clover yellow mosaic virus (CYMV) (Sit et al., 1990). Partial sequences are known for potato aucuba mosaic virus (PaMV) (Bundin et al., 1986), lily virus X (LVX) (Memelink et al., 1990) and strawberry mild yellow edge-associated virus (SMYEAV) (Jelkmann et al., 1990). The organization of potential coding sequences in these viruses is generally well-conserved and supports a model for gene organization proposed by Bendena & Mackie (1986).

Foxtail mosaic virus (FMV), which primarily infects members of the Gramineae (Paulsen & Niblett, 1977), is serologically related to some members of the potexvirus group (Short, 1983) and is structurally similar to all of those examined (Richardson et al., 1981). FMV resembles other potexviruses in that it produces two subgenomic RNAs (sgRNAs) during infection (Mackie et al., 1988), of approximately 0.9 kb and 1.9 kb, the smaller encoding the coat protein. In this report, we continue the description of FMV by specifying the nucleotide sequence of its gRNA and by identifying the origin of its coat protein subgenomic message.

Methods

cDNA synthesis and mapping. FMV was grown in barley (var. Herta) and purified as for CYMV (Bancroft et al., 1979). Viral RNA was extracted as described by Erickson & Bancroft (1978). Double-stranded cDNA was prepared essentially by the method of Gubler & Hoffman (1983). Viral RNA was heated at 90 °C for 1 min and immediately annealed to oligo(dT)₁₁₅ on ice before first-strand synthesis was catalysed by Moloney murine leukaemia virus reverse transcriptase (Bethesda Research Laboratories) for 40 min at 50 °C. After second-strand synthesis, the cDNAs were blunt-ended, ligated to either EcoRI or XhoI linkers and ligated into the vector pSP65 (Melton et al., 1984) prior to transformation of either Escherichia coli MM294 or JM109 strains as described by Mamiatis et al. (1982).

cDNA sequencing. Subclones of overlapping cDNAs were prepared by digestion of larger cloned cDNAs with suitable restriction enzymes and ligation of the smaller cDNA fragments into pSP64 (Melton et al., 1984) or pTZ18U (Mead et al., 1986). The resultant plasmids were isolated either by CsCl centrifugation or polyethylene glycol precipitation (Hattori & Sakaki, 1986) prior to DNA sequencing by the dye deoxynucleotide chain termination method (Sanger et al., 1977) usually using T7 DNA polymerase (Tabor & Richardson, 1987). All nucleotide sequence data were assembled and analysed with the PC Gene programs (IntelliGenetics) or with the University of Wisconsin Genetics Computing Group (UWGCG) programs installed on a VAX computer (Devereux et al., 1984).

The sequence data reported appear in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession number M62730.
RNA sequencing. The identity of the 5'-terminal nucleotides of FMV gRNA was determined by dideoxynucleotide sequencing of viral RNA with avian myeloblastosis virus reverse transcriptase as described by Sit et al. (1990) using the synthetic primer 5' ACTGCCTCGATAGA-

Protein sequence determination. A lyophilized sample (0.5 mg) of FMV capsid protein, prepared by extraction of the virus with 67% glacial acetic acid (Fraenkel-Conrat, 1957), was cleaved by incubation in 0.5 ml 70% formic acid containing 50 μl of 0.7 g/ml CNBr (dissolved in acetonitrile) at room temperature overnight and peptides were separated by HPLC. The sequence of one clearly resolved peptide was determined on a Porter gas phase protein sequencer.

Results and Discussion

Nucleotide sequence and genome organization

Cloned cDNAs in plasmid vectors were prepared from purified FMV gRNA and characterized by restriction mapping. Their sizes were determined to range between 0.5 and 4.8 kb. DNA sequence data were obtained from these plasmids or from smaller subcloned fragments using synthetic primers specific for the vector or for viral sequences already determined. In this way, all but the 5' terminus of the gRNA was sequenced on one or more recombinant plasmids, 85% being on both strands. Dideoxynucleotide sequencing of the gRNA indicated that the first five residues were NGAAA. The first
residue is likely to be an artefact of reverse transcription caused by a cap structure (Ahlquist & Janda, 1984; Allison et al., 1988; Sit et al., 1990) whose presence is expected by analogy to other potexviruses (Sonenberg et al., 1978; AbouHaidar & Bancroft, 1978; Sit et al., 1990). The extreme Y-terminal sequence of FMV RNA, GAAACUCUUCC, is very similar to the terminal sequence of other potexviral RNAs (Zuidema et al., 1989). The complete nucleotide sequence of FMV RNA is presented in Fig. 1. It is 6151 nt long excluding the poly(A) tail, this size being intermediate compared to that of the other members of the potexvirus family.

The organization of the FMV RNA genome, deduced from its nucleotide sequence, is depicted in Fig. 2. The first open reading frame (ORF) initiates at nt 81 and terminates at nt 4088, encoding a protein with an M₀ of 152.3K. This size is consistent with the 160K protein produced by in vitro translation of FMV gRNA (Bendena & Mackie, 1986). Interestingly, the context of the initiation codon for ORF1 (AUGUC) differs from that of all other potexviruses (AUGGC) as well as from the consensus context (AUGGC) in plant mRNAs (Lütcke et al., 1987). ORF1 contains two smaller out-of-frame ORFs, ORF6 (26K) and ORF7 (10K) designated by the dashed-line boxes in Fig. 2. Analogous ORFs are not found in all potexviruses but are present in PMV (Sit et al., 1989) and CYMV (Sit et al., 1990) in which ORF6 encodes a 14K protein. WCIMV RNA of strain M, but not of strain O, encodes a 10K product within ORF1 (Forster et al., 1988) corresponding in size and position to ORF7 found in FMV RNA. The significance of these internal ORFs remains to be established. ORF2, nt 4132 to nt 4842, has a product of predicted M₀ of 26.4K. ORF2 shares its last 68 nt with ORF3 which begins at nt 4775 and finishes at nt 5092, encoding a protein with an M₀ of 11.3K. ORF2, nt 4132 to nt 4842, is translated in vitro from FMV gRNA. ORF5A starts at nt 5527 and leads into the 23.7K coat protein cistron (ORF5) which initiates at nt 5371 and terminates at nt 6018.

The identity of ORF5 as encoding the coat protein was established unambiguously by a direct comparison with the amino acid sequence of 18 residues of an internal peptide of purified FMV coat protein. The sequence of this peptide obtained from a CNBr digest is (M)KSLALACKDADVPHKL and matches the sequence predicted from ORF5 between residues 5626 to 5679 with two exceptions. The seventh and twelfth residues of the sequenced peptide (underlined above) appeared to be A and D, respectively, whereas the nucleotide sequence at residues 5644 to 5646 and 5659 to 5661 clearly predicts R and G. We cannot explain this discrepancy but note that different preparations of virus were employed for protein sequencing and cDNA synthesis. In any event, the residues affected are not 'consensus' residues (see below). The size of the coat protein as well as the amino acid composition predicted from the nucleotide sequence corresponds to that estimated by Short & Davies (1987) from amino acid analyses. It is however approximately 7K smaller than that previously established by electrophoresis (Mackie et al., 1988). The anomalous mobility of the coat protein in the SDS–polyacrylamide gel presumably reflects non-ideal binding of detergent.

The 3' non-coding region of FMV RNA is 133 nucleotides long and is followed by a poly(A) tail of variable length. A polyadenylation signal (AAUAAA) is found in the non-coding region of FMV, 107 nucleotides from the poly(A) tail, as well as in most but not all 3' end sequences of potexviruses (Table 1), as has been deduced independently by Guilford et al. (1991). In addition, a common six-nucleotide sequence, ACUUA, was identified 30 to 60 nt from the 3' terminus of all potexviral sequences (Table 2). It is conceivable that this consensus sequence may be involved in the production of the

Table 1. Putative polyadenylation signals in potexviral RNAs

<table>
<thead>
<tr>
<th>Virus</th>
<th>Putative signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMYEAV</td>
<td>Absent</td>
</tr>
<tr>
<td>NMV</td>
<td>AAUAAAAAUAAA – 11 nt – (A)n</td>
</tr>
<tr>
<td>WCIMV (M)</td>
<td>AAUAAAA – 7 nt – (A)n</td>
</tr>
<tr>
<td>PVX (X3)</td>
<td>Absent</td>
</tr>
<tr>
<td>CYMV</td>
<td>AAUAAA – 104 nt – (A)n</td>
</tr>
<tr>
<td>PMV</td>
<td>AAUAAA – 119 nt – (A)n</td>
</tr>
<tr>
<td>LVX</td>
<td>AAUAAc – 97 nt – (A)n</td>
</tr>
<tr>
<td>PaMV</td>
<td>AAUAAA – 11 nt – (A)n</td>
</tr>
<tr>
<td>FMV</td>
<td>AAUAAAAAA – 107 nt – (A)n</td>
</tr>
</tbody>
</table>

Fig. 2. The genomic organization of FMV. ORF1 to ORF5 are shown by numbered boxes, with the M₀ values of their products. The coding frame of each ORF is indicated by the position of boxes relative to the baseline. The dashed-line boxes represent internal ORFs within ORF1. The asterisks in ORF1 and ORF2 locate the NTP-binding helicase domain. The triangle in ORF1 designates the position of the RNA-dependent RNA polymerase consensus sequence. The arrows correspond to the relative locations of the large (1.9 kb) and small (0.9 kb) sgRNAs.
Table 2. Conserved sequences near the 3' termini of potexviral RNAs

<table>
<thead>
<tr>
<th>Virus</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMYEAV</td>
<td>ACUUA - 34 nt - (A)n</td>
</tr>
<tr>
<td>NMV</td>
<td>ACUUA - 59 nt - (A)n</td>
</tr>
<tr>
<td>WCIMV (M)</td>
<td>ACUUA - 33 nt - (A)n</td>
</tr>
<tr>
<td>PVX (X3)</td>
<td>CAUAA - 58 nt - (A)n</td>
</tr>
<tr>
<td>CYMV</td>
<td>ACUAA - 35 nt - (A)n</td>
</tr>
<tr>
<td>PMV</td>
<td>ACUUA - 32 nt - (A)n</td>
</tr>
<tr>
<td>LVX</td>
<td>ACUAA - 32 nt - (A)n</td>
</tr>
<tr>
<td>PaMV</td>
<td>ACUAA - 60 nt - (A)n</td>
</tr>
<tr>
<td>FMV</td>
<td>ACUAA - 39 nt - (A)n</td>
</tr>
</tbody>
</table>

minus-strand gRNA during viral replication by acting as a recognition sequence for the replicase complex. Another partially conserved sequence near the 3' terminus of some potexvirus RNAs has been identified recently (Orman et al., 1990) but is matched poorly by FMV gRNA (13 of 23 residues).

Several ORFs which might encode proteins of 6-4K to 9-6K are found in the negative RNA strand, as is also the case for PMV, PVX and WCIMV (strain M) minus-strand RNAs. Their significance is unknown.

Similarity of the putative proteins of the FMV genome to corresponding proteins of other potexviruses

The sizes of the anhydrous proteins predicted from the FMV ORFs are generally within the range found for other potexviruses. FMV, however, possesses the smallest ORF3 (11.3K) and ORF4 (5.8K) of all potexviruses (11.9K to 14K and 6.5K to 11K for ORF3 and ORF4, respectively, for other members).

The predicted protein sequences of all potexviruses (including FMV) were compared using the GAP alignment program (UWGCG) with a gap weight of 3.0 and a length weight of 0.1. The percentages of identity and similarity of residues encoded by each ORF with the corresponding ORF of other members of the potexvirus group were obtained and are summarized in Table 3. The ORF1 product from FMV is closely related to that of the other potexviruses, since the percentage of similarity of both the amino and carboxyl ends is in a constant range for all comparisons (Table 3). The C-terminal domain of FMV ORF1 contains both the NTP-binding helicase consensus sequence (Gorbunova et al., 1987; Hodgman, 1988) and the RNA-dependent RNA polymerase consensus motif (Argos, 1988) (Fig. 2), as reported for other potexviruses (Skryabin et al., 1988b). These motifs are well conserved not only in potexviruses but also in most positive-strand RNA plant viruses (Habili & Symons, 1989). The 26K protein (ORF2) of FMV also contains a NTP-binding helicase motif, located at the N terminus as in other potexviruses (Skryabin et al., 1988b). The remainder of ORF2 sequences from all other potexviruses are less related to each other than are ORF1 and ORF3 (11K protein) sequences (Table 3; see also Zuidema et al., 1989). FMV ORF3 and ORF4 are related to those of other potexviruses in that they contain stretches of hydrophobic residues predicted to form transmembrane helices (Morozov et al., 1987; Forster et al., 1987; Skryabin et al., 1988b) bordered by charged residues. Although FMV has the smallest ORF4 of all potexviruses, the common N-terminal hydrophobic region is conserved. Frequent conservative exchanges among hydrophobic amino acids have occurred suggesting that the function of the product of ORF4, if any, is not impaired by such replacements.

The coat proteins (ORF5) of potexviruses are the only translation products with a defined function which may be used to help interpret their amino acid sequences. The amino acid sequence of the FMV coat protein is not as similar to the coat protein sequences of the other potexviruses as these are amongst themselves (Table 3) but maintains consensus regions (Fig. 3). In the optimal alignment in Fig. 3, only 10 residues are absolutely conserved and are predominantly found between amino acids 121 to 145 (on the FMV sequence). The most striking feature in this region is the amphipathic 'core' consensus KFAAFDFDGV. In contrast to the relative divergence of potexviral coat proteins, 25 of 158 residues in seven different tobamoviruses are absolutely conserved (Altshuh et al., 1987). This relatively high number

Table 3. Similarities of FMV ORFs to ORFs in other sequenced potexviruses compared to similarities among all other potexviruses

<table>
<thead>
<tr>
<th>ORF</th>
<th>FMV</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range Average</td>
<td>Range Average</td>
</tr>
<tr>
<td></td>
<td>Sim. (%) Sim. (%) Id. (%) Sim. (%) Sim. (%) Id. (%)</td>
<td></td>
</tr>
<tr>
<td>ORF1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5' end</td>
<td>61-67 63 41</td>
<td>62-72 67 48</td>
</tr>
<tr>
<td>3' end</td>
<td>63-68 66 49</td>
<td>67-73 70 55</td>
</tr>
<tr>
<td>ORF2</td>
<td>48-56 52 31</td>
<td>46-63 53 34</td>
</tr>
<tr>
<td>ORF3</td>
<td>57-64 60 44</td>
<td>48-64 56 38</td>
</tr>
<tr>
<td>ORF4</td>
<td>46-60 53 27</td>
<td>46-55 51 28</td>
</tr>
<tr>
<td>ORF5 (CP)</td>
<td>40-46 43 25</td>
<td>47-65 55 39</td>
</tr>
</tbody>
</table>

* Data for ORF1 and ORF4 were obtained from a six member comparison, data for ORF2 and ORF3 from a seven member comparison and for ORF5, nine potexviral sequences were compared.
† Percentage similarity represents the sum of the percentage identical ('Id.') and similar ('Sim.') residues.
‡ First 400 residues of ORF1.
§ Last 800 residues of ORF1.
SMYE  mgdqprpppvppapgsnplmgstppvp1gp1rtnannpangqdpfvr1tpeel.aapisaaSNKVATr.
NMV  matpqttdpkpanaldspnraPssldikkklyestttatVAtpa
WCIMV  mattrpptPstdlralkytstvsNSpa
PVX  mrsapastqagtstststttaqtatgatapasglft1qdggfttaraiavSNaVAt.n
CYMV  mttdktltfssatptdeqtldtititielSN1VpSis
PMV  mkssmmstptnfaiptqegmmsi1kvdptSNlisp.s.q
LVX  mttrvPdaktwadtaqtaesstvAat
PaMV  mvdskkttetpqvdaskaenkstqagrichflasapkqfasasdvrspslalldelayrttsISap.a
FMV  matgnadvtdatdkkokkctxkqprSNkapSde
Cons  --------------------------------------------P--------------SN-VAS--
  1
SMYE  EqilgivadlnalgfvgD.pAlgLdFLAfChyDiGssSpaqqvpGpspfccsr.mpvaAvvnr..hcTLRQ
NMV  Eqilglgfklklgl1.dAnsvApamwDLARayaDvqASrsvlsGpttspnpatrqa1Arqfvyini1TrpR
WCIMV  EleaikktWtetkiPnDivplaczWDLARafaDaVGSskekGdasaaglyvrsrklAQaiK.ihcTrIQ
PVX  EdlskeiaikmdkmlPdtDmaAqAwDLvrHCaDvGSaqtmtgyngsrsraLAAaiK.eveCTIQ
CYMV  EleaikktWtetkiPnDivplaczWDLARafaDaVGSskekGdasaaglyvrsrklAQaiK.ihcTrIQ
PMV  EqklsstlmuvakvPAsavttvAeLVnFyDnGsSaYttvtrGps.sipeislaQLAsivKasstslRk
LVX  Elqsiai1WsegigipaAnffd.vAfqlAmrcScsGssstvlGcntcvyptttvk1aAglvK.viap1LRLQ
PaMV  EieavcqiWrinte1PAdkvAliAimAaryyaDvGAsrkaadltaptvstvrslAqIqagiaspIQ
FMV  ElvrinaaqrhgltPafqvq.aIvfim..DKGATdstitGkyn..tfpmksIAlarcKdagvypvkh
Cons  E---------------PAD--A--A--DLA--C--D--GAS-------------LA--K--TLRQ
  40
SMYE  lCmfYAPsvWVnkavrdNrPpgWNsnHisqptpKFAADFFDDG1v1PAnqaqve.p..LwRqPtqEiyAaTh
NMV  lCmfYaKAVWV11LdsNVPPAPqWqGqGlpddcFKAfDFFFgeV1sePAa1pdG1RLRpPsgqQlEiqAhsaTh
WCIMV  lCmfYAniWN11mdLdtpPpaWSkIgyKKeaKFAgDFDDFDGVhNPAALmPD1RLRPsdaE1AhqaTh
PVX  lCmkyAPvVWNmlcNNePNPAWaqGfKpehKFADFFFqGvtnPAAlmPkeGL1RLPseaannAqqaTh
CYMV  lCfryAfK1W1vn1LrKNqPPANWsdqyKeaLrFADFFFDGVssSAA1sPplGLRpspnErnAnetN
PMV  lCfryAfpiWN1.LrtdkmAp1WNWASgyKpsaKFAADFFDDGVhENPAlmAqPpLSlRsPTqE1sAnetN
LVX  lCfryAfKvWVN1rsldoi1WNNWdsqFpeaerFADFFFDGVt1sAAPqPpd1RLRppTe1lS1esAqqaTh
PaMV  lCfryAf1W11mLnnNPPANWaniFk1kdyKFADFFF1aV1D1VAsPqP1qVhRPTdKeraAhqy
FMV  lCf1tyKpp1ayrunrvaNqPPAr1Wt1nvenp1kNwAAFDtFDa1l1DPFv.y.PssvpydeTPtd1qve1n
Cons  FC-YAYK-VWN--L--N--PPANW----G--K----KFAADFFDDGV----PAA--P--GLIR-PT--E--A--T--
  103
SMYE  KdVatyRaAskahdrisns.tllT11Kg8t1srtpspallpqpda
NMV  KygaLaRqrymetsfspPwlks1TvGsvstptclplh1qcnrntsk11k1c1g1
WCIMV  KqV1alRdAkp1twkhcrgq1c
PVX  aFvkitkarqsn1d.asldAar1Tr1Grrtqtttatavtt1p
CYMV  KnV1lytqAsr1s1lattsta1c.T1kGaytnasnafqyhrpe
PMV  KqV1lyf1qAqddnntstns.Af1TKG1qst1qfitlpppe
LVX  KfaaLaR...vrgs1fr.tae111h11graevstt11pp
PaMV  nw1sLs1Rrlqeg.ts1ttva1nK11Glghyynlpa1mmp
FMV  K1d1nls1a1srnql1..gtqasiTr1grl1ngp1pnlngyfiea1pq
Cons  K-V-L-R-A----------A-T-G----------
  172

Fig. 3. Comparison of the amino acid sequence of the coat proteins of nine potexviruses. A multiple alignment of the sequences was created using the program GAP (UWGCG) and was manually optimized. Conserved residues are in bold upper case. Underlined residues in the consensus (Cons) line are strictly conserved. The numbering is that of the FMV coat protein sequence. The sequences of strain M of WCIMV and of strain X3 of PVX were used to perform the comparison. SMYE refers to the sequence of the potexvirus-associated strain SMYEAV.
of conserved residues may be related to the fact that tobamovirus members are rigid helical viruses and all strains may require fairly uniform properties encompassed in a rigid structure. Although all potexvirus subunits have the same general size and shape, they are flexuous, relatively 'loose' structures with variable true repeats (Richardson et al., 1981). Consequently, structural options not available to TMV may exist with flexuous viruses and be reflected in a greater permissible amino acid sequence variation than found among tobamoviruses, or among those of other rigid viruses.

Subgenomic termini and 'promoter'
To determine the 5' terminus of the 0.9 kb sgRNA, SI nuclease mapping was performed on RNA from barley leaves infected with FMV. As shown in Fig. 4, polyribosomal RNAs from healthy plants or from plants infected with CYMV were unable to protect the end-labelled probe (lanes 3 and 4, respectively) whereas protection extending to residue 5322 was obtained with polyribosomal RNA enriched in sgRNAs extracted from leaves infected with FMV (lane 1). The polyribosomal RNA fraction enriched for gRNA displayed a small degree of partial protection of the probe to residue 5322 (lane 2) as did gRNA obtained from purified virus (lane 5). Although some 'stutter' was observed, the protected fragment is clearly opposite to the sequence beginning (U)GAAGA on the sequencing ladder, taking into account that the 5' end of subgenomic potexvirus messages is capped (White & Mackie, 1990). The sequence GAAGA (nucleotides 5323 to 5327) corresponds exactly to that found for CYMV (White & Mackie, 1990) and is 43 nt upstream from the first nucleotide of the coat protein initiation codon, compared with 5 nt for CYMV. From the nucleotide sequence and the location of the subgenomic terminus, a theoretical size of 0.83 kb is calculated for the coat protein sgRNA [excluding the poly(A) tail]. This value is in agreement with a size of 0.9 kb estimated from gels (Mackie et al., 1988). The initiation codon for the 'readthrough' protein of ORF5A of FMV (Mackie et al., 1988) is not located within the 0.9 kb sgRNA.

A GAAGA sequence may also denote the possible location of the terminus of the 1.9 kb sgRNA at nts 4116 to 4120. A subgenomic message initiating at this location would have a size of 2.0 kb which corresponds to the experimental value of 1.9 kb (Mackie et al., 1988). The sequence GUUAGG located immediately upstream to the identified and deduced 5' ends for FMV sgRNAs is similar to putative 'promoter' sequences (Skryabin et al., 1988b; Sit et al., 1990; White & Mackie, 1990). Sequences similar to those at the 5' ends of the subgenomic RNAs are also found near the 5' termini of the gRNA in FMV, WCIMV and PMV but are less compelling in the other potexviruses.

No sgRNAs have been detected for the 11K (ORF3) and 6K (ORF4) cistrons of FMV (Mackie et al., 1988). The predicted size of a sgRNA for ORF4 would be close to that encoded by ORF5 (1.0 kb versus 0.83 kb without a tail) and may have escaped detection. A sgRNA for ORF3 would be about 1.4 kb and should be resolved from other viral RNA species. However, no messages other than those of ORF2 and ORF5 have been noted for FMV or for most potexviruses (Mackie et al., 1988). The presence of a series of additional sgRNAs has been reported in two cases, those of PVX (Dolja et al., 1987) and of daphne virus X (Guilford & Forster, 1986). The lack of sgRNAs for ORF3 and ORF4 of FMV is consistent with the absence of a potential 'promoter' sequence or of any other shared motif upstream of the coding regions of the two ORFs. The mechanism by which these ORFs are expressed, assuming that they are, is not understood. Nonetheless, ORF4 (p13) of beet necrotic yellow vein virus which is similar to ORF3 of FMV and of other potexviruses, is expressed in vivo (Niesbach-Klósgen et al., 1990) despite the apparent absence of a corresponding sgRNA. This may also be the case for the FMV ORF3 product.
This work was supported by grants from the National Science and Engineering Research Council (to J.B. and G.A.M.) and the Medical Research Council (to G.A.M.). M.R. was supported by an MRC studentship and L.P. was the recipient of an NSERC summer studentship.

References


(Received 4 March 1991; Accepted 4 June 1991)