IN PLANTA DELETION OF DNA INSERTS IN THE LARGE INTERGENIC REGION OF CAULIFLOWER MOSAIC VIRUS DNA

By

RODNEY EDWARD PENNINGTON II

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1980

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1991

IN PLANTA DELETION OF DNA INSERTS IN
THE LARGE INTERGENIC REGION OF
CAULIFLOWER MOSAIC VIRUS DNA

Thesis Approved:

Ulrich K. Melcher
Thesis Advisor

Richard C Essenberg

Carol D Mitchell

Margaret K. Essenberg

D. W. Thesis

Norman N. Durham
Dean of the Graduate College

PREFACE

Over the course of my graduate career in the Oklahoma State University Biochemistry Department my research objectives have evolved. My original objective was to determine the utility of cauliflower mosaic virus (CaMV) DNA as a vehicle to carry inserted DNA sequences into plants for gene targeting experiments. Specifically, my intention was to construct, in plasmid form, one or more infectious, stable CaMV DNAs bearing fragments of the *Arabidopsis thaliana* alcohol dehydrogenase (ADH) gene coding region. Once stable constructs were obtained, I planned to look for inactivation of the ADH gene in *Arabidopsis* seeds obtained from plants inoculated and systemically infected with the constructs. I would then examine any plants grown from ADH^- seeds for evidence that the observed ADH gene inactivation was due to insertion of CaMV DNA sequences within the gene. I was successful in creating several plasmid CaMV DNAs bearing pieces of ADH DNA. These fragments were added to a portion of the CaMV genome called the large intergenic region (IG) using, as recipient for insertions in the IG, a plasmid CaMV clone constructed by Dr. Melcher. The next step was to inoculate turnips with these plasmids to obtain material for virion inoculation of *Arabidopsis*. This was necessary because plasmid inoculation of *Arabidopsis* is difficult and inefficient. However, after inoculating turnips with the constructs I discovered that the ADH inserts were invariably deleted in plants. Given 1) the instability of the CaMV/ADH constructs, 2) time constraints associated with my degree program, and 3) other anticipated difficulties I decided, in consultation with Dr. Melcher, to change research objectives. We decided that I would continue and complete work on a project of his which was both on-going and relevant to my initial results showing instability of ADH gene fragments in the IG. Specifically, the goal of this project was to discern mechanisms governing stability, or

instability, of insertions in the IG. This paper contains results of initial experiments, performed by myself and by Dr. Melcher (with the able assistance of his two laboratory technicians, Ann Williams and Mary Schatz), as well as results of experiments and analyses performed by me after I took over the project. Following is a breakdown of specific contributions to the manuscript. Dr. Melcher constructed and determined the infectivities of the following plasmid cloned viral DNAs: pCML1, pCaMTL(+), pCaMTL(-), pCaMTL(+)f, and pCML3 through pCML13 (Table 1). He also performed the restriction analyses of viral DNAs recovered from plants infected with these clones (Figs. 4A, 4B, 5A, 5B, and 5C) as well as analyses of viral DNAs recovered from plants inoculated with cloned progeny DNAs and passaged viral DNA (Figs. 7 and 8). I constructed pRPs 1(+), 1(-), 5(-), 5(2+), 5(+), 6(+), and pCML5d (Table 1) and performed the corresponding restriction analyses (Fig. 5D, 5E, and 6). I also performed additional restriction analyses of viral DNAs recovered from pCML1 and pCaMTL(-) infected plants. These results are not shown in any figure but are included in the results and discussion. Cloning and partial DNA sequencing of individual viral DNAs recovered from selected plants was performed by Dr. Melcher (pCML(+), pCML3, pCML5) and by myself (pCML1, pRP1(+)). These results are summarized (Table 2) as are results of dot matrix homology analyses performed by myself. All summaries, conclusions, and speculations in the manuscript are my own.

A decision made during the writing was that our findings would be written in the form of one longer manuscript rather than two or more shorter ones. A consequence of this decision is that rather than having two or more thesis chapters consisting of separate manuscripts to be submitted for publication, I have only one. Thus, I am presenting the parts of the manuscript as individual chapters. The manuscript is intended to be complete and ready, except for minor formatting details, to be submitted to the journal **Virology** under the authorship of myself and Dr. Melcher. As journal article introductions are generally brief, I am including in this preface additional general information about CaMV

which will assist the reader in understanding the manuscript.

CaMV Background Information

Biology and Structure of CaMV. Cauliflower mosaic virus (CaMV) is the type member of the caulimoviruses, one of only two known groups of plant viruses having, for a genome, double-stranded DNA (see review by Shepherd (1989) for information not otherwise attributed). CaMV infects many species of the Brassicaceae (including *Arabidopsis thaliana* and *Brassica rapa*) and some isolates infect some members of the Solanaceae. CaMV-infected plants show symptoms of various types and severities depending on the infecting isolate and host species. Possible symptoms include leaf mosaics or mottles, distortion of leaves, and stunting. The virus is transmitted by aphids in nature and can also be transmitted by mechanical inoculation.

The bulk of CaMV virions in infected cells are localized in large cytoplasmic proteinaceous inclusion bodies which serve as the site of virus assembly (Marsh *et al.*, 1985). The virions themselves are 50 nm icosahedral particles consisting of approximately 84% protein and 16% DNA. The double-strandedness of DNA found in virions is interrupted by two or three (depending on the isolate) site-specific gaps. One of these gaps is invariably found in the minus DNA strand and by convention defines position 1 in the clockwise numbering of the plus strand DNA sequence (12 o'clock position, Fig. 1A).

Genetic Organization of CaMV and ORF Products. The DNA is approximately 8 kilobasepairs (kbp) and contains eight large open reading frames (ORFs), all in the plus strand (see Fig. 1A). ORFS I through V are tightly packed, in some cases abutting or overlapping. ORF VI is separated from ORFs V and VII by intergenic regions of 100 and 700 basepairs (bp), respectively. ORF VIII (not shown) is out of frame with and overlaps the 3' end of ORF IV. Functions have been shown or surmised for most of these ORFs. ORF I encodes a protein similar in sequence, and possibly function, to the cell-to-cell movement proteins of tobacco mosaic virus and other viruses (Hull *et al.*, 1986; Melcher,

1990). ORF II encodes a protein required for aphid transmission (Armour *et al.*, 1983; Woolston *et al.*, 1983). ORF III encodes a protein which has DNA binding activity and may be a structural component of the virion (Mesnard *et al.*, 1990). ORF IV encodes the coat protein precursor (Daubert *et al.*, 1982). All evidence indicates ORF V encodes the viral replicase, in this case an RNA-dependent DNA polymerase (reverse transcriptase) (Takatsuji *et al.*, 1986). ORF VI encodes the inclusion body protein (Xiong *et al.*, 1982) which forms the inclusion body matrix. This ORF also contains symptom (Daubert *et al.*, 1983) and host-range determinants (Schoelz *et al.*, 1986) and has further importance as discussed later. ORFs VII (Dixon and Hohn, 1984) and VIII (Schultze *et al.*, 1990) are dispensable for infection and may be not expressed in infected plants.

CaMV Infection Cycle. Entry into the host cell is accomplished either by aphid feeding or mechanical inoculation. After uncoating, viral DNA in the nucleus has its gaps repaired by host mechanisms. The result of repair is a supercoiled CaMV minichromosome which is transcribed by host RNA polymerase II (Olszewski *et al.*, 1982). Two major polyadenylated RNA species are produced, both by transcription of the CaMV minus-strand [These early events are summarized in Hull and Covey, (1985)]. The smaller of these RNAs, the 19S, spans ORF VI and serves as a messenger RNA for its expression (Covey and Hull, 1981). The larger RNA, the 35S, encompasses the entire genome and its 5'-end maps to a location 100 bp downstream of ORF VI (Guilley *et al.*, 1982). The 3'-termini of both transcripts map to the same position 180 bp downstream of the 5'-end of the 35S RNA . In the 35S RNA this results in a 180 bp terminal redundancy. The function of the 35S RNA is discussed below.

CaMV is a "retroid element". This designation encompasses mammalian retroviruses and other agents which replicate their nucleic acid using reverse transcriptase. Specifically, CaMV DNA replication proceeds with reverse transcription of the 35S RNA by CaMV-encoded reverse transcriptase to produce a minus-strand DNA (see review by Hohn *et al.*, (1985)). Priming of reverse transcription is provided by the binding of the 3'

portion of a methionine initiator tRNA to a complementary sequence in the 35S RNA located 600 bp downstream of its 5' terminus (5' end of tRNA binding site corresponds to position 1). Reverse transcription proceeds counterclockwise to the 5'-end of the 35S RNA where the reverse transcriptase and the nascent minus-strand DNA must switch templates to the 3'-end of the same, or another, 35S RNA molecule. After this obligatory switch, reverse transcription proceeds full-circle to produce a full length, but covalently open, circular DNA strand. After, or concurrently with, minus-strand DNA synthesis, the RNA template is degraded (by an RNase H activity associated with CaMV reverse transcriptase) except for one or two short stretches which map in ORF V and, in some isolates, ORF II. The RNA in these regions of residual DNA/RNA basepairing serves as primer(s) for synthesis, by reverse transcriptase, of the plus-strand DNA. A second obligatory template switch occurs when reverse transcriptase, now synthesizing in the clockwise direction, encounters the gap located at position 1 in the covalent structure of the minus-strand. After this template switch, further plus-strand DNA synthesis completes the CaMV circle, terminating just downstream of the point(s) at which synthesis was primed. The net result is a CaMV DNA molecule with one or two site-specific gaps in the plus-strand DNA and one in the minus strand. This DNA is packaged in virions.

CaMV Translation. As mentioned, the 19S RNA serves as an mRNA for ORF VI. Translation of ORFs I through V is somewhat mysterious, but it is generally believed that these five ORFs are translated utilizing the 35S RNA as a polycistronic mRNA. In this unusual mode of eukaryotic translation, referred to as the "relay race", translation of the 35S proceeds from one ORF to the next in the 3' direction without dissociation of the ribosome from the template (Dixon *et al.*, 1983). Relay-race translation is consistent with, and dependent upon, the tight packing observed for the first five ORFs. Efficient expression of ORFs I to V is also dependent upon a trans-acting effect of the ORF VI product, probably the protein (Bonneville *et al.*, 1989; Gowda *et al.*, 1989). A large stem-loop structure, proposed for the untranslated, 5'-most 600 nucleotides of the 35S RNA,

may be important in determining the balance between translation and reverse transcription of this transcript (Fuetterer *et al.*, 1990).

The information above is intended as a source of background knowledge for those unfamiliar with CaMV molecular biology. The manuscript itself deals with the stability of DNA insertions in the CaMV IG. Background information specifically related to this topic is found in the introduction.

Acknowledgments

I am greatly indebted to the Biochemistry Department for providing me with the financial support and intellectual stimulation necessary to achieve my academic goals. Special thanks go to now-retired department head Dr. Roger Koeppe who offered frequent encouragement and insight and to the current head Dr. Jim Blair who also gave useful advice during the several months our tenures overlapped. I also thank the patient and helpful members of my committee: Dr. Margaret Essenberg, Dr. Richard Essenberg, Dr. Earl Mitchell, and Dr. David Meinke. I am most deeply indebted to my major advisor, Dr. Ulrich Melcher. His advice, encouragement, and friendship were indispensable in the completion of this project. Most importantly, he provided me with a model of what a true scientist should be.

I thank the other inhabitants of the south-end of the fifth floor for putting up with me for more than five years. Most especially I acknowledge the help and friendship of Kelly Hooper, Ann Williams, SueAnn Hudiberg, George Odell, Chuck Gardner, and, of course, my partner in misery, Steve Hartson. Sunnie, Robert, Gordon, Robin, Janet, Dawn, Cheryl, Edgar, Franklin, and others too numerous to list are also appreciated and will be remembered. A special thank you goes to Sue Heil for convincing the Macintosh to disgorge the final version of this thesis.

On a more personal note I would like to thank Bob and Carolyn Hunger for the constant friendship and support they gave to me and my family during my graduate career.

I also thank my wife's parents, Mr. and Mrs. Herbert Green for the many times they came to our aid. My deepest gratitude goes to my parents, Mr. and Mrs. W.E. Pennington. Their love and support has been constant through thick and thin and it is my most cherished hope that they are proud of what I have accomplished. I also thank God, who made all things (including me) possible.

Finally, I thank my loving and persevering wife Janice. It is to her, and our daughters, Katherine Lee and Andrea Jean, that this work is dedicated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

ABSTRACT

Cauliflower mosaic virus (CaMV) is an attractive vehicle for carrying foreign DNA into plants due to its double-stranded, infectious DNA genome. We have examined the insert-bearing potential of the large intergenic region (IG) of CaMV DNA by infecting turnips with a series of DNAs, derived from CaMV isolate CM4-184 and bearing inserts of various sequences and lengths at a unique *Eco*0109 restriction site. This restriction site resides in a region of the IG corresponding to a bulge in the secondary structure predicted for the 5' terminal 600 nucleotides of the 35S RNA, the larger of the two major viral transcripts. Fifteen different CaMV DNAs bearing inserts of up to 0.7 kilobasepairs in length were infectious to turnips producing, with one exception, symptoms indistinguishable from those induced by CM4-184. Insert stabilities were assessed using restriction analysis of viral DNAs recovered from infected plants. One 220 nucleotide insert was stable in one of two plants infected. All other inserts were unstable, undergoing partial or total deletion. The nucleotide sequences around the end-points of deletions were determined for clones of selected recovered viral DNA. Thirty-three such clones (representing 8 different insert deletions from 5 different inoculum clones) had deletions that were bounded by repeats or pseudorepeats residing within, or flanking, the insert. These included 16 examples with up to 46 nucleotides of CM4-184 IG sequence deleted. A relationship between deletion length and the length of repeats associated with deletion ends was noted. Eleven other clones (of progeny DNAs of two inoculum clones) had deletions whose ends were at potential RNA splice signals. These results suggest two mechanisms for deletion of inserts from the IG: 1) splicing of the 35S RNA prior to (-)

strand DNA synthesis from this template by CaMV reverse transcriptase and 2) homologous recombination at direct repeats, probably by template switching during reverse transcription of the 35S RNA.

CHAPTER II

INTRODUCTION

The nature of the cauliflower mosaic virus (CaMV) genome (double-stranded, infectious DNA) makes the virus an attractive vehicle for foreign DNA in plants (Dixon and Hohn, 1985; Gardner, 1983). Strategies for constructing CaMV DNAs carrying DNA inserts must be compatible with features or consequences of CaMV molecular biology. These include the function, alterability, and dispensability of various regions of the CaMV genome and the requirement that open reading frames (ORFs) VII, I, II, and III be tightly packed for the "relay race" mode of translation proposed for the viral mRNA (Dixon and Hohn, 1984; Sieg and Gronenborn, unpublished, cited in Dixon and Hohn, 1984). Of the CaMV ORFs, only II (coding for an aphid aquisition factor) (Armour *et al.*, 1983; Woolston, *et al.*, 1983) and VII (function unknown) (Dixon and Hohn, 1984) are not essential for infection of plants by mechanical inoculation. The dispensability of ORF II has been exploited in the construction of CaMV DNAs carrying inserts within (Gronenborn *et al.*, 1981), or in place of, it. Inserts replacing ORF II include fragments carrying a methotrexate-resistant dihydrofolate reductase (DHFR) gene from an *E. coli* plasmid (Brisson *et al.*, 1984), the coding region of the metallothionein II gene of Chinese hamster (Lefevre *et al.*, 1987), and the coding region of a human interferon gene (De Zoeten *et al.*, 1989). These chimeric CaMV DNAs infected turnips and their protein-encoding inserts were maintained stably and expressed in plants. Consistent with the "relay race" model, stability of the DHFR insert in plants was dependent upon the deletion of nearly all nucleotides between the DHFR ORF and the stop and start codons (respectively) of ORFs I and III. Kovgan (1989) introduced a fragment encoding a hepatitis B virus (HBV)

surface antigen in ORF VII. The resulting CaMV DNA directed HBV antigen synthesis in infected plants.

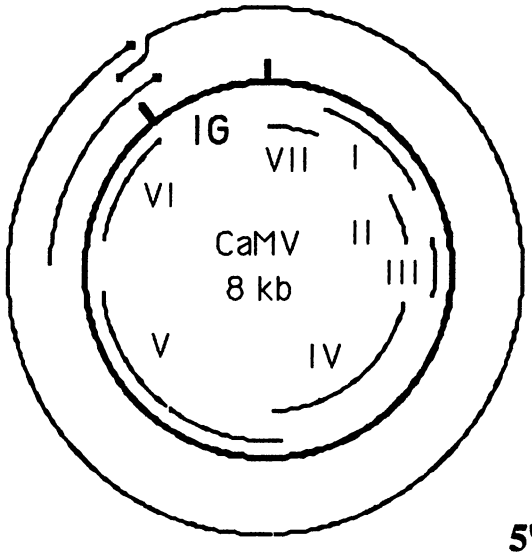
CaMV DNAs bearing short oligonucleotides in the large intergenic region (IG) are infectious and stable (Howell *et al.*, 1981; Dixon *et al.*, 1983; Daubert *et al.*, 1983). The *in planta* use of the IG as a region for the introduction of inserts has not been explored more thoroughly. This region consists of approximately 700 bp between ORFs VI and VII and contains signals necessary for the production and function of a 35S RNA. This transcript serves as template for reverse transcription and, probably, as polycistronic mRNA for the translation of ORFs VII and I through V (Bonneville *et al.*, 1989; Gowda *et al.*, 1989). Synthesis of the 35S RNA (Fig. 1) begins 97 nucleotides downstream of the ORF VI stop codon (position 7431) (Franck *et al.*, 1980) and terminates after transcription of the entire CaMV DNA. The polyadenylation signal terminating transcription is downstream of the transcription start site resulting in a terminal redundancy of 180 nucleotides. Several small ORFs (sORFs) within the IG and downstream of the transcription start site (Fuetterer *et al.*, 1988) are variably conserved in position and sequence between CaMV isolates. Some sORF-reporter gene fusions are translatable after electroporation into protoplasts. Subregions of the IG constitute distinct inhibitory and stimulatory elements in the RNA which may co-operate with cellular factors to regulate translation of downstream ORFs (Fuetterer *et al.*, 1990). The sORFs reside primarily within the inhibitory regions. The 5'-most 600 nucleotides of the 35S RNA have the potential to form a large stem-loop structure (Melcher, 1988; Fuetterer *et al.*, 1988). To examine the insert-bearing potential of the IG we inserted extra nucleotide sequences at a unique *Eco*0109 site in a region of DNA corresponding to a bulge in the hypothetical stem-loop structure. Analyses of viral DNA recovered from turnips infected with the resulting CaMV DNAs sheds light on IG insert maintenance, IG function, and the dynamics of sequence change in CaMV DNA.

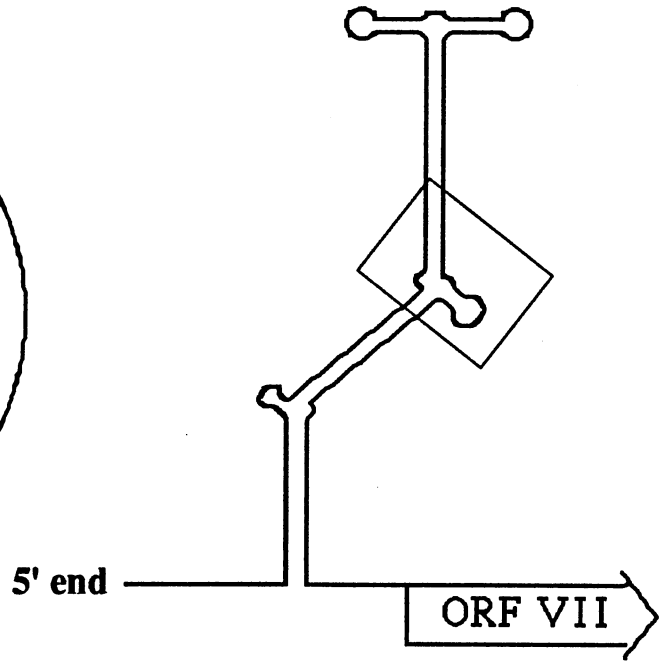
Figure 1. Map of CaMV Genome and Close-ups of Intergenic Region.

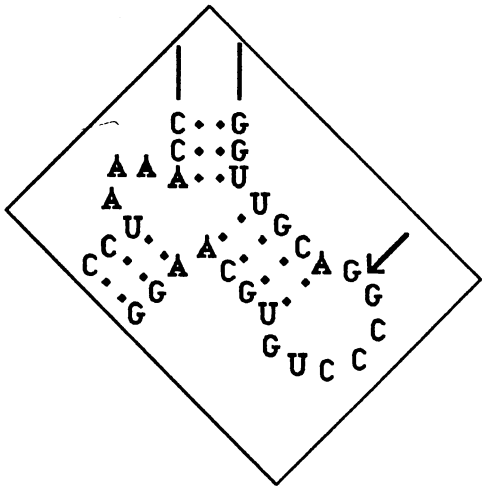
A. Map of CaMV genome. Arcs inside circle and identified with Roman numerals are open reading frames. Large and small concentric arcs outside circle represent 35S and 19S RNA transcripts, respectively. Large intergenic region (IG) lies between spikes pointed outward from circle.

B. Simplified diagram of secondary structure proposed for 5' end of 35S RNA (Fuetterer, *et al.*, 1988). Lines running parallel represent base-paired regions. Boxed-in region is bulge in structure and is shown in more detail in C.

C. Close-up view and nucleotide sequence of region boxed-in in B. Arrow indicates position in RNA sequence corresponding to location of unique *Eco*0109 site in CM4-184 DNA.







A **B** **C**

CHAPTER III

MATERIALS AND METHODS

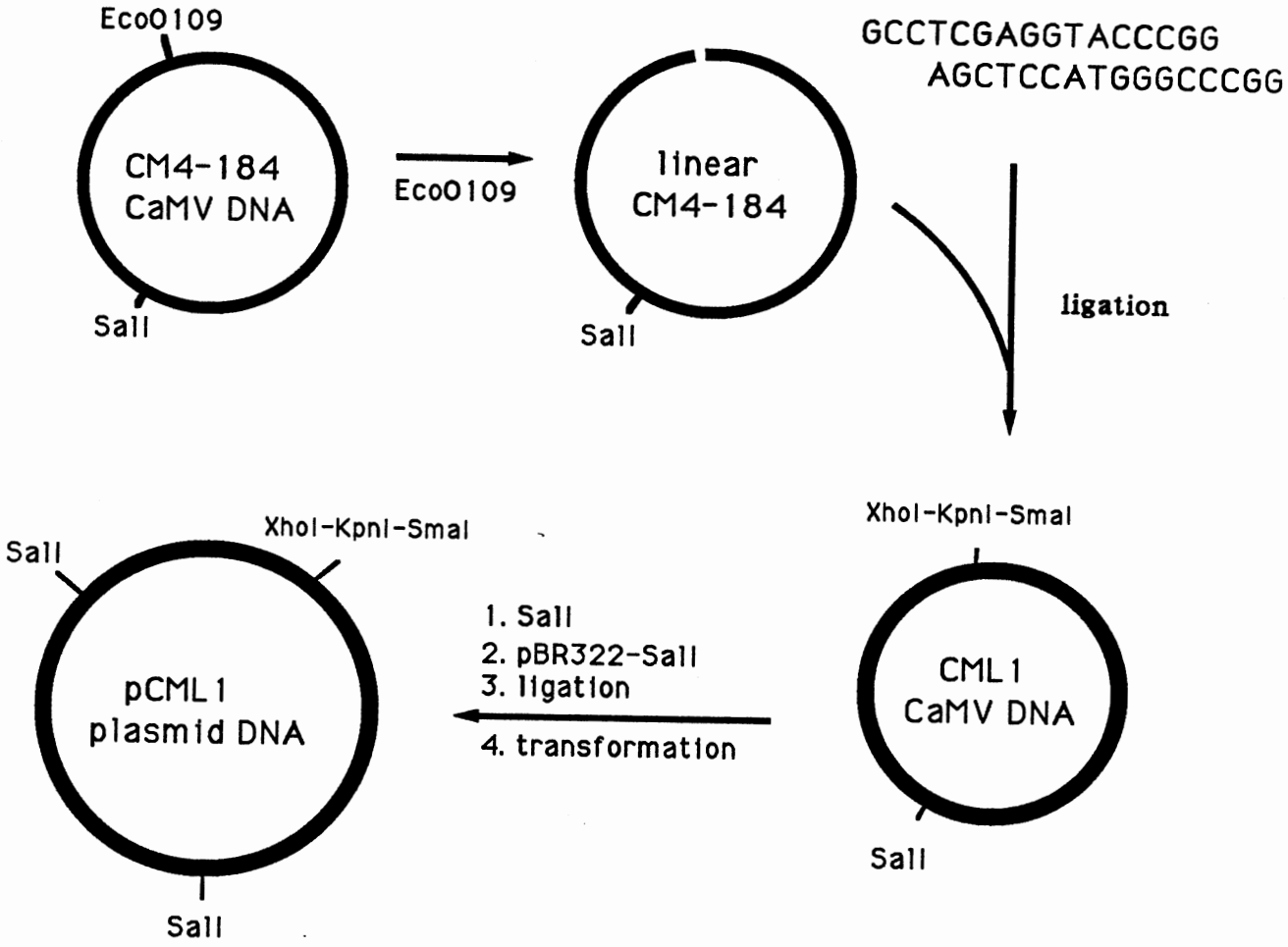
Inoculation of turnips and isolation of viral DNA

Turnips (*Brassica rapa* L., c.v. Just Right) were maintained and inoculated as previously described (Sun *et al.*, 1988). Inocula of cloned viral DNAs were prepared by separating CaMV DNA from plasmids by *Sal*I or *Xho*I digestion as appropriate (Choe *et al.*, 1985). For some experiments, inocula consisted of 0.2 g leaf tissue ground in 1% K_2HPO_4 with 0.4 g sea sand. Inoculated plants were observed over a 5 to 6 week period. Clones which failed to cause systemic disease on at least one of four (or more) inoculated plants by the end of the observation period were judged to be non-infectious. CM4-184 virions were isolated by the method of Hull *et al.* (1976) from turnips inoculated with *Sal*I-digested pLW414, a plasmid clone of CM4-184 (Howell *et al.*, 1981). Viral DNA was prepared from systemically infected leaves and CM4-184 virions as described by Gardner and Shepherd (1980).

Construction of pCML1

pCML1 was constructed as illustrated (Fig. 2). CM4-184 DNA was linearized by digestion at its unique *Eco*0109 (Dixon *et al.*, 1986) site. Linear CM4-184 DNA was recircularized by ligation in the presence of an oligonucleotide containing *Eco*0109 ends and restriction sites for *Xho*I, *Kpn*I, and *Sma*I enzymes. Ligation products were linearized with *Sal*I and ligated in the presence of *Sal*I-digested pBR322. Following transformation of bacteria with the final ligation product, we obtained a clone whose plasmid DNA

Figure 2. Construction of pCML1. Oligonucleotide linker (upper right) was ligated in presence of *Eco*0109-digested CM4-184 DNA giving CML1. *Sal*I-digested CML1 was ligated in presence of *Sal*I-digested pBR322 to give pCML1. Additional inserts were added to pCML1 at unique restriction sites contributed by the linker.


Eco0109
CM4-184
CaMV DNA
SalI
Eco0109
linear
CM4-184
SalI
GCCTCGAGGTACCCGG
AGCTCCATGGGCCCGG
ligation
XhoI-KpnI-SmaI
CML1
CaMV DNA
SalI
1. SalI
2. pBR322-SalI
3. ligation
4. transformation
SalI
XhoI-KpnI-SmaI
pCML1
plasmid DNA
SalI

contained a CaMV DNA and was digestible at single sites for *Xho*I, *Kpn*I, and *Sma*I . The sequence, location, and orientation of the inserted oligonucleotide were confirmed by limited nucleotide sequencing. pCML1 was used as a recipient for all IG insertions.

Construction of pCML1-Derivatives

pCaMTL(+) was constructed from *Xho*I-linearized pCML1 and a 204 bp *Xho*I-*Sal*I fragment of pCHMTII(2) (Lefebvre *et al.*, 1987) containing the Chinese hamster metallothionein II (MT) coding region. Gel electrophoretic analysis of *Xho*I-*Sal*I double digestion products established that the MT coding region was on the the CaMV (+) strand. We also obtained a plasmid with the insert in the opposite orientation, pCaMTL(-).

Derivatives of pCML1 bearing other fragments were constructed. pCc8C5 (Meshi *et al.*, 1981), a cDNA clone of the 3' end of Sunn-hemp mosaic virus (SHMV) (Kassanis and Varma, 1975) RNA in pBR322, was digested with *Dde*I. The fragment ends were made blunt by end-filling and ligated to phosphatase-treated, *Sma*I-digested pCML1. Following transformation, plasmids containing inserts were identified by restriction analysis and nine (Table I) were selected for further study. pCML5 was shown by limited nucleotide sequencing to contain 448 additional bp relative to pCML1, of which 283 derived from SHMV cDNA and the remainder from pBR322 DNA. All others contained single *Dde*I fragments (Table I).

Additional pCML1 derivatives (Table I) were constructed by ligating selected blunt-ended fragments of *Arabidopsis thaliana* alcohol dehydrogenase (ADH) gene DNA into *Sma*I cut and phosphatase-treated pCML1(Table I). A 3.6 kbp fragment containing the majority of the *Arabidopsis* ADH coding sequence was purified from *Hin*dIII digestion products of a lambda clone of the gene (At3101, graciously provided by E. Meyerowitz) (Chang and Meyerowitz, 1986). This fragment was subcloned in the *Hin*dIII site of pGEM3Z (Promega) giving pADH1. pADH5 and pADH6 were derived from pADH1 and

TABLE I

CONSTRUCTION AND INFECTIVITY OF pCML1-DERIVED PLASMIDS

plasmid	insert (orientation)[a]	Insert size	site[b]	infectivity[c]
pCaMTL(+)	metallothionein (+)	204 bp	*Xho*I	+(2-7)
pCaMTL(-)	metallothionein (-)	204 bp	"	+(2-7)
pCML3	pBR322 fragment (-)	412 bp	*Sma*I	+(14)
pCML4	pBR322 fragment (+)	546 bp	"	+(14)
pCML5	SHMV-pBR322 fragments(+,+)	448 bp	"	+(2-7)
pCML6	SHMV fragment (+)	260 bp	"	+(2-7)
pCML7	SHMV fragment (-)	388 bp	"	-
pCML8	SHMV coat protein (+)	647 bp	"	-
pCML10	pBR322 fragment (+)	543 bp	"	-
pCML12	pBR322 fragment (-)	169 bp	"	+(2-7)
pCML13	pBR322 fragment (-)	468 bp	"	+(14)
pRP1(+)	ADH Alu fragment 1 (+)	660 bp	*Sma*I	+(9)
pRP1(-)	ADH Alu fragment 1 (-)	660 bp	"	-
pRP5(-)	ADH Sty fragment 5 (-)	260 bp	"	+(2-8)
pRP5(2+)	ADH Sty fragment 5 (+,+)	540 bp	"	+(0-11)
pRP5(+)	ADH Sty fragment 5 (+)	260 bp	"	+(4)
pRP6(+)	ADH AluI fragment 6 (+)	398 bp	"	+(14)
pCaMTL(+)f	TCGA[d]	4 bp	*Xho*I	+(8)
pCML5d	-(GTAC)[e]	-4 bp	*Kpn*I	+(8)

a. Metallothionein fragment = *Xho*I-*Sal*I fragment of pCHMTII(2) (Lefebvre *et al.*, 1987) carrying the Chinese hamster metallothionein cDNA. pBR322 and SHMV fragments are polished *Dde*I fragments of pCc8C5 (Meshi *et al.*, 1981). SHMV coat protein=fragment of pCc8C5 carrying the coat protein coding region. ADH fragments as identified in figure 3. For all fragments (+) indicates insert and CaMV DNA are of the same polarity. (-) indicates insert and CaMV are of the opposite polarity. Polarity of CaMV, metallothionein, SHMV, and ADH sequences is defined by their protein coding regions. Polarity of pBR322 sequences is defined by numbering of the nucleotide sequence given by Suttcliff (1978). (+,+) = two (+) orientation fragments added. In pCML5, pBR322 fragment is downstream of SHMV fragment.
b. pCML1 restriction site to which fragment was added.
c. + = infectious. - = not infectious. Number of days delay relative to CM4-184 infection are given in parentheses.
d. Parent was pCaMTL(+)
e. Parent was pCML5

differ from it by the addition of a single nucleotide within the 270 bp *Sty*I fragment (pADH5) or the removal of this fragment (pADH6). The single nucleotide addition was achieved by digesting pADH1 at its unique *Tth*111I site, filling in the 5' overhangs, and religating. Removal of the *Sty*I fragment was achieved by digesting pADH1 with *Sty*I and ligating the large fragment. These modifications were confirmed by restriction analysis. pADH1 and pADH5 served as sources of *Alu*I fragments (Fig. 3) ligated into *Sma*I-digested pCML1 to give, respectively, pRP1(+), pRP1(-), pRP5(+), pRP5(2+), and pRP5(-). pADH6 served as source of a *Sty*I fragment (Fig. 3) whose ends were made blunt by end-filling prior to ligation into pCML1 to give pRP6(+).

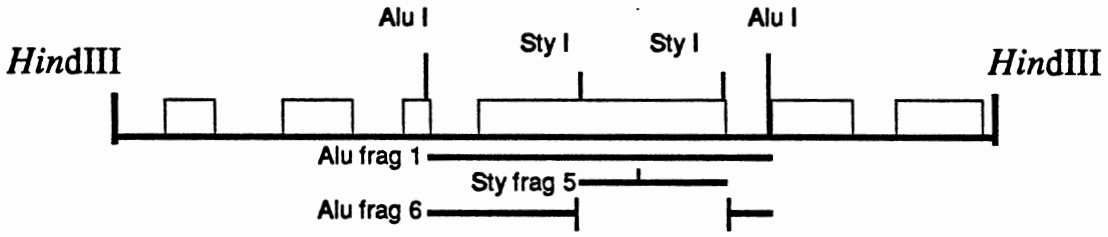
Construction of pCMTL(+)F and pCML5d

We constructed pCaMTL(+)F from pCaMTL(+) to bring the MT ORF in-frame with an upstream start codon. This was achieved by digesting pCaMTL(+) with *Xho*I and filling in the 5' overhangs of the resulting linear molecule prior to recircularization by ligation. pCML5d was created by digesting pCML5 with *Kpn*I and removing the 3' GTAC overhangs by incubation with T4 DNA polymerase and dATP, dGTP, TTP, and dCTP. Following ligation and transformation, clones were obtained and their plasmid DNAs analyzed by restriction analysis. A plasmid DNA, pCML5d, that was undigestible with *Kpn*I but whose *Eco*RI fragments were identical to those of pCML5 was identified. The deletion of the GTAC was confirmed by partial nucleotide sequencing.

Cloning and DNA Manipulations

Viral and plasmid DNAs were analyzed by digestion with restriction endonucleases (BRL, Promega) followed by agarose gel electrophoresis and ethidium bromide staining. Kilobase ladder (Promega) was used as DNA fragment size standard. Complete digestion of some viral DNAs required use of 100 to 400 µl reaction volumes to dilute apparent inhibitors of digestion. In such cases, digested DNA was precipitated with 95% EtOH in

Figure 3. Diagram of *Hin*dIII fragment of lambda At3101 (Chang and Meyerowitz, 1986) cloned into pGEM3Z to facilitate its use as a source of ADH fragments for pRP clones. Open boxes indicate ADH exons, lines in between, introns. Alu frag 1 = *Alu*I fragment of pADH1. Sty frag 5 = *Sty*I fragment of pADH5, identical to pADH1 except for 1 nucleotide addition at location indicated by vertical bar on fragment. Alu frag 6 = *Alu*I fragment of pADH6, identical to pADH1 except *Sty*I fragment has been deleted.


Alu I
Sty I
Sty I
Alu I
*Hin*dIII
*Hin*dIII
Alu frag 1
Sty frag 5
Alu frag 6

the presence of 1 μg/ml carrier tRNA and resuspended in a small volume of TEN buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 10 mM NaCl) prior to application to agarose gels.

For cloning recovered viral DNA, *SalI* or *XhoI* digested viral DNAs were ligated with T4 DNA ligase (BRL) in the presence of *XhoI* or *SalI* digested and phosphatase (calf intestine phosphatase, Promega) treated pGEM3Z (Promega) or *XhoI* digested and phosphatase treated BLUESCRIPT KS+ (Stratagene). These and other ligations were performed using guidelines suggested by Revie *et al.* (1988). Ligation products were used to transform competent *Escherichia coli* DH5α cells (BRL) as per the supplier's protocol. Transformed bacteria were spread on Luria-Bertani medium containing thiamine, ampicillin, 5-bromo-4-chloro-3-indolyl β-D galactopyranoside, and isopropyl-β-D-thiogalactopyranoside to select and identify viral DNA-containing plasmid clones.

Plasmid DNAs were prepared from liquid bacterial cultures by alkaline sodium dodecyl sulfate lysis (Ish-Horowicz and Burke, 1981). 5' End-filling reactions were performed by incubating DNA fragments with reverse transcriptase and dATP, dGTP, dCTP, and TTP as described by Smith and Calvo (1980). *KpnI* 3' overhangs were removed by addition of T4 DNA polymerase and the four deoxyribonucleotides directly to *KpnI* digests followed by 30 min. incubation at 37°C as described by Ausubel *et al.* (1989). Plasmid DNAs were sequenced as denatured, supercoiled DNA (Chen and Seeborg, 1985) using the T7 DNA polymerase (USB) sequencing protocol (Tabor and Richardson, 1987). For priming, we used either 77+, a 14-mer oligonucleotide which primes synthesis of sense-strand CaMV DNA from a site 134 nucleotides upstream of the insert region, or primers supplied by Promega. Other DNA manipulations not specifically described were carried out as suggested by suppliers or as described by Maniatis *et al.*. (1982). The "DNA Strider 1.1" (C. Mark and C.E.A. 1989) program was used on a Macintosh SE computer as an aid to sequence analysis.

Identification of Direct Nucleotide Sequence Repeats

Direct repeats were located in selected CaMV DNAs using dot-matrix homology comparisons of sequences surrounding 5' and 3' CM4-184/insert junctions. Analysis was performed using the Macintosh version of the "MOLGENJR" program (Lowe, 1986). The 5' junction region scanned consisted of 36 bp of CM4-184 sequence followed by 64 bp of insert sequence. The 3' junction region scanned consisted of 52 bp of insert sequence followed by 48 bp of CM4-184 sequence. Comparisons were made twice, once looking for 80% identities in 5 base windows and once looking for 100% identitities in 4 base windows. A visual examination of printed-out results of pairs of comparisons was used to identify all perfect repeats of 6 bp-or-more which could potentially bound deletions which, if they occured, would result in retention or loss of no more than 45 bp relative to CM4-184. Also identified were all perfect and imperfect repeats bounding actual deletions in clones of recovered DNA.

CHAPTER IV

RESULTS

Infectivity of CaMV DNAs With Inserts in the IG

pCML1 and 13 of 17 derivatives of pCML1 bearing 0.2 to 0.7 kbp inserts were infectious to at least one of four inoculated turnips (Table I). Three clones with 0.5 kbp-or-longer inserts [pCML8, pCML10, pRP1(-)] were not infectious. Also not infectious was a fourth clone (pCML7) whose 0.4 kbp insert contains a restriction site for *Sal*I, the enzyme used to separate CaMV from vector DNA in the inoculum preparation. Symptoms produced by infectious clones were indistinguishable from those caused by CM4-184 (Melcher, 1989), the isolate from which pCML1 was derived. Symptoms caused by pCML1 were not delayed relative to symptoms caused by pLW414 (Howell, *et al.*, 1980), the plasmid clone of CM4-184. With all of the pCML1-derivatives, however, onset of symptoms was delayed. Delays in symptoms with clones bearing inserts less than 0.3 kbp [pCaMTL(+), pCaMTL(-), pCML6, pCML12, pRP5(+), pRP5(-)] were 8 days or less. Delays with clones bearing 0.4 kbp-or-longer inserts were similar, longer, or highly variable. Symptoms with pCML5 were delayed 2-7 days. With RP1(+) the delay was 8-9 days. Symptoms with pCML3, pCML4, pCML13, and pRP6(+) were delayed 14 or more days. Delays in symptoms with pRP5(2+) varied between 0 and 11 days.

Insert Stability

pCML1

To test the stability of the 16 bp linker added to CM4-184, we isolated viral DNAs from CML1-infected plants and subjected them to digestion with *Xho*I, *Kpn*I, or *Sma*I. Viral DNA from two plants digested completely with these enzymes. *Xho*I digests of viral DNA from four other plants (from a separate experiment) were incomplete. To confirm the apparent loss of the *Xho*I site from some of the molecules in the populations that digested poorly with *Xho*I, we cloned and partially sequenced individuals. These DNAs were obtained from four different plants from two experiments. Two separately prepared pCML1 plasmid preparations had been used as inocula in these experiments. Seven clones were obtained by ligating *Sal*I digested viral DNA with *Sal*I-digested pGEM3Z. The nucleotide sequences of 6 of the DNAs were identical to CM4-184 in the region of the *Eco*0109 site, indicating precise deletion of the linker. The precise end-points of the deletion were ambiguous because they were within, or adjacent to, GGCC repeats in the CML1 sequence. Limited nucleotide sequencing of the seventh clone revealed a single change, a substitution that destroyed the *Kpn*I site (GGTACC changed to GATACC). Approximately 150 bp of sequence was determined for each of these clones. One further deviation from the expected CM4-184 sequence (Dixon *et al.*, 1986) was detected in one clone. The deviation was a C to T transition at position -72 relative to the 5' CM4-184/linker border.

Strategy for Determining Insert Stability

The *Eco*0109 site of CM4-184 DNA and, therefore, all insertions in the IG described here reside in an *Eco*RI fragment (*Eco*RI-B) common to all viral and cloned viral DNAs used in this study. Thus, the mobility of this fragment in agarose gel electrophoresis varies with the deletion or addition of sequence. We digested viral DNA

samples with *Eco*RI and compared mobilities of *Eco*RI-B fragments of viral DNAs to those of their parental DNAs (insert present) and pCML1 or CML1 (insert absent) to assess insert stability.

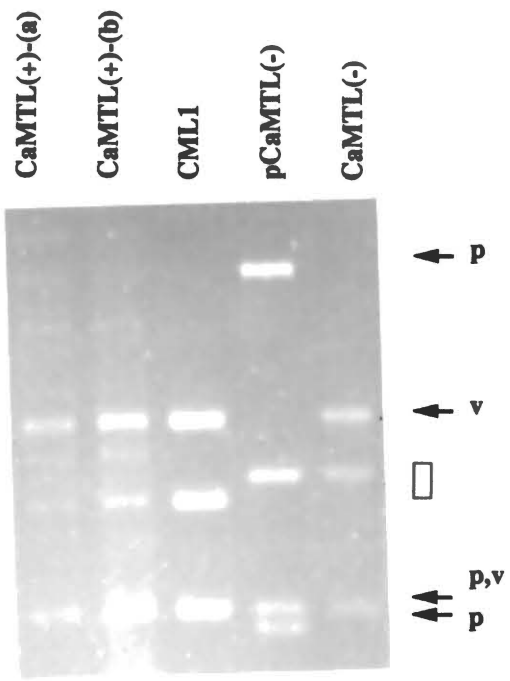
MT Insert

*Eco*RI-B fragments of viral DNA recovered from each of 2 plants (Fig. 4A) inoculated with CaMTL(+) (MT insert) had mobilities nearly identical to mobilities of the *Eco*RI-B fragment of CML1, indicating that this fragment of CaMTL(+) progeny had a deletion roughly the size of the insert. Identical results were obtained in a third plant (data not shown). In a repeat of this experiment, viral DNAs from each of four additional plants had *Eco*RI-B fragments whose mobilities were consistent with deletions of a similar size (Fig. 4B). However, three of the four plants also had a fragment of a size intermediate between those expected for complete retention and for complete loss of the MT insert. Eight cloned representatives of viral DNA pooled from the first three plants were obtained by ligating *Sal*I-digested viral DNA with *Sal*I-digested pGEM3Z. The partial nucleotide sequences of all 8 contained an identical deletion whose ends were 7 bp direct repeat sequences. The 5' end of one repeat was located at position +27 relative to the 5' CM4-184/linker border (5' border) and the other at position +3 relative to the 3' CM4-184/linker border (3' border) (Table II). The resulting viral DNA corresponds to that of the CM4-184 isolate with a 28 bp insert in the *Eco*0109 site. Approximately 110 nucleotides of sequence surrounding the deletion, determined for each clone, agreed with the expected sequence (Dixon *et al.*, 1986).

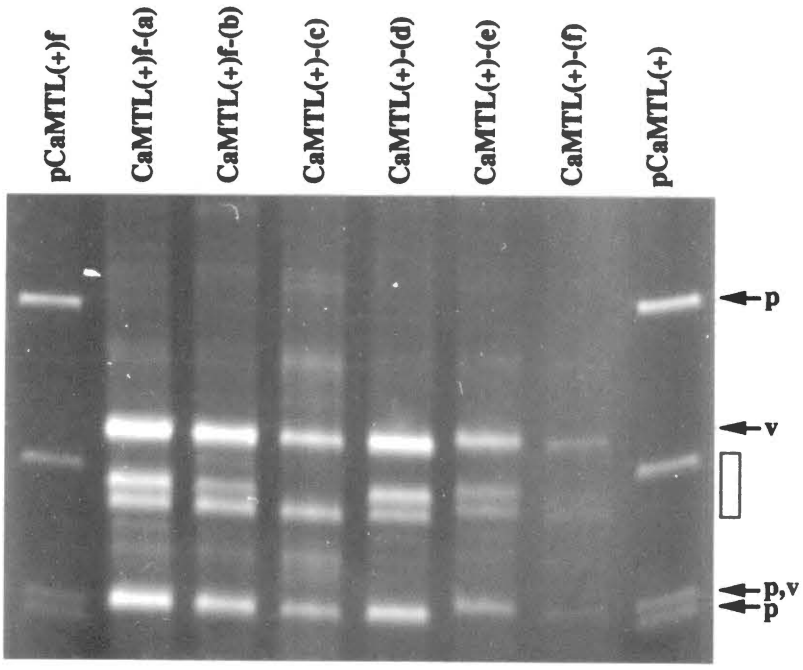
Figure 4. Agarose gel electrophoresis of *Eco*RI digests of viral DNAs recovered from plants infected with CaMTL(+), CaMTL(-), and CaMTL(+)f. "p" prefix in lane label indicates plasmid DNA. Lane labels without prefixes are viral DNA samples from plants inoculated with the corresponding plasmid DNA. Lower case letters in parentheses following identical viral DNA designations indicate samples recovered from separate plants. Arrows to right of gel indicate *Eco*RI fragments common to all construct plasmids (P), all viral DNA samples (V), or both (P,V). Vertical box to right of gel indicates range of mobilities for *Eco*RI-B bands.

A. CaMTL(+)-(a and b) and CaMTL(-) DNA.

B. CaMTL(+)f and CaMTL(+)-(c through f)



A


pCaMTL(+)f
CaMTL(+)f-(a)
CaMTL(+)f-(b)
CaMTL(+)-(c)
CaMTL(+)-(d)
CaMTL(+)-(e)
CaMTL(+)-(f)
pCaMTL(+)
p
v
p,v
p

B

TABLE II

ASSOCIATION OF SELECTED DIRECT REPEATS WITH REPEAT-BOUNDED INSERT DELETIONS

Repeat			Viral DNA					
	Location[b]		16bp	0.2 kb		0.4 - 0.7 kb		
Sequence[a]	5'	3'	CML1	CaMTL(+)	CaMTL(-)	CML3	CML5	RP1(+)
AGGTAAGACGATGGAA	-28	+11				6[c]		10[d]
CCTGTGC	+27	+3	X	8[e]	X	X	X	X
GAGGTAC	+6	+35	X	X	X			
TTGTTG	+37	-12	X	X	X	X		X
GATGGA	+38	+20	X		X	X	X	X
GAAGAC	-26	-26	X	X	X	X		X
GGAAAagTTTG	+48	+23	X	X	X	X	1[f]	X
TGGCAGGG	+15	-7	X	X	X	X	X	1[g]
GGCC	-1	-1	6					

Parental viral DNAs listed with approximate insert sizes above DNA designations. X indicates repeat absent. Numbers indicate number of progeny DNA clones containing deletions at specified repeats. For deletions, equivalent of one repeat deleted except as noted.

a. Upper case-nucleotides in both repeats, lower case - 5' repeat only, **bold face** - 3' repeat only.

b. location of 1st nucleotide of repeat relative to CM4-184/linker border. (-) - upstream of border. (+) - downstream of border.

c. in 2 clones, bold face G and T were deleted. Other 4 clones retained G and T.

d. 9 had T deleted, G retained. 1 had 3'-most 9 nucleotides of 5' repeat and 5'-most 13 nucleotides of 3' repeat deleted.

e. both repeats deleted

f. lower case 'a' retained, lower case 'g' deleted

g. bold face C deleted.

The complete nucleotide sequences of the insert-containing regions of pCML1, pCaMTL(+), pcaMTL(-), pCML3, pCML5, and pRP1(+) are given in appendix A.

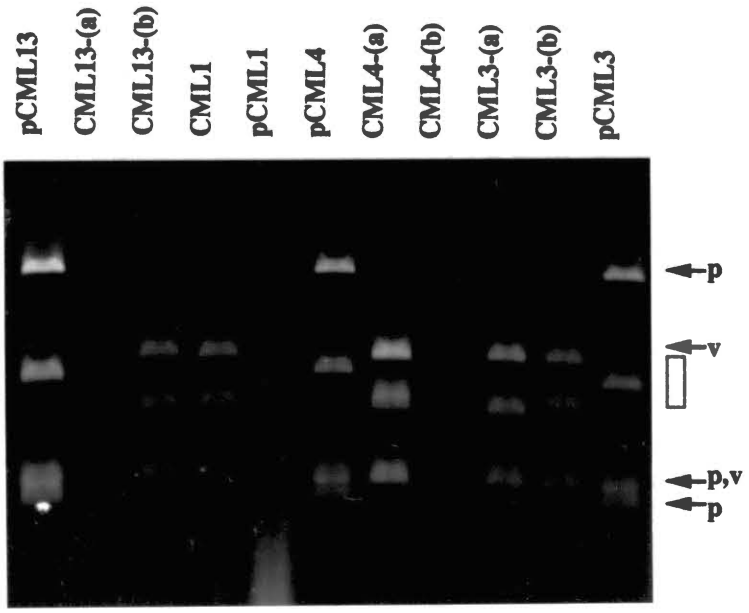
CML1-Derivatives with Other Large Inserts

The stability of pCML1-derivatives bearing other large inserts was examined by isolating and analyzing viral DNAs from plants infected with pCML1-derivatives bearing fragments of pCc8C5 or segments of ADH DNA. Viral DNAs recovered from plants infected with CML4 (Fig. 5A), RP5(2+), RP5(-) (Fig. 5D), RP5(+), RP6(+) (Fig. 5E), and one each of pairs of plants infected with CML3, CML13 (Fig. 5A) and RP1(+) (not shown) had *Eco*RI-B fragments which co-migrated with *Eco*RI-B fragments of pCML1 and CML1 DNA. The mobilities of these fragments were consistent with deletions equal in length to the inserts borne by the inoculum clones. Viral DNAs from the other plants infected with CML3, CML13 (Fig. 5A), and RP1(+) (Fig. 5D) had *Eco*RI-B fragments which migrated slightly faster than those of pCML1 or CML1. These mobilities indicated greater-than-insert-length deletions. Viral DNAs from one CML12-infected plant and two infected with CML6 (Fig. 5B) had *Eco*RI-B fragments which migrated between those of the respective inoculum clones and pCML1 indicating less-than-insert-length (partial) deletions. Viral DNA from a CML5-infected plant (Fig. 5C) had two *Eco*RI-B fragments migrating between those of the inoculum DNA and pCML1 indicating the presence of at least two partial deletions.

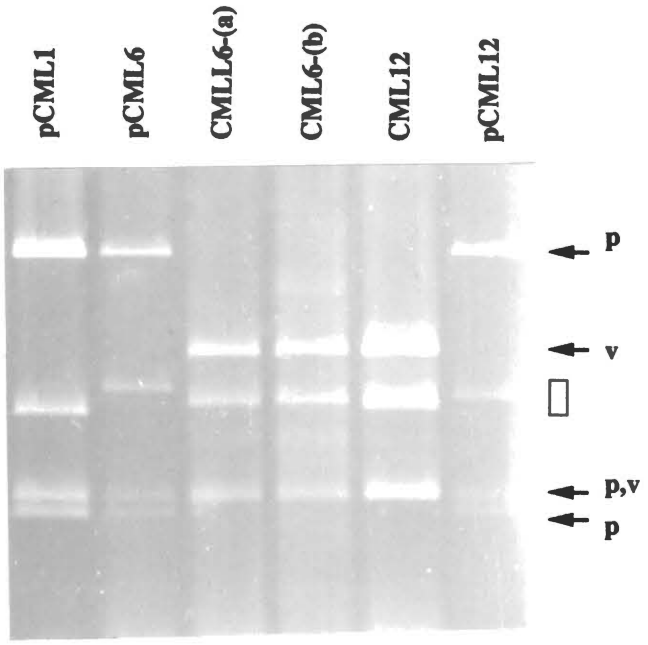
To examine more closely the deletion process, we cloned and partially sequenced viral DNAs from selected plants infected with CML3, CML5, and RP1. These DNA populations were chosen because they had multiple, partial, or greater-than-insert-length deletions. Viral DNA from the CML3-infected plant (greater-than-insert-length deletion) was digested with *Sal*I and cloned into pGEM3Z as described for CaMTL(+). Six clones were obtained. The partial sequences of four of these were consistent with a deletion extending from position -28 (5' border) to position +10 (3' border). The sequences of the other two were consistent with a similar deletion extending from position -20 (5' border) to position +20 (3' border). These deletions resulted in 38 and 40 bp deficits

Figure 5. Agarose Gel Electrophoresis of *Eco*RI Digests of Viral DNA Samples Recovered from Plants Infected with pCML1-Derivitives Bearing pBR322, SHMV, and ADH Fragments. See Figure 4 legend for details.

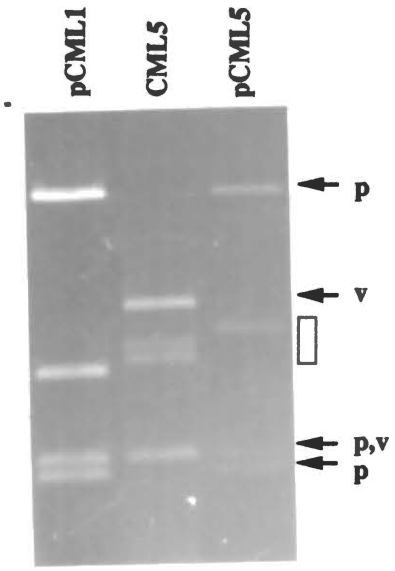
A. CML 3, CML4, and CML13
B. CML6 and CML12
C. CML5
D. RP1(+), RP5(-), and RP5(2+)
E. RP5(+) and RP6(+)


pCML13
CML13-(a)
CML13-(b)
CML1
pCML1
pCML4
CML4-(a)
CML4-(b)
CML3-(a)
CML3-(b)
pCML3
p
v
p,v
p

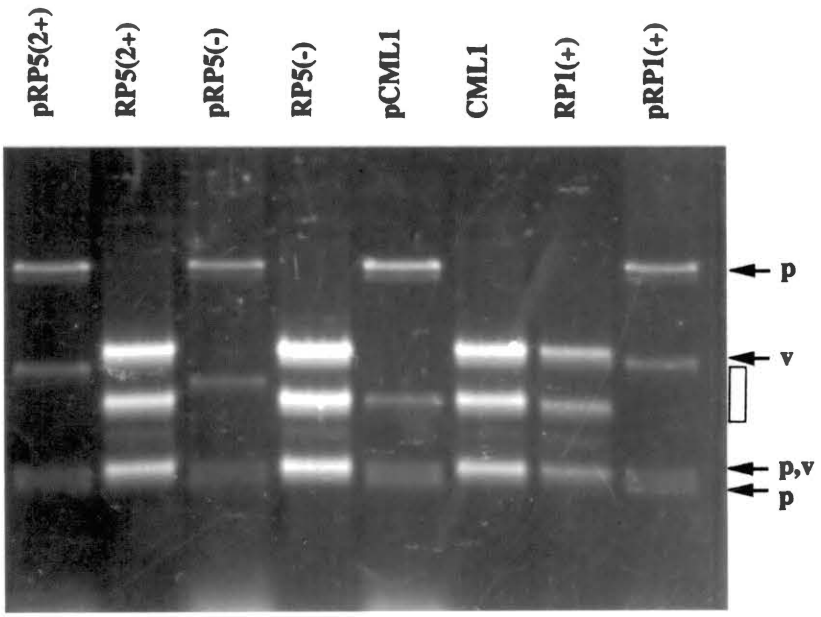
A



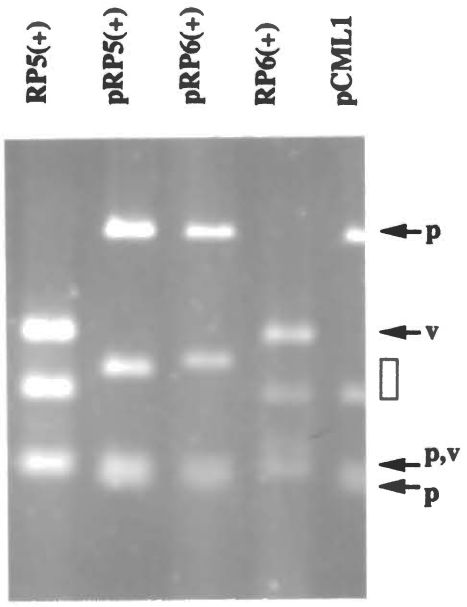
B



C



D



E

relative to the nucleotide sequence of CM4-184. As with the CML1 deletions, the precise assignment of deletion end-points was ambiguous because they were within, or adjacent to, direct repeats. The direct repeats responsible for the ambiguities resided within a set of larger imperfect repeats (pseudorepeats) consisting of a 14 nucleotide sequence upstream of the insert and one downstream of and identical to it but for the presence of two additional nucleotides (Table II). We obtained 9 clones of viral DNAs from an RP1-infected plant (greater-than-insert -length-deletion). Eight were obtained by *Sal*I site cloning as described for CaMTL(+) and CML3. The other was obtained by *Xho*I site cloning. The insertion-region sequences of the first 8 clones were identical and consistent with a deletion extending from position -25 (5' border) to position +14 (3' border) The ends of this deletion were within the pseudorepeat described above and resulted in a 39 bp deficit relative to the nucleotide sequence of CM4-184. The sequence of the *Xho*I site clone was consistent with its having a deletion extending from position +18 (5' border) to position -4 (3' border). These ends resided in a different, shorter pseudorepeat and resulted in a DNA equivalent to CM4-184 with a 20 bp insert in the *Eco*0109 site (Table II). The *Xho*I site used for cloning this DNA is included in the deleted portion of the progeny DNAs with 38 to 40 bp deficits relative to CM4-184. Three clones were obtained by *Sal*I site cloning of viral DNAs from the other RP1-infected plant. The sequence of the first of these had a deletion identical to that in the 8 *Sal*I site clones above. The sequence of the second was consistent with a deletion extending from position -23 (5' border) to position +23 (3' border). The ends were in the longer pseudorepeat described above and resulted in a 46 bp deficit relative to the nucleotide sequence of CM4-184. The sequence of the third clone contained a 99 bp deletion consistent with its extending from a GT dinucleotide residing within the *Kpn*I site (G at position +10) to an AG dinucleotide beginning at position +108 (both positions relative to 5' border). One hundred fifty bp of sequence was determined for clones of CML3 and RP1 DNAs (18 total clones). One deviation from the expected

sequence was found in a pGEM3Z clone of RP1 DNA. The deviation was a T to C transition at position +57 (3' border).

Eleven clones of viral DNAs from a CML5-infected plant (two distinct partial insert deletions) were obtained. These were obtained by cloning *Xho*I digested viral DNA in *Sal*I digested pGEM3Z. The sequences of the clones fell into three groups. Four had deletions of 175 bp consistent with their extending from the GT dinucleotide in the *Kpn*I site to an AG dinucleotide within the SHMV cDNA portion of the insert. Six had deletions of 328 bp consistent with their beginning at the same GT dinucleotide and extended to an AG dinucleotide 49 bp into the portion of the insert derived from pBR322. One had a 436 bp deletion extending from position +51 (5' border), which lies within the SHMV cDNA portion of the insert, to position +24 (3' border) thus effectively replacing a portion of CaMV DNA with SHMV sequences. The sizes of the two smaller deletions agreed closely with deletion sizes calculated from the *Eco*RI-B fragment mobilities observed in gel electrophoresis (Fig. 5C). Four additional clones of viral DNAs from the CML5-infected plant were obtained by *Sal*I site cloning into pGEM3Z. *Eco*RI digestion analysis of these clones showed them to fall into the three deletion size classes represented by the *Xho*I site clones (1 with the full length insert deletion, 2 with the larger partial insert deletion , and 1 with the smaller partial insert deletion).

Importance of Specific Sequence Motifs

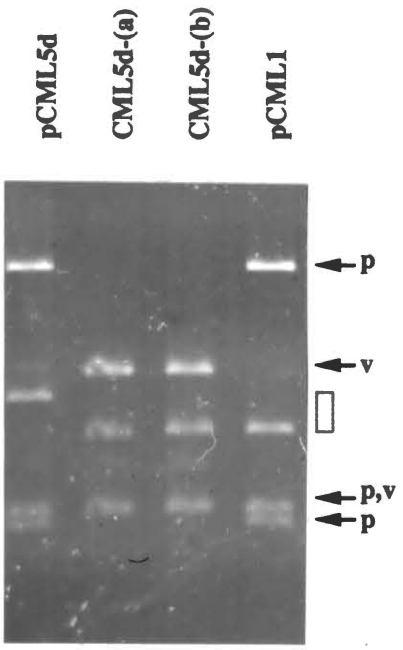
Potential translation of the MT coding region of CaMTL(+) may be bypassed by an overlapping reading frame originating at an upstream, and out-of-frame, start codon. We decided to test whether instability of the insert was due to this bypassing of the coding region. Accordingly we inoculated plants with a clone identical to pCaMTL(+) except that 4 bp had been added to the 5' region of the insert. This addition brought the MT coding region into frame with the upstream start codon. The frameshifted clone, pCaMTL(+)f, was infectious and caused symptoms indistinguishable from those caused by pCaMTL(+).

Viral DNAs from each of two infected plants contained a pair of *Eco*RI-B fragments (Fig. 4B). These fragments migrated faster than the *Eco*RI-B fragment of pCaMTL(+)f and identically with *Eco*RI-B fragments of viral DNAs from three CaMTL(+)-infected plants indicating that the frameshift did not impart stability.

The presence of a common GT dinucleotide and different AG dinucleotides at the ends of deletions in CML5 progeny suggests that deletion had occurred by RNA splicing. If splicing was responsible, then removal of the GT dinucleotide should block the occurrence of these deletions. Therefore, we inoculated plants with a clone identical to pCML5 but for the deletion of four nucleotides, including the GT. The clone, pCML5d, was infectious and caused symptoms indistinguishable from those caused by CML5. *Eco*RI-B fragments of viral DNAs recovered from each of two CML5d-infected plants migrated indistinguishably from the *Eco*RI-B fragment of pCML1, indicating total insert deletion (Fig. 6). The two partial deletion products, observed with viral DNA from the CML5-infected plant (figure 4F) and proposed to be due to RNA splicing, were not seen from CML5d-infected plants.

To further examine the role of direct repeats in deletion we used dot matrix homology plots to identify all direct repeats in the CM4-184/insert border regions of selected inoculum clones (pCML1, pCaMTL(+), pCaMTL(-), pCML3, pCML5, pRP1(+). The complete nucleotide sequences of the insert-containing regions of these clones are listed in appendix A.). All of the inoculum clones contained the 4 bp repeats associated with the CML1 deletions and the long pseudorepeats associated with greater-than-insert-length deletions in CML3 and RP1(+) (Table II). Other direct repeats including a 7 bp repeat found only in the *Sma*I site constructs (*eg.*, pCML3, pCML5, pRP1(+)) and a 7 bp repeat unique to pCaMTL(+) and associated with deletion of the MT insert were construct-specific. pCaMTL(-), identical to pCaMTL(+) except that the MT insert was in the non-sense orientation, was found to be unique in its lack of these, and similar, construct-specific repeats (Table II). To test whether the relative lack of direct repeats would lead to

Figure 6. Agarose Gel Electrophoresis of *Eco*RI Digests of Viral DNA Samples Recovered from CML5d-Infected Plants. (See Figure 4 legend for details.)


pCML5d
CML5d-(a)
CML5d-(b)
pCML1
p
v
p,v
p

insert stability, we inoculated plants with this clone. pCaMTL(-) was infectious producing symptoms indistinguishible from those caused by pCaMTL(+). Viral DNA isolated from one CaMTL(-)-infected plant had a single *Eco*RI-B fragment which co-migrated with the *Eco*RI-B fragment of pCaMTL(+) (Fig. 4A). This co-migration indicated that the non-sense orientation MT insert was stable. Analysis of viral DNA from a second plant (data not shown) was consistent with the full deletion of the insert.

Competition and Further Deletion

Competition occurs between isolates of a virus co-inoculated onto a plant (Zhang and Melcher, 1989). We wanted to assess the relative importance for stability of progeny DNA populations of competition and further deletion. Our first approach was to examine viral DNAs recovered from plants inoculated with tissue from a CML5-infected plant (passsaging). Passaging resulted in infection without delay. Viral DNA recovered from these plants had heterogeneously sized*Eco*RI-B fragments ranging from slightly larger than the *Eco*RI-B fragment of pCML1 to slightly smaller than the smaller *Eco*RI-B fragment seen in digests of CML5 (Fig. 7). The size range of the fragments was consistent with their being the result of multiplication of minor components in the original mixture including, perhaps, the cloned progeny DNA shown to have a near-total insert deletion. Next, cloned viral DNAs with examples of each of the three size classes of CML5 deletions, obtained by *Sal*I site cloning, were inoculated on turnip plants to examine further deletion. Three clones, one representing the CML5 progeny with the largest deletion and two representing the progeny with the smallest one, were infectious. A progeny clone with the intermediate sized deletion was not infectious. As expected, viral DNA from a plant inoculated with cloned DNA with the largest deletion (pCML5pL), did not undergo any further deletions detectable by gel electrophoresis (Fig. 8). The two cloned DNAs with the smallest deletion (pCML5pS1 and pCML5pS2) produced variable results when inoculated on four (total) plants. From one plant, the *Eco*RI-B fragments recovered

Figure 7. Agarose Gel Electrophoresis of *Eco*RI Digests of a Passaged Viral DNA Sample. CML5-2nd Indicates Viral DNA Isolated from Plants Inoculated with Leaf Tissue Homogenate from a CML5-Infected Plant. Viral DNA from the CML5-infected plant shown for comparison. Arrows and labels are as described in Fig. 4 legend.

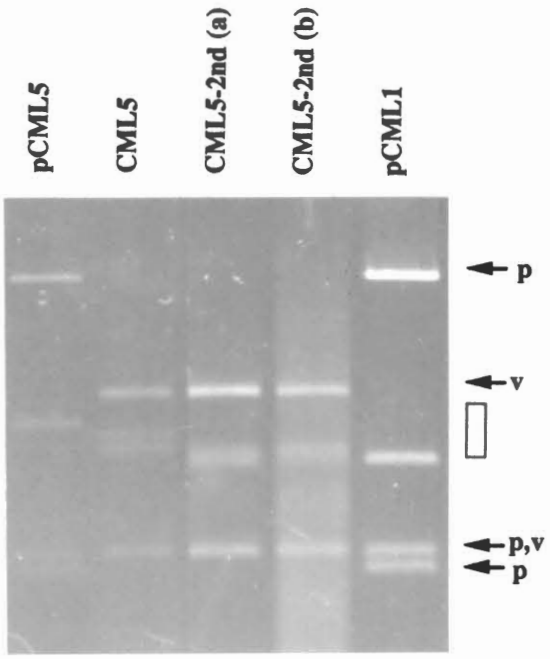
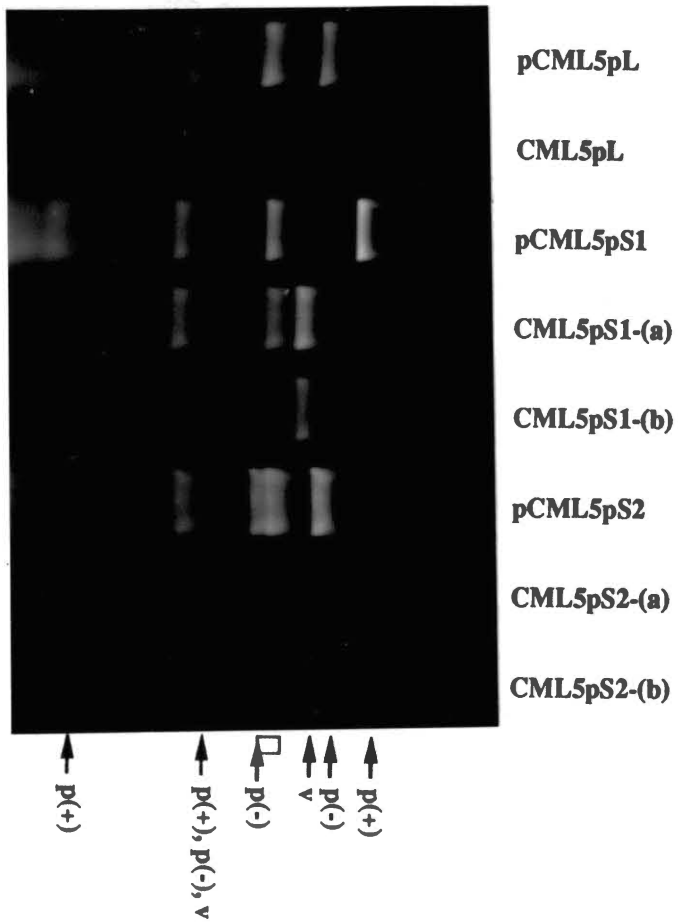

pCML5
CML5
CML5-2nd (a)
CML5-2nd (b)
pCML1
p
v
p,v
p

Figure 8. Agarose Gel Electrophoresis of *Eco*RI Digests of Viral DNAs Recovered from Plants Inoculated with Clones of CML5-Progeny. Due to use of a different plasmid vector, analysis of plasmid lanes differs from those in previous figures. Viral DNA sequences in pCML5pL and pCMLpS2 are reversed relative to vector sequences. Arrows labeled with (P(-)) indicate bands common to these plasmids. Viral DNA in pCML5pS1 is in same orientation as vector [arrows labeled with (P(+))] Arrow labeled (P(-), P(+), V) indicates band common to all plasmid and viral DNA. Other labels are as described in Fig. 4 legend.


pCML5pL
CML5pL
pCML5pS1
CML5pS1-(a)
CML5pS1-(b)
pCML5pS2
CML5pS2-(a)
CML5pS2-(b)
p(+)
p(+), p(-), v
p(-)
v
p(-)
p(+)

were primarily of the size expected from near-complete insert deletion (Fig. 8, CML5pS2-(a)). From another plant, the near-complete insert-deletion fragment was only faintly visible. Instead, the major band corresponded in size to the fragment in the inoculum. Also present was a size-heterogeneous collection of somewhat smaller fragments (Fig. 8, CML5pS2-(b)). In the remaining two plants, the major fragment recovered was of the same size as that in the inoculum. A near-complete insert deletion band was faintly visible but the heterogeneously sized fragments were not. (Fig. 8, CML5pS1-(a) and (b)).

CHAPTER V

DISCUSSION

Addition of large inserts to the unique *Eco*0109 site in the IG of CaMV isolate CM4-184 DNA did not abolish infectivity of the viral DNA or noticeably alter the type of symptoms elicited on turnips (Table I). The largest insertion resulting in an infectious DNA was nearly 0.7 kbp long, greatly exceeding the 8 to 12 bp sizes of previously reported non-lethal insertions in other IG sites (Howell *et al.*, 1981; Dixon *et al.*, 1983; Daubert *et al.*, 1983). The failure of pCML7 (0.4 kbp insert) to produce infection is explainable by the presence in the insert of a *Sal*I site which resulted in the viral DNA being cleaved during inoculum preparation. The other non-infectious clones (pCML8, pCML10, and pRP1(-)) and those producing symptoms after the longest (2 weeks) delays (pCML3, pCML4, and pCML13, pRP6(+)) had inserts of 0.4 kbp or longer. However, pRP1(+) (660 bp insert), pCML5, and pRP5(+) (both with 540 bp inserts) were infectious with moderate or highly variable delays. Thus, some factor(s) other than, or in addition to, insertion length must account for lethality or symptom delays.

Although insertions in the IG site were non-lethal, they were also generally unstable. This instability provides a possible explanation for the similarity in symptoms between the parental and insert-bearing inoculum DNAs. Specifically, it is probable that the symptoms were caused by insert-lacking viral DNAs resulting from deletion in the plant. Deletion cannot, however, explain symptoms produced by CaMTL(-) whose 0.2 kbp insert was stable in one plant (Fig 4A) but deleted in another. In both plants the symptoms were indistinguishable from those elicited by CM4-184. Thus, the site used by us for insert addition, and perhaps the portion of the IG containing this site, probably does

not influence the type of symptoms observed. This is not surprising as symptom determinants have been located elsewhere, principally ORF VI (Daubert, 1988).

Stability is not an inherent insert property. That context is critical is illustrated by the stability of the metallothionein insert in the ORF II site (Lefebvre *et al.*, 1987) and its instability in the IG site (Fig 4A and B). Translation reading frame considerations do not affect the viability or stability of viral DNAs with insertions or deletions in the region of the IG containing the *Eco*0109 site. Viability of DNAs with insertions in other IG sites have also been shown to be reading frame independent (Howell *et al.*, 1981; Dixon *et al.*, 1983; Daubert *et al.*, 1983). This contrasts sharply with translation-related constraints imposed upon insertions and deletions in other regions of the DNA. Short insertions in ORFs II, VI (Daubert *et al.*, 1983), and III (Melcher *et al.*, 1986) were non-lethal only if they did not frameshift the ORF. An out-of-frame ORFII deletion reverted to in-frame by second site mutation (Armour *et al.*, 1983; Melcher *et al.*, 1986). Deletion progeny of an ORF II-insertion construct were stable only if the deletion allowed translation of ORF II to procede in-frame uninterrupted by stop codons into ORF III [K. Sieg and B. Gronenborn, unpublished, cited in Dixon and Hohn, (1985)]. In contrast, we found constructs containing insertions which either maintain the reading frame of the small ORF containing the insertion site or frameshift it +1 or -1 to be infectious. The sense-orientation metallothionein insert was deleted regardless of whether the metallothionein coding region was out-of-frame with the afore-mentioned small ORF (CaMTL(+), Fig.4A), or in frame with it (CaMTL(+)f, Fig. 4B). Further, we found examples of progeny DNAs with 38, 39, and 40 bp deficits (from RP1 and CML3 infection, Table II) relative to CM4-184. These findings indicate that viability and stability of viral DNAs with modifications in the region surrounding the *Eco*0109 site are not reading frame-related.

Mechanisms of Deletion

One progeny DNA from an RP1(+)-infected plant had a partial insert deletion terminating at an AG dinucleotide within the ADH gene fragment. This AG dinucleotide is a splice acceptor utilized in the maturation of ADH mRNA in *Arabidopsis* (Chang and Meyerowitz, 1986, Fig. 3). The 5' end of this deletion was a GT dinucleotide which, in RP1 and CML5, resides within an AGGTA pentanucleotide. This GT is a potential splice donor as the pentanucleotide is the core of the consensus 5' exon-intron splice junction of plant introns [(C/A)GGTAAGT)] (Brown, 1986). Ten of 11 progeny DNAs from a CML5-infected plant had partial insert deletions with this same GT dinucleotide at the 5' end and two different potential splice acceptor AG dinucleotides at the 3' end. The presence of potential splice signals and, in RP1(+), an actual one suggested that these deletions occurred by splicing of the 35S RNA. Additional, but indirect, evidence for the splicing reaction causing CML5 deletions was obtained by analyzing viral DNAs from CML5d-infected plants. CML5d is identical to CML5 except that it lacks the potential 5' splice donor associated with the two supposed splicing events. Two progeny DNAs (detected by *Eco*RI analysis of DNA from a CML5-infected DNA, Fig. 5C) with deletions identical in size with the two putative splice-related deletions were not detected in CML5d-infected plants (Fig. 6). Instead, *Eco*RI analysis revealed a full insert-length deletion product, perhaps identical or similar to a minor progeny species cloned from the CML5-infected plant and not due to splicing. Deletion by splicing, previously suggested (Vaden and Melcher, 1990; Melcher, *et al.*, 1986) or reported in CaMV (Hohn *et al.*, 1986; Hirochika *et al.*, 1985), occurs before reverse transcription of the 35S RNA template and results in the (-)strand DNA, and later the mature double-stranded DNA genome, lacking the spliced-out sequence.

Progeny clones containing deletions not suspected of being splice related, including the clone of the minor species from the CML5-infected plant, were missing the entire insert

plus or minus 45 bp. Further, the deletion end-points were in, or adjacent to, direct nucleotide sequence repeats (Table II) suggesting deletion by homologous recombination. It is possible that the putative recombinational events responsible for the repeat bounded-deletions involved template switching, between direct repeats, by reverse transcriptase. This mechanism ("copy-choice recombination") has been proposed to be responsible for other CaMV deletions (Dixon and Hohn, 1984; Grimsley *et al.*, 1986) and is similar to the obligatory strand switch made when reverse transcriptase reaches the 5' end of the 35S RNA (Grimsley *et al.*, 1986). It is intriguing that all of the near-total, total, and greater-than-insert-length insert deletions found in progeny DNA clones were bounded by direct repeats, whereas the partial deletions were probably due to splicing. If the 5'-most 600 nucleotides of the 35S RNA exist as the stem-loop structure proposed by Fuetterer *et al.*(1988) the repeated sequences associated with the deletions described here would be located near, and on either side of, the neck of the bulge (Fig. 2). The resulting close proximity of the repeats to each other might facilitate full insert or greater-than-insert length deletions (but not partial insert deletions) by copy-choice recombination. On the other hand, the precise crossover points of putative recombinational events responsible for the 8 CaMTL(+) deletions (both repeats deleted, Table II) and the largest RP1(+) deletion (most of both repeats deleted) are difficult to reconcile with template switching.

We observed examples of plant to plant (CaMTL(+), Fig. 4A; CML3, CML13, Fig. 5A), or within-plant (CaMTL(+), CaMTL(+)f, Fig. 4B; CML5, Fig. 5C and Table II; RP1(+), Table II), variability in deletion. The most striking example was the CML5-infected plant which contained more than one partial deletion product and a minor product with a near-complete insert deletion. Competition in the plant tends to favor components of a viral DNA population most closely resembling the parental sequence (Dixon and Hohn, 1985), in this case CM4-184. Thus, the presence of multiple deletion products suggests that these progeny were produced early in infection, perhaps in the initially infected cell or cells. Then, perhaps, the progeny spread systemically without undergoing the same level

of competition experienced by viral DNAs inoculated onto the plant as a mixture. If the progeny DNAs do not undergo extensive competition, minor, but more competitive components, of the original population would be expected to become prominent in subsequent infections after passaging. We tested this by passaging the viral DNA from the CML5 -infected plant to two subsequent plants. Unlike pCML5-inoculated plants, these plants developed symptoms without delay, consistent with replication of minor components more similar to the parental DNA. Viral DNAs from these plants, as shown by *Eco*RI analysis (Fig. 7), contained a mixture of deletions, all larger than those in the two major deletion progeny in the original population. These mixtures contained a major component with a total, or near total, insert deletion such as was found in the clone of the minor component of the original population. We also observed that one of the partial deletion products cloned from the CML5 infected plant (pCML5pS2) was stable in one out of two plants inoculated singly with this DNA (Fig. 8). In two plants inoculated with a second clone of this, or a similar, deletion product (pCML5pS1) further deletion had taken place but the inoculum DNA species (no further deletion) was clearly present (Fig. 8). Taken together, these results suggest that competition, rather than further deletion, is the primary explanation for the predominance after passaging of progeny containing total or near-total insert deletions in the viral DNA. This is consistent with progeny DNAs being produced early, rather than late, in infection.

It has previously been shown that recombination frequency increases with repeat length (Dixon and Hohn, 1985; Pietrzak and Hohn, 1985). Our results, in conjunction with results of previous studies (Dixon and Hohn, 1985), indicate that the relationship between deletion frequency and repeat length may also be linked to the size of the potential deletion. All of the construct clones in this study contain direct repeats and pseudorepeats near the ends of the inserts. While some repeats and pseudorepeats, such as those associated with all of the CaMTL(+) and one each of the CML5 and RP1 deletions, are construct-specific others are common to all of the inoculum constructs (Table II). These

include the 4 bp repeat associated with CML1 deletions and the 14 bp pseudorepeat associated with deletions found in cloned progeny of CML3 and RP1(+). Theoretically, deletion between the 4 bp repeats provides a means by which any of the insert-bearing constructs could revert to the parental sequence (CM4-184). However, our limited survey of individual progeny DNAs uncovered no examples of this potentially advantageous deletion having occurred except with CML1 which has only a 16 nucleotide insertion (Table II). A 16 nucleotide deletion bounded by 4 bp repeats is not surprising, as 19 to 64 bp deletions between 2 and 3 bp repeats have been noted previously (Dixon and Hohn, 1985). In contrast, 25 repeat-bounded deletions of 0.2 kbp or more examined occurred between 7 bp-or-longer repeats (Table II). Further, 15 out of 17 0.4 kb or greater deletions, but none of the shorter ones, had ends within the 14 bp pseudorepeat. These 15 deletions resulted in viral DNAs with 38 to 46 bp deficits relative to CM4-184 DNA. Clones of these DNAs are poorly infectious at best (data not shown). These findings suggest that deletion by homologous recombination between 2 to 4 bp repeats is rarer for long sequences (approximately. 0.2 kbp or more) than for short ones (approximately 0.1 kbp or less).

In summary, our results indicate that the portion of the IG containing the *Eco*0109 site is poorly suited for stably bearing inserts. Variable stability may exist for insert-bearing constructs which lack 6 bp-or-longer repeats very close to insert ends (CaMTL(-), Fig. 4A) or for deletion progeny viral DNAs retaining only 0.1 to 0.3 kbp of insert DNA (clones of CML5 progeny, Fig. 8). However, the apparent dispensability of the sequences flanking the *Eco*O109 site and the presence within them of long repeated sequences makes stability of insertions unlikely.

REFERENCES

Armour, S. L., Melcher, U., Pirone, T. P., Lyttle, D. J., and Essenberg, R. C. (1983). Helper component for aphid transmission encoded by region II of cauliflower mosaic virus DNA. *Virology* **129,** 25-30.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1989). Current Protocols in Molecular Biology. John Wiley & Sons, New York

Bonneville, J. M., Sanfançon, H., Fuetterer, J., and Hohn, T. (1989). Posttranscriptional trans-activation in cauliflower mosaic virus. *Cell* **59,** 1135-1143.

Brisson, N., Paszkowski, J., Penswick, J. R., Gronenborn, B., Potrykus, I., and T., H. (1984). Expression of a bacterial gene in plants by using a viral vector. *Nature (London)* **310,** 511-514.

Brown, J. W. S. (1986). A catalogue of splice junction and putative branch point sequences from plant introns. *Nucl. Acids Res.* **14,** 9549-9559.

Chang, C., Meyerowitz, E. (1986). Molecular cloning and DNA sequencing of the *Arabidopsis thaliana* alcohol degydrogenase gene. *Proc. Natl. Acad. Sci. USA* **83,** 1408-1412.

Chen, E. Y., and Seeborg, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4,** 165-170.

Choe, I. S., Melcher, U., Richards, K., Lebeurier, G., and Essenberg, R. C. (1985). Recombination between mutant cauliflower mosaic virus DNAs. *Plant Mol. Biol.* **5,** 281-289.

Covey, S. N., and Hull, R. (1981). Transcription of cauliflower mosaic virus DNA. Detection of transcripts, properties and location of the gene encoding the virus inclusion body protein. *Virology* **111,** 463-474.

Daubert, S. (1988). Sequence determinants of symptoms in the genomes of plant viruses, viroids, and satellites. *Mol. Plant-Microbe Interactions* **1,** 317-325.

Daubert, S., Richins, R., Shepherd, R. J., and Gardner, R. C. (1982). Mapping of the coat protein gene of cauliflower mosaic virus by its expression in a prokaryotic system. *Virology* **122,** 444-449.

Daubert, S., Shepherd, R. J., and Gardner, R. C. (1983). Insertional mutagenesis of the cauliflower mosaic virus genome. *Gene* **25,** 201-208.

De Zoeten, G. A., Penswick, J. R., Horisberger, M. A., Ahl, P., Schultze, M., and Hohn, T. (1989). The expression, localization, and effect of a human interferon in plants. *Virology* **172,** 213-222.

Dixon, L., Nyffenegger, T., Delley, G., Martinez-Izquierdo, J., and Hohn, T. (1986). Evidence for replicative recombination in cauliflower mosaic virus. *Virology* **150,** 463-468.

Dixon, L. K., and Hohn, T. (1984). Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. *EMBO J.* **3,** 2731-2736.

Dixon, L. K., and Hohn, T. (1985). Cloning and manipulating cauliflower mosaic virus. *Dev. Mol. Virol.* **5,** 247-275.

Dixon, L. K., Koenig, I., and Hohn, T. (1983). Mutagenesis of cauliflower mosaic virus. *Gene* **25,** 189-199.

Franck, A., Guilley, H., Jonard, G., Richards, K., and Hirth, L. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21,** 285-294.

Fuetterer, J., Gordon, K., Bonneville, J. M., Sanfaçon, H., Pisan, B., Penswick, J., and Hohn, T. (1988). The leading sequence of caulimovirus large RNA can be folded into a large stem-loop structure. *Nucleic Acids Res.* **16,** 8377-8390.

Fuetterer, J., Gordon, K., Sanfançon, H., Bonneville, J.-M., and Hohn, T. (1990). Positive and negative control of translation by the leader sequence of cauliflower mosaic virus pregenomic 35S RNA. *EMBO J.* **9,** 1697-1707.

Gardner, R. C. (1983). Plant viral vectors: CaMV as an experimental tool. *In* "Genetic Engineering of Plants An Agricultural Perspective" (T. Kosuge, C. P. Merideth, A. Hollander, eds.), Vol. pp. 121-142. Plenum Press, NY,

Gardner, R. C., and Shepherd, R. J. (1980). A procedure for rapid isolation and analysis of cauliflower mosaic virus DNA. *Virology* **106,** 159-161.

Gowda, S., Wu, F. C., Scholthof, H. B., and Shepherd, R. J. (1989). Gene VI of figwort mosaic virus (caulimovirus group) functions in posttranscriptional expression of genes on the full-length RNA transcript. *Proc. Natl. Acad. Sci.* **86,** 9203-9207.

Griffith, B., Walters, R., Enger, M., Hildebrand, C., and Griffith, J. (1983). cDNA cloning and nucleotide sequence comparison of Chinese hamster metallothionein I and II mRNAs. *Nucleic Acids Res.* **11,** 901-910.

Grimsley, N., Hohn, T., and Hohn, B. (1986). Recombination in a plant virus: template-switching in cauliflower mosaic virus. *EMBO Journal* **5,** 641-646.

Gronenborn, B., Gardner, R. C., Schaefer, S., and Shepherd, R. J. (1981). Propagation of foreign DNA in plants using cauliflower mosaic virus as vector. *Nature (London)* **294,** 773-776.

Guilley, H., Dudley, R. K., Jonard, G., Balazs, E., and Richards, K. E. (1982). Transcription of cauliflower mosaic virus DNA: detection of promoter sequences and characterization of transcripts. *Cell* **30,** 763-773.

Hirochika, H., Takatsuji, H., Ubasawa, A., and Ikeda, J.-E. (1985). Site-specific deletion in cauliflower mosaic virus DNA: possible involvement of RNA splicing and reverse transcription. *EMBO J.* **4,** 1673-1680.

Hohn, B., Balazs, E., Ruegg, D., and Hohn, T. (1986). Splicing of an intervening sequence from hybrid cauliflower mosaic viral RNA. *EMBO J.* **5,** 2759-2762.

Hohn, T., Hohn, B., and Pfeiffer, P. (1985). Reverse transcription in CaMV. *Trend Biochem. Sci.* **10,** 205-209.

Howarth, A. J., Gardner, R. C., Messing, J., and Shepherd, R. J. (1981). Nucleotide sequence of naturally occurring deletion mutants of cauliflower mosaic virus. *Virology* **112,** 678-685.

Howell, S. H., Walker, L. L., and Walden, R. M. (1981). Rescue of *in vitro* generated mutants of cloned cauliflower mosaic virus genomes in infected plants. *Nature* **293,** 483-486.

Hull, R., and Covey, S. (1985). Cauliflower mosaic virus: pathways of infection. *BioEssays* **3,** 160-163.

Hull, R., Sadler, J., and Longstaff, M. (1986). The sequence of carnation etched ring virus DNA: comparison with cauliflower mosaic virus and retroviruses. *EMBO J.* **5,** 3083-3090.

Hull, R., Shepherd, R. J., and Harvey, J. D. (1976). Cauliflower mosaic virus: an improved purification procedure and some properties of the virus particles. *Journal of General Virology* **31,** 93-100.

Ish-Horowicz, D., and Burke, J. F. (1981). Rapid and efficient cosmid vector cloning. *Nucleic Acids Res.* **9,** 2989-2999.

Kassanis, B., and Varma, A. (1975). Sunn-hemp mosaic virus. *In* Commonwealth Mycological Institute/Association of Applied Biologists. Descriptions of Plant Viruses., No. 153

Kovgan, A., and Zhdanov, V. (1989). Potential vector for introducing animal virus genes into cell of higher plants. *Biotekhnologiya (Moscow)* **5,** 148-154.

Lefebvre, D., Miki, B., and Laliberté, J.-F. (1987). Mammalian metallothionein functions in plants. *BIO/TECHNOLOGY* **5,** 1053-1056.

Lowe, J. R. (1986). Transportable microcomputer programs for DNA and protein analysis. *Fed. Proc.* **45,** 1852.

Maniatis, T., Maniatis, E. F., and Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory, NY.

Marsh, L., Kuzj, A., and Guilfoyle, T. (1985). Identification and characterization of cauliflower mosaic virus replication complexes--analogy to hepatitis B viruses. *Virol.* **143,** 212-223.

Melcher, U. (1988). Puppy: a novel nucleotide sequence representation, application to caulimoviral sequences. *Computer Appl. Biosci.* **4,** 93-96.

Melcher, U. (1989). Symptoms of cauliflower mosaic virus infection in *Arabidopsis thaliana* and turnip. *Bot. Gazette* **150,** 139-147.

Melcher, U., Steffens, D. L., Choe, I. S., Lyttle, D. J., Lebeurier, G., Lin, H., and Essenberg, R. C. (1986). Infectious and non-infectious mutants of cauliflower mosaic virus DNA. *J. Gen. Virol.* **67,** 1491-1498.

Meshi, T., Ohno, T., Iba, H., and Okada, Y. (1981). Nucleotide sequence of a cloned cDNA copy of TMV (cowpea strain) RNA, including the assembly origin, the coat protein cistron, and the 3' non-coding region. *Mol. Gen. Genet.* **184,** 20-25.

Mesnard, J.-M., Kirchherr, D., Wurch, T., and Lebeurier, G. (1990). The cauliflower mosaic virus gene III product is a non-sequence-specific DNA binding protein. *Virology* **174,** 622-624.

Olszewski, N., Hagen, G., and Guilfoyle, T. J. (1982). A transcriptionally active, covalently closed, minichromosome of cauliflower mosaic virus DNA isolated from infected turnip leaves. *Cell* **29,** 395-402.

Pietrzak, M., and Hohn, T. (1985). Replication of the cauliflower mosaic virus: role and stability of the cloned D3 discontinuity sequence. *Gene* **33,** 169-179.

Revie, D., Smith, D. W., and Yee, T. W. (1988). Kinetic analysis for optimization of DNA ligation reactions. *Nucl. Acids Res.* **16,** 10301-10321.

Schoelz, J., Shepherd, R. J., and Daubert, S. (1986). Region VI of cauliflower mosaic virus encodes a host range determinant. *Mol. Cell. Biol.* **6,** 2632-2637.

Schultze, M., Jiricny, J., and Hohn, T. (1990). Open reading frame VIII is not required for viability of cauliflower mosaic virus. *Virology* **176,** 662-664.

Shepherd, R. J. (1989). Biochemistry of DNA Plant Viruses. *In* "The Biochemistry of Plants" Vol. 15, pp. 563-616. Academic Press, NY

Smith, D. R., and Calvo, J. M. (1980). Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase. *Nucleic Acids Res.* **8,** 2255-2274.

Sun, T. J., Melcher, U., and Essenberg, M. (1988). Inactivation of cauliflower mosaic virus by a photoactivatable cotton phytoalexin. *Physiol. Molec. Plant Pathol.* **33,** 115-126.

Suttcliff, J. G. (1978). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harbor Symp. Quant. Biol.* **43,** 77-90.

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

Takatsuji, H., Hirochika, H., Fukushi, T., and Ikeda, J.-E. (1986). Expression of cauliflower mosaic virus reverse transcriptase in yeast. *Nature (London)* **319,** 240-243.

Vaden, V. R., and Melcher, U. (1990). Recombination sites in cauliflower mosaic virus DNAs: implications for mechanisms of recombination. *Virology* **177,** 717-726.

Woolston, C. J., Covey, S. N., Penswick, J. R., and Davies, J. W. (1983). Aphid transmission and a polypeptide are specified by a defined region of the cauliflower mosaic virus genome. *Gene* **23,** 15-21.

Xiong, C., Muller, S., Lebeurier, G., and Hirth, L. (1982). Identification by immunoprecipitation of cauliflower mosaic virus *in vitro* major translation products with a specific serum against viroplasm protein. *EMBO J.* **1,** 971-976.

Zhang, X. S. and Melcher, U. (1989). Competition between isolates and variants of cauliflower mosaic virus in infected turnip plants. *J. Gen. Virol.* **70,** 3427-3437.

APPENDIXES

APPENDIX A

DNA sequences of insert-containing regions of selected CaMV DNAs. 5' terminus of each sequence is an A residue at position -36 relative to the *Eco*0109 cleavage site of CM4-184 CaMV isolate. 3' terminus of each sequence is an A residue at position +48 relative to the *Eco*0109 cleavage site. CM4-184 sequences are outside brackets. Inserted sequences are inside brackets. Outlined type shows long pseudo-repeat present in all inoculum clone DNAs. This repeat is associated with deletions in RP1(+) and CML3 (Table II). "ggcc" straddling each CM4-184/insert border and present in all inoculum clones is the direct repeat associated with linker deletion in CML1 (Table II). Underlined sequences are clone-specific repeats associated with insert deletions (Table II). Bold-faced **GT** and two bold-faced **AG** dinucleotides are potential splice sites bounding deletions in RP1(+) and CML5. Plasmid names are shown above sequences. Locations, relative to the secondary structure proposed for the 5' end of 35S RNA by Fuetterer, *et al.* (1988) , of 5' and 3' termini and *Eco*0109 cleavage site are shown in Appendix B.

pCML1

AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGAGGTACCC
Gg]gccCTGTG CAAGGTAAGA CGATGGAAAT TTGATAGAGG TACGTTACTA

pCML3

AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGAGGTACCC
TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT GAGATCCTTT
TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG
CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA
ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC
GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT ACATACCTCG
CTCTGCTAAT CCTGTTACCA GTGGCTGCTG CCAGTGGCGA TAAGTCGTGT
CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG CGCAGCGGTC
GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT
ACACCGAACT GAGg]gccTGT GCAAGGTAAG ACGATGGAAA TTTGATAGAG
GTACGTTACT A

pCML5

AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGAGGTACCC
TGAGGTGTCT AAAATCTCCA CTTTGTTGGC TCCGGAAAAG TTTGTAAAAC
TTTCCGTTTC CGACAAGTTT AAATGGAAGG CACCTTCGAG AGTTTGTAGT
ATAGTACAGA GTGATACCAT ATCTATGACT GCGAACGGAA GATCATTGTT
TACGTTTGAT GTTTTGAAGG ACGTGTTAAA ACACGCAGAG GAGTACACAT
ATGTCGATGT TCTTGGCGTT GTGTTATCTG GACAGTGGTT GCTCCCGAAA
GGGACGCCCG GTTCGGCAGA GATCATTCTC TTATCAGGGT CAATGCCAGC
GCTTCGTTAA TACAGATGTA GGTGTTCCAC AGGGTAGCCA GCAGCATCCT
GCGATGCAGA TCCGGAACAT AATGGTGCAG GGCGCTGACT TCCGCGTTTC
CAGACTTTAC GAAACACGGA AACCGAAGAC CATTCATGTT GTTGCTCAGg
]gccCTGTGCAA GGTAAGACGA TGGAAATTTG ATAGAGGTAC GTTACTA

pRP1 (+)

AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGAGGTACCC
TGGAGGGTAA TAGAAACACT AATCTTCTTT GCTTCGTTTT GGATATTTTT
AAGGTTTTAG AGATTCAAGG TCGTTTTTTT TTTGTTGTGT AGGATTGTTG
AGAGTGTTGG AGAAGGAGTG ACTGATCTTC AGCCAGGAGA TCATGTGTTG
CCGATCTTTA CCGGAGAATG TGGAGATTGT CGTCATTGCC AGTCGGAGGA
ATCAAACATG TGTGATCTTC TCAGGATCAA CACAGAGCGA GGAGGTATGA
TTCACGATGG TGAATCTAGA TTCTCCATTA ATGGCAAACC AATCTACCAT
TTCCTTGGGA CGTCCACGTT CAGTGAGTAC ACTGTGGTTC ACTCTGGTCA
GGTCGCTAAG ATCAATCCGG ATGCTCCTCT TGACAAGGTC TGTATTGTCA
GTTGTGGTTT GTCTACTGGG TTAGGAGCAA CTTTGAATGT GGCTAAACCC
AAGAAAGGTC AAAGTGTTGC CATTTTTGGT CTTGGTGCTG TTGGTTTCGG
CGCTGCAGAA GGTGCTAGAA TCGCTGGTGC TTCTAGGATC ATCGGTGTTG
ATTTTAACTC TAAAAGATTC GACCAAGGTA TTCAAAAAGA TGATAGTCTG
TTTTTGACTA TGTTCTTCTA TAATCTCCCT TCACTTACAT TGAATTTGAT
ATGTTATTGG CAGg]gccCTG TGCAAGGTAA GACGATGGAA ATTTGATAGA
GGTACGTTAC TA

pCaMTL (+)

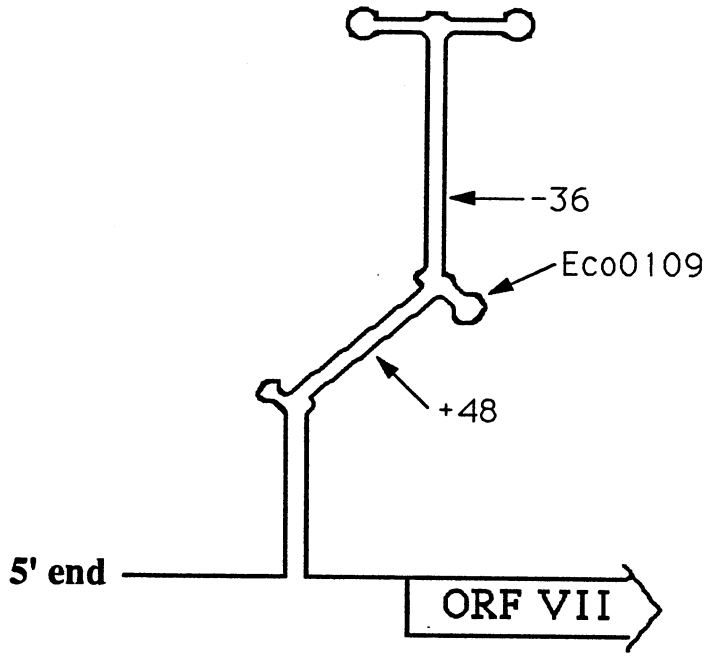
AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGAGCCATGG
ACCCCAACTG CTCCTGTGCT ACAGATGGAT CCTGCTCCTG CGCTGGGTCT
TGCAAATGCA AAGAGTGCAA ATGCACCACC TGCAAGAAAA GCTGCTGCTC
CTGCTGCCCG GTGGGCTGTG CGAAGTGCTC CCAGGGCTGC GTCTGCAAAG
AGGCTTCGGA CAAGTGCAGC TGCTGCGCCT GAAGTGGGAT TGGTCGAGGT
ACCCGGg]gccC TGTGCAAGGT AAGACGATGG AAATTTGATA GAGGTACGTT
ACTA

pCaMTL(-)

AAGGAAGGAG GAAGACATGG AAGGATAAGG TTGCAg[gccT CGACCAATCC
CACTTCAGGC GCAGCAGCTG CACTTGTCCG AAGCCTCTTT GCAGACGCAG
CCCTGGGAGC ACTTCGCACA GCCCACCGGG CAGCAGGAGC AGCAGCTTTT
CTTGCAGGTG GTGCATTTGC ACTCTTTGCA TTTGCAAGAC CCAGCGCAGG
AGCAGGATCC ATCTGTAGCA CAGGAGCAGT TGGGGTCCAT GGCTCGAGGT
ACCCGg]gccC TGTGCAAGGT AAGACGATGG AAATTTGATA GAGGTACGTT
ACTA

APPENDIX B

Diagram of secondary structure proposed for the 5' end of CaMV 35S RNA (Fuetterer, *et al.*, 1988). Parallel lines represent base-paired regions. The arrows indicate, in the RNA, the locations corresponding to the *Eco*0109 restriction site in CM4-184 DNA (Eco0109) and the 5' (-36) and 3' (+48) termini of sequences listed in appendix A.


-36
Eco0109
+48
5' end
ORF VII

2

VITA

Rodney Edward Pennington

Candidate for the Degree of

Doctor of Philosophy

Thesis: IN PLANTA DELETION OF DNA INSERTS IN THE LARGE INTERGENIC REGION OF CAULIFLOWER MOSAIC VIRUS DNA

Major Field: Biochemistry

Biographical:

Personal Data: Born in Tulsa, Oklahoma December 10, 1956, the son of Mr. and Mrs. W.E. Pennington; married in 1982 to Janice Kay Green; father of two daughters: Katherine Lee and Andrea Jean.

Education: Graduated from New Providence High School, New Providence, NJ, 1975; received Bachelor of Science degree in Microbiology from Oklahoma State University, Stillwater, Oklahoma in May 1980; completed requirements for Doctor of Philosophy Degree in Biochemistry at Oklahoma State University in May, 1991.

Professional Experience: Laboratory technician for Microbiology Department, Oklahoma State University, Stillwater, Oklahoma, 1980; graduate research assistant, Microbiology Department, 1980-1983; laboratory technician for Biochemistry Department, 1983-1985; graduate research assistant, Biochemistry Department, 1986-1990, graduate teaching assistant for Biochemistry Department, 1986 and 1987.

Professional Organizations: International Society for Plant Molecular Biology, Oklahoma Academy of Sciences, Phi Lambda Upsilon