THE NUCLEOTIDE SEQUENCE OF THE 3' TERMINUS OF SOYBEAN MOSAIC VIRUS

by

Paul L. Gunyuzlu

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Plant Pathology, Physiology and Weed Science

APPROVED:

S. A. Tolin, Chairman

L. D. Moore, Dept. Head

J. L. Johnson

R. C. Bates

G. L. Lacy

April 18, 1987

Blacksburg, VA

THE NUCLEOTIDE SEQUENCE OF THE 3' TERMINUS OF SOYBEAN MOSAIC VIRUS

by

Paul L. Gunyuzlu

Committee Chairman: Dr. Sue A. Tolin

Plant Pathology

(ABSTRACT)

The nucleotide sequence of the 3' terminus of soybean mosaic virus (SMV) VA/G1 RNA has been determined by dideoxynucleotide sequencing of oligo(dT) primed cDNA cloned into a pUC19 cloning vector via EcoR1 linkers. One recombinant plasmid, pSMV49, identified by colony and dot-blot hybridization to ^{125}I-SMV RNA, contained an insert of 1443 nucleotides and had an open reading frame of 1119 nucleotides terminating 224 nucleotides from the 3' terminal poly(A) tract. The coat protein cistron identified by a glutamine:serine dipeptide cleavage site 792 nucleotides upstream from the termination sequence could potentially code for a 29.8 kDa protein. The amino acid sequence predicted from the nucleotide sequence of this cistron contains regions identical to other potyvirus coat proteins.

Acknowledgements

I would like to thank my major advisor, Dr. Sue Tolin, for providing me with the opportunity to undertake this project and for the patience in completing it.

I extend my deep appreciation and thanks to Dr. John Johnson for allowing me to work in his laboratory and for his technical expertise in completing the research.

I would also like to thank Dr. Robert Bates and Dr. George Lacy for their helpful advise and support for the past two years.

I extend a special thanks to Dr. Laurence Moore for supporting me financially in times of great need.

Finally, I thank my wife, and my two children, and without whose support and love I could not have finished my degree. I dedicate all of this work to them.

Table of Contents

Table of Figures

Biochemistry of the Soybean Mosaic Virus Genome

A Literature Review

Chemical and Structural Properties

Ribonucleic acid (RNA) was demonstrated to be the nucleic acid of soybean mosaic virus (SMV) (Hill & Benner, 1980). The nucleic acid comprises 5.32% of the entire virion mass and consists of 29.9% adenylic acid, 24.4% guanylic acid, 14.9% cytidylic acid and 30.9% uridylic acid as determined by nucleotide analysis of alkaline hydrolyzed RNA.

The molecular weight of SMV RNA was estimated by polyacrylamide electrophoresis using formaldehyde-denatured RNA (Hill & Benner, 1980) and glyoxal-denatured RNA (Vance & Beachy, 1984a) and found to be 3.25 to 3.3 Md, respectfully. This compares well with molecular weight values previously reported for three other Potyviruses; turnip mosaic virus (TuMV); 3.11 Md, tobacco etch virus (TEV); 3.18 Md and maize dwarf mosaic virus (MDMV); 2.98 Md (Hill & Benner, 1976). These similarities in genome sizes indicate that the genome of the known potyviruses contain similar coding capacities.

3′ Polyadenylated Region

Polyadenylation at the 3′ terminus of SMV RNA has been identified by hybridization with ^{3}H-labeled polyuridine and binding in an oligo(dT) column (Vance & Beachy, 1984a). The

poly (A) tail is similar to that described for eukaryotic mRNA's (Kates, 1970) and has been identified in all known potyviruses to date.

A polyadenylated 3' terminus has also been found in poliovirus (Koch & Koch, 1985) and it is similar in length to the poly (A) tail described for SMV. In general; the poly (A) tail has been suggested to be involved in a) processing of precursor RNA to mRNA, b) transport of mRNA from the nucleus to the cytoplasm, c) efficient translation of mRNA, d) protection of mRNA against hydrolysis by exonucleases and e) important for RNA infectivity.

TEV was also determined to contain a polyadenylated 3' terminus by binding in a polyuridine column (Hari et al, 1979). Hari et al (1979) found that this region is variable, with some RNA's containing no polyadenylated region at all or having a variable length. Inoculating the different RNA species onto Chenopodium amaranticolor indicated that both polyadenylated and non-polyadenylated RNA are equally infectious.

The 3' terminus of TVMV (Domier et al, 1986), PeMV (Dougherty et al, 1985) and TEV (Allison et al, 1985) have a poly (A) tail as determined by nucleotide sequence analysis. The genomic RNA's of TEV and PeMV do not contain the [AAUAAA] polyadenylation signal which occurs in cellular

mRNA's (Tucker et al, 1979). Domier et al (1986) state that the RNA of TVMV contains a potential polyadenylation signal 94 nucleotides from the poly (A) site. However, reported polyadenylation signal sequences have been found within 25 nucleotides of the start site of poly (A) addition and this sequence is found at seven other random positions in the TVMV genome. Perhaps the poly (A) tail of potyviruses is transcribed directly off of the RNA template instead of being synthesized post-transcriptionally, as demonstrated with poliovirus (Koch & Koch, 1985).

5' Viral Protein, genome linked.

SMV RNA will aggregate during purification from virions. The aggregates of SMV RNA are converted to monomers only by Proteinase K treatment (Luciano et al, 1984), providing indirect evidence for the presence of a VPg (Viral Protein, genome linked).

VPg's are small proteins, covalently-linked with viral RNA's (Winner, 1982). The proteins are basic, acid soluable and have no helical content. The covalent linkage is a phosphodiester bond between the third amino acid of VPg (tyrosine) and the 5' terminal nucleotide (uridine) of the RNA. The linkage is very energy-rich although the mechanism by which it is formed is not known. VPg is covalently bound to the 5' terminus of all newly synthesized plus and minus

sense polioivirus RNA, the replicative form RNA and all nascent strands of the replicative intermediate suggesting a function as a primer for initiation of viral RNA synthesis (Koch & Koch, 1985).

A covalently-bound 6 Kd protein is associated with the 5' terminus of TEV (Hari, 1981). In this study, the protein was iodinated and then either treated with Proteinase K or treated with phenol; radioactivity was only lost after Proteinase K treatment. Loss of this protein does not result in a loss of infectivity indicating that VPg is not needed during the early stages of TEV infection.

Genetic Map and Translational Strategy of Potyvirus Genomes.

A genetic map of the potyviral genome was constructed from in vitro translation products of TEV and PeMV synthesized in a rabbit reticulocyte lysate system (Dougherty & Hiebert, 1980) (fig. 1). The map is based on isolating in vitro translation proteins representitive of the entire genome and then immunoprecipitating with one or more antibodies made from the four known viral proteins; nuclear inclusion protein 1, nuclear inclusion protein 2, cytoplasmic inclusion protein, and capsid protein. The location of proteins in reference to each other was determined by immunoprecipitating overlapping proteins with two or more

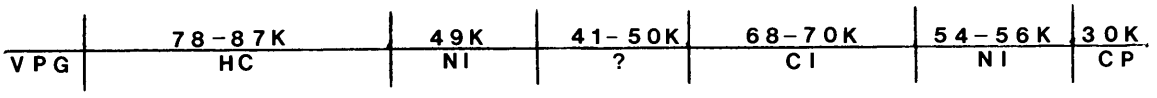

Fig. 1. The putative genetic map of the potyvirus genome based on in vitro translation studies of TEV and PeMV (Dougherty & Hiebert, 1980). VPG = viral protein genome-linked, HC = helper component protein, NI = nuclear inclusion protein, CI = cytoplasmic inclusion protein, CP = capsid protein and ? = unknown protein.

specific antibodies. Cell free translation of TVMV RNA in a rabbit reticulocyte lysate identified the helper component protein as a viral encoded protein and it was assigned to the 5′ terminus of the genetic map (Hellman et al, 1983). These studies were also based on immunoprecipitating overlapping in vitro translation proteins but with antibodies made from the four known viral products listed above along with the helper component protein. The combined information from both studies has resulted in a revised potyviral genetic map (fig. 1). Functions for the 50 Kd, nuclear inclusion, and cytoplasmic inclusion proteins have not been identified to date.

The genetic map of SMV appears comparable to the potyviral general map. Vance & Beachy (1984b) immunoprecipitated high molecular weight in vitro translation proteins with antibodies to the capsid, nuclear inclusion, and cytoplasmic inclusion proteins to determine the genome organization. The results are similar to the general potyviral genetic map. In this report, the capsid protein gene was thought to be on the 5′ terminus of the genome since its the most abundantly synthesized translation protein. More recent studies, however, indicate that the gene is on the 3′ terminus (R. N. Beachy, personnel communication). The presence of high molecular weight in vitro translation proteins that

immunoprecipitated with two or more virus specific antibodies provided indirect evidence that proteolytic processing of a genomic length polyprotein may occurr with SMV (Vance & Beachy, 1984b).

Evidence for proteolytic processing of a genomic-length polyprotein may be found in studies by Yeh et al (1985) with papaya ringspot virus (PRV), another potyvirus. A 330 Kd polyprotein was identified which corresponds to the entire coding capacity of the viral genome. This is the largest polyprotein isolated in a cell-free translation system for potyviruses and was attributed to the absence of dithiothreitol. A time course study with the addition of dithiothreitol revealed that the 330 Kd protein diminished with time while smaller proteins appeared. Yeh et al (1985) proposed that proteolytic processing was a mechanism of potyviral replication and that cleavage may be dependant on a reducing agent. Similarly, Pelham (1979) showed that primary cleavage of the middle (M) RNA proteins of CPMV by a protease encoded by the bottom (B) RNA required an ATP and an easily oxidized thiol group. If dithiothreitol was omitted from the system, the proteins were not cleaved. Perhaps the absence of a reducing agent in both cell-free translation systems inhibits protease activity by altering the primary, secondary or tertiary structure of the protein.

Nucleotide Sequence Analysis of the Potyviral Genome.

A 2.32 Kb cDNA clone of TEV was sequenced to determine if the gene for the capsid protein (30 Kd) contained an initiation codon or whether it was translated as part of a large polyprotein (Allison et al, 1985). There was an open reading frame of 2135 nucleotides beginning at the 5' terminus and ending with an opal termination codon. This was followed by a 189 nucleotide untranslated region containing 14 termination codons in the three possible reading frames and ending in a polyadenylate sequence, indicating the 3' terminus. The open reading frame coded for an 81 Kd protein.

Allison et al (1985) proposed that the capsid protein matured from a larger polyprotein with post-translational cleavage occuring between a glutamine:glycine dipeptide bond. This would result in the production of a 30 Kd capsid protein which is in agreement with Dougherty et al (1980). Similarly, poliovirus polyproteins are post-translationally cleaved into functional proteins by the protease NCVP7c (Koch & Koch, 1985). This enzyme cleaves specifically at glutamine:glycine dipeptide bonds, which is the putative site for cleavage of TEV capsid protein. Franssen et al (1984), showed that both capsid proteins of CPMV (VP37 & VP23) were post-transcriptionally cleaved from their 60 Kd protein precursor by a viral encoded protease and that

cleavage also occured between a glutamine:glycine dipeptide bond.

The complete nucleotide sequence of the TVMV genome was recently obtained by sequencing overlapping cDNA clones (Domier et al, 1986). The length of the genome was determined to be 9471 bases containing a 5′ untranslated region of 205 bases, an open reading frame of 9015 bases encoding a polyprotein of 340 Kd, a 3′ untranslated region of 250 bases followed by a polyadenylate sequence at the 3′ terminus. Domier et al (1986) found stem loop structures in both the 5′ and 3′ untranslated regions. A sequence of [AGGCCAUG] found at the 5′ terminus is similar to the consensus sequence of most eukaryotic mRNA (Kozak, 1984).

Cleavage sites for all known TVMV encoded proteins were proposed based on the approximate positions of the TVMV cistrons as determined by in vitro translation studies (Hellman et al, 1985) and searching through the nucleotide sequence for glutamine:glycine, glutamine:serine and glutamine:alanine potential cleavage sites. These dipeptide bonds are protease cleavage sites for poliovirus (Koch & Koch, 1985), CPMV (Goldbach, 1986) and all potyviruses studied to date (Allison et al, 1985a; Dougherty et al,1985). An amino acid consensus sequence of Val-(Arg or Lys)-Phe-Gln//(Gly,Ser or Ala) was found on the N-terminal

side of the proposed cleavage sites of the polyprotein, with cleavage occuring at the carboxyl side of glutamine in every case. Examination of the nucleotide sequence of the capsid protein gene revealed that proteolytic cleavage occured between a glutamine:serine dipeptide bond, which is in agreement with the proposed cleavage site of the capsid protein of TEV-HAT (Allison et al, 1985b).

The nucleotide sequence of PeMV was also determined but with different results than had been observed with other potyviruses. A discreet open reading frame of 933 nucleotides was identified for the capsid protein gene which started with an initiation codon and ended with an opal termination codon (Dougherty et al, 1985). It has been proposed that the capsid protein gene of PeMV is translated from a sub-genomic mRNA transcribed from the genome, demonstrating an alternative strategy for potyviral translation. However, there is still a post-translational cleavage event. A 37 Kd protein is translated from the sub-genomic mRNA and then processed to a 30 Kd mature capsid protein. Cleavage occurs between a glutamine:alanine dipeptide bond resulting in the removal of N-terminal amino acids. PeMV is the only potyvirus proposed to date that synthesizes a sub-genomic mRNA coding for the expression of capsid protein. Since more capsid protein is needed for

virus replication in relation to other viral proteins, this strategy may allow for the production of proteins in an appropriate proportion.

In summary, the sequence analysis of viruses within the potyvirus group has revealed two primary features of translational strategy. First, post-translational cleavage occurs with all potyviruses examined to date. This can be cleavage of a genomic length polyprotein or cleavage of an immature protein, translated from a sub-genomic mRNA, to its functional form. Second, post-translational cleavage always occurs between the c-terminus of glutamine and the n-terminus of either glycine, serine, or alanine. Poliovirus and CPMV utilize the same sites for proteolytic cleavage of their polyproteins and are known to encode the protease responsible for this process. Potyviruses, however, have not been shown to contain a viral encoded protease although current studies would indicate that there may be.

Objectives of the Thesis Research

The objectives of my research are to obtain a cDNA clone of the 3′ terminus of SMV which encodes the capsid protein gene, 3′ terminal poly (A) tail and 5′ flanking region containing part of the nuclear inclusion protein gene. This clone will be sequenced and analyzed : 1) to determine if

the capsid protein is translated from a sub-genomic mRNA of the gene or if it is post-translationally cleaved from a polyprotein, 2) to determine what the potential cleavage site is and 3) to compare the nucleotide sequence and the deduced amino acid sequence of the SMV capsid protein to the sequences of other potyviruses.

Materials & Methods

Virus purification. SMV was purified from Glycine max (L.) Merr. (cv. Lee 74) 18-21 days postinoculation according to Hunst & Tolin (1982). The isolate used was SMV-Va described by Hunst & Tolin as a G-1 strain (Cho & Goodman, 1979).

RNA isolation. SMV RNA was isolated by a modification of the procedure described by Vance & Beachy (1984b). Purified virions were disrupted in a buffer containing 10mM Tris-HCl (pH 7.2), 1mM EDTA, 1% Sarkosyl and 1% 2-mercaptoethanol by incubating at 95 C for 3 minutes. The RNA was immediately extracted with 1 vol of phenol:chloroform (1:1). The aqueous phase was re-extracted with phenol:chloroform (1:1) until there was no protein interface. An equal vol of Proteinase K [1mg/ml in 10mM Tris-HCl (pH 7.2);1mM EDTA] was added to the aqueous phase and incubated at 65 C for 15 minutes. The RNA was extracted with an equal vol of phenol:chloroform (1:1) and re-extracted as before. The solution was then extracted with 1 vol of chloroform. The RNA was precipitated with 0.5 vols of 7.5M ammonium acetate and 2 vols of cold 95% ethanol overnight at -20 C.

Synthesis of complementary DNA. Double stranded complementary DNA (cDNA) was synthesized by a modification of the procedures of Lapeyre & Amalric (1985) and Gubler & Hoffman (1983). SMV RNA (1ug) was denatured at 95 C for

1 minute in the presence of 0.3 ug oligodeoxythymidylate [oligo (dT)] and allowed to cool slowly to room temperature. The solution was incubated in a final reaction vol of 50 ul containing 50mM Tris-HCl (pH 8.3), 6mM $MgCl_2$, 80mM KCl, 25mM dithiothreitol (DTT), 500uM each of dATP, dCTP, dGTP and dTTP, 4uCi of [^{35}S]dATP, 36 units of RNasin (Boehringer Mannheim Biochemicals, Indianapolis, IN), and 17 units of avian myoblastosis virus (AMV) reverse transcriptase (Life Sciences Inc., St. Petersburg, FL). The reaction was incubated at 42 C for 1 hr, then an additional 17 units of reverse transcriptase were added and incubated an additional 1 hr.

After first strand cDNA synthesis, 9 units of RNase H (Pharmacia, Piscataway, NJ) were added and incubated at 42 C for 5 min. The reaction was diluted five-fold in a buffer containing 20 mM Tris-HCl (pH 7.5), 4 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 100 mM KCl, 0.15 mM B-NAD, 200 ug/ml BSA, 40 uM each of dNTP's, 0.42 ug E. coli ligase (Pharmacia) and 25 units of E. coli DNA polymerase I (Pharmacia) and was incubated 1 hr at 12 C and then 1 hr at 16 C. The double stranded cDNA was blunt-ended by adding 4 units of T4 DNA polymerase (Pharmacia) to the reaction and incubating 15 min at 37 C. The reaction was terminated by adding Na_2EDTA to 25 mM, then extracting with 1 vol of phenol:chloroform (1:1)

followed by two secondary butanol extractions. The cDNA was precipitated with 0.5 vols of 7.5 M ammonium acetate and 2 vols of cold 95% ethanol overnight at -20 C. The cDNA was pelleted by centrifugation at 12,000 x G for 30 min, washed twice with 80% ethanol and dried under vacuum.

The cDNA pellet was resuspended in a reaction vol of 50 ul and incubated 1 hr at 37 C with EcoRI methylase according to the manufacturers instructions (Pharmacia). The methylated cDNA was precipitated and pelleted as described above.

EcoRI linkers were added by incubating the cDNA in a final reaction vol of 25 ul containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5% polyethylene glycol 6000, 5 mM ATP, 5 mM dithiothreitol, 0.5 ug hexadecameric EcoRI linker (Pharmacia) and 2 units of T4 DNA ligase (Pharmacia). Incubation was at room temperature for 4 hrs, then an equal vol of EcoRI restriction enzyme buffer (International Biotechnologies Inc) containing 10 units of EcoRI restriction enzyme (International Biotechnologies Inc.) was added and incubated for 2 hrs at 37 C. The reaction was terminated by a phenol:chloroform (1:1) extraction and the cDNA was precipitated twice with ethanol overnight at -20 C as described previously. The cDNA was ligated with an equal molar amount of pUC 18 which had been EcoRI digested and

dephosphorylated.

Transformation and identification of recombinant clones. E. coli, strain JM109, was made competent by the method of Hanahan (1982) and transformed with the plasmid-insert mix. Transformed bacteria were plated on LB agar containing ampicillin (50 ug/ml). Replicas of the bacterial colonies were fixed onto nitrocellulose membranes and hybridized with an ^{125}I labeled viral RNA probe as described by Maniatis et al (1982). The iodinated viral RNA probe was prepared by the method of Selin et al (1983).

Deletion sub-cloning. Recombinant plasmid (18 ug) was completely digested with SalI and SphI to generate 5′ and 3′ protrusions; respectively and deletions were made according to the manufacturers instructions (Stratagene Cloning Systems, La Jolla, CA). The DNA was resuspended in a final reaction vol of 300 ul containing 50 mM Tris-HCl (pH 8.0), 5mM $MgCl_2$, 10 ug/ml tRNA, 20 mM 2-mercaptoethanol and Exonuclease III (900 units). The reaction was incubated at 37 C ; 3 ug aliquots were removed every 30 sec into a buffer containing 30 mM NaOAc (pH 5.0), 50 mM NaCl, 1 mM $ZnCl_2$, 5% glycerol and put on dry ice. When all aliquots were collected, samples were heated at 68 C for 15 min and put on ice. Mung bean nuclease (9 units) was added to each sample, incubated at 37 C for 30 min, extracted once

with phenol:chloroform (1:1) and precipitated overnight at -20 C with 0.5 vols of 7.5 M ammonium acetate and 2 vols of cold 95 % ethanol.

DNA sequence analysis. Double stranded plasmid containing the cDNA insert was sequenced by the procedure described by Hattori and Sakaki (1986), with the following modifications made to plasmid denaturation and primer hybridization. Plasmid DNA (2 ug) was resuspended in a vol of 7.5 ul with sterile distilled water and heated in a boiling water bath (100 C) for 5 min, then quenched on ice. After cooling, 1.5 ul of 10 x Klenow buffer [100 mM Tris-HCl (pH 8.5), 100 mM $MgCl_2$] and 1 ul of primer (5 ng of pUC18 30 mer primer) were added and the solution was incubated at 37 C for 30 min. Dideoxynucleotide chain termination reactions were according to the method of Sanger et al (1983).

Results

RNA isolation. SMV RNA isolations yielded approximately 25 ug RNA / mg of virus as determined spectrophotometrically. The RNA was diluted to 100 ng / ul in 1 x TE (pH 8.0) and frozen at -20 C. Molecular weight of the RNA was not determined but the RNA did migrate as one band on a 0.7% agarose gel indicating that it was not degraded (data not shown). The RNA was also able to bind in an oligo (dT) cellulose column indicating that the polyadenylate tract was present.

Early observations indicated that RNA degradation was a problem if the purified virus was not used within 3 days or if laboratory materials were not treated with diethylpyrocarbonate at a concentration of 0.2%.

cDNA cloning. Fifty recombinant plasmids containing cDNA inserts were identified by colony hybridization with ^{125}I-labeled SMV genomic RNA (fig. 2). Agarose gel electrophoresis of plasmid mini-prep DNA identified cDNA clones containing inserts between 200 to 1500 base pairs in size (fig. 3). One recombinant plasmid; pSMV-49, exhibited a strong signal when hybridized with ^{125}I-labeled SMV RNA (fig. 2, B) and was determined to contain a 1500 base pair insert by agarose gel electrophoresis (fig. 3, lane 12).

Dot-blot hybridization demonstrated that the ^{125}I-labeled

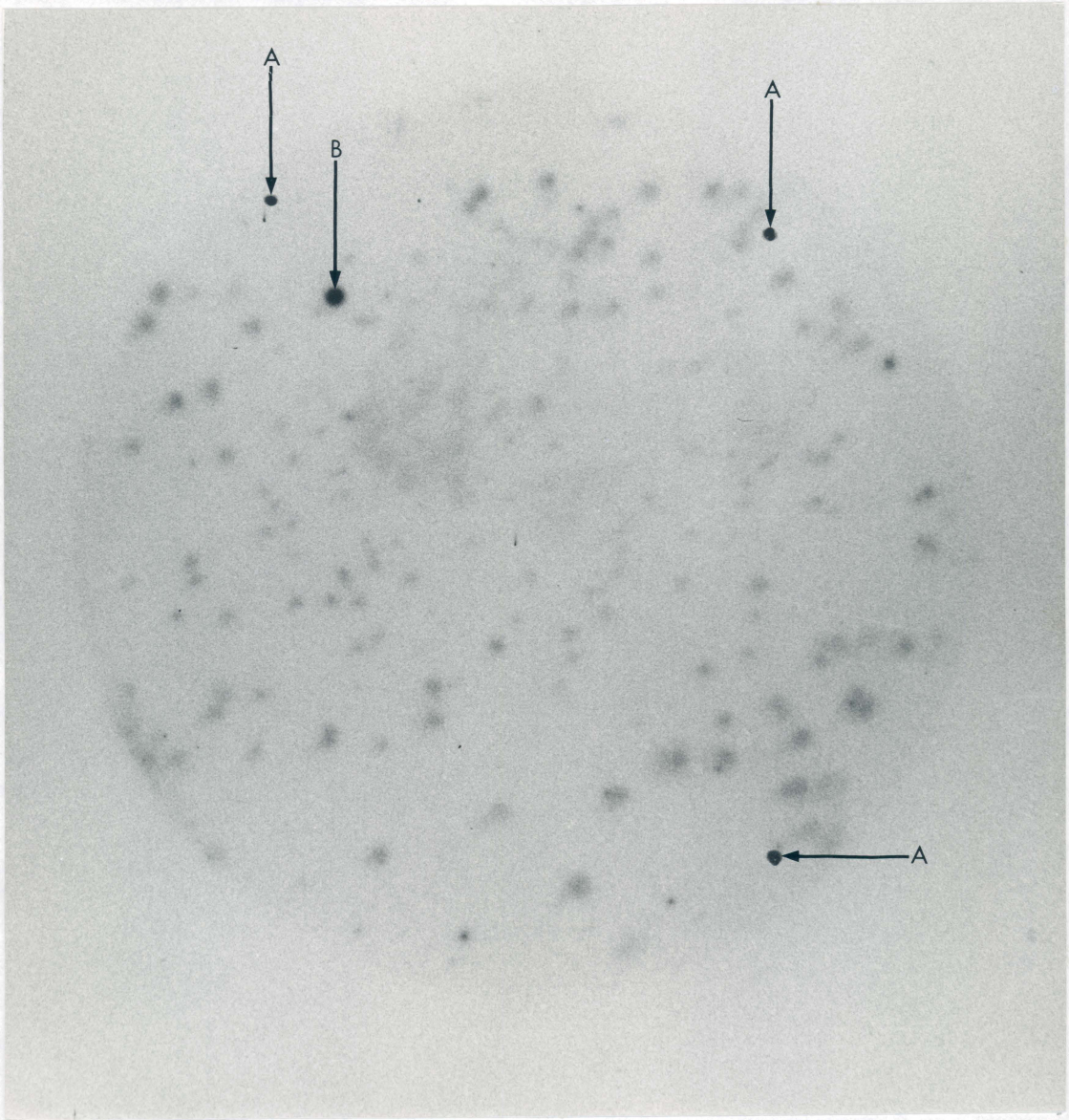

Fig. 2. Colony hybridization. Colonies were fixed onto nitrocellulose membranes and hybridized with an ^{125}I labeled SMV RNA probe. A indicates the key marks used for orientation to the master plate. B identifies pSMV-49.

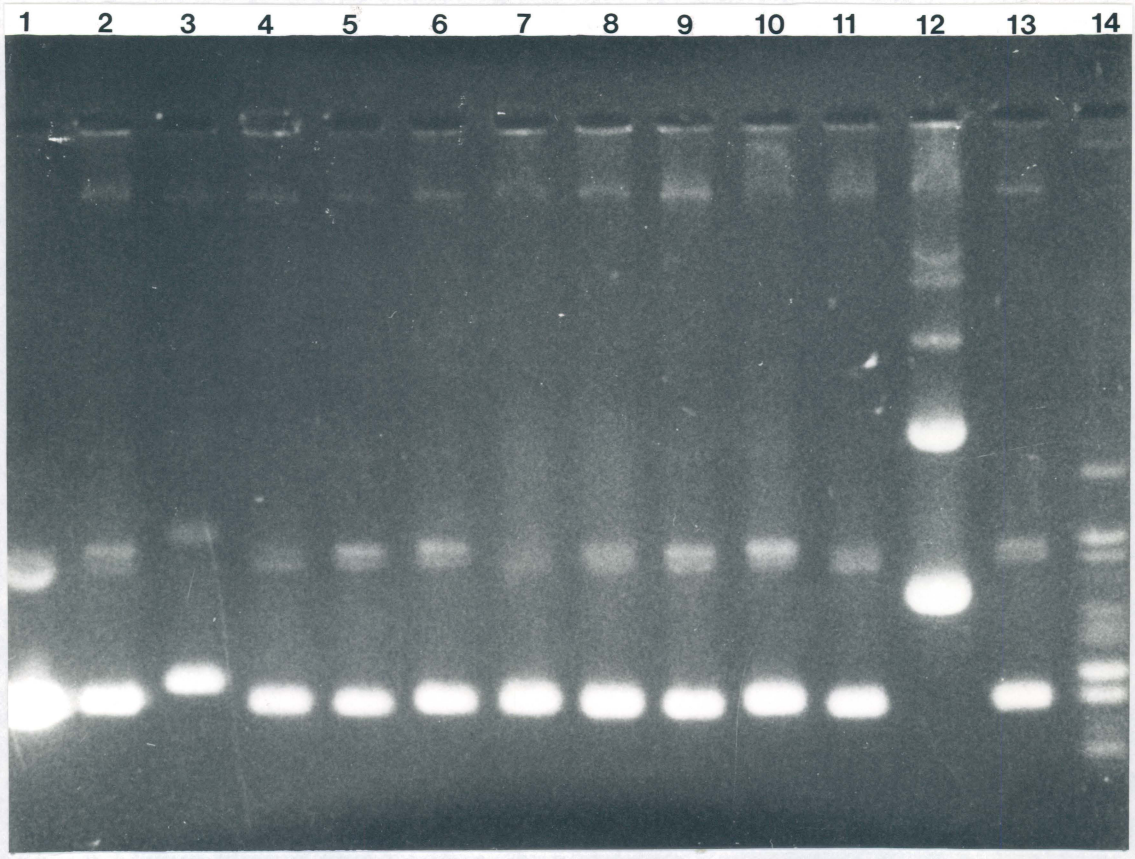

Fig. 3. Agarose gel electrophoresis of recombinant plasmids. Plasmid mini-prep DNA was electrophoresed on a 0.7% agarose gel. Lane 1 = pUC-19, lanes 2 - 13 = pSMV-39 to pSMV-50, respectfully and lane 14 = plasmid V-517.

SMV RNA probe did not hybridize to pUC-19 DNA and had the strongest signal when hybridized to pSMV-49, compared to pSMV-47, pSMV-50 and pSMV-40 containing 100 base pair, 200 base pair and 500 base pair inserts, respectively (fig. 4).

Plasmid sequencing. The strategy for nucleotide sequence analysis of pSMV-49 was based on the generation of a series of 100 - 300 base pair deletion sub-clones using Exonuclease III / mung bean nuclease (fig. 5). Exonuclease III will digest approximately 200 base pairs every 30 sec at 37 C. The aliquots were removed and put on dry ice to stop the reaction. The fragments were then blunt-ended with mung bean nuclease and re-ligated with the plasmid vector. Agarose gel electrophoresis of plasmid mini-prep DNA revealed a population of fragment sizes between 100-300 base pairs for each time interval aliquot collected. The sub-clones allowed for the sequencing of over-lapping fragments in both orientations.

Evaluation of the nucleotide sequence. Recombinant clone pSMV-49 contains a 1443 nucleotide cDNA copy of the 3′ terminus of SMV inclusive of the polyadenylate tract. Examination of the nucleotide sequence identified a large open reading frame (ORF) containing 1119 nucleotides on the positive sense viral RNA (fig. 6). The ORF is separated from the 3′ polyadenylate tail by a 224 nucleotide untranslated

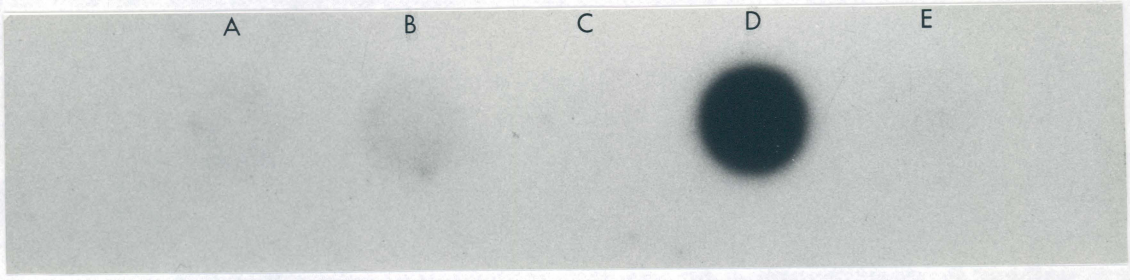

Fig. 4. Dot-blot hybridization. Recombinant plasmids were digested with Eco R1, fixed onto a nitrocellulose membrane and hybridized with an ^{125}I labeled SMV RNA probe. A= pSMV-50, B = pSMV-40, C = pUC-19, D = pSMV-49 and E = pSMV-47.

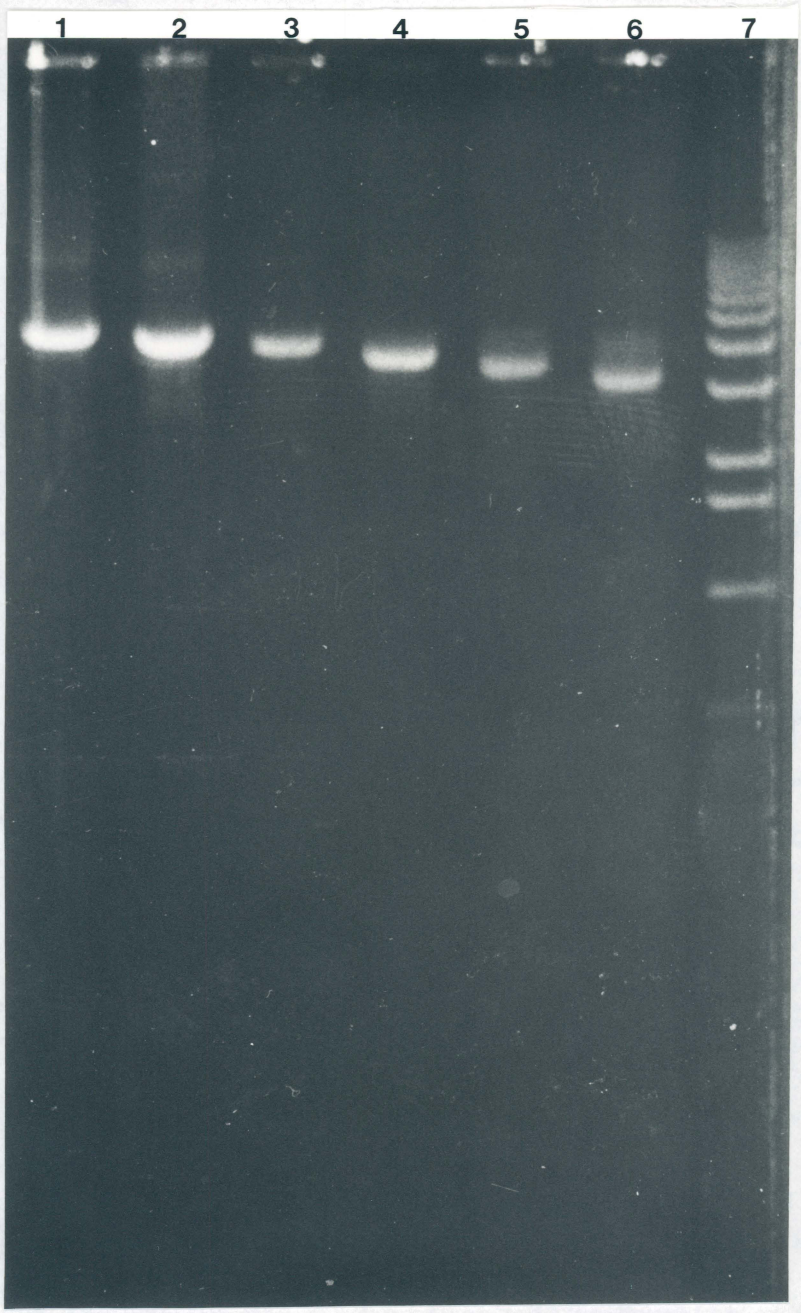

Fig. 5. Exonuclease III / mung bean nuclease deletion subclones. pSMV-49 deletion subclones were generated as described in materials & methods. Lanes 1 - 6 are the 100 to 300 base pair deletions of pSMV-49. Lane 7 is the supercoiled ladder molecular weight standard.

```
 I   P   K   L   E   H   E   R   I   V   S   I   L   E   W 15
AUC CCC AAA CUU GAA CAU GAG AGG AUA GUA UCG AUA CUU GAA UGG

 D   R   S   D   E   P   S   H   R   L   E   A   I   C   A 30
GAC AGG UCG GAC GAG CCC UCC CAC AGA CUA GAA GCU AUC UGU GCC

 S   R   I   E   A   W   G   Y   R   H   L   R   N   E   I 45
UCA AGG AUC GAA GCU UGG GGU UAC AGG CAC CUC AGG AAC GAA AUC

 R   S   F   Y   A   W   V   P   E   P   A   S   Y   S   R 60
CGU UCG UUC UAC GCA UGG GUC CCG GAG CCA GCU UCG UAC UCC AGG

 T   D   V   E   R   L   A   P   Y   L   A   E   P   A   L 75
ACA GAC GUU GAA CGU CUA GCG CCA UAU CUC GCA GAA CCA GCA CUU

 F   L   Y   T   S   Q   A   K   T   N   G   E   L   T   E 90
UUU CUC UAU ACC UCA CAA GCG AAA ACG AAC GGC GAA CUC ACC GAA

 Y   R   K   L   L   C   D   N   H   E   S   Q   R   E   K 105
UAC AGA AAG CUG CUA UGU GAC AAC CAU GAA UCA CAG AGA GAA AAA

 V   H   F   Q ↓ S   N   L   Q   E   V   G   D   V   K   A 120
GUC CAU UUC CAA AGC AAU CUG CAG GAA GUC GGG GAC GUA AAA GCA

 S   A   K   K   H   Q   E   Y   T   N   P   A   L   H   P 135
AGU GCU AAA AAG CAC CAG GAA UAU ACC AAC CCC GCA CUG CAC CCA

 R   K   D   K   D   V   N   A   G   T   S   G   T   F   S 150
AGA AAG GAC AAA GAU GUU AAC GCA GGG ACA AGC GGC ACU UUU UCA

 V   P   R   I   K   I   A   A   P   K   I   T   Y   P   K 165
GUA CCU AGA AUU AAG AUC GCA GCC CCA AAA AUC ACA UAU CCG AAG

 I   N   G   P   P   V   V   N   L   D   H   K   L   E   Y 180
AUC AAC GGG CCU CCA GUC GUA AAC CUG GAC CAC AAG CUU GAA UAC

 K   P   Q   Q   I   D   L   S   N   T   R   A   T   H   S 195
AAA CCA CAA CAG AUC GAU CUA UCC AAC ACG AGG GCU ACU CAC UCU

 Q   F   K   A   W   H   A   A   V   M   D   A   Y   G   I 210
CAA UUU AAG GCG UGG CAU GCA GCU GUC AUG GAU GCC UAU GGC AUC

 N   E   E   D   M   K   I   V   L   N   G   F   M   V   W 225
AAU GAA GAG GAC AUG AAA CUA GUA CUC AAU GGC UUU AUG GUG UGG

 C   I   E   N   G   T   S   P   N   I   N   G   V   W   T 240
UGC AUC GAA AAC GGA UAU AGC CCC AAC AUA AAC GGC GUA UGG ACG

 M   M   D   G   A   E   Q   V   E   Y   P   L   K   P   M 255
AUG AUG GAC GGC GCU GAA CAA GUC GAG UAU CCG CUC AAG CCU AUG

 V   E   N   A   D   P   T   L   R   Q   I   M   D   H   F 270
GUA GAG AAU GCA GAU CCC ACC CUC AGA CAA AUC AUG GAU CAC UUU

 S   D   L   A   E   A   Y   I   E   M   R   N   N   E   W 285
AGC GAC CUC GCG GAA GCU UAU AUC GAA AUG AGA AAC AAC GAA UGG

 P   Y   M   P   R   Y   G   L   Q   R   N   L   G   D   L 300
CCA UAC AUG CCG AGA UAC GGA CUA CAG AGA AAC CUC GGC GAC CUA

 S   L   A   R   I   L   D   F   Y   E   V   T   S   T   T 315
UCG CUC GCA AGG AUC CUU GAU UUC UAU GAA GUC ACC UCA ACG ACA

 P   V   R   A   R   E   A   H   N   Q   M   K   A   A   A 330
CCA GUG AGA GCG AGA GAA GCC CAC AAU CAA AUG AAA GCG GCC GCU

 G   T   Q   A   V   P   R   L   F   G   L   D   G   L   V 345
GGU ACC CAA GCG GUC CCC AGA CUG UUU GGA CUC GAC GGA CUA GUG

 S   T   Q   E   E   N   T   E   R   H   T   T   D   D   V 360
UCG ACG CAG GAA GAA AAC ACG GAG AGA CAC ACG ACU GAC GAU GUA

 N   P   H   M   H   T   L   L   G   V   K   G   M   *     374
AAC CCC CAU AUG CAC ACG CUG CUU GGA GUG AAA GGC AUG UAA ACG

UGCUCUUGUCUAGAUUCCGUUGCUCUGUGUUAAGAUAAUGUUAUGAUUCCUAGCAUGUU
AUCUCUCGCAAUCGUGUUCAAUCAAUAUAUAUUUUUAGCAUUUUAUUCGGUCUAUUAGU
AAAGUGGUCACCUAUUAUUUCUGAAAUUCUCGCUUACAAACGGCGGAGAUCUCUUGUCA
UUCACAGUAAUAAGCAGUUAUACAUGAAUAUAUCGUGGCUA-poly A (100)
```

Fig. 6. The nucleotide sequence of the 3' terminal 1443 nucleotides of the SMV genome. Above the nucleotide sequence is the abreviation of the amino acid encoded for by the codon: A = alanine, C = cysteine, D = aspartate, E = glutamate, F = phenylalanine, G = glycine, H = histadine, I = isoleucine, K = lysine, L =leucine, M = methionine, N = asparagine, P = proline, Q = glutamine, R = arganine, S = serine, T = threonine, V = valine, W = tryptophane and Y = tyrosine. The arrow indicates the putative cleavage site.

region containing 10 termination codons in the three potential reading frames. The sequence does not begin with an initiation codon indicating that it probably lies upstream of the available data. The first initiation codon is at position 205 which would only code for a 19 kDa protein, smaller than the predicted 30 kDa protein described by Vance & Beachy (1984b). The ORF can code for 373 amino acid residues resulting in a polypeptide of 42 kDa.

Discussion

Improvement of cDNA cloning. Initial attempts at cDNA synthesis were not sucessful. The approach was to synthesize first strand cDNA using AMV reverse transcriptase, degrade the RNA and then synthesize second strand cDNA using DNA polymerase I (Klenow fragment). The double stranded cDNA was then blunt ended using S1 nuclease. Gubler & Hoffman (1983) have stated that S1 nuclease treatment often results in the loss of large numbers of clones generated per ug of RNA. Furthermore, the hairpin loop generated after first strand synthesis does not function as a suitable primer for second strand synthesis. Their procedure circumvents both by using RNase H to nick the RNA molecule after first strand synthesis along with DNA polymerase I and DNA ligase. The exonuclease activity of DNA polymerase I will remove the RNA and the enzyme then repairs the duplex by sequentially adding DNA bases. The DNA ligase will repair the gaps that may be present.

Evidence of proteolytic processing. Cell-free translation studies of SMV (Vance & Beachy, 1984b) and other potyviruses (Dougherty & Hiebert, 1980) along with the nucleotide sequence analysis of TEV-HAT (Allison et al, 1985), PeMV (Dougherty et al, 1985) and TVMV (Domier et al, 1986) indicate that capsid protein is expressed as part of a

larger polyprotein which is proteolytically cleaved. The results of this study support those observations. The sequence did not begin with an initiation codon indicating that it lies within a large open reading frame which may code for a genomic length polyprotein. However, since only part of the gene proposed to code for the polyprotein is identified as part of one large ORF, we can not propose the size of the precursor polyprotein.

Specific amino acid dipeptide bonds have been described as protease cleavage sites in the maturation of functional proteins from polyprotein precursors with poliovirus (Koch & Koch, 1985), CPMV (Goldbach, 1986) and all potyviruses studied to date (Allison et al, 1985a ; Dougherty et al, 1985 and Domier et al, 1986). In all cases, cleavage occurs between glutamine:alanine, glutamine:glycine or glutamine: serine dipeptide bonds. A glutamine:serine dipeptide bond was identified at position 109 - 110. This would result in the production of a 30 kDa capsid protein which is in agreement with protein electrophoresis studies described by Vance & Beachy (1984b). Glutamine:serine protease cleavage sites resulting in the maturation of capsid protein have also been described for TEV-HAT (Allison et al, 1985a), TVMV (Domier et al, 1986) and SCMV (Gough et al, 1987).

Previous investigators (Allison et al, 1985a and Dougherty

et al, 1985) have suggested that specific dipeptide bonds are not required for cleavage but that other structural features may be important. The proposed cleavage sites of all potyviruses occur on the carboxyl side of a tripeptide consisting of two ring-structure amino acids followed by a glutamine residue. Specifically, cleavage of TEV-HAT capsid protein occurs at the carboxyl side of a tyrosine:phenylalanine:glutamine tripeptide (Allison et al, 1985a), PeMV cleavage occurs at the carboxyl side of a histadine:histadine:glutamine tripeptide (Dougherty et al, 1985), TVMV capsid protein cleavage occurs at the carboxyl side of a arginine:phenylalanine:glutamine tripeptide and cleavage of the capsid protein of SCMV occurs at the carboxyl side of a glutamate:histadine:glutamine tripeptide. The proposed cleavage site of SMV in this study occurs at the carboxyl side of a histadine:phenylalanine:glutamine tripeptide.

Comparison of the SMV amino acid sequence to other potyvirus amino acid sequences. The derived amino acid sequence of the SMV capsid protein was compared to the predicted amino acid sequences of TEV (Allison et al, 1985), TVMV (Domier et al, 1986), PeMV (Dougherty et al, 1985) SCMV (Gough et al, 1987) and the determined amino acid sequence of PVY (Shukla et al, 1986) (fig. 7). A high degree of

```
          1
SMV   H F Q   S N L Q E V G D V K A S A K K H Q E Y T N P A L H P R K D K
TVMV  R F Q   S D T V D A G K D K A R D Q K L A D K P T L A I D R T K
TEV   Y F Q   S G T V D A G A D A G K K K D Q K D D K V A E Q A S K
PEMV  H H Q A N D T I D T G G N S K K D V K P E Q G S I Q P S S N K G K E
PVY     A N D T I D A G E S S K K D A R P E Q G S I Q V N P N K G K D

SCMV    E H Q   S G N E D A G K Q K S A T P A A N Q T A S G D G K P V Q T
SCMV    A D N K P S S D N T S N A Q G T S Q T K G G G E S G G T N A T A T

         31
SMV       D V N A G T S G T F S V P R I K I A A P K I T Y P K I N G P
TVMV      D K D V N T G T S G T F S I P R L K K A A M N M K L P K V G G S
TEV       D R D V N A G T S G T F S V P R I N A M A T K L Q Y P R M R G E
PEMV      D V N A G T S G T H T V P R I K A I T A K M R M P K S K G A
PVY       D V N A G T S G T H T V P R I K A I T A K M R M P R S K G A
SCMV      D K D V D V G S T G T F V I P K L K K V S P K M R L P M V S N K

         61
SMV       P V V N L D H K L E Y K P Q Q I D L S N T R A T H S Q F K A
TVMV      S V V N L D H L L T Y K P A Q E F V V N T R A T H S Q F K A
TEV       V V V N L N H L L G Y K P Q Q I D L S N A R A T H E Q F A A
PEMV      A V L K L D H L L E Y A P Q Q I D I S N T R A T Q S Q F D T
PVY       T V L H L E H L L E Y A P Q Q I D I S N T R A T Q S Q F D T
SCMV      A I L N L D H L I Q Y K P D Q R D I S N A R A T H T Q F Q F

         91
SMV       W H A A V M D A Y G I N E E D M K L V L N G F M V W C I E N
TVMV      W H T N V M A E L E L N E E Q M K I V L N G F M I W C I E N
TEV       W H Q A V M T A Y G V N E E Q M K I L L N G F M V W C I E N
PEMV      W Y E A V R L A Y D I G E T E M P T V M N G L M V W C I Q N
PVY       W Y E A V R M A Y D I G E T E M P T V M D G L M V W C I E N
SCMV      W Y N R V K K E Y D V D D E Q M R I L M N G L M V W C I E N

        121
SMV       G Y S P N I N G V W T M M D G A E Q V E Y P L K P M V E N A
TVMV      G T S P N I S G V W T M M D G D E Q V E Y P I E P M V K H A
TEV       G T S P N L N G T W V M M D G E D Q V S Y P L K P M V E N A
PEMV      G T S P N I N G V W V M M D G S E Q V E Y P L K P I V E N A
PVY       G T S P N V N G V W V M M D G N E Q V E Y P L K P I V E N A
SCMV      G T S P D I N G Y W T M V D G N N Q S E F P L K P I V E N A
```

Figure 7. A comparison of the putative capsid protein of SMV to the capsid proteins of TVMV, TEV, PeMV, PVY and SCMV. The boxed areas indicate regions of homology.

```
        151
SMV       D P T L R Q I M D H F S D L A E A Y I E M R N N E W P Y M P
TVMV      N P S L R Q I M K H F S N L A E A Y I R M R N S E Q V Y I P
TEV       Q P T L R Q I M T H F S D L A E A Y I E M R N R E R P Y M P
PEMV      K P T L R Q I M A H F S D V A E A Y I E M R N K K E P Y M P
PVY       K P T L R Q I M A H F S D V A E A Y I E M R N K K E P Y M P
SCMV      K P T L R Q C M M H F S D A A E A Y I E M R N L D E P Y M P

        181
SMV       R Y G L Q R N L G D L S L A R I   L D F Y E V T S T T P V R A
TVMV      R Y G L Q R G L V D R N L A P F A F D F F E V N G A T P V R A
TEV       R Y G L Q R N I T D M S L S R Y A F D F Y E L T S K T P V R A
PEMV      R Y G L V P N L R D A S L A R Y A F D F Y E V T S R T P V R A
PVY       R Y G L I R N L R D V G L A R Y A F D F Y E V T S R T P V R A
SCMV      R Y G L L R N L N D K S L A R Y A F D F Y E I N S R T P N R A

        211
SMV       R E A H N Q M K A A A G T Q A V P R L F G L D G L V S T Q E
TVMV      R E A H A Q M K A G R T P Q F A A A M F C L D G S V S G Q E
TEV       R E A H M Q M K A A A V R N S G T R L F G L D G N V G T A E
PEMV      R E A H I Q M K A A A L K S A Q S R L F G L D G G V S T Q E
PVY       R E A H I Q M K A A A L K S A Q P R L F G L D G G I S T Q E
SCMV      R E A H A Q M K A A A I R G S T N H M F G L D G N V G E S S

        231
SMV       E N T E R H T T D D V N P H M H T L L G V K G M
TVMV      E N T E R H T V D D V N A Q M H H L L G V
TEV       E D T E R H T A H D V N R N M M H T L L G V R
PEMV      E N T E R H T T E D V S P S M H T L L G V R N M
PVY       E N T E R H T T E D V S P S M H T L L G V K N M
SCMV      E N T E R H T A A D V S R N V H S Y R G A K I
```

Figure 7 continued.

homology (>60%) is evident within the interior portion and the carboxyl end of the gene between all six viruses. The N-terminal region, however, is highly variable. Allison et al (1985) state that the N-terminus of the capsid protein of potyviruses is strongly hydrophilic and located at or near the external surface of the virion. Dougherty et al (1984) raised monoclonal antibodies to TEV-NAT, TEV-HAT and PeMV and tested them to isolated virions. They found that antibodies from one potyvirus did not react to other potyvirus virions. The variability in amino acid sequence in the N-terminus would account for the non-reactivity of monoclonal antibodies tested among potyviruses.

Hiebert et al (1984), using limited trypsin digestions, demonstrated that a small portion of the potyvirus capsid protein contained approximately 50% of the proteins lysine residues. Examination of the predicted amino acid sequence of SMV along with the sequence of TEV, TVMV and PeMV is in agreement with this observation. Approximately 50% of the capsid proteins lysine residues are in the N-terminal region of all four viruses contributing to it's hydrophilic nature.

Comparison of the nucleotide sequence of SMV to the nucleotide sequences of other potyviruses. The nucleotide sequence of the capsid protein gene of SMV was compared to the nucleotide sequence of the capsid protein genes of TVMV

SMV	321	TTTCCAAAGCAATCTGCAGGAAGTCGGGGACGTAAAAGCA	
TVMV		gagatttcaaAgTgatacaGtAGatGctGggaaggAcaag	20%
TEV		TTattttcagAgTggcactGtgGatGctGgtGctgAcGCt	30%
PEMV		TaTgaAgtGCAtcaccagGcgAaTgatacAatcgAcActg	30%
SCMV		ggatgtAgaacATCaGtcaGgcaatGaGGAtGctgggaaA	30%
SMV	361	AGTGCTAAAAAGCACCAGGAATATACCAACCCCGCACTGC	
TVMV		gcaagggatcAgaAaCtaGctgATAagccaaCactggcaa	20%
TEV		gGTaagAAgAAagAtCAaaAggATgatAAagtCGCtgaGC	47%
PEMV		gagGgaAtAgtaagaaAGatgTgaAaCcAgaaCaaggTag	27%
SCMV		cagaagAgtgcaacaCccGctgcaAatcAaaCaGCAagtg	22%
SMV	401	ACCCAAGAAAGGACAAAGATGTTAACGCAGGGACAAGCGG	
TVMV	459	AtCgcAcAAAGGACAAgGATGTTAAtaCcGGtACAAGCGG	77%
TEV	89	AggCttcAAAGGAtAggGATGTTAAtGCtGGaACttcaGG	62%
PEMV	423	ACaaggGtAAGGAaAAAGATGTaAAtGCtGGtACAtctGG	67%
SCMV	589	caaCAAagAAGGAtAAgGATGTTgACGttGGatCAActGG	65%
SMV	441	CACTTTTTCAGTACCTAGAATTAAGATCGCAGCCCCAAAA	
TVMV		acaTTTagCAtTcCaagGctcaAgaAagcCgcaatgAAtA	35%
TEV	129	aACaTTcTCAGTtCCacGAATaAAtgctatgGCCaCAAAA	62%
PEMV		aACacaTaCtGTgCCaAGAATaAAGgctatcaCggCAAAA	55%
SCMV		aACTTTTgttaTcCCgAaAtTaAAGAagGtttCaCCAAAg	57%
SMV	481	ATCACATATCCGAAGATCAACGGGCCTCCAGTCGTAAACC	
TVMV		tgaAatTgcCaaAgGtTggAgGaagtTCagtTgtcAActt	32%
TEV		AaCttcaATatccAaggatgaGGGgagaggtggtTgtAaact	25%
PEMV		ATgAgAatgCCcAAaAgCAAaGGagCggCcGTgcTgAAt	52%
SCMV		ATgcgtctaCCcAtGgTgAgCaacaaagCcaTacTcAAtt	35%
SMV	521	TGGACCACAAGCTTGAATACAAACCACAACAGATCGATCT	
TVMV		gGatCatCtcctaacAtataAgcCagCAcagGAatttgtT	27%
TEV	209	TGaAtCACctttTaGgATACAAgCCACAgCAaATtGATtT	67%
PEMV	543	TaGACCACttGCTcGAgTAtgcgCCACAACAGATaGATaT	72%
SCMV	709	TGGACCAtctaaTccAATACAAACCAgAtCAGAgaGAcaT	67%
SMV	561	ATCCAACACGAGGGCTACTCACTCTCAATTTAAGGCGTGG	
TVMV	619	tgtgAACACGAGaGCaACaCACTCTCAgTTTAAGGCaTGG	77%
TEV		gTCaAAtgCtcGaGCcACaCAtgagCAgTTTgccGCGTGG	57%
PEMV	583	cTCaAACACtcGGGCaACTCAaTCaCAgTTTgAtaCGTGG	72%
SCMV	749	tTCaAAtgCacGaGCTACaCACaCaCAATTccAGttcTGG	62%

Figure 8. Comparison of the SMV capsid protein gene sequence to the capsid protein gene sequences of TVMV, TEV, PeMV and SCMV.

SMV	601	CATGCAGCTGTCATGGATGCCTATGGCATCAATGAAGAGG	
TVMV	659	CAcaCAaaTGTtATGGcTGaactTGaatTaAATGAAGAGc	62%
TEV	289	CATcagGCaGTgATGacaGCCTATGGagTgAATGAAGAGc	70%
PEMV		tATGaAGCaGTgcgGctTGCaTAcGaCATagggGAAactG	55%
SCMV		tAcaacagaGTCAaGaAaGagTATGatgTtgATGAtGAGc	50%
SMV	641	ACATGAAACTAGTACTCAATGGCTTTATGGTGTGGTGCAT	
TVMV	699	AaATGAAAaTAGTgCTCAATGGCTTcATGaTtTGGTGtAT	82%
TEV	329	AaATGAAAaTAtTgCTaAATGGaTTTATGGTGTGGTGCAT	85%
PEMV	663	AaATGccAactGTgaTgAATGGgcTTATGGTtTGGTGCAT	70%
SCMV	829	AaATGAgAaTttTgaTgAATGGgTTgATGGTtTGGTGtAT	70%
SMV	681	CGAAAACGGATATAGCCCCAACATAAACGGCGTATGGACG	
TVMV	739	aGAgAAtGGAacgtcaCCaAACATcAgtGGtGTgTGGACc	60%
TEV	369	aGAAAAtGGGacTtcCCCaAAttTgAACGGaactTGGgtt	57%
PEMV		tcAAAAtGGAacctcgCCaAAtATcAACGGaGTaTGGgtt	57%
SCMV		tttggtgtatagagaatggcACATcccCtGatataaatcG	17%
SMV	721	ATGATGGACGGCGCTGAACAAGTCGAGTATCCGCTCAAGC	
TVMV	779	ATGATGGAtGGtGaTGAACAAGTgGAGTATCCaaTCgAaC	80%
TEV	409	ATGATGGAtGGtGagGAtCAAGTttcaTAcCCGCTgAAaC	70%
PEMV	743	ATGATGGAtGGaagTGAACAAGTCGAGTATCCatTgAAaC	80%
SCMV	909	ATGgTGGAtGGgaacaAtCAAtcaGAGTtTCCaCTaAAaC	62%
SMV	761	CTATGGTAGAGAATGCAGATCCCACCCTCAGACAAATCAT	
TVMV	819	CaATGGTgaAGcATGCgaATCCCtCatTacGtCAgATaAT	65%
TEV	449	CaATGGTtGAaAAcGCgcAgCCaACaCTgAGgCAAATtAT	70%
PEMV	783	CaATtGTtGAGAATGCAaAaCCaACCCTtAGgCAAATCAT	80%
SCMV	949	CaATaGTgGAaAAcGCAaAaCCaACatTacGACAgtgCAT	62%
SMV	801	GGATCACTTTAGCGACCTCGCGGAAGCTTATATCGAAATG	
TVMV	859	GaAaCACTTTAGtaAtCTtGCtGAgGCTTAcATaaggATG	67%
TEV	489	GacaCACTTcAGtGACCTgGCtGAAGCgTATATtGAgATG	75%
PEMV	823	GGcaCAtTTctcaGAtgTtGCaGAAGCgTATATaGAAATG	67%
SCMV	989	GatgCAtTTTAGtGAtgcCGCaGAAGCaTAcATtGAAATG	70%
SMV	841	AGAAACAACGAATGGCCATACATGCCGAGATACGGACTAC	
TVMV	899	AGgAAttcaGAAcaGgtcTACATaCCcAGgTAtGGcCTgC	60%
TEV	529	AGgAAtAggGAgcGaCCATACATGCCtAGgTAtGGtCTAC	72%
PEMV	863	cGcAACAAaaAggaaCCATAtATGCCacGATAtGGttTAg	62%
SCMV	1029	AGAAAtttgGAtgaGCCgTACATGCCaAGATACGGtCTcC	72%
SMV	881	AGAGAAACCTCGGCGACCTATCGCTCGCAAGGATCCTTGA	
TVMV		gcgcggAttagtagacagaaaCctggcaccGttTgCcTtt	15%
TEV	569	AGAGAAACaTtacaGACaTgagttTgtCAcGctatgcgttcga	42%
PEMV		ttccAAAtCTgcGgGAtgcAagtCTcGCgcGctatgcctt	35%
SCMV		ttAGgAAtCTaaatGACaagagcCTCGCtcGataCGCATT	52%

Figure 8 continued.

SMV	921	TTTCTATGAAGTCACCTCAACGACACCAGTGAGAGCGAGA	
TVMV	982	cTTCTtTGAgGTtAatggggCaACACCAGTccGgGCaAGA	62%
TEV	612	cTTCTATGAgcTaACtTCAAaaACACCtGTtAGAGCGAGg	75%
PEMV	946	cTTtTATGAAGTCACaTCAcgGACACCAGTGAGgGCtAGA	82%
SCMV	1112	TTTCTATGAgaTCAatTCgcgcACACCAaatAGgGCGAGA	70%
SMV	961	GAAGCCCACAATCAAATGAAAGCGGCCGCTGGTACCCAAG	
TVMV	1022	GAAGCgCAtgcaCAAATGAAgGCGGgCcgcacTcCgCAAt	62%
TEV	652	GAgGCgCAtAtgCAAATGAAAGCtGCtGCaGtacgaaAca	60%
PEMV	986	GAAGCgCACAtaCAAATGAAgGCcGCaGCattgAaatcAG	65%
SCMV		GAgGCaCAtgcaCAAATGAAgGCaGCaGCaatTAgagggt	55%
SMV	1001	CGGTCCCCAGACTGTTTGGACTCGACGGACTAGTGTCGAC	
TVMV		tcGcagCagccaTGTTTtGtCTgGAtGGcagcGTGTCagg	47%
TEV	692	gtGgaaCtAGgtTaTTTGGtCTtGAtGGcaacGTGggtAC	52%
PEMV		CccaatCtcGACTtTTcGGgtTgGACGGtggcGTcagtAC	50%
SCMV		CcacgaaCcacaTGTTTGGACTCGACGGgaatGTtggaga	50%
SMV	1041	GCAGGAAGAAAACACGGAGAGACACACGACTGACGATGTA	
TVMV	1102	GCAaGAAGAgAACACaGAacGcCAtACtgtTGACGATGTt	72%
TEV	732	tgcaGAgGAAgACACtGAacGgCACACagCgcACGATGTg	62%
PEMV	1066	aCAaGAgGAgAACACaGAGAGgCACACcACcGAgGATGTt	75%
SCMV	1232	GagctctGAgAACACaGAGcGgCACACagCTGcaGATGTc	62%
SMV	1081	AACCCCCATATGCACACGCTGCTTGGAGTGAAAGGCATGT	
TVMV	1142	AAtgCtCAaATGCACcacCTtCTgGGtGTtAAgGGggTGT	65%
TEV		AACCgtaAcATGCACACaCTatTaGGgGTccgccagtgaT	52%
PEMV	1106	tctCCaagTATGCAtACtCTaCTTGGAGTGAggaaCATGT	67%
SCMV		tcaCggaATgTtCAttCGtacCgTGGgGccAAAatCtaag	42%

Figure 8 continued.

(Domier et al, 1986), TEV (Allison et al, 1985), PeMV (Dougherty et al, 1985) and SCMV (Gough et al, 1987) (fig. 8). The data reveals that there is less than 60% homology in the 5′ and 3′ terminal regions of the capsid protein gene when SMV is compared to the above four viruses. This is in contrast to the amino acid sequence in which only the 5′ terminus of the gene had a low degree of homology. There is a high degree of nucleotide homology (>60%) within the interior portion of the gene.

As described previously, there is a putative proteolytic cleavage site at amino acid positions 109-110 (fig. 6). This would result in a polypeptide of 264 amino acid residues, which is in close agreement with the determined capsid proteins of TEV [263 (Allison et al, 1985)], PeMV [267 (Dougherty et al, 1985)], PVY [267 (Shukla et al, 1986)] and TVMV [262 (Domier et al, 1986)]. These results support the conclusion that clone pSMV-49 contains the entire capsid protein gene and the putative proteolytic cleavage site.

The 5′ flanking region of the capsid protein was compared to TEV (Allison et al, 1985), TVMV (Domier et al, 1986), PeMV (Dougherty et al, 1985) and SCMV (Gough et al, 1987) (fig. 9). The N terminal region of the flanking region revealed a high degree of homology when compared to all four virus sequences. This sequence in TEV is known to encode the

SMV	1	ATCCCCAAACTTGAACATGAGAGGATAGTATCGATACTTG	
TVMV	56	ATaCCaAAgtTaGAACcTGAGcGcATcGTgTCcATtCTTG	70%
TEV		ATaCCaAAgCTaGAAgAaGAGAGGATaGTaTCtATttTgG	70%
PEMV		gTaCCaAAACTcGAAgAaGAaAGaATtGTgatctattcTa	52%
SCMV	63	ATaCCgAAAgTaGAACcaGAGAGGgTtGTcgCcATttTgG	65%
SMV	41	AATGGGACAGGTCGGACGAGCCCTCCCACAGACTAGAAGC	
TVMV	96	AATGGGAtcGtagtGcaGAGCCtcaCCAtAGAtTAGAAGC	67%
TEV		AATGGGACAGaTCcaAaGAGCCgTCaCAtAGgCTtGAAGC	77%
PEMV	46	cAatGGACAGGgCaGAtttGgCCgaaCACAGACTcGAAGC	67%
SCMV	103	AATGGGAtAGaagcattGAGCCagaaCACcGtCTAtcAGC	60%
SMV	81	TATCTGTGCCTCAAGGATCGAAGCTTGGGGTTACAGGCAC	
TVMV	136	aATtTGTGCtTCcAtGATtGAgGCgTGGGGTTACActgAC	72%
TEV		cATCTGTGCaTCAAtGATtGAAGCaTGGGGTTAtgacaAg	72%
PEMV	86	aATtTGcGCagCcAtGATaGAAtCgTGGGGTTACtccgAa	62%
SCMV	143	aATaTGTGCagCgAtcATCGAAtCaTGGGGcTACgaGgAa	65%
SMV	121	CTCAGGAACGAAATCCGTTCGTTCTACGCATGGGTCCCGG	
TVMV		CTactacAgaAcATCaGaagGTTCTAtaaATGGacaatag	45%
TEV		CTggttaAaGAAATCCGcaatTTCTAtGCATGGGTtttGG	65%
PEMV		CTaAcacACcAAATtaGgagGTTtTACtCATGGtTattGc	57%
SCMV	183	tTaAcatAtcAAATCCGacgaTTCTACcaATGGGTCCtcG	62%
SMV	161	AGCCAGCTTCGTACTCCAGGACAGACGTTGAACGTCTAGC	47%
TVMV		AGCaAGaacCtTACagaAGccttGcaGaacAggGcCTtGC	47%
TEV		AaCaAGCgcCGTAtTCacaGcttGcaGaaGAAgGaaagGC	32%
PEMV		AcaacgtTTtGCAtcaataGcgcaggaagGAAaAgCTcct	32%
SCMV		AaCaAGagcCaTACaaggaactAGcacTacAAgGcaaAGC	42%
SMV	201	GCCATATCTCGCAGAACCAGCACTTTTTCTCTATACCTCA	
TVMV		aCCATATtTgtCAGAAgttGCACTaagaagaTtataCaCA	52%
TEV		GCCATATCTgGCtGAgaCtGCgCTTaagtTtTtgtaCaCA	60%
PEMV		aCCAaAaCTCGaAGAAgaAagAaTTgTgaTCTATtCtaCA	62%
SCMV		tCCATActTgtCAGAgaCAGCtCTacggaaaTtatatcct	42%
SMV	241	CAAGCGAAAACGAACGGCGAACTCACCGAATACAGAAAGC	
TVMV		tcgcaaAtAgCcActGaCaAtgaactCacAgAttactAca	27%
TEV		tctcaGcAcggaAcaaaCtctgagAtaGAAgAgtatttaa	22%
PEMV		acAGgGcAgtgGAcgaagaggagCtaaGAgTctttActGa	30%
SCMV		gAtGaatcttgcgAtcaaagtgaactactAcgttattAtg	12%

Figure 9. A comparison of the 5′ flanking region of the SMV capsid protein gene to the nuclear inclusion protein gene sequence of TVMV, TEV, PeMV and SCMV. Capital letters indicate homologous nucleotide bases.

SMV	281	TGCTATGTGACAACCATGAATCACAGAGAGAAAAAGTCCA	
TVMV		aGgagatactCgcaaacaAtgaAttcttgcgtgAAactgt	15%
TEV		aagTgTtgtAtgAttAcGAtattCcaAcgactgAgaatCt	27%
PEMV		aatgATGgttgcAttggatgatgagtttgaAtgtgaTtCt	17%
SCMV		aagagatgtAtAAgaAcGaAttgatGAGtGAAgAtGTggt	37%
SMV	321	TTTCCAAAGCAATCTGCAGGAAGTCGGGGACGTAAAAGCA	
TVMV		gagatttcaaAgTgatacaGtAGatGctGggaaggAcaag	20%
TEV		TTattttcagAgTggcactGtgGatGctGgtGctgAcGCt	30%
PEMV		TaTgaAgtGCAtcaccagGcgAaTgatacAatcgAcActg	30%
SCMV		ggatgtAgaacATCaGtcaGgcaatGaGGAtGctgggaaA	30%

Figure 9 continued.

nuclear inclusion protein (Allison et al, 1985); however, nuclear inclusion bodies have not been found in SMV infected soybean plants (Edwardson, 1974). Nuclear inclusion bodies have not been observed in plants infected with PeMV but they can be immunoprecipitated from in vitro translation products of PeMV RNA (Dougherty & Hiebert, 1980). Gough et al (1987) state that PeMV and SCMV may synthesize nuclear inclusion proteins but they may not be assembled into inclusion proteins. A similar level of expression may occur with SMV suggesting that SMV may encode the nuclear inclusion gene.

In summary, this study supports observations that SMV capsid protein is initially expressed as a high molecular weight polyprotein. The predicted amino acid sequence shares several similarities with other potyviruses; namely, the N-terminal region of the capsid protein is strongly hydrophylic and contains approximately 50% of the proteins lysine residues, the amino acid sequence of the N-terminus of the capsid protein is highly variable when compared to other potyvirus capsid protein sequences while many internal regions are highly conserved and that capsid protein maturation appears to involve proteolytic cleavage at a site containing two ring structure amino acid residues followed by a glutamine residue. We also suggest that SMV may encode a nuclear inclusion gene which shares some homology with other known potyvirus nuclear inclusion genes.

Literature Cited

Allison, R. F., Dougherty, W. G., Parks, T. D., Willis, L., Johnston, R. F., Kelley, M., and Armstrong, F. B. (1985a). Biochemical analysis of the capsid protein gene and capsid protein of tobacco etch virus: N - terminal amino acids are located on the virion surface. Virology 147, 309-316.

Allison, R. F., Sorenson, J. C., Kelly, M. E., Armstrong, F. B., and Dougherty, W. G. (1985b). Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: Evidence for synthesis and processing of a polyprotein in potyvirus genome expression. Proc. Natl. Acad. Sci. USA 82, 3969-3972.

Birnboim, H. C., and Doly, J. (1979). A rapid extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1515-1523.

Cho, E.-K., and Goodman, R. M. (1979. Strains of soybean mosaic virus: Classification based on virulence in resistant soybean cultivars. Phytopathology 69, 467-470.

Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellman, G. M., Overmeyer, J. H., Hirenath, S. T., Siaw, M. F. E., Lomonssoff, G. P., Shaw, J. G., and Rhoads, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. Nucleic Acids Res. 5417-5430.

Dougherty, W. G., and Hiebert, E. (1980). Translation of potyvirus RNA in a rabbit reticulocyte lysate: Cell-free translation strategy and a genetic map of the potyviral genome. Virology 104, 183-194.

Dougherty, W. G., Allison, R. F., Parks, T. D., Johnston, R. E., Feild, M. J., and Armstrong, F. B. (1985). Nucleotide sequence at the 3′ terminus of pepper mottle virus genomic RNA : Evidence for an alternative mode of potyviral capsid protein gene organization. Virology 146, 282-291.

Edwardson, J. R. (1974). Some properties of the potatoe virus Y-group. Monograph Series, Florida Agricultural Experimental Station 4, 85 pp.

Frannsen, H., Moerman, M., Rezelman, G., and Goldbach, R. (1984). Evidence that the 32,000-dalton protein encoded by

bottom-component RNA of cowpea mosaic virus is a proteolytic processing enzyme. Journal of Virology 50, 183-190.

Goldbach, R. W. (1986). "Molecular Evolution of Plant RNA Viruses". Ann. Rev. Phytopathology 24, 289-310.

Gough, K. H., Azad, A. A., Hanna, P. J., and Shukla, D. D. (1987). Nucleotide sequence of the capsid and nuclear inclusion protein genes from the Johnson grass strain of sugar cane mosaic virus RNA. Journal of General Virology 68, 297-304.

Gubler, U., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269.

Hanahan, D. (1985). Techniques for Transformation of E. coli. In "DNA Cloning" (D. M. Glover, ed.), vol. 1 pp. 109-135. IRL Press, New York.

Hattori, M., and Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152, 232-238.

Hari, V., Siegel, A., Rozek, C., and Timberlake, W. E. (1979). The RNA of tobacco etch virus contains poly (A). Virology 92, 568-571.

Hari, V. (1981). The RNA of tobacco etch virus: Further characterization and detection of protein linked to RNA. Virology 112, 391-399.

Hellman, G. M., Thornbury, D. W., Hiebert, E., Shaw, J. G., Pirone, T. P., and Rhoads, R. E. (1983). Cell-free translation of tobacco vein mottling virus RNA. Virology 124, 434-444.

Holmes, D. S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Annal. Biochem. 114, 193-197.

Hill, J. H., and Benner, H. I. (1976). Properties of potyvirus RNA's: Turnip mosaic, tobacco etch and maize dwarf mosaic virus. Virology 75, 419-432.

Hill, J. H., and Benner, H. I. (1980). Properties of soybean

mosaic virus ribonucleic acid. Phytopathology 70, 236-239.

Kates, J. (1970). Transcription of the vaccinia virus genome and the occurance of polyadenylic acid sequences in messenger RNA. Cold Spring Harbor Symp. Quant. Bio. 35, 743-752.

Koch, F. and Koch, G. (1985). " The Molecular Biology of Poliovirus". Springer-Verlag Wien, New York. p 590.

Lapeyre, B., and Amalric, F. (1985). A powerful method for the preparation of cDNA libraries: Isolation of cDNA encoding a 100-kDal nucleolar protein. Gene 37, 215-220.

Luciano, C., Siaw, M., Ballard, S., Xu, Z., and Shaw, J. (1984). Aggregation of potyviral RNA. Phytopathology 74, 859.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). "Molecular Cloning. A Laboratory Manual". Cold Spring Harbor Laboratories, Cold Spring Harbor, New York. 545 p.

Nagel, J., and Hiebert, E. (1985). Complementary DNA cloning and expression of the papaya ringspot potyvirus sequences encoding capsid protein and a nuclear inclusion-like protein in Escherichia coli. Virology 143, 435-441.

Pelham, H. R. B. (1979). Synthesis and proteolytic processing of cowpea mosaic virus proteins in reticulocyte lysates. Virology 96, 463-477.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Selin, Y. M., Harich, B., and Johnson, J. L. (1983). Preparation of labeled nucleic acids (nick translation and iodination) for DNA homology and rRNA hybridization experiments. Current Microbiology 8, 127-132.

Shukla, D. D., Inglis, A. S., McKern, N. M., and Gough, K. H. (1986). Coat protein of potyviruses. Virology 152, 118-125.

Struhl, K. (1985). A rapid method for creating recombinant DNA molecules. Biotechniques 3, 452-453.

Tucker, P. W., Marcu, K. B., Slightom, J. L., and Blattner, R. R. (1979). Structure of the constant and 3' untranslated regions of the murine 2b heavy chain messenger RNA. Science 206, 1299-1303.

Vance, V. B., and Beachy, R. N. (1984a). Detection of genomic-length soybean mosaic virus RNA on polyribosomes of infected soybean leaves. Virology 132, 26-36.

Vance, V. B., and Beachy, R. N. (1984b). Translation of soybean mosaic virus RNA in vitro: Evidence of protein processing. Virology 132, 271-281.

Winner, E. (1982). Genome-linked proteins of viruses. Cell 28, 199-201.

Yeh, S.-D., and Gonsalves, D. (1985). Translation of papaya ringspot virus RNA in vitro: Detection of a possible polyprotein that is processed for capsid protein, cylindrical inclusion protein and amorphous-inclusion protein. Virology 143, 260-271.

Appendix A

Purification of Soybean Mosaic Virus RNA

This protocol is based on a modification of the procedure described by Vance & Beachy (1984b).

1. Pre-experimental preparation.

(a) The virus suspension must be used within 24 hr after the completion of purification.

(b) All eppendorf tubes, pipette tips and glassware must be treated to eliminate ribonucleases by soaking in a solution of 0.2% DEPC for at least 1 hr and then autoclaving.

(c) All buffers and reagents should be prepared using ultra pure chemicals or should be treated with DEPC as desribed above.

(d) Water quality is essential. Distilled water should be treated with 0.2% DEPC overnight and autoclaved.

(e) Gloves must be worn at all times.

2. Disrupt Capsid Protein. Preheat 40 ul of 10% sarkosyl (IBI, ultra pure) 40 ul of 10x TE buffer (pH 7.2), and 4 ul of 2 - mercaptoethanol at 95 C in a water bath. Add 316 ul of purified SMV (1 mg) heat at 95 C for 3 min and put immediately on ice.

3. Remove Protein Contamination. Add 1 vol phenol: chloroform (1:1), mix until solution turns milky white and microfuge for 5 min. Remove supernatant and and save.

Continue phenol:chloroform extractions as described above until the protein interface is gone.

4. Cleave Covalently-Linked Proteins. Add an equal volume of Proteinase K [IBI, 1 mg/ml in 1 x TE (pH 7.2)], mix briefly and incubate at 65 C for 20 min in a water bath.

5. Remove Protein Contamination. Add 1 volume phenol: chloroform (1:1), mix briefly and microfuge for 5 min. Remove supernatant and save. Continue phenol : chloroform extractions until protein interface is gone.

6. Remove Phenol Contamination. Add 1 volume chloroform, mix briefly and microfuge for 5 min. Remove supernatant and save.

7. Precipitate RNA. Add 0.5 vol of 7.5 M ammonium acetate plus 3 vols of cold 95% ethanol overnight at -20 C.

8. Pellet RNA. After overnight precipitation, microfuge for 30 min to pellet the RNA. Gently pipette off ethanol, add an equal vol of 70% ethanol and microfuge for 5 min. Dry pellet thoroughly at room temperature (ca 25 C) in a 37 C incubator or by lyophilization in a speed-vac.

9. Store RNA in Freezer. Short term storage of RNA (less than 1 week) can be accomplished by resuspending the pellet in DEPC treated water at a concentration of 1 ug/ul with 36 units of RNasin (Boehringer Mannheim) at -20 C. Long term storage of RNA should be in TE buffer (pH 7.2) at a

concentration of 1 ug/ul with 36 units of RNasin (Boehringer Mannheim) at -70 C.

Appendix B

cDNA Synthesis

The following procedure is a modification of the procedures described by Gubler & Hoffman (1985) and Lapeyre & Amalric (1985).

1. Thaw RNA. Remove 1 ug of SMV RNA from frozen storage and bring vol up to 11 ul with DEPC treated distilled water.

2. Hybridize Primer to RNA. Add 3 ul oligo-(dT) primer (Boehringer Mannheim, 0.3ug) to the RNA and heat denature at 95 C for 1 min in a heating block, then allow the heating block with sample to cool at room temperature to 40 C.

3. Synthesis of First Strand cDNA. After sufficient cooling, add 10 ul of treated distilled water, 10 ul of 5 x reverse transcriptase buffer, 10 ul of dNTP's (Pharmacia, 5mM each), 1 ul RNasin (Boehringer Mannheim, 36 units), 2.5 ul of 500 mM DTT, 1 ul of labeled ATP (4 uCi), and 1 ul of reverse transcriptase (Life Sciences, 17 units). Incubate at 42 C for 1 hr, add another 17 units of reverse transcriptase and incubate an additional 1 hr at 42 C.

4. Nick First Strand cDNA. Add 4 ul RNase H (Pharmacia, 9 units) and incubate at 42 C for 5 min.

5. Synthesis of Second Strand cDNA. To the reaction, add 143 ul of treated distilled water, 20 ul of 10 x second

strand buffer, 20 ul of 1.4 M B-NAD, 4 ul of BSA (2.5mg/ml), 2 ul of dNTP's (Pharmacia, 5mM each), 2 ul of E. coli ligase (Pharmacia, 0.42 ug), and 5 ul of E. coli polymerase I (Life Sciences, 25 units). Incubate for 1 hr at 12 C and then 1 hr at 16 C in a temperature controlled water bath.

6. Blunt-end Double Stranded cDNA. Add 2 ul of T4 polymerase (Pharmacia, 4 units) to the reaction and incubate for 15 min at 37 C.

7. Precipitate cDNA. Add 0.5 vols of 7.5 M ammonium acetate and 3 vols of 95% ethanol overnight at -20 C.

Buffers

5x Reverse Transcriptase Buffer

250 mM Tris-HCl (pH 8.3 @ 42 C)
400 mM Potassium chloride (KCl)
30 mM Magnesium chloride ($MgCl_2$)

Add 0.3 g KCl and 0.061 g $MgCl_2$ to 10 ml distilled water. Treat overnight with 0.2% DEPC and then autoclave. Add 0.3 g of Tris (IBI, ultra pure) and adjust the pH to 8.3 at 42 C with HCl. Then sterilize the buffer by passage through a 0.22 u filter.

10x Second Strand Buffer

200 mM Tris-HCl (pH 7.5)
100 mM Ammonium sulfate [$(NH_4)_2SO_4$]
40 mM Magnesium chloride ($MgCl_2$)
1 M Potassium chloride (KCl)

Add 0.081 g $MgCl_2$ and 0.75 g KCl to 10 ml of distilled

water. Treat with 0.2% DEPC overnight and then autoclave. Add 0.24 g of Tris (IBI, ultra pure) and 0.132 g of $[(NH_4)_2SO_4]$ then adjust the pH to 7.5 with HCl. Sterilize by passage through a 0.22 u filter.

Appendix C

Competant Cell Preparation and Transformation

This protocol is a slight modification of the procedure described by Hanahan (1985). The competant cells can not be frozen, but must be used immediately for transformation.

1. Start Overnight Culture. Pick one colony of E. coli (strain JM 109) and inoculate 1.5 ml of minimal media. Grow cells overnight at 37 C and 250 rpm agitation.

2. Inoculate and Grow Cells. Inoculate 30 ml of SOB media with 15 drops of the overnight culture and grow cells at 37 C and 250 rpm agitation until they reach a density of 0.45 - 0.55 at O.D. 550 (4-7 x 10 cells/ml). This takes between 2 - 3 hours.

3. Harvest Cells. Remove the cells from the incubator and place on ice for 10 minutes. Harvest the cells by centrifuging at 746 x g (2500 rpm in Sorvall SS34 rotor) for 12 minutes at 4 C. Gently pour off the supernatant and resuspend the cells in 1/3 original volume (10 ml) with TFB buffer by gentle vortexing. Place on ice for 10 minutes.

4. Re-harvest Cells. Centrifuge the cell suspension at 746 x g (2500 rpm) as before, pour off the supernatant and resuspend the pellet in 1/12.5 original volume (2.4 ml) with TFB buffer.

5. Add DMSO. Immediately add freshly-thawed DMSO to 3.5%

(84 ul/2.4 ml cells), swirl gently and place on ice for 5 minutes.

6. Add DTT. Add DTT to a final concentration of 75 mM (84 ul of 2.25 M DTT/2.4 ml cells), gently swirl and place on ice for 10 minutes.

7. Add More DMSO. Add DMSO to 3.5% as before.

8. Add DNA to Competant Cells. For each transformation reaction, add 210 ul of competant cells to each eppendorf tube. Then add up to 10 ul of DNA and place on ice for 30 minutes.

9. Heat Shock. Incubate the competant cell / DNA mixture at 42 C for 2 minutes. Immediately place cells back on ice.

10. Let Cells Recover. Add 800 ul SOC media to the transformed cells and incubate at 37 C and 250 rpm agitation.

11. Plate. Dilute the cells and plate on LB agar containing 5 mg/ml ampicillan. If the plasmid vector contains the lac Z gene, 40 ul of 100 mM IPTG and 40 ul of 2% Bluo-gal (BRL) can be added to the transformed cells just prior to plating.

Buffers and Media

Minimal Media

Component	100 ml Solution
Potassium phosphate (dibasic)	1.05 g
Potassium phosphate (monobasic)	0.45 g
Ammonium sulphate	0.10 g
Sodium citrate	0.15 g

Autoclave the above components in 98 ml distilled water, then add the following filter sterilized components separetly. Store at room temperature.

Magnesium sulfate	1 ml of 2% soln
Glucose	1 ml of 20% soln
Thiamine hydrochloride	0.1 ml of 5mg/ml soln

SOB Media

Component	500 ml solution
Tryptone (2%)	10 g
Yeast Extract (0.5%)	2.5 g
Sodium chloride (10 mM)	1.25 ml of 4 M soln
Potassium chloride (2.5 mM)	1.25 ml of 1 M soln

Autoclave the above components in 490 ml sterile distilled water, then add the following filter sterilized components when the first four ingrediants are cool. Store at 4 C.

Magnesium chloride (10 mM)	5 ml of 1 M soln
Magnesium sulfate (10 mM)	5 ml of 1 M soln

TFB Buffer

Component	250 ml solution
Potassium chloride	1.85 g
Manganese chloride	2.23 g
Calcium chloride	0.38 g
Hexamine cobalt chloride	0.20 g
Potassium-MES (10 mM, pH 6.3)	5 ml of 0.5 M soln

The potassium-MES must be pH adjusted before adding to the other components.

SOC Media

SOB media with filter sterilized glucose added to 20 mM.

DTT Solution

Component	10 ml solution
DTT	3.3 g
Potassium acetate (pH 7.5)	100 ul of 1 M soln

Bring volume to 10 ml with sterile distilled water and sterilize by filtration through a 0.22 u filter.

Appendix D

Colony Hybridization

This protocol is based on a slight modification of the procedure described by Maniatis et al (1982). The iodinated RNA probe was prepared by the method of Selin et al (1983).

1. Lift Colonies onto Nitrocellulose Membranes. Pre-moisten a nitrocellulose membrane by placing onto a blank agar petri plate, then place onto the petri plate containing clones of interest.

2. Prepare Master Plate. Lift the membrane off with blunt end forceps and place on a fresh LB agar plate containing 5mg/ml ampicillan, colony side up. Place another pre-moistened nitrocellulose membrane on the original, key the membranes by punching out three assymetrical holes with a sterile Pastuer pipette and incubate as a sandwich at 37 C for 1 hour. Remove the the upper membrane, place on a fresh LB/amp plate and incubate 3-4 hours to let the cells recover. Store at 4 C for future use.

3. Lyse Cells. Saturate a piece of Whatman 3MM filter paper with denaturing solution (0.5 M NaOH + 1.5 M NaCl) and pour off the excess. *NOTE: This solution must be made just prior to use. Place the original membrane on the filter paper, colony side up, and let stand for 5 minutes at room temperature.

4. Neutralize. Transfer the membrane to a Whatman filter paper saturated with neutralizing solution [1.5 M NaCl + 0.5 M Tris-HCl (pH 8.0)], colony side up. Let stand for 5 minutes at room temperature.

5. Fix DNA to Membrane. Place the membrane onto a piece of dry Whatman filter paper to dry and then bake at 65 C overnight or at for two hours at 80 C under vacuum.

6. Pre-hybridization. Add 30 ml pre-hybridization buffer to a deep-dish, glass petri plate. Float the membrane in the buffer and incubate at 50 C for two hours and 50 rpm agitation.

7. Hybridization. Add ^{125}I RNA probe to 20 ml of pre-hybridization buffer. (In this case, 4 x 10 cpm was used). Gently pour off the first buffer from the membrane and add the buffer containing the RNA probe. *Note: Make sure that the buffer is evenly dispersed about the membrane. Incubate at 50 C for 5 hours.

8. Remove Unbound Probe. Pour off the buffer containing the RNA probe and add 200 ml of 2 x SSC buffer containing 0.5 x Denhardt's solution and 0.1% SDS. Incubate at 50 C and 50 rpm agitation for 5 minutes and pour off. Repeat four more times.

9. Wash Membrane. Add 200 ml of of 2 x SSC buffer and incubate for 5 minutes at 50 C and 50 rpm agitation. Pour

off and repeat 4 more times.

10. Autoradiography. Dry the membrane and place in a cassette with X-ray film. Expose for 36-48 hours at -70 C.

Buffers and Solutions

50 x Denhardt's Solution

Component	500 ml solution
Ficoll (1%)	5 g
Polyvinylpyrrolidone (1%)	5 g
Bovine Serum Albumin (1%)	5 g

20 x SSC

Component	1 litre solution
3 M Sodium chloride	175.3 g
0.3 M Sodium citrate	88.2 g

20 x SSPE

3.6 M Sodium chloride
200 mM Sodium phosphate dibasic (pH 7.4)
20 mM disodium EDTA (pH 7.4)

Pre-hybridization Buffer

50% Formamide
5 x Denhardt's solution
5 x SSPE
0.1% SDS

Appendix E

Exonuclease III Subcloning

This protocol was developed by Stratagene Cloning Systems, LaJolla, CA. For future reference, refer to the lab manual published by the company.

1. Digest Plasmid. Double digest 20-30 ug of the plasmid with restriction enzymes that will generate 5' and 3' 4 base protrussions, respectively. *Note: The restriction digests must be complete to ensure that Exonuclease III will digest in one direction only. When restriction digests are complete, heat innactivate at 65 C for 10 minutes, extract once with phenol : chloroform (1:1) and then two secondary butanol extractions. Precipitate with 0.5 volumes of 7.5 M ammonium acetate and two volumes of cold (-20 C) 95% ethanol, wash the pellet once with 80% ethanol and dry completely.

2. Digest with Exonuclease III. Resuspend the pellet in 1 x Exonuclease III buffer at a final concentration of 1 ug/ul. To 18 ul of double-digested plasmid (18 ug), the following were added; 150 ul of 2 x Exonuclease III buffer, 30 ul of freshly prepared 2-mercaptoethanol, 72 ul of sterile distilled water and 30 ul of Exonuclease III (900 units at 50 units/ ug plasmid). Incubate at 37 C and remove each aliquot (50 ul) at 30 second time intervals.

Immediately place each aliquot in a separate tube containing 350 ul of 1 x Mung Bean Nuclease buffer (dilute 80 ul of 5 x Mung Bean Nuclease buffer in 270 ul sterile distilled water) and place on dry ice. This procedure will result in the digestion of 400 bp / minute.

3. Blunt-end with Mung Bean Nuclease. Dilute Mung Bean Nuclease in 1 x Mung Bean Nuclease Dilution buffer for a final concentration of 9 units/ul. To each aliquot of of Exonuclease III digested plasmid, add 1 ul of Mung Bean Nuclease (9 units at 3 units/ug plasmid) and incubate at 30 C for 30 minutes.

4. Innactivate the Enzyme. Extract once with phenol: chloroform (1:1) followed by two secondary butinol extractions. Precipitate the plasmid, wash once with 80 % ethanol and dry the pellet completely.

5. Electrophoresis. Resuspend each aliquot in 15 ul of TE buffer (pH 8.0). Load 10 ul of each on a 0.7% Seaplaque agarose gel (low melting point agarose) and electrophorese to determine the extent of digestion. *Note: Lack of any plasmid DNA on the gel would indicate insufficient digestion with the restriction enzyme generating 3' protrussions and a high background of undigested plasmid DNA would indicate insufficient digestion with the restriction enzyme generating 5' protrussions. If background due to undigested

plasmid is high, cut out the bands of Exonuclease digested plasmid DNA for in-gel ligations as described by Struhl, (1985).

6. Ligate the Blunt-ended Plasmid. To a separate tube, add 1 ul digested plasmid, 14 ul of sterile distilled water, 2 ul of 10 x Ligation buffer, 1 ul of 10 mM ATP and 2 ul of T4 Ligase (8 units). Incubate at room temperature (ca 25 C) for 4 hours and then overnight at 4 C.

Buffers

2 x Exonuclease III Buffer

100 mM Tris-HCl (pH 8.0)
10 mM magnesium chloride
20 ug/ml t-RNA

5 x Mung Bean Nuclease Buffer

150 mM sodium acetate (pH 5.0)
250 mM sodium chloride
5 mM zinc chloride
25% glycerol

1 x Mung Bean Nuclease Dilution Buffer

10 mM sodium acetate (pH 5.0)
0.1 mM zinc acetate
1 mM Cysteine
0.001 % Triton X-100
50 % glycerol

10 x Ligation Buffer

500 mM Tris-HCl (pH 8.0)
70 mM magnesium chloride
10 mM DTT

Appendix F

Plasmid Sequencing

This protocol is a modification of the procedure developed by Hattori and Sakaki (1986).

1. Heat Denature the Plasmid. Bring 2 ug of plasmid DNA to a final volume of 7.5 ul with sterile distilled water and denature in boiling water (ca 100 C) for 5 minutes. *Note: The eppendorf tube must be completely submerged in the water bath to prevent condensation in the tube. Immediately place on ice for 2-3 minutes.

2. Anneal Primer to Template. Add 1.5 ul of 10 x Klenow buffer [100 mM Tris-HCl (pH 8.5), 100 mM $MgCl_2$] and 1 ul of pUC 30 base primer to the denatured plasmid and incubate at 37 C for 30 minutes. When annealing reaction is complete, add 1 ul of ^{35}S-ATP (20 uCi) and 2 ul of Klenow enzyme (1 unit).

3. Add Nucleotides. Set up 4 eppendorf tubes for each sample to be sequenced and label A,C,G,T. Add 2 ul of nucleotide mixture to each tube according to the label (i.e. add 2 ul of nucleotide mixture A to reaction tube A).

4. Combine Template. Add 3 ul of template/primer/isotope/ Klenow enzyme mixture to each tube as a hanging drop, centrifuge briefly and incubate at 37 C for 20 minutes.

5. Add Cold Chase. Add 2 ul of cold chase solution to each

tube as a hanging drop, centrifuge briefly to mix and incubate at 37 C for 15 minutes.

6. Stop Reaction. Add 4 ul of formamide dye mix to each tube and centrifuge briefly to mix. The reaction mixtures can be kept on ice for up to 4 hours or at -20 C for up to 2 weeks.

Preparing and Running the Sequencing Gel

1. Clean the Plates. Wash the inner face of the plates twice with 95% ethanol. Clean the larger plate once with Windex. Siliconize (100 ul of Silane in 10 ml distilled water) the smaller plate by evenly spreading the solution across the plate with Kimwipes, let stand 5 minutes and wipe dry. *Note: The plate must be siliconized before each use. Place the 0.4mm spacers along each side of the larger plate and lay the shorter plate on top making sure that the spacers are flush along the edge of the plates. Tape the bottom edge of the plates with water-resistant, 3MM Vinyl tape, allowing the tape to overlap the sides by about an inch. Tape the sides in the same manner and then re-inforce the corners with more tape.

2. Prepare the Gel Solution. Add 65 ml of top gel solution to a 250 ml vacuum flask and 15 ml of bottom gel solution to a 50 ml vacuum flask and let them warm to room temperature. Degas the top gel and bottom gel solutions for 2-3 and 1 minute, respectively. Add 284 ul of 10 % ammonium persulfate

and 32.5 ul of TEMED to the top gel solution and add 60 ul of 10 % ammonium persulfate and 14 ul of TEMED to the bottom gel solution. Swirl gently to mix.

3. Pour the Gel. Quickly, pipette up 12 ml of top gel followed by all of the bottom gel in a 25 ml pipette. Hold the plates at an angle and pipette the gel solution down the middle of the plates. Immediately pipette the rest of the top gel solution until it is at the top of the plates. *Note: If any air bubbles form, they can be removed by tapping with the handle of a screwdriver or tilting the plates up and down. Clamp the sides of the plates using 3 clamps for each side. Insert the blank comb at the top of the gel to form one large well and place 3 clamps at the top. Allow the gel to polymerize for 30-45 minutes.

4. Pre-run the Gel. Remove the tape from the bottom and sides of the plates, remove the blank comb and rinse off any excess polyacrylamide. Fill the bottom chamber of the unit with 1 x TBE buffer and place the gel in the chamber, starting at one corner and slowly working to the other. Clamp the gel in place, fill the upper chamber with the same buffer and insert the sharkstooth comb so that the tips just penetrate the surface of the gel. Attach the electrodes and warm the gel by running at 60 W for 30 minutes.

5. Load the Samples. For each reaction (i.e. A,C,G,T), heat

the samples at 80-90 C for 3 minutes in a heat block. Two reactions can be loaded at a time. As the samples heat, turn off the power to the gel and rinse the urea out of the wells to be loaded with a Pasteur pipette. When the samples are ready, load 2 ul of each into the wells, run the samples into the gel for 90 seconds and begin heating the next samples. When the samples have run into the gel, clean out the wells again. When all of the samples are loaded, electrophorese the gel until the lower dye (bromphenol blue) has migrated to the bottom of the gel (usually 90-120 minutes).

6. Fix the Gel. Remove the gel from the apparatus and carefully remove the side spacers and sharkstooth comb. Lay the gel flat with the shorter plate facing up and slowly pry the upper plate free of the gel. Slowly pour a solution of 7% acetic acid / 7% methanol over the gel surface and let it soak 15-20 minutes.

7. Dry the Gel. Drain off the fixing solution, lay a piece of Whatman blotter paper over the gel surface and gently rub smooth. Slowly peel the blotter paper, with the gel adhering to one side away from the larger plate. Cover the gel with plastic wrap and place on a pre-heated 80 C gel dryer. Dry under vacuum for 20 minutes. Note: The blotter paper will begin to fold up on the sides when the gel is sufficiently

dried.

8. Autoradiography. Remove the plastic wrap and place in a cassette with X-ray film. Expose for 18-36 hours at room temperature.

Buffers and Reagents

Nucleotide Mixtures.

dNTP's/ddNTP's: 18 mM stock solutions stored at -20 C.

Cold Chase Solution: 10 ul of each 18 mM dNTP + 320 ul of sterile distilled water.

0.5 mM dNTP's: 2.5 ul of 18 mM dNTP's + 87.5 ul of sterile distilled water.

dNTP mixtures:

	A	C	G	T
0.5 mM dCTP	20 ul	1	20	20
0.5 mM dGTP	20	20	1	20
0.5 mM dTTP	20	20	20	1
sterile distilled water	20	20	20	20

10 mM ddNTP's: 5 ul of 18 mM ddNTP's + 4 ul of sterile distilled water.

ddNTP mixtures:

0.01 mM ddATP: 1 ul ddATP + 999 ul sterile distilled water.
0.1 mM ddCTP: 2 ul ddCTP + 198 ul sterile distilled water.
0.1 mM ddGTP: 2 ul ddGTP + 198 ul sterile distilled water.
0.5 mM ddTTP: 10 ul ddTTP + 190 ul sterile distilled water.

Mix an equal volume of the ddNTP mixture to the appropriate dNTP solution for the final nucleotide mixture used in the sequencing reaction.

Formamide Dye Mixture:

100 ml formamide
5 g Amberlite mixed bed resin

Gently stir for 30 minutes and filter sterilize.

Then add:

0.03 g xylene cyanol
0.03 g bromophenol blue
0.75 g disodium EDTA

Store at room temperature for 1 month.

40% Acrylamide Stock Solution:

38 g acrylamide
2 g bis

Bring volume up to 100 ml with distilled water, add 5 g Amberlite mixed bed resin and gently stir for 30 minutes. Filter sterilize through Whatman #1 filter paper and store in the dark at 4 C.

10 x TBE Buffer:

108 g Tris base
55 g boric acid
9.3 g disodium EDTA

Bring the volume up to 1 liter with distilled water. This gives a stock solution of pH 8.3.

Top Gel (for 3 gels):

93.6 g urea
29.3 ml of 40% acrylamide stock solution
9.75 ml of 10 x TBE

Bring the volume up to 195 ml with warm distilled water (ca 50 C), stir until all components are in solution and

store at 4 C in the dark.

Bottom Gel:

21.6 g urea
4.5 g sucrose
6.75 ml of 40% acrylamide solution
11.25 ml of 10 x TBE

Bring the volume up to 45 ml with warm distilled water (ca 50 C), add enough bromophenol blue dye to turn the solution dark blue and stir until all components are in solution. Store at 4 C in the dark.

Fixing Solution:

Add 100 ml of glacial acetic acid and 100 ml of methanol to 1.5 liters of distilled water.

The vita has been removed from the scanned document