1994

Characterization Of Cowpea Mottle Virus Genome And Expression Of The Viral Rna Replicase Gene

Xue-Juan You
Indiana State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

VITA

Name:		Xue-Juan You
Date of Birth:		August 6, 1956
Place of Birth:		Yantai, China

Education and Experiences

Ph.D Candidate	1991-1994	Microbiology Indiana State University
Research Associate	1987-1990	Plant Quarantine Inst.
Research Assistant	1982-1986	Dept. of Agriculture Beijing, China
Visiting Scientist	1986-1987	Australian National Uni. Canberra, Australia
B. S. in Biochemistry	1978-1982	Wuhan Uni. China

Publication

"Nucleotide Sequence of Cowpea Mottle Virus Genome and the Sequence Homology of Its RNA Replicase with Those of Carmoviruses and Tombusviruses." J. G. Virol. (submitted)

Research Presentations

"Cloning and Sequencing of Cowpea Mottle Virus Genomic RNA." Indiana Academy of Science Annual Meeting, Ball State Uni., 1992.

"Complete Nucleotide Sequence of Cowpea Mottle Virus Genomic RNA." Indiana Academy of Science Annual Meeting, Purdue Uni., 1993.

"Complete Nucleotide Sequence of Cowpea Mottle Virus Genomic RNA and the Sequence Homology of Its RNA Replicase With Those of Carmoviruses and Tombusviruses." American Virology Society Annual Meeting, (Poster) Uni. of Wisconsin, 1994.

Research Grants

Indiana Academy of Sciences, 1994. "Expression and Functional Analysis of Cowpea Mottle Virus RNA-dependent RNA Polymerase."

Sigma XI, 1994. "Expression and Functional Analysis of Cowpea Mottle Virus RNA-dependent RNA Polymerase."

Indiana Academy of Sciences, 1993. "Primary Structure Analysis of Cowpea Mottle Virus."

CHARACTERIZATION OF COWPEA MOTTLE VIRUS GENOME AND EXPRESSION OF THE VIRAL RNA REPLICASE GENE

A Dissertation

Presented to

The School of Graduate Studies

Department of Life Sciences

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Xue-Juan You

December 1994

APPROVAL SHEET

The dissertation of Xue-Juan You, Contribution to the School of Graduate Studies, Indiana State University, Series III, Number 634, under the title Characterization of Cowpea Mottle Virus Genome and Expression of the Viral RNA Replicase Gene is approved as partial fulfillment of the requirements for the Doctor of Philosophy Degree.

December 2, 1994
Date

Committee Chairperson

Committee Member

Committee Member

Committee Member

Committee Member

December 19, 1994
Date

For the School of Graduate Studies

ABSTRACT

Cowpea mottle virus (CPMoV) is a small isometric plant virus which causes severe diseases on cowpea plants. At present, it is a proposed member of Tombusviridae, carmovirus group according to some of its biophysical-chemical properties. In this study, the positive strand RNA of CPMoV composed of 1,029 nucleotides was clone and sequenced. Comparison of the nucleotide and the deduced amino acid sequences of CPMoV to those of other viruses revealed significant sequence homology to the sequences of Tombusviridae, especially the carmovirus group. The deduced amino acid sequence of the second major open reading frame (ORF2), the putative RNA replicase or part of it, has 53% and 41% sequence homology to the replicases of carmovirus and tombusvirus. These data support the assignment of CPMoV to the Tombusviridae, carmovirus group.

The 1.5 kb RNA replicase gene (ORF2) of CPMoV was cloned into the expression vector pGEX-2T for expression as a fusion protein with glutathione S-transferase (GST). However, the recombinant fusion vector failed to express the 83 kD fusion protein. The possible reasons for the failure are discussed.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Robert F. Bozarth, my dissertation advisor, for his guidance and support during my graduate studies and the preparation of this manuscript; I would like to express my appreciation to the other members of my dissertation committee: Drs. William H. Flurkey, Gary W. Stuart, Michael W. King and Patrick Pfaffle for their time, patience and valuable suggestions which led to the success of this project. I would also like to thank my colleague Dr. Anis H. Khimani for his suggestions and friendship.

I would like to express my special appreciation to my husband Feng-Chun Lin, for his great understanding and support, to my son Steven Lin, for his excellent work in school, so as a wife and a mother, I could still be able to dedicate most of my time and energy to this project.

Finally, I would like to express my appreciation to the Indiana Academy of Sciences and Sigma XI for their funding of this research project.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figures Page

INTRODUCTION

Properties of Cowpea Mottle Virus from Previous Studies

Cowpea mottle virus (CPMoV) is a proposed member of Tombusviridae, carmovirus group (Kim and Bozarth, 1992). It was first described by Robertson (1963, 1966). Shoyinka et al. (1978) described the gross biophysical properties of the virus and produced antiserum which has been tested against cowpea virus isolates throughout the world. CPMoV is extremely virulent in cowpea, and yield losses of 50% have been reported (Shoyinka et al., 1978; Rossel, 1977; Allen, 1980). Until recently, this virus was only reported in Nigeria (Bozarth and Shoyinka, 1979). It has now been reported in Senegal, West Africa (Ndiaye et al., 1993), Western Pakistan (Bashir and Hampton, 1993) and in seeds from Botswana, Africa (Hampton, R. O, USDA-ARS, Corvallis, USA; personal communication, 1993). CPMoV is isometric, 30 nm in diameter, and sediments as a single component at 122 S. CPMoV has two natural hosts: cowpea (Vigna unguiculata (L.) Walp.) and bambarra groundnut (Voandzeia suberranea (L.) Thou.) (Rossel, 1977). The bambarra nut isolate is less pathogenic than the cowpea isolate (Allen, 1980). They

also differ in host range but not in serological properties (Shoyinka et al., 1978). CPMoV is principally transmitted through sap, but it is also transmitted by galerucid beetles (Ootheca nutabilis) and through about 10% of seeds from infected plants (Shoyinka et al., 1978). Subsequent study by Allen et al. (1982) suggested that seed transmission was relatively unimportant in the natural infection, occurring in less than 0.4% of seeds. This is probably the main reason for this virus's limited geographical distribution in the past. However seed transmission most likely accounts for the recent appearances outside of West Africa noted above.

CPMoV is serologically distinct from other viruses. Purified CPMoV did not react with any of the antiserum of 19 isometric viruses and none of the 16 isometric viruses reacted with CPMoV antiserum (Shoyinka et al., 1978).

CPMoV has a positive strand RNA genome (Mr 1.4 x 10^6 daltons) and a capsid of 180 protein subunits (Mr 39,611 daltons). The capsid protein gene of 1,104 nucleotides is located near the 3' terminus (Kim and Bozarth, 1992). The postulated amino acid sequence of CPMoV capsid protein had 36% and 27% homology to the S domains of turnip crinkle and carnation mottle virus, but was not serologically related to these or other carmoviruses tested (Kim and Bozarth, 1992). Like other members of carmovirus (Morris and Carrington, 1988; Carrington et al., 1989), CPMoV generates three double strand RNAs which are co-terminal with the 3' end of the

genomic RNA, but no subgenomic RNAs were found encapsidated in the virus particles, and no polyadenylation was found at the 3′ end of the genomic RNA (Kim and Bozarth, 1992). Three species of dsRNA (4, 1.8 and 1.4 kb), all viral in origin, were obtained consistently from CPMoV-infected cowpea leaves. Their sizes were very similar to those of carmoviruses (4, 1.7 and 1.5 kb), but different from those of tombusviruses (4.8, 2.2, and 0.9 kb). The dsRNAs possibly function as replicative forms for synthesis of mRNA as in the case of carmoviruses and tombusviruses (Morris et al., 1983; 1988). Although denatured dsRNA of CPMoV can function as mRNA in in vitro translation system, the translation products of dsRNAs were not as distinct as those from ssRNA and also less efficient than the latter. The capsid protein appeared to be produced most efficiently from the smallest dsRNA segment (Kim and Bozarth, 1992).

Classification of Plant Viruses

Plant viruses were originally classified as groups, based on their physico-chemical properties (Wildy, 1971; Fenner, 1976; Matthews, 1981). As genome sequences of more viruses became available, it became obvious that the similarity of virus genome size and organization provides more reliable information for classification of plant viruses (Goldbach and Wellink, 1988; Habili and Symons, 1989; Martelli, 1992). CPMoV has similar physico-chemical properties, genome size and genomic organization to two groups of viruses, carmovirus and tombusvirus, both in the

family Tombusviridae.

Carmoviruses

The carmovirus group consists of 8 members and 9 possible members. CPMoV is currently listed among the possible members (Morris and Carrington et al., 1988; Morris, 1991). So far, only 4 of these viruses have been sequenced; carnation mottle virus (CarMV) (Guilley et al., 1985), turnip crinkle virus (TCV) (Carrington et al., 1989), melon necrotic spot virus (MNSV) (Riviere and Rochon, 1990), and cardamine chlorotic fleck virus (CCFV) (Skotnicki et al., 1993). These viruses were assigned to the carmovirus group on the basis of their sequence and genomic organization as well as their physicochemical properties. In addition, maize chlorotic mottle virus, which was originally classified as a possible member of the sobemovirus group (Francki et al., 1991), appeared to be a carmovirus from its similar genomic organization and sequence similarity (Nutter et al., 1989; Ward, 1993). Others were considered tentative members of carmovirus, due to the fact that their sequence data was not yet available.

CarMV was the first member of the carmoviruses in which the nucleotide sequence and genomic organization was studied in detail. The genome of carnation mottle virus consists of 4,003 nucleotides, which codes for three major proteins. Starting from the 5′ end, there is a 69 nucleotide non-coding region preceding the first AUG codon at nt 70. The first open reading frame (ORF) encodes a 27

kD (p27) product. If the amber termination codon is suppressed, an 86 kD (p86) read-through product of ORF1/ORF2, which was proposed to be the functional RNA replicase, would be expressed. The 38 kD (p38) capsid protein was located near the 3' end of the genome, and synthesized from one or two encapsidated subgenomic RNA (Guilley et al., 1985). Another well studied carmovirus was turnip crinkle virus (TCV) (Carrington et al., 1989; Hacker et al., 1992). Sequence analysis of TCV genome revealed the presence of four ORFs in addition to the capsid gene. These could potentially encode proteins of 28 kD (p28), 88 kD (p88), 9 kD (p9) and 8 kD (p8). Translational read-through of the amber termination codon of the p28 gene could lead to the synthesis of a read-through protein of 88 kD (p88). The p9 ORF was not noted in the original genetic motif of the TCV genome (Carrington et al., 1989), but was tentatively added later by Riviere and Rochon (1990) after sequence comparisons. They suggested that a similar open reading frame was present in MNSV. Mutational studies of the functions of these genes by Hacker et al. (1992) revealed that the capsid protein gene and p8, p9 products were required for cell-to-cell movement of the virus in the host plant. They postulated that p8 might have an RNA-binding function, while p9 might have a plasmodesmata targeting function.

All carmoviruses share significant sequence similarity at the nucleotide level. The RNA replicases (ORF2) are

similar in length and exhibit 49.6% to 56.3% homology at the nucleotide level and 42% to 53% homology at the amino acid level. There are at least 10 consensus sequences shared by members of carmovirus and a lesser number by members of Tombusviridae within ORF2. However, ORF1 coding regions vary considerably in length, and there is no consensus sequence shared by all members of the carmovirus group or Tombusviridae. The lengths of the capsid proteins of carmovirus are very constant, although they share a lower amino acid sequence similarity (24.1% to 29%). MNSV replicase shares significant sequence similarity with other carmovirus; however MNSV has a p42 capsid protein, which appears more closely related to tombusviruses than carmoviruses (Riviere et al., 1989).

Tombusviruses

Members of tombusvirus have larger genomes. The best studied tombusvirus, tomato bushy stunt virus (TBSV) (Martelli et al., 1971) has a positive strand genomic RNA of 4776 nucleotides and two subgenomic RNAs with molecular weights of 2.2 kb and 0.9 kb (Hayes et al., 1988; Martelli et al., 1988; Morris and Carrington, 1988; Hillman et al., 1989). These subgenomic RNAs are co-terminal at their 3' ends to the genomic RNA and are also encapsidated in the virions (Hayes, et al., 1988). Most members of the tombusvirus group share a measurable degree of serological affinity to one another. The TBSV genome encodes five ORFs, the genomic organization is similar to that of the

carmovirus. Tombusvirus also codes for its putative replicase by read-through of the ORF1 (p33) amber termination codon to give a 92 kD (p92) ORF1/ORF2 replicase product (Hearne et al., 1990). The capsid proteins of tombusviruses are somewhat larger then those of carmoviruses. The 41 kD (p41) capsid protein ORF of tomato bushy stunt virus is downstream and starts 32 nucleotides after the replicase termination codon. Like the carmovirus capsid proteins, it is also translated from subgenomic RNA (Hillman et al., 1989; Hearne et al., 1990). The genomes of tombusvirus differ from those of the carmoviruses in having two additional ORFs at the 3' end of the genome that are translated in overlapping reading frames to give a 19 kD (p19) and a 22 kD (p22) products. These appear to be unrelated to other ORFs encoded by small RNA viruses (Hearne et al., 1990).

MATERIALS AND METHODS

Molecular Cloning and Nncleotide Sequencing of Cowpea Mottle Virus Genomic RNA

Virus Purification

The cowpea strain of CPMoV used in these experiments was originally isolated by S.A. Shoyinka from cowpea near Ibadan Nigeria and sent to R.F. Bozarth who purified the virus, produced antiserum and described the basic physical-chemical properties (Shoyinka et al., 1978). The virus isolate has been stored at -20° C except when propagated to produce purified virus for study. It has been transmitted no more than 30 times over a 20 year period. It was submitted to the American Type Culture Collection in Oct. 1994 (ATCC#: PV-955). The virus was propagated in cowpea (Vigna unguiculata) cv. California No. 5 which was grown in a growth chamber at 27° C and 14 h of light per day. Virus was purified from systemically infected cowpea leaves by the procedure of Bozarth and Shoyinka (1979). CPMoV infected cowpea leaves were harvested 1-2 weeks after inoculation. Virus was purified by 6% polyethylene glycol (PEG) precipitation and two cycles of differential centrifugation.

Further purification was carried out by 10-50% sucrose density gradient centrifugation. The final virus pellets were resuspended in 0.1 M sodium phosphate buffer, pH 7.0, and stored at -70° C.

Viral RNA Extraction and Gel Electrophoresis

Viral RNA was extracted from purified virus particles by boiling 3 min in 10 mM sodium phosphate buffer containing 1% sodium dodecyl sulfate (SDS) and 1 mM ethylene-diamine-tetraacetate-disodium (EDTA), then extracted with an equal volume of phenol. The aqueous phase was extracted once with phenol/chloroform (1:1) and once with chloroform/ isoamylalcohol (24:1); the aqueous phase solution was precipitated with 2.5 volume of 95% ethanol and 0.1 volume of 3 M sodium acetate, pH 5.2, overnight at -20° C. After centrifugation, the pellet was washed with 70% ethanol, lyophilized, and stored at -20° C.

Gel electrophoresis of viral RNA was carried out in 1% agarose gel and run in 0.5X TBE buffer (45 mM Tris-borate, 1 mM EDTA pH 8.0). Since single strand RNAs (ss-RNA) were very sensitive to RNase, they were handled with care, and a RNase free environment was essential. All glassware was baked (70° C overnight), all buffers and solutions were prepared with diethyl pyrocarbonate (DEPC) treated water, and the electrophoresis apparatus was soaked with 3% hydrogen peroxide for 1 h, then rinsed with DEPC-treated water (Sambrook et al., 1989). Approximately 100 ng of RNA was loaded into the gel, electrophoresed, stained with

ethidium bromide and photographed.

cDNA Synthesis

A Promega cDNA synthesis kit (Madison, WI, USA) was used for cDNA synthesis from viral RNA. Two sets of cDNA libraries were generated. In one library, cDNA clones were generated by using random primers. To ensure that the library had clones containing the 3′ end sequence, oligo(dT) was also used for priming after poly(A) tailing of the viral RNA at the 3′ end. In the other library, cDNA clones were generated by using synthetic oligonucleotide primers designed to match specific sites within the genomic RNA. cDNA synthesis was performed by the method of Gubler and Hoffman (1983) using the Riboclone cDNA synthesis kit (Promega Corp.). The first strand cDNA synthesis was performed in a 25 ul reaction which had 5 ul of 5X first strand buffer [50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM spermidine, 10 mM dithiothreitol (DTT), 1 mM each of dATP, dCTP, dGTP, and dTTP], 25 units (U) of rRNasin ribonuclease inhibitor, 2.5 ul of 40 mM sodium pyrophosphate, 2 ug of viral RNA and 30 units of avian myeloblastosis virus (AMV) reverse transcriptase (Promega), and nuclease-free water. The mixture was incubated at 42° C for 1 h. The second strand cDNA synthesis was performed in a 100 ul reaction which included 20 ul of first strand reaction, 10 ul of 10X second strand buffer [50 mM Tris-HCl, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$ 50 ug/ml bovine serum albumin (BSA), 5 mM DTT], 23 U of Escherichia coli (E. coli) DNA

polymerase I (Promega), and 0.8 U of E. coli RNase H (Promega) and incubated at 14° C for 2-4 h. Aliquots containing 1/10 and 1/20 of the first and second strand reaction were transferred to other tubes, and to which aliquots of 0.5 and 1 ul of ^{32}P-dCTP (3,000 Ci/mmol, DuPont NEN) were added. The reactions were then heated to 70° C for 10 min to inactivate the enzyme. The primary reaction of the double strand cDNAs was blunt ended by adding 4 U of T4 DNA polymerase (Promega), then incubated at 37° C for 10 min. The reactions were extracted with an equal volume of phenol/chloroform, the aqueous phase was precipitated with ethanol in presence of 3 M sodium acetate (pH 5.2), and lyophilized. The first and second strand ^{32}P-labeled tracer reactions were subjected to 1.4% alkaline agarose gel [in 50 mM NaCl, 1 mM EDTA and equilibrated for 30 min in alkaline gel running buffer (30 mM NaOH and 1 mM EDTA)], electrophoresed at 4° C for 3-4 h, dried and exposed to Kodak XAR-5 film for autoradiography.

Ligation of EcoRI Adaptors to CPMoV cDNAs

EcoRI adaptors were ligated to the ends of the cDNA using T4 DNA ligase according to manufacturer's protocol (Promega Corp.). The lyophilized double stranded cDNA was resuspended in 10 ul of nuclease-free water in which 3 ul of 10X ligase buffer (300 mM Tris-HCl, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP), 3 ul of BSA (1 mg/ml), 1 ul (10 pmol) of EcoRI adaptors (Promega), 7.5 Weiss units of T4 DNA ligase, and water were added to a final volume of 30 ul.

The reaction was incubated at 16° C overnight (16 h). Then the reaction was heated to 70° C for 10 min to inactivate the ligase.

Kinase Reaction

Since both the vector (pT7/T3 18U, Pharmacia Biotech., Piscataway, NJ) and the EcoRI adaptors used in this cloning experiment were dephosphorylated at the ends, a kinase reaction was required in order to add a phosphate group to each sticky end of EcoRI adaptors. The kinase reaction was carried out in a 40 ul volume. To the 30 ul DNA/adaptor ligation reaction, 4 ul of 10X kinase buffer (700 mM Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 50 mM DDT), 2 ul of ATP (0.1 mM), 10 U of T4 polynucleotide kinase and water were added to a final volume of 40 ul, mixed and incubated at 37° C for 30 min.

Removal of Unligated Adaptors

Sephacryl S-400 spin columns (Pharmacia Biotech.) and Bio 101 (La Jolla, CA, USA) kits were used to remove the unligated EcoRI adaptors from the ligation mixture. In the spin column procedure, 1 ml of Sephacryl S-400 slurry was packed into a 2 ml syringe. The column was washed three times with TE buffer, centrifuged and drained. A 40 ul ligation mixture was applied to the top of the gel bed. The column was then placed in a collection tube and centrifuged in a swinging bucket at 800 x g for 5 min. The column was washed with TE buffer, centrifuged one more time, and the

eluates were analyzed by agarose gel electrophoresis.

In the Geneclean procedure, 40 ul of the product from the above kinase reaction was mixed with 3 volumes of NaI_2 solution and 5 ul of glass beads, incubated for 10 min at 25° C, inverting once every minute. The glass beads, bound with cDNAs containing at least 200 nucleotides, were centrifuged (5 sec), and washed 3 times with "New-wash" solution supplied in the Bio 101 kit. 5 ul of water was added and incubated at 50° C for 5-10 min to elute the cDNAs.

Ligation of CPMoV cDNAs to the Cloning Vectors

Double strand cDNAs with EcoRI ends were ligated to the dephosphorylated vector pT7/T3 18U (Pharmacia, Piscataway, NJ) as described by Sambrook et al., (1989). To the 5 ul CPMoV cDNA eluted from Geneclean, 100 ng of vector pT3/T3 18U DNA and water was added to a volume of 8.5 ul. The mixture was incubated at 45° C for 5 min and chilled on ice. Then 1 ul of T4 DNA ligase buffer, 0.5 Weiss Units of T4 DNA ligase and water were added and incubated at 16° C for 4 h to allow cohesive ends ligation.

Transformation of Competent Cells

The recombinant plasmids were used to transform E. coli strain DH5-alpha-F^- competent cells by heat shock following the manufacture's protocol (GIBCO BRL, Gaithersburg, MD, USA). Competent cells were prepared by the calcium chloride methods (Sambrook et al., 1989). 100 ul of competent cells

were transformed with 1 ul of ligation mixture, directly or diluted 10 fold in water. The mixture was incubated on ice for 15 to 30 min, then heat shocked at 42° C for 45 sec and chilled on ice for 2 min. 0.9 ml of SOB medium (2% tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, adjusted to pH 7.0 with 5 N NaOH) was added and the mixture was incubated at 37° C for 1 h with shaking (225 rpm/min). 100 ul of transformed DH5-alpha-F' cells was plated on each LB/agar (1.5%) plate containing ampicillin (60 ug/ml), X-gal (40 ug/ml) and IPTG (40 ug/ml) and then incubated at 37° C for 16 h.

Screening of White Colonies for Clones Containing CPMoV cDNA

White colonies were selected and tested for inserts by miniprep analysis. The colonies were transferred to tubes containing 2 ml Luria-Bertani medium [LB medium (1% bacto-tryptone, 0.5% yeast extract, and 1% NaCl, pH 7.0)] with 60 ug/ml ampicillin. The cultures were incubated at 37° C overnight (16-20 h) with shaking (225 rpm/min). Plasmid DNAs were isolated by the alkaline lysis method (Sambrook et al., 1989), which was a modification of the methods of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981). The 2 ml cultures were transferred to Eppendorf tubes and centrifuged. The cell pellets were resuspended in 100 ul of solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0 and 10 mM EDTA, pH 8.8). Then 200 ul of solution 2 (0.2 N NaOH and 1% SDS) was added and mixed to lyse the cells. 150 ul of potassium acetate, pH 4.8, was added and mixed to neutralize

the solution. The mixture was centrifuged to remove the pellet, and DNase-free pancreatic RNase was added to a final concentration of 20 ug/ml, then incubated at 37° C for 30 min. The supernatant was extracted with phenol/chloroform, centrifuged, and the DNA in the aqueous phase precipitated with ethanol and 3 M sodium acetate at -70° C for 30 min. The precipitate was washed twice with 70% ethanol and lyophilized. The pellet was resuspended in 20 ul of water, 2 ul of which was subjected to gel electrophoresis and photography.

EcoRI Digestion of Clones with CPMoV cDNA Insert

Clones containing a recombinant plasmid that ran slower than the vector in 1% agarose gel were selected for EcoRI digestion. The recombinant plasmids (0.2 ug/per reaction) were digested with 2 U of EcoRI in a 20 ul reaction at 37° C for 2 h, electrophoresed in 1% agarose gel in 0.5X TBE (45 mM Tris-borate, 1 mM EDTA, pH 8.0) buffer. The DNAs were then stained with ethidium bromide, visualized on a UV illuminator, and photographed.

Southern Blots

Southern blots were carried out as described by Sambrook et al. (1989), modified from the original procedure of Southern (1975). Clones containing plasmid with a CPMoV cDNA insert were digested with EcoRI and analyzed by agarose gel electrophoresis. DNAs in agarose gel were denatured twice for 15 min each in a solution containing 0.5 M NaOH

and 1.5 M NaCl, and neutralized for 15 min in 0.5 M Tris-HCl, pH 7.4. DNA bands in the gel were then blotted onto a nitrocellulose membrane (Schleicher and Schuell, Keene, NH) in 20X SSC buffer for 5 h. The treated membranes were incubated in pre-hybridization buffer [6X SSPE (1X = 0.15 M NaCl, 10 mM Na H_2PO_4, 1 mM EDTA, pH 7.4), 50% formamide, 200 ug/ml yeast tRNA, 5X Denhardt's solution (1X = 0.02% bovine serum albumin (BSA), 0.02% Ficoll, and 0.02% polyvinylpyrrolidone (PVP) (Denhardt, 1966)] for 4 h at 42° C. Hybridization was carried out under the same conditions as pre+wybridization. A first strand cDNA probe prepared by reverse transcription of CPMoV RNA in the presence of ^{32}P dCTP (1×10^8 cpm) was added, and the membrane was incubated at 45° C for 24-36 h. The hybridized membrane was washed with 6X SSC and 0.1% SDS three times at 25° C for 5 min each, once at 42° C for 15 min, and once at 50° C for 2 h. The membranes were then partially air-dried and exposed to Kodak XAR-5 film for autoradiography at -70° C.

Nucleotide Sequencing

DNA sequencing was carried out by Sanger's chain termination method (Sanger et al., 1977). A sequencing kit (Sequenase™ Version-2.0, US Biochemical, Cleveland, OH) was used and the protocols of Tabor and Richardson (1987) and Kraft et al. (1988) were followed. Plasmid DNAs were extracted by the alkaline lysis method as described above. Recombinant plasmid DNAs from 2-5 ml cultures grown for 16 h were denatured in 0.2 N NaOH and 0.2 mM EDTA at 25° C for 10

min. The mixtures were neutralized in 250 mM Tris-HCl, pH 4.5, and 0.3 M sodium acetate. The DNAs were precipitated with 2.5 volumes of 100% ethanol (Kraft et al., 1988) and lyophilized. The DNA pellets were resuspended in 7 ul of water and annealed with 1 ul (0.5 pmol) of either M13 universal forward primer (5'-GTTTTCCCAGTCACGAC-3'), pUC/M13 reverse primer (Promega Corp., 5'-CAGGAAACAGCTATGAC-3'), or specific primers (in 40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl) by heating to 95° C for 1 min and gradually cooling to below 35° C over 30 min. Primers were extended in a 10 ul reaction containing 0.5 uM each of dCTP, dGTP, dTTP, 6.5 uM DTT, ^{35}S-dATP (6 uCi), and 1 ul of diluted (8 fold) sequenase for 2-5 min at room temperature (25° C). Chain termination was carried out in four separate tubes. 3.5 ul extension solution was transferred to each of the four tubes containing 2.5 ul prewarmed ddNTP (dideoxynucleotides), and incubated at 37° C for 5 min. The reactions were inactivated by addition of 4 ul of stop solution (95% formamide, 20 mM EDTA and loading dye) and chilled on ice. The reaction mixtures were denatured at 80° C for 2 min, 2.5 ul of which was loaded into each lane and electrophoresed in a 6% polyacrylamide gel containing 8.3 M urea for 3-6 h at 55-60 W in 1X TBE on IBI Model STS 45 (BIO-RAD) sequencing gel electrophoresis apparatus. After electrophoresis, the gel was fixed in 15% methanol and 10% acetic acid in water for 15 to 30 min and dried at 80° C using a BIO-RAD Model 583 vacuum gel drier, or directly

vacuum dried without fixation. The gel was then exposed to X-ray film overnight (16-24 h) at room temperature. Different approaches were used to increase the number of nucleotides read from each gel. Less than 5-fold dilution of the labeling mixture was used to obtain sequences far from the primer. More than 5-fold dilution of the labeling mixture or addition of $MnCl_2$ was used to obtain sequences close to the primer. Different buffer systems of the anode (0.5X TBE) and cathode (1/2 volume of 0.5X TBE and 1/2 volume of 3M sodium acetate) were used to decrease the distance between the bands at the bottom of the gel. Two or three loadings were almost always employed to increase the number of readable nucleotides. IntelliGenetics (IntelliGenetics, Inc., Mountain View, CA, USA) and MicroGenie (Beckman, Fullerton, CA, USA) programs were used to analyze the sequence data.

5' End Labeling

The 5' end of RNA was labeled using T4 polynucleotide kinase and gamma ^{32}P-ATP. TMV RNA, which is capped at the 5' end (Zimmern, 1975), and MS2 viral RNA, which is not capped, were used as positive and negative controls. In addition, 0.24-9.4 kb RNA markers (GIBCO BRL, Gaithersburg, MD, USA), transcribed and capped at their 5' ends *in vitro*, also served as positive controls. Promega 5' end labeling protocols were followed. One pmol of each RNA and 1 ul of gamma ^{32}P ATP (1 uCi/ul) were used in each of the 10 ul reactions. The reaction mixtures were incubated at 37° C

for 30 min, electrophoresed in 1% agarose gel which was then stained with ethidium bromide and photographed. The same gel was dried and subjected to autoradiography. Then the RNA bands of MS2, CPMoV and TMV were excised from the gel, and counted in a Beckman DP-55000 Scintillation Counter.

5' End RNA Sequencing

The 5' end sequence of CPMoV was obtained by direct sequencing of viral RNA. The reactions were carried out by the method that was used for DNA sequencing described above, except that all the solutions were nuclease free. 1 ug of viral RNA was used as template; synthetic oligonucleotide complementary to the sequences near the 5' end (5'-TCTACAGGTGGCATTGTTCCG-3') was used as primer; and AMV reverse transcriptase was used for extension in the labeling reactions. 1 ul of rRNAsin was included after incubation at 90° C for 5 min for annealing. To obtain the sequences near the primer, 10-fold dilution of labeling mixture and $MnCl_2$ was used in several reactions.

Cloning and Expression of CPMoV RNA Replicase Gene

cDNA Synthesis

Promega cDNA synthesis kits (Madison, WI) and protocols were used for cDNA synthesis of CPMoV RNA replicase gene as described above. First strand cDNAs were synthesized from CPMoV RNA using a specific primer (24 nucleotides) and AMV reverse transcriptase. The primer was complementary to the sequences adjacent to the 3' end of CPMoV replicase gene

with two nucleotide mismatches to introduce an EcoRI restriction site. The primers used for the cDNA synthesis and the polymerase chain reactions below were:

```
                                    EcoRI
3' end primer: 5'-CCG TTG TTG AAT TCA TAA ATG AGC-3'
                              *   *
(CPMoV sequence)  GGC AAC AAC CTA GGT ATT TAC TCG

                                  BamHI
5' end primer: 5'-CGT GGC GGA TCC CCT GAC ATG C-3'
                            *  **
(CPMoV sequence)  CGT GGC GAA CAC CCT GAC ATG C
```

Polymerase Chain Reaction to Amplify CPMoV RNA Replicase Gene

The synthetic first strand cDNAs were used as templates for polymerase chain reaction (PCR) to amplify replicase DNA. The 3' end primer for PCR amplification was the same as for cDNA synthesis (see above). The 5' PCR primer has sequences adjacent to the 5' end of the CPMoV RNA replicase gene with three nucleotide mismatches to introduce a BamHI restriction site at the 5' end, which would make the CPMoV RNA replicase gene in-frame with the vector at the BamHI site. PCR was performed as described by Mullis and Faloona (1987) with modifications (Oste, 1988; Saiki et al., 1988). The reactions were carried out for 35 cycles (in 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 3 mM $MgCl_2$, 0.1% gelatin) for 1 min at 94° C (denaturation), 1 min at 55° C (annealing), and 2 min at 72° C (polymerization). In a pilot experiment, the effect of $MgCl_2$ concentration was tested to optimize the amplification conditions. Tubes containing 15 ul of the complete mixture for PCR, except $MgCl_2$, were set up. $MgCl_2$

was added to a final concentration of 1.5, 2.5, 3,5 and 4.5 mM respectively, and water to a final volume of 20 ul. The reactions were carried out under the same conditions as described above. PCR products were analyzed in 1% agarose gel, stained with ethidium bromide and photographed.

Cloning of PCR Amplified Replicase Gene into TA Cloning™ Vector

It turned out that PCR amplified CPMoV replicase DNA was extremely difficult to digest by restriction enzymes, even with overwhelming amounts of enzyme (25 units enzyme to 1 ug DNA) and overnight incubation (37° C). To solve this problem, an alternative approach was used. The PCR amplified CPMoV RNA replicase DNAs were cloned into TA cloning vectors (Invitrogen Co.). This would convert the external restriction sites into internal sites, which would be digested easily. The PCR amplified RNA replicase DNAs were diluted to a final concentration of 25 ng/ul in sterile water, ligated to 25 ng of vector DNA in an 11 ul reaction containing 1X ligase buffer and 0.5 Weiss units of T4 DNA ligase. This reaction was carried out at 12° C for 16 h. 1 ul of ligation mixture was used to transform 50 ul of E. coli strain INV-alphaF', by the method of heat shock in the presence of 2 ul 0.5 M B-mercaptoethanol (B-ME). Transformed competent cells were grown in SOC (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) medium at 37° C for 1 h, then 100 ul cell culture was plated onto LB/Agar plates containing ampicillin

(50 ul/ml), and incubated at 37° C for 16 h.

EcoRI Digestion of TA Clones Containing Replicase Gene

White colonies were screened by miniprep analysis and clones containing slow moving plasmid DNA were selected, and cultured in 2 ml of LB medium. The cultures were grown overnight, then plasmid DNAs were extracted and digested with EcoRI. A small scale pilot reaction of 20 ul containing 0.2 ug of plasmid DNA, and 2 U of EcoRI was incubated at 37° C for 2 h, then analyzed in 1% agarose gel and photographed. For a large scale preparation, the quantities of the components were increased by 10 fold. The 1.5 kb band, corresponding to the replicase fragment, was excised from the gel, and the DNA was eluted by GeneClean as described above.

EcoRI Digestion and Dephosphorylation of pGEX-2T

The pGEX-2T vector (Smith et al., 1986) DNA was digested with EcoRI under the same conditions as described above. To avoid self-ligation of the pGEX-2T vector DNA, calf intestinal alkaline phosphatase (CIAP) was used to remove the phosphate group from the ends of the linearized plasmid DNA as described by Sambrook et al. (1989). The dephosphorylation reaction was carried out in final volume of 50 ul containing 1 ug of plasmid DNA, 5 ul 10X CIAP buffer (500 mM Tris-HCl, pH 9.0, 10 mM $MgCl_2$, 1 mM $ZnCl_2$ and 10 mM spermidine), 0.5 U of CIAP. The mixture was incubated at 37° C for 30 min. Another 0.2 U of CIAP was added, and

the reaction was incubated for a further 30 min at 37° C. The reaction was extracted once with 1 volume of TE buffer saturated phenol/chloroform and once with chloroform. Then the DNA was precipitated with ethanol and lyophilized. The dephosphorylated and undephosphorylated vector pGEX-2T DNA was then examined by agarose gel electrophoresis. Approximately 50 ng of each of the vector DNA was used for self-ligation in a volume of 10 ul using T4 DNA ligase. 1 ul of the ligation mixture was used to transform 50 ul of E. coli DH-5 alpha competent cells by the heat shock method as described above. Then 100 ul of each was plated onto LB/agar plate containing ampicillin (50 ug/ml). The number of colonies from the two LB/agar plates was compared to determine the efficiency of the dephosphorylation of the vector pGEX-2T.

Ligation of Replicase Gene to Expression Vector pGEX-2T

CPMoV RNA replicase fragments from the TA clones were ligated to the dephosphorylated expression vector pGEX-2T at the EcoRI restriction site using T4 DNA ligase (Sambrook et al., 1989). To avoid the formation of CPMoV replicase gene concatemers, a molar ratio of approximately three to one dephosphorylated vector to CPMoV RNA replicase DNA was used. The reaction mixture containing 100 ng of vector DNA, 20 ng of replicase DNA and sterile water to a final volume of 8.5 ul was incubated at 45° C for 5 min, and chilled on ice. Then 1 ul of T4 DNA ligase buffer and 1 Weiss Unit of T4 DNA ligase were added. The reaction was incubated at 16° C for

2-4 h.

Transformation of E. coli Competent Cells and Screening for Clones Containing Replicase Gene

The expression vector pGEX-2T and replicase DNA ligation mixture was used to transform E. coli strain DH-5 alpha competent cells. 1 ul of ligation mixture was used to transform 100 ul of DH-5 alpha competent cells by the heat shock method as described before. 100 ul of transformation mixture was plated onto LB/agar plates containing ampicillin (50 ug/ml). Since the vector pGEX-2T does not have a LacZ' sequence, blue white selection was not available. Colonies were screened for insertions of CPMoV replicase fragment by miniprep analysis. The slower running recombinant pGEX-2T DNAs were digested by EcoRI and analyzed by agarose gel electrophoresis.

Determination of the Orientation of RNA Replicase Gene in pGEX-2T by BamHI Digestion

Clones containing recombinant pGEX-2T were selected by EcoRI digestion and then digested with BamHI to determine the orientations of the inserts in the vectors. Small scale pilot reactions containing approximately 100 ng of plasmid DNA each were digested with 5 U of BamHI in a volume of 20 ul at 37° C for 2 h, then analyzed by agarose gel electrophoresis. A large scale reaction containing approximately 3 ug of plasmid DNA from a clone with the correct orientation was digested with 45 U of BamHI in a volume of 50 ul, and subjected to agarose gel

electrophoresis to remove the small BamHI fragment (about 36 nucleotides), which otherwise would interrupt the in-frame fusion of the gene. The 6.5 kb band was excised from the gel, and the DNA was eluted by Geneclean. Approximately 50 ng of plasmid DNA was used for self-ligation by T4 DNA ligase at 12° C for 16 h. The recircularized recombinant pGEX-2T DNA was then used to transform E. coli DH5-alpha competent cells as described before.

Expression and Purification of CPMoV Replicase

Growth of transformed E. coli cells and purification of the expressed target protein were performed as described by Sankar and Porter (1991) and Plotch et al. (1989). In a small scale pilot experiment, transformants containing pGEX-2T with a replicase gene inserted in the correct orientation and the pGEX-2T with no insertion were each inoculated into 10 ml of LB medium containing ampicillin (50 ug/ml). The cultures were grown at 37° C with shaking (225 rpm/min). 50 ul of overnight cultures were used to inoculate each of the two tubes containing 3 ml fresh LB medium. The cultures were grown at 37° C until the OD_{600} reached 0.2-0.5 (1.5-2 h). Then IPTG (isopropyl-1-thio-beta-D-galactoside) was added to a final concentration of 0.5 mM. After a further 4-6 h or overnight incubation at 37° C, cells were transferred to Eppendorf tubes, centrifuged and resuspended in 60 ul of buffer 1 (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.25 mM EDTA). The resuspended cells were sonicated at a power setting of 2.5 for 20 sec in a Branson Sonifier 250

(Branson Ultrasonics Co., Danburg, CT). The cell mixtures were centrifuged for 5 min, and 30 ul of the supernatant was transferred to an Eppendorf tube, boiled for 2 min, and chilled on ice. The samples then were mixed with 10 ul of loading buffer (50 mM Tris/HCl, pH 6.8, 100 mM dethiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol) and boiled again for 2 min, then analyzed by 8% SDS-PAGE. The gel was electrophoresed at 56 V for approximately 1.5-2 h, fixed and stained with Coomassie Brilliant Blue solution (0.1%, 45% water, 45% methanol, 10% acetic acid) for 30 min with gentle shaking, then washed in destaining solution (staining solution without dye) overnight and photographed.

Different concentrations of IPTG and different times of addition of IPTG were tested to optimize the conditions for the induction of the glutathine S-transferase (GST) fusion protein. IPTG was added to a final concentration of 0.5, 1.0, 1.5, 2.0 or 2.5 mM at a cell density of OD_{600} = 0.2. In a large scale preparation, 3 ml LB culture was grown at 37° C for 16 h, then 1 ml was transferred to a flask containing 100 ml of fresh LB medium. IPTG was added to a final concentration of 0.5 mM at a cell density of OD_{600} = 0.2 (1-1.5 h). Cells were incubated for another 4 h, then centrifuged and sonicated as described above. After centrifugation, the supernatant of the cell lysate was loaded onto an affinity column (1 ml glutathione-Sepharose 4B) (Pharmacia Biotech., Inc.) which had been previously washed with 3 M NaCl and equilibrated in buffer 1. The

eluate was reapplied to the column one more time. Then the column was washed eight times with 10 ml of buffer 1. After the final wash, the gel slurry containing the absorbed fusion protein was transferred to an Eppendorf tube and washed twice with 500 ul of buffer 1. The gel slurry was resuspended in 70 ul of buffer 1, and then incubated with 100 ng of thrombin (Sigma Chemical Co.) at 4° C for 16 h. The GST-4B slurry was centrifuged, and the supernatant was saved. The slurry was washed two more times with 15 ul of buffer 1, and the washes were combined. An aliquot (30 ul) was analyzed by 8% SDS-PAGE.

RESULTS

Purification of CPMoV

Cowpea Mottle Virus (CPMoV) was purified by the method of Bozarth and Shoyinka (1979) from CPMoV infected cowpea leaves (Fig. 1). A yield of 40 mg was obtained from 200 g or infected cowpea leaves. The virus was stained with 1% uranyl acetate, and photographed in a Hatachi Hu-11 E Electron Microscope (Fig. 2).

Extraction of Viral RNA

CPMoV purified by sucrose density gradient (SDG) centrifugation was used to extract viral RNA. Viral RNA was assayed by UV spectrophotometry to determine the quantity and purity. An OD of 1 at 260 nm corresponds to approximately 40 ug/ml single strand viral RNA (Fig. 3).

Agarose Gel Electrophoresis of Viral RNA

CPMoV RNA extracted from purified virus was subjected to electrophoresis in 1% agarose gel. It appeared in the gel as a single band of approximately 4 kb (Fig. 4). A smear of low molecular weight molecules which ran faster in the gel was probably degraded viral RNAs.

Figure 1. Symptoms Caused by Cowpea Mottle Virus Infection on Cowpea Plants.

Figure 2. Electron Micrograph of Cowpea Mottle Virus Stained with 1% Uranium Acetate.

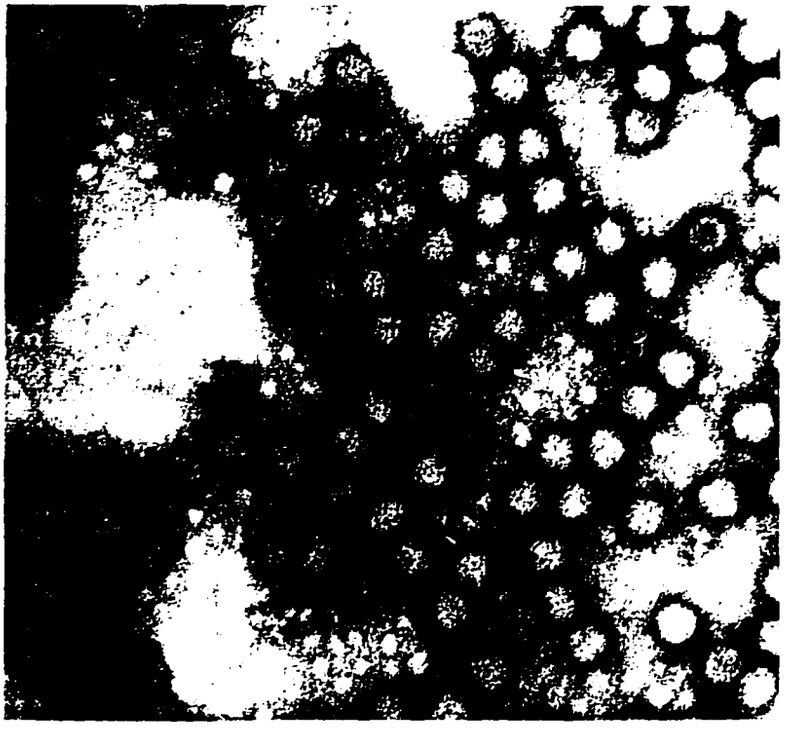


Figure 3. UV Spectrophotometry Profile of Purified Cowpea Mottle Viral RNA. 50 ul of RNA was diluted into 1 ml of water. 1 OD at 260 nm corresponds to 40 ug/ml of ssRNA. The yield of the RNA preparation was approximately 225 ug/ml.

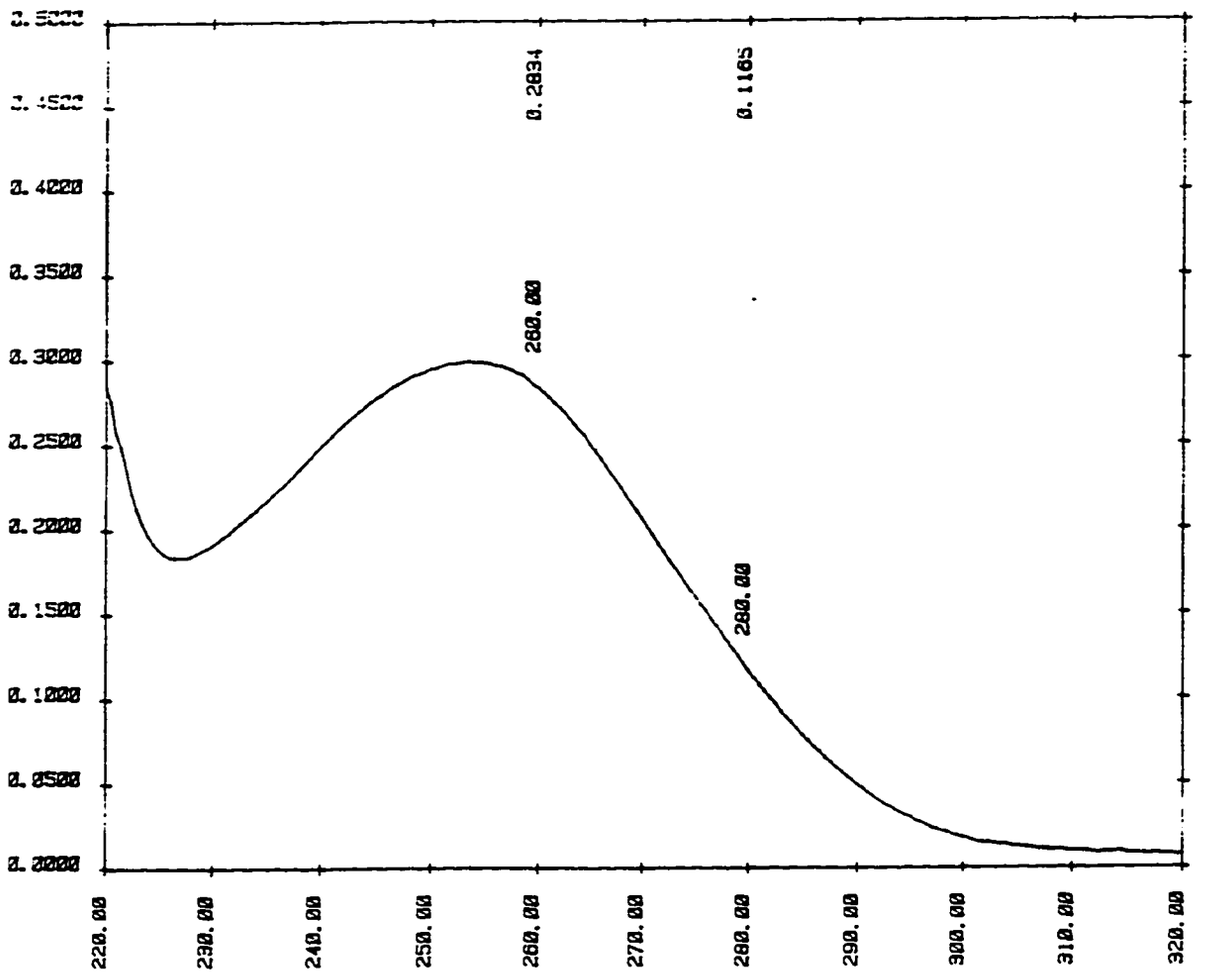
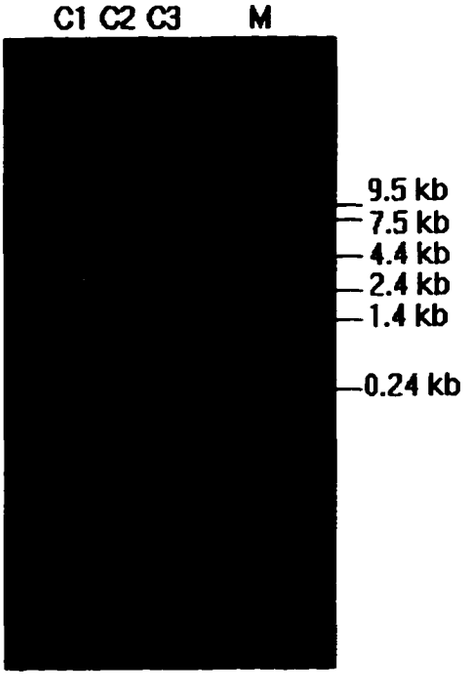


Figure 4. Electrophoresis Profile of Cowpea Mottle Viral RNA in 1% Agarose Gel. Approximately 100 ng of viral RNA extracted from purified virus was electrophoresed for 1.5 h at 60 V in 0.5X TBE buffer. After electrophoresis, the gel was stained with ethidium bromide and photographed under UV light. Lane C1, C2 and C3: CPMoV RNA, Lane M: 0.24-9.4 kb RNA markers (GIBCO BRL).



Molecular Cloning of CPMoV cDNAs

A Promega cDNA synthesis kit and protocols were used for the synthesis of CPMoV cDNA. In the initial experiment, template RNA and primer were heated to 65° C for 5 min, and then slowly cooled to room temperature (25° C) for annealing. The resulting cDNA fragments were very short (300-600 bp), probably due to the secondary structures regenerated by the RNA templates in the process of cooling to room temperature. The procedure was modified by cooling the reaction to 42° C instead of room temperature, after heating to 65° C. The sizes of the resulting cDNAs increased significantly (400-4000 bp). The synthetic CPMoV double strand cDNAs (Fig. 5 and 6) were ligated to pT7/T3 18U vector DNA at the EcoRI site (Fig. 7). Dephosphorylation of the ends of the cloning vector before ligation greatly decreased the efficiency of vector self-ligation. Since the dephosphorylated linear vector DNA lacked 5' end phosphates, self-ligation was minimal. In successful insertion ligations, the phosphate group would be supplied by the CPMoV cDNA. Clones containing CPMoV cDNA were digested with EcoRI and electrophoresed in 1% agarose gel to determine the sizes of the inserts (Fig. 8).

The Nucleotide Sequence of CPMoV

Nucleotide sequencing of the CPMoV cDNA clones was carried out by the chain termination method (Fig. 9), (Sanger et al., 1977). Over 95% of the sequences were determined from sequencing of two or more independent cDNA

Figure 5. Autoradiograph of Cowpea Mottle Virus cDNAs Following Electrophoresis in 1.4% Alkaline Gel. 12 ul of ^{32}P labeled second strand tracer cDNAs of CPMoV were electrophoresed at 30 V at 4° C for 3-4 h. The gels were fixed in 7% trichloroacetic acid solution for 30 min, partially dried and exposed to Kodak film at -70° C with an intensifying screen. Lanes M: End labeled lambda DNA HindIII fragments, lane C, C1 and C2: Double strand CPMoV cDNAs by different primers (primers at the 3′ end, central region, and near the 5′ end of the viral RNA).

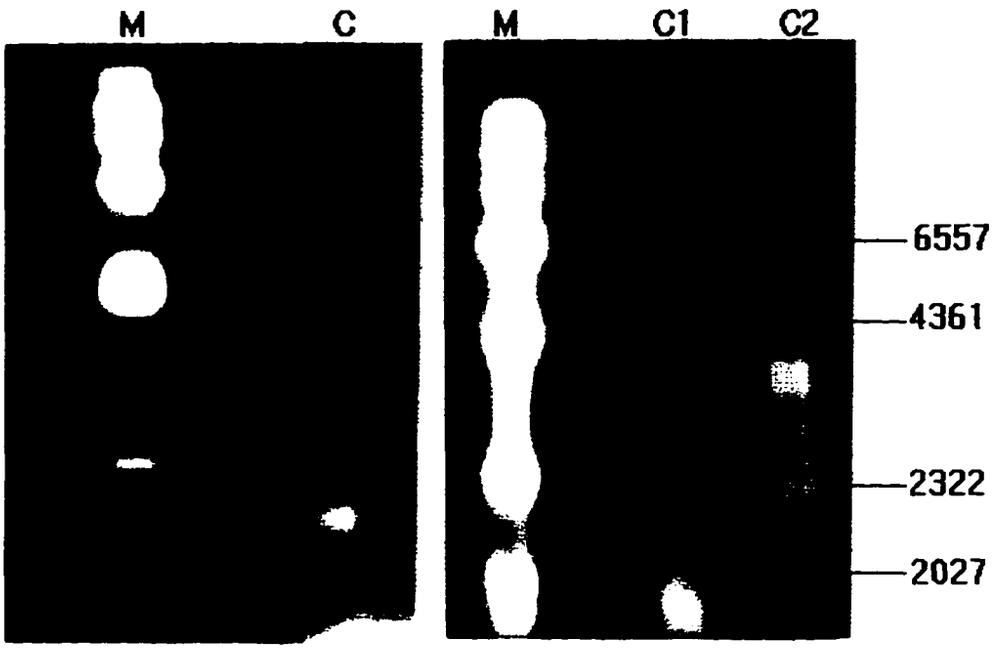


Figure 6. Electrophoresis Profile of Cowpea Mottle Virus cDNAs in 1% Agarose Gels. Lanes M: End labeled lambda DNA HindIII fragments as markers; lanes C and C1: First wash of the spin column; Lane C2: Second wash of the spin column.

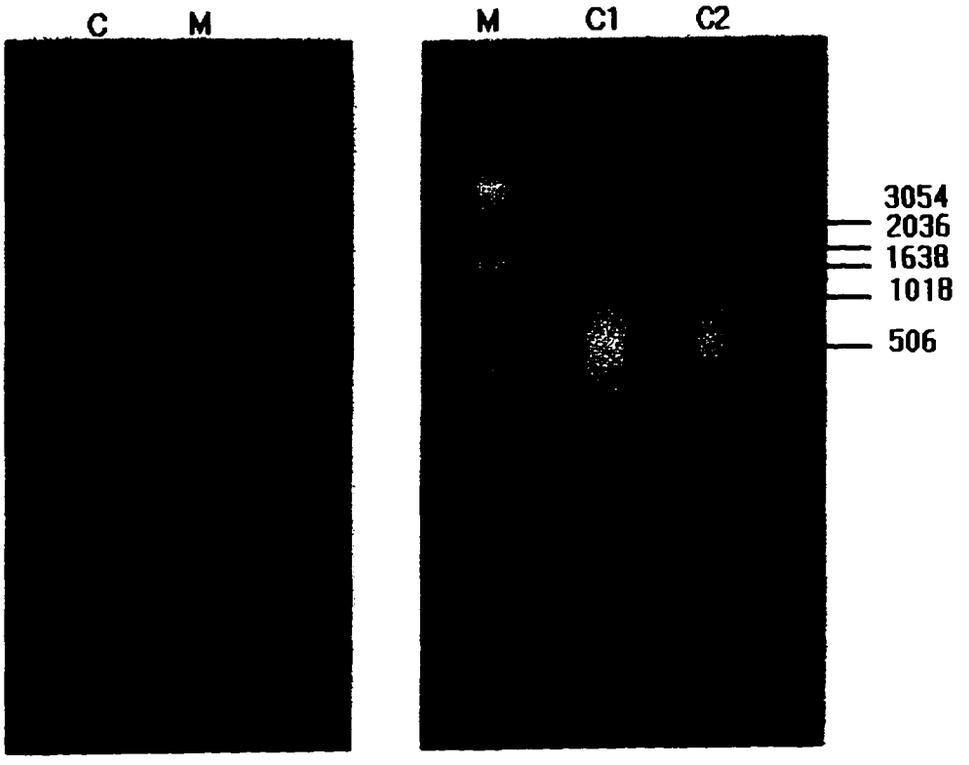
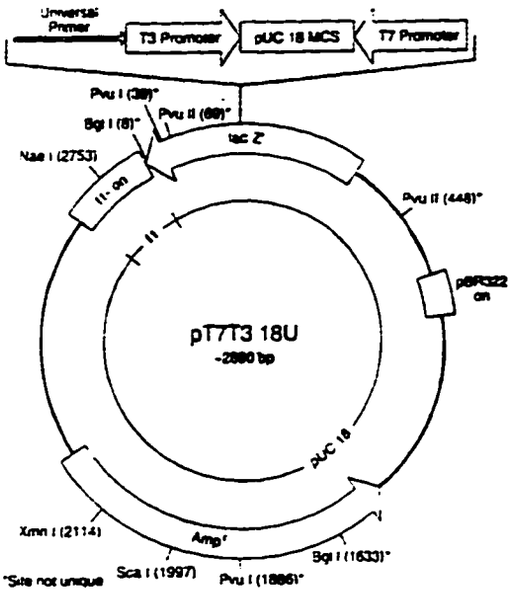


Figure 7. Multiple Cloning Sites of Vector pT7/T3 18U and the Strategy Used to Clone CPMoV cDNAs into the Vector.



CPMoV ssRNA

Specific Primers
AMV Reverse Transcriptase

ss cDNA

DNA Polymerase I
RNase H

ds cDNA

Blunt ended ds cDNA

T4 DNA Polymerase
dNTPs

Add EcoRI Adaptors

T4 DNA Ligase

Remove Unligated Adaptors

Geneclean

Add a phosphate to the ends

T4 Polynucleotide Kinase

ds DNA with EcoRI ends

Ligate cDNA to the Vector

Transform Competent Cells
(E. coli DH5-alpha)

White Colonies on LB/Amp Plate

Figure 8. EcoRI Digestion of pT7/T3 18U Clones Containing CPMoV cDNA. Approximately 100 ng of plasmid DNA prepared from clones containing CPMoV cDNA insert was digested by 6 U of EcoRI in a volume of 20 ul for 2 h, and electrophoresed in 1% agarose gel. The gel was run in 0.5X TBE buffer, stained with ethidium bromide and photographed under UV light. Numbers 1-15 represent different clones and letters u and c represent the uncut and the cut plasmid DNA of the same clone.

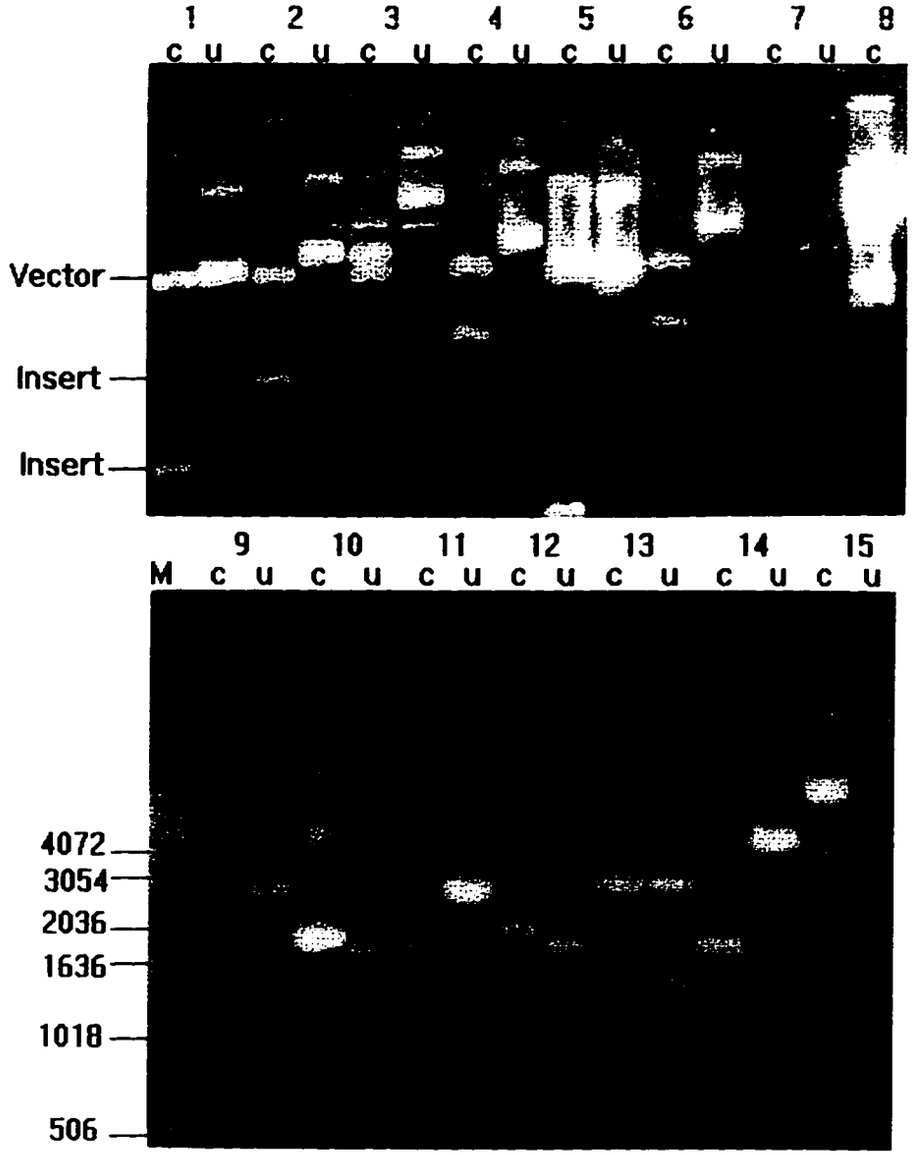
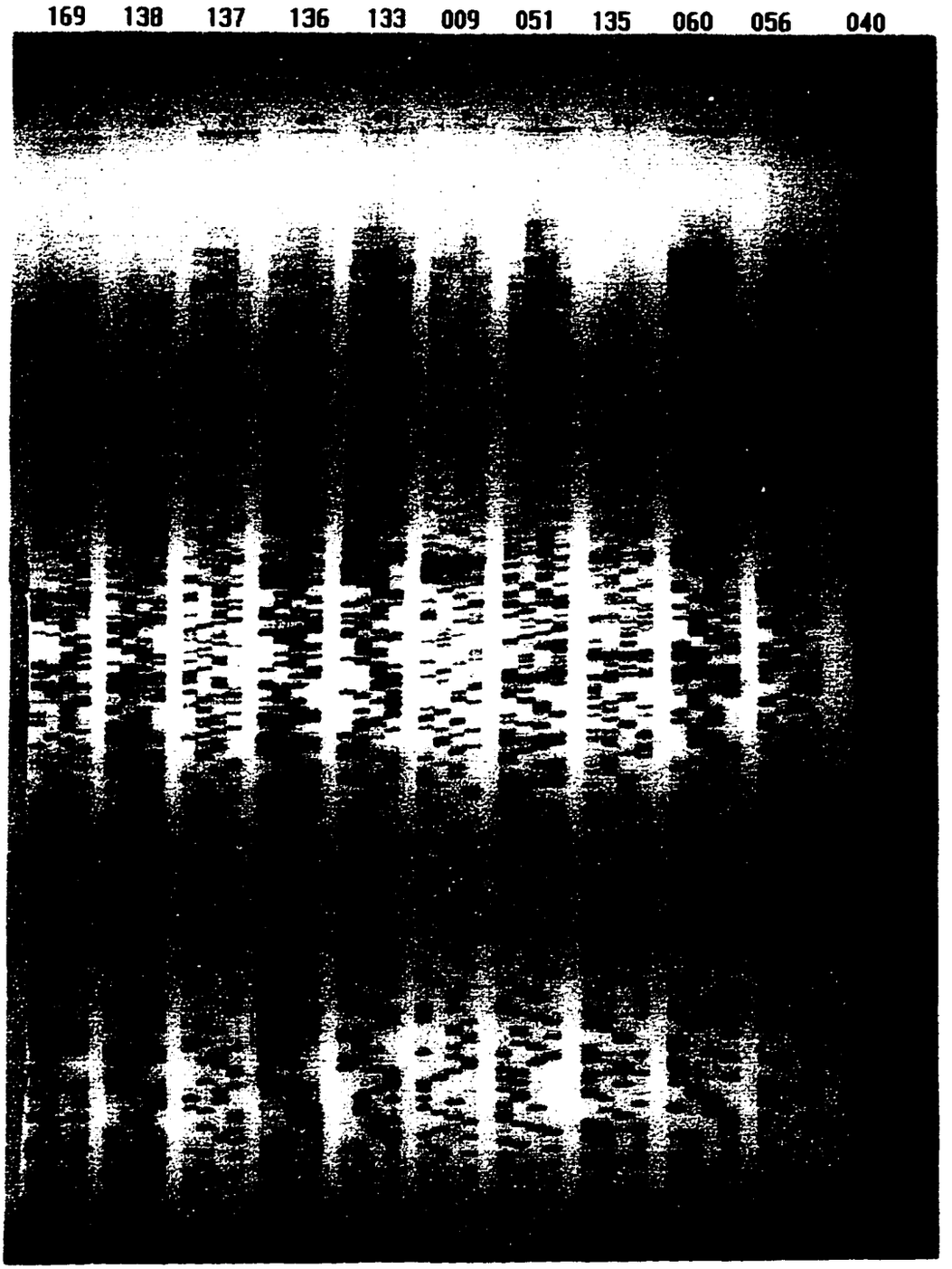


Figure 9. Example Sequencing Gel. Independent cDNA clones of CPMoV were sequenced by Sanger's chain termination method using USB Version 2 DNA sequencing kit. Each sample was loaded in a set of four lanes in the order of G, A, T, C. Sets were separated by a blank lane to make the reading less difficult. Numbers represent specific cDNA clones.



clones. Some of CPMoV cDNA clones were sequenced with the forward primer, some with the reverse primer, and some with both. Several clones were also sequenced using synthetic primers which were complementary to specific areas of CPMoV RNA. The single strand RNA genome of CPMoV consists of 4,029 nucleotides (Fig. 10). Sequence heterogeneity among different clones was found at four positions and may have arisen from variations in the virus population, or from errors introduced by reverse transcriptase in the synthesis of the first strand cDNAs. One of the variations was located in the ORF1. A transition from T to C was found in one of three clones (nt 250). The other variations were located in the ORF2, the putative replicase gene. One transition from T to C was found in one of four clones (nt 874), and another transition from T to C was found in one of three clones (nt 1,816). One transition from A to G was found in two of five clones (nt 1,630), which would result in an amino acid change from N to D. None of the other variations altered the amino acid.

The 5' End Labeling

Approximately equivalent amounts of RNAs were used in the labeling reactions and electrophoresed in an agarose gel. A photograph of the ethidium bromide stained gel shows that the amount of each RNA (TMV, CPMoV and MS2) loaded into the gel was approximately the same (Fig. 11, right). However, the autoradiography of the same gel (Fig. 11, left) shows that the amount of labele incorporated into MS2 RNA,

Figure 10. Nucleotide Sequence of CPMoV Genome. The deduced protein sequences coded by open reading frames 1-6 are shown as p25/83, p56, p7.8, p9.8, p40 and p28 under the ORFs.

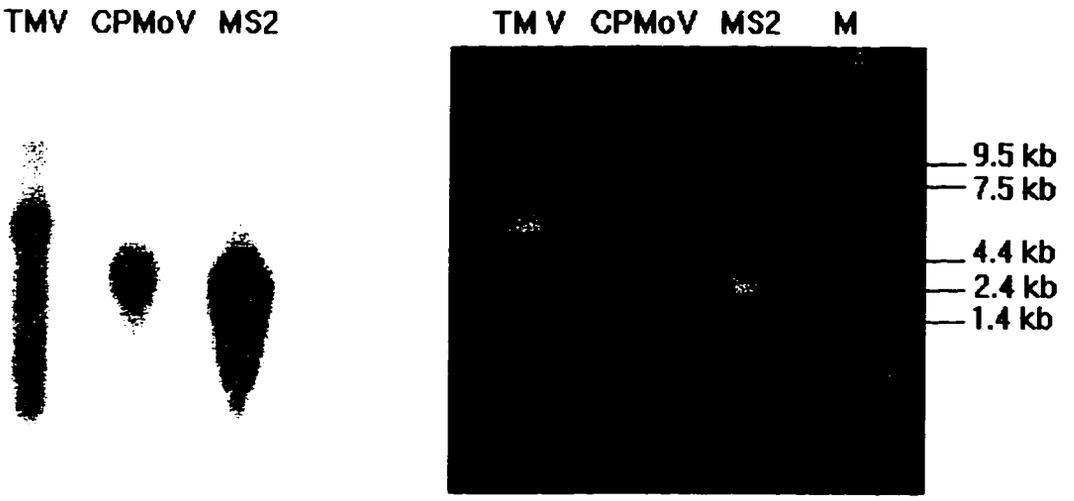
```
   1 NGGTATCGTTAGCCAGATATCCACTTCTCTCAGCATGTTGATCCAATTCGGAACAATGCC
   p25/83                                  M  L  I  Q  F  G  T  M  P

  61 ACCTGTAGAAGTTCAGGGGCCCCCGAGGGTTACCCCTTGGTCGAGCCCTAGCGCCATCGG
       P  V  E  V  Q  G  P  P  R  V  T  P  W  S  S  P  S  A  I  G

 121 CAAATTCCTCGTCGGCGTGTGTGGAGGCCTAGGTGATGCCTGGGCCGACACTGTCACACG
       K  F  L  V  G  V  C  G  G  L  G  D  A  W  A  D  T  V  T  R

 181 GAACCGGACTACGCACATCAACGCTTTCGAGCGGTTTGGTGAATTCCCAATAGAATGGGC
       N  R  T  T  H  I  N  A  F  E  R  F  G  E  F  P  I  E  W  A

 241 AATCAATAGTACCGGGGATGAGGTTTTTGAACCGGAAAGACCCACCAGAGAGGTGGGCAA
       I  N  S  T  G  D  E  V  F  E  P  E  R  P  T  R  E  V  G  K

 301 GGGCAAGCGGGTCATCAATAGGCACGCTAAAGGGTACTTCCTCAAAGAGGCCGTCACTGC
       G  K  R  V  I  N  R  H  A  K  G  Y  F  L  K  E  A  V  T  A

 361 ATGTCGGCTTCAGTTCAACGGAGTGCCCATGTCCACGAACGCCAATCTGCTGGGTGTTCG
       C  R  L  Q  F  N  G  V  P  M  S  T  N  A  N  L  L  G  V  R

 421 GAAGTTCCTTGCAGGTTTTTGCCAGGAACATGGGTTGACCAACCAACAATCTCTGAGGTT
       K  F  L  A  G  F  C  Q  E  H  G  L  T  N  Q  Q  S  L  R  L

 481 GGTCAATGCTGCTACACCCCTCGTGATGACGGCCAACCAAGAGGACGTCGAGTATATGAA
       V  N  A  A  T  P  L  V  M  T  A  N  Q  E  D  V  E  Y  M  K

 541 GCTGTTCAACTGCGTGCAGCTTAATGAGAAGAGGGCTGAGTATCTTGATGCTCAGGTAGT
       L  F  N  C  V  Q  L  N  E  K  R  A  E  Y  L  D  A  Q  V  V

 601 GCACACCCCGATTCGACAACTGGTCGCTCGCCCCTTGTCCGGCGAGAAGTGGTCACAGTG
       H  T  P  I  R  Q  L  V  A  R  P  L  S  G  E  K  W  S  Q  W

 661 GATAGCTCATCTGGGTGGGTTCAGACCTGCCACGGGCTTCCGCGTGTCCAACTAGGGGTG
       I  A  H  L  G  G  F  R  P  A  T  G  F  R  V  S  N  *  G  C

 721 CCTAAGGCGTGTGCCAGGGTTTAATACTAGTATAGTTCGTGGCGAACACCCTGACATGCT
   p56 L  R  R  V  P  G  F  N  T  S  I  V  R  G  E  H  P  D  M  L

 781 GATCACGCCTAGGACAGGAAGCCACAAAGTCCGCAAAATGGGGCATGTAAATTTCATGGG
       I  T  P  R  T  G  S  H  K  V  R  K  M  G  H  V  N  F  M  G

 841 GAACGGCAAGCTCTTCGGGGTCCACAATGCAAGTTTAGTCAACCTTCGAAGAGGGTTGCT
       N  G  K  L  F  G  V  H  N  A  S  L  V  N  L  R  R  G  L  L

 901 TGAGCGGGTGTTCTATGTGGAGCGCGAGGGCAAGTTAGTGCCCCCCCCTGAACCCATTCC
       E  R  V  F  Y  V  E  R  E  G  K  L  V  P  P  P  E  P  I  P

 961 GGGCATATTTAGTAGGTTAGCGGGGTTCAAAAGCAAACTGAAGAAGATTGTAGGGACTCA
       G  I  F  S  R  L  A  G  F  K  S  K  L  K  K  I  V  G  T  H

1021 TTCCAGGATCAGCGACGACGCTTTCGTTGCTTTGTATCATGGCAGGAGGAACACAATCTA
       S  R  I  S  D  D  A  F  V  A  L  Y  H  G  R  R  N  T  I  Y
```

```
1081 TCAGAATGCCGTGGACTCCTTACGTATAGGTGGAGTGCAGAGGAAAGATTCATATTTGAG
       Q  N  A  V  D  S  L  R  I  G  G  V  Q  R  K  D  S  Y  L  S

1141 CACTTTCATCAAAGCTGAGAAAATACCCATCTACCGCAAGCCTGACCCTGCTCCCAGGGT
       T  F  I  K  A  E  K  I  P  I  Y  R  K  P  D  P  A  P  R  V

1201 GATACAACCGCGTAATCCAAGGTACAATGTGGAGGTGGGGAGGTATTTACGGCCCTTTGA
       I  Q  P  R  N  P  R  Y  N  V  E  V  G  R  Y  L  R  P  F  E

1261 GCATCATCTTTACCGAGGCATAGATGAGATTATGGGGGGACCGACTGTGATTAAAGGCTA
       H  H  L  Y  R  G  I  D  E  I  M  G  G  P  T  V  I  K  G  Y

1321 CGACGTGAACCAGATAGGAAATATCATGGAGAAAGCGTCGGGACAATTCGTCCGACCCGT
       D  V  N  Q  I  G  N  I  M  E  K  A  S  G  Q  F  V  R  P  V

1381 AGCCGTGGGCTTTGATATGTCACGATTTGACCAACATGTGTCCTTTGACGCCCTCCAGTT
       A  V  G  F  D  M  S  R  F  D  Q  H  V  S  F  D  A  L  Q  F

1441 TGAACATGGAGTATATTTACAACATTTTGGAGGCGACCGGTTCCTGGCCAAGTTATTGTC
       E  H  G  V  Y  L  Q  H  F  G  G  D  R  F  L  A  K  L  L  S

1501 ATGGCAATTGAACAACAAAGGGTTTGCCAACTTGCCAGAAGGGAAAGTCAAATACACCCG
       W  Q  L  N  N  K  G  F  A  N  L  P  E  G  K  V  K  Y  T  R

1561 CCGAGGGTGCAGGATGAGTGGGGACATGAACACAGCTATGGGCAATTGCCTGCTGGCTTG
       R  G  C  R  M  S  G  D  M  N  T  A  M  G  N  C  L  L  A  C

1621 TGCTATTACGTGGGACCTAATGAAGGGGATCAAGTACCGCTTGCTTAACAACGGAGATGA
       A  I  T  W  D  L  M  K  G  I  K  Y  R  L  L  N  N  G  D  D

1681 CTGTGTGGTTATTACCGAGTCCAAGAACGTTGATTTCGTATGCAAGCAATTGGAACGATT
       C  V  V  I  T  E  S  K  N  V  D  F  V  C  K  Q  L  E  R  F

1741 CCGCGACTATGGGTTTACGTGCATAGCTGAGGAACCAGTGTATGAGTTGGAAAAGCTGGA
       R  D  Y  G  F  T  C  I  A  E  E  P  V  Y  E  L  E  K  L  E

1801 GTTCTGTAATCTACATCCACTGTATGATGGCTCTTCCTGGACGGTGATGCGGAAACCAAG
       F  C  N  L  H  P  L  Y  D  G  S  S  W  T  V  M  R  K  P  S

1861 CGTTTCATTGGCTAAGGACACCTATTGTGTGGCAGGATGGAACAACACCAAGGATGCAGC
       V  S  L  A  K  D  T  Y  C  V  A  G  W  N  N  T  K  D  A  A

1921 CAGTTGGCTCAATGCCATTGGTCAATGTGGTGCAGCCATCACAGGCGGTATACCCATACA
       S  W  L  N  A  I  G  Q  C  G  A  A  I  T  G  G  I  P  I  Q

1981 GCAAGCGTATTATCAGTGTTTAATCCGCAACTTCCCCCGGGGTGGATGGTTACAAAAACA
       Q  A  Y  Y  Q  C  L  I  R  N  F  P  R  G  G  W  L  Q  K  H

2041 CCACCACACCATGGATAGCGGGTTGTATTGGCTTGCTCTCAAAAAGGGACGAAGCGATCC
       H  H  T  M  D  S  G  L  Y  W  L  A  L  K  K  G  R  S  D  P

2101 GGTTCCAGTGGCTCCTGAGGCACGCCATAGTTTTTACTTAACTACTGGCGTCACTCCGGA
       V  P  V  A  P  E  A  R  H  S  F  Y  L  T  T  G  V  T  P  D
```

```
2161 TGCCCAAATAGCGGTGGAAGAGTACTACAATAACCTCCACCTTGCGCTCGAGTTTGAACC
       A  Q  I  A  V  E  E  Y  Y  N  N  L  H  L  A  L  E  F  E  P

2221 GGAGGTATCGCCAACCTCACTCTCGACCACCTCAGCGGACTGATTTAACGTTTGCCAAGC
       E  V  S  P  T  S  L  S  T  T  S  A  D  *

2281 TCATTTATGGATCAAACAACGGAAGTTCGGGGTAGAAGCAGAAGCCGGCAGCGATCAGGG
    p7.8     M  D  Q  T  T  E  V  R  G  R  S  R  S  R  Q  R  S  G

2341 TCCAGCGACTCTCGGAAGCCAAAGAACAATAAGCAGGTGCAGGTAGCCTCGGAGGCCGTT
       S  S  D  S  R  K  P  K  N  N  K  Q  V  Q  V  A  S  E  A  V

2401 CAGCGTCGAAATGATAAACCACTTCGCGGCCATCACGGGGGTGATTTTGTTATTGTGGCT
       Q  R  R  N  D  K  P  L  R  G  H  H  G  G  D  F  V  I  V  A
    p9.8        M  I  N  H  F  A  A  I  T  G  V  I  L  L  L  W  L

2461 CATACGGTCACACTCAATATTAACTTCAATATATGAGCAGAACACCCCCAGGGTAATTAC
       H  T  V  T  L  N  I  N  F  N  I  *
        I  R  S  H  S  I  L  T  S  I  Y  E  Q  N  T  P  R  V  I  T

2521 CTACGCCAATCTCTTGCTTTTGTTGATCGCATATTGCTATTTCTGCTCAACCCAAAGTGT
       Y  A  N  L  L  L  L  L  I  A  Y  C  Y  F  C  S  T  Q  S  V

2581 TAGCTACGTTGTCTCTCACTCAGAACCTGTAGTAAATAGTAAAGTTGTTTACATTTCTGT
       S  Y  V  V  S  H  S  E  P  V  V  N  S  K  V  V  Y  I  S  V

2641 TGGAACTTCTCCTCCCTCTCTTCAAGTGTCGGAATGAATGGAAACGCACTCACCATCACC
    p40                                  M  N  G  N  A  L  T  I  T
    p28 G  T  S  P  P  S  L  Q  V  S  E  *  M  E  T  H  S  P  S  P

2701 CAGCTCGCTAAGAAAGCCGCCACACAGGGGCGTGCCTCACTATCGAAGAATCAAAAGAAG
       Q  L  A  K  K  A  A  T  Q  G  R  A  S  L  S  K  N  Q  K  K
        S  S  L  R  K  P  P  H  R  G  V  P  H  Y  R  R  I  K  R  S

2761 CGGCTGGACAATGCAAACACAAACCCCATCAAGTCCGTCACACGAACCGCACGGATCTCC
       R  L  D  N  A  N  T  N  P  I  K  S  V  T  R  T  A  R  I  S
        G  W  T  M  Q  T  Q  T  P  S  S  P  S  H  E  P  H  G  S  P

2821 ACAAGGAAGGTCCCTGCAGCAACCAGCACCACCATCACAAACGGGGTGCCTAACTATGCG
       T  R  K  V  P  A  A  T  S  T  T  I  T  N  G  V  P  N  Y  A
        Q  G  R  S  L  Q  Q  P  A  P  P  S  Q  T  G  C  L  T  M  R

2881 GTTCGTCAGAACAAGCCTGTTATCCAGCATGTTGAGTTATGGGGTACTCTTATGTCAAAT
       V  R  Q  N  K  P  V  I  Q  H  V  E  L  W  G  T  L  M  S  N
        F  V  R  T  S  L  L  S  S  M  L  S  Y  G  V  L  L  C  Q  I

2941 ACTACTGAGTCCCCTGCTTATATCACCAGGACACTCAACCCAAGTGACCCAGCAACTTTC
       T  T  E  S  P  A  Y  I  T  R  T  L  N  P  S  D  P  A  T  F
        L  L  S  P  L  L  I  S  P  G  H  S  T  Q  V  T  Q  Q  L  S

3001 AACTGGGTCCAACCCCTCTCTACAGGGTATGATATGTACCGACTCGTTCGTTGTGAGATC
       N  W  V  Q  P  L  S  T  G  Y  D  M  Y  R  L  V  R  C  E  I
        T  G  S  N  P  S  L  Q  G  M  I  C  T  D  S  F  V  V  R  S
```

```
3061 ATATACACACCTCGGTGTGCCGTTACGACGACGGGATCAGTGGTGCTGGCTTATGATCCG
      I  Y  T  P  R  C  A  V  T  T  T  G  S  V  V  L  A  Y  D  P
      Y  T  H  L  G  V  P  L  R  R  R  D  Q  W  C  W  L  M  I  R

3121 GACGCTAGTGATGTCAATCCTGATAATGTCACCGATCTTCTCAATATGGCAGGTGCTGAA
      D  A  S  D  V  N  P  D  N  V  T  D  L  L  N  M  A  G  A  E
      T  L  V  M  S  I  L  I  M  S  P  I  F  S  I  W  Q  V  L  N

3181 TCTGGTTCTGCCTATAGCCCTCTATCGCTTGTCCCCAATATCAAGCAGATTGATCGCTAC
      S  G  S  A  Y  S  P  L  S  L  V  P  N  I  K  Q  I  D  R  Y
      L  V  L  P  I  A  L  Y  R  L  S  P  I  S  S  R  L  I  A  T

3241 ATTCGTGACAACAGCACTAGCGACCCCAAGCTTGTGGATGCTGGGAAGATACTTGTTGCT
      I  R  D  N  S  T  S  D  P  K  L  V  D  A  G  K  I  L  V  A
      F  V  T  T  A  L  S  T  P  S  L  W  M  L  G  R  Y  L  L  L

3301 AGCTATGGCCAGCAAACGAGCACTGCACCATTTGCACTCGGGGAAATCCGGTTCGCCTAC
      S  Y  G  Q  Q  T  S  T  A  P  F  A  L  G  E  I  R  F  A  Y
      A  M  A  S  K  R  A  L  H  H  L  H  S  G  K  S  G  S  P  T

3361 ACTCTCCAACTCATTGTCCCGCAACCTCATTCCACTGGGGTGCAGCGGTTGGGGAGTGCC
      T  L  Q  L  I  V  P  Q  P  H  S  T  G  V  Q  R  L  G  S  A
      L  S  N  S  L  S  R  N  L  I  P  L  G  C  S  G  W  G  V  P

3421 CCCCCAGTCGGTAGGGCCTGCATTCGCACGTTAACCAGCGTGGTCAACACTACTTCCAGG
      P  P  V  G  R  A  C  I  R  T  L  T  S  V  V  N  T  T  S  R
      P  Q  S  V  G  T  A  F  A  R  *

3481 TTAGCTGACTACACGCTAACACTGAAGACAGCTGGTTCGTACATGATTTCTGCGGTTGAT
      L  A  D  Y  T  L  T  L  K  T  A  G  S  Y  M  I  S  A  V  D

3541 GCTTCCTCGCATACAAAGCATAAAGCAAACCAACTTTGGAAGGTTTCGAGTGGAAGGAGT
      A  S  S  H  T  K  H  K  A  N  Q  L  W  K  V  S  S  G  R  S

3601 TACTGGAGCCGGGCAACCAATGTTGATATTCGCTCGGCCCTCAAATCCGGGGCCACCATC
      Y  W  S  R  A  T  N  V  D  I  R  S  A  L  K  S  G  A  T  I

3661 ACCTTCGAGGCTGCTACTGATAGTAACGCGGAGGTGTATGTCGTGCGTGACAGTGGCATA
      T  F  E  A  A  T  D  S  N  A  E  V  Y  V  V  R  D  S  G  I

3721 AACGACACTATTGTCGTCCCTAGCCTGCATGTGCCGGTAAACCCAAAATCTCTGTGATAA
      N  D  T  I  V  V  P  S  L  H  V  P  V  N  P  K  S  L  *

3781 ACTTCTGGCCCTCCATGGGGTGCGCGTCCGCAACTTTGCGCAGTACCACGGACATCCAGA

3841 AGTGCCTTGCGGGGCGCGTCACTGAGGCCAAAAATAGAGTGTAGCAACTCCCTCTGTTGG

3901 GGAGCCAAGTAGGACCTACTAACAGGCTGTCTGAGCCTACCTAGCGTACTGGGACAGAAA

3961 TCCTTCGGGTGGGGTGTCCTGGAACATAATCAGATAATGATAGGGGTGGTAGCATGCCAC

4021 CCCCCCCCC
```

Figure 11. Electrophoresis Profile of the 5' End Labeling of CPMoV RNA. RNAs were end-labeled with polynucleotide kinase and gamma ^{32}P-ATP, electrophoresed in 1% agarose gel, dried and autoradiographed. Right panel: Ethidium bromide stained agarose gel. Left panel: Autoradiograph of the same gel. Lane M: Marker RNAs (0.24-9.5 kb, capped); Other lanes from left to right are tobacco mosaic virus (TMV) RNA (capped), cowpea mottle virus (CPMoV) RNA, and MS2 viral RNA (uncapped).

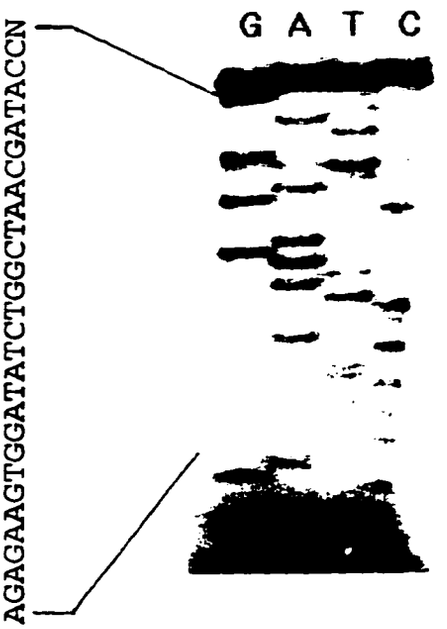


which was not capped at its 5′ end, was much higher then those of either TMV or CPMoV RNA. To confirm the higher level of labeling of MS2 RNA indicated by the autoradiograph, the RNA bands were excised and counted by liquid scintillation. The excised RNA band of MS2 had approximately ten fold more counts than the RNA of either CPMoV or TMV. The marker RNAs (all capped) were not labeled even using ten fold more gamma ^{32}P ATP. This experiment was repeated six times with different amounts of gamma ^{32}P (1 uCi, 2 uCi, 5 uCi, 10 uCi and 20 uCi) with essentially the same results. In all experiments, capped marker RNAs remained unlabeled. This suggested that the 5′ end of CPMoV RNA was blocked, most likely capped, similar to the RNA of TMV and CarMV (Guilley et al., 1985). The very low level of label present in both TMV and CPMoV indicated that a small portion of TMV and CPMoV RNA molecules were uncapped, either because of the heterogeneity of the genomic RNA population or because a small portion of RNA molecules lost their caps during manipulation. However, the 5′ end labeling data alone is not conclusive. Further experimental data is needed to confirm the presence of 5′ end cap structure.

5′ End RNA Sequencing

Direct sequencing of viral RNA identified 18 nucleotides at the 5′ end in addition to the sequences overlapping with the 5′ end cDNA clones (Fig. 12). Addition of $MnCl_2$ in the RNA sequencing reactions resulted in clearer bands than 10 fold dilution of the labeling mixture, when

Figure 12. 5′ End Sequencing of Cowpea Mottle Viral RNA. The 5′ sequences of CPMoV was determined by direct sequencing of CPMoV viral RNA using AMV reverse transcriptase, which extended an additional 18 nucleotides beyond the most 5′ end cDNA clone previously sequenced. The sequences shown are complementary to viral RNA.

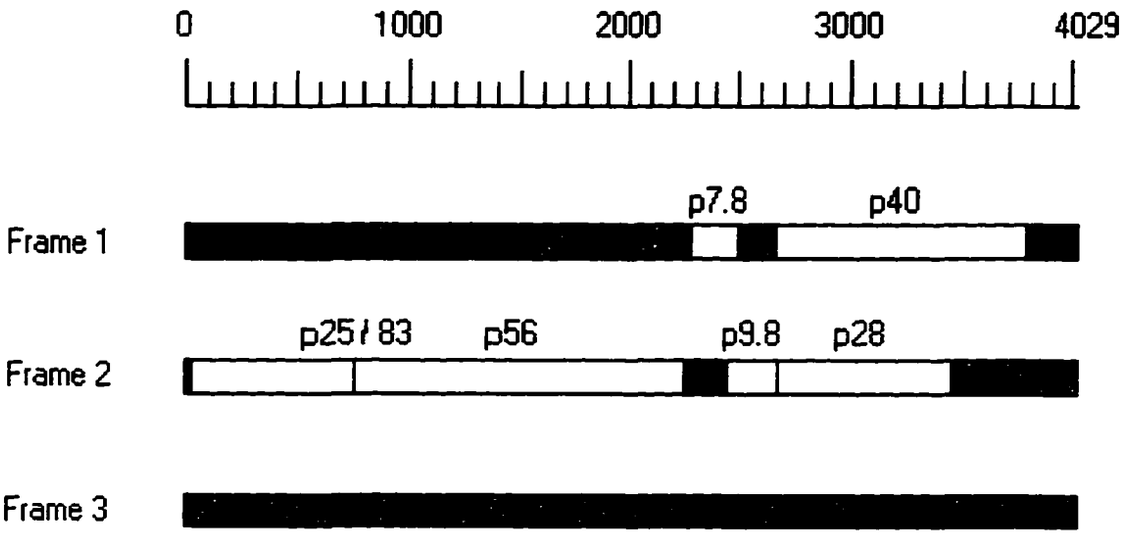


these two methods were used to read nucleotides close to the sequencing primer. Addition of rRNAsin (Promega) after annealing was required in order to control the RNase activity and maintain the integrity of the single strand template RNA. The optimum primer to template ratio (1:2) was also critical.

ORFs in CPMoV

There were six major open reading frames (ORFs) in the CPMoV genome (Fig. 13). ORF1 (nt 35-715) encodes a 25 kD polypeptide of unknown function. ORF2 (nt 776-2,263) encodes a 56 kD protein which is the putative RNA replicase or a part of it. Like other carmoviruses, suppression of ORF1 amber termination codon at 713 would result in a read-through product of 83 kD (Morris and Carrington, 1988). ORF3 and ORF4 are located between the putative replicase and the capsid protein gene, have sequence overlaps with the capsid protein gene, and encode two small proteins (7.8 kD and 9.8 kD). These appear similar to the small proteins of turnip crinkle virus, which were proposed to have movement function (Hacker et al., 1992). ORF5 is located near the 3' end of the RNA (nts 2,674-3,777), and encodes a 40 kD polypeptide which reacts with capsid protein antiserum (Kim and Bozarth, 1992). ORF6 is located within ORF5 but in a different reading frame. This ORF was not present in other carmoviruses.

Figure 13. Open Reading Frames of Cowpea Mottle Virus Genome. ORF1 (35-715) encodes a polypeptide of 25 kD with an unknown function. ORF2 (776-2263) encodes a 56 kD protein which was the putative RNA replicase or part of it. Read-through from ORF1 into ORF2 (35-2263) would produce an 83 kD protin. ORF3 (2287-2496) and ORF4 (2411-2677) encode two small proteins of 7.8 and 9.8 kD with proposed movement function. ORF5 (2674-3777) encodes the 40 kD capsid protein. ORF6 (2680-3454) was within ORF5 but in a different frame.



Sequence Analysis and Comparison

The nucleotide and the derived polypeptide sequences ofCPMoV were compared to the sequences of carmoviruses, tombusviruses (Table 1) and other viruses. The results show extensive sequence similarity between CPMoV and members of carmoviruses; turnip crinkle virus (TCV) (Carrington et al., 1989), cardamine chlorotic fleck virus (Skotnicki et al., 1993), melon necrotic spot virus (MNSV) (Riviere and Rochon, 1990), carnation mottle virus (CarMV) (Guilley et al., 1985), and members of tombusviruses; cucumber necrosis virus (CNV) (Rochon and Tremaine, 1989) and tomato bushy stunt virus (TBSV) (Hearne et al., 1990). The most remarkable homology was within the ORF2 coding regions.

cDNA Synthesis and PCR Amplification of Replicase Gene

Replicase cDNAs were synthesized from viral RNA using a Promega cDNA synthesis kit and a primer complementary to the sequences adjacent to the 3' end of the replicase gene. The two-nucleotide mismatches at the central region of the primer had no obvious effect on the yield of the first strand cDNA. This first strand cDNA was used as a template for PCR amplification. When the PCR reactions were carried out as described in the methods with a $MgCl_2$ concentration of 1.5 mM, the amplified products appeared in 1% agarose gel as a major band of 1.5 kb, which corresponds to the replicase gene, and several minor bands of lower molecular weights (Fig. 14).

Table 1. Percentage Sequence Homology of Cowpea Mottle Virus Versus Other Carmoviruses and Tombusviruses. (NTs: Nucleotide sequences; AAs: Amino acid sequences. CCFV: Cardamine chlorotic fleck carmovirus; TCV: Turnip crinkle carmovirus; MNSV: Melon necrotic spot carmovirus; CarMV: Carnation mottle carmovirus; CNV: Cucumber necrosis tombusvirus; TBSV: Tomato bushy stunt tombusvirus).

	Percentage sequence homology						
CPMoV	Genomic RNA	RNA replicase				Capsid protein	
		(ORF1)		(ORF2)			
Vs	NTs*	NTs	AAs*	NTs	AAs	NTs	AAs
CCFV	49.7	41.4	22.6	56.3	53.3	44.9	28.8
TCV	49.7	48.3	29.1	54.1	53.2	43.9	24.5
MNSV	49.6	40.2	28.9	49.6	46.6	40.7	29.7
CarMV	49.0	42.9	25.4	49.9	42.0	41.2	24.1
CNV	43.8	39.4	18.2	50.9	42.7	43.0	23.4
TBSV	43.8	31.5	15.0	49.8	41.9	42.2	23.5

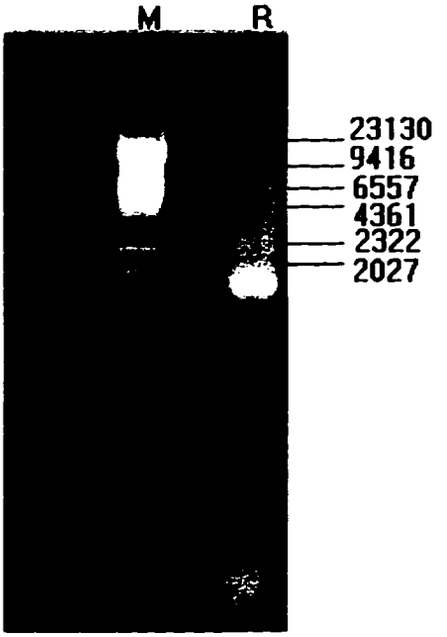


Figure 14. Electrophoresis Profile of PCR Amplified CPMoV RNA Replicase Gene in 1% Agarose Gel. Lane M: Marker (Lambda DNA HindIII fragments); Lane R: PCR amplified CPMoV RNA replicase Gene.

Cloning of Replicase Gene into TA Cloning Vector and Digested with EcoRI

It turned out that PCR amplified replicase DNA was extremely difficult to digest with restriction enzymes, evenwhen reactions were carried out at a very high ratio of restriction enzyme to DNA (5 ug of DNA was digested with 100 units of restriction enzymes in a volume of 100 ul, incubated at 37° C for 16 h would not cleave the DNA). To solve this problem, a TA cloning vector (InvitroGen) was used to convert the external restriction sites into internal sites (Fig. 15). PCR amplified CPMoV RNA replicase fragments were first cloned into the TA vector, then a TA clone containing the 1.5 kb DNA insert was digested with EcoRI. A 20 ul reaction containing 2 ug of plasmid DNA, 2 ul of H buffer and 1 ul of EcoRI (12 U/ul) was incubated at 37° C for 2 h. The complete cleavage resulted in a TA vector fragment of 3.9 kb and an insert fragment of 1.5 kb (Fig. 16).

A large scale preparation of TA vector containing 1.5 kb CPMoV DNA insert was digested with EcoRI in a 100 ul reaction and subjected to 1% agarose gel electrophoresis. The 1.5 kb DNA band was then excised from the gel (Fig. 17) and eluted by Geneclean.

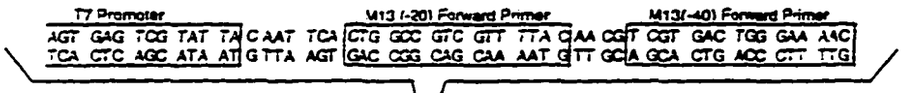
Cloning of Replicase Gene into Expression Vector pGEX-2T

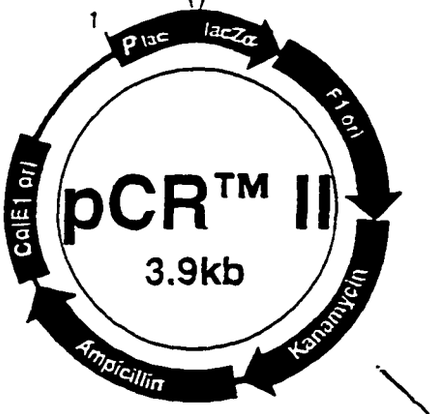
The replicase gene from the TA vector was ligated into the ends of the dephosphorylated expression vector pGEX-2T at the EcoRI site (Fig. 18). Dephosphorylated and

Figure 15. Multiple Cloning Sites of the 3.9 kb TA Cloning Vector and the Strategy Used to Clone PCR Amplified CPMoV RNA Replicase Gene into the TA Vector.

CPMoV ssRNA

Replicase 3′ End Primer
AMV RTase





ss DNA

Replicase Primers
Taq DNA Polymerase
dNTPs, $MgCl_2$

PCR Amplified Replicase DNA

Ligate to TA Vector

Transform Competent Cells
(E. coli INV-alphaF′)

White Colonies on LB/Amp Plate

Figure 16. EcoRI Digestion of TA Vector Containing a CPMoV RNA Replicase Gene. The digestion mixture was electrophoresed in 1% agarose gel. Approximately 1 ug of TA clone DNA, and 4 U of EcoRI in a volume of 20 ul were used for digestion. Lane M: Lambda DNA HindIII fragments. Lane R. Linearized TA vector (3.9 kb) and CPMoV RNA replicase fragment (1.5 kb).

Figure 17. EcoRI Digestion of a Large-Scale Preparation of TA Vector Containing a CPMoV RNA Replicase gene. The digestion mixture was electrophoresed in 1% agarose gel. Lane M: Lambda DNA HindIII fragments. Lane TA-Rep: EcoRI digested TA clones with CPMoV RNA replicase gene inserted.

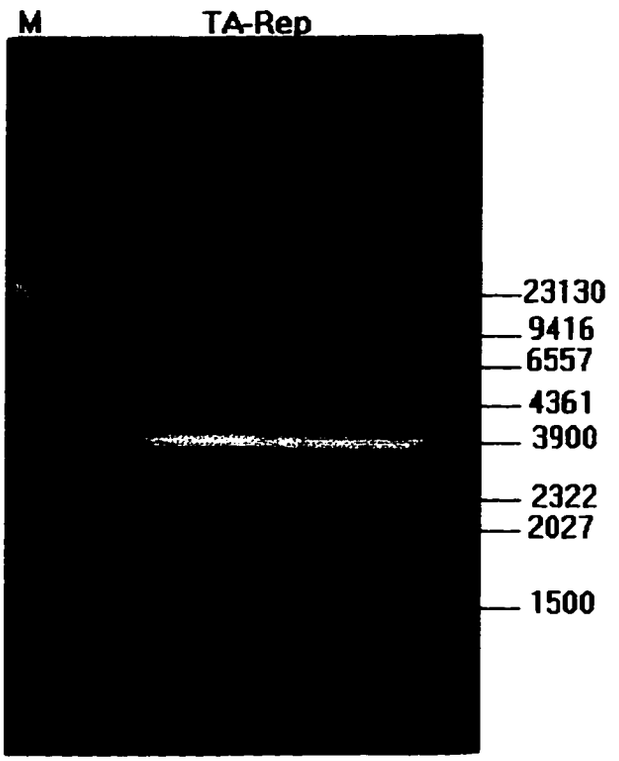
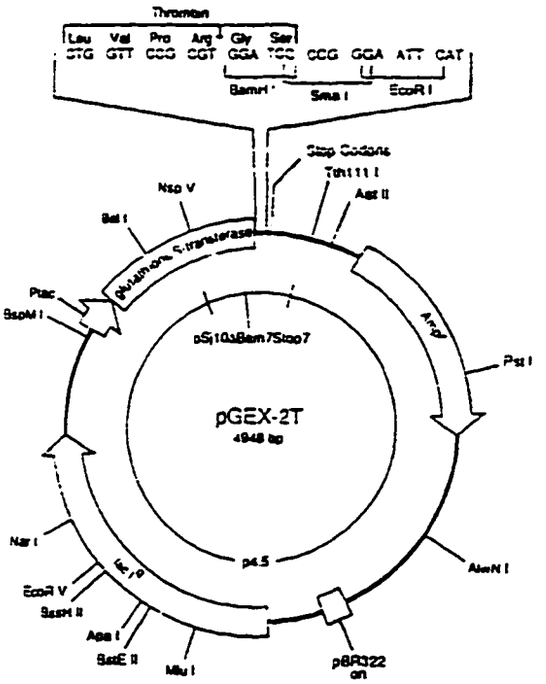


Figure 18. Multiple Cloning Sites of the 5 kb Fusion Vector pGEX-2T and the Strategy Used to Clone CPMoV RNA Replicase Gene into PGEX-2T.



TA Clone

EcoRI Digestion

Geneclean

Ends Dephosphorylation (Calf Intestinal Alkaline Phosphatase)

Replicase Fragments

Ligation to Vector (T4 DNA Ligase)

Transformation (E. coli JM-109)

Colonies on LB/Amp Plate

Miniprep for Screening

EcoRI Digestion

BamHI Digestion

Two Fragments (Wrong Orientation)

One Fragment (Right Orientation)

undephosphorylated vector DNAs were electrophoresed in an agarose gel for quantitation (Fig. 19). Then 50 ng of each was self-ligated using T4 DNA ligase in a 10 ul volume, and 1 ul of this ligation mixture was used to transform E. coli DH5-alpha competent cells. The number of colonies obtained from cells transformed with dephosphorylated plasmid was significantly decreased (Fig. 20).

The replicase gene could be inserted into pGEX-2T in either of two orientations. Only one orientation would result in a fusion which was in-frame with the GST portion of the vector. In one orientation (the correct orientation), the recombinant plasmid would release two fragments by BamHI digestion. A fragment of 6.5 kb, and a very small fragment (about 35 nucleotides) which would be too small to be seen in the agarose gel. The opposite orientation would result in two BamHI sites 1.5 kb apart, thus BamHI digestion would result a 5 kb fragment (pGEX-2T vector) and a 1.5 kb fragment (replicase DNA) (Fig. 21).

A large scale preparation of a pGEX-2T DNAs with a CPMoV RNA replicase fragment inserted in the correct orientation was digested with BamHI in a 100 ul volume. The digestion mixture was electrophoresed in agarose gel to remove the small BamHI fragment, which otherwise would interrupt the in-frame fusion of the gene to the expression vector. A 6.5 kb recombinant plasmid DNA band was then excised from the gel (Fig. 22) and eluted by Geneclean. The linear recombinant plasmid DNA was self-ligated and used to

Figure 19. Determination of the Quantities of the Dephosphorylated and Undephosphorylated pGEX-2T for Transformation. Samples were electrophoresed in 1% agarose gel. Lane M: Lambda DNA HindIII fragments as markers; Lane U: 50 ng undephosphorylated vector DNA; Lane dp: 50 ng vector DNA after dephosphorylation and precipitation procedures.

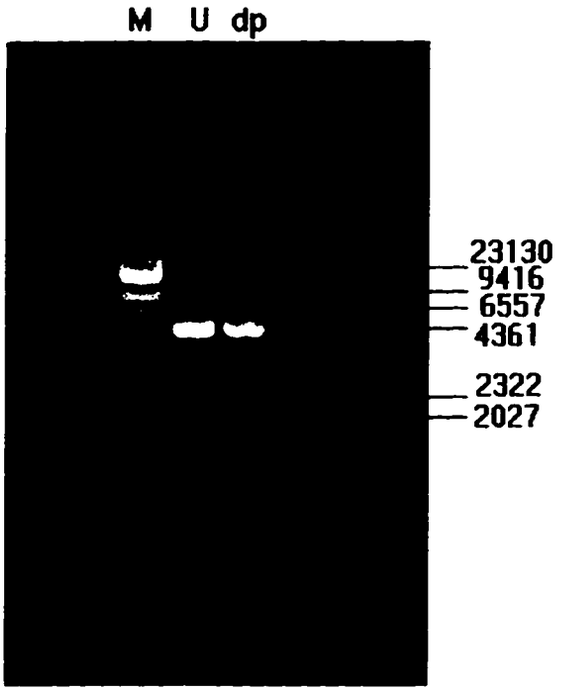
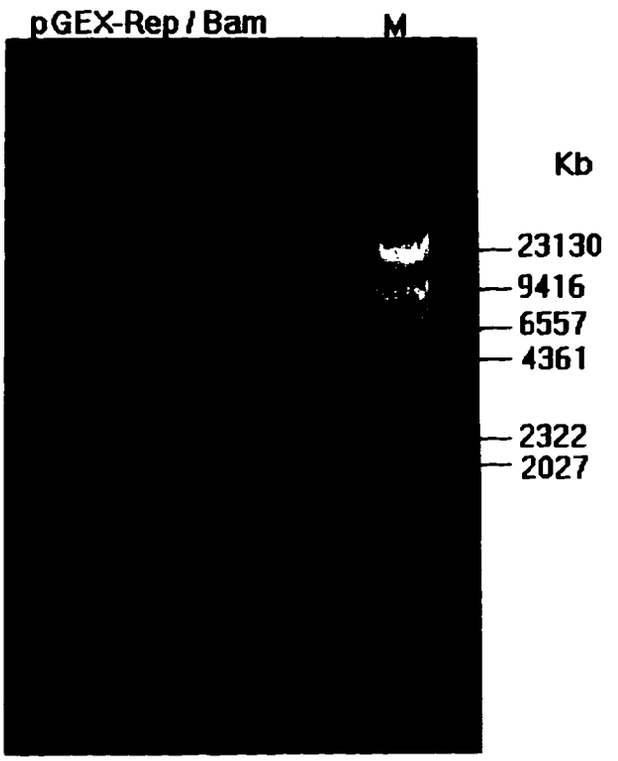


Figure 20. Transformation of E. coli Competent Cells with Dephosphorylated and Undephosphorylated pGEX-2T Self-Ligation Mixtures. 100 ug pGEX-2T DNA was digested with EcoRI, 50 ng of each dephosphorylated (dp) and undephosphorylated vector DNAs were self-ligated separately and used to transform 100 ul of competent cells. Left plate was transformed by 50 ng dephosphorylated vector DNA, and right plate was transformed by 50 ng undephosphorylated vector DNA.

Figure 21. EcoRI and BamHI Digestion of pGEX-2T Clone Containing a CPMoV RNA Replicase Gene. The digestion mixtures were electrophoresed in a 1% agarose gel. Lanes A and B: Undigested pGEX-2T clones from cell lines JM109 and DH5-alpha. Lane M: Lambda DNA HindIII fragments. Lane C: EcoRI digestion. Lane D: BamHI digestion.

Figure 22. BamHI Digestion of a Large-Scale Preparation of pGEX-2T Clone Containing a CPMoV RNA Replicase Gene. The digestion mixture was electrophoresed in 1% agarose gel. Lane M: Lambda DNA HindIII fragments as markers. Lane pGEX/Rep: A pGEX-2T/Replicase recombinant DNA clone digested with BamHI.



transform E. coli competent cells. Transformants were selected by miniprep, BamHI and EcoRI digestion, and agarose gel electrophoresis. The correct construct, which eliminated the small BamHI fragment, would have a singlesite for either EcoRI or BamHI, thus appearing as a single DNA band of 6.5 kb in the agarose gel (Fig. 23). This clone (#2) was selected and used for the expression studies.

Expression and Examination of Fusion Protein and Replicase

The CPMoV replicase gene was cloned into fusion vector pGEX-2T and expressed as a fusion protein to the carboxyl terminus of glutathione S-transferase (GST) from Schistosoma japonicum (Smith et al., 1986). To optimize the expression conditions, E. coli strains DH5-alpha and JM-109 cells containing fusion vector pGEX-2T were used in a small scale pilot experiment to induce the expression of the 27 kD GST fusion protein. The concentration of IPTG and time of addition were tested as described in the methods section. The results showed that final concentrations of 0.5-2.5 mM of IPTG resulted in equal amounts of protein induction (Fig. 24). Addition of IPTG inhibited cell growth about 5-fold. The best time for addition of IPTG was 1-1.5 h after inoculation, the cell density reached OD_{600} = 0.2.

Recombinant pGEX-2T DNA from clone #2 which contained a CPMoV RNA replicase gene in the correct orientation and in which the small BamHI fragment was removed was selected for expression. Growth of cells, induction of expression, and examination of protein by SDS-PAGE was essentially done as

described in the methods. However, no fusion protein in the small scale preparation (Fig. 25) or replicase protein in the large scale preparation (Fig. 26) was detected by 8% SDS-PAGE.

Figure 23. BamHI and EcoRI Digestion of pGEX-2T Clone Containing a Replicase Gene in the Correct Orientation. The digestion mixtures were electrophoresed in 1% agarose gel. Lane M: Lambda HindIII fragments as markers. Lanes 1 and 2: pGEX-2T recombinant clone digested by EcoRI and BamHI.

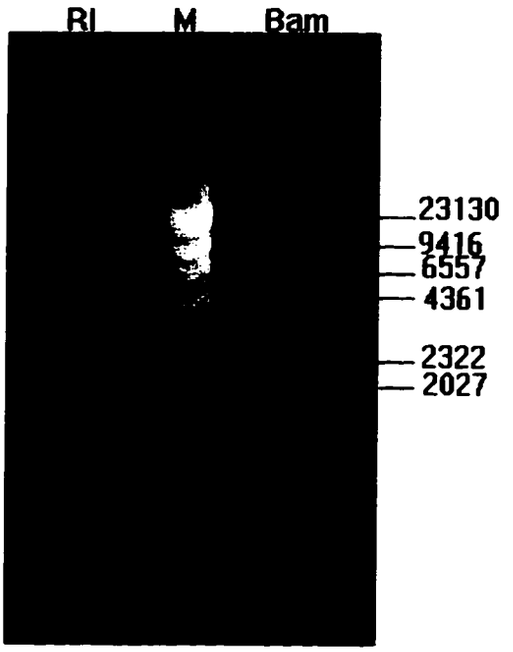


Figure 24. Effect of IPTG Concentrations on the Efficiency of Induction of Glutathione S-transferase (GST) Fusion Protein with Expression Vector pGEX-2T. Supernatant solutions of cell lysates were electrophoresed in 8% SDS-PAGE. Lane M: 2.5-200 kD protein marker (Novex). Lane C: Uninduced control; Lanes 1-5: Induced with a final concentration of 0.5, 1.0, 1.5, 2.0, 2.5 mM of IPTG. The 27 kD protein band at the bottom of the gel is the induced GST fusion protein encoded by pGEX-2T.

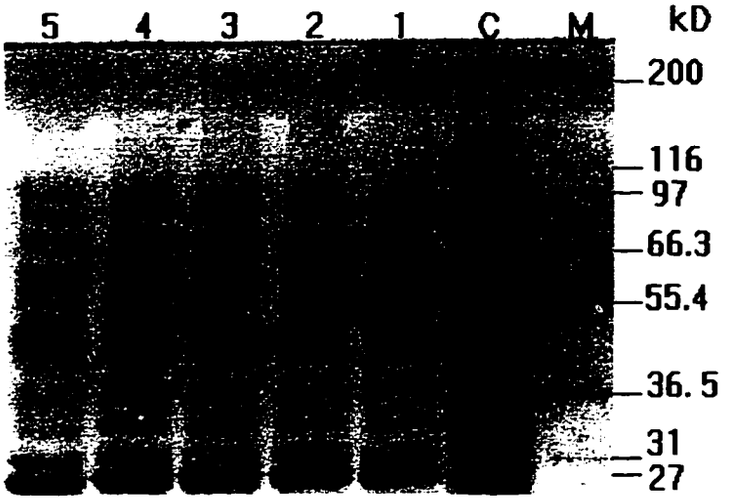


Figure 25. Protein Profile (small scale) of E. coli (strain DH5-alpha and JM-109) Cells Containing pGEX-2T/Replicase Recombinant Clone without and with IPTG Induction. The supernatants of cell lysates were electrophoresed in 8% SDS-PAGE. Lane M: 2.5-200 kD protein marker (Novex). Lanes U1 and In1: DH5-alpha cells without and with IPTG induction; Lanes U2 and In2: JM-109 cells without and with induction.

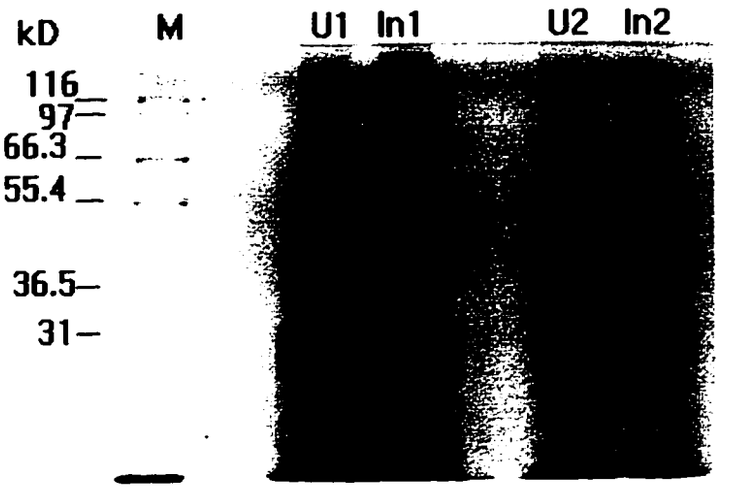
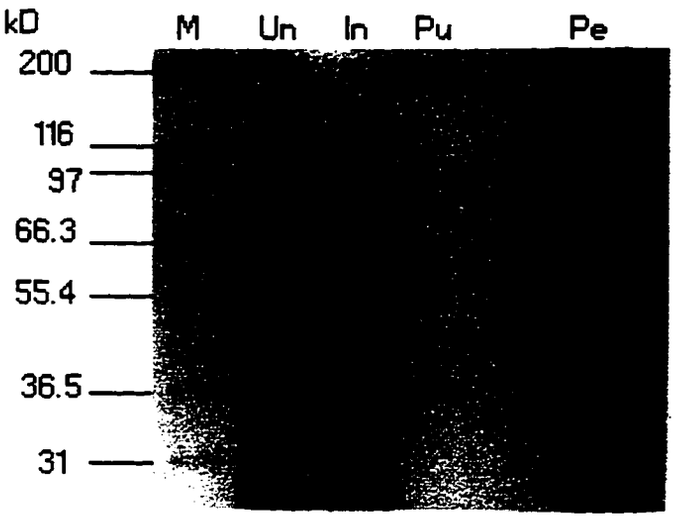


Figure 26. Protein Profile of E. coli (strain DH5-alpha) Cells Containing pGEX-2T/Replicase Recombinant Clone without and with IPTG Induction. The supernatants of cell lysates were electrophoresed in 8% SDS-PAGE. Lane M: 2.5-200 kD protein marker (Novex). Lanes U and In: Supernatant of cell lysates (small scale) without and with IPTG induction. Lane Pu: 30% of total protein sample (30 ul) from a large scale preparation purified on a GST column and cleaved by thrombin. Lane Pe: Pellet fraction of lane In.



DISCUSSION

Cloning and Sequencing of CPMoV Genomic RNA

In the cDNA cloning process, theoretically, when using blue/white selection by IPTG, all white colonies should have an insert. Clones without an insert would have an uninterrupted beta-glucotosidase gene and thus produce blue colonies. However, this was not the case, especially in the early stages of this cloning process when spin columns were used to remove excess EcoRI adaptors. This was perhaps due to the presence of a trace amount of short EcoRI adaptors, which would be cloned more efficiently than the much longer CPMoV cDNA fragments. These small DNA fragments would have been too small to be seen in the 1% agarose gel after EcoRI digestion.

For double strand DNA sequencing using the USB Version 2™ Sequencing Kit, several factors needed to be considered. First, the quantity of the template DNA. Originally, 2 ml overnight LB cultures were used to prepare double strand plasmid DNA as templates for sequencing. For most clones, the amount of plasmid DNA obtained from 2 ml overnight LB culture was sufficient. However, occasionally, and

especially for those clones containing a longer insert, faint bands and compression (band crossing over all four lanes) were observed in the sequencing gel. This was probably due to a high primer/template ratio resulting from a low amount of plasmid DNA. If this was true, using a 5 ml culture would solve the problem. The template DNA purity was not very critical for sequencing. RNA or genomic DNA contaminants had no obvious effect on the quality of sequencing. During early stages of the sequencing, cells were lysed with sodium hydroxide, and plasmid DNA was extracted with phenol/chloroform and precipitated with PEG. Later, when the PEG step was omitted, the DNA preparations still produced good sequencing reactions. Sometimes the compression bands would not disappear even with an increased amount of template DNA. This kind of compression might be due to the secondary structure or the GC rich nature of the template DNA, thus longer denaturation (from 5 min to 10 min), higher extension temperatures (from 37° C to 42° C) or other modifications might be required. To obtain sequences far from the primer, undiluted reaction mixture (which gave a higher ratio of dNTP to ddNTP) was used in the labeling reactions. To obtain sequences near the primer, $MnCl_2$ and a greate than 5-fold dilution of the reaction mixture were used in the labeling reactions. In addition, to increase the number of readable nucleotides from each gel, two or three loadings of the same reaction in a 3 to 4 h interval were used. Different buffer systems have also been used for

the same purpose. Gels run in 0.5X TBE as the anode buffer and 1X TBE (1/2 volume of 0.5X TBE, and 1/2 volume of 3 M sodium acetate) as cathode buffer increased the number of nucleotides readable from each gel, because the distance between the bands was decreased near the bottom of the gel (Sheen and Seed, 1988). In addition, sequencing gels not fixed with 20% methanol and 10% acetic acid before vacuum drying were equal in quality to those fixed.

Prediction of Properties of CPMoV RNA Replicase and Capsid Protein

Predictions of the properties of deduced CPMoV proteins by MicroGenie software are shown in Table 2 and 3. The ORF2 product, the putative RNA replicase, has a molecular weight of 56 kD and a pI of 6.24 (Table 2). The capsid protein (ORF5) has a molecular weight of 40 kD and a pI of 6.24 (Table 3).

The *in vitro* translation of CPMoV RNA produced four major polypeptides (Kim and Bozarth, 1992) with estimated molecular weights of about 50, 40, 27 and 11 kD. The 40 kD polypeptide would correspond to the ORF5 product, the capsid protein, which reacted to CPMoV capsid protein antiserum in previous study. (Kim and Bozarth, 1992). The 50 kD protein was probably encoded by ORF2, the putative replicase gene. The 27 kD protein could have been the product of either ORF1 or ORF6. The 11 kD protein could have been the product of ORF3 or ORF4. However, there was no protein corresponding to a read-through product of ORF1/ORF2 (Kim and Bozarth,

Table 2. Amino Acid Compositions and Other Properties of the CPMoV RNA Replicase (497 amino acids) Deduced from the Nucleotide Sequences (ORF2).

Ala	31	(6.2)	Leu	43	(8.7)
Arg	33	(6.6)	Lya	29	(5.8)
Asn	25	(5.0)	Met	12	(2.4)
Asp	24	(4.8)	Phe	23	(4.6)
Cys	10	(2.0)	Pro	27	(5.4)
Gln	16	(3.2)	Ser	28	(5.6)
Glu	24	(4.8)	Thr	22	(4.4)
Gly	43	(8.7)	Trp	7	(1.4)
His	16	(3.2)	Tyr	22	(4.4)
Ile	26	(5.2)	Val	35	(7.0)
End	1	(0.2)			

Acidic	(Asp + Glu)	48	(9.7)
Basic	(Arg + Lys)	62	(12.5)
Aromatic	(Phe + Trp + Tyr)	52	(10.5)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	168	(33.8)

Molecular weight = 55957 Average pI = 6.24

Table 3. Amino Acid Compositions and Other Properties of the CPMoV Capsid Protein (368 amino acids) Deduced from the Nucleotide Sequences (ORF5).

Ala	36	(9.8)	Leu	28	(7.6)
Arg	21	(5.7)	Lya	17	(4.6)
Asn	23	(6.3)	Met	5	(1.4)
Asp	18	(4.9)	Phe	4	(1.1)
Cys	3	(0.8)	Pro	22	(6.0)
Gln	13	(3.5)	Ser	37	(10.0)
Glu	7	(1.9)	Thr	41	(11.1)
Gly	18	(4.9)	Trp	4	(1.1)
His	5	(1.4)	Tyr	14	(3.8)
Ile	20	(5.4)	Val	31	(8.4)
End	1	(0.3)			

Acidic	(Asp + Glu)	25	(6.8)
Basic	(Arg + Lys)	38	(10.3)
Aromatic	(Phe + Trp + Tyr)	22	(6.0)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	106	(28.8)

Molecular weight = 39610 Average pI = 6.24

1992). This may have been due to either the *in vitro* translation conditions were not favorable for producingread-through polypeptides, or the read-through protein might be a transient product.

CPMoV Sequence Analysis and Comparison

Sequence comparison of CPMoV to other viruses revealed that it has extensive sequence similarity to carmovirus and tombusvirus. The extensive homology obsered for both the nucleotide and the deduced amino acid sequences within the ORF2 coding regions (the putative RNA replicase), is noteworthy. The amino acid sequence of the CPMoV RNA replicase shows greater sequence homology to carmoviruses than to tombusviruses (Fig. 27). There were at least 10 separate consensus sequences among members of the carmovirus group, fewer among tombusviridae (Fig. 28). The lengths of the putative RNA replicase coding regions were variable, especially ORF1 coding region, at the ends among carmoviruses. CPMoV putative RNA replicase has the best sequence homology to CCFV and TCV. Interestingly, at the amino acid level, the sequence homology among the putative RNA replicases is much higher than that of the capsid protein among carmovirus and tombusviruses. Obviously, RNA replicase has been under stringent selection pressure throughout evolution since the conservation of amino acids sequence is vital to the structure and function of this key viral enzyme. Significant sequence homologies have previously been reported among RNA replicases of carmovirus,

Figure 27. Dot Matrix Alignments of CPMoV Putative RNA Replicase Sequence with Those of Carmoviruses: a. TCV; b. MNSV; c. CarMV; and Tombusviruses: d. TBSV; e. CNV; f. TBSV/CNV (two tombusvirus, their replicases share 94.8% sequence homology). Each dot represents a match of 4 consecutive nucleotides.

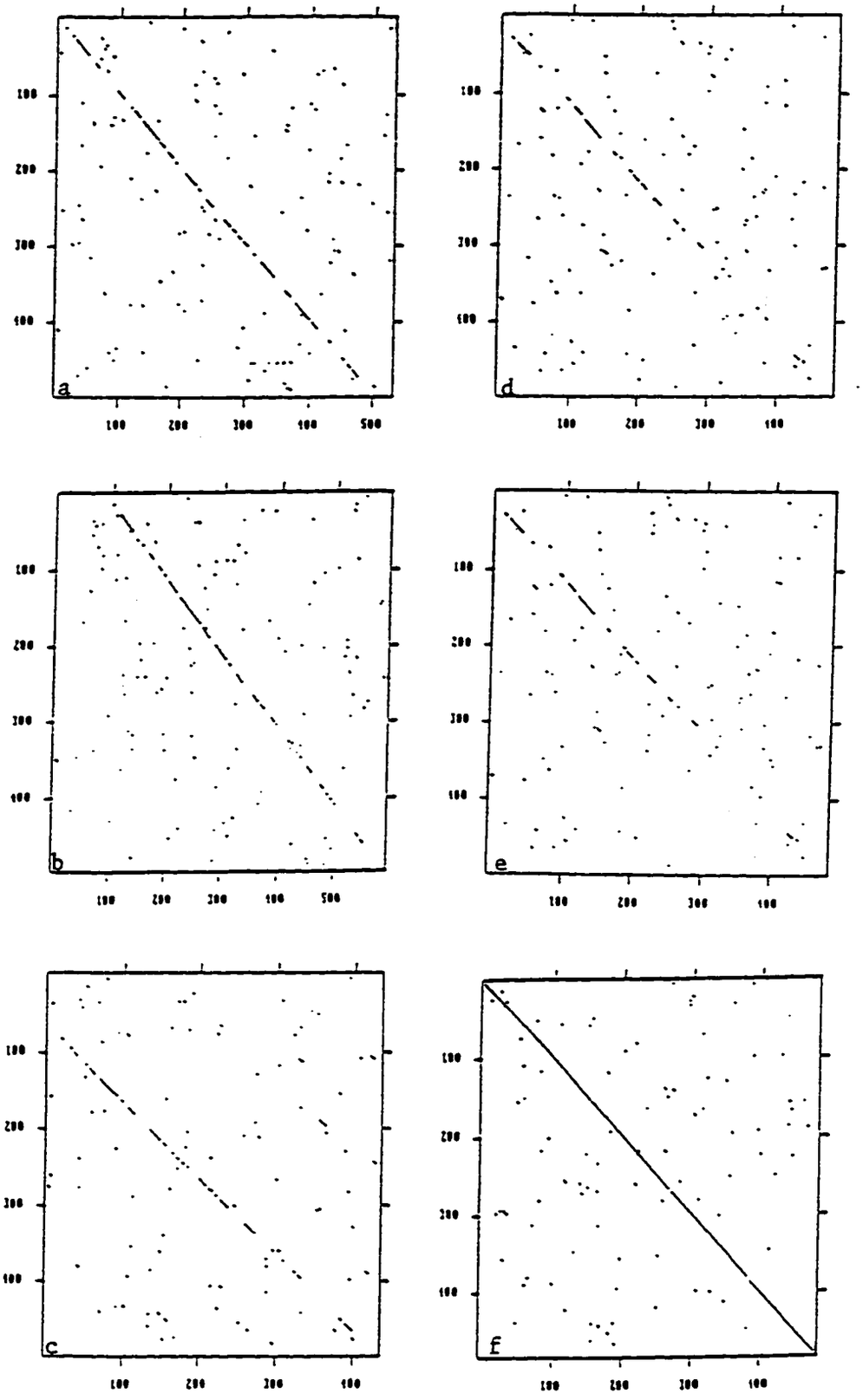


Figure 28. Consensus Sequences of RNA Replicases among Carmoviruses; [turnip crinkle virus (TCV), melon necrotic spot virus (MNSV), cardamine chlorotic fleck virus (CCFV), carnation mottle virus (CarMV)] and Tombusviruses; [cucumber necrosis virus (CNV), tomato bushy stunt virus (TBSV)]. For each amino acid position, homology in all 7 viruses is indicated by capital letters in the virus line and the consensus line; homology among 4-6 viruses is indicated by a capital letter in the virus line and a small letter in the consensus line; homology in 3 of the 7 viruses is indicated by small letter in the virus line and the consensus line; homology in less than 3 viruses is indicated by a dash.

```
MNSV  1  ----------------.....------l-g-P---.....------F-VHNnSL-N
CPMoV 1                          m---p--g--i-r.....mgng--FGVHN-SL-N
CCFV  1                          m---pl-g-Pi-r.....mgng--FGVHNnSL-N
TCV   1                          m---pl---pk-r.....mgng--FGVHNnSL-N
CarMV 0

CNV   1                                                .....--F-VHN--L-N
TBSV  1                                                .....--F-VHN--L-N

consensus ----------------.....m---pl-g-pk-r.....mgng--fgvhnnsl-n


MNSV  123 LRRGLvERVF-VE-----L-PaP-P--GAF-RL--FRRkL---VG-H--Is---FL--Y-G
CPMoV  37 LRRGL-ERVFyVE ----L-P-P-PipG-F-RL-gF--kL---VG-H-RIs-d-F---Y-G
CCFV   38 LRRGL-ERVFyVE-----L-PaP-Pip-AF--LSg-RRkL----GnHt-I----F---Y-G
TCV    38 LRRGL-ERVFyVE-----L-PaP-PipG-F--LSg-RR-L----GnHt--------s-Y-G
CarMV   1                                      --RR------GnHtRIs----L--Y-G

CNV    21 L-RGLvERVF-VE ----L---P-P--GAF-RLS-FR------VG---Rl--d-FLs-Y-G
TBSV   21 L-RGLvERVF-VE ----L---P-P--GAF-RLS-FR------VG---Rl--d-FLs-Y-G

consensus l-rglvervfyve-----l-pap-pipgaf-rlsgfrrkl---vGnhtris-d-fls-Y-G


MNSV  184 RR-TIY--AV-SL----VqR-Da-LKTFVKAEKIN-T-K-DPAPRVIQPRn-RYNVEVGRY
CPMoV  97 RR-TIYqnAV-SL----VqRkD--L-TF-KAEKI----K-DPAPRVIQPRnPRYNVEVGRY
CCFV   98 RR-TIYqnA--SL-----q---a-LKTFVKAEKIN-T-K-DPAPRVIQPR-PRYN-EVG-Y
TCV    98 RR-TIYq-A----------RkDa-LKTFVKAEKIN-T-K-DPAPRVIQPR-PRYN-EVG-Y
CarMV  30 R--TIY-nA--SL------RkD--LKTF-KAEK-N---K-DPAPRVIQPR-PRYNVE-GRY

CNV    81 -----Y--AV-SL----V---D--L-TFVKAEKI -T-K-DPAPRVIQPR-PRYNVE-GRY
TBSV   81 -----Y--AV-SL----V---D--L-TFVKAEKI -T-K-DPAPRVIQPRnPRYNVE-GRY

consensus rr-tiYqnAv-sl----vqrkda-LkTFvKAEKin-t-K-DPAPRVIQPRnpRYNvEvGrY


MNSV  245 LR--EH-LYR-ID--W-GPT-IKGYTVe-iG-I--dAWD-F--PVAIGFDMkRFDQHVS-D
CPMoV 158 LR--EHHLYR-ID---GGPTVIKGY-V--iGnIme-A--qF--PVA-GFDMSRFDQHVS-D
CCFV  159 Lk-yEHHLYRAID--W-GPTV-KGY-VeEiGnIm-dAW-qF-KP-AIGFDMkRFDQHVSVD
TCV   159 Lk-yEHHLYRAID--WGGPTV-KGY-V-E-GnIm---WD-F-K--AIGFDMkRFDQHVSVD
CarMV  91 Lk-yEHH-Y-AlD--WGGPTV-KGYT-eEv------AW-qF--PVAIGFDMSRFDQHVSV-

CNV   141 LR--E--L--A-D-----PT-IKGYT--EvG-I-ed-WD-F-KPVAIG-DMSRFDQH-SV-
TBSV  141 LR--E--L--A-D---G--T-IKGYT--EvG-I-e--WD-F-KPVAIG-D-SRFDQH-SV-

consensus Lr-yEhhlyraiD--wggpTviKGYtvee-gnimedawdqF-kpvAiGfDmsRFDQHvSvd


MNSV  306 AL-wEHSVYL--F--d--LA-LL-WQL-NKGVG-ASDG-IKY-v-GCRMSGDMNTAmGNCL
CPMoV 219 ALqfEH-VYL--F-Gd--LAKLL-WQL-NKG------G--KY ----RMSGDMNTAmG -L
CCFV  220 AL-wEHSVY-A-F---- LAK-L-WQL-NKGVG-ASDG-IKY-vEGCRMSGD-NTAmGNCL
TCV   220 AL-wEHSVY-A-F---- LA-LL-WQL-NKGVG-ASDG-IKY-v-GCRMSGD-NTA-GNCL
CarMV 152 AL-fEHS-YLA-F-Gd--LA-LL--QL-N-GVG-AS-G-l-Y-kEGCRMSGDMNTA-GNCL
```

```
CNV    201 ALqfEH--Y-A---G---L-KLL-WQL-NKGVG---DG-I-Y-kEGCRMSGD-NT -LG-Y
TBSV   201 ALq-EH--Y-A---G---L-KLL-WQL-NKG-G----G-I-Y-kEGCRMSGD-NT -LG-Y

consensus AL--EHsvYla-f-gd--LaklL-wQL-NkGvg-asdG-ikYq-egcRMSGDmNTamgncL


MNSV   367 -AC- IT-------GI-- RL-NNGDDCVvI-E------------W--FGFqC--E-----
CPMoV  278 LAC- IT  ---MKGIk- RL-NNGDDCVvI ------V---l----d-GF-CIAEEPVYe
CCFV   280 LAC- IT  --lMKGIk- RLiNNGDDCVv -------V---l--W-dFGFqCIAEEP-Y-
TCV    280 LAC- IT  --lMKGIk- -LiNNGDDCVL -------V---l--W-dFGFqCIAEEP-Ye
CarMV  213 LAC- IT  --lMK I-- RLiNNGDDCVLI-E-----V------W--FGF-CIAEEPVYe

CNV    261 --------------GI----L-NNGDDCVLI-E---------------FG-----E-PV--
TBSC   261 --------------GI----L-N-GDDCVLI-E---------------FG-----E-PV--

consensus lAC- it----lmkgIk--rLiNnGDDCVli-e-----v---l--w-dfGfqCiaEepvye

MNSV   428 -E--EFCQM-PVyDGe--vMVRNP-VSLSKDSYS---Wn----A--W--a-G-CGL-lTGG
CPMoV  336 LEK-EFC---P-yDG--W---R-P-VSL-KD-Y----Wnn--dAa-WL-aiG-CG-AITGG
CCFV   338 LEKVEFCQM-P-yDGeGWvMVRNP-VSLSKDSYS---W----dAa-WL--iG-CG-AIAGG
TCV    338 LEKVEFCQM-P--DGeGWvMVRNP-VSLSKDSYS---W-n--dAa-WL-aiG-CGLAIAGG
CarMV  271 -EK--FCQM-PV-DG-GW-MVR-P-VSmSKDShS---Wnn---A--WL---G-CGL-IAGG

CNV    321 LE-VEFCQ--PV----GW-MVRN----mSKD--------------W------CG-Al--G
TBSV   321 LE-VEFCQ--PV----GW-MVRN----mSKD--------------W-------GLAl--G

consensus lEkveFCqm-PvydgegWvmvRnp-vslsKDsys---wnn--daa-Wl-aig-cGlaiagG


MNSV   389 IPVVQsYY---iRN  -----------dv-f-SGF--la--G-R-s---Se-ARfSFY-AF
CPMoV  397 IP--Q-YY-cliRNf  -------K---------------G------V---AR-SFY---
CCFV   399 IP-VQsYY-cliRNf---------K--dv-f-SGF--la--G-RGs--VSe-AR-SFY-AF
TCV    399 -PV-QsYY-cl-RNf---------K--dv-f-SGF--l---G-RGs--VS--ARfSFY--F
CarMV  332 -PV   ----------------K-------SGF---a----RG---VSe--RfSFY-AF

CNV    381 IPVV---Y-------------------------G-------------V---AR-SF--AF
TBSV   381 IPVV---Y-------------------------G-------------V----R-SF--AF

consensus iPvvqsyy-clirnf---------k--dv-f-sgf--la--g-rgs-avse-aRfSFy-af


MNSV   548 G-TPD-Q-A-E--Yd---------P-g----------lN--*
CPMoV  456 G-TPD-Q-A-E-yY-nl-L-----P------l------*
CCFV   460 G-TPD-Q--LE-yYdnl-L-----P-g--e-l---wi-N------*
TCV    460 G-TPD-Q-ALE-yYdnl-L-----P-g--e-l---wilN------*
CarMV  390 G-TPD-Q-ALE  -----L----------e-----wilN----*

CNV    421 G-T-D-Q-ALE-------------------------*
TBSV   421 G-T-D-Q-ALE-------------------------*

consensus G-TpD-Q-AlE-yydnl-l-----p-g--e-l---wiln------*
```

tombusvirus, luteovirus (Rochon and Tremaine, 1989), necrovirus (Meulewaeter et al, 1990), dianthovirus and ungrouped maize chlorotic mottle virus (Xiong and Lommel, 1989). More extensive comparative analysis of the amino acid sequences of the putative RNA replicase of positive strand RNA viruses revealed that RNA replicase is the only viral protein containing motifs conserved throughout this class of viruses (Koonin, 1991, Koonin and Dolja, 1993). According to the sequence features of their RNA replicases, Koonin suggested that these viruses could be assigned to three super groups with eight distince conserved motifs among them. Bruenn (1991) compared the replicase sequences of 43 positive strand and 7 double strand viruses. He found homology among these viruses and suggested that all positive strand RNA viruses of eukaryotes except picornaviruses may have evolved from an ancestral dsRNA virus.

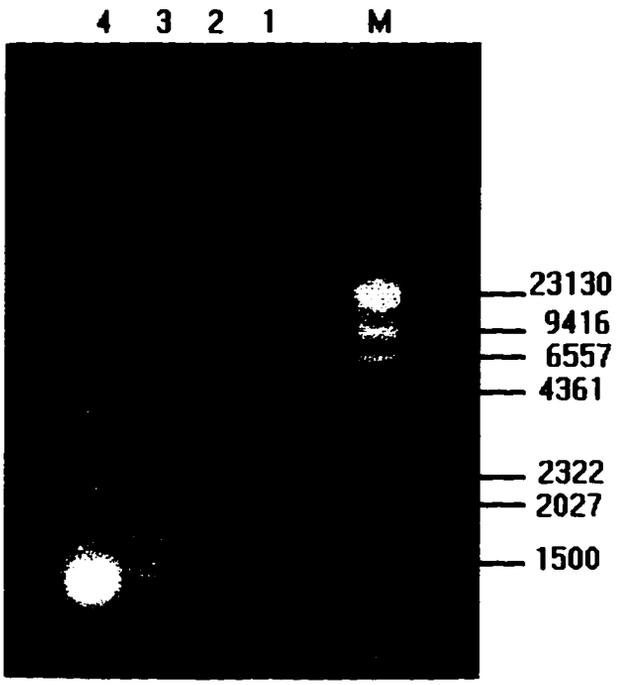
The conservation of RNA replicase sequences among RNA viruses may be significant. Recently, virus resistance mediated by modified viral RNA replicase genes has been reported by three different groups of researchers (Anderson et al., 1992; MacFarlane and Davies, 1992; Longstaff et al., 1993). Since the conserved motifs of RNA replicases are shared by many viruses, resistance induced by insertion of a modified replicase gene might be effective against a large spectrum of viruses.

PCR Amplification and Cloning of CPMoV RNA Replicase Geneinto TA Cloning Vector

cDNAs of CPMoV replicase fragment were amplified by PCR. At an extension temperature of 55° C, PCR amplified CPMoV RNA replicase resulted in a major product of 1.5 kb and some minor bands of low molecular weight (Fig. 14). However, if the extension reaction was carried out at 60° C, the yield of the replicase fragment was greatly decreased and the low molecular weight bands disappeared. The concentrations of $MgCl_2$ had a great impact on the yield and specificity of product. In one experiment, a final concentration of 1.5, 2.5, 3.5, 4.5 mM of $MgCl_2$ was used respectively in four different PCR reactions. The results showed that 1.5 mM $MgCl_2$ gave the best yield and specificity of the 1.5 kb replicase fragment in 1% agarose gel (Fig. 29).

PCR amplified CPMoV replicase fragment turned out to be extremely difficult to digest by restriction enzymes. There were several possible explanations for this difficulty. Taq polymerase might be inefficient for certain terminal sequences, producing frayed ends that can not be cleaved by the restriction endonuclease. Alternatively, the "breathing" of terminal sequences might prevent the stable association of restriction endonucleases with terminal sites. Also, Taq polymerase or other contaminants might bind to the ends of the PCR products, blocking restriction endonuclease activity. Different strategies have been used to solve this problem, which included using proteinase K to

Figure 29. Effect of $MgCl_2$ Concentration on the Efficiency and Specificity of PCR Amplification of CPMoV RNA Replicase Gene. PCR products were electrophoresed in 1% agarose gel. Lane M: Lambda DNA HindIII fragment as markers; Lanes 1-4: 1.5, 2.5, 3.5 and 4.5 mM $MgCl_2$.



deblock, spermidine to stabilize, or T4 polymerase to repair the ends or to convert the terminal restriction sites into internal sites by concatamerization of the PCR product with T4 DNA ligase (Jung et al., 1993). The problem was solved by using commercially available TA cloning vector (Invitrogen) to convert the external restriction sites into internal restriction sites. This vector took advantage of the non-template-dependent activity of the thermostable polymerase used in PCR that adds a single deoxyadenosine to the 3′ end of all duplex molecules produced by PCR. The vector has a T overhang at the 3′end of the linearized double strand DNA which would be specifically annealed with the A overhang generated by the Taq polymerase during PCR amplification. After the external restriction sites were converted into internal sites, they should be easily cut. This was indeed the case.

Cloning of CPMoV Replicase Gene into Fusion Vector pGEX-2T

A TA clone containing a CPMoV RNA replicase fragment was digested by EcoRI and analyzed by agarose gel electrophoresis. The 1.5 kb replicase band was excised from the gel and eluted by Geneclean.

Dephosphorylation of the ends of pGEX-2T efficiently decreased the chance of vector self-ligation. However, the replicase fragment with cohesive EcoRI ends could ligate to the vectors or ligate to each other. Thus, a proper molar ratio of vector to DNA was important to obtain recombinant plasmids with a single copy of replicase fragment. A ratio

of 3 to 1 was used to decrease the chance of replicase fragment self ligation. Ligation products without a copy of vector sequence would be eliminated by ampicillin selection.

Expression and Examination of CPMoV RNA Replicase

The GST gene fusion system has several advantages for foreign gen expression: The tac promoter for inducible, high-level expression; an internal lacIq gene for host-independence; the cleavage site for thrombin to allow cleavage of the desired polypeptide from the fusion product; easy purification using Glutathione Sepharose 4B; and mild elution conditions to maintain the functional activity of the target protein. However, the recombinant fusion vector pGEX-2T, which contained an in-frame CPMoV RNA replicase gene, failed to express the fusion protein. There are several possible reasons for this failure. First, the tag polymerase might introduce errors into the replicase gene during the process of PCR amplification. If errors appeared in the fusion border, or if they resulted in a termination codon anywhere in the gene, it would interrupt the expression. However, the fusion border of the TA clone containing replicase gene was examined by nucleotide sequencing, and there was no such error observed in the fusion border region. Second, there was a possibility that the expressed protein was not soluble in the GST fusion system. This was examined by loading both the supernatant and the pellet of the cell lysate to the SDS-PAGE. The result showed no protein band with the expected molecular

weight of 83 kD in either the supernatant lane or the pellet lane (Fig. 26). Third, the yield of expressed CPMoV RNA replicase might be very low, thus the 3 ml culture used in the small scale preparation would not yield enough fusion protein to be detected by 8% SDS-PAGE. To eliminate this possibility, a large scale preparation was performed. The supernatant of a cell lysate from a 100 ml induced cell culture was passed through a 1 ml Glutathione Sepharose 4B affinity column. After the proper wash with buffer 1, the replicase should be cleaved from the glutathione S-transferase domain by incubating the affinity resin with site-specific protease thrombin at 4° C overnight. 30 ul (15% of the total volume) of this solution was analyzed by 8% SDS-PAGE, but no expected replicase band of 56 kD was observed (Fig. 26). Another possibility was that the expressed protein might be extremely toxic to the cell. If this had been the case, the cell growth would have been inhibited. Visual observation did not indicate that cell growth was inhibited. However, cells did have a tendency to lose the recombinant plasmid, especially when the host was E. coli stain JM-109. JM-109 cells containing recombinant pGEX-2T were found to have lost their plasmid when they were stored at 4° C for 48 h in liquid LB or three weeks on LB/agar plate, both containing ampicillin (50 ug/ml).

REFERENCES

Allen, D. J. (1980). Identification of resistance to cowpea mottle virus. Trop. Agric. (Trinidad) 57, 325-332.

Allen, D. J., Thottappolly, G., and Rossel, H. W. (1982). Cowpea mottle virus: Field resistance and seed transmission in virus-tolerant cowpea. Ann. Appl. Biol. 100, 331-336.

Anderson, J. M., Palukaitis, P., and Zaitlin, M. (1992). A defective replicase gene induces resistance to cucumber mosaic virus in transgenic tobacco plants. Proc. Natl. Acad. Sci. USA 89, 8759-8763.

Bashir, M., and Hampton, R. O. (1993). Natural occurrence of five seed borne cowpea viruses in Pakistan. Plant Diseases 77, 948-951.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523.

Bruenn, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. Nucl. Acids Res. 19, 217-226.

Bozarth, R. F., and Shoyinka, S. A. (1979). Cowpea mottle virus. CMI/AAB Description of Plant Viruses No. 212.

Carrington, J. C., Heaton, L. A., Zuidema, D., Hillman, B., and Morris, T. J. (1989). The genome structure of turnip crinkle virus. Virology 170, 219-226.

Denhardt, D. T. (1966). A membrane filter-technique for the detection of complimentary DNA. Biochem. Biophys. Res. Commun. 23, 641-646.

Fenner, F. (1976). Classification and nomenclature of viruses: Second report of the International Committee on Taxonomy of Viruses. Intervirology 7, 1-116.

Francki, R. I. B., Fauquet, C. M., Knudson, D. L., and Brown, F. (1991). Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses. Archives of Virology Supplementum 2.

Goldbach, R., and Wellink, J. (1988). Evolution of plus-stranded RNA viruses. Intervirology 29, 260-267.

Guilley, H., Carrington, J. C., Balazs, E., Jonard, G., Richards, K., and Morris, T. J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. Nucl. Acids Res. 13, 6663-6677.

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

Habili, N., and Symons, R. H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. Nucl. Acids Res. 17, 9543-9555.

Hacker, D. L., Petty, T. I., Vei, N., and Morris, T. I. (1992). Turnip crinkle virus genes required for RNA replication and virus movement. Virology 186, 1-8.

Hayes, R. J., Brunt, A. A., and Buck, K. W. (1988). Gene mapping and expression of tomato bushy stunt virus. J. Gen. Virol. 69, 3047-3050.

Hearne, P. Q., Knorr, D. A., Hillman, B. I., and Morris, T. J. (1990). The complete genome structure and synthesis of infectious RNA from clones of tomato bushy stunt virus. Virology 177, 141-151.

Hillman, B. I., Hearne, P., Rochon, D., and Morris, T. J. (1989). Organization of tomato bushy stunt virus genome: Characterization of the coat protein gene and the 3' terminus. Virology 169, 42-50.

Ish-Horowicz, D., and Burke, J. F. (1981). Rapid and efficient cosmid cloning. Nucl. Acids Res. 9, 2989.

Kraft, R., Tardiff, J., Krayter, K. S., and Leinwand, L.A. (1988). Using mini-prep plasmid DNA for sequencing double-stranded templates with Sequenase™. BioTechniques 6, 544-547.

KIM, J. W., and Bozarth R. F. (1992). Mapping and sequence analysis of the capsid protein gene of cowpea mottle virus. Intervirology 33, 135-147.

Jung, V., Pestka, S. B., and Pestka, S. (1993). Cloning of polymerase chain reaction-generated DNA containing terminal restriction endonuclease recognition sites. In "Methods in Enzymology", Vol. 218, 357-362.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA

polymerases of positive-strand RNA viruses. J. Gen. Virol. 72, 2197-2206.

Koonin, E. V., and Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences. Critical Rev. Biochem. and Mol. Bio. 28, 375-430.

Longstaff, M., Brigheti, G., Boccard, F., Chapman, S., and Aulcombe, D. (1993). Extreme resistance to potato virus X infection in plants expressing a modified component of the putative viral replicase. EMBO. J. 12, 397-386.

MacFarlane, S. A., and Davies, J. W. (1992). Plants transformed with a region of the 201-kilodalton replicase gene from pea early browning virus RNA1 are resistant to virus infection. Proc. Natl. Acad. Sci. USA 89, 5829-5833.

Martelli, G. P., Russo, M., and Quacquarelli, A. (1971). Tombusvirus (tomato bushy stunt virus) group, in "Maramorosch, The Atlas of Insect and Plant Viruses" 257-279, Academic Press, New York 1977.

Martelli, G. P., Gallitelli, D., and Russo, M. (1988). Tombusviruses. In "The Plant Viruses", Vol. 3 (Eds, Koenig, R.), 13-72, Plenum, N.Y.

Martelli, G. P. (1992). Classification and nomenclature of plant viruses: state of the art. Plant Dis. 76, 436-442.

Matthews, R. E. F. (1981). Classification and Nomenclature of Viruses: Fourth Report of the International Committee on Taxonomy of Viruses in Strasbourg. Intervirology 12, 129-296.

Meulewaeter, F., Seurinck, J., and Emmelo, J. E. (1990). Genome structure of tobacco necrosis virus strain A. Virology 177, 699-709.

Morris, T. J. (1983). Virus-specific double stranded RNA: Functional role in RNA virus infection. In "Plant Infectious Agents: Viruses, Viroids, Virusoids, and satellites", (eds. Robertson, H.D., Howell, S.H., Zaitlih M., and Malmberg, R.L.), 80-83, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Morris, T. J., and Carrington, J. C. (1988). Carnation mottle virus and viruses with similar properties, In "The Plant Viruses: Vol. 3. Polyhedral virions with Monopartite RNA Genomes" (Eds. Koenig, R.). Plenum, New York.

Morris, J. T. (1991). Carnation mottle carmovirus group. in "the classification and nomenclature of plant viruses": Francke, R. I. B., Fauquet, C. M., Knudson, D. L., and Brown, F., eds. Fifth report of the international committee on taxonomy of viruses. Archives of virology supplement 2. pp. 303-308.

Morrow, C. D., Warren B. and Lents, M. R., (1987). Expression of enzymatically active poliovirus RNA-dependent RNA polymerase in Escherichia coli. Proc. Natl. Acad. Sci. USA 84, 6050-6054.

Mullis, K. B., and Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In "Methods in Enzymology", () 155-335.

Nutter, R. C., Scheets, K., Panganiban, L. C., and Lommel, S. A. (1989). The complete nucleotide sequence of the maize chlorotic mottle virus genome. Nucl. Acids Res. 17, 3163-3177.

Ndiaye, M., Bashir, M., Kekker, K. E. and Hampton, R. O. (1993). Cowpea viruses in Senegal, West Africa: Identification, distribution, seed transmission, and sources of genetic resistance. Plant Dis. 77, 999-1003.

Oste, C. (1988). Polymerase chain reaction. BioTechniques 6, 162.

Plotch, S. P., Palant, O. and Gluzman, Y. (1989). Purification and properties of poliovirus RNA polymerase expressed in Escherichia coli. J. Virol. 63, 216-225.

Riviere, C. J., Pot, J., Tremaine, J. H., and Rochon, D. M. (1989). Coat protein of melon necrotic spot carmoviruses is more similar to those of tombusviruses than those of carmoviruses. J. Gen. Virol. 70, 3033-3042.

Riviere, C. J., and Rochon, D. M. (1990). Nucleotide sequence and genomic organization of melon necrotic spot virus. J. Gen. Virol. 70, 935-947.

Robertson, D. G. (1966). Seed-borne viruses of cowpea in Nigeria. BSc. Thesis, Univ. Oxford.

Robertson, D. G. (1963). Cowpea virus research in Nigeria. In Proc. 1st Nigerian Grain Legume Conference, Samaru, Zaria. Inst. Agric. Res., Ahmadu Bello Univ.

Rochon, D. M., and Tremaine, J. H. (1989). Complete

nucleotide sequence of the cucumber necrosis virus genome. Virology 169, 251-259.

Rossel, H. C. (1977). Preliminary investigations on the identity and ecology of legume virus diseases in Northern Nigeria. Trop. Gr. Leg. Bull. 8, 41-46.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

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

Sankar, S., and Porter, A. G., (1991). Expression, purification, and properties of recombinant encephalomyocarditis virus RNA-dependent RNA polymerase. J. Virol. 65, 2993-3000.

Sheen, J. Y., and Seed, B. (1988). Electrolyte gradient gels for DNA sequencing. BioTechniques 6, 944.

Shoyinka, S. A., Bozarth, R. F., Reese, J., and Rossel, H. W. (1978). Cowpea mottle virus: a seed-borne virus with distinctive properties infecting cowpeas in Nigeria. Phytopathology 68, 693-699.

Skotnicki, M. L., Mackenzie, M., Torronen, M., and Gibbs A. J. (1993). The genomic sequence of cardamine chlorotic fleck carmovirus. J. Gen. Virol. 74, 1933-1937.

Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. U., Garcia, E. G., and Mictchell, G. F. (1986). Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase. Proc. Natl. Acad. Sci. USA 83, 8703-8707.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503.

Tabor, S., and Richardson, C. C. (1987). DNA sequencing analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84, 4767-4771.

Ward, W. C. (1993). Progress towards a higher taxonomy of