Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus


CSIRO, Division of Biomolecular Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia and Biochemistry Department, University of Illinois, Urbana, Illinois 61801, U.S.A.

The sequence of the 3'-terminal 1343 nucleotides of the SC strain of the sugarcane mosaic virus (SCMV-SC) genome was compared with the 1376 nucleotides at the 3' terminus of maize dwarf mosaic virus B (MDMV-B). The SCMV-SC sequence includes an open reading frame which codes for the viral coat protein of 313 amino acids (nucleotides 157 to 1116), followed by a 3' non-coding region of 235 nucleotides and a poly(A) tail. The MDMV-B sequence codes for the capsid protein (nucleotides 157 to 1139) of 328 amino acids and has a 3' non-coding region of 236 nucleotides. The coat protein of SCMV-SC has 92% identity with that of MDMV-B except for the region between amino acid residues 27 and 70 of SCMV-SC. This region of SCMV-SC is smaller (44 residues) than the equivalent region in MDMV-B (59 residues) and has only 22% identity with the MDMV-B sequence. Possible mechanisms for the generation of this sequence diversity are discussed. Despite this diversity, the sequence identities of both the major part of the coat proteins and the 3' non-coding regions confirm the proposal, based on previously described serological data, that SCMV-SC and MDMV-B are strains of SCMV.

Introduction

Sugarcane mosaic virus (SCMV), a definitive member of the potyvirus group (Matthews, 1982), infects maize, sorghum, sugarcane and other poaceous species throughout the world (Pirone, 1972; Teakle et al., 1989). Isolates have been designated as strains of either SCMV (Abbott & Tippet, 1966) or maize dwarf mosaic virus (MDMV; Louie & Knoke, 1975), depending on the host from which they were isolated.

On the basis of an apparently distant serological relationship, strains of MDMV were thought to belong to the SCMV group (Shepherd, 1965; Snazelle et al., 1971; Pirone, 1972; Jarjees & Uyemoto, 1984). Recent studies using a set of polyclonal antibodies specific to the amino termini of coat proteins from 11 strains of SCMV and six strains of MDMV have suggested that they belong to four distinct potyviruses (Shukla et al., 1989a). These results confirmed previous studies with four Australian strains of SCMV, i.e. Johnsongrass (SCMV-JG), sugarcane (SCMV-SC), Queensland blue couch grass (SCMV-BC) and Sabi grass (SCMV-Sabi) viruses, which indicated that the JG strain was quite different immunochemically (Shukla & Gough, 1984) and structurally (Gough & Shukla, 1981; Shukla et al., 1987) from the other three strains and should be considered a distinct virus.

Coat protein sequence data have been used to distinguish between viruses and between strains of viruses in the potyvirus group (Shukla & Ward, 1988). In general, coat proteins of distinct members are 38 to 71% identical, and differ markedly in the length and sequence of their amino termini, whereas coat proteins of strains of individual viruses are more than 90% identical and have amino-terminal sequences that are very similar (Shukla & Ward, 1988). We have determined the nucleotide sequences of the coat protein-coding and 3' non-coding regions of SCMV-SC and MDMV-B RNAs and compared them in order to validate the proposal, based on serology (Shukla et al., 1989a) and biological properties (Tosic et al., 1990), that MDMV-B should be considered to be a strain of SCMV.

Methods

Viruses. SCMV-SC was maintained in its perennial host, but was propagated in maize (Zea mays) cv. Iochief for ease of purification. Viral protein was purified as described previously (Gough & Shukla, 1981). MDMV-B was maintained in and propagated from maize cv. Gold Cup. The method of Shukla et al. (1989a) was used to purify the Iowa isolate 66-188 (ATCC PV53) of MDMV-B.

Enzymic digestion of SCMV-SC protein and isolation of peptides. Dried samples (1 to 2 mg) of reduced and S-carboxymethylated SCMV-SC of
coat protein were suspended in 500 to 1000 μl of ammonium bicarbonate (50 mM) by brief sonication, and digested overnight at 37 °C with trypsin (TPCK-treated; Worthington) at an enzyme:substrate ratio of 1:50. Digestion with papain was for 30 min at 37 °C in the above buffer in the presence of 2 mM-DTT. Soluble peptides were fractionated by reverse-phase HPLC as described previously (Shukla et al., 1988a).

**Peptide analysis.** Purified SCMV-SC peptides and intact coat protein were hydrolysed and analysed on an amino acid analysis ion-exchange column (Waters) as described previously (Shukla et al., 1986).

Sequences of tryptic and papain peptides were determined using an Applied Biosystems model 470A protein sequencer. Fractions containing peptide mixtures were rechromatographed as previously described (MeKern et al., 1990) before further analysis.

**Isolation of RNA, and synthesis and screening of cDNA libraries**

(i) SCMV-SC. SCMV-SC RNA was prepared as described previously (Frenkel et al., 1989) and was used to synthesize double-stranded (ds) cDNA essentially according to the method of Gubler & Hoffman (1983) using a kit (Amersham). Following ligation of NotI/EcoRI adaptors (Invitrogen), the ds cDNA was fractionated by passage through a Gen-Pak FAX column (Waters) using a gradient of 50 to 75% 1 M-NaCl in buffer B (25 mM-Tris–HCl pH 7.5, 1 mM-EDTA) and a flow rate of 0.75 ml/min. Fractions containing ds cDNA larger than 1 kb were pooled and ethanol-precipitated in the presence of glycogen.

---

**Fig. 1.** Nucleotide and derived protein sequence of the 3' end of the SCMV-SC genome. The sequence of the 3' 1343 nucleotides of the SCMV-SC genome is shown, together with the derived protein sequence of the open reading frame which extends to the termination codon (*). Amino acids are numbered from the deduced amino terminus of the coat protein, the putative cleavage site being indicated by the vertical arrow. The amino acids identified by sequence analysis are shown by solid lines above the sequence. Peptides placed by amino acid composition are shown by dashed lines. Designation of peptides: T, tryptic; P, papain.
The fractionated DNA was ligated to alkaline phosphatase-treated, EcoRI-digested BlueScript KS + plasmid vector (Stratagene) and transformed into competent *Escherichia coli* DH5α cells according to the suppliers’ instructions (BRL). Single-stranded DNA was prepared from the BlueScript recombinants using the helper phage M13KO7, as described by the supplier, following transformation into *E. coli* DH5αF− cells (BRL).

Plasmid DNA was isolated from overnight cultures by using the boiling method essentially as described by Maniatis et al. (1982). A clone containing an insert of approximately 1-3 kb was digested with EcoRI, and the two fragments derived from the insert (approximately 0-9 kb and 0-4 kb) were ligated to similarly digested M13mpl8. Sequencing from both ends of the original 1-3 kb insert confirmed that it contained a single EcoRI internal site.

(ii) **MDMV-B.** MDMV-B RNA was isolated by the method of Hellman et al. (1980) followed by alcohol precipitation and purification over an oligo(dT) column. ds cDNA was synthesized using the method of Okayama & Berg (1982), digested with EcoRI, and the clone containing an insert of approximately 1.3 kb was digested with HindIII-digested pUC18. Following transformation into *E. coli* MV190 cells, recombinants were screened with radiolabelled oligo GCGUACAUAGAGUACAGAAACUCUACUGAGCGAUAUAUGCCAAGAUACGGACtJUACAGCGCAAUAUCUCAGACUAUAGCUUAGCACGGUAUGCALAJUGAUUUCUAUGAAAUGACLAJCACGC 960

Restriction analysis of DNA from 24 randomly selected clones indicated that all contained inserts larger than 0.8 kb in size (data not shown). A clone containing an insert of approximately 1-3 kb was selected for DNA sequencing studies and the complete RNA sequence is shown in Fig. 1. There is an open reading frame extending for 1095 nucleotides (Fig. 1) which is followed by an untranslated region of 235 nucleotides and a poly(A) tail. Amino acid sequencing of the intact coat protein was used to identify the amino-terminal 20 residues, as well as the sequence of

**Results**

**Sequence of the SCMV-SC coat protein gene**

Restriction analysis of DNA from 24 randomly selected clones indicated that all contained inserts larger than 0.8 kb in size (data not shown). A clone containing an insert of approximately 1-3 kb was selected for DNA sequencing studies and the complete RNA sequence is shown in Fig. 1. There is an open reading frame extending for 1095 nucleotides (Fig. 1) which is followed by an untranslated region of 235 nucleotides and a poly(A) tail. Amino acid sequencing of the intact coat protein was used to identify the amino-terminal 20 residues, as well as the sequence of

**DNA sequencing**

DNA sequencing was carried out using either the T7 Sequenase kit (Pharmacia) or the Taq polymerase kit (Promega) as detailed by the manufacturers. Both DNA strands were sequenced using either universal primers or internal oligonucleotides based on previously determined sequence.

**Fig. 2.** Nucleotide and derived protein sequence of the 3' end of the MDMV-B genome. The sequence of the 3' 1376 nucleotides of the MDMV-B genome are shown, together with the derived protein sequence of the open reading frame which extends to the termination codon (*). Amino acids are numbered from the proposed amino terminus of the coat protein; the putative cleavage site is indicated by the vertical arrow.
a papain peptide corresponding to residues 37 to 51. The amino acid composition of peptides obtained by fragmentation of the coat protein with trypsin or papain enabled the peptide order in the protein to be determined and in every case this was consistent with the sequence derived from the nucleic acid data (Fig. 1).

Sequence of the MDMV-B coat protein gene

Fig. 2 shows the DNA sequence of the 3'-terminal 1376 nucleotides of the MDMV-B genome. There is an open reading frame which is interrupted by a stop codon at position 1141. The 3' non-coding region is 236 nucleotides long and is followed by a poly(A) tail (Fig. 2). The amino terminus of the viral coat protein is blocked (data not shown), however there are data supporting the potential protease cleavage consensus sequences in the 112 amino acid residues towards the N terminus of the SCMV-SC coat protein. This sequence of 44 amino acids shows 86% identity with the equivalent 59 residue sequence recognized by the protease responsible for cleaving the coat protein from the viral polyprotein (Shukla et al., 1991). Whenever the amino-terminal residue in other potyviruses is serine, for example MDMV-B, it is also blocked (Shukla & Ward, 1989b). Based on DNA sequence data, there are no other potential protease cleavage consensus sequences in the 112 amino acid residues towards the N terminus of the site proposed (data not shown). In addition, amino acid analysis of the whole MDMV-B coat protein (data not shown) was in excellent agreement with the composition determined from the derived cDNA sequence (Fig. 2).

Comparison of the SCMV-SC and MDMV-B coat protein sequences

An alignment of the coat protein and immediate upstream sequences of SCMV-SC and MDMV-B is shown in Fig. 3(a). Comparison of the coat protein sequences shows a difference in size (313 and 328 amino acids for SCMV-SC and MDMV-B, respectively). However, although the final three-quarters of the sequences are 94% identical (residues 71 to 313), the amino-terminal regions are less similar, particularly the region corresponding to residues 27 to 70 of the SCMV-SC coat protein. This sequence of 44 amino acids shows only 22% identity with the equivalent 59 residue sequence of MDMV-B coat protein.

Comparison of the SCMV-SC and MDMV-B 3' non-coding sequences

The 3' non-coding portion of the SCMV-SC genome (235 nucleotides) contains one less nucleotide than, and is 86% identical to that of MDMV-B (Fig. 4).
Discussion

In the classification scheme based on serological analyses using virus-specific antibodies (Shukla et al., 1989a), SCMV-SC and MDMV-B were considered to be strains of the same virus. The proposed classification is supported by the assignments of the SCMV and MDMV strains based on the reactions to infection of the different sorghum lines and oat cultivars (Tosic et al., 1990). The relationship between the two viruses has now been examined further by determining the gene sequences encoding the SCMV-SC coat protein and the 3' non-coding region, and comparing these with the corresponding sequences of MDMV-B. Previous investigations have established the value of coat protein sequence comparisons (Shukla & Ward, 1988, 1989a) and 3' non-coding nucleic acid sequence data (Frenkel et al., 1989) in establishing taxonomic relationships among potyvirus members.

The sequences of the SCMV-SC and MDMV-B coat proteins are very similar (92% identity; Fig. 3a) except for the region beginning 27 residues from the amino terminus and spanning 44 residues in SCMV-SC and 59 residues in MDMV-B. Although the regions share a high glycine and relatively high threonine and alanine content, their sequences differ markedly. Possible mechanisms for generating such sequence diversity within otherwise similar genes include deletion, frameshift mutation, gene duplication and recombination. Examination of the nucleic acid sequence data reveals that the sequence diversity is not the result of a simple deletion or a frameshift mutation. Although there is evidence of some gene duplication in this region in both SCMV-SC and MDMV-B (most readily seen by examining the amino acid sequence; Fig. 3b), the duplicated sequence in SCMV-SC is different from that duplicated in MDMV-B, and there is little sequence identity between the common regions. As shown in Fig. 3(b) the sequence coding for the decapeptide Q18 to Q27 is duplicated in SCMV-SC, while in MDMV-B there is a partial duplication of the 16 residue sequence G38 to Q53 in which 11 amino acids are identical, followed by a smaller partial duplication of only six residues. Another possible explanation for the sequence diversity is recombination, but searches of protein and nucleic acid databases with both ‘divergent’ sequences failed to produce any significant identities with other potyviral coat protein or genomic sequences. The origin of this diversity is unknown, but sequence analysis of the corresponding regions of other isolates assigned by serology (Shukla et al., 1989a) to the SCMV group (SCMV-A, -B, -D, -E, -BC and -Sabi) will establish whether one of the two sequence types predominates.

The overall sequence identity of 77% for the coat proteins of SCMV-SC and MDMV-B would appear to invalidate the proposal (Shukla & Ward, 1988) that coat protein sequence data can be used to discriminate between distinct viruses and strains, as it lies between the ranges found for distinct viruses (38 to 71%) and related strains (90 to 99%). However, if the divergent sequences are omitted from the analysis, the overall identity between the remaining sequences of the coat proteins is 92%, suggesting that SCMV-SC and MDMV-B are strains of the same potyvirus. Additional support for this proposal comes from close examination of the coat protein core sequences, which reveals that both proteins have identical changes at residues previously found to be invariant in coat proteins of several potyviruses (Shukla & Ward, 1989a, b). In both coat proteins, residue 132 (numbering as in Shukla & Ward, 1989a) is phenylalanine instead of the histidine found in other potyvirus coat protein sequences. Likewise, residue 160 is isoleucine instead of valine, 187 is cysteine instead of threonine and 277 is lysine instead of arginine.

Further evidence that the two viruses are closely related comes from a comparison of the SCMV-SC and MDMV-B 3' non-coding regions (Fig. 4), which shows that these sequences are 86% identical. Frenkel et al. (1989) have shown that when the 3' non-coding sequences of potyviruses are compared, strains of the same virus are 80% or more identical, whereas those of distinct viruses are less than 50% identical.

Shukla & Ward (1989b) have drawn attention to three types of genetic event (internal deletions, cleavage site mutations and frameshift mutations) which could affect the amino-terminal region of potyviral coat proteins. Such effects might well complicate the use of sequence identity scores or amino terminus-targeted serology in potyvirus identification and detection. Examples of internal deletions have now been described for strains of plum pox virus (Maiss et al., 1989) and watermelon mosaic virus (Yu et al., 1989), but to date there are no reported examples of the other types of potential variations, i.e. cleavage site mutations or frameshift mutations.

In this paper we have described a fourth type of genetic variation in which approximately two-thirds of the immunodominant, surface-exposed region (Shukla et al., 1988b) of the coat protein has been altered. Despite this large change, SCMV-SC and MDMV-B retain strong, virus-specific, serological cross-reactivity (Shukla, 1989a) and induce similar symptoms in selected sorghum lines (Tosic et al., 1990). It will be of interest to see whether this change has any effect on other biological properties of these two strains of SCMV. It is known that SCMV-SC infects sugarcane; MDMV-B does not (Teakle et al., 1989). Perhaps other SCMV strains which do not infect sugarcane, such as SCMV-BC and SCMV-
Sabi (Teakle et al., 1989), resemble MDMV-B in their coat protein sequences. It remains to be seen whether the two SCMV strains, SCMV-SC and MDMV-B, also differ in their cross-protection behaviour and vector specificity.

We thank Nick Bartone for the oligonucleotide syntheses and amino acid analyses performed in relation to the SCMV-SC studies, and Dr R.E. Ford for assistance and advice during the course of the MDMV-B studies.

References


(Received 26 October 1990; Accepted 30 October 1990)