Host factors involved in

Tomato leaf curl virus infection

by

Luke Ashton Selth
B. Biotech. (Hons.), Flinders University

A thesis submitted for the degree of
Doctor of Philosophy

at

The University of Adelaide,
School of Agriculture and Wine,
Discipline of Plant and Pest Science

in collaboration with

CSIRO Plant Industry,
Horticulture Unit

Urrbrae, Adelaide
January, 2005

Table of contents

Abstract

Tomato leaf curl virus (TLCV; family *Geminiviridae*, genus *Begomovirus*) causes a severe disease of tomato in northern areas of Australia. Members of this family of economically-important plant pathogens are characterized by twinned icosahedral particles containing a single-stranded DNA (ssDNA) genome comprising 6-7 genes. Because of their limited coding capacities, geminiviruses rely almost entirely on host factors to replicate their genome. The mechanisms by which they induce and recruit the requisite replicational machinery in differentiated cells have not been well characterised. Similarly, the role of host factors in the intracellular, cell-to-cell and long-distance movement of geminiviruses is largely unknown. In this study, yeast two-hybrid technology was used to identify host partners of TLCV-encoded proteins that may be involved in its replication and/or movement. In addition to yielding important information about these complex processes, this work has identified new host-based resistance targets to counter TLCV infection.

Geminiviral replication enhancer (REn) proteins enhance the accumulation of viral ssDNA approximately 50-fold. TLCV REn was shown to interact with a new member of the NAC domain family of transcription factors, SlNAC1. In a transient TLCV replication system, overexpression of *SlNAC1* significantly increased viral ssDNA accumulation. These and other results obtained from this study imply that SlNAC1 plays an important role in the process by which REn enhances TLCV replication. Transgenic tomato plants partially silenced for *SlNAC1* expression have been generated and their ability to resist TLCV infection is under investigation.

A screen with the V1 protein retrieved a host autocatalytic glycosyltransferase, SlUPTG1, which may function in cell wall biosynthesis. Recent evidence suggests that V1 has a partially-redundant role in cell-to-cell viral movement. Consistent with this idea, overexpression of *SlUPTG1* increased the accumulation of TLCV ssDNA species and this response appears to be a consequence of enhanced viral movement. Thus, down-regulation of *SlUPTG1* may represent another strategy for achieving resistance to TLCV.

Putative binding partners were also identified for the TLCV replication-associated (Rep), C2 and C4 proteins, and the βC1 protein from a satellite molecule that is replicated by TLCV. The C4 screen retrieved three putative kinases, one of which shares significant homology with a receptor-like protein kinase that binds the begomoviral nuclear shuttle protein (NSP), implying that C4 and NSP share a conserved function. The other two kinases are members of the glycogen synthase kinase 3 (GSK-3)/SHAGGY kinase family, and their possible role in the pathogenicity of C4 is discussed. Five putative host binding partners of the βC1 protein from *Cotton leaf curl virus*-associated DNA β were isolated. One of these was able to complement a yeast deficient in ubiquitin conjugating enzyme (UBC) activity. The other four proteins are predicted to function in such diverse processes as ATP production, nitrogen scavenging and protein translation. A screen with the C2 protein, which was found to possess a transcriptional activation domain within its C-terminal 30 amino acids, identified a different UBC and a protein that shares homology with a member of the *Arabidopsis* chloroplast signal recognition particle. This suggests that the ability of βC1 and C2 to suppress post-transcriptional gene silencing occurs by a novel mechanism that involves UBCs. Rep may bind the related DAG and Dal1 proteins that function in chloroplast

differentiation, although the possible role of such an interaction in TLCV replication cannot be predicted at present.

The physical interactions that occur between TLCV proteins were analysed in yeast. A homotypic interaction between the N-terminus of Rep was detected, a finding that supports previous studies of other geminiviral Reps and confirms that this multifunctional protein functions as an oligomer. Putative homotypic interactions were also identified for the C2 and V1 proteins. The possible role of C2 and V1 dimerisation is discussed in relation to their role in TLCV infection.

This work has provided a new insight into the function of TLCV gene products and the role of host factors in facilitating TLCV infection. Furthermore, the identification of host factors required for TLCV replication and movement has provided targets for host-based resistance strategies. The down-regulation of *SlNAC1* in tomato plants is one such strategy currently under investigation.

Declaration

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Luke Selth

January 2005

Acknowledgements

I am indebted to my primary supervisor, Ali Rezaian, for giving a raw Honours student a chance to work in his lab, and then encouraging me to continue my studies by undertaking a Ph.D. His intelligence, helpfulness and work ethic all combine to make him a outstanding example of what a scientist should be. In addition, I have appreciated his friendship and the interesting and insightful discussions we have had, not just about science, but on diverse topics. I also thank my co-supervisor, John Randles, for his interest in my project and his unsurpassed knowledge of plant viruses. I am grateful to the University of Adelaide for awarding me the George Fraser scholarship and CSIRO Plant Industry for providing a Ph.D. top-up scholarship to carry out my studies.

Thanks to all of the members of the virology lab at CSIRO Plant Industry. In particular, I greatly appreciate the supervisory role that Satish Dogra played over the course of my Ph.D., and for the many epic games of table-tennis that we enjoyed. Satish assisted me in carrying out the *in vitro* binding assays and transient replication assays in this study. I also thank Saif Raisheed, whose Ph.D. project bisected mine in many areas. As a result, Saif and I collaborated on many experiments and his constant enthusiasm and optimism was much appreciated. Saif performed the *in situ* hybridisation shown in Chapter 3. I am also grateful to Akbar Behjatnia, with whom I have worked happily alongside over the past four years, for determining the 5'-end of SlUBC (Chapter 5). The time and effort expended by Helen Healy, from the University of Adelaide, to teach me the basics of working with yeast and yeast two-hybrid systems was also much appreciated.

Past and present members of CSIRO Plant Industry deserve equal thanks: Ian Dry, for his constant willingness to answer questions, no matter how stupid, for inviting me to play in his touch football team, which provided a vital outlet from science, and for his friendship; the other members of the 'Vili's Boys' gang, Alan Little and Jamus Stonor, for the many fun times we had together; Dale Mayer, for her constant friendship, the great trip we had to New Zealand, and basically for putting up with seeing me 8 hours a day for over four years; Matty Hayes, for his cooking tips and cheerful outlook on the horrors of molecular biology; Mandy Walker, for reading drafts, answering questions, having me over for dinner, and ruining whisky for me forever; Susan Wheeler, for help with the crossword and helping me understand that my Ph.D. isn't as important as I sometimes think it is; Chris Davies and Paul Boss, two of the funniest (and smartest) scientists I have met, for updating me on cricket scores and making me laugh; Julio Macedo-Rodrigues, a recent addition to CSIRO Plant Industry, for playing tennis and drinking beer with me, two very important pastimes; the administration staff, Julie Powell, Margaret Minter, Marg Amon, Maria Piscioneri and Janet Pratt, for their excellent support; and finally Karina Swann, Matt Tucker, Felix Jaffe, Angela Gackle, Liz Lee, Adam Takos, Jochen Bogs and everyone also at PI for their help and friendship. I must also mention how grateful I was for the unexpected measure of support provided by my workmates after the death of my Mum.

Thanks to Gregory Martin from Purdue University in the US for providing the tomato cDNA library used in yeast two-hybrid screens, Eduardo Bejarano from Malaga University in Spain for providing the *NbSCE1* clone, Stefan Jentsch from the Max Planck Institute of Biochemistry in Germany for the yeast *ubc* mutants, and Ghafar Sarvestani from the The Detmold Family Trust Cell Imaging Centre at the Hanson

Institute in Adelaide for taking such high-quality images of GFP:fusion proteins with the confocal microscope.

To all of my friends from school, university and various sports teams, especially Cleggy, Nick, Groff, Trigger, Hanno and Rhino, thanks for providing me with an escape from science and grounding me in reality. And to my extended family, I am eternally grateful for the home-cooked meals, which provided me with much-needed nourishment, and your constant encouragement.

I owe everything I've achieved to my family. I thank my Mum, whose life ended prematurely in March 2002, for her love and support, and for honing my interest in science. Mum, I wish you could be here to read this thesis (or just the abstract!), but I know you're looking down on me now and are stoked that I've finally finished. Jas, thanks for your encouragement, your efforts to get me to church, and your attempts at improving my social life. Jo, I really appreciated the long-distance support you provided and the interest you always showed in my research. Dad, thanks for providing the means which allowed me to receive such a good education, and your help with some of the more practical things in my life, which generally aren't my forté.

Finally, I want to thank my beautiful and wonderful partner Mera, for putting up with an often cranky Ph.D. student, for making me laugh, and for her genuine interest and concern about my work. Mera, you are an amazing person and the love of my life - I hope sharing my Ph.D. for the past two years hasn't been too painful!

Abbreviations

ACMV	*African cassava mosaic virus*
AD	activation domain
ADK	adenosine kinase
AMP	adenosine monophosphate
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
BCTV	*Beet curly top virus*
BGMV	*Bean golden mosaic virus*
BLAST	basic local alignment search tool
bp	base pairs
CaMV	*Cauliflower mosaic virus*
CBP	calmodulin-binding peptide
cDNA	complementary DNA
CLCuV	*Cotton leaf curl virus*
CaLCuV	*Cabbage leaf curl virus*
CM	common motif
CP	coat protein
CSR	complementary-strand replication
DBD	DNA-binding domain
DIG	digoxygenin
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dpi	days post-inoculation

dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
E1	ubiquitin activating enzyme
E3	ubiquitin ligase
EDTA	ethylenediamine-*tetra*-acetic acid
ER	endoplasmic reticulum
FD	Faraday(s)
g	gram(s)
g	relative centrifugal force
GFP	green fluorescent protein
GRAB	geminivirus RepA-binding
Grx	glutaredoxin
h	hour(s)
HRP	horseradish peroxidase
IR	intergenic region
IPTG	*Iso*-propyl-β-D-thiogalactopyranoside
kb	kilobase pairs
L	litre(s)
LB	Luria broth
M	molar
min	minute(s)
MOPS	3-N-Morpholinopropanesulfonic acid
MP	movement protein
mRNA	messenger RNA
MSV	*Maize streak virus*
Ni-NTA	nickel nitrilotriacetic acid

NLS	nuclear localisation signal
NSP	nuclear shuttle protein
NTP	nucleoside triphosphate
nt	nucleotide
ONPG	*o*-nitrophenyl-β-D-galactopyranoside
ORF	open reading frame
ori	origin of replication
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
pRBR	plant retinoblastoma-related protein
PTGS	post-transcriptional gene silencing
PVDF	polyvinylidene fluoride
RACE	rapid amplification of cDNA ends
RB	retinoblastoma protein
RCR	rolling circle replication
RDR	recombination-dependent replication
REn	replication-enhancer protein (encoded by *C3* or *REn*)
Rep	replication-associated protein
RF	replicative form
RGP	reversibly glycosylated peptide
RNA	ribonucleic acid
RT-PCR	reverse transcription-RCR
s	second(s)
SCE	SUMO conjugating enzyme
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis

siRNA	small interfering RNA
SLW	silverleaf whitefly
SNF1	SUCROSE NONFERMENTING1 protein kinase
SqLCV	*Squash leaf curl virus*
SSC	standard sodium citrate
ssDNA	single-stranded DNA
SS-DNA	salmon sperm DNA
ssRNA	single-stranded RNA
TAR	transcriptional activation region
TBE	tris-borate-EDTA
TGMV	*Tomato golden mosaic virus*
TLCV	*Tomato leaf curl virus* (Australian isolate)
Tris	tris(hydroxymethyl)aminomethane
TYLCV	*Tomato yellow leaf curl virus*
UAS	upstream activating sequence
UBC	ubiquitin conjugating enzyme
UPTG	UDP:glucose protein transglucosylase
V	volt(s)
WDV	*Wheat dwarf virus*
WT	wild-type
YFP	yellow fluorescent protein

Definition of homology, sequence identity and similarity

In this thesis I have followed the recommendations of Thiessen (2002) and Fitch (2000) for usage of terms that describe relationships between genes and proteins. Homology is a relationship between genes or proteins that share a common evolutionary origin. As such, genes or proteins may be homologous or not, but they cannot be partially homologous. Note that a conclusion that two (or more) genes or proteins are homologous is a conjecture, not an experimental fact. The similarity between sequences was determined by alignments using the Gap or Bestfit algorithms. For gene and cDNA sequences, this similarity is expressed as a percent nucleotide sequence identity. For protein sequences, this similarity is expressed as a percent amino acid identity and percent amino acid similarity. This latter calculation takes into account the similarity between specific amino acid residues in protein sequences as determined by their physico-chemical properties (e.g., polarity, size and charge).

Publications

Sections of this thesis have been published in the following articles (see Appendix 2):

Selth, L.A., Randles, J.W., and Rezaian, M.A. (2002). *Agrobacterium tumefaciens* supports DNA replication of diverse geminivirus types. *FEBS Letters*. 516:179-182.

Selth, L.A., Randles, J.W., and Rezaian, M.A. (2004). Expression of the *Tomato leaf curl geminivirus* genes in host plants reveals novel viral:host interactions. *Molecular Plant-Microbe Interactions*. 17:27-33.

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H., Randles, J. W., and Rezaian, M. A. (2005). A NAC domain protein interacts with *Tomato leaf curl virus* replication accessory protein and enhances viral replication. *The Plant Cell*. 17:311-325.

Chapter 1 - General introduction

1.1 - Geminiviruses

1.1.1 *Geminiviridae*

Geminiviruses (family *Geminiviridae*) are pathogens of a range of economically important crops mainly in tropical and subtropical regions (Moffat, 1999). They are characterised by twin icosahedral capsids containing either a monopartite or bipartite single-stranded DNA (ssDNA) genome. In infected cells, replication occurs via a double-stranded DNA (dsDNA) intermediate after viral induction of host replicative machinery (Hanley-Bowdoin et al., 1999).

Four genera of the family have been identified, *Mastrevirus*, *Curtovirus*, *Begomovirus*, and *Topocuvirus*, that differ in respect to insect vector, host range, and genome structure (van Regenmortel et al., 2000; Fauquet et al., 2000). Mastreviruses infect monocotyledonous (monocot) plants, contain single-component (monopartite) genomes, and are transmitted by leafhoppers. Similarly, curtoviruses contain monopartite genomes and are transmitted by leafhoppers, but infect dicotyledonous (dicot) plants. Begomoviruses possess either monopartite or bipartite genomes, infect dicots, and are transmitted by whiteflies. Topocuviruses are transmitted by treehoppers to dicots and contain a monopartite genome.

1.1.2 Geminivirus replication

Geminiviruses are thought to employ a rolling circle replication (RCR) mechanism to amplify their ssDNA genomes and to produce dsDNAs which serve as replicative and

transcriptional templates (Saunders et al., 1991; Stenger et al., 1991; Heyraud et al., 1993). This strategy is analogous to that found in some bacteriophages and a class of eubacterial plasmids (Baas, 1987; Novick, 1998), suggesting that geminiviruses may have evolved from prokaryotic episomal replicons. Supporting this idea, TLCV and other geminiviruses are able to replicate in *Agrobacterium tumefaciens* and *Escherichia coli* (Rigden et al., 1996; Selth et al., 2002).

RCR is a two-step process in which leading- and lagging-strand DNA synthesis are separate events (Kornberg et al., 1992). The initial phase encompasses the synthesis of a "minus" (complementary-sense) strand using the "plus" (virion-sense) strand as a template, to yield a double-stranded, replicative form (RF). Little is known about this step, termed complementary-strand replication (CSR), but it is thought to be primed by an RNA molecule which is generated through RNA polymerase or DNA primase activity (Hanley-Bowdoin et al., 1999). Supporting this hypothesis, a small oligonucleotide complementary to the 3' intergenic region has been isolated from several mastrevirus virions (Hayes et al., 1988; Morris et al., 1992), which can be extended by DNA polymerase *in vitro* and may prime minus-strand synthesis *in vivo*.

In the second phase, shown in Figure 1.1, the RF serves as a template for the synthesis of plus-strand DNA. Priming of this latter step requires the introduction of a site-specific nick. In geminiviruses, this is performed by the Rep protein, which cleaves within a highly conserved nonanucleotide sequence found in the IR of all geminivirus genomes (Section 1.1.9.1.1) (Dry et al., 1993; Laufs et al., 1995a). Upon cleavage, Rep becomes covalently bound to the free 5'-end, whilst the 3'-end is used to prime the synthesis of virion-sense ssDNA.

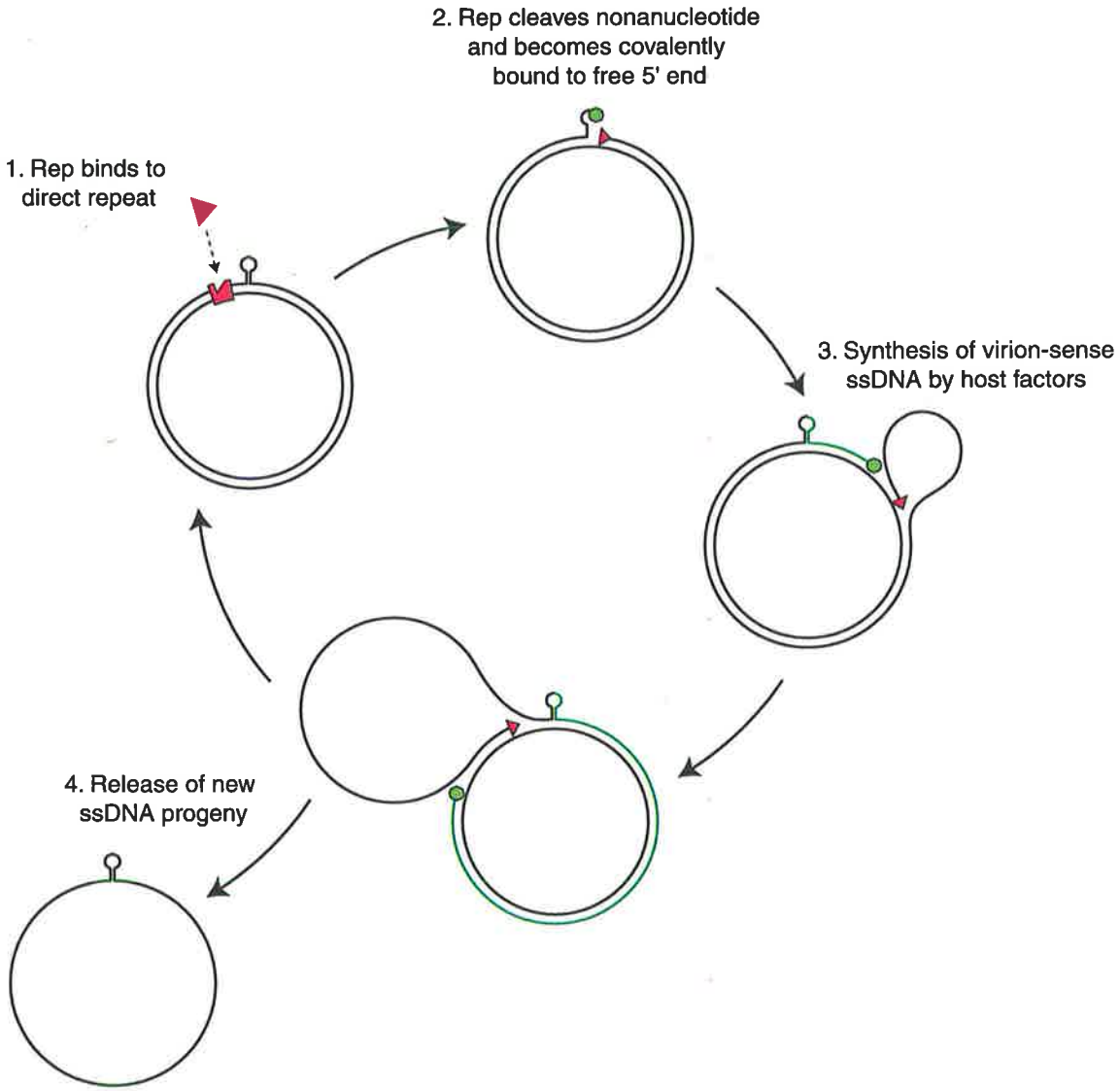


Figure 1.1. Model of rolling-circle replication. After the host-mediated conversion of viral ssDNA to dsDNA (not shown), Rep (filled red triangle) binds the direct repeat in the TLCV origin of replication (1) and subsequently cleaves the conserved nonanucleotide found in the stem-loop (2). A host replication complex (filled green circle) synthesises new virion-sense DNA (3). Rep then mediates termination of virion-strand replication and release of new progeny ssDNA (4).

The subsequent events involved in plus-strand DNA synthesis are not well characterised. Since geminiviruses do not encode a protein with detectable homology to known DNA polymerases, the elongation phase is presumably performed by plant nuclear replication machinery. It is well established, however, that termination of plus-strand replication and resolution of the concatameric DNAs into discrete genome-sized units is mediated by the Rep protein (Section 1.1.9.1.1) (Laufs et al., 1995b).

Recent evidence suggests that geminivirus replication also occurs via a recombination-dependent replication (RDR) mechanism (Jeske et al., 2001; Jeske, 2003; Alberter et al., 2004). In RDR, a viral DNA fragment recombines at a homologous site within an intact covalently closed circular DNA molecule and is extended. This form of replication allows the rescue of damaged and incomplete geminivirus DNA. The newly produced ssDNA molecules would then be converted to dsDNA by a mechanism similar to CSR.

1.1.3 The geminivirus origin of replication

The plus-strand origin of replication (*ori*) for all geminiviruses has been mapped to the 5' IR, and possesses a number of key elements (Figure 1.2). The first is a hairpin motif, containing a GC-rich stem and an AT-rich loop. This loop contains a nonanucleotide sequence, TAATATTAC, that is conserved among all geminivirus genomes (Dry et al., 1993; Laufs et al., 1995a) and which is also found in the plus-strand origins of other rolling circle systems (Baas, 1987). Mutations in this sequence inhibit geminivirus replication *in vivo* and Rep-mediated cleavage *in vitro* (Section 1.1.9.1.1) (Orozco and Hanley-Bowdoin, 1996).

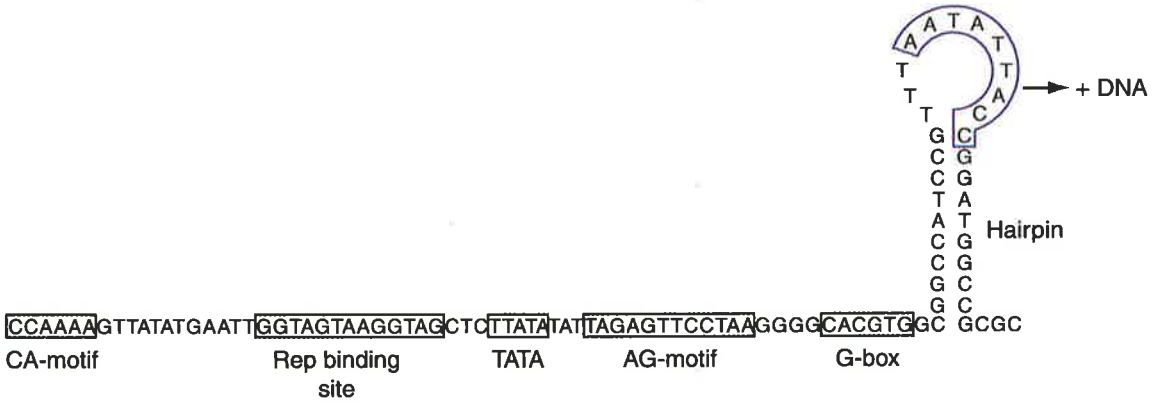


Figure 1.2. The TGMV plus-strand origin of replication and *C1* (Rep) promoter. The DNA sequence corresponding to TGMV A positions 54 to 153 is shown. Only the top strand of the duplex DNA is given. The initiation site and direction of synthesis for plus-strand DNA replication is indicated (+ DNA). Other functional elements are boxed. The hairpin structure is drawn and the conserved nonanucleotide loop sequence is marked (reproduced from Hanley-Bowdoin et al., 1999).

The second crucial element in the plus-strand origin is a directly repeated sequence, GGXXX, found upstream of the hairpin motif. Fontes *et al.* (1992) showed that *Tomato golden mosaic virus* (TGMV) Rep was able to bind to the direct repeat when the viral genome was in the double-stranded RF. Similar motifs have been identified in the IRs of many other begomoviruses and curtoviruses, including TLCV, and *in vitro* experiments have verified that these are sites of Rep binding (Fontes et al., 1994a; Fontes et al., 1994b; Behjatnia et al., 1998). These experiments have also indicated that the direct repeat acts as an origin recognition element to confer virus-specific replication. For example, TGMV Rep, which recognises the sequence GGTAGTAAGGTAG, is unable to bind the *Bean golden mosaic virus* (BGMV) direct repeat, GGAGACTGGAG (Fontes et al., 1994b). Specificity is conferred by variability in the nucleotides directly following the invariant GG, and the spacing within and between the repeats (Choi and Stenger, 1996). The purpose of Rep binding to the direct repeat motif is unclear, but probably plays a role in initiation of replication and transcriptional regulation of the *C1* gene (Section 1.1.9.1.3). However, the importance of these motifs in most geminiviruses has been clearly demonstrated: mutations in the direct repeat that impair Rep binding also interfere with viral replication *in vivo* (Fontes et al., 1994a; Orozco et al., 1998). A possible exception to this rule is TLCV, since recent experiments suggest that replication of TLCV and its satellite DNA is not altered by mutation of the direct repeat (Lin et al., 2002).

The plus-strand origin of begomoviruses overlaps the complementary sense promoter, and they share two common elements in addition to the Rep binding site, a TATA box and a G-box motif (Eagle and Hanley-Bowdoin, 1997). The TATA box, located immediately upstream of the Rep binding site, and the G-box, found at the base of the

hairpin, are required for the efficient transcription of the *C1* and *C4* ORFs, but are not essential for viral replication (Eagle and Hanley-Bowdoin, 1997).

Two other functional elements have been identified in the geminivirus *ori*, a CA and an AG-motif. These are thought to bind host factors that facilitate initiation of plus-strand synthesis (Hanley-Bowdoin et al., 1999; Orozco et al., 1998).

1.1.4 *Tomato leaf curl virus*

Several begomoviruses infect the cultivated tomato, *Solanum lycopersicum*, resulting in the disease termed tomato leaf curl or tomato yellow leaf curl. Symptoms of both diseases include curling of leaves, reduced or no fruit set, chlorotic mottling, and stunted plant growth (Nakhla and Maxwell, 1998). These diseases are of major economic importance in tropical and subtropical regions throughout the world, including the Middle East, India, Mediterranean countries, South East Asia, Africa, and the Americas (Nakhla and Maxwell, 1998). A number of examples highlight the significance of the diseases; in the Mediterranean, losses to the tomato crop range from 28-92%, depending on the age of the plants at the time of infection and the proportion of plants infected, whilst in Egypt production losses may reach 80% in autumn-grown tomatoes (Makkouk and Leterrot, 1983; Nakhla and Maxwell, 1998).

The first report of tomato leaf curl disease in Australia was made by Aldrick (1970) during a survey of plant pathogens in the Northern Territory. Since that time, the tomato crops in this region have suffered losses of 80-100%. Dry *et al.* (1993) isolated the causal agent of this disease, and found it to be a monopartite begomovirus that was subsequently named *Tomato leaf curl virus* (TLCV). An ongoing survey investigating

the spread of TLCV has detected the virus in far northern Queensland (Stonor et al., 2004). TLCV has not yet reached coastal southern Queensland, the major tomato growing area of Australia, but it appears likely that TLCV will spread to these regions and become a significant problem to the tomato industry in Australia.

1.1.5 Genome organisation of begomoviruses

The genomes of begomoviruses consist of one or, more commonly, two circular ssDNA molecules. These DNAs contain divergent coding sequences separated by a 5' intergenic region (IR) (Hanley-Bowdoin et al., 1999). Six to seven open reading frames (ORFs) are encoded within both the virion- and complementary-sense strands of the genome. The products of these genes control viral replication, movement, encapsidation, transmission by insect vector, and host disease symptoms.

1.1.6 Genome organisation of TLCV

The genome of TLCV is monopartite and contains 2,766 nucleotides (nt) (Dry et al., 1993). It comprises six overlapping ORFs, two of which (*V1* and *V2*) are encoded on the virion-sense and four (*C1*, *C2*, *C3* and *C4*) on the complementary-sense strand (Figure 1.3). Alignment of the putative amino acid sequences of their gene products revealed that the Australian TLCV is most closely related to the Sardinian and Israeli isolates of *Tomato yellow leaf curl virus* (TYLCV) and *African cassava mosaic virus* (ACMV) (Dry et al., 1993).

1.1.7 Transcription of geminivirus ORFs

The divergent overlapping ORFs encoded in the geminiviral genome are transcribed bidirectionally, resulting in virion- and complementary-sense mRNAs. Transcription

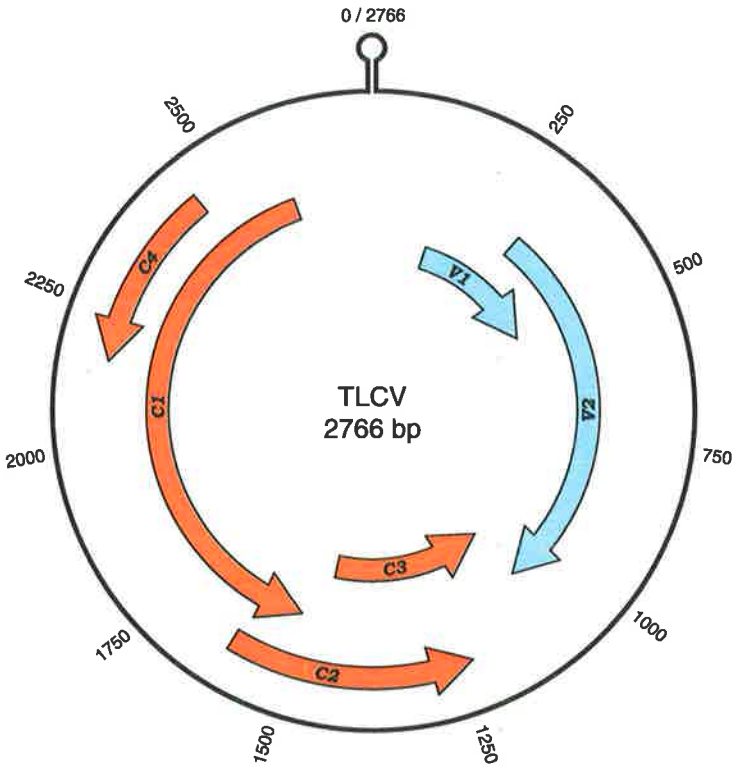


Figure 1.3. Genome organisation of TLCV. ORFs on the virion-sense (clockwise) strand and the complementary-sense (anticlockwise) strand are displayed by arrows. The position of the conserved stem-loop structure (♀) is marked (reproduced from Dry et al., 1993).

initiates downstream of TATA box motifs or initiator elements, suggesting that it is done by host RNA polymerase II. The viral transcripts are polyadenylated and often comprise multiple overlapping mRNA species (Hanley-Bowdoin et al., 1999).

1.1.8 Transcription of TLCV ORFs

The viral transcripts present in host cells during TLCV infection have been characterised (Mullineaux et al., 1993). Nuclease protection assays and rapid amplification of cDNA ends (RACE) identified four major virus-specific RNAs. Two RNA species are produced from the complementary-sense strand; one covers the *C1* (and *C4*), *C2*, and *C3* ORFs and a second internal RNA spans *C2* and *C3* only. The virion-sense strand also produces two transcripts, which initiate on either side of the *V1* start codon.

1.1.9 Function of TLCV genes

The functions of the TLCV genes have not been well characterised. However, detailed biochemical and genetic studies have been performed for a number of other geminiviruses, particularly the bipartite begomovirus TGMV. This section combines data obtained from research into the function of the TLCV genes and studies analysing related sequences in other geminiviruses so as to give a broad outline of the function of TLCV genes.

1.1.9.1 *C1*

The TLCV *C1* gene encodes the geminiviral replication-associated protein (Rep), which is the only viral gene product essential for viral DNA replication (Behjatnia et al.,

1998). Rep has a number of distinct functions, outlined below. Figure 1.4 summarises current knowledge of the Rep protein domains which facilitate these activities.

1.1.9.1.1 DNA binding, nicking, and ligation activities of Rep

Fontes *et al.* (1992, 1994a, 1994b), Lazarowitz *et al.* (1992), and Behjatnia *et al.* (1998) showed that the Rep proteins of TGMV, BGMV, *Squash leaf curl virus* (SqLCV), and TLCV respectively bind direct-repeat sequences in their replicative origins. This dsDNA binding is, in most cases, essential for viral replication and probably plays a role in regulating transcription of the *C1* gene (Section 1.1.9.1.3).

Laufs *et al.* (1995b) demonstrated that the TYLCV Rep also binds single-stranded TYLCV DNA within a highly conserved nonanucleotide sequence, TAATATTAC, which is thought to be a loop in a hairpin structure (Dry et al., 1993; Laufs et al., 1995a) and which appears in the plus-strand origins of other rolling circle systems (Baas, 1987). Upon binding, the Rep protein introduces a nick between the seventh and eighth nucleotides and remains bound to the newly generated 5'-end. The Rep protein also has a joining activity, suggesting that *in vivo* the combination of nicking and ligation would mediate the resolution of the concatameric DNAs into discrete genome-sized units (Laufs et al., 1995b).

1.1.9.1.2 Rep NTPase activity

All geminivirus Rep proteins contain a sequence similar to the consensus nucleotide triphosphate (NTP)-binding motif (Gorbalenya and Koonin, 1989). This homology was confirmed by the detection of adenosine triphosphatase (ATPase) activity in TYLCV (Desbiez et al., 1995) and TGMV (Orozco et al., 1997) Rep proteins. ATPase activity

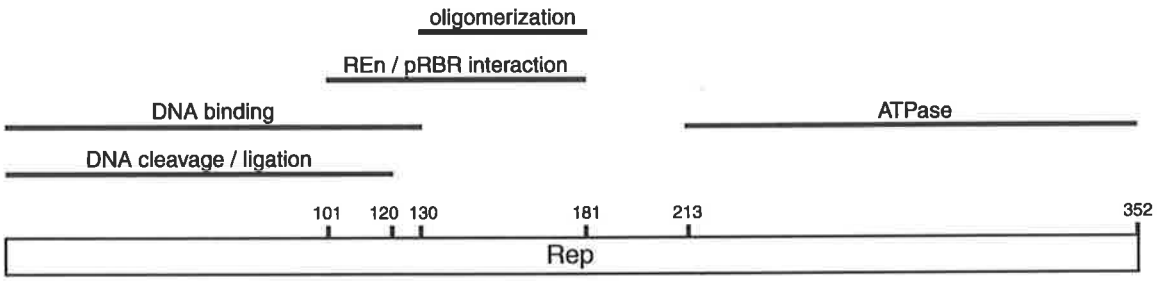


Figure 1.4. Domains of Rep. Solid lines above the protein mark the location of the functional domains for oligomerisation, REn and pRBR interaction, DNA binding, DNA cleavage and ligation, and ATPase activity. The numbers correspond to amino acid positions in TGMV Rep (reproduced from Hanley-Bowdoin et al., 1999).

is often linked to helicase activity, and Rep proteins share weak identity to DNA helicases (Gorbalenya and Koonin, 1993). However, no helicase activity has yet been reported for a geminivirus Rep protein, and the DNA-independent nature of the ATPase activity is not consistent with Rep acting as a helicase (Hanley-Bowdoin et al., 1999). An alternative hypothesis is that hydrolysis of ATP mediates a conformational change in Rep that is necessary for the initiation of plus-strand synthesis (Hanley-Bowdoin et al., 1999). This is supported by the observation that mutation of the NTP binding motif inhibits the ability of Rep to support viral replication but has no effect on Rep-mediated transcriptional repression (Eagle et al., 1994; Desbiez et al., 1995).

1.1.9.1.3 Rep transcriptional regulation

The Rep protein from ACMV has been shown to regulate its own expression at the level of transcription (Hong and Stanley, 1995). It has been proposed that this function is mediated through the binding activity of Rep to the direct-repeat in the *ori*. Since this sequence is located between the TATA box motif and the *C1* transcriptional start-point, it is conceivable that binding of Rep could interfere with the assembly of transcriptional machinery (Hong and Stanley, 1995). This hypothesis is supported by the observation that the regulatory activity of Rep has been delineated to the same region as its DNA binding domain (Hong and Stanley, 1995).

1.1.9.1.4 Induction of host replication machinery by Rep

Geminiviruses replicate in differentiated plant cells, but only supply one or two factors to this process. Furthermore, differentiated plant cells generally express low or no detectable levels of proteins associated with DNA replication. Thus, geminiviruses must modify the host cell cycle and thereby induce the expression of host DNA

replication factors to achieve amplification. The Rep protein plays a critical role in this process by its interaction with host factors involved in cell cycle control (Section 1.2.1).

1.1.9.1.5 Rep toxicity

The first suggestion that Rep may be toxic to host plants came from the observation that repeated attempts to generate transgenic tobacco plants stably expressing TLCV Rep were unsuccessful (L. Krake, I. Dry and M. Rezaian, personal communication). Since that time, two studies have further highlighted the phytotoxicity of Rep. Transient expression of the Rep proteins from TYLCV-China (van Wezel et al., 2002a) and TLCV (Selth et al., 2004) in host plants using viral vectors resulted in the formation of necrotic local lesions on inoculated leaves. In both studies, the region of Rep conferring this toxicity was mapped to the N-terminus of the protein. The N-terminus of Rep contains the putative pRBR-binding domain (Section 1.1.9.1.4), and it was hypothesised that Rep toxicity may be mediated by disruption of pRBR cell cycle control mechanisms.

1.1.9.2 *C2*

C2 (also designated *AC2* and *AL2*) encodes a protein that appears to be required for nuclear transactivation of virion-sense gene expression, which has been shown to occur at the level of transcription (Haley et al., 1992; Sunter and Bisaro, 1992; Hong and Stanley, 1997; Dry et al., 2000). This activity is virus-nonspecific within begomoviruses, and is thought to be mediated by the DNA binding activity of these proteins (Noris et al., 1996).

The *C2* genes from ACMV, TYLCV-China, and TLCV produce a severe phenotype including necrotic lesions and veinal necrosis when transiently expressed in *Nicotiana benthamiana*, a response linked to their ability to suppress post-transcriptional gene silencing (PTGS) (Voinnet et al., 1999; van Wezel et al., 2001; van Wezel et al., 2002b; Selth et al., 2004). PTGS is a sequence-specific RNA-degradation mechanism that involves dsRNA, spreads within the organism from a localised initiating area, and correlates with the accumulation of small-interfering RNAs (siRNA) (Vaucheret et al., 2001). In plants, PTGS acts as an innate host defence mechanism to protect against viruses and transposons (Ding, 2000), and is emerging as a basic host process that regulates endogenous gene expression (Baulcombe, 2002 and Voinnet, 2002). Many plant viruses have evolved proteins that suppress PTGS (Voinnet et al., 1999), highlighting the importance of this counter-defence and suggesting that this role of geminiviral C2 proteins is critical to enable efficient virus infection.

C2 proteins may also target basic metabolic pathways to facilitate virus infection by interacting with, and inactivating, protein kinases involved in the regulation of metabolism (Section 1.2.2).

1.1.9.3 *C3*

The *C3* gene of begomoviruses (hereafter referred to as *REn*) encodes the viral replication enhancer (REn) protein, which is able to enhance viral DNA accumulation through an unknown mechanism (Sunter et al., 1991). Studies of TGMV and BGMV have shown that REn is able to form oligomers and stable complexes with Rep (Settlage et al., 1996). The observation that REn is localised to the nuclei of infected cells at a level similar to that of Rep suggests that the mechanism by which it enhances viral

DNA accumulation may reside in its ability to bind Rep, perhaps by increasing the affinity of Rep for the origin (Hanley-Bowdoin et al., 1999). There is some evidence to support this hypothesis: TGMV Rep overcame an origin binding site mutation in the presence of REn (Fontes et al., 1994a), and BGMV Rep required the presence of REn to support replication from a chimeric origin (Gladfelter et al., 1997).

Alternatively, REn may play a more direct role in induction of a replication-competent environment and/or geminivirus DNA replication because, like Rep, it also interacts with host pRBR and proliferating cell nuclear antigen (PCNA) (Section 1.2.3) (Settlage et al., 2001; Castillo et al., 2003).

1.1.9.4 *C4*

Mutagenesis studies aimed at determining the role of the *C4* gene in TLCV, BCTV and TYLCV infection revealed that it is involved in symptom development (Rigden et al., 1994; Stanley and Latham, 1992; Jupin et al., 1994). However, the pathogenicity of these homologues may arise from different functions. TLCV and BCTV *C4* mutants were able to spread systemically and accumulated viral DNA species to wild-type (WT) levels. In contrast, the systemic spread of a TYLCV *C4* mutant was abolished, suggesting that the role of TYLCV C4 in symptom development relates to its function as a movement protein (MP). Reinforcing the idea that TLCV and BCTV *C4* have a direct role in symptom development, expression of these genes in transgenic plants caused severe developmental abnormalities and enations that were indicative of uncontrolled cell division (Krake et al., 1998; Latham et al., 1997). These findings suggest that, for some viruses, the C4 protein may be involved in creating a cellular environment that is conducive to geminiviral DNA replication (Section 1.2.4).

Derivatives of the bipartite begomoviruses TGMV, ACMV, and BGMV containing defective *AC4* genes (the bipartite homologue of *C4*) moved normally and replicated to WT levels (Elmer et al., 1988; Etessami et al., 1991; Hoogstraten et al., 1996), suggesting that *C4* is only functional in monopartite geminiviruses. A later study by van Wezel et al. (2002a), however, revealed that transient expression of ACMV *AC4* using a *Potato virus X* vector modified the hypersensitive response associated with ACMV Rep expression. Furthermore, Vanitharani et al. (2004) recently found that AC4 proteins from the Cameroon strain of ACMV and Sri Lankan cassava mosaic virus are suppressors of PTGS. It therefore appears as if some, if not all, *AC4* genes of bipartite begomoviruses contribute to viral pathogenicity.

1.1.9.5 *V1*

The *V1* genes of the monopartite begomovirus TYLCV-Sardinia and the mastreviruses *Maize streak virus* (MSV) and *Bean yellow dwarf virus* have been implicated in viral cell-to-cell movement (Rojas et al., 2001; Lazarowitz et al., 1989; Boulton et al., 1993; Liu et al., 1998). Two lines of evidence suggest that TLCV *V1* may possess a similar function. First, a TLCV *V1* mutant accumulated reduced levels of ssDNA (Rigden et al., 1993). Second, *in situ* hybridisation analysis of plants infected with the *V1* mutant indicated that it is deficient in cell-to-cell movement functions (M. S. Raisheed, L. A. Selth, A. M. G. Koltunow, J. W. Randles and M. A. Rezaian, submitted).

In addition, two studies have implicated TLCV *V1* in symptom expression. Plants infected with a TLCV *V1* mutant were asymptomatic (Rigden et al., 1993), and transient expression of *V1* by a *Tobacco mosaic virus*-based viral vector caused severe stunting of *N. benthamiana* and *N. clevelandii* plants (Selth et al., 2004). Since many viral

movement proteins are pathogenicity determinants, these data are consistent with the hypothesis that V1 is able to enhance cell-to-cell movement of TLCV.

1.1.8.6 *V2*

V2 encodes the TLCV coat protein (CP) (Rigden et al., 1993). Geminiviral CPs possess a sequence non-specific ssDNA binding activity, which appears to be important for the accumulation of viral ssDNA (Qin et al., 1998; Palanichelvam et al., 1998; Hartitz et al., 1999; Kirthi and Savithri, 2003). In monopartite geminiviruses, the CP is absolutely required for long-distance movement (Rigden et al., 1993; Noris et al., 1998) but the bipartite begomoviral CP is dispensable for systemic infection (Padidam et al., 1995). It has been proposed that the monopartite CP may act analogously to nuclear shuttle proteins (NSPs) from bipartite geminiviruses by binding and transporting ssDNA across the nuclear envelope (Rojas et al., 2001).

1.2 –Interactions between geminiviruses and their hosts

1.2.1 Rep

Geminiviruses depend on host DNA replication and transcription machinery. The majority of plant cells undergo differentiation and leave the cell division cycle, during which the levels of DNA replication enzymes fall to undetectable levels (Daidoji et al., 1992). Thus, geminiviral replication may be expected to be restricted to actively dividing tissue, such as apical meristems, developing leaves, and the cambium of mature plants (Staiger and Doonan, 1993). While this may be the case for some geminiviruses, such as *Beet curly top virus* (BCTV) (Esau, 1977), SqLCV (Sanderfoot and Lazarowitz, 1996) and TLCV (M. S. Raisheed, L. A. Selth, A. M. G. Koltunow, J. W. Randles and M. A. Rezaian, submitted), which are restricted to phloem tissue and

may replicate in procambial cells, most geminiviruses are found in a variety of tissues. For example, TGMV particles and DNA are present in the nuclei of differentiated cells in root, stem, and leaf tissue of infected *N. benthamiana* (Nagar et al., 1995). MSV and *Bean dwarf mosaic virus* are both found in vascular tissue and throughout the leaf (Hanley-Bowdoin et al., 1999). Thus, it is logical to conclude that geminiviruses are able to modify the host cell cycle and thereby induce the expression of host DNA replication factors.

The means by which geminiviruses achieve this process of dedifferentiation appears to rely, at least partly, on the Rep protein. Recently, it was shown that a number of geminiviral Rep proteins interact with plant homologues of the human retinoblastoma (RB) protein, termed retinoblastoma-related (pRBR) proteins (Xie et al., 1995; Collin et al., 1996; Ach et al., 1997). RB regulates the passage of cells through the G_1 phase and G_1-S transit of the cell cycle by modulating the activity of a family of transcription factors involved in this process (Helin, 1998). More specifically, phosphorylation of RB by cyclin-dependent kinase-cyclin complexes leads to the release of RB-bound E2F-DP factors which activate transcription of genes required for the G_1-S transition and S-phase progression (Mittnacht, 1998; Harbour and Dean, 2000). Geminiviral Rep proteins appear to act analogously to animal oncoviral proteins, for example adenovirus E1A, which are able to bypass the normal RB control pathway by binding to RB (Moran, 1993). In the model developed for mammalian viruses, sequestering of RB by viral proteins releases active E2F-DP complexes, leading to synthesis of host DNA replication machinery.

The theory that Rep induces a permissive state where replicative factors are available is further reinforced by the observation that PCNA, an accessory factor for DNA polymerase δ, was detected in terminally differentiated cells of *N. benthamiana* plants infected with TGMV, but not in healthy plants (Nagar et al., 1995). In addition, expression of Rep alone in transgenic plants was sufficient to induce the accumulation of PCNA in terminally differentiated cells. Two lines of evidence imply that induction of *PCNA* is mediated by the Rep/pRBR interaction. First, analysis of Rep mutants revealed that the ability of Rep to activate *PCNA* expression is linked tightly to its capacity to interact with pRBR (Kong et al., 2000). Second, induction of *PCNA* occurs at the transcriptional level and the *PCNA* promoter is under E2F negative control (Egelkrout et al., 2001).

The recent identification of new host partners for Rep has highlighted the multifunctionality of this protein and the complexity of changes which occur in a host cell following geminivirus entry. The Rep proteins from *Wheat dwarf virus* (WDV) and TYLCV-Sardinia can bind the replication factor C complex and PCNA respectively (Luque et al., 2002; Castillo et al., 2003), implicating Rep in the recruitment of a DNA replication complex to the 3'-OH primer-terminus following Rep-catalysed nicking of the stem-loop. TGMV Rep interacts with a protein kinase and a kinesin (Kong and Hanley-Bowdoin, 2002), possibly to establish and maintain a cellular environment favourable for geminivirus infection. Another host partner of TGMV Rep is histone H3 (Kong and Hanley-Bowdoin, 2002), suggesting that Rep may act to alleviate repression of virus replication and transcription processes induced by the packaging of geminiviral dsDNA species into minichromosomes (Abouzid et al., 1988; Pilartz and Jeske, 1992). Finally, the Rep proteins from TGMV and TYLCV-Sardinia also interact with a

SUMO-conjugating enzyme, suggesting that post-translation modification of Rep may occur *in planta* (Section 5.3.4).

The RepA protein of WDV, which shares approximately 200 N-terminal amino acids with the WDV Rep protein and can bind pRBR (Collin et al., 1996), was shown to interact with a group of proteins termed GRAB (for geminivirus RepA-binding) (Xie et al., 1999). GRAB proteins are members of the NAC domain family, which have roles in such diverse host aspects as flower development (Sablowski and Meyerowitz, 1998) and leaf senescence (John et al., 1997). The functional significance of the RepA/GRAB interaction is unknown, but GRAB expression inhibits WDV DNA replication in cultured cells (Xie et al., 1999), suggesting that RepA may interfere with a GRAB-mediated antiviral response. The potential role of NAC proteins in geminivirus infection is discussed in more detail in Chapter 3.

1.2.2 C2

The C2 proteins from TGMV and BCTV target basic host metabolic pathways to facilitate virus infection. They interact with and inactivate *Arabidopsis* SUCROSE NONFERMENTING1 (SNF1), a protein kinase that plays a central role in the regulation of metabolism (Hao et al., 2003). In response to nutritional and environmental stresses that deplete ATP, SNF1 turns off energy-consuming biosynthetic pathways and turns on alternative ATP-generating systems. Inactivation of SNF1 by C2 leads to an enhanced susceptibility phenotype that can be reproduced by silencing the expression of *SNF1* in transgenic plants, indicating that this gene mediates some level of resistance to geminivirus infection (Hao et al., 2003). This conclusion was supported by subsequent studies which showed that TGMV and BCTV C2 also

interact with and inactivate adenosine kinase (ADK) (Wang et al., 2003). ADK is responsible for recycling adenosine and maintaining intracellular AMP levels, and therefore can activate SNF1 by increasing AMP:ATP ratios. Thus, it appears that geminiviruses have evolved a dual approach to disabling the SNF1-mediated metabolic responses that are an innate antiviral defence. Inactivation of ADK by C2 may serve another purpose. Wang *et al.* (2003) presented preliminary data suggesting that these C2 proteins are able to suppress PTGS, as has been reported for their homologues from ACMV, TYLCV-C, and TLCV (Section 1.1.9.2). Another consequence of ADK inactivation is reduced transmethylation activity (Moffatt et al., 2002). Considering the role of methylation in reinforcing silencing pathways, the inhibition of ADK by C2 may indirectly suppress PTGS (Wang et al., 2003).

1.2.3 REn

The REn protein greatly enhances accumulation of viral ssDNA species *in planta* (Section 1.1.9.3). It binds to Rep and may increase the affinity of this protein for the viral *ori* (Fontes et al., 1994a; Gladfelter et al., 1997), an activity proposed to enhance viral replication (Hanley-Bowdoin et al., 1999). However, accumulating evidence suggests that REn may also possess a more direct role in replication by interacting directly with host factors to reprogramme mature plant cells for DNA replication competency. For example, Settlage *et al.* (2001) showed that, like Rep, TGMV REn can specifically interact with a maize pRBR protein. The observation that significantly higher levels of PCNA accumulate in nuclei of infected plants compared to transgenic plants that express only the Rep protein (Nagar et al., 1995) suggests that the REn/pRBR interaction may also play a role in overcoming pRBR-mediated repression of the *PCNA* promoter (Section 1.2.1).

REn has also been implicated in the recruitment and/or assembly of the geminiviral replication complex by an interaction with PCNA (Castillo et al., 2003). PCNA assembles into a trimer that encircles DNA and acts as a sliding clamp that modulates the interactions of other proteins, including polymerases, with DNA (Hingorani and O'Donnell, 2000).

The findings of Settlage et al. (2001) and Castillo et al. (2003) are similar to those in mammalian systems, where virus replication and virus-mediated cell cycle deregulation often requires several viral-host protein interactions. For example, the papillomavirus E7 protein binds RB (Chellappan et al., 1992), while the E6 protein binds p53 (Lechner and Laimins, 1994), another factor intimately involved in cell cycle control.

1.2.4 C4

To test the function of the TLCV *C4* gene, Krake *et al.* (1998) generated transgenic plants expressing *C4* under the control of the *Cauliflower mosaic virus* (CaMV) 35S promoter. These plants showed virus-like symptoms which included enations in the most severely affected transgenic lines. A study by Latham *et al.* (1997) showed that transgenic *N. benthamiana* plants that express BCTV *C4* develop abnormally and produce tumours. Together, these results suggest that C4 can induce cell division in plants in the absence of other viral proteins, presumably by interaction with host factors (Hanley-Bowdoin et al., 1999). However, the identity of these host components has yet to be determined.

1.2.5 Other geminiviral:host protein interactions

The isolation of host factors involved in geminiviral infections initially focussed on identifying those involved in viral DNA replication. More recently, yeast two-hybrid screens have been carried out using viral movement and coat proteins as bait to identify host factors involved in other viral processes.

The nuclear shuttle protein (NSP), encoded by the *BV1* gene of bipartite begomoviruses, facilitates the transport of viral DNA from the nucleus to the cytoplasm where it interacts with MP, encoded by *BC1*, to promote cell-to-cell spread of viral DNA (Sanderfoot and Lazarowitz, 1995). Recently, the NSP from *Cabbage leaf curl virus* (CaLCuV) was found to interact with an *Arabidopsis* histone acetyltransferase, AtNSI (McGarry et al., 2003). This interaction does not lead to acetylation of NSP, but rather AtNSI specifically acetylates CaLCuV CP. The role of AtNSI-mediated acetylation of CP is unknown, but overexpression of AtNSI enhances the infectivity of CLCV (McGarry et al., 2003) while mutation of the AtNSI-binding site in NSP inhibits infectivity (Carvalho et al., 2004), suggesting that acetylation of CP is important in regulating the nuclear events of viral movement.

A number of geminiviral CPs have been found to accumulate predominantly in nuclei (Liu et al., 1999; Rojas et al., 2001; Unseld et al., 2001). This localisation pattern is not surprising given the role of CP in binding viral ssDNA (Section 1.1.9.6). Kunik *et al.* (1999) elucidated the mechanism of TYLCV CP nuclear import by showing that it specifically interacts with tomato karyopherin α in yeast. Member of the karyopherin α family function as nuclear localisation signal (NLS) receptors (Nigg, 1997).

1.2.6 The yeast two-hybrid system for identifying protein interactions

An important step in elucidating the function of a particular protein is to identify other proteins with which it associates. One assay for detecting novel protein interactions is the yeast two-hybrid system (Fashena et al., 2000). This system relies on the modular nature of transcriptional activators, which are composed of a DNA binding domain (DBD) and a transcription activating domain (AD). These domains are independently non-functional as a transcription factor but, when brought into close proximity via non-covalent interactions, can reconstitute the activity of the intact protein (Stephens and Banting, 2000). In the two-hybrid system (Figure 1.5), a fusion protein that links a protein of interest (X) to the DBD is generated. This hybrid, commonly referred to as the bait, is then co-expressed with plasmid-based constructs encoding a library of independent fusions (Y) with the AD (the prey). If an X-Y interaction occurs, the activity of the transcription factor is recovered, and easily identified utilising a reporter gene (i.e. nutritional markers or enzymatic reporters) regulated in *cis* by a promoter containing cognate recognition sites for the DBD (Fashena et al., 2000).

1.3 - Aims and significance of this project

1.3.1 - Aims of this study

The specific objectives of the work described in this thesis are:

1) Identification of host factors that interact with TLCV-encoded proteins, using the yeast two-hybrid system (Chapters 3, 4, 5 and 6). In addition to providing information regarding the specific roles of TLCV proteins, this will enhance current understanding of the means by which geminiviruses manipulate the host to achieve replication, movement and transmission.

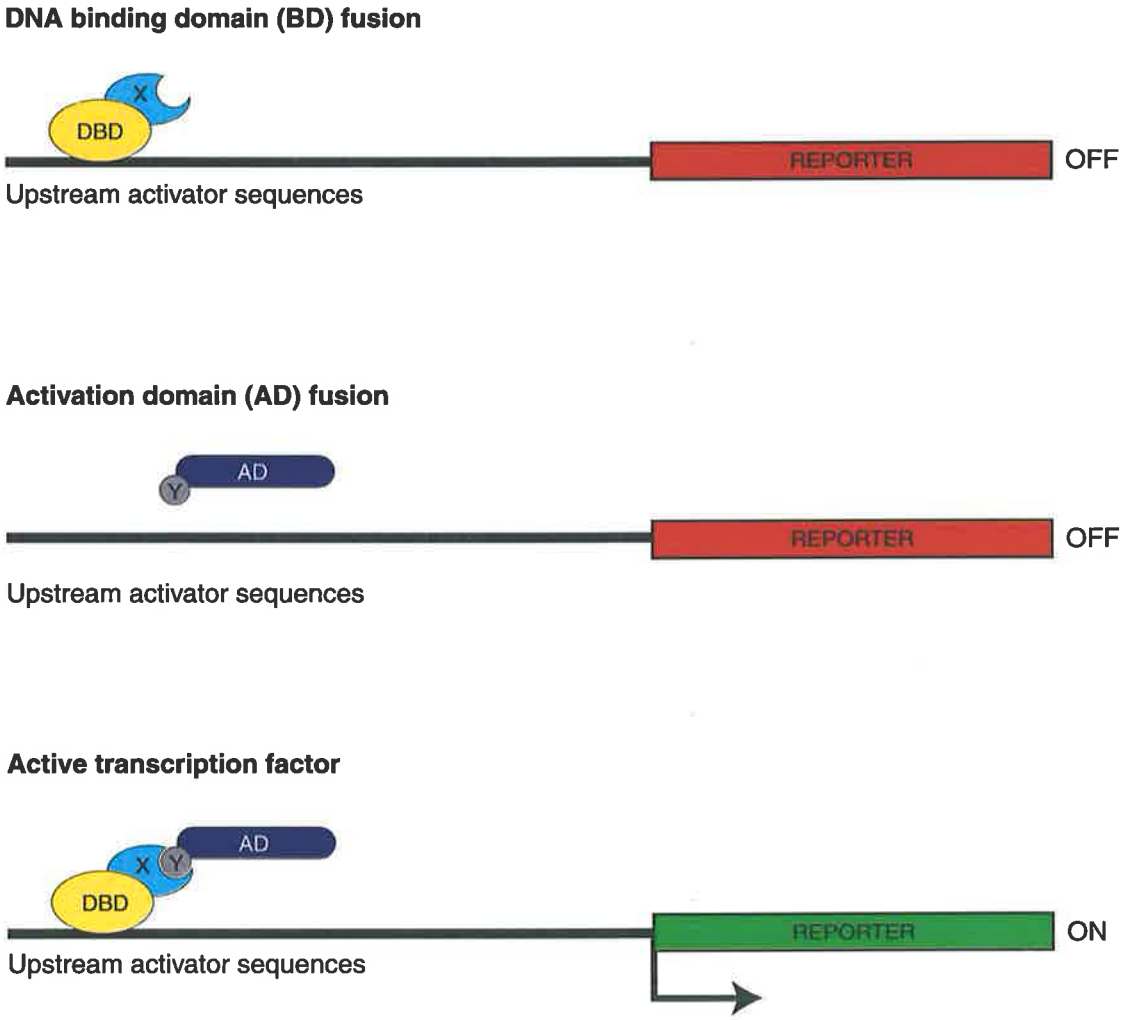


Figure 1.5. Summary of the yeast two-hybrid system for analysing protein:protein interactions. A fusion protein that links a protein of interest (X) to the DBD is co-expressed with a predicted binding partner (Y) fused to the AD. If an interaction between X and Y occurs, the activity of the transcription factor is recovered and can be detected using a reporter gene. This system can also be used to identify unknown proteins which bind X, by creating a library of independent fusions with Y.

2) Verification of putative protein/protein interactions identified in the yeast two-hybrid system by independent means (Chapters 3 and 4).
3) Functional characterisation of host proteins isolated from yeast two-hybrid screens by:
   (i) sequence analysis
   (ii) determining their subcellular localisation
   (iii) analysing their effect on TLCV replication
   (iv) protein-specific experiments to test possible endogenous functions
4) Analysis of the interactions that occur between TLCV-encoded proteins (Chapter 7). This may provide an insight into how multiple viral proteins act in concert to achieve specific functions.

1.3.2 - Significance of project

Geminiviruses are emerging as the most economically important family of plant-infecting viruses worldwide (Moffat, 1999). In particular, tomato-infecting geminiviruses cause devastating crop losses in many world regions (Section 1.1.4) (Nakhla and Maxwell, 1998). The Australian isolate of TLCV was first reported in the Northern Territory (Aldrick, 1970), and has since been spread to northern Queensland by the Australian indigenous biotype of the whitefly *B. tabaci*. Fortunately, this insect has a narrow host range and a very limited capacity to disperse and is not yet found in the intensive horticultural regions of southern, coastal Queensland (Stonor et al., 2003). However, in 1994 the silverleaf whitefly (SLW; syn. *B. tabaci* B biotype and *B. argentifolii*) was recorded in Australia for the first time (Gunning et al., 1995). This insect, a highly efficient vector of TLCV, has since spread to most mainland states and is well established in the major horticultural regions of Queensland, causing significant economic losses through feeding on cotton, cucurbits, eggplants, soyabeans, sunflowers,

sweet potatoes and tomatoes (Stonor et al., 2003). Management of SLW is impractical. Therefore, movement of TLCV to southern Queensland is predicted in the near future, an outcome that could devastate the $200-300 million annual tomato crop (Stonor et al., 2003).

TLCV is not controlled at present. Traditional breeding programs against TLCV have encountered problems including low tomato yield, recessive resistance genes, and resistance-breaking virus isolates, meaning that commercially-acceptable tomato cultivars are still at risk (M. A. Rezaian, personal communication). Efforts to introduce transgenic resistance against TLCV have also been unsuccessful. Expression of virus-derived transgenes has produced resistance to other geminiviruses including TGMV (Day et al., 1991), TYLCV (Bendahmane and Gronenborn, 1997; Brunetti et al., 1997), and ACMV (Sangare et al., 1999), but plants expressing TLCV-derived sequences including the *V2* and *C4* genes and antisense constructs of all six TLCV ORFs were still susceptible to TLCV infection (M. A. Rezaian, personal communication). This work will identify host factors which could have potential as targets for molecular resistance strategies.

In addition to the direct economic importance of this pathogen, studies on TLCV should provide insight into basic cellular plant processes. In animals, several DNA viruses depend on host replication and transcription machinery and can alter their hosts to create an environment that facilitates the infection process (Jansen-Durr, 1996). Research into these viruses has contributed significantly to our understanding of DNA replication, transcription, and cell cycle control in mammalian cells. Geminiviruses afford a similar potential for plant systems.

Chapter 2 - General materials and methods

2.1 Materials

2.1.1 Solutions

The solutions used in this project are described in Table 2.1. All chemicals were analytical or molecular biology grade. Solutions were prepared with nanopure or deionised water and autoclaved where appropriate.

2.1.2 Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were obtained from GeneWorks (Adelaide, Australia). Their nucleotide sequences are shown in Table 2.2.

2.1.3 Bacterial strains

Escherichia coli strain XL1-Blue (Stratagene; Cedar Creek, TX) was used for all routine cloning work. *E. coli* strains M15 (Qiagen; Clifton Hill, Australia) and B834(DE3)pLysS (Novagen; Madison, WI) were used for recombinant protein expression. *Agrobacterium tumefaciens* strain C58 was used both to inoculate plants with TLCV and for infiltration experiments.

Table 2.1. Solutions and their compositions

Solution	Composition
Agarose gel loading dye (DNA) (10x)	78% glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 10 mM ethylenediaminetetraacetic acid (EDTA)
Agarose gel loading dye (RNA) (10x)	50% (v/v) glycerol, 0.2 M EDTA, 0.08% (w/v) bromophenol blue
Binding assay buffer	50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, pH 8.0
Denaturing agarose gel loading buffer (RNA) (5x)	70% (v/v) deionised formamide, 10% (v/v) formaldehyde, 6% (v/v) agarose-gel loading dye (RNA), 14% (v/v) MOPS/EDTA buffer
GUS assay buffer	50 mM phosphate buffer (pH 7.0), 0.1% Triton X-100, 1 mM X-gluc
GUS fixing buffer	5% (v/v) formaldehyde, 5% (v/v) acetic acid, 20% (v/v) ethanol
Hybridisation buffer	0.25 M Na_2HPO_4, 7% SDS (w/v), 1 mM EDTA
LB (liquid growth media)	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.0
LB agar (solid growth media)	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 1.2% (w/v) Bacto-agar, pH 7.0
MOPS/EDTA buffer (10x)	200 mM MOPS, 50 mM sodium acetate, 10 mM EDTA, pH 7.0
Murashige and Skoog (MS) salt solution	4.33 g MS basal salt mixture (comp.)/L
Ni-NTA binding buffer	50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM β-mercaptoethanol, 1% Tween-20
Ni-NTA elution buffer	50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM β-mercaptoethanol, 1% Tween-20, 250mM imidazole, 10% glycerol
Ni-NTA washing buffer	50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM β-mercaptoethanol, 1% Tween-20, 20mM imidazole, 10% glycerol
PBS	138mM NaCl, 10mM sodium phosphate (pH 7.4), 2.7mM KCl
PBS-Tween	138mM NaCl, 10mM sodium phosphate (pH 7.4), 2.7mM KCl, 0.05% Tween-20
RNA extraction buffer	50 mM Tris, 100 mM NaCl, 5 mM EDTA, 2% (w/v) SDS, pH 8.0
SSC	150 mM NaCl, 15 mM tri-sodium citrate, pH 7.0
SDS-PAGE gel fixing buffer	
SDS-PAGE sample loading buffer	0.5 M Tris (pH 6.8), 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.05% bromphenol blue
STE buffer (10x)	500 mM Tris (pH 6.85), 1 M NaCl, 10 mM EDTA
STET buffer	8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM EDTA (pH 8.0), 50 mM Tris (pH 8.0)
TBE buffer	90 mM Tris, 90 mM borate (pH 8.3), 2 mM EDTA
TE buffer	10 mM Tris, 1 mM EDTA, pH 8.0
Electro-transfer buffer	200mM glycine, 25mM Tris, 20% methanol
TSS solution	85 % LB medium (v/v), 10 % PEG-8000 (w/v), 5 % DMSO (v/v), 50mM $MgCl_2$ (pH 6.5)
YT1	10 mM Tris (pH 8.0), 1 mM EDTA, 100 mM lithium acetate
YT2	10 mM Tris (pH 8.0), 1mM EDTA, 100 mM lithium acetate, 45% PEG-4000

Table 2.2. Oligonucleotide primers used in this study

Primer	Description	Sequence (5'→ 3')
C1-F-BamHI	For cloning into pLexA	GGGGATCCTTACTAGACCAAAGTCATTCCG[a]
C1-F-SmaI	For cloning into pART7	TCCCCCGGGATGACTAGACCAAAGTCATTCCGTATA AATGCTAA
C1-R-BamHI	For cloning into pLexA	GGGGATCCTCAATTCTCTTCCTCCGGAT
C1-R-HindIII	For cloning into pART7	CGCAAGCTTTCAATTCTCTTCCTCCGGATGG
$C1_{1-184}$-F-EcoRI	For cloning into pLexA	GGGAATTCACTAGACCAAAGTCATTCCG
$C1_{1-184}$-R-XhoI	For cloning into pLexA	GGTTCTCGAGCTAAAAAGGAGAAACATAAACCT
$C1_{184-362}$-F-BamHI	For cloning into pLexA	GGGGATCCTTTTTTTATCTTCTTCTTTTGA
C2-F-BamHI	For cloning into pLexA	GGGGATCCTTCAGAATTCATCACCCTCAAC
C2-F-pART7	BamHI, start codon, for cloning into pART7	GGGGATCCATGTTCAGAATTCATCACCCTCAAC
C2-F-pQE30	BamHI, for cloning into pQE30	TTGGATCCCAGAATTCATCACCC
C2-R-BamHI	For cloning into pLexA	GGGGATCCTTAAATACCCTCAAGAAACG
C2-R-pQE30	BamHI, for cloning into pQE30	GGGGAAGCTTTTAAATACCCTCAAG
$C2_{1-83}$-R-BamHI	For cloning into pLexA	GGGGATCCCTATGCTTTATGATCTTGAAA
$C2_{33-103}$-F-EcoRI	For cloning into pLexA	GGGAATTCATTGATTTGCCGTGCGGG
$C2_{33-103}$- R-XhoI	For cloning into pLexA	GGTTCTCGAGCTATGATTGAACTGTATCCGG
$C2_{84-136}$-F-BamHI	For cloning into pLexA	TTGGATCCTTCCACAAGCGCCCGTTCAA
CAT-F-BamHI	For cloning into pQE30	ACGGATCCGAGAAAAAAATCACTGGATATACC
CAT-R-HindIII	For cloning into pQE30	GCAAGCTTACGCCCCGCCCTGCCAC
REn-F-BamHI	For cloning into pQE30	TTTGGATCCGATTCACGCACAGGGGAAA
REn-F-EcoRI	For cloning into pLexA	TTGAATTCGATTCACGCACAGGGGAACC
REn-F-pART7	KpnI, start codon, for cloning into pART7	CGGGGTACCATGGATTCACGCACAGGGGAACC
REn-R-HindIII	For cloning into pQE30	GGGGGGGGGAAGCTTTTAATAAAAATTAAATTT
REn-R-XbaI	For cloning into pART7-C'gfp	GGGGGTCTAGATTAATAAAAATTAAATTTTA
REn-R-XhoI	For cloning into pLexA	GGGGCTCGAGTTAATAAAAATTAAATTTTA
REn-R-pART7	XbaI, start codon, for cloning into pART7	TGCTCTAGAGTTAATAAAAATTAAATTTTATATCAT GAT
REn_{1-70}-R-XhoI	For cloning into pLexA	AAGCCTCGAGTCATGTGAAGTCCAGGAA
REn_{40-120}-F-EcoRI	For cloning into pLexA	GGGAATTCAACTACAACCACGAC
REn_{40-120}-R-XhoI	For cloning into pLexA	AAGCCTCGAGTCATGAGTCTAGTACATT
REn_{90-134}-F-EcoRI	For cloning into pLexA	GGGGAATTCAAGTATTTAGATAGT
C4-F-BamHI	For cloning into pQE30	TTGGATCCAGAATGGGGAGCCTCATC
C4-F-EcoRI	For cloning into pLexA	TTGAATTCAGAATGGGGAGCCTCATCTC
C4-R-HindIII	For cloning into pQE30	GGGTTTAAGCTTCTAATTCCCTAAGGACGT
C4-R-XhoI	For cloning into pLexA	TTTTCTCGAGCTAATTCCCTAAGGACGTTA
SlNAC1-F-BamHI	For cloning into pCAL-n-FLAG	TTTGGATCCAACAAAGGAGCAAACGGA
SlNAC1-F-EcoRI	For cloning into pLexA	GGGAATTCAACAAAGGAGCAAACGGA
SlNAC1-F-KpnI	Start codon, for cloning into pART7	CGGGGTACCATGAACAAAGGAGCAAACGGAAATCA G
SlNAC1-R-NotI	For cloning into pLexA	TTGCGGCCGCTTAGTAAGGTTTTTGCAT
SlNAC1-R-XbaI	For cloning into pART7-C'gfp	GGGTCTAGATTAGTAAGGTTTTTGCAT
$SlNAC1_{1-70}$-R-XhoI	For cloning into pLexA	AAGCCTCGAGTCAATACCACTCTTTTTC
$SlNAC1_{1-170}$-R-XhoI	For cloning into pLexA	AAGCCTCGAGTCAATGCTTCTCAAGTGT
$SlNAC1_{1-240}$-R-NotI	For cloning into pLexA	AAAAGCGGCCGCTCAACATGGCGACAAGAC
$SlNAC1_{71-301}$-F-EcoRI	For cloning into pLexA	GGGGAATTCTTTTTCTCACCAAGG
SlNAC1-pN6-1	XbaI/EcoRI	AATCTAGAGAATTCAGTACCGCCTCGCCAACG
SlNAC1-pN6-2	BamHI/KpnI	GGGGATCCGGTACCTTAGTAAGGTTTTTGCAT
SlUBC-F-BamHI	For cloning into pCM vectors	AAAGGATCCATGGCGTCGAAGCGCATA
SlUBC-R-ClaI	For cloning into pCM vectors	GGGATCGATTTATCCCATCGCATATTT
SlUPTG1-F-EcoRI	Cloning into pART7 N'gfp and pART7	TTGAATTCATGGCAGCAGCAACACCA
SlUPTG1-F-pCAL	EcoRI, cloning into pCAL-n-FLAG	TTGAATTCGCAGCAGCAACACCA
SlUPTG1-R-BamHI	For cloning into pART7 N'gfp	GTGGATCCCTTTTTAGTCTTTGCTGG
SlUPTG1-R-XbaI	For cloning into pART7	GGGGTCTAGACTACTTTTTAGTCTT

Table 2.2. continued

Primer	Description	Sequence (5'→ 3')
TGMV_REn-R-XhoI	For cloning into pLexA	GGGGGGCTCGAGTTAATAAAATTTATA
Ubi3-F	For amplification of tomato ubiquitin 3 (SQ RT-PCR)	GGGATGCAGATCTTCGTGAAAACCC
Ubi3-R	For amplification of tomato ubiquitin 3 (SQ RT-PCR)	TCAATCGCCTCCAGCCTTGTTGTAA
V1-F-EcoRI	For cloning into pLexA	GGGAATTCTGGGATCCTTTAGTCCAC
V1-F-SphI	For cloning into pQE30	TTTTTTGCATGCTGGGATCCTTTAGTCCAC
V1-R-HindIII	For cloning into pQE30	TTTTTTAAGCTTTCAGGGCTTCTGAACAGC
V1-R-XhoI	For cloning into pLexA	GGTTCTCGAGTCAGGGCTTCTGAACAGC
V2-F-BamHI	For cloning into pQE30	TTGGATCCAGCAAGCGACCAGCAGAT
V2-F-EcoRI	For cloning into pLexA	GGGAATTCAGCAAGCGACCAGCAGAT
V2-R-BamHI	For cloning into pART7 C'gfp	GGGGATCCTTAATTCTGAATCGA
V2-R-HindIII	For cloning into pQE30	GGGGGGAAGCTTTTAATTCTGAATCGAATC
V2-R-XhoI	For cloning into pLexA	GGGGCTCGAGTTAATTCTGAATCGAATC
βC1-F-EcoRI	For cloning into pLexA	TTGAATTCACACCGAGCGGAACAAACAA
βC1-R-XhoI	For cloning into pLexA	GGGGCTCGAGTTAAACGGTGAACTTTTTAT
pB42AD-F	Sequencing primer for pB42AD	CCAGCCTCTTGCTGAGTGGAGATG
pB42AD-R	Sequencing primer for pB42AD	AGGTAGACAAGCCGACAACCTTGATTGG
pB42-linker-F	pB42AD polylinker	AATTCGAGATCTGGCCCGGGCCATGGGGTACCC
pB42-linker-R	pB42AD polylinker	TCGAGGGTACCCCATGGCCCGGGCCAGATCTCG
pCAL-seq-F	Sequencing primer for pCAL-n-FLAG	TCATCCTCCGGGGCACTT
pCAL-seq-R	Sequencing primer for pCAL-n-FLAG	TTGCTCAGCGGTGGCAGC
pLexA-F	Sequencing primer for pLexA	CGTCAGCAGAGCTTCACCATT
pLexA-R	Sequencing primer for pLexA	GCTGCAGGTCGACTCGAGCGG

[a]Underlined sequences correspond to specific restriction enzyme sites.

2.2 Methods

This section outlines general methods used throughout this project, which are essentially as described by Sambrook and Russell (2001) or according to the manufacturer's instructions. Methods which have been significantly modified from their published form are outlined. Specific protocols are outlined in the relevant chapters.

2.2.1 Restriction enzyme digestion of DNA

DNA was digested with restriction endonucleases from Roche Diagnostics (Indianapolis, IN), Promega (Annandale, Australia), New England Biolabs (Beverly, MA), and Fermentas (Hanover, MD) using buffers supplied by the manufacturers.

2.2.2 Agarose gel electrophoresis

2.2.2.1 Standard agarose gel electrophoresis for DNA and RNA

EasyCast horizontal minigel tanks (OWL Scientific Inc., Cambridge, UK) were used for electrophoresis of DNA. 0.7-2.0% (w/v) agarose gels were prepared using Type I-A low EEO agarose in TBE buffer (Table 2.1), and contained 0.5 μg/ml (w/v) ethidium bromide. Samples were adjusted to 2× agarose loading dye (Table 2.1) before applying to the wells. Gels were electrophoresed at approximately 100 V in TBE running buffer before being visualised and photographed using a short wavelength UV transilluminator.

Electrophoresis of RNA was essentially the same as that described for DNA except that gel tanks, trays, and combs were treated with 0.2 M NaOH for approximately 2 h prior to use (Table 2.1).

2.2.2.2 Denaturing agarose gel electrophoresis for RNA gel blot analysis

Denaturing agarose gels were prepared by adding the appropriate amount of agarose to 105 ml of water, autoclaving, adding 30 ml 37% formaldehyde and 15 ml sterile 10× MOPS/EDTA buffer (Table 2.1), and then pouring into a gel tray pre-treated with 0.2 M NaOH. Samples were adjusted to 3× denaturing agarose gel loading buffer (Table 2.1) and incubated at 65°C for 15 min before loading. Electrophoresis was carried out in 1× MOPS/EDTA buffer (Table 2.1).

2.2.3 Extraction of DNA from agarose gel slices

DNA bands were excised from agarose gels and extracted with QIAquick Gel Extraction Kit (Qiagen).

2.2.4 DNA amplication by the polymerase chain reaction (PCR)

DNA species to be cloned were amplified by PCR with Platinum Taq DNA Polymerase High Fidelity (Invitrogen; Carlsbad, CA). For all other purposes, components of a PCR reaction were as follows: DNA template, oligonucleotide primers (200-500 nM), 1× reaction buffer (Gibco BRL; Rockville, MD), 200 μM dATP, dCTP, dGTP, and dTTP (Promega), 1.5 μM $MgCl_2$, and 0.5 units of recombinant Taq DNA polymerase (Gibco), in a 20 μl or 50 μl reaction volume. Thermal cycling generally consisted of: 3 min at 95°C (one cycle); 45 s at 94°C, 30 s at 50°C, 1 min at 72°C (30 cycles); 7 min at 72°C (one cycle).

2.2.5 Dephosphorylation of DNA 5' termini

5' phosphate groups were removed from DNA fragments using calf intestinal alkaline phosphatase (Roche). The enzyme was inactivated by adding 20 mM EDTA and heating at 65°C for 15 min.

2.2.6 Converting 5' or 3' DNA overhangs to blunt ends

To convert 5' or 3' overhangs to blunt ends, up to 5 μg of the DNA of interest was mixed with 100 μM dNTPs, 1× restriction enzyme buffer B (Promega) and 5 units of T4 DNA polymerase (Promega) per μg of DNA. The reaction mix was incubated for 30 min at room temperature.

2.2.7 Extraction of DNA samples following enzymatic reactions

Extraction of DNA samples after restriction enzyme digestion (Section 2.2.1), PCR (Section 2.2.4), dephosphorylation (Section 2.2.5) and blunting (Section 2.2.6) reactions was achieved using a QIAquick PCR Purification Kit (Qiagen).

2.2.8 DNA ligation

PCR products were ligated into the T-tailed vectors pGEM T-Easy (Promega) or pDRIVE (Qiagen). All other ligations, including blunt-end ligations, were carried out in 20 μl reaction volumes containing a insert:vector molar ratio of approximately 6:1 and 6 units of T4 DNA ligase (New England Biolabs) and incubated overnight at 16°C.

2.2.9 Transformation of bacteria with recombinant plasmids

Electrocompetent *E. coli* XL1-Blue, *E. coli* M15, and *A. tumefaciens* C58 cells (Section 2.2.10) were transformed by electroporation using a Gene-Pulser apparatus (Bio-Rad,

Hercules, CA). Approximately 1 ng of plasmid or 1 µl of ligation reaction mixture was added to a 25 µl aliquot of cells and transferred to an ice-cold electroporation cuvette (path length = 1mm) (Invitrogen). This was then given a single pulse in the Gene-Pulser (1.8 kV, 125 µFD, 200 Ohms), and immediately resuspended in 400 µl of LB (Table 2.1). After incubation at 37°C for 1 h to allow expression of antibiotic-resistance genes, the transformed cells were spread on 1.2% LB agar plates (Table 2.1) with appropriate antibiotic selection and incubated at 37°C overnight.

Chemically-competent *E. coli* B834(DE3)pLysS cells (Section 2.2.11) were transformed by heat-shock. 50 ng of vector DNA was mixed with 100 µl aliquots of cells and incubated on ice for 30 min. Cells were then placed at 42°C for 45 s, resuspended in 500 µl LB, and incubated at 37°C for 1 h. The transformed cells were spread on 1.2% LB agar plates containing 100 µg/ml amipicillin and 25 µg/ml chloramphenicol and incubated at 37°C overnight.

2.2.10 Preparation of electrocompetent *E. coli* cells

500 ml of LB was inoculated with a 5 ml overnight culture of the *E. coli* strain of interest and grown at 37°C with vigorous shaking to an optical density (OD_{600}) of 0.5. Cells were chilled on ice for 10 min and centrifuged for 15 min at 4°C at 5000*g*. The cells were resuspended in 500 ml of sterile ice-cold water and centrifuged again. The cells were washed and centrifuged again with 250 ml sterile ice-cold water and resuspended in 10ml of sterile ice-cold 10% glycerol. Bacterial cells were transferred to a new 50 ml falcon tube, centrifuged again, and resuspended in 2 ml of ice-cold 10%

glycerol. Aliquots of 25 µl were placed into ice-cold microcentrifuge tubes, snap-frozen in liquid nitrogen, and stored at -80°C.

2.2.11 Preparation of chemically-competent *E. coli* cells

100 ml of LB broth was inoculated with a 1 ml overnight culture of the *E. coli* strain of interest and grown at 37°C with vigorous shaking to an optical density (OD_{600}) of 0.5. Cells were chilled on ice for 20 min and centrifuged for 15 min at 4°C at 5000*g*. The cells were resuspended in 10 ml of ice-cold TSS solution (Table 2.1). Aliquots of 100 µl were placed into ice-cold microcentrifuge tubes, snap-frozen in liquid nitrogen, and stored at –80°C.

2.2.12 Growth of bacteria in liquid cultures

Liquid cultures were set up by inoculating LB containing appropriate antibiotics with a single bacterial colony or a loopful of frozen glycerol stock. Cultures were incubated at 37°C overnight with shaking.

2.2.13 Preparation of bacterial plasmid DNA

Preparation of plasmid DNA from 1-5 ml of overnight culture was generally done using a QIAprep Spin Miniprep Kit (Qiagen). Large-scale (20-50 ml cultures) preparation of plasmid DNA was achieved using a Plasmid Midi Kit (Qiagen).

Alternatively, plasmid DNA to be used as template for restriction enzyme digestion or to transform competent *E. coli* cells was prepared using a miniprep boiling method. Briefly, 1 ml of overnight culture was centrifuged at 16,100*g* for 1 min at room

temperature. The supernatant was discarded and the pellet resuspended in 350 µl STET (Table 2.1) buffer. After adding 12.5 µl of 20 mg/ml lysozyme, the cells were boiled for 1 min and centrifuged at 16,100*g* for 20 min. The pellet was removed with a sterile toothpick and 40 µl 3 M sodium acetate and 220 µl isopropanol added to the supernatant. This mixture was centrifuged at 16,100*g* for 10 min. The resulting pellet was washed with 70% ethanol, dried under vacuum, and resuspended in 50 µl of sterile water.

2.2.14 Preparation of yeast plasmid DNA

Yeast plasmid DNA was purified using an RPM Yeast Plasmid Isolation Kit (Q Biogene, Carlsbad, CA) according to the manufacturer's instructions.

2.2.15 Preparation of bacterial glycerol stocks

Bacterial glycerol stocks were prepared by adding 1 volume of 40% or 80% sterile glycerol to an overnight culture, snap-freezing in liquid nitrogen, and storing at –80°C.

2.2.16 Small (S)- and library (L)-scale lithium acetate yeast transformations

5 ml (S) or 50ml (L) yeast cultures were incubated at 30°C overnight with shaking. The following day, 50 ml (S) or 300 ml (L) of fresh media was inoculated with the overnight culture to an OD_{600} of 0.2-03 and grown at 30°C overnight with shaking for 4 h. Cells were pelleted by centrifuging at 1000*g* for 5 min, resuspended in 20 ml sterile water, pelleted again, and resuspended in 1.5 ml of YT1. For small-scale transformations, 200 ng plasmid DNA, 100 µg of denatured salmon sperm DNA (SS-DNA), 100 µl of resuspended cells and 600 µl of YT2 were mixed in microcentrifuge tubes. For library scale transformations, 80-100 µg library DNA, 2 mg SS-DNA, 1 ml of resuspended

cells and 6 ml of YT2 were mixed in a 20 ml falcon tube. After vortexing, tubes were incubated at 30°C with gentle shaking for 30 min. 70 µl (S) or 700 µl (L) of DMSO was added and the tubes mixed by gentle inversion and incubated at 42°C for 15 min. After cooling on ice for 2 min, cells were pelleted at 16,100*g* for 15 s (S) or 1000*g* for 5 min (L) and resuspended in 0.4 ml (S) or 10 ml (L) TE buffer. 200 µl of small-scale transformation mix was spread on 90 mm plates, or ~300 µl of library scale transformation mix was spread each on up to 30 145 mm plates. Plates were incubated at 30°C for 2 (S) or 7 (L) days.

2.2.17 Yeast two-hybrid screening

The vectors pLexA (*HIS3* marker), pB42AD (*TRP1* marker) and pGNG1 (*URA3* marker) (Fig. 2.1) were used in yeast two-hybrid screening (Section 1.2.6). pLexA is used to express bait proteins of interest fused to the *E. coli* LexA DBD. The pB42AD plasmid expresses proteins fused to B42, an acidic peptide which functions as a strong transcriptional AD. A population of pB42AD plasmids containing a library of tomato (v. Rio Grande) cDNAs fused to B42 was kindly provided by Prof. Gregory Martin (Purdue University). The *GFP* gene in pGNG1 is downstream of the recognition sequence for LexA and thus acts as a reporter for interaction between a peptide fused to LexA and a peptide fused to B42. Yeast strain displayYEAST-L (*MATα*, *trp1*, *his3*, *ura3*, *leu2*::*2 LexAop-LEU2*; Display Systems Biotech, Vista, CA) was used in all two-hybrid screens. It contains a genomic leucine biosynthesis gene downstream of the DNA recognition sequence for LexA which is used as a secondary reporter for interaction.

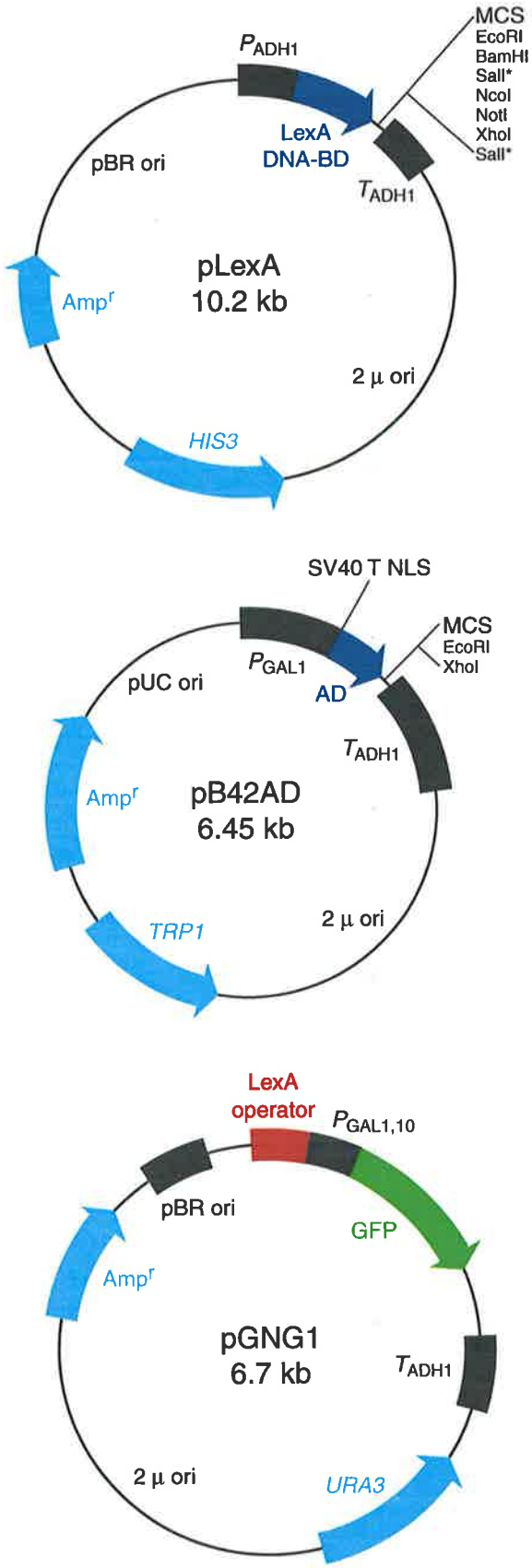


Figure 2.1. Vectors used in yeast two-hybrid screening. pLexA is used to express bait proteins of interest fused to the *E. coli* LexA DNA-binding protein (BD). The pB42AD plasmid expresses proteins fused to B42, an acidic peptide which functions as a strong transcriptional activation domain (AD). The *GFP* gene in pGNG1 is downstream of the recognition sequence for LexA and thus acts as a reporter for interaction between a peptide fused to LexA and a peptide fused to B42.

All viral genes cloned into pLexA as fusions to LexA were subjected to a series of control experiments to assess their suitability as baits in screening. To test for the ability of viral proteins to activate expression of the reporter genes without the presence of B42 (i.e. autoactivation), yeast cells were sequentially transformed (small-scale transformation; Section 2.2.16) with the pLexA vector expressing the protein of interest and pGNG1. Cells were then assessed for prototrophic growth on leucine deficient medium and for *GFP* expression by visualisation using a hand-held UVL-21 Blak-Ray long-wave UV lamp (Ultra-Violet Products, Inc.; San Gabriel, CA). To confirm that the bait LexA fusion protein was being synthesised in yeast, targeted to the nucleus, and binding LexA operator sequences, repression assays were performed (Fig. 2.2). pLexA vectors and pJK101 were transformed sequentially (small-scale transformation; Section 2.2.16) into yeast. pJK101 contains the GAL upstream activating sequence (UAS) from the *GAL1* gene followed by LexA operators upstream of the *lacZ* ORF. Yeast transformed with pJK101 alone will have significant β-galactosidase activity when grown on medium in which galactose is the sole carbon source because of binding of endogenous GAL4 to the GAL UAS. However, LexA-fusion proteins that are made, enter the nucleus, and bind the LexA operators will block activation from the GAL UAS and repress β-galactosidase activity up to 5-fold.

After assessing the suitability of viral proteins for use as bait, yeast two-hybrid screens were carried out. Yeast cells were sequentially transformed by the small-scale lithium acetate method (Section 2.2.16) with pGNG1 and the pLexA vector of interest. Cells were then transformed by the library-scale method (Section 2.2.16) with a pB42AD tomato (v. Rio Grande) cDNA library fused to the B42 activation domain (Zhou et al., 1995). Colonies were selected on agar plates lacking uracil, histidine, tryptophan, and

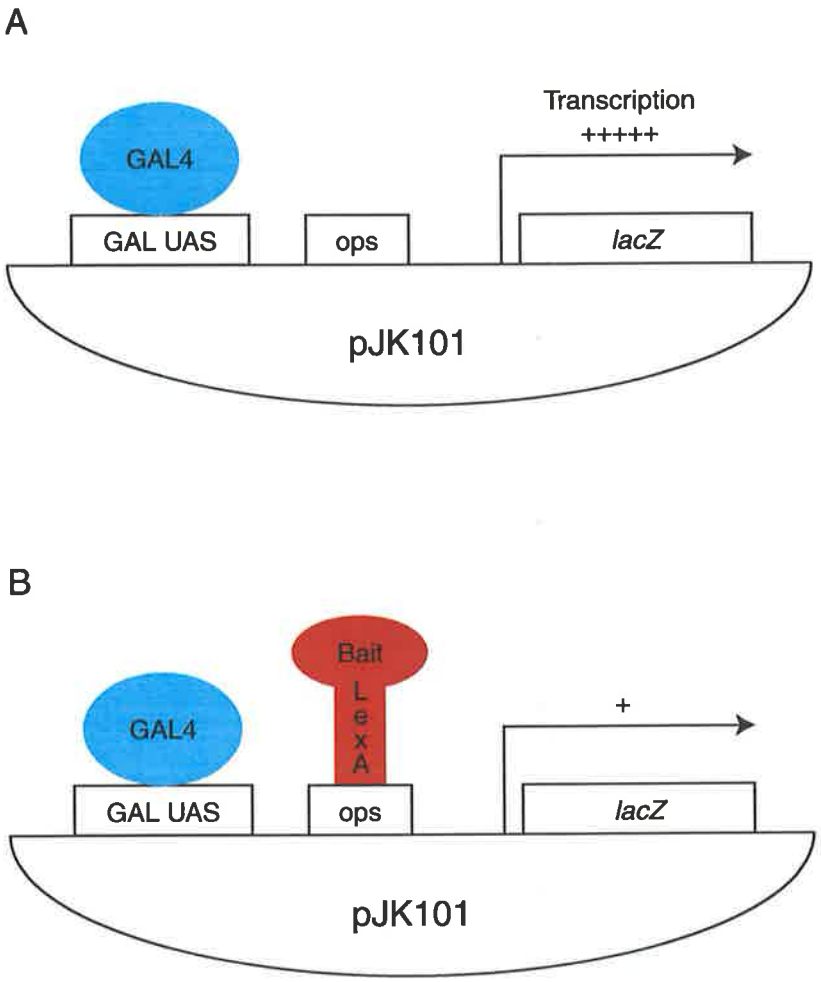


Figure 2.2. Repression assay for DNA binding of LexA-fusion proteins. **(A)** The plasmid pJK101 contains the upstream activating sequence (UAS) from the *GAL1* gene followed by *LexA* operators (ops) upstream of the *lacZ* coding sequence. Yeast transformed with pJK101 alone will have significant β-galactosidase activity when grown on medium in which galactose is the sole carbon source because of binding of endogenous GAL4 to the GAL UAS. **(B)** LexA-bait fusion proteins that are made, enter the nucleus, and bind the ops will block activation from the GAL UAS and repress β-galactosidase activity up to 5-fold (reproduced from Golemis et al., 1996).

leucine, but containing galactose and raffinose to induce the GAL1 promoter driving expression of the tomato cDNA/B42 fusions. Large colonies appearing within 5 days and exhibiting *GFP* expression were spread on plates lacking uracil, histidine, and tryptophan, and then transferred back to plates selecting for activation of the *LEU2* gene to remove false positives. Cells were then grown in media lacking tryptophan to select for pB42AD and yeast plasmid DNA purified (Section 2.2.14). *E. coli* KC8 were transformed with purified pB42AD plasmid DNA, since this strain is *trp1*$^{-}$ and its defect can be complemented by the *TRP1* gene present in pB42AD. To further eliminate false positives, plasmid DNA purified from KC8 was transformed back into displayYEAST-L containing pGNG1 and pLexA-bait plasmids and the activation of *GFP* and *LEU2* reassessed.

2.2.18 Immunoblotting of proteins expressed in bacteria

Immunoblotting was used to monitor 6xHis- or calmodulin binding peptide (CBP)-fusion protein production in bacteria or to detect proteins in reactions from *in vitro* binding assays. Protein samples were first electrophoresed in 4-20% Tris-glycine-SDS polyacrylamide gels (Life-Gels, Clarkston, GA) at 150 V for 1-1.5 h. The gel was assembled into a sandwich as follows: gel-sized sponge, two pieces of gel-sized Whatman 3MM paper, gel, 0.45 μm Immobilon P polyvinylidene fluoride (PVDF) membrane, two pieces of gel-sized Whatman 3MM paper, and another gel-sized sponge. The sandwich was placed into a transfer tank containing 1 L of transfer buffer (Table 2.1) with the gel side facing the negative electrode and transferred at 150-250 mA for 1.5-3 h. After disassembling the sandwich, the membrane was placed into a clean container and washed with distilled water for 5 min with gentle agitation. Blocking was carried out in 20 ml PBS (Table 2.1) containing 5% skim milk powder for at least 2 h

with gentle agitation. After washing the membrane for 3x 5 min with PBS-Tween, anti-polyHistidine (Sigma, St. Louis, MO) antibody, anti-FLAG antibody (Sigma) or a mixture of both antibodies were applied at a concentration of 1:4,000 in PBS-Tween containing 3% milk powder and incubated with the membrane for 1-2 h with gentle agitation. After washing the membrane for 3x 5 min with PBS-Tween, goat anti-Mouse IgG-horseradish peroxidase (HRP) conjugate was applied at a concentration of 1:20,000 in PBS-Tween containing 3% milk powder and incubated with the membrane for 1-2 h with gentle agitation. The membrane was washed again for 3x 5 min with PBS-Tween and HRP activity detected using SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology, Rockford, IL).

2.2.19 Immunoblotting of proteins expressed in yeast

To monitor fusion protein production in yeast, 0.3 ml of yeast culture was pelleted by centrifugation at 16,100*g* for 1 min, resuspended in 500 µl sterile water, pelleted again, and resuspended in 100 µl SDS-PAGE loading buffer (Table 2.1). After boiling the crude protein samples for 5 min, 20-50 µl was size fractionated on 4-20% Tris-glycine-SDS polyacrylamide gels (Gradipore, Frenchs Forest, Australia). Transfer and detection of fusion proteins was performed as described in Section 2.2.19. Rabbit anti-LexA polyclonal antibody (Invitrogen) was used at a concentration of 1:3,000 to detect LexA-fusion proteins and mouse anti-HA monoclonal antibody (Sigma) was used at a concentration of 1:5,000 to detect B42-fusion proteins. The secondary antibodies, used at a concentration of 1:20,000, were donkey anti-Rabbit IgG-HRP conjugate (Sigma) or Goat anti-Mouse IgG-HRP conjugate.

2.2.20 CTAB mini-DNA extraction

A small, new emerging leaf was ground in a microcentrifuge tube by hand using a grinding tip. 100 µl of nucleus lysis buffer (Table 2.1) was added and the mixture was vortexed thoroughly for 10 s. After adding 10 µl 5% sarkosyl and briefly vortexing, tubes were incubate at 65°C for 10 min. 100 µl chloroform:IAA (24:1) was added and tubes vortexed thorougly to form emulsion. After centifugation at 16,100*g* for 10 min, the upper phase was transferred to a new tube containing 100 µl cold isopropanol, mixed, and spun again for 10 min. The pellet was washed with 500 µl 70% ethanol, dried, and resuspended in 100 µl TE. 1 µl was used in PCR reactions.

2.2.21 Preparation of DNA samples for sequencing

DNA sequencing reactions were carried out using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Norwalk, CT). The following day, extension products were precipated by adding 80 µl of 75% isopropanol, incubating for 20 min at room temperature, and centrifuging at 16,100*g* for 20 min. After discarding the supernatant and adding another 250 µl of 75% isopropanol, the tubes were centrifuged at 16,100*g* for 5 min. The samples were then aspirated, dried under vacuum for 15 min, and sent to the Institute of Medical and Veterinary Science (Adelaide, Australia) for analysis.

2.2.22 Synthesis of ^{32}P-labelled nucleic acid probes

Probes for hybridisation analysis were synthesised by random priming using α-^{32}P-dCTP in a Rediprime II DNA Labeling System (Amersham Biosciences, Little Chalfont, England). Prior to hybridisation, the labelled nucleic acid was purified using

MicroSpin S-200 HR Columns (Amersham Biosciences) according to the manufacturer's instructions and denatured by heating at 95°C for 5 min.

2.2.23 RNA gel blot analysis

Total nucleic acid was extracted from plant tissue as described by Selth *et al.* (2004). 2-20 µg of sample was denatured with denaturing buffer (Lewandowski and Dawson, 1998) for 10 min at 65°C, cooled on ice, mixed with loading buffer and analysed on 1.2% agarose MOPS gels containing 2 M formaldehyde (Section 2.2.2.2). Transfer of RNA from the gel to Zeta-Probe membrane (Bio-Rad) was performed overnight in 10× SSC using a TurboBlotter (Schleicher & Schuell, Dassel, Germany). The RNA was stabilised on the membrane by cross-linking using a UV Stratalinker™ 1800 (Stratagene, La Jolla, CA). RNA on membranes was hybridised with specific ^{32}P labelled probes (Section 2.2.22) in 5-15 ml of hybridisation buffer (Table 2.1) at 65°C overnight. Membranes were washed twice at 65°C in 2× SSC/0.1% SDS and once at 65°C in 0.1× SSC/0.1% SDS. Radioactively labeled bands were detected by autoradiography using Biomax MS Scientific Imaging Film (Kodak, Rochester, NY).

2.2.24 Dot blot hybridisation

Dot-blot hybridisation was used as a quick, semi-quantitative method to assess the relative levels of TLCV genomic DNA in plants. 100 mg of tissue obtained from emerging leaves was ground in 200 µl of 0.5 M NaOH with a small amount of sterile sand, left for 30 min at room temperature, and then centrifuged at 16,100*g* for 10 min. 4 µl of the supernatant was dotted onto Zeta-Probe membrane, which was allowed to air dry, washed once with chloroform and twice with 2× SSC, and cross-linked. Detection

of TLCV sequences by hybridisation and autoradiography was performed as described in Sections 2.2.23.

2.2.25 Analysis of GFP-fusion proteins by microprojectile bombardment

pART7 (Gleave, 1992) based vectors were used to transiently express GFP fusion proteins in onion tissue after microprojectile delivery. Onion epidermal strips on agar containing Murashige and Skoog (MS) Salt Mixture (Invitrogen) were bombarded with vectors. For four shots, 400 µg of gold particles in 100 µl ethanol were vortexed for 2 min, spun down for 10 s in a microfuge, drained, washed twice with sterile water, and resuspended in 25 µl 40% glycerol. While gently vortexing, 4 µl of the plasmid solutions (400 ng/µl), 10 µl of cold 0.1 M spermidine and 25 µl 2.5 M $CaCl_2$ were added dropwise and the resulting mixture incubated on ice for 10 min. The particles were spun down, washed with 70% ethanol, resuspended in 24 µl cold 100% ethanol, and 6 µl aliquots were placed onto sterile filter holders. After sterilising the gun chamber with 70% EtOH, plates containing onion strips on MS media were placed inside, covered with a sterile mesh, and bombarded with a pressure of 650 kPa after evacuating the chamber to 90 kPa. After bombardment, tissue was stored in the dark for 48 h and GFP fluorescence visualised using a Bio-Rad Radiance 2100 Confocal Laser Scanning Microscope System (The Hanson Institute Detmold Family Trust Cell Imaging Centre, Institute for Medical and Veterinary Science, Adelaide, Australia). The excitation wavelength used for GFP detection was 488nm.

Chapter 3 - SlNAC1, a novel tomato NAC domain protein, interacts with TLCV replication enhancer protein and facilitates viral replication

3.1 Introduction

To identify host proteins involved in the replication of WDV, Xie *et al.* (1999) used WDV RepA as bait in a yeast two-hybrid screen of a wheat library and isolated two homologous proteins, designated GRAB1 and GRAB2. Overexpression of the *GRAB* genes in cultured cells inhibited WDV DNA replication, suggesting that RepA disrupts a GRAB-mediated response which represses viral infection. Both of the GRABs are members of the recently-identified NAC family found in many plant species but, so far, not in other eukaryotes. The name is derived from the three type members, NO APICAL MERISTEM from *Petunia* (Souer et al., 1996) and ATAF and CUP-SHAPED COTYLEDON (CUC) from *Arabidopsis thaliana* (Aida et al., 1997). Since the identification of these proteins, many more genes encoding the NAC domain have been found; Ooka *et al.* (2003) studied the rice and *Arabidopsis* genomes and found 75 and 105 predicted *NAC* genes in each species respectively. NACs share a common structure consisting of a conserved amino-terminal region (the NAC domain) and a highly variable carboxy terminus. They have roles in such diverse processes as pattern formation in embryos (Souer et al., 1996), flower development (Sablowski and Meyerowitz, 1998), leaf senescence (John et al., 1997), auxin-dependent lateral root formation (Xie et al., 2000) and plant defence (Collinge and Boller, 2001).

Geminiviral REn proteins are able to increase viral DNA accumulation and enhance infectivity and symptom expression (Section 1.1.9.3). While little molecular information regarding this process is available, REn can bind pRBR and PCNA (Section 1.2.3), suggesting that it modifies the host cellular environment to facilitate viral replication. Given the apparent multifunctionality of REn, it may interact with other host proteins to achieve its function. To address this hypothesis, a tomato library was screened for proteins which interact with TLCV REn. This chapter describes a new member of the NAC domain family, SlNAC1, that was found to interact with REn in yeast and *in vitro*. Evidence is presented implicating SlNAC1 in REn-mediated enhancement of viral DNA accumulation. These findings are discussed in relation to our current understanding of the endogenous function of NACs and their possible role in geminivirus pathogenesis.

3.2 Materials and methods

3.2.1 Yeast two-hybrid screening

The *REn* ORF was amplified by PCR using REn-F-EcoRI and REn-R-XhoI, digested with *Eco*RI and *Xho*I and ligated into similarly digested pLexA to generate pLexA-REn. This construct was used as bait to screen a tomato Rio Grande cDNA library (Zhou et al., 1995) as described in Section 2.2.17.

To analyse the domains of the REn and SlNAC1 proteins responsible for their interaction, truncations of their genes were cloned into pLexA and pJG4-5 respectively. The fragments amplified were as follows: REn encoding amino acids 1-70 (REn-F-EcoRI and REn_{1-70}-R-XhoI), REn 40-120 (REn_{40-120}-F-EcoRI and REn_{40-120}-R-XhoI), REn 90-134 (REn_{90-134}-F-EcoRI and REn-R-XhoI), SlNAC1 1-70 (SlNAC1-F-EcoRI

and SlNAC1$_{1-70}$-R-XhoI), SlNAC1 1-170 (SlNAC1-F-EcoRI and SlNAC1$_{1-170}$-R-XhoI), SlNAC1 71-301 (SlNAC1$_{71-301}$-F-EcoRI and SlNAC1-R-NotI). Products were digested with EcoRI/XhoI (REn 1-70, REn 40-120, REn 90-134, SlNAC1 1-70, and SlNAC1 1-170) or EcoRI/NotI (SlNAC1 71-301) and ligated into similarly digested pLexA or pJG4-5.

To generate a pLexA vector expressing TGMV REn fused to LexA, the TGMV *REn* ORF was amplified using primers REn-F-EcoRI and TGMV_REn-R-XhoI, digested with EcoRI/XhoI and ligated into similarly digested pLexA.

In experiments mapping the activating domains of REn and SlNAC1, fusion protein production in yeast was monitored as described in Section 2.2.19.

3.2.2 Analysing the transcriptional activation function of SlNAC1

The *SlNAC1* truncation sequences described above were transferred into pLexA to delineate the putative transcriptional activation domain of SlNAC1 in yeast. Two other *SlNAC1* sequences were cloned into EcoRI/NotI digested pLexA for this yeast one-hybrid study: full length *SlNAC1* (amplified using primers SlNAC1-F-EcoRI and SlNAC1-R-NotI) and a fragment encoding amino acids 1-240 (SlNAC1-F-EcoRI and SlNAC1$_{1-240}$-R-NotI).

pLexA vectors expressing full-length and truncated SlNAC1 peptide sequences fused to LexA were introduced into yeast containing pSH18-34, a reporter plasmid which contains eight LexA operators that direct transcription of the *lacZ* gene (Golemis et al., 1994). Quantitative β-galactosidase assays from liquid cultures were performed

according to the Yeast Protocols Handbook 2001 (Clontech, Palo Alto, CA) using *o*-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate. A Microplate Reader 450 (Bio-Rad, Hercules, CA) was used to measure accumulation of the *o*-nitrophenol product. One unit of β-galactosidase is defined as the amount of activity hydrolysing 1 nmol of ONPG per min per cell. The assay was carried out twice using three independent transformants for each construct. The positive control plasmid used in this study, pSH17-4, expresses a LexA fusion to the GAL4 AD (Golemis et al., 1994).

3.2.3 Production of recombinant proteins and *in vitro* binding experiments

6xHis-tagged REn and C2 proteins were produced using the pQE30 vector (Qiagen, Clifton Hill, Australia). The coding region of *REn* was amplified using oligonucleotides REn-F-BamHI and REn-R-HindIII, digested with BamHI/HindIII, and ligated into similarly digested pQE30 to generate pQE30-REn. pQE30-C2 was constructed in the same way, using oligonucletides C2-F-pQE30 and C2-R-pQE30 to amplify the *C2* gene.

6xHis-REn recombinant protein was purified using a protocol developed by Behjatnia *et al.* (1998) for the preparation of 6xHis-Rep protein, with minor modifications. Briefly, *E. coli* M15 cells were transformed with pQE30-REn, grown to an OD_{600} of 0.9, and induced with 1 mM IPTG for 3 h at room temperature. Cells were harvested, resuspended in Ni-NTA binding buffer (Table 2.1), and lysed by 1 mg/ml lysozyme, freeze/thawing, and sonication. Crude soluble protein was retrieved by centrifugation and incubated with Ni-NTA agarose (Qiagen) with gentle shaking for 1 h at 4°C. The suspension was then loaded onto a 1.2 cm diameter column, washed with 10 volumes of Ni-NTA washing buffer (Table 2.1), and purified 6xHis-REn eluted with Ni-NTA

elution buffer (Table 2.1). For the production of 6xHis-C2, cells were transformed with pQE30-C2, grown to an OD_{600} of 0.7, and induced with 1 mM IPTG for 3 h at 37°C. Total soluble protein was extracted using sarkosyl by the method of Frangioni and Neel (1993) and dialysed against Ni-NTA binding buffer prior to incubation with Ni-NTA agarose and purification of the recombinant protein as described above.

Expression of CBP-tagged SlNAC1 and SlUPTG1 proteins containing a FLAG epitope was achieved using the pCAL-n-FLAG vector (Stratagene, La Jolla, CA). The coding region of *SlNAC1* was amplified by PCR using the oligonucleotides SlNAC1-F-BamHI and SlNAC1-R-NotI, digested with BamHI, and ligated into BamHI/SmaI-digested pCAL-n-FLAG vector to generate pCAL-SlNAC1. The *SlUPTG1* ORF was amplified using the oligonucleotides SlUPTG-F-CAL and SlUPTG-R-CAL, digested with EcoRI/XhoI, and ligated into similarly digested pCAL-n-FLAG to yield pCAL-SlUPTG1.

E. coli B834-pLysS cells were transformed with pCAL-SlNAC1 and pCAL-SlUPTG1, grown to an OD_{600} of 0.7, and induced with 0.5 mM IPTG at 37°C for 3 h. Crude soluble protein was extracted using sarkosyl as described by Frangioni and Neel (1993).

Binding experiments were carried out by adding 50 ng of a purified 6xHis-tagged protein and 200 ng of total soluble protein extracted from cells induced to express the CBP-tagged protein of interest to 10 µl of Ni-NTA agarose in 300 µl of binding assay buffer (Table 2.1) in an eppendorf tube. Tubes were then mixed gently on a rotating platform at 4°C for 40 min. The resin was washed three times by brief centrifugation and resuspension in 400 µl binding buffer, resuspended in 50 µl of SDS-PAGE sample

loading buffer (Table 2.1), and incubated at 94°C for 10 min. Aliquots (10 µl) of eluate from the pelleted beads were size fractionated on 4-20% Tris-glycine-SDS polyacrylamide gels and fusion proteins detected as described in Section 2.2.18.

3.2.4 Analysis of GFP-fusion proteins by microprojectile bombardment

A variant of the shuttle vector pART7 (Gleave, 1992) termed pART7-C'gfp, which contains the full-length *GFP* ORF (lacking the stop codon) upstream of the multiple cloning site (T. Franks, unpublished results), was used to transiently express REn:GFP and SlNAC1:GFP fusion proteins in onion tissue. Full-length *REn* was amplified using primers REn-F-EcoRI and REn-R-XbaI, and the *SlNAC1* ORF was amplified using primers SlNAC1-F-EcoRI and SlNAC1-R-XbaI. After restriction enzyme digestion with EcoRI and XbaI, fragments were ligated into similarly digested pART7-C'gfp to generate C-terminal fusions with GFP. Also used in this experiment were pART7-ATG:GFP, which expresses free GFP (T. Franks, unpublished results), and pBI121-H2B:YFP, which expresses *Arabidopsis* HISTONE 2B (H2B) fused to the GFP yellow variant YFP (H2B:YFP; Boisnard-Lorig et al., 2001).

The localisation of GFP fusion proteins was determined by bombarding GFP vectors into onion epidermal strips and analysing by confocal microscopy as described (Section 2.2.25).

3.2.5 Analysis of *SlNAC1* gene expression

Three week old tomato plants were inoculated with TLCV, TYLCSV, or the TLCV REn-mutant (Rigden et al., 1996) using *A. tumefaciens* (Grimsley et al., 1987). Total nucleic acid was extracted at various time points and subjected to RNA gel blot analysis

as described (Section 2.2.23). To detect replication of the REn-mutant, the same samples were analysed by southern blotting as described previously (Dry et al., 1993).

The binary vector pART27 was used to express individual TLCV genes to analyse their effect on *SlNAC1* mRNA production. Primers to amplify *C1* (C1-F-SmaI and C1-R-HindIII), *C2* (C2-F-pART7 and C2-R-BamHI), and *REn* (REn-F-pART7 and REn-R-pART7) were designed. The PCR products were digested with SmaI/HindIII (*C1*), BamHI (*C2*), and KpnI/XbaI (*REn*) and ligated into similarly digested pART7. Fragments containing the CaMV 35S promoter upstream of the TLCV gene were released from these plasmids by digestion with NotI and ligated into pART27 previously cut with NotI and dephosphorylated. The resultant vectors were designated p35S-C1, p35S-C2, and p35S-REn. Young tomato leaves were infiltrated with *A. tumefaciens* C58 cells containing the p35S constructs as described (Selth et al., 2004). At 5 days post-infiltration, *SlNAC1* expression was analysed by RNA gel blot analysis.

3.2.6 *In situ* hybridisations

Templates for the generation of RNA probes were constructed as follows. A fragment comprising nucleotides 401-906 of *SlNAC1* was amplified using primers SlNAC1-pN6-1 and SlNAC1-pN6-2 and ligated into pGEM-T-Easy (Promega). The full-length TLCV *V2* ORF was amplified using primers V2-F-BamHI and V2-R-HindIII and ligated into pGEM-T-Easy. Plasmids were linearised with NdeI (SlNAC1) or SalI (V2) and transcribed with T7 RNA polymerase. RNA probes labelled with fluorescein-12-UTP (for *SlNAC1* probes) and digoxigenin (DIG)-11-dUTP (for TLCV probes) were prepared using fluorescein or DIG RNA Labeling Mix respectively (Roche Diagnostics).

Plant tissue was collected from TLCV-infected plants three weeks post-inoculation. Preparation of sections and hybridisation of DIG- and fluorescein-labeled probes was carried out as described by Guerin *et al.* (2000). Probes were detected using Fast Red (Roche Diagnostics) or Western Blue substrates (Promega) according to the manufacturer's instructions. For dual-colour in situ hybridisations, probes were applied simultaneously and detected sequentially (Jowett, 2001).

3.2.7 Analysis of TLCV DNA replication

The vector pART27 (Gleave, 1992) was used to transiently overexpress *SlNAC1* in *N. benthamiana* leaf strips. The entire *SlNAC1* ORF was amplified by PCR using primers SlNAC1-F-KpnI and SlNAC1-R-XbaI, digested with KpnI and XbaI, and ligated into KpnI/XbaI-digested pART7 to yield pART7-SlNAC1. A DNA fragment containing the CaMV 35S promoter and the *SlNAC1* ORF was released by NotI digestion and ligated into similarly-digested pART27 to generate p35S-SlNAC1. *A. tumefaciens* strain C58 was transformed separately with p35S-SlNAC1, empty pART27, and a Bin19 construct containing a TLCV 1.1mer (Bin19-TLCV1.1; Rigden et al., 1996). Cultures were grown at 28°C for 48 h and used in leaf strip transient replication assays as described (Dry et al., 1997). *A. tumefaciens* containing Bin19-TLCV1.1 was co-cultivated with leaf strips in combination with *A. tumefaciens* harbouring empty pART27 or p35S-SlNAC1 at a ratio of 1:2. Viral replication in agroinoculated tissues was analysed by southern blotting as described previously (Dry et al., 1997).

3.2.8 Quantitation of *SlNAC1* mRNA expression by semi-quantitative reverse transcription-PCR

Total RNA from *N. benthamiana* leaf strips was prepared using an RNeasy Plant Mini Kit (Qiagen) which includes a treatment with RNase-free DNase. Semi-quantitative reverse transcription (RT)-PCR was carried out using a SuperScript One-Step RT-PCR kit (Invitrogen) and 80 ng of RNA as template. The *SlNAC1* primers (SlNAC1-F-KpnI and SlNAC1-R-XbaI) were used at a final concentration of 0.2 μM. The internal control, ubiquitin, was amplified with primers Ubi3-F and Ubi3-R (Jin et al., 2002) used at a final concentration of 0.05 μM. RT reaction mix without reverse transcriptase served as a negative control. Following the linear phase of DNA amplification (26 cycles), the PCR products were examined by electrophoresis in a 2.0% agarose gel.

3.3 Results

3.3.1 Identification of a novel NAC domain protein that interacts with REn

To identify host proteins interacting with the TLCV REn protein, a yeast two-hybrid screen of a tomato cDNA library fused to the B42 AD-encoding sequence (Zhou et al., 1995) was carried out using REn fused to the LexA DBD as bait. A total of 2 x 10^6 transformants were assayed for leucine prototrophic growth and *GFP* expression. Of these transformants, one was able to activate both reporter genes. Plasmid DNA was rescued, verified by retransformation into yeast with the bait, and the sequence of the insert determined. The plasmid contained a cDNA insert of 1304 bp encoding a predicted full-length translation product of 301 amino acids (Appendix 1). Nucleotides 52-597 of the cDNA are identical to an EST generated from tomato carpel tissue (GenBank accession number AI486942). A BLAST query of the protein sequence revealed that the N-terminal 169 amino acid residues contained the five conserved

blocks of homology that characterise the NAC domain (Fig. 3.1A, boxed). Based on this defining characteristic, the protein was named SlNAC1 for *Solanum lycopersicum* NAC1 (GenBank accession number AY498713).

A recent phylogenetic analysis of the NAC domains from known NAC family proteins and putative *Arabidopsis* and rice NACs separated them into 18 subgroups (Ooka et al., 2003). The NAC domains from SlNAC1 and other known NAC family proteins were compared. According to dendograms obtained by the neighbor-joining method (Fig. 3.1B) and the maximum-parsimony method (data not shown), SlNAC1 falls into the so-called ATAF subgroup. The C-terminal region of NAC proteins, termed the transcriptional activation region (TAR), is highly divergent but Ooka *et al.* (2003) found 13 common motifs (CMs) in 12 of the 18 subgroups. Members of the ATAF subgroup contain the sequence EVQS[E/x]PK[W/l], which is also present in SlNAC1 (Fig. 3.1A, boxed and labelled TAR-CM). This supports our classification of SlNAC1 into this subgroup. Analysis of the primary sequence of SlNAC1 using PSORT II (http://bioweb.pasteur.fr/seqanal/interfaces/psort2.html) identified a putative classical (SV40 large T antigen-type) nuclear localisation signal (NLS) in subdomain C from amino acids 74-80 (Fig. 3.1A, underlined). This sequence, PRDRKYP, was conserved amongst 12 NACs in a study carried out by Kikuchi *et al.* (2000), suggesting that it may be functional *in vivo*.

To map the domains responsible for the interaction between REn and SlNAC1, truncations of the genes encoding both proteins were made and cloned into pLexA and pB42AD to create fusions with the LexA DNA BD and B42 AD respectively. The secondary structure of REn, predicted using PSIPRED

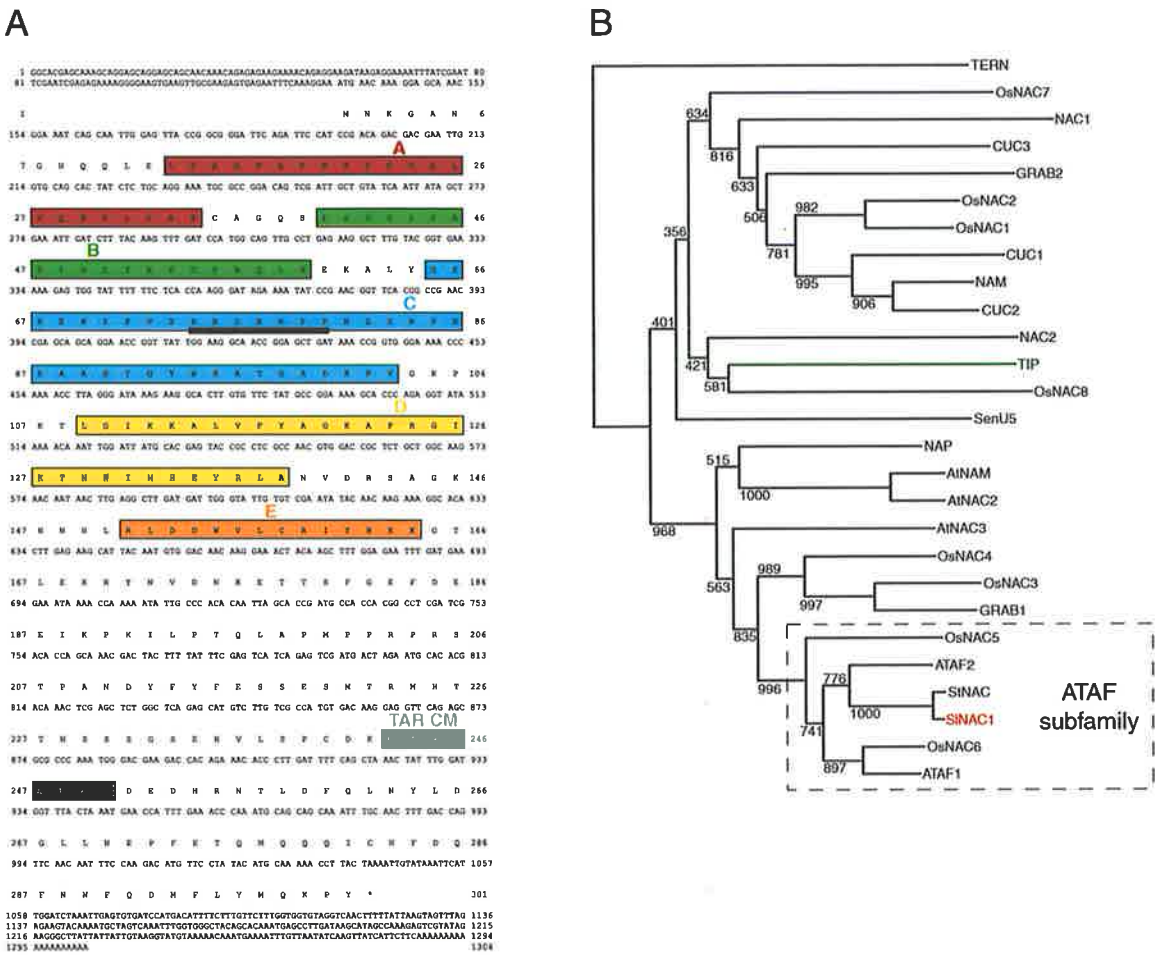


Figure 3.1. Nucleotide sequence of *SlNAC1* and alignment of its putative translation product with other NAC domain proteins. **(A)** Nucleotide and amino acid sequences of *SlNAC1* gene. The five subdomains (A-E) comprising the NAC domain are shown in coloured boxes. A putative nuclear localisation signal is indicated by a bold line under the sequence (PRDRKYP). The transcriptional activation region-common motif (TAR-CM) of the ATAF subgroup is also boxed. **(B)** The predicted amino acid sequence of SlNAC1 (Fig. 1A) and known NAC family proteins were subjected to phylogenetic analysis. Multiple sequence alignment of the proteins was conducted using ClustalX (Thompson et al., 1997), and phylogenetic analysis was carried out by the neighbor-joining method (Saitou et al., 1987). A bootstrap analysis of 1000 resampling replicates was conducted with ClustalX. The rooted phylogenetic tree was displayed using the NJPlot program included with ClustalX. The gene names and references for other NACs are as follows: *Arabidopsis thaliana*: ATAF1 and ATAF2 (Aida et al., 1997), AtNAC2 (Takada et al., 2001), AtNAC3 (Takada et al., 2001), AtNAM (Duval et al., 2002), CUC1 (Takada et al., 2001), CUC2 (Takada et al., 2001), CUC3 (Vroemen et al., 2003), NAC2, NAP (Sablowski et al., 1998), TIP (Ren et al., 2000); *Oryza sativa*: OsNAC1-OsNAC8 (Kikuchi et al., 2000); *Petunia hybdrida*: NAM (Souer et al., 1996); *Solanum lycopersicum*: SenU5 (John et al., 1997); *Solanum tuberosum*: StNAC (Collinge and Boller, 2001); and *Triticum* sp.: GRAB1 and GRAB2 (Xie et al., 1999).

(http://bioinf.cs.ucl.ac.uk/psipred/), contains three α-helices found between amino acids 56-65, 79-95, and 101-116. Since α-helices are frequently important in protein-protein interactions (Settlage et al., 2001), three truncations of REn based on the location of these putative structures were generated (Fig. 3.2A). REn 1-70 contains only the first helix, REn 40-120 contains all three helices, while REn 90-134 contains the third helix and a part of the second. Three truncations of SlNAC1 were made based on the location of NAC subdomains (Fig. 3.2A): SlNAC1 1-70 contains subdomains A, B, and a small part of C; SlNAC1 1-170 contains all of the five subdomains that make up the NAC domain; and SlNAC1 71-301 contains subdomains D and E and the majority of C, and all of the variable C-terminus.

Each of the REn and SlNAC1 truncations as well as the full-length proteins were co-expressed in yeast and their interaction assayed by leucine prototrophic growth and GFP expression. REn 1-70 was able to interact with full-length SlNAC1 (Fig. 3.2C), while the other two REn truncations could not, suggesting that the first putative α-helix of REn may be involved in SlNAC1 binding. None of the three truncations of SlNAC1 was able to interact with REn in yeast. This may indicate the involvement of a larger proportion of SlNAC1 in the interaction or reflect structural constraints imposed on the functional REn-interacting domain. Immunoblot analysis of yeast cells demonstrated that non-interacting REn and SlNAC1 truncations were expressed at levels similar to those of interacting proteins (Fig. 3.2B), confirming that negative results were not due to an absence of protein.

The general significance of SlNAC1 binding to TLCV REn was examined by testing whether SlNAC1 could also interact with REn encoded by TGMV. TGMV is a

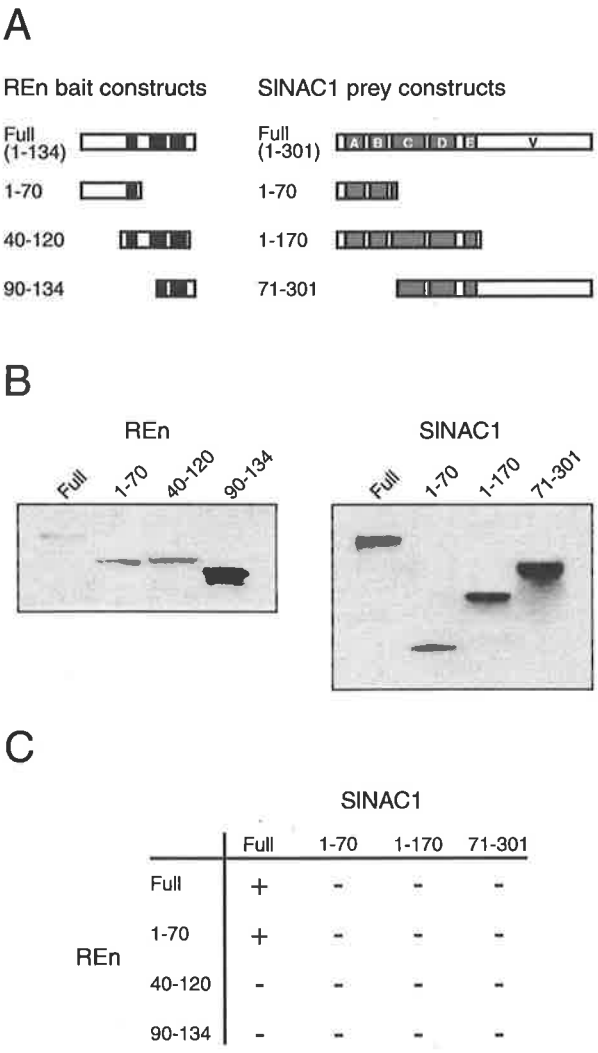


Figure 3.2. Deletion analysis of REn and SlNAC1 to identify regions required for interaction between the two proteins. **(A)** Diagrammatic representation of REn (bait) and SlNAC1 proteins (prey) tested for interaction. The REn proteins were expressed as LexA DNA binding domain fusions, and the SlNAC1 proteins were expressed as B42 activation domain fusions. The positions of three putative α-helices in REn are indicated by filled boxes. In SlNAC1, the positions of the NAC subdomains are shown in grey boxes (A-E) while the variable carboxy-terminus is denoted V. **(B)** Immunoblot analysis of yeast cells demonstrating that non-interacting REn-LexA fusions and SlNAC1-B42 fusions are expressed at levels similar to those of interacting fusion proteins. Total protein from yeast cultures containing different REn and SlNAC1 fusion proteins was extracted, fractionated on 4-20% SDS-polyacrylamide gels, and immunoblotted with anti-LexA (to detect REn-LexA fusions) or anti-HA (to detect SlNAC1-B42 fusions). **(C)** The N-terminal region of REn is important for SlNAC1 binding. Interaction was indicated by the ability of cells transformed with bait, prey, and pGNG1 plasmids to grow on medium lacking leucine. As an additional indicator of interaction, colonies were monitored for *GFP* expression by visualisation under ultraviolet light.

bipartite begomovirus encoding a REn protein that is 54.2% identical in sequence (65.6% similar) to TLCV REn. In the same yeast two-hybrid assay, TGMV REn also interacted with SlNAC1, as shown in Figure 3.3. Neither REn protein interacted with the control protein TLCV C2 (also designated AC2, L2, AL2, or TrAP) or with the AD alone.

3.3.2 SlNAC1 acts as a transcriptional activator in yeast

There is considerable evidence to suggest that NAC domain proteins function as transcription factors. First, the NAC proteins ATAF1 and ATAF2 (Souer et al., 1996), AtNAM (Duval et al., 2002), NAC1 (Xie et al., 2000), TIP (Ren et al., 2000), and a group of *Brassica napus* NACs (Hegedus et al., 2003) are able to activate transcription of a reporter gene in yeast, an activity mediated by the divergent C-terminal sequences. Second, AtNAM and NAC1 bind a specific DNA sequence found in the CaMV 35S promoter (Duval et al., 2002; Xie et al., 2000). Third, overexpression of *NAC1* in *A. thaliana* caused upregulation of the auxin-responsive genes, *AIR*3 and *DBP* (Xie et al., 2000), while *CUC1* activated the expression of genes involved in the development of the shoot apical meristem (Hibara et al., 2003). Finally, a nuclear localisation pattern has been observed for NAC1 (Xie et al., 2000).

Based on these data, the presence of a transcriptional activation domain in SlNAC1 was tested using yeast as an assay system. A SlNAC1 fusion to the LexA DBD was expressed in yeast with pSH18-34, a reporter plasmid which contains eight LexA operators that direct transcription of the *lacZ* gene (Golemis et al., 1994). Cells were assayed for β-galactosidase activity using a liquid culture assay (see Materials and Methods). As predicted, the LexA-SlNAC1 fusion was able to activate expression of

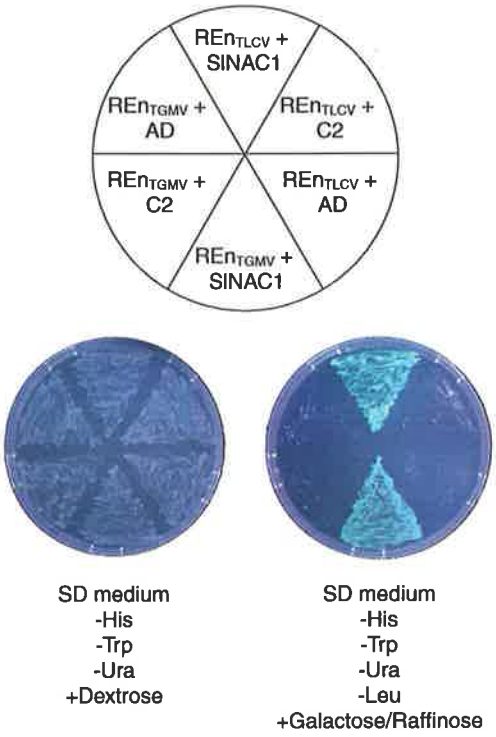


Figure 3.3. SlNAC1 interacts with both TLCV and TGMV REn. Yeast two-hybrid assays testing the ability of SlNAC1 to interact with REn of TLCV (REn_{TLCV}) and TGMV (REn_{TGMV}). Yeast coexpressing proteins as indicated (top) were grown on SD -His -Trp -Ura medium (bottom left plate), and interaction was tested by leucine prototrophy and *GFP* expression on an inductive carbon source (galactose and raffinose) (bottom right plate). REn proteins were fused to the LexA DNA binding domain while SlNAC1 was fused to the B42 activation domain (AD). Negative controls included REn_{TLCV} and REn_{TGMV} coexpressed with TLCV C2 fused to the AD, or coexpressed with AD alone.

the reporter gene, and its transactivation activity was at least as strong as the positive control, a LexA fusion to the GAL4 AD (Fig. 3.4A). Four truncations of the *SlNAC1* gene were fused to LexA to determine the domains required for transcriptional activation (Fig. 3.4A). This deletion analysis revealed that the variable carboxy-terminal region (amino acids 71-301) could activate transcription of *lacZ*, but more weakly than full-length SlNAC1. None of the N-terminal fragments (1-70, 1-170, and 1-240) was able to promote expression of *lacZ*. These data indicate that SlNAC1 has a transcriptional activation domain that is active in yeast and is located near its C-terminus. Immunoblotting confirmed that all LexA-SlNAC1 fusion proteins were produced at similar levels in yeast (Fig. 3.4B).

3.3.3 *In vitro* binding of SlNAC1 to TLCV REn

The specificity of the REn/SlNAC1 protein interaction was tested using an *in vitro* pull down assay. A 6xHis-REn fusion protein was expressed in *E. coli*, purified to homogeneity, and mixed with crude soluble protein extracted from *E. coli* cells induced to express a SlNAC1-CBP fusion protein containing a FLAG epitope (CBP-SlNAC1). The mixture was incubated with Ni-NTA agarose, washed extensively to remove unbound protein, resuspended in loading buffer, electrophoresed, and transferred to PVDF membrane. CBP-SlNAC1 and 6xHis-REn were detected by immunoblotting using antibodies directed against FLAG and polyHis respectively.

Bound CBP-SlNAC1 was detectable when incubated with 6xHis-REn (Fig. 3.5, lane 5). To determine the specificity of CBP-SlNAC1 binding, it was added to Ni-NTA resin alone (lane 8) or in combination with purified 6xHis-C2, another TLCV-encoded protein (lane 7). CBP-SlNAC1 was not detected in the bound fraction from either of

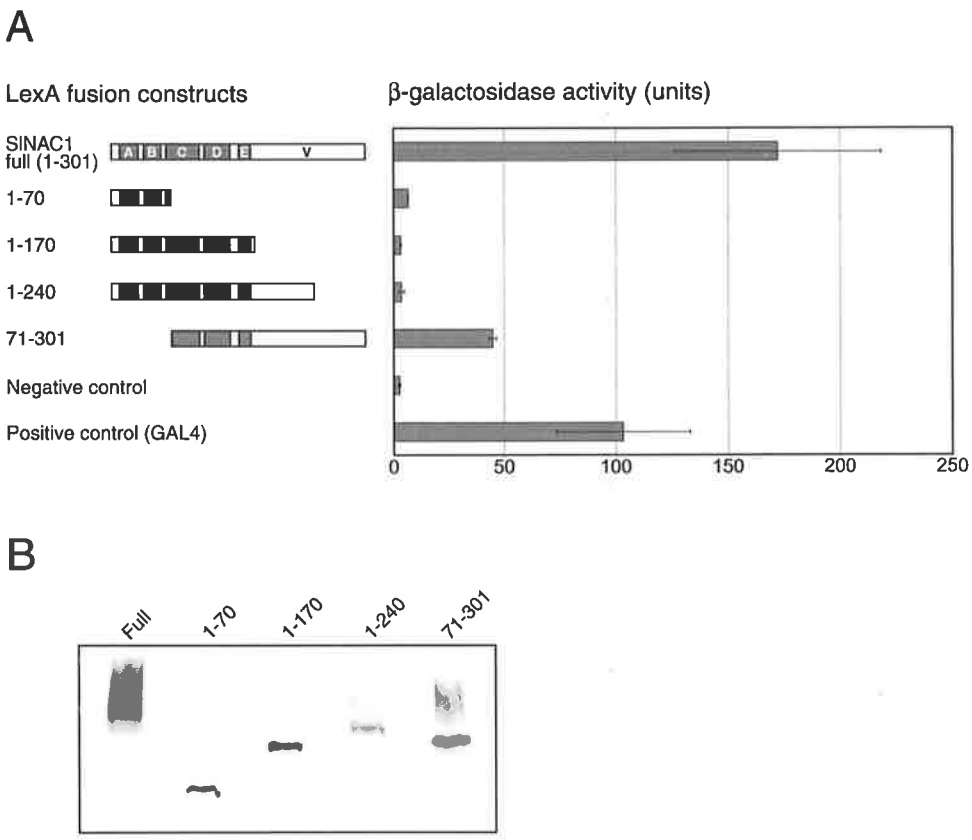


Figure 3.4. The divergent C-terminal region of SlNAC1 can activate transcription in yeast. **(A)** Regions of SlNAC1 able to activate transcription in yeast. The SlNAC1-LexA DNA binding domain fusion proteins are represented diagrammatically on the left, with the positions of the NAC subdomains shown in grey boxes (A-E) and the variable carboxy-terminus denoted V. The ability of SlNAC1-LexA fusion proteins to activate transcription in yeast is shown on the right. Transactivation was indicated by the ability of cells expressing SlNAC1-LexA fusion proteins and transformed with pGNG1 to grow on medium lacking leucine and to express *GFP*. **(B)** Immunoblot analysis of yeast cells demonstrating that non-transactivating SlNAC1-LexA fusions are expressed at levels similar to those of transactivating fusion proteins. Total protein from yeast cultures containing SlNAC1 fusion proteins was extracted, fractionated on 4-20% SDS-polyacrylamide gels, and immunoblotted with anti-LexA.

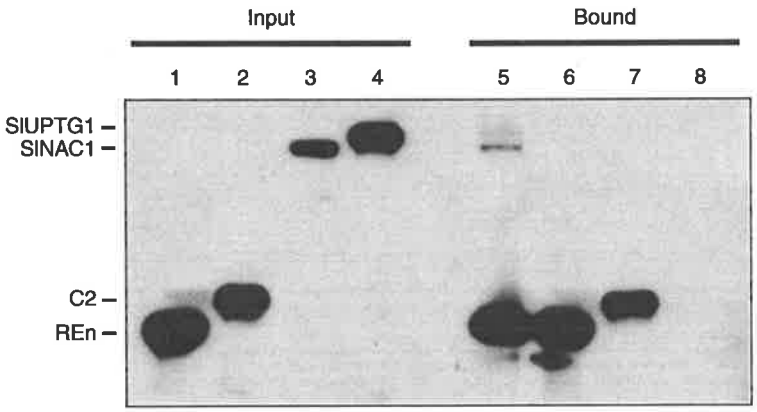


Figure 3.5. REn interacts with SlNAC1 *in vitro*. Purified 6xHis-tagged proteins were mixed with crude CBP-tagged protein mixtures, incubated with nickel-nitrilotriacetic acid agarose and washed extensively to remove any unbound protein. Bound protein was resuspended in loading buffer, fractionated by SDS-PAGE and analysed by immunoblotting using anti-polyHis and anti-FLAG (CBP-tagged proteins also contain a FLAG epitope) antibodies. Reactions were as follows: 6xHis-REn and CBP-SlNAC1 (lane 5), 6xHis-REn and CBP-SlUPTG1 (lane 6), 6xHis-C2 and CBP-SlNAC1 (lane 7), and CBP-SlNAC1 alone (lane 8). Protein inputs for each reaction are shown: 6xHis-REn (lane 1), 6xHis-C2 (lane 2), CBP-SlNAC1 (lane 3), and CBP-SlUPTG1 (lane 4).

these reactions, indicating that it was interacting specifically with 6xHis-REn. To determine if 6xHis-REn was specifically pulling down CBP-SlNAC1, it was mixed with total soluble protein extracted from cells induced to express CBP-SlUPTG1, a control CBP-tagged protein (lane 6). No CBP-SlUPTG1 was detectable in the bound fraction, indicating that 6xHis-REn does not indiscriminately bind abundant proteins in a mixture. SlUPTG1 is a tomato homologue of potato UDP-glucose:protein transglucosylase identified in another of our yeast two-hybrid screens (Chapter 4). All reactions were performed at least twice with the same results.

3.3.4 REn and SlNAC1 are targeted to the nucleus

The subcellular localisation of SlNAC1 and REn in plant cells was examined to investigate their potential roles in viral pathogenesis. Each of the ORFs were cloned in-frame with *GFP* downstream of the CaMV 35S promoter. The fusion proteins (REn:GFP and SlNAC1:GFP) were transiently expressed in onion epidermal cells following biolistic delivery of vector DNA and analysed by confocal microscopy.

Free GFP was distributed in both the cytoplasm and the nucleus of bombarded cells (Fig. 3.6, bottom row). In contrast, both REn:GFP and SlNAC1:GFP localised exclusively to nuclei (top and second row), which were clearly visible as dense ovoid structures when cells were viewed with differential interference optics (middle column). The distribution pattern of REn:GFP and SlNAC1:GFP matched that of the H2B:YFP fusion protein (third row), a control for nuclear localisation (Boisnard-Lorig et al., 2001), confirming the subcellular targeting of these proteins. NAC1 from *Arabidopsis* (Xie et al., 2000) and CmNACP from pumpkin (Ruiz-Medrano, 1999) were also found to be nuclear proteins, implying that this is a general characteristic of NAC proteins and supporting the hypothesis that they function as transcription factors. More importantly,

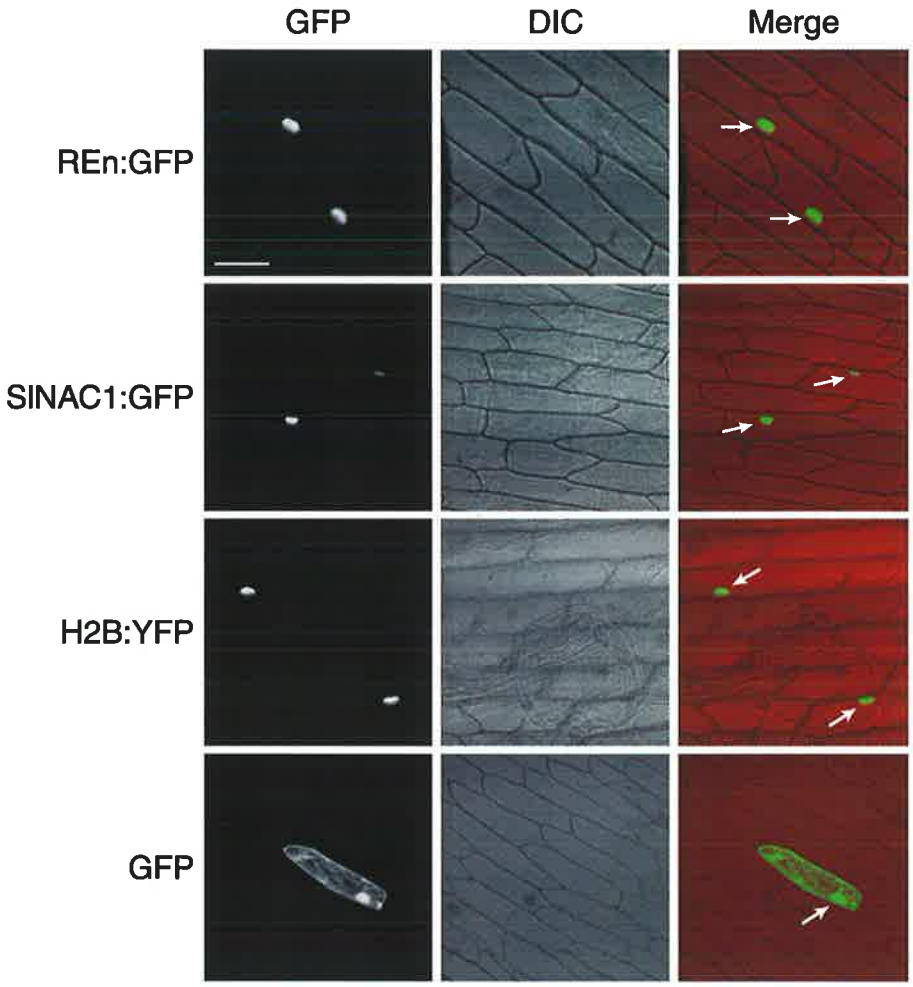


Figure 3.6. REn and SlNAC1 localise to the nucleus of onion cells. REn:GFP (top row) and SlNAC1:GFP (second row), as well as GFP alone (bottom row), were expressed in onion epidermal cells using the CaMV 35S promoter following biolistic delivery of vector DNA. A positive control for nuclear localisation, H2B:YFP, is also shown (third row). Cells were analysed for GFP fluorescence (left column) by confocal microscopy. Differential interference contrast (DIC) images and GFP/DIC merge images are shown in the middle and right columns respectively. Nuclei in GFP/DIC merge images are indicated by arrows. Bar = 100 μm.

however, the localisation of REn and SlNAC1 to the nucleus suggests that an opportunity exists for binding between these proteins in TLCV-infected plants.

3.3.5 TLCV infection induces the expression of *SlNAC1*

To analyse the endogenous expression of *SlNAC1*, a northern blot of total RNA preparations from tomato leaf tissue was done. *SlNAC1* mRNA of the predicted size (~1300 nt) was detectable at low levels in healthy tomato leaves (data not shown). To test whether *SlNAC1* transcription might be regulated by TLCV infection, total RNA samples from new, emerging leaves of infected and healthy plants were analysed by northern blotting in a time-course experiment (Fig. 3.7A). *SlNAC1* expression was strongly induced in infected plants at 10 days post-inoculation (dpi) and maintained to at least 20 dpi, a result observed in three independent experiments. Infection with TYLCV-Sardinia caused a similar increase in the levels of *SlNAC1* transcript (Fig. 3.7B), suggesting that induction of this gene is a general response to geminivirus infection. Some fluctuation in the level of *SlNAC1* gene expression in healthy plants over the course of these experiments was also observed, although this was minimal compared to the induction caused by geminiviral infection.

To test whether REn, given its physical interaction with SlNAC1, plays a role in the regulation of *SlNAC1* gene expression, tomato leaf tissue was infiltrated with *A. tumefaciens* cells harbouring a *REn* expression construct (p35S-REn) and changes in *SlNAC1* transcript accumulation were analysed (Fig. 3.7C). Expression of *REn* induced *SlNAC1* gene expression to levels similar to that observed when tissue was infiltrated with cells containing a replicating TLCV construct. In contrast, tissue which was infiltrated with *A. tumefaciens* containing an empty expression vector or vectors

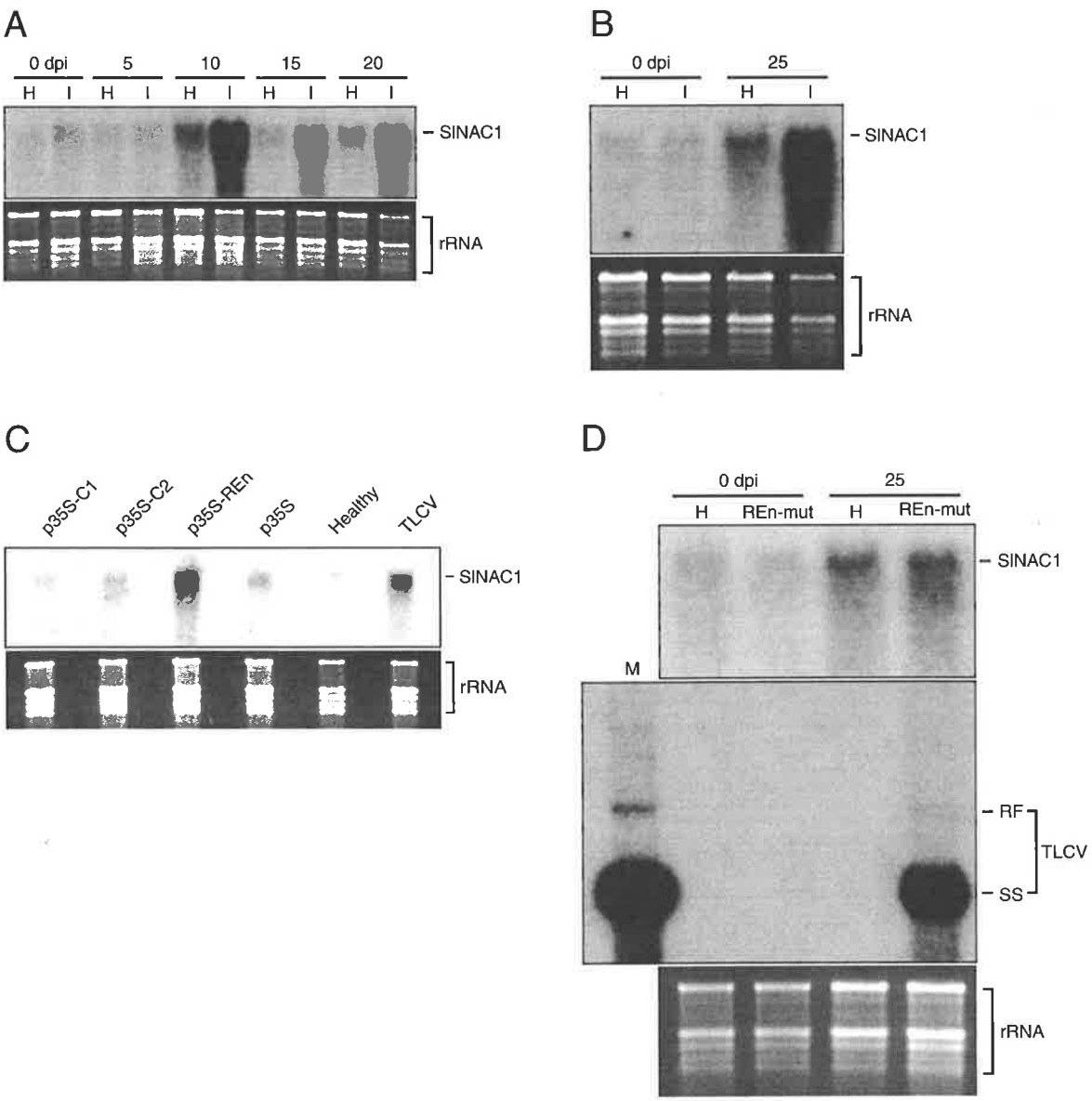


Figure 3.7. *SlNAC1* is induced by TLCV infection. **(A)** TLCV infection causes an upregulation of *SlNAC1* gene expression. RNA gel blot showing the expression of *SlNAC1* in healthy (H) or TLCV-infected (I) tomato plants. Tissue samples were obtained at 0, 5, 10, 15, and 20 days post-inoculation (dpi). **(B)** TYLCSV infection causes an upregulation of *SlNAC1* gene expression. RNA gel blot showing the expression of *SlNAC1* in healthy plants or plants infected with TYLCV-Sardinia. Tissue samples were obtained 0 and 25 dpi. **(C)** Transient expression of *REn* is sufficient to induce *SlNAC1* gene expression. Tomato leaves were infiltrated with *A. tumefaciens* cells containing a replication-competent TLCV 1.1mer, p35S, or p35S expressing the TLCV genes *C1*, *C2*, and *REn*. RNA was extracted from tissues 5 days post-infiltration and *SlNAC1* expression analysed by northern blotting. **(D)** A TLCV *REn*-mutant cannot induce *SlNAC1* gene expression. RNA gel blot showing the expression of *SlNAC1* in healthy plants or plants infected with a TLCV *REn*-mutant (REn-mut) at 0 and 25 dpi (top panel). The presence of replicating TLCV *REn*-mutant was confirmed by southern blotting (middle panel). In this blot, we also ran an extract obtained from plants infected with wild-type virus (left, designated M); the ratio of REn-mutant:wild-type total nucleic acid extracts is 20:1. TLCV DNA species are marked RF (supercoiled double-stranded replicative form) and SS (single-stranded).

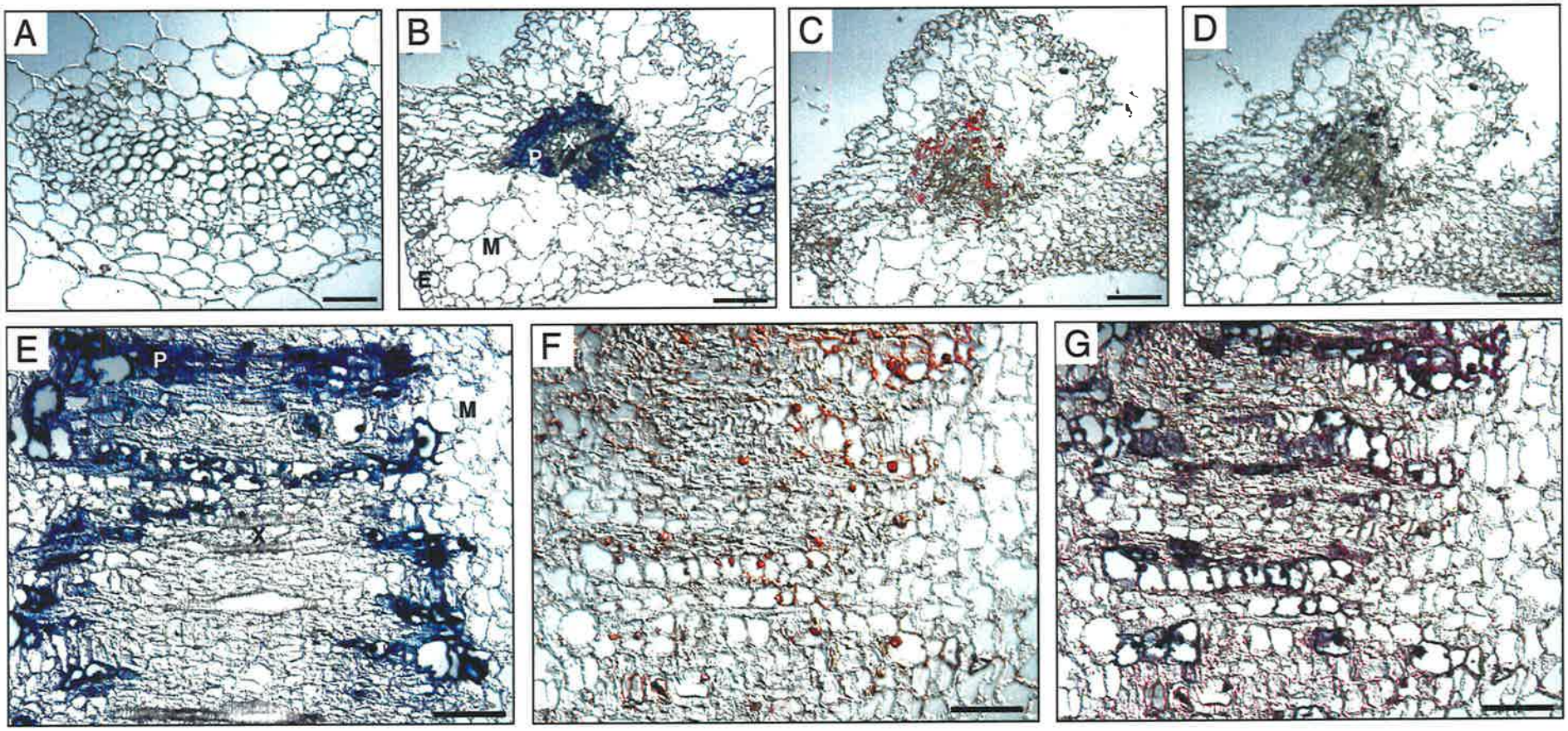
designed to express two other TLCV-encoded genes, *C1* and *C2*, contained levels of *SlNAC1* transcript similar to untreated tissue. These results suggested that induction of *SlNAC1* in response to TLCV infection is mediated by REn, and also demonstrated that *SlNAC1* is not induced non-specifically in response to *A. tumefaciens* infection or wounding associated with the infiltration procedure. In a subsequent experiment, a TLCV derivative containing a mutation in the *REn* gene which prevents translation of the REn protein (Rigden et al., 1996) was tested for its effect on *SlNAC1* expression. The level of *SlNAC1* transcript in tomato plants agroinoculated with the REn-mutant was comparable to that in healthy controls at 25 dpi (Fig. 3.7D, top panel). The presence of replicating REn-mutant virus was confirmed by southern blotting the same total nucleic acid samples and hybridising with a TLCV-specific probe (middle panel). It must be noted that the amount of viral DNA in extracts obtained from REn-mutant infected plants (middle panel, right lane) was much lower than equivalent samples from plants infected with wild-type virus (middle panel, lane M; the ratio of REn-mutant:wild-type total nucleic acid extracts is 20:1). Thus, this experiment does not rule out the possibility that the absence of *SlNAC1* induction in REn-mutant infected plants is due to reduced viral load. However, taken together, these results support the hypothesis that REn alone is responsible for induction of *SlNAC1*.

In all RNA gel blot analyses, the same results were obtained when membranes were hybridised with probes synthesised from the full-length *SlNAC1* gene or from only the divergent 3' sequence (data not shown)), indicating that variation in the expression of other putative *NAC* genes in response to TLCV infection or to transient REn expression was insignificant.

3.3.6 TLCV replication is tissue-specific and *SlNAC1* induction occurs only in TLCV-infected cells

SlNAC1 upregulation may be a systemic stress response, or alternatively TLCV may act to specifically induce expression of this gene in infected cells. To distinguish between these possibilities, *in situ* hybridisation experiments to determine the specific regions of *SlNAC1* mRNA accumulation compared to sites of TLCV infection were carried out. Hybridisation of tomato tissue with TLCV and SlNAC1 probes produced very weak chromogenic signals (data not shown). Since TLCV-derived nucleic acid accumulates to much higher levels in *N. benthamiana*, infected leaf tissue obtained from this host was analysed. A ssRNA complementary to the TLCV *V2* gene produced a strong signal that was observed mainly in phloem cells but also in some xylem parenchyma and bundle sheath cells (Fig 3.8B and 3.8E). This indicates that TLCV is limited to vascular tissue in *N. benthamiana*, a tropism also reported for *Abutilon mosaic virus*, SqLCV, and TYLCV-Dominican Republic (Horns et al., 1991; Sanderfoot et al., 1996; Rojas et al., 2001). No signal was obtained when healthy *N. benthamiana* leaf tissue was hybridised with a probe complementary to the divergent 3' *SlNAC1* sequence, which should not detect unrelated *N. benthamiana* NAC proteins (Fig. 3.8A). However, in TLCV-infected sections a *SlNAC1* homologue was detected in some phloem cells (Fig. 3.8C and 3.8F). To test whether induction of this gene was occurring only in cells infected with TLCV, dual-colour *in situ* hybridisations were carried out (Jowett, 2001). Hybridisation of the TLCV probe to sections exhibiting a *SlNAC1* signal produced a distinctive purple chromogenic output (Fig. 3.8D and 3.8G). This colour is produced by the masking of the red *SlNAC1* signal by the blue viral signal, and confirm that almost every cell that accumulated substantial amounts of *SlNAC1* mRNA also contained TLCV. Thus, induction of a *N. benthamiana SlNAC1* homologue in response

Figure 3.8. Induction of *SlNAC1* by TLCV occurs only in infected cells. Tissue sections derived from mock-inoculated (**[A]**) and TLCV-infected (**[B]-[G]**) leaves of *N. benthamiana* were hybridised with either fluorescein-labeled ssRNA probe complementary to *SlNAC1* (**[A]**, **[C]**, **[D]**, **[F]**, **[G]**) or digoxigenin-labeled ssRNA probe complementary to TLCV (**[A]**, **[B]**, **[D]**, **[E]**, **[G]**). **(A)-(D)** are cross sections and **(E)-(G)** are longitudinal sections taken from the main leaf vein. Bar = 100 μm. Cell types present are indicated: E = epidermal, M = mesophyll, P = phloem, X = xylem.


A
B
P
X
M
E
C
D
E
P
M
X
F
G

to TLCV infection is not a systemic response but rather restricted to cells infected with the virus.

3.3.7 The expression level of *SlNAC1* is a determinant of geminiviral replication

To investigate the possible function of *SlNAC1* in relation to TLCV infection, the effect of constitutive, high-level expression of this gene on TLCV replication was analysed. A transient TLCV replication system, based on *A. tumefaciens*-mediated delivery of an infectious TLCV construct into *N. benthamiana* leaf strips (Dry et al., 1997), was utilised in this study. When *A. tumefaciens* cells harbouring the infectious TLCV construct were mixed with cells containing an *SlNAC1* expression construct, TLCV ssDNA accumulated to a considerably higher level than in the control treatment, where TLCV was combined with an empty expression construct (Fig. 3.9A, compare lanes 1 and 2). This response was observed in four independent experiments in which all treatments were carried out in duplicate. The level of *SlNAC1* expression in all leaf strip samples was concurrently analysed by semi-quantitative real-time PCR, which confirmed that the enhancement of TLCV ssDNA accumulation was associated with *SlNAC1* expression by p35S-SlNAC1 (Fig. 3.9B). Together, these result suggest that SlNAC1 is involved in TLCV replication *in planta*.

3.4 Discussion

Because of their limited coding capacities, geminiviruses depend on host factors to amplify their genomes. In quiescent cells which have exited the cell division cycle and cannot support DNA replication, these pathogens must therefore induce the required replicational machinery. To achieve this, they encode proteins which increase the expression level of growth-promoting genes and/or alter the function of cell-cycle

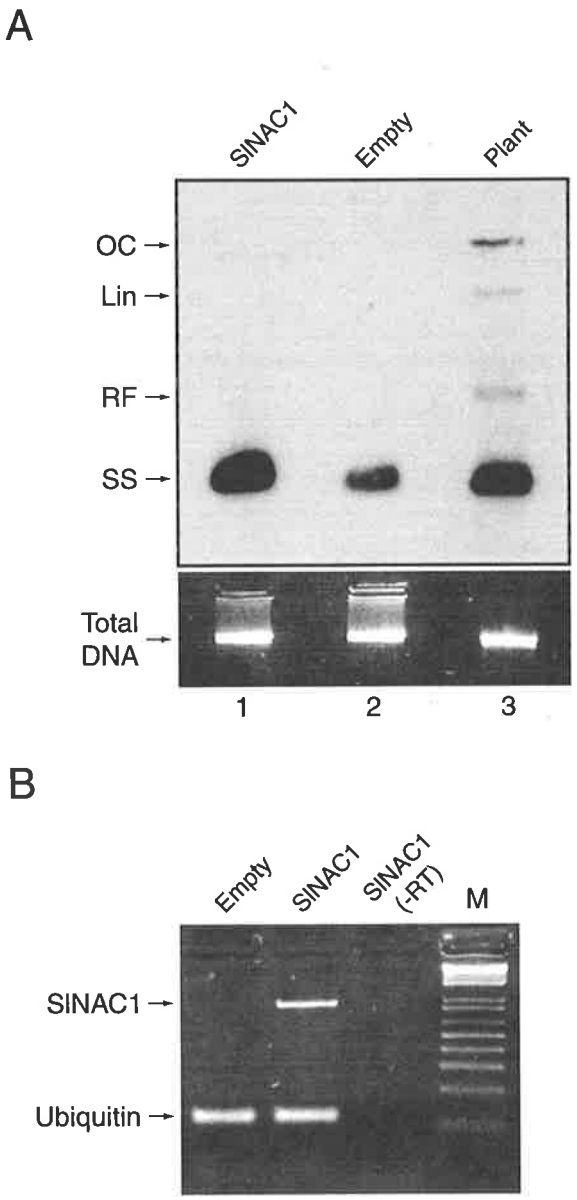


Figure 3.9. *SlNAC1* expression enhances TLCV ssDNA accumulation. **(A)** Expression of *SlNAC1* enhances TLCV ssDNA accumulation in a transient replication assay. *A. tumefaciens* cells harbouring Bin19-TLCV1.1 were combined with *A. tumefaciens* cells containing either an empty expression construct (lane 2) or p35S-SlNAC1 (lane 1) and co-cultivated for 48 h with leaf strips from *N. benthamiana* plants. DNA was extracted from tissue samples 3 days later and replication of TLCV analysed by Southern blotting. Lane 3 ("Plant") is a sample extracted from TLCV-infected *N. benthamiana* used as a marker for TLCV DNA forms, marked OC (open circular double-stranded), Lin (linear double-stranded), RF (supercoiled double-stranded replicative form) and SS (single-stranded). OC, Lin and RF DNA forms were observed in extracts from *N. benthamiana* leaf strips after longer exposures. **(B)** Analysis of *SlNAC1* expression by p35S-SlNAC1 in *N. benthamiana* leaf strips by semi-quantitative RT-PCR. Total RNA was prepared from leaf strips treated with TLCV plus an empty expression construct or TLCV plus p35S-SlNAC1. Ubiquitin mRNA served as an internal control. RT reaction mix without reverse transcriptase was used as a negative control (marked -RT). M, size markers.

regulatory proteins, often by physically interacting with host factors (Sections 1.2.1-1.2.3). The REn protein interacts with pRBR and PCNA, suggesting that it plays a number of distinct roles in geminivirus replication (Settlage et al., 2001; Castillo et al., 2003). Consistent with this idea, results presented in this chapter describe a new tomato protein of the NAC domain family, SlNAC1, that interacts with REn and appears to be involved in viral replication.

3.4.1 The role of *SlNAC1* in TLCV infection

In a transient replication system, expression of *SlNAC1* considerably enhanced the accumulation of TLCV ssDNA, suggesting that this gene facilitates TLCV replication. Although SlNAC1 must possess some essential cellular function, the results presented in this chapter suggest that geminiviruses, through the action of REn proteins, have hijacked its innate role. A number of mechanisms can be envisioned to explain this result. One is that SlNAC1 acts indirectly in TLCV replication as a positive regulator of cellular genes required during viral infection. For example, it may activate transcription of genes required for S-phase functions which are normally absent in differentiated cells, a strategy analogous to the putative release of E2F transcription factors when geminiviral Rep proteins bind Rb. This explanation does not correlate with the proposed function of other NAC proteins in meristem development and plant senescence pathways, where these factors contribute to a decision of cells to leave the proliferative state and take a certain differentiation pathway. For example, NAM is thought to interfere with cell division around the developing shoot apical meristem (Souer et. al, 1996) to drive flower development. Further, the observation that GRAB proteins from wheat interfere with replication of WDV lcd Xie *et al.* (1999) to speculate that these NACs play a role in the pathway leading to cell differentiation. However, the

family of genes encoding NAC domains is very large and members appear to possess highly diverse functions. Thus, some NACs could upregulate genes involved in processes advantageous to geminivirus replication, such as DNA replication, transcription, or the G1/S transition of the cell cycle. Supporting this idea, *NAC1* from *Arabidopsis* is involved in the initiation of lateral root development (Xie et al., 2000), while *CUC1* promotes adventitious SAM formation by maintaining epidermal cells in an undifferentiated state in transgenic *Arabidopsis* (Takada et al., 2001; Hibara et al., 2003).

Another possibility is that SlNAC1 functions directly in geminiviral replication. There are numerous examples where host transcription factors play an important and direct role in activating the DNA replication of mammalian oncoviruses by binding the viral origin of replication and increasing the initiation frequency (Li et al., 1998). Alternatively, REn may recruit SlNAC1 into a DNA replication complex where it could promote amplification of the viral genome. This idea is supported by the observation that REn interacts with PCNA (Castillo et al., 2003), a host factor that acts as a sliding clamp and modulates the interactions of other proteins, including polymerases, with DNA (Hingorani and O'Donnell, 2000).

A third possible scenario is that SlNAC1 activates transcription of viral genes. While the geminiviral C2 protein is responsible for activating virion-sense gene expression (Section 1.1.9.2), expression of the complementary-sense genes is probably controlled by host factors. SlNAC1 mediates expression of a reporter gene in yeast, suggesting that it functions endogenously as a transcription factor and therefore could positively modulate *cis*-acting promoter elements in the geminiviral genome.

3.4.2 Mechanism of TLCV-mediated *SlNAC1* induction

SlNAC1 gene expression was upregulated in response to TLCV infection. Two lines of evidence presented in this study support the idea that this induction is mediated by REn. First, transient delivery of a *REn* expression construct resulted in increased accumulation of *SlNAC1* mRNA, while control constructs were unable to engender this response. Second, a TLCV REn-mutant was unable to upregulate *SlNAC1* despite accumulating to moderate levels in infected tissue. A number of mechanisms by which *SlNAC1* is induced can be envisioned. First, REn could act directly as a transactivator of *SlNAC1* gene expression. Analysis of the peptide sequence of TGMV REn revealed that its acidic N-terminus resembles some transcriptional activation domains (Hanley-Bowdoin et al., 1999). Second, *SlNAC1* induction may be a side-effect of the presence of REn in a plant cell. It is doubtful that it occurs via the putative REn/Rb or REn/PCNA interactions, since Rep, which also binds these host factors, was unable to stimulate *SlNAC1* expression. However, *SlNAC1* induction may occur because REn is affecting other cellular processes, possibly through an as yet unidentified protein interaction(s). This explanation is supported by the observation that REn and Rep produced highly disparate phenotypic effects when transiently expressed in host plants (Selth et al., 2004). A final possibility is that upregulation of *SlNAC1* depends on REn/SlNAC1 binding. For example, by sequestering SlNAC1 through physical interaction, REn may relieve a negative feedback mechanism by which SlNAC1 inhibits transcription of its gene. Such a function is not unprecedented: *At*WRKY6, a member of the large WRKY family of plant-specific transcriptional regulators, is able to suppress its own promoter activity while positively influencing the expression of genes involved in senescence and pathogen defence (Robatzek and Somssich, 2002).

It was originally proposed that NAC proteins could be divided into three subfamilies (Kikuchi et al., 2000). More recently, Ooka *et al.* (2003) carried out a more comprehensive phylogenetic analysis of known NACs and putative *Arabidopsis* and rice NACs and identified 18 subfamilies. Members of the so-called ATAF subfamily, identified in both studies, appear to have a conserved role in the response to stress. Genes belonging to this group are induced by wounding (Collinge and Boller, 2001), fungal infection (Collinge and Boller, 2001; Hegedus et al., 2003), bacterial infection (Mysore et al., 2002), insect damage (Hegedus et al., 2003), and cold shock (Hegedus et al., 2003). *SlNAC1*, which also belongs to the ATAF subfamily, is induced by TLCV (this study) and *Pseudomonas syringae* (Mysore et al., 2002) infection, suggesting that it may play a general role in stress responses. However, numerous lines of evidence presented in this chapter support the idea that stimulation of *SlNAC1* gene expression by TLCV is a specific response and that SlNAC1 plays an active role in TLCV infection. First, neither *A. tumefaciens* infection or wounding associated with the agroinfiltration procedure induced *SlNAC1*. Second, induction of *SlNAC1* by TLCV is restricted to infected cells and appears to be mediated by the REn protein. Third, SlNAC1 interacts with the TLCV-encoded REn protein. Finally, overexpression of *SlNAC1* enhances the accumulation of TLCV DNA species in a transient replication system.

3.4.3 NACs are involved in other viral infections

Xie *et al.* (1999) found an interaction between the WDV RepA protein and two wheat NACs, GRAB1 and GRAB2. The N-terminus (amino acids 1-208) of TLCV Rep shares 39.5% sequence identity with full-length WDV RepA. However, we were unable to detect binding between bacterially-expressed TLCV Rep and SlNAC1 *in vitro* (data not shown). Another apparent difference between the WDV/GRAB and

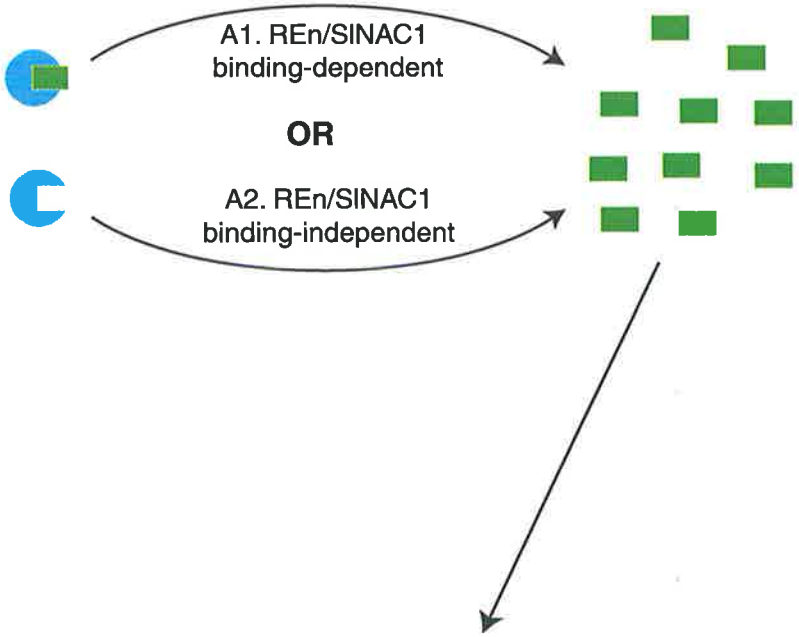
TLCV/SlNAC1 interaction is that while expression of both *GRAB* genes was shown to interfere with WDV DNA replication in cultured wheat cells, *SlNAC1* expression enhanced TLCV ssDNA accumulation in a transient replication assay. The distinct role of GRAB and SlNAC1 in geminivirus infection may again reflect the functional diversity that exists between members of the NAC domain family. Supporting the idea of NACs possessing diverse roles in viral pathogenesis, the NAC domain-containing *Arabidopsis* TIP protein is involved in the *Turnip crinkle virus* (TCV) resistance response pathway by interacting with the TCV coat protein (Ren et al., 2000). Alternatively, it could denote different DNA replication strategies utilised by the highly divergent dicot-infecting TLCV and monocot-infecting WDV. It would be useful to examine the effect WDV infection has on the expression level of *GRAB1* and *GRAB2*, to see whether, in contrast to the situation with TLCV and *SlNAC1*, the virus downregulates these genes.

3.4.4 Models of *SlNAC1* induction and its role in TLCV infection

A summary of the possible mechanisms by which *SlNAC1* is induced and enhances TLCV ssDNA accumulation is shown in Figure 3.10. REn/SlNAC1 binding could either be involved in induction of *SlNAC1* or alternatively play a more direct role in viral replication and/or transcription. To answer this question, the effect of *SlNAC1* overexpression on the replication of TLCV REn-mutant is currently being investigated. Regardless of the exact role of SlNAC1 in TLCV infection, the results presented in this chapter imply that the mechanism by which REn increases viral ssDNA accumulation involves its interaction with SlNAC1.

Figure 3.10. Hypothetical models for the roles of REn and SlNAC1 in enhancement of TLCV replication. TLCV infection causes an upregulation of *SlNAC1* that is mediated by REn **(A)**. This induction may **(A1)** or may not be **(A2)** a consequence of the REn/SlNAC1 interaction. Subsequently, the increased level of SlNAC1 leads to an enhancement of TLCV ssDNA accumulation **(B)**. This may occur via a mechanism that is independent of its interaction with REn **(B1)**. In this binding-independent model, SlNAC1 could activate expression of host genes required for S-phase functions **(B1a)**, or have a more direct role in the TLCV lifecycle by inducing transcription of the viral complementary sense genes **(B1b)**. Alternatively, SlNAC1-mediated enhancement of TLCV ssDNA accumulation may require REn binding **(B2)**. In this model, SlNAC1 could be recruited by REn into a replication complex which may also contain Rep, proliferating cell nuclear antigen (PCNA) and replication factor-C (RFC) **(B2a)**. Alternatively, REn may recruit SlNAC1 to the complementary-sense promoter to activate viral transcription **(B2b)**, a process that may also involve Rep. A key is shown at the bottom of the figure to identify proteins involved in these models.

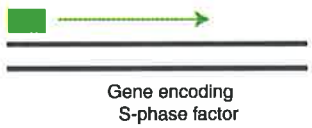
A. *SlNAC1* induction



B. SlNAC1-mediated enhancement of TLCV replication

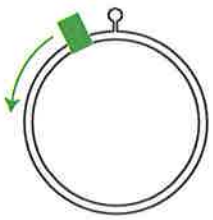
B1. REn/SlNAC1 binding-independent

B1a. Induction of S-phase factors



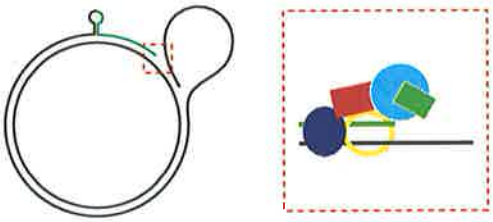
OR

B1b. Facilitation of viral transcription



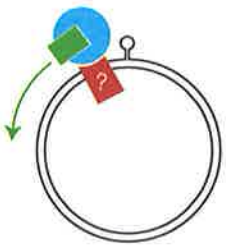
B2. REn/SlNAC1 binding-dependent

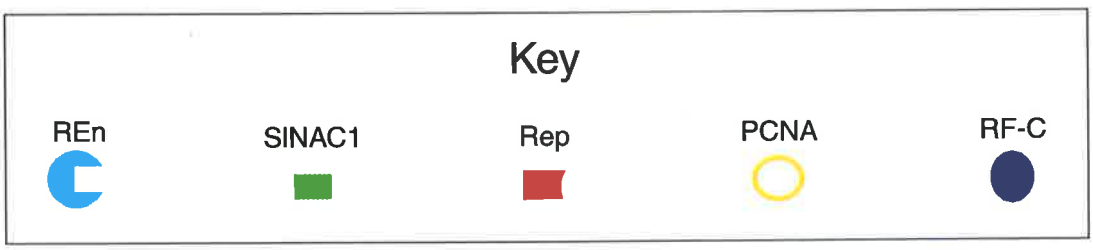
B2a. Involvement in a replication complex



OR

B2b. Facilitation of viral transcription





Chapter 4 – A host reversibly glycosylated peptide is involved in the function of a geminiviral movement protein

4.1 Introduction

The primary cell wall of dicotyledonous plants is composed mainly of polysaccharides, including cellulose, hemicelluloses and pectin. Cellulose is synthesised at the plasma membrane (Delmer, 1999), while pectic and hemicellulosic polysaccharides are assembled in the Golgi apparatus and delivered to the pre-existing wall by vesicles (Driouich et al., 1993). The enzymes responsible for the synthesis of polysaccharides in the Golgi are glycosyltransferases, which catalyse the transfer of a glycosyl residue from a nucleotide sugar to an acceptor molecule (Brummell et al., 1990; Driouich et al., 1993).

Plant reversibly glycosylated peptides (RGPs) are a small multigene family of autocatalytic glycosyltransferases that have been implicated in polysaccharide biosynthesis. In particular, the substrate specificity and subcellular localisation of some RGPs suggests that they have a role in the synthesis of hemicellulosic polysaccharide. For example, the ratio of steady-state glycosylation of PsRGP1 from pea (*Pisum sativum*) by UDP-Glc, UDP-Xyl, and UDP-Gal is approximately 10:7:3, which is similar to the typical sugar composition of β-1,4 glucan (xyloglucan; Dhugga et al., 1997), the predominant hemicellulose in many cell walls. Further, PsRGP1 is localised to trans-Golgi dicytosomal cisternae, and the conditions that stimulate or inhibit β-

glucan synthase activity are the same as those that stimulate or inhibit the glycosylation of PsRGP1 (Dhugga et al., 1997). More recently, other RGPs from maize (*Zea mays*; Epel et al., 1996), *Arabidopsis* (Delgado et al., 1998), wheat (*Triticum aestivum*; Langeveld et al., 2002), rice (*Oryza sativa*; Langeveld et al., 2002) and potato (*Solanum tuberosum*; Bocca et al., 1999a; Bocca et al., 1999b; Wald et al., 2003) have been identified and their functions partially characterised. Like PsRGP1, all of these proteins undergo self-glycosylation in the presence of UDP-Glc and their expression patterns and subcellular localisation provide further evidence for a role in the synthesis of cell wall polysaccharides.

Monopartite begomoviruses possess two virion-sense genes. The *V2* gene encodes the viral coat protein (Dry et al., 1993), but the function of *V1* is largely unknown. Studies of the *V1* gene from TLCV have shown that it is involved in symptom development and may have a role in cell-to-cell movement (Section 1.1.9.5). However, the means by which it mediates these processes is unknown. To further characterise the role of *V1*, a tomato library was screened for proteins which interact with the V1 protein from TLCV. This screen retrieved a protein with sequence similarity to a potato RGP, UDP-glucose:protein transglucosylase (UPTG), that was designated SlUPTG1 (for *S. lycoperisum* UPTG1). The interaction between V1 and SlUPTG1 was confirmed by an *in vitro* binding assay. In this chapter, evidence is presented to suggest that SlUPTG1 enhances the movement of TLCV, and possible mechanisms for the V1/SlUPTG1 interaction in this process are discussed.

4.2 Materials and Methods

4.2.1 Yeast two-hybrid screening

The *V1* ORF was amplified by PCR using primers V1-F-EcoRI and V1-R-XhoI, digested with EcoRI and XhoI and ligated into similarly digested pLexA to generate pLexA-V1. This construct was used as bait to screen a tomato Rio Grande cDNA library (Zhou et al., 1995) as described in Section 2.2.17.

The construction of vectors expressing a fusion of TLCV REn protein to LexA and the isolation of a vector expressing a fusion of SlNAC1 to B42, used as controls to verify the interaction between V1 and SlUPTG1, has been described in Sections 3.2.1 and 3.3.1 respectively.

pLexA vectors expressing LexA fusions to V1 homologues from ACMV and TYLCSV were constructed to examine the general significance of SlUPTG1 binding in geminivirus infections. ACMV *AV2* and TYLCSV *V1* were amplified with primer pairs P3/P4 and P5/P6 respectively, digested with *Eco*RI and *Xho*I, and ligated into similarly digested pLexA.

4.2.2 Production of recombinant proteins and *in vitro* binding experiments

Production of 6xHis-tagged V1 (His-V1) protein was achieved using the pQE30 vector (Qiagen). The coding region of *V1* was amplified using oligonucleotides V1-F-SphI and V1-R-HindIII, digested with SphI/HindIII, and ligated into similarly digested pQE30 to generate pQE30-V1. Chloramphenicol acetyl-transferase (CAT) was used as a control 6xHis-tagged protein. The entire CAT ORF was amplified with primers CAT-F-BamHI and CAT-R-HindIII, digested with BamHI/HindIII, and cloned into similarly

digested pQE30. The construction of pCAL-SlUPTG1 and pCAL-SlNAC1, designed to express SlUPTG1 and SlNAC1 fusions to CBP, respectively, has been described (Section 3.2.3).

Binding experiments were carried out essentially as described in Section 3.2.3, except that inputs in all reactions were crude samples of 6xHis-tagged and CBP-tagged proteins.

4.2.3 Self-glycosylation assays

For analysis of glycosylation activity, 5 µg of purified protein or 50 µg of protein extract was incubated for 30 min at 30°C with 0.2 nmol UDP-[^{14}C]Glc (specific activity 327mCi/mmol; Amersham Biosciences, Castle Hill, Australia), 50 mM Tris/HCl, pH 7.2, and 5 mM $MnCl_2$, in a final volume of 100 µl. The chase experiments included an additional 30 min incubation with 2 mM UDP-Glc or UDP-Gal (Sigma). The samples were spotted onto Whatman 3MM filter paper disks, air-dried, precipitated by agitating in 20% (v/v) trichloroacetic acid (TCA) for 10 min, washed in ice-cold 5% (v/v) TCA four times for 3 min each, and washed in ice-cold ethanol twice for 1 min each. After drying at 37°C for 5 min, the radioactivity on paper disks was measured by scintillation counting. Alternatively, glycosylation reactions were stopped by adding 10% (v/v) TCA, incubating 30 min on ice and centrifuging at 16,100 g for 5 min at 4°C. The precipitated proteins were resuspended in 20 µl of SDS-PAGE loading buffer (0.5 M Tris/HCl, pH 6.8, 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.05% bromophenol blue), boiled for 3 min and loaded onto 4-20% Tris-glycine-SDS polyacrylamide gels (Life Gels, Clarkston, GA). After electrophoresis of the samples, gels were destained and fixed in fixing buffer (7% (v/v) methanol, 7% (v/v) acetic acid),

soaked in Amplify (Amersham Biosciences), dried using a DryEase Mini-Gel Drying System (Invitrogen, Carlsbad, CA) and exposed for fluorography.

4.2.4 Visualisation of GFP-fusion proteins following microprojectile bombardment and agroinfiltration

Variants of the shuttle vector pART7 (Gleave, 1992), termed pC'gfp and pN'gfp, were used to express V1:GFP and SlUPTG1:GFP fusion proteins. The *V1* ORF lacking a start codon was amplified using primers V1-F-EcoRI and V1-R-XhoI, digested with EcoRI, and ligated into EcoRI/SmaI digested pC'gfp to generate an in-frame C-terminal fusion to the *GFP* gene. The *SlUPTG1* ORF lacking a stop codon was amplified with primers SlUPTG-F-EcoRI and SlUPTG-R-BamHI, digested with EcoRI and BamHI, and ligated into similarly digested pN'gfp to generate an in-frame N-terminal fusion to the *GFP* gene. pART7-ATG:GFP, which expresses free GFP (T. Franks, unpublished data), was used as a control in this experiment.

Onion epidermal strips on agar containing Murashige and Skoog Salt Mixture (Invitrogen, Carlsbad, California) were bombarded with each of the vectors as described (Section 2.2.25). After bombardment, tissue was stored in the dark for 48 h and GFP/YFP expression visualised using a Bio-Rad Radiance 2100 Confocal Laser Scanning Microscope System. The excitation wavelength used for both GFP and YFP analysis was 488nm.

For agroinfiltration experiments, the V1:GFP, V2:GFP and ATG:GFP sequences downstream of the CaMV 35 promoter were released from pART7 plasmids by NotI digestion and ligated into NotI-digested pART27. *A. tumefaciens* strain C58 was

transformed separately with each of these binary vectors by electroporation. *N. tabacum* leaf tissue was infiltrated with *A. tumefaciens* cells as described (Selth et al., 2004) prior to visualisation by confocal microscopy.

4.2.5 Analysis of *SlUPTG1* gene expression

Three week old tomato plants were inoculated with TLCV using *A. tumefaciens*. Total nucleic acid was extracted from new leaves at various time points and subjected to RNA gel blot analysis as described (Section 2.2.23). To analyse the spatial expression of SlUPTG1, samples were taken from various tissues of healthy tomato plants and processed as described above.

4.2.6 Analysis of TLCV DNA replication

The vector pART27 (Gleave, 1992) was used to transiently overexpress *SlUPTG1* in *N. benthamiana* leaf strips as described in Section 3.2.7. The entire *SlUPTG1* ORF was amplified by PCR using primers SlUPTG1-F-EcoRI and SlUPTG1-R-XbaI, digested with EcoRI and XbaI, and ligated into EcoRI/XbaI-digested pART7 to yield pART7-SlUPTG1. A DNA fragment containing the CaMV 35S promoter and the *SlUPTG1* ORF was released by NotI digestion and ligated into similarly-digested pART27 to generate p35S-SlUPTG1. *A. tumefaciens* cells containing p35S-SlUPTG1, Bin19-TLCV1.1 and Bin19 harbouring a TLCV derivative with a mutation in the *V2* gene (Bin19-V2mut; Rigden et al., 1993) were co-cultivated with leaf strips alone or in combination (see Results). Extraction and analysis of viral DNA was done as described in Section 3.2.7. Quantitation of *SlUPTG1* mRNA expression by semi-quantitative reverse transcription-PCR was done as described in Section 3.2.8.

4.3 Results

4.3.1 Isolation of SlUPTG1

To identify host proteins interacting with TLCV V1, a yeast two-hybrid screen of a tomato cDNA library fused to the B42 activation domain (AD)-encoding sequence (Zhou et al., 1995) was carried out using V1 fused to the LexA DNA-binding domain (BD) as bait. A total of 2 x 10^6 transformants were assayed for leucine prototrophic growth and *green fluorescent protein* (*GFP*) expression. Of these transformants, three were able to activate both reporter genes. Plasmid DNA was rescued and the sequence of the inserts determined. One of the cDNA sequences was similar to an *Arabidopsis* gene of unknown function, while another shared identity with plant glutaredoxin (Section 6.3.1). The possible involvement of these proteins in V1 function is discussed in Section 6.4.1. The third plasmid contained a cDNA of 1089 bp encoding a predicted full-length translation product of 362 amino acids with a molecular mass of 41.2 kDa. A BLAST query of this protein sequence revealed that it shared a high degree of sequence identity with UDP:glucose protein transglucosylase (UPTG) from potato. Based on this characteristic, the protein was designated SlUPTG1, for *S. lycopersicum* UPTG1 (GenBank accession number AY622990).

The interaction between V1 and SlUPTG1 in yeast was confirmed using various control constructs (Fig. 4.1). When co-expressed, V1 and SlUPTG1 activated *GFP* expression and stimulated leucine prototrophic growth, an activity similar to that observed for TLCV REn (fused to the LexA BD) and SlNAC1 (fused to the B42 AD), which have previously been shown to interact in yeast and *in vitro* (Sections 3.3.1 and 3.3.3). However, the reporter genes were not expressed when V1 was co-expressed with two control proteins, SlNAC1 or the B42 AD alone, indicating that it was specifically

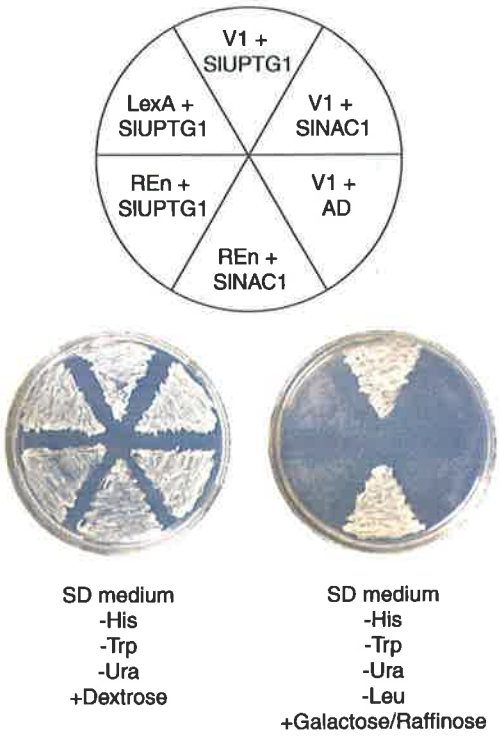


Figure 4.1. V1 interacts with SlUPTG1 in yeast. Yeast coexpressing proteins as indicated (top) were grown on SD -His -Trp -Ura medium (bottom left plate), and interaction was tested by leucine prototrophy on an inductive carbon source (galactose and raffinose) (bottom right plate). V1 proteins were fused to the LexA DNA binding domain while SlUPTG1 was fused to the B42 activation domain. Yeast cells coexpressing LexA:REn and B42:SlNAC1 were used as a positive control. Negative controls included V1 coexpressed with SlNAC1 or B42 alone, and SlUPTG1 coexpressed with REn or LexA alone.

interacting with SlUPTG1. Similary, SlUPTG1 did not interact with TLCV REn or the LexA DNA BD alone.

4.3.2 Sequence analysis of SlUPTG1 and related proteins

SlUPTG1 was similar to two classes of plant peptide sequences (Table 4.1). The first class of sequences, designated RGP1 (Langeveld et al., 2002), shared 77-94% identity at the amino acid level and included proteins putatively involved in cell wall biosynthesis: two UPTGs from potato, StUPTG1 and StUPTG2, and various reversibly glycosylated peptides (RGPs) from maize, cotton (*Gossypium hirsutum*), *Arabidopsis*, wheat, pea and rice. The second class, designated RGP2 (Langeveld et al., 2002), contained two RGPs from rice and wheat (OsRGP2 and TaRGP2) which shared 45-46% sequence identity with SlUPTG1 at the amino acid level. A phylogenetic tree based on mutiple sequence alignment of these related sequences is shown in Fig. 4.2. SlUPTG1 grouped with the potato UPTG proteins, and dicot sequences were separated from those of monocots. As expected, TaRGP2 and OsRGP2 formed a separate branch.

4.3.3 Interaction with SlUPTG1 may be a general feature of geminiviral V1 proteins

The general significance of the V1/SlUPTG1 interaction was examined by testing whether SlUPTG1 could interact with homologues of V1 from TYLCV-Sardinia ($V1_{TYLCSV}$) and from the bipartite ACMV (AV2). $V1_{TYLCSV}$ and AV2 share 58.3% identity (64.3% similarity) and 63.4% identity (72.3% similarity) with TLCV V1 respectively. Although the existence of AV2 in ACMV infections has not been reported to date, its conservation suggests that it possesses some required function (Frey et al., 2001). Supporting this hypothesis, an AV2 protein has been detected by Western

Table 4.1. Homology between RGPs

	AtRGP2	AtRGP3	AtRGP4	GhRGP1	OsRGP1	OsRGP2	PsRGP1	SlUPTG1	StUPTG1	StUPTG2	TaRGP1	TaRGP2	ZmRGP1
AtRGP1	93 (95)[a]	84 (88)	77 (85)	87 (90)	84 (88)	49 (58)	86 (89)	89 (92)	88 (91)	89 (92)	88 (91)	49 (58)	88 (91)
AtRGP2		81 (86)	76 (84)	85 (88)	82 (85)	48 (57)	85 (89)	88 (91)	86 (89)	88 (91)	85 (88)	47 (56)	86 (88)
AtRGP3			73 (81)	84 (88)	79 (84)	45 (55)	84 (89)	86 (89)	85 (89)	87 (91)	83 (89)	45 (56)	83 (88)
AtRGP4				76 (84)	71 (80)	46 (55)	75 (84)	77 (84)	74 (82)	75 (83)	75 (82)	44 (55)	75 (82)
GhRGP1					82 (87)	48 (58)	87 (90)	89 (91)	86 (90)	88 (90)	86 (90)	47 (57)	88 (91)
OsRGP1						45 (55)	80 (85)	83 (87)	83 (88)	83 (86)	87 (89)	45 (56)	87 (89)
OsRGP2							48 (57)	46 (55)	45 (55)	46 (55)	48 (56)	88 (91)	46 (55)
PsRGP1								86 (89)	85 (89)	87 (90)	86 (91)	47 (57)	87 (90)
SlUPTG1									94 (95)	91 (93)	85 (89)	45 (55)	90 (92)
StUPTG1										90 (92)	85 (89)	45 (55)	87 (90)
StUPTG2											86 (90)	46 (56)	88 (91)
TaRGP1												47 (56)	93 (95)
TaRGP2													46 (55)

[a]The percentage identity and similarity (in brackets) as determined by the BestFit algorithm, are shown.

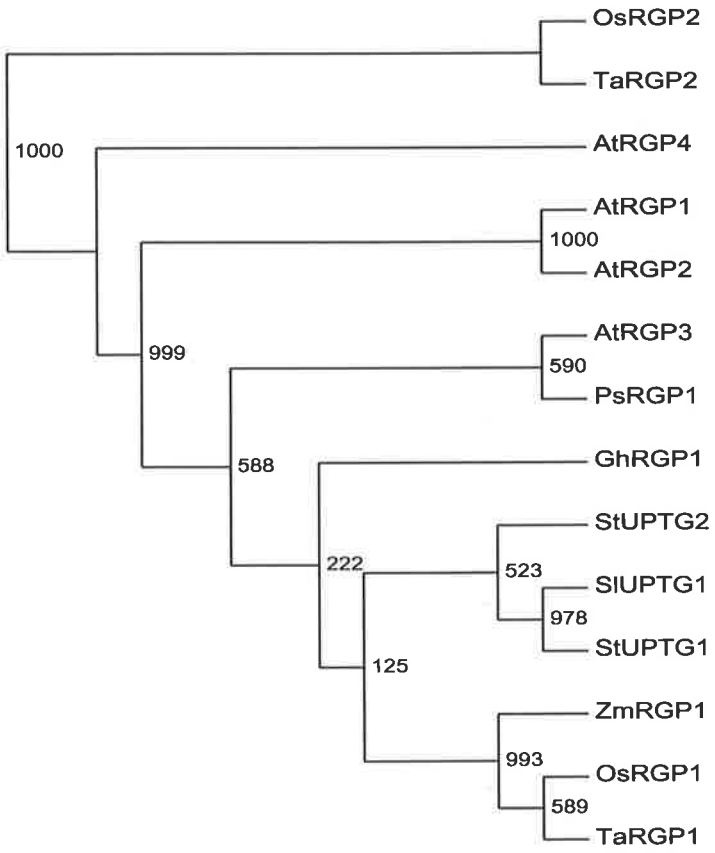


Figure 4.2. Phylogenetic analysis of the predicted amino acid sequences of SlUPTG1 and other RGPs. Multiple sequence alignment of the proteins was conducted using ClustalX (Thompson et al., 1997), and phlyogenetic analysis was carried out by the neighbor-joining method (Saitou et al., 1987). A bootstrap analysis of 1000 resampling replicates was conducted with ClustalX. The rooted phylogenetic tree was displayed using the NJPlot program included with ClustalX. Database accession numbers of the sequences used for the comparison are: OsRGP2 (Y18623), TaRGP2 (Y18625), AtRGP4 (AF329280), AtRGP1 (AF013627), AtRGP2 (AF013628), AtRGP3 (AF034255), PsRGP1 (U31565), GhRGP1 (CAC83750), StUPTG2 (AJ310910), SlUPTG1 (AY622990), StUPTG1 (AJ223252), ZmRGP1 (U89897), OsRGP1 (AJ011078), and TaRGP1 (Y18626).

blotting in plants infected with the bipartite begomovirus TLCV-India, and mutations in its *AV2* gene caused symptom attenuation and reduced accumulation of viral DNAs (Padidam et al., 1996).

The results of this study are shown in Figure 4.3A. In the same yeast two-hybrid assay, ACMV AV2 could also interact with SlUPTG1, as indicated by leucine prototrophic growth and *GFP* expression. $V1_{TYLCSV}$ was able to stimulate some leucine prototrophic growth when co-expressed with SlUPTG1, however this was also observed when $V1_{TYLCSV}$ was co-expressed with the B42 AD alone, suggesting that it may act as a weak transactivator of reporter gene expression in yeast. The synthesis of V1, $V1_{TYLCSV}$, AV2 and SlUPTG1 in yeast was analysed by Western blotting and all proteins were shown to be produced at similar levels (Fig. 4.3B). The finding that the bipartite homologue of V1, AV2, also binds SlUPTG1 in yeast suggests that this interaction occurs in geminivirus infections. In addition, it implies that AV2 is functional in ACMV infection and may act analogously to TLCV V1.

4.3.4 *In vitro* binding of SlUPTG1 to V1

The specificity of the V1/SlUPTG1 protein interaction was tested using an *in vitro* pull down assay. Crude soluble protein samples from bacteria induced to express a recombinant V1 protein fused to six histidines (His-V1) and a recombinant SlUPTG1 protein fused to calmodulin-binding peptide (CBP-SlUPTG1) were mixed and incubated with nickel-nitrilotriacetic acid (Ni-NTA) agarose. After washing, bound protein was analysed by SDS-PAGE and immunoblotting using antibodies directed against the FLAG epitope, found downstream of CBP, and poly-Histidine.

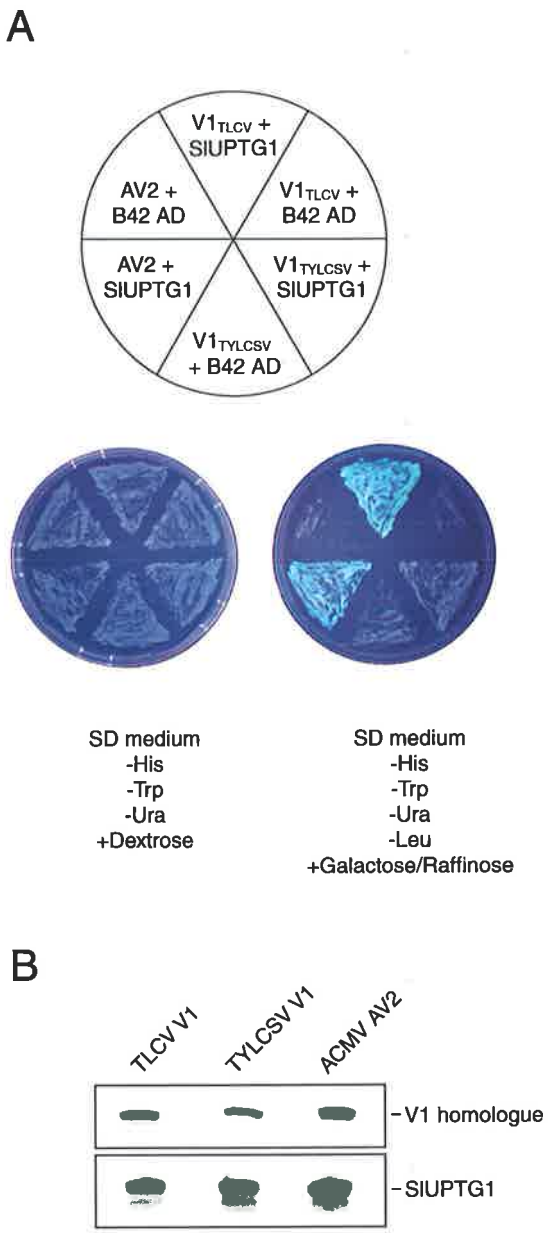


Figure 4.3. SlUPTG1 interacts with a homologue of V1 from ACMV. **(A)** Yeast two-hybrid assays testing the ability of SlUPTG1 to interact with V1 from TYLCSV (V1TYLCSV) and the V1 homologue from ACMV, AV2. Yeast coexpressing proteins as indicated (top) were grown on SD -His -Trp -Ura medium (bottom left plate), and interaction was tested by leucine prototrophy and *GFP* expression on an inductive carbon source (galactose and raffinose) (bottom right plate). V1 proteins were fused to the LexA DNA binding domain while SlUPTG1 was fused to the B42 activation domain. Yeast cells coexpressing a LexA fusion to V1 from TLCV (V1TLCV) and a B42 fusion to SlUPTG1 were used as a positive control. As negative controls, the V1 proteins were coexpressed with B42 alone. **(B)** Immunoblot analysis of yeast cells demonstrating that B42:SlUPTG1 and the different V1:LexA fusion proteins were coexpressed at similar levels. Total protein from yeast cultures containing B42 and LexA fusion proteins was extracted, fractionated on 4-20% SDS-polyacrylamide gels, and immunoblotted with anti-LexA (top panel) and anti-HA (B42-tagged proteins also contain a HA epitope; bottom panel).

CBP-SlUPTG1 was present in the bound fraction when incubated with His-V1 (Fig. 4.4, lane 5). In contrast, bound CBP-SlUPTG1 was not detectable when it was combined with a poly-Histidine fusion to chloramphenicol acetlytransferase (His-CAT) (lane 6) or when it was incubated with Ni-NTA resin alone (lane 8), indicating that it was interacting specifically with His-V1. To determine if His-V1 was specifically pulling down CBP-SlUPTG1, it was mixed with a control protein sample, CBP-SlNAC1 (Sections 3.2.3 and 3.3.3). No CBP-SlNAC1 was detectable in the bound fraction (lane 7), indicating that His-V1 does not indiscriminately bind abundant proteins in a mixture. All reactions were replicated at least three times with similar results.

4.3.5 Subcellular localisation of V1:GFP and SlUPTG1:GFP fusion proteins

To further investigate the potential roles of V1 and SlUPTG1 *in vivo*, the subcellular localisation of these proteins in plant cells was examined. The *V1* and *SlUPTG1* ORFs was fused to *GFP* downstream of the *Cauliflower mosaic virus* (CaMV) 35S promoter. Fusion proteins (V1:GFP and SlUPTG1:GFP) were transiently expressed in onion epidermal cells following biolistic delivery of vector DNA or in *Nicotiana tabacum* (tobacco) leaf tissue following agroinfiltration and detected by confocal microscopy.

Free GFP was distributed in both the cytoplasm and the nucleus of bombarded onion cells (Fig. 4.5A) and infiltrated tobacco cells (Fig. 4.5B). We have previously visualised the production of V1:GFP in onion epidermal cells, and found that it localises primarily to the cell periphery, cytoplasm and around the nucleus but does not enter the nucleus (M. S. Raisheed, L. A. Selth, A. M. G. Koltunow, J. W. Randles and M. A. Rezaian, submitted). A similar pattern of fluorescence was observed in tobacco epidermal cells producing V1:GFP after agroinfiltration of the expression construct.

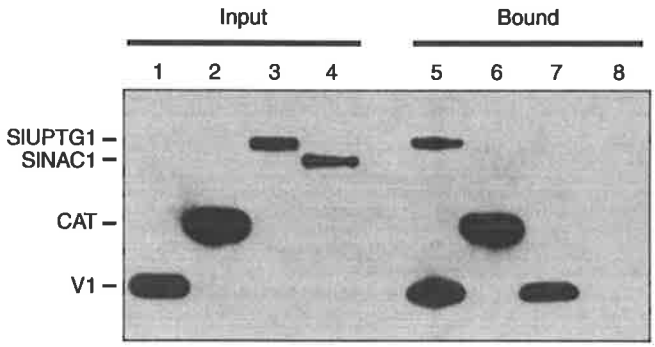


Figure 4.4. V1 interacts with SlUPTG1 *in vitro*. Crude protein samples from cells expressing polyHistidine-tagged and CBP-tagged proteins were mixed, incubated with nickel-nitrilotriacetic acid agarose and washed extensively to remove any unbound protein. Bound protein was resuspended in loading buffer, resolved by SDS-PAGE and analysed by immunoblotting using anti-polyHistidine and anti-FLAG (CBP-tagged proteins also contain a FLAG epitope) antibodies. Reactions were as follows: His-V1 and CBP-SlUPTG1 (lane 5), His-CAT and CBP-SlUPTG1 (lane 6), His-V1 and CBP-SlNAC1 (lane 7), and CBP-SlUPTG1 alone (lane 8). Protein inputs for each reaction are shown: His-V1 (lane 1), His-CAT (lane 2), CBP-SlUPTG1 (lane 3), and CBP-SlNAC1 (lane 4).

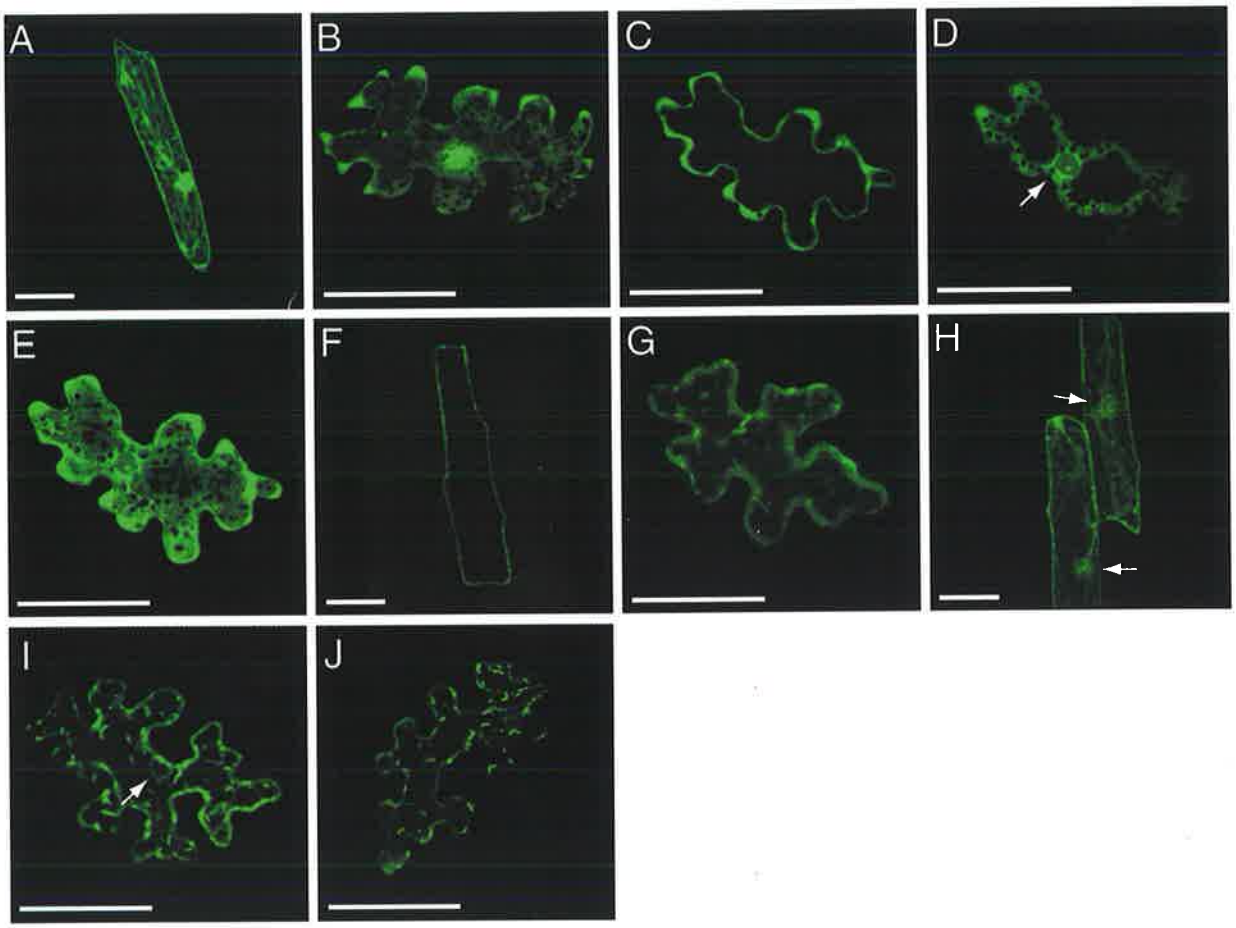


Figure 4.5. Subcellular localisation of V1 and SlUPTG1. V1:GFP (**[C]-[E]**) and SlUPTG1:GFP (**[F]-[J]**), as well as GFP alone (**[A]** and **[B]**), were transiently expressed in onion (**[A]**, **[F]** and **[H]**) or tobacco epidermal (**[B]**, **[C]**, **[D]**, **[E]**, **[G]**, **[I]** and **[J]**) cells. Cells were analysed for GFP fluorescence by confocal microscopy. Nuclei in **(D)**, **(H)** and **(I)** are indicated by arrows. Bars = 50 μm.

Single optical sections of tobacco cells expressing V1:GFP showed peripheral (Fig. 4.5C) and perinuclear (Fig. 4.5D) fluorescence, while a combined image of multiple optical sections also showed fluorescence in the cytoplasm (Fig. 4.5E). The localisation pattern of V1:GFP in onion and tobacco cells resembled that of a GFP fusion to the V1 protein from TYLCV-Israel, which was found to target the endoplasmic reticulum (ER; Rojas et al., 2001).

SlUPTG1:GFP exhibited a similar pattern of fluorescence in both onion and tobacco epidermal cells. Fluorescence was observed at the cell periphery (Fig. 4.5F and 4.5G), around, but not inside, the nucleus (Fig. 4.5H and 4.5I), and associated with cytoplasmic strands (Fig. 4.5H). In tobacco cells, SlUPTG1:GFP was associated with punctate bodies found mainly at the cell periphery and to a lesser extent throughout the cytoplasm (Fig. 4.5G, 4.5I and 4.5J). RGPs from pea (Dhugga et al., 1997) and *Arabidopsis* (Delgado et al., 1998) localise to the Golgi apparatus. However, the fluorescent bodies observed in cells expressing SlUPTG1:GFP do not closely resemble Golgi stacks labelled with GFP-tagged proteins (Brandizzi et al., 2004). Epel *et al.* (1996) immunolocalised a maize RGP to the plasmodesmata (Pd), although the fluorescent punctae observed in tobacco cells expressing SlUPTG1:GFP appear too large to be Pd. Together, these results indicate that SlUPTG1 is associated with cellular membranes and is also found abundantly in the cytoplasm, which is consistent with its primary structure lacking putative transmembrane regions or a signal peptide. In addition, the possibility that it may also target Golgi bodies and/or Pd cannot be ruled out. More important, however, is the finding that SlUPTG1 may co-localise with V1 in the cytoplasm, at the cell periphery, and possibly at the ER.

4.3.6 Glycosylation of SlUPTG1

A number of RGPs from different plant species catalyse the covalent attachment of glycosyl residues from UDP sugars to specific residues in their amino acid sequence, a process termed self-glycosylation (Dhugga et al., 1997; Delgado et al., 1998; Langeveld et al., 2002). To test whether SlUPTG1 is an autocatalytic self-glycosylating protein, extracts of *E. coli* expressing CBP-tagged SlUPTG1 were incubated with UDP-[^{14}C]Glc in the presence of Mn^{2+}, a co-factor for RGPs. Protein was precipitated using TCA, and incorporation of radioactivity into pellets was measured by scintillation counting.

Pellets obtained from SlUPTG1-CBP extracts contained approximately nine times the level of radioactivity of those obtained from a control CBP-tagged protein, SlNAC1 (Table 4.2), suggesting that SlUPTG1 was undergoing self-glycosylation. SlUPTG1-CBP was partially purified using calmodulin-sepharose resin and tested in self-glycosylation assays. This extract incorporated approximately ten times the level of radioactivity compared to the control protein (Table 4.2), discounting the possibility that uptake of labelled glucose by crude SlUPTG1 extracts was catalysed by contaminating *E. coli* components.

The specificity of autoglycosylation by SlUPTG1 was further confirmed by incubating the crude and purified extracts with UDP-[^{14}C]Glc and Mn^{2+} and subjecting these reactions to SDS-PAGE and autoradiography. A radioactive band (Fig. 4.6, lower panel) corresponding to the electrophoretic mobility of SlUPTG1 (upper panel) was observed. This result confirmed that uptake of radiolabeled Glc from UDP-Glc is specifically mediated by SlUPTG1. It must be noted that, while incorporation of radioactivity by crude and pure SlUPTG1 samples was similar when measured by

Table 4.2. Glycosylation of recombinant SlUPTG1 by UDP-[^{14}C]Glc.

Protein	Radioactivity (cpm)	
	Crude	Pure
SlUPTG1	313 (± 13.5[a])	294 (± 24.1)
SlNAC1	35 (± 3.3)	27 (± 4.9)

[a]Numbers shown in brackets are the standard error of the mean cpm.

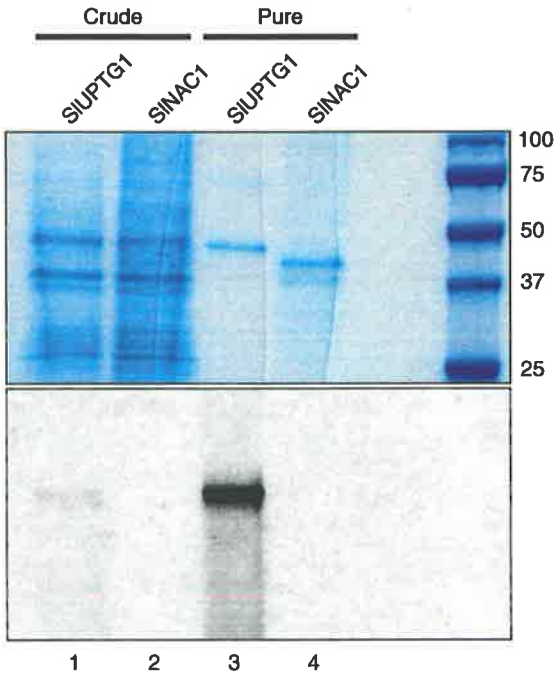


Figure 4.6. Self-glycosylation activity of recombinant SlUPTG1 produced in *E. coli*. 50 mg of crude (lane 1) or 5 mg of partially purified SlUPTG1 (lane 3) was incubated with UDP-[^{14}C]Glc prior to analysis by SDS-PAGE (upper panel) and fluorography (lower panel). Crude (lane 2) and partially purified (lane 4) samples of another bacterially-expressed protein, SlNAC1, were used as negative controls. Molecular mass is indicated in kilodaltons.

precipitation onto filter paper disks (Table 4.2), the radioactivity of these samples following electrophoresis was markedly different. It is likely that this merely reflects the different methods used to precipitate protein in these experiments.

To test whether the self-gycosylation of SlUPTG1 was reversible, as has been observed for other RGPs (Dhugga et al., 1997; Delgado et al., 1998; Bocca et al., 1999a), we incubated purified CBP-tagged SlUPTG1 with UDP-[^{14}C]Glc as the sugar donor and then added unlabelled UDP-Glc and UDP-Gal. Both of these sugars were able to displace radioactivity from the TCA pellets (Table 4.3), indicating that the glycosylation of SlUPTG1 was reversible. This result also suggested that SlUPTG1 can be glycosylated by at least one other UDG sugar, UDP-Gal, which is consistent with previous findings. The donor specificity of SlUPTG1 for other sugars was not tested: however, homologues of SlUPTG1 from pea, potato, and *Arabidopsis* use UDP-Glc, UDP-gal and UDP-Xyl as subtrates but are not glycosylated by UDP-Man (Dhugga et al., 1997; Delgado et al., 1998; Bocca et al., 1999a).

4.3.7 Expression of *SlUPTG1* in tomato

To analyse the endogenous expression of *SlUPTG1*, a northern blot analysis of total RNA preparations from different photosynthetic and non-photosynthetic tissues was carried out. One band of the expected size (~1400nt) was detected in all tissues tested (Fig. 4.7A). The expression of *SlUPTG1* was very low in flowers, leaves and stems, higher in fruit and seed, and higher still in root tissue. A similar expression pattern was observed for *AtRGP1* (Delgado et al., 1998), which was found primarily in roots and suspension-cultured cells, and potato *UPTG1*, which was found mainly in stolons, tubers and roots (Wald et al., 2003). In these studies, the authors identified a correlation

Table 4.3. Ability of different sugar nucleotides to chase out the radioactivity incorporated into SlUPTG1 from UDP-[^{14}C]Glc

Addition (1mM)	Radioactivity (cpm)	
	Mean	Standard error
None	634	± 47.3
UDP-Glc	166	± 22.9
UDP-Gal	109	± 6.1

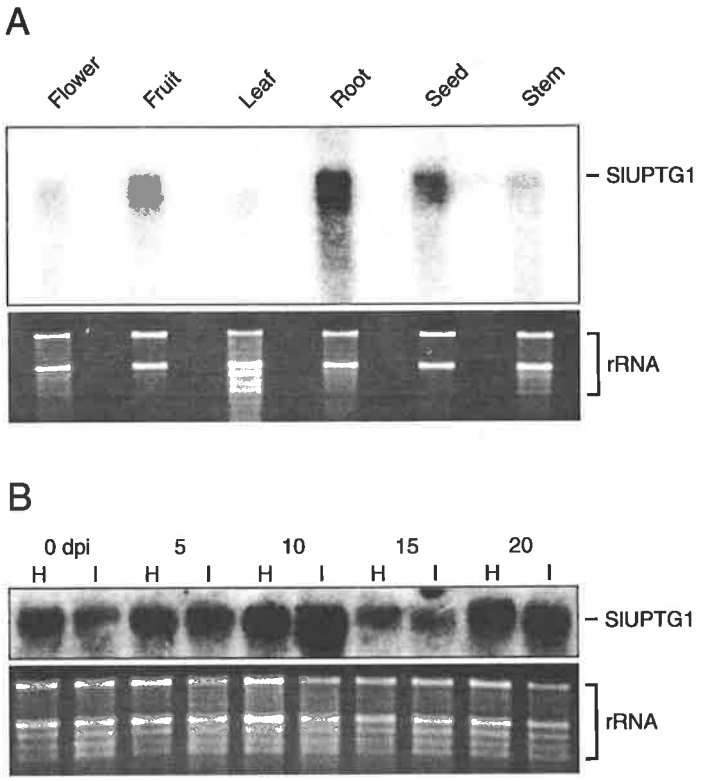


Figure 4.7. Expression of *SlUPTG1* in different tissues and in response to TLCV infection. **(A)** Spatial expression pattern of *SlUPTG1*. RNA gel blot showing that *SlUPTG1* is expressed predominantly in roots, fruit and seed and at much lower levels in flower, stem and leaf tissue. Tissue samples were obtained from 3 week old tomato plants. **(B)** TLCV infection does not effect *SlUPTG1* gene expression. RNA gel blot showing the expression of *SlUPTG1* in healthy (H) or TLCV-infected (I) tomato plants. Tissue samples were obtained at 0, 5, 10, 15, and 20 days post-inoculation (dpi).

between RGP transcript levels and the growing state of tissues. Given that fruit, seed and root tissues are generally more active that flowers, leaves and stems, our results are also indicative of such a link.

To test whether *SlUPTG1* transcription might be regulated by TLCV infection, total RNA from leaf tissue of infected and healthy plants sampled at various time points post-inoculation was analyzed by northern blotting (Fig. 4.7B). The amount of *SlUPTG1* transcript did not appear to change in response to the presence of TLCV. Some fluctuation in the level of *SlUPTG1* expression in healthy plants over the course of this experiment was observed, a finding that correlates with the putative function of SlUPTG1 in plant development by synthesizing cell wall components.

4.3.8 *SlUPTG1* expression increases the accumulation of TLCV ssDNA and may enhance viral movement

To investigate the function of *SlUPTG1* in relation to TLCV infection, the effect of constitutive, high-level expression of this gene on TLCV replication in *N. benthamiana* leaf strips was analysed. *Agrobacterium*-mediated delivery of a 35S-SlUPTG1 construct significantly increased the accumulation of TLCV ssDNA compared to an empty expression construct (Fig. 4.8A). This response was observed in three independent experiments in which all treatments were carried out in duplicate. To determine whether this increase in viral DNA accumulation was due to enhancement of viral replication or movement, an experiment was performed using a TLCV derivative with a mutation in *V2* (V2mut; Rigden et al., 1993). This virus is unable to move cell-to-cell or systemically, and hence an increase in DNA accumulation of V2mut can be attributed wholly to enhanced replication. Expression of SlUPTG1 did increase the

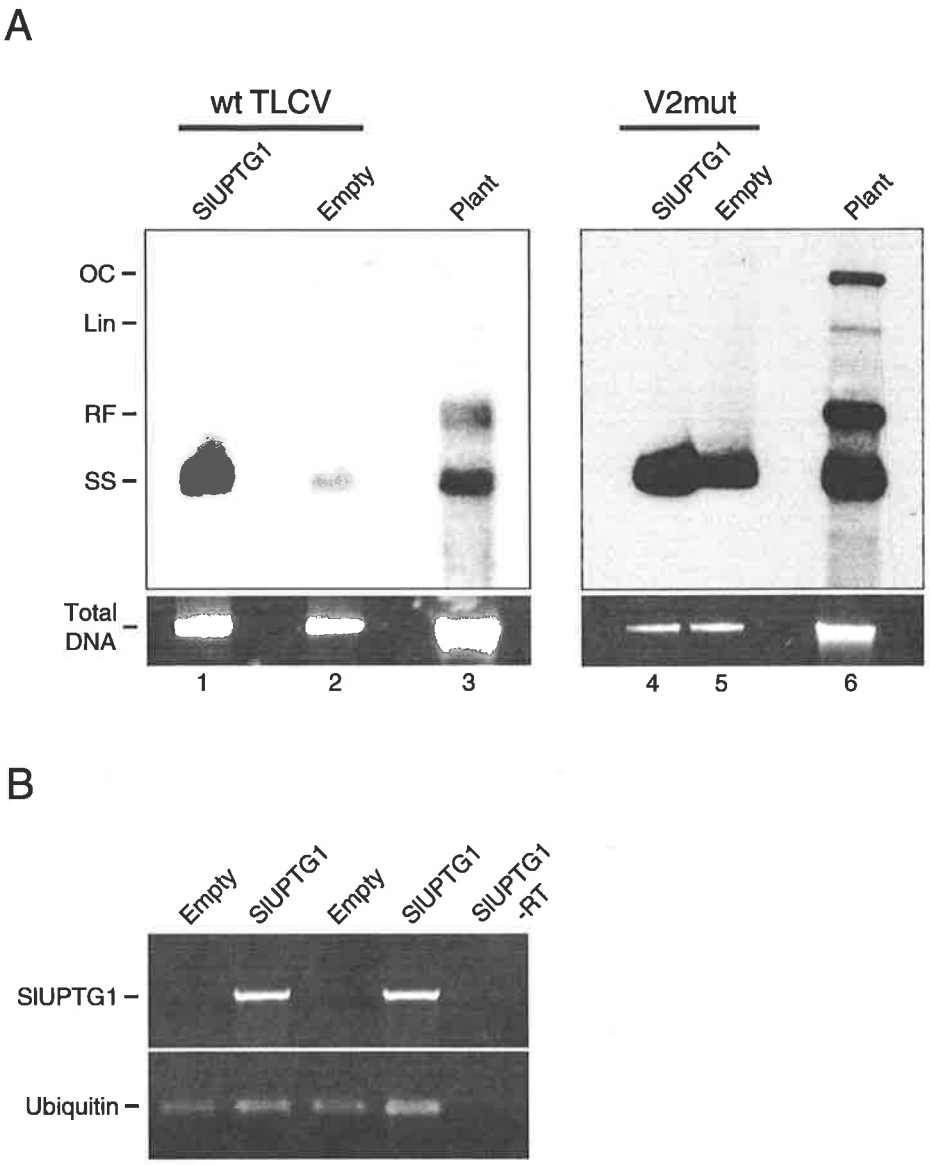


Figure 4.8. Effect of transient *SlUPTG1* expression on TLCV ssDNA accumulation. **(A)** Expression of *SlUPTG1* increases the accumulation of wild-type (wt) TLCV ssDNA but not that of a TLCV *v2* mutant (V2mut) in a transient replication assay. *A. tumefaciens* cells harbouring Bin19-TLCV (lanes 1 and 2) or Bin19-V2mut (lanes 4 and 5) were combined with *A. tumefaciens* cells containing either an empty expression construct (lanes 2 and 5) or p35S-SlUPTG1 (lanes 1 and 4) and co-cultivated for 48 h with leaf strips from *N. benthamiana* plants. DNA was extracted from tissue samples 3 days later and replication of TLCV and V2mut analysed by Southern blotting. Lanes 3 and 6 (labelled "Plant") are samples extracted from TLCV-infected whole *N. benthamiana* plants used as a marker for TLCV DNA forms, marked OC (open circular double-stranded), Lin (linear double-stranded), RF (supercoiled double-stranded replicative form) and SS (single-stranded). **(B)** Analysis of *SlUPTG1* expression by p35S-SlUPTG1 in *N. benthamiana* leaf strips by semi-quantitative RT-PCR. Total RNA was prepared from leaf strips treated with TLCV plus an empty expression construct or TLCV plus p35S-SlUPTG1 (two replicates of each are shown). Ubiquitin mRNA served as an internal control. RT reaction mix without reverse transcriptase was used as a negative control (marked -RT).

level of V2mut ssDNA, but this induction was insignificant compared to that observed when wild-type virus was used. This suggests that SlUPTG1 acts primarily to enhance viral movement. Semi-quantitative RT-PCR revealed high-level *SlUPTG1* expression in leaf strips, verifying the integrity of p35S-SlUPTG1 (Fig. 4.8B).

4.4 Discussion

In this chapter, the identification and characterisation of a tomato protein, SlUPTG1, is described. SlUPTG1 was retrieved from a yeast two-hybrid screen using the TLCV V1 protein as bait. It shares a high degree of similarity with numerous plant RGPs that have recently been grouped into the RGP1 class (Langeveld et al., 2002). RGP1 proteins are autocatalytic glycosyltransferases that may be involved in the synthesis of complex polysaccharides (hemicelluloses) that comprise the cell wall. Three lines of evidence support the idea that SlUPTG1 may have a similar function. First, a recombinant form of the protein produced in *E. coli* underwent self-glycosylation *in vitro*. Second, a SlUPTG1:GFP fusion protein localised primarily to the cytoplasm, cellular membranes and to large cytoplasmic structures (possibly Golgi bodies), suggesting that it could act as a carrier of UDP-sugars from the cytoplasm to the sites of hemicellulose synthesis. A similar localisation pattern has been reported for a RGP1 from *Arabidopsis* (Delgado et al., 1998). Finally, expression of *SlUPTG1* was highest in actively dividing tissues.

Apart from its endogenous function, SlUPTG1 also appears to play an important role in TLCV infection. Overexpression of SlUPTG1 in a transient TLCV replication system promoted the accumulation of TLCV ssDNA. By utilising a movement-defective TLCV derivative, we have shown that this increase in DNA may relate primarily to

SlUPTG1-mediated enhancement of viral movement, presumably via its interaction with V1.

A number of alternative mechanisms could explain this result. First, the partial co-localisation of V1 and SlUPTG1 observed in this study suggests that SlUPTG1 could act to transport V1 to the Golgi, ER and/or plasma membrane where it would execute its movement-associated function. This idea is reinforced by our observation that a V1:GFP fusion protein may partially localise to the ER even though the primary sequence of V1 is not predicted to contain a secretory signal peptide.

A second possibility is that SlUPTG1 serves to glycosylate V1. It is well established that post-translational modification systems are critical to the function of viral proteins (Castillo et al., 2004). N-linked glycosylation, in which oligosaccharides are added to specific asparagine residues in the context of the consensus sequence Asn-X-Ser/Thr (Kornfeld et al., 1985), plays an important role in the function of many viral proteins, particularly those found on the surface of enveloped viruses (Goffard and Dubuisson, 2003). Analysis of the primary sequence of V1 using the YinOYang algorithm (http://www.cbs.dtu.dk/services/YinOYang/) indicates that it has two asparagine residues, at positions 58 and 71, that are potential N-glycosylation sites. This hypothesis is further supported by the proposed co-localisation of V1 and SlUPTG1. However, attempts to detect radiolabelled V1 in glycosylation assays containing recombinant SlUPTG1 and UDP-[^{14}C]Glc were unsuccessful (data not shown). In addition, to our knowledge there have been no reports of glycosylation of substrate proteins by RGPs. Despite this, we cannot discount the possibility that SlUPTG1 acts

to glycosylate V1 and that this process is important in V1-mediated assistance of TLCV cell-to-cell movement.

Finally, the possibility that the enzymatic activity of SlUPTG1 has another role in TLCV infection cannot be excluded. In this scenario, V1-binding could serve to enhance the autoglycosylation of SlUPTG1, or alternatively direct SlUPTG1 to transfer sugars to donor proteins and/or structures which may facilitate virus movement. The presence of recombinant V1 protein, even in great excess, was unable to alter the the uptake of glucose from UDP-[^{14}C]Glc by SlUPTG1 (data not shown). However, these assays were performed *in vitro* and may not reflect the actual events in plant cells.

It must be noted that $V1_{TYLCSV}$ did not appear to interact with SlUPTG1 in the yeast two-hybrid system. One explanation for this result is that, unlike TLCV and ACMV, TYLCSV does not impact on cell wall biosynthesis pathways via SlUPTG1 during infection. This is conceivable given that TLCV and ACMV are more closely related to each other (by sequence homology) than to TYLCSV. Alternatively, the result may have been a false negative, possibly reflecting the disparity between the cellular environment of yeast and plant cells. For example, $V1_{TYLCSV}$ may not be folded properly in yeast. Alternatively, production of $V1_{TYLCSV}$ may impinge on cellular processes in yeast, subsequently affecting its binding to SlUPTG1 and/or visualization of this event.

To summarise, the role of V1 in cell-to-cell movement may relate to its interaction with SlUPTG1. V1 is not essential for viral infection, but is important in symptom expression. This suggests that *SlUPTG1*-silenced host plants might not develop the

symptoms normally associated with TLCV infection, and hence exhibit some tolerance to this pathogen.

Chapter 5 - Characterisation of a ubiquitin conjugating enzyme that interacts with βC1 protein from *Cotton leaf curl virus* DNA β

5.1 Introduction

Several genera of RNA plant viruses have satellite RNAs associated with them. Satellite RNAs depend on a helper virus for their replication but are dispensable for proliferation of the helper virus (Murant and Mayo, 1982; van Regenmortel et al., 2000). The first satellite of a DNA virus was isolated from TLCV infections by Dry *et al.* (1997) in this laboratory. It is a 682 nucleotide covalently-closed ssDNA molecule that has a stem-loop containing the conserved TAATATTAC nonanucleotide typical of geminiviruses (Section 1.1.9.1.1) but no ORFs. The TLCV satellite is not involved in TLCV replication and has no effect on symptom expression.

A new group of DNA satellite molecules, termed DNA βs, that are associated with monopartite begomoviruses was recently identified (Briddon et al., 2003; Mansoor et al., 2003). In contrast to the TLCV satellite and RNA virus satellites, DNA βs affect the replication of their helper viruses and alter the symptoms induced in some host plants (Saunders et al., 2000; Briddon et al., 2001). The organisation of DNA β sequences is conserved and consists of a single complementary-sense ORF (βC1), an adenine rich region, and a satellite-conserved region that carries sequence similarity to the TLCV satellite (Figure 5.1; Briddon et al., 2003). Recent evidence indicates that the

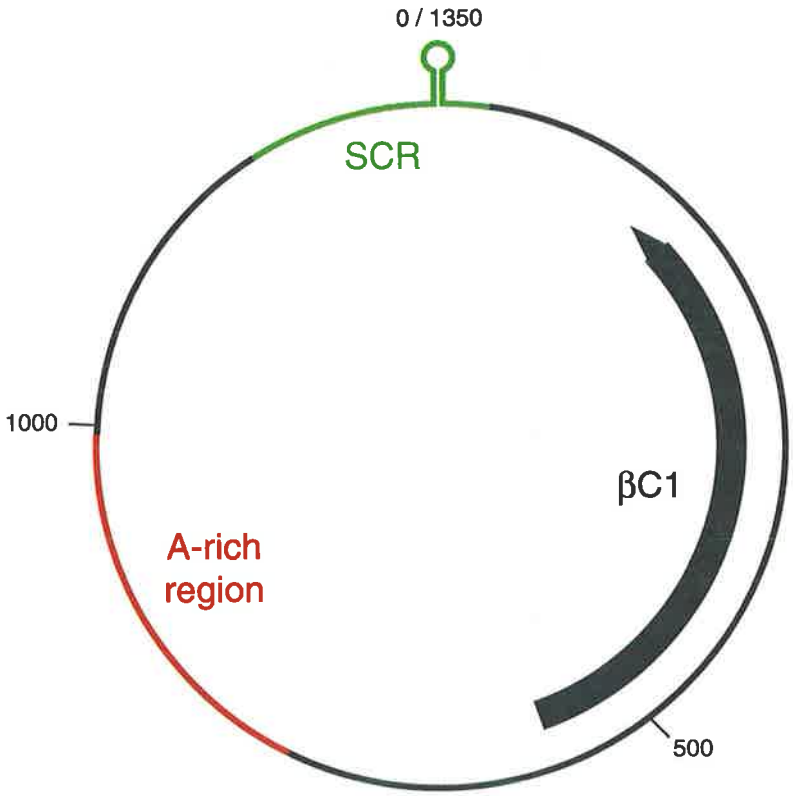


Figure 5.1. Consensus genome map for βDNA molecules. The position and orientation of the conserved β*C1* gene is shown as an arrow within the circle. The relative position of the satellite conserved region (SCR) and A-rich region are coloured (reproduced from Briddon et al., 2003).

pathogenicity of DNA β is mediated by the βC1 gene (Cui et al., 2004; Saunders et al., 2004; Saeed et al., 2004). Expression of βC1 alone, either transiently with a PVX vector or stably in transgenic plants, induces severe developmental abnormalities, vein-greening, and cell proliferation in the vascular bundles. The function of βC1 is unknown, but it has been implicated in suppression of gene silencing (S. Mansoor, personal communication) and movement of monopartite begomoviruses (M. Saeed, personal communication).

The ubiquitin-proteasome pathway is responsible for the selective degradation of abnormal and short-lived regulatory proteins, including transcription factors, cell cycle proteins and signal transducers, and hence plays a central role in many eukaryotic cellular processes (Laney and Hochstrasser, 1999). This pathway involves the attachment of a small protein, ubiquitin, to a target protein which is subsequently recognised and targeted by the 26S proteasome. Covalent bonding of ubiquitin to its substrate generally occurs by: activation of ubiquitin by a ubiquitin activating enzyme (E1); transfer of ubiquitin from E1 to a ubiquitin conjugating enzyme (UBC or E2); and attachment of ubiquitin to the target, a step catalysed by the UBC alone or UBC in combination with an ubiquitin ligase (E3) (Schwartz and Hochstrasser, 2003).

The causative agent of cotton leaf curl disease is a complex between *Cotton leaf curl virus* (CLCuV) and a DNA β molecule (Briddon et al., 2001). This DNA β is also replicated by TLCV and enhances the symptoms of TLCV infection (M. Saeed, personal communication). We have been studying the βC1 protein encoded by this satellite in an effort to better understand its pathogenic effects. This chapter presents evidence that βC1 interacts with the ubiquitin degradation pathway by binding a new

tomato UBC. The possible involvement of this putative interaction in βC1-mediated pathogenesis is discussed.

5.2 Materials and methods

5.2.1 Construction of bait vectors and two-hybrid screens

The βC1 ORF was amplified with primer pairs βC1-F-EcoRI and βC1-R-XhoI respectively, digested with EcoRI/XhoI, and ligated into similarly digested pLexA. Yeast two-hybrid screening of the tomato cDNA library was carried out as described (Section 2.2.17).

5.2.2 5'-RACE

The full SlUBC ORF sequence was determined by 5'-RACE (carried out by A. Behjatnia, CSIRO Plant Industry, Horticulture Unit).

5.2.3 Complementation of yeast *ubc4/ubc5* by SlUBC

S. cerevisiae wild-type, *ubc4*, *ubc5* and *ubc4/ubc5* mutants were kindly provided by Dr. Stefan Jentsch. The SlUBC ORF was amplified with primers βC1-F-BamHI and βC1-R-ClaI, digested with BamHI/ClaI, and ligated into similarly digested pCM188 and pCM190 to generate pCM188-SlUBC and pCM190-SlUBC.

The yeast *ubc4/ubc5* double mutant was transformed separately with pCM188, pCM188-SlUBC, pCM190 and pCM190-SlUBC and selected on media lacking uracil. Yeast cells were grown at 30°C to analyse growth rates.

5.2.4 Analysis of the SUMO conjugating activity of SlUBC

The yeast *ubc9-2* mutant (YWO98) is isogenic to YWO2 (*MATα, his3-Δ200, leu2-3,-112, lys2-801, trp1-1, ura 3-52*) but has a deletion of the *UBC9* gene (Betting and Seufert, 1996). YWO98 was transformed with vectors pCM188, pCM188-SlUBC, pCM190 and pCM190-SlUBC, as well as pCM188-NbSCE1 and pCM190-NbSCE1 that can complement the *ubc9-2* mutation (Castillo et al., 2004). Cells were selected on media lacking uracil and grown at 25°C and 37°C for 5 days.

5.3 Results

5.3.1 Identification of a novel UBC from tomato

Two separate screens of the tomato library retrieved plasmids containing the same 700 bp cDNA, initially termed *17A* (Appendix 1). This sequence contained a putative ORF of 354 bp encoding a predicted translation product of 117 amino acids. A BLAST search revealed that the 17A protein sequence shared sequence identity with tomato UBC4 and with various other plant UBCs, and it was therefore designated SlUBC. SlUBC did not have a methionine residue at its N-terminus and, based on a sequence alignment with UBC4, was truncated by 31 amino acids. 5'RACE was used to determine the missing sequence. The full-length SlUBC ORF was 447 bp in length with a putative translation product of 148 amino acids (Appendix 1).

GAP alignments of the full-length SlUBC protein and ORF with tomato UBC4 sequences showed that these proteins shared 100% sequence identity, while the ORFs were 90% identical. Thus, *SlUBC* probably represents an allele of the tomato *UBC4* gene.

5.3.2 Sequence analysis of SlUBC

The sequence identity and similarity between SlUBC and various UBCs from plants, yeast and mammals is shown in Table 5.1. Apart from tomato UBC4, SlUBC was most closely related to UBCs from cotton, *Arabidopsis*, and rice. SlUBC also shared a high degree of sequence identity, ranging from 77.6 and 80.3% sequence identity (83.0-85.0% similarity), with UBC proteins from humans, mice, yeast, and zebrafish. All of the UBC proteins shown in Table 5.1 are members of the Class I subfamily of UBCs (Jentsch, 1992). The UBC catalytic domain, [FYWLSP]-H-[PC]-[NH]-[LIV]-x(3,4)-G-x-[LIV]-C-[LIV]-x-[LIV] (PROSITE: PDOC00163), was found in SlUBC from amino acids 74 to 86 (FHPNINSNGSICL). The invariant cysteine residue is the active site, accepting ubiquitin from E1 to form a thiol ester (Jentsch, 1992).

5.3.3 *In vivo* UBC activity of SlUBC

To determine whether SlUBC has the same function as yeast UBC4 and UBC5, the ability of SlUBC to complement yeast containing defective *UBC4* and *UBC5* genes was was tested. The growth rate of *ubc4* and *ubc5* single mutants is similar to wild-type yeast, but *ubc4/ubc5* double mutants exhibit markedly reduced growth (Seufert and Jentsch, 1990).

The yeast *ubc4/ubc5* double mutant was transformed with empty plasmids (a centromeric plasmid pCM188 and a multicopy plasmid pCM190) or with the plasmids expressing SlUBC, and the transformants were selected on plates lacking uracil. *ubc4/ubc5* cells expressing SlUBC grew at a rate comparable to wild-type yeast and *ubc4* and *ubc5* single mutants on YPG plates incubated at 30°C, which was considerably greater than the untransformed double mutant or the double mutant

Table 5.1. Homology between SlUBC and UBCs from plants, vertebrates and yeast

Protein	Organism	Accession #	% identity to SlUBC (% similarity)[a]
UBC4	Tomato	P35135	100.0 (100.0)
E2	Cotton	AAL99223	97.3 (98.0)
E2	Tomato	CAA51821	96.6 (97.3)
UBC10	*Arabidopsis*	P35133	95.9 (97.3)
OsUBC5b	Rice	BAB89355	95.9 (97.3)
UBC8	*Arabidopsis*	P35131	95.9 (96.6)
UBC9	*Arabidopsis*	P35132	95.3 (95.9)
UBC11	*Arabidopsis*	P35134	93.9 (95.9)
OsUBC5a	Rice	BAB89354	93.9 (95.2)
HBUCE1	Human	AAD31180	80.3 (85.0)
E2D 2	Mouse	P51669	80.3 (84.4)
UBC4	Yeast	P15731	78.9 (85.7)
E2D 2	Zebrafish	AAH47863	78.2 (83.0)
UBC5	Yeast	P15732	77.6 (84.4)

[a]As determined by the GAP algorithm.

transformed with empty expression vectors (Figure 5.2). This indicates that SlUBC can complement the UBC4 and UBC5 proteins from yeast, which function in the selective degradation of abnormal and short-lived proteins (Seufert and Jentsch, 1990).

5.3.4 SlUBC does not act as a SUMO-conjugating enzyme in yeast

Sumoylation is a post-translational process whereby a target protein is conjugated to a small, ubiquitin-like polypeptide called SUMO (Schwartz and Hochstrasser, 2003). This process is mechanistically similar to ubiquitination, involving separate activation and conjugation events mediated by SUMO activating, SUMO conjugating (SCE) and SUMO ligase enzymes. However, sumoylation does not direct the degradation of the target protein but rather modifies its function, activity, or localisation.

A number of mammalian viral proteins, including cytomegalovirus IE1 and IE2, Epstein-Barr virus BZLF1, and papillomavirus E1, interact with SUMO and/or SCE (Wilson and Rangasamy, 2001). More recently, Castillo *et al.* (2004) found that Rep proteins from TGMV and TYLCV-Sardinia interact with the *N. benthamiana* SCE1 (NbSCE1). The exact role of the Rep/NbSCE1 interaction is unclear, but the replication of TGMV in plants expressing either increased or decreased levels of SUMO is impaired, suggesting that post-translational protein modification of Rep by SUMO plays an important role in geminivirus replication.

SlUBC shares 37% sequence identity (45% similiarity) with NbSCE1. To test whether SlUBC can utilise SUMO as a substrate in addition to ubiquitin, a yeast strain (YWO98) containing a temperature-sensitive mutation in the *NbSCE1* orthologue, *UBC9*, was utilised (Betting and Seufert, 1996; Castillo et al., 2004). Wild-type and YWO98 yeast

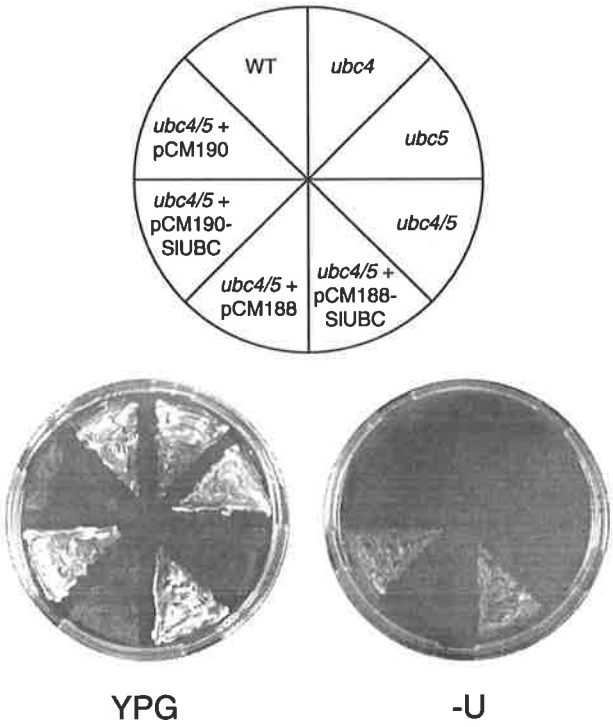


Figure 5.2. Functional complementation of yeast UBC4/UBC5 by tomato SlUBC. The yeast *ubc4*/*ubc5* double mutant was individually transformed with pCM188, pCM188-SlUBC, pCM190 and pCM190-SlUBC. These transformants were streaked with wild type, *ubc4*, *ubc5* and *ubc4*/*ubc5* yeast on YPG media or synthetic dropout media lacking uracil. The location of yeast strains streaked on the plates is shown in the diagram (top). Photographs were taken after a 2-day incubation at 30°C.

were transformed with empty pCM188 and pCM190 or with the plasmids expressing *NbSCE1* (kindly provided by E. Bejerano) or *SlUBC*. Yeast transformed with any of the plasmids grew at 25°C (Fig. 5.3, left panel), but only cells expressing NbSCE1 were viable at 37°C (right panel). This suggests that SlUBC does not act as a SCE in yeast.

5.4 Discussion

In this chapter, the identification and functional characterisation of a new Class I ubiquitin conjugating enzyme, SlUBC, is described. This protein was retrieved from a yeast two-hybrid screen using the βC1 protein from CLCuV DNA β as bait. The role of SlUBC and the ubiquitin degradation pathway in βC1 function is currently unknown, and the studies presented here clearly need to be extended to confirm this putative interaction. However, a number of possiblities can be envisioned. Two basic helix-loop-helix transcription factors, key regulatory proteins controlling cell cycling and differentiation, interact with ubiquitin conjugating enzymes which direct their degradation via the proteasome (Kho et al., 1997; Xu et al., 2000). By analogy, SlUBC may function to regulate turn-over of βC1. Given the phytotoxicity of βC1, this process may be advantageous to the geminivirus/DNA β disease complex by preventing cell death and thereby allowing optimal replication and movement. Alternatively, SlUBC may function as a cellular defence mechanism that minimises βC1-mediated damage to the plant. The observation that wild-type βC1 transcript accumulates to much lower levels than does a mutated copy when expressed by the CaMV 35S promoter in transgenic plants (Saeed et al., 2004), suggesting that the host represses production of βC1 to reduce toxicity, supports this hypothesis.

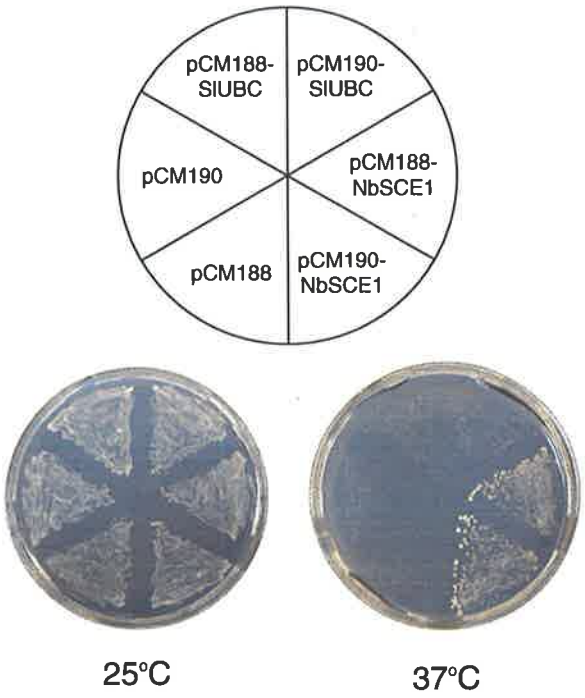


Figure 5.3. Test of SUMO conjugating activity of SlUBC. Transformants of the temperature sensitive *S. cerevisiae ubc9-2* mutant (YWO98) harbouring pCM188, pCM188-SlUBC, pCM190 and pCM190-SlUBC were streaked on selective plates and incubated at 25 and 37°C. YWO98 yeast harbouring pCM188-NbSCE1 and pCM190-NbSCE1 were used as a positive control for complementation. Photographs were taken after 5 days incubation.

Targeting and modification of the host ubiquitination system is an emerging theme in virus research. A number of mammalian oncoviruses utilise E3 enzymes to direct the degradation of undesirable cellular proteins. For example, the E6 protein of papillomaviruses interacts with both the tumor suppressor p53 and a host E3, a bridging process which mediates the proteasomal degradation of p53 and thereby stimulates growth of deregulated cells (Scheffner and Whitaker, 2003). Herpesviruses encode their own E3s that specifically target membrane-associated MHC class I molecules for endosomal destruction (Coscoy and Ganem, 2000). By analogy, βC1 may direct the degradation of target host protein(s), perhaps one or more of the proteins with which it interacts in yeast (Sections 6.3.5 and 6.4.5), by linking them to SlUBC. This idea is further supported by the observation that UBCs belonging to the Class I subfamily are very poor at transferring ubiquitin by themselves and probably require an E3 *in vivo* (S. Jentsch, personal communication). βC1 does not contain either of the motifs found in all E3s so far, namely a HECT domain or a RING finger (Schwartz and Hochstrasser, 2003). However, a nucleoporin which has neither HECT nor RING motifs is capable of acting as an E3 for SUMO, suggesting that the family of E3s may be more diverse than previously thought (Pichler et al., 2002; Pichler et al., 2004).

Ubiquitination is best known as a modification that directs proteins to the proteasome, where they are degraded. However, increasing evidence suggests that attachment of ubiquitin may also have a role in modifying protein structure and/or function. For example, addition of single ubiquitin molecules to histone H2B functions in regulating chromatin structure and transcription by enabling methylation of histone H3 (Briggs et al., 2002). In addition, viruses such as Ebola and Human immunodeficiency virus utilise mono-ubiquitination to orchestrate the budding of enveloped viruses from the

plasma membrane (Bonifacino and Traub, 2003). Thus, the possibility that SlUBC-mediated ubiquitination of βC1 does not cause destruction of βC1 but rather alters its function cannot be discounted.

The function of a number of viral proteins is altered by sumoylation (Wilson and Rangasamy, 2001). For at least three of these proteins, adenovirus E1A (Hateboer et al., 1996), papillomavirus E1 (Rangasamy and Wilson, 2000), and cytomegalovirus IE2 (Ahn et al., 2001), this process involves their interaction with the mammalian SCE Ubc9. Sumoylation of viral proteins has been shown to alter their localisation (Rangasamy et al., 2000) and transactivation activity (Ahn et al., 2001). In this study, SlUBC was unable to complement yeast deficient in SCE activity. Despite this result, it is possible that SlUBC can utilise SUMO, in addition to ubiquitin, as a substrate *in planta*. Supporting this idea, Ubc9 appears to act as both a SUMO and ubiquitin conjugating enzyme (Xu et al., 2000). Thus, SlUBC may sumoylate βC1 to modify its function and/or subcellular localisation.

Finally, although the interaction of the ubiquitination machinary with substrate proteins is likely to be transitory, it is possible that the retrieval of SlUBC from the βC1 yeast two-hybrid screen is simply an artifact of the system. However, no UBCs were isolated from other two-hybrid screens, and plasmids containing cDNAs encoding SlUBC were isolated from two separate experiments. The authenticity of this interaction should be tested using another system, such as an *in vitro* binding assay (Sections 3.3.3 and 4.3.4). Another important future experiment to validate the role of SlUBC in βC1 function would be to analyse the ability of diverse βC1 proteins to interact with this host protein.

Chapter 6 – Putative host partners of Rep, C2, C4, V1 and βC1 proteins

6.1 Introduction

Expression of the *V1*, *C1*, *C2*, *REn* and *C4* genes from TLCV and *βC1* from CLCuV DNA β produces distinct phenotypic effects in various host plants, suggesting that their translation products may interact with host factors to achieve their function (Selth et al., 2004; Saeed et al., 2004). The three previous chapters (3, 4 and 5) all describe new host proteins, identified using the yeast two-hybrid system, which are involved in REn, V1 and βC1 function respectively. This chapter describes the isolation of host factors from yeast two-hybrid screens using the Rep, C2 and C4 proteins as bait. In addition, the sequences of other proteins retrieved from the V1 and βC1 screens are presented. The potential role of these proteins in geminivirus infection is discussed.

6.2 Materials and methods

6.2.1 Construction of bait vectors and two-hybrid screens

The construction of pLexA-REn and pLexA-V1 has been described in Chapters 3 and 4 respectively. Other TLCV sequences used in yeast two-hybrid screens were digested with the appropriate restriction enzymes (Table 6.1), ligated into similarly digested pLexA, and if necessary, clones with inserts in the correct orientation identified by restriction enzyme analysis. The integrity of the inserts was verified by sequencing (Section 2.2.21).

Table 6.1. Strategy for cloning viral genes into pLexA

Sequence	Oligonucleotides used to amplify sequence by PCR (see Table 2.1 for sequences)	Digest prior to ligation with pLexA
V2	V2-F-EcoRI, V2-R-XhoI	EcoRI, XhoI
C1	C1-F-BamHI, C1-R-BamHI	BamHI
$C1_{184-362}$	$C1_{184-362}$-F-BamHI, C1-R-BamHI	BamHI
C2	C2-F-BamHI, C2-R-BamHI	BamHI
$C2_{1-83}$	C2-F-BamHI, $C2_{1-83}$-R-BamHI	BamHI
$C2_{33-103}$	$C2_{33-103}$-F-EcoRI, $C2_{33-103}$-R-XhoI	EcoRI, XhoI
$C2_{84-136}$	$C2_{84-136}$-F-BamHI, C2-R-BamHI	BamHI
$C2_{1-103}$	C2-F-BamHI, $C2_{33-103}$-R-XhoI	BamHI, XhoI
C4	C4-F-EcoRI, C4-R-XhoI	EcoRI, XhoI

Yeast two-hybrid screening of the tomato cDNA library was carried out as described in Section 2.2.17.

6.2.2 Sequence analysis

GeneDoc version 2.4.016, NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and ORF Finder (http://www.cbio.psu.edu/sms/orf_find.html) were used to detect putative ORFs and translation products within cDNA sequences retrieved from yeast two-hybrid screens. Similar protein and gene sequences were identified using BLAST (Altschul et al., 1990). Comparisons between protein sequences were carried out using the Gap (Needleman and Wunsch, 1970) or BestFit (Smith and Waterman, 1981) algorithms through BioManager 2.0 (http://bioman1.angis.org.au).

6.3 Results

The full cDNA, ORF and protein sequences retrieved from each of the yeast two-hybrid screens are included in Appendix 1. Most of the tomato cDNA sequences are truncated at their 5'-ends due to the method used to create cDNA libraries. cDNA synthesis is initiated by oligo (dT)-primed reverse transcription, which ensures the inclusion of the 3'-end of most mRNA species in the library, and helps to orient the clones during subsequent analysis. However, the majority of clones will be truncated at the 5'-end because it is difficult for reverse transcriptase to negotiate long RNA templates and the stable secondary structures often encountered in RNA molecules.

6.3.1 V1

A screen of the cDNA library using the V1 protein as bait identified three putative interacting proteins. One of these, designated SlUPTG1, shares considerable sequence identity with two classes of reversibly glycosylated peptides from plants. Further characterisation of the role of SlUPTG1 in TLCV infection is described in Chapter 4.

The other cDNA sequences retrieved from the V1 screen were designated *9A* and *14A*. The 1096 bp sequence of *9A* encoded a 364 amino acid protein which shared 47% sequence identity and 55% similarity with an expressed protein from *Arabidopsis* (accession number NP_188295) of unknown function. 9A was truncated at the N-terminus and based on an alignment with this *Arabidopsis* protein it probably lacked approximately 320 amino acids.

The 662 bp cDNA of *14A* was retrieved four times from the V1 screen and encoded a predicted translation product of 141 amino acid residues. This peptide shared 57% sequence identity (66% similarity) with a glutaredoxin-like protein from *Arabidopsis* (accession number AAM65800). It lacked a methionine initiation codon but based on sequence comparisons it may only be truncated at the N-terminus by five amino acids. Glutaredoxins (Grxs) are small oxidoreductases of the thioredoxin family that contain an active site sequence CxxC or CxxS required for their redox properties (Fomenko and Gladyshev, 2002). This sequence was found in 14A from amino acids 86-89.

6.3.2 C1

Yeast transformed with pLexA-C1 generally grew more slowly than those containing other bait plasmids in culture and on plates. These yeast colonies could be used to

inoculate new cultures or streaked onto fresh media up to three days post-transformation, but after longer incubations they became abnormally-shaped and their viability decreased. The apparent toxicity of Rep observed in this section was supported by the observation that no LexA-Rep fusion protein was detectable in yeast cells (Section 7.3.1). Furthermore, interacting clones were not recovered from library screens utilising full-length LexA-Rep as bait, despite replication of the experiment.

Transient expression of full-length Rep and Rep amino acids 1-184 (Rep_{1-184}) in plants induced the formation of local lesions resembling a hypersensitive response, but plants expressing $Rep_{184-362}$ were asymptomatic (Section 1.1.9.1.5). The N-terminus of Rep contains the putative pRBR-binding domain, suggesting that the toxicity of Rep_{1-184} is related to deregulation of the cell cycle by disruption of pRBR control mechanisms. Since pathways controlled by RB are highly conserved between kingdoms, it was hypothesised that Rep-mediated toxicity in yeast may occur via a mechanism analogous to that observed in plants and hence C-terminal truncations of Rep lacking the putative pRBR-binding domain would be non-toxic in yeast.

This hypothesis was supported by the apparent lack of toxicity exhibited by the production of a LexA fusion to the C-terminal 179 amino acids of Rep ($Rep_{184-362}$) in yeast cells (this study and Section 7.3.1). A screen of the tomato cDNA library with $Rep_{184-362}$ recovered three cDNA sequences, *1A*, *14A*, and *15A*. The 940 bp *1A* sequence was retrieved from two separate prey plasmids and encoded a 208 amino acid peptide. *14A* and *15A* were 818 and 1007 bp in length encoding predicted translation products of 250 and 232 amino acids respectively. Each of these peptide sequences were similar to plant DAG proteins (Table 6.2). 1A shared a high degree of sequence

Table 6.2. Homology between putative $Rep_{184-362}$ interactors and plant DAG and DAL1 proteins

	14A	15A	AmDAG	AtDAG	AtDAL1
1A/16A	47 (55) [a]	47 (55)	75 (79)	45 (52)	44 (52)
14A		93 (94)	43 (52)	78 (80)	77 (79)
15A			45 (55)	80 (82)	79 (82)
AmDAG				43 (52)	43 (51)
AtDAG					99 (100)
AtDAL1					

[a]The percentage identity and similarity (in brackets) between protein sequences as determined by the GAP algorithm.

identity with DAG from *Antirrhinum majus* (AmDAG; accession number CAA65064), while 14A and 15A were most closely related to a putative *Arabidopsis* DAG (AtDAG; accession number NP_180901). The putative $Rep_{184-362}$ interactors were also similar to *Arabidopsis* DAL1 (AtDAL1; accession number CAA75116) (Table 6.2).

6.3.3 C2

The C2 protein, when fused to the LexA DNA-binding domain, activated transcription of the *GFP* and *leu2* reporter genes in yeast cells that were not expressing the B42 AD. This autoactivation activity supports the hypothesis that C2 positively modulates transcription of the virion-sense genes in begomoviral infections (Section 1.1.9.2). Transcription factors from yeast, animals and plants contain basic domains and zinc-binding motifs, responsible for binding DNA, and acidic domains which activate transcription (Johnson and McKnight, 1989). Similarly, C2 has a modular structure consisting of an acidic C-terminus and a basic N-terminus, and also contains a cysteine/histidine rich region that could form a structure capable of binding zinc (Hartitz et al., 1999; Fig. 6.1). To circumvent the problem presented by C2 autoactivation in yeast, the transcriptional activation domain (TAD) of C2 was mapped. A series of C2 truncations fused to the LexA DNA BD were made and tested for their ability to activate reporter gene expression in yeast. This deletion analysis revealed that the carboxy-terminal region (amino acids 84-136), containing the acidic domain, was responsible for transcriptional activation, whereas neither the N-terminal fragment (amino acids 1-84) nor the fragment encompassing amino acids 33-103 were able to promote transcription (Fig. 6.2). Based on these results, a $C2_{1-103}$ truncation which did not activate transcription in yeast was used for two-hybrid screening.

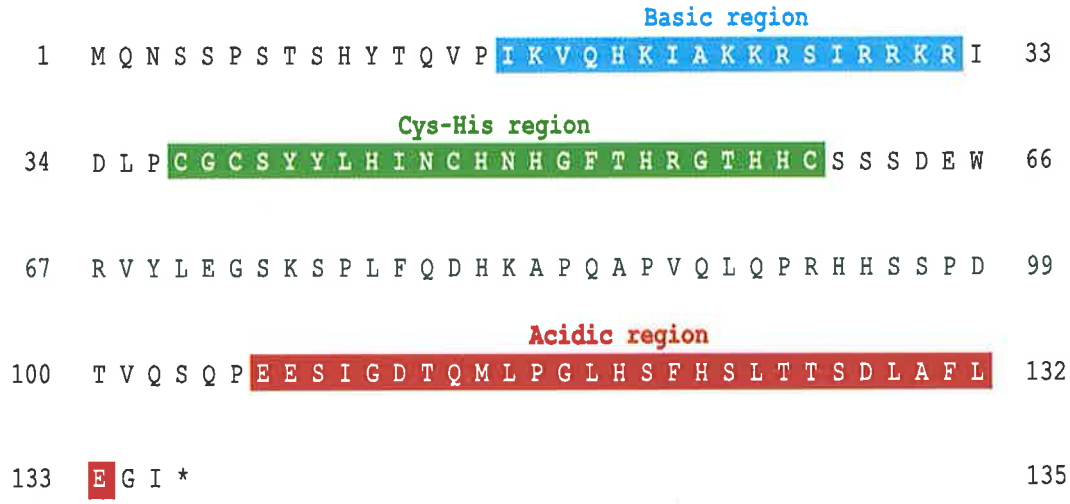


Figure 6.1. Primary structure of C2 showing regions putatively involved in transcriptional activation. The basic, cysteine and histidine (cys-his), and acidic regions are shown in coloured boxes.

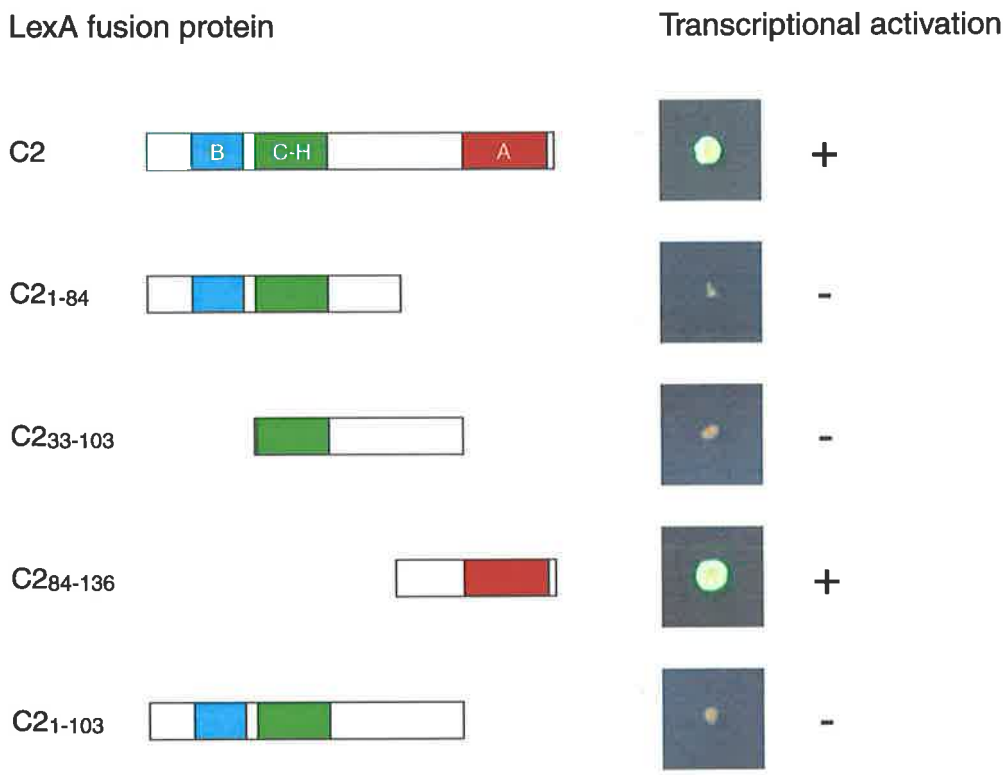


Figure 6.2. Regions of C2 able to activate transcription in yeast. The C2-LexA fusion proteins are represented diagrammatically on the left, with the positions of the basic (B), cysteine-histidine rich (C-H), and acidic (A) regions shown in coloured boxes. The ability of C2-LexA fusion proteins to activate expression of *GFP* and enable protrophic growth on leucine-deficient media is shown on the right, and summarised by + or - designations.

Two putative interactors were retrieved from the $C2_{1-103}$ screen. The cDNA sequence of the first, *13A*, was 1263 bp in length and encoded a predicted translation product of 363 amino acids. This protein shared 61% identity (72% similarity) and 54% identity (63% similarity) with the *Arabidopsis* CAO protein (Klimyuk et al., 1999; accession number AAD01509) and a putative rice CAO respectively (accession number XP_47018). Like CAO, the sequence of 13A contained two segments which are similar to motifs that are known to mediate protein-protein interactions. The first one corresponded to four tandem ankyrin repeats from residues 115 to 244 (Figure 6.3). The second corresponded to two chromatin organisation modifier (chromo) domains found from residues 271 to 368 (Figure 6.3). Chromodomains have been implicated in the regulation of chromatin structure through protein-protein interactions (Paro and Hogness, 1991). The high overall sequence identity between *Arabidopsis* CAO and 13A, and the identical positioning of the four ankyrin repeats and two chromodomains in these proteins, suggests that they are orthologues. Apart from CAO, 13A shared some sequence identity with ankyrin repeat-containing proteins from many and diverse organisms, including plants, *Drosophila*, vertebrates, and *C. elegans*. However, this shared identity only occurred in ankyrin repeat domains and therefore it would be inappropriate to suggest any functional links between 13A and this large group of proteins.

The second interactor, designated 16A, shared 90% identity (93% similarity) with a ubiquitin carrier protein from *Glycine max* (soybean), UBC4, and between 70% and 84% identity (83-89% similarity) with UBC4, UBC5 and UBC6 from *Arabidopsis* (Table 6.3). It was also similar to yeast UBC8, human UBC E2H and murine UBC E2.

A

```
     D  G TPLH-AA-------V-LL---GA-     Consensus
115    EYETPWWNAAKKSDESALRELIEAEDR     13A
147  DDDGRTALLFVSGLGSEPCVKLLAEAGAD     13A
180  RSGGLTALHMAAGYVKPGVAKLLIELGAD     13A
214  DYRGQTPLSLARMVLNQTPKGNPMQFAR      13A
```

B

```
         HHHHHHHH          BBBBBBBB                HHHHHHHHHH
258  EYAQVEEILEKR----GKGENVEYLVKWKDGEDNEWVKAWLISEDLV-RDFEAGL    307   13A
308  EYAEAECILEKREGDNGKG---EYLVKWTDIEEATWEPEENV-----DPLLIEDFEK  356   13A

271  EYAEVDEIVEKR----GKGKDVEYLVRWKDGGDCEWVKGVHVAED-VAKDYEDGL    320   CAO
321  EYAVAESVIGKRVGDDGK--TIEYLVKWTDMSDATWEPQDNV-----DSTLVLLY    368   CAO
```

Figure 6.3. Ankyrin repeats and chromodomains in 13A peptide sequence. **(A)** Similarity to ankyrin consensus motifs (defined by Zhang et al., 1992). Amino acids strictly conserved or closely related are shown in red. **(B)** Similarity to chromodomains in CAO (Klimyuk et al., 1999). A secondary structure prediction generated with PROF secondary structure prediction program (http://www.aber.ac.uk/~phiwww/prof/) is shown. H indicates a putative α helix and B a putative β sheet. Amino acids strictly conserved or closely related are shown in red. Dashes were introduced to optimise alignments.

Table 6.3. Homology between C2 interactor 16A and ubiquitin conjugating enzymes from plants, vertebrates and yeast

Protein	Source	% identity to 1A (% similarity)[a]
UBC4	Soybean (accession number AAF03236)	90.4 (93.4)
UBC4	*Arabidopsis* (NP_568589)	84.0 (89.0)
UBC5	*Arabidopsis* (NP_564817)	83.4 (88.3)
UBC6	*Arabidopsis* (NP_566062)	70.5 (83.1)
UBC E2	Mouse (JC4308)	54.0 (67.7)
UBC E2H	Human (P37286)	54.0 (67.7)
UBC8	Yeast (B53516)	53.9 (69.5)

[a]As determined by the BestFit algorithm.

6.3.4 REn

One cDNA was retrieved from each of two screens of the tomato library using the REn protein as bait. These sequences were identical and encoded a predicted full-length translation product of 301 amino acids, designated SlNAC1. Further characterisation of the role of SlNAC1 in TLCV infection is described in Chapter 3.

6.3.5 C4

A screen with the C4 protein identified eight putative host partners, designated 1A, 5A, 6A, 8B, 11A, 11B, 15B, and 26A.

The cDNA sequences of *1A*, *5A* and *26A* were 1320 nt, 1717 bp and 1416 bp encoding predicted translation products of 352, 380 and 369 amino acids respectively. Each of these proteins contained Serine/Threonine kinase domains. 1A was most closely related to receptor-like protein kinases (RLKs) from soybean (GmRLK2; accession number AAF91323), *Arabidopsis* and rice, while 5A and 26A shared the highest sequence identity with glycogen synthase kinase 3 (GSK-3)/SHAGGY-like kinases (hereafter referred to as SKs) from *Arabidopsis* (AtSKη; accession number AAM63594) and tobacco (NtSK; accession number CAA54803) respectively. The peptide sequences of 1A and 26A did not contain a methionine start and based on alignments to their most closely-related sequences probably lacked approximately 660 and 40 amino acids of amino-terminal sequence respectively. The sequence similarity between 1A, 5A, 26A and related kinases is shown in Table 6.4.

The BLAST results for the other C4 interactors is summarised in Table 6.5. *8B*, encoded by two separate prey plasmids, encoded a 406 amino acid protein which shared

Table 6.4. Homology between putative C4 interactors and plant kinases

	5A	26A	AtSKη	NtSK	GmRLK2
1A	26 (36)[a]	25 (35)	25 (36)	25 (36)	88 (91)
5A		79 (87)	88 (92)	78 (86)	26 (38)
26A			79 (86)	96 (97)	28 (40)
AtSKη				79 (85)	26 (38)
NtSK					25 (36)
GmRLK2					

[a]The percentage identity and similarity (in brackets) between protein sequences as determined by the BestFit algorithm.

Table 6.5. BLAST best hits for C4 interactors 6A, 8B/19A, 11A, 11B, 15B/31A

Interactor	Best hit	% identity (% similarity)[a]
6A	None	NA
8B	Potato isovaleryl-CoA deyhydrogenase (accession number CAC08233)	97 (98)
11A	*Arabidopsis* seh1-like protein, contains WD40 domain (AAG59882)	63 (69)
11B	Wheat heat shock protein Hsp23.5 (AAD03604)	39 (54)
15B	Unknown *Arabidopsis* protein, contains BSA and OMP domains (AAF01515)	57 (69)

[a]As determined by the BestFit algorithm.

97% sequence identity (98% similarity) with the potato isovaleryl-CoA dehydrogenase (IVD) protein and was closely related to IVDs from other plant species. Based on sequence alignments with these proteins, 8B lacked only 5 amino acids at its N-terminus. The 326 amino acid full-length translation product of 11A contained a WD40 domain and shared sequence identity with the nucleoporin proteins seh-1, sec13, and Dal-6. The 604 bp cDNA of *11B* encoded a 101 amino acid peptide which was similar to heat shock proteins from various plant species including wheat, *Arabidopsis*, rice, soybean and pea. *15B*, retrieved twice from the C4 screen, was 1325 bp in length and encoded a predicted full-length protein of 250 amino acids. This protein contained two conserved domains, a bacterial surface antigen domain and an outer membrane protein domain, and was similar to a number of *Arabidopsis* proteins of unknown function. Interestingly, amino acids 1-155 and 156-250 of 15B shared sequence identity with two regions of its *Arabidopsis* relatives which are separated by approximately 265 amino acids, suggesting that these are modular domains that have evolved independently. The 494 bp *6A* sequence encoded a 146 amino acid peptide which did not retrieve any BLAST hits from the GenBank database.

6.3.5 βC1

Two separate screens with the βC1 protein identified six putative host partners, designated 1, 6, 11, 17, 22, and 23.

βC1 interactor 17, encoded by two separate prey plasmids, encoded a protein with homology to plant ubiquitin conjugating enzymes which was designated SlUBC. Further characterisation of SlUBC and its role in βC1 function is described in Chapter 5.

The cDNA sequence of *1* was 1247 bp in length and encoded a predicted translation product of 278 amino acids. This peptide contained a motif found in the GARP family of transcription factors (Riechmann et al., 2000). The closest relative of 1 is the *Arabidopsis* KANADI protein (Kerstetter et al., 2001; accession number AAK59989): these sequences share 93.3 identity (95.6% similarity) over a 92 amino acid stretch containing the GARP domain and an overall sequence identity of 54.6% (59.6% similarity). The nucleotide sequence of *1* lacked an initiation codon and its putative translation product was truncated at the N-terminal end. The number of N-terminal residues missing from 1 is difficult to predict because of the divergence of GARP proteins outside the GARP domain.

The cDNA sequence of *6* was encoded by three separate prey plasmids. It was 1157 bp in length and encoded a predicted full-length translation product of 307 amino acids. This protein shared a high degree of sequence identity with urate oxidases from various plant species including *Lotus corniculatus* (accession number BAB18538), *Cicer arietinum* (accession number CAB77205) and *Medicago sativa* (accession number BAA78607).

The 611 bp cDNA sequence of *11* was retrieved five times from the two βC1 screens. It encoded a 130 amino acid peptide which was similar to a rice protein of unknown function (accession number NP_913510) and 60S acidic ribosomal peptide P3 proteins from various plant species.

The *22* cDNA sequence, retrieved once from each of the two βC1 screens, was 905 bp and encoded a predicted full-length translation product of 175 amino acids. This

protein was most similar (77 % identity, 82 % similarity) to the probable NADH:ubiquinone oxidoreductase 18 kDa subunit mitochondrial precursor protein from *Arabidopsis* (accession number Q9FLX7). A N-terminal transit peptide of 21 amino acids that causes targeting to the mitochondrial matrix was identified in the primary sequence of 1.22. Cleavage of this signal sequence yields a mature protein of 17.5 kDa. Together, these data suggest that 1.22 is the NADH:ubiquinone oxidoreductase 18 kDa subunit of tomato.

1.23 was 711 bp in length and encoded a translation product of 127 amino acids. The closest relative of this protein was the probable *Arabidopsis* NADH:ubiquinone oxidoreductase 17.2 kDa subunit (accession number Q9M9M9), with which it shares 90% sequence identity (93% similarity). The sequence of 1.23 lacked a methionine and based on a sequence alignment with *Arabidopsis* NADH:ubiquinone oxidoreductase it lacked approximately 30 N-terminal amino acids.

6.4 Discussion

6.4.1 Putative host partners of V1

14A was encoded by four separate prey plasmids retrieved from the V1 screen, suggesting that its interaction with V1 was not an artifact of the yeast two-hybrid screen. It shared a high degree of sequence identify with Grx which, along with thioredoxin, is the major reducing molecule in most organisms and is thus involved in many cellular processes (Rouhier et al., 2004). Grxs generally belong to multigenic families of proteins represented by various isoforms. *Arabidopsis* contains at least 31 *Grx* genes divided into three classes based on the sequence of the active site. 14A appears to belong to the third class of Grxs with an active site CCx[C/S/G], which is the

largest class in *Arabidopsis* but forms a minor group in rice, wheat, and maize (Rouhier et al., 2004).

The thiol groups in the active site of Grxs undergo reversible oxidation and reduction and reduce target proteins by dithiol-sulphide exchange and also catalyse protein-glutathione mixed disulphide reduction (deglutathiolation) (Fomenko and Gladyshev, 2002). A number of plant proteins have been identified as Grx targets (Rouhier et al., 2004). In organisms other than plants, they play an important role in many and diverse processes including signal transduction pathways, by regulating transcription factors, kinases, and phosphatases, in stress reponses by regulating antioxidative enzymes, and in cytoskeleton organisation (Rouhier et al., 2004).

A number of roles for the putative V1/Grx interaction in TLCV infection can be envisioned, all of which may relate to function of V1 in cell-to-cell movement (Section 1.1.9.5 and Chapter 4). The most likely scenario is that Grx is required to maintain V1 at the correct redox state. The V1 protein contains six cysteine residues and the DISULFIND program predicts that these may form two disulphide bridges (Vullo and Frasconi, 2004). Formation of these bridges by GRx-mediated reduction may be necessary for correct protein folding, and hence function, of V1. Supporting this idea, a Grx encoded by Vaccinia virus, G4L, is required for virion morphogenesis and it was suggested that it may mediate disulphide bond formation of viral structural proteins (White et al., 2000).

Another possibility is that V1 acts in conjunction with its putative Grx partner in a disulphide bond formation pathway. This hypothesis is based on the recent observation

that the E10R protein of Vaccinia virus can promote disulphide bond formation in G4L which in turn generates disulphide bridges in other viral proteins (White et al., 2002). E10R, a member of the ERV1/ALR protein family which is highly conserved among all poxviruses, is a thiol oxidoreductase that, like Grxs, contains a C-X-X-C motif which is responsible for its enzymatic activity. V1 also contains a C-X-X-C motif (CPHC) from amino acids 86-89 that may be able to promote disulphide bond formation of tomato Grx, although apart from this similarity its sequence is highly divergent to E10R and other members of the ERV1/ALR family.

Finally, we cannot rule out the possibility that by binding Grx, V1 mediates a general disruption of the cellular environment which is advantageous for TLCV infection by changing the redox state or altering the Grx-controlled function of host proteins. In mammals, many viruses, including human immunodefiency virus and hepatitis C, induce a shift towards a pro-oxidant state which contributes to their pathogenicity (Herzenberg et al., 1997; Gong et al., 2001). Further, antioxidants can inhibit the replication of many different viruses (references in Nencioni et al., 2003).

6.4.2 Putative host partners of C1

The three proteins which putatively interact with $Rep_{184-362}$, 1A, 14A and 15A, share a high degree of sequence similarity and one was retrieved from two separate prey plasmids, suggesting that these interactions are not artifactual and may be important in Rep function. They are most closely related the *Antirrhinum majus* and *Arabidopsis* DAG proteins (Chatterjee et al., 1996) and *Arabidopsis* DAL1 (AtDAL1; Bisanz et al., 2003). Both *DAG* and *Dal1* are nuclear genes encoding proteins which contain transit peptides at their N-termini to direct them to chloroplasts. Similarly, chloroplast transit

peptides (CTp) are predicted for 1A and 15A (Emanuelsson et al., 1999). However, 14A, the C-terminal 213 amino acids of which are identical to the the C-terminus of 15A, contains a divergent N-terminus which is not predicted to contain a CTp.

DAG and *Dal1* are both involved in chloroplast differentiation (Chatterjee et al., 1996; Bisanz et al., 2003). Transposon-tagged mutants have white/yellow leaf tissue because chloroplasts fail to develop from proplastids. These mutants are variegated where cells revert back to a wild-type state. Stable mutations of either gene have not been reported, probably because they are lethal. Despite the phenotypic similarities between the *dag* and *dal1* mutants, biochemical analyses indicate that these genes possess different properties. *DAG* acts very early in chloroplast maturation because it is required for expression of plastidial encoded bacterial-like RNA polymerase (PEP) gene which is involved in transcribing the plastid-encoded genes required for photosynthetic light harvesting in functional chloroplasts (Chatterjee et al., 1996). However, PEP is present and active in *dal1* mutants (Bisanz et al., 2003). In addition, in *dal1* mutants proplastids develop into etioplasts and cells contain an increased number of plastids, while in *dag* plants plastids are blocked at the proplastid stage and a slightly smaller number of plastids is present in cells compared to wild-type plants. Hence, these genes are not orthologues.

The exact nature of the DAG protein and its role in controlling plastid development is unknown. In contrast, it has been suggested that AtDAL1 might be a novel protein necessary for the correct assembly of ribosomal subunits in the chloroplast, since the secondary processing of the rRNA operon giving rise to mature rRNA species is altered in the *dal* mutant (Bisanz et al., 2003).

Based on amino acid similarities, 1A and 15A may be the tomato orthologues of DAG and AtDAL1 respectively. The lack of a chloroplast signal peptide in 14A suggests that it possesses another function. The possible role/s of the putative Rep/DAG and Rep/DAL interaction in TLCV infection is difficult to predict. Since TLCV infection, or transient expression of $Rep_{184-362}$, does not cause bleaching of plant tissue (Dry et al., 1993; Selth et al., 2004), it seems probable that Rep does not completely interfere with the function of DAG or DAL. However, ultrastructural analyses of cells infected with TLCV or expressing $Rep_{184-362}$ have not been carried out, and therefore it remains to be determined whether chloroplast development is affected by Rep production. More work, including verification of the Rep/DAG and Rep/DAL interactions by independent means, needs to be carried out to further analyse the potential role of these proteins in Rep function and TLCV infection.

6.4.3 Putative host partners of C2

In this study, C2 was identified as a transcriptional activator in yeast, supporting the idea that it activates expression of the virion-sense genes in begomoviral infections (Section 1.1.9.2). To carry out a yeast two-hybrid screen using the largest segment of the C2 protein possible, we generated a series of truncations of C2 and were able to map the TAD to the acidic carboxy-terminal 30 amino acids. This result supports earlier work by Hartitz *et al.* (1999), who showed that the AL2 protein (C2 homologue) from *Tomato golden mosaic virus* (TGMV) can bind zinc and possesses an acidic C-terminus that functions as a TAD in mammalian and yeast cells.

$C2_{1-103}$ was found to interact with a protein similar to UBCs from soybean, *Arabidopsis*, yeast and mammals. The βC1 protein from the satellite DNA associated with CLCuV

infections also interacted with a functional UBC (Chapter 5), suggesting that interaction with the plant ubiquitination system may be a common strategy that viruses utilise during infection. In addition, the finding that a geminiviral C2 and βC1 protein both putatively interact with tomato UBCs implies that these proteins may be functionally similar. Supporting this idea, TLCV C2 is a suppressor of PTGS (Selth et al., 2004) and βC1 can also suppress PTGS when transiently expressed in *N. benthamiana* (S. Mansoor, personal communication). This raises the intriguing idea that the means by which these proteins mediate suppression of PTGS may be via their interaction with the host cell ubiquitination system. Such an explanation could be tested by mutating the UBC binding sites of C2 and βC1 and analysing the ability of these proteins to suppress PTGS. These experiments are attractive since only two silencing suppressors have been characterised in any detail and in general very little is known about the molecular mechanism of silencing suppression (Silhavy and Burgyan, 2004).

Alternatively, the role of the putative C2/UBC and βC1/UBC interactions may be unrelated and accomplish entirely different outcomes. As described for βC1 (Section 5.4), UBC binding may regulate turn-over of C2, alter C2 function, or mediate the selective degradation of undesirable host proteins. This latter hypothesis, in which C2 would act as a ubiquitin ligase and bridge UBC to a specific cellular protein, is plausible because C2 contains a zinc finger, which is structurally similar to the RING finger found in many ubiquitin ligases (Schwartz and Hochstrasser, 2003). Zinc fingers consist of two pairs of zinc ligands coordinately binding one zinc ion, whereas RING fingers consist of four pairs of ligands binding two ions (Kosarev et al., 2002). The zinc finger of C2, found in a central region of the protein and comprising conserved cysteine and histidine residues (Section 6.3.3), can bind zinc cations and this is important for its

sequence non-specific DNA binding activity (Hartitz et al., 1999). However, zinc fingers are also often involved in protein/protein interactions, and RING fingers appear to mediate interactions with other proteins and in particular UBCs (Freemont, 2000). Despite the differences between the structure of the C2 zinc finger and the consensus RING finger motif, the possibility that C2 acts as a ubiquitin ligase in concert with a UBC to degrade other cellular proteins cannot be ruled out.

The other putative $C2_{1-103}$ interactor, designated 13A, resembles the *Arabidopsis* CAO protein. CAO is part of the chloroplast signal recognition particle (cpSRP) that directs nuclear proteins containing an N-terminal chloroplastic transit peptide from the stroma to the thylakoid membrane (Klimyuk et al., 1999). According to the TargetP algorithm, the $C2_{1-103}$ sequence may contain a chloroplastic signal peptide from amino acids 1-7 (Emanuelsson et al., 1999; Nielsen et al., 1997) to direct it to the stroma where, upon binding CAO, it could be transported into the thylakoid membrane. However, this explanation does not agree with the finding that a C2-GFP fusion protein localises exclusively to the nucleus of onion cells (Selth, 2000) and contains a putative nuclear localisation signal, RRKR, from amino acids 29-32. Further, localisation of C2 to the nucleus is appropriate given its proposed function in transactivation of virion-sense gene expression and suppression of post-transcriptional gene silencing. If C2 does localise to the nucleus in host plants during TLCV infection it is unlikely that an opportunity exists for $C2_{1-103}$/CAO binding. Therefore, the putative $C2_{1-103}$/CAO interaction may be an artifact of the yeast two-hybrid system, possibly due to the unnatural, enforced nuclear localisation of CAO that occurs when it is fused to the SV40 large T antigen NLS located within the B42 sequence.

6.4.4 Putative host partners of C4

The systemic movement of bipartite begomoviruses requires two genes found on the B component (Lazarowitz and Beachy, 1999). *BV1* encodes the nuclear shuttle protein (NSP), while *BC1* encodes the movement protein (MP). NSP facilitates the transport of viral DNA from the nucleus to the cytoplasm where it interacts with the MP to promote cell-to-cell spread of viral DNA (Sanderfoot and Lazarowitz, 1995; Mariano et al., 2004).

Recently, a C-terminal truncation of a RLK from tomato was shown to interact in yeast and *in vitro* with the NSPs from TGMV and *Tomato crinkle leaf yellows virus* (Mariano et al., 2004). This protein was designated LeNIK, for *Lycopersicon esculentum* NSP-interacting kinase. The authors carried out another two hybrid screen of a soybean cDNA library with TGMV NSP and isolated a different RLK, designated GmNIK, suggesting that the NSP:NIK interaction is functionally significant. The full-length GmNIK sequence possesses an internal transmembrane helix and a signal peptide that could target it to the secretory apparatus. Its sequence is most closely related to a group of putative leucine-rich repeat (LRR) receptor serine/threonine protein kinases (LRR RLKs) and is arranged into a modular organisation that resembles a subclass of plant resistance genes, such as *Xa-21* from rice which confers resistance to a bacterial pathogen (Song et al., 1995). Thus, the authors speculated that *NIK* may function as the R gene and *NSP* as the avirulence (*Avr*) gene in induction of resistance to the geminivirus, as predicted in the elicitor-receptor model.

During the preparation of this thesis, Fontes *et al.* (2004) published a report characterising the role of NIK in NSP function. They took advantage of the capacity of

CaLCuV to infect *Arabidopsis*, which allowed reverse genetics studies to be done on *Arabidopsis* NSP-interacting LRR-RLKs. The *Arabidopsis* NIKs are plasma membrane-localised *bona fide* protein kinases with biochemical properties of signaling receptors. NSP is not phosphorylated by NIK, but rather inhibits NIK kinase activity. Furthermore, mutations in *NIK* genes enhanced susceptibility to CaLCuV infection. Together, these results suggest that NIKs are involved in anti-geminiviral defenses and that NSP may repress this mechanism by inhibiting NIK kinase activity.

A yeast two-hybrid screen using C4 as bait isolated three putative serine/threonine protein kinases, 1A, 5A and 26A. Such duplication suggests that these interactions are not artifacts and may be functionally significant *in planta*. The closest relative of 1A is a soybean RLK. This raised the intriguing possibility that C4 may act analagously to NSP and inhibit a RLK-mediated antiviral response. Supporting this idea, 1A and the tomato, soybean and *Arabidopsis* NIKs share sequence identity ranging from 38.1 to 39.8% (44.9-49.6% similarity) and, like the NIKs, 1A is predicted to possess a transmembrane helix (Tusnady and Simon, 1998). The idea of a functional link between C4 and NSP is further supported by a recent observation suggesting that TLCV C4, like NSP, may possess a role in cell-to-cell movement. *In situ* hybridisation analysis of a TLCV *C4* mutant indicates that it is partially deficient in cell-to-cell movement functions (M. S. Raisheed, L. A. Selth, A. M. G. Koltunow, J. W. Randles and M. A. Rezaian, submitted). To test whether the C4/1A and NSP/NIK interactions serve the same function, the ability of TGMV NSP to bind 1A was analysed in yeast. In our system, using *lacZ* as a reporter (Section 3.2.2), binding between these proteins was not detected (results not shown). Despite this negative result, the possibility that C4 is functionally analogous to NSP and acts to inhibit a RLK-mediated antiviral response

cannot be ruled out. 1A is truncated at the N-terminus, possibly by as much as 660 amino acids, and some or all of this region may be involved in an interaction with NSP. Experiments are in progress to determine the full-length sequence of 1A for future binding experiments, which will also reveal whether it contains a signal peptide directing it to the secretory apparatus or a LRR which may group it into the LRR RLK family of resistance proteins. Alternatively, the putative C4/RLK interaction may have some other function. For example, C4 may be regulated by phosphorylation, an idea that is currently under investigation. Alternatively, C4 has been shown to localise to the cell periphery (Raisheed, M. S., Selth, L. A., Koltunow, A. G., and Rezaian, M. A., submitted), and RLK could be involved in "piggy-backing" C4 to this region.

The C4 screen retrieved two other peptides, 5A and 26A, both of which are similar to plant proteins, termed SKs, that belong to the glycogen synthase kinase 3 (GSK-3)/SHAGGY kinase family. 5A and 26A share 79% sequence identity, indicating that they may possess similar functions. Plant SKs are encoded by a multigene family and possess roles in diverse biological processes including hormone signaling, development and stress responses (Jonak and Hirt, 2002). The closest relative of 5A, AtSKη, is a negative regulator of brassinosteroid signaling (Perez-Perez et al., 2002; Li and Nam, 2002), a pathway involved in regulating cell expansion, vascular differentiation, etiolation, reproductive development and stress responses (Clouse and Sasse, 1998). Recently, AtSKθ was implicated in the establishment or maintenance of a vegetative state during male sporogenesis (Tavares et al., 2002). In addition, a mammalian homologue of the AtSKs, GSK-3, represses the progression of G1-S phase progression by phosphorylating cyclin D1 which promotes its nuclear export and subsequent degradation (Diehl et al., 1998). Thus, some SKs also possess a role in cell cycle

control. The *in planta* function of the putative tomato SKs 5A and 26A is unknown. However, a role(s) in regulation of brassinosteroid signalling, development, stress responses, and/or cell cycle control would provide a link to the pathogenicity of C4, since disruption of any of these processes by C4 binding could result in the symptoms associated with C4 expression (Krake et al., 1998). Supporting this idea, AtSKη mutants exhibit leaf curling symptoms that resemble those seen in C4 transgenic plants (Perez-Perez et al., 2002). Alternatively, the enations observed in severely affected C4 transgenic lines could be produced via a disruption of cell-cycle control mediated through C4/SK interaction. In summary, it is conceivable that the mechanism of C4-mediated pathogenesis relies on an interaction with a host SK, but gaining a more precise understanding of the role of SK in C4 function requires further experimentation. Towards this end, the effect of 5A over-expression on TLCV replication has recently been tested, and it appears to enhance accumulation of TLCV ssDNA (S. Dogra, personal communication). The mechanism of this enhancement is currently under investigation.

Other proteins retrieved from the C4 yeast two-hybrid screen included a putative isovaleryl-CoA dehydrogenase, nucleoporin protein, heat shock protein, and two tomato proteins of unknown function, one of which may contain an outer membrane protein domain. The roles, if any, of these proteins in geminivirus infection cannot be predicted at this time. It is sufficient to note that the function of C4 is likely to be complex and may involve protein/protein interactions with many and diverse host proteins.

6.4.5 Putative host partners of βC1

A screen with the βC1 protein retrieved six different putatively interacting proteins, one of which (SlUBC) is described in Chapter 5. Of the others, the peptide sequence designated 1 contains the recently-identified GARP domain (Riechmann et al., 2000) which acts as a sequence-specific DNA binding domain (Sakai et al., 2000). A number of GARP-containing proteins are thought to act as transcription factors (Hosoda et al., 2002). The closest relative of 1, the *Arabidopsis* KANADI protein, is involved in polarity establishment of plant lateral organs (Kerstetter et al., 2001; Eshed et al., 2001). Expression of βC1 in plants causes severe developmental abnormalities (Saeed et al., 2004). Taken together, these findings suggest that the βC1-induced phenotype may be a consequence of its binding to, and subsequently disrupting the function of, 1.

Two of the putative binding partners of βC1, termed 22, were identical and are probably the 18 kDa subunit of complex I, while another interactor, 23, is likely to be the 17.2 kDa subunit of this large multisubunit complex. Complex I is a proton-pumping, NADH:ubiquinone oxidoreductase that oxidises NADH in the electron transport pathway. Could βC1 have a role in disrupting mitochondrial energy production? This hypothesis is supported by the observation that the yeast two-hybrid screen retrieved two putative members of the large complex I polypeptide which do not share any sequence similarity. Furthermore, some mammalian viral proteins inhibit components of the mitochondrial respiratory chain (Koundouris et al., 2000; Chien and Kuo, 2001). It is thought that mitochondria may be a critical target during virus infection because of their central role in regulation of cell death and survival (Kroemer et al., 1998). However, disruption of mitochondrial function in virus infection is generally associated

with cytopathic effects including apoptosis and necrosis (Koundouris et al., 2000), symptoms which have not been observed in response to βC1 expression.

Five of the putative βC1 interactors were identical and likely represent the tomato 60S acidic ribosomal P3 protein. P3 is a plant-specific phophorylated protein associated with the large ribosomal subunit (Szick et al., 1998). Together with P0, P1, and P2 it forms a stalk structure that is present in the active site of the ribosome where interactions between mRNA, tRNA, and translation factors occur during the late initiation, elongation, and termination phases of translation (Szick et al., 1998). The functional significance of a βC1/ribosome interaction is unclear, although it implies that βC1 may regulate, or disrupt, translational processes.

Three of the putative βC1 interactors, grouped as 6, were identical and share significant sequence identity with various plant urate oxidase (uricase) proteins. Legume plants form root nodules in which symbiotic rhizobia are able to fix atmospheric nitrogen (Takane et al., 2000). Uricases are involved in assimilating fixed nitrogen into ureides by catalysing the oxidation of uric acid to allantoin (Takane et al., 2000). However, the expression profiles of a number of uricase genes, as well as the identification of uricase genes in non-legumes, suggest that uricase plays a more common role in plant cellular processes, for example, reutilisation of nitrogen in nucleic acids (Takane et al., 2000). At this time, it is difficult to envision the function of this putative interaction in geminivirus infections.

In summary, βC1 may interact with multiple host proteins involved in such diverse metabolic processes as mitochondrial ATP production, nitrogen scavenging, protein

translation (this Chapter) and protein degradation (Chapter 5). Any of these interactions could severely impact on normal cellular conditions and cause the symptoms associated with βC1 expression. Alternatively, the mechanism of βC1 pathogenicity may be via its putative interaction with a new tomato GARP transcription factor, which could function in regulating plant developmental processes. To determine which, if any, of these host factors function in βC1 pathogenesis, their ability to interact with diverse βC1 proteins should be tested.

Chapter 7 - Interactions between *Tomato leaf curl virus*-encoded proteins

7.1 Introduction

Transcription, replication, and movement processes of geminiviruses often require the combined action of multiple virus-encoded proteins. These collaborative efforts are frequently mediated by physical interactions between the proteins involved. For example, the Rep and REn proteins form heteromers in yeast and *in vitro*, and these complexes may be required for efficient geminiviral replication (Settlage et al., 1996). Binding also occurs between the replication proteins of mastreviruses, Rep and RepA, and this appears to be important for coat protein (CP) gene expression and/or targeting of RepA to the nucleus (Boulton, 2002). The movement of the bipartite begomovirus SqLCV from an initial infected cell to neighbouring cells requires an interaction between its nuclear shuttle protein (NSP) and movement protein (MP) (Sanderfoot and Lazarowitz, 1995). NSP binds newly synthesised viral genomic ssDNA and shuttles it from the nucleus to the cytoplasm (Pascal et al., 1994), where the MP traps the NSP:genome complexes and directs them to and across the plant cell wall (Sanderfoot and Lazarowitz, 1995). An interaction also occurs between the MSV CP and movement protein (Liu et al., 2001), which probably facilitates cell-to-cell movement in a mechanism analogous to that which occurs in SqLCV infections.

In addition, the function of some geminiviral proteins requires their self-association. Various geminiviral Rep proteins form oligomers, a process which may be important for

enzymatic functions and/or DNA binding (Settlage et al., 1996; Castellano et al., 1999). TGMV REn can dimerise *in vitro* (Settlage et al., 1996), although the functional significance of this observation is unclear. Geminiviral CPs also undergo homotypic interactions to produce the virus capsid, which consists of 22 pentameric capsomers and 110 identical protein subunits (Zhang et al., 2001; Hallan and Gafni, 2001).

The aim of the work described in this chapter was to test for interactions occurring between all of the proteins encoded by TLCV. To achieve this, we utilised the LexA yeast two-hybrid system described in previous chapters.

7.2 Materials and methods

7.2.1 Construction of yeast vectors

The construction of pLexA vectors containing *V1*, *C1*, $C1_{184\text{-}362}$, $C2_{1\text{-}103}$, *Ren*, and *C4* sequences fused in-frame to the LexA DNA-BD is described in Chapters 3, 4 and 6. The *V2* and $C1_{1\text{-}184}$ ORFs were amplified with primer pairs V2-EcoRI-F/V2-XhoI-R and C1(1-184)-F-EcoRI/C1(1-184)-R-XhoI respectively, digested with EcoRI/XhoI, and ligated into EcoRI/XhoI digested pLexA.

The *V1*, *V2*, $C1_{1\text{-}184}$, *REn* and *C4* sequences were released from pLexA by digestion with EcoRI/XhoI, separated from the vector by electrophoresis, and eluted from gel fragments. These sequences were ligated into EcoRI/XhoI digested pB42AD.

To assist cloning of the *C1*, $C1_{184\text{-}362}$, and *C2* sequences into pB42AD, a polylinker was inserted into the pB42AD vector. Oligonucleotides pB42-linker-F and pB42-linker-R were diluted to 1μg/ul, combined at a ratio of 1:1, heated to 95°C for 5 min, and

annealed at room temperature for 1 h. This DNA fragment was then ligated into EcoRI/XhoI digested pB42AD to yield pB42ADL. The original polylinker of pB42AD and the modified polylinker of pB42ADL are shown in Figure 7.1. The *C1* and $C1_{184\text{-}362}$ sequences were released from pLexA by digestion with BamHI, purified, and ligated into pB42ADL which had been cut with BglII and dephosphorylated. *C2* was amplified with primers C2-F-BamHI and C2-R-BamHI, digested with BamHI, and ligated into the same vector. Clones containing inserts in the correct orientation were identified by digestion with specific restriction enzymes.

7.2.2 Analysis of LexA and B42 fusion protein production in yeast

Production of fusion proteins in yeast was monitored as described in Section 2.2.19.

7.2.3 Interaction between TLCV-encoded proteins in yeast

Yeast strain displayYEAST-L was transformed sequentially with displayREPORTER, a pLexA vector, and a pB42AD vector. Colonies were combined and streaked onto plates lacking the amino acids uracil, histidine, tryptophan and leucine, but containing galactose and raffinose to induce the GAL1 promoter. After three days, yeast growth and GFP fluorescence was assessed.

7.3 Results

7.3.1 LexA and B42 fusion protein production in yeast

Immunoblots of LexA and B42 peptides fused to each of the viral proteins are shown in Figure 7.2A and 7.2B respectively. The V1, $Rep_{1\text{-}184}$, $Rep_{184\text{-}362}$, C2, REn and C4 proteins fused to either LexA or B42 were all detectable, but the presence of LexA and B42 fusions to both V2 (CP) and Rep could not be visualised by this method.

pB42AD

5′-CCC**GAATTC**GGCCGACTCGAGAAG-3′
EcoRI XhoI

pB42ADL

5′-CCC**GAATTC**GAGATCTGGCCCGGGCCATGGGGTACCCTCGAGAAG-3′
EcoRI BglII SmaI NcoI KpnI XhoI

Figure 7.1. Multiple cloning sites of pB42AD and pB42ADL.

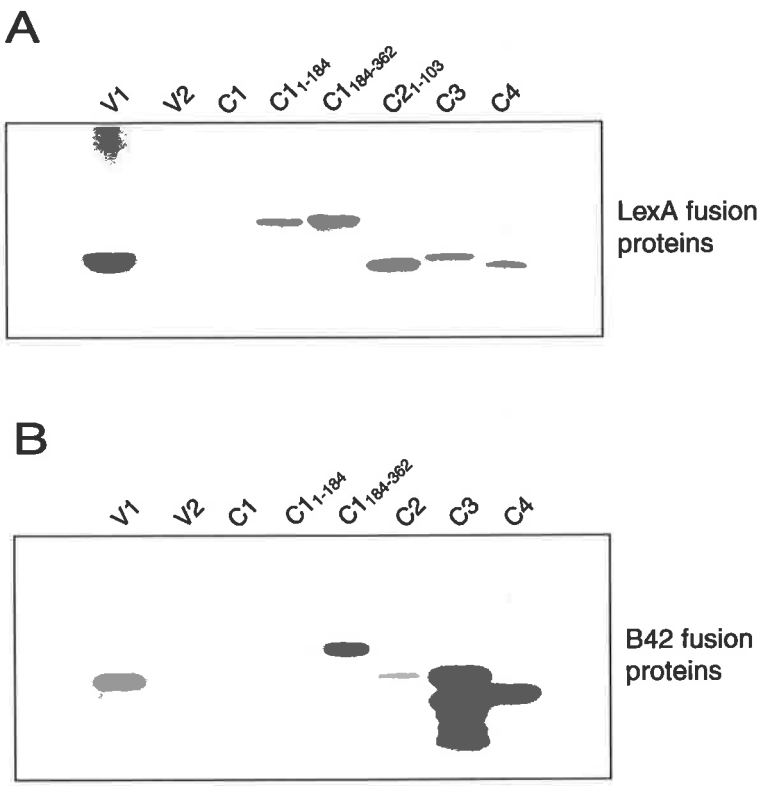


Figure 7.2. Immunoblotting of LexA and B42 fusion proteins in yeast. Total protein from yeast cultures expressing fusion proteins was extracted, fractionated on 4-20% SDS-polyacrylamide gels, and immunoblotted with anti-LexA **(A)** to detect LexA fusion proteins or anti-HA **(B)** to detect B42 fusion proteins.

7.3.2 Interaction between TLCV proteins in yeast

The interaction array of TLCV proteins in yeast is shown in Table 7.1. This assay procedure is semi-quantitative and interaction specificities were assigned values of + or ++ based on a visual assessment of yeast prototrophic growth and the intensity of GFP fluorescence.

A strong interaction was observed in cells expressing Rep fusions to the DNA-BD and AD. To determine the domains of Rep responsible for its self-association, we tested N-terminal (Rep-N, amino acids 1-184) and C-terminal (Rep-C, amino acids 184-362) fragments of Rep. Rep-N interacted with itself and, more weakly, with full-length Rep, while Rep-C was not observed to interact with itself, full-length Rep of Rep-N. These results suggested that the N-terminal sequence of Rep was responsible for its homodimerisation.

Strong homotypic interactions were also observed for the C2 and V1 proteins. The dimerisation domains of C2 and V1 were not mapped. However, the LexA-C2 fusion protein consisted of just the N-terminal 103 amino acids of C2, suggesting that the region responsible for C2 dimerisation lies somewhere within this sequence and is therefore distinct from its activation domain, which is found within the C-terminal 33 amino acid residues (Section 6.3.3).

7.4 Discussion

This chapter describes an analysis of the interactions that may occur between the individual TLCV-encoded proteins. It is important to note, however, that this yeast system has some inherent weaknesses which may have resulted in false negative results.

Table 7.1 Interactions between TLCV-encoded proteins

		LexA fusion proteins							
		Rep	Rep_{1-184}	$Rep_{184-362}$	$C2_{1-103}$	REn	C4	V1	V2
B42 fusion proteins	Rep	++[a]	-	-	-	-	-	-	-
	Rep_{1-184}	-	+++	-	-	-	-	-	-
	$Rep_{184-362}$	-	-	-	-	-	-	-	-
	C2	-	-	-	++	-	-	-	-
	REn	-	-	-	-	-	-	-	-
	C4	-	-	-	-	-	-	-	-
	V1	-	-	-	-	-	-	+++	-
	V2	-	-	-	-	-	-	-	-

[a]Interaction was indicated by the ability of cells transformed with pLexA, pB42AD and pGNG1 plasmids to grow on medium lacking leucine. As an additional indicator of interaction, colonies were monitored for *GFP* expression by visualisation under ultraviolet light.

First, Rep and CP were either absent or below detectable levels when expressed as LexA and B42 fusions in yeast. It is probable, therefore, that any interactions involving these proteins would have gone undetected, although in at least one instance (Rep dimerisation - see below) this is not the case. Second, despite being a eukaryotic system, it is possible that post-translational modifications which occur *in planta* and are required for protein:protein interactions may not occur in yeast. It is therefore possible that interactions between the TLCV proteins that play an important role *in planta* were not identified in this study.

The toxicity of Rep is well documented (Sections 1.1.9.1.5 and 6.3.2), and therefore it is not surprising that its production was repressed in yeast cells. Despite our inability to detect LexA-Rep and B42-Rep fusion proteins, co-transformation of yeast cells with pLexA-Rep and pB42AD-Rep vectors resulted in activation of reporter gene expression. This strongly suggests that the Rep fusion proteins were being produced, but at levels below the detection cut-off of the immunoblot. Further, this implies that very few reconstituted LexA/B42 complexes are required to activate expression of *Leu* and *GFP* such that leucine autotrophic growth is possible and fluoresecence is visible under UV light, highlighting the sensitivity of this yeast two-hybrid system. Using truncations of the *C1* gene, we mapped the oligomerisation domain of TLCV Rep to its N-terminal 1-184 amino acids. A dimerisation activity has also been observed for the Rep proteins from TGMV and BGMV (Settlage et al., 1996), and Orozco *et al.* (1997) later mapped the TGMV Rep dimerisation domain from amino acids 121-181. Together with our results, this suggests that Rep oligomerisation occurs by a domain in its N-terminus that is evolutionarily conserved amongst geminiviruses.

Several mechanisms whereby Rep oligomerisation could contribute to its replication and transcription activities have been proposed. During Rep-mediated DNA cleavage, a tyrosine residue in the active site is covalently cross-linked to the 5'-end of the cleaved DNA (Laufs et al., 1995b). Thus, it is not available for the second cleavage event that resolves the DNA concatemer. Rep dimerisation may provide the second tyrosine required for termination of rolling-circle replication (Settlage et al., 1996). In addition, oligomerisation appears to be required for the DNA binding activity of Rep. For example, Orozco and Hanley-Bowdoin (1998) showed that the DNA-BD of TGMV Rep is functional only when fused to the protein-interaction domain.

To my knowledge, this study is the first to identify a dimerisation activity for a geminiviral C2 protein. A number of mechanisms whereby the putative self-association of C2 may be required for its function can be envisioned. The C2 proteins from ACMV, TYLCV and TLCV C2 have recently been identified as suppressors of PTGS (Section 1.1.9.2). One possibility is that C2 dimerisation is involved in this function. Such an idea is not unprecendented: recently, Vargason *et al.* (2003) resolved the crystal structure of the p19 PTGS-suppressor protein from *Carnation Italian ringspot virus* bound to an siRNA. They found that p19 dimerises and forms a molecular caliper which binds siRNAs based on the length of their duplex region. This prevents incorporation of siRNAs into the RNA-induced silencing complex which mediates targeted degradation of homologous RNA. At this stage the mechanism by which C2 suppresses PTGS is unknown, but the finding that it may dimerise suggests that it could act analogously to p19. Second, the putative self-association of C2 may be involved in its DNA binding activity (Section 1.1.9.2). Many plant transcription factors, including members of the basic/helix-loop-helix family, bZIP family, and MYB family, assemble

as homodimers to bind DNA (Toledo-Ortiz et al., 2003; Jakoby et al., 2002; Martin and Paz-Ares, 1997). Finally, the possibility that the putative homodimerisation of C2 plays a role in its interaction with other host factors, identified in Section 6.3.3, cannot be ruled out. A number of experiments to further characterise the putative homodimerisation activity of C2 should be performed. *In vitro* binding assays using recombinant C2 protein are required to verify this activity. If these support the yeast two-hybrid findings, it would be interesting to delineate more closely the domain in C2 responsible for self-association and generate mutants lacking this activity to assess the effects, if any, this has on its other functions.

A homotypic interaction was also observed for V1, a biochemical activity that has not been reported for other geminiviral V1 proteins. The function of V1 is largely unknown, and thus any discussion on its putative self-association would involve much unwarranted speculation. As suggested for C2, V1 dimerisation should be verified using recombinant V1 protein in *in vitro* binding assays. Confirmation of this activity would allow better planning of future experiments designed to determine the function of V1.

Settlage *et al.* (1996) have previously reported the virus non-specific Rep/REn and REn/REn interactions that occur between TGMV and BGMV replication proteins. However, such interactions were not observed between the TLCV Rep and REn proteins. There are a number of possible explanations for this apparent discrepancy. The most obvious is that TLCV Rep and REn proteins may be functionally different to their homologues from TGMV and BGMV, and therefore Rep/REn and REn/REn interactions may not occur in TLCV infections. Alternatively, some weakness of the

yeast two-hybrid system used in this study, discussed earlier, may have generated a false negative result. For example, post-translational processing of REn may not occur properly in yeast. Another possibility is that the very low expression levels of Rep may not be sufficient to produce a detectable interaction with REn in yeast. This latter explanation is unlikely, however, considering that neither Rep-N, which contains the REn binding domain in TGMV and BGMV Rep proteins, nor Rep-C were found to interact with REn. To further explore the possibility of REn/Rep and REn/REn binding, it would be useful to carry out *in vitro* binding assays using currently available recombinant REn and Rep proteins purified from *E. coli*.

The TYLCV CP is able to self-associate in yeast, an activity mediated by a region around amino acids 129-152 (Hallan and Gafni, 2001). In this study, we were unable to detect such a dimerisation activity for TLCV CP. This could represent a false negative result, since B42- and LexA-V2 fusion proteins could not be detected in yeast and therefore an interaction between these proteins may not be demonstrable in this system. Analysis of the vectors by sequencing revealed that the integrity of the CP coding sequence was intact. This suggests that the TLCV CP may be toxic in yeast. Such an explanation does not agree with results of previous work: in addition to the study mentioned above, two other geminiviral CPs have been expressed in yeast without documented problems associated with toxicity (Morin et al., 2000; McGarry et al., 2003). However, supporting the idea of TLCV CP-mediated toxicity, previous attempts to express a glutathione S-transferase-tagged CP (M. A. Rezaian, personal communication) or 6xHis-tagged CP (M. S. Raisheed, personal communication) in bacteria have been unsuccessful. Elucidating the mechanism by which the TLCV CP

mediates its putative toxicity in yeast and bacterial cells would be a useful step in further characterising its function.

Chapter 8 - General discussion and concluding remarks

Geminiviruses are plant pathogens of increasing economic impact. Members of this family which infect the cultivated tomato are of particular importance, causing significant crop losses in horticultural regions throughout the world (Nakhla and Maxwell, 1998). A cogent example of the virulence of tomato-infecting geminiviruses is provided by the Australian isolate of TLCV, which has reduced fruit yields by up to 80-100% in northern areas of Australia. The progressive spread of this pathogen towards the major tomato-growing areas of Australia represents a significant economic threat.

The strategies used to control diseases caused by plant viruses are summarised in Figure 8.1 (Hull, 2002). To date, none of these have been effective in countering TLCV infection. Removal of sources of TLCV infection is difficult since many weeds are hosts of the virus and its whitefly vector is found ubiqitously throughout coastal Australia (Stonor et al., 2003). Control of *B. tabaci* by chemical or non-chemical means is impractical, and most hosts of TLCV are also susceptible to whitefly feeding (Stonor et al., 2003). Traditional breeding programs to achieve conventional resistance against TLCV have encountered problems including low tomato yield, recessive resistance genes, and resistance-breaking virus isolates (M. A. Rezaian, personal communication). Similarly, pathogen-derived transgenic resistance strategies to control this pathogen, including expression of viral coat protein and antisense sequences, have not proven effective (Section 1.3.2).

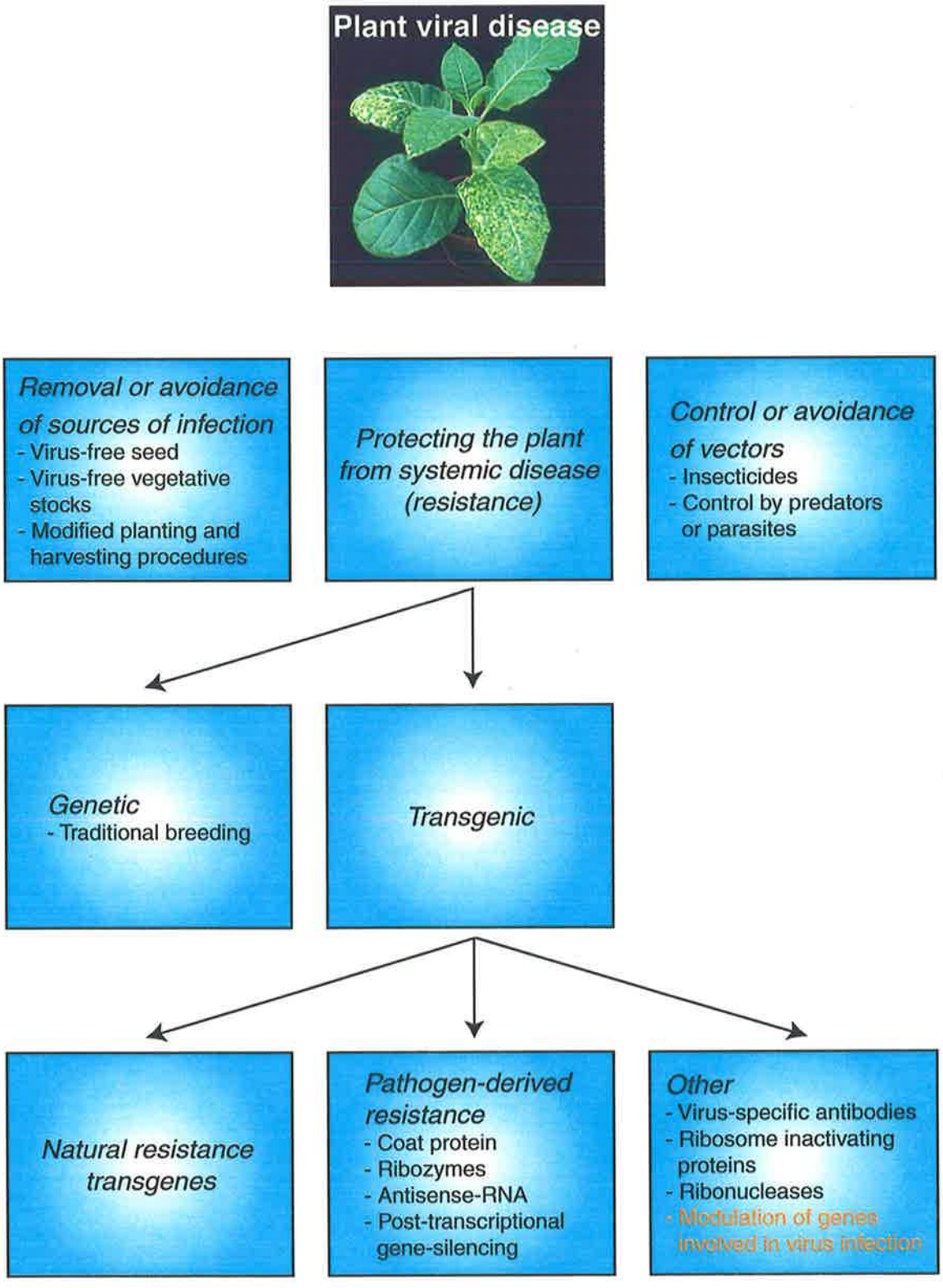


Figure 8.1. Summary of the methods used to control plant viruses (adapted from Hull, 2002).

More recently, other forms of protection using transgenes that are not derived from viruses themselves have been used (Hull, 2002). The basis of this strategy is that, if one understands the molecular interactions involved in the functioning of a pathogen, mechanisms can be devised for interfering with them (Hull, 2002). For example, plants expressing virus-specific antibodies (Tavladoraki et al., 1993), ribosome inactivating proteins (Tumer et al., 1999) and ribonuclease (Watanabe et al., 1995) exhibited resistance to plant viruses. With this in mind, the broad aims of this study were to identify and characterise host factors involved in TLCV infection and develop directed resistance strategies based on this information. For example, resistance could be achieved by down-regulating the expression of host genes encoding proteins vital for infection or by over-expressing genes encoding proteins that inhibit infection. Such a strategy, shown in orange in Figure 8.1, represents a novel form of non-pathogen derived transgenic resistance because it does not use exogenous sequences.

Yeast two-hybrid technology was used to search for host factors involved in TLCV infection. The hypothesis underlying this approach was that host binding partners of TLCV proteins are likely to be resistance targets. This was confirmed by the results obtained in this study: two host factors, SlNAC1 and SlUPTG1, were identified which appear to be required for efficient virus infection and hence their down-regulation or knock-out could allow some level of resistance or tolerance to this pathogen.

Geminiviral REn proteins are able to enhance ssDNA accumulation approximately 50-fold. Elucidating the complex mechanism by which REn achieves this effect has proven difficult, despite the recent findings that REn physically interacts with the viral Rep protein and the host proteins pRBR and PCNA. This study has revealed that REn-

mediated enhancement of TLCV replication probably involves its interaction with a new NAC domain transcription factor, SlNAC1 (Chapter 3). Hence, down-regulating or knocking-out *SlNAC1* may provide tolerance to TLCV infection by reducing the titre of viral DNA. To test this novel HDR strategy, transgenic tomato lines constitutively expressing a *SlNAC1*-hairpin construct and partially silenced for *SlNAC1* expression have been generated (data not shown), and are currently undergoing infectivity assays. The observation that *SlNAC1*-silenced lines appear phenotypically normal is encouraging and suggests that, if they exhibit any tolerance to TLCV infection, a similar resistance strategy could be utilised in other commerically-acceptable tomato cultivars.

Silencing *SlNAC1* may prove an effective means to inhibit TLCV replication. Restricting the movement of TLCV is another strategy to achieve resistance to this pathogen. A screen with the virally-encoded V1 protein retrieved a host autocatalytic glycosyltransferase, SlUPTG1, which may function in cell wall biosynthesis (Chapter 4). Recent evidence suggests that V1 has a partially-redundant role in cell-to-cell viral movement (M. S. Raisheed, L. A. Selth, A. M. G. Koltunow, J. W. Randles and M. A. Rezaian, submitted). Consistent with this idea, overexpression of *SlUPTG1* increased the accumulation of TLCV ssDNA species and this response appears to be a consequence of enhanced viral movement. Thus, silencing *SlUPTG1* may produce tolerance to TLCV by inhibiting viral movement, an idea that will be tested using transgenic plants.

Putative host partners were also identified for the TLCV V1, Rep, C2, C4 proteins as well as the βC1 protein from a satellite that is associated with geminiviral infections (Chapters 5 and 6). Although these interactions were not confirmed by independent

means, in many of the cases they could either be related to the function of the viral protein or have been reported to occur in other viral infections, reinforcing their authenticity. Further characterisation of these host factors may reveal other resistance targets.

In this study, putative interactions that occur between TLCV-encoded proteins were identified. The apparent homodimerisation properties of V1 and C2 has provided new insight into their function, while the suggestion that Rep acts as an oligomer, reported in previous studies, was confirmed.

Apart from providing novel opportunities for TLCV control, this study has revealed key cellular processes involved in geminivirus replication and movement. A number of research lines aimed at elucidating these intricate mechanisms have been initiated as a result of this work.

Reference list

Abouzid, A. M., Frischmuth, T., and Jeske, H. (1988). A putative replicative form of the Abutilon mosaic virus (gemini group) in a chromatin-like structure. *Molecular and General Genetics*. 212:252-258.

Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C., and Gruissem, W. (1997). *RRB1* and *RRB2* encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Molecular and Cellular Biology*. 17:5077-5086.

Ahn, J. H., Xu, Y., Jang, W. J., Matunis, M. J., and Hayward, G. S. (2001). Evaluation of interactions of human cytomegalovirus immediate-early IE2 regulatory protein with small ubiquitin-like modifiers and their conjugation enzyme Ubc9. *Journal of Virology*. 75:3859-72.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The Plant Cell* 9:841-857.

Alberter, B., Rezaian, M. A., and Jeske, H. (2005). Replicative intermediates of *Tomato leaf curl virus* and its satellite DNAs. Virology. 331:441-448.

Aldrick, S. (1970-71). *Plant Pathology Annual Report, Northern Territory Administration* (unpublished).

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. 215:403-410.

Baas, P. D. (1987). Mutational analysis of the bacteriophage f174 replication origin. *Journal of Molecular Biology*. 198:51-61.

Baulcombe, D. (2002). RNA silencing. *Current Biology*. 12:R82-R84.

Behjatnia, S. A. A., Dry, I. B., and Rezaian, M. A. (1998). Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic Acids Research*. 26:925-931.

Bendahmane, M. and Gronenborn, B. (1997). Engineering resistance against tomato yellow leaf curl virus (TYLCV) using antisense RNA. *Plant Molecular Biology*. 33:351-357.

Betting J and Seufert W. (1996). A yeast Ubc9 mutant protein with temperature-sensitive in vivo function is subject to conditional proteolysis by a ubiquitin- and proteasome-dependent pathway. *Journal of Biological Chemistry*. 271:25790-25796.

Bisanz, C., Begot, L., Carol, P., Perez, P., Bligny, M., Pesey, H., Gallois, J. L., Lerbs-Mache, S., and Mache, R. (2003). The *Arabidopsis* nuclear DAL gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Molecular Biology*. 51:651-663.

Bocca, S. N., Kissen, R., Rojas-Beltran, J. A., Noel, F., Gebhardt, C., Moreno, S., du Jardin, P., and Tandecarz, J. S. (1999a). Molecular cloning and characterization of the enzyme UDP-glucose: protein transglucosylase from potato. *Plant Physiology and Biochemistry*. 37:809-819.

Bocca, S. N., Rothschild, A., and Tandecarz, J. S. (1999b). Initiation of starch biosynthesis: purification and characterization of UDP-glucose:protein transglucosylase from potato tubers. *Plant Physiology and Biochemistry*. 35:203-210.

Boisnard-Lorig, C., Colon-Carmona, A., Bauch, M., Hodge, S., Doerner, P., Bancharel, E., Dumas, C., Haseloff, J., and Berger, F. (2001). Dynamic analyses of the expression of the HISTONE::YFP fusion protein in arabidopsis show that syncytial endosperm is divided in mitotic domains. *The Plant Cell*. 13:495-509.

Bonifacino, J. S. and Traub, L. M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annual Review of Biochemistry*. 72:395-447.

Boulton, M. (2002). Functions and interactions of mastrevirus gene products. *Physiological and Molecular Plant Pathology*. 60:243-255.

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.

Brandizzi, F., Irons, S. L., Johansen, J., Kotzer, A., Neumann, U. (2004). GFP is the way to glow: bioimaging of the plant endomembrane system. *Journal of Microscopy*. 214:138-58.

Briddon, R. W., Mansoor, S., Bedford, I. D., Pinner, M. S., Saunders, K., Stanley, J., Zafar, Y., Malik, K. A., and Markham, P. G. (2001). Identification of DNA components required for induction of cotton leaf curl disease. *Virology*. 285:234-243.

Briddon, R. W., Bull, S. E., Amin, I., Idris, A. M., Mansoor, S., Bedford, I. D., Dhawan, P., Rishi, N., Siwatch, S. S., Abdel-Salam, A. M., Brown, J. K., Zafar, Y., and Markham, P. G. (2003). Diversity of DNA beta, a satellite molecule associated with some monopartite begomoviruses. *Virology*. 312:106-121.

Briggs, S. D., Xiao, T., Sun, Z. W., Caldwell, J. A., Shabanowitz, J., Hunt, D. F., Allis, C. D., and Strahl, B. D. (2002). Gene silencing: trans-histone regulatory pathway in chromatin. *Nature*. 418:498.

Brummell, D. A., Camirand, A., and Maclachlan, G. A. (1990). Differential distribution of xyloglucan glycosyl transferases in pea Golgi dicytosome and secretory vesicles. *Journal of Cell Science*. 96:705-710.

Brunetti, A., Tavazza, M., Noris, E., Tavazza, R., Caciagli, P., Ancora, G., Crespi, S., and Accotto, G. P. (1997). High expression of truncated viral Rep protein confers resistance to tomato yellow leaf curl virus in transgenic tomato plants. *Molecular Plant-Microbe Interactions*. 10:571-579.

Carvalho, M. F. and Lazarowitz, S. G. (2004). Interaction of the movement protein NSP and the *Arabidopsis* acetyltransferase AtNSI is necessary for Cabbage leaf curl geminivirus infection and pathogenicity. *Journal of Virology*. 78:11161-11171.

Castellano, M. M., Sanz-Burgos, A. P., and Gutierrez, C. (1999). Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology*. 290:639-652.

Castillo, A., Collinet, D., Deret, S., Kashoggi, A., and Bejerano, E. R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology*. 312:381-394.

Castillo, A. G., Kong, L. J., Hanley-Bowdoin, L., and Bejarano, E. R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. *Journal of Virology*. 78:2758-2769.

Chatterjee, M., Sparvoli, S., Edmunds, C., Garosi, P., Findlay, K., and Martin, C. (1996). DAG, a gene required for chloroplast differentiation and palisade development in *Antirrhinum majus*. *EMBO Journal*. 15:4194-4207.

Chellappan, S., Kraus, V. B., Kroger, B., Munger, K., Howley, P. M., Phelps, W. C., and Nevins, J. R. (1992). Adenovirus E1A, simian virus 40 tumour antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proceedings of the National Academy of Sciences USA*. 89:4549-4553.

Chien, L. F. and Kuo, T. T. (2001). Reduction in mitochondrial respiratory capacity in *Saccharomyces cerevisiae* induced by expression of hepatitis B virus surface antigen. *Microbios*. 105:29-41.

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

Clouse, S. D. and Sasse, J. M. (1998). Brassinosteroids: Essential regulators of plant growth and development. *Annual Reviews in Plant Physiology and Plant Molecular Biology*. 49:427-451.

Collin, S., Fernandez-Lobato, M., Gooding, P. S., Mullineaux, P. M., and Fenoll, C. (1996). The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. *Virology*. 219:324-329.

Collinge, M. and Boller, T. (2001). Differential induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by *Phytophthora infestans* and to wounding. *Plant Molecular Biology*. 46:521-529.

Coscoy, L. and Ganem, D. (2000). Kaposi's sarcoma-associated herpesvirus encodes two proteins that block cell surface display of MHC class I chains by enhancing their endocytosis. *Proceedings of the National Academy of Sciences USA*. 97:8051-8056.

Cui, X., Tao, X., Xie, Y., Fauquet. C. M., and Zhou, X. (2004). A DNAβ associated with *Tomato yellow leaf curl China virus* is required for symptom induction. Journal of Virology. 78:13966-13974.

Daidoji, H., Takasaki, Y., and Nakane, P. K. (1992). Proliferating cell nuclear antigen (PCNA/cyclin) in plant proliferating cells – immunohistochemical and quantitative analysis using autoantibody and murine monoclonal antibodies to PCNA. *Cell Biochemistry and Function*. 10:123-132.

Day, A. G., Bejerano, E. R., Buck, K. W., Burrell, M., and Lichtenstein, C. P. (1991). Expression of an antisense viral gene in transgenic tobacco confers resistance to the DNA virus tomato golden mosaic virus. *Proceedings of the National Academy of Sciences USA*. 88:6721-6725.

Delgado, I. J., Wang, Z., de Rocher, A., Keegstra, K., and Raikhel, N. V. (1998). Cloning and characterization of AtRGP1. A reversibly autoglycosylated arabidopsis protein implicated in cell wall biosynthesis. *Plant Physiology*. 116:1339-1350.

Delmer, D. P. (1999). Cellulose biosynthesis: exciting times for a difficult field of study. *Annual Review of Plant Physiology and Plant Molecular Biology*. 50:245-276.

Desbiez, C., David, C., Mettouchi, A., Laufs, J., and Gronenborn, B. (1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proceedings of the National Academy of Sciences USA*. 92:5640-5644.

Dhugga, K. S. and Ray, P. M. (1994). Purification of the reversibly glycosylated polypeptides from pea. Purified polypeptides exhibit the same properties as the Golgi-bound form (abstract no. 684). *Plant Physiology*. 105:S-126.

Dhugga, K. S., Tiwari, S. C., and Ray, P. M. (1997). A reversibly glycosylated polypeptide (RGP1) possibly involved in plant cell wall synthesis: purification, gene cloning, and trans-Golgi localization. *Proceedings of the National Academy of Sciences USA*. 94:7679-7684.

Diehl, J. A., Cheng, M., Roussel, M. F., and Sherr, C. J. (1998). Glycogen synthase kinase-3β regulates cyclin D1 proteolysis and subcellular localization. *Genes & Development*. 12:3499-3511.

Ding, S. W. (2000). RNA silencing. *Current Opinions in Biotechnology*. 11:152-156.

Driouich, A., Faye, L., and Staehelin, L. A. (1993). The plant Golgi apparatus: a factory for complex polysaccharides and glycoproteins. *Trends in Biochemical Sciences*. 18:210-214.

Dry, I. B., Krake, L. R., Mullineaux, P. M., and Rezaian, M. A. (2000). Regulation of tomato leaf curl viral gene expression in host tissues. *Molecular Plant-Microbe Interactions*. 13:529-537.

Dry, I. B., Krake, L. R., Rigden, J. E., and Rezaian, M. A. (1997). A novel subviral agent associated with a geminivirus: The first report of a DNA satellite. *Proceedings of the National Academy of Sciences USA*. 94:7088-7093.

Dry, I. B., Rigden, J. E., Krake, L. R., Mullineaux, P. M., and Rezaian, M. A. (1993). Nucleotide sequence and genome organization of tomato leaf curl geminivirus. Journal of General Virology. 74:147-151.

Duval, M., Hsieh, T.-F., Kin, S. Y., and Thomas, T. L. (2002). Molecular characterization of *AtNAM*: a member of the *Arabidopsis* NAC domain superfamily. *Plant Molecular Biology*. 50:237-248.

Eagle, P. A. and Hanley-Bowdoin, L. (1997). *Cis*-elements that contribute to geminivirus transcriptional regulation and efficient DNA replication. *Journal of Virology*. 71:6947-6955.

Eagle, P. A., Orozco, B. M., and Hanley-Bowdoin, L. (1994). A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *The Plant Cell*. 6:1157-1170.

Egelkrout, E. M., Robertson, D., and Hanley-Bowdoin, L. (2001). Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. *The Plant Cell*. 13:1437-1452.

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

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*. 8:978-984.

Epel, B. L., van Lent, J. W. M., Cohen, L., Kotlizky, G., Katz, A., and Yahalom, A. (1996). A 41 kDa protein isolated from maize mesocotyl cell walls immunolocalizes to plasmodesmata. *Protoplasma*. 191: 70-78.

Esau, K. (1977). Virus-like particles in the nuclei of phloem cells in spinach leaves infected with the curly top virus. *Journal of Ultrastructure Research*. 65:772-783.

Eshed, Y., Baum, S. F., Perea, J. V., and Bowman, J. L. (2001). Establishment of polarity in lateral organs of plants. *Current Biology*. 11:1251-1260.

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

Fashena, S. J., Serebriiskii, I., and Golemis, E. A. (2000). The continued evolution of two-hybrid screening approaches in yeast: how to outwit different prays with different baits. *Gene*. 250:1-14.

Fauquet, C. M., Maxwell, D. P., Gronenborn, B., and Stanley, J. (2000). Revised proposal for naming geminiviruses. *Archives of Virology*. 145:1743-1761.

Fitch, W. M. (2000). Homology: a personal view on some of the problems. *Trends in Genetics*. 16:227-231.

Fomenko, D. E. and Gladyshev, V. N. (2002). CxxS: fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Science*. 11:2285-2296.

Fontes, E. P. B., Eagle, P. A., Sipe, P. S., Luckow, V. A., and Hanley-Bowdoin, L. (1994a). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *Journal of Biological Chemistry*. 269:8459-8465.

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

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

Fontes, E. P., Santos, A. A., Luz, D. F., Waclawovsky, A. J., and Chory, J. (2004). The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes & Development*. 18:2545-2556.

Frangioni, J. V. and Neel, B. G. (1993). Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Analytical Biochemistry*. 210:179-187.

Freemont, P. S. (2000). Ubiquitination: RING for destruction? *Current Biology*. 10:R84-R87.

Frey, P. M., Scharer-Hernandez, N. G., Futterer, J., Potrykus, I., and Puonti-Kaerlas, J. (2001). Simultaneous analysis of the bidirectional African cassava mosaic virus promoter activity using two different luciferase genes. *Virus Genes*. 22:231-42.

Gladfelter, H. J., Eagle, P. A., Fontes, E. P. B., Batts, L. A., and Hanley-Bowdoin, L. (1997). Two domains of the AL1 protein mediate geminivirus origin recognition. *Virology*. 239:186-197.

Gleave, A. P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology*. 20:1203-1207.

Goffard, A. and Dubuisson, J. (2003). Glycosylation of hepatitis C virus envelope proteins. *Biochimie*. 85:295-301.

Golemis, E. A., Gyuris, J., and Brent, R. (1994). Interaction trap/two-hybrid systems to identify interacting proteins. In: Ausubel, F. M., Struhl. K. (eds). Current Protocols in Molecular Biology. Ch. 13.14. John Wiley & Sons, Inc., New York.

Gong, G., Waris, G., Tanveer, R., and Siddiqui, A. (2001). Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF-kappa B. *Proceedings of the National Academy of Sciences USA*. 98:9599-9604.

Gorbalenya, A. E. and Koonin, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Research*. 17:8413-8440.

Gorbalenya, A. E. and Koonin, E. V. (1993). Helicase: amino acid sequence comparisons and structure-function relationships. *Current Opinions in Structural Biology*. 3:419-429.

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

Guerin, J., Rossel, J. B., Robert, S., Tsuchiya, T. & Koltunow, A. (2000). A *DEFICIENS* homologue is down-regulated during apomictic initiation in ovules of *Hieracium*. *Planta*. 210:914-920.

Gunning, R. V., Byrne, F. J., Conde, B. D., Connelly, M. I., Hergstrom, K., and Devonshire, A. L. (1995). First report of B-biotype Bemisia tabaci (Hemiptera: Aleyrodidae) in Australia. *Journal of the Australian Entomological Society*. 34:116.

Haley, A., Zhan, X. C., 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.

Hallan, V. and Gafni, Y. (2001). Tomato yellow leaf curl virus (TYLCV) capsid protein (CP) subunit interactions: implications for viral assembly. *Archives of Virology*. 146:1765-1773.

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., and Robertson, D. (1999). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Science*. 18:71-106.

Hao, L., Wang, H., Sunter, G., and Bisaro, D. M. (2003). Geminivirus AL2 and L2 proteins interact with and inactivate SNF1 kinase. *The Plant Cell*. 15:1034-1048.

Harbour, J. W. and Dean, D. C. (2000). The Rb/E2F pathway: Expanding roles and emerging paradigms. *Genes & Development*. 14:2393-2409.

Hartitz, M. D., Sunter, G., Bisaro, D. M. (1999). The tomato golden mosaic virus transactivator (TrAP) is a single-stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology*. 263:1-14.

Hateboer, G., Hijmans, E. M., Nooij, J. B., Schlenker, S., Jentsch, S., and Bernards, R. (1996). mUBC9, a novel adenovirus E1A-interacting protein that complements a yeast cell cycle defect. *Journal of Biological Chemistry*. 271:25906-25911.

Hayes, R. J., MacDonald, H., Coutts, R. H. A., and Buck, K. W. (1988). Priming of complementary DNA synthesis in vitro by small DNA molecules tightly bound to virion DNA of wheat dwarf virus. *Journal of General Virology*. 69:1345-1350.

Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., Whitwill, S., and Lydiate, D. (2003). Molecular characterization of *Brassica napus* NAC domain transcriptional activators in response to biotic and abiotic stress. *Plant Molecular Biology*. 53:383-397.

Helin, K. (1998). Regulation of cell proliferation by the E2F transcription factors. *Current Opinion in Genetics and Development*. 8:28-35.

Herzenberg LA, De Rosa SC, Dubs JG, Roederer M, Anderson MT, Ela SW, Deresinski SC, Herzenberg LA. (1997). Glutathione deficiency is associated with impaired survival in HIV disease. *Proceedings of the National Academy of Sciences USA*. 94:1967-1972.

Heyraud, F., Matzeit, V., Kammann, M., Shaefer, S., Schell, J., and Gronenborn, B. (1993). Identification of the initiation sequence for viral-strand DNA synthesis of wheat dwarf virus. *EMBO Journal*. 12:4445-4452.

Hibara, K., Takada, S., and Tasaka, M. (2003). *CUC1* gene activates the expression of SAM-related genes to induce adventitious shoot formation. *The Plant Journal*. 36:687-696.

Hingorani, M. M. and O'Donnell, M. (2000). Sliding clamps: A tail(ored) fit. *Current Biology*. 10:R25-R29.

Hong, Y. and Stanley, J. (1995). Regulation of African cassava mosaic virus complementary-sense gene expression by N-terminal sequences of the replication-associated protein AC1. *Journal of General Virology*. 76:2415-2422.

Hong, Y., Saunders, K., and Stanley, J. 1997. Transactivation of dianthin gene expression by African cassava mosaic virus AC2. *Virology*. 228:383-387.

Hoogstraten, R. A., Hanson, S. F., and Maxwell, D. P. (1996). Mutational analysis of the putative nicking motif in the replication-associated protein (AC1) of bean golden mosaic geminivirus. *Molecular Plant-Microbe Interactions*. 9:594-599.

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

Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., Mizuno, T., and Yamazaki, T. (2002). Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. *The Plant Cell.* 14:2015-29.

Hull, R. (2004). Matthews' Plant Virology. *Academic Press*, London.

Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends in Plant Science.* 7:106-111.

Jansen-Durr, P. (1996). How viral oncogenes make the cell cycle. *Trends in Genetics.* 12:270-275.

Jentsch, S. (1992). The ubiquitin-conjugation system. *Annual Review of Genetics.* 26:179-207.

Jeske, H., Lutgemeier, M., and Preiβ, W. (2001). DNA forms indicate rolling circle and recombination dependent replication of *Abutilon* mosaic virus. *EMBO Journal.* 20:6158-6167.

Jeske, H. and Preiβ, W. (2003). Multitasking in replication is common among geminiviruses. *Journal of Virology.* 77:2972-2980.

Jin, H., Axtell, M. J., Dahlbeck, D., Ekwenna, O., Zhang, S., Staskawicz, B., and Baker, B. (2002). NPK1, an MEKK1-like mitogen-activated protein kinase kinase kinase, regulates innate immunity and development in plants. *Developmental Cell.* 3:291-297.

John, I., Hackett, R., Cooper, W., Drake, R., Farrell, A., and Grierson, D. (1997). Cloning and characterization of tomato leaf-senescence-related cDNAs. *Plant Molecular Biology.* 33:641-651.

Johnson, P. F. and McKnight, S. L. (1989). Eukaryotic transcriptional regulatory proteins. *Annual Reviews in Biochemistry.* 58:799-839.

Jonak, C. and Hirt, H. (2002). Glycogen synthase kinase 3/SHAGGY-like kinases in plants: an emerging family with novel functions. *Trends in Plant Science.* 7:457-461.

Jowett, T. (2001). Double *in situ* hybridization techniques in zebrafish. *Methods.* 23:345-358.

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

Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K., and Poethig, R. S. (2001). KANADI regulates organ polarity in *Arabidopsis*. *Nature.* 411:706-709.

Kho, C. J., Huggins, G. S., Endege, W. O., Hsieh, C. M., Lee, M. E., and Haber, E. (1997). Degradation of E2A proteins through a ubiquitin-conjugating enzyme, UbcE2A. *Journal of Biological Chemistry*. 272:3845-51.

Kikuchi, K., Ueguchi-Tanaka, M., Yoshida, K. T., Nagato, Y., Matsusoka, M., and Hirano, H.-Y. (2000). Molecular analysis of the *NAC* gene family in rice. *Molecular and General Genetics*. 262:1047-1051.

Kirthi, N. and Savithri, H. S. (2003). A conserved zinc finger motif in the coat protein of Tomato leaf curl Bangalore virus is responsible for binding to ssDNA. *Archives of Virology*. 148:2369-2380.

Klimyuk, V. I., Persello-Cartieaux, F., Havaux, M., Contard-David, P., Schuenemann, D., Meiherhoff, K., Gouet, P., Jones, J. D., Hoffman, N. E., and Nussaume, L. (1999). A chromodomain protein encoded by the arabidopsis *CAO* gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting. *The Plant Cell*. 11:87-99.

Kong, L. J., Orozco, B. M., Roe, J. L., Nagar, S., Ou, S., Feiler, H. S., Durfee, T., Miller, A. B., Gruissem, W., Robertson, D., and Hanley-Bowdoin, L. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO Journal*. 19:3485-3495

Kong, L. J. and Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *The Plant Cell*. 14:1817-1832.

Kornberg, A. and Baker, T. A. (1992). DNA Replication. 2nd ed. W. H. Freeman and Company, New York.

Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry*. 54:631-664.

Kosarev, P., Mayer, K. F. X., and Hardtke, C. S. (2002). Evaluation and classification of RING-finger domains encoded by the *Arabidopsis* genome. *Genome Biology*. 3:research0016.1-0016.12.

Koundouris, A., Kass, G. E., Johnson, C. R., Boxall, A., Sanders, P. G., and Carter, M. J. (2000). Poliovirus induces an early impairment of mitochondrial function by inhibiting succinate dehydrogenase activity. *Biochemical and Biophysical Research Communications*. 271:610-614.

Krake, L. R., Rezaian, M. A., and Dry, I. B. (1998). Expression of the tomato leaf curl geminivirus C4 gene produces viruslike symptoms in transgenic plants. *Molecular Plant-Microbe Interactions*. 11:413-417.

Kroemer, G., Dallaporta, B., and Resche-Rigon, M. (1998). The mitochondrial death/life regulator in apoptosis and necrosis. *Annual Review of Physiology*. 60:619-642.

Kunik, T., Mizrachy, L., Citovsky, V., and Gafni, Y. (1999). Characterization of a tomato karyopherin α that interacts with the Tomato yellow leaf curl virus (TYLCV) capsid protein. *Journal of Experimental Botany*. 50:731-732.

Laney, J. D. and Hochstrasser, M. (1999). Substrate targeting in the ubiquitin system. *Cell*. 97:427-430.

Langeveld, S. M., Vennik, M., Kottenhagen, M., van Wijk, R., Buijk, A., Kijne, J. W., and de Pater, S. (2002). Glucosylation activity and complex formation of two classes of reversibly glycosylated polypeptides. *Plant Physiology*. 129:278-289.

Latham, J. R., Saunders, K., Pinner, M. S., and Stanley, J. (1997). Induction of plant cell division by beet curly top virus gene C4. *The Plant Journal*. 11:1273-1283.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. (1995a). Geminivirus replication: Genetic and biochemical characterization of Rep protein function, a review. *Biochimie*. 77:765-773.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995b). Identification of the nicking tyrosine of geminivirus rep protein. *FEBS Letters*. 377:258-262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S., Schell, J., and Gronenborn, B. (1995c). *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences USA*. 92:3879-3883.

Lazarowitz, S. G. and Beachy, R. N. (1999). Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *The Plant Cell*. 11:535-548.

Lazarowitz, S. G., Pinder, A. J., Damsteegt, V. D., and Rogers, S. G. (1989). Maize streak virus genes essential for systemic spread and symptom development. *EMBO Journal*. 8:1023-1032.

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.

Lechner, M. S. and Laimins, L. A. (1994). Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *Journal of Virology*. 68:4262-4273.

Li, J. and Nam, K. H. (2002). Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science*. 295:1299-1301.

Li, R., Yu, D. S., Tanaka, M., Zheng, L., Berger, S. L., and Stillman, B. (1998). Activation of chromosomal DNA replication in *Saccharomyces cerevisiae* by acidic transcriptional activation domains. *Molecular and Cellular Biology*. 18:1296-1302.

Lin, B., Behjatnia, S. A. A., Dry, I. B., Randles, J. W., and Rezaian, M. A. (2002). High affinity Rep-binding is not required for the replication of a geminivirus DNA and its satellite. *Virology*. 305:353-363.

Liu, H. T., Boulton, M. I., Oparka, K. J., and Davies, J. W. (2001). Interaction of the movement and coat proteins of maize streak virus: Implications for the transport of viral DNA. *Journal of General Virology*. 82:35-44.

Liu, H., Boulton, M. I., Thomas, C. L., Prior, D. A., Oparka, K. J., and Davies J. W. (1999). *Maize streak virus* coat protein is karyophyllic and facilitates nuclear transport of viral DNA. *Molecular Plant-Microbe Interactions*. 12:894-900.

Liu, L., Davies, J. W., and Stanley, J. (1998). Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus *Mastrevirus* that is adapted to dicotyledonous plants. *Journal of General Virology*. 79:2265-2274.

Luque, A., Sanz-Burgos, A. P., Ramirez-Parra, E., Castellano, M. M., and Gutierrez, C. (2002). Interaction of geminivirus Rep protein with replication factor C and its potential role during geminivirus DNA replication. *Virology*. 302:83-94.

Makkouk, K. M. and Laterrot, H. (1983). Epidemiology and control of tomato yellow leaf curl virus. In: Plumb, R. T. and Thresh, J. M. (eds). Plant Virus Epidemiology. p. 315-321. Blackwell Scientific Publications, Oxford, UK.

Mansoor S, Briddon RW, Zafar Y, Stanley J. (2003). Geminivirus disease complexes: an emerging threat. *Trends in Plant Science*. 8:128-134.

Mariano, A. C., Andrade, M. O., Santos, A. A., Carolino, S. M., Oliveira, M. L., Baracat-Pereira, M. C., Brommonshenkel, S. H., and Fontes E. P. (2004). Identification of a novel receptor-like protein kinase that interacts with a geminivirus nuclear shuttle protein. *Virology*. 318:24-31.

Martin, C. and Paz-Ares, J. (1997). MYB transcription factors in plants. *Trends in Genetics*. 13:67-73.

McGarry, R. C., Barron, Y. D., Carvalho, M. F., Hill, J. E., Gold, D., Cheung, E., Kraus, W. L., and Lazarowitz, S. G. (2003). A novel *Arabidopsis* acetyltransferase interacts with the geminivirus movement protein NSP. *The Plant Cell*. 15:1605-1618.

Mittnacht, S. (1998). Control of pRB phosphorylation. *Current Opinion in Genetics and Development*. 8:21-27.

Moffat, A. S. (1999). Geminiviruses emerge as serious crop threat. *Science*. 286:1835.

Moffatt, B. A., Stevens, Y. Y., Allen, M. S., Snider, J. D., Pereira, L. A., Todorova, M. I., Summers, P. S., Weretilnyk, E. A., Martin-McCaffrey, L., and Wagner, C. (2002). Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. *Plant Physiology*. 128:812-821.

Moran, E. (1993). Interactions of adenoviral proteins with pRB and p53. *FASEB Journal*. 7:880-885.

Morin, S., Ghanim, M., Sobol, I., and Czosnek, H. (2000). The GroEL protein of the whitefly Bemisia tabaci interacts with the coat protein of transmissible and nontransmissible begomoviruses in the yeast two-hybrid system. *Virology*. 276:404-416.

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

Mullineaux, P. M., Rigden, J. E., Dry, I. B., Krake, L. R., and Rezaian, M. A. (1993). Mapping of the polycistronic RNAs of tomato leaf curl geminivirus. *Virology*. 193:414-423.

Murant, A. F. and Mayo, M. A. (1982). Satellites of plant viruses. *Annual Reviews in Phytopathology*. 20:49-70.

Mysore, K. S., Crasta, O. R., Tuori, R. P., Folkerts, O., Swirsky, P. B., and Martin, G. B. (2002). Comprehensive transcript profiling of Pto- and Prf-mediated host defense responses to infection by *Pseudomonas syringae* pv. *tomato*. *The Plant Journal*. 32:299-315.

Nagar, S., Pedersen, T. J., Carrick, K. M., Hanley-Bowdoin, L., and Robertson, D. (1995). A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. *The Plant Cell*. 7:705-719.

Nakhla, M. K. and Maxwell, D. P. (1998). Epidemiology and management of tomato yellow leaf curl disease. In: Hadidi, A., Khetarpal, R. K., and Koganezawa, H. K. (eds). Plant Virus Disease Control. p.565-583. APS Press.

Needleman, S. B. and Wunsch C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*. 48:443-453.

Nencioni L, Iuvara A, Aquilano K, Ciriolo MR, Cozzolino F, Rotilio G, Garaci E, Palamara AT. (2003). Influenza A virus replication is dependent on an antioxidant pathway that involves GSH and Bcl-2. *FASEB Journal*. 17:758-760.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering. 10:1-6.

Nigg, E. A. (1997). Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature*. 386:779-787.

Noris, E., Jupin, I., Accotto, G. P., and Gronenborn, B. (1996). DNA-binding activity of the C2 protein of tomato yellow leaf curl geminivirus. *Virology*. 217:607-612.

Noris, E., Vaira, A. M., Caciagli, P., Masenga, V., Gronenborn, B., and Accotto, G. P. (1998). Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *Journal of Virology*. 72:10050-10057.

Novick, R. P. (1998). Contrasting lifestyles of rolling-circle phages and plasmids. *Trends in Biochemical Sciences*. 23:434-438.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K. and Kikuchi, S. (2003). Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research*. 10:239-247.

Orozco, B. M., Gladfelter, H. J., Settlage, S. B., Eagle, P. A., Gentry, R., and Hanley-Bowdoin, L. (1998). Multiple cis elements contribute to geminivirus origin function. *Virology*. 242:346-356.

Orozco, B. M. and Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus origin function. *Journal of Virology*. 270:148-158.

Orozco, B. M. and Hanley-Bowdoin, L. (1998). Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminvirus replication protein. *Journal of Biological Chemistry*. 273:24448-24456.

Orozco, B. M., Miller, A. B., Settlage, S. B., and Hanley-Bowdoin, L. (1997). Functional domains of a geminivirus replication protein. *Journal of Biological Chemistry*. 272:9840-9846.

Padidam, M., Beachy, R. N., and Fauquet, C. M. (1996). The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology*. 224:390-404.

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

Palanichelvam, K., Kunik, T., Citovsky, V., and Gafni, Y. (1998). The capsid protein of tomato yellow leaf curl virus binds cooperatively to single-stranded DNA. *Journal of General Virology*. 79:2829-2833.

Paro, R. and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proceedings of the National Academy of Sciences USA*. 88:263-267.

Pascal, E., Sanderfoot, A. A., Ward, B. M., Medville, R., Turgeon, R., and Lazarowitz, S. G. (1994). The geminivirus BR1 movement protein binds single-stranded DNA and localizes to the cell nucleus. *The Plant Cell*. 6:995-1006.

Perez-Perez, J. M., Ponce, M. R., and Micol, J. L. (2002). The *UCU1 Arabidopsis* gene encodes a SHAGGY/GSK3-like kinase required for cell expansion along the proximodistal axis. *Developmental Biology*. 242:161-173.

Pichler, A., Gast, A., Seeler, J. S., Dejean, A., Melchior, F. (2002). The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell*. 108:109-120.

Pichler, A., Knipscheer, P., Saitoh, H., Sixma, T. K., and Melchior, F. (2004). The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type. *Nature Structural and Molecular Biology*. 11:984-991.

Pilartz, M. and Jeske, H. (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology*. 189:800-802.

Qin, S., Ward, B. M., and Lazarowitz, S. G. (1998). The bipartite geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA. *Journal of Virology*. 72:9247-9256.

Rangasamy, D. and Wilson V. G. (2000). Bovine papillomavirus E1 protein is sumoylated by the host cell Ubc9 protein. *Journal of Biological Chemistry*. 275:30487-30495.

Rangasamy, D., Woytek, K., Khan, S. A., and Wilson, V. G. (2000). SUMO-1 modification of bovine papillomavirus E1 protein is required for intranuclear accumulation. *Journal of Biological Chemistry*. 275:37999-38004.

Ren, T. Qu, F., and Morris, T. J. (2000). *HRT* gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. *The Plant Cell*. 12:1917-1925.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., and Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 290:2105-2110.

Rigden, J. E., Dry, I. B., Krake, L. R., and Rezaian, M. A. (1996). Plant virus DNA replication processes in *Agrobacterium*: insight into the origins of geminiviruses? *Proceedings of the National Academy of Sciences USA*. 93:10280-10284.

Rigden, J. E., Dry, I. B., Mullineaux, P. M., and Rezaian, M. A. (1993). Mutagenesis of the virion-sense open reading frames of tomato leaf curl geminivirus. *Virology*. 193:1001-1005.

Rigden, J. E., Krake, L. R., Rezaian, M. A., and Dry, I. B. (1994). ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. *Virology*. 204:847-850.

Robatzek, S. and Somssich, I. E. (2002). Targets of *At*WRKY6 regulation during plant senescence and pathogen defense. *Genes & Development*. 16:1139-1149.

Rojas, M. R., Jiang, H., Salati, R., Xoconostle-Cazares, B., Sudarshana, M. R., Lucas, W. J., and Gilbertson, R. L. (2001). Functional analysis of proteins involved in movement of the monopartite begomovirus, *Tomato yellow leaf curl virus*. *Virology*. 291:110-125.

Rouhier, N., Gelhaye, E., and Jacquot, J. P. (2004). Plant glutaredoxins: still mysterious reducing systems. *Cellular and Molecular Life Sciences*. 61:1266-1277.

Ruiz-Medrano, R., Xoconostle-Cazares, B., and Lucas, W. J. (1999). Phloem long-distance transport of *CmNACP* mRNA: implications for supracellular regulation in plants. *Development*. 126:4405-4419.

Sablowski, R. W. M. and Meyerowitz, E. M. (1998). A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. *Cell*. 92:93-103.

Saeed, M., Behjatnia, S. A. A., Mansoor, S., Zafar, Y., Hasnain, S., and Rezaian, M. A. (2004). A single complementary-sense transcript of a geminivirus DNA beta satellite is determinant of pathogenicity. *Molecular Plant-Microbe Interactions*. **17** (In press).

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 4:406-425.

Sakai, H., Aoyama, T., and Oka, A. (2000). *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. *The Plant Journal*. 24:703-711.

Sambrook, J. and Russell, G. W. (2001). Molecular cloning, a laboratory manual. *Cold Spring Harbour Laboratory Press*, New York.

Sanderfoot, A. A. and Lazarowitz, S. G. (1995). Getting it together in plant virus movement: Cooperative interactions between bipartite geminivirus movement proteins. *Trends in Cell Biology*. 6:353-358.

Sanderfoot, A. A. and Lazarowitz, S. G. (1996). Getting it together in plant virus movement: Cooperative interactions between bipartite geminivirus movement proteins. *Trends in Cell Biology*. 6:353-358.

Sangare, A., Deng, D., Fauquet, C., and Beachy, R. N. (1999). Resistance to African cassava mosaic virus conferred by a mutant of the putative NTP-binding domain of the Rep gene (AC1) in *Nicotiana benthamiana*. *Molecular Biology Reports*. 5:95-102.

Saunders, K., Bedford, I. D., Briddon, R. W., Markham, P. G., Wong, S. M., and Stanley, J. (2000). A unique virus complex causes Ageratum yellow vein disease. *Proceedings of the National Academy of Sciences USA*. 97:6890-6895.

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 Research*. 19:2325-2330.

Saunders, K., Norman, A., Gucciardo, S., and Stanley, J. (2004). The DNA beta satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (betaC1). *Virology*. 324:37-47.

Scheffner, M. and Whitaker, N. J. (2003). Human papillomavirus-induced carcinogenesis and the ubiquitin-proteasome system. *Seminars in Cancer Biology*. 13:59-67.

Schwartz, D. C. and Hochstrasser, M. (2003). A superfamily of protein tags: ubiquitin, SUMO and related modifiers. *Trends in Biochemical Science*. 28:321-328.

Selth, L. A. (2000). Functional analysis of the *Tomato leaf curl geminivirus* genes: An assessment of gene silencing. Honours thesis in Biotechnology, School of Biological Sciences, Flinders University, South Australia.

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H., Randles, J. W., and Rezaian, M. A. (2005). A NAC domain protein interacts with *Tomato leaf curl virus* replication accessory protein and enhances viral replication. *The Plant Cell*. 17:311-325.

Selth, L. A., Randles, J. W., and Rezaian, M. A. (2002). *Agrobacterium tumefaciens* supports DNA replication of diverse geminivirus types. *FEBS Letters*. 516:179-182.

Selth, L. A., Randles, J. W., and Rezaian, M. A. (2004). Host responses to transient expression of individual genes encoded by *Tomato leaf curl virus*. *Molecular Plant-Microbe Interactions*. 17:27-33.

Settlage, S. B., Miller, A. B., Gruissem, W., and Hanley-Bowdoin, L. (2001). Dual interaction of a geminvirus replication accessory factor with a viral replication protein and a plant cell cycle regulator. *Virology*. 279:570-576.

Settlage, S.B., Miller, A.B., and Hanley-Bowdoin, L. (1996). Interactions between geminivirus replication proteins. *Journal of Virology*. 70:6790-6795.

Seufert W and Jentsch S. (1990). Ubiquitin-conjugating enzymes UBC4 and UBC5 mediate selective degradation of short-lived and abnormal proteins. *EMBO Journal*. 9:543-550.

Silhavy, D. and Burgyan, J. (2004). Effects and side-effects of viral RNA silencing suppressors on short RNAs. *Trends in Plant Science*. 9:76-83.

Smith, T. F. and Waterman, M. S. (1981). Comparison of biosequences. *Advances in Applied Mathematics*. 2:482-489.

Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., Gardner, J., Wang, B., Zhai, W. X., and Zhu, L. H. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*. 270:1804-1806.

Souer, E., van Houwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The *No Apical Meristem* gene of *Petunia* is requried for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*. 85:159-170.

Staiger, C. and Doonan, J. (1993). Cell division in plants. *Current Opinions in Cell Biology*. 5:226-231.

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

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. *Proceedings of the National Academy of Sciences*. 88:8029-8033.

Stephens, D. J. and Banting, G. (2000). The use of yeast two-hybrid screens in studies of protein:protein interactions involved in trafficking. *Traffic*. 1:763-768.

Stonor, J., Hart, P., Gunther, M., DeBarro, P., and Rezaian, M. (2003). Tomato leaf curl geminivirus in Australia: occurrence, detection, sequence diversity and host range. *Plant Pathology*. 52:379-388.

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. *The Plant Cell*. 4:1321-1331.

Sunter, G., Hartitz, D. M., Hormuzdi, S. G., Stenger, D. C., and Bisaro, D. M. (1991). Tomato golden mosaic virus gene function. *Phytopathology*. 81:1184-1189.

Szick, K., Springer, M., and Bailey-Serres, J. (1998). Evolutionary analyses of the 12-kDa acidic ribosomal P-proteins reveal a distinct protein of higher plant ribosomes. *Proceedings of the National Academy of Sciences USA*. 95:2378-2383.

Takada, S., Hibara, K., Ishada, T., and Tasaka, M. (2001). The *CUP-SHAPED COTYLEDON1* gene of *Arabidopsis* regulates shoot apical meristem formation. *Development*. 128:1127-1135.

Takane, K., Tajima, S., and Kouchi, H. (2000). Structural and expression analysis of uricase mRNA from *Lotus japonicus*. *Molecular Plant-Microbe Interactions*. 13:1156-1160.

Tavares, R., Vidal, J., van Lammeren, A., and Kreis, M. (2002). AtSKθ, a plant homologue of SGG/GSK-3 marks developing tissues in *Arabidopsis thaliana*. *Plant Molecular Biology*. 50:261-271.

Tavladoraki, P., Benvenuto, E., Trinca, S., De Martinis, D., Cattaneo, A., and Galeffi, P. (1993). Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack. *Nature*. 366:469-472.

Theissen, G. (2002). Secret life of genes. *Nature*. 415:741.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 25:4876-4882.

Toledo-Ortiz, G., Huq, E., and Quail, P. H. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. *The Plant Cell*. 15:1749-1770.

Tumer, N. E., Hudak, K., Di, R., Coetzer, C., Wang, P., and Zoubenko, O. (1999). Pokeweed antiviral protein and its applications. *Current Topics in Microbiology and Immunology*. 240:139-158.

Tusnády, G. E. and Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction. *Journal of Molecular Biology*. 283:489-506.

Unseld, S., Hohnle, M., Ringel, M., and Frischmuth, T. (2001). Subcellular targeting of the coat protein of *African cassava mosaic geminivirus*. *Virology*. 286:373-83.

Vanitharani, R., Chellappan, P., Pita, J. S., and Fauquet, C. M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *Journal of Virology*. 78:9487-9498.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (2000). Virus Taxonomy- Classification and Nomenclature of Viruses. *Academic Press*, San Diego.

van Wezel, R., Liu, H., Tien, P., Stanley, J., and Hong, Y. 2001. Gene C2 of the monopartite geminivirus *Tomato yellow leaf curl virus*-China encodes a pathogenicity determinant that is localized in the nucleus. *Molecular Plant-Microbe Interactions*. 14:1125-1128.

van Wezel, R., Dong, X., Blake, P., Stanley, J., and Hong, Y. (2002a). Differential roles of geminivirus Rep and AC4 (C4) in the induction of necrosis in *Nicotiana benthamiana*. *Molecular Plant Pathology*. 3:461-471.

van Wezel, R., Dong, X., Liu, H., Tien, P., Stanley, J., and Hong, Y. (2002b). Mutation of three cysteine residues in *Tomato yellow leaf curl virus*-China C2 protein causes dysfunction in pathogenesis and postranscriptional gene-silencing suppression. *Molecular Plant-Microbe Interactions*. 15:203-208.

Vargason, J. M., Szittya, G., Burgyan, J., Tanaka Hall, T. M. (2003). Size selective recognition of siRNA by an RNA silencing suppressor. *Cell*. 115:799-811.

Vaucheret, H., Beclin, C., and Fagard, M. (2001). Post-transcriptional gene silencing in plants. *Journal of Cell Science*. 114:3083-3091.

Voinnet, O. (2002). RNA silencing: small RNAs as ubiquitous regulators of gene expression. *Current Opinion in Plant Biology*. 5:444-451.

Voinnet, O., Pinto, Y. M., and Baulcombe, D. C. (1999). Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences*. 96:14147-14152.

Vroemen, C. W., Mordhorst, A. P., Albrecht, C., Kwaaitaal, M. A., de Vries, S. C. (2003). The *CUP-SHAPED COTYLEDON3* gene is required for boundary and shoot meristem formation in *Arabidopsis*. *The Plant Cell*. 15:1563-1577.

Vullo, A. and Frasconi, P. (2004). Disulfide connectivity prediction using recursive neural networks and evolutionary information. *Bioinformatics*. 20:653-659.

Wald, F. A., Kissen, R., du Jardin, P., and Moreno, S. (2003). Characterization of UDP-glucose protein transglucosylase genes from potato. *Plant Molcular Biology*. 52:705-714.

Wang, H., Hao, L., Shung, C. Y., Sunter, G., and Bisaro, D. M. (2003). Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. *The Plant Cell*. 15:3020-3032.

Watanabe, Y., Ogawa, T., Takahashi, H., Ishida, I., Takeuchi, Y., Yamamoto, M., and Okada, Y. (1995). Resistance against multiple plant viruses in plants mediated by a double stranded-RNA specific ribonuclease. *FEBS Letters*. 372:165-168.

White, C. L., Senkevich, T. G., and Moss, B. (2002). Vaccinia virus G4L glutaredoxin is an essential intermediate of a cytoplasmic disulfide bond pathway required for virion assembly. *Journal of Virology*. 76:467-472.

White, C. L., Weisberg, A. S., and Moss, B. (2000). A glutaredoxin, encoded by the G4L gene of Vaccinia virus, is essential for virion morphogenesis. *Journal of Virology*. 74:9175-9183.

Wilson, V. G. and Rangasamy, D. (2001). Viral interaction with the host cell sumoylation system. *Virus Research*. 81:17-27.

Xie, Q., Frugis, G., Colgan, D., and Chua, N.-H. (2000). *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes & Development*. 14:3024-3036.

Xie, Q., Sanz-Burgos, A. P., Guo, H., Garcia, J. A., and Gutierrez, C. (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Molecular Biology*. 39:674-656.

Xie, Q., Suarez-Lopez, P., and Gutierrez, C. (1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO Journal*. 14:4073-4082.

Xu, W., Gong, L., Haddad, M. M., Bischof, O., Campisi, J., Yeh, E. T., and Medrano, E. E. (2000). Regulation of microphthalmia-associated transcription factor MITF protein levels by association with the ubiquitin-conjugating enzyme hUBC9. *Experimental Cell Research*. 255:135-43.

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbanje-McKenna, M., Boulton, M. I., Davies, J. W., and McKenna, R. (2001). Structure of the Maize streak virus geminate particle. *Virology*. 279:471-477.

Zhou, J., Loh, Y.-T., Bressan, R. A., and Martin, G. B. (1995). The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. *Cell.* 83:925-935.

Appendix 1 - Sequences retrieved from yeast two-hybrid screens

Sequences retrieved from each of the yeast two-hybrid screens using viral proteins as bait (Chapters 3, 4, 5 and 6) are shown below. cDNA sequences are shown first, followed by putative ORFs and predicted translation products. GeneDoc version 2.4.016, NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and ORF Finder (http://www.cbio.psu.edu/sms/orf_find.html) were used to predict ORFs and proteins within cDNA sequences retrieved from yeast two-hybrid screens. Putative start and stop codons in cDNAs and ORFs are shown in red. C-terminal ends of protein sequences are indicated by an asterisk (*).

V1

SlUPTG1 (Genbank accession number AY622990)

cDNA

AAAAATCAAAATCTTGAAAATTCATAGAGAATTAGCTATGGCAGCAGCAACACCACTGTTGAAAGATGAG
CTTGATATTGTGATTCCCACAATAAGAAATCTTGATTTTTTGGAGATGTGGAGACCCTTTTTTCAGCCAT
ACCATCTGATTATTGTTCAAGATGGTGATCCTTCAAAGACCATTAAGGTCCCTGAAGGATTTGATTATGA
GCTTTATAATCGTAATGACATTAACAGGATTTTGGGTCCTAAAGCATCTTGTATCTCTTTTAAGGATTCT
GCTTGTAGGTGTTTTGGGTATATGGTGTCTAAGAAGAAGTATATCTACACCATTGATGATGATTGCTTTG
TGGCCAAGGACCCGTCTGGTAAGGATATCAATGCACTTGAGCAGCACATCAAGAACCTCCTGTGCCCATC
TACTCCGCACTTCTTCAACACTCTGTATGATCCATACAGAGATGGTGCAGATTTCGTCCGTGGCTACCCT
TTCAGCATGCGTGAGGGTGCTCCAACAGCTGTTTCTCATGGATTGTGGCTCAACATCCCTGACTACGATG
CTCCCACGCAGCTTGTTAAGCCTCATGAGAGGAACACTAGATATGTTGACGCTGTCATGACGATTCCAAA
GGGCACTTTGTTCCCTATGTGTGGAATGAACTTGGCCTTTGACCGTGATCTCATTGGACCTGCAATGTAC
TTTGGTCTCATGGGTGACGGTCAGCCAATTGGTCGTTACGACGATATGTGGGCCGGCTGGTGTACCAAGG
TCATATGTGACCATTTGGGACTAGGAATCAAGACTGGTCTGCCCTACATATGGCACAGCAAAGCTAGCAA
CCCATTCGTTAACCTCAAAAAGGAGTACAACGGTATCTTCTGGCAAGAGGAGATCATCCCCTTCTTCCAG
ACTGCAACGCTTCCTAAAGAGTGTACAACCGTCCAGCAATGCTACCTTGAGCTCTCAAAGCAGGTTAAGG
AAAAACTTTCCAAGATCGATCCCTATTTCACCAAGCTAGGAGATGCTATGGTCACGTGGATCGAAGCTTG
GGATGAGCTTAACCCTACTGGGGACAACTTGGCTAAGCTGTCCATCGCCGATGGTCCAGCAAAGACTAAA
AAGTAGATGGCTTTTTTGCTCATTTTCTTCGACGACTAAGACACTTATGAGTTAACTAGAACTGGATTTT
ACTTGTTTTAATTTTAGTTGAGGATTTAGGTTTCATATTGCTGCTGTTATGTGGAGCAGCTTTTGATTGC
CTTCTCTGAGACATGATAGGAGTATTTTTATTTTTATTTTTTTATTTATATAAGGCTTAGTTCCAAACTT
TGAGATTTATCTTTTTTATATCAATAAATTATGGTTTAGTGCTTTATGCTTAAAAAAAAAAAAAAAAAACT
CGAG

ORF

ATGGCAGCAGCAACACCACTGTTGAAAGATGAGCTTGATATTGTGATTCCCACAATAAGAAATCTTGATT
TTTTGGAGATGTGGAGACCCTTTTTTCAGCCATACCATCTGATTATTGTTCAAGATGGTGATCCTTCAAA
GACCATTAAGGTCCCTGAAGGATTTGATTATGAGCTTTATAATCGTAATGACATTAACAGGATTTTGGGT
CCTAAAGCATCTTGTATCTCTTTTAAGGATTCTGCTTGTAGGTGTTTTGGGTATATGGTGTCTAAGAAGA
AGTATATCTACACCATTGATGATGATTGCTTTGTGGCCAAGGACCCGTCTGGTAAGGATATCAATGCACT
TGAGCAGCACATCAAGAACCTCCTGTGCCCATCTACTCCGCACTTCTTCAACACTCTGTATGATCCATAC
AGAGATGGTGCAGATTTCGTCCGTGGCTACCCTTTCAGCATGCGTGAGGGTGCTCCAACAGCTGTTTCTC
ATGGATTGTGGCTCAACATCCCTGACTACGATGCTCCCACGCAGCTTGTTAAGCCTCATGAGAGGAACAC
TAGATATGTTGACGCTGTCATGACGATTCCAAAGGGCACTTTGTTCCCTATGTGTGGAATGAACTTGGCC
TTTGACCGTGATCTCATTGGACCTGCAATGTACTTTGGTCTCATGGGTGACGGTCAGCCAATTGGTCGTT
ACGACGATATGTGGGCCGGCTGGTGTACCAAGGTCATATGTGACCATTTGGGACTAGGAATCAAGACTGG
TCTGCCCTACATATGGCACAGCAAAGCTAGCAACCCATTCGTTAACCTCAAAAAGGAGTACAACGGTATC
TTCTGGCAAGAGGAGATCATCCCCTTCTTCCAGACTGCAACGCTTCCTAAAGAGTGTACAACCGTCCAGC

AATGCTACCTTGAGCTCTCAAAGCAGGTTAAGGAAAAACTTTCCAAGATCGATCCCTATTTCACCAAGCT
AGGAGATGCTATGGTCACGTGGATCGAAGCTTGGGATGAGCTTAACCCTACTGGGGACAACTTGGCTAAG
CTGTCCATCGCCGATGGTCCAGCAAAGACTAAAAAGTAG

Protein

MAAATPLLKDELDIVIPTIRNLDFLEMWRPFFQPYHLIIVQDGDPSKIIKVPEGFDYELYNRNDINRILG
PKASCISFKDSACRCFGYMVSKKKYIYTIDDDCFVAKDPSGKDINALEQHIKNLLCPSTPHFFNTLYDPY
RDGADFVRGYPFSMREGAPTAVSHGLWLNIPDYDAPTQLVKPHERNTRYVDAVMTIPKGTLFPMCGMNLA
FDRDLIGPAMYFGLMGDGQPIGRYDDMWAGWCTKVICDHLGLGIKTGLPYIWHSKASNPFVNLKKEYNGI
FWQEEIIPFFQAATLPKECTTVQQCYLELSKQVKKKLSSIDPYFTKLGEAMVTWIEAWDELNLLGTTWLS
CLSPMVQQRLKSRCY*

9A

cDNA

CCCTCATGAACCAGGCAATTTGAAAGTAAAATCGTACAGAAAAGTTAAATCTAATAGAAGGCAGGGTGTG
GTATCCAGTAAACTTGCATCTTTGGCCGCAGAATTTCCTCTTGCAAGATTACATGGTCCCGTTAATGCCG
ACCTCAACGAGATGTGGGAGAGGAAATGTTGTGCTAAGCAAGTCGACTCACAGTCTCCACCACTATTTGA
GAAGCTCCGGGAATCACTTCTTCATGGGGTGAACATTGACTATGATGATTTCTGTACCCCAAATGAAAAG
AATGAAGACAATGACTATGATAGTGCTGATCATGATTTTGGGCCTCCTGATTTTGACATGCCAGAAAATG
CAGACATGAACAGCCATGCTACTCCACATGATGAAAAGCATGATAATTGTGGTCCACTTTTTGATAGTGA
AGCTCATGAAGATCTGAATGGTCAAGAAAACCTTGAAGATCTTTGTCGCTCCCACTTGGATGCTCTTCTT
GCTAACCTTGCTGAAACTGAGAAGCAGACTGAATTGGCTGCTCGGGTTTCAACGTGGAAACAGAGAATTG
ACCAGGAACTTGGAGAACAAGAATCACATCCACCCTTTGACATTCATGAATATGGGGCAAGGGTTTTGAA
CAAGTTATCCCTGGAAGAAAATGATAAAAGCACCATGTCCTTTTCTGATGTTGTCAAGGGTTCGGAGAAG
CATGACATTGCTCGAACATTTTCTGCGCTTCTGCAATTGGTAAACAATGGAGACGTTGCTTTGGAAAGAG
GTGAGGTAGGCGAGTCCACTTGTTACACAGCTGCAAATCCCTTCTCTGTTCAGCTCCTTAGGCATGGCAA
CGATAGGGAGGAAATGCAGTTTCAATCAACAAAAAAGAGAGCAGAATCTCCAATGCACCATCAGAACAAT
AGAAAGGAAAAGAACAAAGGTAAAGCCGTTCATGCTGCTGTTGATTCATCGCCTCCAGGACCCGATTCAG
ATAGCAGATTACCCCTGAAGCTGGGAAAGGTTAATGGGACAAGATGCACACCTGATAGCAAGAAAAGAAG
GAAGTCCAGAATAGCGTTAGCATCGGATGTGCCTACTGCATTGTAGTAAGTCAGGAAAGATTCACCCCAC
CATTGTAATTCTCATCACAAACCAATAANCCCATTGGGTGCAGAAATTTGTAGCATCTGTTTCTTGCTTA
ACTTATTAGGCCTTAGTGTTATACTTAGCAAGGAAGATGTAACAATATTTGTCATATTCTCTTCATTTGT
CTTTGTAGACAAACATGTTCATTAGTGTTACTTGATGTATGGGCTGGTTGGATATAAAAAAAAAAAAAAA
AAACTCGAG

ORF

CCTCATGAACCAGGCAATTTGAAAGTAAAATCGTACAGAAAAGTTAAATCTAATAGAAGGCAGGGTGTGG
TATCCAGTAAACTTGCATCTTTGGCCGCAGAATTTCCTCTTGCAAGATTACATGGTCCCGTTAATGCCGA
CCTCAACGAGATGTGGGAGAGGAAATGTTGTGCTAAGCAAGTCGACTCACAGTCTCCACCACTATTTGAG
AAGCTCCGGGAATCACTTCTTCATGGGGTGAACATTGACTATGATGATTTCTGTACCCCAAATGAAAAGA
ATGAAGACAATGACTATGATAGTGCTGATCATGATTTTGGGCCTCCTGATTTTGACATGCCAGAAAATGC
AGACATGAACAGCCATGCTACTCCACATGATGAAAAGCATGATAATTGTGGTCCACTTTTTGATAGTGAA
GCTCATGAAGATCTGAATGGTCAAGAAAACCTTGAAGATCTTTGTCGCTCCCACTTGGATGCTCTTCTTG
CTAACCTTGCTGAAACTGAGAAGCAGACTGAATTGGCTGCTCGGGTTTCAACGTGGAAACAGAGAATTGA
CCAGGAACTTGGAGAACAAGAATCACATCCACCCTTTGACATTCATGAATATGGGGCAAGGGTTTTGAAC
AAGTTATCCCTGGAAGAAAATGATAAAAGCACCATGTCCTTTTCTGATGTTGTCAAGGGTTCGGAGAAGC
ATGACATTGCTCGAACATTTTCTGCGCTTCTGCAATTGGTAAACAATGGAGACGTTGCTTTGGAAAGAGG
TGAGGTAGGCGAGTCCACTTGTTACACAGCTGCAAATCCCTTCTCTGTTCAGCTCCTTAGGCATGGCAAC
GATAGGGAGGAAATGCAGTTTCAATCAACAAAAAAGAGAGCAGAATCTCCAATGCACCATCAGAACAATA
GAAAGGAAAAGAACAAAGGTAAAGCCGTTCATGCTGCTGTTGATTCATCGCCTCCAGGACCCGATTCAGA
TAGCAGATTACCCCTGAAGCTGGGAAAGGTTAATGGGACAAGATGCACACCTGATAGCAAGAAAAGAAGG
AAGTCCAGAATAGCGTTAGCATCGGATGTGCCTACTGCATTGTAG

Protein

PHEPGNLKVKSYRKVKSNRRQGVVSSKLASLAAEFPLARLHGPVNADLNEMWERKCCAKQVDSQSPPILFE
KLRESLLHGVNIDYDDFCTPNEKNEDNDYDSADHDFGPPDFDMPENADMNSHATPHDEKHDNCGPLFDSE
AHEDLNGQENLEDLCRSHLDALLANLAETEKQTELAARVSTWKQRIDQELGEQESHPPFDIHEYGARVLN
KLSLEENDKSTMSFSDVVKGSEKHDIARTFSALLQLVNNGDVALERGEVGESTCYTAANPFSVQLLRHGN
DREEMQFQSTKKRAESPMHHQNNRKEKNKGKAVHAAVDSSPPGPDSDSRLPLKLGKVNGTRCTPDSKKRR
KSRIALASDVPTAL*

14A

cDNA

GCAAGGCGTTCGCCGTTACTCGCCGTTAACTGACGGCGGCGGAGTAACGCTAGAGCTCACCACGACCACC
AATTCGCCGCTAGCAATCGACGTGACGGAATCGACGGAGATGAGAATCCGGCGACTGATCACCGAGAATC
CCGTCGTTATTTTCACCCGATCCGGCTGCTGCATGTGCCACGTCATGAAACGGCTGTTATTATCCGCCGT
TAGCGTTCACCCAACCGTTATTCAACTGGAAGAAGAAGAGATCGCCGCCCTCCCTGCCGGCGCCGGAGAC
GGCGCGGAGGACGGCGGTGAGATGCCGGCTATGTACATCGGAGGGACACGTGTCGGTGGATTCGAAAGCC
TCGTGGCACTTCACCTTAGTGGTGGCCTTGTTCCTAAGCTTGTGGAAGTTGGTGCGATTACTGAAATGGT
ATTGTAAAGAATTTTAATTATGTTTCTATATTTGTAATTATGTATTTTAATCAAAATTAAAAAATAATAA
TTTTAGTATATTGGTATTAATTAAGTTTTAAGGTATTGTTTAGTCTTCATAATTGTGCTTTTACAGGCTG
TAATCTCTCTATTTTATTCCAAATGACCAAAAAGTTGAATAAAAAAGAAGAGGATGTGCTCCTTTTATAT
ATATAAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

CAAGGCGTTCGCCGTTACTCGCCGTTAACTGACGGCGGCGGAGTAACGCTAGAGCTCACCACGACCACCA
ATTCGCCGCTAGCAATCGACGTGACGGAATCGACGGAGATGAGAATCCGGCGACTGATCACCGAGAATCC
CGTCGTTATTTTCACCCGATCCGGCTGCTGCATGTGCCACGTCATGAAACGGCTGTTATTATCCGCCGTT
AGCGTTCACCCAACCGTTATTCAACTGGAAGAAGAAGAGATCGCCGCCCTCCCTGCCGGCGCCGGAGACG
GCGCGGAGGACGGCGGTGAGATGCCGGCTATGTACATCGGAGGGACACGTGTCGGTGGATTCGAAAGCCT
CGTGGCACTTCACCTTAGTGGTGGCCTTGTTCCTAAGCTTGTGGAAGTTGGTGCGATTACTGAAATGGTA
TTGTAA

Protein

QGVRRYSPLTDGGGVTLELTTTTNSPLAIDVTESTEMRIRRLITENPVVIFTRSGCCMCHVMKRLLLSAV
SVHPTVIQLEEEEIAALPAGAGDGAEDGGEMPAMYIGGTRVGGFESLVALHLSGGLVPKLVEVGAITEMV
L*

C1

1A

cDNA

CCAAAACCTCCCTAAAACCCCAACTTTTCTTCTTTCTCTTTCCTACCCTTTTCTCACCACCTCTCTCTCT
TCAATTCCACCATCAACAAGTTCAAGATTCCCAGTTGTAAGAGCCATTTCTGATGGGGAATACTCTTCTA
AGAGGAGCAGTAACAGCGACGAAAGGGAAACTATTATGTTACCTGGATGTGATTATAACCACTGGCTTAT
TGTTATGGAGTTCCCTAAAGACCCTGCTCCTACTAGAGAACAGATGATTGATACTTATCTTGACACTCTA
GCCACCGTTCTTGGAAGTATGGAAGAGGCAAAGAAGAACATGTATGCCTTTAGTACTACCACTTACACTG
GATTTCAGTGCACTGTTTCGGAAGAAACATCAGAGAAGTTTAAGGGTCTACCTGGAGTTCTATGGGTCCT
GCCCGATTCTTATATAGATGTGAAGAACAAGGACTATGGAGGAGATAAGTACATCAATGGAGAAATAATT
CCTTGTCAGTATCCTACTTACCAACCCAAACAGGCTAACAGAACAAGAAGCAAGAGTAAAGCATATGTAA
GAAGAAGAGATGGTCCTCCGCCTGAACGTACAAGACGAGCAGCAGCTCCCGAGTCTTCCTCCTAAAACTA
TTATGATGCATTTTGGGTTATCTGATCTTAGAGGCGAATTGCCAATATATCATTCATAGTCGCATATTGG
TGGGCAATTTAACAAAGATATGTTGGTGCTTAAAGGTTAGATGGAAATACACTGCTAGTGCAGGTGTTTG
GTTGACTGCATTTGTTGTTTATCTTTCTTTCCTCTAGACTAGCTTTCTGGTTGCATTATTGAGAATGTGA
AAGAAAGCCGTGTCAGTATGTAGTTTTGATCTTACCATTATAATATTCTGCTGGTTCAAAGCCTGAAAAT
GAAAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

CAAAACCTCCCTAAAACCCCAACTTTTCTTCTTTCTCTTTCCTACCCTTTTCTCACCACCTCTCTCTCTT
CAATTCCACCATCAACAAGTTCAAGATTCCCAGTTGTAAGAGCCATTTCTGATGGGGAATACTCTTCTAA
GAGGAGCAGTAACAGCGACGAAAGGGAAACTATTATGTTACCTGGATGTGATTATAACCACTGGCTTATT
GTTATGGAGTTCCCTAAAGACCCTGCTCCTACTAGAGAACAGATGATTGATACTTATCTTGACACTCTAG
CCACCGTTCTTGGAAGTATGGAAGAGGCAAAGAAGAACATGTATGCCTTTAGTACTACCACTTACACTGG
ATTTCAGTGCACTGTTTCGGAAGAAACATCAGAGAAGTTTAAGGGTCTACCTGGAGTTCTATGGGTCCTG
CCCGATTCTTATATAGATGTGAAGAACAAGGACTATGGAGGAGATAAGTACATCAATGGAGAAATAATTC
CTTGTCAGTATCCTACTTACCAACCCAAACAGGCTAACAGAACAAGAAGCAAGAGTAAAGCATATGTAAG
AAGAAGAGATGGTCCTCCGCCTGAACGTACAAGACGAGCAGCAGCTCCCGAGTCTTCCTCCTAA

Protein

SQNLPKTPTFLLSLSYPFLTTSLSSIPPSTSSRFPVVRAISDGEYSSKRSSNSDERETIMLPGCDYNHWL
IVMEFPKDPAPTREQMIDTYLDTLATVLGSMEEAKKNMYAFSTTTYTGFQCTVSEETSEKFKGLPGVLWV
LPDSYIDVKNKDYGGDKYINGEIIPCQYPTYQPKQANRTRSKSKAYVRRRDGPPPERTRRAAAPESSS*

14A

cDNA

ACGGTGGATACCTAGGCACCCAGAGACGAGGAAGGGCGTAGTAATCGACGAAATGCTTCGGGGAGTTGAA
AATAAGCATAGATCCGGAGATTCCCGAATAGGGCAACCTTTCGAACTGTCCTTATCAATCCCAGTACCTA
AACCAACAAATTTCGCCGGCCGTTCATCCACTTTCCGCCGCAACTCCATCCTCCATTCCATCCGGGTCGA
GAATCCAAATTCGGGTCGGGTCAATACTATCGTGTGCAAAGCTAACCGGTCCGCATACTCGCCACTGAAC
TCCGGGTCGAACTACGGCGACCGTCCACCGACTGAGATGGCTCCGCTTTTCCCCGGGTGTGATTACGAAC
ACTGGCTTATAGTTATGGATAAGCCCGGAGGTGAAGGTGCTACTAAACAGCAGATGATTGATTGTTACAT
ACAAACATTGGCTAAAGTAGTTGGGAGTGAAGAAGAGGCCAAGAAGAAGATATACAATGTTTCATGCGAA
AGATACTTCGGTTTTGGATGTGAGATTGATGAGGAGACATCGAATAAGCTTGAAGGTTTGCCTGGTGTTC
TCTTTGTCCTACCAGATTCATATGTCGATCCTGAGAACAAGGATTATGGAGCTGAGCTATTTGTGAACGG
AGAGATAGTTCAAAGATCACCTGAAAGACAAAAGAGAGTGGAGCCAGTACCCCAGAGAGCTCAAGACAGA
CCCAGATATAACGACCGAACACGTTATGTGAGACGCCGTGAGAACACACGGTGAAGATAGTATTGCAACA
ACTATGCCTCAATCCCAAACAAGTTAGGGAAAAAGATCTT

ORF

CGGTGGATACCTAGGCACCCAGAGACGAGGAAGGGCGTAGTAATCGACGAAATGCTTCGGGGAGTTGAAA
ATAAGCATAGATCCGGAGATTCCCGAATAGGGCAACCTTTCGAACTGTCCTTATCAATCCCAGTACCTAA
ACCAACAAATTTCGCCGGCCGTTCATCCACTTTCCGCCGCAACTCCATCCTCCATTCCATCCGGGTCGAG
AATCCAAATTCGGGTCGGGTCAATACTATCGTGTGCAAAGCTAACCGGTCCGCATACTCGCCACTGAACT
CCGGGTCGAACTACGGCGACCGTCCACCGACTGAGATGGCTCCGCTTTTCCCCGGGTGTGATTACGAACA
CTGGCTTATAGTTATGGATAAGCCCGGAGGTGAAGGTGCTACTAAACAGCAGATGATTGATTGTTACATA
CAAACATTGGCTAAAGTAGTTGGGAGTGAAGAAGAGGCCAAGAAGAAGATATACAATGTTTCATGCGAAA
GATACTTCGGTTTTGGATGTGAGATTGATGAGGAGACATCGAATAAGCTTGAAGGTTTGCCTGGTGTTCT
CTTTGTCCTACCAGATTCATATGTCGATCCTGAGAACAAGGATTATGGAGCTGAGCTATTTGTGAACGGA
GAGATAGTTCAAAGATCACCTGAAAGACAAAAGAGAGTGGAGCCAGTACCCCAGAGAGCTCAAGACAGAC
CCAGATATAACGACCGAACACGTTATGTGAGACGCCGTGAGAACACACGGTGA

Protein

RWIPRHPETRKGVVIDEMLRGVENKHRSGDSRIGQPFELSLSIPVPKPTNFAGRSSTFRRNSILHSIRVE
NPNSGRVNTIVCKANRSAYSPLNSGSNYGDRPPTEMAPLFPGCDYEHWLIVMDKPGGEGATKQQMIDCYI
QTLAKVVGSEEEAKKKIYNVSCERYFGFGCEIDEETSNKLEGLPGVLFVLPDSYVDPENKDYGAELFVNG
EIVQRSPERQKRVEPVPQRAQDRPRYNDRTRYVRRRENTR*

15A

cDNA

GAATTCGGCACGAGAGCAGCAGCTATGATGGCCGGCTCAGCCACTCCCTTCCCGTCGACCTCCGTATATC
TCGCCGCTGGCGAACTGTCCTTATCAATCCCAGTACCTAAACCAACAAATTTCGCCGGCCGTTCATCCAC
TTTCCGCCGCAACTCCATCCTCCATTCCATCCGGGTCGAGAATCCAAATTCGGGTCGGGTCAATACTATC
GTGTGCAAAGCTAACCGGTCCGCATACTCGCCACTGAACTCCGGGTCGAACTACGGCGACCGTCCACCGA
CTGAGATGGCTCCGCTTTTCCCCGGGTGTGATTACGAACACTGGCTTATAGTTATGGATAAGCCCGGAGG
TGAAGGTGCTACTAAACAGCAGATGATTGATTGTTACATACAAACATTGGCTAAAGTAGTTGGGAGTGAA
GAAGAGGCCAAGAAGAAGATATACAATGTTTCATGCGAAAGATACTTCGGTTTTGGATGTGAGATTGATG
AGGAGACATCGAATAAGCTTGAAGGTTTGCCTGGTGTTCTCTTTGTCCTACCAGATTCATATGTCGATCC
TGAGAACAAGGATTATGGAGCTGAGCTATTTGTGAACGGAGAGATAGTTCAAAGATCACCTGAAAGACAA
AAGAGAGTGGAGCCAGTACCCCAGAGAGCTCAAGACAGACCCAGATATAACGACCGAACACGTTATGTGA
GACGCCGTGAGAACACACGGTGAAGATAGTATTGCAACAACTATGCCTCAATCCCAAACAAGTTAGGGAA
AAAGATCTTGATTGTGGAAAGATGATTTATCGGNTGTTCTGTTAATTATTGTTCTGTGCTAATGGAACGC
TTGATGATTTTGTTTGAAATCTTTAAAACATAGGTACTTGATAGTCATGCAAGTTCATTGATATGTAAAG
TAATTAGTTTTCTTTGTTTGTTCTACTATTTCTGCTCTTTTAAATCTTTGAGTAAATTCAATTGTGAGAT
AAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGATGGCCGGCTCAGCCACTCCCTTCCCGTCGACCTCCGTATATCTCGCCGCTGGCGAACTGTCCTTAT
CAATCCCAGTACCTAAACCAACAAATTTCGCCGGCCGTTCATCCACTTTCCGCCGCAACTCCATCCTCCA
TTCCATCCGGGTCGAGAATCCAAATTCGGGTCGGGTCAATACTATCGTGTGCAAAGCTAACCGGTCCGCA
TACTCGCCACTGAACTCCGGGTCGAACTACGGCGACCGTCCACCGACTGAGATGGCTCCGCTTTTCCCCG
GGTGTGATTACGAACACTGGCTTATAGTTATGGATAAGCCCGGAGGTGAAGGTGCTACTAAACAGCAGAT
GATTGATTGTTACATACAAACATTGGCTAAAGTAGTTGGGAGTGAAGAAGAGGCCAAGAAGAAGATATAC

AATGTTTCATGCGAAAGATACTTCGGTTTTGGATGTGAGATTGATGAGGAGACATCGAATAAGCTTGAAG
GTTTGCCTGGTGTTCTCTTTGTCCTACCAGATTCATATGTCGATCCTGAGAACAAGGATTATGGAGCTGA
GCTATTTGTGAACGGAGAGATAGTTCAAAGATCACCTGAAAGACAAAAGAGAGTGGAGCCAGTACCCCAG
AGAGCTCAAGACAGACCCAGATATAACGACCGAACACGTTATGTGAGACGCCGTGAGAACACACGGTGA

Protein

MMAGSATPFPSTSVYLAAGELSLSIPVPKPTNFAGRSSTFRRNSILHSIRVENPNSGRVNTIVCKANRSA
YSPLNSGSNYGDRPPTEMAPLFPGCDYEHWLIVMDKPGGEGATKQQMIDCYIQTLAKVVGSEEEAKKKIY
NVSCERYFGFGCEIDEETSNKLEGLPGVLFVLPDSYVDPENKDYGAELFVNGEIVQRSPERQKRVEPVPQ
RAQDRPRYNDRTRYVRRRENTR*

C2

13A

cDNA

TGTCAATTCATCTCTCTCCCGACTCAAACTCAAATTCTCTTCTCAATTTCCTACATTCTCTCCCCTGCCC
CATCACTCTTATCTCCCACTTAAAAAACTTAATTTACCACTTGTATTTGCCACGCTTCAGAACCAACACC
AACAACAATCAGCAGCCGAAGAAGTAGCACAAGAAGAATTTGAAGATTACGATGCCGACGAAACGTATGG
AGAAGTTAACAAAATCATTGGGAGTCGAGCAATTGAAGGTGGGAAAGGAATGGAGTATTTGATTGAGTGG
AAAGATGAACATGCCCCTACCTGGGTTCCCTCTAATTTCATCGCACAAGATGTTGTCGCTGAGTATGAAA
CCCCTTGGTGGAACGCTGCGAAAAAGTCAGATGAATCGGCTCTTCGGGAACTTATTGAAGCCGAGGACGA
CAGAGATGTGGATGCTGTTGATGACGATGGACGAACGGCTTTGCTCTTTGTTTCGGGACTCGGGTCGGAG
CCTTGCGTCAAGCTGCTCGCTGAAGCCGGCGCCGACGTCAATTACCGCGACAGGAGCGGCGGTTTGACGG
CTCTGCATATGGCTGCAGGATATGTGAAGCCTGGAGTCGCTAAGCTGTTAATTGAGCTCGGCGCAGACCC
CGAGGTGCAAGATTACAGAGGGCAGACGCCGCTGAGCTTAGCGAGGATGGTCTTGAATCAAACACCTAAA
GGAAACCCAATGCAATTCGCGAGGAGATTGGGATTAGAGAATGTGATTAGGGTATTGGAGGATGCAATTT
TTGAGTATGCACAAGTGGAGGAGATATTGGAGAAGAGAGGAAAAGGTGAAAATGTTGAGTATTTAGTGAA
ATGGAAGGATGGGGAGGATAATGAATGGGTTAAAGCATGGCTGATATCTGAGGATTTGGTGAGAGATTTT
GAGGCTGGATTGGAATATGCAGAGGCAGAGTGCATCTTGGAGAAGAGAGAGGGTGACAACGGGAAAGGCG
AGTACTTGGTTAAATGGACTGATATTGAGGAGGCTACTTGGGAACCAGAAGAAAATGTTGACCCCCTTCT
GATAGAAGATTTTGAAAAAGGTCAACAGAAAGTAGTAAGTTGAATTCATGGTTTCTCTTTTGGCTTGTAT
TTTTCTTTTTCTGTGGTAATATAATGCTCTTTTGTTTTATAAACTATAGTTGTGGTAATGTATGTAATTC
TGAGGTTGTCTAACAGGTTGAGGATAATTTCCTTTTAGTGATCCAAAAAAAAAAAAAAAAAAAACTCGA
G

ORF

GTCAATTCATCTCTCTCCCGACTCAAACTCAAATTCTCTTCTCAATTTCCTACATTCTCTCCCCTGCCCC
ATCACTCTTATCTCCCACTTAAAAAACTTAATTTACCACTTGTATTTGCCACGCTTCAGAACCAACACCA
ACAACAATCAGCAGCCGAAGAAGTAGCACAAGAAGAATTTGAAGATTACGATGCCGACGAAACGTATGGA
GAAGTTAACAAAATCATTGGGAGTCGAGCAATTGAAGGTGGGAAAGGAATGGAGTATTTGATTGAGTGGA
AAGATGAACATGCCCCTACCTGGGTTCCCTCTAATTTCATCGCACAAGATGTTGTCGCTGAGTATGAAAC
CCCTTGGTGGAACGCTGCGAAAAAGTCAGATGAATCGGCTCTTCGGGAACTTATTGAAGCCGAGGACGAC
AGAGATGTGGATGCTGTTGATGACGATGGACGAACGGCTTTGCTCTTTGTTTCGGGACTCGGGTCGGAGC
CTTGCGTCAAGCTGCTCGCTGAAGCCGGCGCCGACGTCAATTACCGCGACAGGAGCGGCGGTTTGACGGC
TCTGCATATGGCTGCAGGATATGTGAAGCCTGGAGTCGCTAAGCTGTTAATTGAGCTCGGCGCAGACCCC
GAGGTGCAAGATTACAGAGGGCAGACGCCGCTGAGCTTAGCGAGGATGGTCTTGAATCAAACACCTAAAG
GAAACCCAATGCAATTCGCGAGGAGATTGGGATTAGAGAATGTGATTAGGGTATTGGAGGATGCAATTTT
TGAGTATGCACAAGTGGAGGAGATATTGGAGAAGAGAGGAAAAGGTGAAAATGTTGAGTATTTAGTGAAA
TGGAAGGATGGGGAGGATAATGAATGGGTTAAAGCATGGCTGATATCTGAGGATTTGGTGAGAGATTTTG
AGGCTGGATTGGAATATGCAGAGGCAGAGTGCATCTTGGAGAAGAGAGAGGGTGACAACGGGAAAGGCGA
GTACTTGGTTAAATGGACTGATATTGAGGAGGCTACTTGGGAACCAGAAGAAAATGTTGACCCCCTTCTG
ATAGAAGATTTTGAAAAAGGTCAACAGAAAGTAGTAAGTTGA

Protein

VNSSLSRLKLKFSSQFPTFSPLPHHSYLPLKKLNLPLVFATLQNQHQQQSAAEEVAQEEFEDYDADETYG
EVNKIIGSRAIEGGKGMEYLIEWKDEHAPTWVPSNFIAQDVVAEYETPWWNAAKKSDESALRELIEAEDD
RDVDAVDDDGRTALLFVSGLGSEPCVKLLAEAGADVNYRDRSGGLTALHMAAGYVKPGVAKLLIELGADP
EVQDYRGQTPLSLARMVLNQTPKGNPMQFARRLGLENVIRVLEDAIFEYAQVEEILEKRGKGENVEYLVK
WKDGEDNEWVKAWLISEDLVRDFEAGLEYAEAECILEKREGDNGKGEYLVKWTDIEEATWEPEENVDPLL
IEDFEKGQQKVVS*

16A

cDNA

CGATTACAAGGTGGAGATGATTAATGATGGCATGCAAGAGTTTTATGTGCATTTTCATGGACCTGCTGAA
AGTCCTTATCATGGCGGTGTGTGGAAAATAAGGGTGGAACTTCCTGATGCTTACCCATATAAATCTCCAT
CCATTGGTTTTATTAATAAAATCTACCACCCAAATGTTGATGAGATGTCTGGCTCAGTTTGTTTAGATGT
TATCAATCAGACTTGGAGTCCCATGTTTGATTTGGTAAATGTGTTTGAAGTGTTTCTTCCACAACTTCTC
TTGTATCCAAATCCATCAGACCCATTGAATGGAGAGGCTGCTGCCCTTATGATGCGAGACCGGGCTGCCT
ATGAACTAAGAGTTAAAGAATTTTGTCAGAAATATGCCAAGCCTGAAGATGTTGGAGCTGCTGCCCCTGA
GGAGAAGTCAAGTGATGAGGAGTTGAGTGAAGCTGAATATGATTCAGCTGATGATGCAGTTGCAGGCCCT
GTTGATCCATGACCGACTCCCGCACTCATTGACTTTATGTATCTGTAAATGATGTTCTTACTCCTGTTTA
ATATCAAGCTGAACTACAAAGTTCAAGACAATGTTCCCCTTGTTTTATTAGTTTAGCTTCCAATGTAAAT
AATTTCCACCCTTTAAAAAAAAAAAAAAAAAAACTCGAG

ORF

AGCGATTACAAGGTGGAGATGATTAATGATGGCATGCAAGAGTTTTATGTGCATTTTCATGGACCTGCTG
AAAGTCCTTATCATGGCGGTGTGTGGAAAATAAGGGTGGAACTTCCTGATGCTTACCCATATAAATCTCC
ATCCATTGGTTTTATTAATAAAATCTACCACCCAAATGTTGATGAGATGTCTGGCTCAGTTTGTTTAGAT
GTTATCAATCAGACTTGGAGTCCCATGTTTGATTTGGTAAATGTGTTTGAAGTGTTTCTTCCACAACTTC
TCTTGTATCCAAATCCATCAGACCCATTGAATGGAGAGGCTGCTGCCCTTATGATGCGAGACCGGGCTGC
CTATGAACTAAGAGTTAAAGAATTTTGTCAGAAATATGCCAAGCCTGAAGATGTTGGAGCTGCTGCCCCT
GAGGAGAAGTCAAGTGATGAGGAGTTGAGTGAAGCTGAATATGATTCAGCTGATGATGCAGTTGCAGGCC
CTGTTGATCCATGA

Protein

SDYKVEMINDGMQEFYVHFHGPAESPYHGGVWKIRVELPDAYPYKSPSIGFINKIYHPNVDEMSGSVCLD
VINQTWSPMFDLVNVFEVFLPQLLLYPNPSDPLNGEAAALMMRDRAAYELRVKEFCQKYAKPEDVGAAAP
EEKSSDEELSEAEYDSADDAVAGPVDP*

REn

SlNAC1 (accession number AY498713)

cDNA

GAATTCGGCACGAGCAAAGCAGGAGCAGGAGCAGCAACAAACAGAGAGAAGAAAACAGAGGAAGATAAGA
GGAAAATTTATCGAATTCGAATCGAGAGAAAAGGGGAAGTGAAGTTGCGAAGAGTGAGAATTTCAAAGGA
AATGAACAAAGGAGCAAACGGAAATCAGCAATTGGAGTTACCGGCGGGATTCAGATTCCATCCGACAGAC
GACGAATTGGTGCAGCACTATCTCTGCAGGAAATGCGCCGGACAGTCGATTGCTGTATCAATTATAGCTG
AAATTGATCTTTACAAGTTTGATCCATGGCAGTTGCCTGAGAAGGCTTTGTACGGTGAAAAAGAGTGGTA
TTTTTTCTCACCAAGGGATAGAAAATATCCGAACGGTTCACGGCCGAACCGAGCAGCAGGAACCGGTTAT
TGGAAGGCAACCGGAGCTGATAAACCGGTGGGAAAACCCAAAACCTTAGGGATAAAGAAGGCACTTGTGT
TCTATGCCGGAAAAGCACCCAGAGGTATAAAAACAAATTGGATTATGCACGAGTACCGCCTCGCCAACGT
GGACCGCTCTGCTGGCAAGAACAATAACTTGAGGCTTGATGATTGGGTATTGTGTCGAATATACAACAAG
AAAGGCACACTTGAGAAGCATTACAATGTGGACAACAAGGAAACTACAAGCTTTGGAGAATTTGATGAAG
AAATAAAACCAAAAATATTGCCCACACAATTAGCACCGATGCCACCACGGCCTCGATCGACACCAGCAAA
CGACTACTTTTATTTCGAGTCATCAGAGTCGATGACTAGAATGCACACGACAAACTCGAGCTCTGGCTCA
GAGCATGTCTTGTCGCCATGTGACAAGGAGGTTCAGAGCGCGCCCAAATGGGACGAAGACCACAGAAACA
CCCTTGATTTTCAGCTAAACTATTTGGATGGTTTACTAAATGAACCATTTGAAACCCAAATGCAGCAGCA
AATTTGCAACTTTGACCAGTTCAACAATTTCCAAGACATGTTCCTATACATGCAAAAACCTTACTAAAAT
TGTATAAATTCATTGGATCTAAATTGAGTGTGATCCATGACATTTTCTTTGTTCTTTGGTGGTGTAGGTC
AACTTTTTATTAAGTAGTTTAGAGAAGTACAAAATGCTAGTCAAATTTGGTGGGCTACAGCACAAATGAG
CCTTGATAAGCATAGCCAAAGAGTCGTATAGAAGGGCTTATTATTATTGTAAGGTATGTAAAAACAAATG
AAAATTTGTTAATATCAAGTTATCATTCTTCAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGAACAAAGGAGCAAACGGAAATCAGCAATTGGAGTTACCGGCGGGATTCAGATTCCATCCGACAGACG
ACGAATTGGTGCAGCACTATCTCTGCAGGAAATGCGCCGGACAGTCGATTGCTGTATCAATTATAGCTGA
AATTGATCTTTACAAGTTTGATCCATGGCAGTTGCCTGAGAAGGCTTTGTACGGTGAAAAAGAGTGGTAT
TTTTTCTCACCAAGGGATAGAAAATATCCGAACGGTTCACGGCCGAACCGAGCAGCAGGAACCGGTTATT
GGAAGGCAACCGGAGCTGATAAACCGGTGGGAAAACCCAAAACCTTAGGGATAAAGAAGGCACTTGTGTT
CTATGCCGGAAAAGCACCCAGAGGTATAAAAACAAATTGGATTATGCACGAGTACCGCCTCGCCAACGTG
GACCGCTCTGCTGGCAAGAACAATAACTTGAGGCTTGATGATTGGGTATTGTGTCGAATATACAACAAGA
AAGGCACACTTGAGAAGCATTACAATGTGGACAACAAGGAAACTACAAGCTTTGGAGAATTTGATGAAGA

AATAAAACCAAAAATATTGCCCACACAATTAGCACCGATGCCACCACGGCCTCGATCGACACCAGCAAAC
GACTACTTTTATTTCGAGTCATCAGAGTCGATGACTAGAATGCACACGACAAACTCGAGCTCTGGCTCAG
AGCATGTCTTGTCGCCATGTGACAAGGAGGTTCAGAGCGCGCCCAAATGGGACGAAGACCACAGAAACAC
CCTTGATTTTCAGCTAAACTATTTGGATGGTTTACTAAATGAACCATTTGAAACCCAAATGCAGCAGCAA
ATTTGCAACTTTGACCAGTTCAACAATTTCCAAGACATGTTCCTATACATGCAAAAACCTTACTAA

Protein

MNKGANGNQQLELPAGFRFHPTDDELVQHYLCRKCAGQSIAVSIIAEIDLYKFDPWQLPEKALYGEKEWY
FFSPRDRKYPNGSRPNRAAGTGYWKATGADKPVGKPKTLGIKKALVFYAGKAPRGIKTNWIMHEYRLANV
DRSAGKNNNLRLDDWVLCRIYNKKGTLEKHYNVDNKETTSFGEFDEEIKPKILPTQLAPMPPRPRSTPAN
DYFYFESSESMTRMHTTNSSSGSEHVLSPCDKEVQSAPKWDEDHRNTLDFQLNYLDGLLNEPFETQMQQQ
ICNFDQFNNFQDMFLYMQKPY*

C4

1A

cDNA

GGTGAGAAGCTTGTCGACGAATTCAGATTGGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGGTGAGA
TTCCTACTAAGATCACAGGTATGAGAATACTCAACTACTTGAATTTATCGCGAAACCACTTAGTTGGGAG
TATTCCTGCCCCTATTTCTAGTATGCAGAGTTTAACTTCTGTTGATTTCTCGTATAACAACTTTTCTGGT
TTAGTTCCTGGAACCGGGCAATTTAGTTATTTCAATTACACCTCATTTCTAGGCAATCCAGATCTTTGCG
GACCCTATTTGGGCCCTTGCAAAGAGGGCGTTGTTGATGGGGTTAGTCAACCTCACCAACGAGGAGCCTT
AACGCCTTCGATGAAGCTTTTACTTGTTATAGGTTTGCTTGTCTGTTCTATTGTGTTTGCTGTTGCTGCA
ATTATAAAGGCCCGATCTTTAAAGAAGGCAAGTGAAGCTCGTGCCTGGAAGCTAACTGCTAACTGCTTTT
CAGCGCCTGGATTTTACTTGTGATGATATTTTGGATAGCTTGAAGGAGGATAACGTTATTGGAAAAGGAG
GTGCTGGTATTGTCTACAAGGGGGTAATGCCTAGCGGGGAACATGTAGCGGTTAAGAGGTTGCCAGCTAT
GAGCAGGGGTTCCTCTCATGATCATGGGTTCAATGCAGAGATACAGACTCTTGGGAGGATCCGACACAGG
CACATTGTTAGATTATTAGGGTTTTGCTCGAATCATGAGACAAATCTTTTGGTTTACGAGTACATGCCTA
ATGGAAGTCTTGGGGAAATGCTTCATGGCAAGAAAGGCGGTCATTTACATTGGGATACCAGGTATAAGAT
TGCCTTGGAGTCTGCTAAGGGTCTTTGCTATCTCCATCACGATTGCTCTCCTTTGATCCTCCATCGTGAT
GTGAAATCAAACAACATTCTGCTGGACTCCAGCTTTGAAGCTCATGTTGCTGATTTTGGACTTGCTAAGT
TCTTGCAAGATTCAGGGACATCAGAATGCATGTCTGCTATTGCTGGTTCTTATGGGTACATTGCTCCAGA
ATATGCTTACACACTTAAGGTTGATGAGAAGAGTGATGTATATAGCTTCGGTGTGGTGCTACTAGAACTG
GTAAGTGGCCAAAAACCAGTTGGAGAATTTGGTGATGGTGTTGACATAGTCCAATGGGTTAGGAAAATGA
CTGATGGGAAAAAGGATGGAGTTCTCAAGATCCTTGACCCAAGACTCTCAACGGTTCCCCTTAATGAGGT
GATGCATGTCTTCTATGTCGCATTGTTGTGTGTCGAAGAGCAGGCTGTGGAACGTCCCACCATGCGAGAG
GTAGTGCAAATACTAACGGAACTTCCCAAGCCACCAGGTGCAAAATCAGATGACTCAACCGTCACTGATC
AGTCGCCCCCATCAGCCTCTGCATTAGAGTCCCCAACCTCAATTCCCGGGGACACAAAAGACCATCATCA
ACCAACACCTCAATCACCTCCACCTGACCTACTCAGTATCTAATTTGCAATGTTCTTGAAGTAGGAGTGT
TTTATTTAGTTTGATTCTCTAGTTCTATTATGATCAATTGTGCTAAGCTTTATTCCTTTGTTTTAAAAAA
ATTGGGTCTTTCTAGGCTCGGGGGTTTATTCTAACTCTAAGATGGGTTTAATGCTCAGAAGTTTTCCTCT
TGTACAGTAAGATTGGTAGGGTTTTCAAGTGTATTATTAAATGGAAAAAAATTGCCCTTCAAAAAAAAAA
AAAAAAAAAAAAAAACTCGAG

ORF

GGTGAGATTCCTACTAAGATCACAGGTATGAGAATACTCAACTACTTGAATTTATCGCGAAACCACTTAG
TTGGGAGTATTCCTGCCCCTATTTCTAGTATGCAGAGTTTAACTTCTGTTGATTTCTCGTATAACAACTT
TTCTGGTTTAGTTCCTGGAACCGGGCAATTTAGTTATTTCAATTACACCTCATTTCTAGGCAATCCAGAT
CTTTGCGGACCCTATTTGGGCCCTTGCAAAGAGGGCGTTGTTGATGGGGTTAGTCAACCTCACCAACGAG
GAGCCTTAACGCCTTCGATGAAGCTTTTACTTGTTATAGGTTTGCTTGTCTGTTCTATTGTGTTTGCTGT
TGCTGCAATTATAAAGGCCCGATCTTTAAAGAAGGCAAGTGAAGCTCGTGCCTGGAAGCTAACTGCTTTT
CAGCGCCTGGCATTTTACTTGTGATGATATTTTGGATAGCTTGAAGGAGGATAACGTTATTGGAAAAGGAG
GTGCTGGTATTGTCTACAAGGGGGTAATGCCTAGCGGGGAACATGTAGCGGTTAAGAGGTTGCCAGCTAT
GAGCAGGGGTTCCTCTCATGATCATGGGTTCAATGCAGAGATACAGACTCTTGGGAGGATCCGACACAGG
CACATTGTTAGATTATTAGGGTTTTGCTCGAATCATGAGACAAATCTTTTGGTTTACGAGTACATGCCTA
ATGGAAGTCTTGGGGAAATGCTTCATGGCAAGAAAGGCGGTCATTTACATTGGGATACCAGGTATAAGAT
TGCCTTGGAGTCTGCTAAGGGTCTTTGCTATCTCCATCACGATTGCTCTCCTTTGATCCTCCATCGTGAT
GTGAAATCAAACAACATTCTGCTGGACTCCAGCTTTGAAGCTCATGTTGCTGATTTTGGACTTGCTAAGT
TCTTGCAAGATTCAGGGACATCAGAATGCATGTCTGCTATTGCTGGTTCTTATGGGTACATTGCTCCAGA
ATATGCTTACACACTTAAGGTTGATGAGAAGAGTGATGTATATAGCTTCGGTGTGGTGCTACTAGAACTG
GTAAGTGGCCAAAAACCAGTTGGAGAATTTGGTGATGGTGTTGACATAGTCCAATGGGTTAGGAAAATGA

CTGATGGGAAAAAGGATGGAGTTCTCAAGATCCTTGACCCAAGACTCTCAACGGTTCCCCTTAATGAGGT
GATGCATGTCTTCTATGTCGCATTGTTGTGTGTCGAAGAGCAGGCTGTGGAACGTCCCACCATGCGAGAG
GTAGTGCAAATACTAACGGAACTTCCCAAGCCACCAGGTGCAAAATCAGATGACTCAACCGTCACTGATC
AGTCGCCCCCATCAGCCTCTGCATTAGAGTCCCCAACCTCAATTCCCGGGGACACAAAAGACCATCATCA
ACCAACACCTCAATCACCTCCACCTGACCTACTCAGTATCTAA

Protein

GEIPTKITGMRILNYLNLSRNHLVGSIPAPISSMQSLTSVDFSYNNFSGLVPGTGQFSYFNYTSFLGNPD
LCGPYLGPCKEGVVDGVSQPHQRGALTPSMKLLLVIGLLVCSIVFAVAAIIKARSLKKASEARAWKLTAF
QRLDFTCDDILDSLKEDNVIGKGGAGIVYKGVMPSGEHVAVKRLPAMSRGSSHDHGFNAEIQTLGRIRHR
HIVRLLGFCSNHETNLLVYEYMPNGSLGEMLHGKKGGHLHWDTRYKIALESAKGLCYLHHDCSPLILHRD
VKSNNILLDSSFEAHVADFGLAKFLQDSGTSECMSAIAGSYGYIAPEYAYTLKVDEKSDVYSFGVVLLEL
VSGQKPVGEFGDGVDIVQWVRKMTDGKKDGVLKILDPRLSTVPLNEVMHVFYVALLCVEEQAVERPTMRE
VVQILTELPKPPGAKSDDSTVTDQSPPSASALESPTSIPGDTKDHHQPTPQSPPPDLLSI*

5A

cDNA

AAGAGAGCACATGGCCTCGATACCGCTGGGACCTCAGCACCATAATCCGCCGGAAAATCACCATCACCAC
CACCTTCAGCCGCCGCCGCAGCTCGTGCAGCAGCAGCAGCAACAGCCTGTGAATCGTGAACTAGGAGGAG
GAGGAGGAGCTGCAGTAAGGAACGCCGGAGCACGACCGGAAATGGAATCCGAAAAGGAAATGTCAGCTGC
TGTTGTTGAGGGTAATGGTGCCGTCACTGGTCACATAATTTCCACCACCATTGGAGGCAAGAATGGAGAA
CCAAAAAGGACCATCAGTTATATGGCAGAGCGAGTTGTCGGTACAGGGTCCTTTGGGATAGTGTTTCAGG
CAAAATGCTTGGAAACTGGAGAGACTGTGGCCATAAAGAAGGTTTTACAGGACAAACGGTATAAAAATCG
TGAACTACAACTGATGCGCTTGATGGATCACCCAAATGTCATTACTCTAAAGCACTGCTTCTTTTCCACG
ACTAGTAGAGATGAGCTTTTTCTTAATTTGGTCATGGATTATGTCCCTGAAAGTTTATACAAGGTTTTAA
AGCACTATAGCAATTCAAATCAAAGGATGCCACTCATATATGTCAAACTTTACATGTATCAGATATTCAG
AGGGCTGGCTTACATTCATAATGTTCCAAGGATTTGCCATAGAGATGTGAAACCTCAAAATCTTTTGGTT
GATCCTCTGACCCATCAAGTCAAGCTGTGTGATTTTGGAAGTGCAAAAGTCCTGGTAAATGGTGAAGCAA
ATATTTCATACATTTGCTCTCGCTACTACAGAGCTCCAGAACTCATATTTGGTGCCACAGAGTATACAAC
ATCAATTGATATTTGGTCAGCAGGCTGTGTCCTTGCTGAGCTTCTTCTGGGGCAGCCGCTCTTTCCTGGC
GAAAATGCAGTAGATCAACTGGTGGAGATCATCAAGGTCCTTGGTACTCCTACTCGGGAAGAAATTCGAT
GTATGAACCCAAACTACACAGATTTCAGATTCCCACAGATAAAAGCTCATCCTTGGCACAAGGTATTCCA
TAAAAGAATGCCTCCTGAAGCAATTGATCTTGCCTCACGGCTTCTTCAATATTCACCAAGTCTTCGCTGT
ACTGCACTAGAAGCATGTGCACATTCGTTCTTTGATGAGCTTCGTGAGCCCAATGCCCGTCTCCCTAATG
GACGTCCATTTCCACCTCTTTTCAACTTTAAACAAGAGTTAACTGGAGCTTCACCTGATTTGGTCAACAA
GCTGATCCCTGAGCATGTGTGGAGGCAACTTGGTTTGAATTTCCCATTTCCTGGTGCGACGTAATTGTAC
TAATGTAATATAGAGTTGGTGTATCAGAAAAAATATATAGGGAGTCCTTGTGGATTAATGTAACCATGCA
TTTTGGCGTAACCAGTGGCTATGCACGCATGTAAAAGTTGTATGCATATAGCCGCTGTTTGGCTTACAAG
AAGTGATCCTGTCTCTACTTGTGTCCTCTTTTAGCTAATAGCAGACCTGTTGCCTCTTTTGTACATTGTT
TTTGATAAGAAGAATTCAGATAGGAGTATAAAGTGATAAATATTGTAATGTCTAGTTAAGCAGATTTCCT
TGAGAGACCTTCTGGAGCACTAGGACTTGTAGCTTTGGGTTGTAGTTATATTTATATGCAAGTTGTTTTT
AGCTCTAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGGAATCCGAAAAGGAAATGTCAGCTGCTGTTGTTGAGGGTAATGGTGCCGTCACTGGTCACATAATTT
CCACCACCATTGGAGGCAAGAATGGAGAACCAAAAAGGACCATCAGTTATATGGCAGAGCGAGTTGTCGG
TACAGGGTCCTTTGGGATAGTGTTTCAGGCAAAATGCTTGGAAACTGGAGAGACTGTGGCCATAAAGAAG
GTTTTACAGGACAAACGGTATAAAAATCGTGAACTACAACTGATGCGCTTGATGGATCACCCAAATGTCA
TTACTCTAAAGCACTGCTTCTTTTCCACGACTAGTAGAGATGAGCTTTTTCTTAATTTGGTCATGGATTA
TGTCCCTGAAAGTTTATACAAGGTTTTAAAGCACTATAGCAATTCAAATCAAAGGATGCCACTCATATAT
GTCAAACTTTACATGTATCAGATATTCAGAGGGCTGGCTTACATTCATAATGTTCCAAGGATTTGCCATA
GAGATGTGAAACCTCAAAATCTTTTGGTTGATCCTCTGACCCATCAAGTCAAGCTGTGTGATTTTGGAAG
TGCAAAAGTCCTGGTAAATGGTGAAGCAAATATTTCATACATTTGCTCTCGCTACTACAGAGCTCCAGAA
CTCATATTTGGTGCCACAGAGTATACAACATCAATTGATATTTGGTCAGCAGGCTGTGTCCTTGCTGAGC
TTCTTCTGGGGCAGCCGCTCTTTCCTGGCGAAAATGCAGTAGATCAACTGGTGGAGATCATCAAGGTCCT
TGGTACTCCTACTCGGGAAGAAATTCGATGTATGAACCCAAACTACACAGATTTCAGATTCCCACAGATA
AAAGCTCATCCTTGGCACAAGGTATTCCATAAAAGAATGCCTCCTGAAGCAATTGATCTTGCCTCACGGC
TTCTTCAATATTCACCAAGTCTTCGCTGTACTGCACTAGAAGCATGTGCACATTCGTTCTTTGATGAGCT
TCGTGAGCCCAATGCCCGTCTCCCTAATGGACGTCCATTTCCACCTCTTTTCAACTTTAAACAAGAGTTA
ACTGGAGCTTCACCTGATTTGGTCAACAAGCTGATCCCTGAGCATGTGTGGAGGCAACTTGGTTTGAATT
TCCCATTTCCTGGTGCGACGTAA

Protein

MESEKEMSAAVVEGNGAVTGHIISTTIGGKNGEPKRTISYMAERVVGTGSFGIVFQAKCLETGETVAIKK
VLQDKRYKNRELQLMRLMDHPNVITLKHCFFSTTSRDELFLNLVMDYVPESLYKVLKHYSNSNQRMPLIY
VKLYMYQIFRGLAYIHNVPRICHRDVKPQNLLVDPLTHQVKLCDFGSAKVLVNGEANISYICSRYYRAPE
LIFGATEYTTSIDIWSAGCVLAELLLGQPLFPGENAVDQLVEIIKVLGTPTREEIRCMNPNYTDFRFPQI
KAHPWHKVFHKRMPPEAIDLASRLLQYSPSLRCTALEACAHSFFDELREPNARLPNGRPFPPLFNFKQEL
TGASPDLVNKLIPEHVWRQLGLNFPFPGAT*

6A

cDNA

CCGCAACATCTCCGGTGAGTCAAACACCTACTTCATCACCCACAATTTCTCCGGCAACATCTCCGGCGAC
TTCTCCGGCAACATCACCATCAACTTCTCCGACAACATCACCATCAACATCTCCGGCGACGTCGCCGGCA
ACATCTCCGGCGACGTCGCCGGCAACATCTCCGGCGACTACTCCGGCGCCGGCGAGTATATCGCCATCGG
GGGCGATAATTTCACCATCTCCATCGGGTAATTCATCGTCATCGTCTCCAAGTTCACCGAGCACACCAGG
ATCGCCGGTGGAATCACCGCCGGGAGGCAGCGGGAACAGTCCGCCGGCGGATATTCCGGCACCTACCGGC
GTCAAAGAACTCGGCAGTAAAAGCATATTCTACAACTTCAGCTGTTTTTGTGTCAATTTTATTTACTATA
ATATTTTTTATTTCTGTTTAGCGGAAAATAGATAAATTTATTAAAACATATAAAAATAAAAAAAAAG

ORF

GCAACATCTCCGGTGAGTCAAACACCTACTTCATCACCCACAATTTCTCCGGCAACATCTCCGGCGACTT
CTCCGGCAACATCACCATCAACTTCTCCGACAACATCACCATCAACATCTCCGGCGACGTCGCCGGCAAC
ATCTCCGGCGACGTCGCCGGCAACATCTCCGGCGACTACTCCGGCGCCGGCGAGTATATCGCCATCGGGG
GCGATAATTTCACCATCTCCATCGGGTAATTCATCGTCATCGTCTCCAAGTTCACCGAGCACACCAGGAT
CGCCGGTGGAATCACCGCCGGGAGGCAGCGGGAACAGTCCGCCGGCGGATATTCCGGCACCTACCGGCGT
CAAAGAACTCGGCAGTAAAAGCATATTCTACAACTTCAGCTGTTTTTGTGTCAATTTTATTTACTATAAT
ATTTTTTATTTCTGTTTAG

Protein

ATSPVSQTPTSSPTISPATSPATSPATSPSTSPTTSPSTSPATSPATSPATSPATSPATTPAPASISPSG
AIISPSPSGNSSSSSPSSPSTPGSPVESPPGGSGNSPPADIPAPTGVKELGSKSIFYNFSCFCVNFIYYN
IFYFCL*

8B

cDNA

TGCAGTAAGGTCCTTGAGTTCTACCATAGCCAAAAACTTCAAGTCCCTACAAAATCAACAAGCTGCTTTT
TCAACTTCATTACTATTGGATGATACTCAAAAACAGTTTAAAGAAAGTGTAGCAAAATTTGCTCAAGAGA
ATATAGCTCCTTATGCTGAAAAGATTGATAGAACAAACAGTTTCCCAAAGGAGATTAACTTGTGGAAATT
GATGGGAGACTTTAATTTACATGGAATTACAGCACCAGAGGAATATGGTGGTCTCAACCTTGGATATTTA
TATCATTGCATTGCCTTAGAAGAAATTAGTCGTGCATCTGGTGCTGTTGCTGTTTCCTATGGTGTTCAAT
CCAACGTTTGCATTAACCAATTGGTTAGAAATGGAACCCCTGAGCAGAAACAAAAATATTTGCCAAAGCT
TATAAGTGGGGATCACATTGGGGCTCTAGCCATGAGTGAACCAAATGCTGGGTCAGATGTTGTTAGCATG
AAGTGTCGAGCCGATCGTGTCGATGGTGGCTATGTTTTAAATGGAAATAAAATGTGGTGCACCAATGGTC
CTATCGCTAATACTTTGATTGTTTATGCAAAAACGGATACTACTGCTGGTTCTAAAGGAATTACAGCATT
TATAATCGAAAAAGAAATGTCAGGATTCTCAACTGCACAAAAATTGGACAAACTTGGGATGAGAGGAAGC
GATACGTGTGAACTTGTATTCGAGAATTGCTTTGTTCCTAAGGAAAATGTTTTAGGCCACGAAAAAGGAG
TGTATGTTTTAATGTCGGGGCTAGATTTGGAACGACTTATTTTAGCAGCAGGACCTATTGGAATAATGCA
AGCATGTATGGATATTGTTATTCCTTACGTTCAACAAAGGGAGCAATTTGGAAGACCAATCGGCGAACTT
CAACTTATACAGGGAAAACTAGCTGACATGTATACTACTTTACAATCTTCAAGATCATATCGGTATGCTG
TTGCAAAGGACTGTGACAATGGGAAGATTGATCCAAAGGATTGTTCTGGGACTATACTACTGGCAGCTGA
AAGAGCCACTCAAGTAGCTCTCCAGGCAATTCAATGTCTAGGTGGAAATGGATATATAAATGAGTATCCA
ACAGGACGTTTACTGCGAGATGCCAAAATGTATGAGATTGCAGCAGGAACTAGTGAAATAAGAAGAATTA
TAATTGGTCGTGAGCTATTTAAACATCAATAATTAGTTACAAATTAAAATTAATTAGGGGATCCTACATG
ATTTAATTTCTTTAAAAAAAAAATTGTATGTTCAATCAACAAATTATATGTCAGCGTATGTTGTTCTCAA
TGTTTAAAGAAAAGAAGAGTCAAAACTAATAGCACTACTGTTATCNGAAAAAAAAAAAAAAAAAATCTCG
AG

ORF

GCAGTAAGGTCCTTGAGTTCTACCATAGCCAAAAACTTCAAGTCCCTACAAAATCAACAAGCTGCTTTTT
CAACTTCATTACTATTGGATGATACTCAAAAACAGTTTAAAGAAAGTGTAGCAAAATTTGCTCAAGAGAA
TATAGCTCCTTATGCTGAAAAGATTGATAGAACAAACAGTTTCCCAAAGGAGATTAACTTGTGGAAATTG
ATGGGAGACTTTAATTTACATGGAATTACAGCACCAGAGGAATATGGTGGTCTCAACCTTGGATATTTAT
ATCATTGCATTGCCTTAGAAGAAATTAGTCGTGCATCTGGTGCTGTTGCTGTTTCCTATGGTGTTCAATC
CAACGTTTGCATTAACCAATTGGTTAGAAATGGAACCCCTGAGCAGAAACAAAAATATTTGCCAAAGCTT
ATAAGTGGGGATCACATTGGGGCTCTAGCCATGAGTGAACCAAATGCTGGGTCAGATGTTGTTAGCATGA

AGTGTCGAGCCGATCGTGTCGATGGTGGCTATGTTTTAAATGGAAATAAAATGTGGTGCACCAATGGTCC
TATCGCTAATACTTTGATTGTTTATGCAAAAACGGATACTACTGCTGGTTCTAAAGGAATTACAGCATTT
ATAATCGAAAAAGAAATGTCAGGATTCTCAACTGCACAAAAATTGGACAAACTTGGGATGAGAGGAAGCG
ATACGTGTGAACTTGTATTCGAGAATTGCTTTGTTCCTAAGGAAAATGTTTTAGGCCACGAAAAAGGAGT
GTATGTTTTAATGTCGGGGCTAGATTTGGAACGACTTATTTTAGCAGCAGGACCTATTGGAATAATGCAA
GCATGTATGGATATTGTTATTCCTTACGTTCAACAAAGGGAGCAATTTGGAAGACCAATCGGCGAACTTC
AACTTATACAGGGAAAACTAGCTGACATGTATACTACTTTACAATCTTCAAGATCATATCGGTATGCTGT
TGCAAAGGACTGTGACAATGGGAAGATTGATCCAAAGGATTGTTCTGGGACTATACTACTGGCAGCTGAA
AGAGCCACTCAAGTAGCTCTCCAGGCAATTCAATGTCTAGGTGGAAATGGATATATAAATGAGTATCCAA
CAGGACGTTTACTGCGAGATGCCAAAATGTATGAGATTGCAGCAGGAACTAGTGAAATAAGAAGAATTAT
AATTGGTCGTGAGCTATTTAAACATCAATAA

Protein

AVRSLSSTIAKNFKSLQNQQAAFSTSLLLDDTQKQFKESVAKFAQENIAPYAEKIDRTNSFPKEINLWKL
MGDFNLHGITAPEEYGGLNLGYLYHCIALEEISRASGAVAVSYGVQSNVCINQLVRNGTPEQKQKYLPKL
ISGDHIGALAMSEPNAGSDVVSMKCRADRVDGGYVLNGNKMWCTNGPIANTLIVYAKTDTTAGSKGITAF
IIEKEMSGFSTAQKLDKLGMRGSDTCELVFENCFVPKENVLGHEKGVYVLMSGLDLERLILAAGPIGIMQ
ACMDIVIPYVQQREQFGRPIGELQLIQGKLADMYTTLQSSRSYRYAVAKDCDNGKIDPKDCSGTILLAAE
RATQVALQAIQCLGGNGYINEYPTGRLLRDAKMYEIAAGTSEIRRIIIGRELFKHQ*

11A

cDNA

GGGTTTAAGATTTGTTACATTTGGTGAAGAAGGAAGAATGGATAAAGCAATTATTAAGCTTGAAGAAGGT
ACAACATGTACTGCGTGGAATTATTCGGGTCACAGATTAGCTGCTGGTTCAACTGATGGTACTTTGTTTG
TTTTTGATTCTACTGATCCGGCTTCATCTGTTTTCAGTGGTTCTTCTAAATTCAAGGTGCATGAATCCAG
CATTGTAAAAGTTGTTTGGGCTCCACCAGAATATGGAGATGTAGTTGCATGCATTTGTGCTGATGGAAGT
TTGTTGTTGTGGGAGGAGGTAGTTGAAGATTCAGAGCTGCTTCAGTGGAAGCTGTGCAAATGCTTTGACA
GAATCTCATCCCTAGTTCTGGATGTTCAGTTTGGAGTCTCCCAGACAAGTCTGAAATTGGTTGCTGCTTA
TTCGGATGGCCAAGTGAAAGTGTTTGAGCTCCTAGATCCATTTGAATTGAAGAATTGGCAGCTGCAGGCT
GAATTTCAGAATGTTATTGAATCTGTATCTAAATTTGGAAATGTTTCATGCCCGATCGGCTTCTATTGCT
TGGAATCCCTTGAAAGGAGAAATCCAGCAATCAAAGCTTCGTTTTGGGTTTTGATTCCCGATACGCCACA
CCTAAACTCCTCCAAGGTTTGGGAATTCGATCAGGATCATCAGAGATGGCTTCCAGTTGCAGAATTAGCT
TTACCTGCAGATAAAGCTGATCCAGTCTCTACTGTTGCATGGGCACCAAATATTGGAAGGCCGTATGAAT
TAATAGCAGTTGCCACTTGCAAGGAAATTGCATTATGGCATGTTGGATCAAATCCAGACTCTGATGGAAG
GCTCTCGGTGGAGAAGGTTGCAATGCTCTCTACTCATGACAGTGAGGTATGGCAGATGGAATGGGACATG
AGTGGAATGACACTTGCTACTACTGGAAGCGATGGTGTAGTTCGCTTGTGGCAGTGCAACTTGAACGGAG
TTTGGCATGAGCAAGCAACATTACAGCCCACTAGCTAGCCTGCTTATATGATTTTCACATTCAAGCTTGA
ACAAGATTGCAAGAACCCAGTGAAACACAAGATTCTGTGGTTTATCCATGTGACCGATTTGTCGTCAGGC
ATTTCATTTTGTAAAGCTAAGTGACCTTTGCATTGTAAGTGGTGGAATATGAAAAATTCCCAAAATGAAT
ATGTTCATCTGATATACTGGAACTTGTTTGCATGTATGGAGAATTGTGGGGATACAGAGTTTTGATTGAA
TATGAAAAAAAACTCGTATTCGTTGTAAGATTATTGAAAAGTATGCTGAAATGTGAGCATCACTTTTGCT
TTGAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGGATAAAGCAATTATTAAGCTTGAAGAAGGTACAACATGTACTGCGTGGAATTATTCGGGTCACAGAT
TAGCTGCTGGTTCAACTGATGGTACTTTGTTTGTTTTTGATTCTACTGATCCGGCTTCATCTGTTTTCAG
TGGTTCTTCTAAATTCAAGGTGCATGAATCCAGCATTGTAAAAGTTGTTTGGGCTCCACCAGAATATGGA
GATGTAGTTGCATGCATTTGTGCTGATGGAAGTTTGTTGTTGTGGGAGGAGGTAGTTGAAGATTCAGAGC
TGCTTCAGTGGAAGCTGTGCAAATGCTTTGACAGAATCTCATCCCTAGTTCTGGATGTTCAGTTTGGAGT
CTCCCAGACAAGTCTGAAATTGGTTGCTGCTTATTCGGATGGCCAAGTGAAAGTGTTTGAGCTCCTAGAT
CCATTTGAATTGAAGAATTGGCAGCTGCAGGCTGAATTTCAGAATGTTATTGAATCTGTATCTAAATTTG
GAAATGTTTCATGCCCGATCGGCTTCTATTGCTTGGAATCCCTTGAAAGGAGAAATCCAGCAATCAAAGC
TTCGTTTTGGGTTTTGATTCCCGATACGCCACACCTAAACTCCTCCAAGGTTTGGGAATTCGATCAGGAT
CATCAGAGATGGCTTCCAGTTGCAGAATTAGCTTTACCTGCAGATAAAGCTGATCCAGTCTCTACTGTTG
CATGGGCACCAAATATTGGAAGGCCGTATGAATTAATAGCAGTTGCCACTTGCAAGGAAATTGCATTATG
GCATGTTGGATCAAATCCAGACTCTGATGGAAGGCTCTCGGTGGAGAAGGTTGCAATGCTCTCTACTCAT
GACAGTGAGGTATGGCAGATGGAATGGGACATGAGTGGAATGACACTTGCTACTACTGGAAGCGATGGTG
TAGTTCGCTTGTGGCAGTGCAACTTGAACGGAGTTTGGCATGAGCAAGCAACATTACAGCCCACTAGCTA
G

Protein

MDKAIIKLEEGTTCTAWNYSGHRLAAGSTDGTLFVFDSTDPASSVFSGSSKFKVHESSIVKVVWAPPEYG
DVVACICADGSLLLWEEVVEDSELLQWKLCKCFDRISSLVLDVQFGVSQTSLKLVAAYSDGQVKVFELLD
PFELKNWQLQAEFQNVIESVSKFGNVSCPIGFYCLESLERRNPAIKASFWVLIPDTPHLNSSKVWEFDQD

HQRWLPVAELALPADKADPVSTVAWAPNIGRPYELIAVATCKEIALWHVGSNPDSDGRLSVEKVAMLSTH
DSEVWQMEWDMSGMTLATTGSDGVVRLWQCNLNGVWHEQATLQPTS*

11B

cDNA

GAGCAATGAGCAGAAAAACTCTACTGAGCCAATTCTGGTGACGGGTTCTCCACAAGAATTCAAGATGGAG
AACCCGTTTCAGTCAGCAGGACCGAAAGAGGTGCTAGAAGTGGATACACTGAAGGATGGTATACTTGTGA
GAGTGGCAATGCCTAGCGTTGGTGAAGATGGGATTAAGGTCTGGTTAGAGAACAACACGGTTTACTTCAC
TGGCAAAGGAGACATTGAAGTGGAATCTGAGGAATCAGGGAGGAAATACGGAGGGAGTCTCGAGTTTAGC
ACTGATTGCTGTAAAGCTGAGAAGGTTGAAGCACAAATGAAAAATGGTATTCTTAGAATGGTAATTAAAG
GTGAGATGGGAGAAGATTGAATGTCTTAAGGTCTCTTTTTGTTCATCTTGTTGATATATCCTTTGTAGTA
TGAAAACTTTGTACTATTTTGGGAAATGTTTGATCAAAGGTCTGTACTACATGATGGTTGTTGGATTACT
CAATTAGTAGGGATCAAATGCTTGCAAATGTTGTTTGAGAAAAATGGTATTTCTTCTTGTGTGTCCAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

ORF

ATGGAGAACCCGTTTCAGTCAGCAGGACCGAAAGAGGTGCTAGAAGTGGATACACTGAAGGATGGTATAC
TTGTGAGAGTGGCAATGCCTAGCGTTGGTGAAGATGGGATTAAGGTCTGGTTAGAGAACAACACGGTTTA
CTTCACTGGCAAAGGAGACATTGAAGTGGAATCTGAGGAATCAGGGAGGAAATACGGAGGGAGTCTCGAG
TTTAGCACTGATTGCTGTAAAGCTGAGAAGGTTGAAGCACAAATGAAAAATGGTATTCTTAGAATGGTAA
TTAAAGGTGAGATGGGAGAAGATTGA

Protein

MENPFQSAGPKEVLEVDTLKDGILVRVAMPSVGEDGIKVWLENNTVYFTGKGDIEVESEESGRKYGGSLE
FSTDCCKAEKVEAQMKNGILRMVIKGEMGED*

15B

cDNA

TATGGCTTCCTCAAGCAAAGAGCCGATCTCCAATCCCCCAATCCCCGACGAGGAGGAGGAGGAGGAGGAA
GAAGACGATTACGGCGAAATCGATGAAGATGATGAAGAAGAAGAAGAGCGAGAGCGAAGGCCAGTAACTG
AAGAATCTCGAGTGCGTTCCGATAGGGCTAAAATGGAGAATTTATTCCGGCGGCTGTCGTCTGAGAGAGT
GCCAATAAGAGTCCATGATGTGATTATCCAAGGGAATACTAAGACTAAGGAATCACTTATTGAGGCTGAA
ATGGAAGCCCTAAAAAGTGCGACAACTCTTCAGGAGTTGCTTAAGGCTGCAAGTATTGCAAATGCGAGAC
TTCAGCATCTTGATATTTTCGATTCCGTTAAAATTACTTTGGATTCTGGTCCACCTGAGCTGCCTGGGAC
TACAAATGTGGTTGTTGAGATTGTCGAGAGTGAGAATCCACTTACTGCTTTTGCAGACCTTTCTTTTGAT
CTACCACTGCGGGTATTGAGAGAAGCTGGGATTCATGGCCATGCTTTTGCTTGTACAGGAAGCCTGAATA
AATTGACGGAAAATGCGTATAAAGATCTATCTTTACAGAAATTCAAAGAATCATTCCGAGCTTCGGCTGG
ATTTGGTGTCATTGTCCCGACCAAGCTCTTCCGGATGGAGGTTAATTACTGCTATATACTGAAGCAGCAA
GAGCATGACCGAGGGAAGACTGGTGTGCAGTTCAGCTTCTCCTCATCTTTTTAGAAGCTCAACATTGCAT
GTTCTGTTCCCAGCGAATTGATACGGCATGGTTTCCTTCCGGTAAAATTAACATAATTTCATCCTCCGTT
TCTCTGTTTTGTATGAGATGCAAATTTAGCCGGCCCCAACAAGTACACTTCACTTTCAAACTCTGGACAA
TTTTTCAATCCACTGTTACAGGCAAATAATTGGTGGCAAGTTTAGCATACATTCAATACATTGATTTTCA
ATCCGCTGCCTCAGAATTCGCCGTTTGTTATTGGTGATTCATTTCTTATTCTGGTAGTTCAGCTTTGAAA
TTGATGTTTTTGATGTGCCTATACACTTGTATTAAGAAGAGGCTCACTCAATACAGAGGACCATTAAAAA
AAATTTAACGGGCGACCTTACGAGTTGCGATTTGCTTGTAGGAGTCCTTAAAGAGAGCTTTTTTCCTTCT
GCTTGTGTTCATTTCTGAAGCTTTTGGAAATGATGGTTATCTACTTAAACCCCTATGAATTAAGTGTGTT
TTTTTTATGAAGGAAAATGGTTTCTCGAGAAGCTTTGGACTTCTTCGCCAGAGGTTGGTCAAG

ORF

ATGGCTTCCTCAAGCAAAGAGCCGATCTCCAATCCCCCAATCCCCGACGAGGAGGAGGAGGAGGAGGAAG
AAGACGATTACGGCGAAATCGATGAAGATGATGAAGAAGAAGAAGAGCGAGAGCGAAGGCCAGTAACTGA
AGAATCTCGAGTGCGTTCCGATAGGGCTAAAATGGAGAATTTATTCCGGCGGCTGTCGTCTGAGAGAGTG
CCAATAAGAGTCCATGATGTGATTATCCAAGGGAATACTAAGACTAAGGAATCACTTATTGAGGCTGAAA
TGGAAGCCCTAAAAAGTGCGACAACTCTTCAGGAGTTGCTTAAGGCTGCAAGTATTGCAAATGCGAGACT
TCAGCATCTTGATATTTTCGATTCCGTTAAAATTACTTTGGATTCTGGTCCACCTGAGCTGCCTGGGACT
ACAAATGTGGTTGTTGAGATTGTCGAGAGTGAGAATCCACTTACTGCTTTTGCAGACCTTTCTTTTGATC
TACCACTGCGGGTATTGAGAGAAGCTGGGATTCATGGCCATGCTTTTGCTTGTACAGGAAGCCTGAATAA
ATTGACGGAAAATGCGTATAAAGATCTATCTTTACAGAAATTCAAAGAATCATTCCGAGCTTCGGCTGGA
TTTGGTGTCATTGTCCCGACCAAGCTCTTCCGGATGGAGGTTAATTACTGCTATATACTGAAGCAGCAAG
AGCATGACCGAGGGAAGACTGGTGTGCAGTTCAGCTTCTCCTCATCTTTTTAG

Protein

MASSSKEPISNPPIPDEEEEEEEEDDYGEIDEDDEEEEERERRPVTEESRVRSDRAKMENLFRRLSSERV
PIRVHDVIIQGNTKTKESLIEAEMEALKSATTLQELLKAASIANARLQHLDIFDSVKITLDSGPPELPGT

TNVVVEIVESENPLTAFADLSFDLPLRVLREAGIHGHAFACTGSLNKLTENAYKDLSLQKFKESFRASAG
FGVIVPTKLFRMEVNYCYILKQQEHDRGKTGVQFSFSSSF*

26A
cDNA
GGAAGCAGCAATTGTTGACGGTAATGGGACTGAGACAGGCCACATAATAGTGACAACTATTGGTGGTAGA
AATGGCCAGCCAAAGCAGACTATTAGTTACATGGCTGAACGTGTTGTTGGACAAGGATCCTTTGGGGTGG
TGTTTCAGGCAAAATGCTTAGAGACTGGTGAAACTGTTGCTATTAAAAAGGTTCTTCAAGATAAAAGATA
TAAGAACCGAGAATTGCAGACAATGCGCCTTCTTGACCACCCTAATGTTGTGTCTCTCAAACACTGCTTC
TTTTCCACAACTGACAAGGATGAACTGTATCTTAACTTGGTTCTTGAGTATGTACCTGAAACCGTTCACC
GTGTTATCAAACACTACAATAAGCTGAATCAAAGGATGCCGTTGATACTTGTGAAGCTTTATACATATCA
GATTTTCAGGGCATTGTCTTACATACACCGCACTATTGGAGTGTGCCACAGGGACATTAAACCTCAGAAT
CTTTTGGTGAATCCACATACCCACCAGGTTAAATTATGCGATTTTGGGAGTGCCAAAGTTCTGGTTAAAG
GAGAACCAAATATTTCTTACATCTGTTCTAGGTATTATCGAGCACCTGAGCTCATATTTGGTGCAACAGA
GTACACTACTGCTATAGATGTTTGGTCTGCTGGCTGTGTCCTAGCTGAGCTACTTCTTGGTCAGCCATTA
TTTCCAGGTGAAAGTGGAGTTGATCAGCTTGTTGAGATCATTAAGGTTTTAGGCACTCCTACAAGGGAAG
AAATTAAATGTATGAATCCCAACTATAATGAGTTCAAATTTCCTCAAATTAAAGCTCATCCATGGCACAA
GATATTTCATAAGCGCATGCCTCCAGAAGCAGTTGATCTGGTCTCAAGACTGCTTCAGTACTCTCCTAAC
TTGCGCTGTGCTGCTTTGGATGCCTTGGTTCACCCATTTTTTGACGAGCTTCGTGATCCCAATACACGCT
TGCCTAATGGACGCTTTCTTCCTCCCTTGTTTAACTTCAAGGCTCATGAGTTGAAGAATGTGCATGCAGA
GATATTACTGAAGTTGGTTCCAGAGCACGCCAGAAAACAGTGCCCTTCCCTTGGGTTATGAGTTCCTGCC
TTGTGATAGAAATAATTTAAGTGTACCCTTTGAAGCTGTGTTTCTTCTCTGTATCAATTTGTTCCTTCTT
TCCCTTATCTTACCATGTCTCTGTTGTCCCTTTGTATTATTATTATTATATCTTGTTTTGTAAAAGCAGA
TGTAGGGAATTGATCTTTAACAAATACCCGATGTTCAACCTGTGTTCCTTTTTACCCTGTTCCGATATAT
CCTATGCTTGTAACATTATGTGGTTGAGAATTAACAATTTGAGTGAACGTTATTATTATGGGATTGCAAA
AAAAAAAACTCGAG

ORF
GAAGCAGCAATTGTTGACGGTAATGGGACTGAGACAGGCCACATAATAGTGACAACTATTGGTGGTAGAA
ATGGCCAGCCAAAGCAGACTATTAGTTACATGGCTGAACGTGTTGTTGGACAAGGATCCTTTGGGGTGGT
GTTTCAGGCAAAATGCTTAGAGACTGGTGAAACTGTTGCTATTAAAAAGGTTCTTCAAGATAAAAGATAT
AAGAACCGAGAATTGCAGACAATGCGCCTTCTTGACCACCCTAATGTTGTGTCTCTCAAACACTGCTTCT
TTTCCACAACTGACAAGGATGAACTGTATCTTAACTTGGTTCTTGAGTATGTACCTGAAACCGTTCACCG
TGTTATCAAACACTACAATAAGCTGAATCAAAGGATGCCGTTGATACTTGTGAAGCTTTATACATATCAG
ATTTTCAGGGCATTGTCTTACATACACCGCACTATTGGAGTGTGCCACAGGGACATTAAACCTCAGAATC
TTTTGGTGAATCCACATACCCACCAGGTTAAATTATGCGATTTTGGGAGTGCCAAAGTTCTGGTTAAAGG
AGAACCAAATATTTCTTACATCTGTTCTAGGTATTATCGAGCACCTGAGCTCATATTTGGTGCAACAGAG
TACACTACTGCTATAGATGTTTGGTCTGCTGGCTGTGTCCTAGCTGAGCTACTTCTTGGTCAGCCATTAT
TTCCAGGTGAAAGTGGAGTTGATCAGCTTGTTGAGATCATTAAGGTTTTAGGCACTCCTACAAGGGAAGA
AATTAAATGTATGAATCCCAACTATAATGAGTTCAAATTTCCTCAAATTAAAGCTCATCCATGGCACAAG
ATATTTCATAAGCGCATGCCTCCAGAAGCAGTTGATCTGGTCTCAAGACTGCTTCAGTACTCTCCTAACT
TGCGCTGTGCTGCTTTGGATGCCTTGGTTCACCCATTTTTTGACGAGCTTCGTGATCCCAATACACGCTT
GCCTAATGGACGCTTTCTTCCTCCCTTGTTTAACTTCAAGGCTCATGAGTTGAAGAATGTGCATGCAGAG
ATATTACTGAAGTTGGTTCCAGAGCACGCCAGAAAACAGTGCCCTTCCCTTGGGTTATGA

Protein
EAAIVDGNGTETGHIIVTTIGGRNGQPKQTISYMAERVVGQGSFGVVFQAKCLETGETVAIKKVLQDKRY
KNRELQTMRLLDHPNVVSLKHCFFSTTDKDELYLNLVLEYVPETVHRVIKHYNKLNQRMPLILVKLYTYQ
IFRALSYIHRTIGVCHRDIKPQNLLVNPHTHQVKLCDFGSAKVLVKGEPNISYICSRYYRAPELIFGATE
YTTAIDVWSAGCVLAELLLGQPLFPGESGVDQLVEIIKVLGTPTREEIKCMNPNYNEFKFPQIKAHPWHK
IFHKRMPPEAVDLVSRLLQYSPNLRCAALDALVHPFFDELRDPNTRLPNGRFLPPLFNFKAHELKNVHAE
ILLKLVPEHARKQCPSLGL*

βC1

SlUBC
cDNA
CCATTGGCAAGCAACAATCATGGGGCCTACAGATAGCCCTTATGCAGGAGGTGTATTTTTGGTCTCAATC
CATTTCCCCCCTGATTATCCTTTCAAGCCTCCAAAGGTGGCATTTAGAACTAAGGTTTTCCATCCTAACA
TCAATAGCAATGGAAGTATATGTTTGGATATCCTTAAAGAGCAATGGAGTCCGGCTTTGACCATATCTAA
GGTCCTGTTGTCCATTTGCTCTTTATTGACAGATCCAAACCCAGACGACCCACTTGTACCAGAAATTGCT

CATATGTACAAGACTGACAGGGCCAAATACGAGACCACTGCTCGTAGCTGGACTCAAAAATATGCGATGG
GATAACGGCATTGTCCTCGGGTGTGTCTAAGACACTTTCAATTTCAATTACTTCATTTTGCTTTCACTCC
TAGATGCAATATCTTCATTGTGCTTTGGATGAAAGAACAAAATGTTGGTAAGGAAATCAGTTTAAGTAGA
CTCCGATCATGTCAAAGTTGTCAAGACCGAGATCTTTCAATATGTGAATATCTATTTGTTGAATTTTAGT
GAGGGGAGAACCTGAAATGTAATTTTATTTTTACTCCTTCCTGACGTTTCTTGTATGGGTGCTATCTTTT
GACTGGCGAAAATTCCAGTGATTAAACAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGGCGTCGAAGCGCATATTGAAGGAGCTTAAGGATTTGCAGAAGGATCCTCCCACCTCATGCAGCGCTG
GTCCAGTGGCTGAGGATATGTTCCATTGGCAAGCAACAATCATGGGGCCTACAGATAGCCCTTATGCAGG
AGGTGTATTTTTGGTCTCAATCCATTTCCCCCCTGATTATCCTTTCAAGCCTCCAAAGGTGGCATTTAGA
ACTAAGGTTTTCCATCCTAACATCAATAGCAATGGAAGTATATGTTTGGATATCCTTAAAGAGCAATGGA
GTCCGGCTTTGACCATATCTAAGGTCCTGTTGTCCATTTGCTCTTTATTGACAGATCCAAACCCAGACGA
CCCACTTGTACCAGAAATTGCTCATATGTACAAGACTGACAGGGCCAAATACGAGACCACTGCTCGTAGC
TGGACTCAAAAATATGCGATGGGATAA

Protein

MASKRILKELKDLQKDPPTSCSAGPVAEDMFHWQATIMGPTDSPYAGGVFLVSIHFPPDYPFKPPKVAFR
TKVFHPNINSNGSICLDILKEQWSPALTISKVLLSICSLLTDPNPDDPLVPEIAHMYKTDRAKYETTARS
WTQKYAMG*

1

cDNA

AACACGTACAACACATAATAATAGTTTGACTGAGCTTTGTTTAGCTCATCCAACTACTACCAATAATGAA
GAAAATAAGAGTTTTTCAAGAAACCCTTTTCTCCAACAAAGCCAAAACATGAATAGTTATCATGGGGTTT
CTTTATTAGACCCTATAAAGGGTATCCCAATTTATCATCATCATCATCATCAAGATCCAAAGAGATCATC
ATCATTTGGATCCATTTATCATAATAATTTGGATCATATATCATATTCGAATAATTCCTATGTTACAAAT
GTTACTTCCTCTTCTCCTTATAATCGAATACCTATTGTGGCTAATAGATTTCAAAATCAACAACACATAT
ATTATAACGGAGTTGGATTATTAGGTTCACCTTCTTCTCATGAAAGTAATAATTTTTTGATGAGATCAAG
ATTTTTACCTAAATTTCCAACTAAAAGAAGCATGAGGGCACCGCGAATGCGATGGACAACTTCACTTCAT
GCTAGATTTGTTCATGCTGTTGAGCTTCTTGGTGGACATGAAAGGGCTACTCCAAAATCAGTATTGGAGC
TAATGGATGTCAAAGATCTCACTCTAGCTCATGTCAAAAGTCATTTGCAGATGTATCGCACTGTTAAAAC
TACTGACAAACCTGCAGCTTCCTCAGATGGATCAGGTGAAGAAGATTTGTTAGCTATTGACAAGATTTTG
GATCAAAGAGGTCCATTAGATGGATGTGATGAACCTTCTACAACTCCTCAACAAAGATTCAGCCATCATA
TAGAGGAATGTGAATATTCAAGAGCAATGAGCTATGTAGGGTTGCAGTTTGGATCAGAAAAACCCTAGCT
TGGAGTTCACATTAGGAAGATCAGATTGGGTAGAAAAGAACCATGACTAAAATTTGGGCATCCAACACCT
TAAAACATATTAATTTTAATTTTTTTGGGTTTTTGGTTTAAGCTGATTTTTCTAATTTTATTATTTTTTT
TGTAATTTGAGATGAGAAAGGTATGTGACATCAAGATTGAAATATTAAGATCTACTACTAAAAAAAAAAA
AAGGAATGGTTTTCTTAAGGTTCCCATGTAGCTTCAACTAAGGTTGTGACCTACCTATATATATATATAT
ATATATACAACCTAATTTTTTTTCTAGAGTTTTGTTAATTAGACTTTTTTTTTCCTTTGTTTAATTTGAG
TTCTTGGGTTGTTTATGACTATGTAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ACACGTACAACACATAATAATAGTTTGACTGAGCTTTGTTTAGCTCATCCAACTACTACCAATAATGAAG
AAAATAAGAGTTTTTCAAGAAACCCTTTTCTCCAACAAAGCCAAAACATGAATAGTTATCATGGGGTTTC
TTTATTAGACCCTATAAAGGGTATCCCAATTTATCATCATCATCATCATCAAGATCCAAAGAGATCATCA
TCATTTGGATCCATTTATCATAATAATTTGGATCATATATCATATTCGAATAATTCCTATGTTACAAATG
TTACTTCCTCTTCTCCTTATAATCGAATACCTATTGTGGCTAATAGATTTCAAAATCAACAACACATATA
TTATAACGGAGTTGGATTATTAGGTTCACCTTCTTCTCATGAAAGTAATAATTTTTTGATGAGATCAAGA
TTTTTACCTAAATTTCCAACTAAAAGAAGCATGAGGGCACCGCGAATGCGATGGACAACTTCACTTCATG
CTAGATTTGTTCATGCTGTTGAGCTTCTTGGTGGACATGAAAGGGCTACTCCAAAATCAGTATTGGAGCT
AATGGATGTCAAAGATCTCACTCTAGCTCATGTCAAAAGTCATTTGCAGATGTATCGCACTGTTAAAACT
ACTGACAAACCTGCAGCTTCCTCAGATGGATCAGGTGAAGAAGATTTGTTAGCTATTGACAAGATTTTGG
ATCAAAGAGGTCCATTAGATGGATGTGATGAACCTTCTACAACTCCTCAACAAAGATTCAGCCATCATAT
AGAGGAATGTGAATATTCAAGAGCAATGAGCTATGTAGGGTTGCAGTTTGGATCAGAAAAACCCTAG

Protein

TRTTHNNSLTELCLAHPTTTNNEENKSFSRNPFLQQSQNMNSYHGVSLLDPIKGIPIYHHHHHQDPKRSS
SFGSIYHNNLDHISYSNNSYVTNVTSSSPYNRIPIVANRFQNQQHIYYNGVGLLGSPSSHESNNFLMRSR
FLPKFPTKRSMRAPRMRWTTSLHARFVHAVELLGGHERATPKSVLELMDVKDLTLAHVKSHLQMYRTVKT
TDKPAASSDGSGEEDLLAIDKILDQRGPLDGCDEPSTTPQQRFSHHIEECEYSRAMSYVGLQFGSEKP*

6

cDNA

AACGAAGATGGCGGAATTGAGAGGACTGAAGCTGGAGCAGAGGCATGGGAAATCCAGAGTGAGAGTAGCT
AGGGTTTGGAGAGATCGAGATGGAAAACATCACTTTGCTGAATGGAGCGTTAACATCAGTCTCATCTCCG
ATTGTTTACCTGCATATGTCTCCGGCGATAATTCCGATATCGTCGCCACCGATACCATGAAGAACACTGT
ATATGCGAAAGCTAAGGAATGTTCAGAGAAACTTTCAGCTGAAGATTTTGCCATTGTACTTGCAAAGCAC
TTCACCTCCTTTTATCAGCAGGTCACTGCTGCAATTGTTAATATAGTTGAGAAGCCATGCGAACGCATTA
GCATAAAGGGCCAACAACATGAACATGGTTTTAAGCTTGGCTCTGAAAGACATACAACTGAAGTAATGGT
CGATAAATCAGGAACATTGCACCTGACATCTGGTATTGAAGGATTATCAATTCTGAAGACAACTAAGTCA
GGTTTTGAAGGATTCATTAGGGACCGGTACACTATGTTGCCTGAAACACGCGAAAGGATGATGGCTACAG
AGGTCACTGCATCTTGGAGGTACTCTTTTGAATCTCTCTCAAGTCTTCCCCTGAAGCCTCTGTATTTCAC
AGATAGATGCATAGATGTGAAAGAGGTCCTGCTTAACACATTCTTTGGACCATCTAAGGAAGGAGTATAC
AGCCCTTCTGTTCAAGCCACACTTTACGAAATGGCAAAGGCTGTTCTTGGAAGGTCTCCTGACATTTCAT
TCATTCAGTTGAAGATGCCAAATATCCATTTTCTGCCGGTCAATTTGCCAAGCAAAGATAAACCAGACAT
TGTTAAGTTTGCTGACGATGTTTACTTGCCAACGGATGAACCACATGGAACGATTCAGGCGACTCTCAGT
CGTCTAACGTCAAAGATGTGAATTACTGCTGATCAAATGGTACTGCCTATGTTCATTTGCTTACATTTAA
AGTTGTTACCAAACCAAAGCTACACAAGTGGGAGAAATGTTTCTAGCAAGAAGCAACAACATATAACATG
TTTCTGAATCTTTGAAGTTCGTTGNACACATTTAACATATAGGTTAATAAATGTATTGCAACATGAAAAA
AAAAAAAAAAAAAAAAACTCGAG

ORF

ATGGCGGAATTGAGAGGACTGAAGCTGGAGCAGAGGCATGGGAAATCCAGAGTGAGAGTAGCTAGGGTTT
GGAGAGATCGAGATGGAAAACATCACTTTGCTGAATGGAGCGTTAACATCAGTCTCATCTCCGATTGTTT
ACCTGCATATGTCTCCGGCGATAATTCCGATATCGTCGCCACCGATACCATGAAGAACACTGTATATGCG
AAAGCTAAGGAATGTTCAGAGAAACTTTCAGCTGAAGATTTTGCCATTGTACTTGCAAAGCACTTCACCT
CCTTTTATCAGCAGGTCACTGCTGCAATTGTTAATATAGTTGAGAAGCCATGGGAACGCATTAGCATAAA
GGGCCAACAACATGAACATGGTTTTAAGCTTGGCTCTGAAAGACATACAACTGAAGTAATGGTCGATAAA
TCAGGAACATTGCACCTGACATCTGGTATTGAAGGATTATCAATTCTGAAGACAACTAAGTCAGGTTTTG
AAGGATTCATTAGGGACCGGTACACTATGTTGCCTGAAACACGCGAAAGGATGATGGCTACAGAGGTCAC
TGCATCTTGGAGGTACTCTTTTGAATCTCTCTCAAGTCTTCCCCTGAAGCCTCTGTATTTCACAGATAGA
TGCATAGATGTGAAAGAGGTCCTGCTTAACACATTCTTTGGACCATCTAAGGAAGGAGTATACAGCCCTT
CTGTTCAAGCCACACTTTACGAAATGGCAAAGGCTGTTCTTGGAAGGTCTCCTGACATTTCATTCATTCA
GTTGAAGATGCCAAATATCCATTTTCTGCCGGTCAATTTGCCAAGCAAAGATAAACCAGACATTGTTAAG
TTTGCTGACGATGTTTACTTGCCAACGGATGAACCACATGGAACGATTCAGGCGACTCTCAGTCGTCTAA
CGTCAAAGATGTGA

Protein

MAELRGLKLEQRHGKSRVRVARVWRDRDGKHHFAEWSVNISLISDCLPAYVSGDNSDIVATDTMKNTVYA
KAKECSEKLSAEDFAIVLAKHFTSFYQQVTAAIVNIVEKPWERISIKGQQHEHGFKLGSERHTTEVMVDK
SGTLHLTSGIEGLSILKTTKSGFEGFIRDRYTMLPETRERMMATEVTASWRYSFESLSSLPLKPLYFTDR
CIDVKEVLLNTFFGPSKEGVYSPSVQATLYEMAKAVLGRSPDISFIQLKMPNIHFLPVNLPSKDKPDIVK
FADDVYLPTDEPHGTIQATLSRLTSKM*

11

cDNA

CAAAAAGCTCGAGAGAAAGAGAAGAGCAAAAATGGGAGTTTTCACCTTCGTTTGCAAAGGTTCAGGAGAC
GAATGGAGTGCTAAGCAACTCAAGGGTGATCTAGAAGCATCTGCTTCTTGTACCTATGATCTTCAACGCA
AGCTTGTTAAAGCTGCTCTTGCTTCTGACTCTTCTGGTGGTGTTCAGTCATCATTCTCCTTTGTTACACC
TTCCTCTGCTGTTTTCCAGGTGATCATTGGTGGTGGCGGCGGCGGTGGTTTCATTGGTGGTGGAGCAGCA
GCAGCAGCTCCTTCAGGTGGTGGGGCAGCTGCAGAAGCAGCTCCAGCTGAGGAGAAGAAGGAAGAGAAAG
AAGAGAGTGACGACGATATGGGATTCTCACTCTTCGATTAGATTGGGCGTTACTATATAGTAGTACAATT
TTCTGTTGATTTCATCTTTTGTTAGCTTCATTTTGTTGCTTCTAAAGTTTTTGATTAAAACATTGGATTA
TCAGTTTTGGTGTTTAGGTGTCTAAAATGACATTGTGTTTTTATGGTTCTATGAACACATTTTGAGTTAC
TATTATCAAAAAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

AAAAAGCTCGAGAGAAAGAGAAGAGCAAAAATGGGAGTTTTCACCTTCGTTTGCAAAGGTTCAGGAGACG
AATGGAGTGCTAAGCAACTCAAGGGTGATCTAGAAGCATCTGCTTCTTGTACCTATGATCTTCAACGCAA
GCTTGTTAAAGCTGCTCTTGCTTCTGACTCTTCTGGTGGTGTTCAGTCATCATTCTCCTTTGTTACACCT
TCCTCTGCTGTTTTCCAGGTGATCATTGGTGGTGGCGGCGGCGGTGGTTTCATTGGTGGTGGAGCAGCAG
CAGCAGCTCCTTCAGGTGGTGGGGCAGCTGCAGAAGCAGCTCCAGCTGAGGAGAAGAAGGAAGAGAAAGA
AGAGAGTGACGACGATATGGGATTCTCACTCTTCGATTAG

Protein

KKLERKRRAKMGVFTFVCKGSGDEWSAKQLKGDLEASASCTYDLQRKLVKAALASDSSGGVQSSFSFVTP
SSAVFQVIIGGGGGGGFIGGGAAAAAPSGGGAAAEAAPAEEKKEEKEESDDDMGFSLFD*

22

cDNA

CTTCGAGCAAACATCTGTGGAAGTCGAATCGAATAGGAAGTGGCAGAGAAGAAAAATGTTCCTACCGAGG
ATAGCAAGGCCATTGATGATGATGGCGAAAGTGAAGGAAACAACAGGTATAGTAGGTTTGGAGGTAGTCC
CAAATGCAAGGGAAGTTCTGATAAATCTATACAGGAAAACCCTAGAAGAGATCAAGGAGGTACCTGAGGA
CGAAGGATACCGCAAGGCAGTGGAAAGCTTCACCCGCCACCGCCTCAATGTCTGTGTGGAGGAACATGAA
TCTGAAATGATTGAGAAACGGCTTGGTTGTGGTCAGGTTGAAGAGCTCATTGAAGAAAGAAGGCCCAAGA
TGAACTCAAGCTCATTGGTCACATGAACGAGTGGAAACCTTGGGGTATTCCTGATGATTATGAATGTGAA
GTCATTGAAAATGATGCTCCAGTACCAAAACATGTTCCTCTGCACCGTCCTGGTCCTCTCCCTGAGGAAT
TCTATAATACAATGGAGGCGGTTACTTCTGGCAAATTGGATGCTGGTTCAAAGAAGGATGAACCTGCAAT
TTCATCGGGTGACACACAGTCAAAGTAGGGTCTGGCTTCCAGAAGCTCATTTTGGTTATTGATCAGTACA
TGCAAAGTTTTTAAACTGATCATGTGCTACTCTATCATCCTTGTTTTCTTTGTGCATCTTCTATACCTGT
TTACTGGAGCTGGTGGACACAAATGCTTGAATCATTATTGGTTAATAAAACAAAGGTTCCTTTTTTTTGG
GTGGAAGCTAGCAGACGGTTGCTTATTTCTTTGTTTGTAAAAGCATCAACATGGTGTATTTCGTATGAAG
GAAGTTATTTTTCATGAATAATAGTATATCTTCTAAACAAAAAAAAAAAAAAAAAAAACTCGAG

ORF

ATGTTCCTACGGAGGATAGCAAGGCCATTGATGATGATGGCGAAAGTGAAGGAAACAACAGGTATAGTAG
GTTTGGAGGTAGTCCCAAATGCAAGGGAAGTTCTGATAAATCTATACAGGAAAACCCTAGAAGAGATCAA
GGAGGTACCTGAGGACGAAGGATACCGCAAGGCAGTGGAAAGCTTCACCCGCCACCGCCTCAATGTCTGT
GTGGAGGAACATGAATCTGAAATGATTGAGAAACGGCTTGGTTGTGGTCAGGTTGAAGAGCTCATTGAAG
AAAGAAGGCCCAAGATGAACTCAAGCTCATTGGTCACATGAACGAGTGGAAACCTTGGGGTATTCCTGAT
GATTATGAATGTGAAGTCATTGAAAATGATGCTCCAGTACCAAAACATGTTCCTCTGCACCGTCCTGGTC
CTCTCCCTGAGGAATTCTATAATACAATGGAGGCGGTTACTTCTGGCAAATTGGATGCTGGTTCAAAGAA
GGATGAACCTGCAATTTCATCGGGTGACACACAGTCAAAGTAG

Protein

MFLRRIARPLMMMAKVKETTGIVGLEVVPNAREVLINLYRKTLEEIKEVPEDEGYRKAVESFTRHRLNVC
VEEHESEMIEKRLGCGQVEELIEEAQDELKLIGHMNEWKPWGIPDDYECEVIENDAPVPKHVPLHRPGPL
PEEFYNTMEAVTSGKLDAGSKKDEPAISSGDTQSK*

23

cDNA

GAGGTGTCTCGCGGATGGAAATCTTCTGCAAACCAAGATCCACAACATTGGTGCAACACTTGTTGGTGTG
GATAAATTTGGTAACAAATATTATGAGAAGCTAGGCGATACACAATATGGAAGACACCGGTGGGTGGAGT
ATGCACAGAAGAATCGTTACAATGCTTCACAAGTTCCACCAGAATGGCATGGCTGGCTCCATCACGTCAC
TGATCACACTGGTGATGAACTGCTACTGCTGAAACCAAAGAGGTACGGGATTGAGCACAAGGAGAACTTT
TCTGGTGAAGGTGATGCATACATATATCACTCCAAGGGACACACTTTAAATCCTGGACAGAGAGACTGGA
CTAGGTACCAAACTTGGCAACCCACCAAGAAGTAATAGGTCTCAACTGCCATAGTGCAGCTGGACAATTA
TATGGATTTTCTGCCATGCTTCGTTAATAAAAAAGTCCTCGTATTGAACCTTTGGTTATAATCTAAGCAC
ATGTGTACCATCAAANATCAGGTCGANACACTTTGCTGTACAATTTCTCATCTTTGATATGGTTCTCAGG
TTTTGAATTGGATCCAAAATGTAGTCATGCATACTTTATTCCATCGCCGGGTTTGAAGTATGATTGTTAT
TTTTTTTGAAACTAAAGCTATTCAGTGATATGTTATGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAACTCGAG

ORF

AGGTGTCTCGCGGATGGAAATCTTCTGCAAACCAAGATCCACAACATTGGTGCAACACTTGTTGGTGTGG
ATAAATTTGGTAACAAATATTATGAGAAGCTAGGCGATACACAATATGGAAGACACCGGTGGGTGGAGTA
TGCACAGAAGAATCGTTACAATGCTTCACAAGTTCCACCAGAATGGCATGGCTGGCTCCATCACGTCACT
GATCACACTGGTGATGAACTGCTACTGCTGAAACCAAAGAGGTACGGGATTGAGCACAAGGAGAACTTTT
CTGGTGAAGGTGATGCATACATATATCACTCCAAGGGACACACTTTAAATCCTGGACAGAGAGACTGGAC
TAGGTACCAAACTTGGCAACCCACCAAGAAGTAA

Protein

RCLADGNLLQTKIHNIGATLVGVDKFGNKYYEKLGDTQYGRHRWVEYAQKNRYNASQVPPEWHGWLHHVT
DHTGDELLLLKPKRYGIEHKENFSGEGDAYIYHSKGHTLNPGQRDWTRYQTWQPTKK*

Appendix 2 - Publications arising from this work

Sections of this thesis have been published in the following papers:

FEBS 25959

FEBS Letters 516 (2002) 179–182

Agrobacterium tumefaciens supports DNA replication of diverse geminivirus types

Luke A. Selth[a,b], John W. Randles[b], M. Ali Rezaian[a,*]

[a] *CSIRO Plant Industry (Horticulture Unit), PO Box 350, Glen Osmond, SA, Australia*
[b] *Department of Applied and Molecular Ecology, Adelaide University, Adelaide, SA, Australia*

Received 12 February 2002; revised 26 February 2002; accepted 27 February 2002

First published online 11 March 2002

Edited by Jesus Avila

Abstract **We have previously shown that the soil-borne plant pathogen *Agrobacterium tumefaciens* supports the replication of tomato leaf curl geminivirus (Australian isolate) (TLCV) DNA. However, the reproducibility of this observation with other geminiviruses has been questioned. Here, we show that replicative DNA forms of three other geminiviruses also accumulate at varying levels in *Agrobacterium*. Geminiviral DNA constructs that lacked the ability to replicate in *Agrobacterium* were rendered replication-competent by changing their configuration so that two copies of the viral *ori* were present. Furthermore, we report that low-level replication of TLCV DNA can occur in *Escherichia coli* containing a dimeric TLCV construct in a high copy number plasmid. These findings were reinforced by expression studies using β-glucuronidase which revealed that all six TLCV promoters are active in *Agrobacterium*, and two are functional in *E. coli*.** © **2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.**

Key words: Geminivirus; Begomovirus; Replication; β-Glucuronidase; *Agrobacterium tumefaciens*; *Escherichia coli*

1. Introduction

Geminiviruses are plant pathogens which infect a wide range of crops and cause significant economic losses worldwide. Members of this family are characterized by twinned icosahedral virions encapsidating circular, single-stranded DNA genomes of 2.5–3.0 kb (for review, see [1]). Replication of geminiviral DNA is thought to occur by a rolling circle mechanism, analogous to that employed by some bacteriophages [2] and a class of eubacterial plasmids [3]. A hallmark of this replication strategy is the production of supercoiled, open circular, and linear double-stranded (ds) DNA species (for review, see [4]). In addition, recent evidence suggests that a recombination-related process is also involved in geminivirus DNA replication [5].

We have demonstrated in earlier studies that replication of tomato leaf curl virus (TLCV) (*Geminiviridae*: begomovirus) DNA occurs in vivo in *Agrobacterium tumefaciens* carrying a pBin19 plasmid vector harboring tandem copies of the TLCV genome [6]. This process required a functional C1 gene, which encodes the viral replication initiator protein (Rep), and two copies of the viral *ori*. None of the other viral genes were necessary for accumulation of TLCV DNA. The observation that TLCV DNA replication was supported by the bacterial cellular machinery provided the first experimental information supporting the hypothesis that geminiviruses may have evolved from prokaryotic episomal replicons.

Despite the novelty of the original finding, as yet no report of DNA replication in bacteria by any other geminivirus has appeared, suggesting that this ability may be specific to TLCV. This is an intriguing possibility since most characteristics of TLCV, including its genome organization, intergenic region, and gene functions, appear to be typical of the begomovirus genus of *Geminiviridae*. In an effort to elucidate whether DNA replication in *A. tumefaciens* is unique to TLCV or common among geminiviruses, pBin19 constructs carrying other geminiviral sequences were analyzed. Here we report that viral DNA species of two other begomoviruses, the monopartite tomato yellow leaf curl virus (TYLCV) and the bipartite African cassava mosaic virus (ACMV), also accumulate to significant levels in *A. tumefaciens*. In addition, a strain of TLCV recently discovered in the Northern Territory of Australia, termed TLCV-D1 [7], possesses the ability to replicate in *Agrobacterium*. We also tested whether TLCV DNA species could accumulate in *Escherichia coli*, and found that this bacterium could support viral replication when transformed with a pUC8 plasmid construct containing a TLCV dimer.

To provide further evidence for the occurrence of TLCV DNA replication in *Agrobacterium* and *E. coli*, the activity of the TLCV promoters in these bacteria was measured using the β-glucuronidase (GUS) gene as a reporter. All six promoters were active in *Agrobacterium*, while in *E. coli* only the C1 and C2 promoters produced detectable expression of GUS.

2. Materials and methods

2.1. Construction of clones

Tandem repeats of cloned TLCV DNA were inserted into pBin19 or pUC8 to create pBin19-TLCV2.0 and pUC8-TLCV2.0, respectively. The pUC8-TLCV2.0 plasmid was constructed by ligating a *Bam*HI monomer from TLCV clone pTLC4 [8] with *Bam*HI-linearized vector pUC8, and selecting a transformant containing a head-to-tail dimeric insert. Plasmid pBin19-TLCV2.0 is described in Rigden et al. [6]. The method used to create a dimeric clone of TLCV-D1 in pBin19 has been described [7].

To create a pBin19 construct containing a tandem repeat of the ACMV DNA A component, an *Eco*RV monomer from ACMV clone pBinCLV1.3A [9] was ligated into pBluescript SK (Stratagene), which

*Corresponding author. Fax: (1)-618-8303 8601.
E-mail address: ali.rezaian@csiro.au (M.A. Rezaian).

Abbreviations: ACMV, African cassava mosaic virus; GUS, β-glucuronidase; TLCV, tomato leaf curl virus; TYLCV, tomato yellow leaf curl virus

0014-5793/02/$22.00
PII: S0014-5793(02)02539-5

was then linearized with *Pfl*MI and ligated with a full length *Pfl*MI ACMV monomer. The dimeric viral DNA was then moved into pBin19 as a *Xho*I/*Xba*I fragment, generating pBin19-ACMV2.0 (Fig. 1).

The TYLCV construct was created by ligating a *Bgl*II monomer from TYLCV clone pBin19/TYLCV-S1.8 [10] into *Bam*HI-digested pBin19, resulting in loss of the vector *Bam*HI sites. This plasmid was then linearized with *Bam*HI (which cuts at nucleotide 152 of the TYLCV genome) and ligated with a *Bam*HI monomer, to generate pBin19-TYLCV2.0 (Fig. 1).

2.2. Extraction and analysis of DNA

DNA was extracted from *A. tumefaciens* C58 carrying the binary vector pBin19-TLCV2.0 or *E. coli* Dh5α carrying pUC8-TLCV2.0 using the boiling miniprep method [11]. The DNA was purified by either phenol:chloroform extraction or by adsorption onto a spin-column (Qiagen). Approximately 50 ng of samples were electrophoresed in 1.2% w/v agarose gels in Tris/borate/EDTA and blotted onto Zeta-Probe membrane (Bio-Rad) with 0.4 M NaOH, as described [8]. ^{32}P-labelled probes were prepared by a random decamer priming kit (Geneworks) using dimeric viral DNA fragments as the templates.

2.3. Analysis of GUS expression in E. coli

The construction of pBin19 plasmids containing the individual TLCV gene promoter regions fused to the GUS reporter gene has been described [12]. *A. tumefaciens* and *E. coli* cells harboring these constructs were grown for 36 h at 28°C and overnight at 37°C respectively. Cell pellets from 600 μl of culture were frozen in liquid nitrogen, and ground with the aid of sand using an electric screwdriver equipped with a grinding tip in Eppendorf tubes. The homogenate was extracted in 350 μl extraction buffer, spun for 10 min in a microfuge and two 150 μl samples of the supernatant withdrawn for analysis. All measurements were on duplicate starting cultures. GUS activity was determined by a fluorometric assay [13], and corrected for variations in extraction efficiency by measuring protein concentrations using a Bio-Rad protein assay reagent kit according to the manufacturer's instructions.

3. Results and discussion

From studies of TLCV DNA replication in *Agrobacterium* a model has been proposed postulating that the viral construct must contain an intact *rep* gene whose product would cleave the universal geminivirus nicking site (TAATATTAC) present in a structurally conserved hairpin-loop [6]. A replication-competent construct in *Agrobacterium* must also contain two copies of the viral *ori* including nicking sites so that a unit length viral DNA is released and circularized.

With this background in mind, we tested a number of geminiviral constructs which either complied with or were deficient in terms of the criteria outlined above. All viral constructs produced were tested by agroinoculation in tomato (for TYLCV) or tobacco (for ACMV) and viral symptoms were observed after 3 weeks (data not shown). When pBin19 vector constructs containing virus tandem dimers of TLCV-D1, TYLCV, and ACMV DNA were introduced into *Agrobacterium*, replicative viral DNA species were produced in overnight cultures (Fig. 2, lanes 2, 3, and 5 respectively). The accumulation of these viral DNAs was comparable to our control construct, the TLCV dimer (Fig. 2, lane 1). However, *A. tumefaciens* was not able to support DNA replication of the TYLCV 1.8-mer construct (Fig. 2, lane 4), which contains just one viral *ori* (Fig. 1). This construct was infectious in plants following agroinoculation (data not shown). We speculate that failure to reproduce our original findings with TLCV in other laboratories may have been due to the lack of two geminiviral origins of replication in the DNA constructs employed.

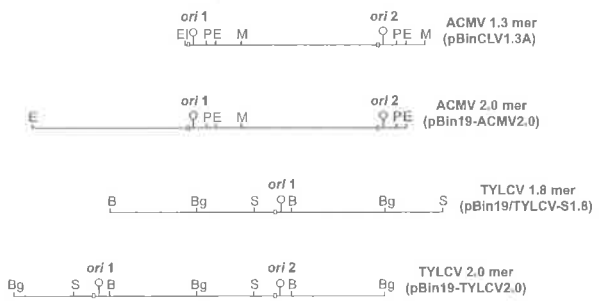


Fig. 1. Linear tandem repeats of ACMV (upper two panels) and TYLCV (lower two panels) DNA cloned into pBin19. The open box (□) upstream of the stem-loop (◦) represents a region containing three iterative elements thought to comprise the Rep protein binding site. Note that the TYLCV 1.8-mer contains only one origin of replication. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RV; EI, *Eco*RI; M, *Mlu*I; P, *Pfl*MI; S, *Sst*I.

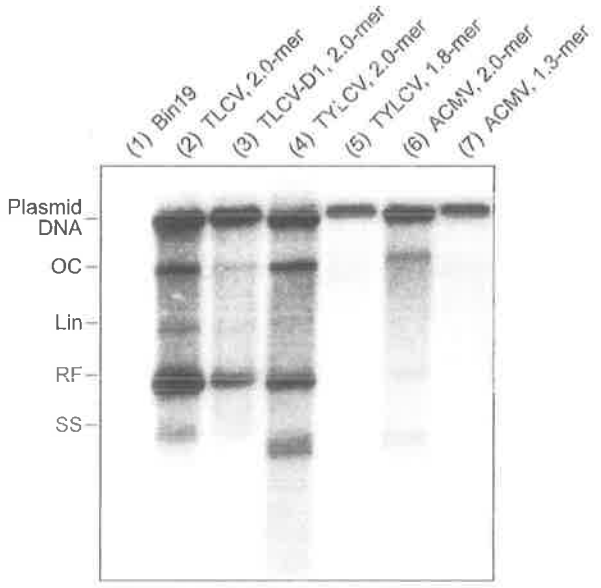


Fig. 2. Virus-specific DNA species produced in *A. tumefaciens* harboring pBin19 containing tandem repeat copies of TLCV, TLCV-D1, TYLCV, and ACMV DNA. DNA extracts from *A. tumefaciens* containing pBin19 geminivirus constructs were analyzed by Southern blotting [4]. Geminiviral DNA forms are marked OC (open circular double-stranded), Lin (linear), RF (supercoiled double-stranded), and SS (single-stranded).

The ACMV 1.3-mer construct (pBinCLV1.3A), which does contain two copies of the stem-loop nicking site, was also unable to replicate in *Agrobacterium* (Fig. 2, lane 6), although in a number of blots trace levels of viral DNA species were visible (data not shown). This result resembles that observed for a TLCV 1.1-mer [6], which lacks one of three iterative elements within the Rep binding region of *ori* 1. Both pBinCLV1.3A and the TLCV 1.1-mer have a relatively short copy of the repeat *ori* 1 (Fig. 1), but unlike the TLCV construct, pBinCLV1.3A appears to contain all of the predicted Rep binding iterons [14]. However, the iterons which make up the high-affinity Rep binding site are only a part of the entire geminivirus *ori* which has a modular structure containing multiple elements [15]. In fact, recent experiments in this laboratory have shown that the removal of these DNA sequences

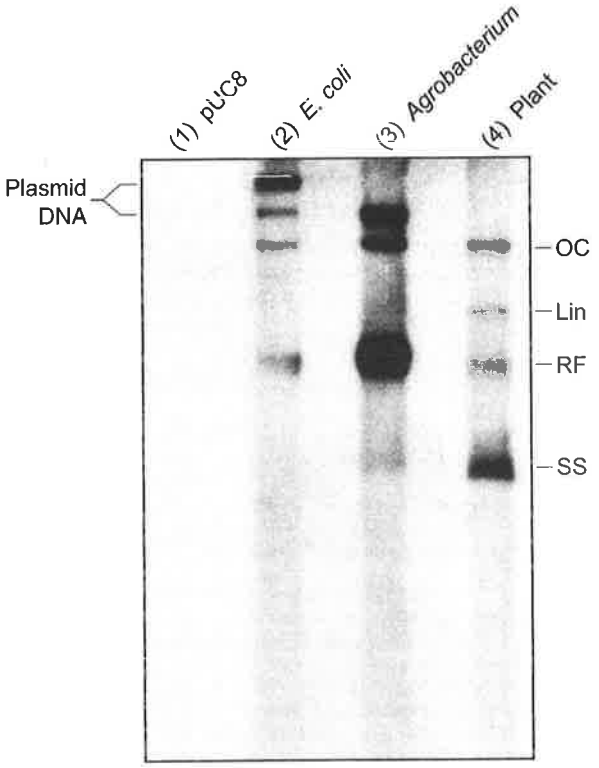


Fig. 3. Accumulation of TLCV DNA species in *E. coli* transformed with pUC8 containing a dimeric head-to-tail insert of TLCV. TLCV DNA obtained from *Agrobacterium* and infected plant tissue is also shown. Geminiviral DNA forms are labelled as in Fig. 2.

has no discernible effect on the replication of TLCV or its satellite in the host plants tested (B. Lin, personal communication). It is therefore possible that the restricted replicative ability of pBinCLV1.3A and the TLCV 1.1-mer is a result of these constructs lacking some specific DNA elements, apart from the Rep binding iterons, in *ori* 1.

In all experiments performed, the quantity of ACMV DNA observed, particularly the ds replicative form (RF), was significantly less than that of the monopartite viruses tested. The A component of ACMV encodes all of the information necessary for viral replication and encapsidation in planta [16]. It seems probable, therefore, that the reduced replicative ability of ACMV reflects some minor differences in the bacterial:viral interaction occurring among the two genera. However, we cannot rule out the possibility that for bipartite geminiviruses, one or both of the B DNA component genes, while unnecessary for efficient replication in planta, may be involved in accumulation of viral DNA in *Agrobacterium*.

The cellular machinery of another bacterium, *E. coli*, is also able to support processes resembling viral DNA replication in plants when transformed with a pUC8 construct containing a tandem dimer of TLCV (Fig. 3, lane 1). In comparison with TLCV DNA replication in *A. tumefaciens* (Fig. 3, lane 2), less replicative form (RF) DNA was observed in relation to the quantity of vector, suggesting that TLCV is less well adapted to the replicative environment of *E. coli*. This is consistent with our previous observations [6] that TLCV replicative DNA forms could not be detected in *E. coli* harboring the low copy number plasmid pBin19 containing a TLCV dimer. It has been proposed that geminivirus progenitors may have arisen from bacterial replicons which were inserted into the host plant genome by *Agrobacterium* and escaped via a DNA release process [6,17]. The finding of TLCV DNA replication in *E. coli*, although not to the level of *A. tumefaciens*, raises the possibility that geminiviruses may also replicate in other bacterial species. It is intriguing to consider that the future evolution of geminiviruses may involve episomal associations with new bacteria, allowing exposure to other species and possibly resulting in host-switching events.

The ability of geminiviruses to replicate in bacteria is further supported by the observation that TLCV promoters are active within these cells. Fig. 4 shows the relative level of expression of GUS by TLCV promoters in *E. coli* and *Agrobacterium*. Each of the promoters exhibit significant activity in *Agrobacterium*, with the putative promoter element of the *rep* gene able to convert the GUS substrate 4-methylumbelliferyl β-glucuronide (MU) at a rate of 7310 μmol/min/mg protein extracted.

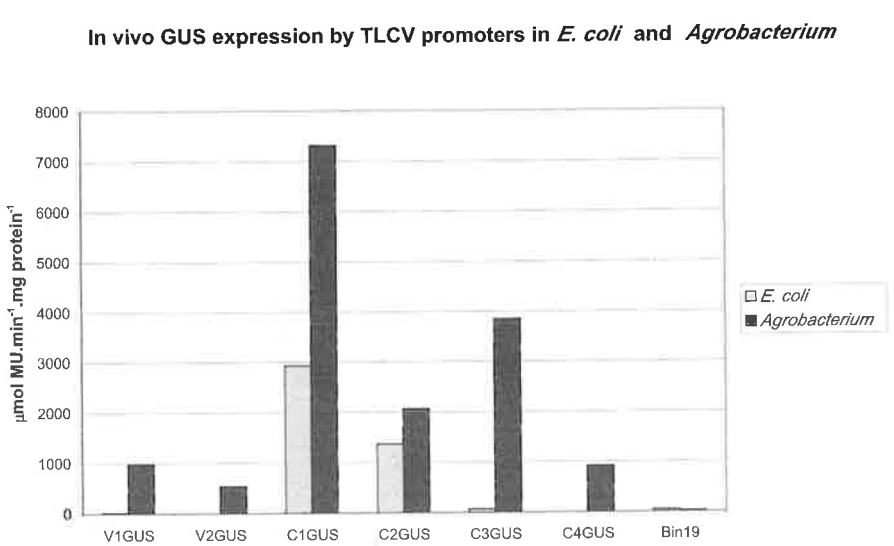


Fig. 4. TLCV promoters are active within bacterial cells. Extracts from *A. tumefaciens* and *E. coli* cells transformed with pBin19 constructs containing GUS-TLCV promoter fusions were analyzed for GUS activity by a fluorometric assay [10].

Detectable levels of GUS expression in *E. coli* were only obtained in cells containing the C1GUS and C2GUS fusions. This is in contrast to an early report which indicated that the tomato golden mosaic virus coat protein promoter actively drives expression of the kanamycin resistance gene in *E. coli* [18], although this discrepancy may simply be a result of the different methods used to measure promoter activity. Since only the C1 gene is required for replication of TLCV DNA in *Agrobacterium*, it is possible that the low-level accumulation of TLCV replicative DNA species in *E. coli* compared to *Agrobacterium* (Fig. 3) is a direct result of the differential expression of this gene in these bacteria. However, it cannot be ruled out that efficient replication of TLCV DNA in *E. coli* requires the presence of other viral gene products. In particular, the lack of GUS expression from the C3 promoter, which drives production of a protein which is known to greatly enhance geminiviral DNA accumulation in planta [19], may be the cause of reduced accumulation of TLCV DNA in *E. coli*.

Our results suggest that the ability of geminiviruses to replicate in *Agrobacterium* is not limited to TLCV and may be a common feature of these plant pathogens. This finding raises the possibility that an in vitro replication system for geminiviruses could be developed using bacterial cell extracts supplemented by viral Rep. Such a tool would prove invaluable for further characterization of geminivirus replication.

Acknowledgements: We thank J. Stonor for excellent technical assistance and D. Pountney and I. Dry for valuable input. We also gratefully acknowledge J. Stanley and B. Gronenborn for providing the ACMV and TYLCV cloned DNA. This work was supported by Australian Research Council Grant A09802106.

References

[1] Hanley-Bowdoin, L., Settlage, S.B., Orozco, B.M., Nagar, S. and Robertson, D. (1999) Crit. Rev. Plant Sci. 18, 71–106.
[2] Gruss, A. and Ehrlich, S.D. (1989) Microbiol. Rev. 53, 231–241.
[3] Baas, P. and Jansz, H. (1988) Curr. Top. Microbiol. Immunol. 136, 31–70.
[4] Gutierrez, C. (2000) EMBO J. 19, 792–799.
[5] Jeske, H., Lütgemeier, M. and Preiß, W. (2001) EMBO J. 20, 6158–6167.
[6] Rigden, J.E., Dry, I.B., Krake, L.R. and Rezaian, M.A. (1996) Proc. Natl. Acad. Sci. USA 93, 10280–10284.
[7] Behjatnia, S.A.A., Dry, I.B. and Rezaian, M.A. (2001) Aust. Plant Pathol. 30, 337–342.
[8] Dry, I.B., Rigden, J.E., Krake, L.R., Mullincaux, P.M. and Rezaian, M.A. (1993) J. Gen. Virol. 74, 147–151.
[9] Klinkenberg, F.A., Ellwood, S. and Stanley, J. (1989) J. Gen. Virol. 70, 1837–1844.
[10] Kheyr-Pour, A., Bendahmane, M., Matzeit, V., Accotto, G.P., Crespi, S. and Gronenborn, B. (1991) Nucleic Acids Res. 19, 6763–6769.
[11] Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193–197.
[12] Dry, I., Krake, L., Mullineaux, P. and Rezaian, A. (2000) MPMI 13, 529–537.
[13] Jefferson, R.A. (1987) Plant Mol. Biol. Rep. 5, 387–405.
[14] Arguello-Astorga, G.R., Guevara-Gonzalez, R.G., Herrera-Estrella, L.R. and Rivera-Bustamante, R.F. (1994) Virology 203, 90–100.
[15] Fontes, E.P., Gladfelter, H.J., Shaffer, R.L., Petty, I.T. and Hanley-Bowdoin, L. (1994) Plant Cell 6, 405–416.
[16] Morris, B., Richardson, K., Eddy, P., Zhan, X.C., Haley, A. and Gardner, R.J. (1991) J. Gen. Virol. 72, 1205–1213.
[17] Bejerano, E.R., Khashoggi, A., Witty, M. and Lichtenstein, C. (1996) Proc. Natl. Acad. Sci. USA 93, 759–764.
[18] Petty, I.T.D., Coutts, R.H.A. and Buck, K.W. (1986) Nucleic Acids Res. 14, 5113.
[19] Sunter, G., Hartitz, M.D., Hormuzdi, S.G., Brough, C.L. and Bisaro, D.M. (1990) Virology 179, 69–77.

Selth, L.A., Randles J.W. & Rezaian, M.A. (2004). Host responses to transient expression of individual genes encoded by Tomato leaf curl virus. *Molecular Plant Microbe Interactions, 17*(1), 27-33.

Selth, L. A., Dogra, S. C., Rasheed, M. S., Healy, H. Randles, J. W. & Rezaian, M. A. (2005). A NAC domain protein interacts with Tomato leaf curl virus replication accessory protein and enhances viral replication. *Plant Cell, 17*(1), 311-325.