The complete nucleotide sequence and genome organization of the
mite-transmitted brome streak mosaic rymovirus in comparison
with those of potyviruses

R. Götz* and E. Maiss

Biologische Bundesanstalt für Land- und Forstwirtschaft, Institut für Biochemie und Pflanzenvirologie,
Messeweg 11–12, 38104 Braunschweig, Germany

A virus isolate, designated as 11-Cal, originating from
southern France has been identified as an isolate of the
mite-transmitted brome streak mosaic rymovirus (BrSMV) by serological and morphological properties.
BrSMV is a member of the genus *Rymovirus* of the family *Potyviridae*. The complete nucleotide sequence of
the RNA genome of BrSMV has been determined. The
assembled RNA is 9672 nucleotides in length, excluding
a 3'-terminal poly(A) sequence. The RNA contains one
open reading frame (ORF) of 9282 nucleotides coding
for a polyprotein of 3093 amino acids. A comparison
with typical potyviruses showed that BrSMV has a
similar genome organization. The predicted cleavage
sites of the polyprotein of BrSMV are similar to those of
potyviruses. Nevertheless, unusual dipeptides are pro-
posed in two cases. Based on the proposed location of
the cleavage sites nine mature proteins are predicted.
Specific motifs, described for potyvirial polyproteins, are
almost all present in the polyprotein of BrSMV, too.
However, only an incomplete zinc-finger motif is present
in the potential helper component and the motif for
aphid transmission in the coat protein is not found.
Several alignments of amino acid sequences showed less
similarity between BrSMV and potyviruses than between
different potyviruses.

Introduction

The genome of potyviruses is a positive-sense ssRNA of
approximately 10 kb, linked covalently at its 5' end to a
virus-encoded protein (VPg) and polyadenylated at its 3'
end (Dougherty & Carrington, 1988). The potyviral
RNA codes for a polyprotein which is proteolytically
processed into nine mature proteins by three virus-
encoded proteinases, the N-terminal protein (P1;
Verchot et al., 1991), the helper component-proteinase
(HC-Pro; Carrington et al., 1989) and the nuclear
inclusion a protein (Nia-Pro; Carrington & Dougherty,
1987a, b; Chang et al., 1988; Hellmann et al., 1988;
Garcia et al., 1989; Dougherty et al., 1989a; Ghabrial
et al., 1990).

The genomes of several potyviruses have been
sequenced completely: tobacco etch virus (TEV; Allison
et al., 1986), plum pox virus (PPV; Maiss et al., 1989;
Lain et al., 1989), potato virus Y (PYY; Robaglia et al.,
1989), zucchini yellow mosaic virus (ZYMV; Balint,
1990), pea seed-borne mosaic virus (PSbMV; Johansen
et al., 1991), papaya ringspot virus (PRSV; Yeh et al.,
1992), soybean mosaic virus (SMV; Jayaram et al.,
1992), turnip mosaic virus (TuMV; Nicolas & Laliberté,
1992), pepper mottle virus (PepMoV; Vance et al., 1992),
potato virus A (PVA; Puurand et al., 1994) and peanut
stripe virus (PStV; Gunasinghe et al., 1994). All these
viruses are transmitted by aphids and have been assigned
to the genus *Potyvirus*.

Furthermore, the complete sequence data of barley
yellow mosaic virus (BaYMV), a member of the genus
*Bymovirus* in the *Potyviridae* have been reported. In
contrast to potyviruses, this virus is not only transmitted
by fungi instead of aphids, but also has a bipartite
genome. Both RNA 1 (Kashiwazaki et al., 1990; Peerenboom et al., 1992) and RNA 2 (Davidson et al.,
1991; Kashiwazaki et al., 1991) have been sequenced.

A filamentous virus, designated 11-Cal, was isolated
from wheat plants in southern France. Based upon
particle morphology and transmission properties it
appeared to be a potyvirus that infects barley, wheat
and several other grasses (W. Huth and others, unpublished).

* Author for correspondence. Fax +49 511 762 3015. e-mail
maiss@mbox.ipp.uni-hannover.de

The nucleotide sequence data reported in this paper have been
submitted to EMBL and assigned the accession number Z48506.
Since 11-Cal could be transmitted by the mite *Aceria tulipae* but not by aphids and reacted strongly in serological tests with an antiserum to brome streak virus (BrSMV), it has been regarded as an isolate of BrSMV and assigned to the genus *Rymovirus* of the *Potyviridae*. BrSMV was first isolated from *Bromus mollis* in Yugoslavia by Milicic *et al.* (1980, 1982).

Of the mite-transmitted viruses of the *Potyviridae*, which all have ssRNA genomes, only the coat protein sequence of wheat streak mosaic virus (WSMV; Niblett *et al.*, 1991) and brome streak mosaic virus (BrSMV; W. Huth and others, unpublished) has been described. We have determined the nucleotide sequence of the RNA of the mite-transmitted BrSMV in order to investigate the genomic differences between a member of the genus *Rymovirus* and viruses belonging to the genus *Potyvirus*. To our knowledge this is the first report of a complete nucleotide sequence of a member of the genus *Rymovirus*.

**Methods**

*Virus preparation.* The isolate 11-Cal (BrSMV) was propagated in wheat by mechanical inoculation and purified by a method devised by W. Huth and others (to be published subsequently).

*RNA isolation.* RNA was extracted from virions by incubation for 45 min in 2 x proteinase K buffer and proteinase K (Maiss *et al.*, 1988).

*cDNA synthesis and cloning.* The viral RNA was used as template for oligo(dT) as well as random primed cDNA synthesis using the Riboclon cDNA synthesis system (Promega). RNA was reverse-transcribed using AMV reverse transcriptase, and the dsDNA was made blunt-ended by T4 DNA polymerase. Subsequently, the DNA was ligated to the *HincII*-cut phagemid vector pT7T3 (19U; Pharmacia) and cloned into *E. coli* NM522 (Pharmacia).

*Nucleotide sequencing.* Sequencing of cDNA clones was done by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). For sequencing, ssDNA templates were used which were prepared with the helper phage M13K07 (Pharmacia). Each nucleotide was determined either by sequencing at least two independent cDNA clones or by sequencing both strands of a cDNA clone.

The extreme 5′ end of the RNA was not present in the cDNA clones. Hence the 5′ Amplifinder Race kit (Clontech) was used to get the 5′ end by using a specific primer and the Amplifinder anchor primer in a polymerase chain reaction.

*Sequence analysis.* Sequence data were assembled and analysed using the GCG program package (Devereux *et al.*, 1984). Nucleotide and protein sequences of BrSMV were analysed by DNA/SIS. For comparison of protein sequences of different viruses ALIGN was used, and the phylogenetic tree was constructed using CLUSTAL with standard parameters.

**Results and Discussion**

*Sequencing*

Screening of the cDNA clones identified a sample of overlapping clones which were estimated to cover almost the entire genome of BrSMV. Eleven independent cDNA clones were selected and sequenced completely or partially to establish the nucleotide sequence of BrSMV (Fig. 1). Four overlapping clones (83, 126, 3 and 9) contained nearly the complete RNA except for 11 nucleotides at the extreme 5′ end of the RNA. There was little nucleotide exchange (six nucleotides) over the complete RNA except for the region where clones 108 and 56 overlapped clones 126 and 3. In this region a nucleotide divergence of 6.7% was obtained. The replacements were almost all in the third codon position and led only in four cases to an amino acid exchange (0.7%). This observation indicated that cDNA clones 108 and 56 originated from a heterogeneous population of viral RNA.

**Genome organization**

The complete nucleotide sequence of BrSMV consists of 9672 nucleotides excluding the 3′-terminal poly(A) tail. The RNA of BrSMV is longer than the RNAs of TEV and TMV but shorter than those of PPV and PSbMV, for example. The RNA has a base composition of 29.7% adenine, 21.0% cytosine, 25.6% guanine and 23.7% uracil. The composition is typical for the potyviruses. The 5′-noncoding region is 145 nucleotides long. This is an average length in comparison with potyviruses.

Translation of the RNA revealed one large open reading frame (ORF). The first possible initiation codon, AUG, at nucleotide positions 146 to 148 is likely to be the start codon of the BrSMV polyprotein. Although there is a second AUG codon at positions 215 to 218, only the first AUG codon is rather similar to the consensus sequence A-A-C-A-AUG-G-C proposed for translation initiation in plants (Lütcke *et al.*, 1987). The termination codon UAG is located at positions 9425 to 9427. Therefore, the ORF consists of 9282 nucleotides and potentially codes for a polyprotein of 3093 amino acids. The ORF is followed by a noncoding region of 245 nucleotides, excluding the poly(A) tail. The length of the 3′-noncoding region is similar to those of the potyviruses.

A tentative map of the polyprotein of BrSMV is shown in Fig. 2, including the predicted cleavage sites and functional proteins. Based on the proposed location of cleavage sites and on alignments of amino acid sequences for each protein, nine mature proteins are predicted for BrSMV (see Fig. 2). Three proteinases are involved in the complete processing of the potyviral polyprotein. The NIa protease, located in the C-terminal part of NIa, has been shown to cleave the potyviral polyprotein at several sites which consist of a conserved sequence of amino acids and hence serve as a NIa protease recognition signal (Dougherty *et al.*, 1988; Robaglia *et al.*, 1989). A block of seven amino acids is sufficient to define the cleavage site recognized by NIa-Pro. A comparison of
putative cleavage sites in the potyvirus polyprotein has shown that NiA-Pro cleaves at Q/A, Q/G, Q/S, Q/T or Q/V dipeptide sequences, and valine was found in most potyviruses at position P4 of the cleavage site (Dougherty & Carrington, 1988; Ghabrial et al., 1990; Dinant et al., 1991). On the other hand, there are several examples of a putative cleavage site for NiA-Pro that differ from the usual consensus sequence of potyviruses. In these cases, position P1 is occupied by glutamic acid instead of glutamine (Johansen et al., 1991; Nicolas & Laliberté, 1992; Yeh et al., 1992). In some other cases, there is an exchange at position P'1 as described by Robaglia et al. (1989) and Vance et al. (1992). Beside cleavage of the functional proteins in the C-terminal half of the potyviral polyprotein, an internal cleavage site delimiting the VPG and proteinase domains was identified by Dougherty & Parks (1991). For this cleavage site, the dipeptides E/S, E/G, E/A and the consensus sequence (V/L)-X-X-E/(S/G/A) were described. Based on these findings the polyprotein of BrSMV was analysed and seven cleavage

Table 1. Putative cleavage sites of the NiA protease in the genome of BrSMV and the amino acid sequence adjacent to the seven proposed sites

<table>
<thead>
<tr>
<th></th>
<th>P6</th>
<th>P5</th>
<th>P4</th>
<th>P3</th>
<th>P2</th>
<th>P1</th>
<th>P'1</th>
<th>P'2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P3/6K1</td>
<td>E</td>
<td>V</td>
<td>V</td>
<td>F</td>
<td>Q</td>
<td>S</td>
<td>K</td>
<td></td>
</tr>
<tr>
<td>6K1/CI</td>
<td>V</td>
<td>G</td>
<td>S</td>
<td>I</td>
<td>Y</td>
<td>Q</td>
<td>S</td>
<td>G</td>
</tr>
<tr>
<td>CI/6K2</td>
<td>A</td>
<td>H</td>
<td>V</td>
<td>M</td>
<td>Y</td>
<td>Q</td>
<td>K</td>
<td>G</td>
</tr>
<tr>
<td>6K2/NiA-VPG</td>
<td>H</td>
<td>E</td>
<td>A</td>
<td>K</td>
<td>F</td>
<td>E</td>
<td>G</td>
<td>K</td>
</tr>
<tr>
<td>NiA-VPG/NiA-Pro</td>
<td>P</td>
<td>Y</td>
<td>A</td>
<td>V</td>
<td>F</td>
<td>E</td>
<td>S</td>
<td>K</td>
</tr>
<tr>
<td>NiA-Pro/NiB</td>
<td>K</td>
<td>L</td>
<td>V</td>
<td>G</td>
<td>F</td>
<td>Q</td>
<td>N</td>
<td>G</td>
</tr>
<tr>
<td>NiB/CP</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td>T</td>
<td>P</td>
<td>Q</td>
<td>S</td>
<td>G</td>
</tr>
<tr>
<td>Consensus</td>
<td>X</td>
<td>X</td>
<td>V/A/S</td>
<td>X</td>
<td>X</td>
<td>Q/E</td>
<td>S/K/N/G</td>
<td>X</td>
</tr>
</tbody>
</table>

sites are predicted (Table 1). The sequences of some putative cleavage sites agree with the NiA-Pro consensus cleavage site sequence of potyviruses, described by Riechmann et al. (1992). Nevertheless, unusual dipeptides are proposed between CI and 6K2 and between NiA-Pro and NiB. These cleavage sites contain glutamine at P1, but atypically contain lysine or asparagine instead
of serine at position P1. However, arginine instead of serine at the junction between 6K1 and CI (Robaglia et al., 1989; Vance et al., 1992) is also not common. The proposed dipeptide E/G between 6K2 and NIa-VPg is unusual in containing glutamic acid instead of glutamine but is similar to the cleavage site in the polyprotein of PSbMV (Johansen et al., 1991) and TuMV (Nicolas & Laliberté, 1992). Furthermore, the NIa-VPg/NIa-Pro cleavage site is similar to that of TuMV (Nicolas & Laliberté, 1992) and SMV (Jayaram et al., 1992). The fact that valine at position P4 of the cleavage sites was sometimes exchanged for alanine or serine is unusual, but replacements in this position were also found in the polyprotein of PVY (Robaglia et al., 1989), for example. The consensus sequence of NIa-Pro of BrSMV, (V/S/A)-X-X-(Q/E)/(S/G/K/N), is based on five different dipeptides (see Table 1). Five different dipeptide cleavage sites for the NIa protease have also been observed for the potyviruses PRSV (Yeh et al., 1992), TuMV (Nicolas & Laliberté, 1992) and PepMoV (Vance et al., 1992).

In addition to NIa-Pro, HC-Pro and P1 are also potyviral proteinases. Carrington et al. (1989) identified a potential cleavage site for HC-Pro of TEV between the dipeptide G/G, and the consensus HC-Pro cleavage sequence is K-X-Y-X-V-G/G (Carrington & Herndon, 1992). A similar sequence exists in the polyprotein of BrSMV at positions 804 to 810. At this location the dipeptide G/G is present, as well as a similar consensus sequence, K-E-Y-E-I-G/G, the only replacement being isoleucine instead of valine.

The P1 cleavage site has been identified by Mavankal & Rhoads (1991). They showed that the dipeptides Y/S or F/S function as the potential cleavage site and that I-V-H-F/S may be the sequence at this cleavage site. The dipeptide Y/S is found at positions 403 to 404 in the BrSMV polyprotein, although the surrounding sequence is not similar to the sequence described by Mavankal & Rhoads (1991).

Comparison of the RNA genome of BrSMV with those of potyviruses

Comparison of noncoding regions. In comparison with the base composition of the complete RNA genome of BrSMV the 5'-noncoding region contains only few guanine residues. This low content of guanine in the 5' leader sequence is typical of potyviruses and other plant viruses (Gallie et al., 1987). An alignment of the extreme 5'-terminal sequence has revealed a high content of adenine and a highly conserved block of 13 nucleotides among potyviruses, starting at position 13 or 14 (Maiss et al., 1989; Riechmann et al., 1989). The corresponding region of the BrSMV genome is also rich in adenine, but in the block of 13 nucleotides only eight are conserved.

Comparison of proposed mature proteins. The conserved amino acids histidine and cysteine, representing the active site of the protease in the C-terminal half of the protein, have been identified (Verchot et al., 1991), and this motif is present in the P1 of all aphid-transmitted potyviruses sequenced so far. The same amino acids are located in the P1 of BrSMV at amino acid positions 311 and 355, respectively. Furthermore, the distance between these two locations is 43 amino acids. This distance is not strictly conserved in the P1 of aphid-transmitted potyviruses.

Oh & Carrington (1989) described a conserved motif G-Y-C-Y-72X-H for HC-Pro. The amino acids cysteine, surrounded by two tyrosine residues, and histidine are essential for the active site of HC-Pro of TEV. The amino acid sequence of BrSMV includes the same motif at the active site of the putative HC-Pro, with cysteine at position 698 and histidine at position 769. However, the distance between C and H is not 72 but 70 amino acids.

In the putative N-terminal region of the HC-Pro of PVY Robaglia et al. (1989) have identified a cysteine cluster which is considered to be similar to the zinc-finger of several nucleic acid-binding proteins (Berg, 1986). This cluster of five cysteines, C-8X-C-13X-C(SMV:V)-4X-C-2X-C, in the HC-Pro of potyviruses is perfectly conserved in the sequence of aphid-transmitted potyviruses. In the polyprotein of BrSMV only a reduced cluster of V-4X-C-2X-C is found, located at positions 446 to 454.

Additionally, a glycosylation motif in HC-Pro has been described for potyviruses (Berger & Pirone, 1986). In the putative HC-Pro of BrSMV two possible glycosylation motifs are present. The motif N-S-T is located at positions 570 to 572, and a second possible amino acid triad, N-S-T, is present at positions 604 to 606.

Protein 3 (P3) seems to have a highly conserved motif as described by Riechmann et al. (1992). This motif, present in all aphid-transmitted potyviruses sequenced, was not observed in the polyprotein of BrSMV.

A conserved motif representing the nucleotide-binding site has been identified in the CIs of TEV and TVMV (Gorbalenya et al., 1989). This motif, G-2X-G-1X-GSK, is not only present in the CIs of all sequenced potyviruses, but is also perfectly conserved in the polyprotein of BrSMV at positions 1228 to 1235.

A conserved triad of the amino acids histidine, aspartic acid and cysteine represents the active site of the NIa protease (Dougherty et al., 1989b). The complete motif, including location and distance between these three amino acids, is strictly conserved in all of the sequenced NIa proteases of aphid-transmitted potyviruses. The amino acid sequence of BrSMV includes the three amino acids representing the active site of the NIa-Pro, but the
Distance is slightly different from that of the potyviruses: H-38X-D-71X-C (BrSMV) compared with H-34X-D-69X-C (potyviruses). The amino acid triad is located at positions 2082, 2121 and 2193 close to the C terminus of the NIa protein.

The N terminus of the NIa protein has been reported to be the viral VPg. For the linkage of the VPg to the 5' end of the viral RNA a tyrosine residue in the N terminal part of the NIa is necessary (Murphy et al., 1991). This tyrosine residue is present in a similar surrounding sequence not only in the NIa of all potyviruses but in the NIa of BrSMV at amino acid position 1915.

The NIb proteins of potyviruses show a high degree of identity indicating that this protein is the most conserved potyvirus protein. The consensus motif G-D-D, important for the putative RNA polymerase function of potyviruses, is present in all potyviruses sequenced. The motif also exists in the polyprotein of BrSMV at amino acid positions 2624 to 2626. Furthermore, the 'GDD' box is part of the conserved motif S-G-3X-T-3X-N-T-(18-37)-X-G-D-D, which is proposed to be the active site of the RNA-dependent RNA polymerase (Kamer & Agros, 1984). The first block of 11 amino acids is located in the sequence of BrSMV at amino acid positions 2584 to 2594 and is highly conserved in comparison with those of potyviruses.

The coat protein (CP) is responsible for the encapsidation of the viral RNA, and is involved in the aphid transmission of the potyviruses. The motif D-A-G in the N-terminal part of the CP is required for aphid transmissibility and is found in almost all of the potyviral coat proteins. A motif similar to D-A-G is not found in the N-terminal part of the CP of BrSMV. This is not surprising, because BrSMV is transmitted by mites but...
not by aphids, as mentioned above. A conserved motif in the coat protein of potyviruses includes a ‘salt bridge’. This motif is composed of the amino acids arginine and glutamine in the core region of the CP and at a distance of 45 amino acids, aspartic acid at the C terminus. The ‘salt bridge’ is probably necessary for the formation of the virus particle (Jagadish et al., 1991). This motif is also present in the polyprotein of BrSMV but the amino acid glutamine is exchanged for alanine. Furthermore, the distance is changed compared to the potyviruses. Both properties, replacement of an amino acid (A instead of Q) and a shorter distance, are also found in the polyprotein of WSMV, the other mite-transmitted potyvirus (Niblett et al., 1991).

**Degree of identity between the mature proteins of BrSMV and other Potyviridae**

An alignment of the polyproteins of several potyviruses, for example PPV, PVY, TEV and TVMV, yields about 50% identity. In contrast, only about 30% identity was obtained for the amino acid sequence of BrSMV and each of the four potyviruses. Furthermore, the amino acid sequences of the nine mature proteins proposed for BrSMV were compared with the corresponding proteins of several potyviruses. The amino acid sequence of PPV was used to compare the degree of identity between potyviruses and BrSMV. The known mature proteins of WSMV and BaYMV were also included in this comparison. The degree of identity to the proteins of PPV is shown in Fig. 3. The most conserved mature proteins are N1b and CI, whereas P1 and P3 show little identity. However, there is a high level of identity when PPV is compared with other potyviruses. In contrast, the degree of identity between proteins of BrSMV and those of PPV is lower than that between different potyviruses (see Fig. 3). This applies also for the highly conserved proteins N1b and CI as well as for the less conserved proteins HC-Pro and P3. Only the sequence of protein P1 shows no significant identity between BrSMV and potyviruses or among potyviruses.

Kashiwazaki et al. (1990) have reported that RNA 1 of BaYMV contains four regions similar to the mature proteins CP, N1b, N1a and CI. These proteins of BaYMV show less identity to the corresponding proteins of PPV than those of BrSMV (see Fig. 3). This leads to the conclusion that rymoviruses are less different from potyviruses than bymoviruses.

The degree of identity between their coat proteins indicates the relationship of potyviruses according to the classification described by Shukla & Ward (1988, 1989). An alignment of the coat proteins of BrSMV and WSMV gives an identity of about 50%. This suggests a close relationship between these two viruses, which are both...
transmitted by mites. The identity between BrSMV and the various aphid-transmissible potyviruses is considerably lower, about 25–30%, similar to that between WSMV and aphid-transmitted potyviruses (Niblett et al., 1991). The phylogenetic tree (Fig. 4) shows the relationship of potyviruses, rymoviruses and bymoviruses based on their coat protein amino acid sequence. The eight potyviruses have different degrees of relationship but they are more closely related to each other than to BrSMV, WSMV or BaYMV. BrSMV is more similar to WSMV than to any other potyvirus. Furthermore, the two rymoviruses, BrSMV and WSMV, are more closely related to the potyviruses than to the bymovirus BaYMV.

We thank Petra Bauer for excellent technical assistance and H. J. Vetten for critical reading of the manuscript. R. G. was awarded a grant by the Deutsche Forschungsgemeinschaft (no. Go 601/1-1).

Note added in proof. The 3′-terminal sequence of a rymovirus, termed the Hordeum isolate of BrSMV, has been published recently by J. Schubert & F. Rabenstein (European Journal of Plant Pathology 101, 123–132, 1995). The coat protein sequence of this isolate is not in accordance with that of the 11-Cal isolate of BrSMV.

References


Maiss, E., Tempe, U., Brisske, A., Jelkmann, W., Casper, R.,


(Received 3 January 1995; Accepted 4 April 1995)