MOLECULAR CHARACTERISATION OF SUBTERRANEAN CLOVER STUNT VIRUS

By

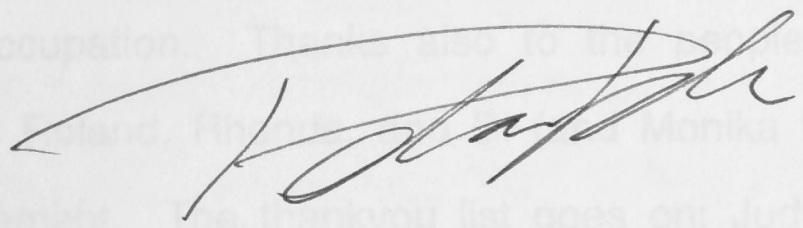
Petra Boevink

A thesis submitted for the degree of Doctor of Philosophy in the Division of Biochemistry and Molecular Biology at the Australian National University.

Submitted May 1995.

A portion of the work presented in Chapters 3 and 4 has been published in a paper entitled "Sequence of Subterranean Clover Stunt Virus DNA: affinities with the geminiviruses". Boevink, P., Chu, P.W.G. and Keese, P (1995). *Virology* 207, 354-361.

I hereby declare that the work presented in this thesis is original research conducted by me (unless otherwise stated) during the course of my enrolment as a PhD student at the Australian National University.



Petra Boevink

CSIRO Division of Plant Industry

GPO Box 1600

Canberra, ACT, Australia

ACKNOWLEDGMENTS

First and foremost, I would like to thank Paul Keese for his understanding and enthusiastic supervision. Thanks also to my advisers Brian Surin and Carol Behm for their support and advice. Heartfelt thanks to Lisa Kelly and Mick Graham for reading the tome, even the gruesome bits. Special thanks to my partner, Warwick, for support, food, and putting up with my stress and preoccupation. Thanks also to the people I've shared the lab with; Marianne, Roland, Rhonda, and Di (and Monika for a bit) for friendship and encouragement. The thankyou list goes on: Judy (for teaching me tobacco transformation), Lynda (oligos that go!) Celia (all that sectioning), Paul Chu (advice on virus biology and organising the colour printer), Paul Whitfeld and Narayana Upadhyaya (for reading the intro), Jeff and Neil (for plant bits), Roger (for glasshouse maintenance); the BBTV group for unpublished information and many discussions; Mum and Dad.

ABSTRACT

Subterranean clover stunt virus (SCSV) is a multicomponent, ssDNA virus associated with an economically important disease of pasture and grain legumes in Australia. This thesis describes an investigation of the molecular characteristics of SCSV. Seven circular DNA components which appear to comprise the genome of the SCSV F isolate were sequenced. Each component is about 1kb in size and contains a potential stem-loop structure, with a conserved nine nucleotide sequence in the loop, 5'TAGTATTAC3'. A similar sequence and a corresponding stem-loop structure are also found in the origin of replication of geminiviruses. A single major open reading frame is found in the virion sense of each component. Two of the components, numbered 2 and 6, encode potential proteins with between 37 and 61% identity to the putative replication associated proteins (RAPs) of other non-geminate ssDNA viruses of plants, such as coconut foliar decay virus (CFDV), banana bunchy top virus (BBTV), and faba bean necrotic yellows virus (FBNYV). Components 2 and 6 appear to be interchangeable, as they were found separately in some isolates of SCSV. Component 4 of SCSV and component 6 of BBTV also encode proteins with 45% sequence identity. The small, non-geminate, ssDNA viruses of plants appear to form a distinct genus, referred to as the SCSV-like viruses. The putative RAPs of SCSV-like viruses share significant homology with the RAPs of the geminiviruses. The SCSV-like viruses and geminiviruses also appear to have similar putative origins of replication, suggesting that they share a common ancestor.

Analysis of the sequence variation of SCSV component 5 in glasshouse and field infected plants revealed two distinct strains of the component which

can co-exist in a single plant. Within each strain of component 5 there were a few nucleotide differences between any two isolates or clones suggesting that SCSV exists as a quasispecies.

An attempt was made to create an infectious clone system for SCSV based on geminivirus agroinfection techniques. Single and double dimeric clones of each of the seven SCSV F isolate components were constructed and introduced into *Agrobacterium tumefaciens*. Inoculations of subterranean clover and pea plants with agrobacteria containing various combinations of the dimeric constructs did not result in any successful SCSV infections.

The regulation of SCSV gene expression was investigated by linking the putative promoter region of each SCSV component to the reporter gene GUS, and testing expression of these GUS fusion constructs in transgenic tobacco. Promoter activity was found in plants transformed with the component 1, 3, 4, 5 and 7 promoter-GUS fusion constructs. The promoters were expressed predominantly in the vascular tissue of the plants. This is consistent with the apparent vascular localisation of the virus.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Plant Viruses

Viruses are obligate intracellular parasites that infect all forms of life. Virus particles are usually composed of a nucleic acid genome encased in a protein coat. Most plant viruses have ribonucleic acid (RNA) genomes, including the largest groups, such as, potyviruses, tymoviruses, luteoviruses, potexviruses and tobamoviruses. There are, however, two plant virus groups with deoxyribonucleic acid (DNA) genomes, the caulimoviruses and the geminiviruses. Caulimoviruses have circular double-stranded (ds) DNA genomes about 8kb in size. They replicate via an RNA intermediate using reverse transcription (for a review see, Hull and Will, 1989). Geminiviruses have small, circular, single-stranded (ss) DNA genomes consisting of one or two components about 2.7kb in size (for reviews see, Davies and Stanley, 1989; Lazarowitz, 1992; Timmermans *et al.*, 1994). They have unusual fused icosahedral particles, are transmitted by either leafhoppers or whiteflies, and replicate via a rolling-circle mechanism (see section 1.5). Geminiviruses have only DNA replicative forms.

This thesis describes the molecular characterisation of a different DNA plant virus, subterranean clover stunt virus (SCSV). Like the geminiviruses, SCSV has a small circular ssDNA genome and appears to have only DNA replicative forms. There are, however, a number of significant differences. SCSV has simple icosahedral particles, more than four genome components which are only about 1kb in size, and it is transmitted by aphids. SCSV, therefore, may be representative of a new group of DNA plant viruses.

1.2 Subterranean clover stunt virus

SCSV infects subterranean clover (*Trifolium subterraneum* L.) and other legumes, such as, peas (*Pisum sativum* L.), broad beans (*Vicia faba* L.) and medics (*Medicago* sp.; Grylls and Butler, 1956; Smith, 1966). In subterranean clover the symptoms of infection may include: stunting, leaf distortion, yellowing of young leaves, and reddening of older leaves (Figure 1.1). SCSV is found throughout Australia and is estimated to cause annual yield losses of 30% in improved pastures (Chu *et al.*, 1995). The disease, subterranean clover stunt, was initially thought to be a nutritional deficiency but attempts to treat it as such failed (Grylls and Butler, 1956). It was then considered to be a viral disease and, on the basis of the symptoms and aphid transmission, probably a luteovirus (Rochow and Dufus, 1981). However, in 1988 Chu and Helms reported the isolation of isometric virus-like particles 17-19nm in diameter, composed of 19,000Mr protein subunits, and a circular, ssDNA genome. The genome was found to consist of at least four, approximately 1kb circular components by restriction analysis (Chu and Helms, 1988; Chu *et al.*, 1993a). Subsequent work indicated that there were seven components. Whether these are sufficient, or all required, for infection is unknown due to the lack of an experimental infection system.

The nucleic acid content of SCSV particles was estimated to be about 17%, on the basis of spectral absorbance and buoyant density in CsCl (Chu and Helms, 1988). Subsequent calculation of nucleic acid content based on the size of the DNA and the particle (assuming a T=1 structure, and thus 60 protein subunits) gave a similar figure, assuming that each particle contained a

Figure 1.1

Healthy and infected subterranean clover plants

Healthy (H) and SCSV infected (F) subterranean clover plants (var. Mt Barker) of the same age. The infected plant shows typical symptoms of infection with the F isolate: stunting, leaf distortion, yellowing of young leaves, and reddening of older leaves.


H
F

single nucleic acid molecule (Chu and Helms, 1988). Therefore, it appears that each component of the SCSV genome is separately encapsidated.

SCSV is obligately aphid transmitted. The most efficient aphid vector is *Aphis craccivora* Koch (Grylls and Butler, 1959). SCSV cannot be mechanically transmitted, nor is it transmitted by leafhoppers or through seed (Grylls and Butler, 1959). Furthermore, aphids can only transmit SCSV from intact infected tissue, and not from purified particle preparations despite the aphids accumulating large amounts of the virus (Chu and Helms, 1988; Chu *et al.*, 1993b). It was suggested that an aphid transmission factor, not present in the particle, or an undetected helper virus, may be required (Chu and Helms, 1988). To demonstrate that the virus was capable of replication, Chu *et al.* (1993b) inoculated pea protoplasts with purified virus particles. Replication was observed after a lag time of 3 days.

1.3 SCSV-like viruses

A number of recently identified viruses of plants have similar characteristics to SCSV (Table 1.1). They have small, multi-component, circular, ssDNA genomes and 18-20nm icosahedral particles. Together with SCSV these viruses probably constitute a new group of ssDNA viruses of plants.

1.3.1 Coconut foliar decay virus

A single, circular ssDNA component associated with coconut foliar disease was first described by Randles *et al.* (1987). Isometric virus-like particles associated with the disease are 20nm in diameter (Randles and Hanold, 1989). Sequencing of the 1291b DNA revealed 6 possible ORFs, the

Table 1.1

Properties of small, circular, ssDNA viruses of eukaryotes

A comparison of the vectors, sizes and number of DNA genome components, the number of open reading frames (ORFs) and the direction of transcription, and the conserved nonanucleotide loop sequences of the small, circular, ssDNA viruses of eukaryotes. The viruses are: subterranean clover stunt virus (Grylls and Butler, 1959; Chu and Helms, 1988; Chu *et al.*, 1993), banana bunchy top virus (Dale, 1987; Thomas and Dietzgen, 1991; Harding *et al.*, 1991; Burns, 1994), coconut foliar decay virus (Julia, 1982; Randles *et al.*, 1987; Rhode *et al.*, 1990) faba bean necrotic yellows virus (Katul *et al.*, 1993; Katul *et al.*, 1995), milk-vetch dwarf virus (Isogai *et al.*, 1992), porcine circovirus (Mankertz *et al.*, 1993), chicken anaemia virus (Noteborn *et al.*, 1991), psittacine beak and feather disease virus (Ritchie *et al.*, 1989), and the geminiviruses (reviewed in Lazarowitz, 1992 and Timmermans *et al.*, 1994).

Virus	Vector	Size of DNA (kb)	Number of DNAs	Transcription	Number of ORFs per DNA	Loop sequence
SCSV	aphid (*Aphis craccivora*)	~ 1	7	ND	ND	ND
BBTV	aphid (*Petalonia nigronervsa*)	1-1.1	> or = to 6	unidirectional[a]	1[a]	TATTATTAC
CFDV	planthopper (*Myndus taffni*)	1.3	>1	bidirectional[a]	>1[a]	TAGTATTAC
FBNYV	aphid (*Acyrthosiphon pisum*)	~ 1	>4	unidirectional[a]	1[a]	TAGTATTAC
MDV	aphid (*Aphis craccivora*)	~ 1	ND	ND	ND	ND
PCV	-	1.7	1	bidirectional	6	TAGTATTAC
CAV	-	2.3	1	unidirectional	3	-[b]
PBFDV	-	~1.7-2	ND	ND	ND	ND
Geminiviruses	*Bemsia tabaci*, leafhoppers	~ 2.7	1-2	bidirectional	2-7	TAATATTAC

ND - not determined

[a] suggested by the authors but not confirmed experimentally

[b] does not have an equivalent stem-loop sequence

largest of which encodes a potential replication associated protein (RAP) (Figure 1.2; Rhode *et al.,* 1990). The DNA sequence flanking the RAP ORF contains a sequence capable of forming a stem-loop. Within the loop of this structure is a nonanucleotide 5'TAGTATTAC3', which is very similar to the conserved nonanucleotide found in the stem-loop in the origin of replication of geminiviruses (see section 1.5). Since the DNA sequenced did not contain an ORF capable of encoding the coat protein, the coconut foliar decay virus (CFDV) genome is anticipated to consist of at least one other DNA component. CFDV differs from the other SCSV-like viruses in that it is transmitted by a planthopper rather than an aphid (Julia, 1982).

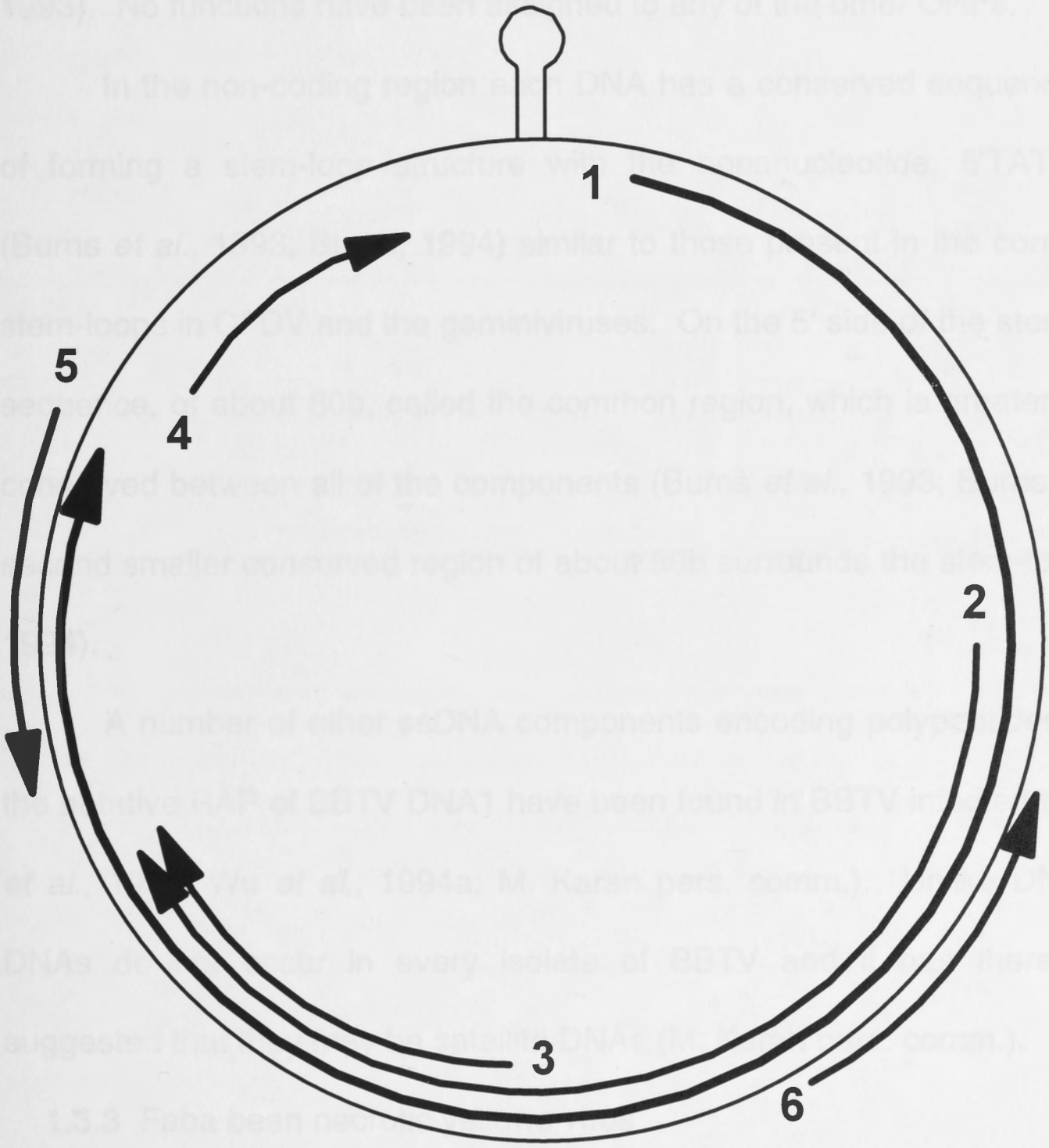
1.3.2 Banana bunchy top virus

Banana bunchy top disease is the most serious viral disease of bananas (Dale, 1987). It is persistently aphid transmitted and is not mechanically transmissible (Dale, 1987). It was initially thought to be due to a luteovirus (Matthews, 1982). This was supported by the isolation of disease associated dsRNA (Dale *et al.*, 1986) and ssRNA associated with 20-22nm virus like particles (Wu and Su, 1990). Subsequently, 18-20nm isometric particles, constructed from coat protein subunits of 20,000Mr, were isolated and found to contain circular, ssDNA about 1kb in size (Thomas and Dietzgen, 1991; Harding *et al.*, 1991; Harding *et al.*, 1993). These ssDNA molecules comprise a virus which is now referred to as banana bunchy top virus (BBTV). As with SCSV, aphids are only able to transmit the disease from intact infected plants and not from purified particle preparations. It was suggested that a helper virus, such as a luteovirus, might be required (Thomas and Dietzgen, 1991).

Figure 1.2

Genome organisation of CFDV

The 1291b genomic component of CFDV that has been sequenced is shown with the 6 potential ORFs represented by arrows. ORF 1 encodes the putative RAP (Rhode *et al.*, 1990).



The BBTV genome consists of at least six DNA components, ranging in size from 1018-1111b (Figure 1.3; Burns *et al.,* 1993; Burns, 1994). With the exception of component 2, each DNA has one major open reading frame flanked by consensus transcription signals (Burns *et al.,* 1993; Burns, 1994). The largest ORF of DNA 1 encodes a potential RAP similar to that of CFDV, which shows distant similarity to the RAPs of the geminiviruses (Harding *et al.,* 1993). No functions have been assigned to any of the other ORFs.

In the non-coding region each DNA has a conserved sequence capable of forming a stem-loop structure with the nonanucleotide, 5'TATTATTAC3' (Burns *et al.,* 1993; Burns, 1994) similar to those present in the corresponding stem-loops in CFDV and the geminiviruses. On the 5' side of the stem-loop is a sequence, of about 80b, called the common region, which is greater than 85% conserved between all of the components (Burns *et al.,* 1993; Burns, 1994). A second smaller conserved region of about 50b surrounds the stem-loop (Burns, 1994).

A number of other ssDNA components encoding polypeptides similar to the putative RAP of BBTV DNA1 have been found in BBTV infected tissue (Yeh *et al.*, 1994; Wu *et al.*, 1994a; M. Karan pers. comm.). Unlike DNA1, these DNAs do not occur in every isolate of BBTV and it has therefore been suggested that they may be satellite DNAs (M. Karan pers. comm.).

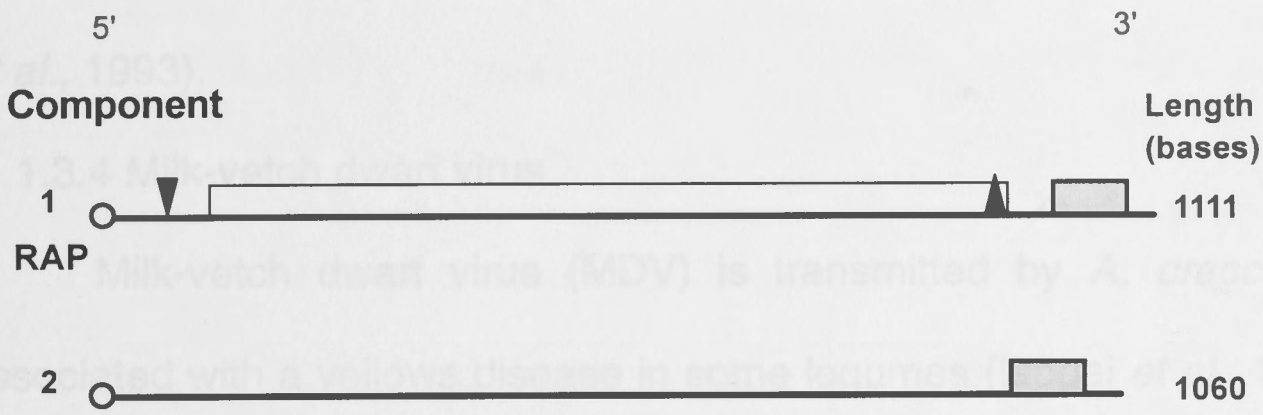
1.3.3 Faba bean necrotic yellows virus

Faba bean necrotic yellows virus (FBNYV) is associated with a yellows disease of crop and pasture legumes. It has 18nm isometric particles, composed of 22kD protein subunits, which contain circular ssDNA of approximately 1kb (Katul *et al.,* 1993). The genome consists of several

Figure 1.3

Genome organisation of BBTV

The circular genomic components of BBTV are shown linearly, starting at the first base of the conserved stem-loop sequence, represented by the small circle. The major ORF in each component is represented by the clear rectangle. The approximately 80b common region is represented by the shaded rectangle. The potential transcription signals are represented by triangles.



components (Katul *et al.,* 1993), one of which has been sequenced (Katul *et al.,* 1995). This component encodes a putative RAP similar to those of BBTV and CFDV. The non-coding region contains a potential stem-loop structure, within which is a nonanucleotide, 5'TAGTATTAC3', identical to that of CFDV. Antibodies raised to FBNYV particles reacted weakly with SCSV in immuno-electron microscopy but not with BBTV (Katul *et al.,* 1993). As with SCSV, the disease is persistently transmitted by aphids (including *A. craccivora*) and cannot be transmitted mechanically or by aphids fed on purified particles (Katul *et al.,* 1993).

1.3.4 Milk-vetch dwarf virus

Milk-vetch dwarf virus (MDV) is transmitted by *A. craccivora* and is associated with a yellows disease in some legumes (Isogai *et al.*, 1992). It has 19-20nm isometric particles and ssDNA about 1kb in size (Isogai *et al.*, 1992; Chu *et al.*, 1995).

1.4 Other small circular ssDNA viruses of eukaryotes

Three animal viruses also have small isometric particles and small, circular, ssDNA genomes. These viruses have been classified in the family circoviridae, genus circovirus. One of them, porcine circovirus, also has some sequence similarities to SCSV-like viruses.

1.4.1 Porcine circovirus

Porcine circovirus (PCV) is a non-pathogenic virus of pigs (Tisher *et al.*, 1982). It has 17nm isometric particles and a positive sense circular, ssDNA genome 1759b in size (Tisher *et al.*, 1982; Mankertz *et al.*, 1993). Six potentially functional open reading frames were identified in the DNA but only

three transcripts were found (Mankertz *et al.*, 1993). At least one of the transcripts appeared to be spliced. ORF 4 shows similarity to the RAP ORF of CFDV (Rohde *et al.,* 1990). The DNA also encoded a potential stem-loop structure with the nonanucleotide 5'TAGTATTAC3' (Mankertz *et al.*, 1993).

1.4.2 Chicken anaemia virus

Chicken anaemia virus (CAV) has 23nm isometric particles composed of 50kD protein subunits (Todd *et al.,* 1990). Its genome is a negative sense circular ssDNA molecule 2319b in size which encodes three overlapping ORFs (Noteborn *et al.,* 1991). None of the encoded polypeptides show any sequence identity to those of the SCSV-like viruses.

1.4.3 Psittacine beak and feather disease virus

Psittacine beak and feather disease virus (PBFDV) is a devastating virus of cockatoos. It has 14-16nm icosahedral particles composed of three proteins of 26.3, 23.7 and 15.9kD, and contains 1.7-2kb circular ssDNA (Ritchie *et al.,* 1989).

1.5 Geminiviruses

When compared to the other two groups of plant DNA viruses, the caulimoviruses and the geminiviruses, SCSV-like viruses share most affinity with the geminiviruses. Geminiviruses have circular ssDNA genomes with an origin of replication containing a conserved stem-loop which is similar in structure and sequence to those of the SCSV-like viruses. Also, some geminiviruses encapsidate a small primer molecule, as does SCSV (Chu and Helms, 1988). Geminiviruses may, therefore, provide insights into the origin and gene functions of SCSV.

1.5.1 Geminivirus classification

Geminiviruses are named after their unusual, fused icosahedral particles, called twinned or geminate particles. They are divided into three subgroups (Rybicki, 1994). The different genome organisations of these three subgroups are illustrated in Figure 1.4.

Subgroup I geminiviruses, of which the type member is maize streak virus (MSV), generally infect monocotyledonous plants, are transmitted by different leafhopper species and have a single genome component (monopartite) with a simple organisation (Figure 1.4). Tobacco yellow dwarf virus (TobYDV), whilst having the genome organisation of a typical subgroup I geminivirus, infects dicotyledonous plants (Morris *et al.*, 1992).

Beet curly top virus (BCTV) is the type member of subgroup II. It has a single genome component, is transmitted by a leafhopper, *Circulifer tenellus* (Baker), and infects dicotyledonous plants (Stanley *et al.*, 1986). Its genome organisation is intermediate between those of subgroup I and III geminiviruses (see below).

Subgroup III geminiviruses, of which the type member is African cassava mosaic virus (ACMV), infect dicotyledonous plants, are all transmitted by the whitefly, *Bemsia tabaci*, and generally have two genomic components. These are designated DNA A and B (bipartite), and are separately encapsidated (Francki *et al.*, 1980; Goodman *et al.*, 1980). DNA A encodes all of the functions necessary for viral replication and encapsidation (Townsend *et al.*, 1985; Rogers *et al.*, 1986; Sunter *et al.*, 1987; Elmer *et al.*, 1988a; Etessami *et al.*, 1991). DNA B encodes functions required for movement of the virus within the infected plant (Brough *et al.*, 1988; Etessami *et al.*, 1988). For the majority

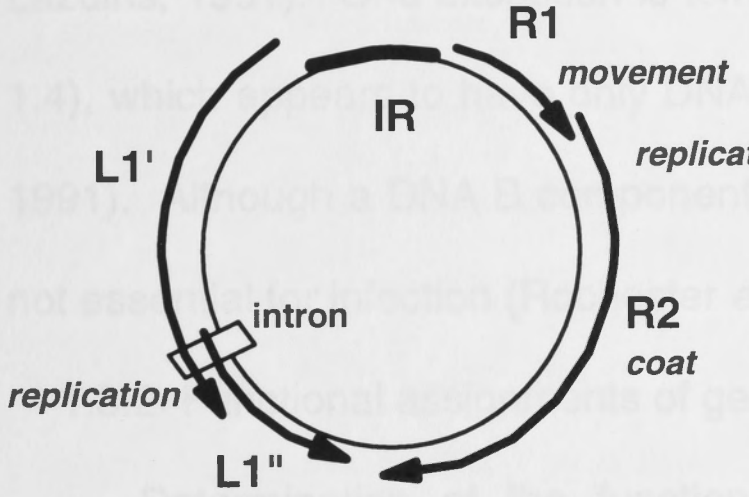
Figure 1.4

Genome organisations of representative geminiviruses.

The maize streak virus (MSV; Mullineaux *et al.*, 1984) genome organisation is typical of subgroup I geminiviruses. Beet curly top virus (BCTV; Hormuzdi and Bisaro, 1993) is the type member of subgroup II. African cassava mosaic virus (ACMV; Stanley and Gay, 1983) DNAs A and B represent the typical, bipartite genome organisation of the subgroup III geminiviruses. Tomato yellow leaf curl virus (TYLCV; Jupin *et al.*, 1994) is an atypical subgroup III virus with a monopartite genome. Open reading frames (ORFs) are represented by arrows indicating the direction of transcription and named according to those directions using rightward (R) and leftward (L) (the designations virion (V) and complementary (C) are also commonly used). The ORF specifications also indicate whether they are from the A or B component in subgroup III. Numbering of the ORFs is according to Timmermans *et al.* (1994). IR represents the intergenic region. Functions of the ORF products are given in *italics* where they are known.

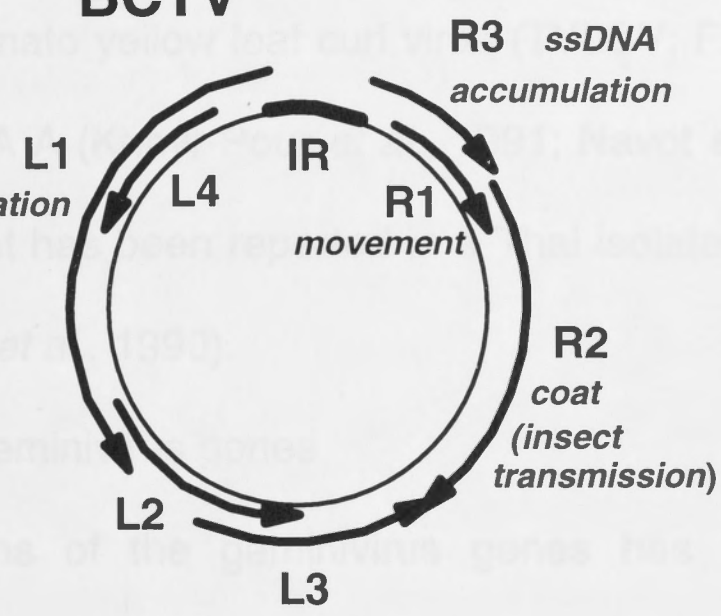
Subgroup I

MSV



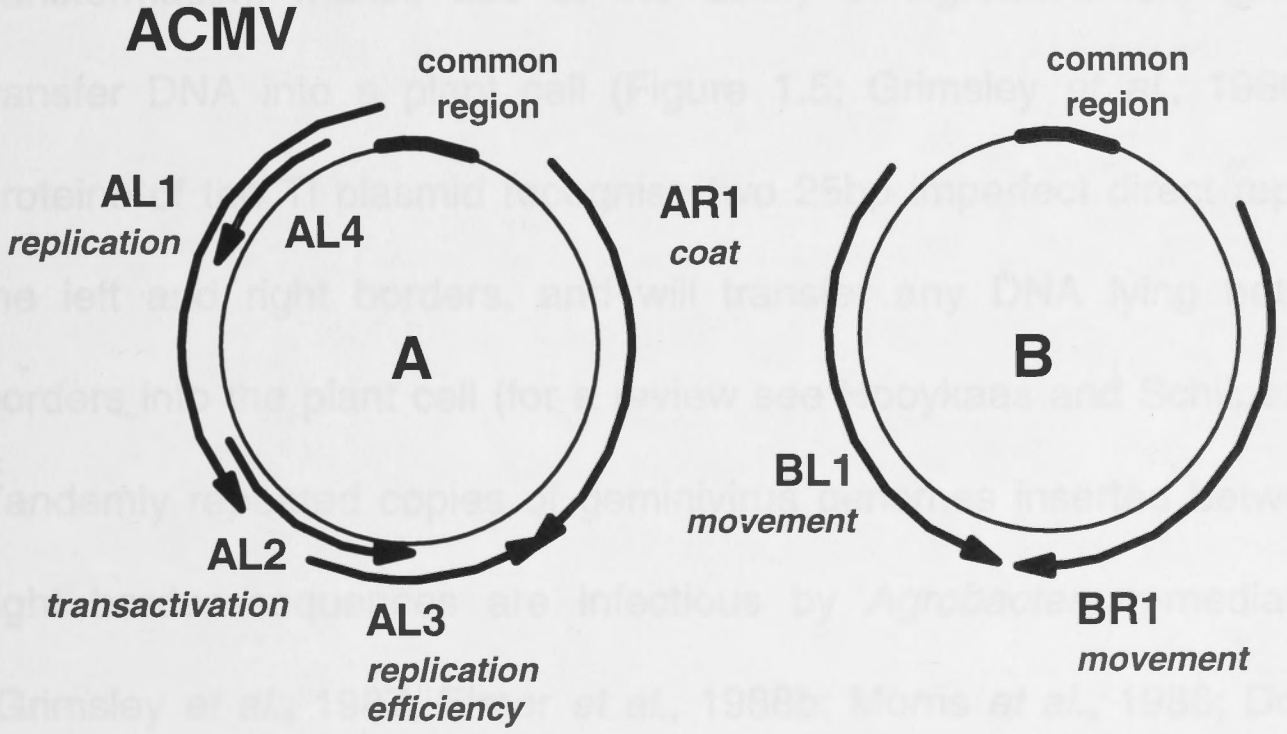
Subgroup II

BCTV

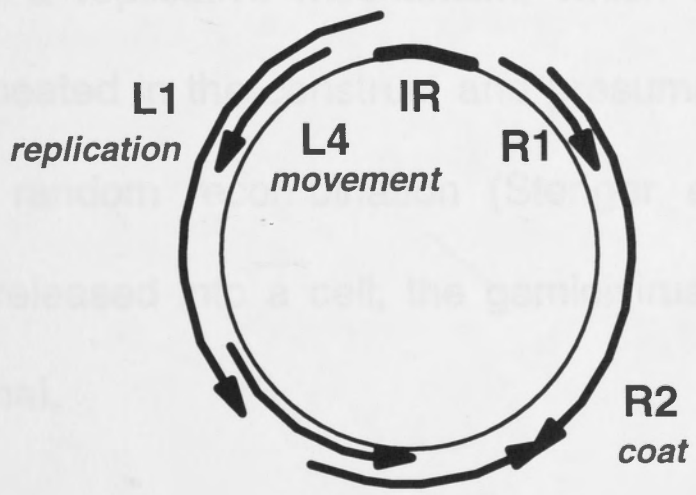


Subgroup III

ACMV



TYLCV



of the subgroup III geminiviruses, both components are essential for infection (Stanley, 1983; Hamilton *et al.*, 1983; Howarth *et al.*, 1985; Lazarowitz and Lazdins, 1991). One exception is tomato yellow leaf curl virus (TYLCV; Figure 1.4), which appears to have only DNA A (Kheyr-Pour *et al.*, 1991; Navot *et al.*, 1991). Although a DNA B component has been reported in a Thai isolate, it is not essential for infection (Rochester *et al.*, 1990).

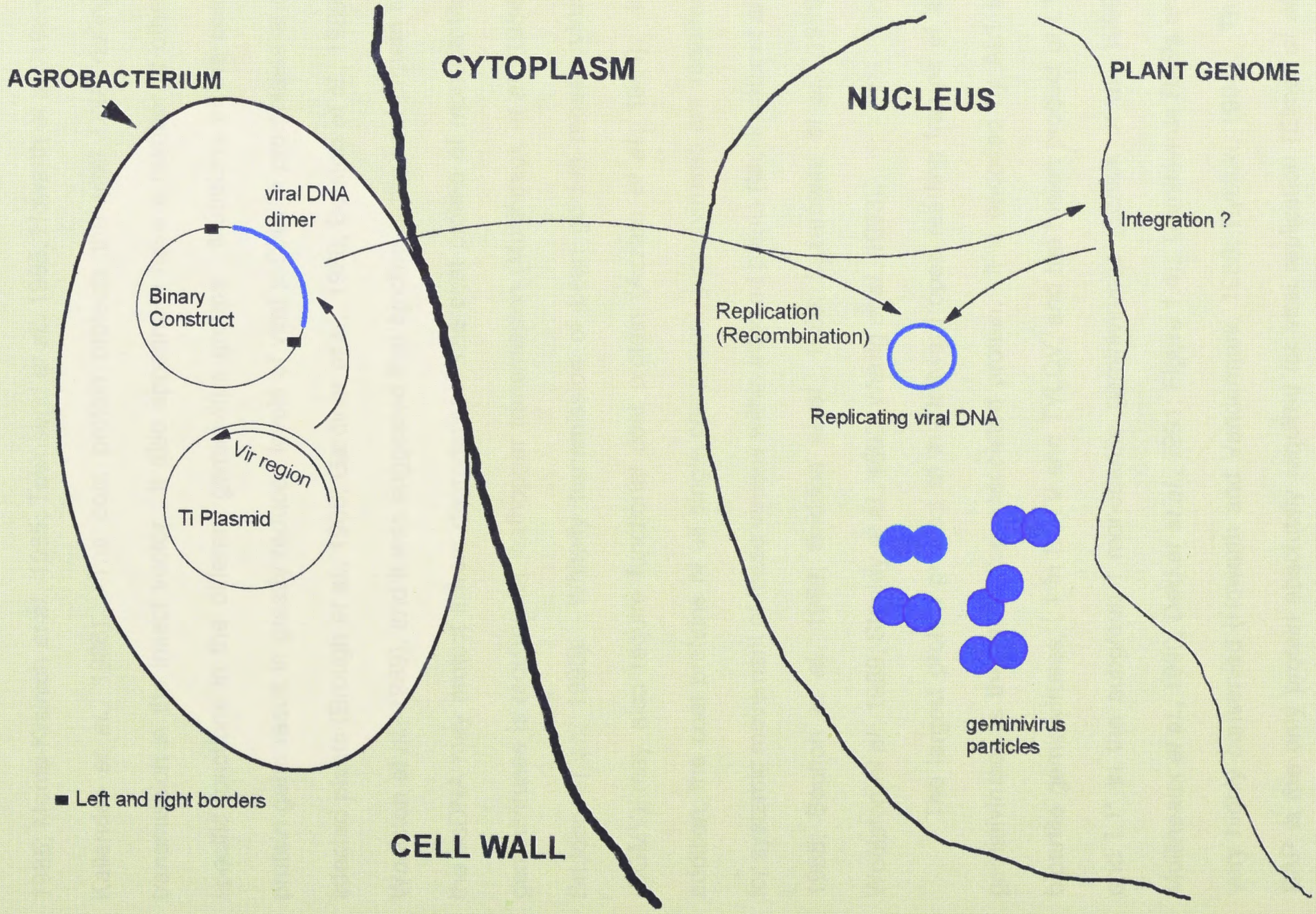
1.5.2 Functional assignments of geminivirus genes

Determination of the functions of the geminivirus genes has been achieved largely by mutation studies. These were made possible by the development of agroinfection (Grimsley *et al.*, 1986). Agroinfection, like plant transformation, makes use of the ability of *Agrobacterium tumefaciens* to transfer DNA into a plant cell (Figure 1.5; Grimsley *et al.*, 1986). The *vir* proteins of the Ti plasmid recognise two 25bp imperfect direct repeats, called the left and right borders, and will transfer any DNA lying between these borders into the plant cell (for a review see Hooykaas and Schilperoort, 1992). Tandemly repeated copies of geminivirus genomes inserted between left and right border sequences are infectious by *Agrobacterium*-mediated delivery (Grimsley *et al.*, 1987; Elmer *et al.*, 1988b; Morris *et al.*, 1988; Donson *et al.*, 1988). Release of viral monomers from the transferred DNA occurs preferentially by a replicative mechanism, which requires that the origin of replication is repeated in the construct and presumably occurs in the nucleus, rather than by random recombination (Stenger *et al.*, 1991). Once viral monomers are released into a cell, the geminivirus infection of the plant can proceed as normal.

Figure 1.5

Agroinfection

The mechanism of agroinfection. The proteins produced by the *vir* region of the helper Ti plasmid recognise the left and right border sequences (represented by small black boxes) surrounding the dimeric viral DNA construct and transfer a copy of it to the nucleus of the plant cell (for review see Hooykaas and Schilperoort, 1992). Once in the nucleus, unit length viral DNA is released from the construct preferentially by a replicative mechanism, although recombination also occurs (Stenger *et al.*, 1991). Integration into the genome is not required for release of viral monomers, although it may occur. Once viral monomers are released, the infection proceeds as normal.


AGROBACTERIUM
CYTOPLASM
NUCLEUS
PLANT GENOME
viral DNA
dimer
Binary
Construct
Vir region
Ti Plasmid
Integration ?
Replication
(Recombination)
Replicating viral DNA
geminivirus
particles
■ Left and right borders
CELL WALL

The major rightward ORF, AR1 or R2, in all geminiviruses encodes the coat protein (Stanley and Gay, 1983; Hamilton *et al.*, 1984; MacDowell *et al.*, 1985; Morris-Krsinich *et al.*, 1985; Townsend *et al.*, 1985; Howarth *et al.*, 1985; Kallender *et al.*, 1988). The coat protein protects the viral DNA during transmission in the insect vector. It also appears to have a number of other specific functions in the different geminivirus groups. Mutations in the coat protein gene result in greatly reduced levels of viral ssDNA in protoplasts and infected plants (Brough *et al.*, 1988; Gardiner *et al.*, 1988; Boulton *et al.*, 1989; Woolston *et al.*, 1989), and it was suggested that binding of the coat protein to the ssDNA may protect it from degradation. The coat protein of monopartite geminiviruses is required for leafhopper transmission (Mullineaux *et al.*, 1984; Briddon *et al.*, 1990). Whitefly transmission of bean golden mosaic virus (BGMV) may also require functional coat protein (Azzam *et al.*, 1994). In addition, the coat proteins of all single component geminiviruses are required for systemic movement of these viruses within infected plants (Mullineaux *et al.*, 1988; Boulton *et al.*, 1989; Briddon *et al.*, 1989; Lazarowitz *et al.*, 1989; Woolston *et al.*, 1989; Stanley *et al.*, 1992; Wartig *et al.*, 1993).

The largest protein product of the leftward open reading frames of all geminiviruses is the replication associated protein. It is encoded by AL1 in bipartite geminiviruses, L1 in BCTV and TYLCV, and the fusion product of L1' and L1" in the subgroup I monopartite geminiviruses (Schalk *et al.*, 1989; Mullineaux *et al.*, 1990; Dekker *et al.*, 1991; Figure 1.4). Geminivirus RAPs are very highly conserved (Howarth and Vandermark, 1989; Rybicki, 1994). The RAP is the only protein absolutely required for viral replication (Elmer *et al.*, 1988a; Accotto *et al.*, 1989; Schalk *et al.*, 1989; Hanley-Bowdoin *et al.*, 1990;

Etessami *et al.*, 1991). The RAP of subgroup I geminiviruses may also be a transactivator (Hofer *et al.*, 1992) analogous to the AL2 protein of subgroup III geminiviruses (Sunter and Bisaro, 1991; Haley *et al.*, 1992; Sunter and Bisaro, 1992).

The movement proteins of the different geminivirus subgroups appear to have quite separate evolutionary origins because they are in different locations in the viral genomes and share little or no homology. DNA B of the bipartite geminiviruses encodes the two proteins required for viral movement, BR1 and BL1 (Brough *et al.*, 1988; Etessami *et al.*, 1988). BL1 appears to be involved in cell-to-cell movement (Hayes and Buck, 1989; Koonin *et al.*, 1991; Von Arnim *et al.*, 1993; Pascal *et al.*, 1993; Ingham *et al.*, 1995). The BL1 protein of bean dwarf mosaic virus (BDMV) increases the size exclusion limit of plasmodesmata (Noueiry *et al.*, 1994). Mutation studies have implicated BR1 in both short (Von Arnim *et al.*, 1993) and long (Lazarowitz, 1991; Pascal *et al.*, 1993; Ingham and Lazarowitz, 1993) distance movement within the plant. The BR1 protein of BDMV also appears to be involved in the export of viral DNA from the nucleus (Noueiry *et al.,* 1994). The bipartite geminivirus BR1 proteins show significant similarity to their coat proteins (AR1; Kikuno *et al.*, 1984; Rybicki, 1994), and Ingham *et al.* (1995) suggested that AR1 and BR1 may share some redundant functions.

TYLCV, which lacks the B component, was found to require the L4 ORF for movement (Jupin *et al.*, 1994). This ORF is also found in the bipartite geminiviruses and BCTV; however, mutations of AL4 of bipartite geminiviruses had no effect on viral infection (Elmer *et al.*, 1988a; Etessami *et al.*, 1991) and

mutations of the BCTV L4 gene affected symptom severity but not movement (Stanley and Latham, 1992).

The subgroup I monopartite geminiviruses and BCTV require the R1 gene (Figure 1.4) for movement (Mullineaux *et al.*, 1988; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989; Woolston *et al.*, 1989; Briddon *et al.*, 1989; Stanley *et al.*, 1992; Hormuzdi and Bisaro, 1993).

1.5.3 Geminivirus replication

Geminiviruses replicate via a rolling-circle mechanism (Stenger *et al.*, 1991; Saunders *et al.*, 1991; Heyraud *et al.*, 1993). The first step, after the virus enters the plant cell nucleus, is the conversion of the viral single-stranded DNA to the double-stranded form (replicative form; RF) by host enzymes (Timmermans *et al.*, 1994). Subgroup I monopartite geminiviruses encapsidate a short complementary DNA fragment of about 80 nucleotides, which is tightly associated with the viral DNA and which primes the synthesis of the complementary strand (Donson *et al.*, 1984; Donson *et al.*, 1987; Hayes *et al.*, 1988). This primer has a few ribonucleotides at its 5' end, which are thought to be the remnants of an RNA primer (Donson *et al.*, 1984; Donson *et al.*, 1987; Hayes *et al.*, 1988) analogous to the priming of Okasaki fragments in lagging strand synthesis in DNA replication (Okasaki *et al.*, 1978). No such encapsidated primer has been found in bipartite geminiviruses. Bipartite geminivirus complementary strand synthesis, which is also primed with RNA (Saunders *et al.*, 1992), was suggested to be initiated by secondary structural features in the common region.

Once the double-stranded DNA has been synthesised, transcription by the host RNA polymerase of the viral genes required for viral strand synthesis

occurs. The model for geminivirus rolling-circle replication is based on that of the bacteriophage øX174 (Saunders *et al.*, 1991; Stenger *et al.*, 1991). Viral strand synthesis is initiated by the RAP which binds to a specific site in the common/large intergenic region (Figure 1.6; Fontes *et al.*, 1992; Lazarowitz *et al.*, 1992) and introduces a nick into the viral strand in the conserved nonanucleotide in the loop of the stem-loop structure (Stenger *et al.*, 1991; Heyraud *et al.*, 1993). This conserved sequence, 5'TAATATTAC3', is similar to the gene A recognition and cleavage sequence of øX174 (Van Mansfield *et al.*, 1979; Rogers *et al.*, 1986; Stenger *et al.*, 1991). The replication complex, which probably consists largely of host proteins, then synthesises viral strand DNA using the 3' OH terminus of the nicked strand as a primer (Saunders *et al.*, 1991). The AL1 protein binds to the 5' phosphoryl end of the nicked viral strand following nicking, and later cleaves the displaced unit-length product of the first round of replication and ligates it into circular form (Laufs *et al.*, 1995).

1.6 Aims of the Project

In contrast to the geminivirus group, the replicative mechanisms and gene functions of which have been extensively studied, there is very little known about the SCSV-like viruses. This project aims to increase the understanding of SCSV.

The specific aims of the project were:

1. To sequence the seven SCSV genome components and compare them to other SCSV-like viruses and the geminiviruses.
2. To examine the molecular basis of symptom variation between SCSV isolates. The symptoms of SCSV infection in the field are variable. This

variation could be caused by the presence of different mixtures of viral genome components in the plant or by sequence differences between individual components.

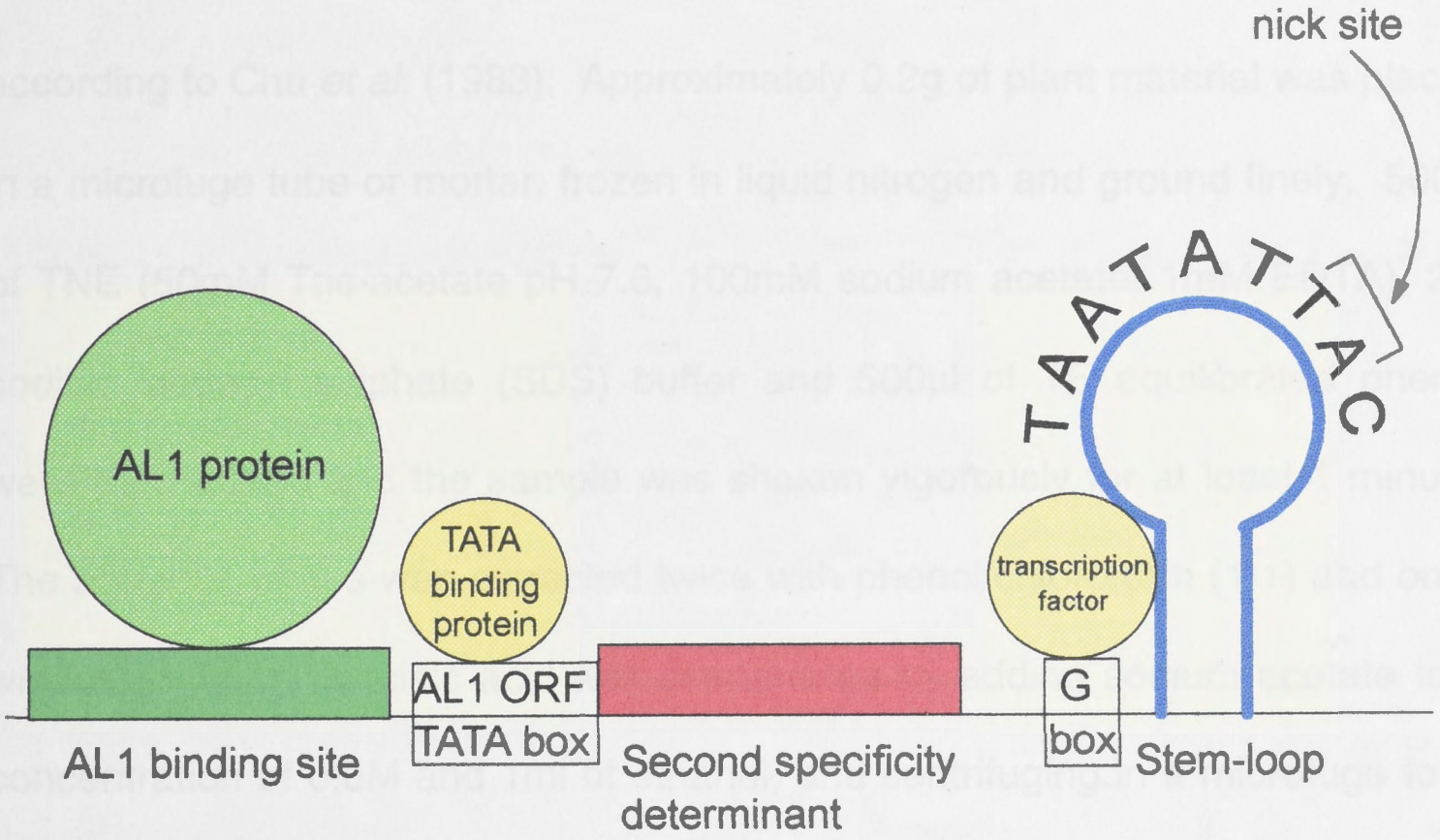
3. To develop an agroinfection system for SCSV, analogous to that used for the geminiviruses. This would not only help to confirm that SCSV causes subterranean clover stunt disease, but also to determine which components are essential for infectivity.

4. To probe the non-coding regions of the SCSV DNAs for promoter activity. Since SCSV is obligately aphid transmitted and not mechanically transmissible, it is inferred that it is phloem limited. This may be reflected in the locations of SCSV DNAs in infected plants and in SCSV gene expression.

Figure 1.6

Structure of the subgroup III geminivirus origin of replication

The proposed structure of the subgroup III geminivirus origin of replication is represented schematically 5' to 3' (left to right) as in Fontes *et al.* (1994) and Argüello-Astorga *et al.* (1994). The AL1 binding site, represented by the green rectangle, is composed of directly repeated sequences, the end of which is less than ten nucleotides upstream of the TATA box of the AL1 ORF. The sequence of the direct repeat is not completely conserved between the bipartite geminiviruses and the AL1 binding site therefore acts as a species specific recognition determinant (Fontes *et al.*, 1994). A second element, represented by the red rectangle, between the AL1 binding site and the stem-loop also determines species specificity (Fontes *et al.*, 1994). Argüello-Astorga *et al.* (1994) proposed that host TATA binding protein and transcription factor (yellow circles) bind to the TATA box and G box respectively and facilitate the folding of the origin DNA to allow the AL1 protein (green circle) to nick the virion strand DNA within the loop sequence. Nicking of the virion strand DNA is the initiation step for rolling circle replication. In the subgroup I geminivirus, wheat dwarf virus, the nicking of the virion strand DNA is proposed to occur at or between the TA residues indicated (Heyraud *et al.*, 1993). Since the nonanucleotide is completely conserved in the geminivirus group, this is likely to be the nick site for all geminiviruses.


nick site
TAATATTAC
AL1 protein
TATA
binding
protein
transcription
factor
AL1 ORF
TATA box
G
box
AL1 binding site
Second specificity
determinant
Stem-loop

CHAPTER 2: MATERIALS AND METHODS

2.1 Plant nucleic acid extraction

Total nucleic acid was extracted from subterranean clover and pea according to Chu *et al.* (1983). Approximately 0.2g of plant material was placed in a microfuge tube or mortar, frozen in liquid nitrogen and ground finely. 500µl of TNE (50mM Tris-acetate pH 7.6, 100mM sodium acetate, 1mM EDTA), 2% sodium dodecyl sulphate (SDS) buffer and 500µl of TE equilibrated phenol were then added and the sample was shaken vigorously for at least 1 minute. The aqueous phase was extracted twice with phenol/chloroform (1:1) and once with chloroform. Nucleic acid was precipitated by adding sodium acetate to a concentration of 0.3M and 1ml of ethanol, and centrifuging in a microfuge for 5 minutes. The precipitate was washed twice with 70% ethanol, dried and resuspended in 200µl of distilled water.

2.2 Cloning

The techniques used for manipulating DNA were generally as described in Sambrook *et al.* (1989). Restriction endonuclease digests were carried out according to the manufacturer's (New England Biolabs) instructions using the buffers provided with the enzymes.

DNA products from PCR reactions were purified with a phenol/chloroform extraction and ethanol precipitation and then either treated with T4 polynucleotide kinase and T4 DNA polymerase (New England Biolabs) to fill in the ends of the fragments, or cut with restriction enzymes. They were purified a second time by phenol/chloroform extraction and ethanol precipitation

and ligated into a blunt vector, such as, pGEM5Z-*Eco*RV, or other suitably cut vectors, using T4 DNA ligase (New England Biolabs).

2.3 Transformation of bacteria

2.3.1 *Escherichia coli*

Escherichia coli strain NM522 (Gough and Murray, 1983) or JM109 (Yanisch-Perron *et al.*, 1985) were transformed by electroporation (Dower *et al.*, 1988). Cells were prepared by growing in 1l of Luria-Bertani broth (LB: 1% bacto-tryptone (Difco), 0.5% bacto-yeast extract (Difco), and 1% NaCl) to an absorbance at 600nm of 0.5-1.0. The cells were harvested by centrifugation at 5000rpm for 15 minutes at 4°C, resuspended in 1l of sterile distilled water and centrifuged as above. The cells were resuspended in 0.5l of water, centrifuged and then resuspended in 20ml of sterile 10% glycerol. After centrifugation the cells were finally resuspended in 10% glycerol to a total volume of 3ml. Aliquots (40µl) were frozen in liquid nitrogen and stored at -80°C.

For transformation, an aliquot was thawed on ice and mixed with a small amount of DNA then electroporated using the Biorad Gene Pulser set at 2.5kV, 200Ω and 25µF. Cells were allowed to recover in LB for 10-20 minutes at 37°C before being plated onto LB plates containing appropriate antibiotics. Several colonies were picked from each plate and grown separately in 2ml of LB medium containing appropriate antibiotics for a minimum of 6 hours at 37°C. The DNA was then extracted by alkaline lysis (Sambrook *et al.*, 1989), analysed by restriction endonuclease digestion and electrophoresis on 0.8-1.2% agarose gels.

2.3.2 *Agrobacterium tumefaciens*

The methods of Nagel *et al.* (1990) were used for electroporation of *Agrobacterium tumefaciens* (strain LBA4404; Hoekema *et al.*, 1983; and strain AGL1; Lazo *et al.*, 1986). To prepare them for electroporation, the cells were grown in 250ml YEP medium (1% bacto-yeast extract (Difco), 1% bacto-peptone (Difco), 0.5% NaCl) to an absorbance of 0.5-1.0 at 600nm. They were collected by centrifugation at 5000rpm for 15 minutes at 4°C, resuspended in 250ml sterile distilled water at 4°C and re-centrifuged. The cells were then resuspended in 125ml of water, centrifuged as before, resuspended in 100ml sterile 10% glycerol at 4°C and re-centrifuged. The cells were finally resuspended in 750µl of 10% glycerol, frozen in liquid nitrogen in 40µl aliquots and stored at -80°C. For transformation, an aliquot of cells was thawed on ice, mixed with 0.5-1µg of DNA and electroporated using the same settings as for *E. coli*. The cells were allowed to recover in YEP broth for 2-4 hours at 28°C and then plated onto YEP plates containing appropriate antibiotics.

A. tumefaciens cells were also transformed by triparental mating, using the helper plasmid RK2013 as described by Ditta *et al.* (1980). The *A. tumefaciens* cells were grown on a YEP plate containing rifampicin 50µg/ml at 28°C for 2-3 days. The *E. coli* strains carrying the binary construct to be transferred and the RK2013 plasmid were grown overnight on LB plates with appropriate antibiotics. Large loopfuls of each type of bacterium were scraped from these plates and mixed together on a LB plate without antibiotics. They were incubated overnight at 28°C and then spread onto YEP plates containing rifampicin 50µg/ml to kill the *E. coli* and the antibiotic for selection of the binary construct. To eliminate contaminating *E. coli* several colonies from each plate

were replated onto selection plates and single colonies from these were again replated.

A few colonies were selected from each plate of either the electroporated or purified mated *A. tumefaciens* and cultured in YEP medium containing appropriate antibiotics for 2 days. The DNA was then extracted by a modified alkaline lysis method. The cells from 1.5ml of culture were pelleted by centrifugation and gently resuspended in 100µl of SET buffer (20% sucrose, 25mM Tris-HCl pH8 5mM di-sodium EDTA). They were lysed with 200µl of 0.2M NaOH 1% SDS for 10 minutes at room temperature. 30µl of phenol, which had been equilibrated with the lysis solution, were then added and the mixture was vortexed briefly. 150µl of 3M sodium acetate were added, the tubes were mixed by inverting, incubated at -20°C for 15 minutes and then centrifuged in a microfuge for 5 minutes. The supernatant was recovered and the DNA precipitated with ethanol. The pellet was resuspended and precipitated with ethanol a second time then washed with 70% ethanol. The final pellet was dried and resuspended in 50µl of water. 10µl were used for endonuclease digestion to check that the constructs were intact. If there was insufficient plasmid DNA for restriction analysis, or the quality of the DNA was poor, it was electroporated back into *E. coli* and DNA from these was analysed.

2.4 Sequencing

The dideoxynucleotide chain termination method (Sanger, 1977) was used to sequence dsDNA templates prepared by the CTAB method (Del Sal *et al.*, 1989) or by the alkaline lysis method (Sambrook *et al.*, 1989) with a phenol extraction step included. It was found that reliable sequence was obtained

when the plasmid was grown in the *E. coli* strain JM109. The Sequenase II DNA sequencing kit (US Biochemicals) was used. The manufacturer's instructions were generally followed except that a boiling method was used to denature the plasmid DNA (Andersen *et al.*, 1992).

Sequence data were analysed using the programs of the University of Wisconsin Genetics Computer Group Inc. sequence analysis software package (Genetics Computer Group, 1991).

2.5 Agroinfection

2.5.1 Injection

Agroinoculation by injecting a solution of agrobacteria containing infectious clones of viral genomes is used routinely for geminiviruses (Grimsley *et al.*, 1987; Morris *et al.*, 1988; Briddon *et al.*, 1989). *A. tumefaciens* (strain AGL1) containing the dimeric constructs were scraped from fresh plates and mixed together in 200µl of sterile distilled water (Grimsley *et al.*, 1987). The mixture was drawn up into a 10µl Hamilton glass syringe and injected into pea (*Pisum sativm* var. Greenfeast) or subterranean clover (var. Mt Barker) plants at various points, particularly the stems and the growing points. The pea plants used were at about the 8 leaf stage, or about 10cm tall. The sub clover seedlings were inoculated within a week of being transplanted.

2.5.2 Vacuum infiltration

This method was based on that used by Bechtold *et al.* (1993) to transform *Arabidopsis*. *A. tumefaciens* (strain AGL1) containing the dimeric constructs were separately grown in flasks containing 150ml of MG/L (0.5% mannitol, 0.1% L-glutamic acid, KH_2PO_4, NaCl, $MgSO_4$, 0.5% bacto-tryptone

(Difco), 0.25% bacto-yeast extract, 1μg/l biotin, pH 7.0; Garfinkel and Nester (1980)) for 1.5-2 days. The bacteria were precipitated by centrifugation at 5000rpm for 10mins, and resuspended in IM (20mM NH_4NO_3, 3mM $CaCl_2$, 19mM KNO_3, 1.5mM $MgSO_4$, 1.25mM KH_2PO_4, 0.1mM H_3BO_3, 0.1mM $MnSO_4$, 0.03mM $ZnSO_4$, 5μM KI, 1μM Na_2MoO_4, 0.1μM $CuSO_4$, 0.1μM $CoCl_2$, 10μg/l 6-benzylaminopurine, 5% sucrose) at one third the initial volume. Two to three week old pea and subterranean clover seedlings were taken out of the soil, immersed in the suspension and subjected to a vacuum for 20-30 mins. The treated plants were replanted in soil and returned to the glasshouse.

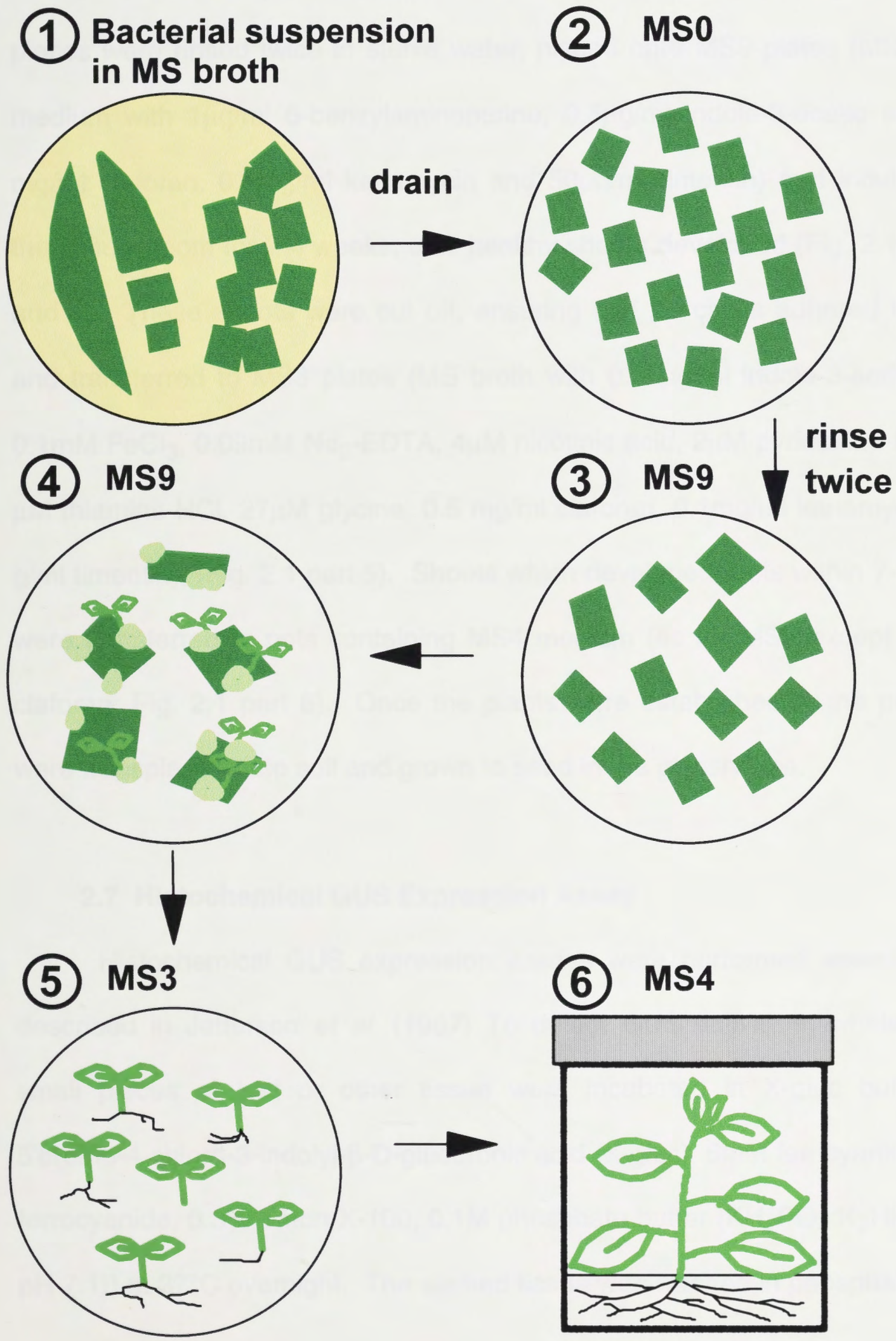
2.6 Tobacco transformation

Nicotinium tabacum cv. Wisconsin 38 plants were transformed by a method based on that of Horsch *et al.* (1985) which is illustrated in Figure 2.1. Agrobacteria containing the construct were scraped from a fresh plate and placed in 100ml of LB broth containing 1.5μg/ml tetracycline. The culture was grown at 28°C for 24 hours. The cell suspension was then centrifuged at 8000rpm for 5 mins. The pellet was resuspended in 25ml of MS broth (3.5% Murashige minimal organics medium powder (Gibco) in water) and centrifuged again. This pellet was resuspended in 10ml MS broth and placed in a high sided petri dish. Leaves from an actively growing tobacco plant raised in tissue culture (about 4 weeks old) were cut into 1cm squares in the suspension (Fig. 2.1 part 1). The pieces were left in the agrobacterial suspension for at least 5 mins, ensuring that all of the cut edges were wetted, and then placed on MS0 plates (20mM NH_4NO_3, 3mM $CaCl_2$, 19mM KNO_3, 1.5mM $MgSO_4$, 1.25mM KH_2PO_4, 0.1mM H_3BO_3, 0.1mM $MnSO_4$, 0.03mM $ZnSO_4$, 5μM KI, 1μM

Figure 2.1

Tobacco transformation

A summary of tobacco transformation, see text for details. The steps are numbered consecutively and the medium used at each step is indicated



Na_2MoO_4, 0.1μM $CuSO_4$, 0.1μM $CoCl_2$, 0.1mM $FeCl_3$, 0.09mM Na_2-EDTA, 87.6mM sucrose, 0.56mM *myo*-inositol, 4μM nicotinic acid, 2μM pyridoxine-HCl, 0.3μM thiamine-HCl, 27μM glycine), after draining off the excess moisture (Fig. 2.1 part 2).

The plates were incubated in a 25°C culture room for 2 days. The leaf pieces were rinsed twice in sterile water, placed onto MS9 plates (MSO plate medium with 1μg/ml 6-benzylaminopurine, 0.5μg/ml indole-3-acetic acid, 0.5 mg/ml claforan, 0.1mg/ml kanamycin and 50μg/ml timentin) and incubated in the culture room for 2-4 weeks, until healthy shoots developed (Fig. 2.1 parts 3 and 4). These shoots were cut off, ensuring that no callus adhered to them, and transferred to MS3 plates (MS broth with 0.05μg/ml indole-3-acetic acid, 0.1mM $FeCl_3$, 0.09mM Na_2-EDTA, 4μM nicotinic acid, 2μM pyridoxine-HCl, 0.3 μM thiamine-HCl, 27μM glycine, 0.5 mg/ml claforan, 0.1mg/ml kanamycin, 50μ g/ml timentin), (Fig. 2.1 part 5). Shoots which developed roots within 7-14 days were transferred to pots containing MS4 medium (as for MS3 except with no claforan; Fig. 2.1 part 6). Once the plants were established in the pots they were transplanted into soil and grown to seed in the glasshouse.

2.7 Histochemical GUS Expression Assay

Histochemical GUS expression assays were performed essentially as described in Jefferson *et al.* (1987) To detect GUS activity in whole tissue, small pieces of leaf or other tissue were incubated in X-gluc buffer (3% 5'bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-gluc), 5mM ferricyanide, 5mM ferrocyanide, 0.3% Triton X-100, 0.1M phosphate buffer (KH_2PO_4:K_2HPO_4, 1:2 pH 7.1)) at 37°C overnight. The stained tissue was washed in phosphate buffer

(0.14M NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4 pH 7.4) and cleared by rinsing in 25%, 50%, 70% ethanol and finally incubating in 95% ethanol for several hours. The gradual increase in ethanol concentration minimised tissue damage. Clearing was also achieved with bleach by gradually increasing the concentration of bleach to 50% and incubating for 10-20 minutes. Ethanol clearing was preferable because it did not remove any of the blue product but bleach was useful if the tissue had browned due to oxidation.

Stained tissue was photographed with a bright field using a Wild stereomicroscope (Heerbrugg, Switzerland) and Fujichrome 64 Tungsten film. For higher magnification, the stained stem pieces were fixed in 3% glutaraldehyde, embedded in LRWhite Resin C° (Basingstoke, UK) and sectioned on a Reichert Ultracut Microtome, by C. Miller. The sections were observed and photographed with a dark field under a Leitz Wetzlar Orthoplan microscope.

2.8 Fluorimetric GUS Expression Assay

GUS expression in transgenic plants was assayed fluorimetrically using methods based on those of Jefferson (1987). Leaf samples (0.1-0.2g) from sterile tissue culture plants were ground with a small amount of acid washed sand in 300-400µl of GUS extraction buffer (GEB): 50mM sodium phosphate pH7, 10mM EDTA, 0.1% Triton X-100, 0.1% Sarkosyl (N-laurylsarcosine), 10mM β-mercaptoethanol. The ground samples were centrifuged in a microfuge and the supernatant was collected.

The protein concentration was determined by a dye-binding method (Bradford, 1976). An aliquot of the supernatant from each sample was diluted

to 50µl with water and mixed with 200µl of Bradford reagent (Bio-rad) which had been diluted 1/4. Samples of bovine serum albumin (BSA) (0.1, 0.2, 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0µg) were used to create a standard curve. The absorbance at 595nm was measured and the protein concentration determined with the standard curve.

Aliquots of the leaf extract containing up to 10µg of protein were placed in wells of black microtitre plates made up to 50µl with GEB. 200µl of GEB containing 0.44mg/ml 4-methylumbelliferyl β-D-glucuronide (MUG) which had been pre-warmed to 37°C were added. Production of 4-methylumbelliferone (MU) was measured with a Labsystems Fluoroskan, which had been pre-warmed to 37°C, at 5 or 10 minute intervals over 60 minutes (excitation wavelength 355nm, emission wavelength 460nm). For quantitation, MU standards in the range 200 to 1000pmoles MU were added to the plate to create a standard curve.

2.9 Neomycin phosphotransferase Assay

Neomycin phosphotransferase (Npt) activity in transgenic tobacco was assayed by a dot blot method (McDonnell *et al.*, 1987). Approximately 0.2g samples of leaf tissue were ground with a small amount of acid washed sand in 200µl of extraction buffer (100mM Tris-HCl pH6.8, 10% glycerol, 5% ß-mercaptoethanol). The samples were centrifuged in a microfuge for 10mins and the supernatant was collected. In a microtitre dish, 15-20µl of reaction mixture (100mM Tris-HCl pH7.5, 50mM $MgCl_2$, 400mM NH_4Cl, 0.5mg/ml neomycin, 9mM KF, 20mM DTT, 1-2µl γ-^{32}P ATP) were mixed with an equal volume of plant extract and incubated at 37°C for 10-15mins. 20µl of each

reaction was blotted onto filter paper which had been blocked with 6.5mg/ml ATP, 60mM tetra-sodium pyrophosphate pH7.5 for 1 hour at 37°C. The blotted paper was washed in 1% SDS/proteinase K (about 0.1%) for 10-20mins at 60°C, then twice in 10mM phosphate buffer pH7.2 at 80°C for 2-5mins. When dry, the filter was scanned with a Molecular Dynamics Phosphorimager.

2.10 Polymerase Chain Reaction (PCR)

The PCR reaction mixture generally contained 1-2mM MgCl, 50mM Tris-HCl, 50mM KCl, 10mM DTT, 0.5mM dNTPs, about 0.1µg of each primer, and 0.2 units of Cetus Amplitaq. The PCR routine used for amplification of relatively abundant target sequences, such as plasmid or viral DNA, was 32 cycles of: denaturation at 95°C for 20 seconds; annealing between 60°C and 42°C (high to low stringency) for 40 seconds; and extension at 72°C for 1.5 to 2 minutes. The Corbett Research FTS-1 thermal sequencer was used.

2.11 ELISA

Enzyme linked immunosorbent assays (ELISA) were performed according to the methods of Chu *et al.* (1993b). Leaf samples (0.2g) were ground in 1ml of extraction buffer (140mM NaCl, 22mM KH_2PO_4, 81mM, 27mM KCl, 3mM NaN_3, 2% polyvinylpyrrolidine(PVP)). A 200µl aliquot of this ground material was added to each well of a microtitre plate, which had been coated in purified γ-globulin (see below), and incubated overnight at 4°C. The plate was then washed three times with PBS (140mM NaCl, 2.2mM KH_2PO_4, 8mM Na_2HPO_4, 3mM KCl, 3mM NaN_3) 0.05% Tween20 pH7.4. A 200µl aliquot of enzyme labelled γ-globulin (see below) in conjugate buffer (PBS 0.05%

Tween20, 2% PVP, 0.2% ovalbumin) was added to each well and the plate incubated for four hours at 37°C. It was then washed three times with PBS and 250µl of freshly prepared substrate solution (0.6mg/ml nitrophenyl phosphate in 3mM NaN_3, 9.7% diethanolamine pH9.8) were added to each well. The plate was incubated for 2 hours at room temperature and the optical density at 405nm was measured. The plate was stored overnight at 4°C and the absorbance was measured again.

The rabbit anti-SCSV γ-globulin (Chu and Helms, 1988) had been purified by precipitation with ammonium sulphate and dialysis against 0.5XPBS. The plates were coated by incubating for 4 hours at 37°C with 200µl of 1µg/ml of the purified γ-globulin in coating buffer (15mM Na_2CO_3, 35mM $NaHCO_3$, 3mM NaN_3, pH9.6) in each well. The enzyme labelled γ-globulin, provided by P. Chu, had been conjugated to alkaline phosphatase (Sigma) in 0.06% glutaraldehyde for four hours at room temperature and then dialysed against PBS.

2.12 Colony Hybridisation

Potential clones were picked and plated in a grid pattern on duplicate plates, one covered with a nitrocellulose filter. The plates were incubated overnight at 37°C. The filter was then removed and the DNA extracted from the colonies by placing it first on filter paper soaked in 10% SDS for 3 mins, then 0.5M NaOH, 1.5M NaCl for 5 mins, then 1.5M NaCl, 0.5M Tris-HCl pH7.4 for 5 mins, and finally, 2XSSC (0.3M NaCl, 30mM sodium citrate) for 5 mins. The filter was dried and then baked for 2 hours at 80°C in a vacuum oven. The filter was then washed in 2XSSC, scraping gently to remove the bacterial debris, and probed overnight at 50°C in 6XSSC, 0.1% Ficoll, 0.1% polyvinylpyrrolidone,

0.1% BSA, 0.5% SDS, 100µg/ml salmon sperm DNA with a mixture of ^{32}P end-labelled oligonucleotides specific to SCSV components 1, 3, 4, 5, and 7 (Table 2.1) using Hybaid tubes and oven. After hybridisation, the filters were washed in 2XSSC, 1% SDS at 42°C for 45 mins.

Table 2.1

Sequences and positions of SCSV specific primers

The first number in the primer name indicates the SCSV component to which it binds.

Primer name	Position of 5' end [a]	Primer sequence (5'-3')
1-40mer	303	TTCCTTCTTCGCATCTCCAGCACCTTCTTCATATGAGTATG
3-40mer	445	AATTTCCTTCAAGTGCTCGTTCATAATCTTCTCCTTCAG
4-40mer	448	CATACTATCACAACAGAAGAAATCCTGTTGTGAAGACGCC
5-40mer	423	CGAATATTTTCTCCTTTGAGACCTCAGACCCTTTCTTCC
7-40mer	452	CACTTCTTCTGCTATCGCTTACAAGAACATCGTCGCTCAC

CHAPTER 3: COMPONENTS 2 AND 6

3.1 Introduction

The sequencing of the SCSV genome was undertaken as a first step in the detailed molecular characterisation of this virus. SCSV infected plant material has been collected from the southern states of Australia and the ACT (Chu *et al.*, 1995). The symptoms of infection varied extensively between and within locations. Some isolates caused severe symptoms including extreme stunting, reddening, leaf deformation, and chlorosis whereas plants infected with mild isolates were virtually indistinguishable from uninfected plants (Grylls and Butler, 1959; Chu *et al.*, 1995). Even a single isolate, taken from one plant, could give variable symptoms when passaged in subterranean clover plants in the glasshouse (Chu *et al.*, 1995). An isolate which caused severe and stable symptoms derived from an infected plant taken from Tamworth (NSW, Australia), called the F isolate, was chosen for sequencing. Initial cloning and sequencing data indicated that there were at least four different DNA components in the SCSV F isolate genome (Chu *et al.*, 1993a). Subsequently seven different components have been identified.

The nucleotide sequences of a DNA component from each of three SCSV-like viruses have been published (Rhode *et al.*, 1990; Harding *et al.*, 1993; Katul *et al.*, 1995). The proteins potentially encoded by these DNAs share 30-40% amino acid identity in pairwise comparisons. In addition, they contain a conserved nucleotide binding motif. This motif is found in a variety of viral replicative proteins, including the essential replication associated proteins (RAPs) of geminiviruses (Koonin and Ilyina, 1992). Two components of SCSV,

numbered 2 and 6, were found which had similarities to these components of the SCSV-like viruses. This chapter presents the sequences of those two components. The analyses of the sequences, and their relationships to the putative RAPs of the other SCSV-like viruses, and the RAPs of the geminiviruses are also presented.

SCSV is unusual in having two components, 2 and 6, which appear to be functionally related. To determine whether both of these components were essential for infection, their presence in other glasshouse and field isolates of SCSV was assessed.

3.2 Results

3.2.1 Cloning and sequencing

Full length and partial clones of genomic DNAs of the SCSV F isolate were created from purified replicative form DNA as described in Chu *et al.* (1993a). Subclones of the full length clones were made to facilitate sequencing of internal regions not readily accessible from the ends of the clones. Sequencing was done on either ssDNA templates from M13 clones, or dsDNA templates from plasmid clones. To obtain the entire sequence of component 6, a full length copy was amplified by PCR with purified RF DNA (Chu *et al.*, 1993a), using the 6Spe primers which extend outwards from the *Spe*I site at nucleotide 66 (Table 3.1). The amplified full length DNA was cloned into the *Sma*I site of pGEM7Z and sequenced. The sequences of both strands of the SCSV DNAs were obtained with an average of seven fold redundancy.

The sequences of components 2 and 6 are presented in Figure 3.1.

Table 3.1

Positions and sequences of, and fragment sizes produced by PCR primers specific to components 2 and 6, and the common region primers, SUNI3 and 5

The primers Bam2 and Bgl2 are specific to component 2. The primers 6SpeS, 6SpeC, Nsi6, and Spe6 are specific to component 6.

Primer 1 name	Position of 5' end[b]	Primer 1 sequence (5'-3')	Primer 2 name	Position of 5' end[b]	Primer 2 sequence (5'-3')	Fragment size (bp)
6SpeS[a]	65	CCACACACTAGTTGGGTGTTCACAC	6SpeC[a]	71	CCACACACTAGTGCTTTTGTCTAG	1035
Bam2	390	CAAGCGGAAGACGATGG	Bgl2	847	GAACTAACTTCCTAATAC	457
Nsi6	550	CTATAATCTGGTTCTTCC	Spe6	42	CCCACTATGCCTACTAGAC	508
SUNI5	93 or 94[c]	TCTGTGATGCTAGTGAAGCG	SUNI3	92 or 93[c]	TCCTGAGCGTCCACGTGTCA	988-1002[c]

[a] non-homologous extra bases added at 5' end

[b] position in F isolate sequence

[c] the SUNI primers are non-specific and bind to several different components (Chapter 4)

Figure 3.1

Sequence of SCSV components 2 and 6.

The sequence of each component is listed separately, with the deduced amino acid sequence of the major ORF shown above. The stem-loop sequence is marked by the arrows, the nonanucleotide in the loop is underlined, the potential TATA boxes and polyadenylation signals are in **bold**, and * denotes the translation stop codon. The amino acids comprising the nucleotide binding motif (Walker *et al.*, 1982) are shown in ***bold italics***.

SCSV Component 2

```
                  <----------
     TAGTATTACCCGACCTTGCCACACCTCCTTGGAACACTTTCTCTCTCTAGAAAGTGTGAG
   1 ---------+---------+---------+---------+---------+---------+ 60

                          M  A  R  R  Y  C  F  T  L  N  Y  A  T
     ACTTTCTCTCTCTAAGCTTATATGGCTAGAAGGTACTGTTTTACATTAAATTACGCTACT
  61 ---------+---------+---------+---------+---------+---------+ 120

     E  I  E  R  E  T  F  L  S  L  F  S  Q  D  E  L  N  Y  F  V
     GAGATAGAGAGAGAAACATTCCTCTCCCTCTTCTCTCAAGACGAATTAAACTATTTCGTT
 121 ---------+---------+---------+---------+---------+---------+ 180

     V  G  D  E  T  A  T  T  G  Q  K  H  L  Q  G  F  V  S  F  K
     GTCGGCGACGAAACTGCAACTACTGGACAGAAACACCTCCAGGGATTTGTATCGTTCAAG
 181 ---------+---------+---------+---------+---------+---------+ 240

     N  K  I  R  L  G  G  L  K  K  K  F  G  N  R  A  H  W  E  I
     AACAAAATTCGTCTTGGTGGATTGAAGAAGAAATTTGGTAATCGAGCTCACTGGGAAATT
 241 ---------+---------+---------+---------+---------+---------+ 300

     A  R  G  S  D  S  Q  N  R  D  Y  C  C  K  E  T  L  I  S  E
     GCGAGAGGCAGCGATTCTCAGAATCGCGATTATTGCTGTAAAGAAACCCTAATTTCTGAA
 301 ---------+---------+---------+---------+---------+---------+ 360

     I  G  I  P  V  M  K  G  S  N  K  R  K  T  M  E  I  Y  E  E
     ATTGGGATTCCGGTCATGAAGGGTTCGAACAAGCGGAAGACGATGGAGATTTATGAAGAG
 361 ---------+---------+---------+---------+---------+---------+ 420

     D  P  E  E  M  Q  L  K  D  P  D  T  A  L  R  C  K  A  K  K
     GATCCCGAAGAAATGCAATTGAAGGATCCAGATACTGCTCTTCGATGTAAGGCGAAGAAA
 421 ---------+---------+---------+---------+---------+---------+ 480

     L  K  E  E  Y  C  S  C  Y  D  F  Q  K  L  R  P  W  Q  I  E
     TTGAAAGAGGAATATTGTTCCTGTTATGATTTTCAGAAACTCCGTCCATGGCAAATTGAG
 481 ---------+---------+---------+---------+---------+---------+ 540

     L  H  A  A  L  M  A  E  P  D  D  R  S  I  I  W  V  Y  G  S
     CTTCACGCGGCTTTAATGGCGGAACCAGATGATCGGAGTATCATCTGGGTCTATGGTTCA
 541 ---------+---------+---------+---------+---------+---------+ 600

     D  G  G  E  G  K  T  S  F  A  K  E  L  I  R  Y  G  W  F  Y
     GACGGAGGAGAAGGAAAGACGAGCTTCGCGAAGGAATTAATCAGGTATGGATGGTTTTAT
 601 ---------+---------+---------+---------+---------+---------+ 660

     T  A  G  G  K  T  Q  D  V  L  Y  M  Y  A  Q  D  P  E  R  N
     ACAGCCGGAGGGAAGACCCAGGACGTATTATATATGTATGCTCAAGACCCAGAGAGGAAT
 661 ---------+---------+---------+---------+---------+---------+ 720
```

```
    I  A  F  D  V  P  R  C  S  S  E  M  M  N  Y  Q  A  M  E  M
    ATTGCGTTTGATGTTCCCAGGTGTTCTTCGGAGATGATGAACTATCAGGCGATGGAGATG
721 ---------+---------+---------+---------+---------+---------+ 780

    L  K  N  R  V  F  A  S  T  K  Y  R  P  V  D  L  C  I  R  K
    TTGAAGAACAGAGTTTTTGCAAGTACAAAATATAGGCCTGTAGATCTTTGTATTAGGAAG
781 ---------+---------+---------+---------+---------+---------+ 840

    L  V  H  L  I  V  F  A  N  V  A  P  D  P  T  R  I  S  E  D
    TTAGTTCATTTAATTGTGTTTGCCAACGTGGCACCTGACCCCACGCGCATAAGTGAGGAC
841 ---------+---------+---------+---------+---------+---------+ 900

    R  L  V  I  I  N  C  *
    AGACTTGTAATTATCAATTGTTGAATAAAAGAATATATATTATTGTTTTAATTTAATTCC
901 ---------+---------+---------+---------+---------+---------+ 960

                                                      ------->
    GCGAAGCGGTAGCCGGTCATAACACTGTTGCCCTTGGAACACTATATATAGCAAGGTCGG
961 ---------+---------+---------+---------+---------+---------+1020

    CT
1021 -- 1022
```

SCSV Component 6

```
                <-------                                  M  P  T  R  Q
    CAGTATTACCGCACCTCGCTTACCCTCCTCGCTTCCCTGGGCCCACTATGCCTACTAGAC
  1 ---------+---------+---------+---------+---------+---------+ 60

      S  T  S  W  V  F  T  L  N  F  E  G  E  I  P  I  L  P  F  N
    AAAGCACTAGTTGGGTGTTCACACTTAACTTTGAGGGCGAAATTCCTATTTTGCCCTTTA
 61 ---------+---------+---------+---------+---------+---------+ 120

      E  S  V  Q  Y  A  C  W  Q  H  E  R  V  G  H  D  H  L  Q  G
    ATGAAAGCGTTCAGTACGCTTGTTGGCAGCATGAGAGAGTGGGACACGATCATTTACAGG
121 ---------+---------+---------+---------+---------+---------+ 180

      F  I  Q  F  K  S  R  N  T  T  L  R  Q  A  K  Y  I  F  N  G
    GATTTATACAATTTAAATCCCGCAACACTACATTGCGTCAGGCTAAGTATATTTTTAATG
181 ---------+---------+---------+---------+---------+---------+ 240

      L  N  P  H  L  E  I  A  R  D  V  E  K  A  Q  L  Y  A  M  K
    GACTGAATCCTCATCTGGAAATTGCTAGGGATGTAGAGAAGGCGCAATTGTACGCGATGA
241 ---------+---------+---------+---------+---------+---------+ 300

      E  D  S  R  V  A  G  P  W  E  Y  G  L  F  I  K  R  G  S  H
    AGGAAGATAGTAGAGTAGCTGGTCCCTGGGAGTATGGGTTGTTTATTAAGAGAGGATCGC
301 ---------+---------+---------+---------+---------+---------+ 360

      K  R  K  L  M  E  R  F  E  E  D  G  E  E  M  K  I  A  D  P
    ATAAGCGTAAGCTGATGGAGAGATTTGAAGAAGATGGAGAAGAGATGAAAATTGCTGATC
361 ---------+---------+---------+---------+---------+---------+ 420

      S  L  Y  R  R  C  L  S  R  K  M  A  E  E  Q  R  C  S  S  E
    CCTCTCTCTATAGGCGTTGTCTATCAAGGAAGATGGCTGAAGAACAACGTTGTTCTTCTG
421 ---------+---------+---------+---------+---------+---------+ 480

      W  N  Y  D  L  R  P  W  Q  E  E  V  M  H  L  L  E  E  E  P
    AGTGGAATTATGACTTACGCCCTTGGCAAGAAGAAGTGATGCATTTGTTAGAGGAAGAAC
481 ---------+---------+---------+---------+---------+---------+ 540

      D  Y  R  T  I  I  W  V  Y  G  P  A  G  N  E  G  K  S  T  F
    CAGATTATAGAACGATAATCTGGGTGTATGGACCTGCTGGTAATGAAGGCAAATCTACAT
541 ---------+---------+---------+---------+---------+---------+ 600

      A  R  H  L  S  L  K  D  G  W  G  Y  L  P  G  G  K  T  Q  D
    TTGCAAGACATCTGTCATTGAAAGATGGTTGGGGTTATCTGCCTGGAGGAAAGACACAAG
601 ---------+---------+---------+---------+---------+---------+ 660
```

```
     M  M  H  L  V  T  A  E  P  K  N  N  W  V  F  D  I  P  R  V
   ATATGATGCATCTTGTGACTGCTGAGCCTAAGAATAATTGGGTATTTGACATACCCAGAG
661 ---------+---------+---------+---------+---------+---------+---------+ 720

     S  S  E  Y  V  N  Y  G  V  I  E  Q  V  K  N  R  V  M  V  N
   TTAGTTCAGAGTATGTGAATTATGGTGTAATAGAACAGGTTAAGAATAGGGTAATGGTGA
721 ---------+---------+---------+---------+---------+---------+---------+ 780

     T  K  Y  E  P  C  V  M  R  D  D  N  H  P  V  H  V  I  V  F
   ATACTAAGTATGAGCCATGTGTAATGCGGGATGATAATCATCCTGTTCATGTAATTGTGT
781 ---------+---------+---------+---------+---------+---------+---------+ 840

     A  N  V  L  P  D  L  G  K  L  S  E  D  R  I  K  L  I  R  C
   TTGCAAATGTACTCCCAGATTTGGGAAAATTAAGTGAAGATAGAATAAAATTAATTCGTT
841 ---------+---------+---------+---------+---------+---------+---------+ 900

     *
   GTTGAAAACTCTGCGAAGGCAGAAGTTATAAAAAAAATGTGTTTTGAGAGAAGTCCCACA
901 ---------+---------+---------+---------+---------+---------+---------+ 960

                                                  ------------>
   TCGGGTAGTTCGCGAAACAGGGTGAGGGAAGCGAGCAATATAAGGCGAGGTGCGTAT
961 ---------+---------+---------+---------+---------+---------+------- 1017
```

3.2.2 Identification of potential ORFs

Components 2 and 6 each contain one major ORF, in the virion sense, of 840 and 855 nucleotides respectively (Figure 3.1). This ORF of component 2 is flanked by a potential TATA box, 5'TATATATA3' (Joshi, 1987a; Messing, *et al.*, 1983), and a polyadenylation signal, 5'AATAAA3' (Messing, *et al.*, 1983). The main ORF of component 6 is associated with the potential TATA box, 5'TATAAG3' (Messing, *et al.*, 1983; Joshi, 1987a), and polyadenylation signal, 5'AATAAA3' (Messing, *et al.*, 1983; Joshi, 1987b). The polyadenylation signal of the component 2 ORF overlaps the translation termination codon and that of component 6 is 13 nucleotides upstream of the termination codon.

Several smaller ORFs were also detected using FRAMES (Figure 3.2). The largest of these in component 2 is 51 amino acids long and is encoded on the complementary strand from nucleotides 529 to 377. In component 6 the largest is 108 amino acids long and is encoded on the sense strand from nucleotides 937 to 243. The smaller ORFs are not associated with any obvious transcription signals. The two largest of these alternative ORFs from each component were compared to the database using FASTA and TFASTA. No significant matches were detected. The analyses presented in this chapter will focus on the main ORF in each component.

3.2.3 Similarities between components 2 and 6 and other viral genomes

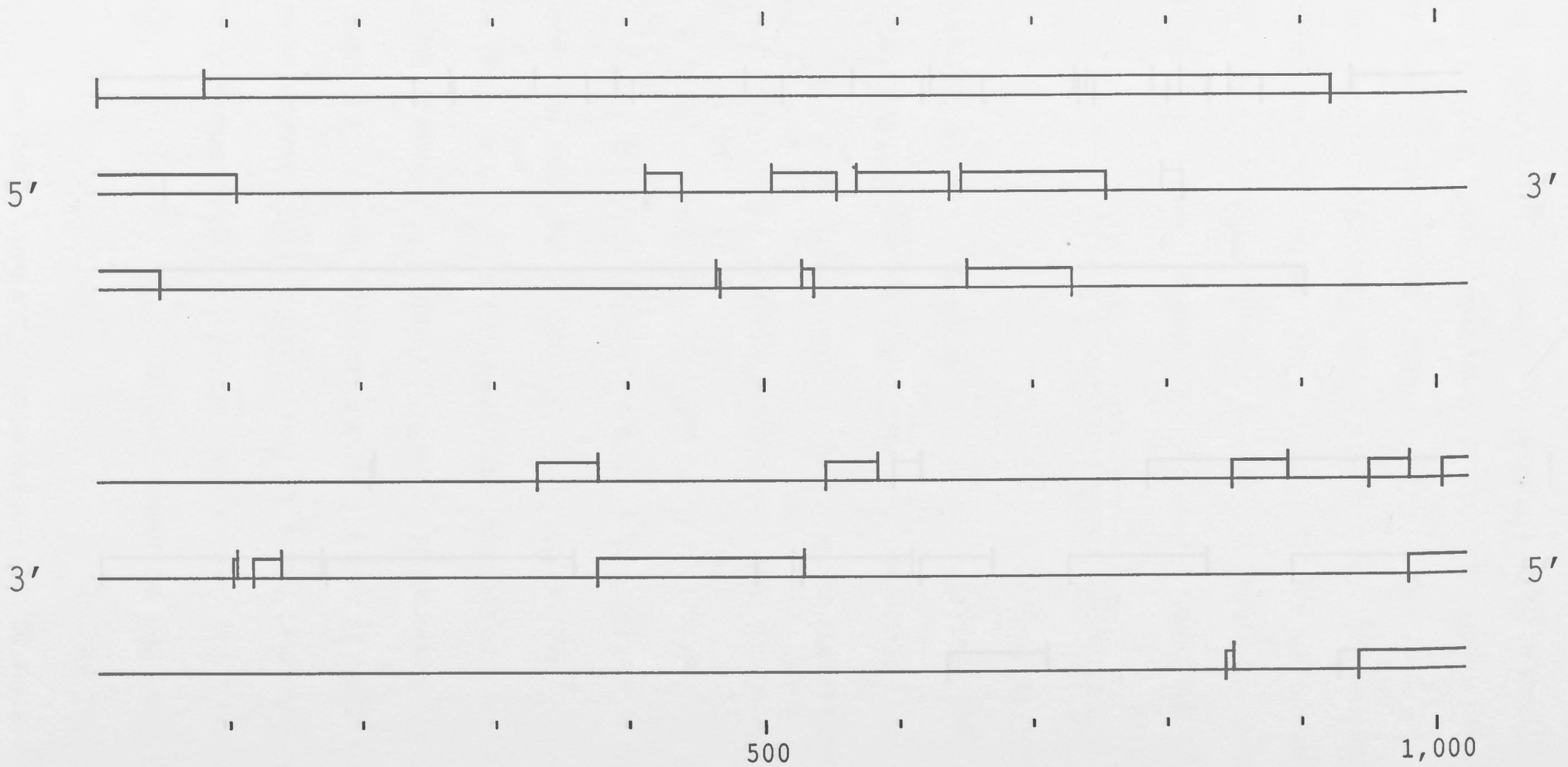
The non-coding regions of components 2 and 6 have a sequence capable of forming a hairpin structure with a GC rich stem and an AT rich loop (Figure 3.1) Within the loop sequences are the nonanucleotides, 5'TAGTATTAC3' (component 2) and 5'CAGTATTAC3' (component 6), which

Figure 3.2

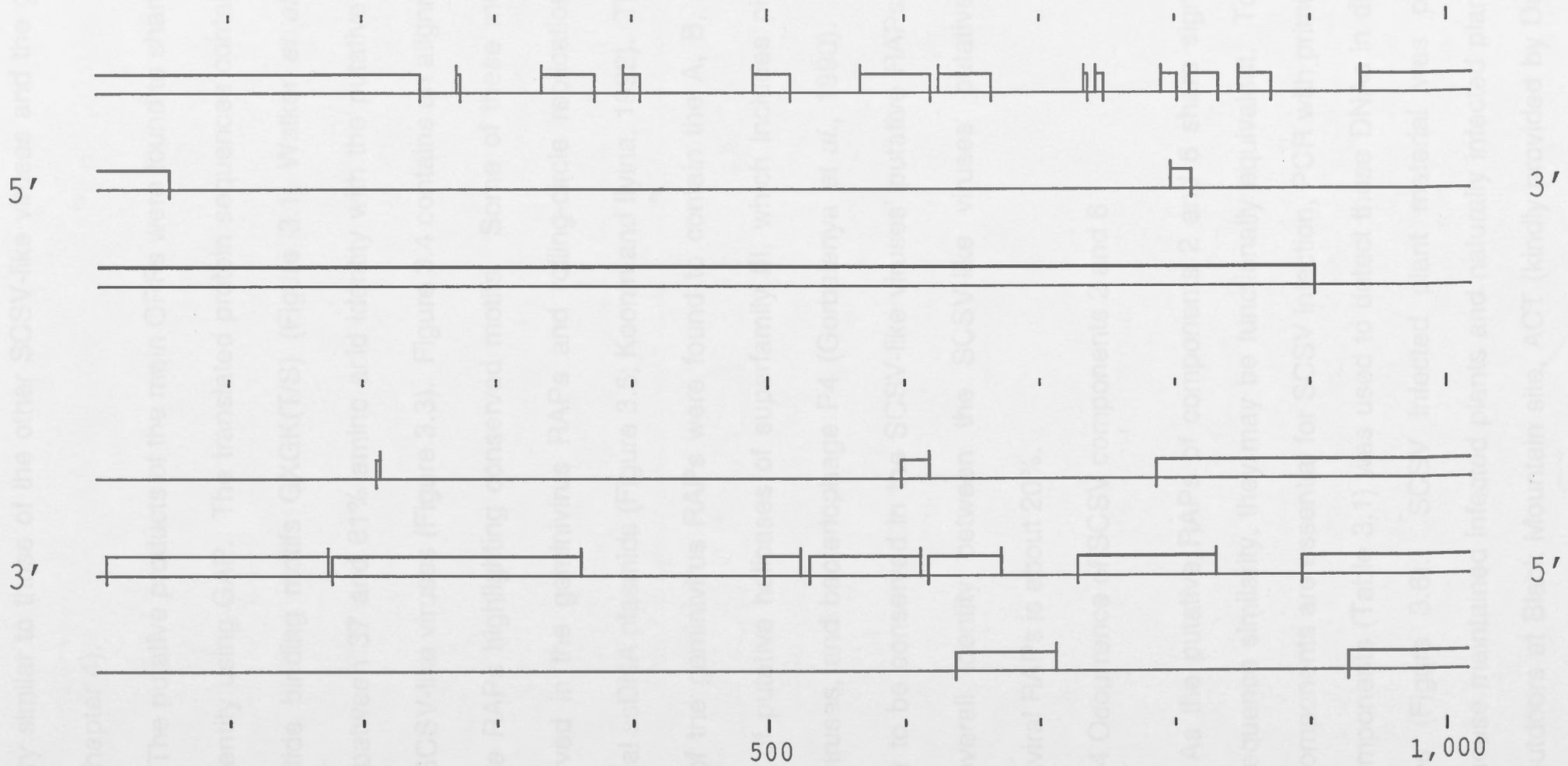
Open reading frames of SCSV components 2 and 6.

Open reading frames, represented by rectangles, in all frames of the virion (top three lines) and complementary (bottom three lines) sense strands of components 2 and 6 were detected using FRAMES. Marks extending above the ends of the rectangles indicate the positions of the initiation codons, and those below the line indicate stop codons. Redundant start and stop codons are not shown.

Open reading frames of SCSV Component 2



Open reading frames of SCSV Component 6



are very similar to those of the other SCSV-like viruses and the geminiviruses (see Chapter 4).

The putative products of the main ORFs were found to share 43% amino acid identity using GAP. The translated protein sequences contain consensus nucleotide binding motifs GXGK(T/S) (Figure 3.1; Walker *et al.*, 1982) and share between 37 and 61% amino acid identity with the putative RAPs of the other SCSV-like viruses (Figure 3.3). Figure 3.4 contains an alignment of these putative RAPs highlighting conserved motifs. Some of these motifs are also conserved in the geminivirus RAPs and rolling-circle replication proteins of bacterial ssDNA plasmids (Figure 3.5; Koonin and Ilyina, 1992). The C-terminal ends of the geminivirus RAPs were found to contain the A, B, and C motifs typical of putative helicases of superfamily III, which includes picornaviruses, parvoviruses, and bacteriophage P4 (Gorbalenya *et al.*, 1990). These motifs appear to be conserved in the SCSV-like viruses' putative RAPs (Figure 3.5). The overall identity between the SCSV-like viruses' putative RAPs and geminiviral RAPs is about 20%.

3.2.4 Occurrence of SCSV components 2 and 6

As the putative RAPs of components 2 and 6 share significant amino acid sequence similarity, they may be functionally equivalent. To determine if both components are essential for SCSV infection, PCR with primers specific to the components (Table 3.1) was used to detect these DNAs in different SCSV isolates (Figure 3.6). SCSV infected plant material was obtained from glasshouse maintained infected plants and naturally infected plants growing in pots outdoors at Black Mountain site, ACT (kindly provided by Dr. John Leigh, and therefore given the prefix JL). To identify which field samples contained

Figure 3.3

Percent identities between putative RAPs of SCSV-like viruses

The percent amino acid identities between the putative RAPs of the SCSV-like viruses: SCSV (components 2 and 6), FBNYV (Katul *et al.*, 1995), CFDV (Rhode *et al.*, 1990), BBTV (Harding *et al.*, 1993) and the 'satellite' BBTV DNAs type 1 and 2 (M. Karan, pers. comm.), were determined by GAP.

	SCSV6	BBTV1	CFDV	FBNYV	BBTVS-1	BBTVS-2
SCSV2	43	37	46	61	54	52
SCSV6		40	39	51	39	52
BBTV1			32	37	37	34
CFDV				38	42	47
FBNYV					42	42
BBTVS-1						62

Figure 3.4

Alignment of RAPs of SCSV-like viruses

Alignment, created using PILEUP, of the putative RAPs of the SCSV-like viruses: SCSV (components 2 and 6), FBNYV (Katul *et al.*, 1995), CFDV (Rhode *et al.*, 1990), BBTV (Harding *et al.*, 1993) and the 'satellite' BBTV DNAs type 1 and 2 (M. Karan, pers. comm.). Conserved residues are indicated in **bold**.

```
          1                                                  50
SCSV2     ....MARRYC FTLNYATEIE RETFLSLFSQ DELNYFVVGD ETA.TTGQKH
FBNYV     ...MACSNWV FTRNFQGAL. ..PLLSF..D ERVQYAVWQH ERG.T..HDH
BBTV-S1   ...MSSFKWC FTLNYSSAAE REDFLALLKE EELNYAVVGD EVAPSSGQKH
BBTV-S2   ...MSSFKWC FTLNYSSAAE REDFLALLKE EELNYAVVGD EVAPSTGRKH
CFDV      .MGSSIRRWC FTLNYETEEE AANVVRRIES LNLVYAIVGD EVAPSTGQRH
SCSV6     MPTRQSTSWV FTLNFEGEIP ILPF.....N ESVQYACWQH ERV...GHDH
BBTV1     .MARYVVCWM FTINNPTTLP .......VMR DEIKYMVYQV ERGQE.GTRH

          51                                                 100
SCSV2     LQGFVSFK.N KIRLGGLKKK FGN.RAHWEI ARGSDSQNRD YCCKETL...
FBNYV     IQGVIQLK.K KARFSTVKEI IGG.NPHVEK MKGTIEEASA YVQKEETRVA
BBTV-S1   LQGYLSLK.K SIKLGGLKKK YSS.RAHWER ARGSDEDNAK YCSKETL...
BBTV-S2   LQGYLSLK.K SIKLGGLKKR YSS.KAHWER ARGTDEQNRG YCSKETL...
CFDV      LQGFIHLK.T GRRLQGLKTV LGNDRIHLEP TRGSDEQNRD YCSKERV...
SCSV6     LQGFIQFKSR NTTLRQAKYI FNGLNPHLEI ARD.VEKAQL YAMKEDSRVA
BBTV1     VQGYVEMKRR S.SLKQMRGF FPG..AHLEK RKGSQEEARS YCMKEDTRIE

          101                                                150
SCSV2     .ISEIGIPVM KGSNKRKTME IYEEDPE...  ...EMQLKDP DTALRCKAK.
FBNYV     GPWSYGDLLK RGSHRRKTME RYLEDPE...  ...EMQLKDP DTALRCNAK.
BBTV-S1   .ILELGFPAS QGSNRRKLSE MVSRSPE...  ...RMRIEQP EIYHRYT...
BBTV-S2   .VLELGTPVV PGSKKRKLLE RFRESPE...  ...ELKMEDP SKYRRCL...
CFDV      .LLEHGVPTR PGVKRPRLAQ RFAEEPD...  ...ELRLEDP GGYRRCVVH.
SCSV6     GPWEYGLFIK RGSHKRKLME RFEEDGE...  ...EMKIADP SLYRRCLSR.
BBTV1     GPFEFGSF..  KLSCNDNLFD VIQDMRETHK RPLEYLYDCP NTFDRSKDTL

          151                                                200
SCSV2     .KLKEEYCSC YDFQKL....  RPWQIELHAA LMAEPDDRSI IWVYGSDGGE
FBNYV     .RLKEDFMKE KTKLQL....  RPWQKELHDL ILTEPDDRTI IWVYGPDGGE
BBTV-S1   ..SVKKLKKF KEEFVHPCLD  RPWQIQLTEA IDEEPDDRSI IWVYGPNGNE
BBTV-S2   ..AVESLNNA RKNSEWVHEL  REWQNKLIQH IEGVPDDRSI IWVYGPNGGE
CFDV      .GASVEWTRW AAENPFPFPY  HNWQLEVLSA IGEPADDRTI LWICGRDGGD
SCSV6     .KMAEE...Q RCSSEWNYDL  RPWQEEVMHL LEEEPDYRTI IWVYGPAGNE
BBTV1     YRVQAEMNKT KAMNSWRTSF  SAWTSEV.EN IMAQPCHRRI IWVYGPNGGE

          201                                                250
SCSV2     GKTSFAKEL.  IRYGWFYTAG GKTQDVLYMY .AQDPERNIA FDVPRCSSEM
FBNYV     GKSMFAKEL.  IKYGWFYTAG GKTQDILYMY .AQDPERNIA FDVPRCSSEM
BBTV-S1   GKSTYAKSL.  MKKDWFYTRG GKKENILFSY VDEGSEKHIV FDIPRCNQDY
BBTV-S2   GKSTFARYLS  LKPGWGYING GKTSDMMH.I ITMDPDNHWI IDIPRSHSDY
CFDV      GKSVFAKYLG  LKPDWFYTCG GTRKDVLYQY I.EDPKRNLI LDVPRCNLEY
SCSV6     GKSTFARHLS  LKDGWGYLPG GKTQDMMHL. VTAEPKNNWV FDIPRVSSEY
BBTV1     GKTTYAKHLM  KTRNAFYSPG GKSLDICRLY NYEDI...VI FDIPRCKEDY
```

```
          251                                                        300
SCSV2     MNYQAMEMLK NRVFASTKYR PVDLCIR.KL VHLIVFANVA PDPTRISEDR
FBNYV     MNYQAMEMMK NRCFASTKYR SVDLCCN.KN VHLVVFANVA YDPTKISEDR
BBTV-S1   LNYDVIEALK DRVIESTKYK PIKL.VELIN IHVIVMANFM PEFCKISEDR
BBTV-S2   LNYGVIEQIK NRVLINTKYE PCVIRKDGQN VHVIVMANVL PDYCKISEDR
CFDV      LNYALLECVK NRAFSSDKYE PLSY.LGFDH VHVLVFANVL PDYLKISRDR
SCSV6     VNYGVIEQVK NRVMVNTKYE PCVMRDDNHP VHVIVFANVL PDLGKLSEDR
BBTV1     LNYGLLEEFK NGIIQSGKYE PVLKIVEY.. VEVIVMANFL PKEGIFSEDR

          301
SCSV2     LVIINC
FBNYV     IVIINC
BBTV-S1   IKIIYC
BBTV-S2   IKIINC
CFDV      IKLWNI
SCSV6     IKLIRC
BBTV1     IKLVSC
```

Figure 3.5

Alignment of RAPs of SCSV-like viruses and geminiviruses

Alignment, created using PILEUP, of the putative RAPs of the SCSV-like viruses and two representative geminiviruses, African cassava mosaic virus (ACMV, bipartite; Stanley and Gay, 1983) and maize streak virus (MSV, Nigerian isolate fusion protein; monopartite; Mullineaux *et al.*, 1984). Conserved motifs are indicated in **bold**. Beneath the first three conserved motifs, which were identified by Koonin and Ilyina (1992) in rolling circle replication proteins, the consensus sequences of the same motifs from the pUB110 family of ssDNA bacterial plasmids are shown (Koonin and Ilyina, 1992). In the consensus sequences, U indicates a bulky hydrophobic residue, upper case letters represent residues conserved in all 10 of the plasmid sequences used whilst lower case letters represent residues conserved in all but one, and x indicates no consensus. The A, B, and C motifs of the putative helicase domain (Gorbalenya *et al.*, 1990) in the C-terminal part of the alignment are indicated below the aligned sequences.

```
          1                                                  50
ACMV      ....MRTPRF RVQAKNV**FLT** **Y**PNCSIPKEH LLSFIQTLSL PSNPKFIKIC
 MSV      MASSSSNRQF SHRNANT**FLT** **Y**PKCPENPEI ACQMIWELVV RWIPKYILCA
SCSV2     .......... .MARRYC**FTL** **N**YATEIERET FLSLFSQDEL NY   FVVGD
FBNYV     .......... MACSNWV**FTR** **N**FQGAL...P LLSF..DERV QY   AVWQH
SCSV6     .......MPT RQSTSWV**FTL** **N**FEGEI   P ILPF..NESV QY   ACWQH
 CFDV     ........MG SSIRRWC**FTL** **N**YETEEEAAN VVRRIESLNL VY   AIVGD
 BBTV     ........MA RYVVCWM**FTI** **N**NPTTLPVM. ......RDEI KY   MVYQV
pUB110                     FlT LTxxN

          51                                                 100
ACMV      RELHQN**G**EP**H** **LHALI**QFEGK ITITNNRLFD CVHPSCSTNF HPNIQGAK.S
 MSV      REAHKD**G**SL**H** **LHALL**QTEKP VRISDSRFFD I......NGF HPNIQSAK.S
SCSV2     ETA.TT**G**QK**H** **LQGFV**SFKNK .IRLGG.... ..LKKKFGN. RAHWEIARGS
FBNYV     ERG.T..HD**H** **IQGVI**QLKKK .ARFST.... ..VKEIIGG. NPHVEKMKGT
SCSV6     ERV...**G**HD**H** **LQGFI**QFKSR NTTLRQ.... ..AKYIFNGL NPHLEIARD.
 CFDV     EVAPST**G**QR**H** **LQGFI**HLKT. GRRLQG.... ..LKTVLGND RIHLEPTRGS
 BBTV     ERGQE.**G**TR**H** **VQGYV**EMKRR SS.LKQ.... ..MRGFFPG. .AHLEKRKGS
pUB110              H UHvLUxVxxy F

          101                                                150
ACMV      SSDVKS**Y**LD**K** DGD...TV.E WG.......Q FQIDGRSARG GQQSANDAYA
 MSV      VNRVRD**Y**IL**K** EPL...AVFE RGTFIPRKSP FLGKSDSEVK EKKPSKDEIM
SCSV2     DSQNRD**Y**CC**K** ETL....ISE IGIPVMKGSN K....RKTME IYEEDPEEMQ
FBNYV     IEEASA**Y**VQ**K** EETRVAGPWS YGDLLKRGSH R....RKTME RYLEDPEEMQ
SCSV6     VEKAQL**Y**AM**K** EDSRVAGPWE YGLFIKRGSH K....RKLME RFEEDGEEMK
 CFDV     DEQNRD**Y**CS**K** ERV....LLE HGVPTRPGVK R....PRLAQ RFAEEPDELR
 BBTV     QEEARS**Y**CM**K** EDTRIEGPFE FGSFKLSCND NLFDVIQDMR ETHKRPLEYL
pUB110      ExxKYxxK xxDU
```

```
          151                                                       200
ACMV      KALNS..GSK SEALNVIREL V.....PKDF VLQFHNLNSN LDRIFQEPPA
 MSV      RDIISHATSK EEYLSMIQKE L.....PFDW STKLQYFEYS ANKLFPEIQE
SCSV2     LKDPDTALRC KAK..KLKEE YC...SCYDF Q.KLRPWQIE LHAALMAEPD
FBNYV     LKDPDTALRC NAK..RLKED FM...KEKTK L.QLRPWQKE LHDLILTEPD
SCSV6     IADPSLYRRC LSR..KMAEE QR...CSSEW NYDLRPWQEE VMHLLEEEPD
 CFDV     LEDPGGYRRC VVH..GASVE WTRWAAENPF PFPYHNWQLE VLSAIGEPAD
 BBTV     YDCPNTFDRS KDTLYRVQAE MNKTKAMNSW RTSFSAWTSE VEN.IMAQPC

          201                                                       250
ACMV      PYVSPFPCSS FDQVPDE.LE EWVADNVRDS AARPWRPNSI VIEGDSRTGK
 MSV      EFTNPHPPSS PDLLCNESIN DWLQPNIFQS SDERSRKQSL YIVGPTRTGK
SCSV2     D......... .......... .......... .....RSIIW VYGSDGGEGK
FBNYV     D......... .......... .......... .....RTIIW VYGPDGGEGK
SCSV6     Y......... .......... .......... .....RTIIW VYGPAGNEGK
 CFDV     D......... .......... .......... .....RTILW ICGRDGGDGK
 BBTV     H......... .......... .......... .....RRIIW VYGPNGGEGK
                                                    ---------A----
          251                                                       300
ACMV      TIWARSLGP. HNYLCGHLDL SPKV...FNN DAWYNVIDDV D..PHYLKHF
 MSV      STWARSLGV. HNYWQNNVDW SS.....YNE DAIYNIVDDI P..FKFCPCW
SCSV2     TSFAKEL.IR YGWFYTAGGK TQDVLYMYAQ DPERNIAFDV PRCSSEMMNY
FBNYV     SMFAKEL.IK YGWFYTAGGK TQDILYMYAQ DPERNIAFDV PRCSSEMMNY
SCSV6     STFARHLSLK DGWGYLPGGK TQDMMHLVTA EPKNNWVFDI PRVSSEYVNY
 CFDV     SVFAKYLGLK PDWFYTCGGT RKDVLYQYIE DPKRNLILDV PRCNLEYLNY
 BBTV     TTYAKHLMKT RNAFYSPGGK SLDICRLY.. NYEDIVIFDI PRCKEDYLNY
          -------                                  --B--
          301                                                       350
ACMV      KEFMGSQRDW QSNTKYGKPV Q.I.KGGIPT IFLCN..PGP TSSYKEFLDE
 MSV      KQLVGCQRDF IVNPKYGKKK K.VQKKSKPT IILAN..SD. ....EDWMKE
SCSV2     QAMEMLKNRV FASTKYRPVD LCIRKL.VHL IVFANVAPDP TRISEDRLVI
FBNYV     QAMEMMKNRC FASTKYRSVD LCCNKN.VHL VVFANVAYDP TKISEDRIVI
SCSV6     GVIEQVKNRV MVNTKYEPCV MRDDNHPVHV IVFANVLPDL GKLSEDRIKL
 CFDV     ALLECVKNRA FSSDKYEPLS .YLGFDHVHV LVFANVLPDY LKISRDRIKL
 BBTV     GLLEEFKNGI IQSGKYEPVL KIVEY..VEV IVMANFLPKE GIFSEDRIKL
                                             --C--
          351                                                  393
ACMV      EKQEALKAWA LKNAIFI.TL TEPLYSGSNQ SQSQTIQEAS HPA
 MSV      MTPGQLEYFE ANCIIYIMSP GEKWYSPPEL PPTEAVHSDR S..
SCSV2     INC
FBNYV     INC
SCSV6     IRC
 CFDV     WNI
 BBTV     VSC
```

Figure 3.6

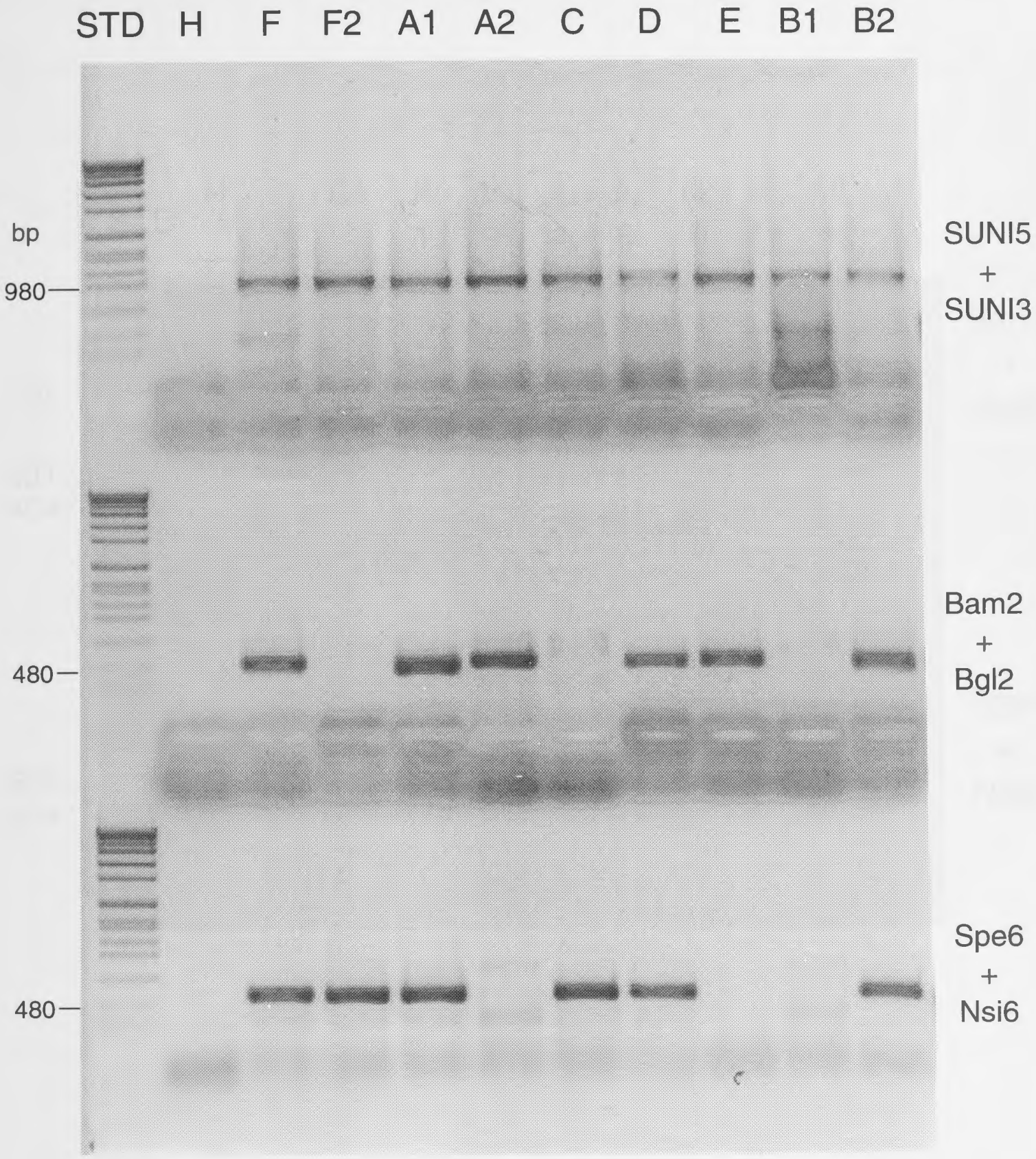
Detection of SCSV components 2 and 6 in different SCSV isolates

a) Detection of components 2 and 6 in glasshouse maintained isolates.

b) Detection of components 2 and 6 in ELISA positive field samples.

PCR was performed on subterranean clover nucleic acid extracts using the component 2 specific primers Bam2 and Bgl2, the component 6 specific primers Spe6 and Nsi6, and the common region primers SUNI3 and SUNI5. (See Table 3.1 for sequences and positions of primers and expected fragment sizes.) The PCR program was: one cycle of 30s at 96°C then 32 cycles of: 20s at 96°C, 40s at 52°C, 90s at 72°C. 6µl of each reaction were electrophoresed on a 1.2% agarose gel, stained with ethidium bromide and visualised under UV. 0.5µg samples of the DNA size markers, pUC19 digested with *Hpa*II or bacteriophage SPP-1 digested with *Eco*RI were loaded in the lanes marked STD. Relevant fragment sizes (bp) of the standards are indicated. The test samples are: uninfected (H), F isolate (F), glasshouse isolates F2, A1, A2, C, D, E, B1, and B2, ELISA positive field samples JL3, JL4, JL8, JL9, JL11, JL12, and JL13, and an ELISA negative sample JL1.

a.


STD
H
F
F2
A1
A2
C
D
E
B1
B2
bp
980
480
480
SUNI5
+
SUNI3
Bam2
+
Bgl2
Spe6
+
Nsi6

b.

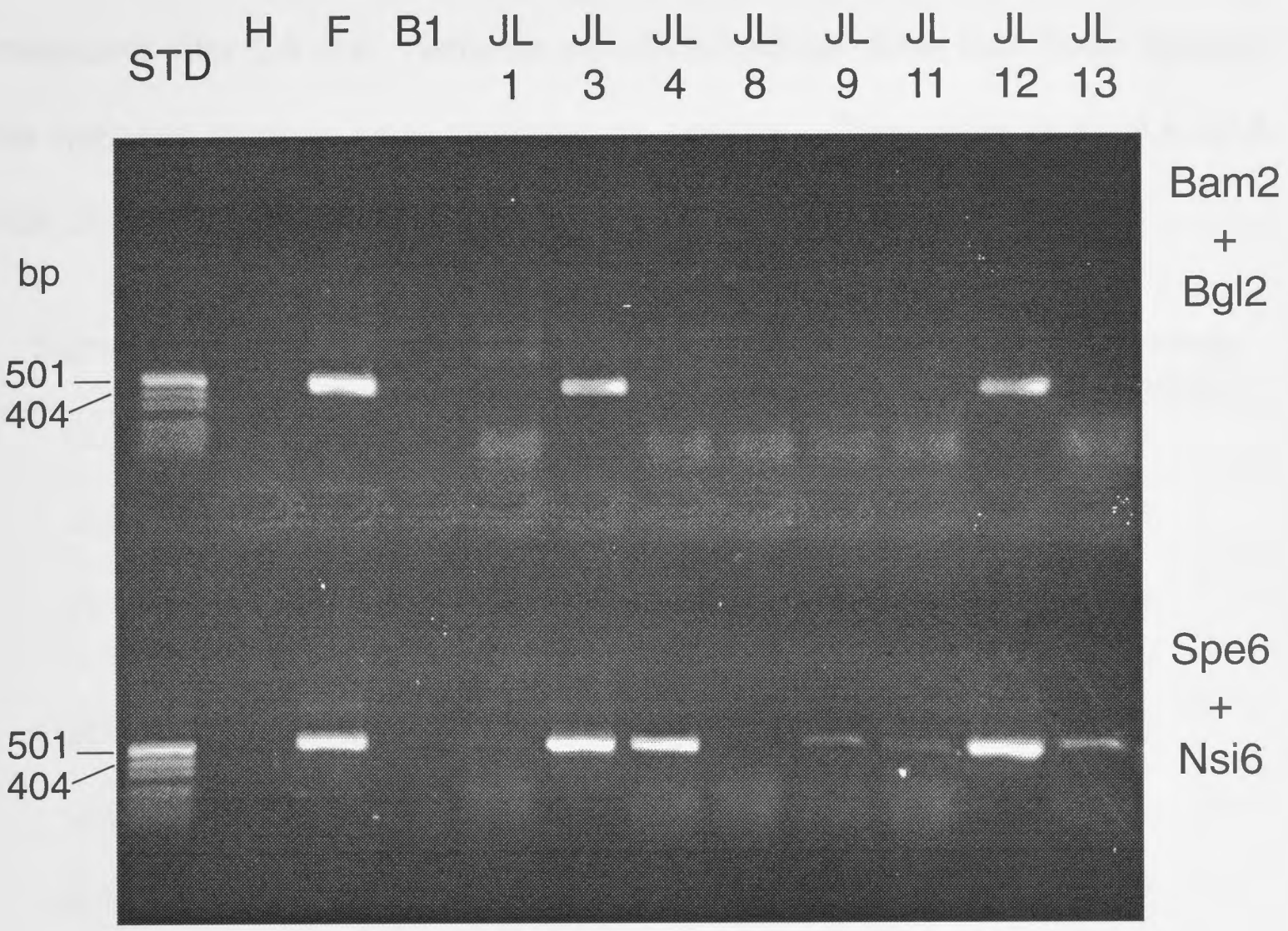

STD
H
F
B1
JL 1
JL 3
JL 4
JL 8
JL 9
JL 11
JL 12
JL 13
bp
501
404
501
404
Bam2 + Bgl2
Spe6 + Nsi6

Table 3.2

Symptoms of field infected plants and ELISA results

Symptoms and ELISA values are given for the field samples and controls: F, B1, and healthy from the glasshouse, which were used for PCR detection of components 2 and 6 (and component 5, see Chapter 5). ELISA was performed as described in Methods 2.11. The optical densities (OD) at 405nm were measured after 2.5 hrs. Samples with OD readings more than twice those of the negative controls were regarded as positive. They were, JL3, JL4, JL8, JL9, JL11, JL12, JL13, B1, and F.

Sample	Symptoms	ELISA reading (OD 405nm)
JL1	marginal reddening of old leaves, yellowing of young leaves	0.05
JL3	reddening and rolling of old leaves, stunting and pinching of young leaves	0.53
JL4	reddening of old leaves, small cupped young leaves	0.09
JL8	clumping and bright reddening of old leaves, stunting of shoots	0.56
JL9	bright reddening and elongation of leaflets, stunted growth	0.16
JL11	reddening of old leaves, puckering and marginal chlorosis of young leaves, stunting of shoots	1.57
JL12	new infection, some reddening, stunting, puckering and yellowing on new shoots	1.57
JL13	reddening of a few older leaves	0.21
-ve control	healthy	0.04
+ve control (B1)	clumping, deformation and yellowing of leaves, stunted growth	0.51
+ve control (F)	leaf deformation, reddening of old leaves and yellowing of young leaves, extremely stunted growth	1.53

SCSV, part of each sample was used for enzyme linked immunosorbent assays (ELISA) as described in Methods 2.11 (Table 3.2). DNA for use in PCR was extracted from the ELISA positive samples, JL3, JL4, JL8, JL9, JL11, JL12, and JL13, and one ELISA negative sample as a control (JL1).

In the glasshouse samples components 2 and 6 were not always present together in the same plant. Isolate B1 appeared to have neither 2 nor 6, although it was determined to be infected with SCSV by ELISA (Table 3.2) and by PCR with primers to other components (Figure 3.6; see also Chapter 5). From the PCRs with naturally infected plant samples, it appeared that several of the field isolates had only component 6 (Figure 3.6), indicating that isolates with only one of the two components also occur in nature. The lower amount of PCR product from isolates JL9, JL11, and JL13 may have been due to the presence of greater amounts of inhibitors in those preparations, or to sequence mismatches in the primer binding sites making the PCR inefficient. No amplified fragment was produced from JL8 with either set of primers. This was probably also due to the quality of the JL8 preparation as no other PCRs with it were successful (Chapter 5).

3.3 Discussion

Components 2 and 6 of SCSV F isolate were sequenced and found to each contain a single large ORF in the virion sense. These ORFs had the potential to encode related RAPs similar to those of the other SCSV-like viruses and the geminiviruses (Figure 3.3, 3.4 and 3.5).

It is unlikely that any of the smaller ORFs detected by FRAMES encode functional proteins. The smaller ORFs were not flanked by obvious

transcription signals. Some of the sense ORFs could be expressed from the same transcript as the main ORF, but the largest alternative ORF of component 6 is partially outside of the predicted main RNA transcript. This and the complementary sense ORFs would require separate transcripts. Furthermore, the encoded amino acid sequences of these small ORFs do not match any sequences in the database. Finally, none of the small ORFs were conserved in sequence or position between the two components, which may have indicated functional importance. Whether or not any of the smaller ORFs encode functional proteins requires analysis of transcription and translation products of the components.

The SCSV DNA components 2 and 6 appear to encode complete genes and, must therefore, include all of the *cis*-regulatory signals for transcription within the 1kb sequence of each component. As SCSV genes have evolved to function in plants, exploiting the host transcription machinery, many of these *cis*-regulatory elements must be equivalent to plant signals. This makes SCSV genes some of the most compact eukaryotic-like genes. The main ORF of each of components 2 and 6 occupy most of the available DNA sequence; the noncoding regions are only 179 and 158b long, respectively. It is likely, therefore, that the coding region also contains some regulatory sequences. Indeed, the polyadenylation signal of component 6 appears to be within the coding sequence, and that of component 2 overlaps the termination codon. Regulatory sequences consisting of repeated motifs, 5'UUUGUA3' and 5'CAGUGU3', have been found upstream of the polyadenylation signal of the CaMV transcription unit (Sanfaçon *et al.*, 1991; Rothnie *et al.*, 1994) and maize 27-kDa zein gene (Wu *et al.*, 1994b), respectively. These sequences were

found to be important determinants of the mRNA 3'-end processing efficiency of these genes. If equivalent upstream regulatory elements are present in the SCSV genes in components 2 and 6, they would have to be within the coding regions of both components. The related components of the other SCSV-like viruses also have potential polyadenylation signals upstream of the termination codon (Rhode *et al.*, 1990; Harding *et al.*, 1993; Katul *et al.*, 1995).

The transcripts produced from components 2 and 6 would not be expected to be much larger than the coding sequence. Polyadenylation signals are generally 10-40nts upstream from the polyadenylation site (Mogen *et al.*, 1990; MacDonald *et al.*, 1991; Sanfaçon *et al.*, 1991; Mogen *et al.*, 1992; Wu *et al.*, 1993; Sanfaçon, 1994); the poly-A tails of the main transcripts of components 2 and 6 would be added shortly after the termination codons. TATA boxes were found to be present an average of 32±7nts upstream from the transcription initiation sites of more than 75 plant genes (Joshi, 1987a). The potential TATA boxes of components 2 and 6 are only 95 and 61b upstream of the initiation codon; therefore, the leader sequences of the transcripts would be about 65 and 30b respectively.

The similarity between the SCSV-like viral and geminiviral RAPs (Figure 3.5) and the conservation of the stem-loop structure and nonanucleotide, which in geminiviruses is part of the origin of replication (Heyraud *et al.*, 1993; Stanley *et al.*, 1995), suggest that SCSV-like viruses may use a similar mechanism of rolling-circle replication as the geminiviruses (Chapter 1; Saunders *et al.*, 1991). The N-terminal half of the proteins contain the three motifs identified in a variety of rolling circle replication proteins (Koonin and Ilyina, 1992). The conserved Tyr is thought to be involved in the nicking of the DNA (Koonin, 1993). The N-

terminal half of the TYLCV RAP and the sequences 5' of the stem-loop have been shown to be involved in strain specific replication (Jupin *et al.*, 1995). Therefore, the N-terminal half of the geminivirus RAP (and by analogy, the SCSV-like viruses' putative RAPs) may also be a sequence specific binding domain.

By homology with the geminiviruses, SCSV is predicted to localise in the nucleus of infected plant cells. The proteins required for SCSV replication would have to be imported into the nucleus. Nuclear localisation signals (NLSs) are generally found in nuclear targeted proteins (for a review see, Raikhel, 1992). Both SCSV RAPs have potential NLSs of the bipartite type (Robbins *et al.*, 1991); **K**N**K**I**R**LGGL**KKK** in the component 2 RAP and **KR**GSH**KRK** in the component 6 RAP (the basic residues which define NLSs are indicated in bold). Patrick (1994) fused the N-terminal 149 amino acids of the component 2 ORF, which contains the NLS mentioned above, to the ß-glucuronidase reporter gene (GUS). When this construct was electroporated into tobacco protoplasts, GUS expression localised to the nucleus (Patrick, 1994). Neither the sequences nor the positions of the NLSs of the two SCSV RAPs are conserved; however, this may be of no functional consequence as the sequences of NLSs are not well conserved (Raikhel, 1992). Furthermore, there are other groups of basic amino acids within the SCSV RAPs which may also function as NLSs and their presence may be redundant. Nuclear targeted proteins commonly have several potential NLSs which may not all be functional or which differ in efficiency (Raikhel, 1992; Varagona *et al.*, 1992; Ingham *et al.*, 1995). In the *Agrobacterium* VirE2 protein, two NLSs act together to direct the protein to the nucleus (Citovsky *et al.*, 1992).

The relatedness of the potential products of components 2 and 6 is no greater than the relatedness of either of them to the putative RAPs of other SCSV-like viruses. Indeed, the similarity of the FBNYV putative RAP to both SCSV RAPs is greater than the similarity shared between the SCSV RAPs themselves (Katul *et al.*, 1995; Figure 3.3). Furthermore, there is a single nucleotide difference between components 2 and 6 in the otherwise highly conserved nonanucleotide in the stem-loop. This suggests that components 2 and 6 may have come from two different viruses in the SCSV-like group. Another possibility is that the different RAP encoding DNAs have diverged from a common ancestor and are maintained because they have developed slightly different functions or are advantageous in different conditions, for example, in different hosts. A third possibility is that one of them may be a satellite DNA which encodes an RAP specific for its own replication and relies on the host virus for other functions.

From the results of the PCRs used to detect components 2 and 6 in different SCSV isolates, components 2 and 6 may be functionally interchangeable. Isolate C, for example, had only component 6 and isolate E had only component 2 (Figure 3.6). Glasshouse isolate B1 may have a different RAP-encoding DNA as it contains neither component 2 nor 6. The inability to detect these components by PCR does not rule out the possibility that closely related RAP type components do exist in this isolate. Sequence mismatches in the binding sites of PCR primers, particularly those at the 3' end of the primers, may disrupt the binding sufficiently that no product can be amplified. Other experiments indicate that some mismatches within primer binding sites will be tolerated but mismatches at the 3' end of a primer were

generally disruptive (see Chapter 5). Attempts to amplify a related sequence from this isolate using degenerate primers based on the amino acids and corresponding nucleotides which were highly conserved between SCSV components 2 and 6 and the FBNYV RAP-type DNA were unsuccessful (data not shown). This may have been due to inappropriate PCR conditions, the quality of the preparations, the sequence of the RAP component being too different, or to the absence of an SCSV RAP-type component. The last possibility seems unlikely given that the related geminivirus RAPs are essential for replication of those viruses (Chapter 1). If isolate B1 does contain a third type of RAP component, then other isolates may have this type in addition to component 2 and/or 6.

The glasshouse isolates have been passaged for up to 9 years and some may have lost either components 2 or 6 during passaging. Since some of the field isolates contained only component 6, the detection of just one of the two components in some isolates was not necessarily an artefact of glasshouse conditions. Although isolates containing only component 2 were not among these field samples, the samples were taken from one area. The glasshouse maintained isolates have come from different states of Australia and may be more representative of the range of possible RAP component mixtures.

Components 2 and 6 appear to be functionally equivalent. They are similar in size, structurally related in their non-coding regions, and encode similar size proteins which have extensive amino acid identity throughout the polypeptide. Although they encode proteins that are presumably essential for viral DNA replication, they also appear to be redundant for SCSV infection as either may be absent in different isolates. This apparent coexistence of

functionally equivalent and redundant genes in a viral genome contradicts current notions that successful virus evolution is based on minimalist genomic design.

The presence of several DNAs encoding related putative RAPs in one SCSV isolate may be a feature of SCSV-like viruses. At least three different potentially RAP-encoding DNAs have been found in BBTV isolates (Harding *et al.*, 1993; Yeh *et al.*, 1994; Wu *et al.*, 1994a; M. Karan, pers. comm.). Some BBTV isolates contain two or more of these DNAs (Wu *et al.*, 1994a; M. Karan, pers. comm.).

CHAPTER 4: COMPONENTS 1, 3, 4, 5, AND 7

4.1 Introduction

As well as the RAP-encoding DNAs described in the previous chapter, five other components were identified in the F isolate (Chu *et al.*, 1993). This chapter presents the sequences of these additional components, numbered 1, 3, 4, 5, and 7, which are associated with SCSV F isolate. A number of functions required for the virus life cycle may be encoded by these components, including: the viral coat protein, aphid transmission factor, movement protein(s), and factor(s) associated with pathogenicity. Insect transmission may be dependant on the coat protein, as it appears to be in the geminiviruses (Mullineaux *et al.*, 1984; Briddon *et al.*, 1990; Azzam *et al.*, 1994), or on a distinct protein as in CaMV (Lung and Pirone, 1974), or on both the coat protein and another protein, as in the potyviruses (Pirone, 1991). Similarly, viral movement within the infected plant may be one of the functions of the coat protein, as it is in the monopartite geminiviruses (Mullineaux *et al.*, 1988; Briddon *et al.*, 1989; Lazarowitz *et al.*, 1989; Woolston *et al.*, 1989; Stanley *et al.*, 1992; Wartig *et al.*, 1993). By analogy with the geminiviruses, one or more specific movement proteins are likely to be present, whether or not the coat protein is involved in movement (Chapter 1). Pathogenicity is not strictly a viral function but a consequence of the virus disrupting the host systems. The BL1 movement protein of squash leaf curl geminivirus (SqLCV) was found to cause the symptoms of viral infection when expressed in transgenic tobacco (Pascal *et al.*, 1993). This was suggested to be due to disruption of the regulation of phloem transport.

Fundamental questions which arise from the presence of at least seven viral components in SCSV include: what makes up the complete genome of SCSV F isolate and which components are essential for the viral life cycle?

4.2 Results

4.2.1 Cloning and sequencing

The sequences of components 1, 3, 4, 5, and 7 of the F isolate are presented in Figure 4.1. The cloning and sequencing were done as described in Chapter 3. To complete the sequencing, full-length clones of components 4 and 7 were created by PCR on purified RF DNA (Chu *et al.*, 1993) with the 4Mlu, 4-850 and 7Hind primers listed in Table 4.1

A few nucleotide variations were found between different clones of the same DNA component in the F isolate. In component 4 position 132, in the non-coding region, there was a G rather than an A in one clone, and position 588, in the coding region, there was a T instead of C in another clone, which changes the encoded amino acid from Pro to Arg. In component 7 position 775, in the coding region, there was an A rather than G in one clone, which changes the encoded amino acid from Gly to Asp. In the variable positions the sequences presented in Figure 4.1 show the nucleotide most commonly found.

4.2.2 Identification of potential ORFs

Each DNA contains one large ORF, translated in Figure 4.1, and other smaller ORFs detected by FRAMES (Figure 4.2). The main ORF of each DNA is flanked by consensus transcription signals 5'TATAAA3' (TATA box) and 5'AATAAA3' (polyadenylation signal) (Messing *et al.*, 1983; Joshi *et al.*, 1987a and b), except component 1, which has the potential polyadenylation signal,

Figure 4.1

Sequence of SCSV components 1, 3, 4, 5, and 7.

The sequence of each component is listed separately, with the deduced amino acid sequence of the major ORF shown above. The stem-loop sequence is marked by the arrows, the nonanucleotide in the loop is underlined, the potential TATA boxes and polyadenylation signals are in **bold**, and * denotes the translation stop codon.

SCSV Component 1

```
                 <----------
    TAGTATTACCCCCGTGCCGGGATCAGAGACATTTGACCAATAGTTGACTAGTATAATAGC
  1 ---------+---------+---------+---------+---------+---------+ 60

    CCTTGGATTAAATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT
 61 ---------+---------+---------+---------+---------+---------+ 120

    GAACGAATCTGACGGAAGAGCGTTCACACTTAGATCTAGTTAGCGTACTTAGTACGCGTT
121 ---------+---------+---------+---------+---------+---------+ 180

                                                          M
    GTCTTGGGTCTATAAATAGAGTGCTTCTGAACAGATTGTTCAGAATTTCATAGCGAGATG
181 ---------+---------+---------+---------+---------+---------+ 240

    D  S  G  D  G  Y  N  T  Y  S  Y  E  E  G  A  G  D  A  K  K
    GATTCTGGTGATGGTTACAATACATACTCATATGAAGAAGGTGCTGGAGATGCGAAGAAG
241 ---------+---------+---------+---------+---------+---------+ 300

    E  V  L  Y  K  I  G  I  I  M  L  C  I  V  G  I  V  V  L  W
    GAAGTTTTATATAAAATAGGTATTATTATGTTATGTATTGTAGGGATTGTAGTTTTATGG
301 ---------+---------+---------+---------+---------+---------+ 360

    V  L  I  I  L  C  C  A  V  P  R  Y  A  K  S  T  M  D  A  W
    GTTTTAATTATATTATGTTGTGCTGTTCCTCGCTATGCTAAATCAACGATGGACGCTTGG
361 ---------+---------+---------+---------+---------+---------+ 420

    L  S  S  S  S  I  M  K  R  K  M  A  S  R  I  T  G  T  P  F
    TTATCTTCGTCTTCTATTATGAAGAGGAAGATGGCTTCAAGGATTACTGGTACTCCGTTT
421 ---------+---------+---------+---------+---------+---------+ 480

    E  E  T  G  P  H  R  E  R  R  W  A  E  R  R  T  E  A  T  N
    GAAGAAACTGGTCCTCATCGTGAAAGAAGATGGGCTGAAAGAAGAACTGAAGCGACGAAC
481 ---------+---------+---------+---------+---------+---------+ 540

    Q  N  N  N  D  N  V  N  R  F  S  *
    CAGAATAATAATGATAATGTAAATAGATTTAGTTGATATGTTGTAATTTTATATGGATTA
541 ---------+---------+---------+---------+---------+---------+ 600

    ATGAGAATTATTATTATTCTGTTCTTCGTCTGTGTTTTTTAAGCTTTTTCTGTGTTTTAA
601 ---------+---------+---------+---------+---------+---------+ 660

    TGGCGTCTGGAGAGAGAAAGGAATAATTGTAAGGTAGACGACGATGTAGTGGATTACAGT
661 ---------+---------+---------+---------+---------+---------+ 720

    TGTCTTTACTTCGCCTCGAAGAAAGACACATTTCAAGTTGTGAGTGTTATTGCTTTTGAG
721 ---------+---------+---------+---------+---------+---------+ 780

    GAAGCTTCCTCGAAGCAGCGTATAACTTTAATTTGAATTTGGTTTTGGCGCGTTAGTGAA
781 ---------+---------+---------+---------+---------+---------+ 840

    ATTGCGGCTGTAAACGTGTCAAGTTGTGAGTGGCTGAAATAAGATAATAGATATATTATT
841 ---------+---------+---------+---------+---------+---------+ 900

    ATTGTTTTAATTTAATTCCGCGAAGCGATATGTTAAGTGATAAATGAAACGAAGCGTTTT
901 ---------+---------+---------+---------+---------+---------+ 960

                                          ---------->
    GATGACGTCATATGTCTCCGTGCCTACGTCAGCACGGGGCT
961 ---------+---------+---------+---------+- 1001
```

SCSV Component 3

```
                                  <---
    TAGTATTACCCCCGTGCCGGGATCAGAGACATTTGACCAATAGTTGACTATGAATAATAG
  1 ---------+---------+---------+---------+---------+---------+ 60

    CCCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGC
 61 ---------+---------+---------+---------+---------+---------+ 120

    TGAACGAATCTGACGGAAGAGCGGACATACGCACATGGATTATGGCCCACATGTCTAAAG
121 ---------+---------+---------+---------+---------+---------+ 180

    TGTATCTCTTTACAGCTATATTGATGTGACGTAAGATGCTTTACTTCGCCTCGAAGTAAA
181 ---------+---------+---------+---------+---------+---------+ 240

    GTAGGAAATTGCTCGCTAAGTTATTCTTTTCTGAAAGAAATTAATTTAATTCTAAATTAA
241 ---------+---------+---------+---------+---------+---------+ 300

    ATTAAATGAGTGGCTATAAATAGATGTTTCGTCTTCGTTGTTTTACAACGAAGCTTAGAA
301 ---------+---------+---------+---------+---------+---------+ 360

              M  A  L  R  Y  F  S  H  L  P  E  E  L  K  E  K  I
    TCTTGTGTTAATGGCGTTAAGGTATTTCTCTCATCTTCCTGAAGAACTGAAGGAGAAGAT
361 ---------+---------+---------+---------+---------+---------+ 420

     M  N  E  H  L  K  E  I  K  K  K  E  F  L  E  N  V  I  K  A
    TATGAACGAGCACTTGAAGGAAATTAAGAAGAAGGAATTTCTAGAGAATGTAATTAAAGC
421 ---------+---------+---------+---------+---------+---------+ 480

     A  C  A  V  F  E  G  L  T  K  K  E  S  V  E  E  D  D  I  L
    TGCGTGTGCTGTGTTCGAAGGTTTAACAAAGAAGGAGTCTGTTGAAGAAGACGACATACT
481 ---------+---------+---------+---------+---------+---------+ 540

     R  F  S  G  F  L  E  G  L  S  A  Y  Y  A  E  A  T  K  K  K
    ACGCTTCTCTGGGTTTCTGGAAGGTCTGTCTGCATATTATGCAGAGGCGACGAAGAAGAA
541 ---------+---------+---------+---------+---------+---------+ 600

     C  L  V  R  W  K  K  S  V  A  I  N  L  K  W  R  V  M  E  E
    GTGTTTAGTTAGATGGAAGAAGAGCGTTGCAATAAATCTGAAATGGAGAGTTATGGAGGA
601 ---------+---------+---------+---------+---------+---------+ 660

     M  H  Y  K  L  Y  G  F  A  D  M  E  D  L  Y  C  S  E  L  G
    GATGCATTACAAGCTTTATGGATTTGCAGACATGGAAGATTTATATTGTTCAGAGTTAGG
661 ---------+---------+---------+---------+---------+---------+ 720

     F  P  N  Y  G  E  D  D  V  A  Y  H  D  G  A  I  V  N  C  K
    GTTTCCTAATTACGGTGAAGACGATGTAGCTTATCACGATGGTGCAATTGTAAATTGTAA
721 ---------+---------+---------+---------+---------+---------+ 780

     Q  L  E  V  V  F  D  D  L  G  I  E  F  M  S  I  V  I  D  R
    GCAATTAGAAGTTGTATTTGATGATTTAGGTATTGAGTTTATGTCTATTGTAATTGATAG
781 ---------+---------+---------+---------+---------+---------+ 840

     G  S  I  K  I  E  L  *
    AGGTTCTATTAAGATAGAATTATGAGATGTAATTGTGATTAATGAATAAAGAGTTGTTAT
841 ---------+---------+---------+---------+---------+---------+ 900

    TATTCTTTGAATTACTCCGCGAAGCGGTGTGTTATGTTTTTGTTGGAGACATATGACGTC
901 ---------+---------+---------+---------+---------+---------+ 960

                        --------->
    ATATGTCTCGCCGACAGGCTGGCACGGGGCT
961 ---------+---------+---------+- 991
```

SCSV Component 4

```
                 <----------
    TAGTATTACCCCGTGCCGGGATCAGAGACATTTGACTAAATGTTGACTTGGAATAATAGC
  1 ---------+---------+---------+---------+---------+---------+ 60

    CCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT
 61 ---------+---------+---------+---------+---------+---------+ 120

    GAACGAATCTGACGGAAGAGCGGACAAACGCACATGGACTATGGCCCACTGCTTTATTAA
121 ---------+---------+---------+---------+---------+---------+ 180

    AGAAGTGAATGACAGCTGTCTTTGCTTCAAGACGAAGTAAAGAATAGTGGAAAACGCGTA
181 ---------+---------+---------+---------+---------+---------+ 240

    AAGAATAAGCGTACTCAGTACGCTTCGTGGCTTTATAAATAGTGCTTCGTCTTATTCTTC
241 ---------+---------+---------+---------+---------+---------+ 300

                                                  M  A  D  W  F  H
    GTTGTATCATCAACGAAGAAGTTAAGCTTTGTTCTGCGTTTTAATGGCGGACTGGTTTCA
301 ---------+---------+---------+---------+---------+---------+ 360

     S  A  L  K  T  C  T  H  V  C  D  F  S  D  I  K  A  S  S  Q
    CAGTGCGCTTAAGACATGTACTCATGTCTGTGATTTTTCAGATATTAAGGCGTCTTCACA
361 ---------+---------+---------+---------+---------+---------+ 420

     Q  D  F  F  C  C  D  S  M  R  G  K  L  S  E  P  R  K  V  L
    ACAGGATTTCTTCTGTTGTGATAGTATGCGAGGTAAATTATCTGAACCTAGGAAGGTGTT
421 ---------+---------+---------+---------+---------+---------+ 480

     L  V  S  C  F  V  S  F  T  G  S  F  Y  G  S  N  R  N  V  R
    GTTAGTTAGTTGTTTTGTAAGTTTTACTGGTAGTTTTTATGGAAGTAATAGGAATGTTAG
481 ---------+---------+---------+---------+---------+---------+ 540

     G  Q  V  Q  L  G  M  Q  Q  D  D  G  V  V  R  P  I  G  Y  I
    AGGTCAAGTTCAGTTGGGTATGCAGCAAGATGATGGCGTTGTTCGTCCAATAGGATATAT
541 ---------+---------+---------+---------+---------+---------+ 600

     P  I  G  G  Y  L  Y  H  D  D  Y  G  Y  Y  Q  G  E  K  T  F
    TCCTATTGGGGGTTATTTGTATCATGATGATTATGGATATTATCAAGGAGAGAAGACGTT
601 ---------+---------+---------+---------+---------+---------+ 660

     N  L  D  I  E  S  D  Y  L  K  P  D  E  D  F  W  K  R  F  T
    CAATCTGGACATCGAGTCAGATTATCTGAAGCCTGATGAAGATTTTTGGAAGAGATTTAC
661 ---------+---------+---------+---------+---------+---------+ 720

     I  N  I  V  N  D  K  G  L  D  D  R  C  D  V  K  C  Y  V  V
    AATTAATATTGTAAATGATAAAGGATTAGATGATAGGTGTGATGTAAAATGTTATGTAGT
721 ---------+---------+---------+---------+---------+---------+ 780

     H  T  M  R  I  K  V  *
    TCATACGATGCGTATTAAGGTGTAATTGTTATTATCAATAAAAGAATTTTTATTGTTATT
781 ---------+---------+---------+---------+---------+---------+ 840

    GTGTTATTTGGTAATTTATGCTTATAAGTAATTCTATGATTAATTGTGAATTAATAAGAC
841 ---------+---------+---------+---------+---------+---------+ 900

    TAATGAGGATAATAATTGAATTTGATTAAATTAACTCTGCGAAGCTATATGTCTTTCACG
901 ---------+---------+---------+---------+---------+---------+ 960

                                          ---------->
    TGAGAGTCACGTGATGTCTCCGCGACAGGCTGGCACGGGGCT
961 ---------+---------+---------+---------+-- 1002
```

SCSV Component 5

```
                 <----------
    TAGTATTACCCCGTGCCGGGGTCAGAGACATTTGACTAAATATTGACTTGGAATAATAGC
  1 ---------+---------+---------+---------+---------+---------+ 60

    CCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT
 61 ---------+---------+---------+---------+---------+---------+ 120

    GAACGAATCTGACGGAAGAGCGTCATGGTCCACATGTCTAAAGAATAATGCTTTACAGCT
121 ---------+---------+---------+---------+---------+---------+ 180

    GTATTGATTTGACTTTACGCGCTTTACTTTAATTGCTTTAAGTAAAGTAAGATGCTTTAC
181 ---------+---------+---------+---------+---------+---------+ 240

    TTTGCTCGCGACGAAGCAAAGTGATTGTAGCTGCAGAAATTGATGCTTTAATTACCGGGT
241 ---------+---------+---------+---------+---------+---------+ 300

    AACACGGTTTGATTGTGGGTATAAATATGTTCTGTTCGTTTTCTTCGTTGTCATTTTACA
301 ---------+---------+---------+---------+---------+---------+ 360

          M  V  A  V  R  W  G  R  K  G  L  R  S  Q  R  R  K  Y
    ACGAAGATGGTTGCTGTTCGATGGGGAAGAAAGGGTCTGAGGTCTCAAAGGAGAAAATAT
361 ---------+---------+---------+---------+---------+---------+ 420

    S  R  I  A  Y  K  P  P  S  S  K  V  V  S  H  V  E  S  V  L
    TCGCGAATTGCTTACAAACCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAGTCTGTTCTG
421 ---------+---------+---------+---------+---------+---------+ 480

    N  K  R  D  V  T  G  A  E  V  K  P  F  A  D  G  S  R  Y  S
    AATAAGAGAGATGTTACTGGAGCGGAGGTTAAGCCATTCGCTGATGGTTCAAGGTATAGT
481 ---------+---------+---------+---------+---------+---------+ 540

    M  K  K  V  M  L  I  A  T  L  T  M  A  P  G  E  L  V  N  Y
    ATGAAGAAGGTAATGTTGATTGCAACATTAACTATGGCTCCTGGAGAATTAGTTAATTAT
541 ---------+---------+---------+---------+---------+---------+ 600

    L  I  V  K  S  N  S  P  I  A  N  W  S  S  S  F  S  N  P  S
    CTTATTGTGAAGAGTAATTCGCCTATTGCGAATTGGAGTTCGTCTTTCAGTAATCCTTCG
601 ---------+---------+---------+---------+---------+---------+ 660

    L  M  V  K  E  S  V  Q  D  T  V  T  I  V  G  G  G  K  L  E
    TTGATGGTGAAAGAGTCTGTTCAAGATACAGTTACGATTGTTGGAGGAGGAAAGCTTGAG
661 ---------+---------+---------+---------+---------+---------+ 720

    S  S  G  T  A  G  K  D  V  T  K  S  F  R  K  F  V  K  L  G
    TCTTCTGGTACTGCTGGTAAAGATGTAACTAAGTCTTTTAGGAAGTTTGTTAAGCTGGGT
721 ---------+---------+---------+---------+---------+---------+ 780

    S  G  I  S  Q  T  Q  H  L  Y  L  I  I  Y  S  S  D  A  M  K
    TCAGGTATTAGTCAGACCCAGCATTTGTATTTAATTATTTATTCCAGTGATGCGATGAAG
781 ---------+---------+---------+---------+---------+---------+ 840

    I  T  L  E  T  R  M  Y  I  D  V  *
    ATCACACTGGAGACGAGAATGTATATTGATGTATAATTGTGATGATTAATGAATAAAGAG
841 ---------+---------+---------+---------+---------+---------+ 900

    TTGTTTTTATTCTTTGAATTACTCCGCGAAGCGGTGTGTTATGTTTTTGTTGGAGACATA
901 ---------+---------+---------+---------+---------+---------+ 960

                                  ---------->
    TGACGTCATATGTCTCCGCGACAGGCTGGCACGGGGCT
961 ---------+---------+---------+-------- 998
```

SCSV Component 7

```
     <--------
     TAGTATTACCCCGTGCCGGGATCAGAGACATTTGACTAAATATTGACTTGGAATAATAGC
   1 ---------+---------+---------+---------+---------+---------+ 60

     CCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT
  61 ---------+---------+---------+---------+---------+---------+ 120

     GAACGAATCTGACGGAAGAGCGGACATACGCACATGGATTATGGCCCACATGTCTAAAGT
 121 ---------+---------+---------+---------+---------+---------+ 180

     GTATCTCTTTACAGCTATATTGATGTGACGTAAGATGCTTTACTTCGCTTCGAAGTAAAG
 181 ---------+---------+---------+---------+---------+---------+ 240

     TAGGAAATTGCTCGCTAAGTTATTCTTTTCTGAAAGAAATTAATTTAATTCTAATTAAAT
 241 ---------+---------+---------+---------+---------+---------+ 300

     TAAATGAGTGGCTATAAATAGTGTCGATGCTGCCTCACATCGTATTCTTCTTCGCATCGT
 301 ---------+---------+---------+---------+---------+---------+ 360

                       M  V  S  F  S  F  P  E  I  Y  D  V  S  D
     CTGTTCTGGTTTTAAGCGATGGTCAGTTTTAGTTTTCCTGAGATATACGATGTGAGCGAC
 361 ---------+---------+---------+---------+---------+---------+ 420

     D  V  L  V  S  D  S  R  R  S  V  A  V  E  V  E  E  K  V  Q
     GATGTTCTTGTAAGCGATAGCAGAAGAAGTGTAGCTGTTGAGGTCGAAGAGAAGGTTCAA
 421 ---------+---------+---------+---------+---------+---------+ 480

     V  I  N  V  K  V  L  R  L  I  E  A  V  D  E  D  R  V  G  V
     GTGATTAACGTGAAGGTACTGAGGTTGATTGAAGCTGTTGATGAAGATAGAGTTGGAGTG
 481 ---------+---------+---------+---------+---------+---------+ 540

     K  V  M  F  R  L  C  Y  R  Y  R  R  E  L  K  I  T  L  L  G
     AAGGTTATGTTTCGTCTGTGTTACAGATACAGACGAGAACTGAAGATTACGTTGTTGGGT
 541 ---------+---------+---------+---------+---------+---------+ 600

     C  K  M  E  L  W  T  S  L  K  S  S  G  K  Y  S  V  Q  S  L
     TGTAAGATGGAGCTATGGACTTCGTTGAAGTCTTCAGGCAAGTATTCAGTTCAATCTTTG
 601 ---------+---------+---------+---------+---------+---------+ 660

     L  Q  R  K  L  N  G  I  C  V  S  N  Y  C  I  G  I  D  M  F
     TTGCAGAGGAAGCTTAATGGTATATGTGTTAGTAATTACTGTATAGGTATTGATATGTTT
 661 ---------+---------+---------+---------+---------+---------+ 720

     V  S  N  V  K  E  L  I  N  R  C  K  W  I  T  S  V  Q  G  V
     GTAAGTAATGTTAAAGAGTTGATTAATAGATGTAAATGGATTACATCTGTTCAAGGTGTT
 721 ---------+---------+---------+---------+---------+---------+ 780

     N  P  I  C  C  L  Y  H  M  D  E  E  *
     AATCCTATATGTTGTTTGTATCATATGGACGAAGAGTAATTAATAGTAATTATGATTAAT
 781 ---------+---------+---------+---------+---------+---------+ 840

     TATGAGATAAGAGTTGTTATTAATGCTTATGAGGAATAAAGAATGATTAATATTGTTTAA
 841 ---------+---------+---------+---------+---------+---------+ 900

     TTTTATTCGCGAAGCGGTGTGTTATGTTTTTGTTGGAGACATCACGTGACTCTCACGTGA
 901 ---------+---------+---------+---------+---------+---------+ 960

                          ---------->
     TGTCTCCGCGACAGGCTGGCACGGGGCT
 961 ---------+---------+-------- 988
```

Table 4.1

Sequences and positions of and fragment sizes produced by PCR primers to components 1, 3, 4, 5 and 7.

Primer 1 name	Position of 5' end[b]	Primer 1 sequence (5'-3')	Primer 2 name	Position of 5' end[b]	Primer 2 sequence (5'-3')	Fragment size (bp)
4MluS[a]	233	CCACCAACGCGTAAAGAATA AGCG	4MluC[a]	239	CCACCAACGCGTTTTCCAC TATTC	1020
4-850S	692	CCTGATGAAGATTTTTGGAAG	4-850C	691	CTTCAGATAATCTGACTC	1002
5HinPI	181	GTATTGATTTGACTTTACGCG CTTT	5Hind	737	CCAGCAGTACCAGAAGAC	556
7HindS[a]	670	CCACCAAAGCTTAATGGTATA TGTG	7HindC[a]	675	CCACCAAAGCTTCCTCTGC AACAAAG	1008

[a] non-homologous extra bases added at 5' end

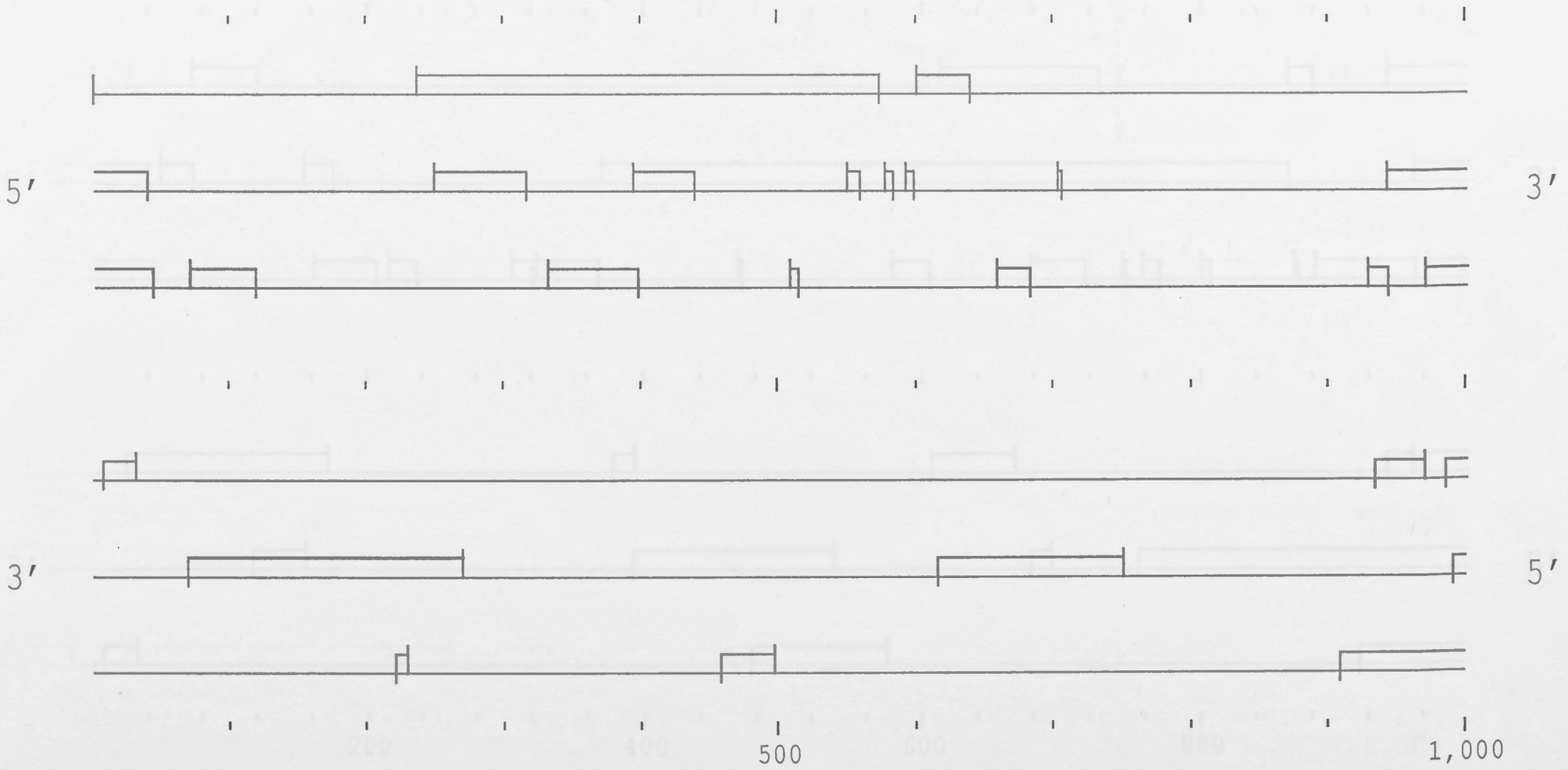
[b] position in F isolate sequence

Figure 4.2

Open reading frames of SCSV components 1, 3, 4, 5, and 7.

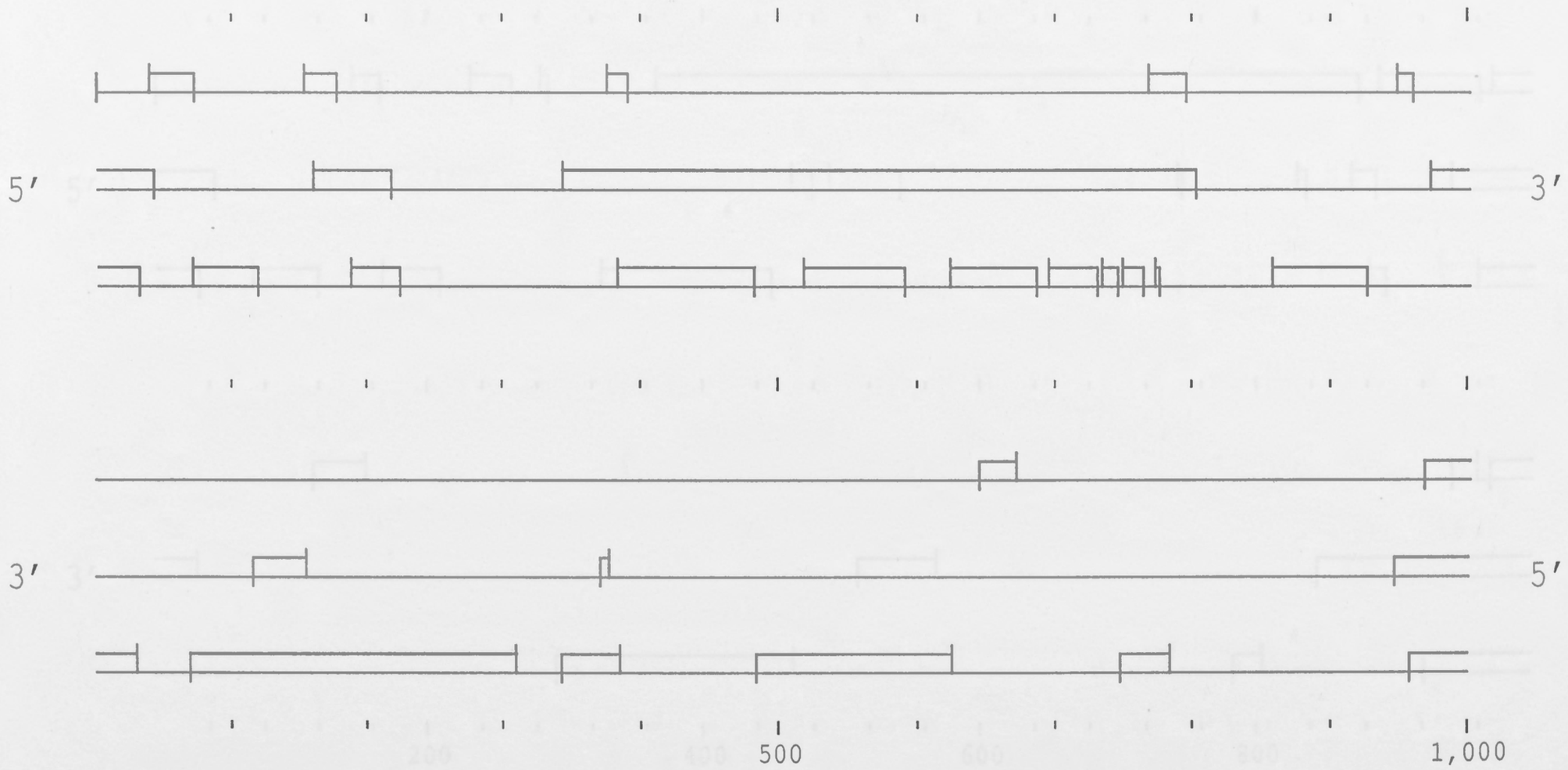
Open reading frames, represented by rectangles, in all frames of the virion (top three lines) and complementary (bottom three lines) sense strands of components 1, 3, 4, 5, and 7 were detected using FRAMES. Marks extending above the ends of the rectangles indicate the positions of the initiation codons, and those below the line indicate stop codons. Redundant start and stop codons are not shown.

FRAMES of SCSV component 1

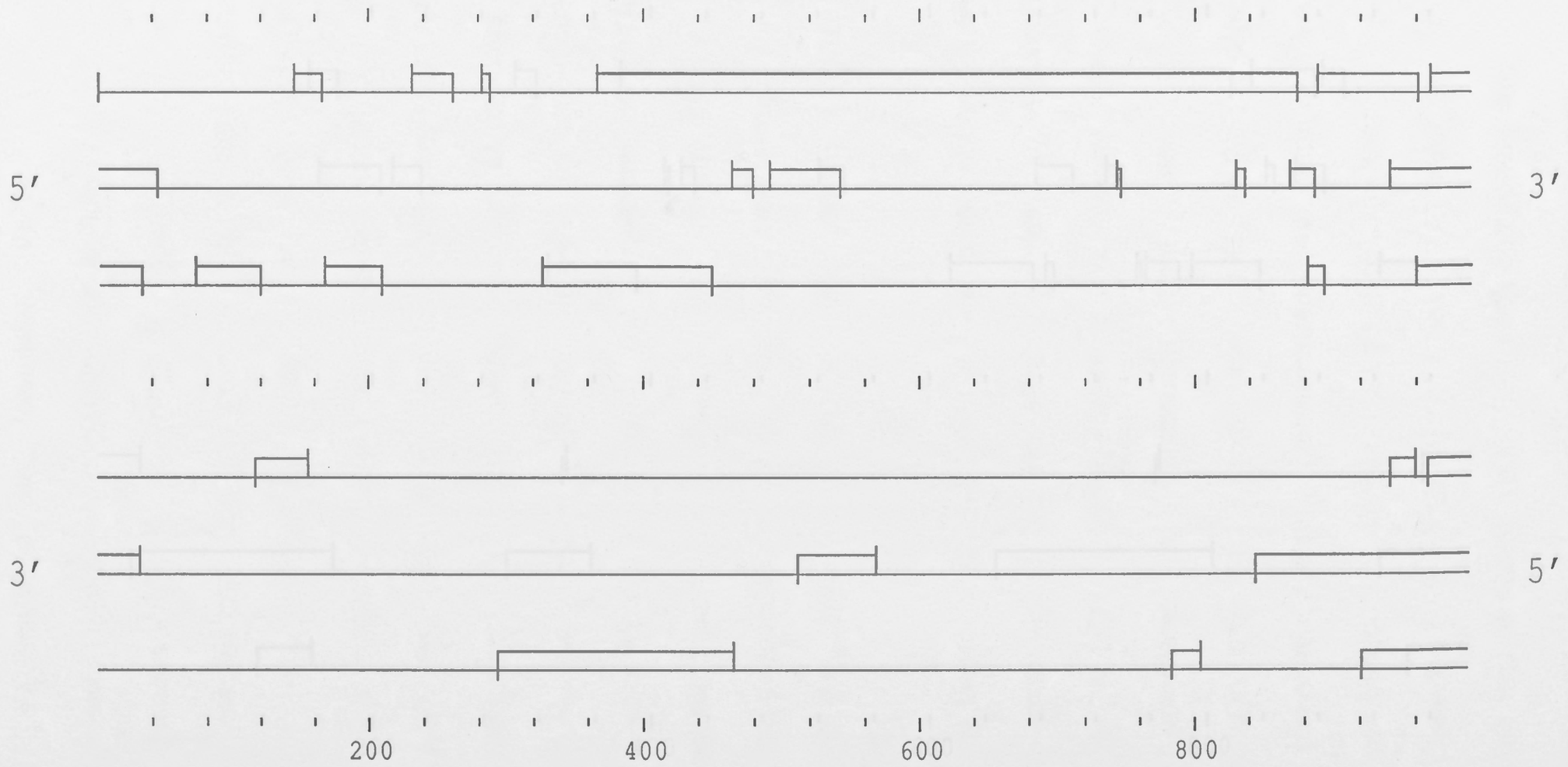


FRAMES of SCSV component 3

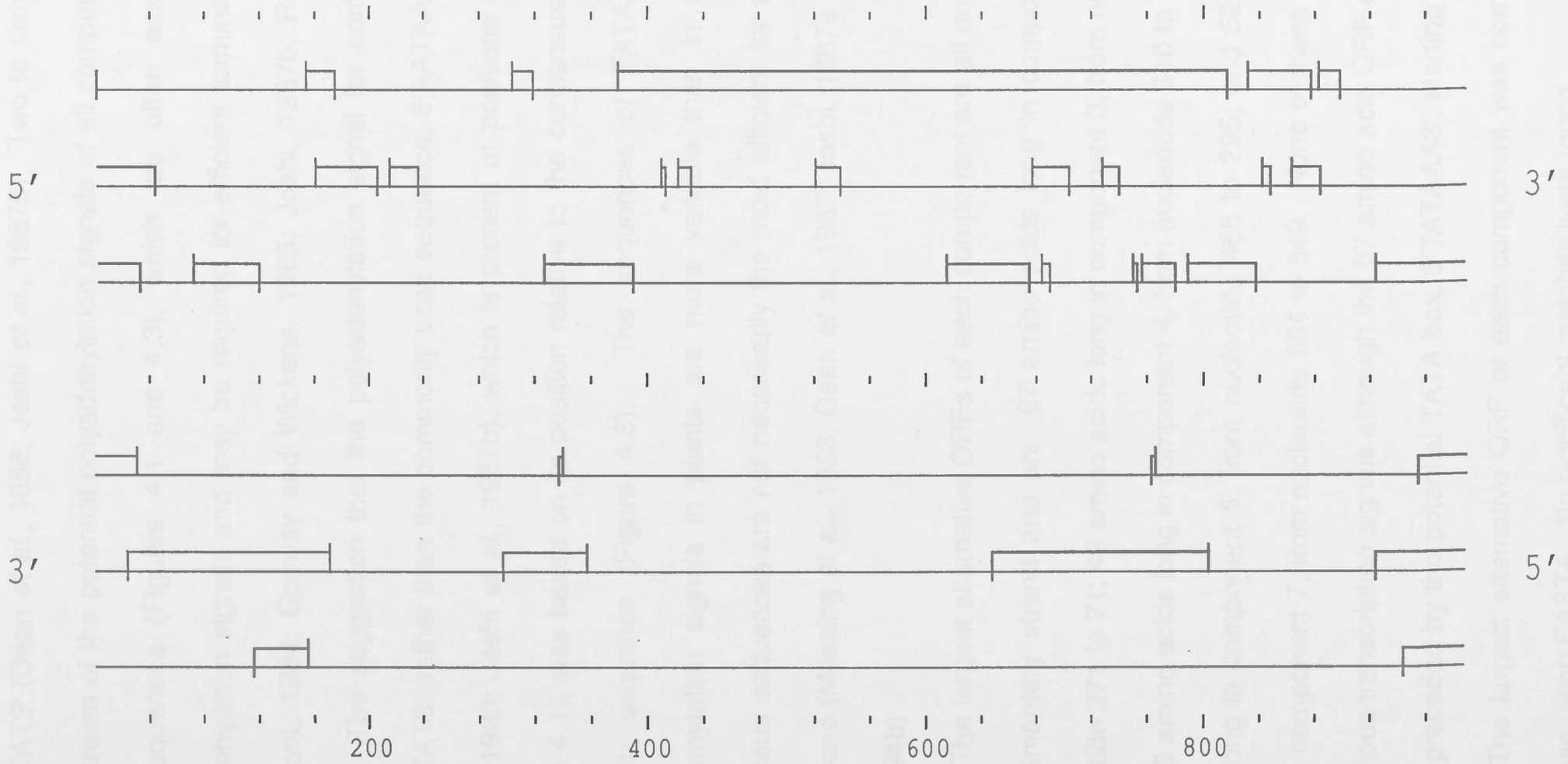
FRAMES of SCSV component 4



FRAMES of SCSV component 5



FRAMES of SCSV component 7



5'AATTAT3' (Dean *et al.*, 1986; Joshi *et al.*, 1987b). Two to sixty nucleotides downstream of the potential polyadenylation signals of all components are TG rich sequences (Figures 4.1 and 4.3), which are often associated with polyadenylation signals and may be required for efficient termination (Gil and Proudfoot, 1984; Conway and Wickens, 1985; Joshi, 1987b; Rothnie *et al.*, 1994). The suggestion that the polyadenylation signal for component 1 is, 5'AATTAT3', rather than the commonly used sequence, 5'AATAAT3' (Messing *et al.*, 1983; Joshi *et al.*, 1987b), which is present at positions 544 and 682 (Figure 4.1), was based on its position relative to the consensus downstream TG rich sequence (Figure 4.3). The sequences of TATA boxes and polyadenylation signals in plants are more variable than in animals, and consensus sequences are not necessarily the most efficient for any particular plant gene (Messing *et al.*, 1983; Dean *et al.*, 1986; Joshi, 1987a and b; Wu *et al.*, 1994b)

The largest alternative ORFs of each component are all encoded on the complementary strand and are: 67 amino acids long in component 1, from nucleotide 271 to 71; 49 amino acids long in component 3, from nucleotide 172 to 6; 80 amino acids long in component 4, from nucleotide 310 to 71; 57 amino acids long in component 5, from nucleotide 465 to 295; and 52 amino acids long in component 7, from nucleotide 804 to 649. None of these were flanked by obvious transcription signals although the 67 amino acid ORF of component 1 was preceded by the potential TATA box, 5'TATAAA3', starting at nucleotide 311. The largest alternative ORF of each component was compared to the database using FASTA. No significant matches were found.

Figure 4.3

Alignment of termination signals of SCSV components 1, 3, 4, 5, and 7

Alignment, using PILEUP, of the putative polyadenylation signals (in **bold**) and downstream TG rich sequences (underlined) of the SCSV components 1, 3, 4, 5, and 7. The component 1 sequence from the region equivalent to the termination signals of components 3, 5, and 7, which shows some homology to those components (884-950), is also included. The positions of the sequence fragments are indicated in brackets.

```
Component
   3   (885) AATAAAGAGT TGTTATTATT CTTTGAATTA CTCCGCGAAG CGGTGTGTTA TGTTTTTGTT GGAGACA (951)
   5   (892) AATAAAGAGT TGTTTTTATT CTTTGAATTA CTCCGCGAAG CGGTGTGTTA TGTTTTTGTT GGAGACA (958)
   7   (875) AATAAAGAAT GATTAATATT GTTTAATTTT ATTCGCGAAG CGGTGTGTTA TGTTTTTGTT GGAGACA (942)
   1   (884) ATAATAGATA TATTATTATT GTTTTAATTT AATTCCGCGA AGCGATATGT TAAGTGATAA ATGAAAC (950)
   1   (606) AATTATTATT ATTCTGTTCT TCGTCTGTGT TTTTTAAGCT TTTTCTGTGT TTTAATGGCG TCTGGAG (672)
   4   (817) AATAAAAGAA TTTTTATTGT TATTGTGTTA TTTGGTAATT TATGCTTATA AGTAATTCTA TGATTAA (883)
```

4.2.3 Properties of the putative gene products

Component 5 encodes the coat protein (Chu *et al.*, 1993a). Direct sequencing of the viral coat protein gave one unambiguous sequence of 15 amino acids which matched perfectly with the N-terminal sequence of the protein encoded by component 5. The molecular weight of the protein encoded by the main ORF of component 5 matches that of the coat protein. No functions have been assigned to any of the ORFs encoded by the other components.

The polypeptides encoded by the main ORFs were examined for structural features and hydrophobicity using PLOTSTRUCTURE. The only polypeptide with any striking features is that encoded by component 1 (Figure 4.4). This protein has a hydrophobic stretch approximately 25 amino acids long which has a potential β-sheet structure (Figure 4.4).

No significant similarities were detected between the putative gene products and any sequences in the Swissprot and Genbank databases by FASTA and TFASTA searches. The only exception was component 4, which shares 45% amino acid identity with component 6 of BBTV (Figure 4.5; Burns, 1994).

4.2.4 Features of non-coding regions

As in components 2 and 6, the non-coding region of each of the components presented in this chapter contains a sequence capable of forming a hairpin structure with a GC rich stem and an AT rich loop (the stem-loop). The nonanucleotide in the loop of these five components is 5'TAGTATTAC3'. An alignment of the stem and loop sequences of the SCSV-like viruses and a few representative geminiviruses (Figure 4.6) shows that the stems are

Figure 4.4

Predicted secondary structure of the putative product of SCSV 1

Secondary structural features of the putative protein of SCSV component 1 were predicted by PEPTIDESTRUCTURE and displayed with PLOTSTRUCTURE. PEPTIDESTRUCTURE predicted hydrophilicity by the Kyte-Doolittle (KD) method and secondary structure by the Chou-Fasman (CF) and Garnier-Osguthorpe-Robson (GOR) methods (Genetics Computer Group, 1991).

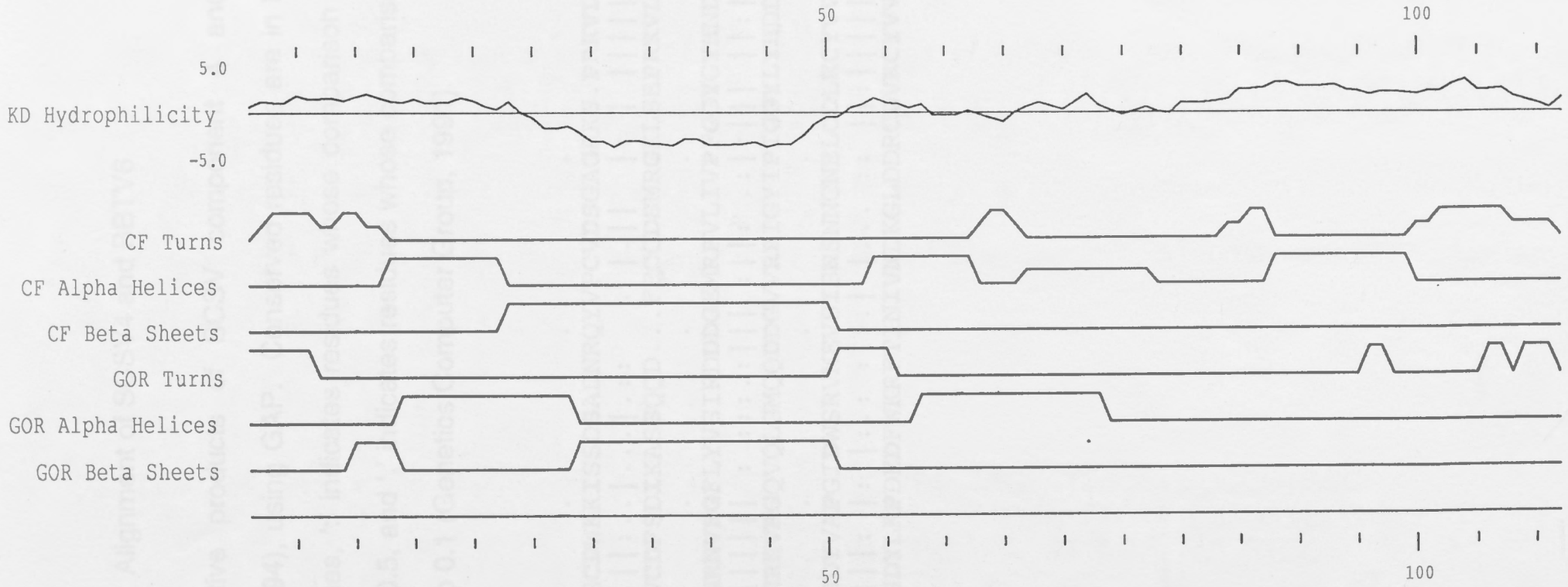

50
100
5.0
KD Hydrophilicity
-5.0
CF Turns
CF Alpha Helices
CF Beta Sheets
GOR Turns
GOR Alpha Helices
GOR Beta Sheets
50
100

Figure 4.5

Alignment of SCSV4 and BBTV6

Alignment of the putative products of SCSV component 4 and BBTV component 6 (Burns, 1994), using GAP. Conserved residues are in **bold**. 'I' indicates identical residues, ':' indicates residues whose comparison value is greater than or equal to 0.5, and '.' indicates residues whose comparison value is greater than or equal to 0.1 (Genetics Computer Group, 1991).

```
                .         .         .         .         .
  1 .MDWAESQFKTCTHGCDWKKISSDSADNRQYVPCVDSGAGRKS.PRKVLL 48
     || .|.:||||||.||:..|...|.::    . |.||  |: | ||||||
  1 MADWFHSALKTCTHVCDFSDIKASSQQD...FFCCDSMRGKLSEPRKVLL 47
                .         .         .         .         .
 49 RSIEAVFNGSFSGNNRNVRGFLYVSIRDDDGEMRPVLIVPFGGYGYHNDF 98
     |. . |.||| |.|||||| : ::::.:||| :||: .:|:||| ||:|:
 48 VSCFVSFTGSFYGSNRNVRGQVQLGMQQDDGVVRPIGYIPIGGYLYHDDY 97
                .         .         .         .         .
 99 YYFEGKGKVECDISSDYVAPGIDWSRVMEVSISNSNNCNELCDLKCYVVC 148
     |::|. ..: ||.|||: |: |:.: :.:.| |... :: ||:|||||
 98 GYYQGEKTFNLDIESDYLKPDEDFWKRFTINIVNDKGLDDRCDVKCYVVH 147

149 SLRIKE 154
    .:|||
148 TMRIKV 153
```

Figure 4.6

Alignment of stem and loop sequences

Alignment of the stem and loop sequences of the SCSV-like viruses, SCSV, CFDV (Rhode *et al.*, 1990), FBNYV (Katul *et al.*, 1995), and BBTV (Harding *et al.*, 1993), and three representative geminiviruses, ACMV, MSV, and BCTV (Stanley and Gay, 1983; Mullineaux *et al.*, 1984; Stanley *et al.*, 1986). The sequences are presented 5' to 3', right to left. The conserved nonanucleotide sequences are boxed and the sequences making up the stems are underlined (in SCSV components 4, 5, and 7, and CFDV, the last nucleotide of the nonanucleotide is within the stem sequence).

SCSV components:	1	AGCACGGGGCT	TAGTATTAC	CCCCGTGCC
	3	GGCACGGGGCT	TAGTATTAC	CCCCGTGCC
	4, 5, 7	GGCACGGGGCT	TAGTATTAC	CCCGTGCC
	2	GCAAGGTCGGCT	TAGTATTAC	CCGACCTTGC
	6	GCGAGGTGCGTAT	CAGTATTAC	CGCACCTCGC
CFDV		AGCCGCGGGGGC	TAGTATTAC	CCCCGCGGCT
FBNYV		CCAAGGCGGGTA	TAGTATTAC	CCCGCCTTGG
BBTV		AGCGCTGGGGCT	TATTATTAC	CCCCAGCGCT
Geminiviruses:	ACMV	GGGGCCAACCGTA	TAATATTAC	CGGTTGGCCCC
	BCTV	GGGCCATCCGTTA	TAATATTAC	CGGATGGCCC
	MSV	GCAGGAAAAGAAGGCGCGCAC	TAATATTAC	CGCGCCTTCTTTTCCTGC

conserved structurally but not in sequence. The nonanucleotides, however, are highly conserved.

The stem-loop in components 1, 3, 4, 5, and 7 is flanked by sequence, we have called the common region, which is greater than 95% conserved between the components. An alignment of the non-coding regions reveals the minimal common region is 153b long and the maximum length of shared sequence, between components 3 and 5, is 258b, (Figure 4.7).

4.2.5 Search for other SCSV genomic components

A unique *Eco*47III site is present in the common region of SCSV components 1, 3, 4, 5, and 7. Other non-RAP-encoding SCSV components were expected to contain this common region. To determine whether the F isolate had any so far undetected components, RF DNA was digested with *Eco*47III, end-labelled with [α-^{32}P]dCTP using the Klenow fragment of DNA polymerase I and electrophoresed on a 4% polyacrylamide gel. The band corresponding to approximately 1kb-sized fragments was extracted and cloned into the *Eco*RV site of pGEM5Z (Promega). Colonies potentially containing clones were probed with a mixture of ^{32}P labelled oligonucleotides complementary to the known SCSV components (Table 2.1; Figure 4.8). Colonies containing clones of the five known common region containing components were detected by the probe, and colonies which were not detected potentially contained clones of hitherto undetected SCSV components. Plasmid DNA was extracted from cultures of bacterial colonies whose DNA did not hybridise to the probe, and analysed by restriction digestion. The plasmids isolated from these bacteria did not contain inserts; therefore, no new components were identified amongst the 134 clones screened.

Figure 4.7

Alignment of non-coding sequences of components 1, 3, 4, 5, and 7

Alignment, using PILEUP, of the non-coding region sequences of the common region-containing components, 1, 3, 4, 5, and 7. The component numbers are indicated on the left in *italics*. The left (5') and right (3') boundaries of the minimal common region are indicated by CR→ and ←CR respectively, the conserved nonanucleotide sequences are in green and the sequences making up the stems are underlined.

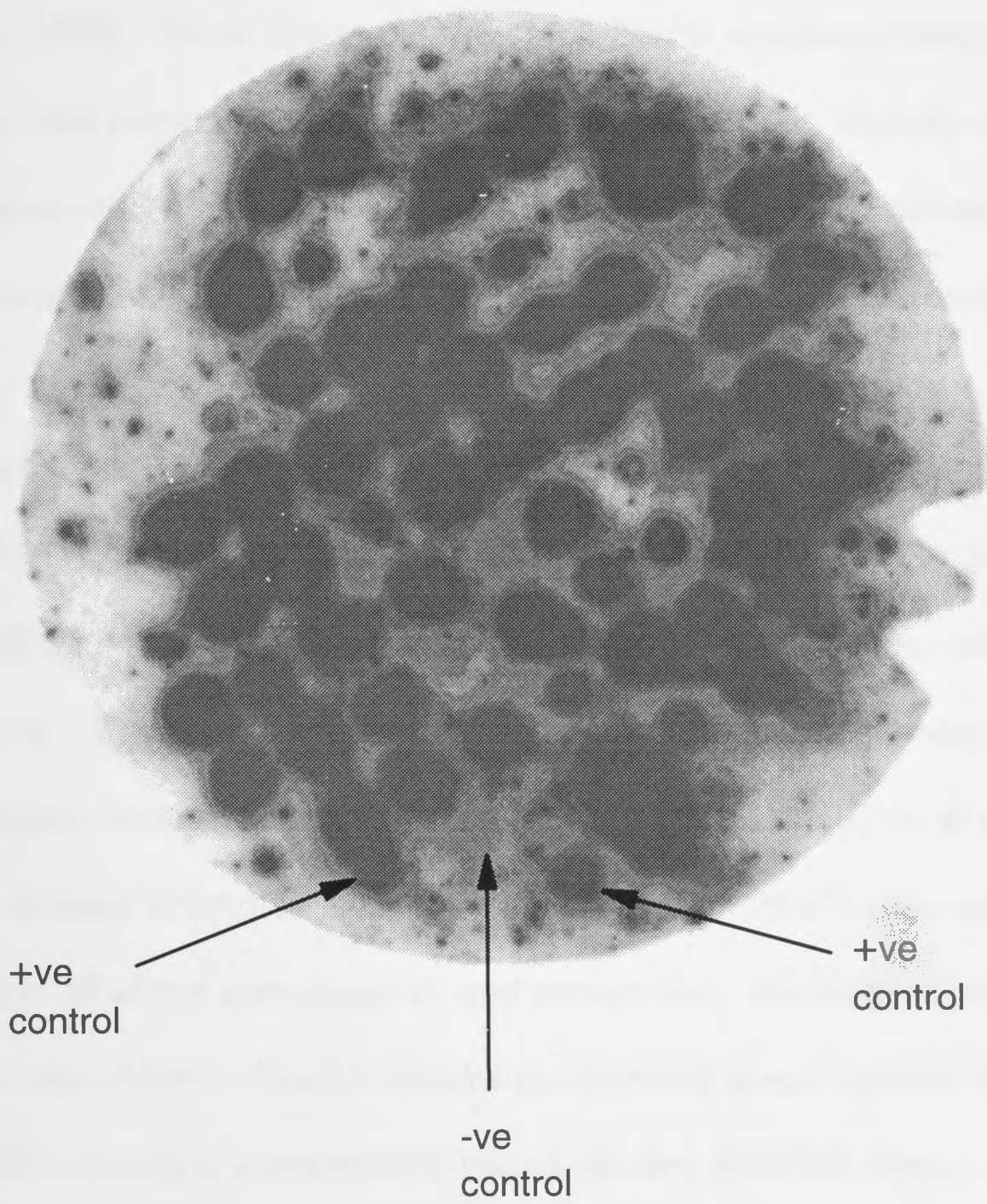
```
    1                                                   50
3   .......... ....GATGTA ATTGTGATTA ATGAATAAAG AGTTGTTATT
7   ATGAGATAAG AGTTGTTATT AATGCTTATG AGGAATAAAG AATGATTAAT
5   .......... ........TT GTGATGATTA ATGAATAAAG AGTTGTTTTT
1   TAAACGTGTC AAGTTGTGAG TGGCTGAAAT AAGATAATAG ATATATTATT
4   TTATGCTTAT AAGTAATTCT ATGATTAATT GTGAATTAAT AAGACTAATG

    51                                                 100
3   ATTCTTTGAA TTAC..TCCG CGAAGCGGTG TGTTATGTTT TTGTTG.GAG
7   ATTGTTTAAT TTTA..TTCG CGAAGCGGTG TGTTATGTTT TTGTTG.GAG
5   ATTCTTTGAA TTAC..TCCG CGAAGCGGTG TGTTATGTTT TTGTTG.GAG
1   ATTGTTTTAA TTTAATTCCG CGAAGCGATA TGTTAAGTGA TAAATGAAAC
4   AGGATAATAA TTGAATTTGA TTAAATTAAC TCTGCGAAGC TATATGTCTT

    101                                      CR→        150
3   ACAT......  .ATGACGTCA TATGTCTCGC CGAC.AGGCT GGCACGGGGC
7   ACATCACGTG ACTCTCACGT GATGTCTCCG CGAC.AGGCT GGCACGGGGC
5   ACAT......  .ATGACGTCA TATGTCTCCG CGAC.AGGCT GGCACGGGGC
1   GAAGCGTTTT GATGACGTCA TATGTCTCCG TGCCTACGTC AGCACGGGGC
4   TCA...CGTG AGAGTCACGT GATGTCTCCG CGAC.AGGCT GGCACGGGGC

    151                                                200
3   TTAGTATTAC CCCCGTGCCG GGATCAGAGA CATTTGACCA ATAGTTGACT
7   TTAGTATTA. CCCCGTGCCG GGATCAGAGA CATTTGACTA AATATTGACT
5   TTAGTATTA. CCCCGTGCCG GGGTCAGAGA CATTTGACTA AATATTGACT
1   TTAGTATTAC CCCCGTGCCG GGATCAGAGA CATTTGACCA ATAGTTGACT
4   TTAGTATTA. CCCCGTGCCG GGATCAGAGA CATTTGACTA AATGTTGACT

    201                                                250
3   ATGAATAATA GCCCTTGGAT TAGATGACAC GTGGACGCTC AGGATCTGTG
7   TGGAATAATA GCCCTTGGAT TAGATGACAC GTGGACGCTC AGGATCTGTG
5   TGGAATAATA GCCCTTGGAT TAGATGACAC GTGGACGCTC AGGATCTGTG
1   A.GTATAATA GCCCTTGGAT TAAATGACAC GTGGACGCTC AGGATCTGTG
4   TGGAATAATA GCCCTTGGAT TAGATGACAC GTGGACGCTC AGGATCTGTG

    251                                      ←CR       300
3   ATGCTAGTGA AGCGCTTAAG CTGAACGAAT CTGACGGAAG AGCGGACATA
7   ATGCTAGTGA AGCGCTTAAG CTGAACGAAT CTGACGGAAG AGCGGACATA
5   ATGCTAGTGA AGCGCTTAAG CTGAACGAAT CTGACGGAAG AGCG......
1   ATGCTAGTGA AGCGCTTAAG CTGAACGAAT CTGACGGAAG AGCGTTCACA
4   ATGCTAGTGA AGCGCTTAAG CTGAACGAAT CTGACGGAAG AGCGGACAAA
```

```
     301                                                        350
3    CGCACATGGA TTATGGCCCA CATGTCTA.A AGTGTATCTC TTTACAGCTA
7    CGCACATGGA TTATGGCCCA CATGTCTA.A AGTGTATCTC TTTACAGCTA
5    .......... TCATGGTCCA CATGTCTA.A AGAATAATGC TTTACAGCTG
1    CTTAGATCTA GTTAGCGTAC TTAGTACG.C GTTGTCTTGG GTCTATAAAT
4    CGCACATGGA CTATGGCCCA CTGCTTTATT AAAGAAGTGA ATGACAGCTG

     351                                                        400
3    TATTGATGTG ACGTAAGATG CTTTACTTCG CCTCGAAGTA AAGTAGGAAA
7    TATTGATGTG ACGTAAGATG CTTTACTTCG CTTCGAAGTA AAGTAGGAAA
5    TATTGATTTG ACTTTACGCG CTTTACTTTA ATTGCTTTAA GTAAAGTAAG
4    TCTTTGCTTC AAGACGAAGT AAAGAATAGT GGAAAACGCG TAAAGAATAA

     401                                                        450
3    TTGCTCGCTA AGTTATTCTT TTCTGAAAGA AATTAATTTA .ATTCTAAAT
7    TTGCTCGCTA AGTTATTCTT TTCTGAAAGA AATTAATTTA .ATTCT.AAT
5    ATGCT..TTA CTTTGCTCGC GACGAAGCAA AGTGATTGTA GCTGCAGAAA
4    GCGTACTCAG TACGCTTCGT GGCTTTATAA ATAGTGCTTC GTCTTATTCT

     451                                                        500
3    TAAATTAAAT GAGTGGCTAT AAATAG.... ..ATGTTTCG TCTTCGTTGT
7    TAAATTAAAT GAGTGGCTAT AAATAGTGTC GATGCTGCCT CACATCGTAT
5    TTGATGCTTT AATTACCGGG TAACACGGTT TGATTGTGGG TATAAATATG
4    TCGTTGTATC ATCAACGAAG AAGTTAAGCT TTGTTCTGCG TTTTA.....

     501                                        537
3    TTTACAACGA AGCTTAGAAT CTTGTGTTA. .......
7    TCTTCTTCGC ATCGTCTGTT CTGGTTTTAA GCG....
5    TTCTGTTCGT TTTCTTCGTT GTCATTTTAC AACGAAG
```

Figure 4.8

Colony hybridisation of full length *Eco*47III clones of common region containing components

The DNA extracted from colonies potentially containing *Eco*47III clones of common region containing SCSV components was probed with a combination of ^{32}P end-labelled component specific primers (Table 2.1). DNA from plasmid clones of components 1 and 5 was blotted onto the filter as positive controls and pGEM5Z plasmid DNA was blotted as a negative control.


+ve
control
-ve
control
+ve
control

4.3 Discussion

The seven DNA components sequenced herein (Chapters 3 and 4) may represent the complete genome of the F isolate of SCSV. No other components have been detected. More than 20 restriction endonucleases were used to create the full length clones of the identified components. Some partial length clones were also created with *Sau*3AI, *Hpa*II and *Hin*PI digests (Chu *et al.*, 1993). Given the small size of the DNAs, however, there may have been other viral components present which did not contain sites for any of the enzymes used and would, therefore, have been missed. Two approaches were taken previously to determine whether more components were present in the F isolate.

The first approach using back-to-back primers, 37A (5'AGAACAGACGATGCG3') and 37B (5'GGTTTTAAGCGATGG3'), based on a random SCSV library obtained from the original field isolate lead to the discovery of component 7 (P. Keese, unpublished). Secondly, a more comprehensive library was constructed from the RF DNA (Chu *et al.*, 1993), using the 4 base restriction enzyme *Hin*PI (because *Hin*PI sites occurred at least once in all of the components), and sequenced. No novel SCSV specific sequences were found. Finally, no new components were found in the search for non-RAP encoding components based on the *Eco*47III library described herein. However, infectivity studies will be required to ascertain the number of essential genome components.

A few sequence variations were encountered among the clones of the F isolate DNAs. It is possible that some of the differences were artefacts of PCR. However, sequence variations in a population of viral genomes are not

unexpected (Domingo *et al.*, 1985). Nothing can be said at this stage as to whether any of the variations affect the viral phenotype, although it is interesting that the amino acid which is potentially (from a proline to an arginine) changed in component 4 is conserved (as a proline) between SCSV 4 and BBTV 6 (Figure 4.5).

The SCSV components are all predicted to have single major ORFs in the virion sense. The alternative ORFs detected in the SCSV DNAs were not flanked by transcription signals, were generally much smaller than the main ORF, and did not match any sequences in the database. Analysis of the transcripts produced by the SCSV DNAs will be required to determine which ORFs are expressed. The apparent unidirectional transcription of SCSV distinguishes it from the geminiviruses, which have bidirectional transcription. The same genome organisation is predicted for BBTV and FBNYV (Chapter 1). However, unidirectional transcription may not be a feature of the SCSV-like virus group. Bidirectional transcription was predicted for CFDV (Rhode *et al.*, 1990), although it has not been confirmed experimentally. Furthermore, PCV, which may be an SCSV-like virus, has bidirectional transcription (Mankertz *et al.*, 1993).

The only ORF which has been firmly identified is the major ORF of SCSV component 5. From N-terminal amino acid sequence it was found to encode the coat protein (Chu *et al.*, 1993). Components 2 and 6 are predicted to encode RAPs (Chapter 3).

Hydrophobic stretches of amino acids, such as that found in the putative gene product of component 1 (Figure 4.4), indicate a potential for the protein to be trans-membrane or membrane associated. This was suggested for the

putative movement protein of MSV (R1) which has a 25 amino acid hydrophobic region (Boulton *et al.*, 1993). The movement proteins of the bipartite geminiviruses, BL1 and BR1 also have hydrophobic domains (Lazarowitz, 1992; Timmermans *et al.*, 1994), and have been shown to be membrane associated (von Arnim *et al.*, 1993; Pascal *et al.*, 1993). This structural similarity with the movement proteins of geminiviruses suggests a possible function for the product of component 1.

The conservation of component 4 and its equivalent in BBTV (component 6) suggests that this type of component has an essential function in the life cycles of these viruses. The observation that no other homologous components (besides the RAP DNAs) have been identified between SCSV and BBTV at this stage suggests that some of these components may have arisen independently.

In addition to viral replication, geminivirus particle assembly takes place in the nucleus of infected cells (Kim *et al.*, 1978). This requires that the coat protein be imported into the nucleus. The BR1 protein, which transports the viral DNA out of the nucleus (Noueiry *et al.*, 1994; Ingham *et al.*, 1995) is also imported into the nucleus and contains several nuclear localisation signals (Ingham *et al.*, 1995). SCSV replication and assembly are predicted to be similar to the geminiviruses. Potential NLSs were found in the putative RAPs of SCSV (see Chapter 3) and the sequences of the coat protein and the other putative proteins were examined for potential NLSs. The coat protein has a potential bipartite type NLS (Raikhel, 1992), **RK**GLRSQ**RRK**, at the N-terminus. The component 1 putative protein has several groups of basic amino acids ^{52}RYAK, ^{79}KRK and ^{88}RERRWAERR (numbered as in the protein sequence, see

Figure 4.1), which may form less compact bipartite type NLSs than that found in the coat protein. The putative component 3 protein (see Figure 4.1) also has several groups of basic residues, including ^{14}KEK, ^{26}KKK, ^{47}KK, ^{75}KKK, ^{83}KK, and ^{91}KWR, the first two and the last three groups are particularly well spaced to form bipartite type NLSs (Robbins *et al.*, 1991). The component 4 and 7 proteins did not have any obvious bipartite type NLSs however the basic residues around positions 43 and 40 respectively are embedded in short hydrophobic regions and could therefore form MAT α2-like NLSs (Raikhel, 1992). The presence of NLS type sequences does not guarantee that a protein will be imported into the nucleus, and more than one NLS may be required, as in *Agrobacterium* VirE2 (Citovsky *et al.*, 1992). Only functional studies can reveal which proteins are localised in the nucleus. It is interesting however that all of the putative SCSV proteins have the potential to be imported into the nucleus.

As discussed in Chapter 3, the conservation of the putative stem-loop structure between geminiviruses and SCSV, and in particular the conservation of the 3' end of the nonanucleotide, 5'TAGTATTAC3', which contains the nick site for geminiviral plus strand replication (Heyraud *et al.*, 1993; Stanley, 1985), suggests that the SCSV-like viruses may also use rolling-circle replication. Studies showing that ACMV and MSV will tolerate a C instead of a T in position 6 of the nonanucleotide (Schneider *et al.*, 1992; Roberts and Stanley, 1994; Stanley, 1985) are surprising considering the complete conservation of the 3' six nucleotides of the nonanucleotide (and the following C) between the two groups (Figure 4.6). With both the ACMV and MSV mutants, however, the appearance of symptoms was delayed. Unlike the SCSV-like viruses and

monopartite geminiviruses, the stem sequences of the bipartite geminiviruses are generally conserved (Lazarowitz, 1987; Roberts and Stanley, 1994).

Common regions, like those found in SCSV components 1, 3, 4, 5, and 7, are also found in BBTV and bipartite geminivirus DNAs (Burns, 1994; Stanley and Gay, 1983; Hamilton *et al.*, 1984; Howarth *et al.*, 1985; Lazarowitz and Lazdins, 1991). Bipartite geminivirus common regions are highly conserved between the A and B components of one virus but are not conserved between viruses (Davies and Stanley, 1989). Like SCSV, the common region sequences of BBTV (which share no sequence identity with that of SCSV) are of variable length when any two DNAs are compared; the minimal common region is split into two sections about 80 and 50 nucleotides in size, and the maximum shared sequence between two components is about 260b (Burns, 1994). The extensions to the minimal common regions in SCSV and BBTV may have arisen by recombination between their genomic components. The observation that the noncoding regions of SCSV components 3, 5, and 7 outside of the common region are more similar to each other than to the noncoding regions of components 1 and 4 supports this possibility (Figures 4.3 and 4.7). The homology between components 3 and 7 is more extensive on the 3' side of the common region whereas the homology between components 3 and 5 is more extensive on the 5' side. This suggests that the similarities between these components represent independent recombination events. The homology on the 5' side of the common region between component 1 and the component 3, 5 and 7 group may represent another recombination (Figures 4.3 and 4.7). Interestingly, this region of homology includes the putative polyadenylation signals and downstream TG rich sequences of components 3,

5, and 7 (Figure 4.3). These signals are mutated in component 1, suggesting that they may not be functional in that component. This is in agreement with the major ORF of component 1 being in a different position than those of components 3, 5, and 7, although it may be that the component 1 transcript has a long 3' untranslated sequence. Recombination between the DNAs of ACMV which repair introduced mutations and deletions occur at a high frequency (Etessami *et al.*, 1989; Klinkenberg *et al.*, 1989; Roberts and Stanley, 1994). Etessami *et al.*, (1989) identified a recombinational hot-spot at the 3' end of the nonanucleotide.

The bipartite geminivirus common regions contain the binding site for the AL1 protein (Figure 1.6; Fontes *et al.*, 1992; Lazarowitz *et al.*, 1992; Thommes *et al.*, 1993). AL1 proteins are virus specific (Lazarowitz *et al.*, 1992; Frischmuth *et al.*, 1993; Gilbertson *et al.*, 1993; Fontes *et al.*, 1994; Sunter *et al.*, 1994; Stenger, 1994). The species specificities of the TGMV and BGMV AL1 proteins are determined by both the AL1 binding site and another sequence element between the binding site and the stem-loop (Figure 1.6; Fontes *et al.*, 1994). The geminivirus common regions therefore function as viral self-recognition determinants. This may be part of their function in SCSV and BBTV. The geminivirus common regions also contain promoter elements for transcription in both directions (Townsend *et al.*, 1985; Sunter and Bisaro, 1989; Sunter *et al.*, 1989; Frischmuth *et al.*, 1991; Haley *et al.*, 1992). Preliminary evidence indicates that SCSV common regions also contain promoter elements (B. Surin, unpublished). The extensions to the common region sequences of some of the SCSV DNAs may allow coordinated regulation of the replication and expression of those DNAs.

CHAPTER 5: VARIATION OF SCSV COMPONENT 5

5.1 Introduction

The symptoms of SCSV infection in the field vary from severe to virtually undetectable (Grylls and Butler, 1959; Chu *et al.*, 1995). The symptom variation may be due to sequence differences in particular components, or the presence or absence of different components.

From mutation studies of geminivirus genes (Chapter 1), and analyses of virus isolates causing different symptoms, it is clear that complex interactions between a number of different viral proteins are responsible for generating and modulating the symptoms of infection in plants.

Mutations in proteins involved in viral DNA replication, such as the AL3 protein of TGMV, generally result in delayed and attenuated symptoms probably due to reduction of the levels of virus (Elmer *et al.*, 1988a; Brough *et al.*, 1988; Etessami *et al.*, 1991; Sunter *et al.*, 1990). Similarly, deletions of the coat proteins of TGMV and BGMV cause a reduction in the levels of viral ssDNA, and a concomitant delay and attenuation of symptoms (Brough *et al.*, 1988; Gardiner *et al.*, 1988; Sunter *et al.*, 1990; Azzam *et al.*, 1994).

Viral movement proteins have also been found to be important determinants of symptom development. Changes in the size and regulation of the plasmodesmatal channels caused by viral movement proteins (Noueiry *et al.*, 1994; Wolf *et al.*, 1989; Deom *et al.*, 1990) are thought to disrupt the regulation of phloem transport, and thus plant growth (Pascal *et al.*, 1993). Indeed, tobacco plants transformed with the BL1 movement protein of SqLCV had symptoms typical of SqLCV infection (Pascal *et al.*, 1993). Symptoms of viral infection were also induced in transgenic tobacco expressing the VI gene

of CaMV; this gene encodes the P6 protein, a translational transactivator and the major constituent of the viroplasm, where particle assembly takes place (reviewed in Pfeiffer and Mensard, 1995).

Viral coat proteins are also associated with symptom development. The potato virus X coat protein is involved in both cell-to-cell and systemic movement of the virus, and deletions in the N-terminal region of the protein caused attenuation of symptoms (Chapman *et al.*, 1992). A single amino acid change in the tobacco mosaic virus (TMV) coat protein, which is involved in systemic movement (reviewed in Maule, 1991), was found to be responsible for a change in symptoms from a light green mosaic to a severe yellow mosaic (Banerjee *et al.*, 1995). Analysis of the electron transport rates of chloroplasts from tobacco infected with severe and mild strains of TMV indicated that the coat protein may affect symptoms by interacting with the thylakoid membranes and disrupting photosynthesis (Reinero and Beachy, 1989). By making recombinants of different CaMV strains and testing them in turnip, Anderson *et al.* (1991) found that the CaMV coat protein determines the severity of stunting.

Sequence variations in the non-coding regions of viruses can also alter symptom expression. The severe and mild Nigerian strains of MSV, differ in the width, length, and severity of chlorosis of the streaking they cause in infected plants (Boulton *et al.*, 1991a). All of these differences, except streak width, were found to be determined by a single nucleotide change in the promoter region of the RAP ORF (Boulton *et al.*, 1991 a, b). Streak width was determined by a single nucleotide change in the R1 movement protein ORF, which did not change the encoded amino acid (Boulton *et al.*, 1991b).

Determination of the molecular basis for SCSV symptom variation has important implications for understanding the viral life cycle and for the development of pathogen derived resistance strategies. Investigating the sequence variation of component 5, which encodes the coat protein, was chosen as the starting point for the study of SCSV symptom variation because it contains the only gene that has been assigned a definite function and, as mentioned above, changes in the coat protein sequences of viruses can affect symptom expression.

A number of symptom variants of SCSV derived from field samples have been maintained in the glasshouse by passaging with *A. craccivora* or propagation of infected plants from cuttings (Figure 5.1). Previous work (Boevink, 1991) indicated that there was significant variation between partial component 5 sequences of isolates F, D, and B1.

Initially the aim was to determine whether the high degree of symptom variability was mirrored by sequence variation. This was done by sequencing component 5 DNA from the glasshouse isolates and analysing field samples by PCR and sequencing. The non-coding sequence of component 5 was the most thoroughly examined as it was the most variable, and this allowed the identification of sequences which were highly conserved between isolates and therefore may contain important regulatory signals.

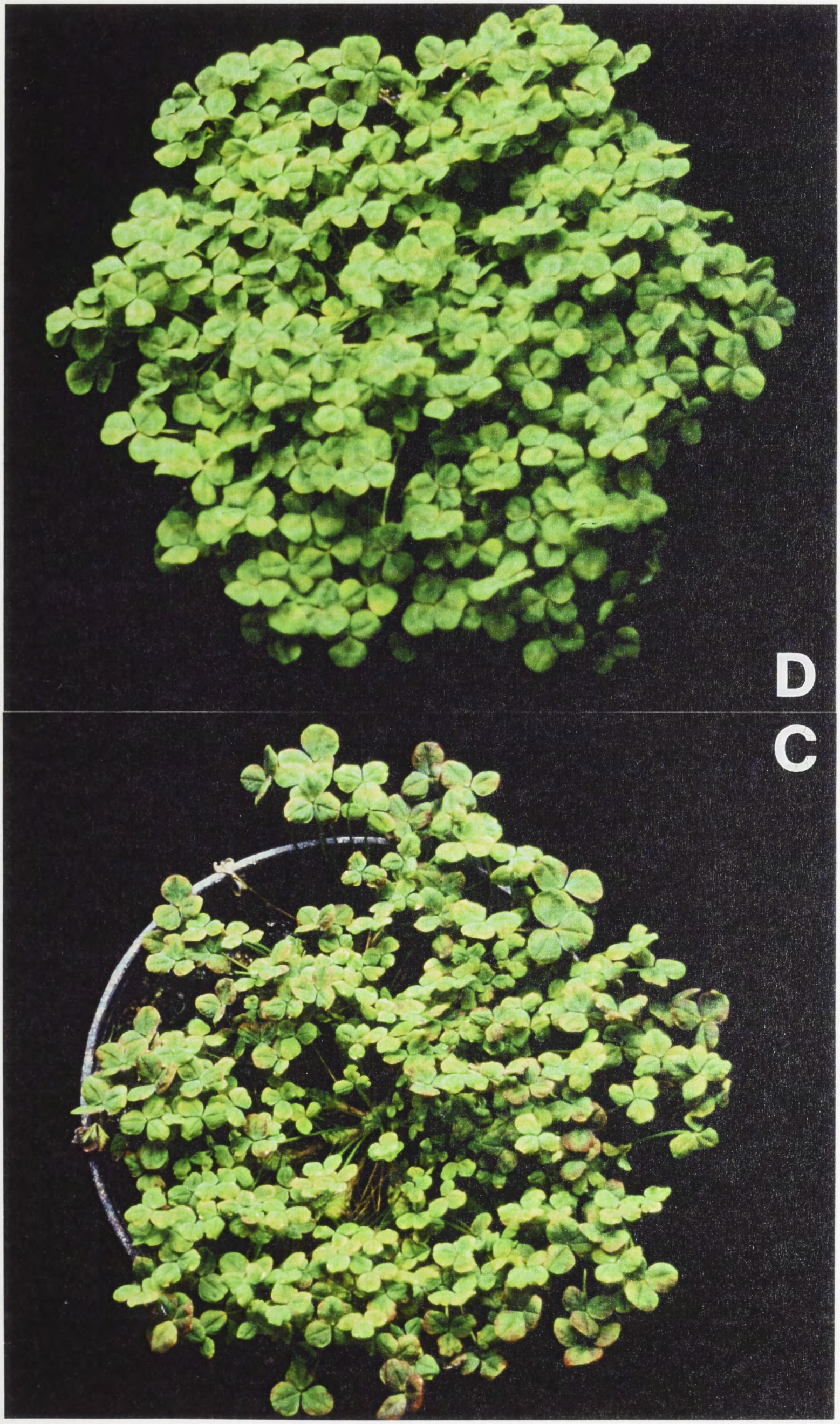
5.2 Results

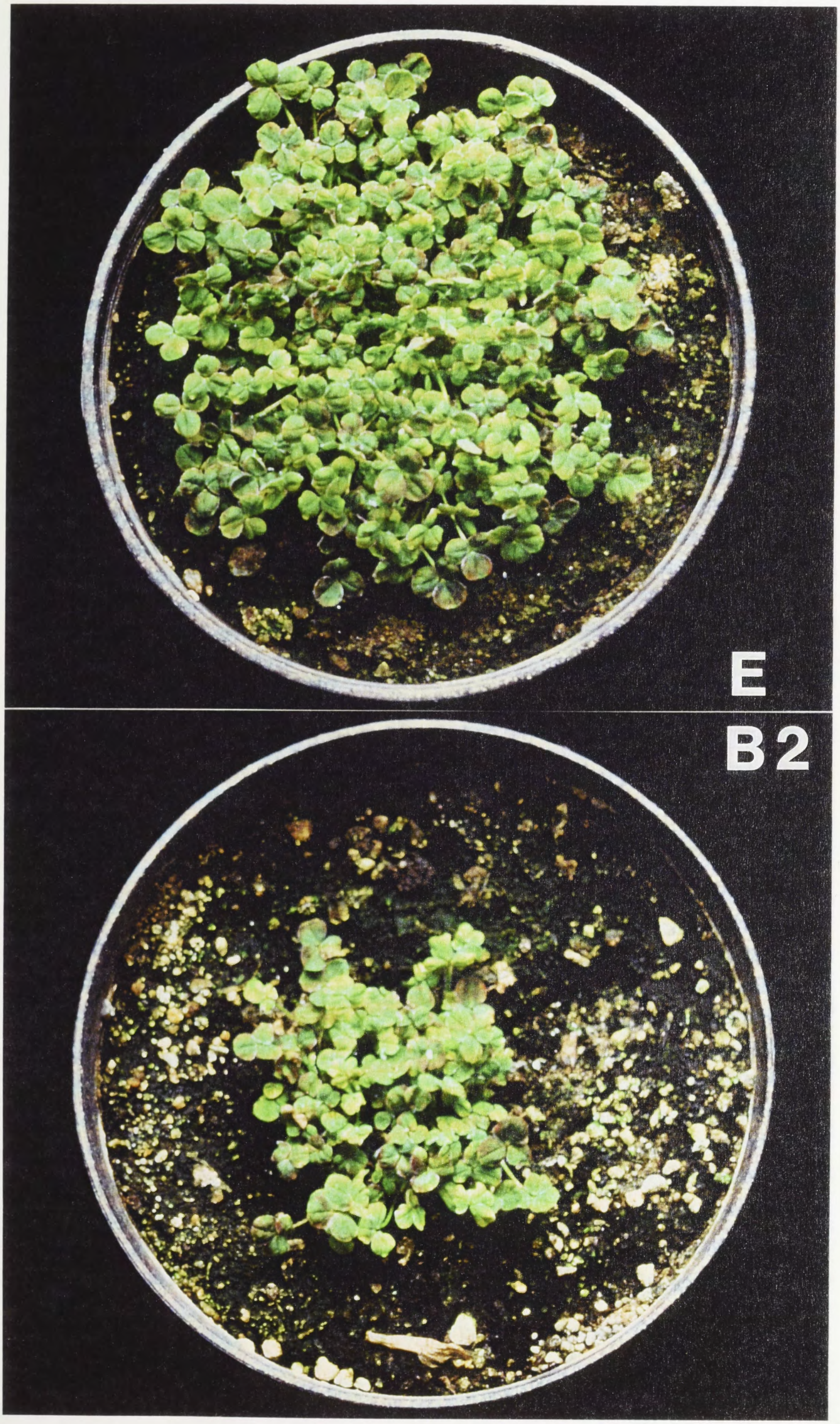
Component 5 sequences from ten symptom variants from the glasshouse and six field isolates were amplified by PCR and sequenced. Different primer sets were required to obtain sequences from the same region

Figure 5.1

Glasshouse maintained isolates of SCSV

Subterranean clover plants infected with different SCSV isolates (C, D, E, and B2). Symptom differences between the isolates include the degree of stunting and reddening.


D
C


E
B2

of component 5 due to the high degree of sequence variability between the different isolates. The sequences compiled were therefore different lengths and were amplified from slightly different positions. The sequences fell into two groups or strains, which were called F and JL3 strains after the isolates from which the entire component 5 sequences were obtained.

5.2.1 Differential PCRs of SCSV component 5

DNA from the ELISA positive field samples, JL3, JL4, JL8, JL9, JL11, JL12, and JL13, and the ELISA negative control sample, JL1, was obtained as described in Chapter 3. Three component 5 specific primer combinations; SUNI5+5-40mer, 5Hind+5HinPI, and 5Hind+5Pst (Table 5.1), were used for PCRs on the field isolate samples and the control samples; F isolate, isolate B1, and healthy (Figures 5.2 and 5.3; Table 5.2). B1 was included as a control because it was known from previous work (Boevink, 1991) that isolate B1 had a variant of component 5 to which the latter two primer combinations would not bind.

The SUNI5 and 5-40mer primers were expected to detect any variant of component 5 as SUNI5 binds in the highly conserved common region (see Chapter 4) and 5-40mer is a large primer which binds within the coding region, which was expected to be reasonably well conserved. The expected size product (331bp) was amplified from all except the healthy, JL1, and JL8 samples. The healthy and JL1 samples were ELISA negative and, therefore, were not expected to contain any virus. JL8 was ELISA positive, however, PCRs with other primers on JL8 also did not give any product (Chapter 3). This preparation may have contained inhibitors to the PCR, or the DNA may have been lost or degraded during the extraction procedure.

Table 5.1

Sequences and positions of, and fragment sizes produced by, PCR primers to component 5.

The sequences, positions of and PCR product sizes from the primers used to amplify component 5 sequences from the glasshouse and field isolates of SCSV are given in tabular (a) and diagrammatic (b) form. In b., the F and JL3 component 5 sequences (green and blue rectangles respectively) are represented linearly, with nucleotide 1 at the left. Therefore, the product of PCR with 5DL and 5DR primers, which amplify a full length fragment from the JL3 type component 5, is represented with a small gap to indicate where the primers are.

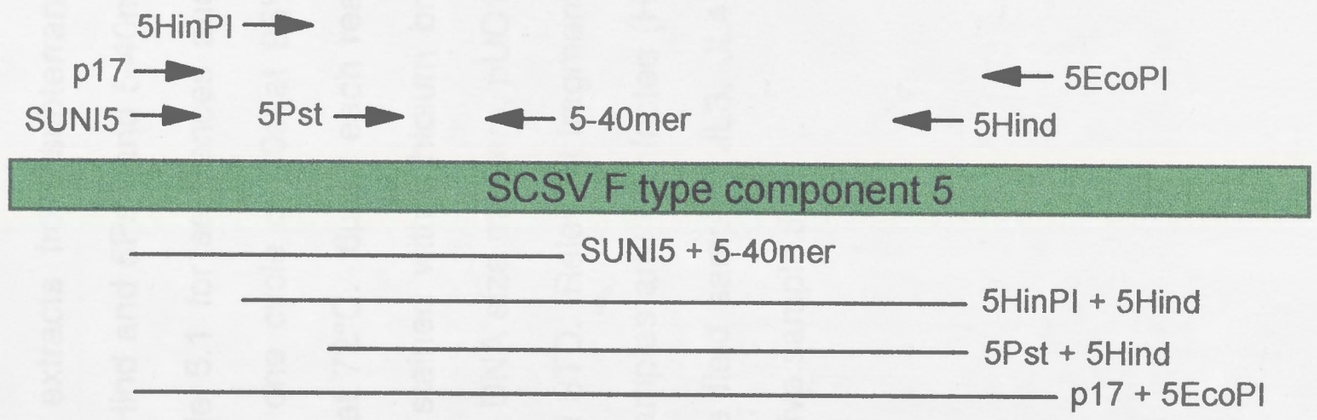
Table 5.1 a.

Primer 1 name	Position of 5' end	Primer 1 sequence (5'-3')	Primer 2 name	Position of 5' end	Primer 2 sequence (5'-3')	Fragment size (bp)
5-40mer	423	CGAATATTTTCTCCTTTGAGACCTC AGACCCTTTCTTCC	SUNI5	93	(see Table 3.1)	331
5HinPI	181	GTATTGATTTGACTTTACGCGCTTT	5Hind	737	CCAGCAGTACCAGAAGAC	557
5Pst	249	CGACGAAGCAAAGTGATTG	"	"	"	489
p17	97	TGATGCTAGTGAAGCG	5EcoPI	802	GCTGGGTCTGACTAATACCTG	706
5DR	187[a]	TGGATGACAGCTGTAGATGC	5DL	186[a]	CTTCTTTATTAAAGTAGTGGGC	~1kb[b]

[a] position in JL3 component 5 sequence

[b] the exact size of the JL3 type component 5 was unknown

Table 5.1 b.


5HinPI
p17
SUNI5
5Pst
5-40mer
5EcoPI
5Hind
SCSV F type component 5
SUNI5 + 5-40mer
5HinPI + 5Hind
5Pst + 5Hind
p17 + 5EcoPI

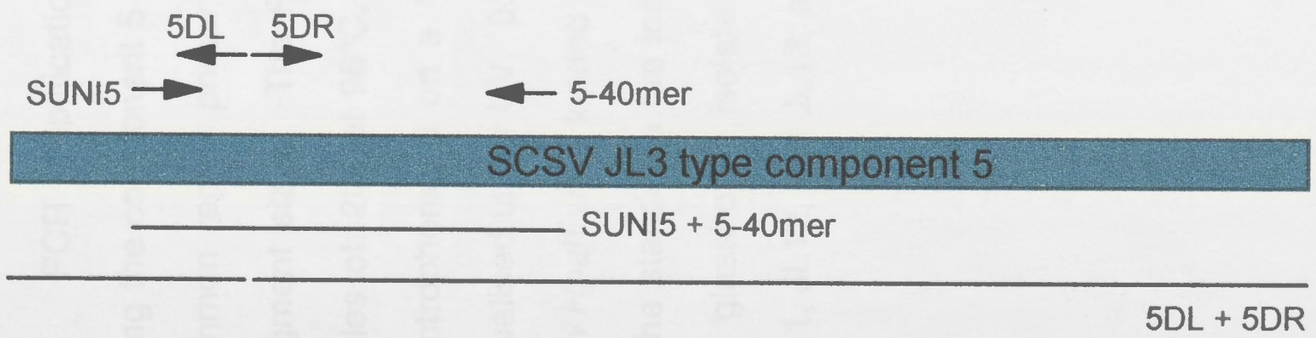

5DL
5DR
SUNI5
5-40mer
SCSV JL3 type component 5
SUNI5 + 5-40mer
5DL + 5DR

Figure 5.2

Component 5 PCRs with different primer sets

PCR amplification of nucleic acid extracts from subterranean clover using the component 5 specific primers, 5Hind and 5Pst, and 5-40mer and the common region primer SUNI5. See Table 5.1 for sequences and expected fragment sizes. The PCR program was: one cycle of 30s at 96°C then 32 cycles of: 20s at 96°C, 40s at 52°C, 90s at 72°C. 6µl of each reaction were electrophoresed on a 1.2% agarose gel, stained with ethidium bromide and visualised under UV. 0.5µg samples of the DNA size markers, pUC19 digested with *Hpa*II were loaded in the lanes marked STD. Relevant fragment sizes (bp) of the standards are indicated. The test samples are: uninfected (H), F isolate (F), glasshouse isolate B1, ELISA positive field samples JL3, JL4, JL8, JL9, JL11, JL12, and JL13, and an ELISA negative sample JL1.

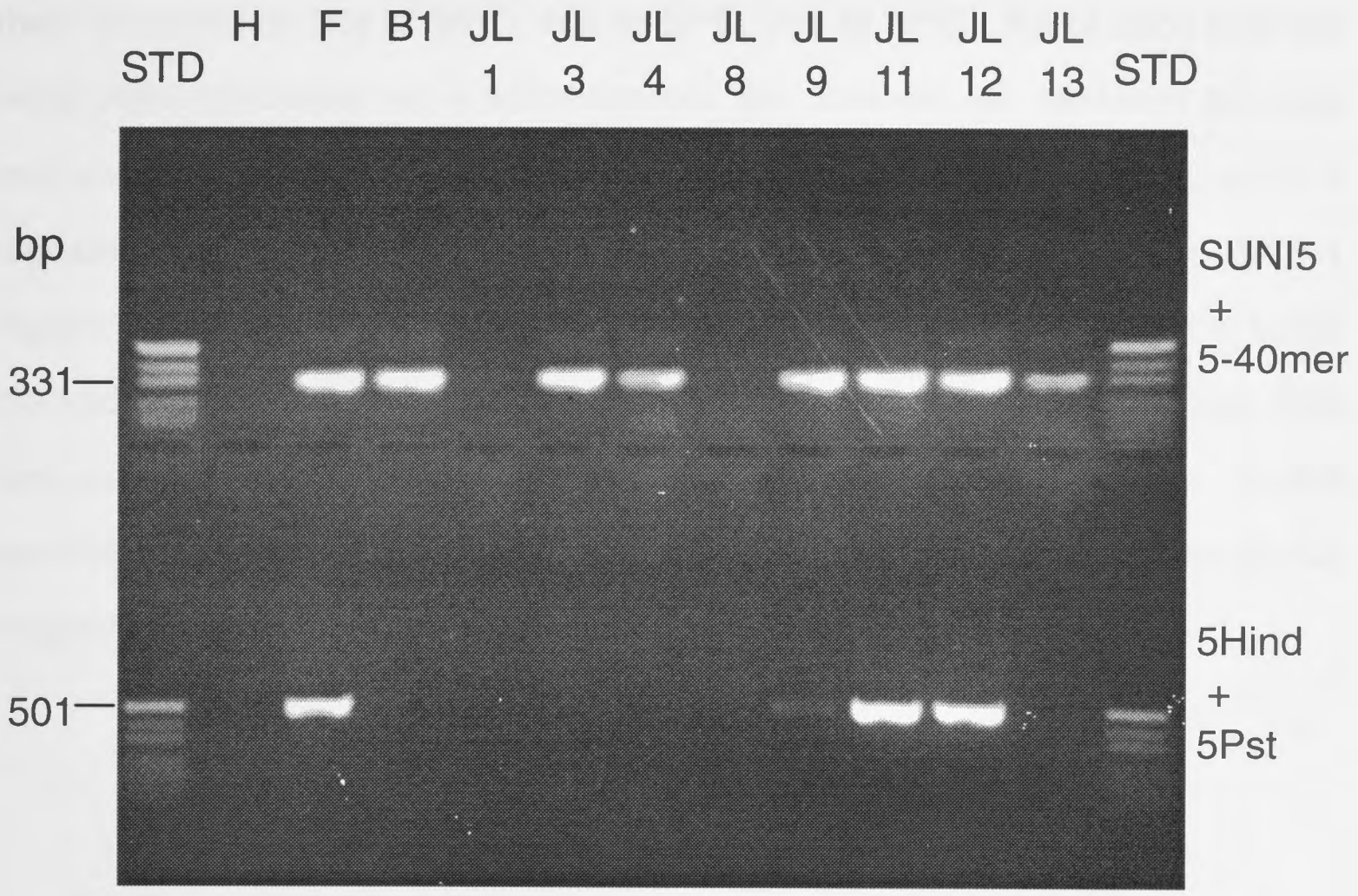

STD
H
F
B1
JL 1
JL 3
JL 4
JL 8
JL 9
JL 11
JL 12
JL 13
STD
bp
331—
501—
SUNI5 + 5-40mer
5Hind + 5Pst

Figure 5.3

Component 5 PCRs with different primer sets

PCR amplification of nucleic acid extracts from subterranean clover using the F type component 5 specific primers 5Hind and 5HinPI, and the JL3 type component 5 primers 5DL and 5DR. See Table 5.1 for sequences and expected fragment sizes. The PCR program was: one cycle of 30s at 96°C then 32 cycles of: 20s at 96°C, 40s at 52°C, 90s at 72°C. 6µl of each reaction were electrophoresed on a 1.2% agarose gel, stained with ethidium bromide and visualised under UV. 0.5µg samples of the DNA size markers, pUC19 digested with *Hpa*II (in the upper half of the gel) or bacteriophage SPP-1 digested with *Eco*RI (in the lower half of the gel) were loaded in the lanes marked STD. Relevant fragment sizes (bp) of the standards are indicated. The test samples are: uninfected (H), F isolate (F), glasshouse isolate B1, ELISA positive field samples JL3, JL4, JL8, JL9, JL11, JL12, and JL13, and an ELISA negative sample JL1.

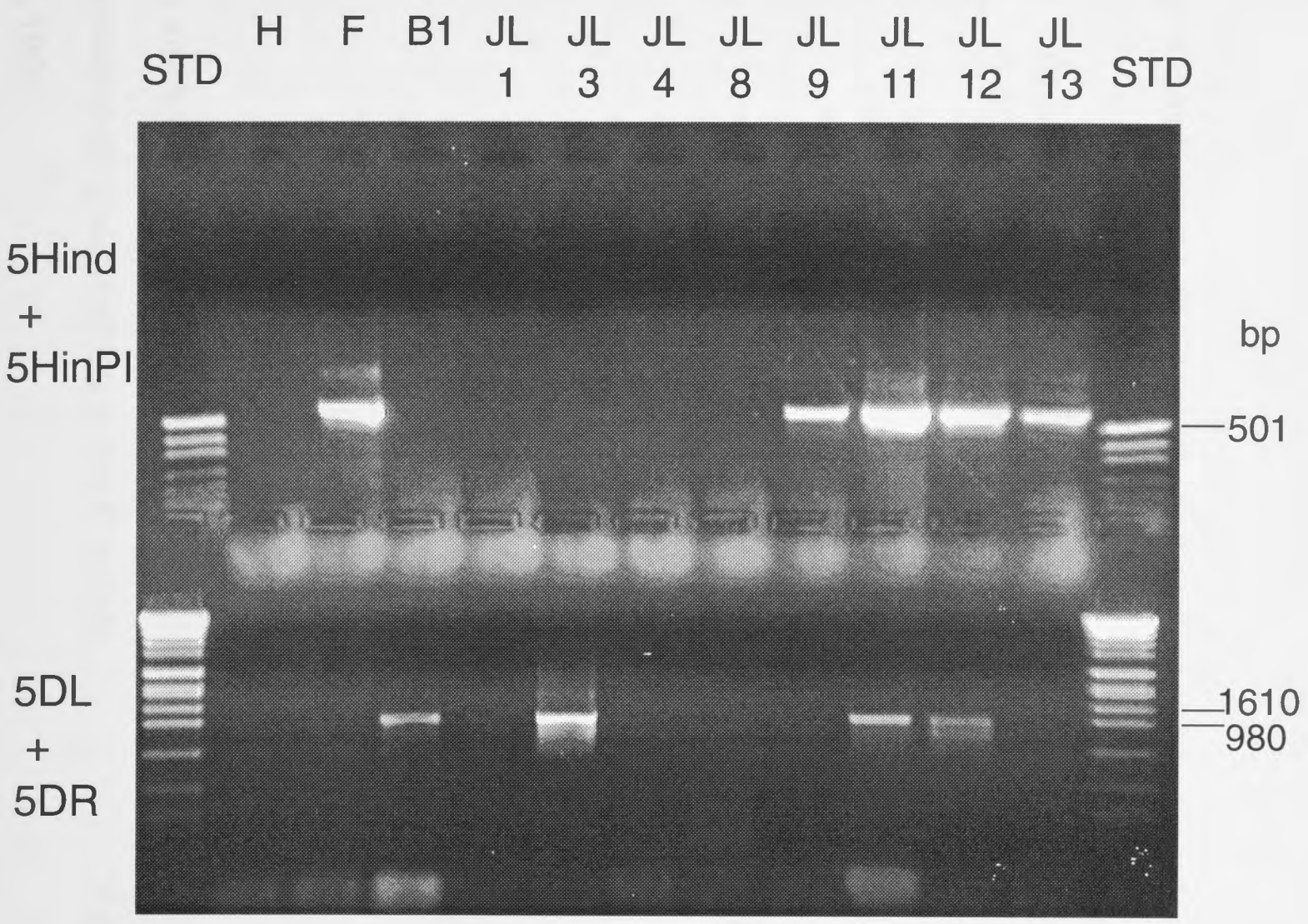

STD
H
F
B1
JL 1
JL 3
JL 4
JL 8
JL 9
JL 11
JL 12
JL 13
STD
5Hind + 5HinPI
5DL + 5DR
bp
501
1610
980

Table 5.2

Summary of component 5 PCR results and sequenced fragments of the glasshouse and JL field isolates.

Isolate	Component 5 strain present (F/JL3)	PCR primer combinations				Sequence obtained [a, b]	Length of sequence
		SUNI5 + 5-40mer	5Hind +5Pst	5Hind + 5HinPI	5DL + 5DR		
F	F	+	+	+	-	1-998	998
F2[c]	F	+	+	+	ND	93-412	321
A1[c]	F	+	+	-	ND	95-247	121
A2[c]	F	+	+	-	ND	93-414	290
C[c]	F	+	+	+	ND	93-295	204
D[c], D12	JL3	+	-	-	ND	97-802,99-410	715, 318
E[c]	F	+	+	+	ND	93-417	324
B1	JL3	+	-	-	+	97-802	714
B2[c]	F	+	+	+	ND	93-341	248
JL3	JL3	+	-	-	+	1-1012	1012
JL4	ND	+	-	-	-	ND	ND
JL9f, d	F + JL3(?)	+	$+_w$	+	-	181-420 & 508-736,117-411	470, 296
JL11	F + JL3	+	+	+	+	ND	ND
JL12	F + JL3	+	+	+	+	114-384	272
JL13	F	+	-	+	-	ND	ND

$+_w$: weak band on gel due to inefficient PCR

ND: not done

[a] numbering according to corresponding nucleotides in the F isolate

[b] the coding region is from nucleotides 367 to 873

[c] PCR results from Boevink, 1991

The 5Hind primer binds within the coding region, and like the 5-40mer primer, was expected to detect most variants of component 5. The 5HinPI and 5Pst primers, however, bind within the noncoding region and previous work showed that they did not bind to some isolates (Boevink, 1991). This was also the case with the field isolates when they were used in combination with the 5Hind primer (Figures 5.2 and 5.3; Table 5.2). The differences in DNA amplification by PCR with DNA from the various field isolates indicated that there were differences between the component 5 sequences of these isolates.

5.2.2 Cloning and sequencing of component 5 fragments

The fragments amplified from JL3, JL9, and JL12 with the SUNI5 and 5-40mer primers were chosen for initial sequencing to examine the sequence variation present among the field isolates. The same fragments from the glasshouse isolates F2, A1, A2, C, D (clone D12), E, and B2 were also chosen to further the analysis of those isolates. The fragments amplified by SUNI5 and 5-40mer were used because they encompassed the most variable region of component 5, that is, between the common region and the start of the coat protein ORF. The fragments amplified by primers 5Hind and 5HinPI from isolate JL9 (clone JL9f), and primers 5EcoPI and p17 (Table 5.1) from isolates D and B1, were also cloned and sequenced.

The ends of the PCR fragments to be sequenced were repaired with T4 DNA polymerase and the blunt-ended fragments were ligated into pGEM5Z-*Eco*RV (Promega) as described in Methods 2.2. The ligated plasmids were electroporated into *E. coli* strain JM109 and putative clones were selected by restriction analysis. Clones were sequenced with the T7 and SP6 promoter primers (Promega) as described in Methods 2.4.

The component 5 of the JL3 field isolate was completely sequenced from clones of the fragments amplified by SUNI5 and 5-40mer, and 5DL and 5DR (full length). Two representative full length clones were sequenced with the T7 and SP6 promoter primers and the SCSV specific primers 5Hind (Table 5.1), p16, 5seq520, 5seq630, and Jlseq920 (Table 5.3). The two full length clone of the JL3 component 5 had two nucleotide differences. These may have been due to errors by *Taq* DNA polymerase, although the error rate of this enzyme makes it unlikely that both differences were due to incorporation errors (Tindal and Kunkel, 1988).

5.2.3 Component 5 relationships between different isolates

The sequenced component 5 fragments from the different isolates were compared using GAP (Figure 5.4). The sequences fell into two groups, which I have called the F and JL3 strains. Within each group the sequences are greater than 90% similar. Some of the similarity values were lower because only a short sequence entirely within the highly variable part of the noncoding region was obtained (eg. B2), or because of the 28b deletion in isolates A1 and A2. Between the two groups, the sequences were an average of 68.5% similar. Figures 5.5 and 5.6 show the sequence variations found within each group.

To distinguish detection of the JL3 strain component 5 from the F strain, the outward pointing primers, 5DL and 5DR (Table 5.1), were designed to sequence from the JL9d clone of the JL9 isolate, which was closely related to the JL3 sequence. These primers amplified the expected size (approximately 1kb) fragments from B1, JL3, JL11, and JL12 (Figure 5.3; Table 5.2). They did not amplify this fragment from JL9 despite the fact that they had been created using JL9 sequence. The clone JL9d may have been a result of cross

Table 5.3

Sequences and positions of sequencing primers for JL3 component 5

Primer name	Position of 5' end [a]	Primer sequence (5'-3')
p16	97	CAGATCCTGAGCGTCC
5seq520	480	CAGAACAGACTCCACATG
5seq630	622	CTATTGCGAATTGGAGTTC
JLseq920	920	GATTAATTGTACTCTGCG

[a] position in JL3 component 5 sequence

Figure 5.4

Percent similarities between component 5 sequences of different SCSV isolates

The percent similarities between the sequenced component 5 fragments from the different glasshouse and field isolates, were determined by GAP. The sequences fall into 2 groups, related to the F or JL3 sequences. The shaded areas contain comparisons between component 5 sequences in the same group. D12 is a separate clone from isolate D which had a number of nucleotide differences from the clone represented by D (Figure 5.6).

	F	F2	A1	A2	C	E	B2	JL12	JL9f	D	D12	B1	JL3	JL9d
F		99.7	91.7	97.2	96.0	96.0	96.8	98.5	97.4	79.4	69.5	78.8	79.7	68.0
F2			91.7	99.3	95.6	95.9	96.4	98.2	95.7	69.7	69.6	67.3	68.2	66.8
A1				99.2	94.2	92.6	87.6	93.1	*	71.2	70.7	70.6	69.4	62.6
A2					98.8	95.8	97.2	99.6	83.8	73.0	73.2	71.9	72.8	71.9
C						95.6	93.1	97.3	98.3	66.8	67.0	67.0	67.6	63.3
E							92.7	97.0	95.7	68.4	68.1	68.3	71.0	69.7
B2								95.6	91.2	63.5	63.5	64.2	65.3	61.6
JL12									96.6	64.8	64.8	65.5	65.3	64.9
JL9f										67.4	66.4	67.4	68.2	67.9
D											98.0	94.5	95.6	93.9
D12												90.5	93.6	93.9
B1													96.5	95.6
JL3														99.2

* Due to the deletion in A1, there was not enough overlap with the JL9f sequence to obtain an accurate percent similarity.

Figure 5.5

Variation of F strain component 5 sequences

The partial sequence of component 5 of the F isolate is shown; nucleotide differences found in the other F-strain component 5 sequences are shown below it. Deletions are represented by dashes and insertions indicated by spaces in the consensus F sequence (with the inserted sequence given below). The initiation codon is in **bold** and the SUNI5, 5-40mer and 5HinPI primer binding sites are underlined.

```
    ---------+---------+---------+---------+---------+---------+ 120
    CCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGATGCTAGTGAAGCGCTTAAGCT
                                          (SUNI5)
    ---------+---------+---------+---------+---------+---------+ 180
    GAACGAATCTGACGGAAGAGCGTCATGGTCCACATGTCTAAAGAATAATGCTTTACAGCT
A1  ............................C................................
A2  ............................C................................
B2  ............................C.............TG..TCT..........
C   ............................C................................
E   .......C....................C.............T..................
JL12............................C................................
    ---------+---------+---------+---------+---------+---------+ 240
    GTATTGATTTGACTTTACGCGCTTTACTTTAATTGCTTTAAGTAAAGTAAGATGCTTTAC
A1  (5HinPI)G....----------------------------.....................
A2  ........G....----------------------------T.....................
B2  ....................................................T.....................
F2  ....................................................G.....................
C   ........G.................A...GC..A.....................
E   ........G.........A.A..........C.....................
JL9f........................A...GC.....................
JL12.......G.....................C.....................
    -----.----+---------+---------+---------+---------+---------+ 300
    TTTGC.TCGCGACGAAGCAAAGTGATTGTAGCTGCAGAAATTGATGCTTTAATTACCGGGT
A1  .....T.....................................................
A2  .....T.....................................................
B2  .....T.....................................................
F2  .....T.....................................................
C   .....T......A.................................................
E   ..............................................TAT....
JL9f.....T......A..............................T.T.A..
    ---------+---------+---------+---------+---------+---------+ 360
    AACACGGTTTGATTGTGGGTATAAATATGTTCTGTTCGTTTTCTTCGTTGTCATTTTACA
B2  .............--.....................................
E   ..................T.....................C....-.....................
JL9f.....-.....................................C.....................
JL12.....................................C.....................
    ---------+---------+---------+---------+---------+---------+ 420
    ACGAAGATGGTTGCTGTTCGATGGGGAAGAAAGGGTCTGAGGTCTCAAAGGAGAAAATAT
JL9f.............................C.............(5-40mer).....
    ---------+---------+---------+---------+---------+---------+ 480
    TCGCGAATTGCTTACAAACCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAGTCTGTTCTG
    ---------+---------+---------+---------+---------+---------+ 540
    AATAAGAGAGATGTTACTGGAGCGGAGGTTAAGCCATTCGCTGATGGTTCAAGGTATAGT
JL9f.....................................................C.....................
    ---------+---------+---------+---------+---------+---------+ 600
    ATGAAGAAGGTAATGTTGATTGCAACATTAACTATGGCTCCTGGAGAATTAGTTAATTAT
    ---------+---------+---------+---------+---------+---------+ 660
    CTTATTGTGAAGAGTAATTCGCCTATTGCGAATTGGAGTTCGTCTTTCAGTAATCCTTCG
JL9f.....................................................C.....................A
```

Figure 5.6

Variation of JL3 strain component 5 sequences

The partial sequence of component 5 of the JL3 isolate is shown: nucleotide differences found in the other JL3-strain component 5 sequences; from isolates D, B1, and JL9 (clone d), are shown below it. Deletions are represented by dashes and insertions indicated by spaces in the JL3 sequence (with the inserted sequence given below). The initiation codon is in **bold** and the SUNI5 and 5-40mer primer binding sites are underlined.

```
         ---------+---------+---------+---------+---------+---------+ 130
         AAATAATAAATGGACGCTCAGGATCTGTGTTGCTAGTGAAGCGCTTAAGCTGAACGAAAC
D12      ..........................(SUNI5).....................G.
         ---------+---------+---------+-----...----+--...-------+----.184
         TGACGGAAGAGCGGACATACGCAC...ATGGATTA...TGGCCCACTACTTTAATAAAGA
D        ..................TGT.T.TTTC.CTC.C.........................
D12      ..................TGT.G.TTT..CTC.C.........................
B1       ..............................C..CCG.......................
         -----+---------+---------+---------+---------+---------.+-.-- 243
         AGTGTATGACAGCTGTAGATGCTTTACTTAGCTTCGTCTCGAAGCAAAGTAAGAT.GC.TT
JL9      ....G...................................................
D        ......C.....................T.............T..............
D12      ....G.C.....................T.............T..............
B1       ....G.................................T.............A..C..
         ------+-..--------+---------+---------+--------.-+---------+ 300
         TGTCACTT..TTTACTTTACTCTATTATATGCGTGTAGCTGTAGAAA.TTGATTCATTAA
D        ........TA.............T...........G.....................
D12      ........TA..............................................
B1       ......-.....................................T...........
         ---------+---------+---------+---------+---------+---------+ 360
         TTTCTGGGTAACACGGTTTGAATTTTAGTATAAATAGAGTTCTTCTTCAATTGTTTTCAC
D        ..........................................G.....................
D12      .....T.....................................G.....................
B1       ..............-.....................................................
         ---------+---------+---------+---------+---------+---------+ 420
         AATTGAAGATGGCGCAGTTACGATGGGGACGAAAGGGTGTGAGGTCTCAGAGGAGGAAAT
D        ............T.........................(5-40mer).............
D12      ............T.....................................................
B1       ...C.........T.....................................................
         ---------+---------+---------+---------+---------+---------+ 480
         ATTCACGGCCAGTTGCTTACAAACCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAAACTG
D        ..........................................C..........C.............
B1       ..........................................C..........C.............
         ---------+---------+---------+---------+---------+---------+ 540
         TTCTGAATAAGAAGGATGTTACTGGCGCGGAGATGAAGCCGTTTAATGATGGTTCAAGGT
D        ......................A.....T.................................
B1       ....................A.....................................
         ---------+---------+---------+---------+---------+---------+ 600
         ATAGCATGAAGAAGATAATGGTGTCCGCAACATTAACGATGGCTCCTGGAGAGTTGCTGA
D        ......................TG.T..G.........................A.......
B1       ......................TG.T..G.........................A.....T.
         ---------+---------+---------+---------+---------+---------+ 660
         ACTATCTGATTGTGAAGAGTAATTCGCCTATTGCGAATTGGAGTTCTTCGTTCAGTAATC
B1       ............T.....................................................
         ---------+---------+---------+---------+---------+---------+ 720
         CGTCGTTAATGGTGAAAGAGTCTGTGCAAGATACGGTAACGATTCTGAGAAGAGGAAAGC
D        ..........................................................C........
         ---------+---------+---------+---------+---------+---------+ 780
         TGGAGTCTTCTGGGACTGCTGGTAAAGATGTAACTAAGTCTTTTAGTCGATTTGTTAATT
D        ..................................C.............A.............
B1       .T..................................................................
         ---------+---------+---------+---------+---------+---------+ 840
         TGGGTCTAGGGATTAGTCAGACCCAGCATTTGTATTTAATTATTTATTCCAGTGATGCAA
D        .....TC...T.....................................................
```

contamination of the PCR reaction mixture by one of the other samples. The JL9 sample did contain an F strain component 5 (clone JL9f). The JL11 and JL12 samples appeared to contain both F and JL3 strain component 5 sequences as products were amplified by both the F-specific primers 5HinPI and 5Pst, and JL3-specific primers 5DL and 5DR (Table 5.2).

When the complete component 5 sequences of F and JL3 were compared using GAP (Figure 5.7) it could be seen that the majority of the sequence differences are within the noncoding region. The values for sequence similarities are: 68.8% in the noncoding region, 87.8% in the coding region, and 79.7% overall. The protein sequences encoded by the major ORF of F and JL3 component 5s were compared using GAP (Figure 5.8) and found to share 88.2% identity. There are 21 differences in the minimal common region, which was defined by the alignment of the F isolate components (Figure 4.6). This included differences in the stem of the conserved stem-loop structure.

5.3 Discussion

Sequences of component 5 from a number of SCSV isolates, which elicit a variety of symptoms in subterranean clover, were obtained to determine if the phenotypic variation was reflected in the viral genome sequence. Indeed, the component 5 sequences from different isolates contained a large number of nucleotide variations. The sequence variants fell into two groups, or 'strains', related to either F or JL3. Both strains of component 5 were found in both the glasshouse and the field isolates.

Figure 5.7

Alignment of F and JL3 component 5 sequences

The sequences of component 5 of the F and JL3 isolates were aligned using GAP. The conserved nonanucleotide in the stem-loop is in bold, the initiation and termination codons of the major ORF are bold and underlined, and the ends of the minimal common region (see Chapter 4) are indicated by CR→ and ←CR.

```
                .         .         .         .         .
  1 TAGTATTACCCCCGCCCCGGGGTCAGAGACATTTGACTAATTGTTGACTA 50
    ||||||||| |||||  |||||||||||||||||||||||| | ||||||
  1 TAGTATTA.CCCCGTGCCGGGGTCAGAGACATTTGACTAAATATTGACTT 49
                .         .         .         .         .
 51 TGCTTAATAGCCGTTAGATGAAATAATAAATGGACGCTCAGGATCTGTGT 100
      |  |||||||| || ||| | || | |  ||||||||||||||||||||
 50 GGAATAATAGCCCTTGGATTAGATGACACGTGGACGCTCAGGATCTGTGA 99
                .         .         .       ←CR       .
101 TGCTAGTGAAGCGCTTAAGCTGAACGAAACTGACGGAAGAGCGGACATAC 150
    ||||||||||||||||||||||||||| ||||||||||||||
100 TGCTAGTGAAGCGCTTAAGCTGAACGAATCTGACGGAAGAGCGTCATGGT 149
                .         .         .         .         .
151 GCACATGGATTATGGCCCACTACTTTAATAAAGAAGTGTATGACAGCTGT 200
      ||||||  | | |    |   ||  |      |    | |||   || |
150 CCACATGTCTAAAGAATAATGCTTTACAGCTGTATTGATTTGA...CTTT 196
                .         .         .         .         .
201 AGATGCTTTACTTAGCTTCGTCTCGAAGCAAAGTAAGATGCTTTGTCACT 250
    |   |||||||||   ||   ||  ||| ||||||||||||||   |||
197 ACGCGCTTTACTTTAATT..GCTTTAAGTAAAGTAAGATGCTTT...ACT 241
                .         .         .         .         .
251 TTTTACTTTACTCTATTATATGCGTGTAGCTGTAGAAATTGATTCATTAA 300
    ||   |   ||       |  ||  |||||||| |||||||||| | ||||
242 TTGCTCGCGACGAAGCAAAGTGATTGTAGCTGCAGAAATTGATGCTTTAA 291
                .         .         .         .         .
301 TTTCTGGGTAACACGGTTTGAATTTTAGTATAAATAGAGT.......TCT 343
    ||  | |||||||||||||||| | |  |||||||||    |       | |
292 TTACCGGGTAACACGGTTTGATTGTGGGTATAAATATGTTCTGTTCGTTT 341
                .         .         .         .         .
344 TCTTCAATTGTTTTCACAATTGAAGATGGCGCAGTTACGATGGGGACGAA 393
    |||||  |    ||     |  ||||||||       | |||||||||  |||
342 TCTTCGTTGTCATTTTACAACGAAGATGGTTGCTGTTCGATGGGGAAGAA 391
                .         .         .         .         .
394 AGGGTGTGAGGTCTCAGAGGAGGAAATATTCACGGCCAGTTGCTTACAAA 443
    |||||  |||||||||| |||||  |||||||||   || | ||||||||||
392 AGGGTCTGAGGTCTCAAAGGAGAAAATATTC...GCGAATTGCTTACAAA 438
                .         .         .         .         .
444 CCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAAACTGTTCTGAATAAGAA 493
    |||||||||||||||||||||||||||||||||  ||||||||||||||||
439 CCTCCTTCGTCTAAGGTTGTAAGTCATGTGGAGTCTGTTCTGAATAAGAG 488
```

```
494 GGATGTTACTGGCGCGGAGATGAAGCCGTTTAATGATGGTTCAAGGTATA 543
    ||||||||||| |||||| | ||||| ||   ||||||||||||||||||
489 AGATGTTACTGGAGCGGAGGTTAAGCCATTCGCTGATGGTTCAAGGTATA 538

544 GCATGAAGAAGATAATGGTGTCCGCAACATTAACGATGGCTCCTGGAGAG 593
    | |||||||||| ||||| ||   |||||||||||| ||||||||||||||
539 GTATGAAGAAGGTAATGTTGATTGCAACATTAACTATGGCTCCTGGAGAA 588

594 TTGCTGAACTATCTGATTGTGAAGAGTAATTCGCCTATTGCGAATTGGAG 643
    ||  | || ||||| ||||||||||||||||||||||||||||||||||||
589 TTAGTTAATTATCTTATTGTGAAGAGTAATTCGCCTATTGCGAATTGGAG 638

644 TTCTTCGTTCAGTAATCCGTCGTTAATGGTGAAAGAGTCTGTGCAAGATA 693
    ||| || |||||||||||| ||||| |||||||||||||||| |||||||
639 TTCGTCTTTCAGTAATCCTTCGTTGATGGTGAAAGAGTCTGTTCAAGATA 688

694 CGGTAACGATTCTGAGAAGAGGAAAGCTGGAGTCTTCTGGGACTGCTGGT 743
    | || ||||||  |  || |||||||||| |||||||||| |||||||||
689 CAGTTACGATTGTTGGAGGAGGAAAGCTTGAGTCTTCTGGTACTGCTGGT 738

744 AAAGATGTAACTAAGTCTTTTAGTCGATTTGTTAATTTGGGTCTAGGGAT 793
    ||||||||||||||||||||||    |||||||| ||||| ||| ||
739 AAAGATGTAACTAAGTCTTTTAGGAAGTTTGTTAAGCTGGGTTCAGGTAT 788

794 TAGTCAGACCCAGCATTTGTATTTAATTATTTATTCCAGTGATGCAATGA 843
    |||||||||||||||||||||||||||||||||||||||||||| ||||
789 TAGTCAGACCCAGCATTTGTATTTAATTATTTATTCCAGTGATGCGATGA 838

844 AAATTACACTGGAGACGAGAATGTATATTGATGTATAATTATGTTTAATG 893
    | || |||||||||||||||||||||||||||||||||| || | | |
839 AGATCACACTGGAGACGAGAATGTATATTGATGTATAATTGTGATGATTA 888

894 ATTAATAAAGTGTGATTATTGAATTTGATTAATTGTACTCTGCGAAGCGA 943
    || |||||||  ||  || |    | |  ||      ||||| |||||||
889 ATGAATAAAGAGTTGTTTT....TATTCTTTGAATTACTCCGCGAAGCGG 934

944 TATGTCTTATGTTTTGTGGAGGCATGTGAGTGACGTCATATGTCTCCGTG 993
    | |||  | | |||  ||||| |||     ||||||||||||||||||| |
935 TGTGTTATGTTTTTGTTGGAGACAT....ATGACGTCATATGTCTCCGCG 980
        CR→
994 CTGACGTAGGGGCGGGGCT 1012
       | |  ||  |||||||
981 .ACAGGCTGGCACGGGGCT 998
```

Figure 5.8

Alignment of F and JL3 component 5 translated protein sequences

The translated protein sequences of the major ORF of component 5 of the F and JL3 isolates were aligned using GAP. 'I' indicates identical residues, ':' indicates residues whose comparison value is greater than or equal to 0.5, and '.' indicates residues whose comparison value is greater than or equal to 0.1 (Genetics Computer Group, 1991).

```
                .         .         .         .         .
  1 MAQLRWGRKGVRSQRRKYSRPVAYKPPSSKVVSHVETVLNKKDVTGAEMK 50
    |..:|||||||:|||||||||| :|||||||||||||||.||||:|||||||:|
  1 MVAVRWGRKGLRSQRRKYSR.IAYKPPSSKVVSHVESVLNKRDVTGAEVK 49

               .         .         .         .         .
 51 PFNDGSRYSMKKIMVSATLTMAPGELLNYLIVKSNSPIANWSSSFSNPSL 100
    ||.||||||||||:|: ||||||||||:||||||||||||||||||||||||
 50 PFADGSRYSMKKVMLIATLTMAPGELVNYLIVKSNSPIANWSSSFSNPSL 99

               .         .         .         .         .
101 MVKESVQDTVTILRRGKLESSGTAGKDVTKSFSRFVNLGLGISQTQHLYL 150
    ||||||||||||:  ||||||||||||||||||.:||.|| ||||||||||
100 MVKESVQDTVTIVGGGKLESSGTAGKDVTKSFRKFVKLGSGISQTQHLYL 149

                    .         .
151 IIYSSDAMKITLETRMYIDV 170
    ||||||||||||||||||||
150 IIYSSDAMKITLETRMYIDV 169
```

The presence of the F or JL3 strains of component 5, however, did not correlate with any obvious differences in symptoms. The lack of correlation may be due in part to co-infections of the two strains (see below) and to the large number of sequence variations between isolates of each strain in different plants. Most of these variations were in the noncoding region. The coding region of the F strain was not included in the PCR fragments amplified from the other isolates. There were three positions in the JL3 strain at which the variations found in different isolates affected the protein sequence; these were at nucleotides 564, 712 and 786 and the corresponding amino acid changes were Ser→Ala, Arg→Thr, and Ser→Leu (Figure 5.6). The deletion present in the noncoding region of the F strain isolates A1 and A2 (Figure 5.5) did not correlate with a change in symptom expression compared to F, as A1 produces severe symptoms whereas A2 produces mild symptoms. The sample size of this study may have been too small to observe any significant correlations between sequence differences and symptom variation. Alternatively, the SCSV coat protein may not be involved in pathogenesis. Deletions of the coat proteins do not affect the infectivity or symptom expression of the bipartite geminiviruses ACMV (Stanley and Townsend, 1986) and ToLCV (tomato leaf curl virus India strain; Padidam *et al.*, 1995a). For other bipartite geminviruses, such as TGMV and BGMV, the coat protein is also not required for infectivity, but the symptoms of viral coat protein mutants are delayed and attenuated (Brough *et al.*, 1988; Gardiner *et al.*, 1988; Azzam *et al.*, 1994). To determine if these variations affect symptom expression, an infectious clone system is required which would enable mutagenesis to be carried out, as was done for the geminiviruses. Studies of naturally infected plants would not be feasible as

they often appear to contain mixtures of component strains as well as a population of sequence variants.

From both the PCR and sequencing, isolates F, F2, A1, A2, C, E, B2, JL9 and JL12 were found to contain F strain component 5 sequences. From just the PCR, isolates JL11 and JL13 also appeared to contain F strain sequences. Sequencing revealed JL3 strain component 5 sequences in isolates D, B1, JL3, and JL9 (although the sequence from JL9 may have been a contaminant), and PCR indicated isolates JL11 and JL12 also contained JL3 strain component 5 sequences (Table 5.2). The strain of component 5 present in isolate JL4 was not determined. The F and JL3 strains of component 5 appear to co-exist in isolates JL9, JL11 and JL12.

As was found previously with the glasshouse maintained isolates (Boevink, 1991), the field isolates could be differentiated by PCR with different primer sets. Some reactions amplified smaller amounts of product than others on the same sample, for example, the 5Hind+5Pst PCR on JL9 and the SUNI5+5-40mer PCR on JL13. This may have been due to sequence mismatches in one of the primer binding sites (there was one mismatch in the binding site of 5Pst in JL9) or different reaction requirements for different primer sets.

In hindsight, the assumption that the 5Hind primer would bind to all strains of component 5 because its site was in the coding region was not entirely justified. An attempt at sequencing JL3 with the primer 5EcoPI, which also binds in the coding region, was unsuccessful and it was later evident that there were mismatches at the 3' end of the primer which disrupted binding. This could also have been the case for 5Hind.

Within the F and JL3 strains, the component 5 sequence variation between different isolates was less than 10% (the small size of some of the fragments distorted some of the values). As well as nucleotide substitutions, insertions and deletions were found in sequences from both strains. The nucleotide sequences between members of the two strains differed by an average of 31.5%. This value largely reflects nucleotide differences within noncoding regions of the two strains. The complete component 5 nucleotide sequences of isolates F and JL3 showed 31.2% difference in the noncoding region and 12.2% difference in the coding region. In the comparison of the F and JL3 isolate sequences (Figure 5.7), blocks of conserved sequence could be seen in the noncoding region between the common region and the start of the ORF, for example, nucleotides 205-213, 230-244, and 306-321 (numbered as in the JL3 sequence). These conserved elements may constitute important regulatory signals and would therefore be obvious targets for mutational analysis of the component 5 promoter region.

No intermediates between the F and JL3 related component 5 sequence variants were found. This suggests that F and JL3 represent distinct strains of component 5 rather than extremes of a spectrum of variation. This is supported by the presence of greater number of differences in the common region of component 5 of the F and JL3 isolates, than between the different components of one isolate (and apparently one strain, F; Chapter 4).

If the F and JL3 component 5s are regarded as strains then the sequence difference between them (20% at the nucleotide level, and 12% at the amino acid level) is fairly high. Strains of potyviruses were found to have 1-10% difference between their coat protein sequences, and there was 25-45%

difference between the coat proteins of different species within a genus, although there were also sets of related species with around 20% difference (Ward *et al.*, 1994). Interspecies comparisons of the geminivirus replication proteins clustered around 20-34% difference, while comparisons of coat proteins revealed around 18-34% difference (Rybicki, 1994). Padidam *et al.* (1995b) proposed that geminivirus isolates with greater than 90% sequence identity should be considered the same species. Indeed, the analysis of 12 MSV strains found 10.9% and 2% difference at the nucleotide and amino acid levels respectively (Briddon *et al.*, 1994). The so-called strains of BGMV and TYLCV which fall outside of this range have been suggested to be distinct species (Gilbertson *et al.*, 1993; Padidam *et al.*, 1995b).

There were nearly twice as many differences between the 'common region' sequences of the F and JL3 isolate component 5s as between any of the common region containing F isolate components (Chapter 4). By analogy with the geminiviruses (Chapter 1), the common region is predicted to contain *cis* elements controlling replication of the components. The differences between F and JL3 could therefore lead to differences in the replication of the components. F and JL3 component 5 strains may be replicated specifically by different RAP strains. The replication of the Logan and CFH strains of BCTV is strain specific, that is, the replication protein of Logan will not *trans*-replicate CFH and vice versa (Choi and Stenger, 1995). This lead Choi and Stenger (1995) to suggest that Logan and CFH be regarded as different species, since the definition of a species usually involves replicative compatibility. The intergenic region sequences of Logan and CFH have diverged (around 35% difference) (Choi and Stenger, 1995). The virion sense genes of CFH and

Logan are highly conserved, with 0-5% amino acid difference, except for a unique 24 amino acid extension to R2 in CFH, but the complementary sense genes have diverged (14-43% amino acid difference). A similar situation of replicative incompatibility exists between the Sardinian and Israeli strains of TYLCV (Jupin *et al.*, 1995), and the Brazillian (BGMV-Bra) and other strains of BGMV (Gilbertson *et al.*, 1993; Faria *et al.*, 1994). By the criteria of Padidam *et al.* (1995b), however, the Sardinian and Israeli strains of TYLCV are distinct species, and BGMV-Bra is a different species to the other BGMV strains.

In SCSV there was no correlation between the presence of F or JL3 component 5 strains and the RAP components 2 and 6 (see Chapter 3). However, strain specific replication could be carried out by different strains of one of the RAPs, which may be too similar to be differentiated by the primers used. If they do occur, differences in replication of the two strains may explain the co-existence of the two strains in one plant because they would not be competing for replication proteins. More sequence information and studies on replication between different SCSV strains are required before any conclusions can be made as to what the F and JL3 component 5 sequences represent.

The discovery of two distinct strains of component 5 leads to the question of whether other components also occur as strains, and whether any strains of other components are found together with the JL3 strain of component 5. There may be a complete set of JL3 strain components distinguished by highly conserved common regions, like the components of the F isolate presented in Chapter 4. This might be expected if the different common region sequences did lead to strain specific replication as suggested above. Alternatively, the isolation of a set of apparently closely related

components, which make up the F isolate, may have been serendipitous, or a consequence of glasshouse conditions. Natural infections may be a complex mixture of different components, and different strains of those components.

Different viral strains are not usually found in the same plant. Indeed, mild strains of viruses, such as tomato mosaic virus, citrus tristeza virus, and papaya ringspot virus, have been used to cross-protect plants against more destructive viral strains (reviewed in Urban *et al.*, 1990). The existence of such different SCSV component strains, and the ability of them to co-infect plants, may have serious consequences for the use of cross-protection and the development of resistant plants using viral sequences (pathogen derived resistance, PDR; Sanford and Johnstone, 1985). Both cross-protection and PDR are often specific to closely related strains. Tobacco plants transformed with the nuclear inclusion b (NIb) gene of potato virus Y strain O (PVY-O) were resistant to infection by PVY-O but not resistant to infection by the closely related strain PVY-N (93.4% amino acid identity between the NIb proteins; Audy *et al.*, 1994). Transgenic tobacco expressing the TMV coat protein from the type strain were well protected against viruses with at least 85% amino acid homology and only moderately protected against viruses with 60-85% homology (Nejidat and Beachy, 1990). Studies of transgenic tobacco containing the plum pox virus coat protein gene and the tomato spotted wilt nucleocapsid protein gene indicated that whether PDR is effective against only closely related strains or also against related viruses may depend on whether the protective mechanism involves the RNA transcribed from the viral gene and/or the viral protein (Ravelonandro *et al.*, 1993; Pang *et al.*, 1994). If all of the essential SCSV genome components exist as strains as dissimilar as F and

JL3, PDR may be easily overcome by SCSV under field conditions. Wild legume species which serve as over-wintering hosts of the SCSV vector *A. craccivora*, such as *Medicago hispida* (Johnson, 1957), are thought to provide a reservoir of SCSV from which subterranean clover fields are newly infected upon aphid migration (Grylls and Butler, 1959). They may also be a reservoir of SCSV strains which could overcome any introduced resistance genes.

CHAPTER 6: AGROINFECTION

6.1 Introduction

Many plant viruses can successfully infect their hosts when either their purified particles or genomic nucleic acid are rubbed onto a leaf assisted by an abrasive material such as carborundum. SCSV is not mechanically transmissible, neither as DNA or purified particles, nor can aphids transmit the virus when they are fed purified particles (Grylls and Butler, 1959; Chu and Helms, 1988). These limitations make it impossible to prove that SCSV causes the stunt disease by normal methods in fulfilment of the requirements of Koch's postulates (Agrios, 1988). Furthermore, studies of the functions of the genome components are hampered by the inability to manipulate the types of components present, or the sequences of the components, *in vivo*.

Most geminiviruses are similarly limited by not being mechanically transmissible (Rybicki, 1994). The technique which was developed to overcome the problem is agroinfection (Grimsley *et al.*, 1986; Elmer *et al.*, 1988b; Hayes *et al.*, 1988; Morris *et al.*, 1988). The mechanism of agroinfection is briefly described in Chapter 1 and Figure 1.5. The copies of the geminiviral DNA dimers, cloned between the border sequences of the T-DNA are transferred into the plant cell as ssDNA attached to the VirD2 protein of the *Agrobacterium* (Tinland *et al.*, 1994). Once in the nucleus of the plant cell, the ssDNA is rapidly converted to dsDNA (Rodenburg *et al.*, 1989). Transcription from this dsDNA must then occur to produce the RAP. This enables the release of ssDNA monomers of the geminiviral DNA by a replicative mechanism similar to rolling-circle replication which requires that the origin of replication

(*ori*) sequences are repeated (Stenger *et al.*, 1991). The repetition of the *ori* sequences presumably mimics the circular DNA normally encountered by the RAP. It can be inferred that the *ori* region also contains sequences which control the cleavage and ligation into circular form, of the unit length DNAs. The released circular monomers would again be converted to dsDNA allowing further transcription and rolling-circle replication, and the development of a normal viral life cycle. The agrobacteria therefore act as alternative vectors for the delivery of the viral DNA into the nucleus of the plant cell.

This Chapter presents an attempt at artificially infecting subterranean clover and pea plants with SCSV using an agroinfection system like those used for the geminiviruses. Infectious geminivirus clones used for agroinfection are dimeric constructs of viral DNA components in which the *ori* sequences are repeated. If SCSV uses rolling-circle replication, as was suggested by the similarities between the genome structure and RAPs of SCSV and the geminiviruses (Chapter 3), then repetition of the putative ori sequences around the stem-loop of the SCSV DNAs in the agroinfected constructs is probably required for efficient release of viral DNA monomers. To meet this requirement, dimeric constructs of all of the SCSV components were made.

6.2 Results

6.2.1 Construction of dimeric clones

The circular nature of the SCSV components allowed dimeric clones to be constructed by inserting one full length clone into another full length clone, which had been generated using a different restriction endonuclease (Figure 6.1). The first full length clone, therefore, had an internal site for the enzyme

used to create the full length clone which was inserted. Dimers of components 1, 2, 3, 4, 5, and 6 were created in this manner.

After attempts to create a dimer of component 7 in the same way as the other components proved unsuccessful, a different approach was taken (Figure 6.2). One full length and one nearly full length copy of component 7 were created by PCR with the primers 7dZA + 7dY'ext, and 7dXext + 7dA'ext (Table 6.1). One *Hind*III site in each fragment was maintained such that the when the PCR products were digested with *Hind*III and ligated together a tandem dimer was created. The ligated fragment was digested with *Eco*RI and *Bam*HI, whose sites had been engineered into the primers 7dY'ext and 7dXext, and cloned into pGEM3Z (Promega).

The dimeric constructs, present in pGEM3Z, were inserted into the binary vector pGA470 (An *et al.*, 1985) at the *Eco*RI site and transformed into *A. tumefaciens* strain AGL1.

6.2.2 Construction of double dimers

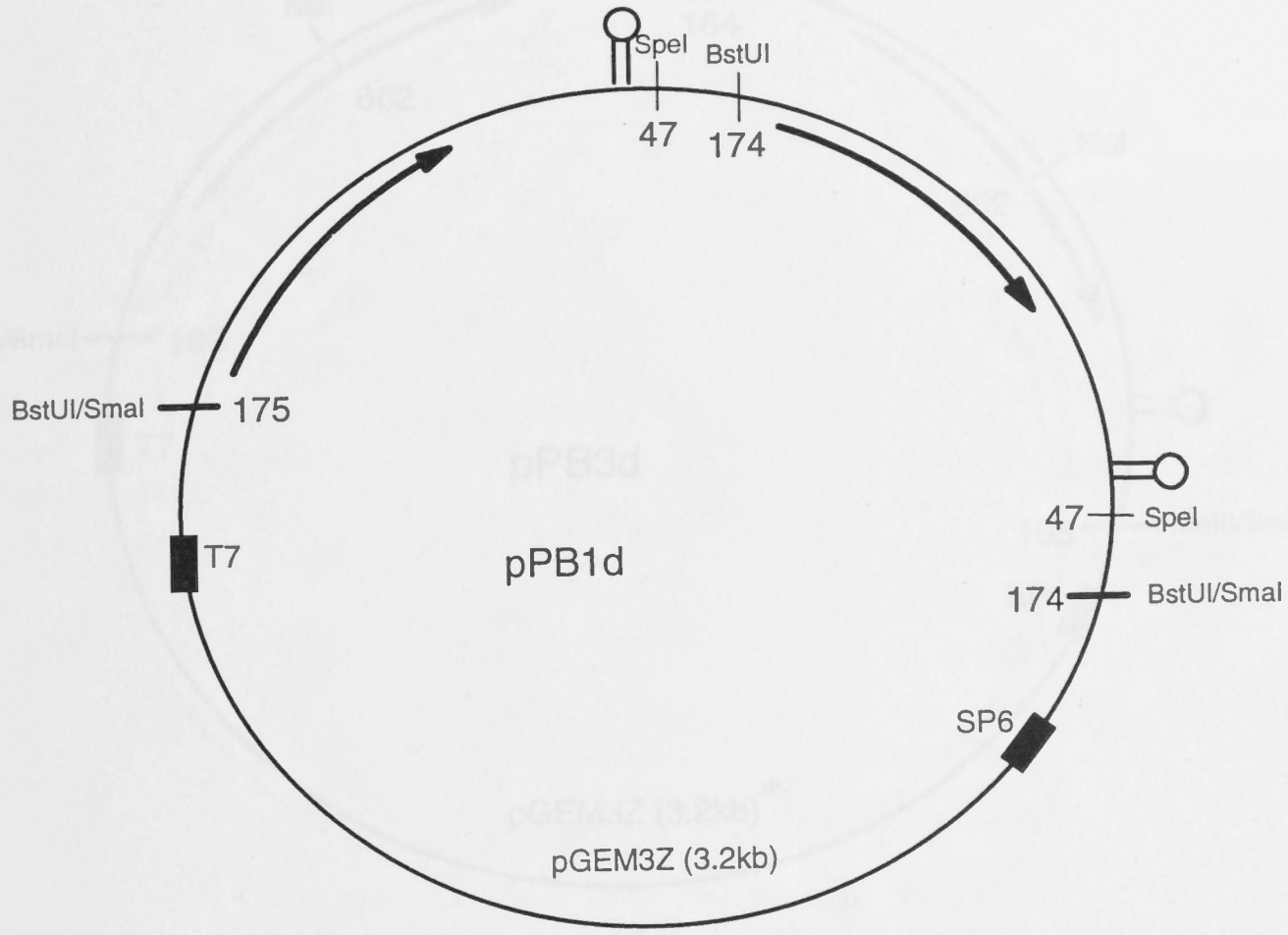
To reduce the number of agrobacteria required to infect a single cell, the dimeric clones were combined to make three double dimers (Figure 6.3 and 6.4). The double dimers combined components 2 with 4, 1 with 7, and 5 with 6. Figure 6.3 shows the creation of the double dimer containing components 2 and 4 in pGA470. The double dimer of components 1 and 7 was created in the same way. Figure 6.4 shows the construction of the double dimer containing components 5 and 6 in pJITΔHind. The vector pJITΔHind was created from pJIT119 (Guerineau *et al.*, 1990) by digestion with *Hind*III, religation, and

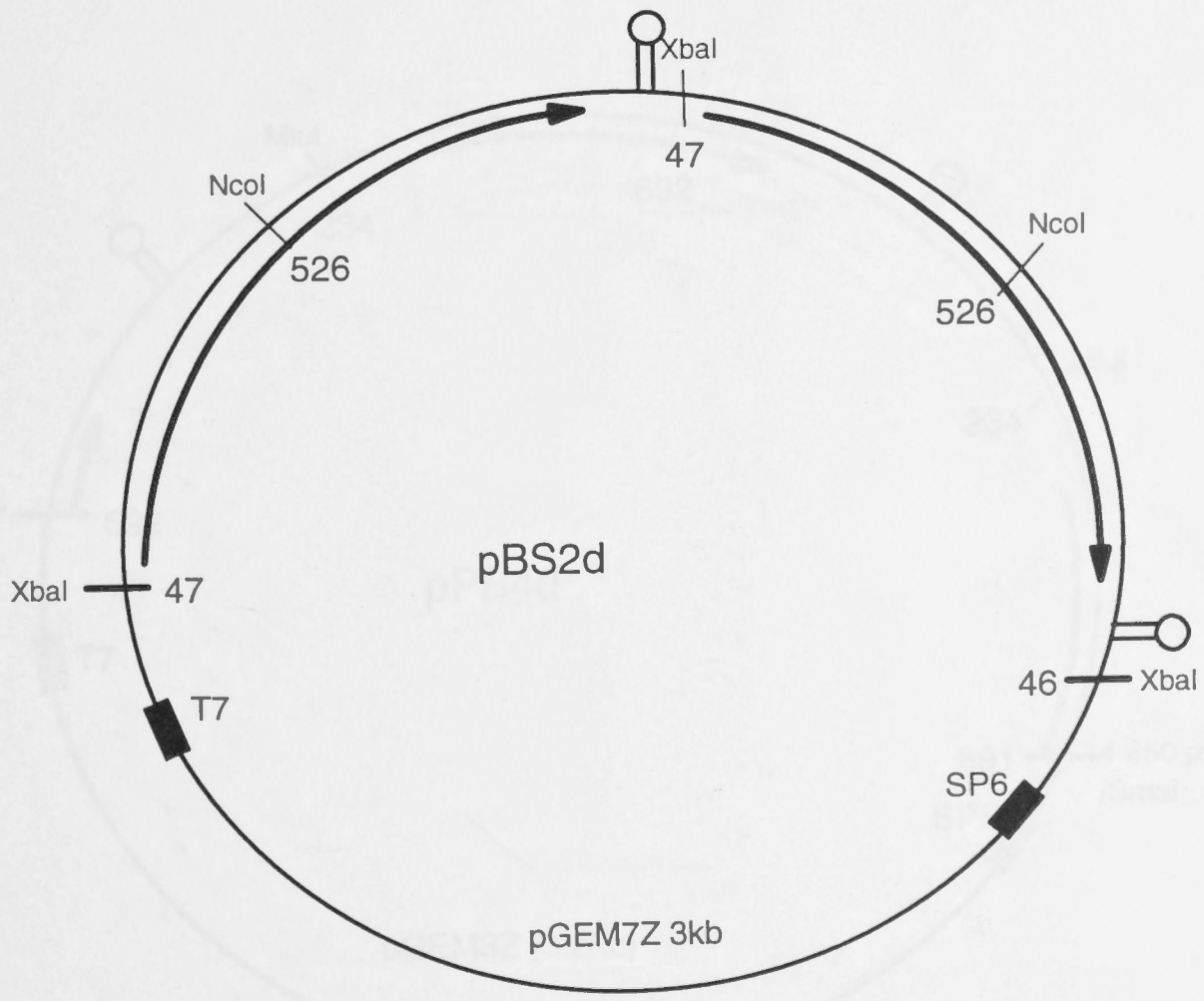
Figure 6.1

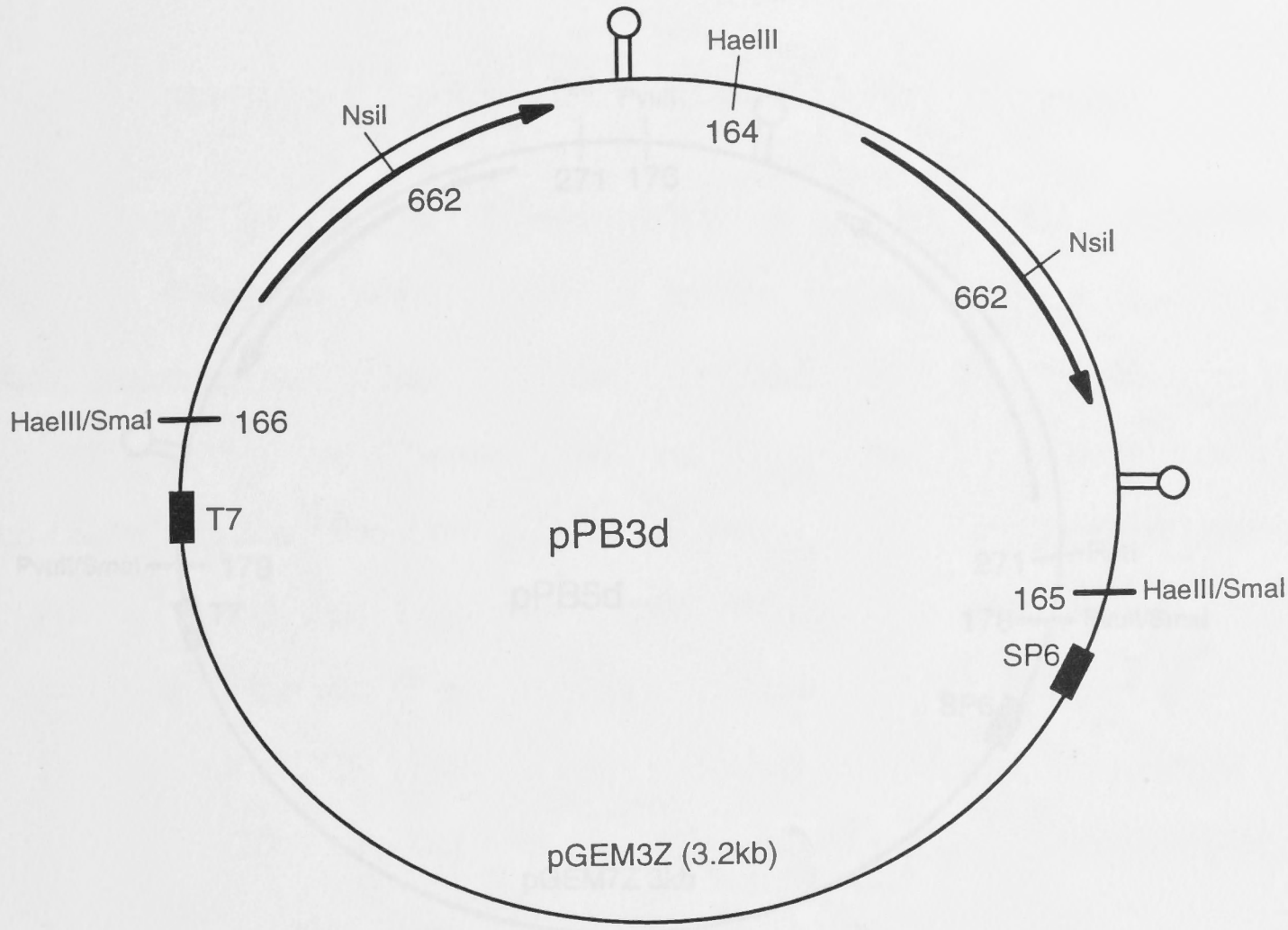
Construction of dimeric clones of components 1 to 6

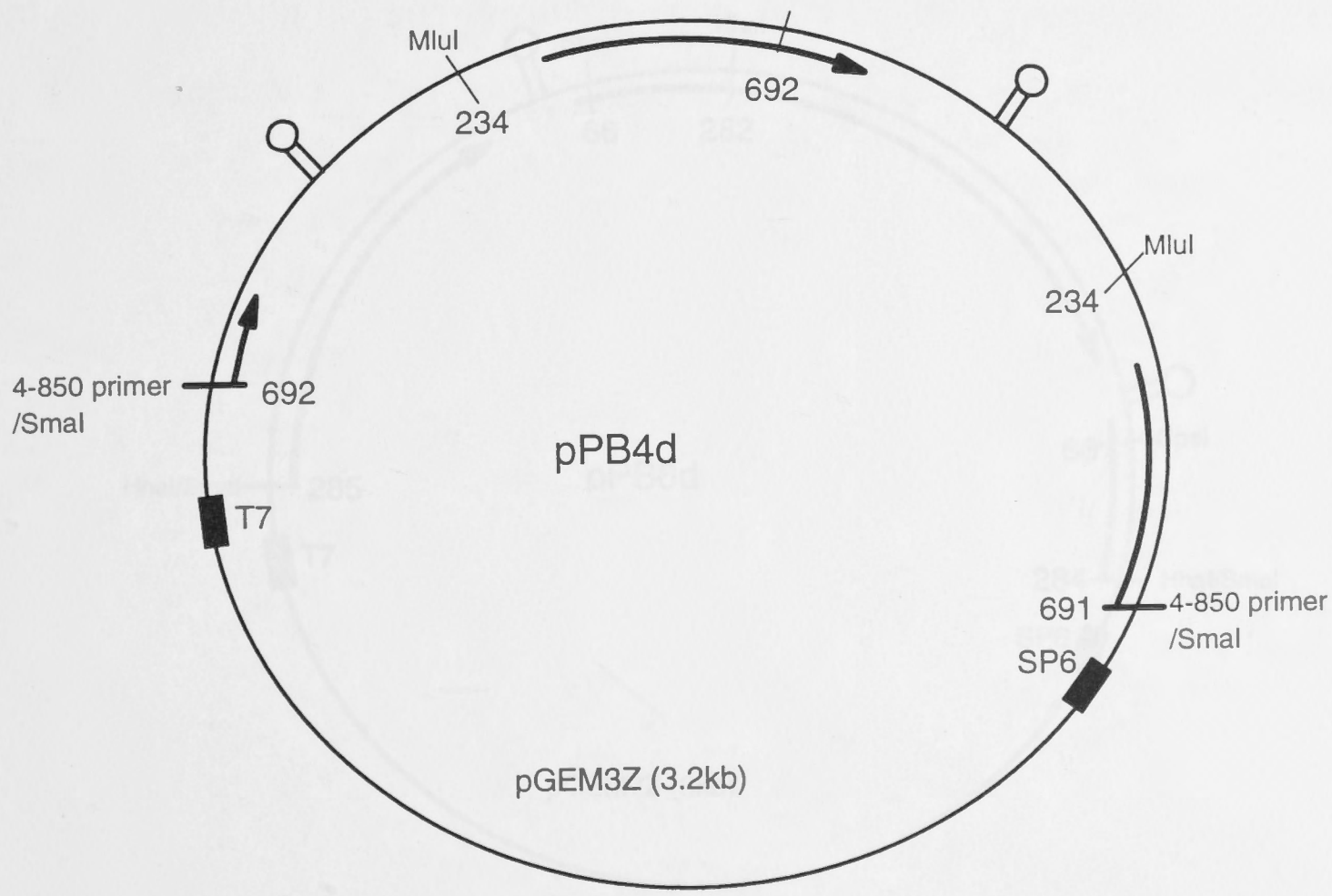
The component 1 dimer (pPB1d) was created by inserting a full length *Spe*I fragment into a full length *Bst*UI clone of component 1. The component 2 dimer (pBS2d) was created by Brian Surin by inserting a full length *Nco*I fragment into a full length *Xba*I clone of component 2. The component 3 dimer (pPB3d) was created by inserting a full length *Nsi*I fragment into a full length *Hae*III clone of component 3. The component 4 dimer (pPB4d) was created by inserting a full length *Mlu*I fragment into a full length clone of component 4 created by PCR with the 4-850 primers (Table 4.1). The component 5 dimer (pPB5d) was created by inserting a full length *Pst*I fragment into a full length *Pvu*II clone of component 5. The component 6 dimer (pPB6d) was created by inserting a full length *Spe*I fragment into a full length *Hha*I clone of component 6.

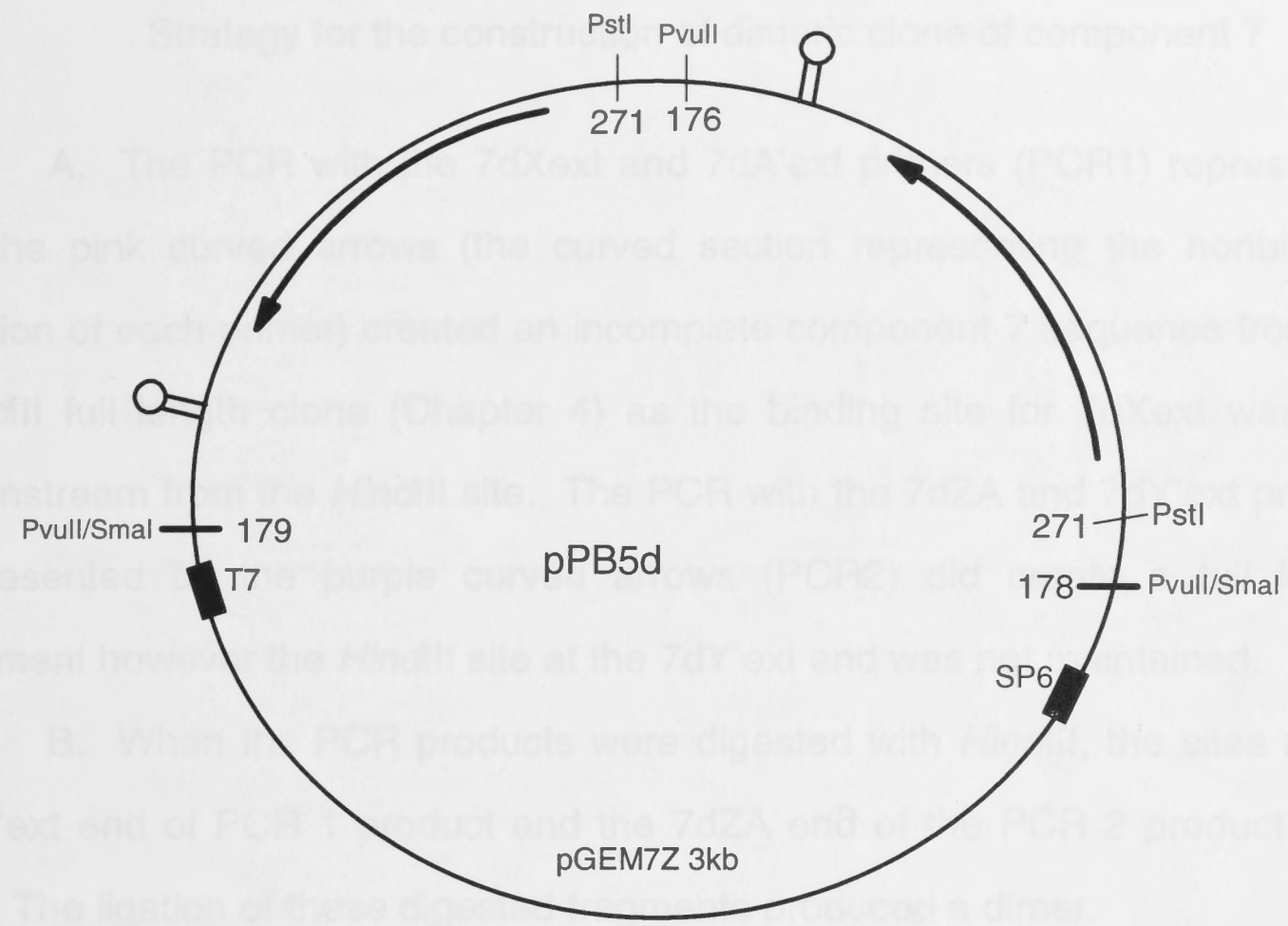
The constructs are not drawn with the vector sequences to scale, however, the positions of the features of the SCSV sequences are proportional. The stem-loop sequences are represented by ball-and-stick symbols, the ORFs by arrows, and the T7 and SP6 promoter sequences flanking the polycloning sites of the pGEM vectors by black rectangles. The positions of restriction sites are as in the sequences in Figures 3.1 and 4.1, with nucleotide 1 in the conserved loop sequence.


SpeI
BstUI
47
174
BstUI/SmaI
175
T7
pPB1d
47
SpeI
174
BstUI/SmaI
SP6
pGEM3Z (3.2kb)


XbaI
47
NcoI
526
NcoI
526
pBS2d
XbaI
47
T7
46
XbaI
SP6
pGEM7Z 3kb


HaeIII
164
NsiI
662
NsiI
662
HaeIII/SmaI
166
T7
pPB3d
165
HaeIII/SmaI
SP6
pGEM3Z (3.2kb)


MluI
234
692
MluI
234
4-850 primer
/SmaI
692
T7
pPB4d
691
4-850 primer
/SmaI
SP6
pGEM3Z (3.2kb)


PstI
PvuII
271
176
PvuII/SmaI
179
T7
pPB5d
271
PstI
178
PvuII/SmaI
SP6
pGEM7Z 3kb

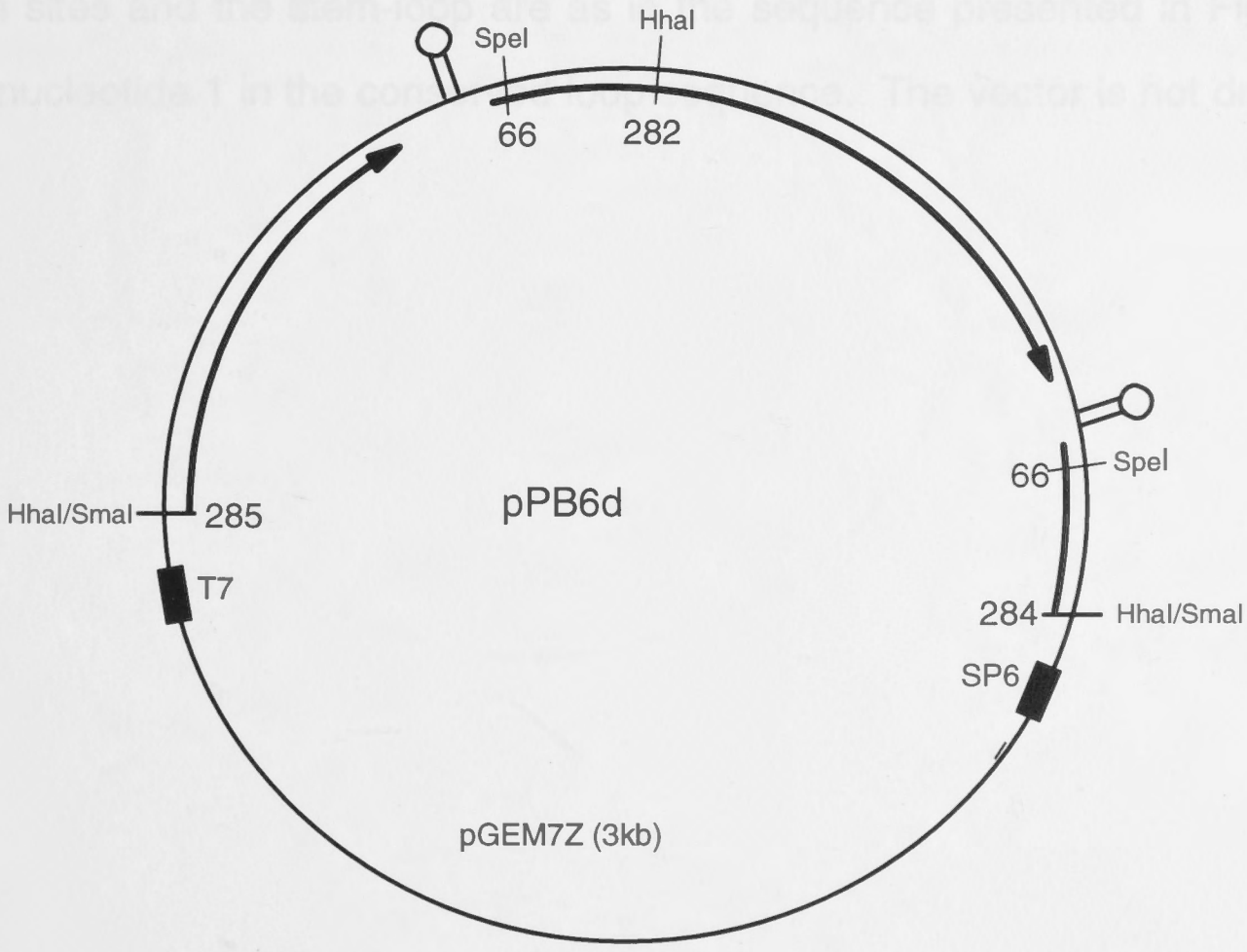

SpeI
HhaI
66
282
HhaI/SmaI
285
T7
pPB6d
66
SpeI
284
HhaI/SmaI
SP6
pGEM7Z (3kb)

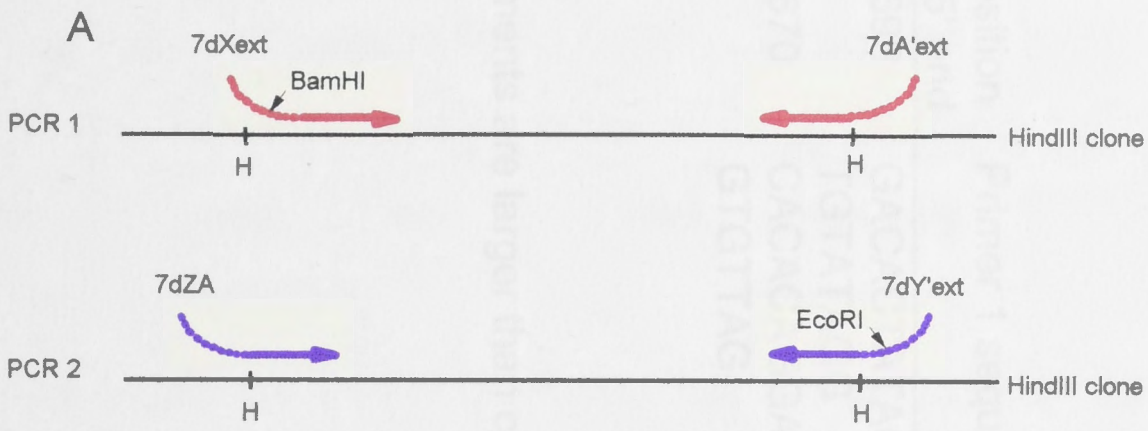
Figure 6.2

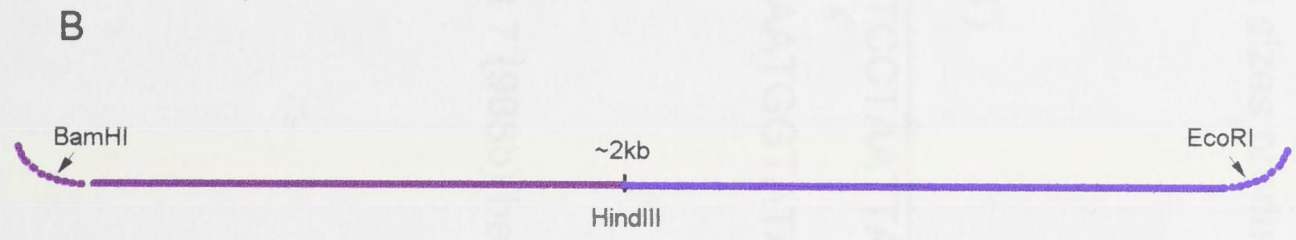
Strategy for the construction of dimeric clone of component 7

A. The PCR with the 7dXext and 7dA'ext primers (PCR1) represented by the pink curved arrows (the curved section representing the nonbinding section of each primer) created an incomplete component 7 sequence from the *Hind*III full length clone (Chapter 4) as the binding site for 7dXext was 17b downstream from the *Hind*III site. The PCR with the 7dZA and 7dY'ext primers represented by the purple curved arrows (PCR2) did create a full length fragment however the *Hind*III site at the 7dY'ext end was not maintained.

B. When the PCR products were digested with *Hind*III, the sites at the 7dA'ext end of PCR 1 product and the 7dZA end of the PCR 2 product were cut. The ligation of these digested fragments produced a dimer.

C. This was digested with *Eco*RI and *Bam*HI, which cut at the engineered restriction sites in primers 7dXext and 7dY'ext, and cloned into pGEM3Z (Promega). The ORF (represented by the green arrows) is repeated but the second copy is disrupted by the cloning site. The repeated stem-loop sequence is represented by the ball-and-stick symbol. The positions of the restriction sites and the stem-loop are as in the sequence presented in Figure 4.1, with nucleotide 1 in the conserved loop sequence. The vector is not drawn to scale.


A
7dXext
BamHI
7dA'ext
PCR 1
H
H
HindIII clone
7dZA
7dY'ext
EcoRI
PCR 2
H
H
HindIII clone


B
BamHI
~2kb
EcoRI
HindIII

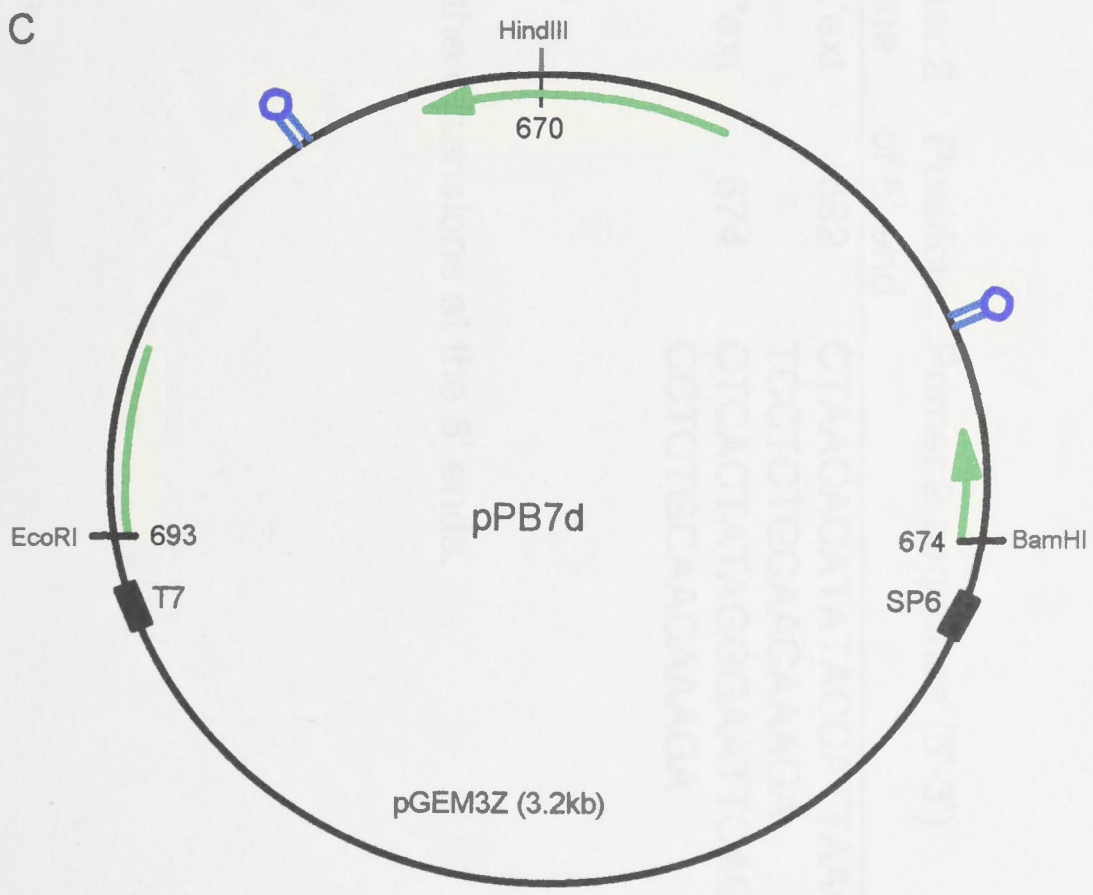

C
HindIII
670
pPB7d
EcoRI
693
T7
674
BamHI
SP6
pGEM3Z (3.2kb)

Table 6.1

Sequences and positions of and fragment sizes produced by PCR primers to component 7, used to create the component 7 dimer.

Primer 1 name	Position of 5' end	Primer 1 sequence (5'-3')	Primer 2 name	Position of 5' end	Primer 2 sequence (5'-3')	Fragment size (bp)
7dXext	693	GACACTATAGAAGGATCCTAATTAC TGTATAGG	7dA'ext	692	CTAACACATATACCATTAAGCT TCCTCTGCAACAAAGA	1006[a]
7dZA	670	CACACAGGAAAGCTTAATGGTATAT GTGTTAG	7dY'ext	674	CTCACTATAGGGAATTCAGCTT CCTCTGCAACAAAGA	1016[a]

[a] the PCR fragments are larger than component 7 (988b) because of the extensions at the 5' ends.

Figure 6.3

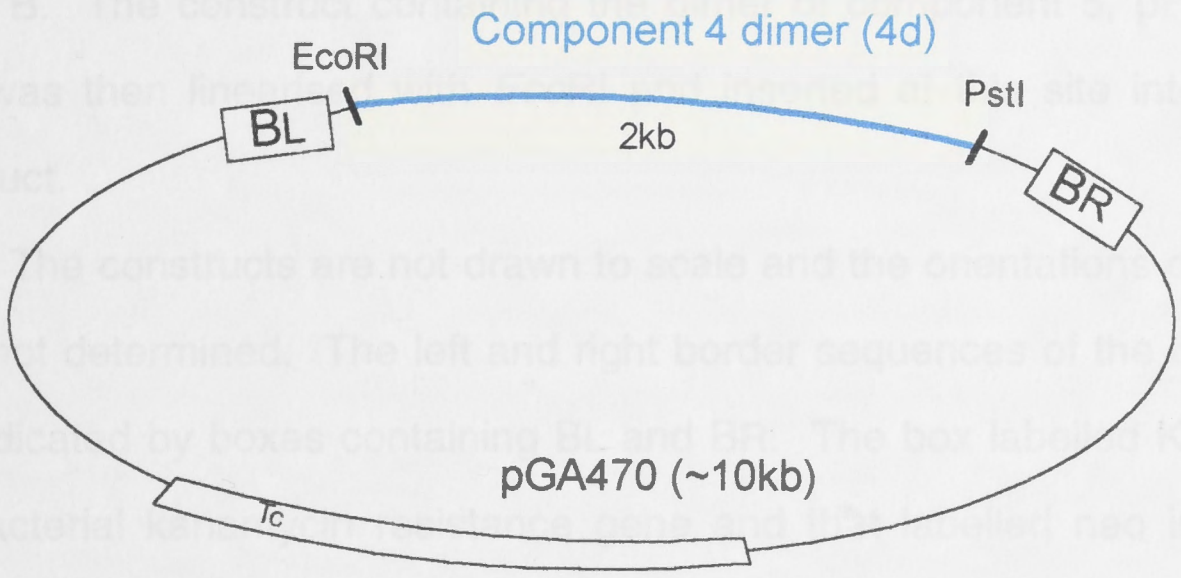
Construction of double dimer containing components 2 and 4

A. The dimer of component 4 was excised from pPB4d (Figure 6.1) with *Eco*RI and *Pst*I and inserted into pGA470 (An *et al.*, 1985) at these sites.

B. The construct containing the dimer of component 2, pBS2d (Figure 6.1), was then linearised with *Eco*RI and inserted at this site into the above construct.

The constructs are not drawn to scale and the orientations of the dimers were not determined. The left and right border sequences of the binary vector are indicated by boxes containing BL and BR. The box labelled Tc indicates the relative position of the tetracycline resistance gene of the vector.

A


Component 4 dimer (4d)
EcoRI
PstI
BL
2kb
BR
pGA470 (~10kb)
Tc

B

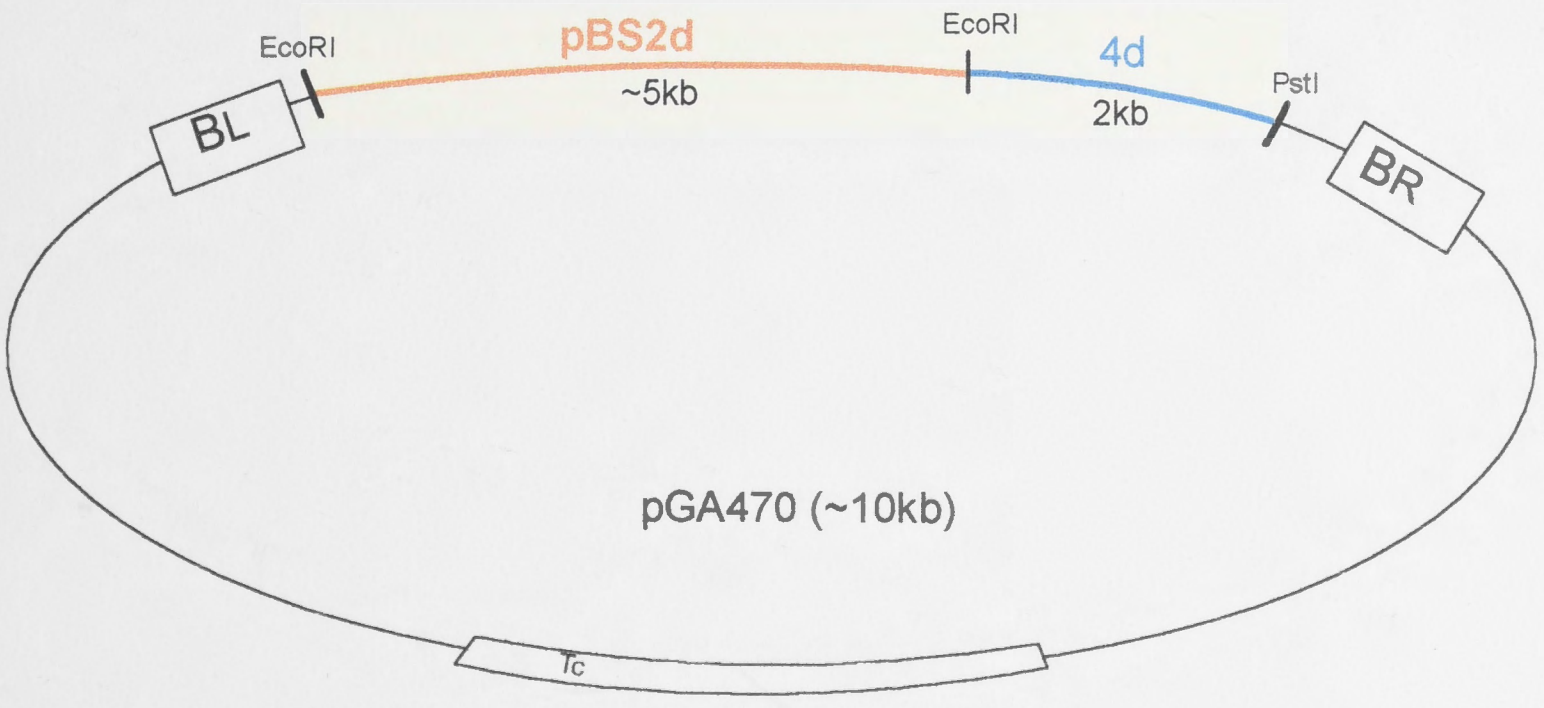

EcoRI
pBS2d
~5kb
EcoRI
4d
2kb
PstI
BL
BR
pGA470 (~10kb)
Tc

Figure 6.4

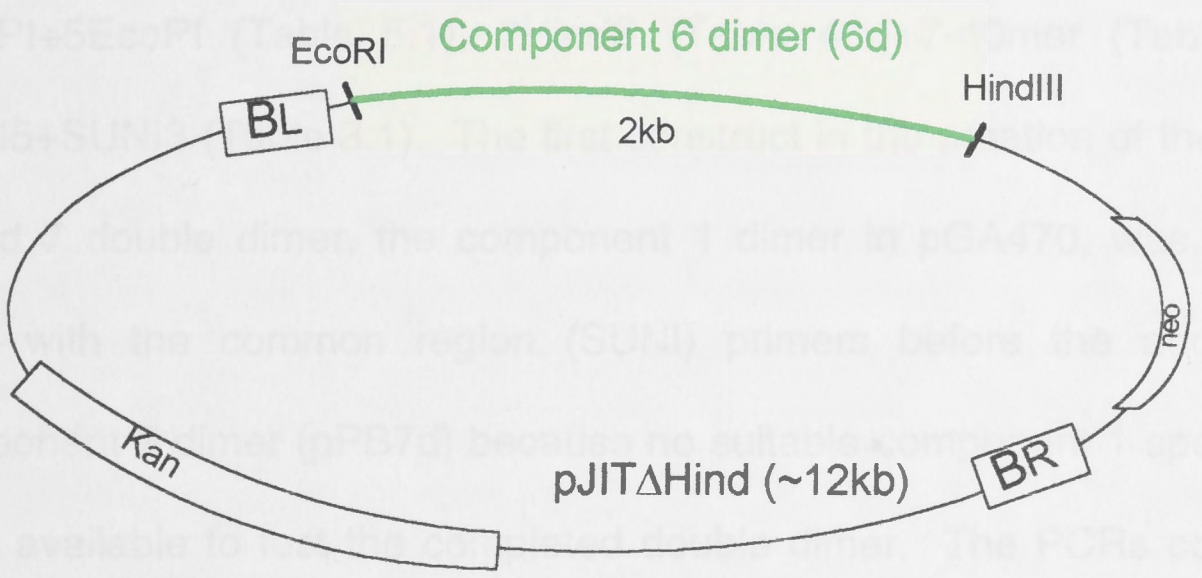
Construction of double dimer containing components 5 and 6

A. The dimer of component 6 was excised from pPB6d (Figure 6.1) with *Eco*RI and *Hind*III and inserted into pJITΔHind at these sites.

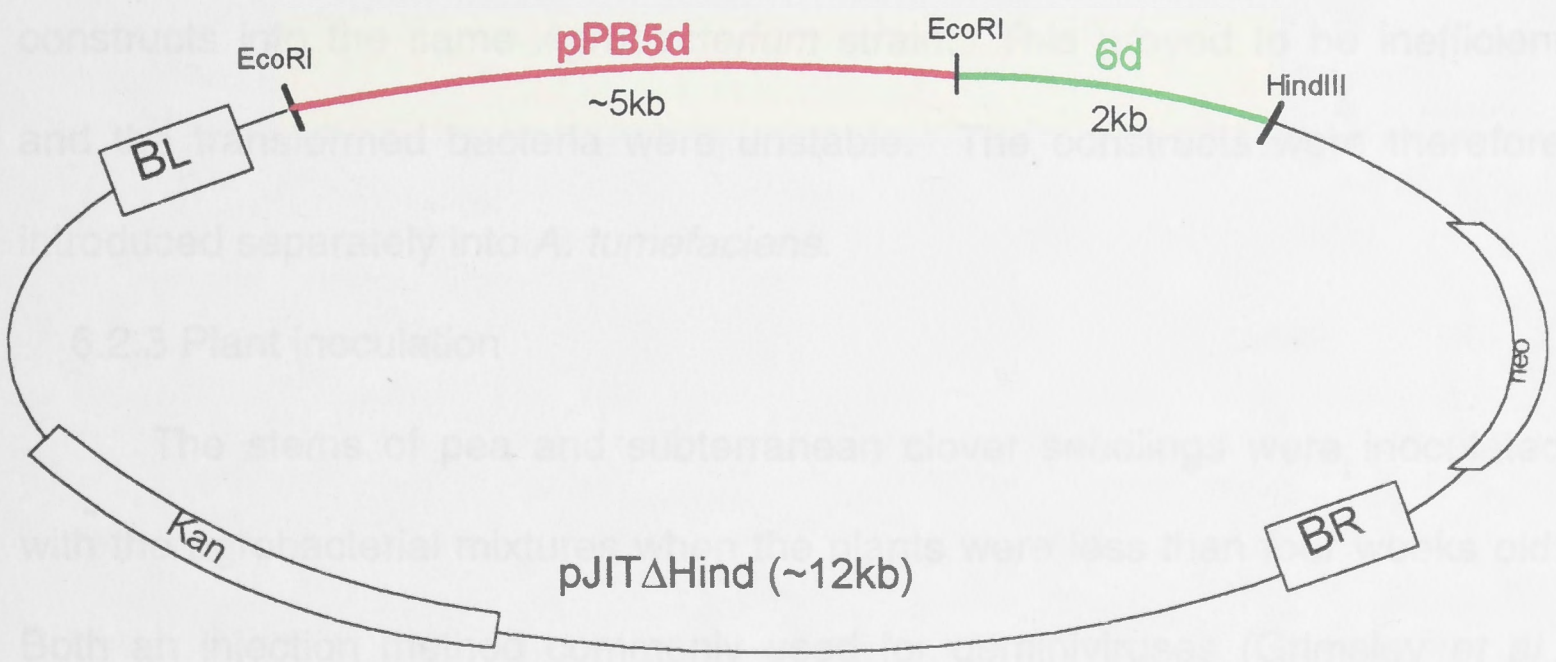
B. The construct containing the dimer of component 5, pPB5d (Figure 6.1), was then linearised with *Eco*RI and inserted at this site into the above construct.

The constructs are not drawn to scale and the orientations of the dimers were not determined. The left and right border sequences of the binary vector are indicated by boxes containing BL and BR. The box labelled Kan indicates the bacterial kanamycin resistance gene and that labelled neo indicates the plant kanamycin resistance gene.

A


EcoRI
Component 6 dimer (6d)
HindIII
BL
2kb
neo
Kan
pJITΔHind (~12kb)
BR

B


EcoRI
pPB5d
~5kb
EcoRI
6d
2kb
HindIII
BL
neo
Kan
pJITΔHind (~12kb)
BR

selection of the plasmids without the *Hind*III fragment. The component 3 dimer (pPB3d; Figure 6.1) was also cloned into pJITΔHind at the *Eco*RI site.

The double dimer constructs were analysed by PCR with primers specific to each of the components (results not shown). The primer sets used were: 4-850S+4-850C (Table 4.1), Bam2+Bgl2 (Table 3.1), Spe6+Nsi6 (Table 3.1), 5HinPI+5EcoPI (Table 5.1), 7HindS (Table 4.1)+7-40mer (Table 2.1), and SUNI5+SUNI3 (Table 3.1). The first construct in the creation of the component 1 and 7 double dimer, the component 1 dimer in pGA470, was analysed by PCR with the common region (SUNI) primers before the addition of the component 7 dimer (pPB7d) because no suitable component 1 specific primers were available to test the completed double dimer. The PCRs confirmed that both components were present in each construct and that no obvious deletions had occurred as each PCR gave the expected fragment sizes.

The binary vectors, pGA470 or pJITΔHind, have resistance genes to tetracycline and kanamycin, respectively. They were used with a view to further reducing the required number of agrobacteria by introducing two different binary constructs into the same *Agrobacterium* strain. This proved to be inefficient and the transformed bacteria were unstable. The constructs were therefore introduced separately into *A. tumefaciens*.

6.2.3 Plant inoculation

The stems of pea and subterranean clover seedlings were inoculated with the agrobacterial mixtures when the plants were less than four weeks old. Both an injection method commonly used for geminiviruses (Grimsley *et al.*, 1987; Morris *et al.*, 1988; Briddon *et al.*, 1989) and a vacuum infiltration method adapted from Bechtold *et al.* (1993) were tried. No plants with symptoms

typical of SCSV infection were obtained, and SCSV DNA could not be detected by PCR on plant extracts.

6.3 Discussion

This attempt at the development of an agroinfection system for SCSV was not successful. No infected plants were obtained after inoculations with mixtures of either; seven agrobacterial strains, each containing a different SCSV component dimeric construct, or three agrobacterial strains carrying the double dimers and the component 3 dimer, or combinations of the single and double dimer-carrying strains. There are several possibilities as to why this attempt was not successful.

SCSV may not be as amenable to agroinfection as the geminiviruses because of its larger number of genome components. For a systemic infection to develop, all of the SCSV DNAs would probably have to be introduced into the same plant cell. In this system there may have been too many agrobacteria required to infect a single plant cell to deliver all seven components. The creation of double dimers was an attempt to reduce this number; however, it was only possible to reduce the number to four. It is obvious from geminivirus work that having two different agrobacterial strains to deliver the required dimeric constructs does allow infection. The four strains required for SCSV, however, may have been too many. Further reduction of the number of agrobacteria was attempted by transforming them with both the pGA470 and pJITΔHind-based constructs, and selecting for resistance to both kanamycin and tetracycline. This was not practical as the transformants were not stable. Another possibility was to create constructs with three dimers. This idea was

not pursued as triple dimers were expected to be extremely unstable. Difficulties were experienced in creating some of the simple dimeric constructs. Furthermore, the double dimer containing components 1 and 7 was more difficult to create, and less stable than the double dimers of components 2 and 4, and 5 and 6, which did not combine repeated DNA in the form of two common region-containing components.

A general instability of the agrobacteria transformed with the binary constructs was also noticed. Glycerol stocks of agrobacteria stored at -70°C generally lasted only a couple of months and the plasmids were frequently lost. One reason for this may have been the instability of the tetracycline resistance. Several colleagues have noted an apparent spontaneous tetracycline resistance of the *Agrobacterium* strain AGL1. The pGA470 based constructs did appear to be more easily lost from transformed AGL1 than the pJIT119 constructs based on kanamycin selection. Before further attempts are made at agroinoculating SCSV, the component dimers should all be transferred to a kanamycin selection vector. Alternatively, a different agrobacterial strain should be used.

As mentioned above, difficulties were experienced in creating the initial dimeric constructs. There was no selectable marker to distinguish the dimers from monomers and numerous colonies had to be analysed before each dimer was found. This may have been due to a high recombination frequency of the dimers. If there was a high recombination frequency of the constructs, then there was a risk that they would recombine in the agrobacteria. Testing constructs in agrobacteria by the miniprep method (Methods 2.3.2) seldom produced easily digested DNA, and only a rough estimate of the size of the

cloned fragment was possible. Construction of partial dimers of the SCSV DNAs, such as 1.8mers, may be a way to overcome some of the instability apparent in the dimeric constructs. Partial dimers are generally used for the geminiviruses, and efficient agroinfection appears to require that only the ori sequences are repeated (Stenger *et al.*, 1991).

Further complications may be spontaneous mutations in the sequences of the cloned components causing a loss of infectivity, or the cloning of non-infectious variants of any of the components from the original source material. Limited sequencing of some of the dimeric constructs did not detect any mutations. Finally, all of the essential components of the SCSV F isolate may not have been isolated (Chapter 4), and therefore some essential components could have been missing in the constructs prepared for agroinfection.

Two inoculation techniques were used, injection and vacuum infiltration. As no positive control, such as a set of geminivirus constructs, was available, it was not determined which was the better technique. The injection method was not well suited to subterranean clover because the stems are thin and the meristems are difficult to locate. Neither of these delivery systems may be appropriate for SCSV if they do not allow the agrobacteria to come in contact with sufficient phloem cells. As SCSV appears to be strictly phloem limited it may be unable to move into the phloem from surrounding cells into which the SCSV DNA is delivered by the agrobacteria. Phloem-limited geminiviruses, such as AbMV (Horns and Jeske, 1991), TYLCV (Cherif and Russo, 1983) and WDV (Tomenius and Oxfelt, 1981), however, have been successfully agroinoculated (Evans and Jeske, 1993; Kheyr-Pour *et al.*, 1991; Navot *et al.*, 1991; Hayes *et al.*, 1988; Woolston *et al.*, 1988). A method in which the insect

vector is fed on the agrobacteria containing the DNA constructs, which was used for TYLCV (Zeidan and Czosnek, 1994), could perhaps be used to increase the number of phloem cells receiving DNA as the vector should inject large numbers of agrobacteria directly into the phloem. Another inoculation method reported for the geminivirus BGMV (Gilbertson *et al.*, 1991) used a particle gun, with the particles coated in linearised, double-stranded copies of the DNAs. This would not be as specific as the insect method in delivering the DNA into the phloem tissue.

Future development of this system should include; transfer of the dimeric constructs into binary vectors which encode a more stable resistance gene than that to tetracycline, for example, the kanamycin resistance gene, more testing of the vacuum infiltration method, and testing of other methods such as feeding aphids on agrobacteria containing the SCSV constructs.

* Whereas the GUS coding and 3' non-coding regions were common to all constructs, the 5' non-coding region and transcription control elements were specific for each SCSV component. Therefore, variation in gene expression patterns due to different SCSV non-coding sequences may involve differences in expression of both cell type and amount, mRNA stability and efficiency of translation initiation. The term, promoter, will be used more generally in this chapter to cover all of these aspects as measured by GUS activity. As some variation of GUS expression is also likely to arise due to transgene position or gene silencing effects, several independent transformants were examined for all constructs.

CHAPTER 7: STUDY OF SCSV NON-CODING REGIONS ON GENE EXPRESSION

7.1 Introduction

Each SCSV component appears to encode a single complete gene. All of the *cis* regulatory signals involved in transcription and translation, must be encoded within each of the 1kb components. This makes SCSV genome components some of the most compact in eukaryotic systems, analogous to the single gene encoding minichromosomes in the macronuclei of hypotrichous ciliates (Hoffman *et al.*, 1995). Furthermore, SCSV genes are probably transcribed and translated almost entirely by host enzymes. These features suggest that SCSV may be a useful system to study plant gene expression.

The aim of the work presented in this chapter was to probe the noncoding region sequences of the SCSV components for promoter activity, by linking them to the reporter gene GUS, and transforming plants with these GUS-fusion constructs.* Two of the SCSV promoters had been tested in protoplasts of tobacco and subterranean clover and found to be active in both, although they gave higher expression in subterranean clover protoplasts (B. Surin, unpublished). The expression levels of the promoters in whole plants may not be the same as in protoplasts, and other features of the promoters, such as the timing and tissue specificity, can only be determined by testing the promoters in transgenic plants. Tobacco was chosen for the initial testing of the SCSV promoters in whole plants because current subterranean clover transformation protocols are inefficient and time consuming.

The GUS gene, *uid*A, which encodes β-glucuronidase, was developed as a marker gene system by Jefferson (1987) and Jefferson *et al.* (1987). GUS

has proven very useful as a marker gene in plant systems as the activity of β-glucuronidase is easily detected against the cellular background. It is routinely detected by hydrolysis of compounds such as X-gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide) which produces a blue dye, 5-bromo-4-chloro-3-indigo, or 4-methylumbelliferyl β-D-glucuronide (MUG) which is hydrolysed to 4-methylumbelliferone (MU). MU is measured fluorimetrically (Jefferson, 1987). Staining of intact plant tissue with X-gluc shows where the GUS gene is being expressed. The levels of expression can be quantified spectrophotometrically when X-gluc is used, or fluorimetrically when MUG is used as the substrate.

GUS has been used to study the structures and functions of putative promoter regions of CFDV and the geminiviruses. CFDV is limited to the phloem tissue of coconuts (Randles *et al.*, 1992). The promoter region from the CFDV component which encodes the putative RAP was found to drive GUS expression predominantly in the vascular tissue of transgenic tobacco (Rhode *et al.*, 1995). A great deal more work has been done on geminivirus promoters. As geminiviruses are structurally similar to SCSV and may also be phylogenetically related (Chapter 3 and 4), analyses of geminivirus promoters may provide good models or comparative studies for the analysis of SCSV promoters. The promoter regions of ACMV were linked to GUS and the promoter-GUS fusions were analysed in protoplasts, leaf disks, and transgenic plants alone (Zhan *et al.*, 1991) and in combination with the ORFs of ACMV expressed from the CaMV 35S promoter (Haley *et al.*, 1992). The strongest promoter was that of the AL1 gene, which encodes the RAP. Coexpression of the AL1 ORF with the AL1 promoter linked to GUS reduced expression by about 50%. This demonstrated that AL1 is self-regulating (ie. the AL1 protein

down-regulates its own promoter) and indicated that AL1 is an early gene, consistent with its involvement in viral replication. The coat protein promoter alone gave weak expression, but was transactivated threefold by coexpression of the AL2 ORF and by about 60% by the AL3 ORF. The B component promoters were also transactivated by the AL2 and AL3 proteins. The coat protein gene and the B component movement genes were therefore suggested to be late genes. The gene products of viral late genes are those which are only required later in the infection of a cell and the genes are therefore down-regulated until the proteins produced early in infection (from early genes) are sufficiently abundant to activate them. The promoter regions of TGMV have been analysed in a similar way to those of ACMV (Sunter and Bisaro, 1992; Sunter *et al.*, 1993; Gröning *et al.*, 1994). As with ACMV, expression of the AL1 ORF was found to repress expression from the AL1 promoter (Sunter *et al.*, 1993), and AL2 was found to transactivate the coat protein promoter and the promoter of BR1 but not that of BL1 (Sunter and Bisaro, 1992). The expression of a small ORF overlapping AL1, AL4, was also found to contribute to the suppression of expression from the AL1 promoter (Gröning *et al.*, 1994). The promoter regions of the subgroup I geminivirus CSMV have also been analysed using GUS fusions in *Zea mays* protoplasts (Zhan *et al.*, 1993). As with ACMV and TGMV, the coat protein promoter of CSMV was transactivated by expression of the RAP ORF (there is no AL2 homologue in CSMV). The activity of the coat protein promoter was also enhanced to a lesser extent by the coat protein.

Like the promoters of these geminiviruses, the SCSV promoters were anticipated to have different levels and tissue specificities that may be modified

by different viral proteins. Furthermore, SCSV is obligately aphid transmitted and not mechanically transmissible (Chapter 1) and thus is likely to be limited to the phloem. The phloem limitation may be reflected in vascular specific activity of the SCSV promoters.

7.2 Results

7.2.1 Construction of noncoding region-GUS fusions

The whole of the noncoding region sequence of each SCSV component was amplified by PCR with the primers listed in Table 7.1. The PCR fragments were separately cloned in front of the promoterless GUS gene in pHW9 (constructed by R. Dolferus) using the *Nco*I and *Bam*HI restriction sites engineered into the primers (Figure 7.1). One or two nucleotides 5' of the native ATG initiation codons were altered to generate the *Nco*I site. The initiation codon of the GUS gene was therefore in the same position as the start of each component's main ORF would be. The junctions of the clones were checked by sequencing. The pHW9 GUS-fusion constructs were then separately cloned into the binary vector pGA470 (An *et al.*, 1985) at the *Eco*RI site and electroporated into *A. tumefaciens* LBA4404. A promoterless GUS control construct was also made by cloning unmodified pHW9 into the *Eco*RI site of pGA470.

7.2.2 Transformation and analysis of tobacco

N. tabacum var. cv. Wisconsin 38 plants were separately transformed with the GUS fusion constructs and the promoterless GUS control as described in Methods 2.6. Whilst the plants were still in tissue culture, leaf and stem pieces were taken from each kanamycin resistant plant and stained with X-gluc.

Table 7.1

Sequences and positions of PCR primers for cloning non-coding regions and corresponding PCR fragment sizes.

Primer 1 name	Position of 5' end	Primer 1 sequence (5'-3')	Primer 2 name	Position of 5' end	Primer 2 sequence (5'-3')	Fragment size (bp)[a]
S1nc3	241	GGCGTGCGTCGGCCATGG CGCTATGAAATTCTGAAC	S1nc5	577	GGCGTGCGTCGGGGATCCTA TGTTGTAATTTTATATGG	665
S2nc3	79	GGCGTGCGTCGGCCATGG AAGCTTAGAGAGAGAAAG	S2nc5	924	GGCGTGCGTCGGGGATCCA ATAAAAGAATATATATTATTG	177
S3nc3	374	CTCACTATAGAACCATGGA CACAAGATTCTAAG	S3nc5	863	CTCACTAAAGGGGGATCCTG AGATGTAATTGTG	502
S4nc3	347	CTCACTATAGAACCATGGA AACGCAGAACAAAG	S4nc5	803	CTCACTAAAGGGGGATCCTA ATTGTTATTATCA	546
S5nc3	372	CTCACTATAGAACCATGGT CGTTGTAAAATGAC	S5nc5	874	CTCACTAAAGGGGGATCCTA ATTGTGATGATT	523
S6nc3	50	CTCACTATAGAACCATGGT GGGCCCAGGGAAGCGA	S6nc5	903	CTCACTAAAGGGGGATCCTG AAAACTCTGCGAA	164
S7nc3	383	CTCACTATAGAACCATGGC TTAAAACCAGAACA	S7nc5	817	CTCACTAAAGGGGGATCCTA ATTAATAGTAATTATG	554

[a] not including the spacer nucleotides at the 5' ends of the primers

Figure 7.1

Construction of GUS fusions

PCR amplified fragments representing the entire noncoding region (containing the putative promoters) of each SCSV component (DNAs 1 to 7) were each digested with *Bam*HI and *Nco*I, whose sites are represented by B and N, and separately cloned into pHW9 (created by R. Dolferus) at those sites. The GUS coding region (blue rectangle) was followed by the nopaline synthase termination sequence (3'NOS; Depicker *et al.*, 1982). The lengths of the SCSV noncoding region fragments are indicated. The sizes of the inserts, GUS gene, and NOS terminator are not proportional. The arrows labelled T7 and SP6 indicate the T7 and SP6 RNA polymerase initiation sequences flanking the polycloning site of the pGEM3Z plasmid (Promega) on which pHW9 was based.

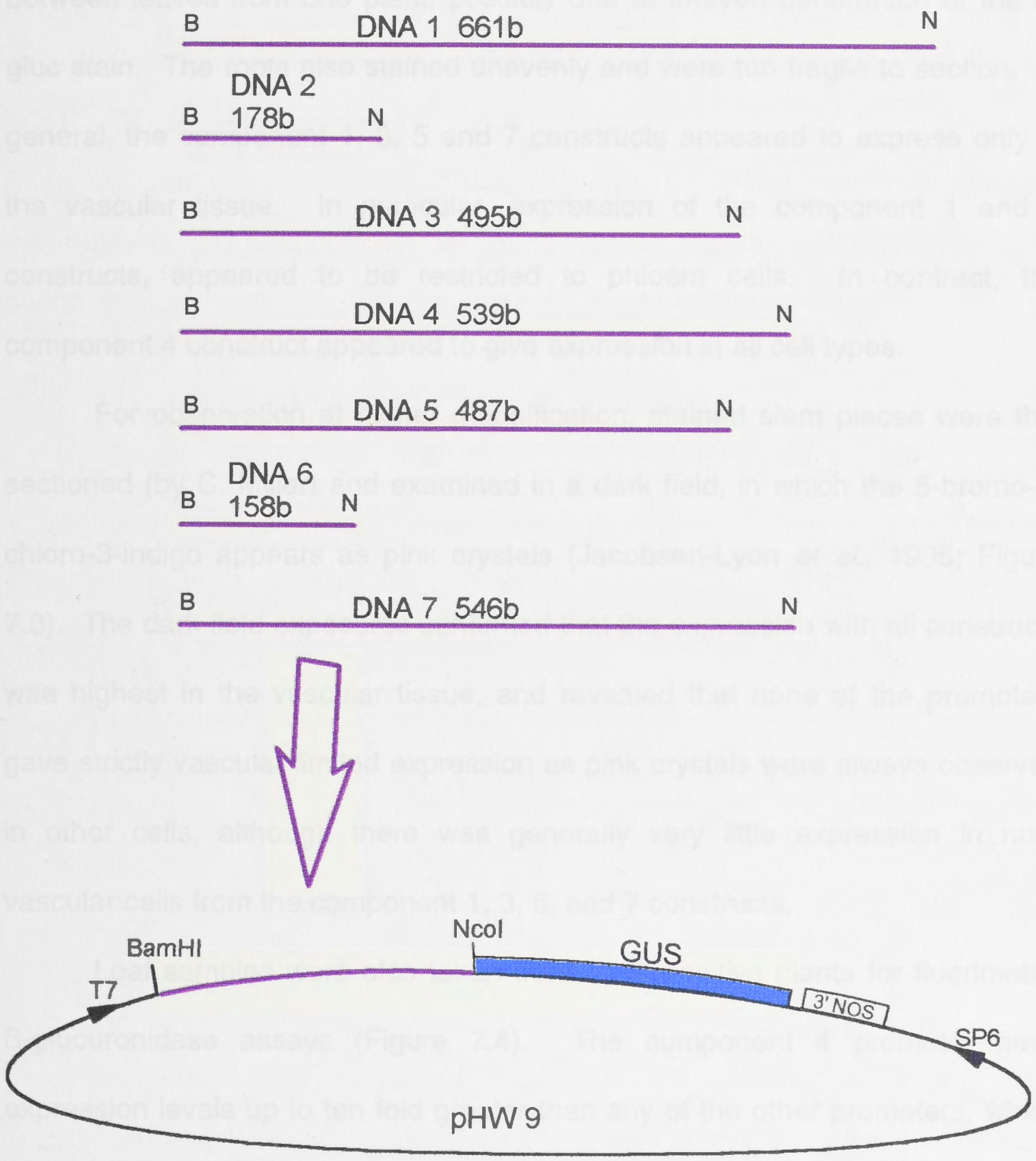

B
DNA 1 661b
N
DNA 2
B 178b N
B
DNA 3 495b
N
B
DNA 4 539b
N
B
DNA 5 487b
N
DNA 6
B 158b N
B
DNA 7 546b
N
NcoI
BamHI
GUS
T7
3' NOS
SP6
pHW 9

Figure 7.2 shows patterns of X-gluc staining observed in a bright field in leaves (L), stems (S) and roots (R) of plants transformed with the different constructs. The expression patterns were judged largely on the basis of staining in stem sections. Staining of the leaves was frequently blotchy and variable, even between leaves from one plant, possibly due to uneven penetration of the X-gluc stain. The roots also stained unevenly and were too fragile to section. In general, the component 1, 3, 5 and 7 constructs appeared to express only in the vascular tissue. In particular, expression of the component 1 and 3 constructs, appeared to be restricted to phloem cells. In contrast, the component 4 construct appeared to give expression in all cell types.

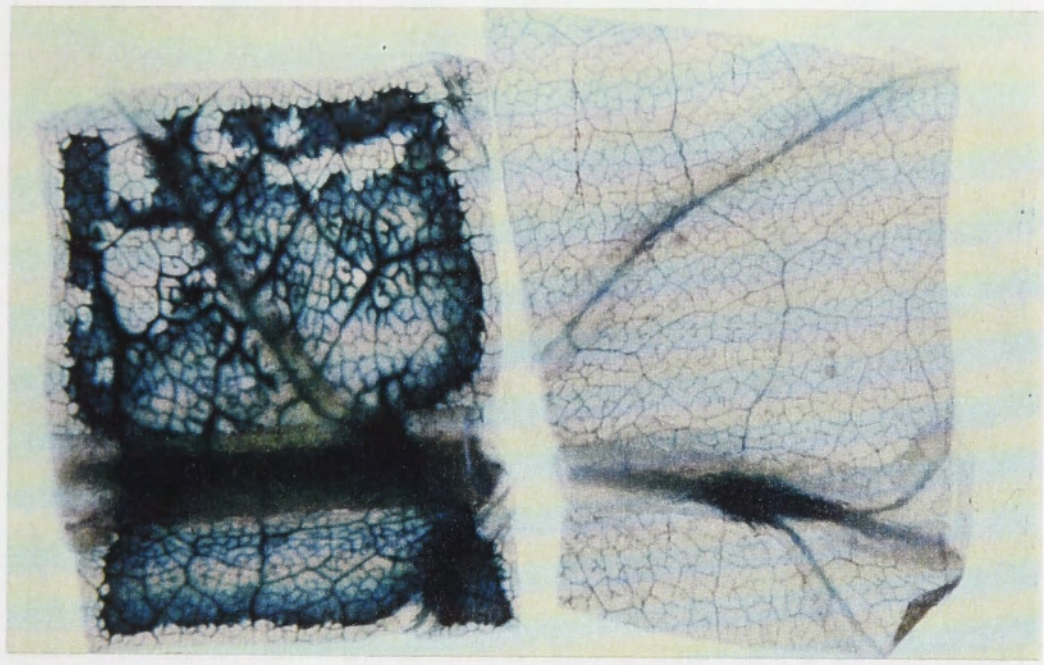
For observation at higher magnification, stained stem pieces were thin sectioned (by C. Miller) and examined in a dark field, in which the 5-bromo-4-chloro-3-indigo appears as pink crystals (Jacobsen-Lyon *et al.*, 1995; Figure 7.3). The dark field exposures confirmed that the expression with all constructs was highest in the vascular tissue, and revealed that none of the promoters gave strictly vascular limited expression as pink crystals were always observed in other cells, although there was generally very little expression in non-vascular cells from the component 1, 3, 5, and 7 constructs.

Leaf samples were also taken from GUS positive plants for fluorimetric ß-glucuronidase assays (Figure 7.4). The component 4 promoter gave expression levels up to ten fold greater than any of the other promoters, which all had similar expression levels. Some transgenic tobacco containing the CaMV 35S promoter fused to GUS were available (from M. Graham). The levels of GUS expression in leaf samples of these were also measured fluorimetrically and were generally found to be about four times stronger than

Figure 7.2

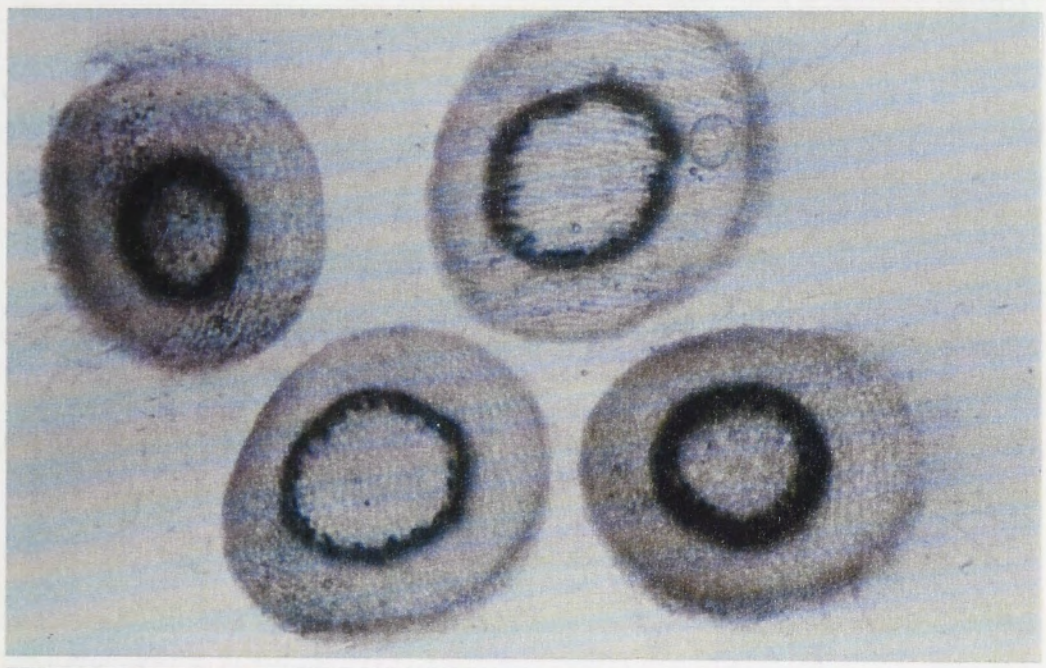
Histochemical staining for GUS activity in transgenic plants - bright field

Bright field exposures of stained leaf pieces (L), stem sections (S), roots (R) and pollen (P) from tobacco plants transformed with the GUS fusion constructs containing the component 1 (S1nc), 3 (S3nc), 4 (S4nc), 5 (S5nc), and 7 (S7nc) promoter regions. Blue colouration indicates GUS expression. Each leaf, stem or root piece represents an independent transformant (except the top two root pieces of S5nc which could not be separated easily), and the pollen samples were mixtures from two or more transformants.

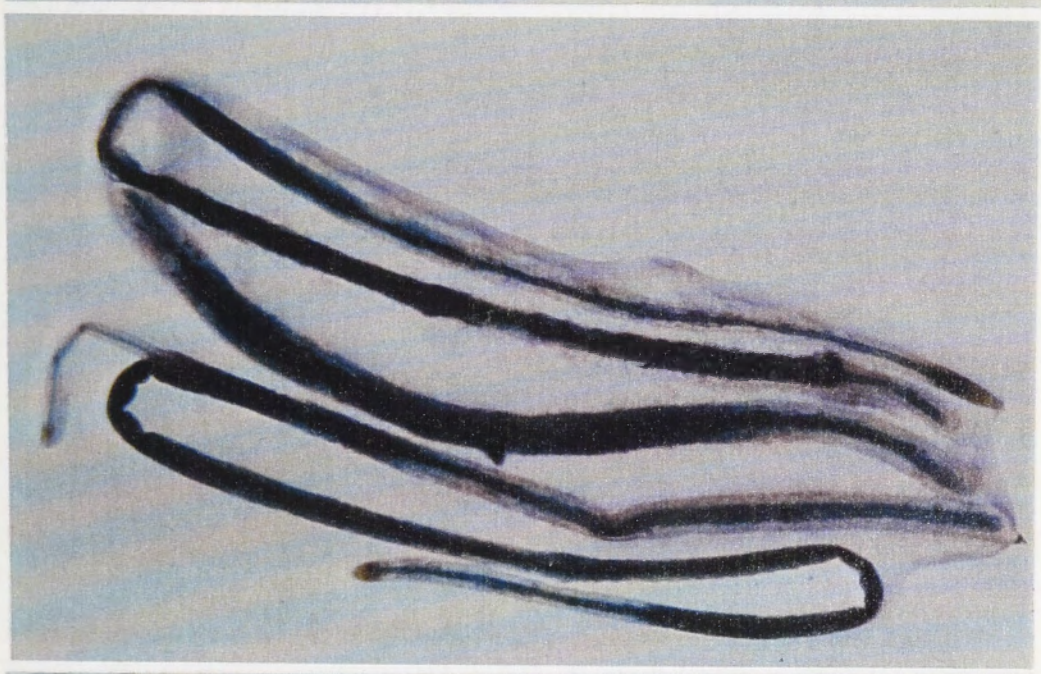


SInc

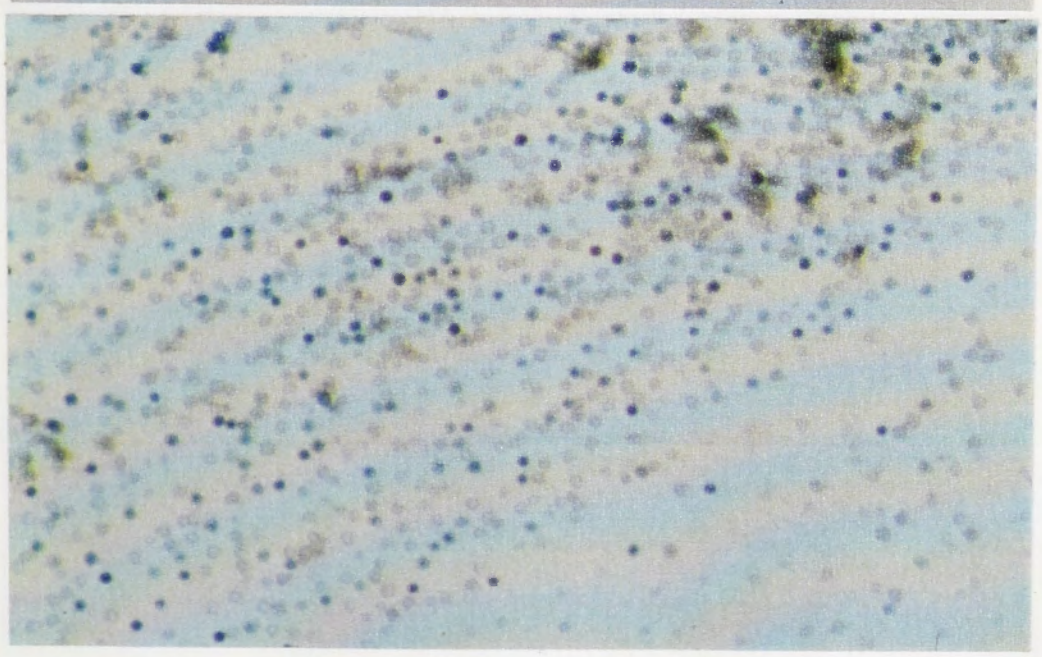
L



S

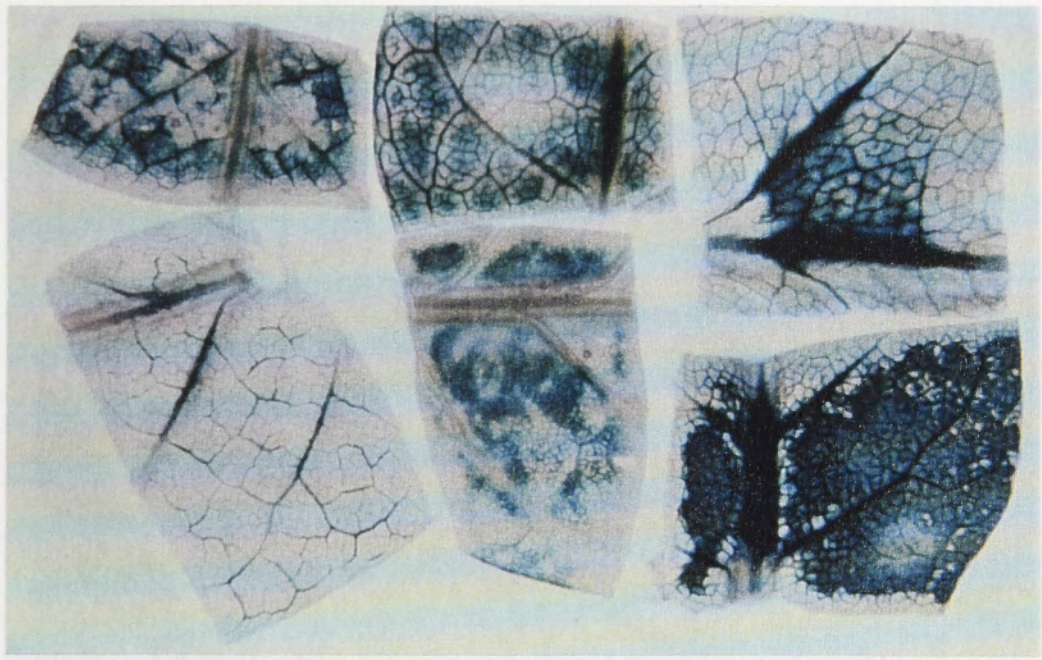


R

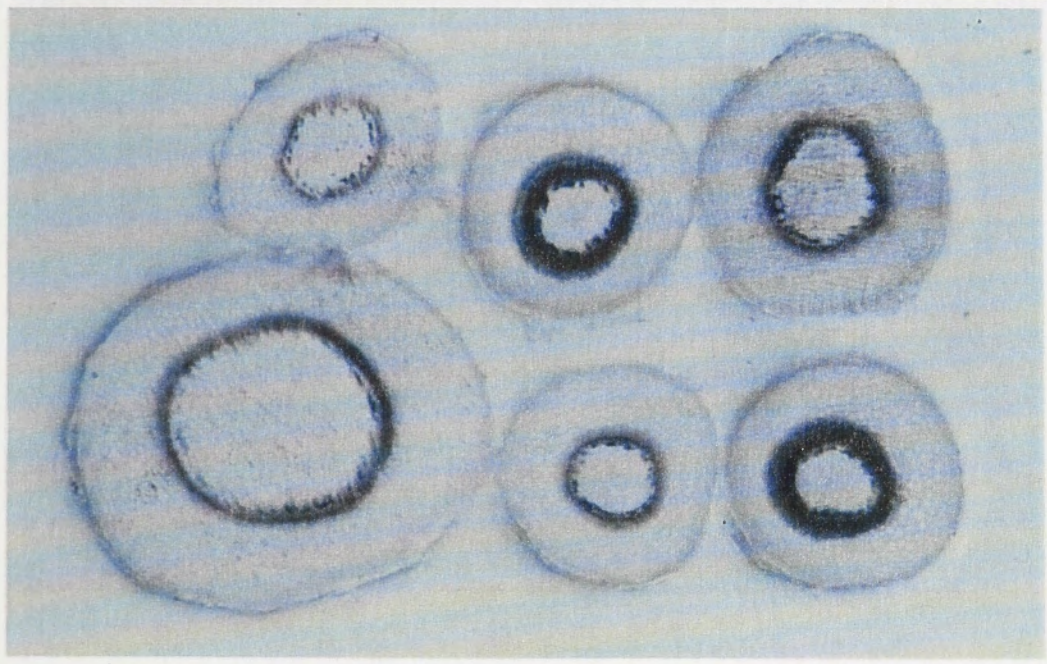


P

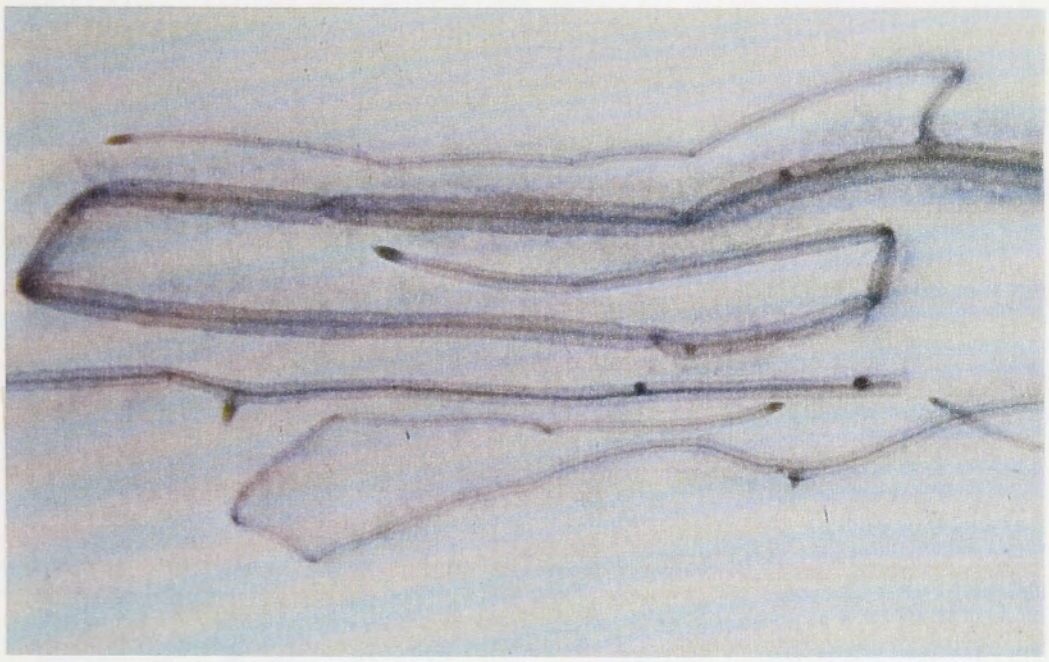
S3nc



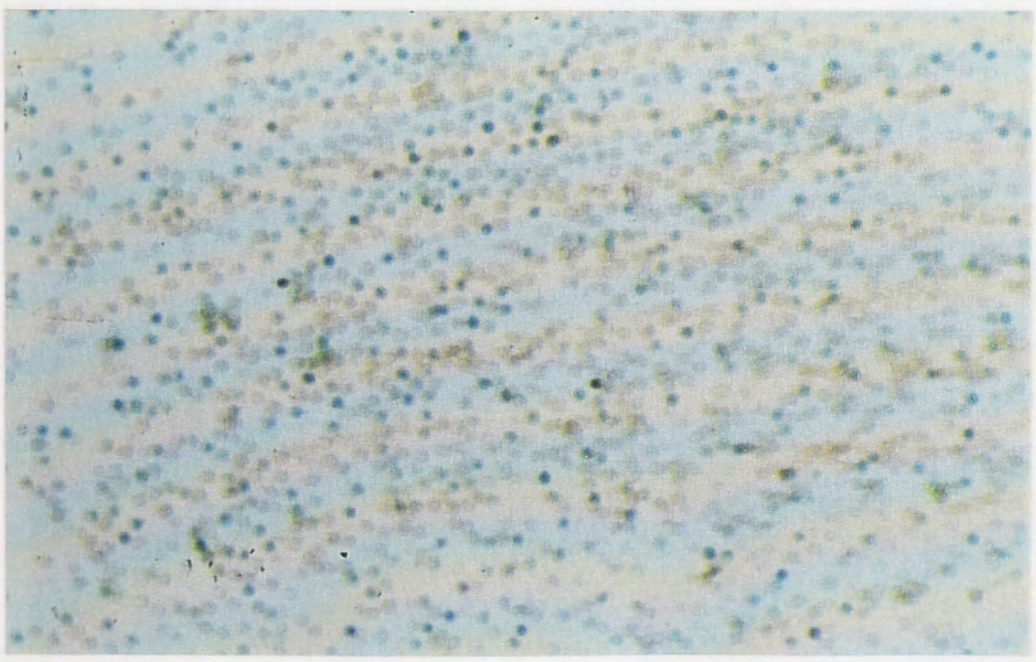
L



S

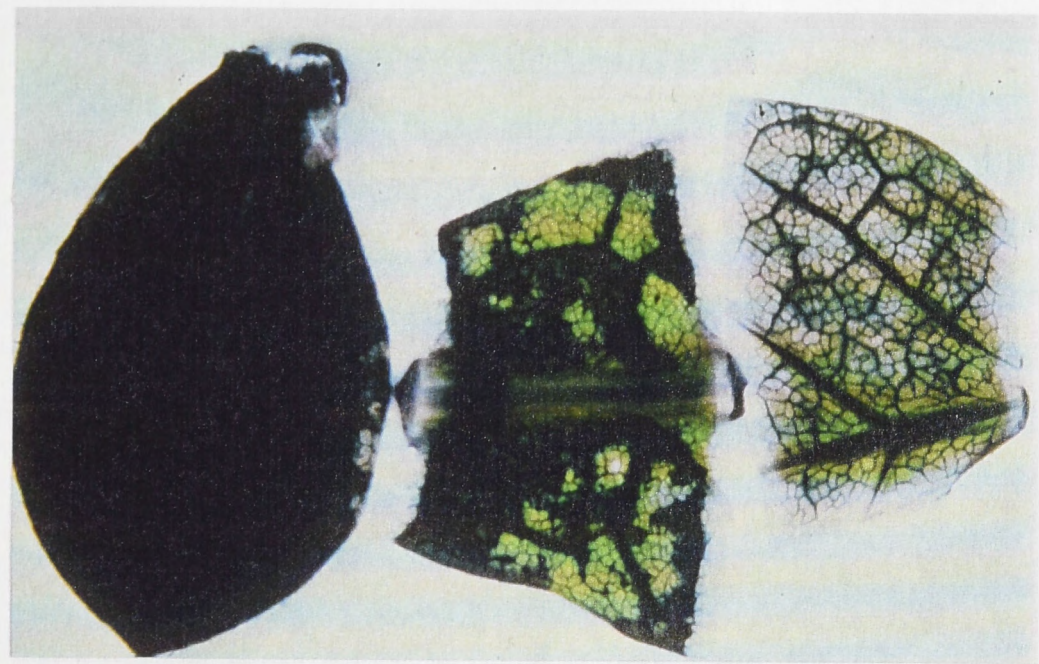


R

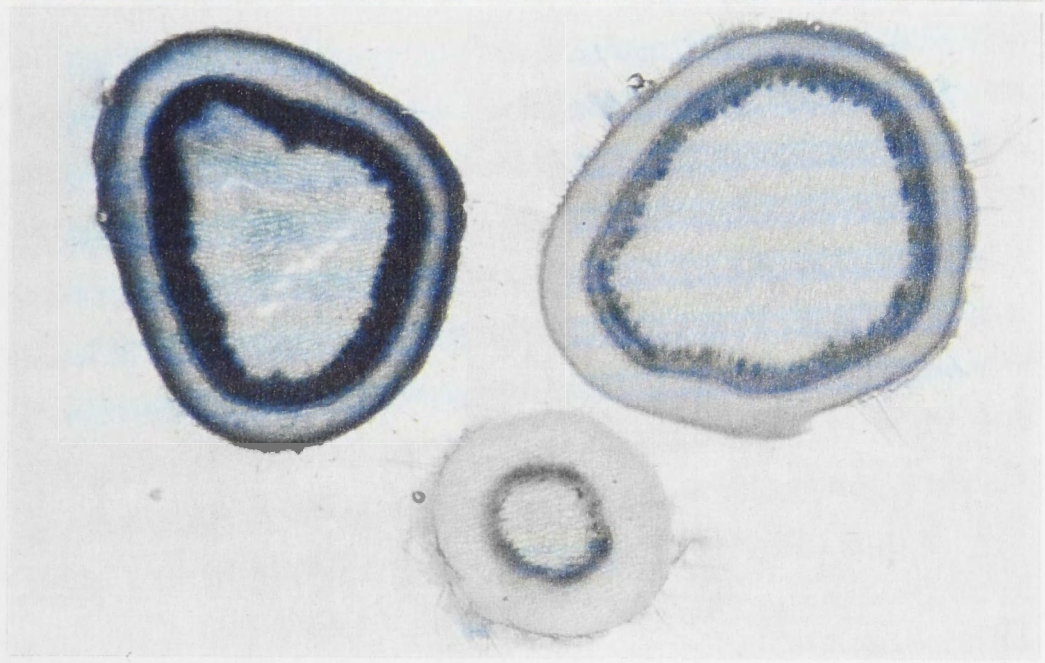


P

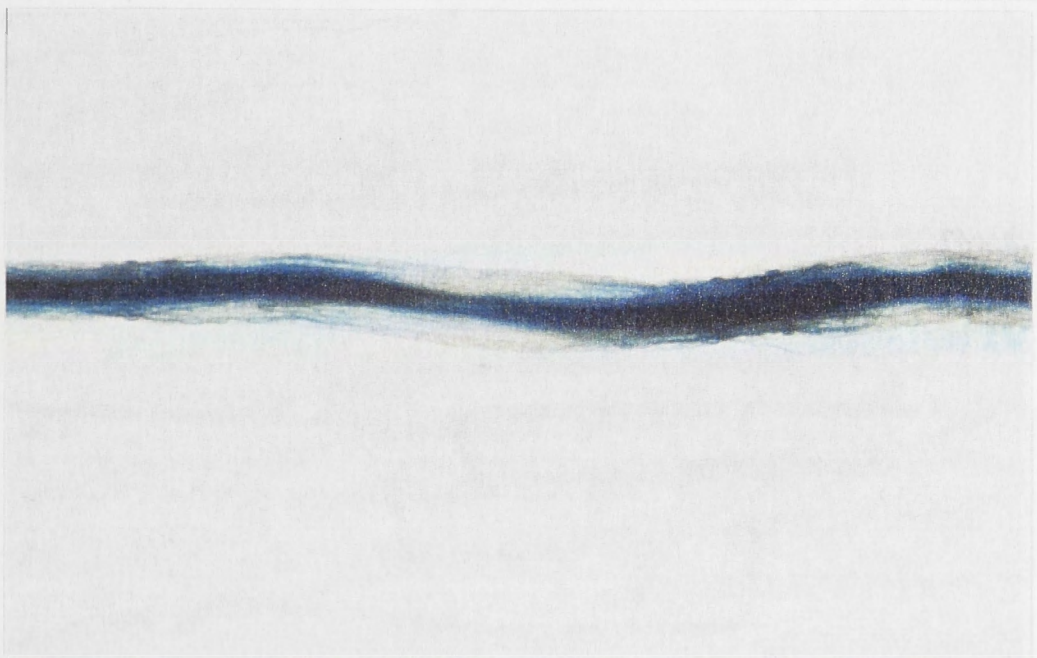
S4nc



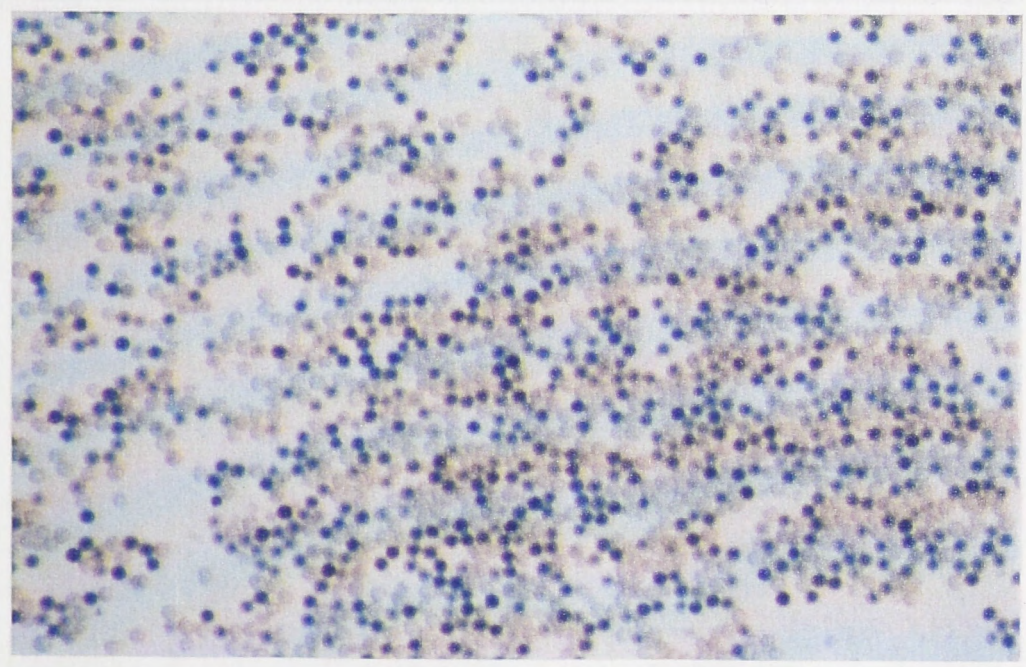
L



S

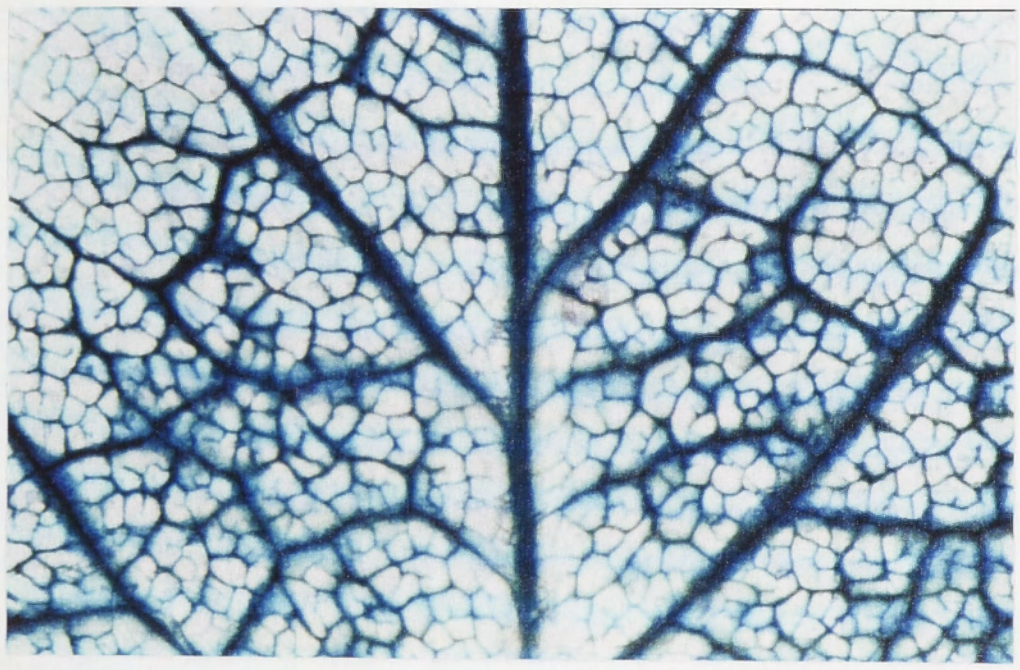


R

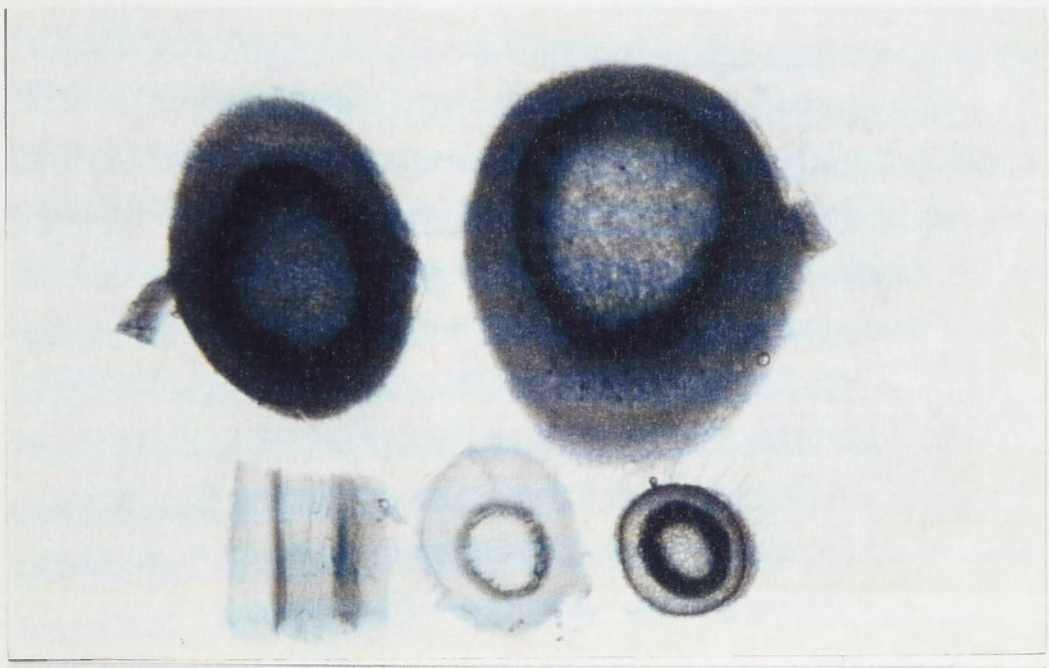


P

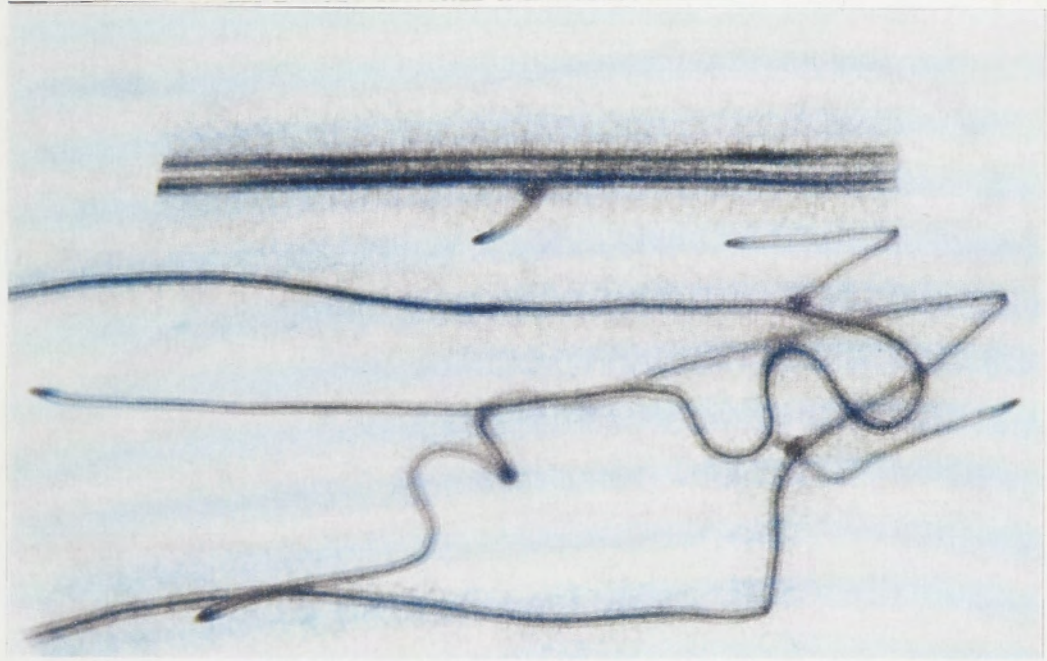
S5nc



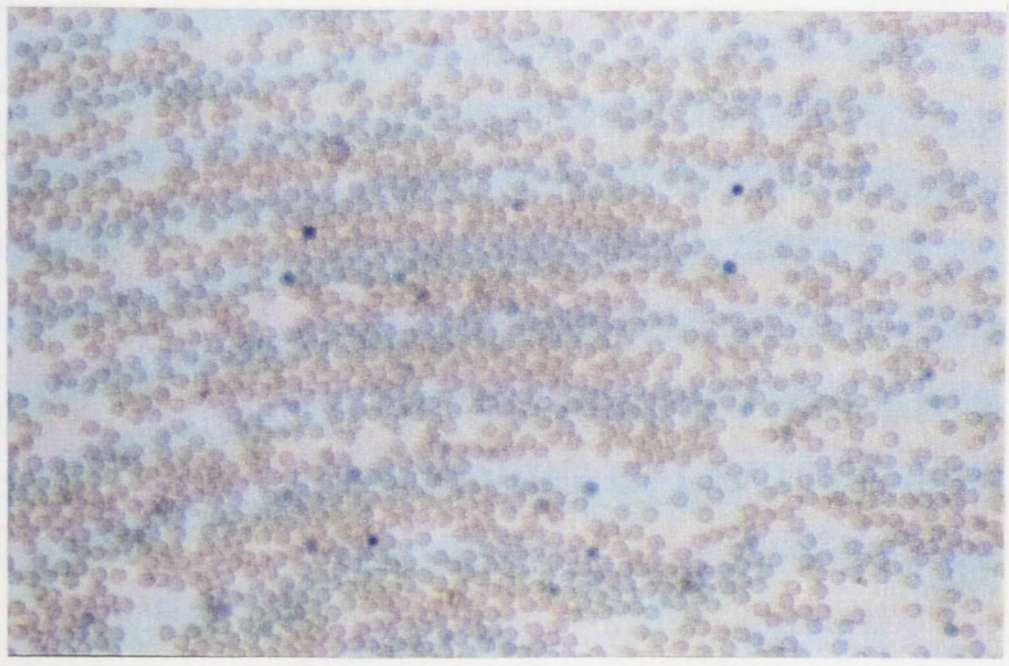
L



S

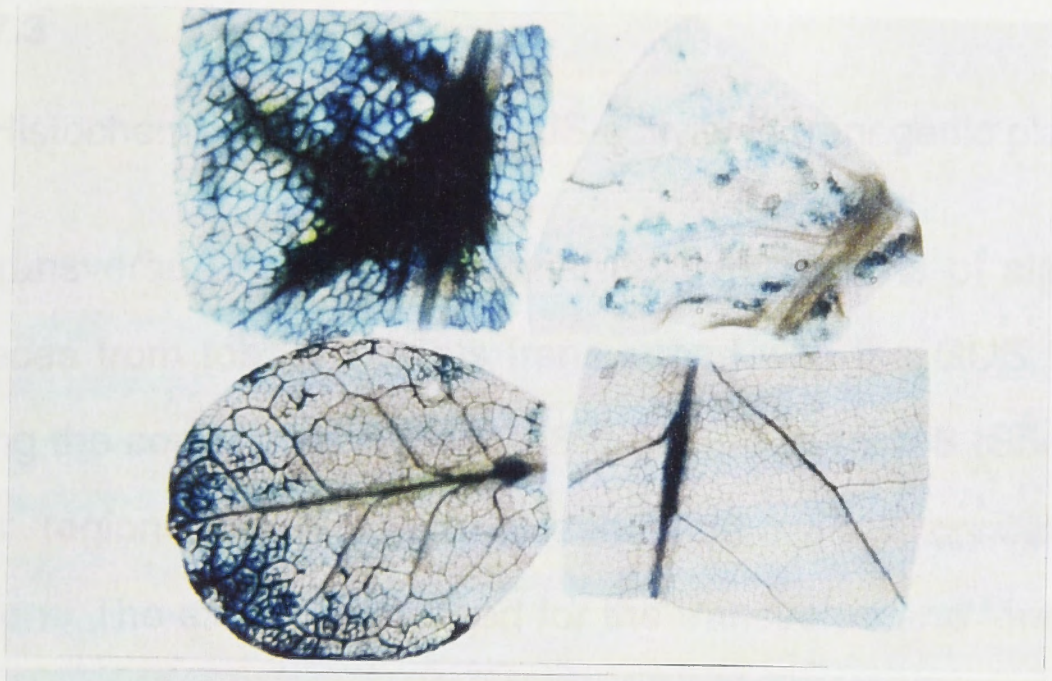


R

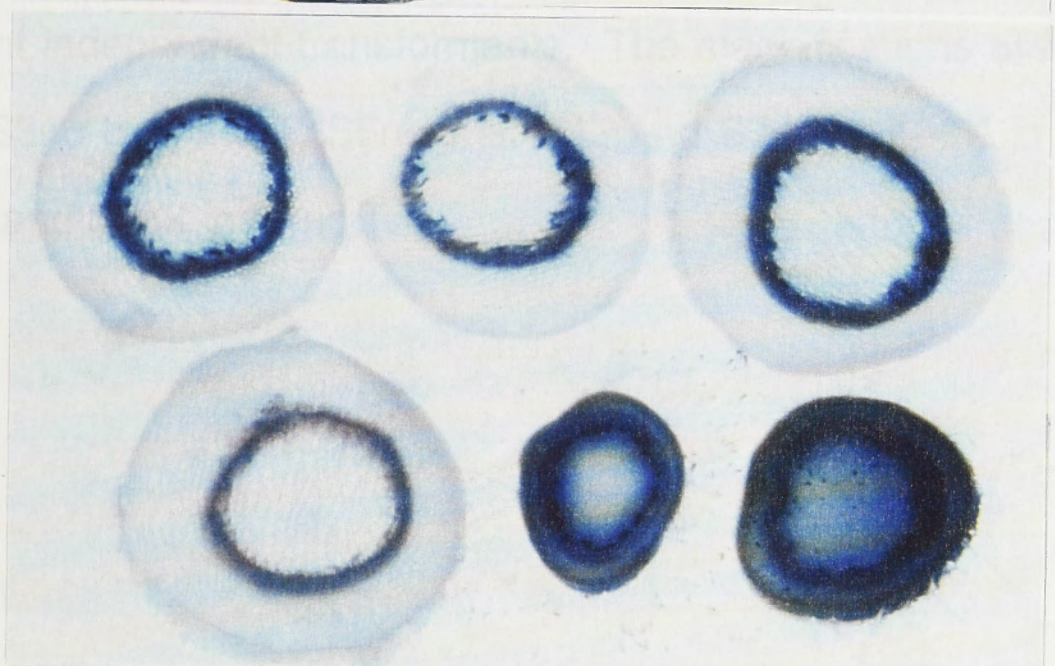


P

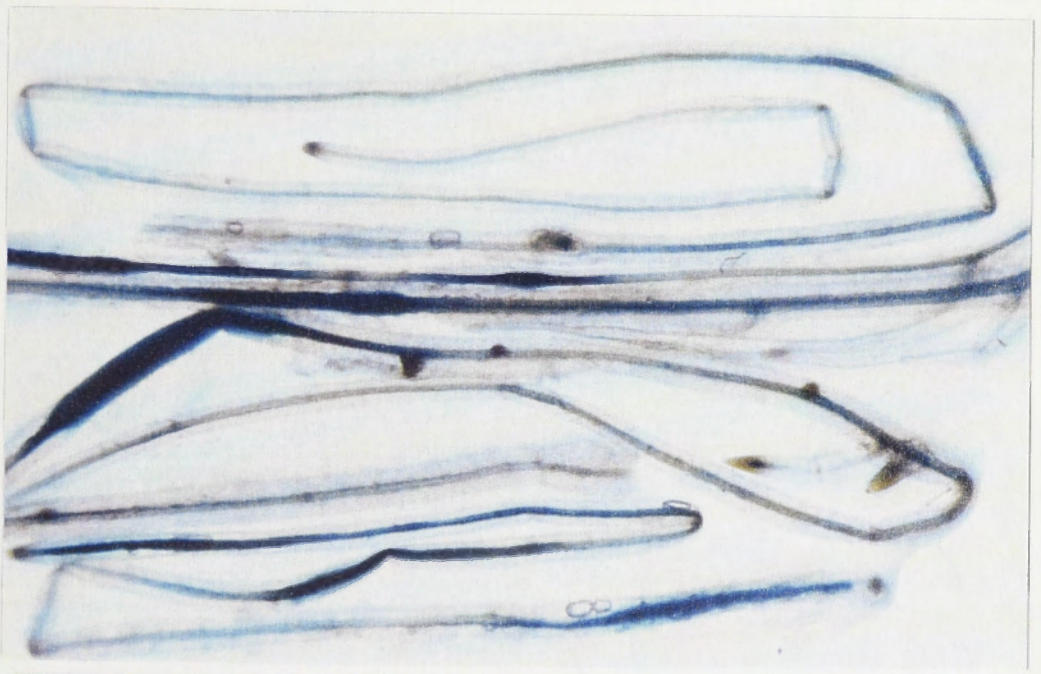
S7nc



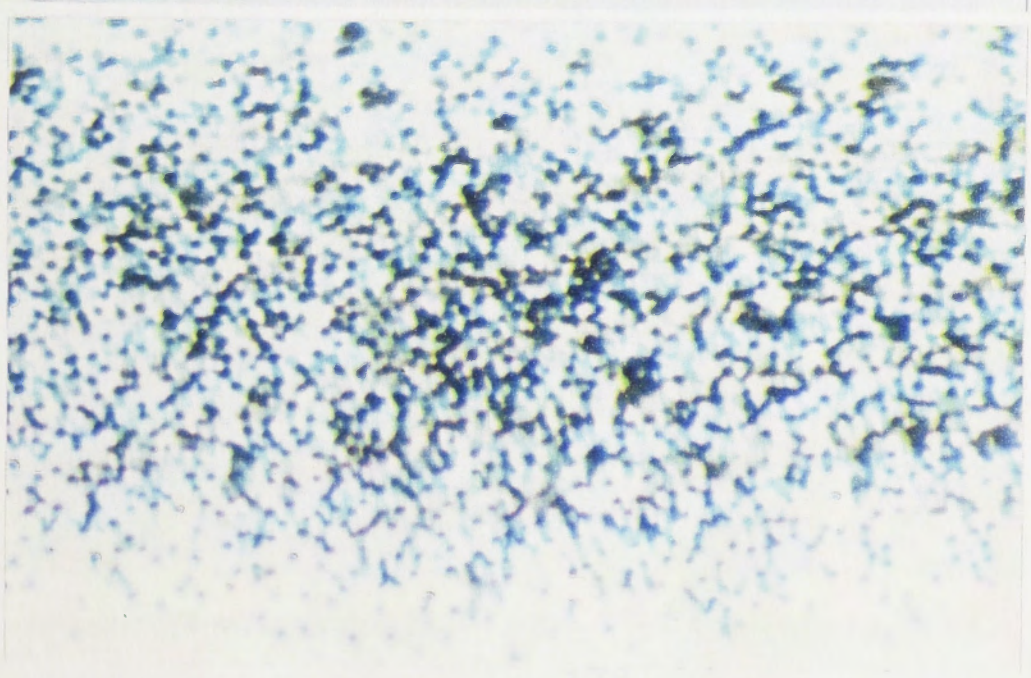
L



S



R



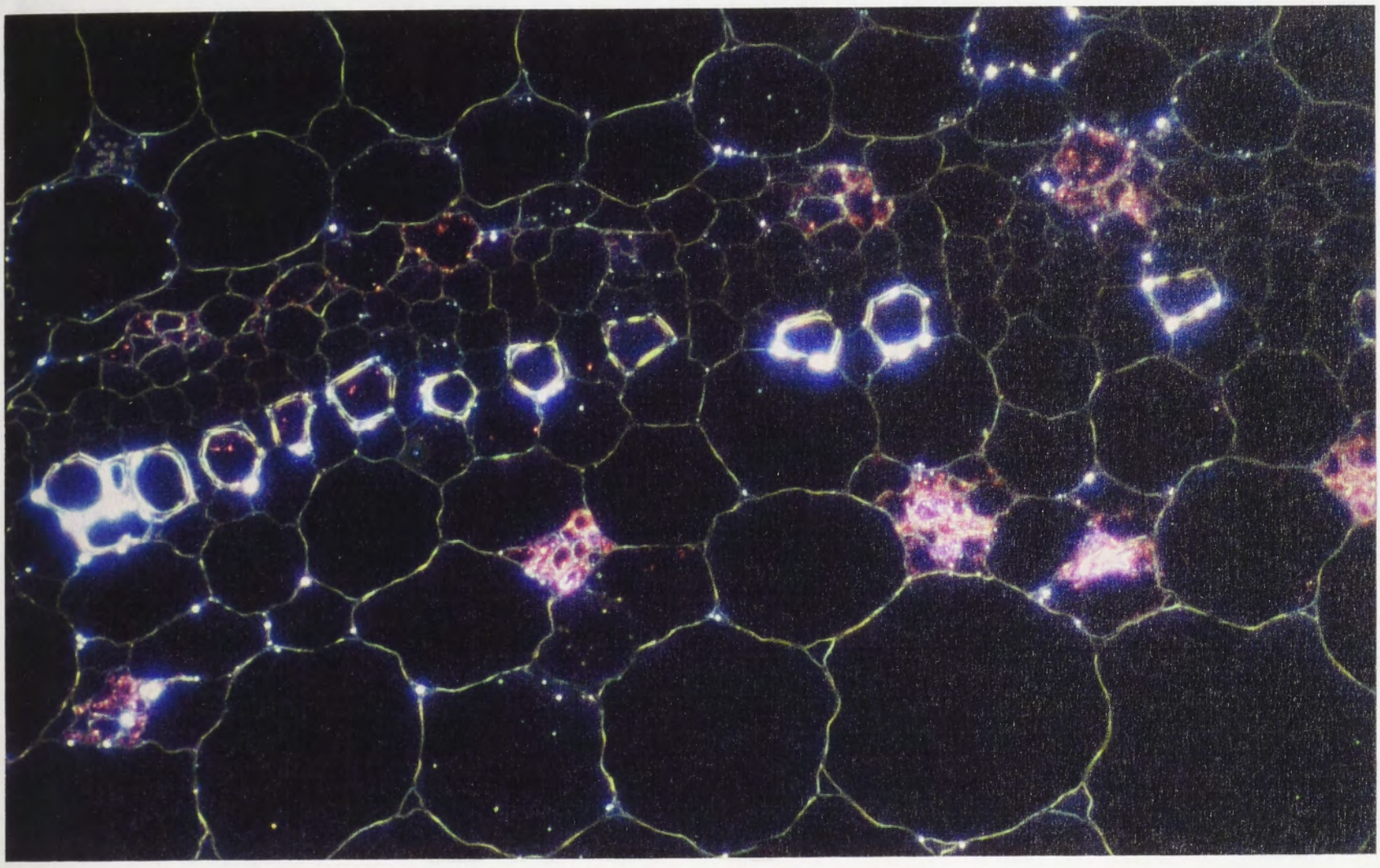
P

Figure 7.3

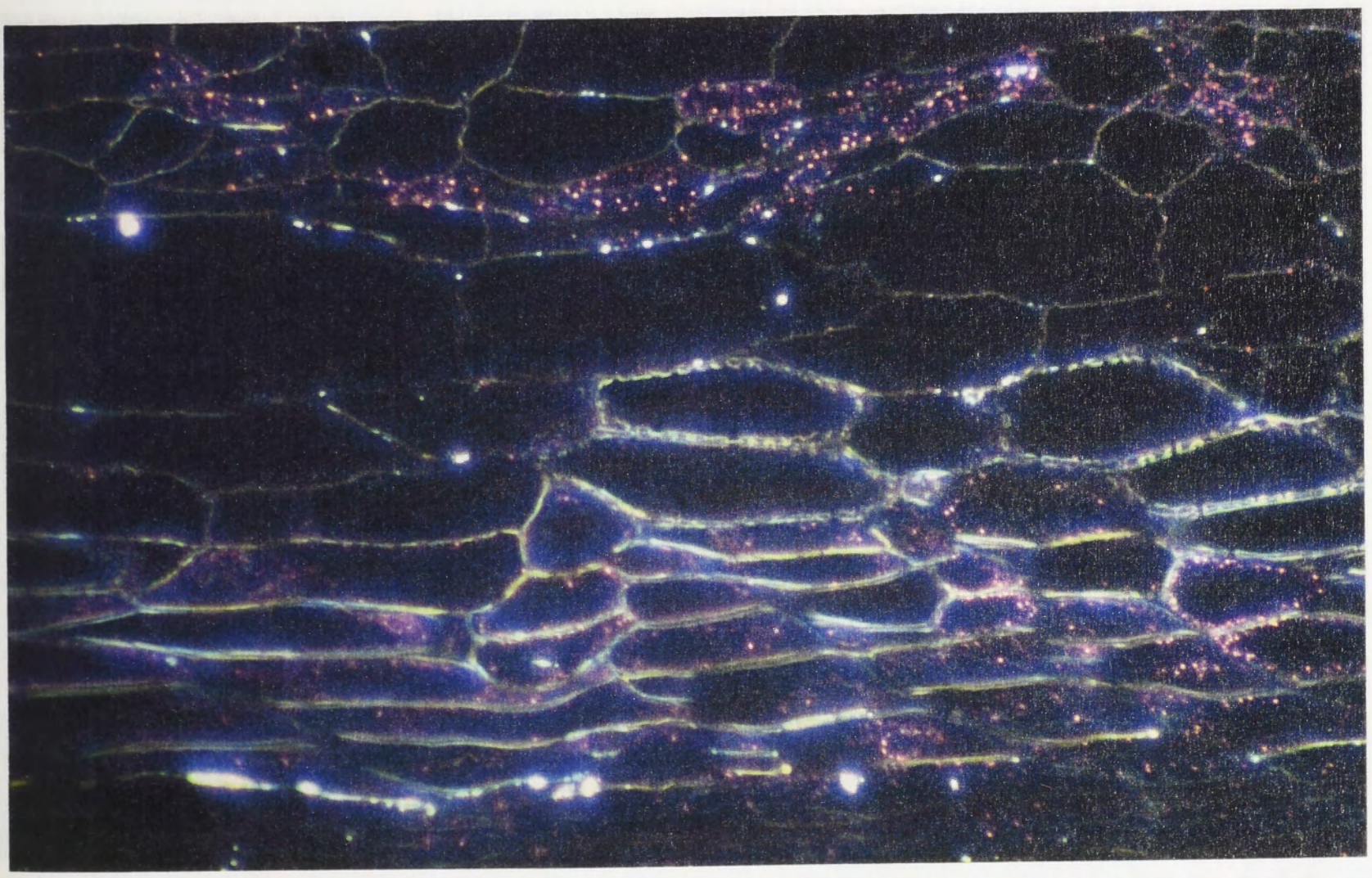
Histochemical staining for GUS activity in transgenic plants - dark field

Transverse (T) and longitudinal (L) thin sections of stained, embedded stem pieces from tobacco plants transformed with the GUS fusion constructs containing the component 1 (S1nc), 3 (S3nc), 4 (S4nc), 5 (S5nc), and 7 (S7nc) promoter regions, viewed with a dark field. Pink crystals indicate GUS expression. The stem pieces used for the transverse and longitudinal sections represent independent transformants. The magnifications are: 375X in S1nc T and L, S3nc L, S4nc T, S5nc T and L, and S7nc L; 480X in S3nc T; 200X in S4nc L; and 300X in S7nc T.

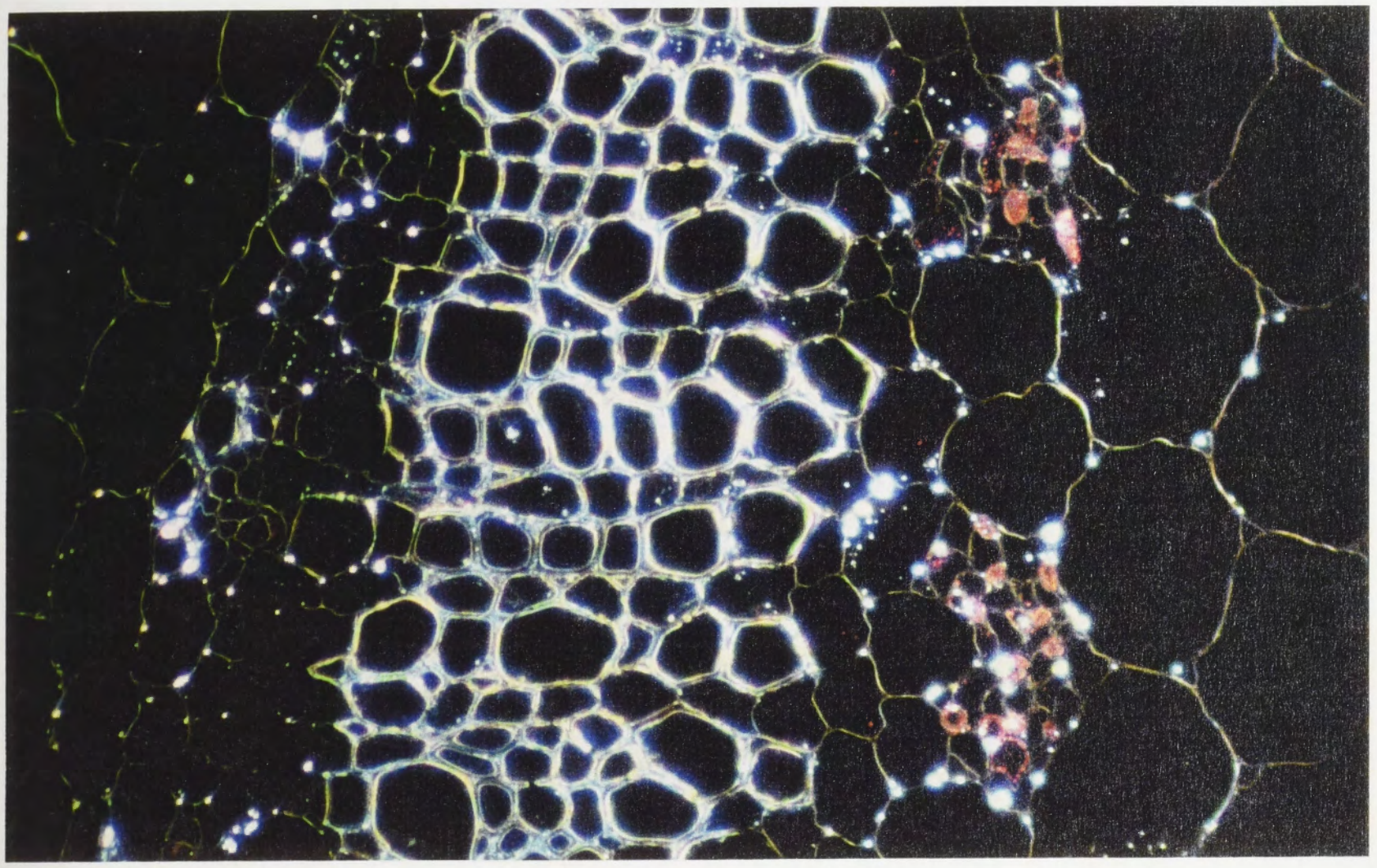
S1nc



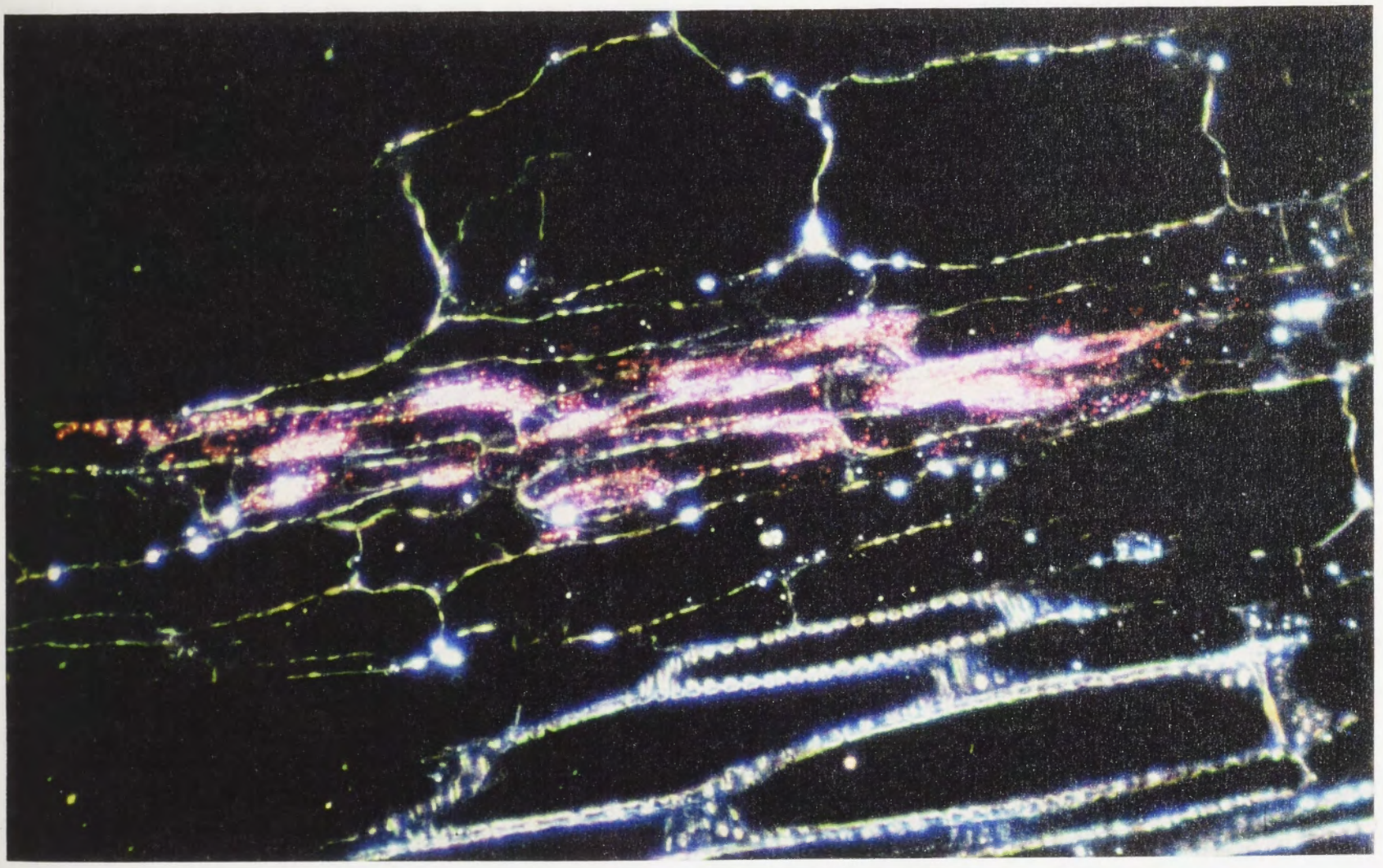
T



L

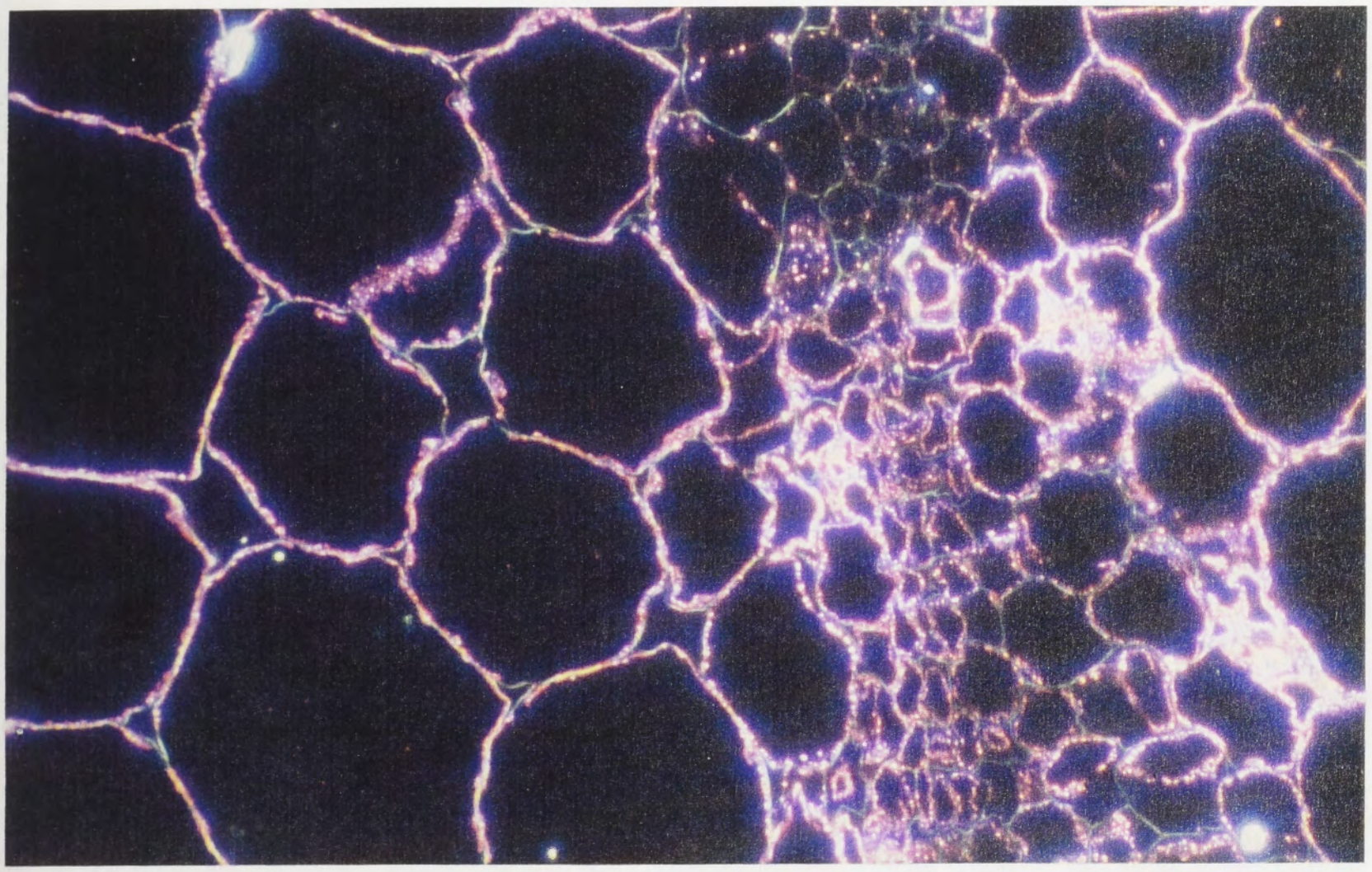


T



L

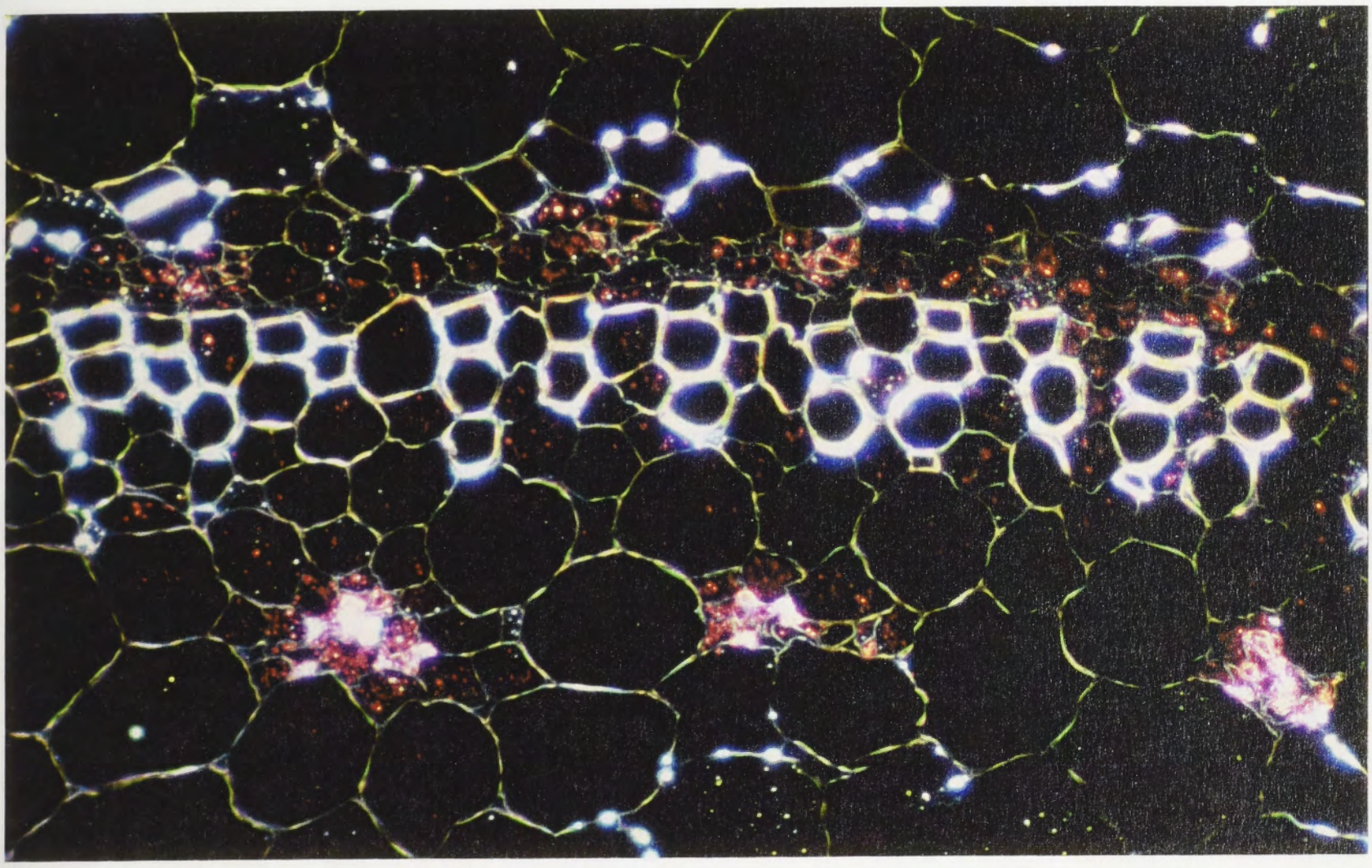
S4nc



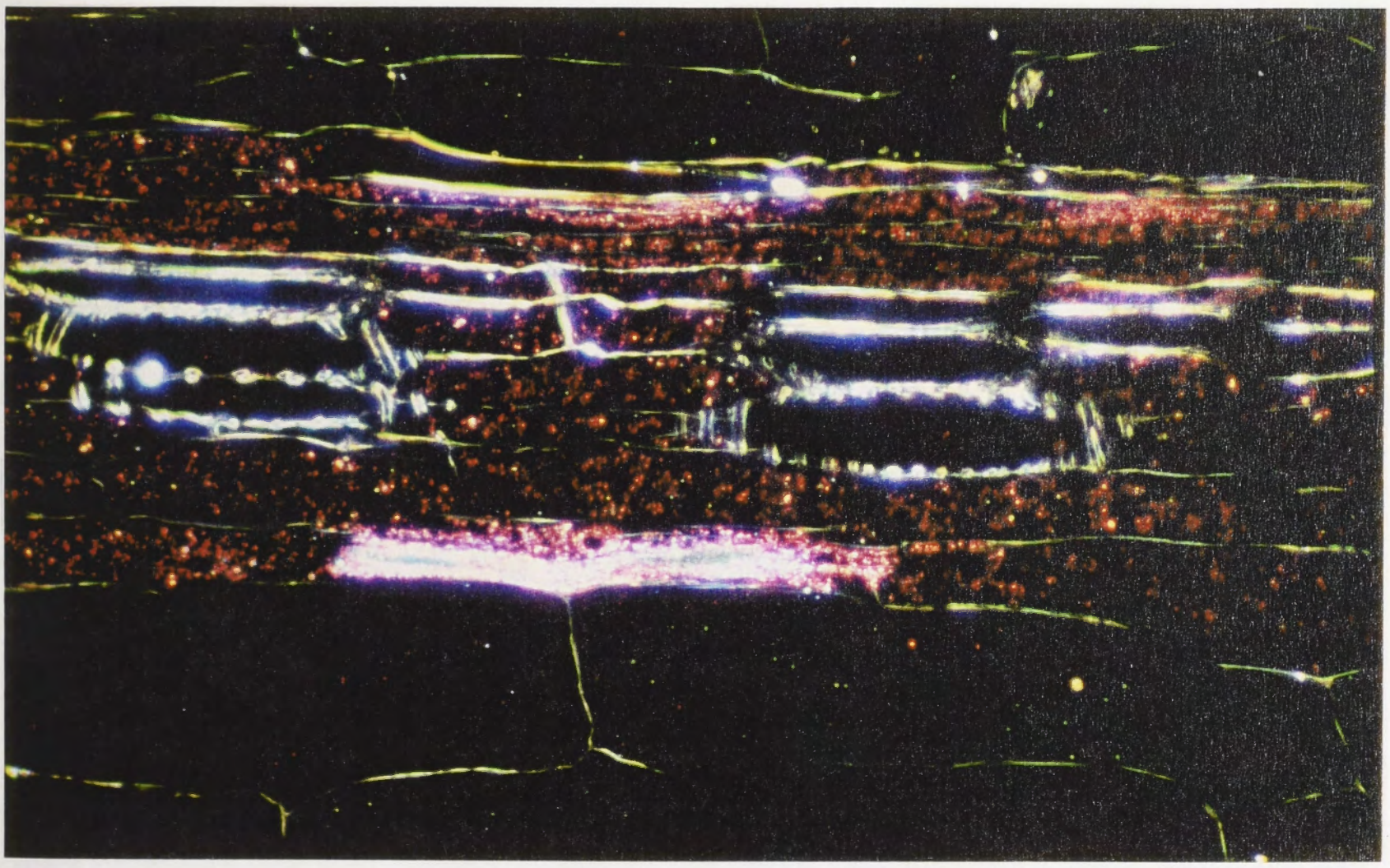
T

L

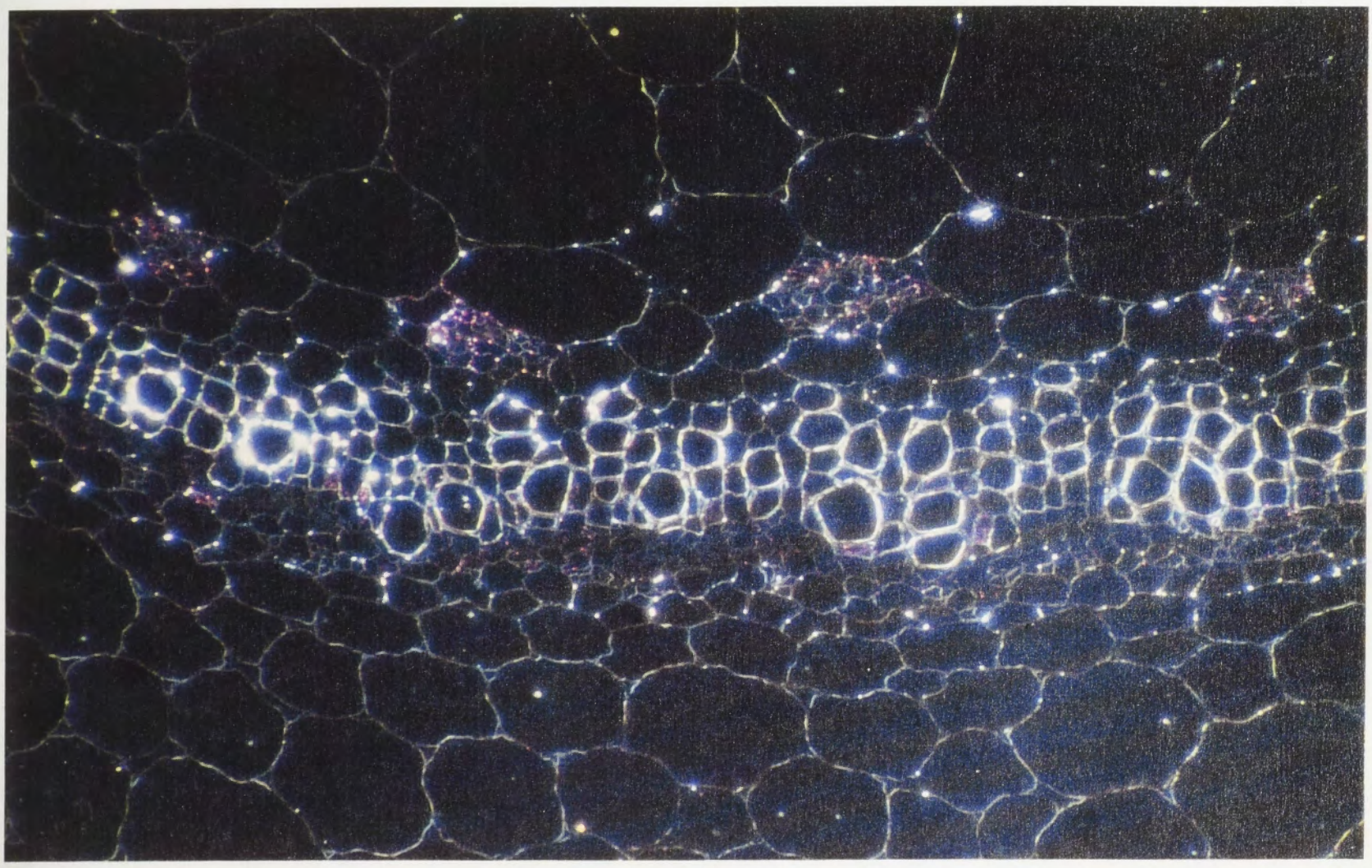
S5nc



T



L



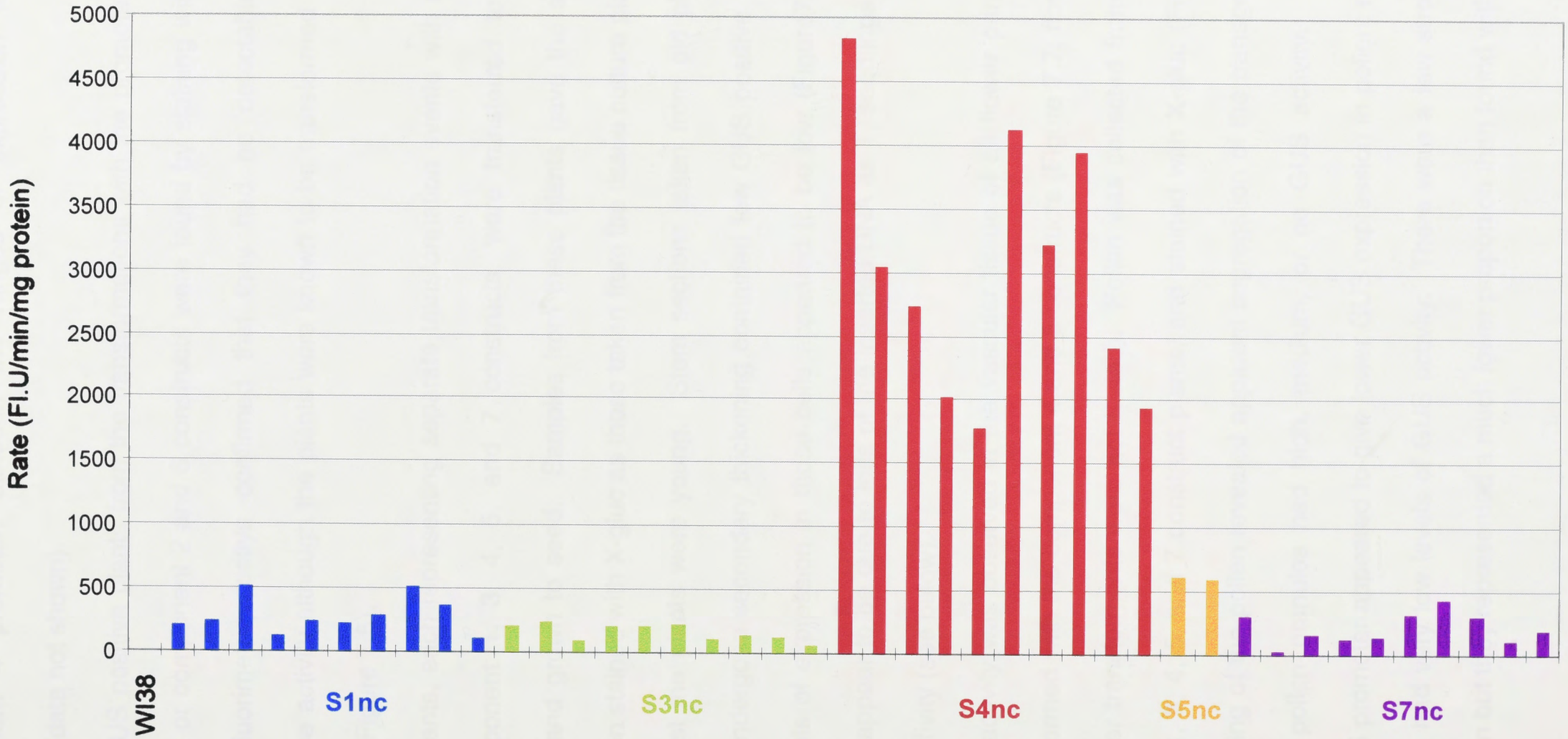
T

L

Figure 7.4

Fluorimetric assay for GUS expression

Fluorimetric assay results from leaf extracts of tobacco plants transformed with the GUS fusion constructs containing the component 1 (S1nc), 3 (S3nc), 4 (S4nc), 5 (S5nc), and 7 (S7nc) promoter regions. Each column represents an independent transformant. GUS activities were measured with a Labsystems Fluoroskan at 5 or 10 minute intervals over 60 minutes (or 30 minutes for S4nc) using 4-methylumbelliferyl β-D-glucuronide (MUG) as the substrate. The rates of GUS activity are expressed as fluorimetric units (Fl.U.) per minute per mg of protein. 1000 Fl.U is approximately equal to 825 pmoles of 4-methylumbelliferone (MU).



Independent tobacco transformants containing noncoding region constructs

the component 4 promoter, although the ranges of expression levels overlapped (data not shown).

No GUS positive plants from the transformations with the promoterless GUS control or component 2 and 6 constructs were found by staining with X-gluc, and fluorimetric assays confirmed that they had no detectable ß-glucuronidase activity although the plants were known to be transformed from Npt assays (Figure 7.5).

Six plants, each representing separate transformation events with each of the component 1, 3, 4, 5, and 7 constructs, were transferred to the glasshouse and grown to seed. Samples from these plants gave the same patterns when stained with X-gluc as those taken from the tissue culture plants, at least whilst the plants were young. Stem sections taken from old plants which had undergone secondary thickening contained few GUS positive cells and the levels of expression in those cells appeared to be low (Figure7.6a). This did not appear to be due to loss of the inserted DNA as T_1 seedlings had high GUS activity (see below).

GUS activity was observed in the vascular tissue of all flower parts of plants transformed with the component 4 and 5 constructs (Figure 7.7; flowers from the other transformants were not tested). Pollen was collected from the component 1, 3, 4, 5, and 7 construct plants, and stained with X-gluc (Figure 7.2P). Staining of the pollen revealed apparent segregation of the transgenes as different pollen granules had high, medium, or no GUS activity. The component 5 promoter appeared to give lower GUS expression in pollen; most of the pollen had very low levels of GUS activity. There were a few strongly stained pollen but they represented a much lower proportion than found with

Figure 7.5

Npt Assay

An example of the results from Npt assays. A 20µl aliquot of each reaction mixture was blotted onto the filter paper. Samples 2-16 were from reaction mixtures containing extracts from different GUS positive plants containing the component 4 construct, samples 17-24 were from plants transformed with the component 2 construct (all GUS negative), and sample 1 was from untransformed tobacco. The negative control (-ve ctrl) was from a reaction mixture containing no plant extract.

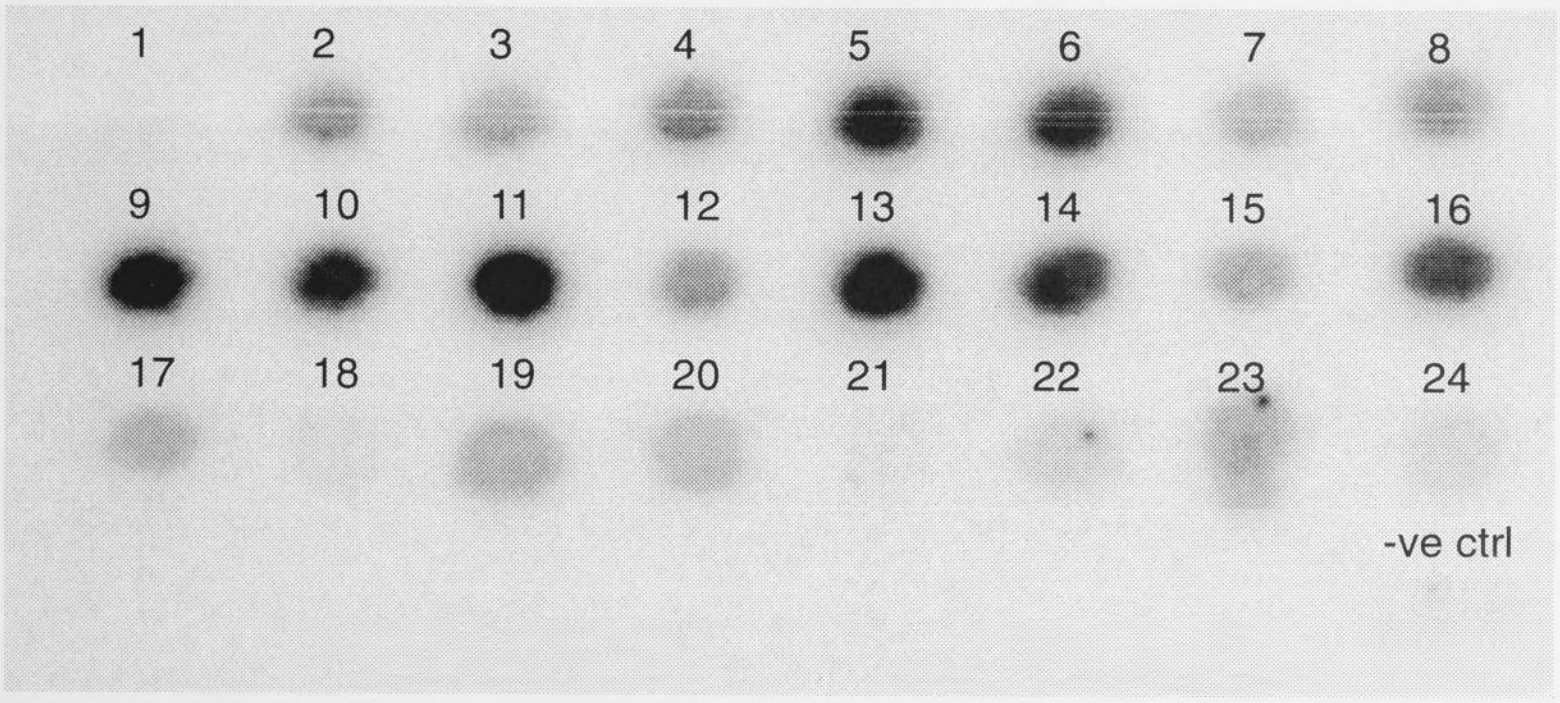
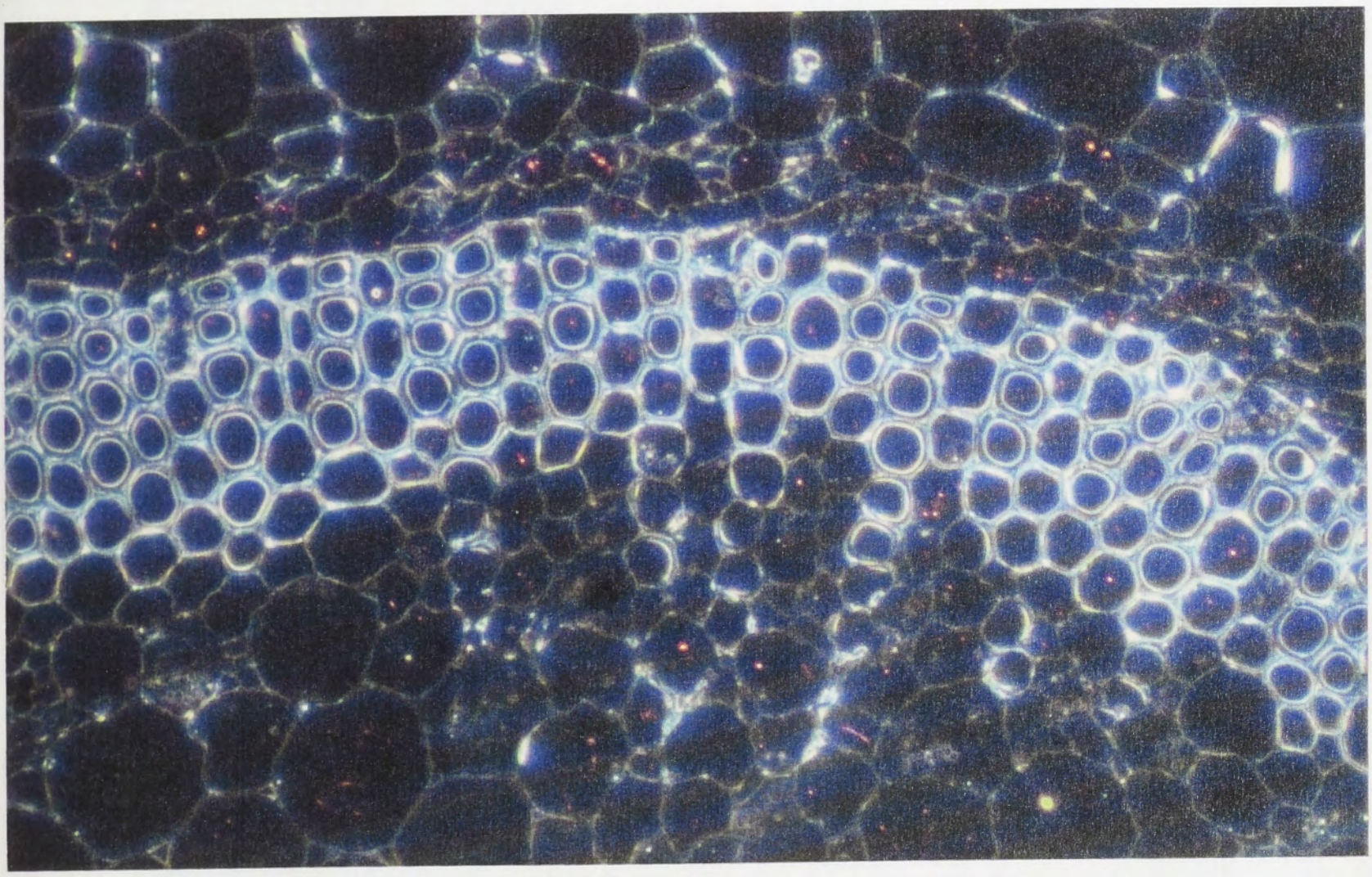


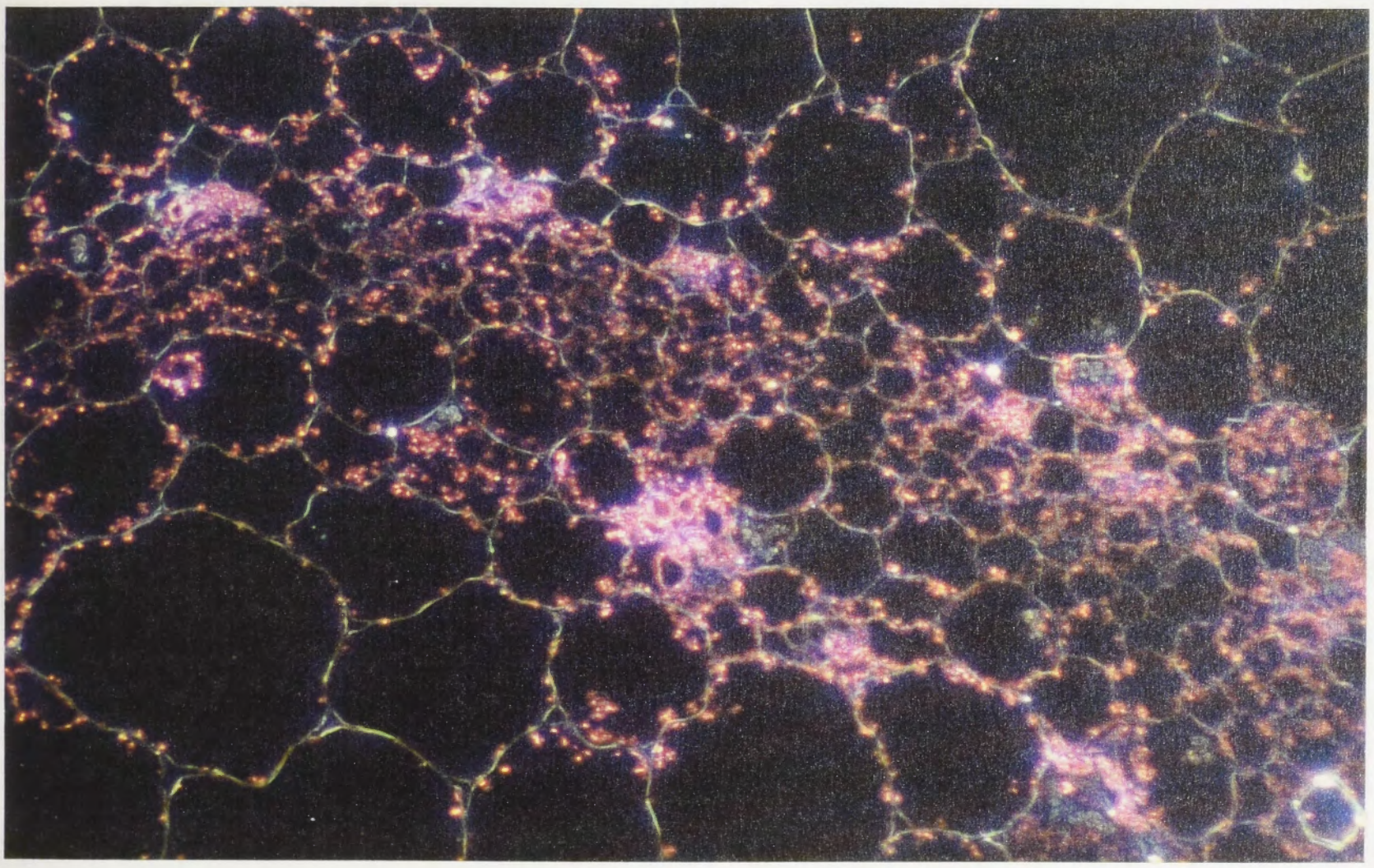
Figure 7.6

Histochemical staining for GUS activity in old and young transgenic plants - dark field

Transverse thin sections of stained, embedded stem pieces from an old T_0 tobacco plant (a) and a young T_1 plant (b) transformed with the GUS fusion constructs containing the component 3 and 7 promoter regions respectively, viewed with a dark field. Pink crystals indicate GUS expression. The magnifications are about 480X in (a) and 375X in (b).



a.



b.

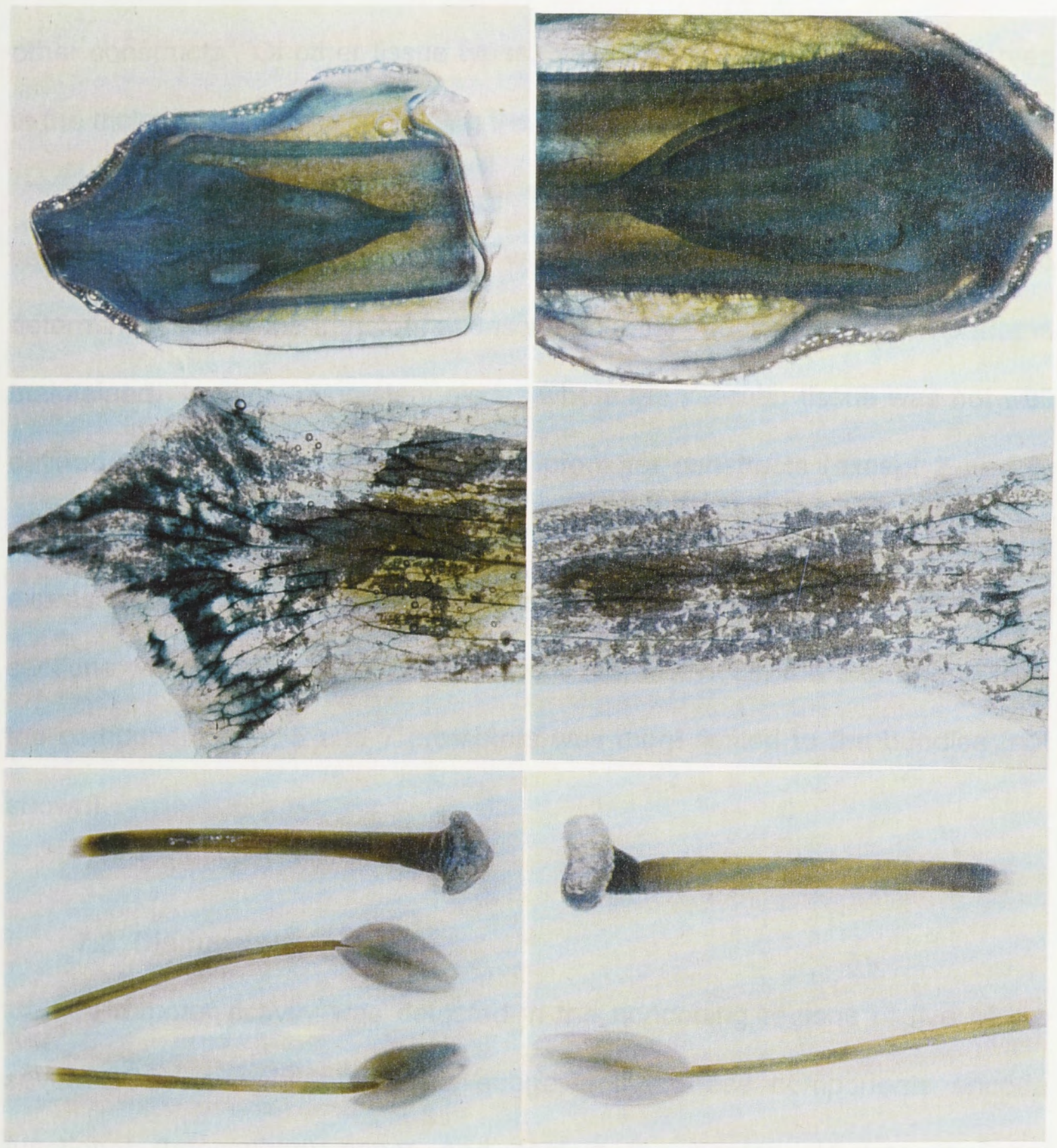


Figure 7.7

Histochemical staining for GUS activity in flower parts of transgenic plants

Stained flower parts (flower body - top; petal - centre; and sigma and stamen - bottom) from tobacco plants transformed with the GUS fusion constructs containing the component 4 (left column) and 5 (right column) promoter regions, viewed with a bright field. Blue colouration indicates GUS expression. The flower bodies (top) were cut in half to allow penetration of the stain.

other constructs. Of other tissue types, staining was also commonly observed in the trichomes of plants containing the component 4, 5, and 7 constructs.

Some seeds from two lines of each construct were planted and stem sections from these (T_1) seedlings were taken and stained with X-gluc to determine whether the transformations were stable and the expression patterns maintained. In very young stem pieces where the vascular tissue was not well defined, the expression from all of the promoter constructs (except 2 and 6) was particularly strong in the region of the developing vascular cells and expression was seen throughout the stem (Figure 7.6b). In slightly older stem sections where the vascular bundles were more obvious, the expression from the component 1, 3, 5 and 7 promoters was more limited to the bundles (not shown).

7.3 Discussion

Promoter activity was detected in the noncoding regions of five of the seven SCSV components. This suggests that these components encode functional (transcribed) genes and contain the appropriate transcriptional signals. The expression driven by the promoters of components 1, 3, 5, and 7 was predominantly in the vascular tissue of the transformed tobacco. The component 4 promoter appeared to give constitutive expression, although expression in the vascular tissue was much higher than in other tissues. This has also been observed for the constitutive CaMV 35S promoter, and was suggested either to be due to the greater density of the vascular tissue, or to indicate greater metabolic activity of vascular cells (Jefferson *et al.*, 1987).

The expression patterns were not completely consistent between different transgenic lines of the same construct. At least one plant from each transformation gave apparently constitutive expression, and some plants gave unusually low expression. The most commonly observed expression patterns were therefore the ones described. Variation in expression patterns between transgenic lines have also been observed with the 35S promoter (for example, Hanley-Bowdoin *et al.*, 1989), and are thought to be due to different effects of the different positions of insertion of the construct into the genome in independent transformants.

The vascular dominant expression of the SCSV promoters is consistent with the apparent vascular localisation of the virus. Although promoters are generally not host specific, it remains to be seen whether these expression patterns also occur in legumes. The CaMV 35S promoter is expressed in monocotyledonous plants although CaMV is limited to dicotyledonous hosts (Dekeyser *et al.*, 1990; Terada and Shimamoto, 1990). Similarly, the promoters of CFDV and commelina yellow mottle virus (CoYMV) are active in tobacco although these viruses infect the monocotyledonous plants, coconut (Randles *et al.*, 1992; Rhode *et al.*, 1995) and *Commelina diffusa* (Lockhart, 1990; Medberry *et al.*, 1992) respectively.

The expression of the SCSV component 1, 3, 5, and 7 promoters appeared to become progressively more limited to the vascular tissue as the plants aged. In the stem sections of young T_1 seedlings, GUS activity was apparent throughout the stems (Figure 7.6). In sections from slightly older T_1 seedlings, the GUS activity was reduced in non-vascular cells (not shown). In the T_0 plants which were several weeks old the GUS activity was almost

completely limited to the vascular tissue. In the old, woody plants, there was little GUS activity. A similar progressive vascular limitation was observed for the CoYMV promoter (Medberry *et al.*, 1992). The expression of these promoters may therefore be linked to the replicative or transcriptional activity of the cells. In the very young stems the majority of the cells would be actively growing and dividing, and hence be transcriptionally active, whereas in the old woody tissue many of the cells are inactive (for review see Jensen and Salisbury, 1984). The dominance of vascular expression therefore, may be a reflection of the greater metabolic activity of the phloem associated cells compared to non-vascular cells.

From the fluorimetric assays of GUS activity (Figure 7.4) it is obvious that the component 4 promoter gives the highest GUS expression. This may be due in part to the higher level of expression from the component 4 promoter in non-vascular cells compared to the other promoters. Alternatively, the higher expression of the component 4 promoter may have allowed ready detection of GUS activity in non-vascular cells. The strong expression of the component 4 promoter may indicate that this is an early gene, as was suggested for the RAP gene of ACMV (Zhan *et al.*, 1991; Haley *et al.*, 1992). Of the other promoters, that of component 5 (the coat protein encoding component) appeared to be the strongest although only two plants were sampled. The expression levels of the component 1, 3, 5 and 7 promoters were comparable to those of the phloem specific *rolC* promoter in transgenic tobacco (Schmülling *et al.*, 1989; Sugaya *et al.*, 1989). The levels of expression of the constitutive component 4 promoter were low compared to the commonly used, constitutive CaMV 35S promoter in tobacco; however, the expression levels of the SCSV promoters may be higher

in legumes than in tobacco. In legume protoplasts, expression of the component 5 and 7 promoters was higher than the corresponding expression levels in tobacco protoplasts, whereas the activity of the 35S promoter is lower in legume protoplasts compared to tobacco (B. Surin, unpublished).

Studies in tobacco protoplasts have indicated that the coat protein promoter and the component 7 promoter of SCSV are transactivated about 2 fold by the RAP encoded by component 2 (the component 6 RAP was not tested; B. Surin, unpublished). This is similar to the transactivation described for the monopartite geminivirus CSMV (Zhan *et al.*, 1993). SCSV components 5 and 7 may therefore be late genes.

The component 2 and 6 constructs gave no detectable expression when tested in transgenic tobacco. The noncoding regions of these components are only 179 and 159 nucleotides long respectively, and it is likely that some of the promoter elements are within the coding region. Subsequent work by B. Surin has shown that this is indeed the case for component 2. A fragment of component 2 from nucleotides 823 to 46 fused to GUS gave similar levels of GUS activity to the component 7 promoter in tobacco protoplasts. Given the small size of the SCSV components, the other components (besides 2 and 6) may also have promoter elements within their ORFs.

CHAPTER 8: GENERAL DISCUSSION

The aim of this thesis was to increase our knowledge of the molecular biology of SCSV. Sequencing seven DNA components of the SCSV F isolate defined the genome structure of SCSV (Figure 8.1) and confirmed its close relationship to other plant ssDNA viruses. The investigation of SCSV isolates revealed that its sequence variability may be similar to that found with RNA viruses, and indicated that natural variation may be an important consideration for the development of virus derived resistance to SCSV. The analysis of the SCSV promoter regions indicated that SCSV components 1, 2, 3, 4, 5, and 7 encode functional genes, and that the promoter sequences of these components are functional. The promoters were found to generally drive vascular expression, which is consistent with the apparent vascular localisation and obligate aphid transmission of the virus. In this Chapter I will discuss the conclusions, consequences, and questions which arise from this new information.

8.1 Defining SCSV-like viruses

8.1.1 The SCSV genus

On the basis of their similar particle structure, nucleic acid structure, genome organisation, and amino acid sequences, I propose that SCSV, BBTV, FBNYV, and CFDV form a distinct genus which probably does include MDV, and may include PCV, but which does not include CAV. SCSV, BBTV, FBNYV, and CFDV have small isometric particles and genomes consisting of several circular ssDNA components each about 1kb in size (Burns, 1994; L. Katul, pers.

Figure 8.1

SCSV Genome Organisation

Summary of the genome organisation of the F isolate of SCSV. The sequences are represented linearly with the first nucleotide of the conserved nonanucleotide in the stem-loop sequence (green circle) at the left (as in Figures 3.1 and 4.1). The minimal common region sequences are represented by the red rectangles and the main ORF in each sequence is represented by the blue rectangle. The positions of the predicted TATA box and polyadenylation signal sequences are indicated by the green and black triangles respectively. The predicted hydrophobic region in the putative product encoded by component 1 is represented by the yellow region in the component 1 ORF. The putative binding/nicking and helicase domains (see Chapter 3) in the putative RAP ORFs are indicated by the green and pink regions respectively.


Component
Length (bases)
1
112
1001
5'
3'
2
RAP
280
1022
3
164
991
4
153
1002
5
coat
169
998
6
RAP
285
1017
7
146
988
common region
open reading frame
hydrophobic region
'binding/nicking' domain
TATA box
polyadenylation signal
conserved stem-loop sequence
'helicase' domain

comm.). The putative RAPs share greater than 35% identity. SCSV and BBTV have a second related ORF, encoded by components 4 and 6 respectively (Chapter 4; Burns, 1994). Each of the viral DNA components has a sequence capable of forming a strong stem-loop structure containing a conserved nonanucleotide similar to that found in the geminiviruses. Except for CFDV, each DNA component appears to encode a single ORF in the virion sense.

The CFDV DNA which has been sequenced has the potential to encode several ORFs (Figure 1.2). An ORF of 153 amino acids (ORF 2) is encoded within the main ORF (ORF 1) in a different reading frame. ORF 2 is not flanked by consensus transcription signals but could be translated from the main transcript by ribosomal frameshifting. Following the main ORF there is a 62 amino acid ORF in a different reading frame (ORF 4). This ORF is not preceded by an obvious TATA box, however, there is an open region (contains no stop codons) of 56 amino acids preceding it which overlaps the main ORF. ORFs 5 and 6 in the complementary sense were not flanked by obvious transcription signals, therefore CFDV transcription may be unidirectional. Of the ORFs encoded by the CFDV DNA, none appears to encode an amino acid sequence related to the SCSV coat protein. The CFDV genome, therefore, may contain at least one more component and, by analogy with the other viruses of this group, may contain several more.

PCV has a similar particle size to SCSV and the DNA encodes a potential stem-loop of the SCSV/geminivirus type; however, the single infectious DNA component encodes multiple ORFs which are transcribed bidirectionally (Tisher *et al.*, 1982; Mankertz *et al.*, 1993). PCV has been assigned to the same genus as CAV (Murphy *et al.*, 1995). Determination of

whether PCV is included in the same genus as the SCSV-like plant viruses, as I have suggested, rather than the CAV genus, awaits the release of the PCV sequence to confirm the reported similarity between the putative RAP of CFDV and one of the PCV ORFs (Rohde *et al.,* 1990). The CAV particle is composed of protein subunits more than twice the size of the SCSV coat protein (50K as opposed to 19K) (Todd *et al.*, 1990) and the encapsidated DNA is negative sense (Noteborn *et al.*, 1991). The sequence of the single, infectious CAV DNA shares no similarity with SCSV-like viruses and does not encode a potential stem-loop of the SCSV/geminivirus type (Noteborn *et al.*, 1991). Instead, the noncoding region of the CAV DNA has an unusual series of five 21b direct repeats and several stretches of G or C greater than 5b in length (Noteborn *et al.*, 1991).

8.1.2 Defining the virus species

It may prove difficult to define what constitutes a distinct virus species in the SCSV genus (which I have referred to as the SCSV-like viruses). The large number of DNA components in the SCSV genome and the lack of a manipulable infection system has meant that the minimum number of components required to cause an infection is unknown. A similar situation exists with BBTV. Some minimal requirements would be: an RAP (either component 2 or 6 in SCSV), the coat protein (SCSV component 5), at least one movement protein (SCSV component 1 is a possible candidate), and an aphid transmission factor. As discussed in Chapter 4, the SCSV coat protein may be multifunctional, like those of monopartite geminiviruses, and may have aphid transmission and movement functions. The monopartite geminiviruses, however, have a specific movement protein in addition to the coat protein

(Chapter 1). Determining what are the essential components in the SCSV genome is complicated by the existence of apparently redundant RAP encoding components and the co-existence of the different strains of component 5 (F and JL3).

8.2 Origins and functions of components 2 and 6

It appears that components 2 and 6 have come from different viruses in the SCSV-like group rather than having diverged from a common ancestor. This is the simplest way to explain why their RAPs are as closely related to each other as to the putative RAPs of other SCSV-like viruses, and why there is a nucleotide difference in the highly conserved nonanucleotide loop sequence; which is 5'**T**AGTATTAC3' in component 2 and 5'**C**ATGATTAC3' in component 6. The only other difference in the nonanucleotides of the SCSV-like viruses is a T instead of a G at position 3 in BBTV.

What is striking, however, is that SCSV maintains at least two functionally equivalent genes. The conservation of replication associated motifs in both the N- and C-terminal ends of the proteins (Chapter 3) suggests that they may have the same function. One explanation for their maintenance may be that the RAPs from components 2 and 6 have slightly different functions or may be more advantageous in different hosts. Alternatively, one of the RAP DNAs may be acting as a parasitic satellite DNA, which either relies entirely on the host virus, or encodes an RAP which directs replication of only its own component, and relies on its host for other functions, such as movement, encapsidation, and aphid transmission. This last possibility seems unlikely given that some SCSV isolates appear to have only one of the two RAP

encoding components, although, as discussed in Chapter 3, there may be a third RAP encoding component.

Further questions about the origins and functions of components 2 and 6 arise from the lack of any common region type sequence in these components. As mentioned in Chapter 7, the noncoding regions of these components are only 179 and 159 nucleotides long respectively. When the polyadenylation signals and the sequence downstream of the TATA boxes are excluded, about 50 and 90 nucleotides are left, which is significantly less than the minimal common region of 153b in the other SCSV components (Chapter 4). The BBTV minimal common region, however, is only 80b long, and the geminivirus RAP binding site in TGMV is only 13b long (Fontes *et al.*, 1994), therefore it is conceivable that SCSV components 2 and 6 could support a much reduced common region consisting of little more than an RAP binding site. The absence of any common region in SCSV components 2 and 6 may simply reflect size constraints on the SCSV components due to packaging, or indicate that the regulation of their replication is different from the other DNAs. It is interesting with regard to the regulation of replication that the proposed TATA boxes of SCSV components 2 and 6 are upstream of the stem-loop whereas those of the other components are downstream. TATA binding proteins were suggested to be involved in the initiation of geminivirus replication (Argüello-Astorga *et al.*, 1994).

The BBTV component 1 (RAP) has a 286 nucleotide noncoding region which contains both the 80b and the 50b common region sequences, although they are separated by fewer nucleotides than in the other BBTV DNAs. A PILEUP of the noncoding regions of the 6 standard BBTV DNAs (Burns, 1994)

and 2 'satellite' DNAs (from M. Karan) shows that the 'satellite' DNAs (encoding putative RAPs) appear to have sequences related to the 80b common region but not the 50b common region surrounding the stem-loop (results not shown). As all isolates of BBTV have the component 1 (M. Karan, pers. comm.), the other RAP DNAs may indeed be satellites, which came from another SCSV-like virus and have obtained the 80b common region by recombination (which may allow them to be maintained by the BBTV RAP). It is suggested that the BBTV 'satellites' have come from another SCSV-like virus because there is more similarity between the putative RAPs encoded by the satellites and that of SCSV component 2, than between the satellite RAPs and the BBTV component 1 RAP (Chapter 3). This evidence lends weight to the suggestions that one of the two SCSV RAP type components may also have come from a different SCSV-like virus and may be a satellite.

8.3 SCSV variation

The structure of the SCSV genome and the sequence variation uncovered in Chapter 5 indicate that molecular variation of SCSV isolates may occur at three levels: the presence or absence of different components, the presence or absence of different strains of each component, and intra-strain sequence variation.

8.3.1 Component mixture

From the analysis of the occurrence of components 2 and 6 it appears that SCSV can vary in the types of components present. The presence of components other than 2 and 6 may also vary. If there is a distinct component required for aphid transmission, for example, then it is probably absent in the

glasshouse isolates A2 and B1, which have lost their aphid transmissibility. Furthermore, some initial cloning of SCSV DNA directly from field isolates indicated that components other than the seven in the F isolate exist. The presence of non-essential or redundant components would be expected to vary in natural infections.

8.3.2 Component strains

The analysis of sequence variation of component 5 revealed that there are distinct strains of that component. There may be distinct strains of other, if not all of the components. If all strains can co-exist, like F and JL3, and be replicated by any RAP, then any one infection of SCSV is likely to be a complex mixture of component types as well as strains of those components. This would allow the evolution of many different genome arrangements by reassortment and recombination.

8.3.3 Component sequences

Nucleotide substitutions, insertions, and deletions were found between clones from different isolates of both strains of component 5. No two sequences from different plant samples were completely identical and furthermore, any two clones from one plant were seldom identical. The two clones of component 5 from the JL3 isolate, for example, differed by 2 nucleotides (see Section 5.2.2). Limited information from the sequencing of the F isolate, and also of component 2 sequence fragments from a few different isolates, indicated that a few percent variation probably occurs between isolates and clones of all components. SCSV therefore, appears to exist as a quasispecies similar to RNA viruses (Domingo *et al.*, 1985). A quasispecies was defined as a heterogeneous mixture of related genomes which share a

consensus sequence, but all individual genomes differ by at least one nucleotide (Domingo *et al.*, 1985). Domingo *et al.* (1985) found that a second feature of viruses, such as Qβ, which exist as quasispecies populations, is that they rapidly attain the quasispecies state due to their high mutation rate and large population size. Most RNA viruses have since been suggested to be quasispecies on the basis of the first property alone, as I have done for SCSV. The term quasispecies does not appear to have been applied to DNA viruses. DNA viruses are generally thought to be more highly conserved than RNA viruses because DNA polymerases have proofreading capabilities whereas RNA polymerases do not. The apparent quasispecies nature of SCSV may be due to; the multi-component nature of the SCSV genome, a large pool of variability in infected weed plants surrounding infected fields, increased mutation rates, or low selection pressures. The mechanisms which allow the maintenance of the several essential genomic components may also allow the maintenance of a number of non-essential, redundant, and defective DNAs. The significance of a quasispecies population is that, unlike a more conserved population, it can rapidly respond to environmental change, because in any quasispecies population there is a greater range of variants, which is more likely to include sequences that are advantageous in the new conditions.

8.4 SCSV-like viruses and the geminiviruses

The SCSV-like viruses and the geminiviruses appear to share a common ancestor. A number of similarities between these groups have been identified in this thesis. In summary:

- They have small, single-stranded, circular DNA genomes.

- They encapsidate the positive strand of the DNA in small particles composed of a single type of protein subunit.
- They have related RAPs and similar stem-loop structures, with a conserved nonanucleotide sequence.
- SCSV and subgroup I geminiviruses encapsidate oligonucleotide primers required for second strand synthesis.
- SCSV, BBTV, and bipartite geminiviruses have common regions, which are highly conserved between components from the same species.
- CFDV and subgroup I and II geminiviruses are transmitted by related insect vectors.

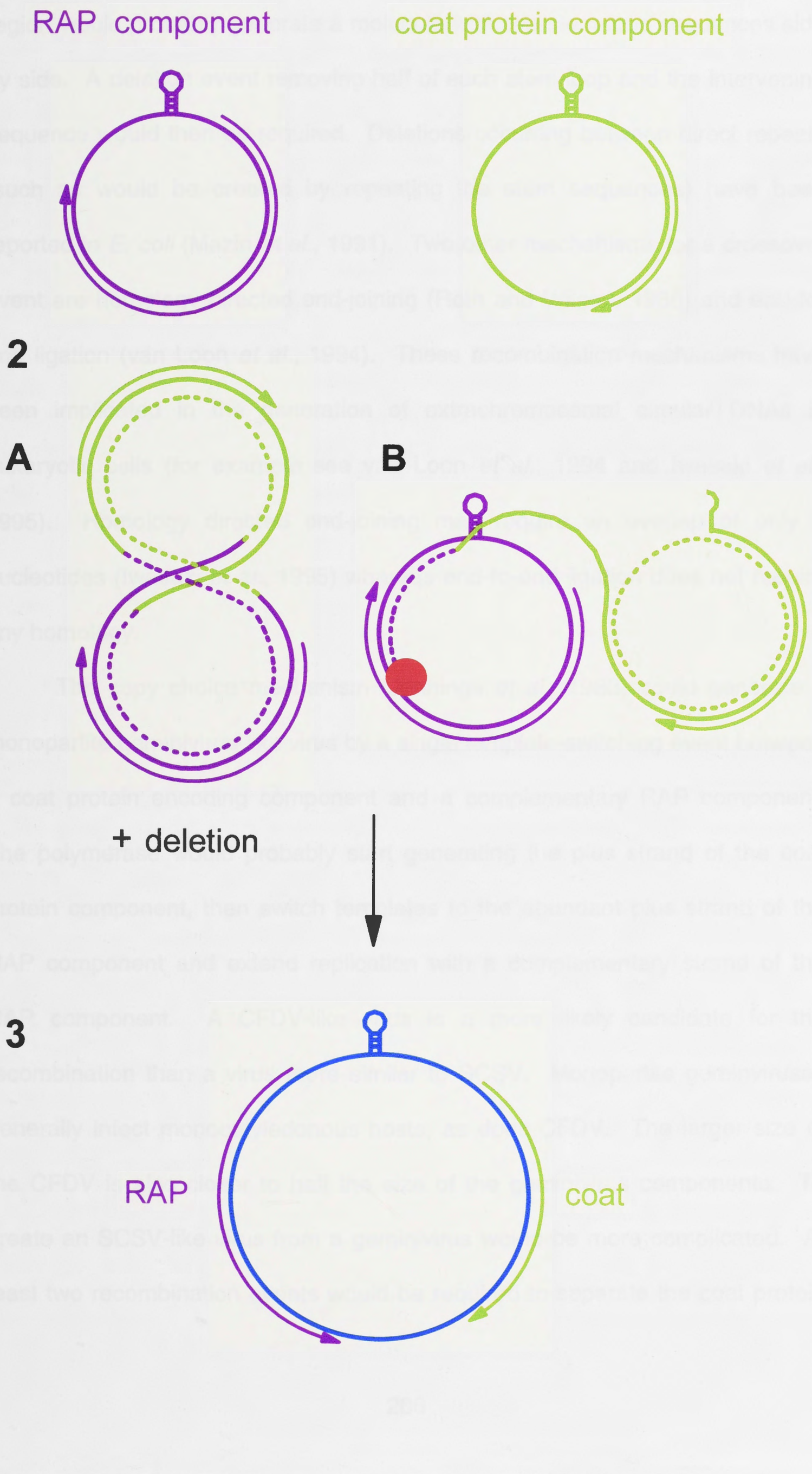
I propose that the common ancestor of the SCSV-like viruses and the geminiviruses was more like SCSV in genome size and structure. The geminivirus particle is highly unusual. Most viruses have icosahedral/spherical or rod-like particles. The geminivirus particle looks like the partial fusion of two icosahedra. The geminivirus coat protein is capable of forming simple icosahedra, which have been found in ACMV and CSMV preparations, as well as multimers (Stanley, 1985). The geminate particle structure, therefore, appears to be a derivation of a basic icosahedral particle and may have arisen in response to a doubling in genome component size. A possible scenario for this doubling in genome size is illustrated in Figure 8.2. Recombination between the negative strand of an RAP encoding component, and the positive strand of a coat protein encoding component, centred around the stem-loop, could give rise to a component similar to that of a monopartite geminivirus. Two possible recombination mechanisms, a single crossover and template switching, are presented in Figure 8.2. The homologous recombination type

Figure 8.2

Evolution of the geminiviruses

SCSV-like viruses encode their RAPs (purple) and coat proteins (green) on separate components (1). A recombination event between these DNAs (2) could produce a DNA similar to a monopartite geminivirus (3), if the recombination is such that the relative directions of the ORFs are changed. Two possible mechanisms for this are represented. A single crossover event, depicted in A, could result in reversal of the RAP ORF because of the circularity of the DNAs. The crossover event would have to then be followed by a deletion of one of the stem-loop sequences. Deletions flanked by direct repeats, such as would occur with the repetition of the stem sequences, have been reported (Mazin *et al.*, 1991). Template switching (B) (Jennings *et al.*, 1983), in which the polymerase complex (red dot) swaps from replicating the positive strand of the coat protein component to replicating the negative strand of the RAP component. The positive strand template of the RAP component would need to be in close proximity to the coat protein component for the polymerase to transfer.

The virion sense strands are represented by filled lines, and the complementary sense strands are represented by dotted lines. The ORFs are represented by arrows. The positions of the stem-loop sequences are represented by the circle and ladder symbol.


1
RAP component
coat protein component
2
A
B
+ deletion
3
RAP
coat

single crossover event at self-complementary sequences of the stem-loop region would probably generate a molecule with two stem-loop sequences side by side. A deletion event removing half of each stem-loop and the intervening sequence would then be required. Deletions occurring between direct repeats (such as would be created by repeating the stem sequences) have been reported in *E. coli* (Mazin *et al.*, 1991). Two other mechanisms for a crossover event are homology directed end-joining (Roth and Wilson, 1986) and end-to-end ligation (van Loon *et al.*, 1994). These recombination mechanisms have been implicated in the generation of extrachromosomal circular DNAs in eukaryotic cells (for example see van Loon *et al.*, 1994 and Iwasaki *et al.*, 1995). Homology directed end-joining may require an overlap of only 3 nucleotides (Iwasaki *et al.*, 1995) whereas end-to-end ligation does not require any homology.

The copy choice mechanism (Jennings *et al.*, 1983) could generate a monopartite geminivirus-like virus by a single template-switching event between a coat protein encoding component and a complementary RAP component. The polymerase would probably start generating the plus strand of the coat protein component, then switch templates to the abundant plus strand of the RAP component and extend replication with a complementary strand of the RAP component. A CFDV-like virus is a more likely candidate for the recombination than a virus more similar to SCSV. Monopartite geminviruses generally infect monocotyledonous hosts, as does CFDV. The larger size of the CFDV is also closer to half the size of the geminivirus components. To create an SCSV-like virus from a geminivirus would be more complicated. At least two recombination events would be required to separate the coat protein

and RAP ORFs to ensure the resultant components both have stem-loop sequences. The orientation of the encapsidated DNA would also have to be reversed in one of the components to achieve unidirectional transcription. Furthermore, several new components would probably have to be created (although the number of components in CFDV is unknown).

Sequence comparisons have revealed abundant evidence of recombination in viruses. In geminiviruses, BCTV appears to be a recombinant between subgroup I and III viruses because its complementary sense genes are more closely related to those of the subgroup III viruses, whereas its coat protein is more closely related to those of subgroup I viruses (Howarth and Vandermark, 1989). Other geminiviruses which show evidence of recombination include SqLCV and pepper huasteco virus (PHV) (Torres-Pacheco *et al.*, 1993; Rybicki, 1994). PHV appears to derive its RAP from Old World subgroup III geminiviruses, while the rest of its genome is related to New World subgroup III geminiviruses. SqLCV appears to derive its common region and RAP from within one group of New World viruses and the rest of its genome from another New World virus. Recombination between the geminiviral DNAs appears to be very efficient. Lethal mutations in the stem-loop sequence of ACMV DNA B were corrected by recombination with co-inoculated DNA A, and in the reciprocal experiment, a stem-loop deletion mutant of DNA A was corrected by recombination with co-inoculated DNA B (Roberts and Stanley, 1994). The recombinations correcting the point mutations in the B component stem-loop were rapid enough to result in wild type infectivity with no delay or attenuation of symptoms. Other experiments involving DNA A deletion mutants of ACMV also found rapid reversion of the

mutants to wild-type size by recombination during infection (Etessami *et al.*, 1989; Klinkenberg *et al.*, 1989). Analysis of these size revertants indicated that the 3' end of the conserved nonanucleotide, which is the predicted site of nicking (Heyraud *et al.*, 1993), may be a hot spot for recombination (Etessami *et al.*, 1989). This supports the suggestion above, that the recombination event which may have produced a gemini-type virus from an SCSV-like virus may be centred around the stem-loop sequence.

Rybicki (1994) proposed that the subgroup I and III geminiviruses diverged from a common ancestor more than 80 million years ago. The greater sequence divergence between subgroup I viruses and their greater range of vector species, may indicate that the subgroup I viruses (monopartite) are the ancestral type. This would be predicted from the scenario outlined above, for the evolution of a monopartite geminivirus-like virus from an SCSV-like virus. The alternative explanation for the lower level of diversity and single vector species of the subgroup III geminiviruses is that this group has been through an evolutionary bottleneck (Rybicki, 1994).

The ancestor of both the geminiviruses and the SCSV-like viruses has been suggested to be a bacteriophage or plasmid (Koonin and Ilyina, 1992; Rybicki, 1994). This is due to the similarity between their RAPs, the similarity between the nonanucleotide initiation sequence and the gene A recognition and cleavage sequence of øX174, and the fact that geminiviruses use rolling circle replication (see Chapters 1 and 3).

REFERENCES

Accotto, G.P., Donson, J., and Mullineaux, P.M. (1989). Mapping of *Digitaria* streak virus transcripts reveals different RNA species from the same transcription unit. *EMBO J.* 8, 1033-1039.

Agrios, G. (1988). Plant pathology, 3rd edition. Academic Press, California.

An, G., Watson, B.D., Satchel, S., Gordon, M.P., and Nester, E.W. (1985). New cloning vehicles for transformation of higher plants. *EMBO J.* 4(2), 277-284.

Andersen, A.S., Pettersson, A.F., and Kjeldsen, T.B (1992). A fast and simple technique for sequencing plasmid DNA with Sequenase using heat denaturation. *BioTechniques* 13(5), 678-679.

Anderson, E.J., Qui, S.G., and Schoelz, J.E. (1991). Genetic analysis of determinants of disease severity and virus concentrations in cauliflower mosaic virus. *Virology* 181, 647-655.

Argüello-Astorga, G., Herrera-Estrella, L., and Rivera-Bustamante, R. (1994). Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol. Biol.* 26, 553-556.

Audy, P., Palukaitis, P., Slack, S.A., and Zaitlin, M. (1994). Replicase mediated resistance to potato virus Y in transgenic tobacco plants. *Molec. Plant-Microb. Interactn.* 7, 15-22.

Azzam, O., Frazer. J., De La Rosa, D., Beaver, J.S., Ahlquist, P., and Maxwell, D.P. (1994). Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. *Virology* 204, 289-296.

Banerjee, N., Wang, J-Y., and Zaitlin, M. (1995). A single nucleotide change in the coat protein of tobacco mosaic virus is involved in the induction of severe chlorosis. *Virology* 207, 234-239.

Bechtold, N., Ellis, J., and Pelletier, G. (1993). *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Paris, Life Sci.* 316, 1194-1199.

Boevink, P (1991). Honours thesis. The Australian National University, Canberra, Australia.

Boulton, M.I., Buchholz, W.G., Marks, M.S., Markham, P.G., and Davies, J.W. (1989). Specificity of *Agrobacterium*-mediated delivery of maize streak virus DNA to members of the Gramineae. *Plant Mol. Biol.* 12, 31-40.

Boulton, M.I., King, D.I., Markham, P.G., Pinner, M.S., and Davies, J.W. (1991a). Host range and symptoms are determined by specific domains of the maize streak virus genome. *Virology* 181, 312-318.

Boulton, M.I., King, D.I., Donson, J., and Davies, J.W. (1991b). Point substitutions in a promoter-like region and the V1 gene affect the host range and symptoms of maize streak virus. *Virology* 183, 114-121.

Boulton, M.I., Pallaghy, C.K., Chatani, M., MacFarlane, S., and Davies, J.W. (1993). Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology* 192, 85-93.

Bradford, M.M. (1976). A rapid and sensitive method for quantification of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254.

Briddon, R.W., Watts, J., Markham, P.G., and Stanley, J. (1989). The coat protein of beet curly top virus is essential for infectivity. *Virology* 172, 628-633.

Briddon, R.W., Pinner, M.S., Stanley, J., and Markham, P.G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* 177, 85-94.

Briddon, R.W., Luness, P., Chamberlin, L.C.L., and Markham, P.G. (1994). Analysis of the genetic variability of maize streak virus. *Virus Genes* 9, 93-100.

Brough, C.L., Hayes, R.J., Morgan, A.J., Coutts, R.H.A., and Buck, K.W. (1988). Effects of mutagenisis *in vitro* on the ability of cloned tomato golden mosaic virus DNA to infect *Nicotiana benthamiana* plants. *J. Gen. Virol.* 69, 503-514.

Brough, C.L., Sunter, G., Gardiner, W.E., and Bisaro, D.M. (1992). Kinetics of tomato golden mosaic virus DNA replication and coat protein promoter activity in *Nicotiana tabacum* protoplasts. *Virology* 187, 1-9.

Burns, T. (1994). PhD thesis. Queensland University of Technology, Brisbane, Australia.

Burns, T.M., Harding, R.M., Hafner, G., Beetham, P., and Dale, J.L. (1993). Single-stranded DNA genome organisation of banana bunchy top virus. IXth International Congress of Virology, Abstract W62-8.

Chapman, S., Kavanagh, T., and Baulcombe, D. (1992). Potato virus X as a vector for gene expression in plants. *Plant J.* 2, 549-557.

Cherif, C. and Russo, M. (1983). Cytological evidence of the association of a geminivirus with the tomato yellow leaf curl disease in Tunisia. *Phytopathol. Z.* 108, 221-225.

Choi, I-R. and Stenger, D. (1995). Strain-specific determinants of beet curly top geminivirus DNA replication. *Virology* 206, 904-912.

Chu, P.W.G., Francki, R.I.B., and Randles, J.W. (1983). Detection, isolation and characterisation of high molecular weight double-stranded RNAs in plants infected with velvet tobacco mottle virus. *Virology* 126, 480-492.

Chu, P.W.G., and Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. *Virology* 167, 38-49.

Chu, P.W.G., Keese, P., Qiu, B., Waterhouse, P.M., and Gerlach, W.L. (1993a). Putative full-length clones of subterranean clover stunt virus and identification of the component coding for the viral coat protein. *Virus Research* 27, 161-171.

Chu, P.W.G., Qui, B., Zhong-yi, L., and Larkin, P. (1993b). Replication of subterranean clover stunt virus in pea and subterranean clover protoplasts. *Virus Res.* 27, 173-183.

Chu, P.W.G., Boevink, P., Surin,B., Larkin,P., Keese,P., and Waterhouse, P. (1995). Non-geminated ssDNA Plant Viruses. In: Singh, R.P., Singh, U.S., and Kohmoto, K. (eds). Pathogenisis and Host Parasite Specificity in Plant Diseases. Vol. III, Viruses and Viroids. Pergamon Press, Oxford, U.K.

Citovsky, V., Zupan, J., Warnick, D., and Zambryski, P. (1992). Nuclear localisation of *Agrobacterium* VirE2 protein in plant cells. *Science* 256, 18802-18804.

Conway, L., and Wickens, M. (1985). A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc. Natl. Acad. Sci. USA* 82, 3949-3953.

Dale, J.L. (1987). Banana bunchy top: an economically important tropical plant disease. *Adv. Virus Res.* 33, 301-325.

Dale, J.L., Phillips, D.A., and Parry, J.N. (1986). Double-stranded RNA in bananas with bunchy-top disease. *J. Gen. Virol.* 67, 371-375.

Davies, J.W., and Stanley, J. (1989). Geminivirus genes and vectors. *Trends Genet.* 5, 77-81.

Dean, C., Tamaki, S., Dunsimuir, P., Favreau, M., Katayama, C., Dooner, H., and Bedbrook, J. (1986). mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucleic Acids Res.* 14, 2229-2240.

Dekeyser, R.A., Claes, B., De Rycke, R.M.U., Habets, M.E., and Van Montague, M.C. (1990). Transient gene expression in intact and organised rice tissues. *Plant Cell* 2, 591-602.

Dekker, E.L., Woolston, C.J., Xue, Y., Cox, B., and Mullineaux, P.M. (1991). Transcript mapping reveals different expression strategies for the bicistronic

RNAs of the geminivirus wheat dwarf virus. *Nucleic Acids Res.* 19, 4075-4081.

Del Sal, G., Manfioletti, G., and Schnider, C. (1989) The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *Biotechniques* 7, 514-158.

Deom, C.M., Schubert, K.R., Wolf, S., Holt, C.A., Lucas, W.J., and Beachy, R.N. (1990). Molecular characterisation and biological function of the movement protein of tobacco mosaic virus in transgenic plants. *Proc. Natl. Acad. Sci. USA* 87, 3284-3288.

Depicker, A., Satchel, S., Dhaese, P., Zambryski, P., and Goodman, H. (1982). Nopaline synthase: transcript mapping and DNA sequence. *J. Mol. Appl. Genet.* 1, 561-575.

Ditta, G., Stanfield, S., Corbin, D., and Helsinki, D.R. (1980). Broad host range cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium melilotti*. *Proc. Natl. Acad. Sci. USA* 77, 7347-7351.

Domingo, E., Martinez-Salaz, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., VandePol, S., Steinhaur, D., DePolo, N., and Holland, J. (1985). The quasispecies (extremely heterogeneuos) nature of viral genome populations: biological relevance - a review. *Gene* 40, 1-8.

Donson, J., Morris-Krsinich, B.A.M., Mullineaux, P.M., Boulton, M.J., and Davies, J.W. (1984). A putative primer for second-strand synthesis of maize streak virus is virion-associated. *EMBO J.* 3, 3069-3073.

Donson, J., Accotto, G.P., Boulton, M.I., Mullineaux, P.M., and Davies, J.W. (1987). The nucleotide sequence of a geminivirus from *Digitaria sanguinalis*. *Virology* 161, 160-169.

Dower, J.W., Miller, F.J., and Ragsdale, W.C. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16, 6127-6145.

Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M., and Rogers, S.G. (1988a). Genetic analysis of the tomato golden mosaic virus II. The product of the AL1 coding sequence is required for replication. *Nucleic Acids Res.* 16, 7043-7060.

Elmer, J.S., Sunter, G., Gardiner, W.E., Brand, L., Browning, C.K., Bisaro, D.M., and Rogers, S.G. (1988b). *Agrobacterium*-mediated inoculation of plants with tomato golden mosaic virus DNAs. *Plant Mol. Biol.* 10, 225-234.

Etessami, P., Callis, R., Ellwood, S., and Stanley, J. (1988). Delimination of essential genes of cassava latent virus DNA 2. *Nucleic Acids Res.* 16, 4811-4829.

Etessami, P., Watts, J., and Stanley, J. (1989). Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of *Nicotinia benthamiana*. *J. Gen. Virol.* 70, 277-289.

Etessami, P., Saunders, K., Watts, J., and Stanley, J. (1991). Mutational analysis of complementary-sense genes of African cassava mosaic virus DNA A. *J. Gen. Virol.* 72, 1005-1012.

Evans, D. and Jeske, H. (1993). DNA B facilitates, but is not essential for, the spread of abutilon mosaic virus in agroinoculated *Nicotiana benthamiana*. *Virology* 194, 752-757.

Faria, J.C., Gilbertson, R.L., Hanson, S.F., Morales, F.J., Ahlquist, P. Loniello, A.O., and Maxwell, D.P. (1994). Bean golden mosaic geminivirus type II isolates from the Dominican Republic and Guatemala: nuceotide sequences, infectious pseudorecombinants, and phylogenetic relationships. *Phtopathology* 84, 321-329.

Fontes, E.P.B., Luckow, V.A., and Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence specific DNA binding protein. *Plant Cell* 4, 597-608.

Fontes, E.P.B, Gladfelter, H.J., Schaffer, R.L., Petty, I.T.D., and Hanley-Bowdoin, L. (1994). Geminivirus replication origins have a modular organization. *Plant Cell* 6, 405-416.

Francki, R.I.B., Hatta, T., Boccardo, G., and Randles, J.W. (1980). The composition of chloris striate mosaic virus, a geminivirus. *Virology* 101, 233-241.

Frischmuth, S., Frischmuth, T., and Jeske, H. (1991). Transcript mapping of abutilon mosaic virus, a geminivirus. *Virology* 185, 596-604.

Frischmuth, T., Roberts, S., von Arnim, A., and Stanley, J. (1993). Specificity of bipartite geminivirus movement proteins. *Virology* 196, 666-673.

Gardiner, W.E., Sunter, G., Brand, L., Elmer, J.S., Rogers, S.G., and Bisaro, D.M. (1988). Genetic analysis of tomato golden mosaic: the coat protein is not required for systemic spread or symptom development. *EMBO J.* 7, 899-904.

Garfinkel, D.J. and Nester, E.W. (1980). *Agrobacterium tumefaciens* mutants affected in crown gall tumorigenesis and octopine catabolism. *J. Bacteriol.* 144, 732-743.

Genetics Computer Group (1991). Program Manual for the GCG Package, Version 7. Madison, Wisconsin, USA.

Gil, A. and Proudfoot, N.J. (1984). A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3'-end formation. *Nature* (London) 312, 473-474.

Gilbertson, R.L., Faria, J.C., Hanson, S.F., Morales, F.J., Ahlquist, P., Maxwell, D.P., and Russell, D.R. (1991). Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. *Phytoathology* 81, 980-985.

Gilbertson, R.L., Faria, J.C., Ahlquist, P., and Maxwell, D.P. (1993). Genetic diversity in geminiviruses causing bean golden mosaic disease: the nucleotide sequence of the infectious cloned DNA components of a Brazillian isolate of bean golden mosaic geminivirus. *Phytopathology* 83, 709-715.

Goodman, R.M., Shock, T.L., Haber, S., Browning, K.S., and Bowers, G.R. (1980). The composition of bean golden mosaic virus and its single-stranded DNA genome. *Virology* 106, 168-172.

Gough, J.A. and Murray, N.E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* 166, 1-19.

Gorbalenya, A.E., Koonin, E.V., and Wolf, Y.I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* 262, 145-148.

Grimsley, N., Hohn, B., Hohn, T., and Walden, R. (1986). "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. *Proc. Natl. Acad. Sci. USA* 83, 3282-3286.

Grimsley, N., Hohn, T., Davies, J.W., and Hohn, B. (1987). *Agrobacterium*-mediated delivery of infectious maize streak virus into maize plants. *Nature* 325, 177-179.

Gröning, B.R., Hayes, R.J., and Buck, K.W. (1994). Simultaneous regulation of tomato golden mosaic virus coat protein and AL1 gene expression: expression of the AL4 gene may contribute to suppression of the AL1 gene. *J. Gen. Virol.* 75, 721-726.

Grylls, N.E., and Butler, F.C. (1956). An aphid transmitted virus affecting subterranean clover. *J. Aust. Inst. Agric. Sci.* 22, 73-74.

Grylls, N.E., and Butler, F.C. (1959). Subterranean clover stunt, a virus disease of pasture legumes. *Aust. J. Agric. Res.* 10, 145-159.

Guerineau, F., Brooks, L., Meadows, J., Lucy, A., Robinson, C., and Mullineaux, P. (1990). Sulfonamide resistance gene for plant transformation. *Plant Mol. Biol.* 15, 127-136.

Haley, A., Zhan, X., Richardson, K., Head, K., and Morris, B. (1992). Regulation of the activities of African cassava mosaic virus promoters by the AC1, AC2, and AC3 gene products. *Virology* 188, 905-909.

Hamilton, W.D.O., Bisaro, D.M., Coutts, R.H.A., and Buck, K.W. (1983). Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned components. *Nucleic Acids Res.* 21, 7387-7396.

Hamilton, W.D.O., Stein, V.E., Coutts, R.H.A., and Buck, K.W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences. *EMBO J.* 3, 2197-2205.

Hanley-Bowdoin, L., Elmer, J.S., and Rogers, S.G. (1989). Functional expression of the leftward open reading frames of the A component of tomato golden mosaic virus in transgenic tobacco plants. *Plant Cell* 1, 1057-1067.

Hanley-Bowdoin, L., Elmer, J.S., and Rogers, S.G. (1990). Expression of functional replication protein from tomato golden mosaic virus in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* 87, 1446-1450.

Harding, R.M., Burns, T.M., and Dale, J.L. (1991) Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *J. Gen. Virol.* 72, 225-230.

Harding, R.M., Burns, T.M., Hafner, G., Dietzgen, R.G., and Dale, J.L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *J. Gen. Virol.* 74, 323-328.

Hayes, R.J. and Buck, K.W. (1989). Replication of tomato golden mosaic virus DNA B intransgenic plants expressing open reading frames (ORFs) of DNA A: requirement of ORF AL2 for production of single-stranded DNA. *Nucleic Acids Res.* 17, 10213-10222.

Hayes, R.J., Coutts, R.H.A., and Buck, K.W. (1988). Agroinfection of *Nicotinia* spp. with cloned DNA of tomato golden mosaic virus. *J. Gen. Virol.* 69, 1487-1496.

Heyraud, F., Matzeit, V., Schaefer, S., Schell, J., and Gronenborn (1993). The conserved nonanucleotide motif of the geminivirus stem-loop sequence promotes replicational release of virus molecules from redundant copies. *Biochimie* 75, 605-615.

Hoekema, A., Hirsch, P., Hooykaas, P., and Schilperoort, R. (1983). A binary plant vector strategy based on separation of *vir* and T-region of the *Agrobacterium tumefaciens* Ti plasmid. *Nature*, 303, 179-180.

Hofer, J.M.I., Dekker, E.L., Reynolds, H.V., Woolston, C.J., Cox, B.S., and Mullineaux, P.M. (1992). Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *Plant Cell* 4, 213-223.

Hoffman, D.C., Anderson, R.C., DuBois, M.L., and Prescott, D.M. (1995). Macronuclear gene-sized molecules of hyptrichs. *Nucleic Acids Res.* 23, 1279-1283.

Hooykaas, P.J.J. and Schilperoort, R.A. (1992). *Agrobacterium* and plant genetic engineering. *Plant Mol. Biol.* 19, 15-38.

Hormuzdi, S.G. and Bisaro, D.M. (1993). Genetic analysis of beet curly top virus: evidence for three virion sense genes involved in movement and

regulation of single- and double-stranded DNA levels. *Virology* 193, 900-909.

Horns, T. and Jeske, H. (1991). Localisation of abutilon mosaic virus (AbMV) DNA within leaf tissue by *in situ* hybridisation. *Virology* 181, 580-588.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eichholtz, D., Rogers, S.G., and Fraley, R.T. (1985). *Science* 227, 1229-1231.

Howarth, A.J., Canton, J., Bossert, M., and Goodman, R.M. (1985). Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in geminiviruses. *Proc. Natl. Acad. Sci. USA* 82, 3572-3576.

Howarth,, A.J. and Van de Mark, G.J. (1989). Phylogeny of geminiviruses. *J. Gen. Virol.* 70, 2717-2727.

Hull, R. and Wil, H. (1989). Molecular biology of viral and non-viral retroelements. *Trends Genet.* 5, 357-359.

Isogai, M., Sano, Y., and Kojima, M. (1992). Identification of the unique DNA species in the milk-vetch dwarf virus-infected leaves. *Ann. Phtopath. Soc. Japan* 58, 631-632.

Ingham, D.J. and Lazarowitz, S.G. (1993). A single missense mutation in the BR1 movement protein alters the host range of the squash leaf curl geminivirus. *Virology* 196, 694-702.

Ingham, D.J., Pascal, E., and Lazarowitz, S.G. (1995). Both bipartite geminivirus movement proteins define viral host range, but only BL1 determines viral pathogenicity. *Virology* 207, 191-204.

Iwasaki, T., Ohki, R., Kiyama, R., and Oishi, M. (1995). Analysis of recombination junctions in extrachromosomal circular DNA by in-gel competitive reassociation. *FEBS Lett.* 363, 239-245.

Jacobsen-Lyon, K., Jensen, E.Ø., Jørgensen, J-E., Marcker, K.A., Peacock, W.J., and Dennis, E.S. (1995). Symbiotic and nonsymbiotic hemoglobin genes of *Casurina glauca*. *Plant Cell* 7, 213-223.

Jefferson, R.A. (1987). Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* 5, 387-405.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: β-glucuronidase as a sensitive and versatile gene marker in higher plants. *EMBO J.* 6, 3901-3907.

Jennings, P.A., Finch, J.T., Winter, G., and Robertson, J.S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* 34, 619-627.

Jensen, W.A. and Sailsbury, F.B. (1984). Botany, second edition. Wadsworth Publishing Co., California.

Johnson, B. (1957). Studies on the dispersal by upper winds of *Aphis craccivora* Koch in New South Wales. *Proc. Linn. Soc. NSW* 82, 191-198.

Joshi, C.P. (1987a). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15, 6643-6653.

Joshi, C.P. (1987b). Putative polyadenylation signals in nuclear genes of higher plants: a compoilation and analysis. *Nucleic Acids Res.* 15, 9627-9640.

Julia, J.F. (1982). *Myndus taffini* (Homoptera Cixiidae), vector of foliar decay of coconuts in Vanuatu. *Oléagineaux* 37, 409-414.

Jupin, I., de Kouchkovsky, F., Jouanneau, F., and Gronenborn, B. (1994). Movement of tomato yellow leaf curl geminivirus (TYLCV): involvement of the protein encoded by the ORF C4. *Virology* 204, 82-90.

Jupin, I., Hericourt, Benz, B., and Gronenborn, B. (1995). DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. *FEBS Letts* 362, 116-120.

Kallender, H., Petty, I.T.D., Stein, V.E., Panico, M., Blench, I.P., Etienne, A.T., Morris, H.R., Coutts, R.H.A., and Buck, K.W. (1988). Identification of the coat protein of tomato golden mosaic virus. *J. Gen. Virol.* 69, 1351-1357.

Katul, L., Vetten, H.J., Maiss, E., Makkouk, K.M., Lesemann, D-E, and Casper, R. (1993). Characterisation and serology of virus-like particles associated with faba bean necrotic yellows. *Ann. appl. Biol.* 123, 629-647.

Katul, L., Maiss, E., and Vetten, H.J. (1995). Sequence analysis of a faba bean necrotic yellows virus DNA component containing a putative replicase gene. *J. Gen. Virol.* 76, 475-479.

Kikuno, R., Toh, H., Hayashida, H., and Miyata, T. (1984). Sequence similarity between putative gene products of geminivirus DNAs. *Nature* 308, 562-563.

Kim, K.S., Shock, T.L., and Goodman, R.M. (1978). Infection of *Phaseolus vulgaris* by bean golden mosaic virus: ultrastructural aspects. *Virology* 89, 22-33.

Kheyr-Pour, A., Bendahmane, M., Matzeit, V., Accotto, G.P., Crespi, S., and Gronenborn, B. (1991). Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus. *Nucleic Acids Res.* 19, 6763-6769.

Klinkenberg, F.A., Elwood, S., and Stanley, J. (1989). Fate of African cassava mosaic virus coat protein deletion mutants after agroinfection. *J. Gen. Virol.* 70, 1837-1844.

Koonin, E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiationof eukaryotic replication. *Nucleic Acids Res.* 21, 2541-2547.

Koonin, E.V., Mushegian, A.R., Ryabov, E.V., and Dolja, V.V. (1991). Diverse groups of plant RNA and DNA viruses share related mmovement proteins that may possess chaperone-like activity. *J. Gen. Virol.* 72, 2895-2903.

Koonin, E.V., and Ilyina, T.V. (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J. Gen. Virol.* 73, 2763-2766.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, and Gronenborn, B. (1995). *Proc. Natl. Acad. Sci. USA* (in press).

Lazarowitz, S.G (1987). The molecular characterisation of geminiviruses. Plant Mol. Biol. Rep. 4, 177-192.

Lazarowitz, S.G. (1991). Molecular characterisation of two bipartite geminiviruses causing squash leaf curl disease: role of viral replication and movement functions in determining host range. *Virology* 180, 70-80.

Lazarowitz, S.G., Wu, L.C., Rogers, S.G., and Elmer, J.S. (1989). Maize streak virus genes essential for systemic spread and symptom development. *EMBO J.* 8, 1023-1032.

Lazarowitz, S.G (1992). Geminiviruses: genome structure and gene function. *Critical Reviews in Plant Sciences* 11, 327-349.

Lazarowitz, S.G., and Lazdins, I.B. (1991). Infectivity and complete nucleotide sequence of the cloned genomic components of a bipartite squash leaf curl geminivirus with a broad host range phenotype. *Virology* 180, 58-69.

Lazarowitz, S.G., Wu, L.C., Rogers, S.G., and Elmer, J.S. (1992). Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *The Plant Cell* 4, 799-809.

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1986). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology* 9, 936-976.

Lockhart, B.E.L. (1990). Evidence for a double-stranded circular genome in a second group of plant viruses. *Phytopathology* 80, 127-131.

Lung, M.C.Y. and Pirone, T.P. (1974). Aquisition factor required for aphid transmission of purified cauliflower mosaic virus. *Virology* 60, 260-264.

MacDowell, S.W., MacDonald, H., Hamilton, W.D.O., Coutts, R.H.A., and Buck, K.W. (1985). The nucleotide sequence of cloned wheat dwarf virus DNA. *EMBO J.* 4, 2173-2180.

Mankertz, A., Bläß, ., Mankertz, J., Walter, A., and Buhk, H-J. (1993). Analysis of transcription and replication of porcine circovirus (PCV). IXth International Congress of Virology, Abstract W45-8.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses. *Intervirology* 17, 1-199.

Maule, A.J. (1991). Virus movement in infected plants. *Crit. Rev. Plant Sci.* 9, 457-473.

Mazin, A.V., Kuzminov, A.V., Dianov, G.L., and Salganik, R.I. (1991). Mechanisms of deletion formation in *Escherichia coli* plasmids. II. Deletions mediated by short direct repeats. *Mol. Gen. Genet.* 228, 209-214.

McDonnell, R.E., Clark, R.D., Smith, W.A., and Hinchee, M.A. (1987). A simplified method for the detection of neomycin phosphotransferase II activity in transformed plant tissues. *Plant Mol. Biol. Rep.* 5, 380-386.

Medberry, S.L., Lockhart, B.E.L., and Olszewski, N.E. (1992). The commelina yellow mottle virus promoter is a strong promoter in vascular and reproductive tissues. *Plant Cell* 4, 185-192.

Messing, J., Geraghty, D., Heidecker, G., Hu, N-T., Kridl, J., and Rubenstein, I. (1983). Plant gene structure, pg. 211-277. In: "Genetic engineering of plants. An agricultural perspective." Kosuge, T., Meredith, C.P., and Hollaender, A. (eds). Plenum Press, New York.

Mogen, MacDonald, M.H., Leggewie, G., and Hunt A.G. (1992). Several disinct types of sequence elements are required for efficient mRNA 3'-end formation in a pea *rbcS* gene. *Mol. Cell. Biol.* 12,5406-5414.

Morris, B.A.M., Richardson, K.A., Andersen, M.T., and Gardner, R.C. (1988). Cassava latent virus infections mediated by the Ti plasmid of *Agrobacterium tumefaciens* containing either monomeric or dimeric viral DNA. *Plant Mol. Biol.* 11, 795-803.

Morris, B.A.M., Richardson, K.A., Haley, A., Zhan, X., and Thomas, J.E. (1992). The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants. *Virology* 187, 633-642.

Morris-Krsinich, B.A.M., Mullineaux, P.M., Donson, J., Boulton, M.I., Markham, P.G., Short, M.N., and Davies, J.W. (1985). Bidirectional transcription of maize streak virus DNA and identification of the coat protein gene. *Nucleic Acids Res.* 13, 7237-7256.

Mullineaux, P.M., Donson, J., Morris-Krsinich, B.A.M., Boulton, M.I., and Davies, J.W. (1984). The nucleotide sequence of maize streak virus DNA. *EMBO J.* 3, 3063-3068.

Mullineaux, P.M., Boulton, M.I., Bowyer, P., van der Vlugt, R., Marks, M., Donson, J., and Davies, J.W. (1988). Detection of a non-structural protein of

Mr 11000 encoded by the virion DNA of maize streak virus. *Plant Mol. Biol.* 11, 57-66.

Mullineaux, P.M., Guerineau, F., and Accotto, G.P. (1990). Processing of the complementary sense RNAs of *Digitaria* streak virus in its host and in transgenic tobacco. *Nucleic Acids Res.* 18, 7259-7265.

Murphy, F.A., Fauquet, C.M., Mayo, M.A., Jarvis, A.W., Ghabrial, S.A., Summers, M.D., Martelli, G.P., and Bishop, D.H.L. (eds) (1995). The classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses. Archives of Virology, Springer Verlag, Wein New York.

Nagel, R., Elliot, A., Masel, A., Birch, R.G., and Manners, J.M. (1990). Electroporation of binary Ti plasmid vector into *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *FEMS Microbiology Letters* 67, 325-328.

Navot, N., Pichersky, E., Zeidan, M., Zamir, D., and Czosnek, H. (1991). Tomato yellow leaf curl virus: A whitefly-transmitted geminivirus with a single genome component. *Virology* 185, 151-161.

Nejidat, A. and Beachy, R.N. (1990). Transgenic tobacco plants expressing a coat protein gene of tobacco mosaic virus are resistant to some other tobamoviruses. *Molec. Plant-Microb. Interactn.* 3, 247-251.

Noteborn, M.H.M., de Boer, G.F., van Roozelaar, D.J., Karreman, C., Kranenburg, O., Vos, J.G., Jeurissen, S.H.M., Hoeben, R.C., Zantema, A., Koch, G., van Ormondt, H., and van der Eb, A.J. (1991). Characterisation of cloned chicken anaemia virus DNA that contains all elements for the infectious replication cycle. *J. Virology* 65, 3131-3139.

Noueiry, A.O., Lucas, W.J., and Gilbertson, R.L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmatal transport. *Cell* 76, 925-932.

Okasaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A.A., Kohara, Y., Machida, Y., Tamanoi, F., and Hozumi, T. (1978). Structure and metabolism of the RNA primer in the discontinuous replication of prokaryotic DNA. *Cold Spring Harbor Symp. Quant. Biol.* 43, 203-219.

Padidam, M., Beachy, R.N. and Fauquet, C.M. (1995a). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *J. Gen. Virol.* 76, 25-35.

Padidam, M., Beachy, R.N., and Fauquet, C.M. (1995b). Classification and identification of geminiviruses using sequence comparisons. *J. Gen. Virol.* 76, 249-263.

Pang, S.Z., Bock, J.H., Gonsalves, C., Slightom, J.L., and Gonsalves, D. (1994). Resistance of transgenic *Nicotiana benthamiana* plants to tomato spotted wilt and impatiens necrotic spot tospoviruses: evidencce of

involvement of the N protein and N gene RNA in resistance. *Phytopathology* 84, 243-249.

Pascal, E., Goodlove, P.E., Wu, L.C., and Lazarowitz, S.G. (1993). Transgenic tobacco plants expressing the geminivirus BL1 protein exhibit symptoms of viral disease. *Plant Cell* 5, 795-807.

Patrick, C. (1994). Honours thesis. The Australian National University, Canberra, Australia.

Pfeiffer, P. and Mensard, J. (1995). The interplay of host and virus genes in the specificity and pathogenicity of cauliflower mosaic virus. In: Singh, R.P., Singh, U.S., and Kohmoto, K. (eds). Pathogenisis and Host Parasite Specificity in Plant Diseases. Vol. III, Viruses and Viroids. Pergamon Press, Oxford, U.K.

Pirone, T.P. (1991). Viral genes and gene products that determine insect transmissibility. *Seminars in Virology* 2, 81-87.

Randles, J.W., Hanold, D., and Julia, J.F. (1987). Small circular single-stranded DNA associated with foliar decay disease of coconut palm in Vanuatu. *J. Gen. Virol.* 68, 273-280.

Randles, J.W., and Hanold, D. (1989). Coconut foliar decay virus particles are 20nm icosahedra. *Intervirology* 30, 177-180.

Randles, J.W., Miller, D.C., Morin, J.P., Rohde, W., and Hanold, D. (1992). Localisation of coconut foliar decay virus in coconut palm. *Ann. Appl. Biol.* 121, 601-617.

Raikhel, N. (1992). Nuclear targetting in plants. *Plant Physiol.* 100, 1627-1632.

Ravelonandro, M., Monsion, M., Delbios, R., and Dunez, J. (1993). Variable resistance to plum pox virus and potato virus Y infection in transgenic *Nicotiana* plants expressing plum pox virus coat protein. *Plant Science* 91, 157-169.

Reinero, A. and Beachy, R.N. (1989). Reduced photosystem II activity and accumulation of viral coat protein in chloroplasts of leaves infected with tobacco mosaic virus. *Plant Physiol.* 89, 111-116.

Ritchie, B.W., Niagro, F.D., Lukert, P.D., Steffens, W.L., and Latimer, K.S. (1989). Characterisation of a new virus from cockatoos with psittacine beak and feather disease. *Virology* 171, 83-88.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependant basic domains in nucleoplasmin nuclear targetting sequence: Identification of a class of bipartite nuclear targetting sequence. *Cell* 64, 615-623.

Roberts, S. and Stanley, J. (1994). Lethal mutations within the conserved stem-loop of African cassava mosaic virus DNA are rapidly corrected by genomic recombination. *J. Gen. Virol.* 75, 3203-3209.

Rochester, D.E., Kositratana, W., and Beachy, R.N. (1990). Systemic movement and symptom production following agroinoculation with a single DNA of tomato yellow leaf curl geminivirus (Thailand). *Virology* 178, 520-526.

Rochow, W.F. and Dufus, J.E. (1981). Luteoviruses and yellows diseases. In: Kurstak, E. (ed.), Handbook of plant virus infections and comparative diagnosis. Elsevier/North-Holland Biomedical, Amsterdam.

Rodenburg, K.W., de Groot, M.J.A., Schilperoort, R.A., and Hooykaas, P.J.J. (1989). Single-stranded DNA used as an efficient new vehicle for transformation of plant protoplasts. *Plant Mol. Biol.* 13, 711-719.

Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffman, N.L., Brand, L., Elmer, J.S., and Lloyd, A. (1986). Tomato golden mosaic virus A component DNA replicates autonomously in transgenic plants. *Cell* 45, 593-600.

Rohde, W., Randles, J.W., Langridge, P., and Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* 176, 648-651.

Rohde, W., Becker, D., and Randles, J.W. (1995). The promoter of coconut foliar decay-associated circular single-stranded DNA directs phloem-specific reporter gene expression in transgenic tobacco. *Plant Mol. Biol.* 27, 623-628.

Roth, D.B. and Wilson, J.H. (1986). Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell. Biol.* 6, 4295-4304.

Rothnie, H.M., Reid, J., and Hohn, T. (1994). The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. *EMBO J.* 13, 2200-2210.

Rybicki, E.P. (1994). A phylogenetic and evolutioary justification for three genera of geminiviridae. *Archives of Virology* 139, 49-77.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). "Molecular Cloning: a laboratory manual." Second edition. Cold Spring Harbor Laboratory Press, NY.

Sanfaçon, H. (1994). Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. *Virology* 198, 39-49.

Sanfaçon, H., Brodmann, P., and Hohn, T. (1991). A dissection of the cauliflower mosiac virus polyadenylation signal. *Genes Dev.* 5, 141-149.

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

Sanford, J.C. and Johnstone, S.A. (1985). The concept of parasite-derived resistance-deriving resistance genes from the parasite's own genome. *J. Theor. Biol.* 113, 395-405.

Saunders, K., Lucy, A., and Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Res.* 19, 2325-2330.

Saunders, K., Lucy, A., and Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African casssava mosaic virus. *Nucleic Acids Res.* 20, 6311-6315.

Schalk, H-J., Matzeit, V., Schiller, B., Schell, J., and Gronenborn, B. (1989). Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. *EMBO J.* 8, 359-364.

Schmülling, T., Schell, J., and Spena, A. (1989). Promoters of the *rolA*, *B*, and *C* genes of *Agrobacterium rhizogenes* are differentially regulated in plants. *Plant Cell* 1, 665-670.

Schneider, M., Jarchow, E., and Hohn, B. (1992). Mutational analysis of the conserved region of maize streak virus suggests its involvement in replication. *Plant Mol. Biol.* 19, 601-610.

Smith, P.R. (1966). A disease of French beans (*Phaseolus vulgaris* L.) caused by subterranean clover stunt virus. *Aust. J. Agric. Res.* 17, 875-883.

Stanley, J. (1983). Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature* 305, 643-645.

Stanley, J. (1985). The molecular biology of geminiviruses. *Adv. Virus Res.* 30, 139-177.

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle replication. *Virology* 206, 707-712.

Stanley, J. and Gay, M.R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature* 301, 260-262.

Stanley, J. and Latham, J.R. (1992). A symptom variant of beet curly top geminivirus produced by mutation of open reading frame C4. *Virology* 190, 506-509.

Stanley, J., and Townsend, R. (1986). Infectious mutants of cassava latent virus generated *in vivo* from intact recombinant DNA clones containing single copies of the genome. *Nucleic Acids Res.* 14, 5981-5998.

Stanley, J., Markham, P.G., Callis, R.J., and Pinner, M.S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO J.* 5, 1761-1767.

Stanley, J., Latham, J.R., Pinner, M.S., Bedford, I., and Markham, P.G. (1992). Mutational analysis of the monopartite geminivirus beet curly top virus. *Virology* 191, 396-405.

Stenger, D.C. (1994). Strain-specific mobilization and amplification of a transgenic defective-interfering DNA of the geminivirus beet curly top virus. *Virology* 203, 397-402.

Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proc. Natl. Acad. Sci. USA* 88, 8029-8033.

Sugaya, S., Hayakawa, K., Handa, T., and Uchimiya, H. (1989). Cell-specific expression of the *rolC* gene of the TL-DNA of Ri plasmid in transgenic tobacco plants. *Plant Cell Physiol.* 30, 649-653.

Sunter, G. and Bisaro, D.M. (1989). Transcription map of the B genome component of tomato golden mosaic virus and comparison with A component transcripts. *virology* 173, 647-655.

Sunter, G. and Bisaro, D.M. (1991). Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. *Virology* 180, 416-419.

Sunter, G. and Bisaro, D.M. (1992). Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell* 4, 1321-1331.

Sunter, G., Gardiner, W.E., Rushing, A.E., Rogers, S.G., and Bisaro, D.M. (1987). Independent encapsidation of tomato golden mosaic virus A component DNA in transgenic plants. *Plant Mol. Biol.* 8, 477-484.

Sunter, G., Gardiner, W.E., and Bisaro, D.M. (1989). Identification of tomato golden mosaic virus-specific RNAs in infected plants. *Virology* 170, 243-250.

Sunter, G., Hartitz, M.D., Hormuzdi, S.G., Brough, C.L., and Bisaro, D.M. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology* 179, 69-77.

Sunter, G., Hartitz, M.D., and Bisaro, D.M. (1993). Tomato golden mosaic virus leftward gene expression: autoregulation of geminivirus replication protein. *Virology* 195, 275-280.

Sunter, G., Stenger, D.C., and Bisaro, D.M. (1994). Heterologous complementation by gemeinivirus AL2 and AL3 genes. *Virology* 203, 203-210.

Terada, R. and Shimamoto, K. (1990). Expression of CaMV 35S-GUS gene in transgenic rice plants. *Mol. Gen. Genet.* 220, 389-392.

Tinland, B., Hohn, B., and Puchta, H. (1994). *Agrobacterium tumefaciens* transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. *Proc. Natl. Acad. Sci. USA* 91, 8000-8004.

Tischer, I., Gelderblom, H., Vettermann, W., and Koch, M.A. (1982). A very small porcine virus with circular single-stranded DNA. *Nature* 295, 64-66.

Timmermans, M.C.P., Prem Das, O., and Messing, J. (1994). Geminiviruses and their uses as extrachromosomal replicons. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 45, 79-112.

Tindal, K.R. and Kunkel, T.A. (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* 27, 6008-6013.

Thomas, J.E., and Dietzgen, R.G. (1991). Purification, characterisation and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *J. Gen. Virol.* 72, 217-224.

Thommes, P., Osman, T.A.M., Hayes, R.J., and Buck, K.W. (1993). TGMV replication protein AL1 preferentially binds to single-stranded DA from the common region. *FEBS Lett.* 319, 95-99.

Todd, D., Creelan, J.L., Mackie, D.P., Rixon, F., and McNulty, M.S. (1990). Purification and biochemical characterisation of chicken anemia agent. *J. Gen. Virol.* 71, 819-823.

Tomenius, K. and Oxfelt, P. (1981). Preliminary observations of viruslike particles in nuclei in cells of wheat infected with the wheat dwarf disease. *Phtypath. Z.* 101, 163-167.

Torres-Pacheco, I., Garzón-Tiznado, J.A., Herrera-Estrella, L., and Rivera-Bustamante, R.F. (1993). Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. *J. Gen. Virol.* 74, 2225-2231.

Townsend, R., Stanley, J., Curson, S.J., and Short, M.N. (1985). Major polyadenylated transcripts of cassava latent virus and location of the gene encoding the coat protein. *EMBO J.* 4, 33-37.

Urban, L.A., Sherwood, J.J., Rezende, J.A.M., and Melcher, U. (1990). Examination of mechanisms of cross protection of cross protection with non-transgenic plants. In: Fraser, R.S.S. (ed.). Recognition and response in plant-virus interactions. Springer Verlag, Berlin.

van Loon, N., Miller, D., and Murane, J.P. (1994). Formation of extrachromosomal circular DNA in HeLa cells by nonhomologous recombination. *Nucleic Acids Res.* 22, 2447-2452.

van Mansfield, A.D.M., Langeveld, S.A., Weisbeek, P.J., Baas, P.D., van Arkel, G.A., and Jansz, H.S. (1979). Cleavage site of ϕX and G4 RFI DNA. *Cold Spring Harbor Symp. Quant. Biol.* 43, 331-334.

Varagona, M.J., Schmidt, R.J., and Raikhel, N.V. (1992). Nulear localisation signal(s) required for nuclaer targetting of the maize regulatory protein, Opaque-2. *Plant Cell* 4, 1213-1227.

von Arnim, A., Frischmuth, T., and Stanley, J. (1993). Detection and possible functions of African cassava mosaic virus DNA B gene products. *Virology* 192, 264-272.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the - and -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945-951.

Ward, C.W., Weiller, G.F., Shukla, D.D., and Gibbs, A.J. (1994). Molecular evolution of the Potyviridae, the largest plant virus family. In: Gibbs, A.J., Calisher, C.H., and Garcia-Arenal, F. (eds). Molecular basis of virus evolution. Cambridge University Press, Cambridge.

Wartig, L., Ramasso, E., Kheyr-Pour, A., de Kouchkovsky, F., Jouanneau, F., Luisoni, E., Accotto, G-P., and Gronenborn, B. (1993). Tomato yellow leaf curl virus (TYLCV) encodes a cytoplasmic non-structural protein V1. IXth International Congress of Virology, Abstract P62-11.

Wolf, S., Deom, C.M., Beachy, R.N., and Lucas, W.J. (1989). Movement protein of tobacco mosaic virus modifies plasmodesmatal size exclusion limit. *Science* 246, 377-379.

Woolston, C.J., Barker, R., Gunn, H., Boulton, M.I., and Mullineaux, P.M. (1988). Agroinfection and nucleotide sequence of cloned wheat dwarf virus DNA. *Plant Mol. Biol.* 11, 35-43.

Woolston, C.J., Reynolds, H.V., Stacey, N.S., and Mullineaux, P.M. (1989). Replication of wheat dwarf virus DNA in protoplasts and analysis of caot protein mutants in protoplasts and transgenic plants. *Nucleic Acids Res.* 17, 6029-6041.

Wu, R.Y. and Su, H.J. (1990). Purification and characterisation of banana bunchy top virus. *J. Phytopathol.* 128, 153-160.

Wu, L., Ueda, T., and Messing, J. (1993). 3'-end processing of the maize 27-kDa zein mRNA. *Plant J.* 4, 535-544.

Wu, R.Y., You, L.R., and Soong, T.S. (1994a). Nucleotide sequence of two circular single-stranded DNAs associated with banana bunchy top virus. *Phytopath.* 84, 952-958.

Wu, L., Ueda, T., and Messing, J. (1994b). Sequence and spatial requirements for the tissue- and species- independant 3'-end processing mechanism of plant mRNA. *Mol. Cell Biol.* 14, 6829-6838.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119.

Yeh, H-H., Su, H-J., and Chao, Y-C. (1994). Genome characterisation and identification of viral-associated dsDNA component of banana bunchy top virus. *Virology* 198, 645-652.

Zeidan, M. and Czosnek, H. (1994). Aquisition and transmission of agrobacterium by the whitefly *Bemsia tabaci*. *Mol. Plant-Microbe Interact.* 7, 792-798.

Zhan, X., Haley, A., Richardson, and Morris, B. (1991). Analysis of the potential promoter sequences of African cassava mosaic virus by transient expression of the β-glucuronidase gene. *J. Gen. Virol.* 72, 2849-2852.

Zhan, X., Richardson, K., Haley, A., and Morris, B. (1993). The activity of the coat protein promoter of chloris striate mosaic virus is enhanced by its own and C1-C2 gene products. *Virology* 193, 498-502.