# MOLECULAR CHARACTERIZATION AND NUCLEOTIDE SEQUENCE OF THE COAT PROTEIN GENE OF THE NL-3 STRAIN OF BEAN COMMON MOSAIC VIRUS

---

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE GRADUATE SCHOOL OF THE
TEXAS WOMAN'S UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

BY
MARY JEAN SPARKS, B.S.,M.S.

---

DENTON, TEXAS
MAY 1993

TEXAS WOMAN'S UNIVERSITY
DENTON, TX

4/13/93
DATE

To the Associate Vice President for Research and Dean of Graduate Studies:

I am submitting herewith a dissertation written by Mary Jean Sparks entitled "Molecular characterization and nucleotide sequence of the coat protein gene of the NL-3 strain of Bean common mosaic virus". I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Molecular Biology.

John E. Knesek
Dr. J. Knesek, Major Professor

We have read this dissertation and recommend its acceptance.

Nathaniel Mills
Alan Corkalis
Michael J. Rudick
F. Silee
Department Chairperson

Accepted
Leslie M. Thompson
Associate Vice President for Research and Dean of Graduate Studies

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Knesek for his endless patience and understanding, to my committee members for their endurance and to Dr. Louise Higgins for her moral strength and support.

I would especially like to extend my profound thanks to my parents, Max and Lorene Sparks, who instilled in me the will to pursue academic excellence, for their unwavering moral support and love which gave me the strength to pursue an advanced education.

I also would like to extend my thanks and love to my children, Shane and Erin, who endured numerous sacrifices on my behalf and who will always remain an inspiration to me.

ABSTRACT

MOLECULAR CHARACTERIZATION AND NUCLEOTIDE SEQUENCE
OF THE COAT PROTEIN GENE OF THE NL-3 STRAIN
OF BEAN COMMON MOSAIC VIRUS

MARY JEAN SPARKS
MAY 1993

Bean common mosaic virus (BCMV) is a member of the potyvirus group, causing severe damage to leguminous crops. The NL-3 strain of the virus was purified and viral RNA extracted. Complementary DNA was synthesized from the viral RNA using Moloney-murine leukemia reverse transcriptase and Xho I oligo-dT priming. The single-stranded cDNA was converted into ds DNA by DNA polymerase and directionally cloned into lambda Uni-ZAP XR vector. The coat protein gene of the potyvirus is located at the extreme 3' end of the viral RNA, so clones of approximately 1 kb were selected, subcloned, and sequenced by the dideoxy chain termination method. Results from the nucleotide sequencing indicate that areas of homology and non-homology exist between the NL-3 strain and other published potyviral coat protein sequence data.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . . . . . . . . . . . . . . . . ii

ABSTRACT . . . . . . . . . . . . . . . . . . . . . . iii

LIST OF TABLES . . . . . . . . . . . . . . . . . . . vii

LIST OF FIGURES . . . . . . . . . . . . . . . . . . . viii

LIST OF ABBREVIATIONS . . . . . . . . . . . . . . . . . ix

CHAPTER

I. INTRODUCTION . . . . . . . . . . . . . . . . . 1

II. MATERIALS AND METHODS . . . . . . . . . . . . . 8

Materials . . . . . . . . . . . . . . . . . . . 8

Enzymes . . . . . . . . . . . . . . . . . . . 8

Kits . . . . . . . . . . . . . . . . . . . . 8

Reagents . . . . . . . . . . . . . . . . . . . 8

Media . . . . . . . . . . . . . . . . . . . . 9

Bacterial Strains . . . . . . . . . . . . . . . 10

Uni-ZAP XR Vector . . . . . . . . . . . . . . . 10

Methods . . . . . . . . . . . . . . . . . . . . 12

Gel Electrophoresis . . . . . . . . . . . . . . 12

Restriction Enzyme Digestions . . . . . . . . . 13

Geneclean Kit Protocol . . . . . . . . . . . . 14

Preparation of Competent Bacterial Cells . . 14

Propagation of Virus . . . . . . . . . . . . . 15

Purification of Bean Common Mosaic Virus . . 16

Bean Common Mosaic Virus RNA Extraction . . . 17
cDNA Synthesis of BCMV . . . . . . . . . . . 18
First strand synthesis . . . . . . . . . . . 18
Second strand synthesis . . . . . . . . . . 19
Molecular Cloning of BCMV ds DNA . . . . . . 19
Blunting the cDNA termini with T4 DNA polymerase . . . . . . . . . . . . . . . . . 19
Ligation and kinasing of Eco RI adaptors . 20
Xho I digestion and size separation of BCMV ds DNA . . . . . . . . . . . . . . . . . . . 21
Ligation of cDNA into Uni-ZAP XR vector . . 22
Packaging of BCMV vector DNA in lambda virions . . . . . . . . . . . . . . . . . . . 22
Titering of the BCMV recombinant library . 23
Amplification of BCMV library . . . . . . . 23
*In vivo* excision of pBluescript from Uni-ZAP XR vector . . . . . . . . . . . . . . . . 24
Plasmid Preparations . . . . . . . . . . . . 25
Mini-lysates . . . . . . . . . . . . . . . 25
Large scale . . . . . . . . . . . . . . . . 26
Preparation of DNA for Probe Synthesis . . . 28
Preparation of DNA Hybridization Probe . . . 28
Isolation of Single-Stranded Plasmid DNA . . 29

Sequencing of BCMV DNA . . . . . . . . . . . . 30

Sequencing of BCMV RNA . . . . . . . . . . . . 33

III. RESULTS . . . . . . . . . . . . . . . . . . . . 35

Molecular Cloning . . . . . . . . . . . . . . . . 35

Restriction Enzyme Analysis . . . . . . . . . . 36

Nucleic Acid Sequence Analysis of NL3-1 . . . . 42

Identification of Capsid Protein Coding Sequence . . . . . . . . . . . . . . . . . . . . 48

Comparison of BCMV Coat Protein Sequence with PSbMV and BYMV Coat Protein Sequences . . . . . 49

IV. DISCUSSION . . . . . . . . . . . . . . . . . . . 52

REFERENCES . . . . . . . . . . . . . . . . . . . . . . 59

APPENDIX . . . . . . . . . . . . . . . . . . . . . . . 63

Amino Acid Symbol List . . . . . . . . . . . . . 63

Restriction Enzyme Site Map . . . . . . . . . . . 64

## LIST OF TABLES

Table 1. Primers used in DNA and RNA sequencing of BCMV . . . . . . . . . . . . . . . . . . . . . . 11

LIST OF FIGURES

Fig. 1. Dot blot autoradiograph of the NL3-1 probe hybridization to BCMV NL3 RNA . . . . . . . . . . . 37

Fig. 2. Restriction endonuclease mapping of pBluescript NL3-1 on a 5% polyacrylamide gel digested with Eco RI, Hind III, Pst I and Xho I . . . . . . . . 39

Fig. 3. Restriction endonuclease mapping of pBluescript NL3-1 on a 5% polyacrylamide gel digested with Xba and Sau 3A1 . . . . . . . . . . . . . . . . . . 40

Fig. 4. Restriction endonuclease mapping of pBluescript NL3-1 on a 5% polyacrylamide gel digested with Taq I . . . . . . . . . . . . . . . . . . . . . . 41

Fig. 5. Restriction endonuclease map of the BCMV NL3 1267 bp fragment . . . . . . . . . . . . . . . . . 43

Fig. 6. Autoradiograph depicting the nucleotide sequence sequence of ssDNA NL3-1 electrophoresed on an 8% polyacrylamide gel . . . . . . . . . . . . . . . 44

Fig. 7. Sequencing strategy for clones NL3-1, NL3-19 and NL3-38 . . . . . . . . . . . . . . . . . . . . . . 46

Fig. 8. Nucleotide sequence of the 3'-terminal 1267 nucleotides of NL3 cDNA . . . . . . . . . . . . . 47

Fig. 9. Dot matrix comparison of nucleotide sequence homology of BYMV and BCMV NL3-1 coat protein genes . . . . . . . . . . . . . . . . . . . . . . . 50

Fig. 10. Dot matrix comparison of nucleotide sequence homology of PSbMV and BCMV NL-3 coat protein genes . . . . . . . . . . . . . . . . . . . . . . . 51

## LIST OF ABBREVIATIONS

| | |
|---|---|
| TAE | Tris-acetate/EDTA |
| TBE | Tris-borate/EDTA |
| EtBr | ethidium bromide |
| DTT | dithiothreitol |
| TEMED | N,N,N',N'-tetramethylethylenediamine |
| IPTG | isopropylthiogalactoside |
| EDTA | disodium ethylenediaminetetraacetate |
| BSA | bovine serum albumin |
| dNTP | deoxynucleotide triphosphates |
| SDS | sodium dodecylsulfate |
| CsCl | cesium chloride |
| PEG | polyethylene glycol |
| BCMV | Bean common mosaic virus |
| X gal | 5-bromo-4-chloro-3-indolyl-beta-D-galactoside |
| M-MuLVRT | Moloney-murine leukemia virus reverse transcriptase |
| RT | reverse transcriptase |
| AMVRT | Avian myeloblastosis virus reverse transcriptase |

# CHAPTER I

# INTRODUCTION

The potato virus Y (potyvirus) group of plant virus pathogens, constitute the largest and most widely distributed group of plant viruses, containing over 180 distinct members (Ward and Shukla 1991). The viruses within this group are considered to have the most serious economic impact on many crop plants causing devastating diseases and losses of $15 billion per year (Ward and Shukla 1991). Potyviruses can be mechanically transmitted but most members of this group are transmitted in nature by aphids and through seeds of infected plants. These two properties along with the diversity of crops the potyviruses infect assure the continuous presence of this group in the agricultural field.

Potyvirus particles are long flexous rods approximately 680-900 nm long (Ward and Shukla 1991). Their single-stranded, positive sense RNA genome is about 10,000 nucleotides in length with a small genome-linked virion protein (VPg) covalently attached to the 5'-terminus. A polyadenylate region has been located at the 3'-terminus and varies in length from 20 to 120 adenosines (Hari et al. 1979). The genome of all potyviruses encodes for a protein

characteristic pinwheel type inclusion bodies. The presence of these cylindrical inclusion bodies has been used in potyviral identification and taxonomy.

The majority of the viruses have been included in the potyvirus group on the basis of their particle morphology, biological properties, and transmission mechanism. However, with so many members and possible members in this group and the continuous descriptions of new members, there is a desperate need to define new criteria for identifying and classifying the potyviruses. This criteria should include molecular parameters such as: (1) genome structure and organization, (2) coat protein structure and sequence, and (3) virus replication and assembly mechanisms.

Until recently, little information was available regarding the characteristics of the potyviruses at the molecular level. However, with new developments in purification procedures, improved techniques in nucleic acid analyses and growing worldwide interest in potyviruses because of their economic importance, rapid progress has been achieved in the characterization of potyviral structure. The complete nucleotide sequences of five potyvirus RNA's, (1) tobacco etch virus (TEV), (2) potato virus Y (PVY), (3) tobacco vein mottling virus (TVMV), (4) plum pox virus (PPV) and (5) pea seed-borne mosaic virus (PSbMV) have now been determined along with partial sequences (1018-3332 bases) from the 3'-end of an additional

31 strains of 18 viruses (Ward and Shukla 1991, Timmerman et al 1991).

The potyviral genome is translated as a large molecular weight precursor polyprotein. This molecule is cleaved at susceptible glutamine-serine, glutamine-glycine and glutamine-alanine residues by one or two virus-encoded proteinases into eight polypeptides (Dougherty and Carrington 1988). A genetic map of the genomic organization has been established and positions of the virus encoded proteins have been assigned, although precise boundaries are still uncertain. The order of these proteins from 5' to 3' is: first protein (P1), helper component (HC), third protein (P3), cylindrical inclusion protein (CI), small nuclear inclusion protein (NIa) which includes the VPg at its N terminus, large nuclear inclusion protein (NIb), and coat protein (CP). Only the coat protein and the small VPg are found in the mature virion. Four of the remaining six proteins (HC, CI, NIa, NIb) are present in infected plants (Gough et al. 1987), while the last two proteins, P1 and P3, have not been detected in vivo (Hellman et al 1986).

Biological functions are known for four of the gene products. These are the helper component, a 106 to 116 K protein which is required for aphid transmission (Thornbury et al 1985), the CI protein (42K) which is an RNA helicase (Lain et al. 1990), the NIa protein (49K) which is a protease and is responsible for several site-specific

cleavages within the polyprotein including its own autocatalytic release (Carrington and Dougherty 1987) and the coat protein (33 - 34K) which encapsidates the viral RNA. The P1 protein (28 - 34K) is located at the N-terminus of the precursor and has been implicated in cell-to-cell movement of virus particles (Domier et al. 1987). The P3 protein (42K) may be another proteinase or utilized as a regulatory protein (Vos et al. 1988). The NIb protein (54K) may be an RNA-dependent RNA polymerase or a viral replicase (Allison et al. 1986; Domier et al. 1987) based on sequence homology data and the genome-linked VPg (6 - 24K) is thought to be involved in virus replication by serving as a primer for viral nucleic acid synthesis (Shahabuddin et al. 1988). The VPg may also play a role in aphid transmission but it is not involved in infectivity or translation of viral RNA in vitro (Hari 1981; Siaw et al. 1985).

The coat protein of potyviruses is the major gene product in the virion and accounts for 95% of potyvirus particles. Sequence data from distinct potyvirus indicate a considerable variation in size due to differences in the length of their N-termini (263 - 329 amino acids). These variations in length at the N-terminal region may reflect different locations of the specific cleavage sites (Allison et al. 1986; Domier et al. 1986) in this highly variable region of the potyvirus polyprotein. In contrast, the highly conserved C-terminal ends vary in length by only one

or two residues. After studying the coat protein sequence data between different potyviruses and between strains of viruses in the potyviral group, it was found that distinct potyviral members show 38 to 78% homology and vary considerably in the length and sequence of their amino-termini. The coat proteins of strains of individual viruses exhibit approximately 90% or greater homology and have very similar amino-terminal sequences (Shukla and Ward 1988). In a comparison of the extent of sequence similarity between the coat protein of turnip mosaic virus (TMV) with the coat proteins of sugarcane mosaic virus (SCMV), PPV, pepper mottle virus (PeMV), PVY and TVMV, it became clear that the similarity drastically decreased when the amino-terminal portion was considered (Tremblay et al 1990). A closer examination of the N-terminal nucleotide sequence of the coat protein of a number of potyviruses reveals an interesting finding. The N-terminal amino acid can be either a serine, alanine or glycine. When the amino acid is alanine or glycine, the capsid protein can be chemically sequenced. If serine is the N-terminal amino acid, it is not possible to directly determine the N-terminus of the particular coat protein, as is evident in TVMV, zucchini yellow mosaic virus (ZYMV), Johnson grass mosaic virus (JGMV), sugarcane mosaic virus (SCMV) and passion fruit woodiness virus (PWV) (Domier et al. 1986; Grumet and Fang 1990; Shukla et al. 1987, Frenkel et al. 1991; Shukla et al. 1988).

Pea seed-borne mosaic virus (PSbMV) is an important member of the potyviral group because it has a broad geographical distribution, it infects a wide variety of plant species and it is spread by infected seed (Timmerman et al 1990). Bean yellow mosaic virus (BYMV), another economically significant potyvirus, infects leguminous crops and a variety of cultivated ornamental species (Bos 1970).

Bean common mosaic virus (BCMV) is a member of the potyvirus group of plant pathogens and exhibits worldwide distribution due mainly to its transmission by seed. The virus is also easily spread by aphids in a non-persistant manner. BCMV is found in all parts of the world where beans are grown and economic damage consists of a severe reduction in crop yield and quality of the harvested crop.

Numerous strains of the virus have been reported, the most prevalent in the United States being the Type strain described in 1917. In 1934, a more detailed description of the biological properties of the virus was made and added the term 'common' to distinguish it from bean yellow mosaic virus to which BCMV is serologically closely related. Various BCMV strains were described in the mid-fifties in the Netherlands, which were later given NL or Netherland numbers in the order of description, NL-1 through NL-8 (Drijshout, 1978). Currently, eighteen strains of BCMV have been reported.

Since BCMV causes devastating damage to crops and

because it is seed-borne and probably coexistent with the host, *Phaseolus vulgaris*, farmers need a relatively quick diagnostic tool to determine whether their crops are infected by BCMV or by another serologically related potyvirus which would not cause quarantine of the crop resulting in monetary and market losses to the farmer. By determining the coat protein gene sequence of the viral strain and by examining these sequences for variable and conserved regions, it would be possible to engineer diagnostic probes for detecting infected plants. The published data of other potyviral coat proteins indicate that the N-terminal regions are potyvirus-specific, sharing in one case less than 12% homology with other potyviral N-terminal regions (Grumet and Fang 1990). This research was initiated to provide a foundation at the molecular level for the characterization of bean common mosaic virus and represents the first step towards understanding the initial genome organization of BCMV. The specific aims of this research are as follows:

1) to clone the DNA complementary to the RNA of the coat protein gene of the Netherland strain, NL-3 of bean common mosaic virus

2) to characterize by restriction endonuclease mapping the coat protein gene from the NL-3 strain

3) to sequence the nucleic acid of the cloned coat protein gene of the BCMV NL-3 strain

4) to investigate the extent of sequence homology between the coat protein of the NL-3 strain of BCMV with the coat protein sequence of pea seedborne mosaic virus and bean yellow mosaic virus.

# CHAPTER II

# MATERIALS AND METHODS

## MATERIALS

### Enzymes

Restriction endonucleases were purchased from Bethesda Research Laboratories (BRL), Gaithersburg, MD or New England Biolabs, Beverly, MA. Lysozyme was available from Sigma Chemical Co., St. Louis, MO., and proteinase K was obtained from Ambion, Austin, TX.

### Kits

Geneclean kits were from Bio 101, La Jolla, CA. The Random Primers DNA Labeling System was purchased from BRL and the cDNA Synthesis Kit was acquired from Stratagene, La Jolla, CA. The DNA sequencing kit, Sequenase Version 2.0 was from USB Corporation, Cleveland, OH and the RNA sequencing kit, GemSeq Transcript Sequencing System, was purchased from Promega, Madison, WI.

### Reagents

Reagents commonly used in these experiments were obtained from commercial suppliers and were of the best reagent grade available. Polyacrylamide-bis solution was

purchased from AMRESCO, Solon, OH. Long Ranger acrylamide was from AT Biochem, Malvern, PA and SeaKem GTG agarose was obtained from FMC Bioproducts. NAP columns, nick translation columns and Sephacryl spin columns were acquired from Pharmacia, Uppsale, Sweden; Whatman 3MM filter paper from Hoefer Scientific Instruments, San Francisco, CA and nitrocellulose filters from Schleicher and Schuell, Keene, NH. Isotopes were purchased from Amersham Corporation, Arlington Heights, IL. Primers used in the sequencing reactions were obtained from either Pharmacia, Uppsala, Sweden or from Biosynthesis, Inc., Lewisville, TX.

## Media

NZY powdered medium was purchased from GIBCO, Grand Island, NY and was prepared according to the manufacturers' directions (21 g NZY media per liter). Bacto-agar, tryptone and yeast extract were from DIFCO Laboratories, Detroit, MI. Agar plates were prepared by adding 16 g of agar per liter of media. The 2X YT media (16 g tryptone, 10 g yeast extract, 5 g NaCl per liter) was made for the isolation of single-stranded plasmid DNA according to Messing (1983) and for the growth and maintenance of bacterial strains according to Stratagenes' protocol. Ampicillin, tetracycline, kanamycin and chloramphenicol were acquired from AMRESCO, Solon, OH. All media and reagents were prepared with ultrapure water.

Bacterial Strains

The XL1-Blue and PLK-F' bacterial strains used in the cloning experiments were provided by Stratagene, La Jolla, CA. These strains were maintained by streaking each onto Luria-Bertani (10 g tryptone, 5 g yeast extract, 10 g NaCl per liter) tetracycline (40 ug/ml) agar plates and incubating the plates overnight at 37 $^0$C. A liquid culture can be started from a single colony and grown overnight with vigorous shaking at 30 $^0$C in LB medium. The cells were centrifuged at 1000g for 10 min, the supernatant removed and the pellet resuspended in 0.5 volumes of 10 mM $MgSO_4$. These freshly prepared cells can be used for all manipulations required during cloning and construction of the cDNA library. The DH5αF' bacterial strain was provided by BRL and maintained by streaking it onto NZY agar plates. Each of the bacterial strains was stored until needed at -70 $^0$C after mixing 100 ul of each bacterial strain grown to 0.8 $A_{600nm}$ with 200 ul of 80% glycerol solution.

Uni-ZAP XR Vector

The lambda Uni-ZAP XR vector has been double-digested with Eco RI and Xho I and will accomodate DNA inserts from 0 - 10 kb in length. This particular vector was designed to allow *in vivo* excision and recircularization of any cloned inserts contained within the lambda vector allowing the insert to be characterized in a plasmid system. The

PRIMERS

| Primers | Sequence |
|---|---|
| 3-2 | 5' -TCA GAC AGA GCC AGA GA- 3' |
| 3-4 | 5' -GTA CGC TTT CAA TTC CA- 3' |
| 3-5 | 5' -AAG CAC CAT ACA TAG CA- 3' |
| 3-2R | 5' -TCT CTG GCT CTG TCT GA- 3' |
| 3-5R | 5' -TGC TAT GTA TGG TGC TT- 3' |
| 3-6R | 5' -TTA GAG AAT ATT CAT AC- 3' |
| 40 | 5' -GAG AAA GAG AAG AAG AAA G- 3' |
| 41R | 5' -CTG CGT AGG CTT GCA TTC T- 3' |
| 3R801 | 5' -ACA GAA AAC TAT GAA CG- 3' |
| Reverse | 5' -AAC AGC TAT GAC CAT G- 3' |
| Universal | 5' -GTA AAA CGA CGG CCA GT- 3' |

Table 1. Primers used in DNA and RNA sequencing of BCMV.

polylinker of the pBluescript phagemid contains 21 unique cloning sites flanked by T3 and T7 promoters in addition to six different primer sites for DNA sequencing. The phagemid has the bacteriophage f1 origin of replication which permits the single-stranded DNA to be rescued for use in DNA sequencing.

## METHODS

### General Procedures

Gel electrophoresis

DNA was analyzed on either acrylamide or agarose gels. The 5% polyacrylamide gels were prepared by adding together 38 ml Ultrapure $H_2O$, 6.4 ml of 40% acrylamide-bis solution, 5 ml of 5X TBE buffer (0.5X Final Concentration-89 mM Tris, 89 mM Boric acid, 2 mM EDTA), 500 ul of fresh 10% ammonium persulfate, 50 ul of TEMED and pipetted into a vertical 14 x 16 cm Hoefer Model SE 400 gel unit. The gels were electrophoresed in 0.5X TBE buffer at 80 volts for 1.5 to 3 h. After electrophoresis, the gels were stained with ethidium bromide (EtBr, 1 ug/ml) for 30 min. Stained gels were photographed with Polaroid 57 film under 302 nm UV illumination with an Orange 15 filter.

Agarose gels were prepared with SeaKem GTG agarose at a concentration of 1%. The agarose gels were prepared and electrophoresed with 1X TAE buffer (40 mM Tris acetate,

2 mM EDTA). Varying amounts of DNA samples mixed with water and tracking dye (0.25% Bromophenol blue, 20 mM EDTA pH 7.0, 50% glycerol) were electrophoresed at 40 V for 1.5 to 3 h on either the Hoefer Scientific Instruments Model HE 33 unit or on the BRL Model H6 unit. After electrophoresis, the gels were stained with EtBr (1 ug/ml) for 15 - 30 min and photographed using UV transillumination.

Restriction enzyme digestions

Restriction endonucleases are enzymes, derived mainly from prokaryotes, that recognize specific sequences in double-stranded DNA (Maniatis 1982). Each restriction enzyme has a set of optimal reaction conditions which include incubation temperature and buffer composition. A battery of restriction enzymes was utilized following BRL specific requirements. Reactions typically contained volumes of 50 ul or less of the DNA, UP water, the appropriate digestion buffer to a final concentration of 1X and the selected restriction enzyme. One unit of enzyme is sufficient to digest 1 ug of DNA in one h. The reaction mixture was incubated at the optimum temperature for the required period of time. The reaction was stopped by immediate analysis on a gel, by heating at 68 $^{0}$C for 10 min to inactivate the enzyme or by utilizing the Geneclean protocol on the sample.

Geneclean Kit protocol

The DNA was extracted, purified and concentrated by the Geneclean Kit procedure. Two to three times the volume of saturated sodium iodide were added to the DNA solution. Five microliters of glassmilk (silica matrix in water) were added to the tube and the contents were mixed and incubated at room temperature for 5 min to allow binding of the DNA to the silica matrix. The bound DNA was microcentrifuged briefly and the supernatant was discarded. The pellet was washed three times with 400 ul of NEW wash (50% ethanol, 100 mM EDTA, 100 mM Tris-HCl pH 7.8). After the last NEW wash, the pellet was resuspended in sterile water and this was incubated in a 50 $^0$C waterbath for 2-3 min. The heated sample was centrifuged for 30 sec and the supernatant containing the DNA was transfered to a clean Eppendorf tube. All DNA samples were stored at 4 $^0$C until needed.

Preparation of Competent Bacterial Cells

All bacterial cell lines were rendered competent by using a protocol devised by Bio 101, Inc. (La Jolla, CA) involving Rubidium calcium transformation (RbCa TXN) salts. An important step in the procedure is to grow the cells in a rich Super-Comp media which is prepared by adding the contents of one Super-Comp capsule to 25 ml distilled water and autoclaving for 10 min. The bacterial cells were streaked for isolation on an appropriate agar plate and

incubated at 37 $^{0}$C overnight. An isolated colony was inoculated into 50 ml media and grown to 0.8 $A_{600nm}$. One milliliter of the fresh bacterial cells was added to the 25 ml Super-Comp media and incubated with shaking at 37 $^{0}$C until an 0.6 $A_{600nm}$ was reached. The cell culture was transferred to a sterile tube and centrifuged at 3400 rpm for 5 min at 4 $^{0}$C in a Beckman JS-7.5 rotor. The supernatant was discarded and the cell pellet was resuspended in 15 ml of cold 1X RbCa TXN salts and incubated on ice for 15 min. The solution was again centrifuged at 3400 rpm for 5 min at 4 $^{0}$C. The supernatant was discarded and the pellet was resuspended in 2 ml 1X RbCa TXN salts. The cells were immediately ready for DNA uptake or they were stored at 4 $^{0}$C for 12-24 h to increase competency 4-6 fold.

**Propagation of Virus**

Phaseolus vulgaris cultivar Bountiful seeds were supplied by Harpool Nursery, Denton, TX. The seeds, five per pot, were planted in 6 inch plastic pots filled with commercial potting soil. These seeds were germinated and maintained in growth chambers at a temperature of 20 - 26 $^{0}$C with a 16 h photoperiod. After 5 to 7 days, the primary leaves were inoculated with the appropriate virus strain. The Netherland strain (BCMV-NL3) was provided by Dr. M. J. Silbernagel, Prosser, WA. The leaves were inoculated by first dusting the upper surface with 600 mesh carborundum

(Silicon Carbide, Buehler, Ltd, USA) to mechanically break open the cell walls. Viral infected leaves were ground with a mortar and pestle in 0.02 M sodium phosphate buffer pH 8.0. The virus containing solution was manually applied to each leaf. Nine to eleven days following inoculation, the virus infected leaves were harvested for virus purification.

**Purification of Bean Common Mosaic Virus**

Each virus was purified by a modification of the method developed by Morales, (1979). Nine to eleven days after post infection, 50 to 100 g of infected leaf tissue were homogenized using a Waring blender and a mixture of 50 ml chloroform, 50 ml carbon tetrachloride and 200 ml of 0.5 M potassium phosphate buffer pH 7.5, containing 0.02 M sodium sulphite for 1 min. The homogenized mixture was centrifuged for 20 min at 10,000 rpm in a type 19 Beckman rotor. The resulting supernatant was filtered through glasswool. The filtered supernatant was layered onto a 10 ml solution containing 6% polyethylene glycol (PEG FW 6000, w/v), 30% sucrose (w/v) and 0.5 M potassium phosphate pH 7.5 and centrifuged at 25,000 rpm for 2 h at 4 $^{0}$C in a SW 27 rotor. The supernatant was discarded and the pellet of virus was resuspended by standing overnight in 0.65 ml of 0.25 M potassium phosphate pH 7.5 at 4 $^{0}$C. The solution of virus was centrifuged for clarification at 10,000 rpm for 15 min at 4 $^{0}$C in a type 30 rotor. The pellets were

discarded. The virus was further purified by adding 4.8 g of cesium chloride (CsCl, 40% F.C., w/v) to every 12 ml of virus solution. This suspension was placed in Quick seal polyallomer tubes and centrifuged to equilibrium at 35,000 rpm for 17 h at 20 $^0$C in an 80 Ti rotor. The visible band of virus was collected very slowly using an 18 gauge needle on a one ml syringe to prevent fragmentation of virus particles.

**Bean Common Mosaic Virus RNA Extraction**

Viral RNA was obtained by adding 3 ml of RNA extraction buffer (6 M guanidine isothiocyanate, 5 mM sodium citrate, 0.1 M beta-mercaptoethanol, 0.5% sarkosyl) to each 0.5 ml of viral solution and incubated for 1 h at room temperature. After this incubation period, the RNA solution was diluted to 10 ml and 3.5 g of CsCl were added and centrifuged at 32,000 rpm for 18 h at 20 $^0$C in an 80 Ti rotor. The supernatant was discarded and the pellet is stored at -70 $^0$C or is resuspended in 40 ul of UP water to which 10 ul of 10 M ammonium acetate was added. This solution was transferred to a Beckman G-max tube and 150 ul of absolute ethanol was added and stored at -20 $^0$C overnight. The tube was then filled with 80% ethanol and centrifuged at 25,000 rpm for 1 h at 20 $^0$C in a SW 27 rotor. The ethanol was discarded and 1.5 ml 80% ethanol was added to the resultant RNA pellet, mixed well, microfuged for 10 min and dried in a

Speed Vac (Savant Instruments) for 1 h. The RNA was resuspended in an appropriate volume of UP water (25 ul). The concentration of the nucleic acid can be determined spectrophotometrically by diluting 10 ul of the sample to 300 ul with UP water and determining the absorbance at 260 nm. The concentration of RNA was calculated according to the formula:

$A_{260nm}$ x 30 x 40 = micrograms RNA/ml.

Each preparation of RNA was analyzed for integrity on an agarose gel.

**cDNA Synthesis of BCMV**

First Strand Synthesis

The synthesis reactions were performed by using the ZAP-cDNA Synthesis kit protocol provided by Stratagene. First strand cDNA synthesis begins when reverse transcriptase finds a template and primer in the presence of nucleotides and the appropriate buffer. A first strand cDNA reaction, consisting of 5 ul of 10X first strand buffer (0.5 M Tris pH 8.3, 0.75 M KCl, 0.03 M $MgCl_2$, 0.1 M DTT), 2.5 ul of 10 mM dNTP (dATP, dGTP, dTTP, 5-me-dCTP), 2 ul of an oligo dT Xho primer (2.8 ug), 1 ul of RNase II block (1 unit), 5 ug of poly $(A)^+$ mRNA template, and UP water to a volume of 47.5 ul, was incubated at room temperature for 10 min. After the template and primer were annealed, 2.5 ul of Moloney-murine

leukemia virus reverse transcriptase (M-MuLVRT) was added, mixed thoroughly and incubated at 37 $^{0}$C for 1 h.

Second Strand Synthesis

Second strand synthesis was initiated when 40 ul of 10X second strand buffer (0.19 M Tris pH 8.3, 0.9 M KCl, 0.04 M $MgCl_2$, 0.04 M DTT), 7 ul of 10 mM dNTP (dATP, dCTP, dGTP, dTTP), 0.8 ul RNase H (3.2 units), 10 ul of DNA polymerase I (110 units) and 297 ul of UP water were added to the 45 ul first strand cDNA, vortexed and incubated 2.5 h at 16 $^{0}$C. A phenol:chloroform (1:1) extraction was performed and the upper aqueous layer was transfered to a new microfuge tube where an equal volume chloroform extraction was done. The upper aqueous phase, containing the dsDNA was ethanol precipitated overnight at -20 $^{0}$C by adding 33.3 ul of 3 M sodium acetate and 867 ul of absolute ethanol, then microfuged for 1 h at room temperature and the subsequent pellet was washed with 80% ethanol, lyophilized to dryness and resuspended in 43.5 ul sterile water.

**Molecular Cloning of BCMV ds DNA**

Blunting the cDNA Termini with T4 DNA Polymerase

The following reagents were added to 39 ul of the cDNA; 5 ul of 10X T4 DNA polymerase buffer (0.03 M Tris acetate pH 7.9, 0.07 M potassium acetate, 0.01 M Mg acetate, 0.5 mM DTT, 0.1 mg/ml BSA) 2.5 ul of 2.5 mM dNTP mixture and 3.5 ul

of T4 DNA polymerase and incubated at 37 $^0$C for 30 min. An equal volume of sterile water was added to the sample and a phenol:chloroform (1:1 v/v) extraction was performed followed by an equal volume chloroform extraction. The supernatant was ethanol precipitated by adding 7 ul 3 M sodium acetate and 226 ul absolute ethanol and stored at -20 $^0$C overnight. The sample was microfuged at room temperature for 1 h at 10,000 rpm and the supernatant discarded. The pellet was washed with 1.5 ml 80% ethanol, microfuged for 10 min at 10,000 rpm and lyophilized to dryness.

Ligation and Kinasing of Eco RI Adaptors

The sample pellet was resuspended in 7 ul of the Eco RI adaptors and 1 ul of 10X ligation buffer (500 mM Tris-HCl pH 7.5, 70 mM $MgCl_2$, 10 mM DTT). Following a 10 min incubation at 37 $^0$C, 1 ul of 10 mM ATP and 1 ul of T4 DNA ligase (4 U/ul) were added and the sample was incubated overnight at 4 $^0$C. After heat inactivating the ligase at 70 $^0$C for 30 min, the adaptor ends were kinased by adding 1 ul of 10X ligation buffer, 2 ul of 10 mM ATP, 1 ul of T4 polynucleotide kinase (10 U/ul), 6 ul of sterile water and incubating 30 min at 37 $^0$C. The kinase was then heat inactivated for 30 min at 70 $^0$C.

Xho I Digestion and Size Separation of BCMV ds DNA

Digestion with the restriction enzyme Xho I will cleave Xho I linker-primer and Eco RI adaptors from the ds DNA. Twenty-eight microliters of Xho I buffer (187.5 mM NaCl, 12.9 mM $MgCl_2$, 1.8 mM DTT) and 2 ul Xho I (45 U/ul) were added to the sample and incubated at 37 $^0$C for 1 h. The reaction was cooled to room temperature and 5 ul of 10X STE (1 M NaCl, 200 mM Tris pH 7.5, 0.1 M EDTA) was added. The BCMV DNA was size fractionated on a precentrifuged Sephacryl spin column. The DNA sample was loaded onto the top of the prepared resin bed and centrifuged at 1400 rpm for 2 min in an IEC PR-J centrifuge. The eluant removed was labeled Fraction 1 and should contain the larger DNA molecules. Fifty microliters of 1X STE was loaded onto the top of the same column and centrifuged at 1400 rpm for 2 min. This second DNA sample was labeled Fraction 2. The size range of the DNA will decrease with each additional collection. Each DNA containing fraction was extracted with phenol:chloroform (1:1 v/v) followed by a second extraction with an equal volume of chloroform. The upper aqueous phase containing the DNA was transfered to a clean tube and ethanol precipitated at -20 $^0$C overnight after adding 2X volume absolute ethanol. The 1X STE buffer contained sufficient NaCl for precipitation. The sample was microfuged at maximum speed for 60 min at 4 $^0$C. The pellet was washed

with 80% ethanol and lyophilized to dryness. The DNA was resuspended in 10 ul of sterile water.

Ligation of cDNA into Uni-ZAP XR Vector

The lambda Uni-ZAP XR vector has been double-digested with Eco RI and Xho I and can accomodate DNA inserts from 0 - 10 kb in length. To 2.5 ul of the resuspended DNA, add 0.5 ul of 10X ligation buffer, 0.5 ul of 10 mM ATP, 1 ul of Uni-ZAP XR vector (1 ug/ul) and 0.5 ul of T4 DNA ligase (4 U/ul). A test ligation is simultaneously set up utilizing 1 ul of Uni-ZAP XR/Eco RI and Xho I prepared arms (1 ug), 1.6 ul of test insert (0.4 ug pBR322 digested with Sal I and Eco RI), 0.5 ul of 10X ligation buffer, 0.5 ul of 10 mM ATP pH 7.5, 0.5 ul of T4 DNA ligase (4 U/ul) and 0.9 ul sterile water. Both reactions were incubated overnight at 12 $^{0}$C, followed by 2 h at room temperature.

Packaging of BCMV Vector DNA in Lambda Virions

Two packaging extracts, a Sonic extract and a Freeze/Thaw extract were used to package the recombinant lambda phage. Five microliters of recombinant lambda DNA was added immediately to the Freeze/Thaw extract just as it was beginning to thaw and placed on ice. Immediately, 15 ul of Sonic extract was added to the sample tube. After a 2 h incubation at room temperature, 500 ul of SM buffer (5.8 g NaCl, 2 g $MgSO_4$, 50 ml 1 M Tris-HCl pH 7.5, 5 ml 2% gelatin

per liter) was added followed by 20 ul of chloroform. The sample was mixed, centrifuged briefly to sediment any debris and stored at 4 $^0$C.

Titering of the BCMV Recombinant Library

To determine the quality and quantity of the packaged ligation sample, a titer of virion concentration was performed. The BCMV library was diluted 1:10 and 1 ul of this dilution and 1 ul of the undiluted sample was added to 200 ul each of 0.5 $A_{600nm}$ PLK-F' cells and incubated 15 min at 37 $^0$C to allow the phage time to attach to the bacteria. Three milliliters of top agar (NZY broth with 0.7% agarose) was added to the phage/bacteria sample and was immediately plated onto prewarmed (37 $^0$C) NZY agar plates. The test insert suspension was plated, both diluted and undiluted as stated previously. All plates were incubated overnight at 37 $^0$C.

Amplification of BCMV Library

The recombinant bacteriophage library was amplified in order to obtain a large, stable quantity of a high titer stock of the library. Several aliquots of the phage library containing approximately 50,000 recombinant phage were added to 600 ul of 0.5 $A_{600nm}$ PLK-F' host cells and incubated 15 min at 37 $^0$C. Each aliquot was then mixed with 6.5 ml of top agar and poured evenly onto prewarmed (37 $^0$C) NZY agar

plates. The plates were placed at 37 $^0$C for 5 to 8 h. The viral plaques should not be allowed to touch. The plates were overlayed with 10 ml of SM buffer and incubated overnight at 4 $^0$C with gentle shaking. The bacteriophage suspension from each plate was pooled into sterile polypropylene tubes and each plate rinsed with an additional 2 ml of SM buffer. Chloroform was added to 5%, mixed and incubated at room temperature 15 min. All cell and agar debris was removed by centrifugation for 5 min in an IEC PR-J centrifuge at 2000 rpm. The supernatant was transfered to sterile polypropylene tubes and chloroform was added to 0.3%. The stable bacteriophage library was stored at 4 $^0$C. Using serial dilutions of the library, the titer of the amplified library was checked using PLK-F' host cells. The titer was determined to be 5 x $10^9$ for the NL3 strain of BCMV.

*In vivo* Excision of pBluescript from Uni-ZAP XR Phage

The Uni-ZAP XR vector system has been designed to allow *in vivo* excision and recircularization of any cloned insert within the lambda vector to form a phagemid containing the insert of interest. This excision process is dependent on specific DNA sequences strategically placed in the lambda phage genome and upon the presence of a variety of proteins. In a 50 ml conical tube, 200 ul of 1.0 $A_{600nm}$ XL1-Blue cells, 200 ul of Uni-ZAP XR recombinant stock and 1 ul of R408

helper phage were combined and incubated 15 min at 37 $^0$C. Five milliliters of 2X YT broth were added and incubated 3 h at 37 $^0$C with constant shaking. The tube was heated to 70 $^0$C for 30 min and centrifuged at 2000 rpm in an IEC PR-J centrifuge for 5 min. The supernatant was decanted into a sterile tube and stored at 4 $^0$C. This stock contains the pBluescript phagemid packaged as a filamentous phage particle. In order to plate the rescued phagemid, combine in two separate 15 ml conical tubes, a) 200 ul of newly prepared phage stock and 200 ul 1.0 $A_{600nm}$ XL1-Blue host cells and b) 20 ul of a $10^{-2}$ dilution of phage stock and 200 ul of 1.0 $A_{600nm}$ XL1-Blue host cells. Both tubes were incubated at 37 $^0$C for 15 min. Each of the two samples were plated on LB/Ampicillin (50 ug/ml) plates and incubated overnight at 37 $^0$C. Any colonies which appear contain the pBluescript double-stranded phagemid with the cloned DNA insert. To maintain each recombinant, each colony was streaked onto new NZY/Ampicillin (50 ug/ml) agar plates and stored at 4 $^0$C.

## Plasmid Preparations

### Mini-lysates

Each recombinant XL1-Blue colony was inoculated into 5 ml NZY broth and shaken at 200 rpm at 37 $^0$C until an estimated 0.8 $A_{600nm}$ was reached. Chloramphenicol (50 mg/ml) was added to each flask and incubated overnight at 37 $^0$C with continuous shaking. After incubation, the cultures were

centrifuged in a microcentrifuge to pellet the bacterial cells. The cell pellets were resuspended in 100 ul of STET (8% sucrose, 50 mM Tris-HCl pH 8, 50 mM EDTA, 5% Triton X 100) and 10 ul of lysozyme (10 mg/ml) and incubated on ice for 10 min. Each tube was placed in a boiling waterbath for 2 min and centrifuged 15 min in a microfuge. The supernatants were removed and transfered to a clean 1.5 ml Eppendorf tube. The DNA was concentrated and impurities were removed by the Geneclean method.

Large Scale

XL1-Blue host cells, containing the plasmid plus insert of interest, were streaked onto NZY ampicillin plates and incubated overnight at 37 $^0$C. One isolated colony was selected and inoculated into 250 ml of NZY media in a 1 liter flask, and incubated in a 37 $^0$C shaker (200 rpm). When the culture reached an $A_{600nm}$ between 0.8 and 1.0, chloramphenicol was added to the flask to a final concentration of 200 ug/ml. The culture was incubated overnight in a 37 $^0$C shaker. The bacterial culture was centrifuged at 5000 rpm for 15 min at 4 $^0$C in a JA-14 rotor. The resulting pellet was lysed by resuspending it in a 5 ml solution containing 20% sucrose, 50 mM Tris-HCl pH 8.0. Two milliters of lysozyme (10 mg/ml) was added to the solution. After a 10 min incubation on ice, 25 ml of a solution containing 50 mM Tris-HCl pH 8.0, 20 mM EDTA and 2.5 mg of

proteinase K was added plus 1.5 ml 10% SDS. The solution was incubated for 1 h in a 56 $^{0}$C waterbath. After incubation, 38 ml of 50 mM Tris-HCl pH 8.0 and 20 mM EDTA solution, 17.5 ml 5 M NaCl and 2 ml of 10% SDS were added and the solution was incubated 12 to 24 h at 37 $^{0}$C. The mixture, following an incubation for 1.5 h at -20 $^{0}$C was centrifuged at 12,000 rpm for 20 min at 4 $^{0}$C in a type 19 rotor. The supernatant was transferred to a 250 ml screw top bottle and two volumes of absolute ethanol were added. This solution was stored at -20 $^{0}$C overnight to precipitate the nucleic acid. After it was removed from the freezer, the sample was warmed to room temperature and centrifuged at 6000 rpm for 20 min at 20 $^{0}$C. The supernatant was discarded and the pellet was washed with 70% ethanol at room temperature for 1 to 2 h. The sample was centrifuged at 6000 rpm for 20 min at 20 $^{0}$C and the supernatant discarded. The pellet was dissolved in 9 ml of UP water and 1 ml of 0.05 M sodium citrate. The solution was incubated for 1 h at 68 $^{0}$C. A cesium chloride gradient was prepared by adding 9.43 g of CsCl and 400 ul of EtBr (10 mg/ml) to the 10 ml of plasmid DNA solution. The gradient was centrifuged at 55,000 rpm for 17 h at 20 $^{0}$C in an 80 Ti rotor. The band, which was visible under UV illumination was collected with a 1 ml syringe fitted with an 18 gauge needle. The plasmid DNA was extracted four times with an equal volume of isopropyl alcohol. The upper phase was discarded each time.

Following extraction the DNA was passed through a NAP-5 column which had previously been rinsed with sterile water. The DNA concentration was determined by absorption spectrophotometry at 260 nm.

**Preparation of DNA for Probe Synthesis**

The pBluescript plasmid containing the cloned insert was linearized by digesting the DNA overnight with Eco RI at 37 $^0$C. The sample was gene cleaned to remove the React 3 buffer (50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM NaCl) and recovered in 25 ul of UP water. The linearized plasmid was digested with Xho I to cleave the insert of interest from the plasmid and was separated from the plasmid by agarose gel electrophoresis. The 1.3 kb band was removed from the gel and liquified at 50 $^0$C for 10 min in the presence of NaI. The DNA was concentrated and impurities removed by the Geneclean procedure.

**Preparation of DNA Hybridization Probe**

In order to synthesize a DNA probe, a Random Primers DNA Labeling System was utilized. A minimum of 25 ng of DNA contained in a volume of 5-20 ul was denatured and quick cooled on ice. The following reagents were added to the DNA sample on ice, 2 ul each of dATP, dGTP, dTTP, 15 ul of random primers buffer (0.67 M HEPES, 0.17 M Tris-HCl, 17 mM $MgCl_2$, 33 mM $\beta$-mercaptoethanol, 1.33 mg/ml BSA, 18 $A_{260nm}$/ml

oligodeoxyribonucleotide primers pH 6.8), 5 ul [ $^{32}$P]dCTP Amersham-10 Ci/ml, 6000 Ci/mmole) and distilled water to a final volume of 49 ul. The contents were mixed and 1 ul of Klenow fragment (3 U) was added and incubated overnight at room temperature. Five microliters of 0.2 M $Na_2$EDTA, pH 7.5, was added and the unincorporated nucleoside triphosphates were removed by passing the sample through a nick translation column which had previously been rinsed with sterile water. Various amounts of the probe were added to scintillation vials containing Betaphase (5 ml-Safety Solve Research Products International Corp.) and counted in a Beckman LS 9000 scintillation counter. The specific acitivity of the probe was 10 million cpm/ug.

**Isolation of single-stranded (ss) Plasmid DNA**

The bacterial strain containing the recombinant plasmid were grown in 30 ml of 2XYT broth in a 250 ml flask at 37 $^0$C with shaking to an 0.4 $A_{600nm.}$ The culture was infected with 100 ul of an overnight K13 stock solution (7 X $10^{11}$ pfu/ml). After a 30 min incubation period, kanamycin was added to a final concentration of 50 ug/ml. The infected cells were again incubated for an additional 5 h at 37 $^0$C with shaking. After the incubation period, 20 ml of the bacterial culture was transfered to a 30 ml Corex tube and centrifuged at 9000 rpm for 20 min in an IEC Model HT centrifuge. The supernatant was transfered into a clean 30 ml Corex tube and

centrifuged an additional 20 min at 9000 rpm. Thirteen milliliters of the solution were carefully removed from the top of the tube and added to 13 ml of a 20 % PEG (polyethylene glycol-Fisher 8000) and 3.5 M ammonium acetate solution and set at room temperature for 15 min or overnight at 4 $^0$C. The phage was centrifuged again in an IEC Model HT centrifuge at 9000 rpm for 20 min at 4 $^0$C. The supernatant was discarded and the pellet was resuspended in 400 ul of a solution of 10 mM Tris-HCl, 1 mM EDTA pH 8.0. The phage containing solution was extracted twice with an equal volume of phenol and three times with an equal volume of chloroform. After the final chloroform extraction, the ssDNA in the aqueous phase was precipitated at -20 $^0$C overnight with 0.2 volumes of 5 M NaCl and 3 volumes of absolute ethanol. The mixture was microfuged for 10 min, the supernatant discarded and 1.5 ml of 80% ethanol was added to the pellet. This solution was incubated at room temperature for 1 h and microfuged again for 10 min. The supernatant was discarded and the pellet was dried under vacuum for 1.5 h. The dried ssDNA was resuspended in 50 ul U.P. water and the DNA concentration was determined spectrophotometrically at 260 nm.

**Sequencing of BCMV DNA**

Sequencing was performed using the dideoxy chain termination method of Sanger et al (1977). This method

involves the synthesis of a complementary DNA strand by DNA polymerase using a single-stranded DNA as a template. Synthesis is initiated at the unique site where an oligonucleotide primer binds to the template. Termination of the synthesis reaction occurs when a nucleotide analog (2',3'-dideoxynucleoside 5'-triphosphate-ddNTP) is incorporated. These analogs which lack the 3'-OH group required for DNA chain elongation do not allow continued DNA synthesis. When appropriate mixtures of dNTP's and one of the four ddNTP's are used, enzyme-catalyzed polymerization will be terminated at a site where a ddNTP can be incorporated. Four separate reactions are used to give complete sequence information. A radioactive nucleotide, usually dATP, is included in the synthesis, so the labeled chains can be visualized by autoradiography after separation of DNA fragments by denaturing polyacrylamide gel electrophoresis.

The sequencing reactions were performed by using the Sequenase$^R$ Version 2.0 kit protocol. A 10 ul annealing mixture, consisting of 1-2 ug of single-stranded template DNA, 4 to 10 ng of an appropriate sequencing primer and 2 ul of 5X sequencing buffer (200 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl) was incubated at 65 $^0$C for 2 min and slowly cooled to room temperature. Dideoxy termination mixtures were prepared by distributing 2.5 ul of the appropriate termination mixture (80 uM each of dATP, dCTP,

dGTP and dTTP; 8 uM of appropriate ddNTP, 50 mM NaCl) into 4 eppendorf tubes and prewarmed to 37 $^{0}$C. After the annealing mixture had cooled, 1 ul 0.1 M DTT, 2 ul of diluted dNTP mix (1.25 uM dCTP, 1.25 uM dGTP, 1.25 uM dTTP), 2 ul (2 U) of 1:8 dilution of Sequenase$^{R}$ Version 2.0 enzyme and 0.5 ul of alpha [$^{35}$S] dATP (Amersham->10mCi/ml, 1000 Ci/mmol) were added and the solution was mixed well and incubated for 2 to 5 min at room temperature. When the labeling incubation is complete, 3.5 ul of the reaction was added to each prewarmed termination tube, mixed well and incubated at 37 $^{0}$C for 5 min. The reactions were terminated by adding 4 ul of a stop solution (95% formamide, 20 mM EDTA, 0.05% Bromphenol blue, 0.05% xylene cyanol FF). The sequencing reactions were stored at -20 $^{0}$C until needed.

The sequencing reactions were electrophoresed on 6% polyacrylamide-urea denaturing gels. The ssDNA, which had been labeled internally by the incorporation of [$^{35}$S] dATP, were separated on the gel on the basis of size. A sealing gel was prepared by mixing 1.2 ml 50% acrylamide, 2.4 ml 5X TBE buffer and UP water to a total volume of 10 ml. The polyacrylamide-urea running gel was prepared by adding 5 ml 50% acrylamide, 21 g urea, 6 ml 5X TBE and UP water up to 50 ml. Eighty microliters of freshly prepared 10% ammonium persulfate and 5 ul of TEMED were added to the 10 ml sealing gel. The solution was immediately pipetted between the siliconized glass plates of the gel unit (Hoefer SE 1500

Sequencer, Hoefer Scientific Instruments, San Francisco, CA). After this gel was polymerized, 400 ul of 10% ammonium persulfate and 25 ul of TEMED were added to the 50 ml running gel. This solution was mixed and pipetted between the glass plates. A comb was placed into the top of the gel, clamped into place and the gel was allowed to polymerize for 20-30 min. Four liters of 0.6X TBE buffer was prepared for electrophoresis. The upper reservoir of the unit was filled with prewarmed (70 $^0$C) buffer and the lower reservoir contained buffer at room temperature. The gel was pre-electrophoresed with 0.6X TBE buffer at 50 mA. The sequencing samples were heated at 75 $^0$C for 2 min, loaded in 5 ul aliquots and electrophoresed at 50 mA and constant voltage of 1800 volts for 2 to 3 h at 45-55 $^0$C. After electrophoresis, the gel was transfered to filter paper, covered with a plastic sheet to prevent sticking and dried under vacuum at 80$^0$C for 2 h (Bio-Rad dual temperature slab gel drier Model 1125B) and exposed to X-ray film (AGFA Curex MR4) at room temperature for 24 h. Autoradiographs were developed according to Kodak's protocol.

**Sequencing of BCMV RNA**

RNA sequencing was performed by utilizing an appropriate oligonucleotide primer which was annealed to the RNA and served as the priming site for AMV reverse transcriptase.

The primed transcripts are truncated during the reaction by using dideoxy nucleotide analogs.

The sequencing reactions were performed by using the GemSeq Sequencing System protocol. A 10 ul annealing reaction, consisting of 1 ug of RNA transcript, 30 ng of an appropriate sequencing primer and 1 ul 10X RT buffer (340 mM Tris-HCl, pH 8.3, 500 mM NaCl, 60 mM $MgCl_2$, 50 mM DTT) was incubated at 65 $^0$C for 2 min and slowly cooled to 42 $^0$C. Dideoxy nucleotide mixtures were prepared by distributing 2.5 ul of the appropriate GemSeq System nucleotide mix into four Eppendorf tubes and placed at 4 $^0$C. After the annealing mixture has cooled and was centrifuged briefly, 5 units AMV reverse transcriptase was added to each reaction, and mixed well, followed by the addition of 4 ul of alpha $^{35}$S dATP. Three microliters of the labelling reaction was then added to each of the nucleotide mixtures and incubated at 42 $^0$C for 20 min. One microliter of chase solution was then added to each reaction tube and incubation at 42 $^0$C continued for an additional 15 min. All reactions were stopped by the addition of 5 ul of stop solution and samples were stored at -20 $^0$C. All reaction mixtures were heated at 65 $^0$C for 10 min before loading onto the sequencing gel.

# CHAPTER III

# RESULTS

The purpose of this research was to molecularly clone and sequence the coat protein gene of the NL3 strain of bean common mosaic virus and to compare the nucleotide sequence of this gene with other known potyviral coat protein sequences. Currently, there are no sequence data published on any strain of BCMV.

## Molecular Cloning

Viral RNA was isolated and purified using a guanidinium isothiocyanate extraction method and cesium chloride gradients. DNA complementary to BCMV RNA was synthesized using the Stratagene cDNA synthesis system, and inserted into a lambda Uni-ZAP XR vector. cDNA inserts that were cloned into the Uni-ZAP XR vector were excised with R408 helper phage and recircularized to form subclones in the pBluescript SK- phagemid vector. For production of double-stranded DNA, the recircularized pBluescript DNA was mixed with fresh E. coli cells and spread on LB/Amp plates to produce colonies. The DNA, isolated through mini-plasmid preparations of these colonies was used for the initial screening of recombinants. cDNA clones of 600 to 1300 bp

were obtained. The sizes of the inserts in the clones were initially estimated by digestion of the plasmid DNA with Xho I to linearize it, followed by electrophoresis on submerged 1% agarose gels or 5% polyacrylamide gels.

The molecular weights of several coat protein subunits have been determined by SDS-PAGE and these values range from 28K to 40K, with most estimates in the 33K to 34K range. With this molecular weight information and the location of the coat protein genes of potyviruses at the extreme 3' ends of the viral RNAs, which contains a poly(A) tail, clones of approximately 1 kb from the BCMV-NL3 strain were sufficiently long to encode the coat protein gene, because the cDNA's were synthesized by oligo (dT) priming. Three clones, NL3-1, NL3-19 and NL3-38, ranging in size from approximately 1.3 kb to 0.6 kb, were selected for characterization and sequence analysis.

A randomly primed DNA hybridization probe was synthesized from the linearized plasmid pBluescript NL3-1. A dot blot hybridization was performed to confirm that the NL3-1 recombinant did hybridize to the purified NL3 genomic RNA. No hybridization was obtained with tRNA (20mg/ml) which was utilized as a negative control (Fig. 1).

**Restriction Enzyme Analysis**

To determine the presence of different restriction endonuclease sites in the NL3-1 insert, 0.5 ug plasmid was

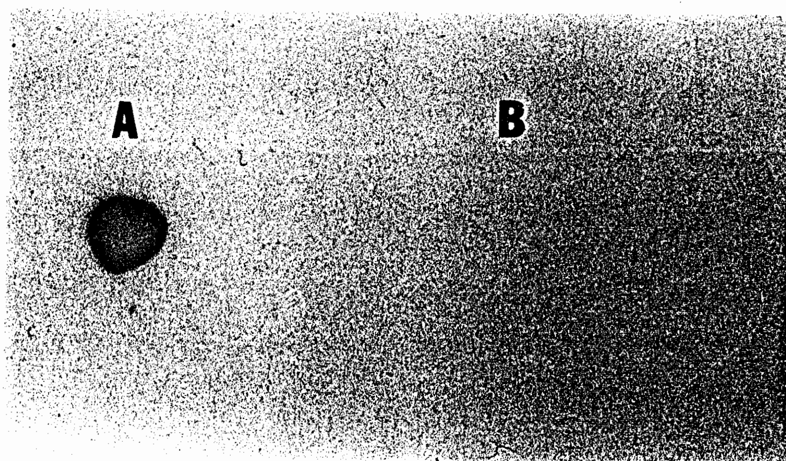


Fig. 1. A dot blot autoradiograph of the linearized pBluescript NL3-1 probe hybrization to BCMV NL3 RNA. (A) 5 ug NL3 RNA (B) 40 ug tRNA utilized as a negative control.

digested with selected restriction enzymes. Single digestions were performed with Bam HI, Eco RI, Hind III, Pst I, Xho I and Xba, followed by electrophoresis on a 5% polyacrylamide gel. Analysis of the pBluescript NL3-1 plasmid indicated that the 1.3 kbp insert contained no Pst I, Bam HI or additional Xho I sites. Cleavage was observed with Xba, Hind III and Eco RI (Fig. 2).

The NL3-1 insert was then purified from the pBluescript plasmid by digestion with Bam HI and Kpn I and single restriction enzyme digestions were performed using Xba, Sau 3AI and Taq I (Fig. 3, Fig. 4). Xba was determined to have one cleavage site in the insert resulting in two discrete bands measuring approximately 350 bp and 950 bp. Digestion with Taq I showed two restriction sites resulting in three bands measuring approximately 580 bp, 500 bp and 250 bp. However, no restriction sites were observed following treatment with Sau 3A1.

After sequencing the NL3-1 coat protein gene, restriction site analysis was performed using the DNASIS program (Hitachi Software Engineering) and the NL3-1 nucleotide sequence data. The restriction maps from the computer program confirmed the presence of a single Xba site at +923 bases from the 5'end of NL3-1 gene. This divides the gene into two bands measuring 923 bp and 344 bp. Eco RI cuts the NL3-1 insert once at +746 resulting in two distinct bands, one measuring 746 bp and the other measuring 521 bp.

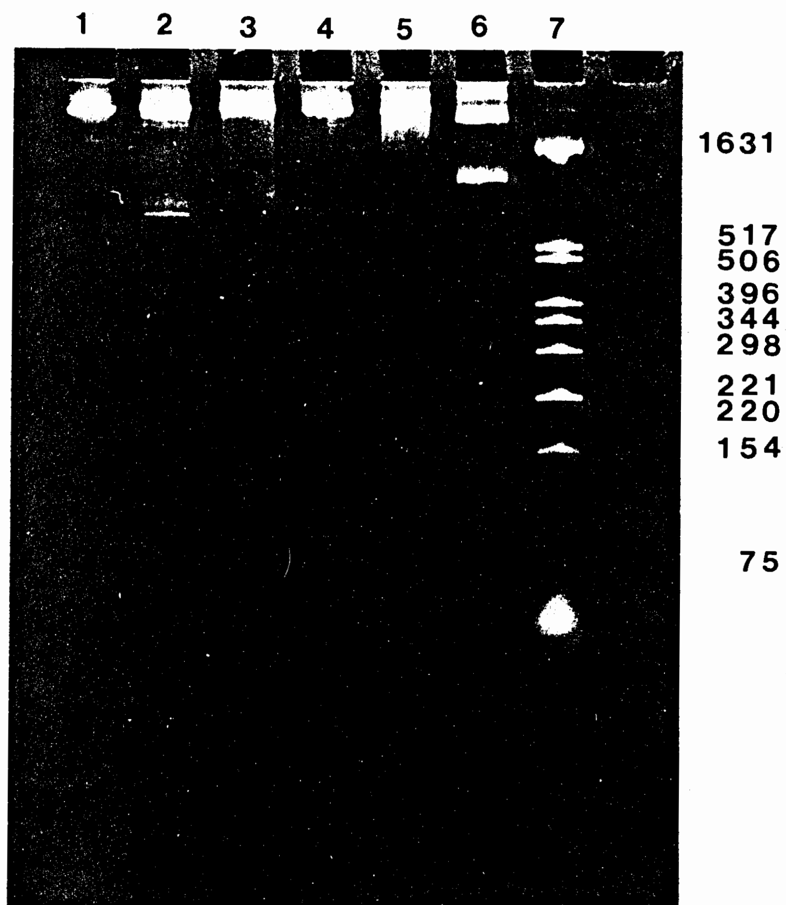


Fig. 2. Restriction endonuclease mapping of pBluescript NL3-1 on a 5% polyacrylamide gel electrophoresed with 0.5X TBE buffer for 3 h at 80V and stained with 1 ug/ml EtBr. Lane 2: 0.5 ug NL3-1 digested with Eco R1. Lane 3: 0.5 ug NL3-1 digested with Hind III. Lane 4: 0.5 ug NL3-1 digested with Pst 1. Lane 5: 0.5 ug NL3-1 digested with Xho I. Lane 6: 0.5 ug NL3-1 digested with Xba. Lane 7: 0.5 ug pBR322 Hinf I digested marker.

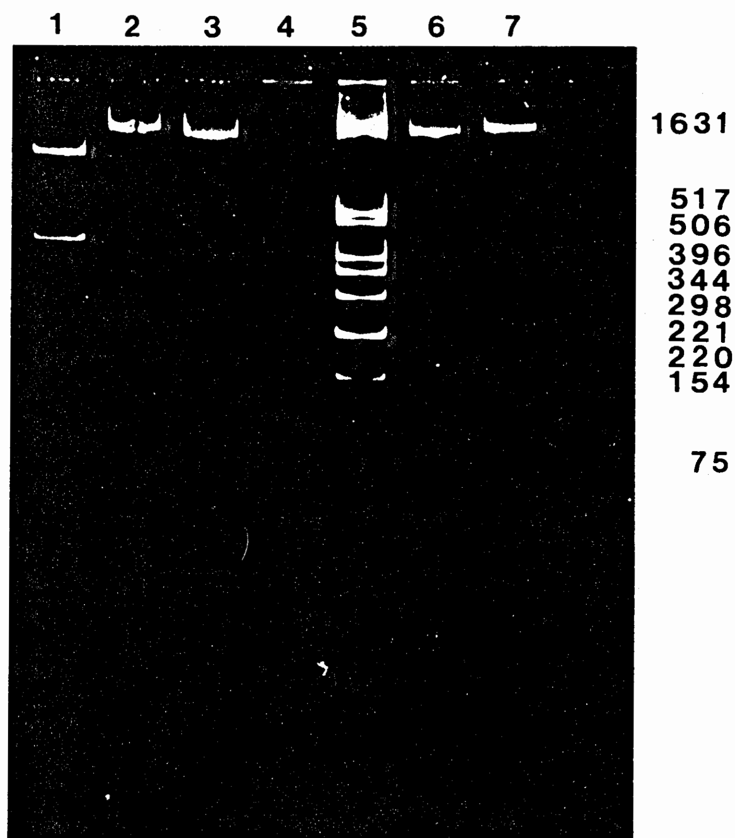


Fig. 3. Restriction endonuclease mapping of the 1.3 Kbp NL3-1 fragment on a 5% polyacrylamide gel electrophoresed with 0.5X TBE buffer for 3 h at 80V and stained with 1 ug/ml EtBr. Lane 1: 0.5 ug NL3-1 digested with Xba. Lane 3: 0.5 ug NL3-1 digested with Sau 3A1. Lane 5: 0.5 ug pBR322 Hinf I digested marker. Lane 6: 0.5 ug NL3-1 uncut DNA.

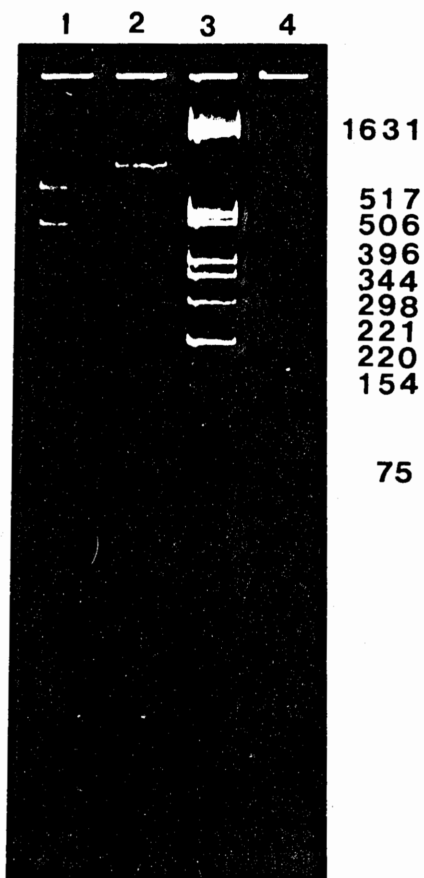


Fig. 4. Restriction endonuclease mapping of the 1.3 Kbp NL3-1 fragment on a 5% polyacrylamide gel electrophoresed with 0.5X TBE buffer for 3 h at 80V and stained with 1 ug/ml EtBr. Lane 1: 0.5 ug NL3-1 digested with Taq I. Lane 3: 0.5 ug pBR322 Hinf I digested marker.

Digestion with Hind III results in a single site located at 802 bp which would indicate fragments measuring 802 bp and 465 bp. The restriction enzyme Taq I cleaves the DNA at two distinct sites, +588 and +822 bp, yielding fragments of 588 bp, 465 bp and 214 bp. From these data, a fine restriction map indicating the enzyme recognition sites within the NL3-1 insert was proposed and constructed (Fig. 5).

**Nucleic Acid Sequence Analysis of NL3-1**

Nucleotide sequence studies on the BCMV NL3 strain were commenced to determine the coat protein gene sequence. The NL3-1 fragment was first sequenced using Sanger's dideoxy chain termination DNA sequencing method. The NL3-1 DNA fragment was first made single-stranded and was sequenced from the junction of the cDNA and the vector using the M13 universal primer and synthetic oligonucleotides. Primers were radiolabelled with $\alpha(^{35}S)ATP$, annealed to the BCMV single-stranded NL3 and extended with reverse transcriptase in the presence of dideoxynucleotides to obtain the sequence of the minus or anti-sense strand. After each sequencing gel (Fig. 6) was read and analysed, a new synthetic primer was designed from the sequences obtained and the NL3-1 minus strand was sequenced by "walking" the single-stranded DNA until the poly(A) tail was reached. Clones NL3-1, NL3-19 and NL3-38 were each sequenced according to the strategy

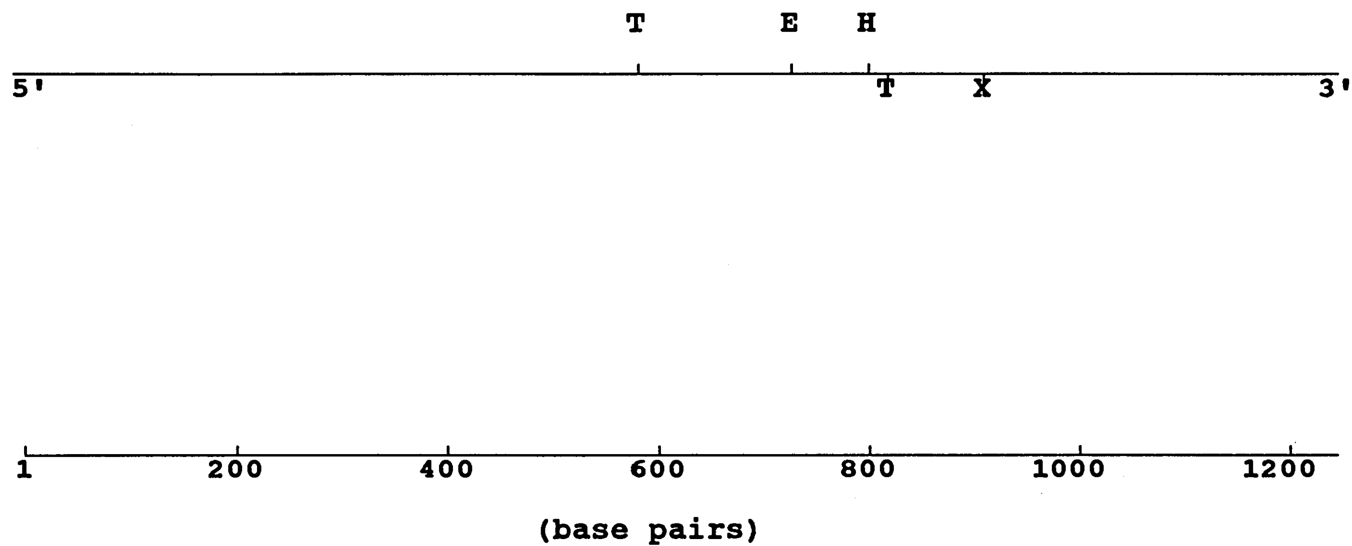


Fig. 5. Restriction endonuclease map of the BCMV NL3 1267 bp fragment. Restriction sites are as follows: T = Taq I, E = Eco RI, H = Hind III, X = Xba I.

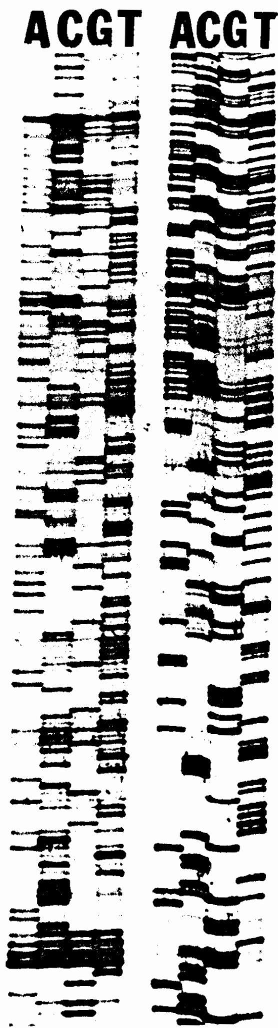


Fig. 6. Autoradiograph depicting the nucleotide sequence of a partial segment of the ssDNA NL3-1 electrophoresed on a denaturing 8% polyacrylamide gel at 1500V for 4 h.

shown in Fig. 7. Initially, the plus or sense strand was designed to be sequenced by digesting the purified insert with Taq I which cut the insert into three smaller fragments. These Taq fragments were directionally subcloned into M13 and sequenced. However, only partial sequences were obtained, so direct RNA sequencing was employed to obtain the remaining sequences. The genomic virus is a single-stranded positive sense RNA which was purified to use as a template for sequencing. Sequencing was initiated using synthetic oligonucleotides which were designed from the known sequences of the minus strand.

The NL3-1 insert was determined to be 1267 nucleotides. All three of the clones examined, NL3-1, NL3-19 and NL3-38, contained 3' poly(A) tracts 28 nucleotides long as well as identical sequences adjacent to the polyadenylate tail. The resulting sequence was assembled and analysed on an IBM PC computer using the DNASIS software package (Hitachi Software Engineering).

The sequence of the 1267 nucleotides is at the 3' end of BCMV genomic RNA (Fig. 8). One open reading frame (ORF) was found on the positive strand starting with the initiation codon methionine (AUG) at nucleotide 24 and ending with a termination codon (UAA) at nucleotide 1028. There is a 3' untranslated region of 239 nucleotides which ends with 28 adenylate residues.

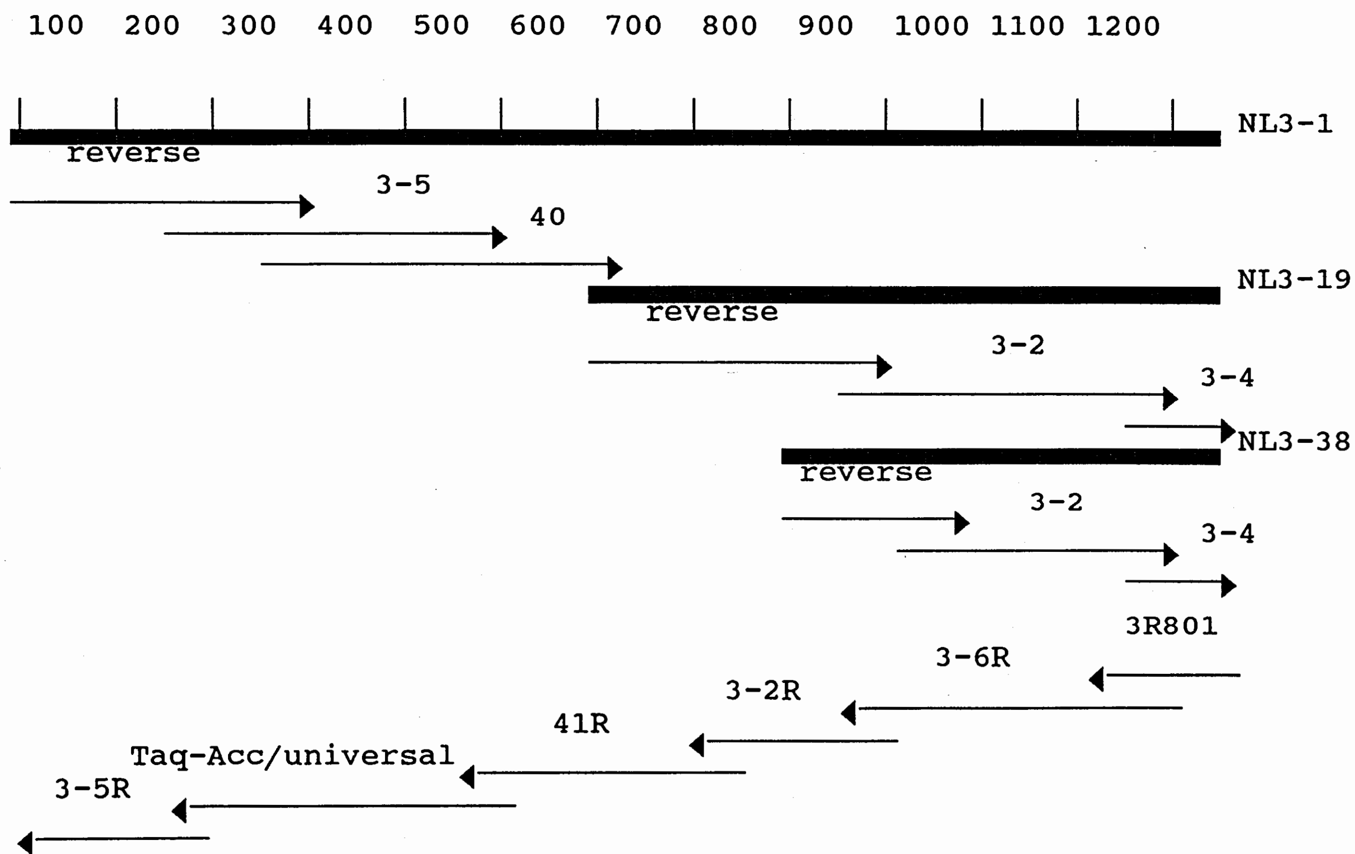


Fig. 7. Sequencing strategy for clones NL3-1, NL3-19, NL3-38. Each primer used is listed above the arrow which depicts the length of the sequence information in nucleotides.

```
*** DNA TRANSLATION ***

   1   GA ACC GAA GCT ATA TGT GCA GCA ATG ATT GAA GCC TGG GGG CAT ACT     47
   1       T   E   A   I   C   A   A   M   I   E   A   W   G   H   T     15

  48  GAA CTC CTT ACA GAA ATA AGG AAA TTC TAC TTG TGG CTC ATG GTA AAG     95
  16   E   L   L   T   E   I   R   K   F   Y   L   W   L   M   V   K     31

  96  ACA ATT TCA GGA ATT AGC TTT GAA TGG AAA AGC ACC ATA CAT AGC AGA    143
  32   T   I   S   G   I   S   F   E   W   K   S   T   I   H   S   R     47

 144  AAC AGC CCT TCG CAA\/GCT CTA TAC GGA CAA GAT GCC AAA ATG GAG GAA   191
  48   N   S   P   S   Q   A   L   Y   G   Q   D   A   K   M   E   E     63

 192  ATG CAA GAG TAC CTG AAA CAG CTT GAA TTT GAT TCT GAT GAT GAG GTG    239
  64   M   Q   E   Y   L   K   Q   L   E   F   D   S   D   D   E   V     79

 240  TAT GAA TCC GTG TCA ACA CAA TCC AGC AAG AAA GAA GAA GAG AAA GAC    287
  80   Y   E   S   V   S   T   Q   S   S   K   K   E   E   E   K   D     95

 288  GCT GGG GCC GAT GAG AGA GAG AAG GAC AAA GGC AAA GGA ACA GCG GAT    335
  96   A   G   A   D   E   R   E   K   D   K   G   K   G   T   A   D    111

 336  AAA GAC GTT GGA GCT GGC TCA AAA GGA AAA GTA GTG CCA AGA TTG CAG    383
 112   K   D   V   G   A   G   S   K   G   K   V   V   P   R   L   Q    127

 384  AAA ATC ACC AAA AAG ATG AAT TTG CCT ATG GTT GGC GGT AGG ATG ATT    431
 128   K   I   T   K   K   M   N   L   P   M   V   G   G   R   M   I    143

 432  CTA AAC TTG GAC CAC CTA ATT GAG TAC AAA CCG CAG CAG ACG GAC TTG    479
 144   L   N   L   D   H   L   I   E   Y   K   P   Q   Q   T   D   L    159

 480  TAC AAC ACA AGA GCT ACC AGG CAC AAT TTG AAG CAG TCA AGA CTG AAT    527
 160   Y   N   T   R   A   T   R   H   N   L   K   Q   S   R   L   N    175

 528  ATG AGC TTA ATG ACC AGC GAA ATG ACG GTA GTA ATG GAA TGG CGT TCA    575
 176   M   S   L   M   T   S   E   M   T   V   V   M   E   W   R   S    191

 576  TGG TGT GGT GCA TCG ATA ATG GAC ATC TCC GAT GTG AAT GGA GTG TGG    623
 192   W   C   G   A   S   I   M   D   I   S   D   V   N   G   V   W    207

 624  GTG ATG ATG GAT GGA GAT GAG TCA AAT AGA TAC CCA TTG AAG CCA TGG    671
 208   V   M   M   D   G   D   E   S   N   R   Y   P   L   K   P   W    223

 672  TTG AGA ATG CAA AGC CTA CGC TGC GAC AAT TAT GCA TCA TTT TTC AGA    719
 224   L   R   M   Q   S   L   R   C   D   N   Y   A   S   F   F   R    239

 720  TGC AGC GGA GGC TAT ATA GAG ATG AGG AAT TCT GAA GGG TTC TAC ATG    767
 240   C   S   G   G   Y   I   E   M   R   N   S   E   G   F   Y   M    255

 768  CCT AGG TAT GGA CTT CTT CGG AAT TTG AGG ATA AAA GCT TGG CTC GCT    815
 256   P   R   Y   G   L   L   R   N   L   R   I   K   A   W   L   A    271

 816  ATG CAT TCG ATT TCT ATG AGG TTA ACT CAA AAA ACC TCA GAC AGA GCC    863
 272   M   H   S   I   S   M   R   L   T   Q   K   T   S   D   R   A    287

 864  AGA GAG CGT TGC TCA GAT GAA AGC GCC AGC TCG CTA ACG TTA ACA CTA    911
 288   R   E   R   C   S   D   E   S   A   S   S   L   T   L   T   L    303

 912  GAT TGG TTT GGT CTA GAT GGT ACA CGT GGC AAC AAC CAG CGA GAA TAC    959
 304   D   W   F   G   L   D   G   T   R   G   N   N   Q   R   E   Y    319

 960  GAA AGG CAC ACT GCA CGG GAC GTC AAT CAA AAC ATG CAT CAT TTG CTT   1007
 320   E   R   H   T   A   R   D   V   N   Q   N   M   H   H   L   L    335

1008  GGT ATG ACT TCT GGG CAG TAA AGG AGT GGG GCA ACC CCT CTA CAG TTA   1055
 336   G   M   T   S   G   Q   *   R   S   G   A   T   P   L   Q   L    351

1056  GCA TCT CGC GTC GTT CAT AGT TTT CTG TAT TAG ATA GTA CGC TTT CAA   1103
 352   A   S   R   V   V   H   S   F   L   Y   *   I   V   R   F   Q    367

1104  TTC CAG TGT GTT ATA CCA CCT TGT GTC TAT GTA AGT TAG TAG TGG CTA   1151
 368   F   Q   C   V   I   P   P   C   V   Y   V   S   *   *   W   L    383

1152  TCA AGT ATG TTA TCT ATT CAG TTT ATG CGA GCA GGA GGA GCC ATT CCA   1199
 384   S   S   M   L   S   I   Q   F   M   R   A   G   G   A   I   P    399

1200  ACA CCG GAG CTG CCA GCG TGG TTA GTA CAT GAG GTG ACT GTC CGA GGT   1247
 400   T   P   E   L   P   A   W   L   V   H   E   V   T   V   R   G    415

1248  GAC GGT TAG AAT ATT CTC TA                                        1267
 416   D   G   *   N   I   L                                            421
```

Fig. 8. Nucleotide sequence of the 3'-terminal 1267 nucleotides of NL3 cDNA. Below the nucleotide sequence is the single letter abbreviation for the amino acid encoded for by the ORF codons. Nucleotides and amino acids are numbered from the 5'-terminus. The arrow indicates the putative cleavage site resulting in alanine as the amino-terminal amino acide of the capsid protein.

**Identification of Capsid Protein Coding Sequences**

Potyviral coat proteins are known to be released from longer precursors by proteolytic cleavage at glutamine-glycine, glutamine-serine or glutamine-alanine dipeptides (Dougherty et al., 1989). Based on this information, it is possible that the mature capsid protein is produced following cleavage of a glutamine-alanine dipeptide at position 156 within the 1267 bp NL3-1 sequence. Cleavage at this site produces a coat protein of 290 amino acids, with a calculated molecular weight ($M_r$) of 32,428. This agrees well with the results of SDS-polyacrylamide gel electrophoresis of the purified BCMV-NL3 strain coat protein which produces one protein band having a mobility corresponding to 30 - 32K (George, personal communication). There is a glutamine-serine site at position 261 which could be a possible cleavage site, but was ruled out due to lack of correspondence with molecular weight evidence, SDS-PAGE results and current literature expectations.

The 3' untranslated region was analysed and found to contain 239 nucleotides, excluding the polyadenylate tract which measured 28 adenosine residues. This is consistent with other potyviral 3' non-coding regions which vary in length from 147 bases to 667 bases (Ward and Shukla, 1991). These regions are reported to contain AU-rich segments (Turpen, 1989). A base composition analysis which was run

on the NL3-1 239 nucleotide 3' non-coding region shows 24% adenosine, 32% uridine, 24% guanosine and 20% cytosine.

**Comparison of BCMV Coat Protein Sequence with PSBMV and BYMV Coat Protein Sequences**

An alignment of the BCMV-NL3 coat protein nucleotide sequence with those of two other distinct members of the potyviral group has been evaluated by homology analysis utilizing the DNASIS software (Fig. 9, Fig. 10). This evaluation indicates that the BCMV-NL3 coat protein has extensive sequence similarity to the other potyviral coat proteins throughout its middle and C-terminal regions but, like other potyviruses, has less similarity in its N-terminal region. The amount of homology between the entire BCMV-NL3 coat protein and other potyviral coat proteins varies from 60% nucleotide sequence identity with BYMV to 57% with PSBMV. The percentage of sequence homology between these proteins is greater when only their middle and C-terminal regions are compared. The homology between the two-thirds C-terminal nucleotides of BCMV-NL3 and the aligned nucleotides from the other two potyviral coat proteins varies from 63.5% for BYMV to 58.5% for PSBMV. The 5' one-third nucleotide sequence shows less homology, 55% for PSBMV and 54% for BYMV.

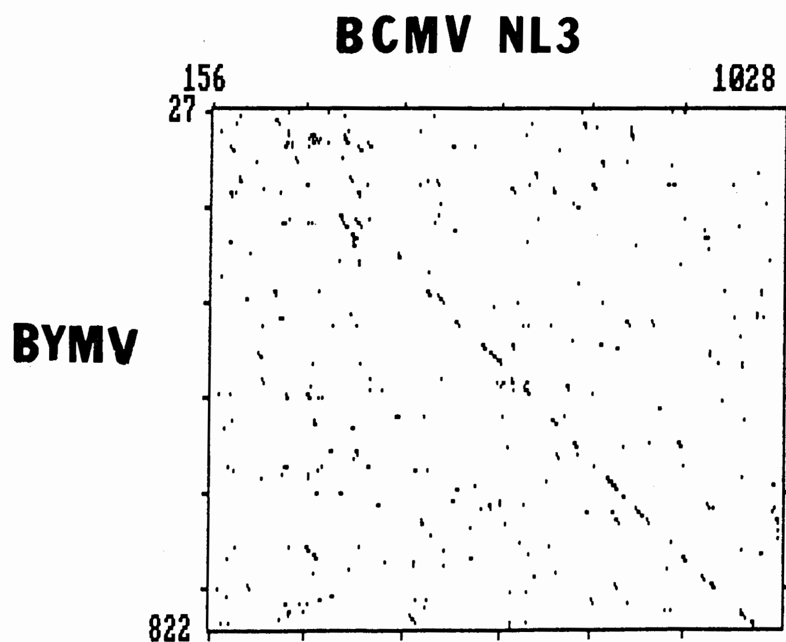


Fig. 9. Dot matrix homology comparison of the nucleotide sequence of the BYMV coat protein gene with the BCMV NL3 coat protein gene. The formation of a straight line is indicative of homologous sequences indicating a greater extent of homology at the 3' end of the gene.

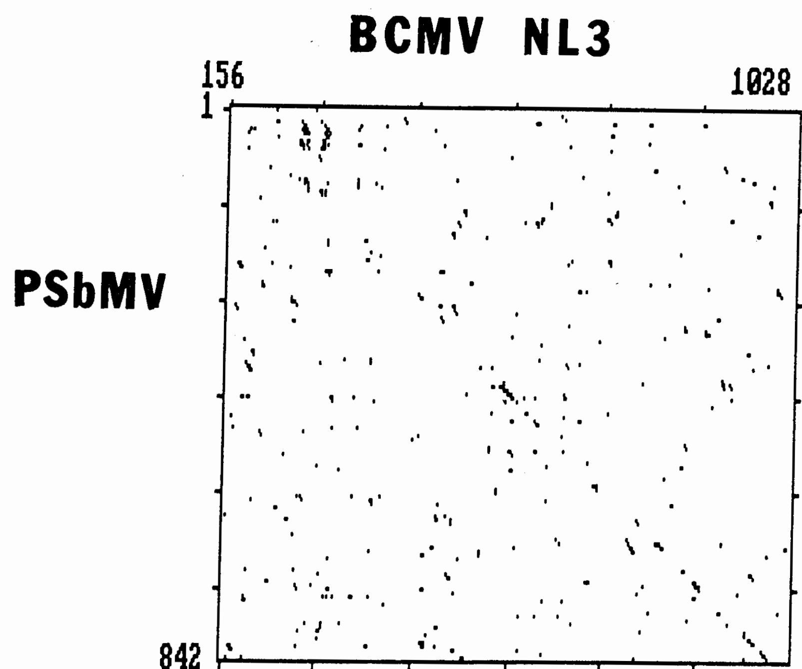


Fig. 10. Dot matrix homology comparison of the nucleotide sequence of the PSbMV coat protein gene with the BCMV NL3 coat protein gene. The formation of a straight line is indicative of homologous sequences indicating a greater extent of homology at the 3' end of the gene.

# CHAPTER IV

# DISCUSSION

A breakthrough in potyviral research was achieved in 1986 when the complete genome sequences of two members of this group, Tobacco etch virus (Allison et al., 1986) and Tobacco vein mottling virus (Domier et al., 1986) were reported for the first time. Since then, several additional potyviruses have been sequenced completely and intensive research has led to a greater understanding of potyviral genome structure allowing functions to be proposed for most of the potyviral gene products.

The gene product which has been the most extensively characterized in terms of sequence data is the capsid protein (Ward and Shukla, 1991). These data have been utilized to distinguish between viruses and between strains of viruses in the potyviral group. To date, 43 potyviruses have had their coat protein genes cloned and sequenced. These sequence data are being used by plant taxonomists to revise classifications of some viruses and strains. NL3 is one of eight described Netherland strains of Bean common mosaic virus (Drijshout, 1978) and there are no available sequence data published on any of the strains of BCMV.

Pea seed-borne mosaic virus (PSbMV) and Bean yellow

potyviral group of plant pathogens. PSbMV infects a variety of plant species, with the greatest economic impact occuring when peas are infected (Hampton and Mink, 1975). Like BCMV, the virus has a broad geographical distribution and has probably been spread throughout the world by infected seed. Bean yellow mosaic virus infects leguminous crops including beans and clovers and a variety of cultivated ornamental species (Bos, 1970). When BYMV infects a bean crop, the symptoms are indistinguishable from those produced by BCMV, and thus the crops will be quarantined, because BCMV is spread by infected seed, whereas BYMV is not. Farmers suffer great monetary and market losses due to BCMV infection, so it would be advantageous to them to have a quick diagnostic tool to determine which potyvirus was infecting their crops. Determination of the coat protein sequence of the NL3 strain of BCMV would make it possible to engineer a diagnostic probe for detecting BCMV infected plants quickly.

The sequence of the 1267 3'-terminal nucleotides of BCMV-NL3 cDNA is presented in Figure 6. Computer analysis revealed one large ORF of 1028 nucleotides in only one of the reading frames of the plus strand (virion polarity). The other reading frames of the plus strand and all three reading frames in the minus strand contained numerous stop codons and few extended ORF's. The large ORF was terminated

by the stop codon UAA located 239 residues upstream of the 3' poly(A) tail.

Potyviral coat proteins are known to be released from longer precursors by proteolytic cleavage at glutamine-glycine, glutamine-serine and glutamine-alanine dipeptide sequences. Similar dipeptide sequences have been described as polyprotein cleavage sites for comoviruses and picornaviruses (Wellink and van Kammen, 1988). The putative cleavage site for the production of BCMV-NL3 capsid protein has been predicted by examining the amino acid sequence of the long ORF. The most likely precursor cleavage site for the coat protein is between the glutamine-alanine dipeptide found at nucleotide 156. Cleavage at this site produces a coat protein of 290 amino acids with a calculated molecular weight of 32,428. These data are in agreement with the results of SDS-polyacrylamide gel electrophoresis of purified NL3, which produces one protein band with a mobility corresponding to 30 - 32K (George, personal communication). These results are in agreement with other published potyviral coat protein data. The coat proteins from eighteen distinct potyviruses vary considerably in size, ranging from 263 - 329 amino acids (Ward and Shukla, 1991). The SDS gel values range from 28K to 40K with most estimates in the 33 - 34K range (Shukla and Ward, 1989). PSbMV has a coat protein of 287 amino acids with a molecular weight of 32,651, while BYMV has a coat protein of 273 amino

acids with a molecular weight of 30,910 (Timmerman et al., 1990; Hammond and Hammond, 1989).

A second putative cleavage site was examined at nucleotide 261 which was comprised of a glutamine-serine dipeptide. This site would produce a coat protein of 255 amino acids with a calculated molecular weight of 28,886. This site probably is not utilized as the cleavage site because the resulting molecular weight would not be in agreement with published values. There were no other potential cleavage sites in the ORF.

A computer search of the aligned potyviral nucleotide sequences of BCMV-NL3, BYMV and PSbMV all showed considerable homology. Shukla and Ward (1988) have used coat protein sequence data to discriminate between distinct viruses which show 38 - 71% homology and between related strains of viruses which show 90 - 99% homology. In order to determine the degree of homology, the genome was divided into two regions, the N-terminal region and the C-terminal region. Research has revealed major differences in the sequence of the N-terminal region, but considerable sequence homology in the C-terminal two-thirds of the coat protein molecule (Shukla and Ward, 1989). Alignment of the N-terminal one-third BCMV-NL3 sequences with the PSbMV or BYMV N-terminal region displayed 56% or 53% homology, respectively. Alignment of the C-terminal two-thirds of the BCMV-NL3 sequences with the PSbMV or BYMV C-terminal region

displayed 59% or 62% homology, respectively. Computer alignment of the sequences was maximized by the insertion of gaps so that the longest homologous sequences could be located and aligned for identity.

In a comparative study of the nucleotide sequence of Watermelon mosaic virus II (USA) with other poytviruses, the closest relative to WMVII (USA) is WMVII (Aust.) which shares 93% homology. In contrast, the nucleotide sequences for the coat protein region of the other potyviruses were found to differ distinctly with most sharing approximately 65% identity (Quemada et al., 1990).

Analysis of the coat protein gene and the predicted nucleotide sequence indicate that BCMV-NL3 coat protein is similar to those of other potyviruses analysed, including BYMV and PSbMV which each exhibits overall sequence homology of approximately 57%. These results obviously indicate that these three viruses are distinct members of the potyvirus group, as was expected.

An alternative approach to discriminating between independent potyviruses and strains involves the 3' untranslated regions. Frenkel et al. (1989) have indicated that the nucleotide sequence of the 3' non-coding region of the potyvirus genome can serve as an aid in the identification and classification of potyviruses, especially when used in conjunction with the coat protein sequence data. By examining the 3' non-coding regions of distinct

viruses, the degree of sequence homology ranged from 39 to 53%. This is comparable to that obtained when the 3' untranslated regions of unrelated viruses from other plant groups are compared with potyviruses. However, the 3' untranslated regions of related strains are very similar both in length and nucleotide sequence. Frenkel determined that strains of the same virus will have 80% or more homology and distinct viruses will be less than 50% identical.

The BCMV genome is characterized by a 3' non-coding region of 239 nucleotides that is relatively (A + U) rich (55%). Many potyviruses have non-coding tails which range between 159 nucleotides in PSbMV and 667 nucleotides in TuMV (Timmerman et al., 1990). Bean yellow mosaic virus has a 3' non-coding region of 166 nucleotides which is 61% (A + U) rich (Hammond and Hammond, 1989). When the nucleotide sequences of these regions are compared, BYMV shows 52% homology and PSbMV shows 51% homology with BCMV-NL3, thus identifying BCMV-NL3 as a distinct virus.

The potyviruses are a group of plant viruses for which a significant amount of information, including amino acid and nucleotide sequences, is being assimilated, especially for their coat protein genes. Additional information concerning the coat protein genes of other potyvirus types will allow for further classification of viruses in this group. Furthermore, as these coat protein sequences are

examined and compared with other potyvirus coat protein sequences, it should become evident whether there are conserved and variable regions present which could be useful as diagnostic probes in the detection of infected plants. When the nucleotide sequences of BCMV, BYMV and PSbMV cDNA are compared, regions of extensive homology are observed (Fig. 9, Fig. 10) in the C-terminal portion. These areas of homology are up to 14 nucleotides long and may permit the synthesis of a nucleotide probe useful for the detection of several potyviruses in infected plants. A lesser extent of homology, 54-55%, is noted in the N-terminal portion of the capsid gene and may permit the synthesis of a nucleotide probe useful for the detection of a specific potyvirus in infected plants.

In conclusion, the nucleotide sequence of BCMV-NL3 cDNA encoding the coat protein is similar to that reported for BYMV and PSbMV, two other distinct potyviruses, except for its N-terminal region.

## REFERENCES


Allison, R.; Johnston, R.E.; Dougherty, W.G. The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for the synthesis of a single polyprotein. Virology 154: 9-20; 1986.

Bos, L. The identification of three new viruses isolated from Wisteria and Pisum in the Netherlands and the problem of variation within the potato virus Y group. Neth. J. Plant Path. 76: 8-46; 1970.

Carrington, J.C.; Dougherty, W.G. Processing of tobacco etch virus 49KD protease requires autoproteolysis. Virology 160: 355-362; 1987.

Domier, L.L.; Franklin, K.M.; Shahabuddin, M.; Hellman, G.; Overmeyer, J.H.; Hirimath, S.T.; Siaw, M.E.E.; Lomonossoff, G.P.; Shaw, J.G.; Rhoads, E. The nucleotide sequence of tobacco vein mottling virus. Nuc. Acids Res. 14: 5417-5430; 1986.

Domier, L.L.; Shaw, J.G.; Rhoads, R.E. Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. Virology 158: 20-27; 1987.

Dougherty, W.G.; Carrington, J.C. Expression and function of potyviral gene products. Annual Rev. Phytopath. 26: 123-143; 1988.

Dougherty, W.G.; Cary, S.M.; Parks, T.D. Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. Virology 71: 356-364; 1989.

Drijshout, E. Genetic interaction between Phaseolus vulgaris and bean common mosaic virus with implications for strain identification and breeding for resistance. Wageningen; Center for Agricultural Publishing and Documentation; 1978.

Frenkel, M.J.; Ward, C.W.; Shukla, D.D. The use of 3' non-coding sequences in the taxonomy of potyviruses: Application to watermelon mosaic virus 2 and soybean mosaic virus N. J. Gen. Vir. 70: 2775-2783; 1989.

Frenkel, M.J.; Jilka, J.M.; McKern, N.M.; Strike, P.M.; Clark Jr. J.M.; Shukla, D.D.; Ward, C.W. Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus. J. Gen. Vir. 72: 237-242; 1991.

Gough, K.H.; Azad, A.A.; Hanna, P.J.; Shukla, D.D. Nucleotide sequence of the capsid and nuclear inclusion protein genes from the Johnson grass strain of sugarcane mosaic virus RNA. J. Gen. Vir. 68: 297-304; 1987.

Grumet, R.; Fang, G. cDNA cloning and sequence analysis of the 3'-terminal region of zucchini yellow mosaic virus RNA. J. Gen. Vir. 71: 1619-1622; 1990.

Hammond, J.; Hammond, R.W. Molecular cloning and expression in Escherichia coli of the bean yellow mosaic virus coat protein gene. J. Gen. Vir. 70: 1961-1974; 1989.

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

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

Hellman, G.M.; Hiremath, S.T.; Shaw, J.G.; Rhoades, R.E. Cistron mapping of tobacco vein mottling virus. Virology 151: 159 - 171; 1986.

Lain, S.; Reichmann, J.L.; Garcia, J.A. RNA helicase: A novel activity associated with a protein encoded by a positive strand RNA virus. Nuc. Acids Res. 18: 7003-7006; 1990.

Maniatis, T.E.; Fritsch, E.F.; Sambrook, J. Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Press; 1982.

Messing, J. New M13 vectors for cloning. Methods and Enzymology. 101: 201-278; 1983.

Quemada, H.L.; L'Hostis, B.; Gonsalves, D.; Reardon, I.M.; Heinrikson, R.; Hiebert, E.L.; Sieu, L.C.; Slightom, J.L. The nucleotide sequences of the 3'-terminal regions of papaya ringspot strains W and P. J. Gen. Vir. 71: 203-210; 1990.

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

Shahabuddin, M.; Shaw, J.G.; Rhoads, R.E. Mapping of the tobacco vein mottling virus VPg cistron. Virology 163: 635-637; 1988.

Siaw, M.F.; Shahabuddin, M.; Ballard, S.; Shaw, J.G.; Rhoads, R.E. Identification of a protein covalently linked to the 5'-terminus of tobacco vein mottling virus RNA. Virology 142: 134-143; 1985.

Shukla, D.D.; Gough, K.H.; Ward, C.W. Coat protein of potyviruses. 3. Comparison of amino acid sequences of the coat proteins of four Austrailian strains of sugarcane mosaic virus. Arch. Vir. 96: 59-74; 1987.

Shukla, D.D.; Ward, C.W. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. J. Gen. Vir. 70: 1961-1974; 1988.

Shukla, D.D.; McKern, N.M.; Ward, C.W. Coat protein of potyviruses. 5. Symptomatology, serology and coat protein sequences of three strains of passion fruit woodiness virus. Arch. Vir. 102: 221-232; 1988.

Shukla, D.D.; Ward, C.W. Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. Adv. in Vir. Res. 36: 273-314; 1989.

Thornbury, D.W.; Hellman, G.M.; Rhoads, R.E.; Perone, T.P. Purification and characterization of potyvirus helper component. Virology 144: 260-267; 1985.

Timmerman, G.M.; Calder, V.L.; Bolger, L.E.A. Nucleotide sequence of the coat protein gene of pea seed-borne mosaic potyvirus. J. Gen. Vir. 71: 1869-1872; 1990.

Tremblay, M.F.; Nicolas, O.; Sinha, R.C.; Lazure, C.; Laliberte, J.F. Sequence of the 3'-terminal region of turnip mosaic virus RNA and the capsid protein gene. J. Gen. Vir. 71: 2769-2772; 1990.

Turpen, T. Molecular cloning of a ptotato virus Y genome: Nucleotide sequence homology in non-coding regions of potyviruses. J. Gen. Vir. 70: 1951-1960; 1989.

Vos, P.; Verver, J.; Jargle, M.; Wellink, J.; van Kammen, A.; Goldbach, R. Two viral proteins involved in the proteolytic processing of the cowpea mosaic virus polyproteins. Nuc. Acids Res. 16: 1967-1986; 1988.

Ward, C.W.; Shukla, D.D. Taxonomy of potyviruses: Current problems and some solutions. Intervirology 32: 269-296; 1991.

Wellink, J.; van Kammen, A. Proteases involved in the processing of viral polyproteins. Arch. Vir. 98: 1-26; 1988.

# APPENDIX A

## AMINO ACID SYMBOL LIST

| Amino Acid | 1-character abbreviation |
|---|---|
| Alanine | A |
| Arginine | R |
| Asparagine | N |
| Aspartic Acid | D |
| Cysteine | C |
| Glutamine | Q |
| Glutamic Acid | E |
| Glycine | G |
| Histidine | H |
| Isoleucine | I |
| Leucine | L |
| Lysine | K |
| Methionine | M |
| Phenylalanine | F |
| Proline | P |
| Serine | S |
| Threonine | T |
| Tryptophane | W |
| Tyrosine | Y |
| Valine | V |

# APPENDIX B

## RESTRICTION ENZYME SITE MAP FOR BCMV NL3-1

| Enzyme Name | Sequence | Start Position | | # of Sites |
|---|---|---|---|---|
| Bam HI | G!GATCC | | | 0 |
| Eco RI | G!AATTC | 746 | | 1 |
| Hind III | A!AGCTT | 802 | | 1 |
| Pst I | CTGCA!G | | | 0 |
| Sau 3AI | !GATC | | | 0 |
| Taq I | T!CGA | 588 | 822 | 2 |
| Xba I | T!CTAGA | 923 | | 1 |