8-1-1995

Site Directed Mutagenesis of a Putative Protease Cleavage Site Within the 105 Kilodation Protein of Southern Bean Mosaic Virus

Jingyu Shi
Western Kentucky University

SITE DIRECTED MUTAGENESIS OF A PUTATIVE PROTEASE CLEAVAGE SITE WITHIN THE 105 KILODALTON PROTEIN OF SOUTHERN BEAN MOSAIC VIRUS

A Thesis

Presented to

the Faculty of the Department of Biology

Western Kentucky University

Bowling Green, Kentucky

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

Jingyu Shi

August, 1995

SITE DIRECTED MUTAGENESIS OF A PUTATIVE PROTEASE CLEAVAGE SITE WITHIN THE 105 KILODALTON PROTEIN OF SOUTHERN BEAN MOSAIC VIRUS

Date Recommended May 26, 1995

Claire A. Rinehart

Director of Thesis

Cheryl D. Davis

Volgene L. Dunkan

Elmer Gray 6/26/95

Director of Graduate Studies Date

Acknowledgement

I would like to express gratitude to my advisor, Dr. Claire A. Rinehart, for his guidance, kindness and special patience. I would like to thank Dr. Valgene Dunham for his kindness and precise guidance. Also, I would like to thank Dr. Cheryl Davis for her guidance, kindness and meticulate help. Thanks are also due to the graduate college for providing me a graduate student research grant. Also, I would like to thank the other faculty and staff members of the Department of Biology for helping to make my years at Western informative and enjoyable.

This thesis is dedicated to my parents, Haichou Shi and Xianzhen Xu, and my wife, Hong Shen. Their love and support enabled me to complete this work.

Table of Contents

List of Figures

SITE DIRECTED MUTAGENESIS OF A PUTATIVE PROTEASE CLEAVAGE SITE WITHIN THE 105 KILODALTON PROTEIN OF SOUTHERN BEAN MOSAIC VIRUS

Jingyu Shi August, 1995 32 Pages

Directed by: Claire A. Rinehart, Valgene L. Dunham, and Cheryl D. Davis

Department of Biology Western Kentucky University

There are four open reading frames (ORF) in three positive-sense translational phases of Southern Bean Mosaic Virus (SBMV). ORF2 codes for a 105 KDa polyprotein which is proposed to be autocatalytically cleaved into smaller functional proteins. Amino acid sequence homology between the SBMV and poliovirus, foot and mouth disease virus, and cowpea mosaic virus reveal that SBMV has a similar genomic organization to picornaviruses with a conserved order: Vpg (Viral protein, genome-linked)-protease-replicase. Two Gln-Ser (QS) amino acid pairs within the 105 KDa polyprotein, C75= 930-936, C60= 1305-1310, are proposed to be the cleavage sites based on 1) similar structural arrangement around the two QS pairs to that of known cleavage sites of picornaviruses and potyviruses; 2) computer prediction of the cleavage products from

the SBMV RNA sequence; 3) observed *in vitro* translation products. In order to investigate the role of the C60 QS pair in the proteolytic processing of the 105 KDa polyprotein, a mutagenic oligo was designed to create various amino acid substitutions in the QS pair. A coupled *in vitro* transcription/translation system was used to produce the protein products of both wildtype and mutants. During the procedure, tritiated leucine was incorporated into the protein products. The protein products were separated by SDS-PAGE and the cleavage patterns were detected after autoradiography. Two substitutions of the C60 QS pair--Pro-Val (PS) in $mutant_6$, Gln-Pro (QP) in $mutant_{19}$--reduced the yield of the 60 KDa protein by different levels. This result supports the hypothesis that the C60 QS is a cleavage site.

Introduction

The genome of southern bean mosaic virus (SBMV), type member of the sobemovirus group of plant viruses, consists of a single, positive-sense RNA. Turnip rosette virus (TRoSV), rice yellow mottle virus (RYMV), cocksfoot mottle virus (CfMV), and sowbane mosaic virus (SoMV) are four other members of the sobemovirus family. Among the four strains of SBMV, the cowpea strain (SBMV-cp) and the bean strain (SBMV-b) have been studied extensively with respect to their host range and other biological properties. As a pathogen of cultivated legumes, SBMV is distributed in the tropical and warm temperate parts of the world. SBMV-cp and SBMV-b have restricted host range. SBMV-cp infects Vigna sinensis (cowpea) systemically but is not transmissible to the garden bean, while SBMV-b readily infects garden bean but not cowpea. The other two strains of SBMV, Mexican and Ghana, have wider host range and cause severe diseases (Sehgal, 1981).

The virons of SBMV are icosahedral with 180 copies of a single capsid protein (molecular weight 28,214 Da) occupying a T=3 lattice (Hermodson *et al.,* 1982). Ghosh *et al.* (1979) reported that SBMV viral RNA has a 5`-linked protein, but lacks a 3` terminal poly(A). The 5`-protein, VPg (viral protein genome-linked), is covalently linked to genome RNA (Mang *et al.* 1982) and is required for full infectivity of

SBMV RNA (Veerisetty and Sehgal, 1979).

The genomic RNA sequence of SBMV-cp was determined by Wu *et al.* (1987). The 4194 nucleotide genome has three translational phases and four open reading frames (ORF) as predicted by the sequence using computer modeling (see Figure 1). ORF 1 (base 49 to 603) translates a product of 20 KDa with undetermined function. ORF4 (base 3217 to 4053), in the same phase as ORF1, produces the SBMV capsid protein via a subgenomic RNA. Extending from base 1895 to 2380, ORF3 is predicted to code for a 18.4 KDa protein in the second phase. ORF 2, in the third phase, codes for a large protein of 105 KDa, and it overlaps the other three ORFs (Wu *et al.,* 1987). Mang *et al.* (1982) used cell-free extracts to translate SBMV virion-extracted RNAs and 4 proteins were detected, CP1 (100 KDa), CP2 (70 KDA), CP3 (30 KDa), and CP4 (20 KDa). Wu *et al.* (1987) reported the following *in vitro* translation products: P1 (105 KDa), P2 (60 KDa), P3 (28 KDa), and P4 (3 proteins with molecular weights ranging 21-25 KDa). The correlation between the predicted ORFs by sequence and the observed translation products indicates that P4 may be coded by ORF1 since it codes 185 amino acids with a molecular weight of 21 KDa. A 18.4 KDa protein, related to the product of ORF3, has never been observed in proteins from *in vitro* translation. Protein P2 (60 KDa) is considered to be a subset of P1 by proteolytic processing (Wu *et al.,* 1987). The sequence coding for P1 should reside in ORF2 which is the only ORF large enough to produce a protein of approximate molecular weight of 105 KDa. The 105 KDa protein is

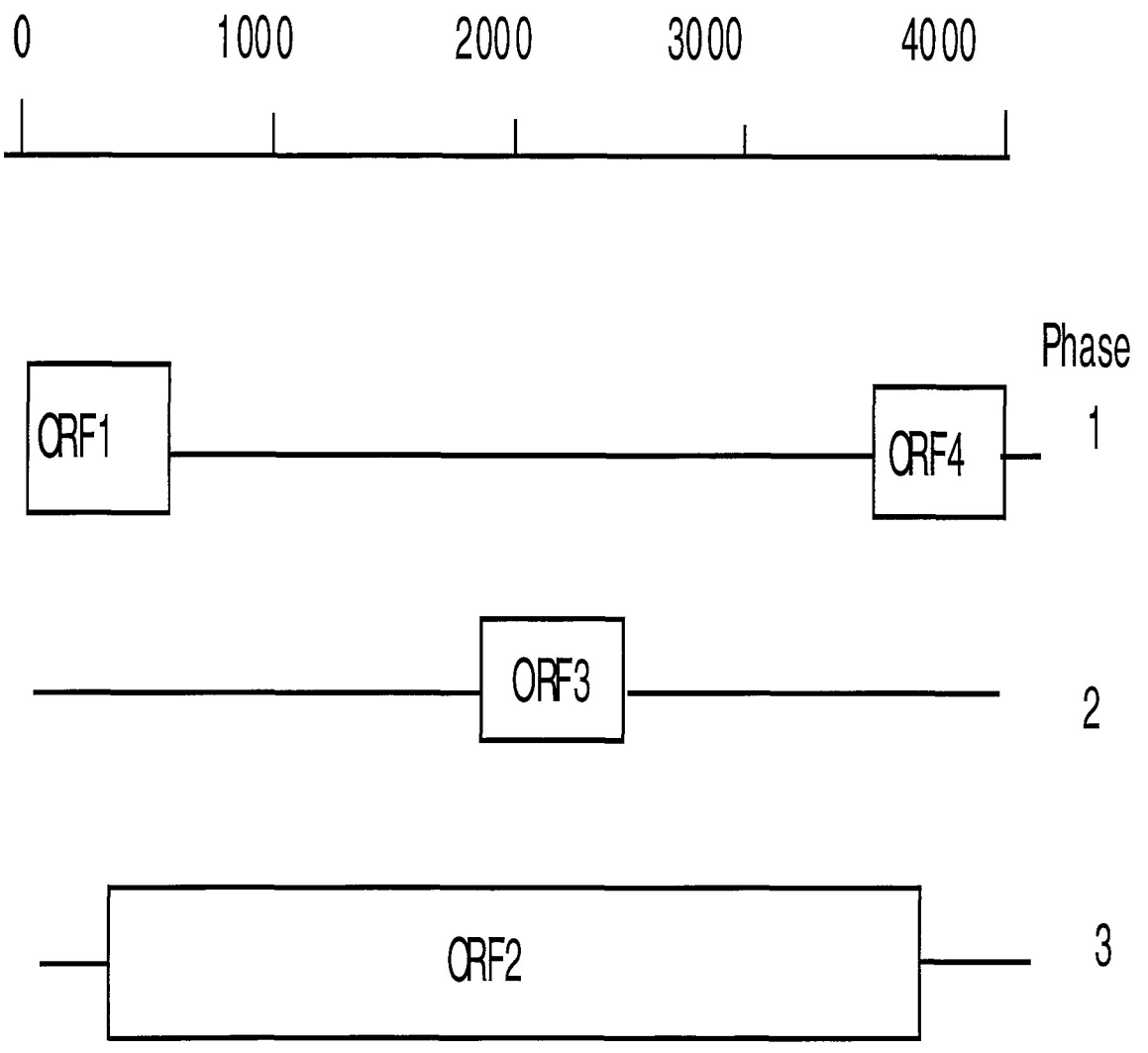

Figure 1. Open reading frames of SBMV-cp. Open reading frames in the three positive-sense translational phases. Each line shows one of the translational phases. The boxes show the open reading frames. According to Wu *et al.* (1987).

the major concern in this study, and it is proposed to be a polyprotein which undergoes autocatalytic cleavage to produce smaller functional proteins (Wu *et al.,* 1987; Shannon,1991). Correlation in the genomic organization between the putative 105 KDa SBMV sequence and the polyprotein sequence of picornavirus as well as comovirus has been found by computer comparison. Bases 2067-2126 of SBMV show homology to the VPg sequence from Cowpea Mosaic Virus (CPMV) and 17 picornaviruses (Fig.2A; Wu *et al.,* 1987). Bases 2946-2954 of SBMV code for a tripeptide, Gly-Asp-Asp (GDD) (Wu *et al.,* 1987), found in all RNA virus polymerases (Fig. 2B; Argos *et al.,* 1984). By aligning the 105 KDa protein to polyproteins of CPMV (comovirus), TVMV (potyvirus) and picornaviruse, Gorbalenya and his colleagues (1988) hypothesized that the SBMV protease was located between VPg and RNA polymerase (from bases 161 to bases 295) within the polyprotein. The putative protease with a molecular weight of 20 KDa is derived by cleaving at 2 Glu-Ser (ES) sites and an alternate Glu-Tyr (ET) site. The molecular weight of the known poliovirus 3C protease is 20 KDa (Hanecak *et al.,* 1984). This cleavage also produces a 60 KDa protein which is consistent with the *in vitro* products of SBMV RNA reported by Wu *et al.* (1987). In a deletion study of SBMV-cp RNA at the 3` end, the molecular weight of *in vitro* translation products shifted as deletion size increased. The 60 KDa protein remained the same until a large deletion was made in the RNA at the 3` end (C.A. Rinehart, unpublished results), indicating that the 60 KDa protein is in the N terminus and thus

A.

```
                                    *         *
                                    * * *   * *
CONSENSUS                 E V D F S K F D - S
                          .   .   .   . .   .
SBMV      (2661-2690)     E A D I S G F D W S
                            *       *       *
                          * *       *       * * * *

CONSENSUS                 S G - S   T F I I N T I L
                          . .       .       . .   .
SBMV      (2859-2897)     S G S Y C T S S T N S R I
                                *   *   *
                          *     *   *   *   * *

CONSENSUS                 I - - G D D S L I
                          .     . . . . .
SBMV      (2937-2963)     I A M G D D S V E
```

B.

```
                        *
CONSENSUS          G P Y A G P - Q R Q - P L V K V A K - P
                   .   . .     .           .       .
SBMV               S L F P P K P R A T S S K P I T T S S P
(2067-2126)
```

Figure 2. Amino acid alignment of nonstructural domains identified with putative functions. The numbers following SBMV represent the range of nucleic acid sequence coding for the amino acids displayed. The letters in the consensus lines represent the most prevalent amino acid with a homology of at least 60% and the hyphens show where the homology is less than 60%. The dots between the lines mark identical or equivalent residues between two sequence. A single asterisk indicates that all the aligned residues are equivalent. (**A**) The putative RNA dependent RNA polymerase domain. The consensus sequence was obtained from Argos *et al.* (1984). (**B**) The putative Vpg domain, taken from Wu *et al.* (1987).

the protease would not be located in the region proposed by Gorbalenya *et al.* (1988).

The organization of the SBMV 105 KDa polyprotein resembles the polyproteins from the picornavirus superfamily (Figure 3) based on amino acid sequence homologies (Shannon, 1991). The order of the conserved organization is nucleotide binding domain, VPg, protease, and replicase. The protease is hypothesized to cleave at two QS (glutamine-serine) and one ES (glutamic-serine) sites which border functional proteins within the polyprotein. Context analysis of the two putative sites, the C75 QS pair and the C60 QS pair, indicate that the two QS pairs have a high possibility of being in the surface of a pocket structure (Shannon, 1991). This structure is consistent with the cleavage domains observed in picornaviruses and potyviruses (Palmenberg, 1990; Dougherty *et al.,* 1990).

The putative protease is a trypsin-like serine protease and serine 558 proposed to be the catalytic site (Shannon, 1991). Site-directed mutagenesis to change the codon for serine 558 to a glycine indicated that serine 558 was not the proposed catalytic site because the *in vitro* translation products of the mutant cleaved the same as that of the wild type (Simmons, 1993).

A QS amino acid pair was defined as a proteolytic cleavage site within the 200 Kda polyprotein of CPMV (Wellink *et al.,* 1986; Garcia *et al.,* 1987). A model for the polyprotein cleavage site of tobacco etch virus (TEV) was proposed by Dougherty *et al.* (1989). In their

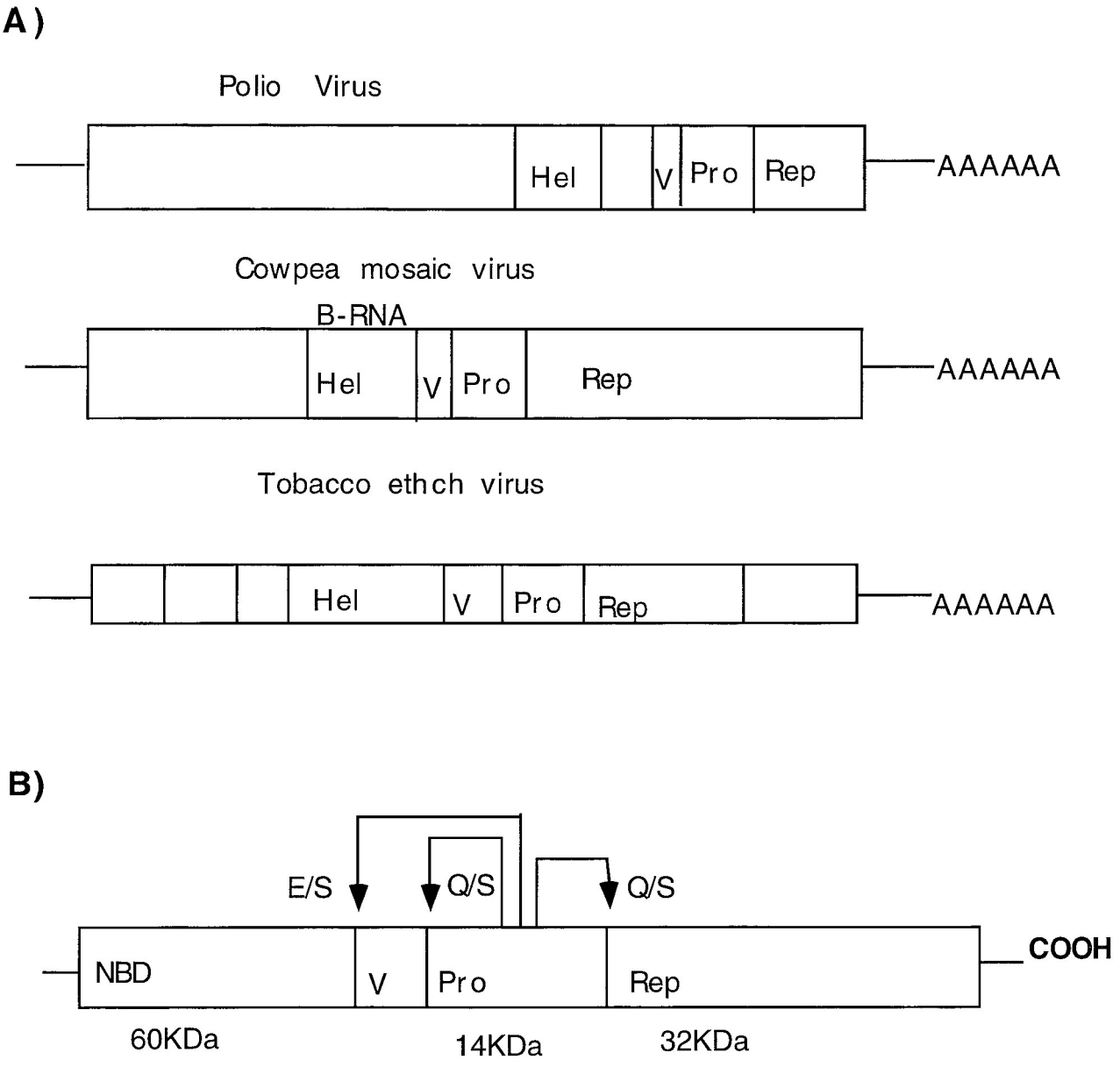

Figure 3. A). Genomic organizations of viruses belonging to the picornavirus superfamily. Notations are as follows: Hel=helicase; Pro=protease; Rep=replicase, V=Vpg, NBD=nucleotide binding domain. According to Turnbull-Ross *et al.* (1993). **B).** Hypothesized organization of the 105 Kda polyprotein of SBMV. Taken from Shannon (1991).

model, the polyprotein expressed from the RNA genome of TEV was co- and post-translationally processed by a viral encoded 49 KDa protease. There are five cleavage sites each defined by the seven amino acid consensus sequence (* is the cleavage location and / designates alternative amino acids).

```
   P6  P5   P4   P3   P2    P1   P1`
-Glu--Xaa--Xaa--Tyr--Xaa--Gln*Ser/Gly
```

The QS (Gln*Ser) pair is observed as a cleavage site in picornaviruses and is a legitimate candidate for proteolytic processing by viral encoded protease (Palmenberg, 1990).

The current study of viral polyprotein processing focuses on two aspects: finding the active site of the proposed protease and mapping the cleavage sites and sequences flanking the putative cleavage sites.

The complex pattern of proteins produced by SBMV from such a small genome requires a mixture of expression strategies. These strategies include overlapping reading frames, production of subgenomic RNA, internal initiation or reinitiation of translation and possible post-translational proteolytic processing and modification (Wu *et al.,* 1987). This researcher used site-directed mutagenesis to change the putative C60 QS site to various amino acid pairs to determine the role of the QS pair in the proteolytic processing. Two mutants with the substitutions of PV (Pro-Val) and QP (Gln-Pro) resulted in reduced levels of cleavage products, suggesting that the C60 QS site is indeed one of the cleavage sites.

Materials and Methods

Computer Analysis

MacVector (IBI) was used to analyze the genomic and amino acid sequences of the putative 105 KDa polyprotein, predict fragment sizes of restriction digests and read the DNA sequences of both wild type and mutants.

SBMV(cp) Clones

pSLORF18, previously described by Shannon (1991), contains the cDNA of SBMV which codes for all amino acids of ORF1 and ORF2 except for 2 amino acids. After Hind III digestion of pSLORF18, a 3462 bp fragment was isolated from an agarose gel and subcloned into the HindIII cloning site of the pSELECT-1 plasmid (Promega). The total 9142 base recombinant construct, pSELORF, was used for ssDNA synthesis, mutagenesis, and *in vitro* transcription and translation (Figure 4).

Preparation of Competent Cells

The bacterial cell lines JM109 and BMH71-18 were made competent by treatment with either 0.1M $CaCl_2$ (Maniatis *et al.,* 1982) or TSS (20% polyethylene glycol, 100mM $MgCl_2$, and 10% dimethyl sulfoxide) buffer (Chung *et al.,* 1989).

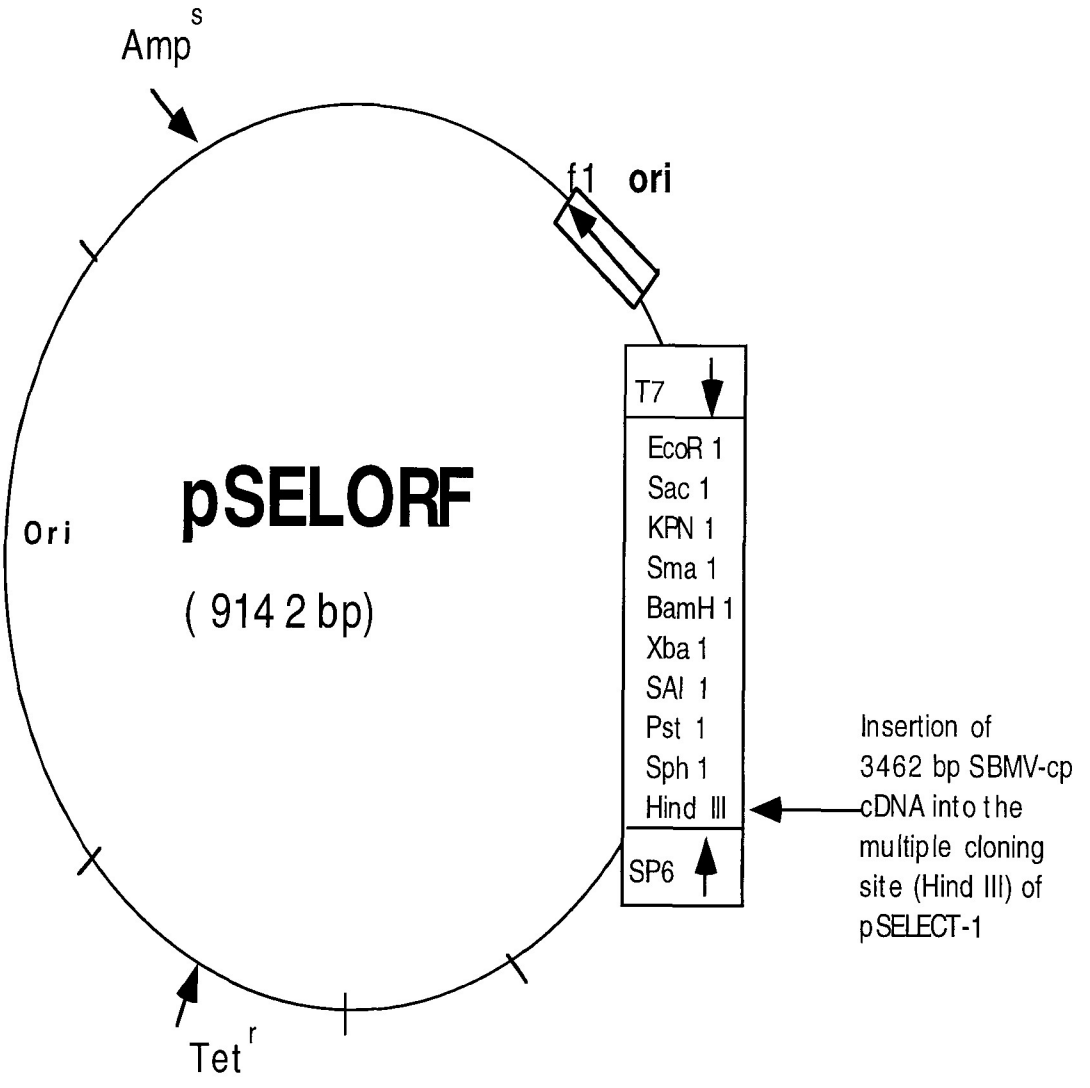

Figure. 4. Construction of pSELORF. The 3462 bp SBMV-cp cDNA fragment from pSLORF18 was subcloned to the multiple cloning site of pSelect-1. The arrow of the f1 ori shows the orientation of the insertion and transcription. SP6 RNA polymerase was used to transcribe the template DNA.

Transformation

Plasmids pSELECT-1, pSLORF18, recombinant pSELORF, and mutants pSELORF were amplified by transformation into competent bacterial cells, JM 109 or BMH71-18 (Promega, 1991a).

Restriction Endonuclease Digestion

Restriction analysis was performed according to the special instructions described by the supplier. Hind III (USB), Nde1 (Phamcia), and Xho1 (USB) were used to subclone and to screen recombinants and mutants.

Ligation

Ligation of HindIII digested pSELECT-1 and the 3462 bp fragment of Hind III digested pSLORF18 was performed by using T_4 DNA ligase. The reaction was set up according to the supplier`s specifications (USB) with the exception that incubation was at 12°C for 20 hours instead of 14°C overnight.

Screening of Recombinant pSELORF

Recombinant DNA was screened according to the Blue/White Screening method as outlined in the Protocols and Applications Guide (Promega, 1991a).

Preparation of Single Stranded Plasmid DNA

Single stranded pSELORF was prepared by superinfecting phage R408 into pSelorf- transformed JM109 at m.o.i (multiplicity of infection) 10. Phage precipitation solution (3.75M ammonium acetate, pH7.5, 20% polyethylene glycol) was used to precipitate the R408 (Promega 1991b).

Mutagenic Oligonucleotide Primer Design and Phosphorylation

A 30-mer mutagenic oligo (Fig.5) was designed to change the C60 QS site to various other amino acids pairs. The mutagenic primer allowed for the mutation of the first 2 bases of serine and all three bases of the glutamine. The primer was designed so that mutants showed an additional new Ava1 or Xho1 restriction site at the primer location. The 5` end of the mutagenic primer was phosphorylated using T_4 polynucleotide kinase as described in the Promega Protocols and Application Guide (Promega, 1991b).

Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the Altered Sites™ System (Promega 1991b). Several modifications were made in the mutagenesis annealing reaction: 1) ampicillin repair primer was 0.125 pmol (0.5 µl) instead of 0.25 pmol (1 µl); 2) mutagenic primer was increased to 5 pmol instead of 1.25 pmol.

Plasmid Isolation and Purification

A large scale alkaline lysis preparation of mutagenized pSELORF DNA was prepared and purified by centrifugation in cesium chloride-ethidium bromide gradients (Maniatis *et al.,* 1982). Before ultracentrifugation, the DNA pellet was dissolved in 4.5 ml of 1X TE buffer (25 mM Tris-cl, pH 8.0, 10 mM EDTA). The dissolved DNA was mixed with 5.1 grams of CsCl and 0.75 ml of 10 mg/mL ethidium bromide. A 2.7 ml aliquot of the mix was then put into thick-walled polycarbonate tubes. The gradients were run at 180,000Xg for 20 hours.

(A): 2172...5`-CCA GGA AGC TGC GTT AGA AGC TCG TTC AAC-3`...
... GGT CCT TCG ACG CAA TCT TCG AGC AAG CCG...2201
Q * S

(B): ... 5`-CCA GGA AGC TGC GNY NN***G AGC*** TCG TTC AAC-3`
Xho1 or Ava1 site

Figure 5. Mutagenic primer designed for site-directed mutagenesis. (A) is the original sequence in the C60 region in pSELORF. (B) is the mutagenic primer designed to generate mutants. The *italicised* and bolded ***G AGC*** indicates a new Xho1 or Ava1 restriction site after mutation.

Agarose Gel Electrophoresis

Agarose gels were prepared according to the method developed by McDonnell *et al.* (1977). Agarose gels contained 0.9% agarose for isolating the Hind III fragment and 1.5 % agarose for screening the mutants. All gels were run for 1.5 hours at 10 volts/cm in 1X TBE electrophoresis buffer (0.089 M Tris pH7.2, 0.089 M Boric acid, 2mM EDTA, and 0.5µg/ml of Ethedium bromide). DNA was visualized in the UV transilluminator. Photos of the gels were taken using polaroid 667 film (F=5.6, Tiffen orange filter) with 1 second exposure for the 0.9% gel and 2 second exposure for the 1.5% gel.

Alkali Denaturation of Double-Stranded plasmid DNA for Sequencing

Both wildtype pSELORF and mutants were denatured according to Hattori *et al.* (1986).

DNA Sequencing

The dideoxy nucleotide chain termination method (Sanger *et al.,* 1977) was used to sequence plasmid DNA. The sequencing kit was supplied by USB. The sequencing primer, 5`-cccaggtcgaagccccttgc-3`, was synthesized by UK using the Applied Biosystems 394A DNA synthesizer. The chain termination products were labeled with [a-thio-^{35}S]dATP (1000-1500 Ci/mmol) (Amersham). Reaction products were loaded onto a sequencing gel (20X40cm) which contained 6% polyacrylamide, 1XTBE and 8.3M urea. The gel was run at 35 watts for 1.5 hours. The sequencing gel was then fixed in a solution containing 10% methanol and 10% glacial acetic acid for 30 minutes and dried at 80°C for 60 minutes in a gel drier (Model 583, Biorad). Kodak X-OMAT AR X-ray film was used to carry out autoradiography. After 48 hours,

the films were developed for 3 minutes followed by a 2 minute rinse in water and 5 minutes in fixer.

Coupled *in vitro* Transcription and Translation

The T_NT^{TM} coupled reticulocyte lysate transcription/translation system (Promega, 1993) was used to translate protein products of wildtype pSELORF and six mutants. SBMV-cp RNA, pSLORF 18 and M_{22} (Ser 558 muted to Gly) from Simmons (1993) were used as controls to test the activity of the lysate. Tritiated L-leucine (ICN 122Ci/mmol) was used to label translation products. All reactions were performed according the Promega Technical Bulletin 126 (Promega, 1993) except that the total reaction volume was 25 µl instead of 50µl.

SDS-PAGE and Visualization of Protein Products

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (Laemmli, 1970) in a mini-gel format was used to separate all translation products. The stacking gel and separating gel contained 4.5% and 12.5% polyacrylamide, respectively. Three microliters of prestained low molecular weight marker (Bio-rad) ranging from 106 KDa to 18.2 KDa were loaded in one of the wells for molecular weight comparison of the products. The stacking gel was run at 20mA and was increased to 35mA for the separating gel. Following a 30 minute fixation (10% methanol and 10% glacial acetic acid), the gel was soaked in water for 30 minutes and then in Fluro-Hance (RPI). Fluro-Hance was used to enhance the resolution images of the weak beta emitter, tritium, during the process of fluorography.

The fluorography was performed at -70°C for three weeks. The film was developed as described previously.

Results

The 105 KDa polyprotein of SBMV-cp was proposed to possess autocatalytic function which resulted in the release of smaller functional products. Two QS pairs were hypothesized to be cleavage sites within the 105 KDa protein. Site-directed mutagenesis was used to create various amino acid replacements in the C60 QS site to identify the role of the QS site in proteolytic processing and to determine functional and nonfunctional amino acids substitutions.

The 3462 bp fragment isolated from the agarose gel after Hind III digestion was inserted into the multiple cloning site of pSELECT-1. Recombinant colonies were white after growing in the LB plate containing 25 µg/ml of tetracycline, 0.5mM IPTG and 40 µg/ml X-gal for 16 hours (Promega, 1991b). The orientation of the insert was confirmed by Nde1 restriction digestion. The required orientation has three fragments of 3588 bp, 3317 bp, and 2237 bp while the reverse insertion would result in different fragments of 5537 bp, 2237 bp, and 1368 bp.

Site-directed mutagenesis was not successful on the first attempt following the protocol outlined in the Promega Protocol and Application (Promega, 1991c). Sixty ampicillin resistant colonies were generated, but none of them appeared to be mutant. Since the mutagenic primer (Figure 5) has 5 mismatch bases and therefore

could result in poor annealing to the DNA template, the mutagenic primer`s concentration was increased from 1.25pmol to 6pmol. The ampicillin oligonucleotide primer`s concentration was reduced in half (from 0.05pmol to 0.025pmol. In addition, 8 units instead of 4 units of T_4 polynucleotide kinase were used to ensure proper phosphorylation of the mutagenic primer. Following the modified reaction, forty ampicillin resistant colonies were selected for screening. Plasmid DNA was isolated by plasmid miniprep and approximately 1μg of DNA from each sample was used to set up Xho 1 restriction digests. Prediction of the restriction fragment size was assisted by the restriction analysis program in MacVector. The MacVector DNA analysis program predicted 8070 bp and 1072 bp for the wildtype while the predicted fragments for mutant would be 8070 bp, 979 bp and 93 bp. The 1072 bp of wildtype was 54 bp larger than the 1018 bp band of the 1 Kb marker and the mutants` 979 bp band was only 39 bp less. Since it was difficult to distinguish the differences between 50 bases in the agarose gel, a combination of the wildtype and possible mutant gave the most credible method to verify the mutant (Figure 6). If there were two bands near the 1072 marker, the clone was designated as a mutant (lane 1, 3, 4, 7, 9, 10, 11, 14, 16, and 18). If there was only one band approximately 1072 bp, the clone was designated as a wildtype (lane 13, and 15) . After screening of the forty colonies, 12 mutants were found. Six of them were isolated and purified by CsCl gradients for further analysis.

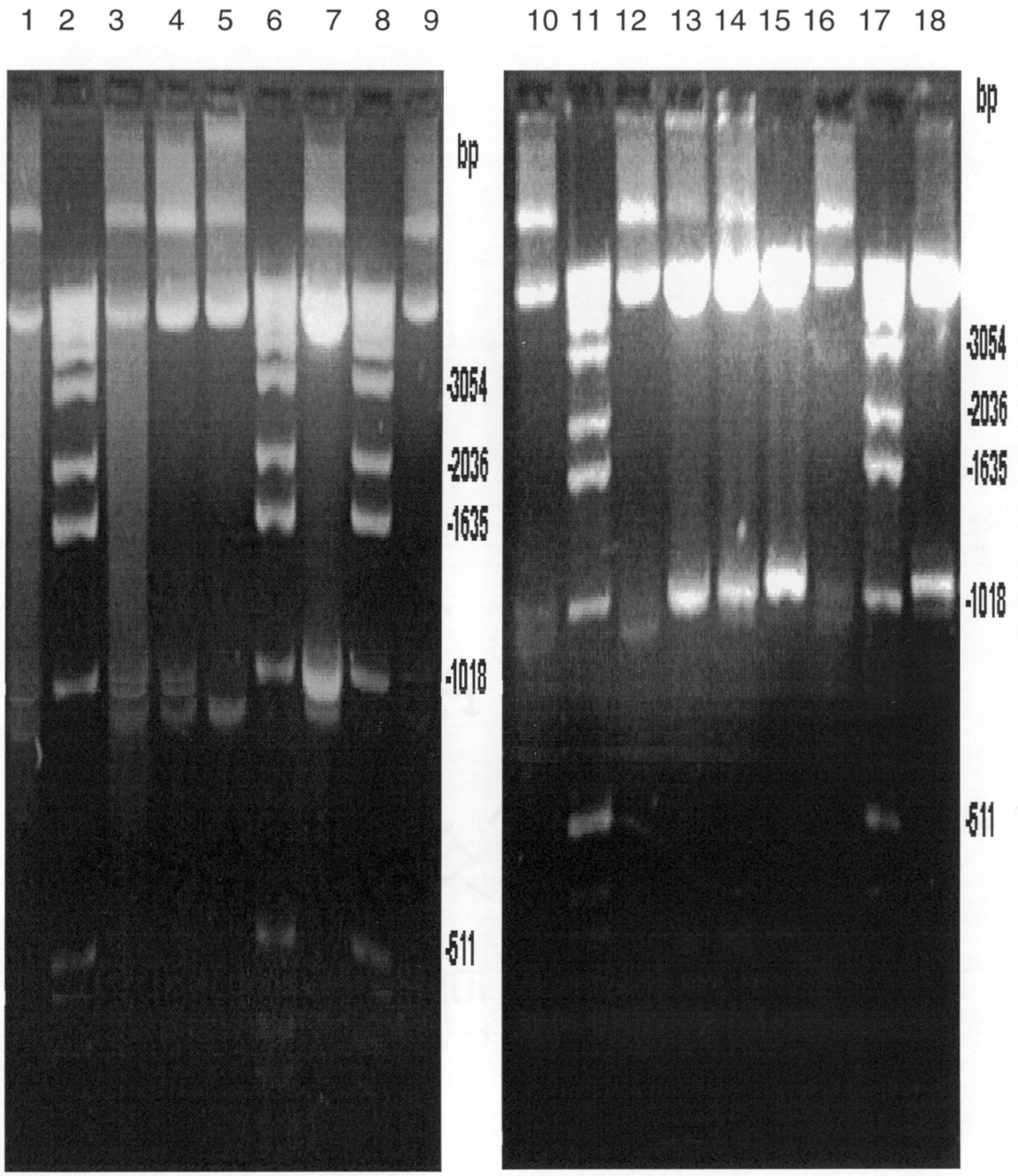

Figure 6. Agarose gel of mutants and wildtype after Xho1 restriction digestion. Each sample contained approximately 1μg of DNA. Mutants were loaded together with wild-type. Lane 1, 3, 4, 7, 9, 10, 12, 14, 16, and 18 were mutants. Lane 13 and 15 were wildtypes. Lane 2, 6, 8, 11, and 17 were 1 Kb DNA markers. Lane 5 was loaded with two unknown clones, the single band looks like 979 but was not confirmed.

The T_NT^{TM} coupled reticulocyte lysate transcription/translation system was used to synthesize labeled protein products of both wild type and the six mutants. A control was set up without DNA but compensated with DEPC-treated ddH_2O to determine the purity of the lysate system. During the reaction, tritiated leucine was incorporated into the polyprotein and the translation products were then separated on a 12.5% SDS-PAGE gel. After fixing the gel, fluorhance was used to increase detection of the radioactivity and facilitate the visualization of the translation products.

The wildtype and mutants generated different band patterns on the SDS-PAGE gel (Figure 7). A very large band (molecular weight approximately 160 KDa) was found in wildtype, mutant M_6 (lane 1), mutant M_{19} (lane 4) and SBMV RNA (lane 6). The wildtype (lane 2) had 4 additional bands, 75 KDa, 60 KDa, 40 KDa, and three proteins in the 20 KDa range. For mutant M_6, a very light 60 KDa was detected while mutant M_{19} has a relatively dense 60 KDa band and a 40 KDa band. The translation products of RNA included identical 60 KDa and 40 KDa bands but did not include the 20 KDa range proteins. The pSLORF 18 (lane 3) had only one 60 KDa product. No product was detected in the control reaction (lane 7) and in Simmons` M_{22} (lane 5).

In order to verify the fidelity of the mutation and to determine the specific amino acid substitutions in the C60 QS site, three mutants (M_6, M_{18}, and M_{19}) and wild-type pSELORF were sequenced using the dideoxy chain termination method. Reading of the sequencing gel was assisted by IBI`s MacVector. In M 6, the glutamine was replaced by proline; the serine was replaced by

valine. For M_{18}, although the bases that code for serine were changed to TCC instead of the original TCT, the TCC still coded for serine due to replacement in the wobble position. The QS site on the protein remains the same. The C60 QS pair was replaced by QP (glutamine-proline) in the $mutant_{19}$ (Figure 8).

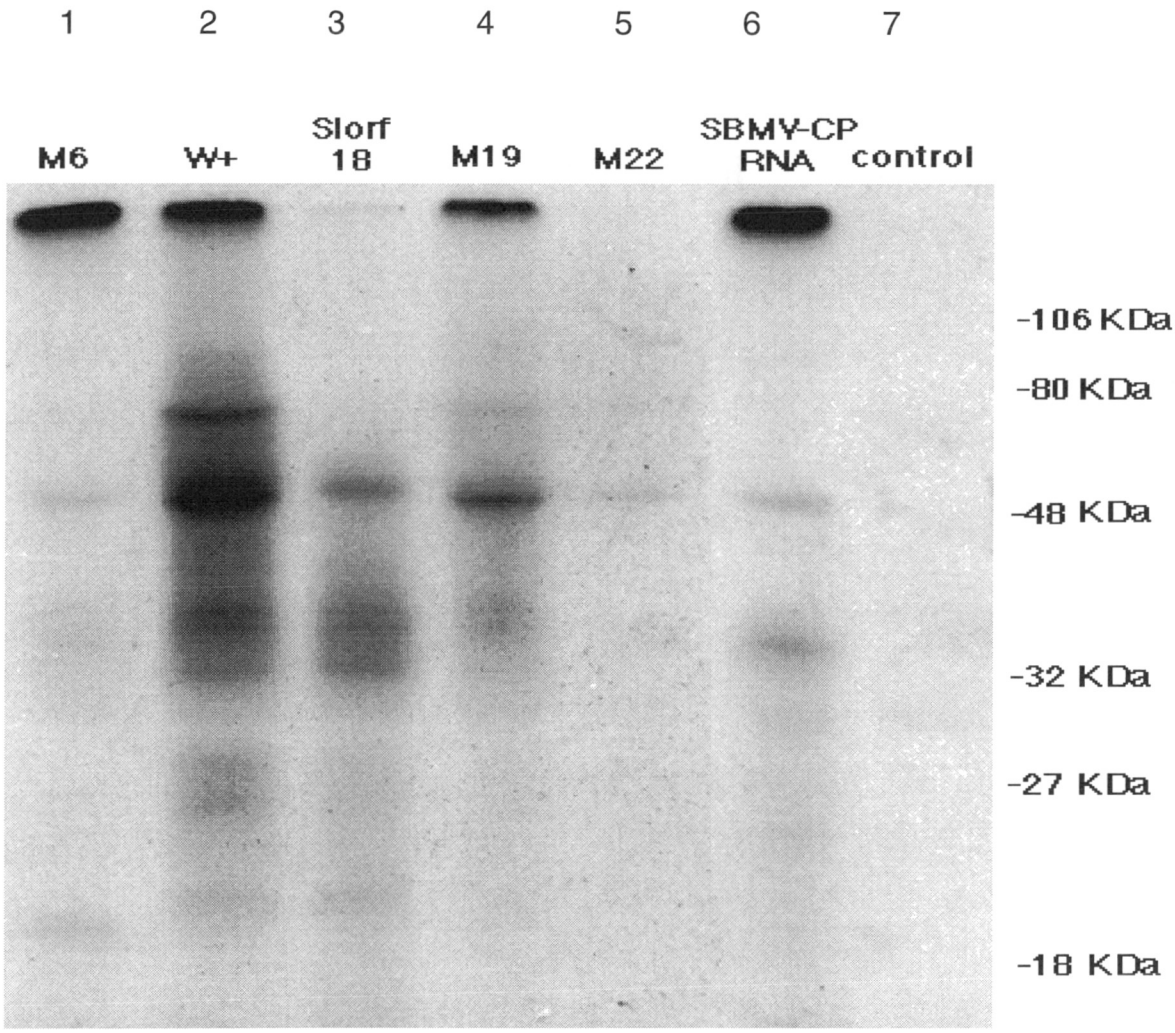

Figure 7. Autoradiography of *in vitro* translation products separated on 12.5% SDS-PAGE. Each sample contained approximately 200 ng labled proteins. Lane 1, mutant M_6; Lane 2, wildtype; Lane 3, pSLORF 18; Lane 4, mutant M_{19}; Lane 5, mutant M_{22}; Lane 6, SBMV-cp RNA, and Lane 7, control.

Q * S

Original: 2172--GGT CCT TCG ACG CAA TCT TCG AGC AAG TTG--2201

P V

Mutant 6: 2172--GGT CCT TCG ACG CCA GTC TCG AGC AAG TTG--2201

Q S

Mutant 18: 2172--GGT CCT TCG ACG CAA TCC TCG AGC AAG TTG--2201

Q P

Mutant 19: 2172--GGT CCT TCG ACG CAA CCC TCG AGC AAG TTG--2201

Figure 8. Interpretation of sequencing gel of wild-type and mutants. M_6 represents a PV replacement of the C60 QS site. M_{18} is a mutant but due to the wobble, the amino acid pair is the same. M_{19} is a QP substitution.

Discussion

The 105 KDa polyprotein produced by the largest ORF of SBMV-CP has been hypothesized to undergo autoproteolytic cleavage to release functional proteins. Comparison studies between the SBMV genome and other known viral proteases indicates that the 105 KDa polyprotein may contain a protease located between the putative VPg and the replicase. Two QS (glutamine-serine) and one ES (glutamate-serine) are proposed to be cleavage sites based on the observed *in vitro* cell-free translation products and homologies between the SBMV sequence and known polyproteins (Shannon, 1991). The current study used site-directed mutagenesis to determine if the C60 QS is a true cleavage site and if so which substitutions of amino acid pairs are functional and which are not.

Twelve mutants were found out of forty colonies screened. Upon sequencing of both the wildtype and three mutants, the mutations were confirmed. In the mutant $M_{6,}$ the QS pair was replaced by a PV pair. In the mutant $M_{19,}$ the QS pair was substituted by a QP pair. For the mutant M_{18}, although the codon coding for serine was changed from TCT to TCC, it still coded for serine.

Several *in vitro* translation products have been observed with molecular weights of 75 KDa, 60 KDa, 40 KDa and a group of three proteins ranging from 21-29 KDa in the wildtype (lane 2, Figure 7).

Mang *et al.* (1982) reported similar *in vitro* translation products 105 KDa, 60 KDa, 40 KDa and a group of three protein with the molecular weights ranging from 21-25 KDa. In the mutant M_6, the 60 KDa protein was synthesized at lower levels than in the wildtype. The 60 KDa protein in M_{19} was present in higher levels than in M_6 but in lower levels than in the wildtype. The 60 KDa protein is the cleavage product of the 105 KDa polyprotein. The reduced yield of the 60 KDa protein indicates that the proteolytic processing is inhibited or at a reduced level. This result indicated that the two amino acid replacement (M_6) had a greater effect on proteolytic cleavage than with one amino acid substitution (M_{19}).

The low level of protein products obtained from SBMV RNA as compared to the wildtype cDNA may suggest that the coupled transcription/translation system favored the DNA template.

Two results in this study remain to be explained. First, the high molecular weight band (possible multimers of the 105 KDa polyproteins) appeared in M_6, M_{19}, wildtype, and SBMV RNA. These proteins were not soluble even when the SDS in the loading buffer was increased to 4%. No published papers have reported such an aggregated large band. Second, there was little translation product for M_{22} (Serine 558 mutated to glycine, Simmons, 1993). Site-directed mutagenesis is a powerful tool for locating the active sites of proteases and determining functional cleavage sites. Using site-directed mutagenesis, one Tyr-Gly junction and 9 Gln-Gly junctions were determined to be the proteolytic cleavage sites of poliovirus 2A and 3C protease (Baum *et al.,* 1991). All of the cleavage sites in

the encephalomyocarditis viral capsid region were found to be invariable between Gln/Glu * Gly/Ser (* indicates a cleavage site; / stands for an alternative amino acid) (Parks *et al.,* 1988). The protease of tobacco vein mottling virus (TVMV) recognizes the sequence Val-Arg/Lys-Phe-Gln*Ser/Gly (Domier *et al.,* 1986; Hellmann *et al.,* 1988). A consensus amino acid motif of Val-Xaa-Thr/Hls-Gln*Ser/Ala-Thr is found in plum pox virus (Maiss *et al.,* 1989). In the cleavage model proposed by Dougherty *et al.,* (1989), the five cleavage sites for 49 KDa TEV protease were defined as

```
P6   P5   P4   P3   P2      P1   P1`
-Glu--Xaa--Xaa--Tyr--Xaa--Gln*Ser/Gly.
```

Mutations in P5, P4, and P2 had no effects on cleavage. Amino acid substitutions at P6, P3, P1, and P1` resulted in elimination or a major reduction in cleavage. Dougherty and his colleagues also pointed out that no replacement at any of the five conserved sites could totally eliminate a cleavage.

Results of the present study support Dougherty`s model. The substitution of the QS pair with PV and QP yielded reduced levels of the 60 KDa cleavage product. Neither mutation totally eliminated the cleavage, but followed a trend where more substitution resulted in less cleavage. The results also support the hypothesis that the C60 QS pair is a cleavage site. However, further study is necessary to generate all possible amino acid substitutions at C60 QS site to determine the specificity of the QS site. It will be important to investigate the prolonged incubation of the translation reaction to determine if substitutions reduced the rate of the proteolytic

processing only. The sequence flanking the QS site should also be analyzed by site-directed mutagenesis and deletion studies to determine all the sequence required for a cleavage.

Literature Cited

Argos, P., Kamer, G., Nicklin, M. J. H., Wimmer, E. (1984). Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these families. Nucleic Acid Research **12,** 7251-7267.

Baum, E. Z., Bebernitz, G. A., Palant, O., Mueller, T. (1991). Purification, properties, and mutagenesis of 3C protease. Virology. **185**, 140-150.

Chung, C. T., Niemela, S. L., Miller, R. H. (1989). One-step preparation of competent *E. coli* : Transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences U.S.A. **86**, 2172-2175.

Domier, L.L., Franklin, K. M., Shahabuddin, M., Hellmann, G. M., Overmyer, J. H., Hiremath, S.T., Siaw, M. F., Lomnossoff, G. P., Shaw, J. G., Rhoads, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. Nucleic Acid Research **14**, 5417-5430.

Dougherty, W. G., Gary, S. M., Parks, T. D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: A model. Virology **171**, 356-364.

Dougherty, W. G., Parks, T. D., Smith, H. A., Lindbo, J. A. (1990). Expression of the potyvirus genome: the role of proteolytic processing in viral genes and plant pathogenesis. (T.P. Pirone, and J.G. Shaw, Eds.) pp 124-139, Springer-Verlag, New York.

Garcia, J. A., Schrijvers, L., Tan, A., Vos, P., Wellink, J., Goldbach, R. (1987). Proteolytic activity of the cowpea mosaic virus encoded 24KDa protein synthesized in *E. coli.* Virology **159**: 67-75.

Ghosh, A., Dasgupta, R., Salerno-Rife, T., Rutgers, T., Kaesberg., G. (1979). Southern bean mosaic viral RNA has a 5`-linked protein but lacks 3` terminal poly (A). Nucleic Acid Research **7**, 2137-2146

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M., Donchenko, A. P. (1988). Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. FEBS Letters **236**, 287-290.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., Koonin, E. V. (1989). Cysteine proteases of positive strand RNA virus and chymotrypsin-like serine proteases. FEBS Letters **234**, 103-114.

Hanecak, R., Semler, B. L., Ariga, H., Anderson, C. W., Wimmer, E. (1984). Expression of a cloned segment of poliovirus in *E . coli:* Evidence for autocatalytic production of the proteinase. Cell **37**, 1063-1073.

Hattyori, M., Sakaki, Y. (1986). Dideoxy DNA sequencing method using denatured plasmid templates. Annual Biochemistry **152**, 232-238.

Hellmann, G. M., Shaw, J. G., Rhoads, R. E. (1988). *In vitro* analysis of tobacco vein mottling virus nia cistron: Evidence for a a virus-encoded protease. Virology **163**, 554-562.

Hermodson, M. A., Abad-Zapatero, C., Abdel-meguid, S. S., Pundak, S., Rossmann, M. G., Tremaine, J. H. (1982). Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three dimentional structure of the virus. Virology **119**, 133-149.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriiophage T_4. Nature (London) **227**, 680-685.

Maiss, E., Timpe, U., Brisske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D., Katinger, H. W. D. (1989). The complete nucleotide sequence of plum pox virus RNA (PPV). Journal General Virololgy **70**, 513-524.

Mang, K. Q., Ghosh, A., Kaesberg, P. (1982). A comparative study of the Cowpea and Bean Strains of Southern Bean Mosaic Virus. Virology **116**, 264-274.

Maniatis, T., Fritsch, E. F., Sambrook, J. (1982). In *Molecular Cloning* : *A Laboratory Manual.* Pp152-159 Cold Spring Harbor, New York.

McDonnell, M. W., Simon, M. N., Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA determination of molecular weight. Journal Molecular Biology **110**, 119-146.

Palmenberg, A. C., Kirby, E. M., Janda, M. R., Drake, N. L., Duke, G. M., Potratz, K. F., Collet M.S. (1984). The nucleotide and deduced amino acid sequences of the encephalomyocarditis viral polyprotein coding region. Nucleic Acid Research **12,** 2969-2985.

Palmenberg, A. C. (1990). Proteolytic processing of picornaviral polyprotein. Annual Review Microbiology **44,** 606-623.

Parks, G. D., Baker, J. C., Palmenberg, A. C. (1988). Proteolytic cleavage of encephalomyocarditis virus capisid region substrates by precursors to the 3C Enzyme. Journal of Virology **63**, 1054-1058.

Promega Protocols and Application Guide (1991a). (Second Edition). Plasmid cloning and transcription *in vitro.* pp 45-72. Promega Corporation, Madison, WI.

Promega Protocols and Application Guide (1991b). (Second Edition). Nucleic acid sequencing and mutagenesis pp 75-121 Promega Corporation, Madison, WI.

Promega Technical Bulletin 126 (1993). The T_nT^{TM} coupled reticulocyte lysate transcription/translation system. Promega Corporation, Madison, WI.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-termination inhibitors. Proceedings of the National Academy of Sciences U.S.A. **74**, 5463-5467.

Sehgal, O. P. (1981). Sourthern Bean Mosaic Virus. In Handbook of Plant Virus Infections and Comparative Diagnosis. (E. Kurstak, Ed.), pp 95-121. Elsevier/North Holland Biomedical Press.

Shannon, T. E. (1991). Evidence for the location of a viral protease and protease cleavage sites in the 105 KDa plyprotein of Southern Bean Mosaic Virus. M. S. Thesis, Western Kentucky University, Bowling Green, Kentucky.

Simmons, A. J. (1993). Site directed mutagenesis of southern bean mosaic virus (Cowpea Strain): In search of a viral protease. M. S. Thesis, Western Kentucky University, Bowling Green, Kentucky.

Turnbull-Ross, A. D., Mayo, M. A., Reavy, B., Murant, A. F. (1993) Sequence analysis of the parsnip yellow fleck virus polyprotein: Evidence of affinities with picornaviruses. Journal General Virology **74,** 555-561.

Veerisetty, V., Sehgal, O. P. (1979). Genome-linked proteinase K-sensitive factor essential for the infectivity of southern bean mosaic virus. Phytopathology **69**, 1048.

Wellink, J., Rezelman, G., Goldbach, R., Beyreuther, K. (1986). Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component RNA of cowpea mosaic virus. Journal of Virology **59**, 50-58.

Wu, S., Rinehart, C. A., Kaesberg, P. (1987). Sequence and organization of Southern Bean Mosaic Virus genomic RNA. Virology **161**, 73-80.