



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Nucleotide Sequence of the Bean Strain of Southern Bean Mosaic Virus

YASMIN OTHMAN¹ AND ROGER HULL²

John Innes Institute, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, United Kingdom

Received August 9, 1994; accepted September 20, 1994

The genome of the bean strain of southern bean mosaic virus (SBMV-B) comprises 4109 nucleotides and thus is slightly shorter than those of the two other sequenced sobemoviruses (southern bean mosaic virus, cowpea strain (SBMV-C) and rice yellow mottle virus (RYMV)). SBMV-B has an overall sequence similarity with SBMV-C of 55% and with RYMV of 45%. Three potential open reading frames (ORFs) were recognized in SBMV-B which were in similar positions in the genomes of SBMV-C and RYMV. However, there was no analog of SBMV-C and RYMV ORF 3. From a comparison of the predicted sequences of the ORFs of these three sobemoviruses and of the noncoding regions, it is suggested that the two SBMV strains differ from one another as much as they do from RYMV and that they should be considered as different viruses.

© 1995 Academic Press, Inc.

INTRODUCTION

Southern bean mosaic virus (SBMV) is the type member of the sobemovirus group of small icosahedral positive-strand RNA viruses (for reviews see Sehgal, 1981; Tremaine and Hamilton, 1983; Hull, 1988). Four major strains of SBMV are recognized which are serologically related (Grogan and Kimble, 1964). The type strain (bean strain), SBMV-B, infects several *Phaseolus* spp., including many cultivars of *Phaseolus vulgaris*, but does not systemically infect cowpeas (*Vigna unguiculata*), whereas the cowpea strain (SBMV-C) infects *V. unguiculata* and *Pisum sativum* but fails to systemically infect beans (Shepherd and Fulton, 1962). The Ghana and Mexican or severe bean mosaic strains (SBMV-G and SBMV-M) (Lampthey and Hamilton, 1974; Yerkes and Patino, 1960), which are transmissible to several cultivars of both bean and cowpea, are distinguished by their reaction in certain cultivars and also serologically.

The SBMV genome consists of a single molecule of positive sense RNA (M_r 1.4×10^6) which constitutes approximately 21% of the total particle weight. The 5' terminus of SBMV RNA has a covalently linked protein (VPg) (Ghosh *et al.*, 1979) of molecular weight between 10 (SBMV-C) and 12 kDa (SBMV-B) (Mang *et al.*, 1982), which is essential for infectivity (Veerisetty and Sehgal, 1980). The 3' terminus of SBMV does not have a poly(A)

tail (Mang *et al.*, 1982) but contains a free hydroxyl group (Ghosh *et al.*, 1979). A subgenomic RNA of 0.4×10^6 encodes the gene for viral coat protein (Rutgers *et al.*, 1980; Ghosh *et al.*, 1981).

The full nucleotide sequence of the cowpea strain, SBMV-C, has been determined (Wu *et al.*, 1987). The genome consists of 4194 nucleotides with four open reading frames designated ORFs 1–4 (Fig. 1). In addition, limited sequence data are available for the 3' end of SBMV-B (Mang *et al.*, 1982), which showed little similarity to the sequence of SBMV-C. The recently published nucleotide sequence (4450 nt) of rice yellow mottle virus (RYMV), another sobemovirus (Ngon A Yassi *et al.*, 1994), shows that it has a similar genome organization to SBMV-C (Fig. 1).

The RNAs of SBMV-B, SBMV-C, and SBMV-GH have been translated in the rabbit reticulocyte lysate and wheat germ extract *in vitro* systems (Salerno-Rife *et al.*, 1980; Ghosh *et al.*, 1981; Rutgers *et al.*, 1980; Brisco *et al.*, 1985b) to give four major polypeptides: polypeptide P1 of molecular weight 100–105 kDa, P2 (60–75 kDa), P3 (28–29 kDa), and P4 of molecular weight between 14 and 25 kDa. Wu *et al.* (1987) related these products to ORFs 1 (P4), 2 (P1 and P2), and 4 (P3, coat protein) of the SBMV-C sequence.

We have determined the complete sequence of the genomic RNA of SBMV-B. Detailed comparison of the SBMV-B sequence with that of SBMV-C RNA reveals significant differences in the organization and deduced amino acid sequences of the predicted major open reading frames. These comparisons have been extended to a comparison with other plant viruses. Our observations suggest that while SBMV-B and SBMV-C are related they should be considered as distinct viruses.

Sequence data from this article have been deposited with the GenBank Data Libraries under Accession No. L34672.

¹ Present address: Department of Genetics and Cellular Biology, Universiti Malaya, 59100 Kuala Lumpur, Malaysia.

² To whom correspondence and reprint requests should be addressed.

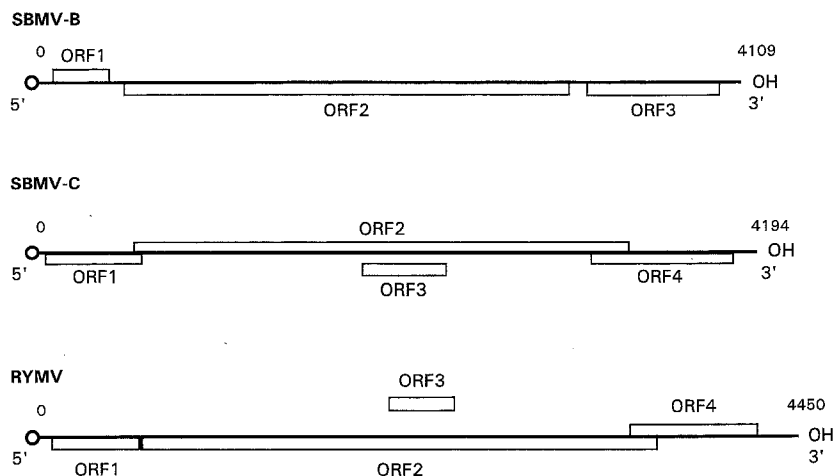


FIG. 1. Genome organizations of sobemoviruses. The viral RNA is represented by the thick line with the VPg (circle) at the 5' end. The open reading frames (ORFs) are shown as boxes.

MATERIALS AND METHODS

Virus purification and RNA extraction

SBMV-B, originally obtained from Dr. J. P. Fulton (Arkansas) and maintained as dried leaves since 1974, was propagated in *P. vulgaris* (cv. The Prince) plants and the virus was purified essentially as described by Hull (1985). Viral RNA was extracted using phenol-chloroform by the method of Zimmern (1975) from virus particles dissociated in the presence of 1% SDS.

cDNA cloning and sequencing

First- and second-strand cDNA synthesis was performed according to Gubler and Hoffman (1983) using the cDNA Cloning System Plus kit (Amersham) with Superscript reverse transcriptase (BRL). Second-strand products were treated briefly with T4 DNA polymerase (Sambrook *et al.*, 1989) to remove any 3' protruding ends. The blunt-ended double-stranded cDNA was size fractionated using a Sephacryl column (Size-sep 400; Pharmacia) and products >400 bp in size were cloned into the *Sma*I site of pBluescript II SK(+) vectors (Stratagene) using standard techniques (Sambrook *et al.*, 1989). The resulting set of overlapping cDNA clones were used as templates for sequencing.

Cloning of the 3' end. Clones representing the 3' end of SBMV-B RNA used in the confirmation of the 3' terminal sequences were obtained by polyadenylation of the RNA using poly(A) polymerase (Pharmacia), and oligo(dT)-primed clones were generated using the Amersham cloning kit as described above. These cDNA's were ligated into the pBluescript vectors at the *Hinc*II site.

Sequencing of SBMV-B. The dideoxynucleotide chain termination procedure (Sanger *et al.*, 1977) was used to sequence the double-stranded DNA clones with Sequenase (USB Corp.). Ambiguities in the sequence were re-

solved by using PCR cycle sequencing (Murray, 1989) or using dITP with Sequenase. The sequence of the 5' end of the RNA was established by the dideoxynucleotide chain termination method directly on the RNA, with primer extensions using terminal transferase (de Borde *et al.*, 1986) and oligonucleotide primers complementary to nucleotides 41 to 58 of SBMV-B. General cloning techniques were as in Sambrook *et al.* (1989).

Computer analysis

Sequence data were assembled and analyzed using the UWGCG programs (Devereux *et al.*, 1984). The COMPARE and DOTPLOT algorithms were used for RNA and protein sequence comparisons. Alignment of homologous nucleotide or amino acid sequences were obtained using the GAP and BESTFIT algorithms and amino acid alignments were refined using the SOMAP program (Parry-Smith and Atwood, 1990). AHOMOL (Hull *et al.*, 1986) was used for presentation of some alignments. Database searches were carried out either through the FASTA program (Devereux *et al.*, 1984) or through the DAPJOB program at the University of Kent in Canterbury (PROSRCH).

RESULTS AND DISCUSSION

Nucleotide sequence and organization of the SBMV-B genome

The main strategy employed in the determination of the sequence of SBMV-B involved the use of the dideoxy-chain termination method together with a selection of overlapping cDNA clones which spanned the full-length of the genome of SBMV-B, excluding approximately 40 nucleotides at the 5' end. The use of synthetic oligonucleotide primers designed on both strands of the clones as the sequences were elucidated enabled systematic

sequencing along the whole length of the clone. All of the genome was sequenced from at least two independent clones and each clone was sequenced on both strands.

Attempts at cloning the 5' terminal sequences into the bacterial plasmids pUC18 or pBluescript proved unsuccessful and all clones obtained lacked at least 40 terminal nucleotides. The sequence of the 5' terminal region of the RNA that was not contained within the cloned cDNA was therefore analyzed directly on the viral RNA using specific oligonucleotide primers and reverse transcriptase and by additional extension using terminal transferase. The reason for the inability to clone the 5' terminal region in pBluescript and related plasmids is unknown but it did prove possible to clone in plasmids in *Escherichia coli* behind the cauliflower mosaic virus 35S promoter. It appeared that this portion of the SBMV-B genome could be toxic to *E. coli* cells if behind the T7 or T3 promoters.

The 3' terminal sequence was obtained from independent poly(A)-tailed cDNA clones and shows identity in all but eight bases when compared to the published 393 nucleotide terminal sequence of SBMV-B (Mang *et al.*, 1982). Six of the changes were U-C, C-U, A-U, or U-A substitutions, the 3' nucleotide was a G-U substitution and nucleotide 91 on the Mang *et al.* (1982) sequence was absent in our sequence. The differences were conserved in all the clones that were sequenced and could represent isolate heterogeneity rather than sequencing artifacts.

The complete nucleotide sequence and the deduced amino acid sequence of the predicted translation products of SBMV-B are shown in Fig. 2. The genome contains 4109 nucleotides and thus is slightly shorter than that of SBMV-C. The base composition of SBMV-B RNA comprised 23.7% A, 24.5% C, 25.6% G, and 26.0% U residues, which is similar to that reported from direct analysis (Tremaine, 1966; Ghabrial *et al.*, 1967), and the RNA has an overall sequence identity with SBMV-C of only 55%. The identity of SBMV-B with RYMV at the nucleotide level was 45.1%; sequence homology between RYMV and SBMV-C was 45.6%.

Computer analysis of SBMV-B RNA and its complementary strand in all six possible reading frames revealed three potential open reading frames (ORFs) of >10 kDa on the messenger-sense strand of the RNA, referred to as ORF 1 to ORF 3 in order of their proximity to the 5' end of the sequence (Fig. 1); there were no ORFs of >10 kDa on the complementary strand. Thus, the overall genome arrangement of ORFs 1, 2, and 3 appears similar to ORFs 1, 2, and 4 of SBMV-C (Fig. 1) and RYMV. However, there are differences in detail. In SBMV-B and RYMV there are intercistronic regions between ORFs 1 and 2 whereas these ORFs overlap in SBMV-C. In SBMV-B there is also an intercistronic region between ORFs 2 and 3 whereas the analogous ORFs of

SBMV-C and RYMV (ORFs 2 and 4) overlap. No potential internal coding region within ORF 2 similar to the ORFs 3 of SBMV-C and RYMV could be identified in SBMV-B. An increasing number of viral genes are now known to utilize non-AUG codons, which allow translation of multiple related proteins from a single ORF (Beccera *et al.*, 1985; Prats *et al.*, 1989; Gordon *et al.*, 1992). The efficiency of translation is generally much lower than initiation at AUG. However, even with these initiation codons no ORF could be defined which was similar to SBMV-C ORF3.

Noncoding regions of SBMV-B RNA

An A + U rich (61%) 5' noncoding region of 92 bases precedes the first ORF of SBMV-B. The leader sequence of SBMV-B was considerably longer than that of SBMV-C (46 nt) and of a similar length to that of RYMV (79 nt). The alignment of the 5' noncoding region of SBMV-B with those of SBMV-C and RYMV (Fig. 3) shows a region of close similarity between the 5' 43 nucleotides of SBMV-B and the leader of SBMV-C but less similarity to the leader of RYMV; there was no increase in similarity between the leader sequences of SBMV-B and RYMV in the part not shown in Fig. 3. There are various suggestions for significant motifs in the leader sequences of SBMV-C and RYMV (Wu *et al.*, 1987; Ngon A Yassi *et al.*, 1994). These include suggested complementarity to the 3' terminus of 18S rRNA for both viruses (highlighted in Fig. 3) and a repeat of the sequence ACAAUUG in RYMV. The alignment of the three leader sequences does not lend much support to any significance of these motifs. Keese *et al.* (1990) noted similarities between the 5' terminal sequences of various luteoviruses and SBMV-C. This does appear to be reflected in the consensus sequence between the three sobemoviruses (Fig. 3).

The 3' terminus of SBMV-B is nonpolyadenylated and the 129 nucleotides which constitute the noncoding region exhibit no strong secondary structures using the computer folding programs of Zuker and Steigler (1981). Although the 3' noncoding sequence of SBMV-C was of similar length (136 nt) and conformation (no tRNA-like structures), very little sequence similarities (34%) were found between the two noncoding regions. The somewhat longer 3' noncoding region of RYMV (245 nucleotides) also shows no significant sequence similarity (40%).

Although the subgenomic mRNAs of sobemoviruses have not been mapped in detail they do encode the 3' ORF (Rutgers *et al.*, 1980) and most likely resemble many other viruses in being 3' coterminal with the genomic RNA. This would mean that they are promoted from a sequence upstream of the 3' ORF which in SBMV-B is an intergenic region. No significant sequence similarity could be found between this intergenic region and the

1 CACAAAUAUAAGAAGGAAAGCUGGAUUUCCUACCUUUGUUUCCAUUGUCGAGCAUUGGUC AACGAAUACAAAACGGUGCAUUUUCUGCAUGAGCUA
ORF1 M S Y

101 UCGUUUCCUAACAGUUAGAGCAUUCGGCUUUAACGGUUUCCAUUGCGACGCUACGCGCUUGCUAAGUGAAACAGAAUCUAGAGCAUUCUACGUCCAUG
4 R F L T V R A F G F T G F H C D A T R L L S E T E V I D V P T S M

201 ACUUGUUGGCGAGACCGAACUCAGACUUGAAACUGCUUGGCCCCAGUGUGAAGAGAAUUGUUACACGAUUCUCCUUGAUUCAACGUUCAAGUUGACUU
37 T L L A R P N S D L K L L G P S V K R I V T R F S L D S T F K L T S

301 CGAGUAUACCCUGUGCGUGUCGAGAUUGUGUGCCGAGUCUGCGCUACAUCUCCUAUCUGUCAUUUUCAGCAAGUGGGACUUCUACUGUAGUAGGAGAGGC
71 S I T L C V S R L C A E S A L H P Y L S F S A S G T S T V V G E A

401 CAUUUUGUUCUGUAGAUCAAGCGGGAUCUGUUUAGGAUUGAACGUCUCCGGGAGACGGGAGAGAAAUCUUCUACUUCUGUGUAAAUCUACUGUA
104 I L F L * i r t g i c l g l e r s g r r e r n t s t s v i n l s v

501 GACAAUGUAUCAUCCAGGCCGUCACCAUCAUCCUGAUAAACGUUGGCAAACGUUAUCUGCGCGGCAUUUUGUUGGACAUCUCCAGGGGGUUUACCAAC
d n v s s r p l t i i p d n v g k r y l r g d f v r h p s g g y q p
ORF2 M Y H P G R S P S F L I T L A N V I C A A I L F D I H R G V T N

601 CCGGUCACUAUCCAUAUGUGCCUGGAUACCCCGUUCGUUACUCUGCCUGGUUGAGCGCGUCAUUCGGCACAUAACUUAUAAGUAUGUAAGAUCUGGC
g s l s n s g l d t p f v t l p g *
33 P A H Y P I V A W I P R S L L C L V E R V I R D I P Y K Y V R T R L

701 UGCUCAGAGGAGAAGGUUGCCAGGCUAUAUCGCGCAGUCUGCCGUAUUUGUCCCGUGGUGUAUACAGACUUGAUCGUCGAAUGCGGAGACCGCAUG
67 L Q R R R L P G Y I A Q S A V F V P L V Y T D L I A R M S E T A W

801 GACCACGGGACCGGUAUCAACUGAGAGCACUGGUUAUCGCAACGUGACAUUGUUAUACGAGUACAGGCAGUUCACCGAGUCUGCGGACGGUUAUC
100 T T G T G Y Q L M S T G I R N A D I V D T S T G S F T E S A D G Y

901 CAACAGUCACGGUCAGUCCUGUCGAGACAUUGGCCUGAAGAGCGUCUUCGACUUGCCGUGUAUUCGGAUAGCGAACCACGAAGCAUUGGUCCUCCUGCGA
133 Q Q S R S V L C R H W P E E R L R L A V Y S D S E P R S I G P P R T

1001 CUGUCAUGAUGAGAUUGAUGGUGAUACUUCUUUACUGGUCACCAUGUUUGGAAUAAUUCUAGGCGGUCCUACCGCUCUAGCUAAAGCUGGCAAGCA
167 V M M R L M V I L L L P T G H H V W N N S M A A P T A L A K A G K Q
SP 1

1101 GGUCGUCAGUCGACGGAUUGGCAAACCCACUCUCUUGCGACCAUAGAUGCUUGAUUUGUAGUGGUGCGUGGCCAAAACAGUGUGGUCCAAGCUAGGA
200 V A D S T D W Q T P L S C D H R C L I C E G A C A K T V W S K L G
SP 2

1201 GUGAAAGCGACUCAUUGGUUUUGUCCAUCAGAUAAAGGAUGCCGUAACCGUUAUCGUGGUAUCUAGUUCAGACAACUUCUGCGGGACGGGCGUUUGUA
233 V K A T Q L V C P S D K D A V T C Y G G S S S D N L L S G T G V C S

1301 GUAAGGUCGAUUUCUCAUCCAAGUUAACCCACUCAUGCCCCACGGCAGCUUGGUCGGAGUGGAACACCGAUUCACUCGAGAGGUGUGUAGGAAUGCA
267 K V D F S S K L T H S C P T A A W S E W N T D L L C R G V V G M E
SP 3 SP 4

1401 CGUUGGGUUGAAGAUUCGGAAAACUCAUUCGUGGCGUGAACCGUUCUUAUGUGUCAAAUUAUUAUUGAGGUCUCAAGAGACUUCUCCAGAGUUG
300 V G F E D I G K L N R G V N A F Y V S N Y L L R S Q E T L P P E L

1501 UCCGUGAUGCAAUUCUUGCAAGACGUUGAGACUCGAGUUAUGAGUUAUUGAGGUUGAGAUCAAAGGCAGAGGUAAGGCCAAAUCUGGUAGGCGUG
333 S V I E I P F E D V E T R S Y E F I E V E I K G R G K A K L G K R E
AB 1

1601 AGUUCGCUUGGAUCCAGAAUCAGGAAAUAUCUGGGCUGAUGAUGACGACGAUUCUUGCCCCCACCUCUAAAGGUGGUAGACGGCAAGUUGUUGGAG
367 F A W I P E S G K Y W A D D D D D S L P P P P K V V D G K M V W S

1701 UUCUGUCAGGAAACUGUUGCGGACGCUUUAACUACCAGCGGGCGGAGGUCAAGGCCUUGCCGCCUUCUCAAUUGCAGGCUACGACUUCAAAGA
400 S A Q E T V A D A L N Y Q R A A G S R P L P P F S T C R L R L Q R
AB 2

1801 AGGAGAAGCAGCCUACAAGAGGAUUGCCUUAAGUUUGGCGAGUCGGCUUGCAAGUUUAGAGUGUGUAGAAAGAUUACUCCAGAUGAAGUCCUAG
433 R R S S L Y K R N A L R F V G Q S A C K F R V C R K I L Q M K S L E

1901 AGCUACUAGGGUUCUCCAGAAUUGUCAGACUUCUUGCGCCGAGGCUCCAAAGCAGAGCUUCACUCCUGCUACGCCAAGCAGGAAAGUUUAAU
467 L L G S S Q N C Q T S P G P S E A P K Q S F T P C Y A K Q E S L I

2001 CCCACUGGAAUCCCAAGGAUUCUUGAAGGAGCUUGUCAAAAACUCCUUGAAAGCAGCUACCCACCUCUCAAUUCUGUUAUCUGUCCUGGAGAAGCUG
500 P L E S Q G I L K E L V K T S L K A A T P P P N P V T V S V E K L

2101 GUGCUUCGACAGCAGUCUACAAGAAGUCUGCAAGAAGAGCGCAAUCGGCGGAAUACAACGAGAAAGUCCAGUCCAGGAGUCCCGCCUCUCCGCCUCGCC
533 V L R Q Q S T R S L Q E E R N R R K S T R K S S P G V P P L P P R L

2201 UCAACCAACAAAGACCUCUUGAGGCACCUAGAGUUAGUUGCUCUGUGUUUACCGAGAGAUUUUUCUACUCAGCGAAGCCGAGAUUUACUUGACGA
567 N Q Q R P P L R H L E L V A L C V T E R L F L L S E A E D L L D E

2301 AUCUCCUGGACUUGGUUCGAGGGUUUGUGUACCCGGUUCGUUGUUUUAAGCAGGAGCCCCACGCCUCCGAAAGGUGAGAGAGGGUCGCUUC
600 S P V D L V R R G L C D P V R L F V K Q E P H A S R X V R E G R F
P I P II

2401 CGCUUGAUUAUCUGUUCGUGGUGAUCAGCUUGUGGAGCGUAUGCUCUUGGGCCUCAAACACAGCUUAAUUCGUGAGUGGGAGCAUUAUCCCU
633 R L I S S V S L V D Q L V E R M L F G P Q N Q L E I A E W E H I P E
P III

2501 CAAAACCCGGCAUGGGCUUAUCGUCGCGGCAACAGGCCAAAAGCUUGUUGACGACUUGAGAGUUAACACUCUCGUUGUCCUGCAGCUGAAGCUGACAU
667 F P G M G L E L R Q Q A K S L F D D L R V K H S R C P A A B A D I
P IV

2601 UUCAGGUUUUGAUUGGUCUGUUCAGGACUGGGAGUUGUGGGCUGAUGUAGAGAUAUAUCGUUCUGGGGGCUUGGCCACAAACUGGCUAAAGCCGCC
700 S G F D W S V Q D W E L W A D V E M I I V L G G F G H K L A K A A

2701 AGGAAUAGGUUUUCGUGUUUCAUGAACUCAGUCUCCAGCUCUCGGAUGGCACACUCAUAGAACAGCAAUUGCCUGGUUAUUGAAGCGGUCAUUGAGCC
733 R N R F S C F M N S V F Q L S D G T L I E Q Q L P G I M F R S L D L
P V

2801 UCCUCACCAACUCCAGAAUUCGUCGCCUUAUGGUCGAGCUCAUUGGCUCGCCUGGUGUAUCGCCAUGGGUGAUGAUCUGUUGAAGGUUGGGUUGAUGG
767 L T N S F I F C L M A E L I G S P W C T A M Q E D S V S G W V D G
P VI

2901 GGCAAAGGACAAGUACAUGAGACUAGACACACGUGCAGAUUAUAAACCCUGUGCAACAACAUUCCGGUCGCUUAUACGAGGUAGAGUUCUGCUCAC
800 A K D K Y M R L D T R A D Y K P C A T T I S G R L Y E V E F C S H
P VII

3001 GUUAUAAGGAAGAUGAUUGGUGUGGCGUCGUGGCCUAAAACUCUGUUUAAAUAUCUUGUCUGAGGGCAAGUGGUUCUUGAGGACUAGAGCGAGAGCU
833 V I E K I D V G W R R G L K L L C L N T C L R A S G S L R T *
P VIII

3101 CAGUUCGUCACCCACUGGCCAGAAUCAGACACUAUGUAGUGGGGAUAUCUCAUCGCCCCACAAAUAUUAUAGAAAUAACAAGUCGCCGCUAUGGC
ORF3 M A

3201 UACCCGUCUGACCAAAAAACAGCUGGCUCAGGCUAUCCAGAACACCCUGCCGAACCCGCGCGUCGUAACGUCGUGCUAACGUCGUGCUCAGGUU
3 T R L T K K Q L A Q A I Q N T L P N P P R R K R R A K R R A A Q V

3301 CCGAAACCGACCCAGGCUUGGGUUAUCAUGGCCCUUAUUGCUCAGGGGACUAUGGUUCGCUUACGUGAACCAUCGCUUAGAACGGCUGGAGGUGUGACAG
36 P K P T Q A G V S M A P I A Q G T M V R L R E P S L R T A G G V T V

3401 UCCUACGCACUCUGAGCUCUCAACUGAGCUCUCAGUGACGAAUGCGAUAGUCAUACCCUCUGAGCUCGUUUGCCCUACACAAUGGGCACCUGGCUUCG
70 L T H S E L S T E L S V T N A I V I T S E L V M P Y T M G T W L R

3501 AGGCGUUGCGCUAAUUGGUCGAAAUUAUAGUUGCUGUCGGUCACAUUAUCGUAUCUCCUCUUGUCCUUAACGACAUCUGGGUCCAUCACAUGGGC
103 G V A A N W S K Y S L L S V T Y T Y L P S C P S T T S G S I H M G

3601 UUCAUAUGAUUAGGUCGACACCCUUCGCUUACCGUUAACCGUUAACCGUUAUCCAAUAGAGGUUAUGUAUCAGGGCAGGUCUGGUCUGGAUCCUCUGGC
136 F Q Y D M A D T L P V S V N Q L S N L R G Y V S G Q V W S G S S G L

3701 UGUGCUACAUAUAGGACAGGAGGUCUUGACACGGCCAAUGCUAUCACGACCACACUGGACGUUGGACAGCUUGGCAAGAAGUGGUUAUCCUUAAGAC
170 C Y I N G T R C L D T A N A I T T T L D V G Q L G K K W Y P F K T

3801 UAGCACAGACUUCACAACGGCUGUUGCGUAAAUGUCAACAUAUGCCACUCCCCUGGUCGCGGAGGCUAAAUAUAGCCAUGCUGGAUGGGUCGAGUUC
203 S T D F T T A V G V N V N I A T P L V P A R L I I A M L D G S S S

3901 ACGGCUUGAGUACUGGACGCCUUAUCGUGUCGUUAUUAUUAACUGAUAGAGCCACUGCUUUGGCCUUGAACACUGAAGGAGUUGUAUUAUAUAC
236 T A V S T G R L Y V S Y T I Q L I E P T A L A L N N *

4001 CUGCACCCUCUCCUUGGGGGAGGGUUGUUCGUUUCACAUGCCACGUGUUGAGGGGAAUGCAGCUUAUCAUCCUCGCGUAGUGAUGGAGCGUAA
4101 UCCAAAAGU 4109

FIG. 2. Nucleotide sequence of SBMV-B. Below the line of sequence is the amino acid sequence of the three ORFs given in single letter code, that in lowercase being the possible readthrough product mentioned in the text. AB1 and 2 indicate the ATP-binding domains, SP1-4 the serine protease domains, and PI-VIII the polymerase domains as discussed in the text.

region immediately upstream of the coat protein ORFs of the other two sobemoviruses.

Coding regions for nonstructural proteins.

The first AUG in the SBMV-B sequence at nucleotide 93 marks the start of the coding region (ORF 1) which terminates at nucleotide 415 and potentially encodes a protein of *M*_r 11,684. This is somewhat smaller than the product(s) of the equivalent ORF of SBMV-C (21K) and of RYMV (18 and 19.5K), but is in accord with the *in vitro* translation product (14K) attributed to this ORF (Mang *et al.*, 1982). Ngon A Yassi *et al.* (1994) suggested that the UGA stop codon of ORF 1 of RYMV might be read

through to give a product with an extra 16 residues. Readthrough of the UAG stop codon of SBMV-B would give a product with an extra 79 amino acids (Fig. 2). However, as *in vitro* translation did not show a product with the expected molecular weight (about 20,500), readthrough is thought unlikely to occur. There is little similarity between the nucleotide or amino acid sequences of this ORF from SBMV-B, SBMV-C, or RYMV (Table 1). Data base searching did not reveal any other proteins similar to the product of this ORF and currently no firm function can be attributed to it. However, most viruses, which like sobemoviruses give a full systemic infection of susceptible host, encode a protein which potentiates cell-to-cell movement. Thus, one possible

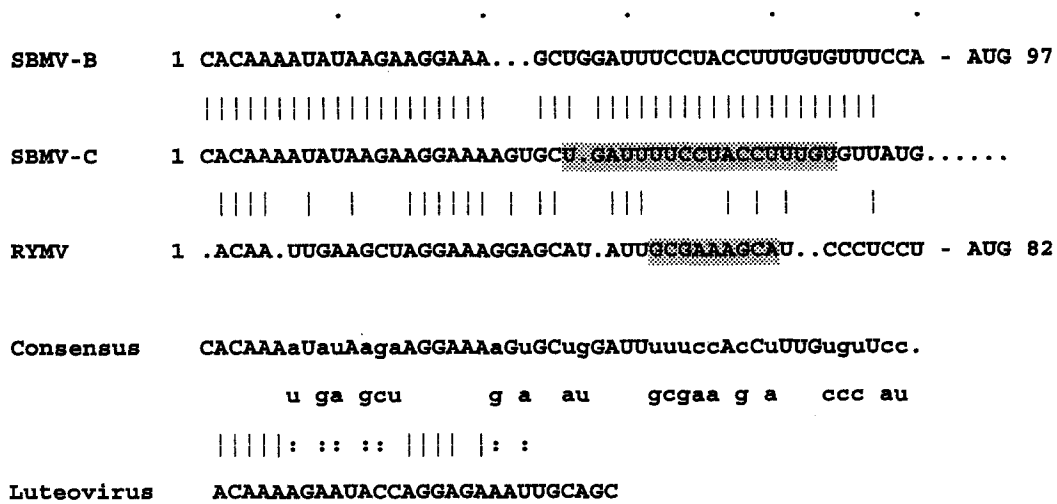


Fig. 3. Alignment of 5' noncoding regions of SBMV-B, SBMV-CP, and RYMV. The consensus sequence is compared with the consensus sequence of the 5' end of luteovirus RNA (Keese *et al.*, 1990). The regions with suggested complementarity to the 3' terminus of 18S rRNA are highlighted.

function of the product of sobemovirus ORF 1 is that it is a movement protein.

The second ORF of SBMV-B extends from nucleotide 505 to nucleotide 3087 and encodes the largest potential protein product of *M_r* 96,481. Overall the product of this ORF has an intermediate similarity to ORF 2 of SBMV-C and RYMV (Table 1) but detailed comparison (Fig. 4) shows that there are regions of high homology. These correspond in the main to motifs found in ORF 2 of SBMV-C and RYMV which are suggestive of the product being a polyprotein containing polymerase and other activities. Toward the 5' end are the motifs characteristic of serine proteases (Bazan and Fletterick, 1990) shown boxed in Fig. 2 (SP 1 - SP 4) with the catalytic serine site at residue 284 (underlined). It has been suggested (Gorbalenya *et al.*, 1988) that this protease could process the polyprotein to functional products.

The C-terminal domain of SBMV-B ORF 2 was identified as the putative RNA-dependent RNA polymerase as

it contained the GDD motif (Fig. 2, underlined). This region shows high homology to those of SBMV-C (83%) and RYMV (71%) (Fig. 4). The C-terminus also exhibited obvious consensus with the proposed RNA-dependent RNA polymerases of the luteoviruses beet western yellows virus and potato leafroll virus (Veidt *et al.*, 1988; Mayo *et al.*, 1989) and the eight conserved motifs (P I - P VIII), defined by Koonin (1991), are identified in Fig. 2. Such similarities have been used to evaluate the taxonomic position of SBMV-C in relation to other positive-strand RNA viruses (Habibi and Symons 1989; Poch *et al.*, 1989; Koonin, 1991; Koonin and Dolja, 1993) and are supported by the sequence of SBMV-B.

Biochemical processes such as replication, recombination and repair, translation, and transcription are coupled to nucleoside triphosphate hydrolysis and viral genomes are thought to have specific domains which have nucleotide binding functions. We could not find any obvious nucleotide-binding domain in the SBMV-B sequence

TABLE 1

NUCLEIC ACID AND AMINO ACID SIMILARITIES BETWEEN CODING REGIONS OF SBMV-B, SBMV-C, AND RYMV

		ORF 1		ORF 2		ORF 3/4 ^a	
		SBMV-C	RYMV	SBMV-C	RYMV	SBMV-C	RYMV
SBMV-B	NA ident ^b	24.8 ^b	24.5	42.6	36.2	51.2	33.2
	Pr ident	25.7	21.7	46.0	37.7	74.2	29.8
	Pr simil	47.6	44.3	62.4	55.0	81.7	46.2
SBMV-C	NA ident	— ^c	35.7	—	49.3	—	39.9
	Pr ident	—	14.7	—	37.4	—	29.2
	Pr simil	—	39.3	—	53.7	—	46.8

^a Comparison of ORF 3 of SBMV-B with ORF 4 of SBMV-C and RYMV.

^b % Nucleic acid identity, protein identity, protein familial similarity (using amino acid grouping of Dayhoff *et al.* (1983)).

^c 100% identity.

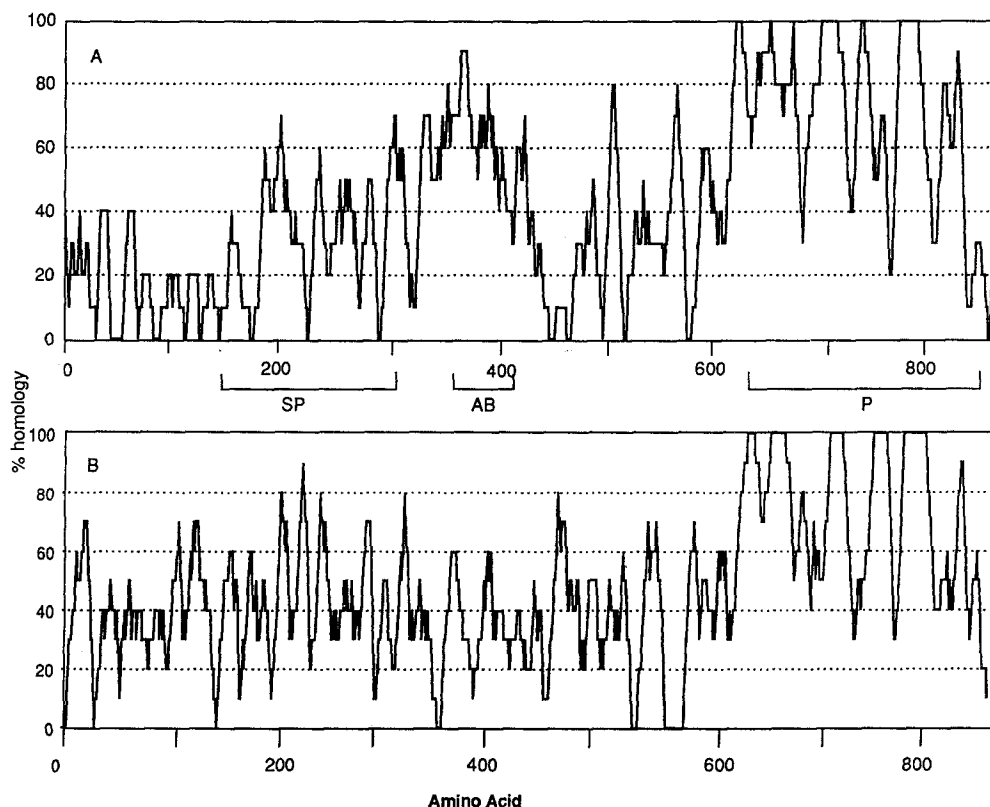


Fig. 4. AHOMOL plot of the direct similarities between the aligned amino acid sequences of ORF 2 of (A) SBMV-B and SBMV-CP and (B) SBMV-B and RYMV. The window size of the plots was 10. SP, AB, and P refer to the regions identified in the legend to Fig. 2.

resembling the hydrophobic stretches proposed as NTP-binding proteins in the potyviruses, comoviruses, and picornaviruses (Gorbalenya *et al.*, 1988; Gorbalenya and Koonin, 1989). Wu *et al.* (1987) proposed a putative ATP-binding domain in SBMV-C based on a loose homology to an ATP-binding consensus sequence derived mainly from an alignment of picornaviral sequences. While this designation is fairly speculative, this motif is well conserved between SBMV-B and SBMV-C sequences in two regions (AB 1 and AB 2) (Fig. 2). On the other hand, possibly significantly, the sequence lacks the consensus motif GKS/T, thought to be the general "signature" sequence of the helicase domain in viral RNA polymerase. However, the absence of a clear helicase domain in SBMV sequences is consistent with the observation that no such viral-encoded function has been identified in viruses with a compact (<5 kb) genome (Gorbalenya and Koonin, 1989).

For the majority of VPg-containing viruses, it has been shown that this protein is encoded within a polyprotein sequence of (Helicase)-VPg-protease-RdRp (super-group I type) (Dolja and Carrington, 1992). Wu *et al.* (1987) identified a putative VPg sequence at residues 500-519 of SBMV-C, based on a loose homology to several known sequences. As noted by Gorbalenya *et al.* (1988) the location of the VPg in this region is inconsistent with the

proposed model of the genome organization of ORF 2 taking into account the putative identification of the protease domain as described earlier. Should the VPg cistron be located at the NH₂-terminal end as proposed, the lack of homology in this region between the two strains may indicate that the viruses possess different species of the protein. The dissimilar sizes of the VPg protein between the two strains (Mang *et al.*, 1982) may support this assumption.

Coat protein

ORF 3, from nucleotide 3195 to nucleotide 3981, encodes a putative protein of M_r 28,107 and its allocation as the coat protein cistron is supported by the close similarity of its amino acid content with that published for isolated SBMV-B capsid proteins (Tremaine, 1966; Ghabrial *et al.*, 1967).

Detailed structural information based on crystallographic studies at resolutions of up to 2.9 Å is available for the coat protein subunit of SBMV-C (Abad-Zapatero *et al.*, 1980; Rossmann, 1984; Silva and Rossmann, 1987). The SBMV particle consists of 180 coat protein subunits arranged in an icosahedral $T = 3$ geometry (Johnson *et al.*, 1976; Silva and Rossmann, 1987) with each protein subunit comprising two domains, the R (Random) and

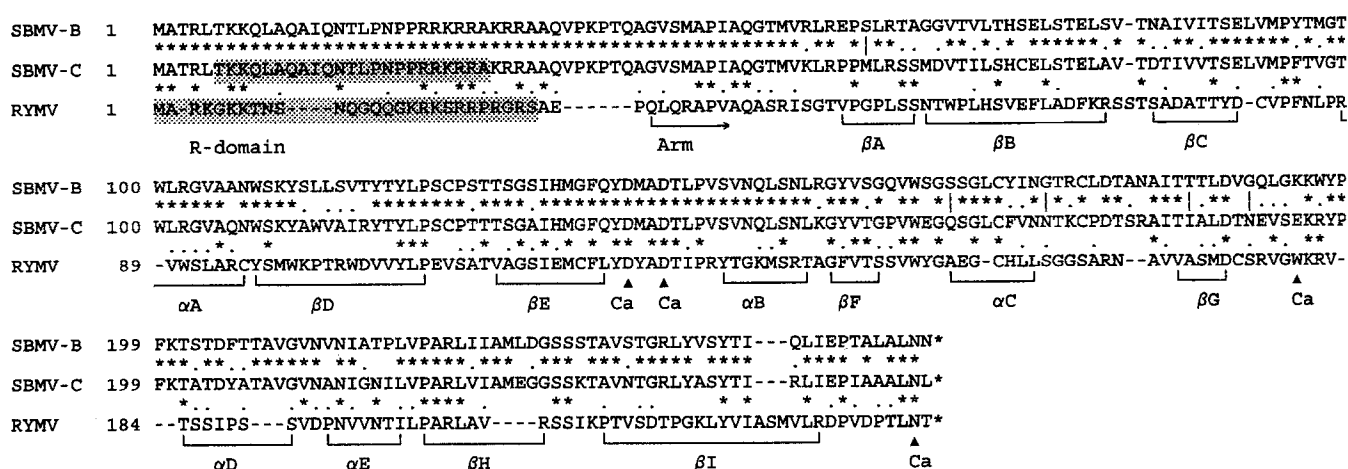


Fig. 5. Alignment of coat protein sequences of SBMV-B, SBMV-CP, and RYMV. Asterisks indicate direct similarity and periods familial similarity. The tertiary structure features as determined for SBMV-CP coat protein (see Rossmann, 1984) are shown as are the amino acids implicated in calcium binding (Ca). The highlighted residues in the SBMV-CP and RYMV sequences are those suggested to be similar to the bipartite nuclear targeting motif (Dingwall and Laskey, 1991).

the S (Shell or surface) domain connected by an "arm." The R domain, made up of the N-terminal portion of the polypeptide chain, is rich in arginine, lysine, proline, and glutamine and penetrates into the interior of the particle where it interacts with the RNA (Abad-Zapatero *et al.*, 1980; Tremaine *et al.*, 1981, 1982; Hermodson *et al.*, 1982; Kruse *et al.*, 1982). The S domain consists of a core made up of a bundle of eight stranded anti-parallel β sheets (β barrel) together with five α helical regions. The interactions between the protein subunits and the RNA also involve the basic amino acids on the inner surface of the S domain (Hermodson *et al.*, 1982).

Comparison of the amino acid sequences predicted from SBMV-B ORF 3 and the corresponding ORF 4 of the other sobemoviruses (Table 1) shows that it is more closely related to SBMV-C than to RYMV. When the amino acid sequence of SBMV-B ORF 3 is compared as far as tertiary structure considerations are concerned with that of SBMV-C coat protein there are many similarities but also some differences in potentially structurally important residues that maintain the integrity of the capsid.

Tremaine *et al.* (1981) showed that the NH_2 -terminal polypeptide binds RNA, DNA, and sodium dextran sulfate. Comparison of the 66 residues which compose the NH_2 -terminal arm of the two strains of SBMV upstream of the first β -barrel (Fig. 5) shows that, while there are differences between the two sequences (89.4% identity), the residues in this region remain primarily basic, a factor thought to be important in the interaction between protein and RNA. Significantly, the trypsin cleavage site at Arg 61 (Erickson and Rossmann, 1982) is also conserved. Rossmann (1984) proposed several possible docking sites for specific RNA-protein interaction in SBMV-C, including a polar polypeptide (188 to 196) which forms a large bulge in strand β G and protrudes into the RNA.

Interestingly there is little conservation between SBMV-B and SBMV-C in this region, which may suggest subtle differences in the strategies of their protein-RNA interactions.

Factors affecting stabilization of the capsid of SBMV represent a common link with other members of the sobemovirus group and have been extensively studied (Hull, 1977, 1978; Sehgal *et al.*, 1979; Abdel-Meguid *et al.*, 1981; Kruse *et al.*, 1982; Brisco *et al.*, 1986; Sehgal, 1990). The function of metal ions, in particular of calcium and magnesium, in relation to the assembly and stabilization of SBMV protein shells has been investigated in both the bean and cowpea strains (Hsu *et al.*, 1976; Hull, 1977). The major calcium-binding site that lies in the quasi-threefold axis between the three subunits (A, B, C) of the SBMV-C capsid has been identified at Glu 194 (see Rossmann, 1984) (Fig. 5). The amino acid sequence of SBMV-B shows the substitution of a lysine residue at the equivalent site which could indicate an alternative mode for subunit interaction. It is also interesting to note that, while other residues interacting with Glu-194 in SBMV-C may mediate subunit contacts and are important in Ca^{2+} binding (Silva and Rossmann, 1987), no metal ion has been detected in the refined crystallized virus at this site. The minor calcium-binding sites proposed for SBMV-C at Asp 138 and Asp 141 and the main chain carbonyls at residues 199 and 259 (Rossmann, 1984) (Fig. 5) are, on the other hand, conserved in the SBMV-B sequence. The proposed magnesium-binding sites at residues His 132, Glu 229, and Glu 77 (Rossmann, 1984) (Fig. 5) are also conserved in both strains.

Ngon A Yassi *et al.* (1994) noted that the sequence of the N-terminal regions of RYMV and SBMV-CP coat proteins resembles a bipartite nuclear targeting motif (Dingwall and Laskey, 1991). There is full homology be-

tween this region of SBMV-C and that of SBMV-B (Fig. 5), which could relate to the occurrence of SBMV-B in the nucleus (de Zoeten and Gaard, 1969).

Translational strategies

The putative genome organization of the SBMV-B genome (Fig. 1) raises interesting questions in terms of its overall translational strategies, which are also relevant to the other sobemoviruses. ORF 3, the capsid protein cistron, is thought to be expressed via a subgenomic mRNA (Ghosh *et al.*, 1981; Mang *et al.*, 1982). The expression of ORF2 is, however, less certain as there is no evidence for a subgenomic mRNA associated with it which would suggest that it is a "closed" ORF. There are three possibilities for its expression, frameshift, read through of a weak stop codon, or internal initiation. Frameshift would be feasible for SBMV-C and, after read through of a weak stop codon (indicated by amino acids in lowercase letters in Fig. 2), in SBMV-B. However, RYMV ORFs 1 and 2 are in the same frame separated by an amber and ochre stop codon (Ngon A Yassi *et al.*, 1994), which are unlikely to be both read through at a significant level. The AUG codon for ORF 1 of each of the three sobemoviruses is the first in the sequence and is in a poor context for initiation of translation (Lutcke *et al.*, 1987; Kozak, 1989). In SBMV-B and SBMV-C there is only one other AUG in any reading frame before that of ORF2 and that is also in a poor context; in RYMV there are no AUGs between those of ORFs 1 and 2. For all three sobemoviruses the AUG of ORF 2 is in a favorable context. This raises the possibility is that the expression of ORF2 may involve a mechanism of internal initiation which would resemble the expression strategy suggested for some of the cistrons of the coronaviruses (reviewed by Spann *et al.*, 1988).

Classification of SBMV

Despite their differing host specificities SBMV-B and SBMV-C are classified as two strains of the same virus on the basis of morphological and physicochemical similarities and their serological cross-reactivity (Shepherd and Fulton, 1962). Detailed structural studies, including extensive studies on their virion assembly and disassembly under various physicochemical conditions (Brisco *et al.*, 1985a; 1986; Shields *et al.*, 1989) again highlighted the close similarity of the two strains. However, as with the earlier immunological evidence, these data related primarily to the properties of the virion proteins. The availability of the complete nucleotide sequence of SBMV-B raises the interesting question as to whether it and SBMV-C are truly strains of the same virus or whether or not they should be reclassified as unique viruses within the sobemovirus group.

Various molecular criteria, such as 3' noncoding se-

quences (Frenkel *et al.*, 1989) or coat protein sequences (Shukla and Ward, 1989), have been used to classify viruses either as distinct entities or as strains. On the basis of 3' noncoding sequence SBMV-B and SBMV-C would be recognized as distinct viruses but on coat protein sequence they would probably be considered as strains. One of the main functions of classification is communication and entities distinguished as distinct viruses indicate that the differences are distinct. The molecular differences between SBMV-B and SBMV-C are notable and indicate that there might be significant functional differences especially between the ORF 1 products. Therefore, we suggest that consideration should be given to separating SBMV-B and SBMV-C as distinct viruses.

ACKNOWLEDGMENTS

Y.O. was supported by The University of Malaya, Kuala Lumpur, the Public Services Department, the Government of Malaysia, and the British High Commissioners Award Scheme. We thank Mike Harvey for the preparation of Fig. 1. SBMV was held under MAFF License No. PHF 1419A/951/24 and the genetic manipulation work was conducted under MAFF License No. 1185B/15/111.

REFERENCES

- ABAD-ZAPATERO, C., ABDEL MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMAN, M. G., SUCK, D., and TSUKUHARA, T. (1980). Structure of southern bean mosaic virus at 2.8 angstrom resolution. *Nature* **286**, 33-39.
- ABDEL-MEGUID, S. S., YAMANE, T., FUFUYAMA, K., and ROSSMAN, M. G. (1981). The location of calcium ions in southern bean mosaic virus. *Virology* **114**, 81-85.
- BAZAN, J. F., and FLETTERICK, R. J. (1990). Structural and catalytic models of trypsin-like viral proteases. *Semin. Virol.* **1**, 311-322.
- BECCERA, S. P., ROSE, J. A., HARDY, M., BAROUDY, B. M., and ANDERSON, C. W. (1985). Direct mapping of adeno-associated virus capsid protein B and C; a possible AUC initiation codon. *Proc. Natl. Acad. Sci. USA* **80**, 7919-7923.
- BRISCO, M. J., HULL, R., and WILSON, T. M. A. (1985a). The effect of extraction protocol on the yield, purity and translation products of RNA from an isometric plant virus. *J. Virol. Methods* **10**, 195-202.
- BRISCO, M. J., HULL, R., and WILSON, T. M. A. (1985b). Southern bean mosaic virus-specific proteins are synthesised in an *in vitro* system supplemented with intact, treated virions. *Virology* **143**, 392-398.
- BRISCO, M. J., HANIFF, C., HULL, R., WILSON, T. M. A., and SATTELLE, D. B. (1986). The kinetics of swelling of southern bean mosaic virus: A study using photon correlation spectroscopy. *Virology* **148**, 210-217.
- DAYHOFF, M. O., BARKER, W. C., and HUNT, L. T. (1983). Establishing homologies in protein sequences. *Methods Enzymol.* **91**, 524-549.
- DE BORDE, D. C., NAEVE, C. W., HERLOCHER, M. L., and MAASSAB, H. F. (1986). Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotide transferase. *Anal. Biochem.* **157**, 275-282.
- DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.
- DE ZOETEN, G. A., and GAARD, G. (1969). Possibilities for inter- and intracellular translocation of some icosahedral plant viruses. *J. Cell Biol.* **40**, 814-823.
- DINGWALL, C., and LASKEY, R. A. (1991). Nuclear targeting sequence — A consensus. *Trends Biochem. Sci.* **16**, 478-481.

- DOLJA, V. V., and CARRINGTON, J. C. (1992). Evolution of positive-strand RNA viruses. *Semin. Virol.* **3**, 315–326.
- ERICKSON, J. W., and ROSSMANN, M. G. (1982). Assembly and crystallization of a T = 1 icosahedral particle from trypanized southern bean mosaic virus coat protein. *Virology* **116**, 128–136.
- FRENKEL, M. J., WARD, C. W., and SHUKLA, D. D. (1989). The use of the 3' non-coding sequence in the taxonomy of potyviruses: Application to watermelon mosaic virus and soybean mosaic virus-N. *J. Gen. Virol.* **70**, 2775–2783.
- GHABRIAL, S. A., SHEPHERD, R. J., and GROGAN, R. G. (1967). Chemical properties of three strains of southern bean mosaic virus. *Virology* **33**, 17–25.
- GHOSH, A., DASGUPTA, R., SALERNO-RIFE, T., RUTGERS, T., and KAESBERG, P. (1979). Southern bean mosaic virus has a 5'-linked protein but lacks 3' terminal poly(A). *Nucleic Acids Res.* **7**, 2137–2146.
- GHOSH, A., RUTGERS, T., MANG, K. Q., and KAESBERG, P. (1981). Characterization of the coat protein mRNA of southern bean mosaic virus and its relationship to the genomic RNA. *J. Virol.* **39**, 87–92.
- GORBALENYA, A. E., and KOONIN, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* **17**, 8413–8440.
- GORBALENYA, A. E., KOONIN, E. V., BINOV V. M., and DONCHENKO, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornoviruses. *FEBS Lett.* **236**, 287–290.
- GORDON, K., FÜTTERER, J., and HOHN, T. (1992). Efficient initiation of translation at non-AUG triplets in plant cells. *Plant J.* **2**, 809–813.
- GROGAN, R. G., and KIMBLE, K. A. (1964). The relationship of severe bean mosaic from Mexico to southern bean mosaic virus and its related strain from cowpea. *Phytopathology* **54**, 75–78.
- GUBLER, U., and HOFFMAN, B. J. (1983). A simple and effective way for generating cDNA libraries. *Gene* **25**, 263–269.
- HABILI, N., and SYMONS, R. H. (1989). Evolutionary relationships between luteoviruses and other plant RNA viruses. *Nucleic Acids Res.* **17**, 4543–4555.
- HERMODSON, M. A., ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., PINDEK, S., ROSSMAN, M. G., and TREMAINE, J. H. (1982). Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three dimensional structure of the virus. *Virology* **119**, 133–149.
- HSU, C. H., SEHGAL, O. P., and PICKETT, E. E. (1976). Stabilizing effect of divalent metal ions on virions of southern bean mosaic virus. *Virology* **69**, 587–595.
- HULL, R. (1977). The stabilization of the particles of turnip rosette virus and of other members of the southern bean mosaic virus group. *Virology* **79**, 58–66.
- HULL, R. (1978). The stabilization of the particles of turnip rosette virus. III. Divalent cations. *Virology* **89**, 418–422.
- HULL, R. (1985). Purification, biophysical and biochemical characterization of viruses with special reference to plant viruses. In "Virology: A Practical Approach" (B. W. J. Mahy, Ed.), pp. 1–24. IRL Press, Oxford.
- HULL, R. (1988). Southern bean mosaic virus. In "The Plant Viruses" (R. Francki, Ed.), Vol. 4, pp. 113–146. Plenum, New York.
- HULL, R., SADLER, J., and LONGSTAFF, M. (1986). The sequence of carnation etched ring virus DNA: Comparison with cauliflower mosaic virus and retroviruses. *EMBO J.* **5**, 3083–3090.
- JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT I., and ROSSMANN, M. G. (1976). The structure of southern bean mosaic virus at 2.5 angstrom resolution. *Virology* **75**, 394–400.
- KEESE, P., MARTIN, R. R., KAWCHUK, L. M., WATERHOUSE, P. M., and GERLACH, W. L. (1990). Nucleotide sequences of an Australian and a Canadian isolate of potato leafroll luteovirus and their relationships with two European isolates. *J. Gen. Virol.* **71**, 719–724.
- KOONIN, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2206.
- KOONIN, E. V., and DOLJA, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Revs. Biochem. Mol. Biol.* **28**, 375–430.
- KOZAK, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* **108**, 229–241.
- KRUSE, J., TIMMINS, P. A., and WITZ, J. (1982). A neutron scattering study of the structure of compact and swollen forms of southern bean mosaic virus. *Virology* **119**, 42–50.
- LAMPTEY, P. N. L., and HAMILTON, R. I. (1974). A new cowpea strain of southern bean mosaic virus from Ghana. *Phytopathology* **64**, 1110–1104.
- LUTCKE, H. A., CHOW, K. C., MICKEL, F. S., MOSS, K. A., KERN, H. F., and SCHEELLE, G. A. (1987). Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6**, 43–48.
- MANG, K. Q., GHOSH, A., and KAESBERG, P. (1982). A comparative study of the cowpea and bean strains of southern bean mosaic virus. *Virology* **116**, 264–274.
- MAYO, M. A., ROBINSON, D. J., JOLLY C. A., and HYMAN, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. Gen. Virol.* **70**, 1037–1051.
- MURRAY, V. (1989). Improved double stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**, 8889.
- NGON A YASSI, M., RITZENHALER, C., BRUGIDOU, C., FAUQUET, C., and BEACHY, R. N. (1994). Nucleotide sequence and genome characterization of rice yellow mottle virus RNA. *J. Gen. Virol.* **75**, 249–257.
- PARRY-SMITH, D. J., and ATTWOOD, T. K. (1990). SOMAP: a novel approach to multiple sequence alignment. *Comput. Appl. Biosci.* **7**, 233–235.
- POCH, O., SAUVEGET, I., DELARUE, M., and TORDO, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874.
- PRATS, A. C., DEBILLY, G., WANG, P., and DARLIX, J. L. (1989). CUG initiation codon used for the synthesis of cell surface antigen coded by the murine leukaemia virus. *J. Mol. Biol.* **205**, 363–372.
- ROSSMANN, M. G. (1984). The structure of southern bean mosaic virus. In "Biological Macromolecules and Assemblies" (A. McPherson, Ed.), Vol. 1, p. 46. Wiley, New York.
- RUTGERS, T., SALERNO-RIFE, T., and KAESBERG, P. (1980). Messenger RNA for the coat protein of southern bean mosaic virus. *Virology* **104**, 506–509.
- SALERNO-RIFE, T., RUTGERS, T., and KAESBERG, P. (1980). Translation of southern bean mosaic virus RNA in wheat germ embryo and rabbit reticulocyte extracts. *J. Virol.* **34**, 51–58.
- SAMBROOK, J., FRITSCH, E. F., and MANIATIS, T. (1989). "Molecular Cloning: A Laboratory Manual." 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- SEHGAL, O. P. (1981). Southern bean mosaic virus group. In "Handbook of Plant Virus Infections and Diagnosis" (P. Kurstak, Ed.), pp. 91–121. Elsevier, North Holland.
- SEHGAL, O. P. (1990). Structural transitions in southern bean mosaic virus and their correlation with infectivity and ribonuclease activity. *J. Phytopathol.* **129**, 69–82.
- SEHGAL, O. P., HSU, C. H., WHITE, J. H., and MAI VAN (1979). Enzymic sensitivity of conformationally altered virions of southern bean mosaic virus. *Phytop. Z.* **95**, 167–177.
- SHEPHERD, R. J., and FULTON, R. W. (1962). Identity of a seed-borne virus of cowpea. *Phytopathology* **52**, 489–493.
- SHIELDS, S., BRISCO, M. J., WILSON, T. M. A., and HULL, R. (1989). Southern bean mosaic virus RNA remains associated with swollen virions during translation in wheat germ cell-free extracts. *Virology* **171**, 602–606.
- SHUKLA, D. D., and WARD, C. W. (1989). Identification and classification of potyvirus on the basis of coat protein sequence data and serology. *Arch. Virol.* **106**, 171–200.

- SILVA, A. M., and ROSSMANN, M. G. (1987). Refined structure of southern bean mosaic virus at 2.9Å resolution. *J. Mol. Biol.* **197**, 69–87.
- SPAAN, W., CAVANAGH, D., and HORZINEK, M. (1988). Coronavirus: Structure and genome expression. *J. Gen. Virol.* **69**, 2939–2952.
- TREMAINE, J. H. (1966). The amino acid and nucleotide composition of the bean and cowpea strains of southern bean mosaic virus. *Virology* **30**, 348–354.
- TREMAINE, J. H., and HAMILTON, R. I. (1983). Southern bean mosaic virus. "CMI/AAB Descriptions of Plant Viruses," No. 274.
- TREMAINE, J. H., RONALD, W. P., and KELLY, E. M. (1981). Comparison of highly basic cyanogen bromide peptides from strains of southern bean mosaic virus. *Can. J. Microbiol.* **27**, 654–663.
- TREMAINE, J. H., RONALD, W. P., and KELLEY, E. M. (1982). Intermediates of the assembly and disassembly of southern bean mosaic virus. *Virology*, **118**, 35–44.
- VEERISETTY, V., and SEHGAL, O. P. (1980). Proteinase K-sensitive factor essential for the infectivity of southern bean mosaic virus. *Phytopathology* **70**, 282–284.
- VEIDT, I., LTO, H., LEISER, M., SCHEIDECKER, D., GUILLEY, H., RICHARDS K., and JONARD, G. (1988). Nucleotide equence of beet western yellows virus RNA. *Nucleic Acids Res.* **16**, 9917–9932.
- WU, S., RINEHART C. A., and KAESBERG, P. (1987). Sequence and organization of southern bean mosaic virus genomic RNA. *Virology* **161**, 73–80.
- YERKES, W. D., and PATINO, G. (1960). The severe bean mosaic virus, a new bean virus from Mexico. *Phytopathology* **50**, 334–338.
- ZIMMERN, D. (1975). The 5' end group of tobacco mosaic virus RNA is m⁷G^{5'}ppp^{5'}Gp. *Nucleic Acids Res.* **2**, 1189–1201.
- ZUKER, M., and STEIGLER, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133–148.