The nucleotide sequence and genome organization of strawberry mild yellow edge-associated potexvirus

Wilhelm Jelkmann, 1* Edgar Maiss 2 and Robert R. Martin 3

1Federal Biological Research Centre for Agriculture and Forestry, Institute for Plant Protection in Fruit Crops, Schwabenheimer Strasse 101, D-6915 Dossenheim, 2Institute of Biochemistry and Plant Virology, Messeweg 11/12, D-3300 Braunschweig, Germany and 3Agriculture Canada Research Station, 6660 N.W. Marine Drive, Vancouver, British Columbia V6T 1X2, Canada

The nucleotide sequence (5966 nucleotides) of cDNA clones of strawberry mild yellow edge-associated potexvirus was determined. The genome contains six open reading frames (ORFs) encoding putative proteins with Mr's of 149,423, 25,344, 11,576, 8,079, 25,714 and 11,216. In the first three putative proteins and the coat protein considerable similarity was found to comparable polypeptides of the potexviruses potato virus X, clover yellow mosaic virus, narcissus mosaic virus, papaya mosaic virus, white clover mosaic virus and lily virus X.

Potexviruses are flexuous filamentous particles with a modal length of 470 to 580 nm (Koenig & Lesemann, 1978), and contain a positive-sense ssRNA which is capped at the 5' end (Sonnenberg et al., 1978) and polyadenylated at the 3' terminus (Morozov et al., 1983). Strawberry mild yellow edge-associated potexvirus (SMYEAV) has been identified recently in strawberry plants. The viral RNA is polyadenylated and the 3'-terminal open reading frame (ORF) encodes the coat protein (CP). Virus particles have been detected by immunoelectron microscopy using an antisera raised against a CP fusion protein produced in Escherichia coli. Based on the sequence of CP, particle morphology and its serological relationship with other potexviruses, SMYEAV has been proposed to be a member of the potexvirus group (Jelkmann et al., 1990).

Complete nucleotide sequences are available for potato virus X (PVX) strains Russian (Skryabin et al., 1988) and X3 (Huisman et al., 1988), clover yellow mosaic virus (CYMV; Sit et al., 1990), narcissus mosaic virus (NMV; Zuidema et al., 1989), papaya mosaic virus (PMV; Sit et al., 1989), and strains M and O of white clover mosaic virus (WC1MV; Beck et al., 1990). Partial sequences have been reported for lily virus X (LVX; Memelink et al., 1990) and potato aucuba mosaic virus (Bundin et al., 1986). In this study we present the genome organization and sequence of SMYEAV, and compare them to those of other potexviruses.

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases under the accession number D01227.

SMYEAV MY-18 was propagated in Rubus rosifolius (Martin et al., 1989). DsRNA was purified from these plants and used as template for cDNA cloning as described previously (Jelkmann et al., 1989).

Sequence information for both strands was obtained from cDNA clones that were unidirectionally deleted with exonuclease III and religated after mung bean nuclease treatment according to Henikoff (1984). When convenient, subclones were generated by restriction enzyme digestion and subcloned into the multiple cloning site of Bluescript M13+ and M13- (Stratagene), or pT7T3 18U and pT7T3 19U (Pharmacia). Sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977) was either from dsDNA or ssDNA templates prepared with the helper phage M13-K07 (Pharmacia). Reactions were performed with modified bacteriophage T7 DNA polymerase (Sequenase; USB) (Toneguzzo et al., 1988) or the Klenow fragment of DNA polymerase I (BRL). Direct dsRNA sequencing with reverse transcriptase (Boehringer) and two synthetic oligodeoxynucleotide primers [5' d(TGTTGTGTACGTCAGGA) 3', 5' d(TGCACTGCGGCAGGCAGTTGTTGT) 3'] was as described by Maiss et al. (1989). Another strategy for sequencing the 5' end using the polymerase chain reaction with the 39 base primer 5' d(GGGGGGATCCGGTACCATTTAGCTACACTATAGAAAAC) 3', kindly provided by R. L. S. Forster, was described in Beck et al. (1990). Sequence data were collected, assembled and analysed using the GCG program package (Devereux et al., 1984). Amino acid similarities were analysed with the FASTA computer program (Pearson & Lipman, 1988).
The SMYEAV sequence shown in Fig. 1 was determined from both strands of the cDNA clones shown in Fig. 2. Assembling the sequence data identified 5966 nucleotides excluding the poly(A) tail at the 3' end. Since attempts to purify virus particles failed, viral ssRNA could not be obtained. Several attempts to determine the exact 5' terminus from dsRNA using the methods described were not successful. From clone p398 74 non-coding nucleotides were identified at the 5' terminus. In other potexviruses the non-coding region at the 5' terminus contains 99, 80, 84, 107 and 94 nucleotides for PMV, NMV, PVX, WCIMV and CYMV, respectively. Thus the 5'-terminal nucleotides not determined are probably less than 0-5% of the total sequence.

Computer-assisted translation of the sequence revealed a genome organization (Fig. 2) similar to that of the other potexviruses for which complete nucleotide sequences have been reported. ORF 1 [nucleotides (nt)

\[ \text{SMYEAV cDNA.} \]

\[ \text{p24, p417, p463, p326 and p43.} \]

\[ \text{Within this sequence, beginning at the residue encoded by Sit et al., 1990). The codon encoding the first amino acid of this conserved sequence is located 12 to 14 amino acids 3' of the ATG start codon in potexviruses, and in potex- and carlaviruses, CYMV, PMV, NMV, PVX, and is a known sequence motif for putative viral RNA polymerases (Argos, 1988; Morozov et al., 1990). Other motifs conserved among RNA-dependent RNA polymerases (for a review see Poch et al., 1990). Asterisks indicate the alignment of nucleotide sequences of the putative promoter for subgenomic mRNA. Pu, denotes purine residues in positions \(-3(\text{A})\) and \(+4(\text{G})\) in relation to the proposed initiation codon. \]

\[ \text{Fig. 3. Possible ATT (CTG) initiation of translation at position 4096 (4070) of a polypeptide having similarity with the 24K to 26K products of other potexviruses (Memelink et al., 1990). Boxed amino acids indicate the highly conserved area mentioned in the text (Sit et al., 1990).} \]

\[ \text{An early start point of translation is indicated by the alignment of the SMYEAV sequence with the putative promoter for subgenomic mRNA.} \]

\[ \text{Fig. 2. The genome organization of SMYEAV and cDNA clones used to determine the nucleotide sequence. The dashed line in the CP-encoding ORF indicates a second methionine codon that possibly could initiate CP synthesis. The dashed line at the beginning of the proposed 25K protein ORF indicates a possible ATT start position. To exclude a possible mutation in a single cDNA clone leading to the elimination of an ATG start codon, this area was sequenced in the five independent cDNA clones p24, p417, p463, p326 and p43.} \]

\[ \text{The SMYEAV sequence shown in Fig. 1 was determined from both strands of the cDNA clones shown in Fig. 2. Assembling the sequence data identified 5966 nucleotides excluding the poly(A) tail at the 3' end. Since attempts to purify virus particles failed, viral ssRNA could not be obtained. Several attempts to determine the exact 5' terminus from dsRNA using the methods described were not successful. From clone p398 74 non-coding nucleotides were identified at the 5' terminus. In other potexviruses the non-coding region at the 5' terminus contains 99, 80, 84, 107 and 94 nucleotides for PMV, NMV, PVX, WCIMV and CYMV, respectively. Thus the 5'-terminal nucleotides not determined are probably less than 0-5% of the total sequence.} \]

\[ \text{Computer-assisted translation of the sequence revealed a genome organization (Fig. 2) similar to that of the other potexviruses for which complete nucleotide sequences have been reported. ORF 1 [nucleotides (nt) 75 to 4044] encodes a protein of M, 149423 (150K). In the C-terminal domain of ORF 1, the motif GDD (position 3660) occurs; this is highly conserved among potexviruses and is a known sequence motif for putative viral RNA polymerases (Argos, 1988; Morozov et al., 1990). Other motifs conserved among RNA-dependent RNA polymerases (for a review see Poch et al., 1989) can be detected in the SMYEAV sequence (data not shown); Brunnen (1991) used these to produce a dendrogram showing the relationships between positive-stranded and double-stranded RNA viruses.} \]

\[ \text{Beginning with the consensus sequence of the putative subgenomic promoter at position 4048, which is common to ORF 2 (encoding proteins of 24K to 26K) of various potexviruses (Fig. 3), an amino acid sequence was deduced to be encoded, and this terminates at nt 4783. Within this sequence, beginning at the residue encoded from position 4134, a span of 30 amino acids shares significant similarity with a conserved region in the potex- and carlaviruses, CYMV, PMV, NMV, PVX, WCIMV, LVX and PV and potato virus M, described by Sit et al. (1990). The codon encoding the first amino acid of this conserved sequence is located 12 to 14 amino acids 3' of the ATG start codon in potexviruses, and in} \]
SMYEAV it is at position 13 relative to an ATT codon. Unlike the other potexviruses, no ATG start codon was observed between the putative subgenomic promoter sequence and the sequence encoding the highly conserved amino acid sequence of SMYEAV. The nucleotide sequence between nt 3820 and nt 4250 was found to be consistent in five independent cDNA clones (p24, p43, p362, p417 and p463), excluding reading errors or the heterogeneity of the cDNA population.

Proteins encoded by potexvirus ORFs 2 to 4 are thought to be functional in cell-to-cell spread of virus (Huisman et al., 1988), but none of these proteins has yet been identified in vivo. Most recently, using infectious transcripts of WC1MV with mutations in ORFs 2 to 4, it has been proposed that this triple gene block constitutes a new class of transport proteins (Beck et al., 1991). The corresponding ORF 2 of SMYEAV either does not exist or is translated by some unknown mechanism, one possibility for which is the use of another initiation codon. Although it has been established that the ATG codon initiates protein synthesis by eukaryotic ribosomes, other codons can be used at low frequency (Kozak, 1989; Clements et al., 1988; Prats et al., 1989; Mehdii et al., 1990). If the ATT codon at position 4096 (Fig. 3) could initiate translation the Mr of the resulting translation is found at position 4078 (PuAACTGPu); the residue. Another start codon that could possibly initiate polypeptide would be 25344 (25K). This codon is in positions -3 and +4 (PuCTATTPu) relative to the A reasonable context for translation, with purines at positions -3 and +4 (PuCTATTPu) relative to the A residue. Another start codon that could possibly initiate translation is found at position 4078 (PuAACTGPu); the protein encoded by this ORF would have an Mr of 26054 (26K).

ORF 3 (nt 4763 to 5087) encodes a protein of Mr 11576 (12K). It shares similarities with ORF 3 of other potexviruses (data not shown). ORF 4 (nt 5016 to 5241) overlaps ORF 3 and encodes a protein of Mr 8079 (8K). ORF 5 (nt 5113 to 5839) was shown to encode the CP of Mr 25714 (26K) or Mr 23810 (24K), depending on the functional ATG codon. The close relationship in Mr and amino acid sequence between potexvirus CPs has been demonstrated by Jelkmann et al. (1990) and Memelink et al. (1990), the latter including three carlaviruses. ORF 6, located within the ORF encoding CP (nt 5048 to 5726), encodes a protein of Mr 11216 (11K).

Although the extent of the similarity in sequence and genome organization among potexviruses is high, there are also distinctive features. LVX lacks the small ORF 5' to the coat protein cistron present in all other potexviruses (Memelink et al., 1990). A small ORF within the coat protein cistron occurs in NMV, WC1MV (Zuidema et al., 1989), SMYEAV and LVX (Memelink et al., 1990), but has not been reported for PVX, PMV and CYMV. Whether they are functional is not known.

The results presented in this paper confirm the conclusion of Jelkmann et al. (1990) that SMYEAV is a member of the potexvirus group. To our knowledge this is the first case in which the sequence of a ssRNA virus has been determined from cDNA clones obtained from a dsRNA template.

References


(Received 12 July 1991; Accepted 29 October 1991)