Volume 14 Number 9 1986 Nucleic Acids Research

The complete nucleotide sequence of RNAβ from the type strain of barley stripe mosaic virus

Gary Gustafson and Susan L.Armour

Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, IN 46285, USA

Received 24 December 1985; Revised 28 March 1986; Accepted 1 April 1986

Abstract

The complete nucleotide sequence of RNAβ from the type strain of barley stripe mosaic virus (BSMV) has been determined. The sequence is 3289 nucleotides in length and contains four open reading frames (ORFs) which code for proteins of M_r 22,147 (ORF1), M_r 58,098 (ORF2), M_r 17,378 (ORF3), and M_r 14,119 (ORF4). The predicted N-terminal amino acid sequence of the polypeptide encoded by the ORF nearest the 5'-end of the RNA (ORF1) is identical (after the initiator methionine) to the published N-terminal amino acid sequence of BSMV coat protein for 29 of the first 30 amino acids. ORF2 occupies the central portion of the coding region of RNAβ and ORF3 is located at the 3'-end. The ORF4 sequence overlaps the 3'-region of ORF2 and the 5'-region of ORF3 and differs in codon usage from the other three RNAβ ORFs. The coding region of RNAβ is followed by a poly(A) tract and a 238 nucleotide tRNA-like structure which are common to all three BSMV genomic RNAs.

INTRODUCTION

Barley stripe mosaic virus (BSMV) is a member of a small group of plant virus pathogens known as the hordeiviruses (1). This rod-shaped virus primarily infects members of the Gramineae; however, several dicot species can also act as hosts (2). The genome of BSMV consists of three separately encapsidated, single-stranded RNAs designated α, β, and γ in order of decreasing molecular weight (3). The genomic RNAs have a 7-methylguanosine cap at the 5'-end (4) and a tRNA-like structure at the 3'-end (5) which can be aminoacylated with tyrosine (6,7). A poly(A) sequence of variable length is located between the coding region and the tRNA-like structure in each genomic RNA (8,9).

In other tripartite viruses which have been sequenced, the coding capacity of the genomic RNAs is apparently limited to four polypeptides. The two largest genomic RNAs of brome mosaic virus, cucumber mosaic virus and alfalfa mosaic virus each encode a single high molecular weight polypeptide (10,11,12,13,14) while the smallest genomic RNA codes for two polypeptides (15,16,17); one of which (coat protein) is expressed only from

a subgenomic RNA (18,19,20). In vitro translation and hybridization studies (21,22,23) suggest that BSMV differs from other tripartite viruses in at least two ways. First, although the smallest genomic RNA (RNAγ) codes for two polypeptides (including one which is expressed only from a subgenomic RNA), it does not code for BSMV coat protein. Instead, coat protein is translated in vitro from the second largest genomic RNA (RNAβ). Second, because the coat protein gene occupies only about 25% of the theoretical coding capacity of RNAβ, it is likely that RNAβ codes for at least one additional polypeptide. If so, then the number of polypeptides encoded by the BSMV genome would be greater than that of other tripartite viruses which have been studied.

In order to elucidate the exact nature of the genetic organization of BSMV RNAβ, we have now determined its complete nucleotide sequence. The analysis of the sequence confirms that BSMV is a unique tripartite virus.

MATERIALS AND METHODS

Virus Isolation and RNA Purification

The Type (ATCC-PV43) strain of BSMV was maintained on and isolated from "Black Hulless" barley (C.1.666) grown in controlled climate chambers (21). Viral RNA was isolated from purified virus as previously described (21).

Synthesis and Cloning of Double-Stranded cDNA

Two independent sets of cDNA clones were prepared from unfractionated BSMV RNA. The first set of clones was prepared as previously described (22) using oligo(dT) to prime the synthesis of the first strand of cDNA from the internal poly(A) sequence present in all three BSMV genomic RNAs. The second set was generated by a different method (24) and used a synthetic oligonucleotide (5'-TGGTCTTCCCTTGGG) complementary to the sequence at the extreme 3'-end of all three BSMV genomic RNAs (5,25) to prime first-strand synthesis. Double-stranded cDNAs were (dC)-tailed with terminal deoxynucleotidyl transferase (BRL) and cloned into Pst 1 digested/(dG)-tailed pBR322 as described (22).

Isolation of BSMV RNA cDNA Clones

cDNA clones were screened to determine insert size by digestion with Pst 1. Those with the largest inserts were nick translated and hybridized with BSMV RNAs which had been separated by electrophoresis on agarose/formaldehyde gels (26) and transferred to nitrocellulose (27). Inserts from clones which hybridized with RNAβ were mapped with restriction enzymes and three clones (Fig. 1) containing the entire coding and

3'-noncoding regions of RNAβ were chosen for sequencing.

Subcloning and Sequencing

The 1688 bp and 1532 bp Pst 1 fragments from clones pBSM41 and pBSM191, respectively, and the 730 bp Sal 1/Eco RI fragment from pBSM41 (Fig. 1) were isolated on 3.5% polyacrylamide gels and digested with Sau 3a or Taq 1. The fragments generated were ligated into M13mp19 which had been restricted with Bam HI or Acc I and treated with calf intestinal phosphatase (Boehringer Mannheim). In addition to the random Sau 3a and Taq 1 fragments, a number of specific restriction fragments from pBSM23, pBSM41 and pBSM191 were isolated from acylamide gels and subcloned into an appropriate M13 bacteriophage vector. BSMV cDNA fragments subcloned in M13 were sequenced by the dideoxynucleotide chain termination method (28). In some cases sequencing reactions were labeled with [α-^{35}S]dATP (NEN) and electrophoresed on 6% buffer gradient gels (29). One DNA fragment from pBSM41 was also sequenced by the chemical degradation method (30).

The sequence at the extreme 5'-end of RNAβ (which was not contained within any of our cDNA clones) was determined with the aid of an oligonucleotide primer (β1). This oligonucleotide (5'-GTTCGGCATACTGTGAAGCT) is complementary to a region of RNAβ that is less than 100 nucleotides from its 5'-end. The oligonucleotide was end-labeled, annealed with total BSMV RNA, and extended with reverse transcriptase. The cDNA synthesized was isolated and sequenced by the chemical degradation method (30).

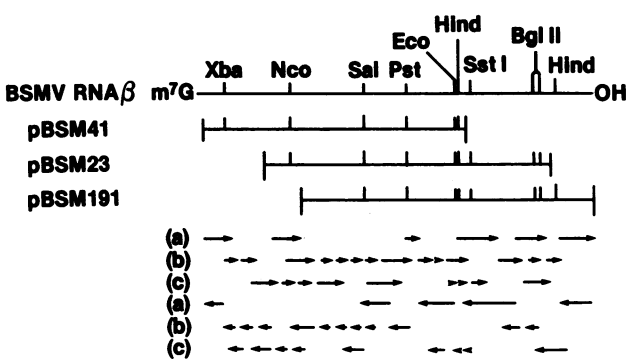

Figure 1. Restriction map and sequencing strategy for BSMV RNAβ. The majority of the sequence was determined by dideoxy sequencing of Sau 3a (b) and Taq 1 (c) restriction fragments generated from the cloned RNAβ sequences in pBSM41 and pBSM191. Specific restriction fragments (a) were used to complete the sequence. Clone pBSM23 was used to confirm sequences in critical areas. Arrows indicate the location, direction and length of each sequence.

RESULTS

Construction of the Sequence

The alignment of cDNA clones pBSM41, pBSM23 and pBSM191 with RNAβ and the strategy employed to determine the sequence of the RNA from those clones are shown in Figure 1. Greater than 98% of the portion of the RNA represented in the cDNA clones was sequenced in both strands. The cloned sequence includes the coding region (2939 nucleotides), the 3'-noncoding region (261 nucleotides), and 22 nucleotides of the 5'-noncoding region of BSMV RNAβ.

The sequence of the remaining portion of the 5'-noncoding region was determined with the aid of an oligonucleotide primer (β1) complementary to a

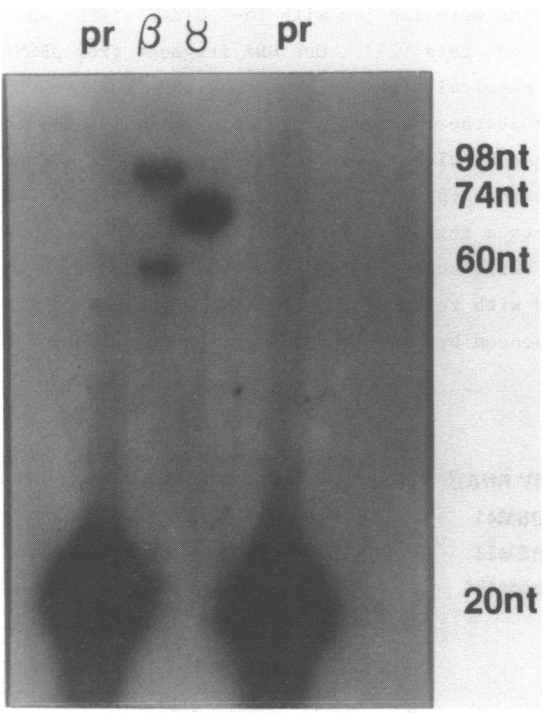

Figure 2. cDNA fragments synthesized by reverse transcription of total RNA from the Type strain of BSMV primed with an end-labeled oligonucleotide (pr) complementary to a region near the 5'-end of RNAβ(β) or RNAγ(γ). cDNA reactions were incubated at 42°C for 30 min, extracted with phenol and precipitated. RNA was then hydrolyzed in 0.3M NaOH/5mM EDTA for 30 min at 70°C. End-labeled cDNA was precipitated, resuspended in TBE buffer, and electrophoresed on a 20% acrylamide/7M urea gel. Labeled bands were detected by autoradiography.

```
                                                                                                                  MET PRO
 (m7GpppG)UAAAAGAAAAGGAACAACCCUGUUGUUGUUCGACGCUAUACUAAAUAUAUAUUAUCUUAUUAGUGCAUUUCUUUUACCGCUUCACAGU AUG CCG
96
ASN VAL SER LEU THR ALA LYS GLY GLY GLY HIS TYR ILE GLU ASP GLN TRP ASP THR GLN VAL VAL GLU ALA GLY VAL PHE
AAC GUU UCU UUG ACU GCU AAG GGU GGA GGA CAC UAC AUC GAG GAU CAA UGG GAU ACA CAA GUU GUG GAA GCC GGA GUA UUU
177
ASP ASP TRP TRP VAL HIS VAL GLU ALA TRP ASN LYS PHE LEU ASP ASN LEU ARG GLY ILE ASN PHE SER VAL ALA SER SER
GAC GAU UGG UGG GUC CAC GUA GAA GCC UGG AAU AAA UUU CUA GAC AAU UUA CGU GGU AUC AAC UUU AGC GUU GCU UCC UCU
258
ARG SER GLN VAL ALA GLU TYR LEU ALA ALA LEU ASP ARG ASP LEU PRO ALA ASP VAL ASP ARG ARG PHE ALA GLY ALA ARG
CGG UCG CAA GUC GCU GAA UAC UUA GCU GCG UUA GAU CGU GAU CUA CCU GCU GAU GUA GAC AGA CGG UUU GCA GGU GCU AGG
339
GLY GLN ILE GLY SER PRO ASN TYR LEU PRO ALA PRO LYS PHE PHE ARG LEU ASP LYS ARG THR ILE ALA GLU LEU THR ARG
GGA CAA AUU GGU UCA CCC AAU UAU CUU CCU GCG CCA AAA UUC UUU CGU CUC GAU AAG CGA ACU AUC GCU GAA CUG ACU AGA
420
LEU SER ARG LEU THR ASP GLN PRO HIS ASN ASN ARG ASP ILE GLU LEU ASN ARG ALA LYS ARG ALA THR THR ASN PRO SER
CUC UCU CGU CUU ACG GAU CAG CCG CAC AAC AAU CGC GAU AUA GAA CUU AAC CGA GCG AAA AGA GCC ACA ACU AAC CCA UCU
501
PRO PRO ALA GLN ALA PRO SER GLU ASN LEU THR LEU ARG ASP VAL GLN PRO LEU LYS ASP SER ALA LEU HIS TYR GLN TYR
CCC CCG GCG CAG GCA CCG UCG GAG AAU CUC ACU CUU CGU GAU GUU CAA CCG UUA AAG GAU AGU GCG UUG CAU UAU CAA UAC
582
VAL LEU ILE ASP LEU GLN SER ALA ARG LEU PRO VAL TYR THR ARG LYS THR PHE GLU ARG GLU LEU ALA LEU GLU TRP ILE
GUG UUG AUU GAC CUA CAG AGU GCG AGA CUC CCG GUG UAU ACC AGG AAG ACU UUC GAA CGU GAA CUC GCU UUG GAA UGG AUC
663
ILE PRO ASP ALA GLU GLU ALA ???
AUU CCA GAU GCC GAG GAA GCG UGACCUGCUGUCGAAGCGGUAAAAGGAUGUACAUAUGUAUCUUAUUUAUUUUGUUUAUCUAUUUCUUUUGCUUUUAGU
763
                                               MET ASP MET THR LYS THR VAL GLU GLU LYS LYS THR ASN GLY THR ASP
UUUUGCUUUUUACGCAUUAACUAGAUGUAUUGACUUUAGCC   AUG GAC AUG ACG AAA ACU GUU GAG GAA AAG AAA ACA AAU GGA ACU GAU
852
SER VAL LYS GLY VAL PHE GLU ASN SER THR ILE PRO LYS VAL PRO THR GLY GLN GLU MET GLY GLY ASP GLY SER SER THR
UCA GUG AAA GGU GUU UUU GAA AAC UCG ACG AUU CCC AAA GUU CCG ACU GGA CAG GAA AUG GGU GGU GAC GGU UCU UCU ACU
933
                                   [------------------------------------------------*********************..
SER LYS LEU LYS GLU THR LEU LYS VAL ALA ASP GLN THR PRO LEU SER VAL ASP ASN GLY ALA LYS SER LYS LEU ASP SER
UCU AAA UUA AAG GAA ACU CUG AAA GUU GCC GAU CAG ACU CCA UUG UCC GUU GAC AAU GGU GCC AAA UCC AAA UUG GAU UCC
1014
--------------------------[------------------------------------------------*********************......
SER ASP ARG GLN VAL PRO GLY VAL ALA ASP GLN THR PRO LEU SER VAL ASP ASN GLY ALA LYS SER LYS LEU ASP SER SER
UCU GAU AGA CAA GUU CCU GGA GUU GCC GAU CAG ACU CCA UUG UCC GUU GAC AAU GGU GCC AAA UCC AAA UUG GAU UCC UCU
1095
----------------------[
ASP ARG GLN VAL PRO GLY PRO GLU LEU LYS PRO ASN VAL LYS LYS SER LYS LYS LYS ARG ILE GLN LYS PRO ALA GLN PRO
GAU AGA CAA GUU CCU GGA CCU GAG UUG AAA CCC AAC GUU AAG AAG UCC AAG AAG AAA AGA AUC CAA AAA CCU GCU CAA CCG
1176
SER GLY PRO ASN ASP LEU LYS GLY GLY THR LYS GLY SER SER GLN VAL GLY GLU ASN VAL SER GLU ASN TYR THR GLY ILE
AGU GGG CCC AAU GAC CUU AAA GGC GGG ACU AAG GGA UCA UCU CAA GUG GGU GAA AAU GUG AGU GAG AAC UAU ACU GGG AUU
1257
SER LYS GLU ALA ALA LYS GLN LYS GLN LYS THR PRO LYS SER VAL LYS MET GLN SER ASN LEU ALA ASP LYS PHE LYS ALA
UCU AAG GAA GCA GCU AAG CAA AAG CAG AAG ACG CCC AAG UCU GUG AAA AUG CAA AGC AAU CUG GCC GAU AAG UUC AAA GCG
1338
ASN ASP THR ARG ARG SER GLU LEU ILE ASN LYS PHE GLN GLN PHE VAL HIS GLU THR CYS LEU LYS SER ASP PHE GLU TYR
AAU GAU ACU CGU AGA UCG GAA UUA AUU AAC AAG UUU CAG CAA UUU GUG CAU GAA ACC UGU CUU AAA UCU GAU UUU GAG UAC
1419
THR GLY ARG GLN TYR PHE ARG ALA ARG SER ASN PHE PHE GLU MET ILE LYS LEU ALA SER LEU TYR ASP LYS HIS LEU LYS
ACU GGU CGA CAG UAU UUC AGA GCU AGA UCA AAU UUC UUU GAA AUG AUU AAG CUC GCA UCC UUG UAU GAC AAA CAU CUA AAG
```

```
1500
GLU CYS MET ALA ARG ALA CYS THR LEU GLU ARG GLU ARG LEU LYS ARG LYS LEU LEU LEU VAL ARG ALA LEU LYS PRO ALA
GAA UGU AUG GCG CGA GCC UGC ACC CUA GAA CGU GAA CGA UUG AAG CGU AAG UUA CUC CUA GUA CGA GCU UUG AAA CCA GCA
1581
VAL ASP PHE LEU THR GLY ILE ILE SER GLY VAL PRO GLY SER GLY LYS SER THR ILE VAL ARG THR LEU LEU LYS GLY GLU
GUU GAC UUC CUU ACG GGA AUC AUC UCU GGA GUU CCU GGC UCA GGA AAA UCA ACC AUU GUG CGU ACU UUG CUC AAA GGU GAA
1662
PHE PRO ALA VAL CYS ALA LEU ALA ASN PRO ALA LEU MET ASN ASP TYR SER GLY ILE GLU GLY VAL TYR GLY LEU ASP ASP
UUU CCG GCU GUU UGU GCU UUG GCC AAU CCU GCC UUA AUG AAC GAC UAU UCU GGU AUU GAA GGC GUU UAC GGG UUA GAU GAC
1743
LEU LEU LEU SER ALA VAL PRO ILE THR SER ASP LEU LEU ILE ILE ASP GLU TYR THR LEU ALA GLU SER ALA GLU ILE LEU
CUG UUG CUU UCU GCA GUU CCG AUA ACG UCU GAU UUA UUG AUC AUA GAU GAA UAU ACA CUU GCU GAG AGC GCG GAA AUC CUG
1824
LEU LEU GLN ARG ARG LEU ARG ALA SER MET VAL LEU LEU VAL GLY ASP VAL ALA GLN GLY LYS ALA THR THR ALA SER SER
UUG UUA CAA CGA AGA CUC AGA GCC UCU AUG GUG UUG UUA GUC GGG GAU GUA GCU CAA GGA AAA GCC ACC ACU GCU UCC AGU
1905
ILE GLU TYR LEU THR LEU PRO VAL ILE TYR ARG SER GLU THR THR TYR ARG LEU GLY GLN GLU THR ALA SER LEU CYS SER
AUU GAG UAU UUA ACU CUG CCG GUG AUC UAC AGA UCA GAG ACG ACU UAU CGU UUG GGA CAA GAG ACU GCU UCG CUU UGC AGC
1986
LYS GLN GLY ASN ARG MET VAL SER LYS GLY GLY ARG ASP THR VAL ILE ILE THR ASP TYR ASP GLY GLU THR ASP GLU THR
AAG CAG GGU AAC AGA AUG GUU UCA AAG GGU GGA AGG GAC ACA GUG AUC AUU ACU GAU UAC GAU GGC GAA ACA GAU GAA ACG
2067
GLU LYS ASN ILE ALA PHE THR VAL ASP THR VAL ARG ASP VAL LYS ASP CYS GLY TYR ASP CYS ALA LEU ALA ILE ASP VAL
GAG AAA AAU AUC GCU UUU ACU GUC GAU ACA GUU CGA GAU GUG AAA GAU UGC GGG UAC GAU UGU GCC CUG GCA AUU GAU GUG
2148
GLN GLY LYS GLU PHE ASP SER VAL THR LEU PHE LEU ARG ASN GLU ASP ARG LYS ALA LEU ALA ASP LYS HIS LEU ARG LEU
CAA GGG AAA GAA UUC GAU UCA GUG ACU UUA UUC CUA AGG AAC GAA GAC CGG AAA GCU UUA GCA GAU AAG CAU UUG CGU UUA
2229
VAL ALA LEU SER ARG HIS LYS SER LYS LEU ILE ILE ARG ALA ASP ALA GLU ILE ARG GLN ALA PHE LEU THR GLY ASP ILE
GUC GCU UUG AGC AGA CAU AAG UCG AAG UUA AUC AUC AGG GCC GAC GCG GAA AUU CGU CAA GCA UUC CUG ACA GGU GAU AUU
2310
                                                                  M   K   T   T   V   G   S   R   P   N   K
ASP LEU SER SER LYS ALA SER ASN SER HIS ARG TYR SER ALA LYS PRO ASP GLU ASP HIS SER TRP PHE LYS ALA LYS ???
GAC UUG AGC UCU AAG GCG AGU AAC UCU CAU CGU UAU UCU GCA AAA CCG GAU GAA GAC CAC AGU UGG UUC AAG GCC AAA UAA
2391
 Y   W   P   I   V   A   G   I   G   V   V   G   L   F   A   Y   L   I   F   S   N   Q   K   H   S   T   E   S   G   D   N   I   H   K   F   A
GUAUUGGCCAAUUGUCGCCGGAAUCGGUGUCGUUGGAUUGUUUGCGUAUUUGAUCUUUUCAAAUCAAAAACAUUCUACGGAAUCCGGUGAUAAUAUUCACAAAUUCG
2498
  N   G   G   S   Y   R   D   G   S   K   S   I   S   Y   N   R   N   H   P   F   A   Y     G   M   A   S   S   P   G   M   L   L
                                                                  MET ALA MET PRO HIS PRO LEU GLU CYS CYS
CCAACGGAGGUAGUUACAGAGACGGGUCAAAGAGUAUAAGUUAUAAUCGUAAUCAUCCUUUUGCCU  AUG GCA AUG CCU CAU CCC CUG GAA UGU UGU
2594
  P   A   M   L   T   I   I   G   I   I   S   Y   L   W   R   T   R   D   S   V   L   G   D   S   G   G   N
CYS PRO GLN CYS LEU PRO SER SER GLU SER PHE PRO ILE TYR GLY GLU GLN GLU ILE PRO CYS SER GLU THR GLN ALA GLU
UGC CCG CAA UGC UUA CCA UCA UCG GAA UCA UUU CCU AUU UAU GGC GAA CAA GAG AUU CCG UGC UCG GAG ACU CAG GCG GAA
2673
  N   S   C   G   E   D   C   Q   G   E   C   L   N   G   N   S   R   R   S   L   L   C   D   I   G   ?
THR THR PRO VAL GLU LYS THR VAL ARG ALA ASN VAL LEU THR ASP ILE LEU ASP ASP HIS TYR TYR ALA ILE LEU ALA SER
ACA ACU CCU GUG GAG AAG ACU GUC AGG GCG AAU GUC UUA ACG GAC AUU CUC GAC GAU CAU UAC UAU GCG AUA UUG GCU AGU
2954
LEU PHE ILE ILE ALA LEU TRP LEU LEU TYR ILE TYR LEU SER SER ILE PRO THR GLU THR GLY PRO TYR PHE TYR GLN ASP
CUU UUU AUC AUU GCU CUA UGG CUA UUG UAU AUA UAU CUA AGC AGU AUA CCU ACG GAG ACU GGU CCC UAC UUC UAU CAA GAU
2835
LEU ASN SER VAL LYS ILE TYR GLY ILE GLY ALA THR ASN PRO GLU VAL ILE ALA ALA ILE HIS HIS TRP GLN LYS TYR PRO
CUG AAC UCU GUG AAG AUC UAU GGA AUA GGG GCU ACG AAC CCA GAA GUU AUU GCG GCC AUC CAC CAU UGG CAG AAG UAC CCU
```

2916
PHE GLY GLU SER PRO MET TRP GLY GLY LEU ILE SER VAL LEU SER ILE LEU LEU LYS PRO LEU THR LEU VAL PHE ALA LEU
UUU GGG GAA UCU CCG AUG UGG GGA GGU UUA AUC AGU GUU UUG AGU AUU CUU CUU AAA CCG CUG ACA UUA GUU UUU GCG UUA
2997
SER PHE PHE LEU LEU LEU SER SER LYS ARG ???
AGC UUU UUU CUC UUA CUU UCU UCA AAA AGG UAAAAAAAAAAAAAAAAAAAAAAUGUUUGAUCAGAUCAUUCAAAUCUGAUGGUGCCCAUCAACCAUA
3096
UGAUGGGAGUGUUUGCAAGUCCACUAUAAUCGAACUUGAAAACAAUGCCUGAAUUGGAAACCAUGAAUCUUAACGGAUUCUGGAGAGAAAAUUUAGGAAUUGGUAUG
3203
UAAGCUACAACUUCCGGUAGCUGCGUCACACUUUAAGAGUGUGCAUACUGAGCCGAAGCUCAGCUUCGGUCCCCCAAGGGAAGACCA

Figure 3. Nucleotide sequence of the coding and 3'-non-coding regions of BSMV RNAβ. The deduced amino acid sequences of three open reading frames are given in three letter codes. One letter codes are used for the predicted amino acid sequence of a fourth, overlapping open reading frame. Stop codons are represented by ??? or ?. A 78 nucleotide direct tandem repeat (I-----I) containing a 15 nucleotide palindrome (****) is also shown.

segment of RNAβ which is less than 100 nucleotides from its 5'-end. When the β1 oligonucleotide was end-labeled and used to prime reverse transcription of RNA from the Type strain of BSMV, two distinct cDNA fragments were synthesized (Fig. 2). These two cDNA fragments were also produced when the β1 oligonucleotide was used to prime reverse transcription of RNA from the ND18 strain of BSMV or of RNA from the Type strain which had been denatured in 5mM methylmercury hydroxide (data not shown). In contrast, only a single cDNA fragment was synthesized when an oligonucleotide complementary to a region near the 5'-end of RNAγ was used to prime reverse transcription of Type RNA (Fig. 2).

The larger cDNA fragment synthesized from the β1 primer was purified and sequenced by the chemical degradation method (30). The deduced sequence of the complementary (viral RNA) strand is similar to a previously published partial sequence of the 5'-noncoding region of RNAβ from the Norwich strain of BSMV (31). The sequencing data do not unequivocally establish the identity of the initial nucleotide of RNAβ or the presence of a cap sructure at its 5'-end. However, it is known that BMSV RNAs α and γ are capped with 7-methylguanosine (4) and that guanosine is the initial nucleotide of RNAγ (32).

The smaller cDNA fragment synthesized from the β1 primer has not yet been purified and sequenced. However, sequencing data obtained using total cDNA synthesized from the β1 primer suggest that the small cDNA fragment is a truncated form of the larger cDNA fragment which lacks the final 37 or 38 nucleotides of the large fragment. The sequence at the 3'-terminus of the short cDNA fragment could not be resolved and may differ from the sequence

determined for the corresponding portion of the large cDNA fragment. The sequences of the two cDNA fragments synthesized from the β1 primer are otherwise identical.

Sequence Analysis

The complete nucleotide sequence of RNAβ from the Type strain of BSMV is presented in Figure 3. The (+)-stranded (virion polarity) RNA is 3289 nucleotides in length and contains four open reading frames (Fig. 4) which code for polypeptides with molecular weights of greater than 14,000. If translated, the next largest ORF on the (+)RNA strand would encode a polypeptide with a molecular weight of approximately 6000. The largest open reading frame on the (-)RNA strand would code for a protein comprised of 87 amino acids.

Identification of the BSMV Coat Protein Gene

The first open reading frame (ORF1) in our sequence extends from the AUG codon at position 90-92 to the termination codon at position 684-686 (198 amino acids). Two lines of evidence suggests that this is the gene for BSMV coat protein. First, the amino acid composition of the BSMV coat protein as determined by acid hydrolysis (33) agrees very closely with the amino acid composition predicted from the nucleotide sequence (Table 1). Second, following the initiator methionine, the predicted N-terminal amino acid sequence of the ORF1 translation product is identical to that determined by direct sequencing of purified BSMV coat protein (31) for 29 of the first 30 amino acids (Fig. 5).

Identification of Other RNAβ ORFs

Three additional ORFs capable of coding for polypeptides with molecular weights greater than 14,000 were identified on RNAβ (Fig. 4). The second ORF is separated from the coat protein gene by a 117 nucleotide intercistronic region that is low in G (14%) and high in U (42%). ORF2 codes for a polypeptide of unknown function that has a molecular weight of 58,098. There is a 78 nucleotide direct tandem repeat located about 150 nucleotides into the ORF (Fig. 3). Within each repeat is a 15 nucleotide

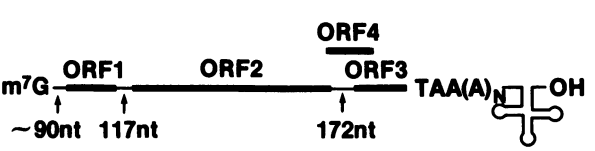

Figure 4. A genetic map of BSMV RNAβ. The four major open reading frames, the intergenic regions, the 5'-non-coding region and the 3' internal poly(A) tract and tRNA-like structure of BSMV RNAβ are depicted.

Table 1. Amino acid composition of BSMV coat protein.

Amino acid	Amount Determined by acid hydrolysis of coat protein (33)	Amount predicted from the sequence of ORF1 of RNAβ
Asx	25-26	26
Glx	19-20	21
Ala	20	21
Leu	21	21
Arg	17	17
Pro	12-13	13
Thr	9	10
Ser	9	11
Val	10	12
Gly	8	8
Lys	7	7
Phe	7	7
Tyr	8	8
Ile	6	8
Trp	5	5
His	4	4
Cys	0	0
Met	0	1
Total	187-190	198

palindromic sequence. This direct repeat has been shown to be present in all three of the independently derived clones depicted in Figure 1. It is therefore unlikely that the repeat results from a copying error by the reverse transcriptase enzyme.

ORF3 begins 173 nucleotides downstream from the termination codon for ORF2 and codes for a polypeptide with a molecular weight of 17,378. The first two adenine residues of the internal poly(A) sequence are used to form the stop codon (TAA) for this ORF (Fig. 4).

ORF4 overlaps the last 29 nucleotides of ORF2 and the first 188 nucleotides of ORF3 (Fig. 4) and could potentially code for a protein with a molecular weight of 14,119.

(A)		PRO	GLN	VAL	SER	LEU	THR	ALA	LYS	GLY	GLY
(B)	MET	PRO	ASN	VAL	SER	LEU	THR	ALA	LYS	GLY	GLY
(A)	GLY	XXX	TYR	XXX	GLU	ASP	GLN	TRP	ASP	THR	XXX
(B)	GLY	HIS	TYR	ILE	GLU	ASP	GLN	TRP	ASP	THR	GLN
(A)	VAL	VAL	GLX	ALA	XXX	VAL	PHE	ASP	ASP		
(B)	VAL	VAL	GLU	ALA	GLY	VAL	PHE	ASP	ASP		

Figure 5. Comparison of the N-terminal amino acid sequence of BSMV coat protein as determined by direct sequencing (31) of the purified polypeptide (A) with the predicted N-terminal amino acid sequence of the protein encoded by the first open reading frame on BSMV RNAβ (B).

Table 2A. Usage of degenerate codons in BSMV RNAβ open reading frames (ORFs).

	Percent of third postion bases			
ORF	G	A	U	C
1	22	24	35	19
2	23	26	33	18
3	26	22	35	17
4	14	27	33	27

Table 2B. Percent base composition of BSMV RNAβ open reading frames (ORFs).

ORF	G	A	U	C
1	24	26	26	24
2	24	30	26	20
3	22	25	32	21
4	23	28	28	21

Homology Between Polypeptides Encoded by RNAβ and Other Proteins

Each of the polypeptides encoded by the four major ORFs on RNAβ was analyzed for homology with other proteins using the FASTP and RDF computer programs (34). No significant amino acid sequence homology was found between the polypeptides encoded by BSMV RNAβ and proteins contained in the National Biomedical Research Foundation library.

Analysis of Translation Initiation Sites

The sequences surrounding the initiation codons of the four RNAβ open reading frames agree to various extents with the concensus sequence (CCA/GCCAUGG) for translation initiation sites in eukaryotic mRNAs (35). The ORF2 sequence (UAGCCAUGG) matches the concensus sequence in four of the six positions, while the ORF1 sequence (ACAGUAUGC) and the ORF3 sequence (UGCCUAUGG) match the concensus in two of six positions. The sequence surrounding ORF4 (ACCGGAUGA) has only one nucleotide in common with the concensus. Only the ORF1 and ORF2 initiation sites have a purine in the -3 position; a feature which is highly conserved in eukaryotic mRNAs (35).

Codon usage of RNAβ ORFs. The distribution of degenerate codons is decidedly non-random in all four RNAβ ORFs. In each case U is the preferred third position base while either C (ORFs 1,2 and 3) or G (ORF4) is avoided (Table 2A). These preferences do not simply mirror the base composition (Table 2B), which is nearly identical in each ORF. A preference for U in

the third position has also been observed in coding regions of brome mosaic virus (10), cucumber mosaic virus (11,12) and alfalfa mosaic virus (13,14). The difference in codon usage between ORF4 and ORFs 1,2 and 3 suggests that ORF4 may not be translated in vivo; however, this can only be verified through further experimentation.

The 3'-Untranslated Region

A 3'-noncoding region consisting of a poly(A) tract of variable length followed by a tRNA-like structure is present in all BSMV genomic RNAs (8). The poly(A) tract in clone pBSM191 (Fig. 1) is composed of 22 adenine residues which is within the previously determined size range of the poly(A) region (9). The tRNA-like structure in pBSM191 is 238 nucleotides in length and contains two base additions (positions 3126 and 3135) and three base substitutions (positions 3048, 3071, and 3186) when compared to the sequence previously published for the Argentina Mild and Norwich strains of BSMV (5). These substitutions and additions have no effect on the proposed secondary structure for this region.

DISCUSSION

The analysis of the nucleotide sequence RNAβ in conjunction with in vitro translation studies (21,23) confirms that there are several differences between BSMV and other tripartite viruses. One significant difference is the location of the coat protein gene. In BSMV, this gene is located at the 5'-end of RNAβ and is expressed directly from that genomic RNA. In other tripartite viruses, the coat protein gene is located at the 3'-end of RNA3 and is expressed only from a subgenomic (sg)RNA.

A second important difference is the number of polypeptides which are encoded by the genomic RNAs. Discounting small open reading frames on the plus and minus-stranded RNAs of brome mosaic virus, cucumber mosaic virus and alfalfa mosaic virus, the coding capacity of these tripartite viruses would appear to be limited to only four polypeptides. In comparison, the tripartite genome of BSMV could code for as many as six or perhaps even seven polypeptides. BSMV RNAα directs the synthesis of an M_r 120,000 protein in vitro (23) and preliminary sequencing data suggest that this is the only polypeptide encoded by that RNA (36). BSMV RNAγ encodes two polypeptides (32), one with a molecular weight of approximately 75,000 or 85,000 (depending on the strain of BSMV) which is translated directly from the genomic RNA (21,23), and another with a molecular weight of about 17,000 which is expressed from an sgRNA generated from the 3'-end of the genomic

RNA (37). BSMV RNAβ codes for at least one and potentially as many as four polypeptides. The data presented here clearly indicate that the open reading frame nearest the 5'-end of RNAβ (ORF1) codes for BSMV coat protein. This gene is highly expressed in both in vitro and in vivo (21). Both ORF2 and ORF3 (Fig. 4) are likely to be expressed based on the similarity in codon usage between these two ORFs and the coat protein gene and on the assumed necessity for viruses to make maximum use of their genetic material. In addition, ORF4 might also be expressed even though there are differences in codon usage between it and the other RNAβ ORFs. However, at this time no direct evidence exists for expression in vivo of specific translation products or sgRNAs from ORF 2,3 or 4 of RNAβ.

Based on the nucleotide sequence of ORF1 of RNAβ, the BSMV coat protein would be predicted to contain a single methionine residue located at its N-terminus. However, it has been shown by direct sequencing that the N-terminal amino acid of the BSMV coat protein is actually proline (31). Furthermore, analysis of the amino acid composition of the BSMV coat protein (Table 1) confirms that the N-terminal methionine is not present. These data indicate that the methionine residue, which originally precedes the proline residue in the BSMV coat protein, is at some point cleaved or modified. This process must occur rapidly even in vitro because the coat protein is not labeled when BSMV RNAs are translated in vitro in the presence of [^{35}S]-methionine (21). The type of modification occurring at the N-terminus of BSMV coat protein is not known. However, the situation could be similar to that observed in brome mosaic virus in which the N-terminal Met-Ser of the BMV coat protein is modified to acetylserine (38,39).

Coat protein is the only polypeptide which is consistently translated from RNAβ in vitro. However, under certain ionic conditions RNAβ also directs the synthesis of an M_r 25,000 polypeptide in both the wheat germ and rabbit reticulocyte lysate systems (23). A protein with an Mr of 28,000, which appears to correspond to the M_r 25,000 polypeptide, was detected in independent experiments in which unfractionated BSMV RNA was translated in wheat germ extracts (21). This 28 kd product is apparently unrelated to BSMV coat protein because it cannot be precipitated with coat protein antibodies and because it contains methionine (21). Under the proper ionic conditions, the amounts of the 25-28 kd polypeptide and coat protein which are synthesized in vitro are approximately equal (21,23). However, unlike coat protein, this polypeptide has never been detected in vivo (21). Although none of the four open reading frames which we have identified on

RNAβ code specifically for a 25-28 kd polypeptide, the protein could possibly be produced within ORF2 either by premature termination or by internal initiation at the AUG codon located at positions 1698-1700. None of the remaining ORFs on RNAβ are large enough to encode a polypeptide of this size. However, other initiation sites and mechanisms of expression cannot be ruled out until corroborating data on the N-terminal amino acid sequence of this protein is available.

BSMV RNAβ appears to be characteristic of multigenic RNAs from other tripartite viruses in that only the gene nearest the 5'-end (coat protein gene) is translated from the genomic RNA in vitro. The most likely mechanism for the in vivo expression of the remaining ORFs on RNAβ is through the production of sgRNAs. Although several BSMV-related sgRNAs have been detected, none have been specifically associated with RNAβ. The only sgRNA detected in RNA isolated from purified BSMV virions is generated from the 3'-end of RNAγ (22,37). This sgRNA terminates with a poly(A) tract and lacks the tRNA-like structure that is found at the 3'-end of the genomic RNAs (25). Two sgRNAs (0.65×10^6 and 0.28×10^6), which could be aminoacylated with tyrosine and so presumably contain the tRNA-like structure, were detected in BSMV-infected barley protoplasts (7). However, the origin of these sgRNAs has not been determined.

The failure to detect sgRNAs originating from RNAβ in BSMV RNA preparations (22) may indicate that these sgRNAs are not encapsidated. It is possible, in a situation analogous to the fate of the sgRNA which codes for coat protein in the common versus the cowpea strain of tobacco mosaic virus (40), that the BSMV encapsidation nucleation site is present in the RNAγ sgRNA but not in the putative RNAβ sgRNAs. Another possibility is that sgRNAs from RNAβ are encapsidated, but at a much lower level than the RNAγ sgRNA. In RNA prepared from BSMV virions, the quantity of RNAγ greatly exceeds that of the sgRNA which is generated from it. The RNAγ sgRNA was originally detected only after exposing blots containing 75 ng of total BSMV RNA hybridized with labeled cDNA for 30 hours (22). Subgenomic RNAs which are encapsidated to a lesser extent than the RNAγ sgRNA may not have been detected under similar conditions.

ACKNOWLEDGEMENTS

We would like to thank Dr. Ramamoorthy Belagaje for synthesis of the oligonucleotides used in this project and Dr. George Long for his help with Maxam and Gilbert sequencing.

REFERENCES

1. Matthews, R.E.F. (1979) Intervirology 12, 132-296.
2. Jackson, A.O. and Lane, L.C. (1981) In "Handbook of Plant Virus Infections" (E. Kurstak, ed.), pp. 565-624. Elsevier/North Holland, Amsterdam/New York.
3. Jackson, A.O., McFarland, J.E., Gustafson, G.D., Dawson, J.R.O. and Stanley, J. (1983). In "Plant Molecular Biology", UCLA Symposia on Molecular and Cellular Biology, Vol. 12, (R. B. Goldberg, ed.), pp 67-79, Alan R. Liss, Inc., New York, New York.
4. Aganovsky, A.A., Dolja, V.V., Kagramanova, V.K. and Atabekov, J.G. (1979) Virology 95, 208-210.
5. Kozlov, Yu.V., Rupasov, V.V., Adyshev, D.M., Belgelarskaya, S.N., Agranovsky, A.A., Mankin, A.S., Morozov, S. Yu, and Atabekov, J.G. (1984) Nucl. Acids Res. 12, 4001-4009.
6. Agranovsky, A.A., Dolja, V.V., Gorbulev, V.G., Kozlov, Yu.V. and Atabekov, J.G. (1981) Virology 113, 174-187.
7. Loesch-Fries, L.S. and Hall, T.C. (1982) Nature 298, 771-773.
8. Agranovsky, A.A., Dolja, V.V. and Atabekov, J.G. (1982) Virology 119, 51-58.
9. Agranovsky, A.A., Dolja, V.V. and Atabekov, J.G. (1983) Virology 129, 344-349.
10. Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984) J. Mol. Biol. 172, 369-383.
11. Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984) Eur. J. Biochem. 143, 277-284.
12. Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985) Eur. J. Biochem. 150, 331-339.
13. Cornelissen, B.J.C., Brederode, F.Th., Moormann, R.J.M. and Bol, J.F. (1983) Nucl. Acids Res. 11, 1253-1265.
14. Cornelissen, B.J.C., Brederode, F.Th., Veeneman, G.H., van Boom, J.H. and Bol, J.F. (1983) Nucl. Acids Res. 11, 3019-3025.
15. Ahlquist, P., Luckow, V. and Kaesberg, P. (1981) 153, 23-38.
16. Gould, A.R. and Symons, R.H. (1982) Eur. J. Biochem. 126, 217-226.
17. Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S. and Hall, T.C. (1983) Nucl. Acids Res. 11, 2881-2891.
18. Shih, D.S. and Kaesberg, P. (1976) J. Mol. Biol. 103, 77-88.
19. Schwinghammer, M.W. and Symons, R.H. (1977) Virology 79, 88-108.
20. Brederode, F. Th., Koper-Zwarthoff, E.C. and Bol, J.F. (1980) Nucl. Acids Res. 8, 2213-2223.
21. Gustafson, G.D., Larkins, B.A. and Jackson, A.O. (1981) Virology 111, 579-587.
22. Gustafson, G.D., Milner, J.J., McFarland, J.E., Pedersen, K., Larkins, B.A. and Jackson, A.O. (1982) Virology 120, 182-193.
23. Dolja, V.V., Lunia, N.A., Leiser, R.-M, Stanarius, T., Belzhelarskaya, S.N., Kozlov, Yu.V. and Atabekov, J.G. (1983) Virology 127, 1-14.
24. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269.
25. Stanley, J., Hanau, R. and Jackson, A.O. (1984) Virlogy 139, 375-383.
26. Rave, N., Crkvenjakov, R. and Boedtker, H. (1979) Nucl. Acids Res. 6, 3559-3567.
27. Thomas, S.P. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
28. Sanger, R., Nicklen, S. and Coulsen, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
29. Biggen, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
30. Maxam, A. and Gilbert W. (1980) Methods Enzymol. 65, 499-560.
31. Dolja, V.V., Lunina, N.A., Smirnov, V.D., Karpov, V.A., Chudiakov,

Yu.E., Kozlov, A.A. and Atabekov, J.G. (1982) Dokl. Acad. Nauk. SSSR 265, 474-477.
32. Gustafson, G.D., Hunter, B., Hanau, R., Armour, S.L. and Jackson, A.O. In preparation.
33. Atabekov, J.G., Novikov, V.K., Kiselev, N.A., Kaftanova, A.S. and Egporov, A.M. (1968) Virology 36, 620-638.
34. Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435-1441.
35. Kozak, M. (1984) Nucl. Acids Res. 12, 857-872.
36. Gustafson, G.D. and Armour, S.A.; Unpublished results.
37. Jackson, A.O., Dawson, J.R.O., Covey, S.N., Hull, R., Davies, J.W., McFarland, J.E. and Gustafson, G.D. (1983) Virology 127, 37-44.
38. Shih, D.S. and Kaesberg, P. (1973) Proc. Natl. Acad. Sci. USA 70, 1799-1803.
39. Moosie, J., McKean, D., Shih, D.S. and Kaesberg, P. (1983) J. Virol. 129, 517-520.
40. Fukuda, M., Okada, Y., Otsuki, Y. and Takebe, I. (1980) Virology 101, 493-502.