Characterisation and functional analysis of pollen-specific cDNAs encoding putative transcriptional regulators from *Nicotiana tabacum*.

Thesis submitted for the degree of Doctor of Philosophy
Justin Paul Sweetman.
Department of Botany,
University of Leicester.
1996.

UMI Number: U641628

UMI

Dissertation Publishing

UMI U641628

ProQuest

Acknowledgements

Abbreviations

Ab	absorbance
bp	base pair
BSA	bovine serum albumin
$^{\circ}C$	degrees centigrade
cDNA	complimentary DNA
Ci	curie
cm	centimetre
cpm	counts per minute
Da	daltons
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanidine triphosphate
dTTP	deoxythymidine triphosphate
dNTPs	deoxynucleotides
ddNTPs	dideoxynucleotides
DMSO	dimethyl sulpoxide
DNA	deoxyribonucleic acid
dH_2O	sterilised double distilled water
DTT	dithiothreitol
EDTA	ethylenediaminotetraacetic acid
g	gram
g	gravity
GUS	β–Glucoronidase
hr	hour
IPTG	iso-propyl-ß-D-thiogalactoside
kb	kilobase pair
kDa	kilo daltons
l	litre
μg	microgram
μl	microlitre
mA	milliamperes
mg	milligram
min	minutes
ml	millilitre

M	molar
mM	millimolar
mRNA	messenger RNA
ng	nanogram
nm	nanometer
OD	optical density (measured in nm)
PAGE	polyacrylamide gel electrophoresis
pfu	plaque forming units
poly A$^+$ RNA	polyadenylated RNA
psi	pounds per square inch
RNA	ribonucleic acid
mRNA	messenger RNA
rpm	revolutions per minute
rRNA	ribosomal RNA
s	seconds
SDS	sodium dodecyl sulphate
TEMED	N, N, N', N', Tetramethylene diamine
T$_m$	melting temperature
Tris	2-amino 2-hydroxymethyl propane-1, 3-diol
Triton-X-100	iso-octylphenoxypolyethoxy ethanol
U	enzyme units
uv	ultraviolet
V	volts
var	plant variety
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo 4-chloro 3-indolyl B-D-galactoside

Contents   Page

List of figures

List of tables

Characterisation and functional analysis of pollen-specific cDNAs encoding putative transcriptional regulators from *Nicotiana tabacum*.

Ph.D. Thesis by Justin Paul Sweetman.

The significance of putative transcriptional regulators and their expression patterns in the developing male gametophyte (or pollen grain) have been investigated in *Nicotiana tabacum*. A reverse genetic approach was used to isolate genes involved with the development of the pollen grain, after initial attempts to isolate transcription factors directly by expression library screening proved unsuccessful. Oligonucleotides corresponding to a part of the MYB and LIM domain sequences from other plant proteins were used as probes to screen a cDNA library prepared from poly A^+ RNA isolated from mature tobacco pollen. Two clones from each class isolated were found to have unique nucleotide sequences (*myb.Nt1, myb.Nt2, Ntl1 and Ntl2*). Homology searches revealed the presence of MYB and LIM domain repeats in each pair of clones respectively. Detailed Northern blot analysis showed that the *myb.Nt1* and *Ntl1* transcripts were present specifically in mature and germinating pollen. Both transcripts were detectable immediately before the asymmetrical division Pollen Mitosis I, reaching a maximal level in mature pollen.

Transgenic plants containing the whole *myb.Nt1* and *Ntl1* cDNA clones in sense and antisense orientations to the pollen-specific *lat52* promoter and in sense orientation to the widely and stronghly expressing cauliflower mosaic virus 35S promoter (but whose expression in pollen is relatively weak) were generated. Analysis of transgenic seed from second generation plants, revealed significantly reduced transmission through the pollen grain. Analysis of these plants showed a reduction in endogenous *myb/lim* transcript abundance. These results suggest that the MYB.NT1 protein is crucial for pollen function.

Plants harbouring the *35S CaMV-myb.Nt1* construct exhibited a stably inherited dwarf phenotype, ranging from partial to severe dwarfism in second generation plants. Analysis of these plants showed among other traits a reduction in internodal growth and a change in leaf morphology. These results suggest that MYB.NT1 is involved with plant hormone biosynthesis and/or hormone induced gene expression in the pollen grain.

Chapter One.

Introduction.

1.1 The scope of this thesis

The pollen grain of flowering plants (Angiosperms) is a highly specialised two- or three-cell gametophyte, whose function is to deliver sperm cells to the ovule. In order to achieve this function the pollen grain must be able to survive as a free organism from the sporophyte plant, from the time it is released at anthesis, until it is delivered to the stigma of the pistil by a vector such as wind or insects. However before fertilisation is accomplished, a complex developmental programme is necessary, which includes meiotic divisions, the production of a unique extracellular matrix, a process of vacuolation and nuclear migration and the determinative asymmetric cell division, pollen mitosis I (PMI). Once the pollen grain has reached the stigma, it germinates to produce a highly polarised pollen tube that rapidly grows through the style to the ovule, thereby delivering sperm cells to the embryo sac. Thus, the pollen grain provides an interesting subject for the study of mitotic events, cellular organisation, cell-to-cell interactions and polar growth both at the cytological and molecular levels. Furthermore, pollen is particularly suited to molecular research due to its synchronous development and its easy accessibility, representing a very pure tissue type composed of only two or three cell types (vegetative and generative sperm cells).

Molecular approaches have been used to clone and characterise the expression patterns of a range of microspore and pollen specific genes. Based on these expression patterns, two broadly defined groups of genes were recognised, the 'early' genes activated prior to PMI and the 'late' genes first activated after PMI. Detailed analysis of the promoters of a group of pollen-specific genes have identified a number of *cis*-regulatory elements responsible for the specificity and level of expression of these genes.

The first approach adopted in this thesis to facilitate the isolation of pollen-specific *trans*-acting factors was to utilise concatenated DNA binding site probes of defined *cis*-element sequences present in the promoters of a number of pollen-specific genes. A cDNA expression library was made from poly A$^+$ RNA from mature pollen *Nicotiana tabacum (N. tabacum)* var. Samsun. By screening the library with the DNA binding site probes it was hoped that *trans*-acting factors binding specifically to the *cis*-element sequences would be isolated. In this way the mechanism of activation of pollen-specific gene expression could be analysed. A second approach which was used to isolate pollen expressed *trans*-acting factors was heterologous screening of the

pollen cDNA library using probes with sequence identity to the DNA binding domains of known classes of plant *trans*-acting factors.

The function of the cDNAs isolated from the latter approach was investigated *in planta* by generating transgenic plants containing antisense, overexpression and ectopic pollen cDNA expressing constructs. If a pollen expressed gene plays an important (and in the case of the antisense experiment) non-redundant role during pollen development or germination, the pollen of these transgenic plants might exhibit an altered phenotype. Similarily, ectopic expression of a pollen expressed gene might produce altered plant growth.

1.2 Microsporogenesis and microgametogenesis in *N. tabacum*

The life cycle of plants alternates between a haploid (gametophytic) and diploid (sporophytic) phase. Flowering plants make up the vast majority of higher plant species (250 000 species) compared to the evolutionary more primitive gymnosperms (700 species). In higher plants, the male and female gametophytes (pollen grain and embryo sac respectively), are reduced to microscopic structures which are dependant on the tissues of the sporophyte for their development.

The evolutionary success of flowering plants is attributed in part to mechanisms which select for pollen vigour ((Mulcahy, 1979); (Snow and Spira, 1991)) and ensure self-incompatability and favouring outcrossing (Singh and Kao, 1992). Most flowering plants contain hermaphrodite flowers containing both anthers and pistil in which the male and female gametophytes respectively are formed. The following discussion of male gametophyte development focuses on flowering plants.

Early in flower development, the stamen primordium consists of a mass of undifferentiated archisporal cells surrounded by a layer of epidermal cells. These archisporal cells divide mitotically to produce the primary parietal cells and the sporogenous cells. The primary parietal cells give rise to the endothecium, middle layer and tapetum of the anther. The tapetum encloses a fluid-filled locule within which the sporogenous cells are situated. These sporogenous cells develop into the pollen mother cells (PMC), which after two meiotic divisions gives rise to four haploid microspores surrounded by a callose wall (β-1, 3 glucan; Figure 1.1). Two types of cytokinesis are seen in conjunction with meiosis in the PMC to produce the microspore tetrad, successive and simultaneous. In successive cytokinesis, eg. lily, a cell plate is laid down after each of the two meiotic nuclear divisions (Steiglitz, 1974).

3

Figure 1.1

Microsporogenesis and microgametogenesis in *N. tabacum*.

(adapted from (Twell, 1994)).

Schematic pollen developmental pathway. The relationship between bud length and developmental stage is shown.

Abbreviations: VN, vegetative nucleus; GN, generative nucleus; N, haploid nuclear content; 2N, diploid nuclear content.

Shading along the scale bar represents the range over which a particular developmental stage occurs.

MEIOSIS

MICROSPORE RELEASE

MICROSPORE MITOSIS

POLLEN RELEASE

POLLEN MOTHER CELL

TETRAD

MICROSPORE

VACUOLATION NUCLEAR MIGRATION

NUCLEAR MIGRATION DEVACOULATION SYNTHESIS

Vegetative cell

Generative cell

BICELLULAR POLLEN

MATURE POLLEN

MICROSPOROGENESIS

MICROSPORE DEVELOPMENT

POLLEN DEVELOPMENT

BUD LENGTH (mm)

MICROSPORE

BICELLULAR POLLEN

9 10 11   8   6   14 15 16   4   2   45 48 50   0

Days before anthesis

In contrast, in the simultaneous type of cytokinesis, eg. *N. tabacum*, meiosis results in the production of four nuclei in a common cytoplasm, and it is only at the end of the second meiotic division that the haploid microspores are separated from one another by the cell plates (Horner, 1977). These meiotic divisions are synchronous within the anther, probably due to the fact that the PMC's are connected by cytoplasmic bridges. During the meiotic divisions of the PMC's the tapetum undergoes a differentiation into binucleate polar secretory cells lacking a primary cell wall. These cells become filled with ribosomes, mitochondria, endoplasmic reticulum, Golgi and vesicles located at the locular face of the plasma membrane. In addition, the tapetal cells are connected by cytoplasmic bridges which is thought to coordinate tapetal cell activity. The tapetal cells are metabolically highly active during early microspore development and supply the developing microspores with nutrients (Pacini, 1990). However, late microspore and pollen stages can complete their normal development *in vitro* if supplied with the necessary nutrients (Tupy *et al.*, 1983). This suggests that pollen and to a lesser extent the late microspore can direct its own developmental programme, reflecting the specialised development, structure and function of the pollen grain.

Upon secretion of β 1, 3-glucanases (callase) by the tapetal cells, young haploid microspores are released from the callose wall of the meiotic tetrad (Steiglitz, 1977). Prior to the release of the microspores from the callose wall, a patterning structure called the primexine is deposited on the microspore cell surface forming a template for the deposition of the exine extracellular matrix. The exine is largely composed of sporopollenin, a very resistant biopolymer which protects the pollen grain when it is released from the environment. After their release from the tetrads the microspores rapidly increase in size and become vacuolate over a period of approximately 3.5 days. These small vacuoles eventually coalesce into a single large vacuole, which compresses the cell cytoplasm and moves the centrally located nucleus to a peripheral position. In orchids, a microtubule system has been detected at the site of nuclear migration (Brown and Lemmon, 1991), suggesting that the nuclear migration may be dynamic rather than passive. Concurrent to this development, the tapetal cells undergo cell death but still contribute to pollen development by depositing cell remnants called tryphine, flavanoids and an oily coating of carotenes and lipids called pollenkitt. The tryphine is thought to be important for subsequent pollen-stigma interactions.

At about 4.5 days before anthesis, the microspore divides asymmetrically in a mitotic division called pollen mitosis I (PMI), producing a smaller generative cell and a larger vegetative cell enclosed within the confines of the original microspore cell wall. PMI marks the end of microsporogenesis and the beginning of microgametogenesis.

5

Whereas the nucleus of the generative cell is highly condensed, the vegetative nucleus is larger, has more nuclear pores and is less condensed, indicating that it is transcriptionally active (Wagner *et al.*, 1990). Studies of RNA and protein populations (Bedinger and Edgerton, 1990a) and of protein synthesis (Mandaron *et al.*, 1990) indicate that PMI marks a major developmental switch in gene expression patterns. Indeed, PMI is thought to represent a determinitive switch which commits the young pollen cells to a process of maturation resulting in the production of a mature pollen grain. Anther and microspore culture experiments suggest that microspores of *Brassica napus* can be more easily induced than post-meiotic young pollen to enter a sporophytic developmental program and develop into haploid plants (Gaillard *et al.*, 1991). When the divisional asymmetry of PMI is disrupted by centrifugation (Terasaka and Niitsu, 1987) or by using drugs to inhibit microtubule formation (Zaki and Dickinson, 1991), two similar cells are produced which process the nuclear characteristics of the vegetative cell. This implies that the asymmetrical division is essential for the different characteristics of the generative nucleus after PMI.

During the next 4.5 days to anthesis, the young pollen grain undergoes a process of maturation involving production of large amounts of stored RNA, proteins, metabolic reserves and starch granules in the vegetative cell only (Mascarenhas, 1990). The pollen grain secretes an inner pollen wall, called the intine, which is cellulosic and pectic in character (Bedinger, 1992). During the final stages of maturation the pollen dehydrates to a water content ranging from 6 % to 35 % of dry weight at anthesis (Heslop-Harrison, 1987). In this way the pollen grain is prepared for its function of survival after dispersal from the anther and germination on the stigma surface of a compatible flower.

The generative cell undergoes a further mitotic division to produce two sperm cells. In 70 % of flowering plant species, such as *N. tabacum*, this second division takes place in the pollen tube as it grows through the pistil. In the other species, such as *Arabidopsis thaliana*, this second division occurs soon after PMI in the early pollen grain.

Pollen tube growth occurs by a process of polar cell growth, in which cellular growth is confined to a specific domain of an elongating cell. The vegetative nucleus and cytoplasm and the generative cell (sperm cells) migrate at a constant position relative to the elongating tube. Callose plugs form at intervals behind the cytoplasm of the tube as it grows. In this way, the pollen tube represents a dramatic example of cell polarity. Both actin and myosin have been shown to be active in pollen tube growth by the use of drugs and immunofluorescence ((Perdue *et al.*, 1985); (Tang *et al.*, 1989)). In

6

contrast, drugs that inhibit microtubule growth do not affect pollen tube growth. The pollen tube wall is also rich in hydroxy-proline rich glycoproteins which are deposited in the cell wall during wall formation (Heslop-Harrison, 1987) and which may be important for structural integrity or pollen tube-pistil interactions.

1.3 Gene expression during pollen development
1.3.1 The complexity of gene expression during microsporogenesis and microgametogenesis

Estimates of the number of genes expressed in the male gametophyte have been based on the kinetics of hybridisation of ^{3}H-cDNA with mRNA in excess in *Tradescantia* and maize ((Willing and Mascarenhas, 1984); (Willing *et al.*, 1988)). These studies indicate that some 20 000 to 24 000 differential RNAs are present in the pollen grain at maturity. In contrast the shoot has approximately 30 000 different mRNAs, which probably reflects the reduced structure and restricted functions of the pollen grain. For comparison the anther of *N. tabacum* is estimated to have approximately 26 000 different transcripts of which 11 000 are anther specific (Kamaley and Goldberg, 1980). An analysis of the three abundance classes of mRNAs within the pollen grain and shoot, revealed that the pollen mRNAs in all classes were approximately 100 fold more abundant than their corresponding classes in shoot mRNA. This is most likely due to mRNAs which are stored in the mature pollen grain for rapid translation during pollen germination. Indeed, *in vitro* translation of mRNA isolated from mature pollen is very similar to the pattern of proteins synthesised during pollen germination *in vivo* (Mascarenhas *et al.*, 1984). In addition, blocking translation in germinating pollen blocks pollen tube growth (Mascarenhas, 1975). Taken together these studies indicate that large numbers of stored mRNAs are translated rapidly in the germinating pollen grain. Heterologous hybridisations of pollen cDNA to shoot cDNA and *visa versa* from maize (Willing *et al.*, 1988), and the screening of maize and *Tradescantia* pollen cDNA libraries with pollen and vegetative tissue probes (Stinson *et al.*, 1987) has been used to estimate the overlap of gene expression in the sporophyte and gametophyte. About 80 % to 90 % of the pollen sequences are shared with the sporophyte. Of the 10 % to 20 % remaining sequences that are pollen specific, the expectation is that most are unique, reflecting the specialised structure and function of the pollen grain. However, several examples of pollen specific transcripts have been identified which are the result of expression of genes encoding different isoforms of proteins present in the vegetative tissue. For example, β-glucosidase (Frova *et al.*,

7

1987) and ADP-glucose pyrophosphorylase (Bryce and Nelson, 1979) in maize. This situation can make the differential screening of libraries for anther and pollen specific genes more complex, as the sequence encoding these shared proteins might be similar and thus ignored.

1.3.2 Gene expression during microsporogenesis

A number of genes have been identified, mainly through differential screening of cDNA libraries from poly A^+ RNA from whole anthers at different developmental stages, which are preferentially transcribed during microsporogenesis (Table 1.1). With the exception of the *ntm19* gene isolated from *N. tabacum* (Dr. J.A.M. Schrauwen, Nijmegen), which is specifically expressed in developing microspores, all the other microspore genes have some overlap of expression in the sporophytic tissue (mainly tapetal). This anomaly might be due to the fact that the cDNA libraries screened were constructed from anther poly A^+ RNA and not from isolated microspores. The small size of the microspore and the high metabolic activity of the tapetum very likely saturates anther cDNA libraries with anther derived transcripts (Dickinson, 1993). The application of 'cold plaque' screening is an alternative approach to differential cDNA library screening and involves picking non-hybridising plaques after probing a cDNA library with labelled cDNA from the source tissue. In this way low abundance mRNA transcripts can be cloned (Hodge *et al.*, 1991). The *apg* (Roberts *et al.*, 1993a) and *cex* (Mascarenhas, 1988) genes are expressed within the tapetum and microspores and encode proline rich proteins which may be secreted to the microspore and pollen grain wall. The *Brassica napus Bp19* cDNA (Albani *et al.*, 1991) bears sequence homology to pectin esterases, thought to act by being secreted from the extending pollen tube into the stylar tissues during germination.

Analysis of the proteins translated *in vitro* from mRNA isolated from pollen at different developmental stages showed that a large number of new transcripts appear soon after the developmental switch PMI (Schrauwen *et al.*, 1990). Considering this evidence it was suggested that two sets of genes are active in the developing male gametophyte: 'early' genes switched on soon after meiosis, with transcripts declining in abundance before pollen maturity and 'late' genes activated after PMI with their transcripts accumulating to pollen maturity (Stinson *et al.*, 1987). A large number of pollen specific genes have had their temporal expression patterns investigated and the vast majority have a characteristic 'late' pattern of expression, such as the lat or late anther tomato cDNA clones *lat51, lat52, lat56, lat58* and *lat59* ((Twell *et al.*, 1989a);

8

(Ursin *et al.*, 1989)). However a *P6* transcript from *Onethera organensis* accumulates in mid to late pollen stages of development (Brown and Crouch, 1990). In addition, promoter-GUS fusions have been used to define more precisely the temporal regulation of pollen gene expression *in planta*. Two microspore expressed genes from *Arabidopsis* have had the activities of their promoters investigated. A *pall* gene promoter fusion to the GUS reporter gene enabled detection of GUS activity in microspores prior to PMI, with increasing levels after PMI accumulating to pollen maturity (Ohl *et al.*, 1990). Similarly the promoter of the *Tua1* gene directed low level gus expression in early microspores with higher levels accumulating up to pollen maturity (Carpenter *et al.*, 1992). These patterns of gene expression represent an intermediate pattern overlapping those of 'early' and 'late' genes.

1.3.3 Gene expression during microgametogenesis

In contrast to the microspore expressed genes, a much larger group of late pollen genes have been isolated (Table 1.2). This is probably due to the accessibility of mature pollen from dehisced anthers. Before molecular cloning of pollen genes, it was already evident from genetic analysis of dimeric enzymes such as alcohol dehydrogenase (Schwartz, 1971) that haploid gene transcription and translation occurred in pollen. The pollen specific clones were isolated either by differential screening of cDNA libraries, or in the case of pollen allergens, immunological screening of cDNA expression libraries.

Pollen specificity has been established mainly by Northern blot analysis. A few clones have also been investigated by *in situ* hybridisation to tissue sections (*Zmc13*, (Hanson *et al.*, 1989); *P2*, (Brown and Crouch, 1990); *ntp303*, (Reijen *et al.*, 1991); *Lolplb*, (Singh *et al.*, 1991)). All the clones are pollen specific except *Bcp1* (Theerakulpisut *et al.*, 1991), *Tua1* (Carpenter *et al.*, 1992) and the *lat* genes ((Twell *et al.*, 1989a); (Ursin *et al.*, 1989)) which have a sporophytic overlap (tapetum (*Bcp1*); trichomes (*Tua1*), tapetum (*Tua1*); roots (*lat59*); testa (*lat59*); endosperm (*lat52* and *lat59*)). The *ntp303* gene expression pattern has been characterised further by confocal laser scanning microscopy, and established expression in the vegetative cell of the pollen grain immediately after PMI. The generation of transgenic *N. tabacum* plants containing a nuclear targeted *gus* gene driven by the *lat52* promoter, established that expression of *lat52* is also restricted to the vegetative cell of the developing pollen grain (Twell, 1992). This work was extended by demonstrating that

9

Table 1.1
List of microspore preferential genes.

Overlap indicates that either through northern blot analysis, *in situ* localisation or analysis of stable transformants, expression of the respective gene could be detected in microgametophyte or sporophytic tissue of the anther.

expression overlap

plant species	name	microspore	sporophyte	homology/function	reference
Arabidopsis	*apg*	+	+	*CEX*	(Roberts *et al.*, 1993a)
Brassica napus	*Bp4*	+	?	-----	(Albani *et al.*, 1991)
Brassica napus	*Bp19*	+	?	pectin esterase	(Albani *et al.*, 1991)
Brassica napus	*I3/C98*	+	-	oleosin	((Hodge *et al.*, 1991); (Roberts *et al.*, 1991); (Roberts *et al.*, 1993b))
Brassica napus	*E2*	-	+	phospholipid transfer protein	(Foster *et al.*, 1992)
Brassica napus	*CEX*	+	+	*apg*	(Roberts *et al.*, 1993a)
Brassica napus	*BA54*	+	-	-----	(Shen and Hsu, 1992)
Brassica napus	*BA42*	+	+	chalcone synthase	(Shen and Hsu, 1992)
Brassica napus	*BA73*	?	?	-----	(Shen and Hsu, 1992)
Oryza sativa	*Osc6, Osc4*	+	?	-----	(Tsuchiya *et al.*, 1992)
Petunia hybrida	*Adh-1*	+	+	Alcohol dehydrogenase	(Gregerson *et al.*, 1991)
Nicotiana tabacum	*ntm19*	+	-	-----	(Oldenhof *et al.*, 1996)
Zea mays	*Adh-1*	+	+	Alcohol dehydrogenase	(Stinson and Mascarenhas, 1985)

Table 1.2
List of microgametophyte-specific genes.

All genes isolated have been shown to be expressed specifically in pollen by northern blot analysis, *in situ* hybridisation or analysis of stable plant transformants.

plant species	name	homology\function	reference
Ambrosia artemisiifolia	*AmbaI.1/AmbaI.2/ AmbaI.3*	pectate lyases; *lat56, lat59, TP10, Zmc58*	((Rafner et al., 1991); (Mascarenhas, 1992))
Ambrosia trifolia	*AmbtV*	allergen	(Ghosh et al., 1991)
Aradidopsis thaliana	*TUA1*	a-tubulin	(Carpenter et al., 1992)
Bettula verrucosa	*BetVI*	pea disease resistance gene	(Breiteneder et al., 1989)
Bettula verrucosa	*BetVII*	profillin	(Valenta et al., 1991)
Brassica campestris	*Bcp1*	*Bgp1*	(Theerakulpisut et al., 1991)
Brassica campestris	*Bgp1*	*Bcp1*	(Xu et al., 1993)
Brassica napus	*BMP1*	profilin	(Kim et al., 1993)
Brassica napus	*Bp10*	ascorbate oxidase; *ntp303; lat51*	(Albani et al., 1992)
Corylus avellana	*Cor aI/S*	allergen; *BetVI*	(Breiteneder et al., 1989)
Dactylis glomerata	*7.8*	allergen	(Walsh et al., 1989)
Helianthus annus	*sf3*	Zinc-finger protein; LIM motif; *Ntl1*; *Ntl2*	((Baltz et al., 1992a.); (Baltz et al., 1992b.); (Steinmetz et al., 1993))
Helianthus annus	*sf16*	------	(Steinmetz et al., 1993)
Easter lily	*LMP131A*	profilin; *BMP1*	(Kim et al., 1993)
Lilium longiflorum	*gH2B*	histone	(Ueda and Tanaka, 1995)
Lilium longiflorum	*gH3*	histone	(Ueda and Tanaka, 1995)
Lolium perenne	*Lolpla*	------	(Griffith et al., 1991)
Lolium perenne	*Lolplb*	amylase inhibitor; KBG 41/60/30	((Singh et al., 1991); (Davies et al., 1992))
Lycoperisum esculentum	*lat51*	ascorbate oxidase; *Bp10, ntp303*	((Ursin et al., 1989); (McCormick, 1991))
Lycoperisum esculentum	*lat52*	kunitz trypsin inhibitor *Zmc13*	((Twell et al., 1989b); (McCormick, 1991)
Lycoperisum esculentum	*lat56*	pectate lyase; *lat59, TP10, Zmc58,* Ambrosia allergens	(Wing et al., 1989)
Lycoperisum esculentum	*lat58*	------	(Ursin et al., 1989)
Lycoperisum esculentum	*lat59*	pectate lyase; *lat56*	(Wing et al., 1989)
Nicotiana tabacum	*Npg1*	polygalacturonase;	(Tebbutt et al., 1994)

Npg1, 3C12, P2

Nicotiana tabacum	*Nt.myb1*	MYB-domain protein	(Sweetman and Twell, in prep.)
Nicotiana tabacum	*Nt.myb2*	MYB-domain protein	(Sweetman and Twell, in prep.)
Nicotiana tabacum	*Ntl1*	LIM-motif protein; *sf3*	(Sweetman and Twell, in prep.)
Nicotiana tabacum	*Ntl2*	LIM-motif protein; *sf3*	(Sweetman and Twell, in prep.)
Nicotiana tabacum	*ntp303*	ascorbate oxidase; *Bp10*	((Reijen et al., 1991); (Weterings et al., 1992))
Nicotiana tabacum	*Tac25*	actin	(Thangavelu et al., 1993)
Nicotiana tabacum	*TP10*	pectate lyase; *lat56, lat59, Zmc58*	(Rogers and Lonsdale, 1992b)
Oenothera organensis	*P1/P3/P6*	--------	(Mascarenhas, 1992)
Oenothera organensis	*P2/P22*	polygalacturonase; *Npg1, 3C12*	(Brown and Crouch, 1990)
Oryza sativa	*Act1*	actin	(Tsuchiya et al., 1992)
Oryza sativa	*H3*	histone	(Ragavan, 1989)
Oryza sativa	*Ps1*	*Zm13*	(Zou et al., 1994)
Oryza sativa	*ORY S 1*	allergen	(Xu et al., 1995)
Petunia hybrida	*ChiA*	chalcone flavanone isomerase	(van Tunen et al., 1989)
Poa pratensis	*KGB7.2*	allergen	(Mohapatra et al., 1990)
Poa pratensis	*KGB41/KGB60/ KGB31*	*Lolp1b*	(Silvanovich et al., 1991)
Trad escantia paludosa	*Tpc44/Tpc70*	------	(Stinson et al., 1987)
Zea mays	*cdpk*	Calcium dependant calmodulin independant protein kinase	(Estruch et al., 1995)
Zea mays	*Pex1*	Extensin domain protein	(Rubenstein et al., 1995)
Zea mays	*PG1*	polygalacturonase; *P2 Npg1*	(Niogret et al., 1991)
Zea mays	*3C12 Npg1*	polygalacturonase; *P2*	((Rogers et al., 1991); (Allen and Lonsdale, 1992); (Allen and Lonsdale, 1993)
Zea mays	*3A6/4H3/4H7*	------	(Rogers et al., 1991)
Zea mays	*Zm13*	kunitz trypsin inhibitor *lat52*	(Stinson et al., 1987)
Zea mays	*Zm58*	pectate lyase; *lat56; lat59, Tp10, Amba1*	(Stinson et al., 1987)

vegetative cell specific activity of the *lat52* promoter is maintained in *Arabidopsis thaliana*(Eady *et al.*, 1994). The transcriptional difference between the two cells in the pollen grain immediately after their formation in PMI, highlights the significance of PMI and the strict regulation of gene expression after this division. It is likely that many other of the pollen specific genes isolated to date are specifically activated in the vegetative cell considering the higher transcriptional activity of this cell in the pollen grain (section 1.2; (Wagner *et al.*, 1990)).

Most of the pollen specific genes isolated share homology to previously characterised genes. Genes showing identity to the cytoskeletal proteins actin (*tac25*, (Thangavelu *et al.*, 1993)), α-tubulin (*Tua1*, (Carpenter *et al.*, 1992)) and profilin (*Bet VII*, (Valenta *et al.*, 1991)) have been isolated, which most likely play a role in pollen tube growth. A number of genes bearing sequence identity to catabolic enzymes like polygalacturonases (*tp27*, (Rogers *et al.*, 1992a); *P2*, (Brown and Crouch, 1990); *3C12*, (Rogers *et al.*, 1991)) and pectate lyases (*Arb aI 1* to 3, (Rafner *et al.*, 1991); *lat56* and *lat59* (Wing *et al.*, 1989); *Zmc58*, (Stinson *et al.*, 1987)) and inhibitors of catabolic enzymes such as kunitz trypsin inhibitor (*Zmc13*, (Stinson *et al.*, 1987)) and amylase inhibitor (*Lolp lb*, (Singh *et al.*, 1991)) are thought to function in growth of the pollen tube through the style.

Regarding regulatory proteins a number of cDNAs with sequence identity to *myb* and *lim* putative zinc finger genes have now been isolated (*myb.Nt1* and *myb.Nt2*, section 4.2.1.1; *Ntl1* and *Ntl2*, section 4.2.1.2; *sf3*, (Baltz *et al.*, 1992a.)) which are the first pollen expressed transcription factors to have been isolated.

1.4 Regulation of gene expression
1.4.1 *Cis*-Regulatory elements which regulate gene expression in pollen

The use of deletion and mutation analysis of promoter-gus fusions in transgenic analysis in combination with transient expression assays has led to the identification of functional *cis*-regulatory sequences within the promoters of a small group of pollen expressed genes. The identification of groups of 'early', 'late' and 'intermediate' pollen expressed genes (section 1.3.3) has prompted the search for common *cis*-regulatory elements in the promoters of these groups of genes, which could direct their co-ordinate expression during pollen development.

Promoter sequences directing microspore specific expression have not yet been studied in detail. However a number of late pollen expressed promoters have been

investigated. One of the first was the identification of a 440 bp region in the *P. hybrida*
chi-A gene promoter which directed pollen specific expression of the *gus* reporter
gene, with activity appearing after PMI and increasing to maturity. However in this
case, the *chi-A* gene was regulated by another promoter region immediately
downstream of the pollen promoter to drive higher level expression in petals ((van
Tunen *et al.*, 1989); (van Tunen *et al.*, 1990)).

The promoters of three *lat* genes from tomato, *lat52*, *lat56* and *lat59* have been studied
in most detail ((Twell *et al.*, 1989a); (Twell *et al.*, 1990); (Twell *et al.*, 1991)). Two
regulatory *cis*-elements were identified as being important for promoter activity. A
'TGTGGTTATAT' sequence was present in the promoters of *lat52* and *lat56*
(Lat52/56 box). The 'TGTGGTT' core sequence called the PB-motif was repeated
three times in the *lat52* promoter (PBI, PBII and PBIII). A second *cis*-element
'GAAATTGTGA' was present in the promoters of *lat56* and *lat59* (Lat56/59 box).
These elements are located within 100 bp of the transcription start site and were
sufficient to drive pollen specific expression. The upstream regions of these promoters
controlled the level of reporter gene expression. Sequences identical to one or both of
the Lat52/56 and Lat56/59 boxes have been found in the promoters of the *Chi-A* gene
from *P. hybrida* (van Tunen *et al.*, 1990), *Zmc13* from maize (Guerrero *et al.*, 1990),
α-tubulin from *Arabidopsis* (Carpenter *et al.*, 1992) and *ntp303* from *N. tabacum*
(Weterings *et al.*, 1992).

A -97 deletion of the *Tua1* promoter still showed pollen-specific activity in transient
assays but a -39 deletion (removing a Lat56/59 box, but not a Lat52/56 box) abolished
activity (Carpenter *et al.*, 1992). Deletion of the Lat52/56 box had no effect on the
level of promoter activity. Removal of the Lat 56/59 box in the *lat59* promoter still
allowed pollen specific activity. These data suggests that the Lat52/56 and Lat56/59
boxes are not solely responsible for pollen specific activity. As the Lat52/56 box is not
present in the *lat59* promoter nor the Lat56/59 box in the *lat52* promoter this indicates
that these boxes do not by themselves coordinate expression of these genes in pollen.
Analysis of more pollen promoters will need to be done before it will be possible to
determine which sequences are necessary for pollen specificity.

Another functional *cis*-element was defined in the *ntp303* promoter (Weterings *et al.*,
1992). Two regions which when deleted resulted in a sequential loss of promoter
activity *in vitro* and *in vivo*, shared a 'AAATGA' sequence. These sequences are also
present in the same positions relative to the TATA box in the *Bp10* promoter from
Brassica napus (Albani *et al.*, 1992). Thus different pollen genes may be regulated by

different sequences and therefore presumably *trans*-acting factors, despite very similar 'late' gene expression patterns.

1.4.2 *Trans*-acting factors which regulate gene expression in pollen

The molecular mechanisms of pollen gene expression are unknown. Coordinately expressed genes may be regulated by a single *trans*-acting factor or different pollen genes may be regulated by different factors. In addition, different *trans*-acting factors may control early and late patterns of gene expression in the microspore and pollen grain. It is likely that the *trans*-acting factors that interact with pollen specific promoters are evolutionarily conserved in different species. The *lat52*, *lat56* and *lat59* tomato promoters function in *N. tabacum* and *Arabidopsis thaliana*(Twell *et al.*, 1990). Also, the maize *Zmc13* promoter and the *Brassica Bp4* and *Bp10* promoters function in *N. tabacum* ((Albani *et al.*, 1990); (Guerrero *et al.*, 1990)).

The most direct way to isolate *trans*-acting factors involved in pollen gene expression is to use defined *cis*-element sequences as binding site probes to pollen cDNA expression libraries (Singh *et al.*, 1989). Since the copy number of transcripts in the pollen grain is higher than other tissues this should increase the chance of isolating possible low abundance transcription factor cDNA clones. A more indirect method would be to isolate low abundance messages through cold plaque screening (Hodge *et al.*, 1991) or by heterologous screening using DNA sequences with identity to conserved DNA binding domains of known classes of transcription factors (section 1.4.4). It should also be possible to use the yeast one hybrid system to identify *trans*-acting factors that interact with pollen specific promoter fragments (Fields and Sternglanz, 1994).

A biochemical isolation of pollen factors would be difficult due to the difficulty of isolating large amounts of pollen nuclei. A genetic approach could be used by generating mutagenised *Arabidopsis* plants homozygous for a pollen promoter-*gus* fusion (McCormick, 1993). Plants harbouring mutations in genes that affect expression from these promoters could be identified by looking for plants exhibiting 50 % blue GUS staining of pollen rather than 100 % staining. Assuming the mutation is not also required for female gametophyte function then male lethal mutations could be recovered through the female line. Mutations in the promoter-*gus* fusion could then be distinguished from *trans*-acting factor mutations by outcrosses.

14

Prior to the work described in this thesis, a pollen specific cDNA, *sf3*,encoding a lim-only zinc-finger putative DNA-binding protein was isolated from sunflower inflorescence by differential screening (Baltz *et al.*, 1992a.). Heterologous screening of an *N. tabacum* pollen cDNA library using probes with sequence identity to the *myb* DNA-binding domain and the *sf3* LIM-domain was used to isolate two pairs of cDNA clones; *myb.Nt1, myb.Nt2* and *Ntl1, Ntl2* respectively.

1.4.3 Regulation of gene expression during pollen germination

In the light of the rapid growth of pollen tubes once hydration and germination of pollen has begun on the stigma surface (1cm/hr for *in vivo* germinated maize pollen tubes; (Barnabus and Fridvalszky, 1984)) and the demonstration that germination and early tube growth are independent of transcription, many mRNA species appear to be stored in the mature pollen grain and are translated during pollen germination. *Ntp303* transcript abundance increased two to three fold during germination *in vitro*, reaching a maximal level after 2 hours, declining after 5 hours to undetectable levels at 20 hours. However this increase in transcripts was correlated with transcription as determined by pulse labelling assays. Thus despite the fact that most *ntp303* transcript is presynthesised prior to dehydration of pollen, transcription still plays a role in germination (Weterings, 1994).

Studies have shown that late pollen genes are also translated during pollen development (Brown and Crouch, 1990). Dehydration in the later stages of pollen development would increase the cytoplasmic and hence ion concentration inside the pollen grain and stabilise 2^O and 3° RNA structures. In order to maintain translational efficiency under these conditions, transcripts with low secondary structure in their 5' untranslated leader sequences would be at an advantage. Indeed, a number of late pollen transcripts have low indices of secondary structure despite their relatively large size (80 to 273 bp compared to 30 to 50 bp average; (Joshi, 1987)). Alternatively, binding of protein factors to the untranslated leader sequence or an increase in translation initiation factor concentration in the cytoplasm could maintain translational efficiency under conditions of dehydration (Twell, 1994). The 5' untranslated leader of *lat52* is relatively long (110 bp) with a low propensity for secondary structure formation. Further, the leader has been shown to function as a translational enhancer in *in vitro* translation systems and in pollen (Twell, 1994).

1.4.4 Two classes of putative plant transcription factors which might be involved in pollen growth and development

A number of different classes of transcription factors have been characterised at a molecular and biochemical level in plants, including basic leucine zipper (bZIP), homeodomain, MADS box, MYB and MYC factors (for review (Ramachandran *et al.*, 1994)). In this thesis, it was decided to investigate the role of MYB-related factors in pollen development because they had been demonstrated to play a role in a variety of plant developmental processes including the differentiation and maintenance of the differentiated state of specialized cell types (eg. plant trichomes). In addition, DNA probes were available for use in the laboratory at the time this study was begun.

A novel plant LIM-domain gene encoding a zinc-finger protein, SF3, had been isolated from sunflower and was found to be expressed specifically in pollen ((Baltz *et al.*, 1992a.); (Baltz *et al.*, 1992b.)). It was decided to look for these proteins in *N. tabacum* pollen, using the *sf3* cDNA as a heterologous DNA probe, as a second class of putative transcription factor which might regulate the development of the male gametophyte.

1.4.4.1 MYB-related proteins

Myb genes were first identified in animal systems where it was shown that a *v-myb* oncogene present in two avian retroviruses, E26 and AMV were responsible for blood born leukemias in chickens (Graf, 1992). It was soon recognised that *v-myb* was derived from a cellular counterpart *c-myb* (Klempnauer *et al.*, 1982). *C-myb* genes have been isolated from a wide variety of organisms using the chicken *c-myb* probe, including mice, humans and more distantly related *Drosophila* (Katzen *et al.*, 1985) and yeast (Tice-Baldwin *et al.*, 1989). *C-myb* is unusual for a proto-oncogene in that it has a narrow pattern of expression being largely restricted to the nuclei of immature hematopoetic cells of the monocyte lineage (Graf and Beug, 1978). Subsequently two cDNAs isolated from a human T-cell library (A-myb and B-myb; (Nomura *et al.*, 1988)) are expressed more widely in kidney, colon and lymphoid tissues (A-myb) and all tissues tested (B-myb).

Two pieces of evidence suggested that *c-myb* was crucial for hematopoesis. First, antisense *c-myb* oligonucleotides were shown to inhibit normal regeneration of human bone marrow derived erythroid and myeloid colonies (Gerwirtz and Calabretta, 1988) and second, inactivation of *c-myb* in mice by homologous recombination showed

16

normal development until 14 days after which a severe impairment of erythro- and myelopoeisis was seen (Mucenski *et al.*, 1991). In addition, over-expression in mouse erythroleukemia (MEL) cell lines inhibits their terminal differentiation induced by dimethylsulphoxide (DMSO) (Weber *et al.*, 1990). These results indicate that *c-myb* affects the proliferation and differentiation of hematopoetic cells.

The first major evidence which suggested that c-MYB was a transcription factor was the demonstration that v-MYB bound specifically and as a monomer in filter binding assays to DNA fragments at a consensus 'AACG/TG' sequence called the myb recognition element (MRE) (Biedenkapp *et al.*, 1988). This element was shown to confer MYB inducibility to a test promoter in transient assays (Weston and Bishop, 1989).

Detailed deletion analysis of *c-myb* sequences defined a DNA-binding domain, transactivation domain and a negative regulatory domain (Figure 1.2; (Sakura *et al.*, 1989)). The DNA-binding domain consisted of three imperfect repeats, R1, R2 and R3 each 50-52 amino acids in length. R2 and R3 were essential for DNA-binding, whereas R1 was redundant. Subsequent alignment of the *myb* repeats from various *myb* genes led to the identification of three α-helical regions within each repeat similar to homeodomain proteins (Frampton *et al.*, 1991). Based on model building and mutational analysis, it was predicted that these α-helices are packed over a hydrophobic core with helices 2 and 3 on the R2 and R3 repeats, forming a tandem helix-turn-helix motif. Three conserved tryptophan residues in each repeat form the hydrophobic core over which the helices are positioned (Kanei-Ishii *et al.*, 1990). This model was confirmed and refined using nuclear magnetic resonance analysis of c-MYB bound to a MRE (Ogata *et al.*, 1994). The two third helices, one from each of the R2 and R3 repeats, were closely packed in the major groove of DNA and contacted each other to bind to the 'AACNG' element cooperatively.

The negative regulatory domain of c-MYB was defined from carboxy deletions which resulted from an increase in transactivation of the reporter gene in transient assays by four to six times. These region was found to contain a putative leucine zipper. At high concentrations c-MYB was shown to dimerise to form homodimers, which were unable to transactivate in transient assays, suggesting a novel autoregulation mechanism (Nomura *et al.*, 1993). In addition, c-MYB can bind to unrelated factors via the leucine zipper to abolish DNA-binding and acting as inhibitors of c-MYB function (Frova *et al.*, 1987).

17

Figure 1.2
Schematic representation of C-MYB protein structure.

Protein domains are indicated with double-headed arrows. R1, R2 and R3 represent
the MYB repeats with the MYB domain; $\alpha 1$, $\alpha 2$ and $\alpha 3$ represent α-helical regions
within each MYB repeat.

Scale bar indicates polypeptide length in amino acids.

Abbreviations: 'ACTIVATION' represents an acidic activation domain; 'NEGATIVE
REGULATION' represents a region within which are sequences such as a leucine
zipper which repress DNA binding.
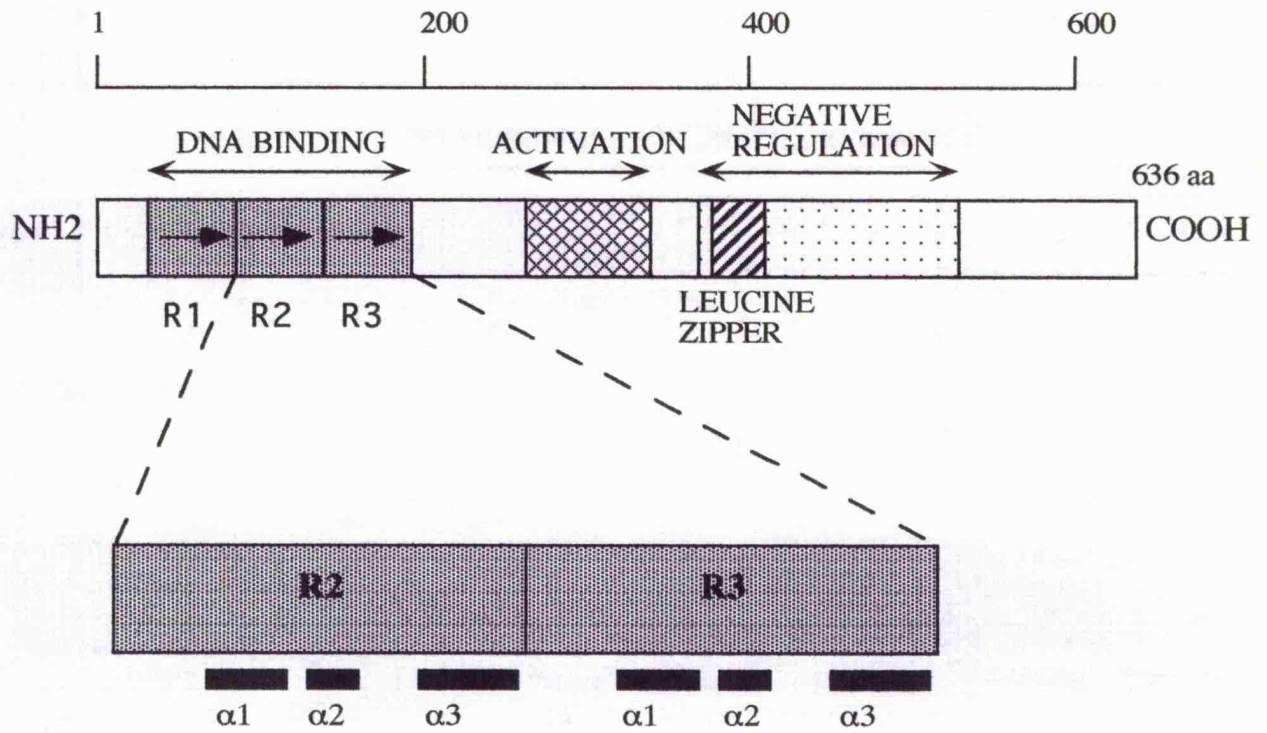
Another region of c-MYB which has been identified as being important for regulation of DNA-binding, are two conserved serine residues (serines 11 and 12) which are targets for casein kinase II (CKII) (Luscher and Eisenmann, 1990). Phosphorylation of the CKII site strongly inhibited sequence specific DNA-binding and removal of the phosphate increased DNA-binding. A comparison of the sequences of *v-myb* and *c-myb* revealed amino- and carboxy-truncations of *v-myb* which involved loss of the conserved CKII site and the negative regulatory domain respectively (Shen-Ong *et al.*, 1986).

The MRE sequence was identified in a number of animal promoters including *c-myb* and *c-myc* (Zobel *et al.*, 1991), *hsp70* (Foos *et al.*, 1993), DNA polymerase α (Sudo *et al.*, 1992), *CD34* (He *et al.*, 1992), *CD4* (Siu *et al.*, 1992) and *cdc2* (Ku, 1993). The presence of MREs in the promoters of cell cycle regulated genes suggests that c-MYB may play a part in the cell cycle. Indeed, downregulation of *c-myb* using antisense oligonucleotides blocked the cycling of human T-lymphocytes at the G1/S stage of the cell cycle (Gewirtz *et al.*, 1989), suggesting that c-MYB is needed as a progression factor to the S phase. Downregulation of the low level expression of *c-myb* in non-hematopoetic systems such as human embryonic fetal tissue, gut epithelial tissue and breast mammary tissue suggest an analogous role for *c-myb* in cell proliferation, differentiation and transformation (Thompson and Ramsey, 1995).

More recently, *myb* genes have been isolated from plants (Table 1.3), the first of which was the cloning of the *c1* locus in maize, previously identified from genetic studies as being important for anthocyanin biosynthesis, which contained a gene encoding a MYB protein (Paz-Ares *et al.*, 1987). Previously no transcriptional regulators had been discovered in plants. Plant MYB proteins differ from their animal counterparts in that their MYB DNA-binding domains are composed of two imperfect repeats R2 and R3. Exceptions to this rule are the MYBST1 protein isolated from potato (Baranowski *et al.*, 1994) and LIBP from parsley (Feldbrugge *et al.*, 1994) which each contain a single repeat and the CPM10 protein from the resurrection plant (*Craterostigma plantagineum*) which has a MYB domain composed of one and a half repeats (Iturriaga *et al.*, 1994). Putative transactivating regions are also present in plant MYB proteins, although the presence of a negative regulatory domain has not been identified.

Myb gene families have been isolated from plants, the first of which regulate phenylpropanoid biosynthesis (*C1, Pl, P, Zmc1, Zmc38* from maize (Cone *et al.*, 1986); (Cone *et al.*, 1993); (Franken *et al.*, 1994); (Grotewold *et al.*, 1994); (Paz-Ares

Table 1.3

List of plant *myb* genes isolated to date and their function where known.

species	name	function	reference
Antirrhinum majus	*Am305*	Regulation of phenylpropanoid biosynthesis	(Jackson *et al.*, 1991)
Antirrhinum majus	*Am306*		(Jackson *et al.*, 1991)
Antirrhinum majus	*Am308*		(Jackson *et al.*, 1991)
Antirrhinum majus	*Am315*		(Jackson *et al.*, 1991)
Antirrhinum majus	*Am330*		(Jackson *et al.*, 1991)
Antirrhinum majus	*Am340*		(Jackson *et al.*, 1991)
Antirrinum majus	*mixta*	Controls petal epidermal cell shape	(Noda *et al.*, 1994)
Arabidopsis thaliana	*atmyb1*		(Shinozaki *et al.*, 1992)
Arabidopsis thaliana	*atmyb2*	Gene regulation during dehydration	(Urao *et al.*, 1993)
Aradidopsis thaliana	*athmyb3*		(Heazlewood *et al.*, *unpubl.*)
Arabiopsis thaliana	*athmyb5*		(Heazlewood *et al.*, *unpubl.*)
Arabidopsis thaliana	*athmyb6*		(Li and Parish, 1995)
Arabidopsis thaliana *athmyb7*			(Li and Parish, 1995)
Aradidopsis thaliana	*gl1*	Initiation of trichome development and mantainance of the differentiated state	(Oppenheimer *et al.*, 1991)
Craterostigma plantagineum	*cpm10*		(Iturriaga *et al.*, 1994)
Hordeum vulgare	*GAmyb*	Gibberellin activated transactivation of aleurone a-amylase gene promoter	(Gubler *et al.*, 1995)
Hordeum vulgare	*Hv1*		(Marocco *et al.*, 1989)
Hordeum vulgare	*Hv33*		(Marocco *et al.*, 1989)
Nicotiana tabacum	*myb.nt1*		(Sweetman and Twell, *in prep.*)
Nicotiana tabacum	*myb.nt2*		(Sweetman and Twell, *in prep.*)
Petroselinum crispum	*Libp*		(Feldbrugge *et al.*, 1994)
Petunia hybrida	*Ph1*		(Avila *et al.*, 1993)
Petunia hybrida	*Ph2*		(Avila *et al.*, 1993)
Petunia hybrida	*Ph3*		(Avila *et al.*, 1993)
Petunia hybrida	*Pmyb92*		(Mur *et al.*, 1994)
Physcomitrella patens	*Pp1*		(Leech *et al.*, 1993)
Physcomitrella patens	*Pp2*		(Leech *et al.*, 1993)

species	name	function	reference
Potato	*MybSt1*		(Baranowski *et al.*, 1994)
Zea mays	*P*	Regulation of phenylpropanoid biosynthesis	(Grotewold *et al.*, 1991)
Zea mays	*Zm1*	Regulation of phenylpropanoid biosynthesis	(Marocco *et al.*, 1989)
Zea mays	*Zm38*		(Marocco *et al.*, 1989)
Zea mays	*ZmC1*	Regulation of phenylpropanoid biosynthesis	(Paz-Ares *et al.*, 1987)
Zea mays	*ZmP1*	Regulation of phenylpropanoid biosynthesis	(Cone *et al.*, 1993)

et al., 1987) and *Am305* from *Antirrinhum* (Sablowski *et al.*, 1994)). Genetic studies have defined a number of regulatory alleles that are required for the spatial and temporal expression of anthocyanin (purple) and phlobaphene (red) pigments in maize (Dooner *et al.*, 1991). As well as the C1/Pl gene family, which has sequence homology to c-MYB an R/B gene family (R, B and Lc) was identified which has sequence homology to another class of unrelated proto-oncogenes, MYC basic helix-loop-helix proteins. Microprojectile bombardment transient expression assays were used to show that one member of each of the two gene families is required for activation of the *bronze-1 (Bz1)* and *bronze-2 (Bz2)* gene promoters which are genes that form part of the anthocyanin biosynthetic pathway ((Bodeau and Walbot, 1992); (Goff *et al.*, 1990); (Goff *et al.*, 1991)). These proteins bind as heterocomplexes at composite DNA sites (Roth *et al.*, 1991). The combinatorial interaction of MYB and MYC plant factors is an interesting feature of plant transcription factor research.

Bz2 promoter activation by R and C1 was found to be sensitive to the ratio of both proteins. Over-expression of R resulted in a reduction of transactivation and increasing C1 expression overcame this interference. These observations suggest that R and C1 bind as a heterocomplex (Bodeau and Walbot, 1992). The heterocomplex formed between C1 and B proteins was investigated using deletions and GAL4 fusions ((Goff *et al.*, 1990); (Goff *et al.*, 1991)). The acidic activation domain of C1 could independently activate a *Bz1* promoter fused to the GAL4 DNA-binding domain. Fusions of the C1 DNA-binding domain with the GAL4 activator domain were dependant on functional B protein for coactivation. Further the amino-terminal acidic domain of B was necessary but not sufficient for coactivation when fused to the GAL4 DNA-binding domain. A study of B and C1 binding at a composite MYB/MYC binding site in the *Bz1* promoter suggests that B/C1 dimers bind with possible synergism at their targets. Neither homodimer could bind at the target site.

In contrast, the maize *P* gene binds alone at a 'CCT/AACC' sequence in the *A1* gene promoter, which is a gene common to both the phlobaphene and anthocyanin biosynthetic pathways. This sequence is different from the 'C/TAACGG' sequence bound by C1/R in the *A1* promoter, which is similar to the MRE defined in animal systems (Grotewold *et al.*, 1994). An *Am305* gene from *Antirrinhum* was shown to transactivate a reporter gene from a promoter containing the 'CCTACC' sequence identical to that used by P in the *A1* promoter. In protoplasts another core conserved sequence 'CACGTG' was needed for transactivation (Sablowski *et al.*, 1994). This sequence is found in the promoters of genes expressed in response to stimuli such as light, anerobisis and abscisic acid. These studies illustrate how complex *myb* gene

regulatory networks appear to be involved in the regulation of phenylpropanoid biosynthetic genes.

Two other *myb* genes have been assigned a function in plants. The *glabrous-1 (Gl-1)* gene from *Arabidopsis* is involved with trichome cell differentiation (Oppenheimer *et al.*, 1991). In addition, *Gl-1* is thought to maintain trichome cell differentiation as *Gl-1* transcripts are present in the differentiating trichome cells after expression in the surrounding leaf cells have decreased (Larkin *et al.*, 1993). The *Antirrinhum* gene *mixta*, has been shown to control shape in corolla epidermal cells (Noda *et al.*, 1994). The cells from a stable mixta plant line are flat in appearance rather than the conical shape seen in wild type corolla epidermal cells. This conical shape gives petals their brightness of colour by reflecting incident light. It is possible the *mixta* gene has an affect on epidermal cell shape through expression of phenylpropanoid derivatives which are present in the cell wall, in the light of *myb* induction of phenylpropanoid gene expression.

Myb genes have been isolated from a range of other plant species including barley (Marocco *et al.*, 1989), *P. hybrida* (Avila *et al.*, 1993), *Physcomitrella patens* (Leech *et al.*, 1993), *N. tabacum* (Sweetman and Twell, section 4.2.1), Loblolly pine (Campell *et al.*, 1994) and tomato (Lin *et al.*, 1996). However their physiological functions remain unknown. Clues to gene function can be obtained from their expression patterns; for instance the *Arabidopsis myb* gene *Atmyb-2* is induced by dehydration and salt stress, suggesting it may be involved in the regulation of water stress response genes (Urao *et al.*, 1993b), or by demonstrating transactivation from reporter genes from promoters containing identified binding sites; MYB.PH3 was shown to activate transcription from the *chsJ*, chalcone synthase gene promoter of *P. hybrida* in *N. tabacum* protoplasts from predicted *myb.Ph3* binding sites (Solano *et al.*, 1995a.), which suggests a role in regulation of phenylpropanoid biosynthetic genes.

1.4.4.2 LIM-related proteins

Genes encoding LIM-domain proteins were first isolated relatively recently in animal systems. The term 'LIM' is an acronym of the first letter of each of the three *lim* genes isolated: *lin-11* from *Caenorhabitis elegans (C. elegans)* (Freyd *et al.*, 1990), *Isl1* from rat (Karlsson *et al.*, 1990) and *mec-3* from *C. elegans* (Way and Chalfie, 1988). These proteins were found to contain a homeodomain motif and two LIM-domain

Figure 1.3
The LIM domain as a double zinc finger structure in linear
presentation.

'*': The third and fourth metal-coordinating residues differ from those shown in
ENIGMA, PAXILLIN, PINCH and LRG1 (H to C or D; C to H or E respectively).
'o': Spacing varies from the standard two to four in SF3, NTL1 and NTL2.

C-xx-C-x16-23-H-xx-C-xx-C-x16-21-Cxx-D-x
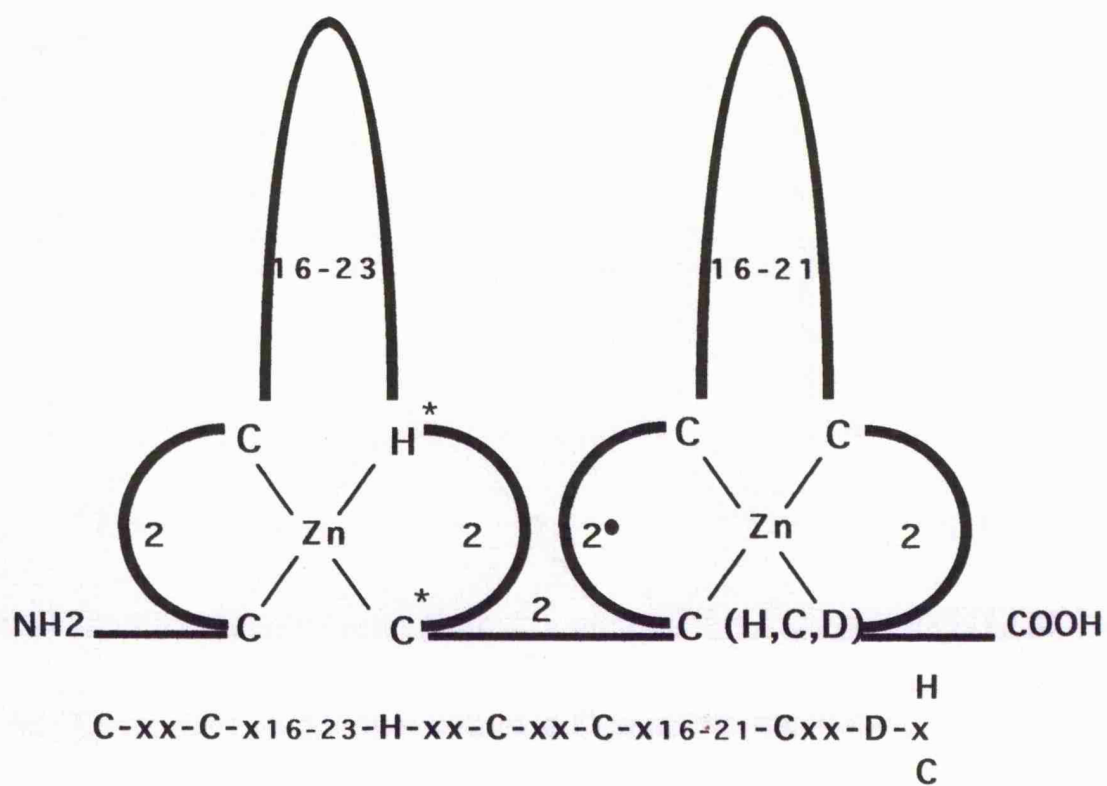
motifs pocessing a unique pattern of cysteine residues. Another class of LIM-domain proteins isolated later did not possess an accompanying homeodomain motif.

The LIM-domain is a 50 to 60 amino acid structural motif, defined by a conserved pattern of cysteine, histidine and in some cases an aspartate residue at the last metal chelating residue position (Figure 1.3). Each LIM-domain was found to coordinate two zinc ions via nuclear magnetic resonance and ultra-violet spectroscopy and reconstruction studies ((Michelson *et al.*, 1993); (Archer *et al.*, 1994); (Kosa *et al.*, 1994)), forming two zinc fingers separated by a two amino acid residue spacer. Subtle variations in this structure occur between LIM proteins isolated to date mainly in the length of the loops and the spacing between the two zinc fingers. The LIM-domain while similar to the RING-finger domain (Lovering *et al.*, 1993), forms a distinct sixth class of zinc finger structure (Sanchez-Garcia and Rabbitts, 1994).

Of the forty or so *lim* family genes isolated to date (Table 1.4), the proteins they encode have been classified into three main groups (Figure 1.4; (Taira *et al.*, 1995)) based on the presence and nature of the associated domains and the sequence similarities between the LIM domains. Most of the proteins identified so far have been described in animals, but *lim* genes have now been isolated from plants ((Baltz *et al.*, 1992a.); (Sweetman and Twell, unpublished) as well as in yeast (Tice-Baldwin *et al.*, 1989). However, no *lim* gene has been isolated from prokaryotes. The high degree of evolutionary conservation between proteins from such diverse species suggests an important and conserved functions for these domains.

Group 1 proteins contain a pair of LIM domains at an amino-terminal position. The LIM domains are separated by a relatively short interlim distance (about 8 residues). The amino- and carboxy-terminal LIM-domains fall into two distinct sequence classes (type A and B respectively), which are clearly seen on a dendrogram constructed from computer sequence analysis of LIM-domain sequences (Dawid *et al.*, 1995). Further, the A and B type LIM-domains are always associated with one another on the same protein, which suggests an important biological role for these domains. Of the group 1 proteins, most are associated with a homeodomain motif positioned at a carboxy-position to the LIM-domains. Most of the LIM-homeodomain (LHX) proteins have been implicated either through mutational or direct biochemical analysis to control differentiation of different cell types.

The *lhx* group of genes contains the three defining *lim* genes. The *lin-11* gene encodes a protein that controls certain asymmetric cell divisions during vulval development in *C. elegans* (Freyd *et al.*, 1990). In a lin-11 mutant, the daughter cells are identical and

Table 1.4

List of *lim* genes isolated to date and their function where known.

species	name	homology/class/function	reference
Acanthocheilonema vitae	*l3-1 (avl3-1)*	3	(Oberlaender *et al.*, unpubl.)
Caenorhabditis elegans	*ceh14*	1	(Burglin and Ruvkun, unpubl.)
Caenorhabditis elegans	*lin-11*	1, control of asymmetric cell division in vulval cells	(Freyd *et al.*, 1990)
Caenorhabditis elegans	*mec-3*	1, formation of mechano-sensory nuerons in 1 of 2 daughter cells of an asymm-etric division. Cooperation with POU domain protein	(Way and Chalfie, 1988)
Chicken	*crp*	*hcrp*, 2	(Crawford *et al.*, 1994)
Chicken	*isl1*	1	(Tsuchida *et al.*, 1994)
Chicken	*isl3*	1	(Tsuchida *et al.*, 1994)
Chicken	*lim1*	1	(Tsuchida *et al.*, 1994)
Chicken	*lim3*	1	(Tsuchida *et al.*, 1994)
Chicken	*zyxin*	3, interacts with cyto-skeletal proteins and CRP	(Sadler *et al.*, 1992)
Chinock Salmon	*isl-2a, isl-2b, isl-3, sl-1*, 1		(Gong and Hew, 1994)
Drosophila melangaster	*apterous*	*rlh2, hlh2*, 1, wing pattern and a subset of abdominal muscle formation	(Cohen *et al.*, 1992)
Drosophila melangaster	*dmlp1*	*mlp*, 2, muscle development	(Arber *et al.*, 1994)
Helianthus annus	*sf3*	*ntl1,ntl2*, 2	(Baltz *et al.*, 1992a.)
Homo sapiens	*enigma*	3	(Wu and Gill, 1994)
Homo sapiens	*hcrp*	*ccrp*, 2	(Liebhaber *et al.*, 1990)
Homo sapiens	*limk*	*limk-1*, 1	(Mizuno *et al.*, 1994)
Homo sapiens	*lh-2*	*apterous*, 1	(Wu and Gill, 1994)
Homo sapiens	*paxillin*	3, associated with cytoskeleton	(Salgia *et al.*, 1995)
Homo sapiens	*pinch*	3	(Rearden, 1994)
Homo sapiens	*lmo1 (rbtn/ttg)*	1	((McGuire *et al.*, 1989); (Boehm *et al.*, 1990))
Homo sapiens	*lmo3*	1	(Foroni *et al.*, 1992)
Homo sapiens	*lmx1*	1, activation of insulin gene promoter in cooperation with a bHLH protein	(German *et al.*, 1992)

species	name	homology/class/function	reference
Mouse, Rat	*crip*	2	(Birkenmeier and Gordon, 1986)
Mouse	*gsh-4*	1, postnatal survival	(Li *et al.*, 1994)
Mouse	*lhx1*	1	(Fujii *et al.*, 1994)
Mouse	*lhx3*	1	(Zhadanov *et al.*, 1995)
Mouse	*lhx5*	*xlim-5*, 1	(Bertuzzi *et al.* unpubl.)
Mouse, *Homo sapiens*	*lmo2*	1, erythroid development	(Royer-Pakora *et al.*, 1991)
Mouse	*p-lim*	1	(Bach *et al.*, 1995)
Nicotiana tabacum	*ntl1, ntl2*	*sf3*, 2	(Sweetman and Twell, *in prep.*)
Quail	*crp*	2	(Weiskirchen and Bister, 1991)
Rabbit	*esp2*	2	(Wang, *unpubl.*)
Rat	*crp2*	2	(Okano *et al.*, 1993)
Rat	*esp1*	2	(Nalik, *unpubl.*)
Rat	*isl-1*	1	(Karlsson *et al.*, 1990)
Rat	*isl-2*	1	(Tsuchida *et al.*, 1994)
Rat	*rlim1*	1	(Furuyama *et al.*, 1994)
Rat	*rlim2/rlim5*	1	(Tsuchida *et al.*, 1994)
Rat	*limk-1*	*limk*, 1	(Nunoue *et al.*, 1995)
Rat	*limk-2*	1	(Nunoue *et al.*, 1995)
Rat	*lh2*	*apterous*, 1	(Xu *et al.*, 1993)
Rat	*mlp*	*dmlp*, 2, muscle development	(Arber *et al.*, 1994)
Rat	*ril*	3	(Kiess *et al.*, 1995)
Rat	*ril18*	3	(Kiess, *unpubl.*)
Saccharomyces cerevisiae	*lrg1p*	Ungrouped	(Muller *et al.*, 1994)
Xenopus laevis	*xlim1*	1, negative regulator of neural cell differentiation in Spemann organizer	(Taira *et al.*, 1992)
Xenopus leavis	*xlim2 (xlim5)*	*lhx5*, 1	(Taira *et al.*, 1992)
Xenopus laevis	*xlim3*	1	(Taira *et al.*, 1993)
Zebrafish	*isl1, 2, 3*	1	(Gong *et al. unpubl.*)
Zebrafish	*lim1, 3, 5, 6*	1	(Dawid *et al. unpubl.*)

Figure 1.4

Schematic representation of LIM proteins (adapted from (Taira *et al.*, 1995)).

LIM domains (shaded), glycine-rich sequence (black), distinct LIM domain (asterisk). Kinase and GAP domains are indicated.

Abbreviation: HD, homeodomain.

Group 1

ISL-1, LHX, MEC-3

TTG (LMO)

kinase   LIM kinase

A   B

Group 2

CRIP, DMLP1

NTL, SF3, CRP, MLP

C   C

Group 3

RIL

YKR090W

ENIGMA

★   ZYXIN

PAXILLIN

PINCH
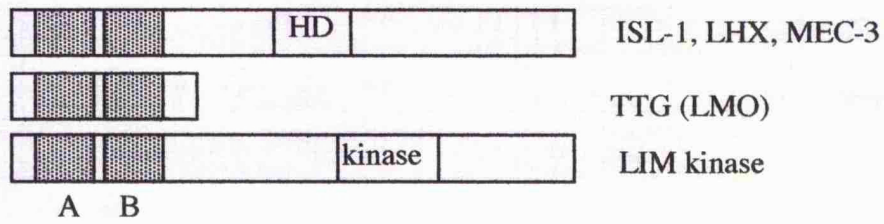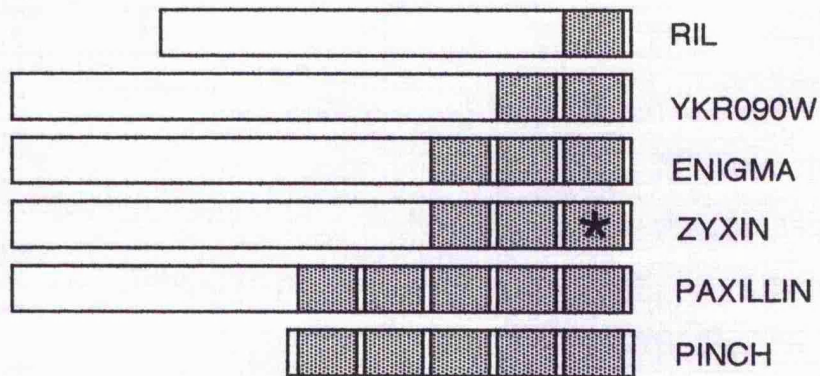
Ungrouped (yeast GAP proteins)

GAP   DBM1

GAP   LRG1
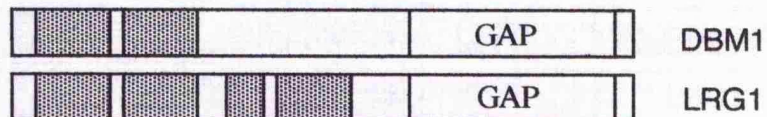
as a result fail to form a functional vulva (Ferguson *et al.*, 1987). However, the *lin-11* gene is expressed in other tissues and implies this gene has other functions in *C. elegans*. The *Isl-1* gene is expressed in developing and mature islet cells and was isolated through expression library screening of a rat pancreatic cell library using the insulin enhancer complex as a DNA-binding site probe (Karlsson *et al.*, 1990). This suggests that the ISL-1 protein is involved with the cell specific expression of the insulin gene. *Isl-1* is also expressed widely in other tissues (Dong *et al.*, 1991) and studies suggest that the ISL-1 protein may also be involved in establishing motor neuron fate in the developing rat brain ((Thor *et al.*, 1991); (Korzh *et al.*, 1993)). The *mec-3* gene in contrast to *lin-11* and *Isl-1* has a restricted pattern of spatial expression, being limited to touch neuron precursor cells in *C. elegans* (Way and Chalfie, 1988). Synergistic activation of the *mec-3* promoter was observed in transient assays with expression of MEC-3 and UNC-86 test plasmids (Lichtsteiner and Tijan, 1995). UNC-86 is a DNA-binding POU-domain protein also expressed in touch neuron precursor cells. Whereas UNC-86 could bind alone to the promoter to elicit low level expression of a reporter gene, MEC-3 could not bind DNA. Further, both proteins could interact with each other to form a heterocomplex combining the two activation domains from each protein to form a strong activator.

Synergistic interactions between LHX proteins and unrelated proteins were also seen in activation of the insulin gene promoter. Weak activation was seen with the LMX1.1 LIM-domain protein and stronger activation seen with a truncated LMX1.1 protein which had the LIM-domains removed. This suggests that the LIM-domains have a negative regulatory function. However, much higher activation was achieved with co-expression of full length LMX1.1 and a PAN-1 helix-loop-helix protein (HLH; (German *et al.*, 1992)). Evidence that the LIM-domains suppress DNA binding of the homeodomain motif was also seen in the case of ectopic expression of the XLIM-1 LHX protein from *Xenopus laevis* (Taira *et al.*, 1994). Expression of full-length XLIM-1 protein in *Xenopus laevis* embryos elicited a small effect on the conversion of ectodermal cells into cells with organising properties (Spemann organiser region). However, ectopic expression of protein with mutations of the LIM-domains elicited large scale conversion of the ectodermal cells with organising properties which induced neighboring ectodermal cells to neural differentiation and neighbouring ventral mesodermal cells to muscle differentiation. This study also suggests that LIM-domains inhibit the binding of the homeodomain. Interactions with other proteins may activate XLIM-1 and other LHX proteins presumably by altering the conformation of the LIM-domains in relation to the homeodomain.

27

Another theme of LHX proteins seems to be their widespread expression suggesting they elicit multiple functions. This is seen in the case of APTEROUS, a LHX protein expressed in *Drosophila* ((Cohen *et al.*, 1992); (Williams *et al.*, 1994)). APTEROUS is expressed in wing and haltere imaginal discs and has been shown to be responsible for wing pattern formation. In addition, APTEROUS is expressed in muscle precursor cells in the embryo. In apterous mutants, a subset of abdominal muscles fail to form and ectopic expression of APTEROUS generates muscle differentiation. Expression is also seen in interneural cells of the nervous system (Lundgren *et al.*, 1995). In apterous mutants these neurons choose incorrect pathways and fail to form connections with one another.

Other classes of Group 1 proteins include LIM-kinase proteins expressed mainly in the brain in rat and humans ((Nunoue *et al.*, 1995); (Mizuno *et al.*, 1994)). The unique structure of these proteins suggests that they may have a role in an uncharacterised signally pathway in the brain. The LIM-only group 1 proteins LMO1 and LMO2 (formally TTG1 and TTG2) are associated with chromosomal translocations in T-cell leukemias ((McGuire *et al.*, 1989); (Boehm *et al.*, 1990)). In mice, disruption of the LMO2 gene by homologous recombination causes death as a result of an early block in erythropoesis (Warren *et al.*, 1994).

Group 2 proteins contain one or two copies of a single sequence type LIM-domain with a relatively larger interlim region of about 55 amino acids. Most of these proteins contain a glycine rich region in their LIM-domains, except the small group of proteins isolated from plants (SF3, (Baltz *et al.*, 1992b.); NTL1 and NTL2, Sweetman and Twell, unpublished). Group 2 proteins have been shown to be involved with interactions with identical or different LIM-domain proteins in order to regulate the function of other proteins. Thus the human cysteine rich protein (hCRP) has been shown to homo- and hetero-dimerise in cell-based systems and in *in vitro* blot overlay assays via the LIM-domains (Feuerstein *et al.*, 1994). This dimerisation activity could be transferred to an unrelated protein via fusion of a single LIM-domain of hCRP. Dimerisation was independent of DNA but dependant on coordination of zinc, providing evidence of the involvement of the LIM-domains as a protein-protein dimerisation motif.

The hCRP protein was also capable of heterodimerisation with ZYXIN, a Group 3 LIM-only protein which possesses three tandem LIM domains at a carboxy position (Schmeichel and Beckerle, 1994). Both of these proteins are associated with actin microfilaments in adherent fibroblast cells. *In vitro* blot overlay assays were used to show a specific interaction between hCRP and ZYXIN. Again zinc was crucial for

heterodimerisation, illustrating that the secondary structure of the LIM-domain is necessary for the protein-protein binding function.

A positive role in muscle development has been demonstrated for the Group 2 LIM proteins MLP from rat and its *Drosophila* homolog DMLP 1 (Arber *et al.*, 1994). These genes are expressed in the nuclei and cytoplasm of muscle cells during myogenic differentiation. Overexpression and antisense repression of MLP expressed in cultured myogenic cells provides evidence for a positive role for MLP in myogenesis. Other LIM-only proteins could substitute for MLP in overexpression experiments, suggesting that the specificity of the LIM-domain interactions does not play a part in myogenesis. Specificity could be provided here by the precise temporal regulation of LIM-only protein in muscle cells.

A group of Group 2 LIM-only proteins have been isolated in plants; SF3 from sunflower ((Baltz *et al.*, 1992a.); (Baltz *et al.*, 1992b.)) and NTL1 and NTL2 from *N. tabacum* (Sweetman and Twell, unpublished). These proteins are specifically expressed in pollen. In the light of the functions of other Group 2 proteins the plant proteins probably function in pollen via protein-protein interactions. However, non-sequence specific DNA binding has been demonstrated for SF3 via Southwestern blot overlay assays (personal communication, Dr. A. Steinmetz).

Group 3 LIM-only proteins contain LIM-domains that are more heterogenous in sequence than other Group 1 or 2 domains and are localised at a carboxy-terminal position. In addition, Group 3 proteins have putative cytoplasmic functions. For instance, ZYXIN and PAXILLIN are located at focal adhesion plaques, are associated with the cytoskeleton ((Sadler *et al.*, 1992); (Turner, 1994)) and possess three and five LIM-domains respectively. ZYXIN has been shown to interact with hCRP (see above) and α-actinin (Crawford *et al.*, 1992), which suggests that ZYXIN may be involved with cytoskeletal connections between adhesion plaques via interaction with CRP. PAXILLIN has been shown to bind with other proteins including vinculin, pp125^{FAK} and proteins with SH2 and SH3 domains (Turner and Millar, 1994), although the participation of the LIM-domains in these interactions has not been demonstrated. The three LIM-domain ENIGMA protein has been shown to bind the endocytosis recognition protein motif of the insulin receptor via the LIM-domains (Wu and Gill, 1994).

In conclusion, LIM proteins have a range of different functions, from transcription factors (LHX proteins) to cytoplasmic components (Group 3 LIM-only proteins) and are combined in the same protein molecule with very different motifs. LIM-domains are responsible for homo- and hetero-dimerisation with LIM-domains and other

protein motifs. In addition, functional cooperation is seen between LIM-domain and HLH proteins. Further the LIM-domain can inhibit the function of the homeodoamin. Together, these pieces of evidence suggest a role for the LIM-domain in protein-protein interactions. Although LHX proteins act as DNA-binding transcription factors, LIM-only proteins probably function entirely by modulating the activity of other proteins or as structural bridging components. While the LIM zinc fingers are unique in not being able to bind to DNA, no direct evidence for sequence specific binding exists. However the possibility that LIM-domains do function through DNA-binding cannot be ruled out.

1.5 An overview of thesis aims

The overall aim of this project was to isolate and characterise *trans*-acting protein factors involved in the expression of genes during development of the male gametophyte in *N. tabacum*. The first part of this investigation involved the isolation of cDNA clones encoding such protein factors. A cDNA library was made from poly A$^+$ RNA from mature pollen of *N. tabacum*. An initial attempt at isolating *trans*-acting factors directly by using the cDNA library as an expression library, and screening it with double-stranded DNA binding site probes containing pollen-specific *cis*-element sequences proved unsuccessful. A second attempt, using the more conventional approach of DNA homology screening, resulted in the isolation of two classes of cDNA clones encoding putative MYB and LIM DNA-binding protein domains. The first section of this thesis (Chapters 3 and 4) describe the production of the cDNA library and the isolation of cDNA clones. The spatial and temporal expression patterns of the two classes of cDNA clones in relation to pollen development are discussed.

The second part of the project was to investigate the function of the protein factors in pollen development. To achieve this, one of each of the *myb* and *lim* cDNA clones was introduced into *N. tabacum* linked to either a pollen-specific promoter or a highly active 'constitutive' promoter. The information derived from the analysis of transgenic plants in relation to the function of the protein factors in pollen development is discussed.

Chapter Two.

Materials and methods.

All solutions containing ingredients quoted as percentages are weight to volume or volume to volume ratios.

2.1 Sources of molecular biology reagents, enzymes and plant tissue culture chemicals

Chemicals and reagents were purchased from Sigma Chemical Company Ltd., or from BDH Ltd. Enzymes were obtained from BRL, Boehringer Mannheim, Stratagene, Perkin Elmer, Pharmacia or NBL. Tissue culture salts and hormones were purchased from Flow Laboratories. Agar and agar based media were from Difco Laboratories.

2.2 Plant material

Material for nucleic acid isolation was obtained from *Nicotiana tabacum (N. tabacum)* var SR1, *N. sylvestris* and *N. tomentosiformis* plants grown under normal greenhouse conditions.

Extraction buffer:

25 mM	sodium citrate, pH 7.0
4 M	guanidinium isothiocyanate
1.5 %	sodium lauryl sarcosine
100 mM	ß-mercapto-ethanol

Material for the RNA extraction of microspores and pollen was obtained from *N. tabacum* cv. Samsun grown outside in soil. Buds were collected and sorted into defined length groups. The contents of the anthers were squeezed out between fingers into an isotonic 0.3 M mannitol solution until all the anthers had been collected. Anthers from each length group were homogenised in extraction buffer by mechanical disruption with a ground glass cylinder and plunger. The suspensions were filtered through a nylon sieve (125 μm), microcentrifuged at 150 xg for 1 minute and the isolated cells stored on ice. Aliquots of 10 μl were taken and the number of cells estimated by microscopy. The remaining parts were stored at -80 °C until RNA extractions were performed (section 2.7.1).

2.3 Bacterial culture and storage
2.3.1 Media for the growth of bacteria

NA and NB were made according to the manufacturers instructions and then bottled and autoclaved.

NZY medium (per litre):

5 g	sodium chloride
2 g	magnesium sulphate (hydrated)
5 g	Bacto-yeast extract
10 g	caesine hydrolysate

Made to pH 7.5 with sodium hydroxide, bottled and autoclaved.

NZY medium was solidified by the addition of 1.5 % agar (technical number 3). The agar was dissolved by steaming the medium before autoclaving.

Top agar: the above NZY medium was solidified by the addition of 0.7 % agarose.

2xYT medium (per litre):

16 g	Bacto-tryptone
10 g	Bacto-yeast extract
5 g	sodium chloride

Made to pH 7.0 with sodium hydroxide, bottled and autoclaved.

2.3.2 Strains and genotypes
2.3.2.1 Eschericia coli (E. coli)

XL1-Blue: recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, lac, {F', proAB, lacI^q, ZDM15, Tn10, (tet^r)}.

SURE: e14$^-$(McrA$^-$) Δ(mcrCB-hsdSMR-mrr)171 endA1 supE44 thi-1 gyrA96 relA1 lac recB recJ sbcC umuC::Tn5 (Kan^r) uvrC [F' proAB lacI^{q}ZΔM15 Tn10 (Tet^r) Amy Cam^r]

SOLR: e14⁻ (mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, uvrC, umuC::Tn5(kanʳ), lac, gyrA96, relA1, thi-1, endA1, λ^R [F'proAB lacI^{q}ZΔM15, Tn10 (tetʳ)] Su⁻ (non-suppressing)

DH10B: F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) ø80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ(ara, leu)7697 galU galK rpsL nupG

2.3.2.2 Agrobacterium tumefaciens (A. tumefaciens)

LBA4404: containing Ti plasmid pAL4404.

2.3.2.3 Lambda vector

λZAPII: λsbhlλ1^o, chiA131 (T, amp, ColE1, ori, lacZ, T3 promoter-polycloning site-T7 promoter), I, srlλ3^o, clts857, srlλ4^o, nin5, srlλ5^o, Sam100

2.3.3 Antibiotics for bacterial selection

Name	Concentration (μg/ml) for :-	
	E.coli	*A. tumefaciens*
Ampicillin	100	-
Kanamycin	100	50
Rifampicin	-	50
Tetracyclin	12.5	-

2.3.4 Growth of cultures

Using a sterile micropipette tip a single colony from a bacterial plate was transferred to 5 ml of NB or NZY medium containing the appropriate concentration of selective antibiotics (table in section 2.3.3) in a sterile universal. The culture was placed in an orbital shaker at 37 °C *(E. coli)* or 30 °C *(A. tumefaciens)* overnight or until the culture reached the required optical density.

Single colonies were obtained by dipping a flame sterilised and cooled into an overnight culture and streaking across the surface of solid medium, containing selective antibiotics. The medium was allowed to cool to 50 °C before the antibiotics

34

were added and plates poured. Plates were incubated upside down, at the appropriate temperature until colonies formed. Cultures could be kept on a plate for about a month if kept at 4 °C.

2.3.5 Long term storage of cultures

750 μl of an overnight culture was mixed with 250 μl of sterile 100 % glycerol in NB medium in a Cryogenic storage tube. After labelling the solution was flash frozen in liquid nitrogen and stored at -80 °C. To revive cells stored in this way, a portion of the frozen mixture was scraped off the surface using a flame sterilised loop and resuspended in 5ml of medium and grown as described in section 2.3.3.

2.3.6 Preparation of competent *A. tumefaciens* cells

$T_{10}E_1$:

10 mM	Tris-CI, pH 7.5
1 mM	EDTA

A colony of *A. tumefaciens* taken from a streak plate at 4 °C was used to inoculate 10 ml of 2xYT media and grown with shaking at 30 °C overnight. The overnight culture was used to seed 200 ml of 2xYT medium and grown with shaking at 30 °C for 3-4 hours. Cells were harvested by centrifugation at 3, 000 xg for 10 minutes at 4 °C. The supernatent was poured off and the pellet washed in 20 ml of ice cold $T_{10}E_1$, pH 7.5 and finally resuspended in 20 ml ice cold 2xYT and dispensed in 500 μl aliquots into cold microcentrifuge tubes. The cells were flash frozen in liquid nitrogen and stored at -80 °C. Cells treated in this way remained competent for 6-12 months.

2.3.7 Transformation of *A. tumefaciens* with plasmid DNA

pMOG402 (Mogen) is a binary vector based on a wide host range replicon. It can therefore replicate in both *E. coli* and *A. tumefaciens*, which means that foreign DNA can be inserted into the vector and then screened in *E. coli* prior to transfer to *Agrobacterium*. Competent *A. tumefaciens* cells prepared by the procedure above were thawed on ice and approximately 0.5 μg of plasmid DNA was added to an aliquot of cells and mixed by inverting the tube. The cells were incubated on ice for 5 minutes and in liquid nitrogen for 5 minutes. The cells were then thawed at 37 °C, 1 ml of

prewarmed 2xYT medium added and incubated with shaking at 30 °C for 2 hours. Cells were plated out in 200 μl aliquots onto LB media plates with 50 μg/ml kanamycin to select for cells transformed with the recombinant binary vector. Plates were incubated for 2 days at 30 °C, by which time colonies could be picked and grown up in liquid culture and checked for the presence of binary vector by carrying out a small scale isolation of plasmid DNA (section 2.3.7).

2.3.8 Small scale isolation of *A. tumefaciens* plasmid DNA

The method followed was essentially the same as the plasmid mini-prep procedure described in Section 2.8.8, with the exception that 4.5 ml of overnight culture was used in the preparation of plasmid DNA. A relatively large volume of culture was used because binary vectors are present at low copy numbers in bacteria.

2.4 Plant tissue culture
2.4.1 Media for plant cell and tissue culture

MSO: Based on MS salts (Murashige and Skoog, 1962). Preweighed packets marketed by Flow Lab contained all the ingredients for this medium with the exception of sucrose. Sucrose was added at a concentration of 30 g/l. The pH was adjusted to 5.8 and the medium bottled and autoclaved.

MSD4x2: MSO salts with the following additions, NAA (0.1 mg/l), 6-BAP (1 mg/l), sucrose (30 mg/l). The pH was adjusted to 5.8 and the medium bottled and autoclaved.

The above media based on MSO salts was solidified by the addition of 0.8 % agar which was dissolved by steaming the medium before autoclaving.

2.4.2 *N. tabacum* transformation

Young *N.tabacum* (SR1) leaves were placed in a sterile Pyrex casserole dish and covered in 10 % bleach (Domestos). After 10-15 minutes the leaves were thoroughly rinsed with 3x1L of sterile tap water. Working in a sterile petri dish half filled with liquid MSD4x2 medium in a lamina flow cabinet, the leaves were cut into 1 cm squares, avoiding the major veins, using a sterile scapal blade and flame sterilised

forceps. The leaf squares were put into another sterile petri dish containing a solution of *A. tumefaciens* diluted 1/20 with liquid MSD4x2 medium for 5 minutes. After draining on the lid of the sterile petri dish the squares were transferred, lower leaf surface against the agar onto MSD4x2 medium. The plates were sealed with Nescofilm and incubated for two days in a growth room with continuous light. After this incubation period, when colonies of *A. tumefaciens* were visible, the leaf discs were transferred to MSD4x2 medium containing kanamycin (50 μg/ml) and cefertaxin (200 μg/ml), to select for transformed plant cells and inhibit the growth of *A. tumefaciens*. The plates were sealed with Nescofilm and returned to the growth room. After approximately 16 days, kanamycin resistant shoots and callus formed along the cut edges of the leaf disc explants. When the shoots had developed 2-3 well formed leaves, they were cut away from the callus with flame sterilised scapal blade and forceps and the stem pushed into MSO medium with kanamycin (50 μg/ml) and cefertaxin (200 μg/ml), contained within small sterilised pots. The shoots were cultured under the same conditions as before until they had developed a good root system. The shoots were then transferred to soil and grown under greenhouse conditions. The trays of young shoots were covered with a transparent plastic hood for 1-2 days to keep the atmosphere humid until they had become more established.

2.4.3 Collection and germination of transgenic seed

Transgenic *N. tabacum* flowers that had just opened were either crossed with wild type pollen, allowed to self naturally, or pollen from dehisced anthers of transgenic flowers used to outcross to wild type flowers. The non-dehisced anthers were removed with forceps prior to crossing. The flower heads were then bagged to prevent pollination by other plants.

Transgenic seed was surface sterilised by submerging for 1 minute in 70 % ethanol and then 15 minutes in 10 % bleach (Domestos). The seeds were then rinsed thoroughly with sterile tap water before spreading 50-100 seeds over the surface of MSO medium containing kanamycin (50 μg/ml) in a sterile petri dish.

2.5 Analysis of DNA-protein interactions via Southwestern blotting
2.5.1 Micro protein extraction

Extraction buffer:

0.1 M	Tris-HCl, pH 8.0
5 mM	EDTA
5 mM	DTT

3xSDS sample buffer:

200 mM	Tris-HCl, pH 6.8
6 %	SDS
30 %	glycerol
15 %	β–mercatoethanol
0.05 %	bromophenol blue

500 mg of desiccated pollen, stored at -80 °C was ground to a slurry with 90 μl of extraction buffer in a ground glass cylinder and plunger. The slurry was transferred to a sterile eppendorf tube, mixed by vortexing for 15 seconds and extracted repeatedly with an equal volume of phenol/chloroform (section 2.7.1), until a clear interface was obtained. The supernatent was combined with SDS sample buffer and boiled for 3 minutes prior to storage at -20 °C or separation of proteins by SDS-polyacrylamide gel electrophoresis (see section 2.5.3).

2.5.2 Approximate protein quantification via Coomassie staining

Coomassie stain:

0.1 %	Coomassie blue dye
25 %	methanol
10 %	acetic acid

Destain:

25 %	methanol
10 %	acetic acid

Total protein concentration was determined via a Coomassie blue dot binding assay. 5 μl of sample was dotted at neat, 1:1 and 1:9 dilutions onto Whatmann 3mm paper. A range of BSA standards at 1-10 mg/ml were dotted next to the samples. When dry the filter was immersed in Coomassie stain for 10-15 minutes on a slow shaking platform.

The filter was destained for 10 minutes in destain. Protein concentrations were estimated by comparison to the BSA standards.

2.5.3 Separation of proteins via polyacrylamide electrophoresis

SDS running buffer:

25 mM	Trisma base
192 mM	glycine
0.1 %	SDS

Native loading buffer:

50 %	glycerol containing 0.05 % bromophenol blue

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins according to molecular weight. SDS sample buffer (section 2.5.2) was combined with an aliquot of protein and the solution boiled for 5 minutes. The sample could then be stored at -20 ^{o}C or separated on a gel. Frozen samples required boiling for 1 minute prior to loading. Gel mixture was made up according as follows. 15 ml of gel mix was sufficient to cast two gels in the Anachem Protein system.

% gel	volume of stock solution (ml)			
	30 % acrylamide 0.8 % bis acryl.	water	1.5 M Tris-HCl pH 8.8	10 % SDS
10	5	6.1	3.75	0.15
11	5.51	5.59	3.75	0.15
12.5	6.25	4.85	3.75	0.15

100 μl of ammonium sulphate (100 mg/ml) and 10 μl TEMED were added to polymerise the gel.

The percentage of acrylamide in the gel could be varied depending upon the expected size and/or the required resolution of the particular protein of interest. Once the

resolving gel had been poured into the apparatus, propanol:water (1:1) was used as an overlay to exclude oxygen and help the gel set. Once set, the propanol:water was rinsed off the top of the gel and the surface dried. The stacking gel, made according to the recipe below, was added and the well forming comb pushed into place.

volume of stock solution (ml)

30 % acrylamide 0.8 % bis acryl.	water	1.5 M Tris-HCl pH 8.8	10 % SDS
1.67	6.9	1.25	0.1

Denaturing polyacrylamide gels were run in SDS running buffer at 150 V until the dye front had reached the bottom of the gel. Molecular weight size markers (10 kDa, Gibco-BRL) were included on the gel to determine the sizes of the proteins of interest. Proteins were stained by submerging the gel in Coomassie stain (section 2.5.2) for approximately 30 minutes on a rocking platform and then destained for 1 hour. Alternatively the proteins could be blotted from the gel onto an immobilising membrane for immunoblotting.

2.5.4 Electroblotting proteins after separation by SDS-PAGE

Anode buffer 1:

	0.3 M	Trisma base
	10 %	methanol

Anode buffer 2:

	25 mM	Trisma base
	10 %	methanol, pH 10.4

Cathode buffer:

	25 mM	Trisma base
	40 mM	β–aminohexanoic acid
	20 %	methanol, pH 9.4

Ponceau stain:

0.6 %	Ponceau S
1 %	acetic acid

TBS:

50 mM	Tris-HCl, pH 7.4
200 mM	sodium chloride

TBS-Tween 20:

TBS containing 0.1 % Tween 20

Proteins separated by SDS-PAGE were electroblotted onto a nitrocellulose membrane (Sartorius) using a Milliblot SDE system (Millipore). A sandwich of Whatmann 3MM paper, gel and membrane was set up on the anode. Six pieces of filter paper were cut to the same size as the gel and two pieces were soaked in anode buffer 1 and placed on top of the anode, one filter soaked in anode buffer 2 was placed on top followed by the pre-wet membrane and then the gel and finally three pieces of filter paper soaked in cathode buffer. Transfer was accomplished after 45 minutes at 250 mA.

Correct transfer of proteins to the membrane could be determined by Poncaeu staining. The stain was poured onto the membrane and left for a few seconds. The stain was then rinsed off with water to reveal the bands of the protein, the molecular weight markers were marked with a pencil and the stain was washed off with TBS-Tween.

2.5.5 Screening with DNA binding-site probes

Blocking buffer:

2.5 %	dried milk powder
25 mM	Hepes, pH 8.0
1 mM	DTT
10 %	glycerol
50 mM	sodium chloride
0.05 %	LDAO
1 mM	EDTA

Binding buffer:

10 mM	Tris-HCl, pH 7.5
50 mM	sodium chloride

41

1 mM	EDTA
1 mM	DTT
10 μg/ml	sheared and denatured
	herring sperm DNA

Washing buffer:

same as above without the herring sperm DNA

Nitrocellulose filters containing 10 μg of electroblotted protein extracts were immersed in 50 ml blocking buffer and incubated at 4 °C overnight.

The filters were laid protein side up, one at a time in 5 ml of binding buffer containing ^{32}-P end-labelled double-stranded concatenated binding-site probe at a concentration of approximately 1×10^6 cpm/ml (section 2.6.5.2). The filters were incubated for 1 hour at room temperature with gentle shaking. The filters were then removed to a tray containing 50 ml washing buffer and washed at room temperature for 10 minutes with gentle shaking. The washing step was repeated twice more and the filters wrapped damp in cling film and exposed to X-ray film (Hyperfilm MP, Amersham) at -80 °C in a cassette with an intensifying screen.

2.6 Identification of pollen specific mRNAs via cDNA cloning

2.6.1 Construction of an *N. tabacum* pollen cDNA-expression library

Mature pollen was collected from *N. tabacum* var. Samsun plants grown in a glasshouse. A cDNA library was constructed in lambda ZAPII (Strategene) from 5 μg of mature pollen polyA$^+$ RNA (section 2.7.2) according to the manufacturer's instructions. The primary cDNA library was amplified once in SURE™ strain.

2.6.2 Plating out the cDNA library

E. coli cells (XL-1 blue strain) competent for phage adsorption were produced by inoculating 50 ml of NZY medium (containing 0.2 % maltose and 10 mM magnesium sulphate) with a colony from a stock streak plate. Once at an OD_{600} of 0.5 this culture was centrifuged for 10 minutes at 1,000 xg and the pellet resuspended in 50 ml of 10 mM magnesium sulphate and stored for up to 5 days at 4 °C. The lambda phage was

plated out on a lawn of competent cells by mixing 3 ml of cells with the desired number of plaque forming units (pfu) from the library and incubated at 37 °C for 15 minutes to permit adsorption of the phage particles to the cells. 30 ml of top agar at 50 °C was added to the cell/phage suspension, mixed and immediately poured onto solid NZY medium in 22.5 cm square petri plates (Nunc), ensuring an even coating over the surface. Once dry the plates were inverted and incubated for 8-16 hours at 37 °C for library screening with DNA probes.

2.6.3 Screening the pollen cDNA library with DNA probes
2.6.3.1 Screening with double-stranded DNA probes

Denaturing solution:

	0.5 M	sodium hydroxide
	1.5 M	sodium chloride

Neutralising solution:

	3.0 M	sodium chloride
	0.5 M	Tris-HCl, pH 7.4

20X SSC:

	3 M	sodium chloride
	0.3 M	sodium citrate, pH 7.0

Hybridisation solution:

	0.25 M	sodium phosphate, pH 7.2
	7 %	sodium dodecyl sulphate (SDS)

Wash solution I:

	20 mM	sodium phosphate, pH 7.2
	5 %	SDS

Wash solution II:

	20 mM	sodium phosphate, pH 7.2
	1 %	SDS

Plates were cooled to 4 °C in the refrigerator to harden the top agar prior to plaque lifts being taken. Squares of "Zetaprobe" (Bio-Rad) nylon membrane were laid onto plates for 2 minutes, during which time orientation marks were made by tracing three asymmetric crosses made in pencil on the membrane, onto the back of the plate. Membranes were laid plaque-side-up on filter paper soaked in denaturing solution for 7 minutes, then transferred similarly to paper soaked in neutralising solution for 3 minutes and then again in fresh neutralising solution for another 3 minutes. The membranes were finally rinsed briefly in 2xSSC, and DNA bound to the filter by baking at 80 °C under vacuum for 30 minutes. Plaque lifts were prehybridised for 5-30 minutes at 40, 55 or 65 °C (depending on whether low, medium or high stringency was required), in order to get even wetting of the membrane. Boiled radioactively labelled probe was then added and hybridised overnight at the same temperature. The filters were washed twice for 30 minutes each in wash solution I at the hybridisation temperature. A further 2 washes with wash solution II at the hybridisation temperature were carried out for high stringency hybridisation, or if a more stringent wash was required. A hand held radiation monitor was used to measure the removal of radioactivity. After washing the filters were wrapped damp in cling film and exposed to X-ray film (Hyperfilm MP, Amersham) at -80 °C in a cassette with an intensifying screen. Autoradiograms were developed after 2-16 hours depending on the signal. Hybridising plaques were lined up to the corresponding position on the original plate.

2.6.3.2 Expression library screening with DNA binding-site probes
((Singh *et al.*, 1989); (Vinson *et al.*, 1988))

Blocking buffer:

2.5 %	dried milk powder
25 mM	Hepes, pH 8.0
1 mM	DTT
10 %	glycerol
50 mM	sodium chloride
0.05 %	LDAO
1 mM	EDTA

Binding buffer:

10 mM	Tris-HCl, pH 7.5
50 mM	sodium chloride

1 mM	EDTA
1 mM	DTT
10 μg/ml	sheared and denatured
	herring sperm DNA

Washing buffer:

same as above without the herring sperm DNA

The pollen cDNA library was plated out at a density of $5x10^4$ pfu/nunc plate as described in section 2.6.2. This lower plaque density is needed so that plaque size and thus signal intensity is maximised. The plates were incubated at 42 °C for about 3 hours until tiny plaques were visible. Hybond-C nitrocellulose filters pre-soaked in 10 mM IPTG for 30 minutes and air dried, were overlaid onto each phage plate and the plates incubated further for 6 hours at 37 °C. Non-duplicate lifts were taken of each plate as described in section 2.6.3.1. 10 filters were immersed in 500 ml blocking buffer and incubated at 4 °C overnight.

The filters were laid protein side up, one at a time in 100 ml of binding buffer containing ^{32}P end-labelled double-stranded concatenated binding-site probe at a concentration of approximately $1x10^6$ cpm/ml (section 2.6.5.2). The filters were incubated for 1 hour at room temperature with gentle shaking. The filters were then removed to a large tray containing 500 ml washing buffer and washed at room temperature for 10 minutes with gentle shaking. The washing step was repeated twice more and the filters wrapped damp in cling film and exposed to X-ray film (Hyperfilm MP, Amersham) at -80 °C in a cassette with an intensifying screen.

In order to recover DNA binding activity of potential transcription factors that may have been lost due to insolubility or incorrect protein folding of the bacterially produced fusion protein, the filter were then processed through a denaturation/renaturation procedure. The filters (up to 10) were immersed one at a time in 500 ml 6M guanidinium hydrochloride and incubated for 10 minutes at room temperature with gentle shaking. Half of the denaturing solution was removed and replaced with an equal volume of washing buffer. The filters were then incubated for a further 10 minutes. This dilution was repeated every 10 minutes a further 7 times, finally discarding the solution and replacing with washing buffer. The filters were then blocked, probed and washed as above.

Putative hybridising phage plaques identified from either of the two probing steps were cored and eluted as described in section 2.6.4. A second round of screening,

45

using 9 cm petri plates and the same radioactive probe from the primary screening, was carried out to identify true positively hybridising plaques.

2.6.3.3 Screening with oligonucleotide probes

Hybridisation solution:

6x	SSC (section 2.6.3.2)
0.1 %	sodium dodecyl sulphate (SDS)
0.02 %	ficoll
0.02 %	poly vinyl pyrrolidine (PVP)
50 μg/ml	sheared and denatured herring
	sperm (DNA added fresh before use)

Wash buffer:

3x	SSC
0.5 %	SDS

Duplicate lifts of phage plates were processed as above. Filters were prehybridised for 1 hour in hybridisation solution at 37 °C. Purified ^{32}P end-labelled oligonucleotide(s) (section 2.6.5.3.2) were added to 5 ml of hybridisation solution at a concentration of 25 pm/ml and the probe mixture added to each of the filters, and hybridised at 50 °C overnight. Filters were then washed in wash buffer at 50 °C, once for 5 minutes and twice for 30 minutes respectively. After washing the filters were processed for autoradiography as above.

2.6.4 Storage of isolated lambda phage clones

SM buffer:

100 mM	sodium chloride
17 mM	magnesium sulphate
50 mM	Tris-HCI, pH 7.5
0.01 %	gelatin

Clones were isolated by coring plaques out of plates using a sterile Pasteur pipette and ejecting phage particles into 500 μl SM buffer containing a drop of chloroform to prevent bacterial growth. These phage suspensions can be stored for several years at 4

^{o}C. In cases where cores were taken from densely plated phage, a second and even third round of plating and screening was carried out in order to obtain a single hybridising plaque.

2.6.5 Radioactive labelling of DNA probes for hybridisation
2.6.5.1 Radioactive labelling of a double-stranded DNA probe for hybridisation

Random priming (Feinberg and Vogelstein, 1983) was used as the method of radioactively labelling double-stranded DNA molecules for use as probes in hybridisation. A Prime IT™ II Random Primer Labelling Kit (Strategene) was used which involved annealing random nonamer (9-mer) primers to the probe DNA (100-1000 nucleotides in length) and incorporation of alpha ^{32}P-dCTP in the extension products using a 3-prime exonuclease-deficient mutant of the Klenow fragment of DNA polymerase I [Exo(-) Klenow]. Probes with a specific activity of approximately 1×10^{9} dpm/μg were routinely made. Probes were purified using NucTrap™ (Strategene) probe purification columns which removed unincorporated nucleotides.

2.6.5.2 Radioactive labelling of a double-stranded DNA binding-site probe for hybridisation

10 μg of a concatenated double-stranded DNA binding-site fragment was used in the following Nick translation reaction:

0.5 M	Tris-HCL, pH 7.5
0.1 M	magnesium sulphate
200 μCi	α^{32}P-dATP (5000 Ci/mM)
1.0 mM	dCTP
1.0 mM	dGTP
1.0 mM	dTTP
1.0 mM	dATP
10 U	*E.coli* DNA polymerase I (Klenow fragment)

The above reaction was incubated at room temperature for 30 minutes and stopped by adding 1 μl of 0.5 M EDTA. The DNA was extracted with phenol/chloroform and

precipitated 1/10 volume 3 M sodium acetate and 2.5 volumes ethanol. The DNA pellet was resuspended in 50 μl distilled water and used in the expression screening of the pollen cDNA library (approximately 1×10^8 cpm/μg).

2.6.5.3 Radioactive labelling of an oligonucleotide probe for hybridisation
2.6.5.3.1 Purification of oligonucleotides

Deprotected oligonucleotides were purified from an ammonium solution using butanol extraction. 1 ml of n-butanol was added to 100 μl of oligonucleotide solution and mixed by vortexing for 15 seconds. The precipitated oligonucleotide was collected by microcentrifugation for 1 minute and the smear resuspended in 100 μl distilled water. The oligonucleotide was butanol extracted again as before, the smear dried and resuspended in 100 μl distilled water. The concentration of oligonucleotide in solution was estimated by spectrophotometric determination of absorbance at 260 nm.

2.6.5.3.2 5-prime end labelling of oligonucleotides

10x Polynucleotide kinase buffer:

500 mM	imadozole chloride pH 6.4
180 mM	magnesium chloride
50 mM	DTT
1 mM	spermidine chloride
1 mM	EDTA pH 8.0

T4 Polynucleotide Kinase (T4 PK) was used to add γ^{32}P-dATP to the 5-prime terminal phosphate position of a purified oligonucleotide probe. The reaction was set up and incubated at 37 °C for 60 minutes. 1 μl of 0.5 M EDTA was added to stop the reaction.

Typical Reaction:

5 μl	γ^{32}P-dATP (250 μCi)
5 μl	10x buffer
2.2 μl	T4 PK (20 Units)
10 μl	deprotected oligonucleotide (1-50 pM termini)

48

<div align="center">

27.8 μl   distilled water

</div>

Probes were purified using Chroma Spin-10 probe purification columns (Clontech) which removed unincorporated nucleotides.

2.6.6 *In vivo* excision of pBluescript plasmids

Plaques of interest from the amplified cDNA library were cored from the agarose plate and transferred to a sterile microcentrifuge tube containing 500 μl of SM buffer (section 2.6.4) and 20 μl of chloroform. The tube was vortexed to release the phage particles and incubated at room temperature for 1-2 hours. In a sterile 50 ml conical flask the following reagents were combined;

200 μl	OD$_{600}$ XL1-Blue cells
100 μl	phage stock
1 μl	ExAssist helper phage
	(Strategene, >1 x 10^6 pfu/ml)

The mixture was incubated with shaking for 15 minutes. 3 ml of 2xYT media (section 2.3.1) was added and the flask incubated for 2 hours with shaking. The culture was heated at 70 °C for 20 minutes and spun for 15 minutes at 4, 000 xg. The resulting filamentous phage stock supernatent was decanted into a sterile tube and stored at 4 °C. 1 μl of the phage stock was added to 200 μl of SOLR cells (Strategene, OD$_{600}$ =1) in a microcentrifuge tube and incubated at 37 °C for 15 minutes. 100 μl was plated on LB agar plates containing 100 μg/ml ampicillin and incubated overnight at 37 °C. Colonies which appeared contained pBluescript SK(-) double-stranded phagemid. For long term storage of the pBluescript plasmid, glycerol stocks were made from liquid cultures grown from bacterial colonies from the selection plate (section 2.3.5).

2.7 Analysis of gene expression by Northern blotting
2.7.1 Pollen total RNA extraction

Extraction buffer:

25 mM	sodium citrate, pH 7.0
4 M	guanidinium isothiocyanate
1.5 %	sodium lauryl sarcosine

<div align="center">

49

</div>

100 mM		ß-mercapto-ethanol

The stock solution minus ß-mercapto-ethanol was filter sterilised and stored at 4 °C.

Phenol/chloroform:	1	Tris equilibrated phenol (Fisions)
(volume to volume ratios)	1	chloroform
	24	iso-amyl alcohol

The stock solution was mixed and left to separate overnight at 4 °C. The bottom organic phase was used in subsequent extractions.

The method followed was essentially that described in (Verwoerd et al., 1989). 250-500 mg of desiccated pollen was ground to a slurry with 500 μl of extraction buffer in a ground glass cylinder and plunger. The slurry was transferred to a sterile eppendorf tube, mixed by vortexing for 15 seconds and extracted with an equal volume of phenol/chloroform (section 2.7.1). The aqueous phase was transferred to a fresh tube and extracted twice more with phenol/chloroform. 400 μl of 6 M lithium chloride was mixed with 400 μl of aqueous phase and incubated on ice for 1 hour. Precipitated nucleic acids were pelleted by microcentrifugation at 4 °C, resuspended in 1 ml 3 M lithium chloride and pelleted again. The wash was repeated. The pellet was resuspended in 400 μl 2 % potassium acetate and heated at 55 °C for 10 minutes to dissolve the RNA. Insoluble matter was pelleted by microcentrifugation and the supernatant transferred to a fresh tube. RNA was precipitated with the addition of 1 ml ethanol and incubation at -80 °C for 15 minutes. The RNA was pelleted by microcentrifugation, dried under vacuum, resuspended in DEPC treated water and stored at -80 °C. Approximately 3 mg of total RNA was extracted from 1 g of dessicated mature tobacco pollen.

2.7.2 Preparation of poly A+ RNA

Poly A+ RNA or messenger RNA (mRNA) was prepared using Dynabeads (Dynal UK) according to the manufacturers instructions. The procedure involved the binding of poly A+ RNA 3-prime residues with oligo (dT) residues covalently coupled to magnetic beads. Other RNA species without poly A+ RNA tails such as rRNA and tRNA are not bound. The beads are captured using a magnet and the poly A+ RNA washed. The poly A+ RNA is eluted into distilled water yielding an essentially pure

fraction. Spectrophotometric determination of poly A$^+$ RNA concentration was achieved by measuring the absorbance of the pure poly A$^+$ RNA fraction at 260 nm, using cuvettes that had been soaked in concentrated hydrochloric acid and methanol (1:1) for 1 hour and then washed extensively in DEPC treated water. An approximate yield of 2.4 % poly A$^+$ RNA was obtained from mature dessicated tobacco pollen.

2.7.3 Scanning spectrophotometry of nucleic acids

Nucleic acids absorb light maximally at ultra violet wavelengths, with an absorption peak at 260 nm. The measurement of OD_{260} can therefore be used to determine the concentration. An OD_{260} of 1.0 is equivalent to 40 μg/ml RNA solution, 50 μg/ml double-stranded DNA solution and 33 μg/ml oligonucleotide solution. Nucleic acid solutions were diluted into 200 ul of distilled water, placed in a quartz cuvette and scanned across the range 200-300 nm. A sharp peak at 260 nm preceded by a trough at 220-240 nm indicated a pure nucleic acid preparation. A reading at 320 nm indicated background in the scan, which could be subtracted from the 260 nm reading to gain a more accurate estimation of concentration.

2.7.4 Separation of RNA by agarose gel electrophoresis

10x MOPS buffer:

	0.2 M	3-[N-Morpholino]-propane-sulphonic acid
	50 mM	sodium acetate, pH 7.0
	10 mM	EDTA

TAE buffer:

	0.4 M	Tris-acetate, pH 8.0
	10 mM	EDTA

Loading buffer:

	50 %	glycerol containing 0.1 mg/ml bromophenol blue

RNA was denatured at 65°C for 5 minutes in the following solution:

6 μl	RNA (5-10 ug)
12.5 μl	formamide (deionised)
2.5 μl	10x MOPS buffer
4 μl	formaldehyde (37 %)

The solution was chilled on ice before combining with 2.5 μl loading buffer and loading onto a 1.5 % agarose gel made with TAE buffer containing 0.2 μg/ml ethidium bromide. The gel tank, well forming comb and gel tray were cleaned thoroughly and wiped out with alcohol before use. The RNA was electrophoresed in TAE buffer until the bromophenol dye front was about 3 cm from the bottom of the gel. The ethidium bromide intercalates with the nucleic acid and therefore allows visualisation by fluorescence when illuminated with a u. v. transilluminator. The RNA could therefore be visualised before capillary blotting onto Zetaprobe.

2.7.5 Northern gel blotting

The ethidium bromide stained gel was equilibrated for 20 minutes in 20x SSC (section 2.6.3) before capillary blotting onto Zetaprobe. A capillary blot was set up as follows: the gel was placed on top of two pieces of Whatman 3 MM filter paper on a sponge standing in a tray of 2x SSC buffer. The remaining surface of the sponge and filter paper was covered with plastic cling film. The Zetaprobe membrane was then placed on top of the gel ensuring that no air bubbles were trapped underneath. Two pieces of filter paper soaked in transfer buffer were placed on top of the membrane, and two dry pieces of filter paper on top of these. A stack of paper towels and a weight was laid over the gel. Transfer of the buffer from the tray through the gel to the towels was left to occur overnight. The filter was then dried and the DNA bound to the filter by u. v. cross linking for 90 seconds.

2.7.6 Quantification of RNA blot signals

RNA blot signals were quantified using an Molecular Dynamics Image Quant scanning densitometry system. Gel blots wrapped in cling film were exposed to a scanning densitometry screen in a cassette for a time period of hours to overnight, and the gel image scanned, visualised and the individual signals quantified using the scanning densitometer.

52

2.7.7 Hybridisation of DNA probes to RNA gel blots

RNA gel blots were hybridised to DNA probes and washed as described for hybridisation of double-stranded DNA probes to phage lifts in section 2.6.3.1.

2.8 Restriction enzyme and Southern blot analysis of DNA
2.8.1 Digestion of DNA with restriction endonucleases

Restriction enzymes were bought with an accompanying buffer and used as recommended by the manufacturers. Up to 2 μg of plasmid DNA was incubated for 1-16 hours at the appropriate temperature with the enzyme and buffer, and 1 μg RNase A. The enzyme activity was destroyed by phenol/chloroform (section 2.7.1) extraction or by heat denaturation.

2.8.2 Separation of DNA by agarose gel electrophoresis

Loading buffer:

0.25 %	bromophenol blue
0.25 %	xylene cyanol
25 %	Ficoll (type 400)

DNA fragments can be separated on agarose gel matrices. 2 μl loading buffer was added to the DNA sample prior to loading on the gel. The agarose gels were made by dissolving a variable concentration of agarose (0.8-4.0 %, depending on the size range of fragments expected) in TAE buffer (section 2.7.3), by heating to 95 $^\circ$C. The agarose was allowed to cool before adding 0.2 mg/ml ethidium bromide and pouring into a casting tray. The gels were submerged in TAE buffer and run at 80 V, with the appropriate molecular weight markers included on the gel. The separation of the DNA fragments could be monitored by visualisation with a u. v. transilluminator.

2.8.3 Southern blotting

Depurinating solution:

0.25 M	hydrochloric acid

53

Before blotting, agarose gels containing DNA fragments greater than 1 kb in size were equilibrated by submerging in depurinating solution for 7 minutes, denaturing solution (section 2.6.5) for 30 minutes and neutralising solution (section 2.6.5) for 30 minutes. The gel was rinsed in distilled water between treatments. The capillary blot was then set up as described for northern blotting (section 2.7.4). After baking under vacuum for 30 minutes at 80 °C, hybridisation with radioactively labelled probes was carried out as described in section 2.6.3.

2.8.4 Hybridisation of DNA probes to DNA gel blots

DNA gel blots were hybridised with DNA probes as described for hybridisation of double-stranded DNA probes to phage lifts in section 2.6.3.1.

2.9 Manipulation of DNA for cloning of fragments into plasmid vectors
2.9.1 Large scale DNA preparation

Solution I:

50 mM	sucrose
25 mM	Tris-HCl, pH 8.0
10 mM	EDTA, pH 8.0

Solution II:

0.2 M	sodium hydroxide
1 %	SDS

Solution III:

5 M	potassium acetate, pH 5.2

A 500 ml overnight culture of the bacterial strain carrying the plasmid of interest was pelleted by centrifugation at 1,000 xg for 10 minutes. The cells were resuspended in 10 ml of solution I, to lyse the cells. To this suspension 2 volumes of solution II were added and mixed by gentle swirling. To this 1.5 (original) volumes of solution III were added. The bacterial debris was spun down at 3,000 xg for 10 minutes and the plasmid precipitated from the supernatant with an equal volume of isopropanol.

Nucleic acids were collected by centrifugation at 3,000 xg for 10 minutes and resuspended in 0.5 ml distilled water per 100 ml culture.

High molecular weight RNA's were precipitated by the addition of an equal volume of 5M lithium chloride and removed by centrifugation at 3,000 xg for 5 minutes. The remaining nucleic acids were precipitated with an equal volume of isopropanol, collected by centrifugation for 10 minutes, rinsed in 70 % ethanol and the pellet dried under vacuum. The pellet was then dissolved in 100 μl of distilled water containing 10 μg/ml RNase A and left for 1 hour at 37 $^\circ$C. 100 μl of 1.6 M sodium chloride containing 13 % polyethylene glycol 8000 was added to the DNA solution and left for 45 minutes on ice. The DNA solution was extracted with phenol/chloroform (section 2.7.1) until a clear interface was observed, and the DNA finally precipitated by the addition of an equal volume of 10 M ammonium acetate and 2 volumes of ethanol. The plasmid was collected by microcentrifugation, rinsed in 70 % ethanol and dried before dissolving in the required volume of sterile distilled water. The concentration of plasmid DNA was calculated after the OD_{260} had been measured.

2.9.2 Purification of DNA from agarose gels

DNA fragments could be used for cloning and preparation of radioactive labelled probes after purification from agarose gel slices following electrophoresis. DNA fragments greater than 200 bp in size were recovered from agarose gels using a "Geneclean II" DNA purification kit (Bio 101 Ltd.), while fragments less than 200 bp were isolated using a "MERmaid" DNA purification kit (Bio 101 Ltd.). These procedures involved the binding the of DNA to silica glass suspensions in high salt conditions, washing the glass fog to remove traces of agarose and eluting the DNA into water. The kit was used as per manufacturers instructions. Fragments purified in this way were ready for ligation into vectors, or enzyme modifications prior to ligation.

2.9.3 Dephosphorylation of vector 5-prime termini

After digestion of the plasmid DNA with restriction endonuclease(s), a portion of the digest was run on an agarose gel to confirm complete digestion of the plasmid DNA. To the remaining DNA was added;

1 μl	calf intestinal phosphatase (0.5 U)
10 μl	10x phosphatase buffer
	distilled water to 100 μl

The reaction was incubated for 30 minutes at 37 °C. The plasmid DNA was purified by extracting with phenol/chloroform, precipitation with sodium acetate and ethanol and washing with 70 % ethanol. DNA pellets were vacuum dried for 5 minutes and resuspended in an appropriate volume of distilled water to approximately 100 ng/μl.

2.9.4 Ligation of DNA fragments

Purified DNA was mixed in approximately equimolar ratios of vector and insert. 2 units of T4 DNA ligase and 5x ligation buffer (supplied with the enzyme from GIBCO) were added.

Typical ligation reaction:

Vector DNA	x μl
Insert DNA	y μl
5x Ligation Buffer	2 μl
T4 DNA Ligase	1μl (2 U)
Water	to 10 μl

Ligations were left at 15 °C overnight before transformation into competent *E. coli*.

2.9.5 Transformation of *E. coli* with plasmid DNA
2.9.5.1 Preparation of electro-competent cells

A single colony of *E. coli* strain DH10B from a streak plate at 4 °C was used to inoculate 10 ml of 2xYT medium with 100 μg/ml ampicillin and grown overnight. The overnight culture was used to seed 2x400 ml flasks of 2xYT media with ampicillin as above and the cultures grown to an OD_{600} of 0.3 to 0.4 to ensure the cells are growing rapidly (log phase). The cultures were chilled on ice for 15 minutes and the cells pelleted at 3,000 xg for 10 minutes at 4 °C. The pellets were resuspended in 5 ml of chilled sterile 10 % glycerol and the volume made up to 400 ml with 10 % glycerol. The cells were washed in this way thrice more in a 400, 40 and 20 ml volume and

56

finally resuspended in the equivalent of 1 packed cell volume of 10 % glycerol (approximately 500 μl). The cells were dispensed in 20 μl aliquots into chilled eppendorf tubes and frozen in a dry ice and ethanol bath and stored at -80 °C.

2.9.5.2 Transformation procedure

SOC recovery medium (per litre):

20 g	Bacto-tryptone
5 g	yeast extract
0.5 g	sodium chloride
0.1 %	glucose
0.001 M	magnesium chloride
0.001 M	magnesium sulphate

Both magnesium salts were added after autoclaving.

Transformation of *E. coli* was achieved by electroporation using a BioRad Gene Pulser. A 20 μl aliquot of competent DH10B cells (section 2.9.5.1) was thawed on ice and DNA added in a 1 μl volume. DNA from a ligation mix was purified, by extracting with phenol/chloroform (section 2.7.1), ethanol precipitated and resuspended in 1 μl of distilled water. The DNA cell mix was placed in a chilled 1 cm electroporation cuvette and placed in the Gene Pulser electroporation chamber. With the Pulser set at 25 μF, 1.6 kV and the pulse controller set at 200 Ω, the cells were pulsed once which yielded a time constant of around 4.5 msec. 1 ml of SOC medium was added immediately to the cuvette and the cells resuspended. The contents of the cuvette was transferred to a disposable culture tube and the tube shaken for 1 hour at 37 °C to let the cells recover. Fractions of the cell suspension were plated on LB plates with the appropriate antibiotic selection and the plates incubated inverted at 37 °C overnight. Colonies were picked and grown up in liquid culture or plates lifted in colony hybridisation's to check for the presence of the required recombinant plasmid.

2.9.6 Identification of recombinant plasmids in transformed colonies

A colour assay using X-GAL and IPTG is the most convenient method of identifying colonies containing recombinant plasmids. Vectors which can be used contain a multiple cloning site within the *lacZ* gene. If the plasmid vector is present in a selected

host strain which lacks the *lacz* gene, the plasmid *lacZ* gene can complement for it. The *lacZ* gene product is a subunit of ß-galactosidase and this facilitates the metabolism of the indigogenic substrate X-GAL when IPTG is present. The resulting blue product is easily visible in colonies. When the *lacZ* gene in the vector is interrupted by an insert, recombinant colonies are white and thus can be distinguished from non-recombinants. For colour selection, 20 μl of 40 mg/ml IPTG in water, and 20 μl of 40 mg/ml X-GAL in dimethylformamide were spread onto plates prior to spreading cells.

2.9.7 Colony hybridisation

When colour selection with IPTG and X-GAL was not possible, large numbers of numbers of colonies were screened by transferring them onto a Zetaprobe membrane, lysing the bacteria and probing with the required radioactively labelled DNA. A nylon membrane disc was laid onto the bacterial plate for 2 minutes, during this time orientation marks were made. The membrane was then transferred colony side up onto Whatman 3MM filter paper soaked in 2xSSC containing 0.5 % SDS and left for 2.5 minutes. The membrane was then baked under vacuum at 80 $^{\circ}$C for 30 minutes to fix the DNA. The membrane was then hybridised as for plaque lifts (section 2.6.3). The colonies were recovered by replacing the bacterial plate in the 37 $^{\circ}$C incubator for about 6 hours. Putative recombinants which gave a radioactive signal were checked by small scale isolation of DNA and subsequent restriction enzyme analysis.

2.9.8 Small scale isolation of plasmid DNA

Recombinant colonies were picked with a sterile tooth pick and placed into 5 ml of medium containing the appropriate selective antibiotic and grown overnight. Cells were then pelleted from 1.5 ml of culture in a microcentrifuge and plasmid DNA prepared by a scaled down alkaline lysis method as described in section 2.9.1. The cell pellet was resuspended in 100 μl of lysis solution and lysed by mixing with 200 μl of alkaline SDS and incubating on ice for 1 minute. Cell debris and chromosomal DNA was precipitated with 150 μl of 3 M potassium acetate pH 5.2 and removed by microcentrifugation for 10 minutes. The remaining solution was extracted with phenol/chloroform (section 2.7.1) and the plasmid DNA precipitated by the addition of 2 volumes of ethanol. The nucleic acids were collected by microcentrifugation for 10

minutes, rinsed in 70 % ethanol and dried before dissolving in 30 μl of sterile distilled water. Restriction digestion analysis was carried out on 5 μl of the DNA solution.

2.9.9 Large scale isolation of genomic DNA from *N. tabacum*

2x cetyl triethylammonium
bromide (CTAB) buffer:

100 mM	Tris-HCl, pH 8.0
1.4 M	sodium chloride
20 mM	EDTA
2 %	CTAB

10 % CTAB buffer:

10 %	CTAB
0.7 M	sodium chloride

CTAB precipitation buffer:

50 mM	Tris-HCl, pH 8.0
10 mM	EDTA
1 %	CTAB

Tobacco tissue was collected, weighed, frozen in liquid nitrogen and stored at -80 °C until use. The collected tissue was ground to a fine powder in liquid nitrogen in a mortar and pestle. The powder was placed in a 50 ml screw cap tube and 1-2 ml per gram of 2x CTAB buffer pre-heated to 95 °C was added. The tissue was mixed thoroughly and the tube transferred to a 56 °C waterbath. The mixture was incubated for 1 hour with occasional mixing. The tube was allowed to cool and the mixture extracted with an equal volume of chloroform:octanol (24:1). The upper aqueous phase was removed after centrifugation in a benchtop Sorvall for 10 minutes and filtered through muslin Miracloth into a new 50 ml tube. A 1/10th volume 10 % CTAB pre-heated to 56 °C was added and mixed thoroughly with the aqueous phase until the solution cleared. Chloroform:octanol extractions were carried out until a clear interface was obtained. An equal volume of CTAB precipitation buffer was added, mixed and allowed to stand at room temperature overnight. The precipitates were spun down for 10 minutes in a bench top Sorval and the pellets resuspended in 1 M sodium

chloride (1 ml per gram tissue used). Two volumes of ethanol was added to precipitate nucleic acids, which were collected by centrifugation. The DNA pellet was washed in 70 % ethanol, vacuum dried and resuspended in 100 μl of distilled water per gram of starting tissue. 1 μl of the DNA preparation was run on an 0.7 % agarose gel to check the quality of the DNA.

2.10 Amplification of DNA by the Polymerase Chain Reaction (PCR)
2.10.1 Amplification of plasmid DNA

1x PCR buffer:

44 mM	Tris-HCl, pH 8.8
11 mM	ammonium sulphate
4.5 mM	magnesium chloride
7 mM	ß-mercaptoethanol
4.5 mM	EDTA, pH 8.0
1 mM	of each dNTP
113 μg/ml	BSA

The PCR conditions used to amplify cDNA inserts in pBluescript SK plasmid were as follows:

Reactants:

Plasmid target DNA (1 ng/μl)	1 μl
PCR Buffer (from a 10x stock)	2 μl
Primers (100 ng/μl)	2 μl and 2 μl
Taq (Perkin Elmer) DNA Polymerase I (5 U/μl)	0.2 μl

Sterile distilled water up to 20 μl.

Cetus DNA thermal cycling machine programme:
The PCR programme consisted of 25 amplification cycles each consisting of: 1 minute denaturing step at 94 °C; 1 minute annealing step at a temperature determined by the sequence of the primers used (55 °C in this case) and an extension step at 72 °C for 1 minute (timing adjusted for length of DNA fragment to be amplified, 1 minute per kb DNA).
Reactions were run and the products visualised by electrophoresis.

2.10.2 PCR analysis of bacterial colonies

A sterile plastic micropipette tip was used to touch the surface of a transformed bacterial colony and the resulting cells picked up were transferred to a 0.5 ml microcentrifuge tube containing the following PCR reagent mix;

PCR Buffer (from a 10x stock)	2 μl
Primers (100 ng/μl)	2 μl and 2 μl
Taq (Perkin Elmer) DNA Polymerase I (5 U/μl)	0.2 μl

Sterile distilled water up to 20 μl.

A PCR programme essentially the same as for the amplification of plasmid DNA (section 2.10.3) was carried out, except that a preliminary incubation of 5 minutes at 94 $^\circ$C was used in order to break open the bacterial cells. Reactions were run and the products visualised by electrophoresis.

2.10.3 PCR analysis of transgenic *N. tabacum* plants

PCR extraction buffer:

200 mM	Tris-HCl, pH 7.5
250 mM	sodium chloride
25 mM	EDTA
0.5 %	SDS

A small piece of *N. tabacum* leaf was removed from the plant and DNA extracted by a rapid method described by (Edwards *et al.*, 1991). The leaf disc was homogenized with a disposable plastic rod for 15 seconds before 400 μl PCR extraction buffer was added. The solution was vortexed for 5 seconds and stored at room temperature until all of the samples were at this stage. The debris was pelleted by microcentrifugation for 1 minute and 300 μl of supernatant removed to a fresh tube. The DNA was precipitated by the addition of an equal volume of isopropanol and then pelleted, after a 2 minute incubation at room temperature, by centrifugation for 10 minutes. The DNA pellet was dried and resuspended in 100 μl of distilled water. 2μl of the resulting DNA solution was used in a PCR reaction, using the lat52 N-terminal and *lim/myb*

internal oligonucleotides as primers, both at a concentration of 1 μM. The PCR program consisted of 30 amplification cycles, each consisting of 1 minute and 20 seconds at 94 °C, 2 minutes at 55 °C and 2 minutes at 72 °C. The amplified DNA was then visualised by ethidium bromide staining after electrophoresis of samples on a 0.8 % agarose gel.

2.11 DNA Sequencing
2.11.1 Preparation of double-stranded DNA template for sequencing

DNA prepared from the large scale isolation procedure (section 2.9.1) was used in sequencing experiments. To obtain good quality sequence the DNA was purified via Wizard™ (Promega) DNA purification columns according to the manufacturers instructions. 1 μl of this DNA solution was used for spectrophotometric determination of DNA concentration.

Approximately 10 μg of this DNA was denatured by the following method: the DNA solution was made up to 30 μl and 2 μl of 2 M sodium hydroxide was added. The solution was left to stand at room temperature for 5 minutes. The DNA solution was neutralised by the addition of 4 μl of 4 M ammonium acetate (pH 4.8). The denatured DNA was precipitated with 2.5 volumes of ethanol at placed at -80 °C for 15 minutes. The denatured DNA was recovered by microcentrifugation and the pellet washed with 70 % ethanol. The pellet was resuspended in 10 μl of sterile water and used for annealing to the primer.

2.11.2 Sequencing using T7 DNA polymerase

10x TBE (per litre):

121 g	Tris base
53.4 g	boric acid
7.4 g	EDTA, pH 8.3

6 % Acrylamide solution:

6 %	acrylamide
3 %	bisacrylamide
7 M	urea
1x TBE	

Ammonium persulphate solution:

250 mg/ml   ammonium persulphate

Templates were sequenced using the T7 Polymerase kit (Pharmacia) which employs dideoxynucleotide chain-terminating mixes (Sanger *et al.*, 1977). Templates primed with the appropriate oligonucleotide were sequenced with ^{35}S-labelled dATP, and the products separated by denaturing polyacrylamide gel electrophoresis. An Anachem V4 sequencing apparatus was used to pour the gel, with the acrylamide solution polymerised by the addition of 50 μl ammonium persulphate solution and 50 μl TEMED. Once poured, the gel was left to polymerise for 1 hour. The gel was pre-run to 50 ^{o}C in 1x buffer before loading pre-heated samples. Gels were dried onto Whatman 3MM filter paper and exposed to autoradiographic film at room temperature overnight without the use of an intensifying screen.

2.11.3 Sequencing using the automated ABI Prism™ procedure

Double-stranded DNA used in automatic sequencing was prepared as for manual sequencing (section 2.11.1) with the additional purification step of an ethanol precipitation after the use of the Wizard™ (Promega) DNA purification column. The following reagents were then mixed in a 0.5 ml microcentrifuge tube;

9.5 μl   ABI Prism Terminator Premix

5.0 μl   1 μg purified double-stranded DNA

3.2 pM   primer

Distilled water to 20 μl.

The reaction mixture was overlaid with a drop of mineral oil and the sequencing reaction placed in a Perkin-Elmer Cetus Model 480 PCR cycling machine preheated to 96 ^{o}C. The following PCR cycle was used for 25 cycles;

96 ^{o}C   30 seconds

50 ^{o}C   15 seconds

60 ^{o}C   4 minutes

The PCR extension products were then purified by two extraction's with 100 μl phenol:H2O:chloroform (ABI P/N 400765), precipitation with sodium acetate and ethanol and washing with 70 % ethanol. The DNA pellet was air dried and the samples run on an Applied Biosystems Model 373A DNA Sequencing System.

2.11.4 Data handling

Sequence data was compiled and analysed using the GeneJockey II and DNA STRIDER Macintosh programmes. Data from the ABI automated sequencing service was compiled and analysed using the Seq.Ed v 1.0.3 Macintosh programme. Multiple alignments of deduced amino acid sequences were performed using the CLUSTAL programme (Higgins and Sharp, 1989). RNA 2° structure predictions, protein 2° structure predictions and dendrograms were compiled using the 'SQUIGGLES', 'PLOTSRUCTURE' and 'PILEUP' programs respectively, contained in the Version 8, GCG suite of programs (Devereux et $al.$, 1984).

2.12 Cytological analysis of pollen
2.12.1 Staining pollen grains with Alexander Stain (Alexander, 1969)

Alexander stain mixture:

95 %	alcohol
10 mg	malachite green (1 ml of 1 % solution in 95 % alcohol)
50 ml	distilled water
25 ml	glycerol
5 g	phenol
5 g	chloral hydrate
50 mg	acid fucsin (5 ml of 1 % solution in water)
5 mg	orange G (0.5 ml of 1 % solution in water)
2 ml	glacial acetic acid

Pollen grains from dehisced *N. tabacum* anthers were mounted on a drop of stain, covered with a coverslip and then viewed with a Nikon microscope.

2.12.2 Staining pollen nuclei with Fluorescene Di-Acetate (FDA)

FDA stain:

$$100 \ \mu g/ml \quad \text{FDA (in acetone)}$$

Pollen grains were resuspended in 100 μl of 0.25 M mannitol in water and mixed. 0.25 μl of FDA stain was mixed and the grains and visualised after 1 minute by pipetting onto a microscope slide, placing over a cover slip and illuminating under uv light.

2.12.3 Staining pollen grains with 4', 6-diamidino-2 phenylindole (DAPI) (Coleman and Goff, 1985)

Stock DAPI solution:

$$1 \ mg/ml \quad \text{DAPI (in distilled water)}$$

Working DAPI solution:

$$10 \ \mu g/ml \quad \text{DAPI (in distilled water)}$$

To visualise pollen nuclei, 20 μl of pollen was mixed with 5 μl working DAPI solution in a microcentrifuge tube and visualised after 5-10 minutes by pipetting onto a microscope slide, placing over a cover slip and illuminating under uv light.

2.12.4 *In Vitro* pollen germination

Tupys pollen
germination medium:

0.3 M	sucrose
1.6 mM	boric acid
3 mM	calcium nitrate
0.8 mM	magnesium sulphate
1 mM	potassium nitrate

| | 25 mM | (2-[N-Morpholino]ethane-sulphonic acid (MES) |

pH 5.9 with potassium hydroxide solution.

Dessicated pollen grains were resuspended in Tupys pollen germination medium in a microcentrifuge tube and visualised after 30 minutes to overnight by pipetting onto a microscope slide, placing over a cover slip and illuminating under uv light.

2.13 Transient expression analysis by microprojectile bombardment

2.13.1 Preparation of plant material for bombardment

Leaves of young *N.tabacum* SR1 plants grown in the greenhouse were taken and cut into 1 cm squares with a scapal and forceps. Each leaf square was placed adaxial side down onto the surface of an MS0 plate.

2.13.2 Equilisation of plasmid DNA

Test, reference and standard plasmids were equalised to each other prior to microprojectile bombardment at a concentration of 1 μg/μl. An aliquot of a large scale plasmid preparation of each plasmid (section 2.9.1) was used as a working stock solution. A 100x dilution of each working stock was made and 5 μl and 10 μl of these dilution's of each plasmid were run with 1 μl of loading buffer (section 2.8.2) on a 0.8 % agarose gel. The 5 μl sample was mixed with 5 μl of distilled water prior to loading. The concentration of the working stock solutions were adjusted until 10 μl of a 100x dilution of each plasmid were indistinguishable from each other and from 100 ng of marker reference DNA. Plasmids were thus equalised at 1 μg/μl.

2.13.3 Precipitation of plasmid DNA onto microprojectiles

60 mg of dry tungsten microprojectiles were resuspended in 1 ml of ethanol by vortexing for 6 minutes in a microcentrifuge tube. The tube was spun in a microcentrifuge for 1 minute and the pellet resuspended in 1 ml of sterile distilled water by vortexing. The microprojectiles were washed in 1 ml of water again and the tungsten pellet finally resuspended in 1 ml of water. 50 μl aliquots were taken while vortexing the suspension continually. aliquots were stored at -20 °C.

A 50 μl aliquot was vortexed and 5 μg each of equalised test, reference and standard plasmids were added immediately and the mixture votexed. 50 μl of 2.5 M calcium chloride was added and the mixture vortexed. 20 μl of 0.1 M spermidine (free base, tissue culture grade) was then added and the microcentrifuge tube vortexed for 4 minutes. The microprojectiles were then collected by centrifugation for 10 seconds and the supernatent removed. The pellet was resuspended in 250 μl of ethanol by vortexing. The microcentrifuge tube was then spun for 10 seconds and the microprojectiles resuspended in 60 μl ethanol by vortexing.

2.13.4 Microprojectile bombardment

Microprojectile bombardment was carried out with a Bio-Rad Biolistic PDS-1000/He helium gun (Figure 2.1). Helium pressure was set at 1750 Psi with 1550 Psi rated rupture discs used. A petri dish containing the plant material was placed second shelf down from the macrocarrier shelf in the chamber of the particle gun with the lid removed. A rupture disc was placed in the rupture disc holder and a stopping plate in the macroprojectile shelf. The DNA-tungsten microprojectile mixture was vortexed and 8 μl pipetted onto the surface of a clean macroprojectile. The ethanol was allowed to evaporate and the macroprojectile containing the dried microprojectile/DNA mix was placed into the macroprojectile shelf of the gun (DNA side towards the stopping grid). The gun was fired and this procedure repeated twice more with the same DNA/tungsten mix. Bombarded plates were sealed with Nescofilm and incubated for 24 hours at 25 °C in a culture room.

2.13.5 Quantitative determination of luciferase and ß-glucuronidase activities
2.13.5.1 Preparation of cell extracts

Cell extraction buffer (CEB):

0.1 M	potassium phosphate buffer, pH 7.5
1 mM	DTT (added fresh)

A bombarded leaf square was placed in a clean mortar and ground with 500 μl cold CEB for 1 minute. The extract was collected in a pre-chilled microcentrifuge tube and

Figure 2.1
Schematic diagram illustrating a biolistic gun.

blank charge

macroprojectiles

microprojectiles

stopping plate

biolistic
chamber

target tissue
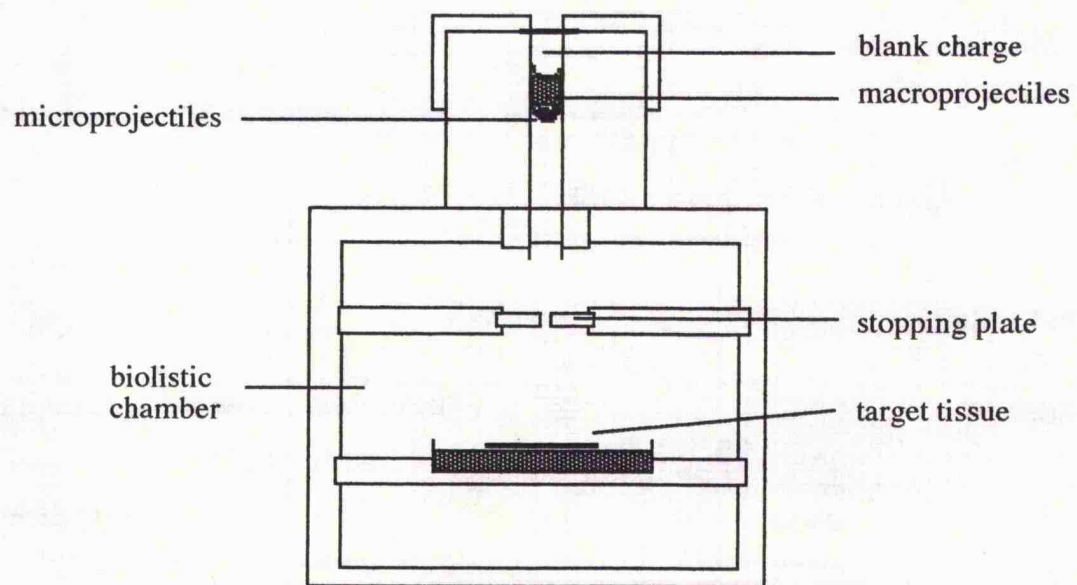
the samples centrifuged for 2 minutes at 3,000 xg. The supernatents were transferred to clean tubes and chilled on ice.

2.13.5.2 Assay of luciferase (Ow *et al.*, 1986).

dATP buffer:

10 mM	dATP
50 mM	potassium phosphate, pH 7.5
	(section 2.12.5.1)
50 mM	HEPES, pH 7.5
20 mM	magnesium chloride

Luciferin:

0.05 M	luciferin
50 mM	potassium phosphate, pH 7.5

Luciferase activity was measured in a luminometer (Berthold LB9502 Clinilumat). 25 μl of extract was added to the bottom of a sample cuvette and 250 μl luciferin and 100 μl dATP buffer injected automatically into the cuvette by the luminometer. A reading of total light emission over a period of 10 seconds and background over 1 second was taken. The assay was repeated using a fresh cuvette for each extract.

2.13.5.3 Assay of ß-glucuronidase (GUS) (Jefferson *et al.*, 1987).

GUS extraction buffer:

50 mM	sodium phosphate, pH 7.0
10 mM	β–mercapto ethanol
1 mM	EDTA
0.1 %	Triton X-100

0.4 ml of 1 mM 4-methyl umbelliferyl β-D-glucuronide (MUG) in GUS extraction buffer was added to a microcentrifuge tube and incubated for 5 minutes in a 37 °C water bath. 100 μl of extract was added to the reaction tube, mixed and incubated at 37 °C. 100 μl aliquots were sampled at 10, 20 , 30 and 40 minutes, adding each directly to 0.1 ml of 0.2 M sodium carbonate in a microtitre plate. Fluorescence at an excitation wavelength of 365 nm and an emission wavelength of 455 nm was measured.

The relative activity of each bombardment was calculated as a ratio of test plasmid activity to reference plasmid activity.

Chapter Three.

Expression library screening of a mature pollen cDNA library to identify proteins which bind to pollen regulatory sequences.

3.1 Introduction and aims.

Promoter analysis of pollen-specific genes has led to the identification of a number of cis-acting elements (section 1.4.1). Deletion and mutational analysis of three pollen-specific lat (late anther tomato) genes (Twell et al., 1991) in transient expression assays and transgenic plants revealed two functional cis-elements. The first, a Lat52/56 box, 'TGTGGTTATATA', was present in the promoters of lat52 and lat56. From this element a core motif (TGTGGTT; PB motif) was deduced and found to be reiterated three times in the lat52 promoter (PBI, PBII and PBIII). A second cis-element the Lat56/59 box, 'GAAATTGTGA', was present in the promoters of lat56 and lat59. Both of these box elements are positioned 100 bp upstream of the transcription initiation site of the respective genes. Deletion of the Lat52/56 box from the lat52 promoter in transient assays, resulted in a 10-fold drop in promoter activity and deletion of the Lat56/59 box in the promoter of lat59, gave a 6-fold drop in activity. Deletion of the PBII region, containing the PB core motif, in the lat52 promoter gave a 5-fold drop of activity of the promoter in transient assays.

Sequences similar to both the Lat52/56 and the Lat56/59 boxes have been found in the promoters of a number of pollen expressed genes including the pectate lyase (tpl) gene from N. tabacum (Rogers et al., 1992a) and the ntp303 gene from N. tabacum (Weterings et al., 1992). The lat52/56 box and the PB core motif show high nucleotide sequence similarity (8/12 and 7/7 base pairs respectively to the binding site for GT-1 (TCTTGTGTGGTTAATAT)). GT-1 is a trans-acting factor activating transcription of the gene coding for the small subunit of ribulose bisphosphate carboxylase (rbcS-3A; (Gilmartin et al., 1992)). Detailed pair-wise mutation of the 'GG' doublet in the PB core motif to 'CA' in -100 deleted lat52 promoter in transient assays in N. tabacum pollen resulted in a 12-fold drop in promoter activity, which suggests a functional importance for this doublet in protein binding. The importance of the 'GG' doublet was also seen in mutational studies of the DNA-binding site of GT-1 (Green et al., 1988), which suggests that GT-1 may also bind to the PB motif.

Another putative pollen-specific cis-acting element has been defined in the ntp303 promoter (Weterings et al., 1992). Deletion analysis of 2 regions, -106 to -89 and -89 to -64 resulted in a 7- and 6-fold loss of promoter activity in transient assays respectively. Comparison of both regions revealed they shared a 'AAATGA' sequence.

Previous analysis of the expression patterns of the lat52, lat56, lat59, ntp303 and tpl genes by temporal Northern analysis showed a late pattern of pollen gene expression,

with steady-state transcript levels appearing at pollen mitosis I (PMI) and increasing to maturity. In the light of these observations, mature pollen was considered a potential source of material for cDNA library construction, likely to yield cDNAs encoding *trans*-acting factors involved with pollen gene expression.

It was decided to use fragments containing these *cis*-element sequences from the above pollen-specific promoters as probes to directly screen for sequence-specific transcription factors using expression library screening (Singh *et al.*, 1988). Previously this method had been used to isolate a number of *trans*-acting factors in mammalian systems including, the CCAAT-box binding factor, C/EBP (Vinson *et al.*, 1988), NFkB (Singh *et al.*, 1988) and the octamer binding protein OCT2 (Stauldt *et al.*, 1988). In plant systems, a large group of basic 'leucine zipper ' type proteins had been isolated, including the TGA1a and TGA1b proteins from *N. tabacum* (Katagiri *et al.*, 1989) and CPRF family of factors from parsley (Weisshaar, 1991). Other plant factors which had been isolated by expression library screening included the GT-1 protein from *N. tabacum* (Gilmartin *et al.*, 1992) and GT-2 from rice (Dehesh *et al.*, 1992). Factors cloned in this way can then be used in a wide variety of experiments to characterise the cloned gene and the corresponding protein and its activity.

Here, cDNA is cloned into the β-galactosidase (LacZ) gene of bacteriophage expression vectors which enables cloned cDNAs to be expressed as fusion proteins upon addition of an inducer of transcription such as IPTG. These fusion proteins bind DNA with the same specificity as the original factor. Hence a cDNA clone encoding a particular factor can be identified in a library by screening with a radioactive oligonucleotide containing the binding site (Figure 3.1).

An advantage of using such defined DNA binding site probes is that no prior purification or characterisation of the factor or factors is required before screening. However, this technique cannot detect factors that bind DNA as part of a heterodimeric or heteromeric protein complex. The factor must bind DNA in a stable manner in the absence of other components of a multisubunit complex. Detection can be limited also in cases where a factor requires post-translational modifications such as phosphorylation or specific proteolytic cleavage for efficient DNA-binding activity. These requirements will not be met in the *E. coli* host cell and the factor may bind DNA with low efficiency or not at all. In this case optimisation of binding and washing conditions may circumvent this problem. An addition to the protocol involves treating the bacteriophage filter lifts with a guanidinium denaturation/renaturation regime (Vinson *et al.*, 1988). This treatment was shown to enhance detection of some

Figure 3.1

Outline of the strategy for expression library screening using recognition site DNA probes.

'X' is a recognition site DNA probe. 'Y' is a control DNA probe that lacks the given recognition site or contains a mutant version of it.

 represents a ^{32}P labelled cognate DNA-binding site probe.

λ ZAP II cDNA library

Transfer proteins from plaques
onto nitrocellulose filter

Block non-specific binding sites
on filter

Probe filter with 32P-cognate

in the presence of excess non-specific
competitor DNA

Wash filter

Process for autoradiography

Plaque purify

X   Y

Generate lysogens

Assay extracts of lysogens for
X-specific DNA binding activity
via gel retardation. X-DNA probe.

λY   λX

—   Bound X-DNA

—   —   Free X-DNA

protein factors, possibly by facilitating the correct folding of a larger fraction of the *E. coli* expressed fusion protein. Alternatively it may help to dissociate insoluble aggregates of the fusion protein that form as a result of overexpression. This modification of the procedure was used to isolate clones encoding *Oct1* (Sturm *et al.*, 1988) and *Pit-1* (Ingraham *et al.*, 1988) amongst others.

Another modification of the basic protocol was the demonstration that multi-site probes can increase signals intensities of positive clones (Singh *et al.*, 1989). A three to four fold increase in signal intensity was obtained by using a trimer form of a NFkB binding site probe over a monomer probe. Such probes can be prepared by concatenation of a binding site fragment prior to radioactive labelling by nick translation. Multi-site probes may allow a single DNA molecule to be tethered at more than one protein binding site, making the probe more stable.

Preliminary characterisation of the conditions necessary for protein binding to DNA binding-site probes, either by gel retardation or Southwestern blot analysis is preferred. Factors such as binding buffer composition, length of washing and concentration of non-specific competitor DNA can be optimised. Once specific protein bands have been identified then by definition the factor or factors must be able to bind as a single protein monomer or homodimer. More than one hybridising band would indicate a family of DNA binding factors.

The first aim of the project was the production of a pollen cDNA library. Mature desiccated pollen from *Nicotiana tabacum* cv. Samsun was chosen as a source of poly A+ RNA, from which a cDNA population would be made, because of it's availability and the ease with which it is collected (approximately 1.5 to 2 g of pure pollen can be collected from twenty mature *N.tabacum* cv. Samsun plants in a single session). Moreover, mature pollen from *N. tabacum* consists of a highly purified population of cells at a specific stage of development. In this way, transcription factors binding to promoters of 'late' pollen genes will be present in a greater abundance than would be the case in a mixed cell population where transcript encoding 'late' expressed transcription factors might not be present in cells at earlier stages of development.

In summary, this chapter investigates the use of double-stranded DNA binding site probes from the promoters of a group of pollen-specific genes in a direct approach to the identification of proteins involved with pollen development and/or function.

3.2 Results.

3.2.1 An attempt to isolate DNA binding proteins to pollen cis-element sequences in N.tabacum.

3.2.1.1 Production and characterisation of a pollen cDNA library.

A cDNA library was made from 5 μg of poly A^+ RNA from mature desiccated pollen of N. tabacum cv. Samsun. The lambda ZAP II bacteriophage expression vector (Short et al., 1988) was used in the library construction, as it allows the in vivo excision and recircularisation from the lambda XR vector of pBluescript SK vector and cDNA insert. This step obviates the need for a separate subcloning step prior to cDNA sequencing.

A poly dT linker-primer was annealed to the poly A tails of the poly A^+ RNA population and first-strand cDNA made by malony murine leukemia virus reverse transcriptase (MLVRT; Figure 3.2). MLVRT has less DNA dependant DNA polymerase activity and therefore less likely to produce unclonable hairpin loops. It also has less inherent RNase H activity than other reverse transcriptases, such as AMV RT and is therefore less likely to hydrolyse the RNA template. As first-strand template produces the template for second-strand synthesis the length and integrity of the cDNA are important. For this reason the efficiency of cDNA synthesis was assessed by including a small amount of radiolabelled ^{32}P dCTP in the first-strand cDNA reaction and running a small aliquot of resulting cDNA on an alkaline agarose gel. The average size of the cDNA was 1-2 kb, but cDNA product was visible upto 6 kb. First-strand cDNA nucleotide mix contained methyl-dCTP in place of dCTP, which, when incorporated into cDNA provided protection from restriction enzymes used in subsequent cloning steps.

Second-strand cDNA was made using DNA polymerase I and RNase H. RNase H nicks the RNA strand of the RNA-DNA duplex and DNA polymerase I nick translates the second-strand cDNA. T4 DNA polymerase is then used to end-fill the double-stranded cDNA 3' ends and EcoR1 adapters ligated onto the blunt ends via T4 DNA ligase. After ligation the EcoR1 ends are kinased with T4 polynucleotide kinase and Xho1 restriction digestion used to create directional cDNA. At this stage the cDNA was run twice through a sephacryl size exclusion column, and the two fractions pooled. This procedure excluded cDNA fragments shorter than 200 bp from the cDNA population. If these smaller fragments were not removed they would clone with

Figure 3.2
Construction of a pollen cDNA library and PCR analysis from randomly picked library clones.

a. Flow diagram showing construction of a λZAPII cDNA library.
b. PCR products separated on a 0.8 % agarose gel.

Molecular size standards are in kilobases.

a.

mRNA   Primer
                                            TTTTTTTGAGCTC
█████████████████████████████   AAAAAAA

1st strand cDNA | Malony reverse trnscriptase
synthesis | CH3-dCTP
                    ↓   dATP, dCTP, dGTP

CH3   CH3   CH3   CH3
─────────────────────────────   TTTTTTTGAGCTC
█████████████████████████████   AAAAAAA

2nd strand cDNA | RNaseH
synthesis | DNA PolI
                    ↓ dNTPs

CH3   CH3   CH3   CH3
─────────────────────────────   TTTTTTTGAGCTC
─────────────────────────────   AAAAAAA

                    1. T4 DNA Pol
                    2. T4 DNA ligase
                    ↓ EcoR1 adapters

13mer primer   13mer primer

EcoR1   Xho1   EcoR1
AATTC... CH3   CH3   CH3   CH3
         ─────────────────────────── TTTTTTTGAGCTC...CTTAA
G...   ─────────────────────────── AAAAAAACTCGAG...G

9mer primer   9mer primer

                    1. T4 polynucleotide
                    ↓   kinase
                    2. Xho1

EcoR1   CH3   CH3   CH3   CH3   Xho1
AATTC... ──────────────────────────
         ────────────────────────── TTTTTTTGAGCT
G...   ────────────────────────── AAAAAAAC

                 Directional cDNA (0-10 kb)

                    DNA ligase
                    ↓ Uni-ZAP XR vector

AATTC...──────────────────────────TTTTTTTGAGCT
TTAAG...──────────────────────────AAAAAAACTCGA

                 Uni-ZAP XR library

b.

greater efficiency than larger fragments into the lambda vector and represent a large proportion of the final cDNA inserts. Ligation of the cDNA into the Uni-ZAP XR vector completed the cloning procedure.

At this stage the cDNA was still in a hemi-methylated state. After packaging into lambda phage particles, the library was amplified once in a methylase⁻ host strain (SURE) which removed the methyl groups and allowed the library to be stored in a stable state. The preamplified library had a titre of 1.5×10^6 pfu/ml, and after amplification in SURE strain was increased to 4×10^9 pfu/ml. Ten randomly chosen recombinant plaques (distinguished as white plaques by IPTG/X-gal selection) and two non-recombinant blue plaques were analysed both by restriction digestion of the phage DNA and by *in vivo* excision and amplification of inserts from the pBluescript plasmids by PCR using forward and reverse primers (Figure 3.2b). Insert sizes of the clones ranged from 1.7-0.1 kb. One of the clones had a very small insert size of approximately 0.1 kb (ie. 0.1 kb larger than the two non-recombinant plaques which gave an insert size by PCR of approximately 0.2 kb) and another multiple inserts of 3.5, 0.5 and 0.2 kb as revealed by *Eco*R1/*Xho*1 restriction digestion. The library was tested for the presence of known pollen-specific sequences by screening with the *lat52* cDNA (Twell *et al.*, 1989a) and *ntp303* cDNA (Weterings *et al.*, 1992). Both of these probes gave positive hybridising plaques at their expected transcript frequencies in mature pollen of 0.3 % and 0.2 % of the library clones respectively.

3.2.1.2 Preparation of double-stranded DNA binding-site probes.

PCR fragments containing putative protein binding sites from four pollen-specific promoters were made (Figure 3.3a. to e.; section 1.4.1) for use as probes in expression library screening. A 58 bp fragment from the *lat52* promoter containing a 52/56 box; a 49 bp from the *lat52* promoter containing the PBII motif sequence; a 67 bp fragment from the *lat59* promoter containing a 56/59 box and a 56/59 box core sequence; a 39 bp fragment from the *ntp303* promoter containing two 'AAATGA' sequences and a 50 bp fragment from the *tpl* promoter containing a 56/59 box core sequence and a 52/56 box core sequence. These fragments were synthesised with terminal restriction enzyme sites which enabled them to be concatamerised with DNA ligase, after restriction digestion to generate sticky ends, to produce a probe mixture consisting of predominately monomer, dimer and trimer forms (Figure 3.4).

Figure 3.3

Schematic diagram representing PCR fragments used in the preparation of DNA-binding site probes in expression library screening.

Numbers and arrows indicate the position and orientation of the binding site motifs relative to the relevant transcription start sites.

a. PCR fragment amplified from the *lat52* pollen-specific promoter containing a 52/56 box sequence. Bold sequence indicates the PB core motif.

b. PCR fragment amplified from the *lat52* pollen-specific promoter containing the PBII sequence. Bold sequence indicates the PB core motif.

c. PCR fragment amplified from the *lat59* pollen-specific promoter containing a whole 56/59 box sequence and a core 56/59 box sequence.

d. PCR fragment amplified from the *ntpg303* pollen-specific promoter containing two putative cis-element sequences.

e. PCR fragment amplified from the *tp1* pollen-specific promoter containing a 56/59 box core sequence and a 52/56 box core sequence.

f. Complimentary oligonucleotides containing the PBI sequence. Boxed sequence indicates the PB core motif. Abbreviations: T, top oligonucleotide; B, bottom oligonucleotide.

g. Complimentary oligonucleotides containing the PBII sequence. Boxed sequence indicates the PB core motif. Abbreviations: T, top oligonucleotide; B, bottom oligonucleotide.

a.

Asp718   52/56 box   Sal1

TGTGGTTATATA

-100   -42

b.

Sal1   PBII   Sal1

TTTTTAC**AACCACA**

-225   -176

c.

HinD111   56/59   Sal1

GAAATTGTGA   GTGA

-115   -48

d.

Sal1   Sal1

AAATGA   TCATTT

-106   -67

e.

Asp718   Sal1

GTGA   CACC

-103   -53

f.

5'   PBI   3'

T   ATATTA TGTGGTTAT

           TACACCAATATATA A **B**

g.

5'   PBII   3'

T   TT ACAACCACA AAAA

          TGGTGT TTTTAATGT **B**
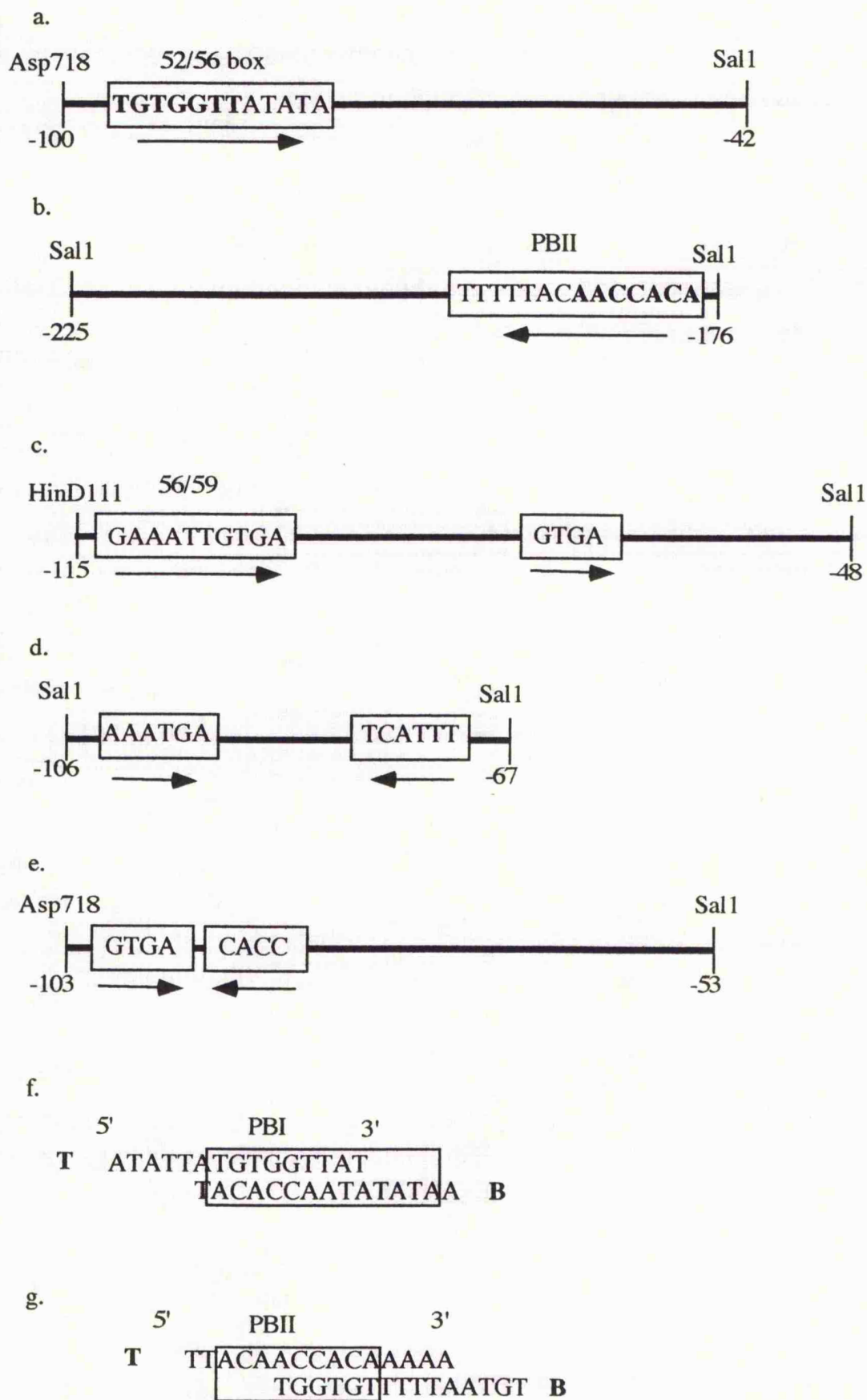
Figure 3.4

Photo-autoradiograph showing concatenated DNA binding-site probes.

Lanes 1. and 4., ^{32}P dATP free probe.

Lane 2., 52/56 box (long fragment) binding-site probe.

Lane 3., PBII (long fragment) binding-site probe.

a. Trimer form of probes.

b. Dimer.

c. Monomer.

d. Un-incorporated nucleotides.
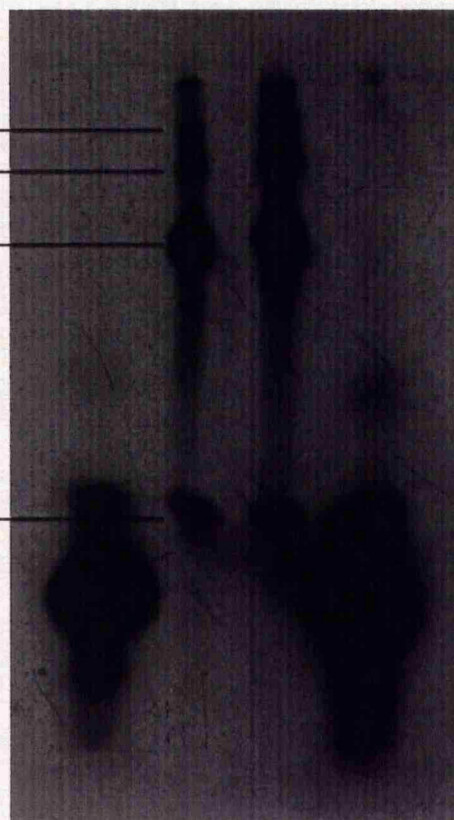
In addition, two pairs of complimentary 15 mer oligonucleotides were made containing a 52/56 box and a PBII sequence in isolation (Figure 3.3f. and g.). By using these shorter binding-site probes it was hoped any non-specific background hybridisation obtained by screening with relatively long fragment probes would be eliminated. The oligonucleotide pairs were annealed and ligated to produce concatamers. The overhanging ends were then end-filled with T4 DNA polymerase. 10 μg of each of the concatamerised binding-site probes were labelled by nick translation to a specific activity of approximately 1×10^7 cpm/μg. These probe mixtures were purified from unincorporated nucleotides by passing down a Sephadex column.

3.2.1.3 Characterisation of DNA-binding conditions by Southwestern blot analysis.

Preliminary Southwestern blot analysis was performed using crude *N. tabacum* pollen and leaf nuclear (Dr. L. Muir, Leicester) protein extracts. A mixed concatamerised probe containing the 52/56 and PBI (long fragment) fragments were used to investigate the conditions for protein/DNA interaction. Using 50 μg/ml sonicated herring sperm and probe concentrations of 2×10^7 cpm/ml, a nitrocellulose filter containing one lane each of pollen total and leaf nuclear protein extract was overlayed with the probe mixture in 5 ml of binding buffer and incubated for one hour at room temperature. Washing was carried out at room temperature for 3×10 minute intervals with 50 ml binding buffer each. Four strong and one weakly hybridising protein bands were visualised in the pollen extract (15 kDa, 17 kDa, 39 kDa, 37 kDa and 23 kDa respectively) and three strongly hybridising protein bands in the leaf nuclear extract (10 kDa, 34 kDa and 36 kDa) at these conditions (Figure 3.5; bands are indicated by arrows). The strongly hybridising 10 kDa protein band in the leaf extract is more intense than the other two bands in the extract suggesting it is composed of two proteins of similar size. Binding demonstrated that proteins present in the extracts were capable of binding to the probes as monomers or homodimers. Repetition of this experiment with 150 μg/ml herring sperm DNA and 2×20 minutes washing produced the same pattern of hybridising bands.

3.2.1.4 Expression library screening with double-stranded DNA probes.

81

Figure 3.5

Photo-autoradiograph showing an interaction between PBI and PBII concatenated DNA binding-site probes and proteins in pollen and leaf extracts.

Lane 1. *N. tabacum* pollen total protein extract overlayed with 52/56 box (long fragment) binding-site and PBII (long fragment) binding-site DNA probes.

Lane 2. *N. tabacum* leaf nuclear protein extract overlayed with 52/56 box (long fragment) binding-site and PBII (long fragment) binding-site DNA probes.

Molecular weight standards are given in kDa. Arrows indicate bands referred to in text.
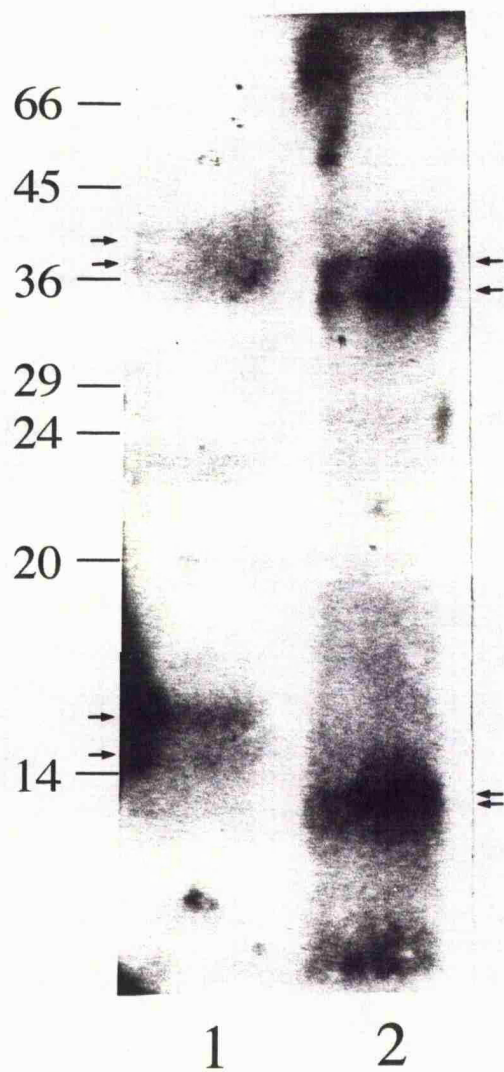
Since hybridising bands had been detected in protein extracts from *N. tabacum* pollen and leaf nuclei with concatenated 52/56 box and PBII motif probes, it was anticipated that these probes using the conditions of Southwestern blot analysis would identify proteins expressed from cDNAs in the pollen cDNA expression library.

For expression library screening the cDNA library was plated out on medium without IPTG and incubated at 42 °C to ensure that no fusion proteins toxic to the host cells were synthesised until plaque formation was underway. After 3 hours filters impregnated with IPTG were laid onto of the developing plaques. This induced the production of fusion proteins and also produced an imprint of the plate position on the plate. After further incubation, the filters were probed with a mixture containing all six concatenated binding site probes using identical conditions to that used in Southwestern blotting above. Approximately 10^6 pfu from the pollen cDNA library were screened using a regime that consisted of using the standard protocol of Singh *et al.* (Singh *et al.*, 1988), and then rescreening the same filters in the guanidinium denaturation/renaturation procedure and probing a second time (Vinson *et al.*, 1988), as the denaturation step results in dissociation of the DNA probe that may be bound to the filter from the first screen (duplicate filters were not used in these primary screens because they appeared to be unreliable in previous studies; (Singh *et al.*, 1988)). Fifty putative positive plaques were identified in these non-duplicated primary screens. These plaques were cored and lysogens made. Secondary screening of the lysogens however, did not identify any duplicating hybridising positive clones, despite screening 10^3 pfu per lysogen (Figure 3.6).

3.3 Discussion.
3.3.1 An unsuccessful attempt at isolating pollen DNA-binding factors.

The work described in section 3.2.1.3 showed that proteins of different molecular weight were present in leaf nuclear and pollen protein extracts that bound to fragments containing pollen *cis*-elements in Southwestern analysis. The four hybridising bands in the pollen extract would indicate two interacting proteins or protein complexes per diploid genome of *N. tabacum*, considering that the *N.tabacum* genome is amphidiploid (Goodspeed, 1954). If the assumption is made that a single protein binds at each of the 52/56 box and PBII motif *cis*-elements, then the genes encoding the trans-acting factors are single copy.

Figure 3.6

Photo-autoradiographs of primary and secondary phage lifts used in expression library screening.

a. Primary phage lift signal after probing with concatenated binding-site probes; putative positive plaques are circled; b. Secondary phage lifts after rescreening of putative positive plaques from a.

a.
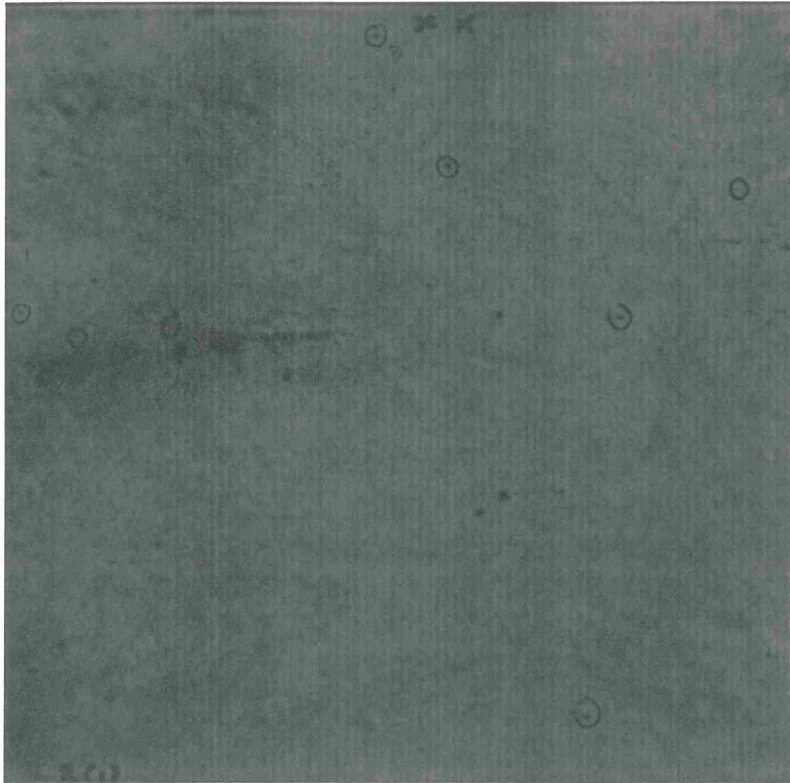
b.
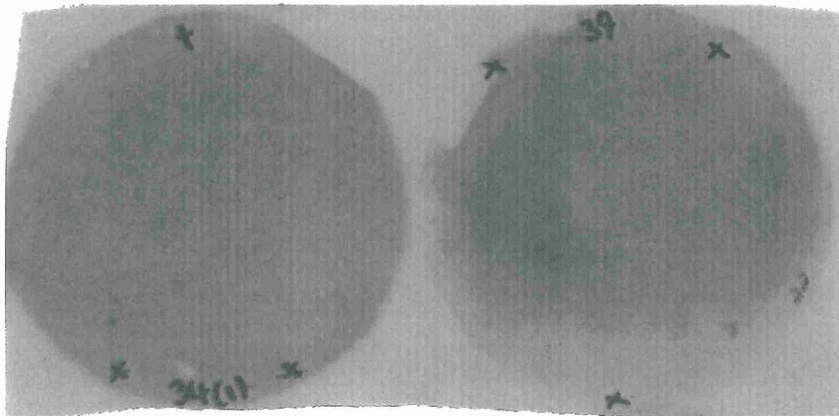
The presence of different proteins which can apparently bind to the 52/56 box or PBII motif (or both) in both the leaf and pollen extracts could be explained by two mechanisms by which pollen-specific gene expression is achieved. The simplest explanation would be one where a sequence-specific binding factor in pollen transactivates pollen gene transcription, but those different factors that bind in leaf do not transactivate, due to the absence of a transactivation domain or a different conformation on the part DNA-binding protein concealing the transactivation domain. If these factors were present in pollen they could act as negative inhibitors of pollen gene transcription, competing for cis-elements on the promoter. Alternatively, factors binding to the pollen cis-elements in leaves would be unable to activate transcription due to the absence of other trans-acting factors, present in pollen, which would be needed for activation. This situation is seen with the Drosophila hsp70 heat-shock gene transcription (Pelham, 1982). In non-heat shocked cells the promoter is bound at two sites by a GAGA factor bound at a 'GAGA' element and the TFIID basal transcription factor bound at the TATA box. A heat shock factor (HSF) produced in heat-shocked cells which binds in proximity to the GAGA and TFIID factors at a heat-shock element (HSE) to activate transcription of the hsp70 gene.

Screening the N. tabacum expression library under the same conditions used for the Southwestern blot with identical probes should in theory have resulted in the isolation of cDNAs encoding trans-acting DNA binding proteins. A potential problem may have been connected with the choice of λZAP as the vector in expression library construction. Here cDNA is expressed as a fusion protein to a small N-terminal portion of the β-galactosidase protein, as opposed to it being inserted into the coding region of the complete lacZ gene, as with the λgt11 vector, which results in its translation as part of a bigger protein. As a result the λZAP fusion proteins might be unstable under the conditions which were chosen for library screening. However, the λZAP vector has been used successfully to isolate a number of DNA-binding proteins such as CPRF-1 from parsley (Weisshaar, 1991), GT-1a from N. tabacum (Gilmartin et al., 1992) and OBF3 from maize (Foley et al., 1993).

A further possibility is that more plaques needed to be screened if the transcripts encoding factors were present at a frequency less than 1 copy per 10^6 transcripts. The preamplified library complexity was 1.5×10^6 clones, rare cDNA clones encoding these trans-acting factors might be absent from this population resulting from incomplete methylation protection from EcoR1 or Xho1 digestion, used in subsequent cloning steps and resulting in partial cleavage of the cDNA. This can result either in the disruption of a cDNA segment encoding a DNA binding domain or in a decrease in

85

the frequency of recombinants containing in-frame fusions of the DNA binding domain to the bacterial protein segment.

Amplification of the cDNA population may have discriminated against slower growing clones. If these clones encoded the *trans*-acting factors being screened for, then they would have been significantly under represented in the library. A future screening regime could be envisaged where an unamplified library would be screened, any positive clones isolated and the library subsequently amplified for stable storage for future use. Under representation of cDNA clones encoding *trans*-acting factors may have also resulted from their expression being toxic to the *E. coli* host cell, despite the occurrence of this phenomenon being minimised by initial growth of the bacteriophage plates at 42 °C to establish the growth of plaques before protein expression is induced.

Concurrent to the expression library screening, it was decided to use homology screening of the pollen cDNA library using probes designed to detect two classes of putative transcription factors, the MYB-like proteins (Dr. C. Martin, Norwich) and LIM-domain containing factors (Dr. A. Steinmetz, Strasbourg). Use of these probes resulted in the isolation of a number of duplicating positive plaques. It was decided therefore, that pursuing work with these cDNA clones would be more rewarding, and would have been more likely to contribute significantly to our understanding of regulatory factors expressed in developing pollen.

3.4 Summary.

The first route taken to clone a pollen transcription factor was to utilise DNA binding-site probes and screen a pollen expression library. Although this approach did not lead to the isolation of cDNAs encoding *trans*-acting factors it did result in the production of a mature *N. tabacum* pollen cDNA library and demonstrated the presence of potential factors in a crude pollen extract. It should be possible therefore to isolate sequence specific DNA binding factors via expression library screening. However, rather than spend more time optimising binding conditions and rescreening the expression library this approach was put to one side, when cDNAs encoding two classes of putative transcription factors were isolated directly by sequence hybridisation screening of the pollen cDNA library.

Chapter Four.

Identification of pollen-specific cDNAs encoding MYB and LIM-domain proteins from *Nicotiana tabacum*.

4.1 Introduction and aims

A cDNA library was prepared from poly A$^+$ RNA isolated from mature pollen of *N. tabacum* cv. Samsum. The expression library was screened with concatenated DNA recognition site probes containing putative binding sites from four different pollen specific promoters (Chapter 3). No positive duplicating plaques were isolated using this approach. An alternative screening strategy was adopted using heterologous DNA hybridisation probes to look for specific groups of transcription factors involved with pollen development and/or function. Primary structure analysis of different classes of transcription factors have revealed the modular nature of these proteins, with functional domains such as DNA binding motifs being highly conserved between different members of the same class of transcription factors. The sequence conservation of the MYB and LIM domain sequences were used to isolate cDNA clones related to these two classes of putative transcriptional regulators.

Two oligonucleotides were available containing sequences homologous to conserved regions of the MYB-domain found in a large family of proteins isolated from both animal and plant systems, and shown to function as a sequence-specific DNA-binding interface (section 1.4.4.1) These oligonucleotides had been used previously to isolate a group of six different types of *myb*-like cDNA clones from *Antirrhinum* (Jackson *et al.*, 1991) and two different types of cDNA clones from the moss, *Physcomitrella patens* (Leech *et al.*, 1993).

A pollen-specific cDNA, *sf3*, from sunflower had been isolated (section 1.4.4.2; (Baltz *et al.*, 1992a.)) which encoded sequences with homology to the LIM-domain, a zinc finger motif previously seen only in proteins isolated from animal systems. LIM-domain containing proteins had been shown to act as developmental regulators, although DNA-binding activity had not been demonstrated (Dawid *et al.*, 1995). The *sf3* cDNA and sequence specific PCR primers, which could be used to amplify a fragment containing the *lim*-domain, were made available (Dr. R. Baltz and Dr. A. Steinmetz, IBMP, Strasbourg) for screening the *N. tabacum* pollen cDNA library.

The two classes of *N. tabacum* cDNA clones encoding *myb*- and *lim*-domain sequences were identified at the same time as the expression library was being screened with the binding site probes. The expression library screening was abandoned and work concentrated on the characterisation of these clones. This chapter details the isolation and sequence analysis of the *N. tabacum* pollen-specific *myb.Nt1*, *myb.Nt2*, *Ntl1* and *Ntl2* cDNA clones encoding MYB and LIM domain amino-acid motifs respectively.

4.2 Results

4.2.1 Identification of pollen-expressed cDNA's encoding MYB and LIM domain motifs

4.2.1.1 *Myb.Nt1* and *myb.Nt2*

Two 42mer oligonucleotides, O1 and O2, were made available (Dr. C. Martin, Norwich); O1 (TTGCCIGGICGIACIGATAATGAIATTAAIAATTATTGGAAT; $T_m = 65$ °C) was a degenerate oligonucleotide corresponding to a *myb*-domain consensus sequence derived from a comparison of the most highly conserved *myb*-domain sequences from human c-MYB (Majello *et al.*, 1986), *Drosophila* DM (Katzen *et al.*, 1985) and the maize C1 (Paz-Ares *et al.*, 1987) sequences; O2 (CTGCCTGGCCGAACAGACAATGAAATCAAGAACTACTGGAAC; $T_m = 69.9$ °C) corresponded to a region of the maize C1 *myb*-domain sequence. The pair of oligonucleotides were used as a mixed probe to screen duplicate filters lifted from 4×10^5 pfu of the λZAPII cDNA library prepared from poly A^+ isolated from mature pollen of *N. tabacum* var. Samsun (section 2.6.3.3).

Four positively hybridising cDNA clones were isolated. Their inserts were isolated by a process involving *in vivo* excision and recircularisation from the λZAPII XR vector (section 2.6.6). The 'zapped' circular molecules contained phagemid sequences, pBluescript vector DNA and the cDNA insert. The circularised DNA was then packaged and used to infect *E. coli* cells which were spread on selection plates to produce colonies. DNA was prepared from mini cultures of these colonies and the cDNA insert sizes were determined by *XhoI/EcoRI* restriction digestion. The cDNA inserts were visualised via separation of the digests by agarose gel electrophoresis. Clones 1, 2 and 4 had inserts of 1.8 kb and clone 3 an insert of 1.7 kb. The inserts of all four cDNA clones were sequenced completely on both DNA strands using primer walking of double-stranded DNA templates (Figure 4.1a.). Computer sequence analysis revealed that clones 1 and 3, and clones 2 and 4 had identical nucleotide sequences.

EMBL and Swissprot data base searches were carried out using the deduced amino-acid sequences derived from the two unique cDNA clones, *myb.Nt1* and *myb.Nt2* (Figures 4.2 and 4.3). Regions of significant similarity to MYB-domain containing proteins were discovered. The cDNA library was rescreened using the *myb.Nt1* cDNA insert as a probe at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1) to determine the abundance of cDNAs with a more

a.
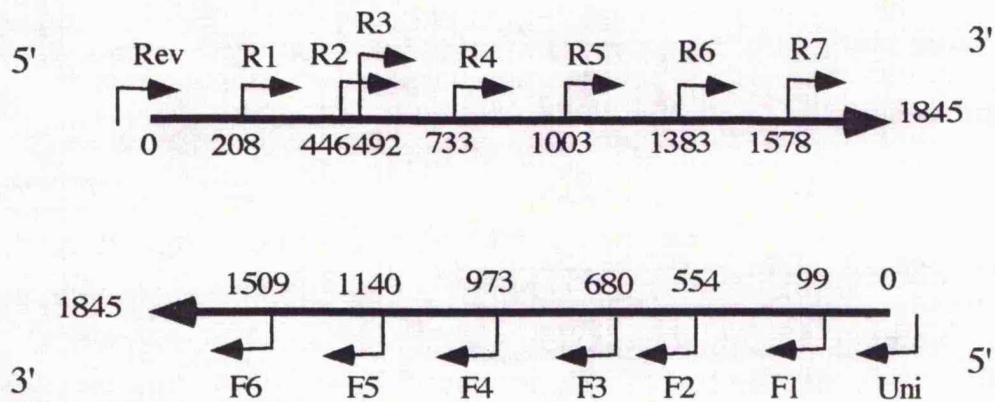
b.
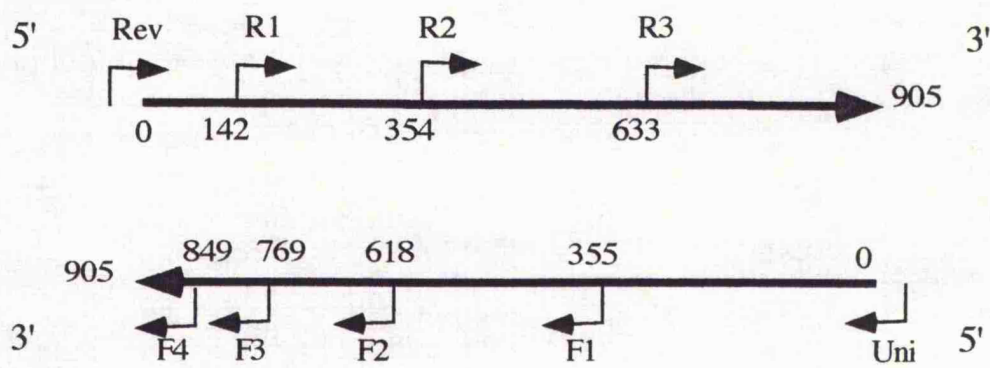
Figure 4.1
Sequencing primer positions on the *myb.Nt1* and *Ntl1* cDNA clones.

Abbreviations: R, sense strand; F, antisense strand; Rev, reverse primer; Uni, universal primer. Numbers refer to the 5' priming sites.

a. Position of sequencing primers on the *myb.Nt1* cDNA (same primers were used for *myb.Nt2*).

b. Position of sequencing primers on the *Ntl1* cDNA (same primers were used for *Ntl2*).

Figure 4.2

Nucleotide and deduced amino-acid sequence of the _myb.Nt1_ cDNA.

Amino-acid sequence is shown in three-letter code above the coding sequence. Compacted DNA sequence indicates 5' and 3' untranslated regions.

The predicted ATG initiation codon is underlined. The nucleotide region corresponding to oligonucleotides O1 and O2 is underlined. Predicted polyadenylation sites are underlined.

```
  1                                                              GTCTTTTTGGCTTATTTTT
 20 CCATCCATATTGATTCTTGAAGATAAAGAAGAAACCCCCAGATAAAGAATTGAAATAAAAAGGATTA
    Met Ala Pro Asp Gly Gly Gly Leu Lys Ala Arg Asn Asn Gly Gly Thr Arg      17
 87 ATG GCA CCA GAT GGA GGA GGA TTG AAA GCA AGA AAT AAT GGA GGG ACA AGG
    Gln Val Leu Lys Lys Gly Pro Trp Thr Ala Ala Glu Asp Ala Ile Leu Met      34
138 CAA GTA TTA AAA AAA GGT CCA TGG ACA GCA GCA GAA GAT GCA ATT TTA ATG
    Glu Tyr Val Lys Lys Asn Gly Glu Gly Asn Trp Asn Ala Val Gln Arg Asn      51
189 GAA TAT GTG AAG AAA AAT GGA GAA GGA AAT TGG AAT GCA GTA CAA AGG AAT
    Ser Gly Leu Met Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Ala Asn His      68
240 TCA GGA TTA ATG AGA TGT GGG AAA AGT TGT AGG TTA AGA TGG GCT AAT CAT
    Leu Arg Pro Asn Leu Lys Lys Gly Ala Phe Ser Leu Glu Glu Glu Arg Phe      85
291 TTG AGG CCT AAT CTT AAA AAA GGT GCA TTT TCT TTA GAG GAA GAA AGG TTT
    Ile Val Glu Leu His Ala Lys Leu Gly Asn Lys Trp Ala Arg Met Ala Ala     102
342 ATT GTG GAA CTT CAT GCT AAG CTT GGA AAT AAA TGG GCT CGT ATG GCT GCT
    Gln Met Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Tyr Trp Asn Thr Arg     119
393 CAG ATG CCA GGT AGA ACA GAC AAT GAA ATC AAA AAT TAC TGG AAT ACA AGG
    Leu Lys Arg Arg Gln Arg Ala Gly Leu Pro Ile Tyr Pro Gln Asp Ile Gln     136
444 CTA AAA AGA AGG CAG AGA GCT GGT TTG CCT ATA TAC CCT CAA GAC ATA CAA
    Pro Gln Leu Gln Gln Asn Ile Ser Ile Pro Ser Pro Phe Asp Asn Asn        153
495 CCA CAA TTA AAC CAA CAA AAT ATT TCT ATC CCT TCA CCA TTT GAT AAT AAT
    Pro Gln Asn Ser Asn Tyr Ile Asn Asn Pro Pro Leu Ser Leu Leu Asp Ile     170
546 CCT CAA AAT TCC AAT TAC ATT AAT AAC CCC CCT CTT TCC CTT TTA GAT ATT
    Phe Asn Pro Ser Thr Met Lys Pro Ser Asn Ile Ser Asn Gln Tyr Gln Ser     187
597 TTC AAT CCT TCA ACC ATG AAA CCT AGT AAT ATT TCA AAT CAA TAC CAA TCC
    Asn Asn Asn Pro Ser Pro Tyr Leu Thr Asn Asn Asn Asn Asn Asn Gln Phe     204
648 AAT AAT AAC CCT TCT CCT TAT CTT ACA AAT AAC AAT AAC AAC AAC CAA TTC
    Lys Phe Phe Arg Asp Pro Arg Val Arg Leu Ser Leu Thr Leu Ala Ser Ser     221
699 AAG TTT TTT CGC GAC CCT CGA GTT CGC CTT TCT TTA ACA TTA GCA TCA TCG
    Ile Arg Asn Ser Gln His Ser Ser Met Val Ala Pro Val Pro Asn Asn Phe     238
750 ATA AGA AAT TCA CAA CAT TCT TCA ATG GTA GCA CCC GTG CCT AAT AAT TTT
    Asn Gln Ser Tyr Ser Asn Ser Met Pro Val Pro Pro Leu Gln His Asn Tyr     255
801 AAT CAG AGC TAC TCT AAT TCA ATG CCA GTG CCA CCA CTT CAA CAT AAT TAT
    Pro Asn Phe Gly Ser Thr Thr Arg Pro Phe Thr Gly Ile Pro Ser Asn Pro     272
852 CCA AAT TTT GGC TCC ACT ACA AGA CCT TTT ACG GGG ATT CCC TCG AAC CCA
    Asn Gly Leu Ile Leu Gly Met Gly Val Gln Asn Asn Pro Ser Val Gln Ser     289
903 AAT GGT TTA ATC TTA GGA ATG GGC GTG CAA AAC AAC CCT TCA GTC CAA TCG
    Ser Met Pro Glu Thr Lys Thr Cys Ser Arg Asn Thr Gly Ser Asp Phe Met     306
954 TCC ATG CCT GAG ACT AAA ACA TGT TCT AGG AAT ACA GGG AGT GAT TTT ATG
    Asn Thr Thr Ser Ser Asp Gly Ala Asp Asn Tyr Asp Ile Asn Pro Gly Leu     323
1005 AAT ACA ACA TCA AGT GAT GGT GCT GAT AAT TAT GAT ATT AAT CCT GGA TTA
    Ser Arg Gly Asn Ser Gly Leu Leu Glu Asp Leu Leu Glu Glu Ser Gln Ala     340
1056 TCG CGA GGA AAC AGT GGA TTA TTG GAA GAT TTG TTA GAG GAG TCT CAG GCT
    Leu Thr Arg Val Glu Lys Ile Glu Glu Asn Cys Pro Ile Glu Asn Glu Ala     357
1107 TTA ACT AGA GTT GAG AAA ATA GAA GAG AAT TGT CCA ATT GAA AAC GAG GCT
    Ile Lys Gly Lys Leu Val Trp Glu Glu Tyr Gly Leu Ser Glu Glu Ala Glu     374
1158 ATT AAA GGG AAA TTG GTG TGG GAA GAA TAT GGA TTA AGT GAA GAG GCA GAA
    Asp Ile Ile Leu Thr Glu Glu Ser Thr Phe Ser Phe Ala Gln Glu Gly Gly     391
1209 GAT ATT ATT TTA ACT GAA GAA TCA ACA TTT TCT TTT GCT CAA GAA GGT GGT
    Asp Asn Ala Thr Pro Asn Lys His Ser Glu Asp Ser Ser Ser Leu Asn Ser     408
1260 GAT AAT GCC ACC CCA AAT AAG CAT TCT GAA GAC TCC AGC TCT CTC AAT TCA
    Ser Ser Gly Ile Thr Thr Lys Glu Gly Ser Leu Glu Leu Ala Asn Gln Val     425
1311 TCC TCA GGG ATA ACA ACA AAG GAA GGT TCA TTA GAG CTG GCT AAT CAA GTG
    Asp Glu Asp Ile Met Arg Phe Leu Asp Asn Phe Pro Val Gly Val Pro Val     442
1362 GAT GAG GAT ATT ATG AGA TTT CTT GAT AAT TTT CCA GTA GGA GTA CCA GTT
    Pro Asp Trp Cys Asn Asp Glu Asn Asp Glu Gln Asn Thr Ser Asn Gly Gln     459
1413 CCT GAT TGG TGT AAT GAT GAA AAT GAT GAG CAG AAT ACT TCT AAT GGC CAA
    Ser Phe Glu Cys Asp Gln Ile Gln Ser His Cys Ser Lys Ser Gly Stop      474
1464 TCT TTC GAA TGT GAC CAA ATA CAA TCA CAT TGC TCA AAA TCA GGC TGA TGA
1515 AAATTCAAATTCATGAATATGGCAAAATTGGAGCATGTTATTGTAGTGACGTGCTTGTTTTTTTTTG
1582 GTTAATTAGAACTAGTAAATTAAGACATTTATATTCATAAAAAAAGACCGCTAAGTAAATAGCAAAG
1649 ATTCTTGGATTTTGTTGTTACTTTTTTTTAAAAAAACTTTAGGTTAATATGTAGGTCTAATGATAAT
1716 TGGCCACTCGATTGTGTTGGCTTAATTATCTGTTGTGTACGATTCCTATGAATCATGGAATTTTTTT
1783 GAGTTAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 4.3
Nucleotide and deduced amino-acid sequence of *myb.Nt2* cDNA.

Amino-acid sequence is shown in three-letter code above the coding sequence. Compacted DNA sequence indicates 5' and 3' untranslated regions.

The predicted ATG initiation codon is underlined. The nucleotide region corresponding to oligonucleotides O1 and O2 is underlined. Predicted polyadenylation sites are underlined.

```
   1                           CCCCCCTCTCATATTCTCCCTCACCATCTTTGTCTATACAAGATAAA
  48 CAACAAAAATATACTAAAAGAAAAAAATAAAAAAGAGAGAAAAACAAAGCCTTTTTGGCTTATTTTT
 115 CCATCCAAATTCATTCTTGAAGATAAATAAACCCCAGATAAAGAATTGGAAAAGAAAGGAATA
       Met Ala Pro Asp Gly Gly Gly Leu Lys Ala Arg Asn Asn Gly Gly Thr Arg      17
 182 ATG GCA CCA GAT GGA GGA GGA TTG AAA GCA AGA AAT AAT GGA GGG ACA AGG
       Gln Val Leu Lys Lys Gly Pro Trp Thr Ala Thr Glu Asp Ala Ile Leu Met      34
 233 CAA GTA TTA AAA AAA GGT CCA TGG ACA GCA ACA GAA GAC GCG ATT TTA ATG
       Glu Tyr Val Lys Lys Asn Gly Glu Gly Asn Trp Asn Ala Val Gln Arg Asn      51
 284 GAA TAT GTG AAG AAA AAT GGA GAA GGA AAT TGG AAT GCA GTT CAA AGG AAT
       Ser Gly Leu Met Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Ala Asn His      68
 335 TCA GGA TTA ATG AGA TGT GGG AAA AGT TGT AGG TTA AGA TGG GCT AAT CAT
       Leu Arg Pro Asn Leu Lys Lys Gly Ala Phe Ser Leu Glu Glu Glu Arg Ile      85
 386 TTG AGG CCT AAT CTT AAA AAA GGT GCA TTT TCT TTA GAA GAA GAA AGG ATT
       Ile Val Glu Leu His Ala Lys Leu Gly Asn Lys Trp Ala Arg Met Ala Ala     102
 437 ATT GTG GAA CTT CAT GCT AAA CTT GGA AAC AAA TGG GCT CGT ATG GCT GCT
       Gln Met Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Tyr Trp Asn Thr Arg     119
 488 CAG ATG CCA GGT AGA ACA GAC AAT GAA ATC AAG AAT TAC TGG AAC ACA AGG
       Leu Lys Arg Arg Gln Arg Ala Gly Leu Pro Ile Tyr Pro Gln Asp Ile Gln     136
 539 CTA AAA AGA AGA CAA AGA GCT GGT TTG CCT ATA TAC CCT CAA GAC ATA CAA
       Leu Gln Leu Asn Gln Gln Glu Asn Gln Leu Gln His Ser Thr Ile Pro Ser     153
 590 CTA CAA CTA AAC CAA CAA GAA AAC CAA TTA CAA CAT AGT ACT ATC CCT TCA
       Pro Phe Asp Asn Asn Pro Gln Asn Ser Asn Tyr Ile Asn Pro Pro Leu Ser     170
 641 CCA TTT GAT AAT AAT CCC CAA AAC TCT AAC TAC ATT AAT CCC CCT CTT TCC
       Leu Leu Asp Ile Phe Asn Pro Ser Thr Met Lys Pro Ser Asn Ile Ser Asn     187
 692 CTT TTA GAT ATC TTC AAT CCT TCA ACC ATG AAA CCT AGT AAT ATT TCA AAT
       Gln Tyr Gln Phe Asn Asn Pro Ser Pro Tyr Leu Thr Thr Thr Asn Asn Asn     204
 743 CAA TAC CAA TTC AAT AAT CCT TCT CCA TAT CTC ACA ACT ACA AAC AAT AAC
       Asn Gln Leu Lys Leu Phe Arg Asp Pro Arg Val Ser Leu Ser Leu Thr Leu     221
 794 AAC CAA CTC AAG CTT TTT CGC GAT CCT CGG GTT AGC CTT TCT TTA ACA TTA
       Ala Ser Ser Ile Arg Asn Ser Gln Leu Ser Ser Met Val Ala Pro Val Pro     238
 845 GCA TCA TCG ATA AGA AAT TCA CAA CTT TCT TCA ATG GTA GCA CCC GTG CCA
       Asn Asn Phe Ser Gln Ser Tyr Ser Asn Ser Met Pro Val Pro Pro Leu Gln     255
 896 AAT AAT TTT AGT CAG AGC TAC TCT AAT TCA ATG CCA GTG CCA CCA CTT CAA
       His Asn Tyr Pro Asn Phe Gly Ser Thr Thr Arg Pro Phe Thr Gly Ile Pro     272
 947 CAT AAT TAT CCA AAT TTT GGC TCC ACT ACA AGA CCT TTT ACG GGG ATT CCC
       Ser Asn Pro Asn Gly Leu Ile Leu Gly Met Gly Val Gln Asn Asn Pro Ser     289
 998 TCG AAC CCA AAT GGT TTA ATC TTA GGC ATG GGC GTA CAA AAC AAC CCT TCA
       Val Gln Ser Ser Ile Pro Glu Thr Arg Ile Cys Ser Arg His Thr Thr Ser     306
1049 GTC CAA TCC TCC ATC CCC GAG ACC AGA ATA TGT TCT AGG CAT ACA ACA TCA
       Ser Asp Asp Ala Asp Asn Tyr Ala Val Asp Pro Gly Leu Ser Arg Gly Asn     323
1100 AGT GAT GAT GCT GAT AAT TAT GCT GTT GAT CCT GGA TTA TCG CGA GGA AAC
       Ser Gly Leu Leu Glu Asp Leu Leu Glu Glu Ser Gln Thr Leu Thr Arg Ala     340
1151 AGT GGA TTA TTG GAA GAT TTA TTG GAG GAG TCT CAG ACT TTG ACA AGA GCT
       Glu Lys Ile Glu Glu Asn Cys Pro Ile Glu Asn Glu Ala Gly Lys Gly Lys     357
1202 GAG AAA ATA GAG GAG AAT TGT CCA ATT GAG AAT GAG GCT GGT AAA GGG AAA
       Leu Val Trp Glu Glu Tyr Gly Leu Ser Glu Glu Ala Glu Asp Ile Ile Leu     374
1253 TTA GTG TGG GAA GAA TAT GGA TTA AGT GAA GAG GCA GAA GAT ATT ATT TTA
       Thr Glu Glu Ser Thr Phe Ser Phe Ala Gln Gln Gly Gly Gag Gat Gcc Acc     391
1304 ACT GAA GAA TCA ACA TTT AGT TTT GCT CAA CAA GGT GGT GAG GAT GCC ACC
       Pro Ile Lys His Ser Glu Asp Ser Thr Ser Leu Asn Ser Ser Ser Gly Ile     408
1355 CCA ATA AAG CAC TCT GAA GAC TCT ACC TCT CTC AAC TCA TCC TCA GGG ATA
       Thr Thr Lys Glu Gly Ser Leu Glu Leu Ala Asn Gln Val Asp Glu Asp Ile     425
1406 ACA ACA AAG GAA GGT TCA TTA GAG CTG GCT AAT CAA GTG GAT GAG GAT ATT
       Met Arg Phe Leu Asp Asn Phe Pro Val Gly Val Pro Val Pro Asp Trp Cys     442
1457 ATG AGA TTT CTT GAT AAT TTT CCA GTA GGT GTA CCA GTT CCT GAT TGG TGT
       Asn Asp Glu Asn Asp Gln Gln Asn Thr Ser Asn Gly Gln Ser Phe Glu Cys     459
1508 AAT GAT GAA AAT GAT CAG CAG AAT ACT TCT AAT GGC CAA TCT TTC GAA TGT
       Asp Gln Ile Gln Ser His Cys Ser Lys Ser Gly Stop                        470
1559 GAC CAA ATA CAA TCA CAT TGC TCA AAA TCA GGC TGA TGAGGATTCAAATTCATCA
1614 ATATGGGAAATTGGAGCATGTTATTGTAGTGACATTTTGTTTTTGTGTGTTAATTAGAACTAGTAAA
1681 TTAAGGCACTTATATTCATCAAAAAGCCGTCAAGTAAATAGCAAAGATTCTTGGATTTTGTTGTTAC
1748 CTTTTTTTCTTCTTATAAAGTTTAGGTTAATATGTAGATCTAATGATAATTGGCCATTCAATTGTGT
1815 TGGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

homologous probe. A group of 50 positively hybridising clones were visualised from a screen of 1×10^5 pfu. This represented a mRNA transcript abundance of 0.05 % of the pollen transcripts, assuming unbiased representation of pollen transcripts in the cDNA library.

4.2.1.2 *Ntl1* and *Ntl2*

Probes used to screen for a second class of putative transcriptional regulator were generated using PCR to amplify a region of the *sf3* LIM-domain containing cDNA from sunflower (Dr. A. Steinmetz, Strasbourg). In this way possible pBluescript DNA contamination of the probe was avoided which would hybridise to the library λZAP vector sequences to give a high background in the screening. Two primers were used to amplify a DNA fragment of 660 bp containing the two contiguous LIM-domain sequences and the pentapeptide C-terminal repeat of *sf3*. In order to ensure that the probe fragment was specific for the *lim*-domain sequences, the pentapeptide repeat fragment was removed by *Xho*I restriction digestion, producing a 530 bp DNA fragment containing only the two SF3 LIM-domains. This DNA fragment was labelled by oligonucleotide random priming ((Feinberg and Vogelstein, 1983), section 2.6.5.1) and used for screening (section 2.6.3.1). From an initial screen of 4×10^5 pfu, approximately 2000 positively hybridising plaques were obtained. Assuming that all 2000 plaques represented SF3 homologous transcripts this represented a mRNA transcript abundance of 0.5 % of pollen transcripts, assuming unbiased representation of the pollen transcripts in the cDNA library. A group of 29 positive clones were isolated and the cDNA inserts recovered as 'phagemid' by *in vivo* excision, and sized by restriction digestion agarose gel electrophoresis. Most of the clones had an insert size of 0.9 kb. A smaller group of 4 clones were randomly selected for further analysis. Clone 1 had an insert size of 0.9 kb, clones 2 and 3, 0.8 kb and clone 4 was found to generate multiple fragments of 0.9, 0.6 and 0.1 kb. Further restriction mapping and Southern analysis of clone 4 revealed that only the 0.9 kb insert had sequence homology to the LIM-domain probe. The inserts of all four cDNA clones were sequenced completely on both DNA strands using primer walking of double-stranded DNA templates (Figure 4.1b.). Computer sequence analysis revealed that clone 1 and the a 0.9 kb 5'-terminal portion of clone 4 had identical nucleotide sequences, as did clones 2 and 3.

Figure 4.4

Nucleotide and deduced amino-acid sequence of *Ntl1* and *Ntl2* cDNAs.

Amino-acid sequences are shown in three-letter code above the coding sequences. Compacted DNA sequences indicates 5' and 3' untranslated regions.

The predicted ATG initiation codons are underlined.

a. *Ntl1*

b. *Ntl2*. The predicted polyadenylation sites are underlined.

to determine the abundance of cDNAs with a more homologous probe. A similar frequency of 0.5 % of hybridising clones within the library was obtained.

a.

```
  1                          GGAAGAGGGGAGAGAGAGAACAGAGTAGAAAGGAATTATTGATTGTCT
 49 TCAAATATCTGTTCTCCTGCCTTCGCCTCTTTTCAACAATCGTAGAATTTACTTTCAACAATCAAAG
    Met Thr Phe Ala Gly Thr Thr Gln Lys Cys Ser Ala Cys Glu Lys Thr Val    17
116 ATG ACG TTC GCA GGA ACA ACG CAA AAG TGC AGT GCT TGT GAG AAA ACG GTG
    Tyr Leu Val Asp Arg Leu Ala Ala Asp Asn Arg Ile Tyr His Lys Ala Cys    34
167 TAT CTG GTG GAT CGT CTT GCT GCT GAT AAT CGC ATT TAT CAC AAG GCT TGC
    Phe Arg Cys Tyr His Cys Lys Ser Thr Leu Lys Leu Ser Asn Phe Asn Ser    51
218 TTT AGG TGC TAC CAC TGC AAA AGT ACT CTC AAG CTC AGT AAT TTC AAC TCC
    Phe Glu Gly Val Ile Tyr Cys Arg Pro His Phe Asp Gln Leu Phe Lys Arg    68
269 TTT GAG GGG GTA ATT TAC TGT AGG CCT CAC TTT GAT CAG CTT TTT AAG AGA
    Thr Gly Ser Leu Asp Lys Ser Phe Glu Gly Thr Pro Lys Val Thr Lys Pro    85
320 ACT GGC AGT TTG GAC AAG AGT TTT GAA GGA ACT CCG AAA GTC ACA AAG CCA
    Glu Lys Pro Val Asp Asn Glu Asn Gly Ser Gly Thr Lys Val Ser Ser Leu   102
371 GAA AAA CCT GTG GAC AAC GAG AAT GGT AGC GGG ACG AAA GTC TCA AGT TTA
    Phe Ala Gly Thr Arg Glu Lys Cys Val Gly Cys Thr Lys Thr Val Tyr Pro   119
422 TTT GCA GGC ACG AGG GAA AAA TGT GTG GGC TGC ACT AAA ACT GTG TAT CCG
    Ile Glu Lys Val Ser Val Asn Gly Thr Ala Tyr His Lys Ala Cys Phe Lys   136
473 ATT GAG AAG GTA AGT GTG AAC GGA ACA GCA TAC CAT AAG GCC TGC TTC AAA
    Cys Ser His Gly Gly Cys Thr Ile Ser Pro Ser Asn Tyr Ile Ala His Glu   153
524 TGC AGT CAT GGC GGC TGT ACA ATA AGC CCA TCA AAT TAC ATT GCA CAT GAA
    Gly Arg Leu Tyr Cys Lys His His His Ile Gln Leu Phe Lys Glu Lys Gly   170
575 GGA AGG CTT TAT TGC AAG CAC CAC CAT ATT CAA CTC TTC AAG GAG AAA GGC
    Asn Tyr Ser Gln Leu Glu Ser Asp His Glu Thr Asp Pro Ala Leu Ser Thr   187
626 AAT TAC AGC CAG CTT GAG TCT GAC CAT GAA ACT GAC CCT GCT CTT TCC ACT
    Gln Ser Leu Thr Stop                                                   191
677 CAA TCA TTA ACT TAA AACAATATTCCTCAAATTGTTTTCTCTTTCATTTCTCCCTATCTTTT
739 CCTCGTCACTGATTCTGGATTTTGTTCTTTCTTACGGGGTAAATTCAGTTATATCTATCCATTTTGG
806 TTGTACTTCTATGTATTCTTGTCGGTTTCTGTACATTGTAGCATATTGAAGGTTGTTTTAAAAATTG
873 AACCAGAATCATTGTAAAAAAAAAAAAAAAAAA
```

b.

```
  1                        AGAGAGAGAGGGGAGAGAGAGAACAGAGTGGAAAG
 36 GAATTATTGATTGTCTTCAAATATCTGTTCTTCTCTTGCCTTTGCCTCTTTTCAACAATAATCAACG
    Met Thr Phe Ala Gly Thr Thr Gln Lys Cys Ser Ala Cys Glu Lys Thr Val    17
103 ATG ACG TTC GCA GGA ACA ACG CAA AAG TGC AGT GCT TGT GAG AAA ACG GTG
    Tyr Leu Val Asp Arg Leu Ala Ala Asp Asn Arg Ile Tyr His Lys Ala Cys    34
154 TAT CTG GTA GAT CGC CTT GCT GCT GAT AAT CGT ATT TAT CAC AAG GCT TGC
    Phe Arg Cys Tyr His Cys Lys Ser Thr Leu Lys Leu Ser Asn Phe Asn Ser    51
205 TTC AGG TGC TAC CAT TGC AAA AGT ACT CTC AAG CTC AGT AAT TTC AAC TCC
    Phe Glu Gly Val Ile Tyr Cys Arg Pro His Phe Asp Gln Leu Phe Lys Arg    68
256 TTT GAG GGG GTA ATT TAC TGT AGG CCT CAC TTT GAT CAG CTT TTT AAG AGA
    Thr Gly Ser Leu Asp Lys Ser Phe Asp Gly Thr Pro Lys Val Thr Lys Pro    85
307 ACT GGC AGT TTG GAC AAG AGT TTT GAC GGA ACT CCG AAA GTC ACA AAG CCA
    Glu Lys Ser Val Glu Asn Glu Asn Gly Ser Gly Ser Lys Val Ser Ser Leu   102
358 GAA AAA TCT GTG GAG AAC GAG AAT GGT AGC GGG AGC AAA GTC TCA AGT TTA
    Phe Ala Gly Thr Arg Glu Lys Cys Val Gly Cys Thr Lys Thr Val Tyr Pro   119
409 TTT GCA GGC ACG AGG GAA AAA TGT GTG GGC TGC ACT AAA ACT GTC TAT CCG
    Ile Glu Lys Val Ser Val Asn Gly Thr Ala Tyr His Lys Gly Cys Phe Lys   136
460 ATT GAG AAG GTT AGT GTG AAC GGG ACA GCA TAC CAT AAG GGC TGC TTC AAA
    Cys Ser His Gly Gly Cys Thr Ile Ser Pro Ser Asn Tyr Ile Ala His Glu   153
511 TGC AGT CAT GGA GGC TGT ACA ATA AGC CCA TCA AAT TAC ATT GCA CAT GAG
    Gly Arg Leu Tyr Cys Lys His His His Ile Gln Leu Phe Lys Glu Lys Gly   170
562 GGA AGG CTT TAT TGC AAG CAC CAC CAT ATT CAA CTC TTC AAG GAG AAA GGC
    Asn Tyr Ser Gln Leu Glu Ser Asp His Glu Thr Stop                       181
613 AAT TAC AGC CAG CTT GAG TCT GAC CAT GAA ACA TAA AGTTAAAATGCCAAATAGA
668 AACAAACAAAAAAAAAAAAAAAATCTCTCTTTCATTTCTCCCTATCTTTTCCGCAGCTCGTTTTCAACT
735 GATTCTTGATTTTGTTCTTTCTTACGGCGTAAATTCAGTTGTCTCTATCCATTTTGGTTGTAAAAAA
    AAAAAAAAAAA
```

EMBL and Swissprot data base searches were carried out using the deduced amino-acid sequences derived from the two unique cDNA clones, *Nt1l* and *Nt2l* (Figure 4.4). Regions of significant similarity to LIM-domain containing proteins were discovered. The cDNA library was rescreened using the *Nt1l* cDNA insert as a probe at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1) to determine the abundance of cDNA's with a more homologous probe. A similar frequency of 0.5 % of hybridising clones within the library was obtained.

4.2.2 Determination of the nucleotide sequence of *myb.Nt1*, *myb.Nt2*, *Nt1l* and *Nt2l*

4.2.2.1 *Myb.Nt1*

The *myb.Nt1* cDNA insert was 1810 bp in length (Figure 4.2). The longest open reading frame was 1419 bp in length, starting from the first ATG codon of the cDNA sequence at position 87 bp and terminating with two TGA stop codons in tandem at positions 1509 bp and 1512 bp. The open reading frame codes for a predicted protein of 474 amino acids with a calculated molecular weight of 52.2 kDa. The sequence surrounding the initiator methionine was in good agreement with the consensus derived from eukaryotic genes (81 % agreement; Kozak, 1986). The initiation codon was unambiguously identified, since it was preceded 4 codons upstream and in phase, by a termination codon TAA. The 5' and 3' untranslated regions were 86 and 302 bp in length respectively. No consensus polyadenylation signal (AATAAA) was identified upstream of the 23 bp poly A tail (Joshi, 1987), however two sequences showing a 1 bp mismatch from the consensus sequence was present at positions 1600 (AATTAA) and 1618 (CATAAA). These predicted polyadenylation signal sequences are in an unusual position however as they lie more than 20-30 bp 5' upstream of the poly A addition site. This feature has been seen in a group of pollen-specific genes including *Zmc13* from maize (Hanson *et al.*, 1989), *adh-1* from maize (Sachs *et al.*, 1986) and *P2* from *Onethera organensis* (Brown and Crouch, 1990). The distance between the AATAAA motif and the addition site on most plant mRNAs is 27± 9 nucleotides (Joshi, 1987).

4.2.2.2 *Myb.Nt2*

Determination of the complete *myb.Nt2* cDNA sequence revealed an insert length of 1845 bp (Figure 4.3). The longest open reading frame was 1307 bp in length, starting

from the first ATG codon of the cDNA sequence at position 182 bp and terminating with two tandem TGA stop codons at positions 1592 and 1595 bp. The open reading frame codes for a predicted protein of 470 amino acids with a calculated molecular weight of 52.7 kDa. The sequence surrounding the initiator methionine was in good agreement with the consensus derived from eukaryotic genes, although its context was not as favourable for translation initiation as that of *myb.Nt1* (60 % agreement). The initiation codon was unambiguously identified, since it was preceded 4 codons upstream and in phase, by a termination codon TGA. The 5' and 3' untranslated regions were 181 and 353 bp in length respectively. No consensus polyadenylation signal was identified upstream of the 25 bp poly A tail, however three sequences showing a 1 bp mismatch from the consensus signal are present at positions 1675 (AGTAAA), 1714 (AGTAAA) and 1761 (TATAAA). These predicted polyadenylation signal sequences again lie more than 20-30 bp 5' upstream of the poly A addition site. Both the O1 and O2 oligonucleotides recognised a region corresponding to nucleotides 491 to 532 of *myb.Nt1* and 396 to 437 of *myb.Nt2* respectively (Figure 4.5). The percentage sequence identity between the O1 and O2 oligonucleotides and the *myb.Nt* cDNA clones over these regions was 71 % and 88 % for *myb.Nt1* and 73 % and 83 % for *myb.Nt2* respectively.

4.2.2.3 *Ntl1*

The *Ntl1* cDNA insert was 904 bp in length (Figure 4.4a.). The longest open reading frame was 573 bp in length, starting from the first ATG codon of the cDNA sequence at position 116 and terminating with a TAA stop codon at position 689. This open reading frame codes for a putative protein of 191 amino acids with a calculated molecular weight of 21.4 kDa. The sequence surrounding the initiator methionine was in good agreement with the consensus derived from eukaryotic genes (51 % agreement). The initiation codon was unambiguously identified, since it was preceded 30 triplets upstream and in phase by a termination codon TAG. The 5' and 3' untranslated regions were 115 and 212 bp in length respectively. No consensus polyadenylation signal (AATAAA) was identified upstream of the 17 bp poly A tail.

4.2.2.4 *Ntl2*

The *Ntl2* cDNA insert was 813 bp in length (Figure 4.4). The longest open reading frame was 543 bp in length, starting from the first ATG codon of the cDNA sequence

a.

```
                                                                O1
  1      TTGCCXGGXCGXACXGATAATGAXATTAAXAATTATTGGAAT      42
         ···· ··   ·  ·· ··  ···· ··  ··  ···· ····
487 TCAGATGCCAGGTAGAACAGACAATGAAATCAAGAATTACTGGAACACAA 536
    ···· ··   ··················· ·· ········
  1      CTGCCTGGCCGAACAGACAATGAAATCAAGAACTACTGGAAC      42
                                                                O2
```

b.

```
        1    LXXXXDNXIXNYWN   14   O1
        104  MPGRTDNEIKNYWN  117
        1    LPGRTDNEIKNYWN   14   O2
             lPgRTDNeiKNyWn       MYB.PLANT CONSENSUS
```

c.

```
  1      TTGCCXGGXCGXACXGATAATGAXATTAAXAATTATTGGAAT      42 O1
         ···· ··   ·  ·· ··  ···· ··  ··  ···· ····
392 TCAGATGCCAGGTAGAACAGACAATGAAATCAAAAATTACTGGAATACAA 441
    ···· ··   ···················· ·· ········
  1      CTGCCTGGCCGAACAGACAATGAAATCAAGAACTACTGGAAC      42 O2
```

d.

```
        1    LXXXXDNXIXNYWN   14   O1
        104  MPGRTDNEIKNYWN  117
        1    LPGRTDNEIKNYWN   14   O2
             lPgRTDNeiKNyWn       MYB.PLANT  CONSENSUS
```

Figure 4.5

Sequence alignment of O1 and O2 oligonucleotides with *myb.Nt1* and *myb.Nt2* partial sequences.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

a. O1 and O2 nucleotide sequences are shown above and below the partial *myb.Nt1* nucleotide cDNA sequence. Numbers refer to nucleotide positions.

b. O1 and O2 predicted amino acid sequences are shown above and below the partial predicted amino acid MYB.NT1 sequence. Upper case letters in the MYB.PLANT CONSENSUS sequence refer to fully conserved residues in plant MYB-domains, lower case letters refer to residues conserved in approximately 80 % of plant MYB-domains. Numbers refer to amino acid positions.

c. O1 and O2 nucleotide sequences are shown above and below the partial *myb.Nt2* nucleotide cDNA sequences. Numbers refer to nucleotide positions.

d. O1 and O2 predicted amino acid sequences are shown above and below the partial predicted amino acid MYB.NT2 sequence. Upper case letters in the MYB.PLANT CONSENSUS sequence refer to fully conserved residues in plant MYB-domains, lower case letters refer to residues conserved in approximately 80 % of plant MYB-domains. Numbers refer to amino acid positions.

at position 103 and terminating with a TAA stop codon at position 646. This open reading frame codes for a predicted protein of 181 amino acids with a calculated molecular weight of 20.3 kDa. The sequence surrounding the initiator methionine was in good agreement with the consensus derived from eukaryotic genes (60 % agreement). The initiation codon was unambiguously identified, since it was preceded 3 codons upstream and in phase, by a termination codon TAA. The 5' and 3' untranslated regions were 102 and 164 bp in length respectively. No consensus polyadenylation signal (AATAAA) was identified upstream of the 18 bp poly A tail, however a sequence showing a 1 bp mismatch to the consensus sequence was present at position 662 (AATAGA). This predicted polyadenylation signal sequence is in an unusual position however as it lies more than 20-30 bp 5' upstream of the poly A addition site.

4.2.3 Prediction and analysis of the encoded polypeptides
4.2.3.1 MYB.NT1 and MYB.NT2

The hydropathy plot of Kyte and Doolittle (Kyte and Doolittle, 1982) predicts a hydrophilic character of the MYB.NT1 and MYB.NT2 proteins (Figure 4.6), which is reflected in the predicted pI values of both proteins of 5.2. Assuming the cDNA sequence is fuul length, the absence of any hydrophobic signal peptide sequence at the amino-terminus, suggests that the proteins are intracellular (cytoplasmic or nuclear). The secondary structure predictions of Chou-Fasman (Figures 4.7) predicts the presence of six possible N-glycosylation sites (NXS/T) on MYB.NT1 (at amino acid positions 175, 184, 193, 287, 402, 409 and 449) and ten on MYB.NT2 (positions 143, 173, 180, 190, 239, 284, 307, 393, 407 and 454). This prediction also highlights the lack of any hydrophobic signal peptide sequence.

The predicted amino-acid sequences of MYB.NT1 and MYB.NT2 were aligned (Figure 4.8). An overall sequence identity of 88 % is seen with tracts of high similarity over the entire length of the proteins. EMBL and Swissprot data base homology searches using the complete MYB.NT1 amino acid sequence revealed homology to a range of MYB-domain proteins. Overall sequence homology was higher to other plant MYB-domain proteins (20-48 %) than to animal proteins (10-17 %). The MYB.NT1 protein shared highest homology to the MYB.PH3 and MYB.PH2 proteins from *Petunia hybrida* (48 and 32 %; (Avila *et al.*, 1993)), the maize MYB.ZMP1 (30 %; (Grotewold *et al.*, 1994)) and MYB.GA from barley (27 %; (Gubler *et al.*, 1995).

Detailed alignment of the amino acid sequences enables a more informative identification of residues which are conserved among different sequences. Since it is more likely for sequences adjacent to essential amino acids to be conserved in related proteins, important regions can easily be identified in such alignments. Alignment of the amino acid sequences of the two *N. tabacum myb* clones with the *P. hybrida* MYB.PH3 protein revealed extensive sequence homology to two N-terminal contiguous MYB repeats, which together constitute the MYB domain in plants (Figure 4.9). In addition to the MYB-domain, islands of homology exist both centrally and at the carboxy-terminus. These largely correspond to regions predicted to be α-helical by the algorithms of Chou-Fasman (Chou and Fasman, 1978), Garnier (Garnier *et al.*, 1978) and acidic by the algorithm of Eisenberg (Eisenberg *et al.*, 1984), and are therefore likely to be potential acidic transactivation domains. These regions are at positions 124 to 133 (10 amino acids), 320 to 335 (8 amino acids) and 367 to 384 (17 amino acids) in the MYB.NT1 protein. The first of these regions is rich in glutamine and proline residues and the last of which are rich in glutamate and aspartate residues. In addition two potential protein kinase C phosphorylation sites are present in the MYB proteins at positions 179 and 409, the first of these sites being conserved with MYB.PH3, and several serine/threonine rich regions are present in all three proteins which are also potential sites of phosphorylation.

The clustering of sequence homology to the MYB-domains is readily seen when regions of the MYB.NT1 and MYB.NT2 predicted polypeptides are compared to the equivalent regions of other MYB-domain proteins (Figure 4.10). Here on average, 80 % of the overall sequence homology in plant and animal proteins is constituted by the MYB-domain repeats with little significant sequence similarity, with the exception of the *P. hybrida* proteins, found outside the N-terminal region.

The MYB-domain repeats of the MYB.NT proteins were aligned with repeats from selected members of the MYB protein family (Figure 4.11). Particular conserved tryptophan residues are apparent, except for the first tryptophan of the second repeat in the plant proteins, which is replaced by phenylalanine, isoleucine or leucine. The three tryptophan residues are regularly spaced by 18-19 residues, and are implicated in DNA binding (Saikumar *et al.*, 1990), by forming hydrophobic cores to the MYB repeats over which the three α-helices are situated, which contact DNA (Ogata *et al.*, 1994). The plant MYB repeats bear more sequence similarity over the carboxy-terminal half where the helix-turn-helix motif may occur. This is less apparent in the case of the second repeat, although conservative substitutions maintain the basic nature

99

Figure 4.6

Hydrophobicity plot of the deduced MYB.NT1 and MYB.NT2 peptides using the Kyte and Doolittle prediction.

Hydropathy plot according to Kyte and Doolittle (Kyte and Doolittle, 1982); window size: 9 residues. R2 and R3 indicate MYB-domain repeats. Bold lines underline putative acidic activator domains.

a. MYB.NT1.
b. MYB.NT2.

a.

Hydrophobic
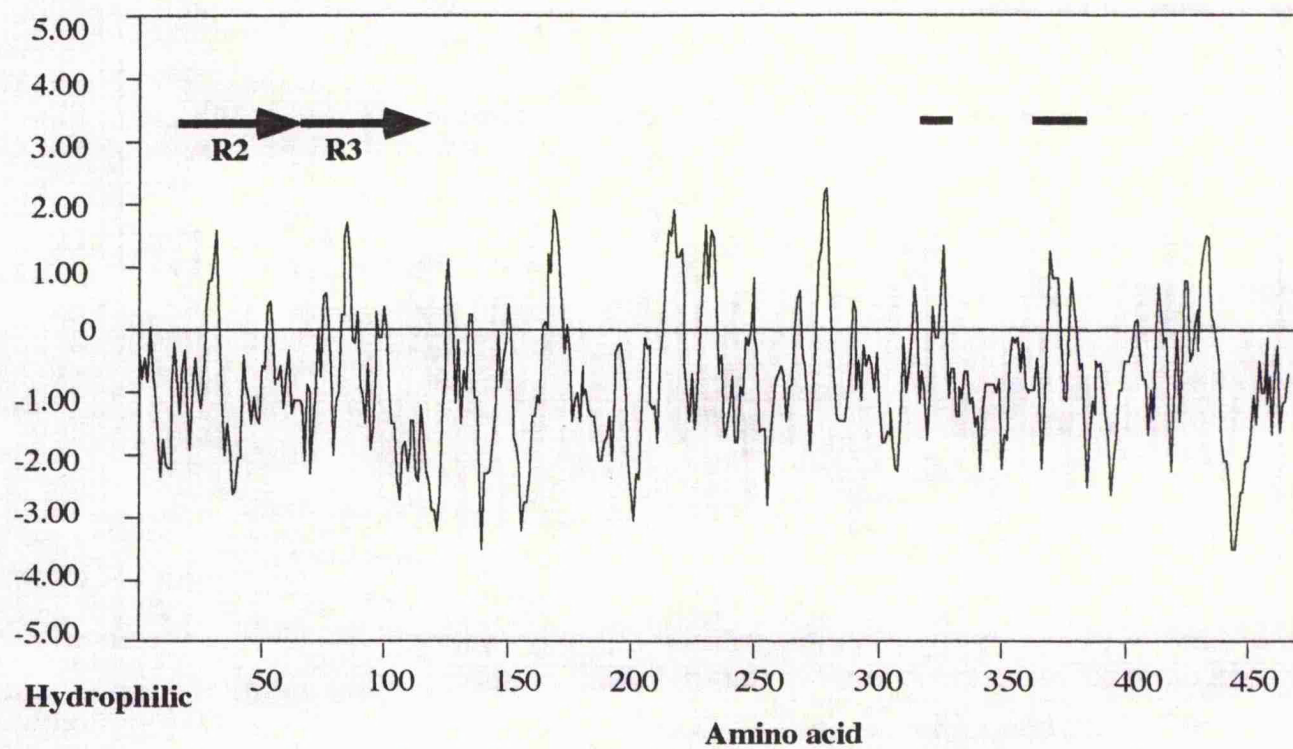
b.

Hydrophobic
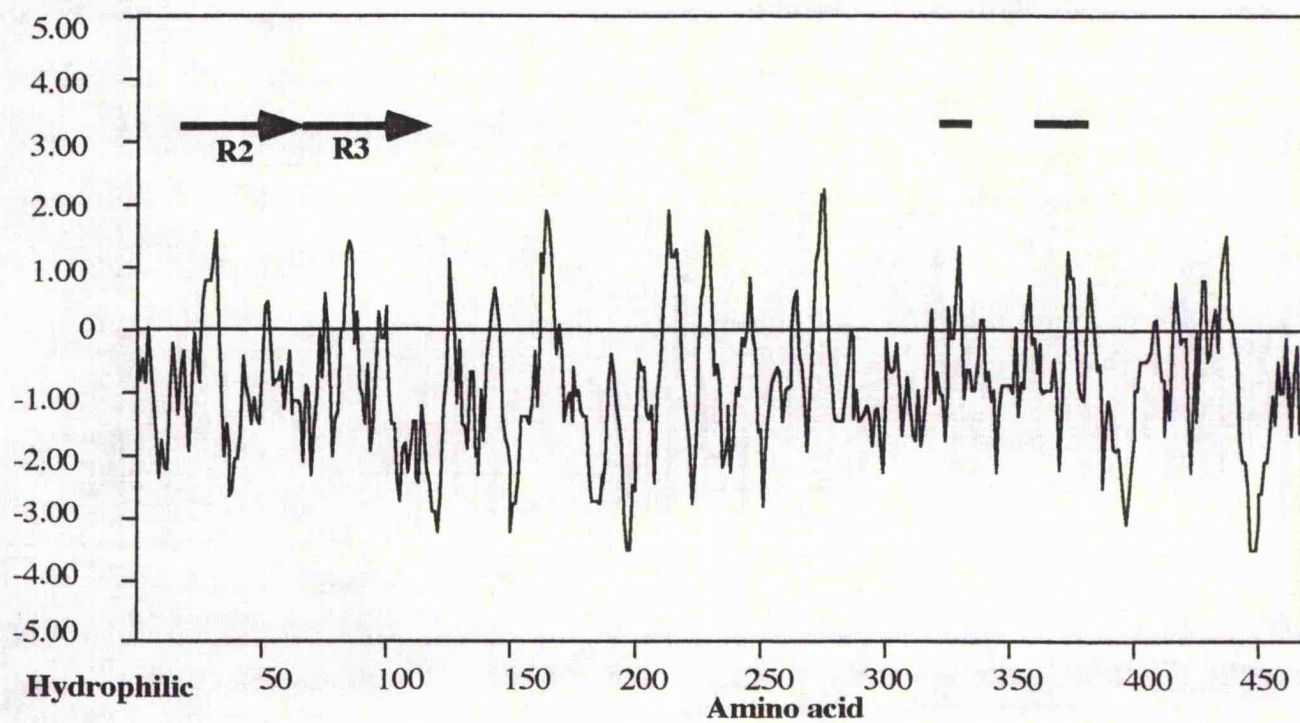
Figure 4.7

Secondary structure illustration of the deduced MYB.NT1 peptide using the Chou-Fasman prediction.

Performed using the PLOTSTRUCTURE program from the GCG suite of programs (Devereux *et al.*, 1984).

Putative glycosylation sites are indicated with a 'lollipop'. Hydrophilic and hydrophobic regions are represented with oval and diamond symbols, respectively.

PLOTSTRUCTURE of: MYB.NT1 ck: 3024

MYB.NT1 translated from 182 to 1594

Chou-Fasman Prediction
November 25, 1995 14:21

◯ KD Hydrophilicity >=1.3
◇ KD Hydrophobicity >=1.3

NH2

COOH

50
100
150
200
250
300
350
400
450
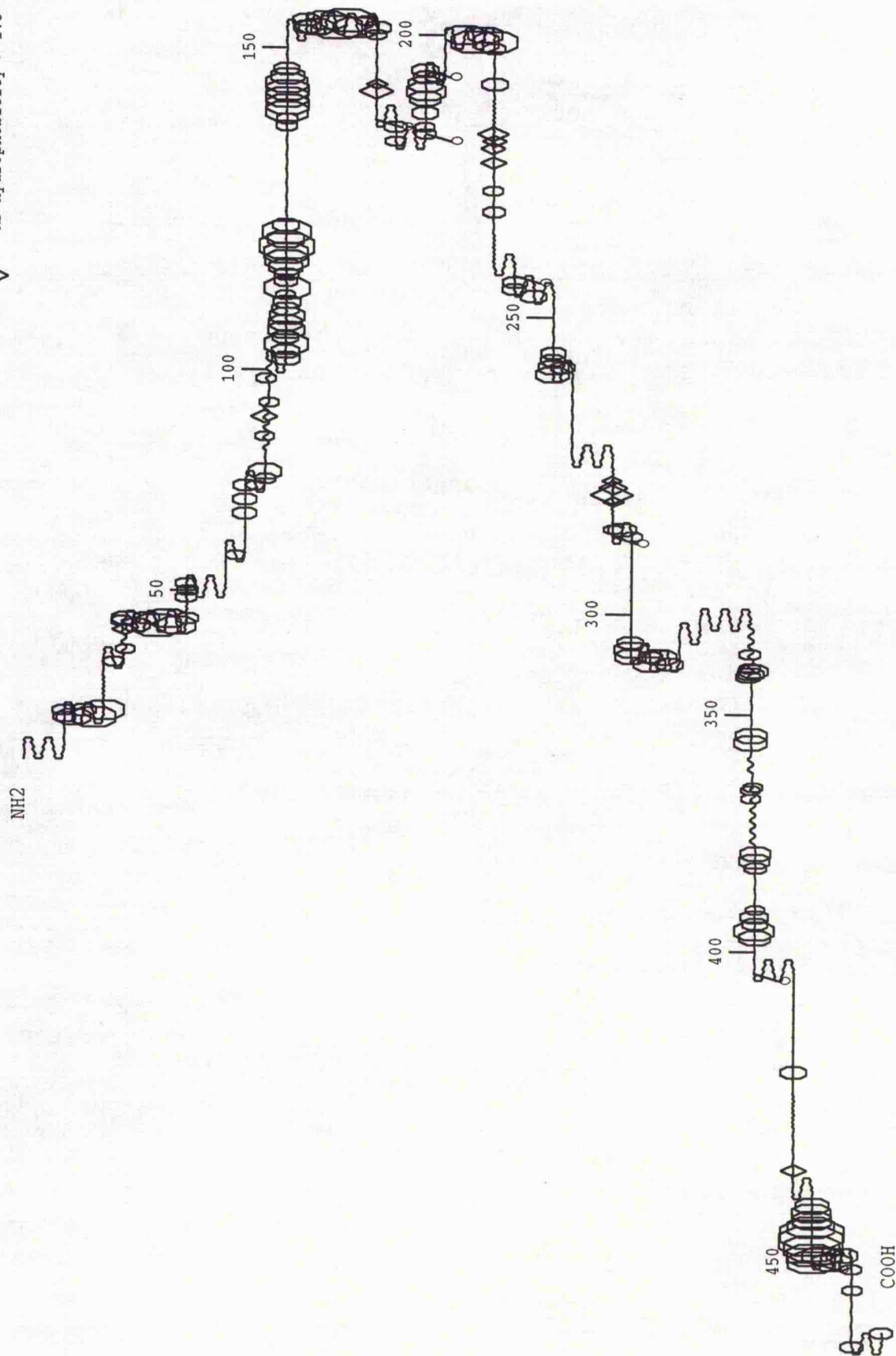
Figure 4.8

Sequence alignment of the deduced MYB.NT1 and MYB.NT2 proteins.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

```
MAPDGGGLKARNNGGTRQVLKKGPWTATEDAILMEYVKKNGEGNWNAVQRNSGLMRC  57
·············································
MAPDGGGLKARNNGGTRQVLKKGPWTAAEDAILMEYVKKNGEGNWNAVQRNSGLMRC  57

GKSCRLRWANHLRPNLKKGAFSLEEERIIVELHAKLGNKWARMAAQMPGRTDNEIKN  114
·············································
GKSCRLRWANHLRPNLKKGAFSLEEERFIVELHAKLGNKWARMAAQMPGRTDNEIKN  114

YWNTRLKRRQRAGLPIYPQDIQLQLNQQENQLQHSTIPSPFDNNPQNSNYI-NPPLS  170
·············································
YWNTRLKRRQRAGLPIYPQDIQPQLNQQNIS-----IPSPFDNNPQNSNYINNPPLS  166

LLDIFNPSTMKPSNISNQYQF-NNPSPYLTTTNNNNQLKLFRDPRVSLSLTLASSIR  226
·············································
LLDIFNPSTMKPSNISNQYQSNNNPSPYLTNNNNNNQFKFFRDPRVRLSLTLASSIR  223

NSQLSSMVAPVPNNFSQSYSNSMPVPPLQHNYPNFGSTTRPFTGIPSNPNGLILGMG  283
·············································
NSQHSSMVAPVPNNFNQSYSNSMPVPPLQHNYPNFGSTTRPFTGIPSNPNGLILGMG  280

VQNNPSVQSSIPETRICSRH-------TTSSDDADNYAVDPGLSRGNSGLLEDLLEE  333
·············································
VQNNPSVQSSMPETKTCSRNTGSDFMNTTSSDGADNYDINPGLSRGNSGLLEDLLEE  337

SQTLTRAEKIEENCPIENEAGKGKLVWEEYGLSEEAEDIILTEESTFSFAQQGGEDA  390
·············································
SQALTRVEKIEENCPIENEAIKGKLVWEEYGLSEEAEDIILTEESTFSFAQEGGDNA  394

TPIKHSEDSTSLNSSSGITTKEGSLELANQVDEDIMRFLDNFPVGVPVPDWCNDEND  447
·············································
TPNKHSEDSSSLNSSSGITTKEGSLELANQVDEDIMRFLDNFPVGVPVPDWCNDEND  451

QQNTSNGQSFECDQIQSHCSKSG  470
·············································
EQNTSNGQSFECDQIQSHCSKSG  474
```

Figure 4.9

Sequence alignment of the deduced MYB.NT1, MYB.NT2 and MYB.PH3 proteins.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

'o' signifies an amino acid identity;

'|' signifies a conservative amino acid substitution;

'·' mark every tenth amino acid residue.

The two imperfect MYB repeats are boxed. Putative protein kinase C phosphorylation sites are boxed and shaded. Putative acidic activator domains are underlined in bold (predicted to be α–helical by the methods of both (Chou and Fasman, 1978) and (Garnier et al.,, 1978) and amphipathic by the algorithm of (Eisenberg et al., 1984)).

```
              ....|.|.              .         .||.      .||.....|......|...|.
MYB.NT1   MAPDGGGLK---------ARNNGGTRQVLKKGPWTATEDAILMEYVKKNG 41
MYB.NT2   MAPDGGGLK---------ARNNGGTRQVLKKGPWTAAEDAILMEYVKKNG 41
MYB.PH3   MAPDDRGMKNGGASTGRSNGAGSSRQVLKKGPWTAAEDSILMEYVKKHG 49

          ..............................   ..|||.|..|.||
MYB.NT1   EGNWNAVQRNSGLMRCGKSCRLRWANHLRPNLKKGAFSLEEERIIVELH 90
MYB.NT2   EGNWNAVQRNSGLMRCGKSCRLRWANHLRPNLKKGAFSLEEERFIVELH 90
MYB.PH3   EGNWNAVQRNSGLMRCGKSCRLRWANHLRPNLKKGAFTVEEERIIIELH 98

          ...........................................||..
MYB.NT1   AKLGNKWARMAAQMPGRTDNEIKNYWNTRLKRRQRAGLPIYPQDIQLQL 139
MYB.NT2   AKLGNKWARMAAQMPGRTDNEIKNYWNTRLKRRQRAGLPIYPQDIQPQL 139
MYB.PH3   AKLGNKWARMAAQLPGRTDNEIKNYWNTRLKRRQRAGLPIYPQELQQQN 147

          ||  .  |   ||.||.  ....  .  .|......|  TMKPSNISN
MYB.NT1   NQQENQLQHSTIPSPFDNNPQNSNYIN-PPLSLLDIFNPSTMKPSNISN 187
MYB.NT2   NQ-----QNISIPSPFDNNPQNSNYINNPPLSLLDIFNPSTMKPSNISN 183
MYB.PH3   QHENNNQPHSLLSSSYD--PQNSTNYNSPSLSLLDIFNPSTMKPS-ITQ 193

          .|  .  ..  .  ...  |..|  .....|.|.....|
MYB.NT1   QYQFNN-PSPYLTTTNNNNQLKLFRDPRVSLSLTLASSIRNSQLSSMVA 235
MYB.NT2   QYQSNNNPSPYLTNNNNNNQFKFFRDPRVRLSLTLASSIRNSQHSSMVA 232
MYB.PH3   QFPINT-PSLCLPSTNNNN---IFRNTPKGLSLTLPSSMRNSQFSSLPN 238

          .  .|   .  .|. |.....|.....
MYB.NT1   PVPNNFSQSYS-NSMPVPPLQHNYPNFGSTTRPFTGIPSNPNGLILGMG 283
MYB.NT2   PVPNNFNQSYS-NSMPVPPLQHNYPNFGSTTRPFTGIPSNPNGLILGMG 280
MYB.PH3   ---NNFTQGLSSNSIQLPPFQHNYPNF-NINRPFTGISSNPNGLICGMG 283

          ||  .  ..|.  .  |.  ...|.|..  ......
MYB.NT1   VQ--NNPSVQSSIPETRICSRHT-------TSSDDADNYAVDPGLSRGN 323
MYB.NT2   VQ--NNPSVQSSMPETKTCSRNTGSDFMNTTSSDGADNYDINPGLSRGN 327
MYB.PH3   INTINYPSGQSSMPVTASSSENTGSDF---GSSDNANNYANTNGLSRGN 329

          .  |    ...|.  ||..  ...  |....
MYB.NT1   SGLLEDLLEESQTLTRAE-KIEENC---PIENEAG-KGK-LVWEEYGLS 366
MYB.NT2   SGLLEDLLEESQALTRVE-KIEENC---PIENEAI-KGK-LVWEEYGLS 370
MYB.PH3   SGLLEDLLEESQTLNRPGMKIEDNFLDLKEDQEADYKGKSMLWEDYGLV 378

          .|.|  |.. .|...  |....||.....|.|..|.|  .  TTK
MYB.NT1   EEAEDIILTEESTFSFAQQGGEDATPIKHSEDSTSL---NSSSGI-TTK 411
MYB.NT2   EEAEDIILTEESTFSFAQEGGDNATPNKHSEDSSSL---NSSSGI-TTK 415
MYB.PH3   EDAEEAILTEESAYSFAH-GVDHVAQNKNSESSSPHSPPNSSSGIFMKK 426

          .|.||.|...|...|....||....|.|..|.|..|  |||.|
MYB.NT1   EGSLELANQVDEDIMRFLDNFPVGVPVPDWCNDENDQQNTS--NGQSFE 458
MYB.NT2   EGSLELANQVDEDIMRFLDNFPVGVPVPDWCNDENDEQNTS--NGQSFE 462
MYB.PH3   EDSFHGTNQADDDIMCLLDNFPLAVPVPEWYEDEDDKNNCNGQSSNVTN 475

          ..|.  |                             .    |
MYB.NT1   CDQIQSH--------------------------CSKSG    470
MYB.NT2   CDQIQSH--------------------------CSKSG    474
MYB.PH3   CDHIAENQAEDSKSPALTLNSGTRNHDWEFGGCCWNNMPSFC 517
```

Figure 4.10

The *Myb.Nt1* and *Myb.Nt2* sequences are more closely related to the *Myb.Ph3* sequence.

The table illustrates the percentage identity at the nucleotide level between the *N. tabacum Myb.Nt1* cDNA sequence and *Myb.Nt2*, *Myb. Ph3* (Avila *et al.*, 1993) from *P. hybrida*, *Myb.ZmC1* (Paz-Ares *et al.*, 1987) from maize, *Myb.HsC* (Majello *et al.*, 1986) from humans and *Myb.Reb1* from yeast. Only the regions common to all sequences are compared.

MYB.NT1

	R2	R3	C-terminus	Overall
MYB.NT2	98	98	84	88
MYB.PH3	94	92	36	48
MYB.ZMC1	60	54	13	27
MYB.HSC	38	51	9	17
MYB.REB1	20	—	10	10

Figure 4.11

Sequence alignment of the deduced MYB.NT1 MYB-domain with selected MYB domains.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

MYB proteins compared are the MYB.ZMC1 (Paz-Ares *et al.*, 1987), MYB.ZMP1 (Grotewold *et al.*, 1991), MYB.ZM1 and MYB.ZM38 form maize, MYB.HV1 from barley (Marocco *et al.*, 1989), MYB.AM305, MYB.AM306, MYB.AM315, MYB.AM330 and MYB.AM340 from *A. majus* (Jackson *et al.*, 1991), MYB.GL1 from *A. thaliana* (Oppenheimer *et al.*, 1991), MYB.ATHMYB (ATHMYB1) from *A. thaliana* (Shinozaki *et al.*, 1992), MYB.HSC (Majello *et al.*, 1986), MYB-HSA and MYB.HSB from humans (Nomura *et al.*, 1988), MYB.DM from *D. melanogaster* (Katzen *et al.*, 1985), MYB.PH2 and MYB.PH3 from *P. hybrida* (Avila *et al.*, 1993) and MYB.REB1 (REB1).

Numbers indicate amino acid residue positions of the MYB.NT1 and MYB.NT2 proteins. Upper case in the consensus of plant MYB domains (MYB.PLANT) indicates identical residues. Lower case indicates conservation in 80 % or more of the plant MYB proteins compared. Other symbols are: +, basic amino-acid, - acidic amino-acid, #, hydrophobic amino-acid. The three tryptophan residues present in each repeat of animal MYB proteins are labelled with asterisks. The proposed three α-helices in each repeat are shown in the lower part of the figure.

```
Repeat II
                   .                           .                          .       72
MYB.HSC      LIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLK-GRIGKQCRERWHNHLNPE
MYB.HSA      LIKGPWTKEEDQRVIELVQKYGPKRWSLIAKHLK-GRIGKQCRERWHNHLNPE
MYB.HSB      LVKGPWTKEEDQKVIELVKKYGTKQWTLIAKHLK-GRLGKQCRERWHNHLNPE
MYB.DM       LIKGPWTRDEDDMVIKLVRNFGPKKWTLIARYLN-GRIGKQCRERWHNHLNPN
MYB.ZMC1     VKRGAWTSKEDDALAAYVKAHGEGKWREVPQKAGLRRCGKSCRLRWLNYLRPN
MYB.ZMP1     LKRGRWTAEEDQLLANYIAEHGEGSWRSLPKNAGLLRCGKSCRLRWINYLRAD
MYB.ZM1      LNRGSWTPQEDMRLIAYIQKHGHTNWRALPKQAGLLRCGKSCRLRWINYLRPD
MYB.ZM38     TNRGAWTKEEDERLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINYLRPD
MYB.HV1      TNKGAWTKEEDDRLTAYIKAHGEGCWRSLPKAAGLLRCGKSCRLRWINYLRPD
MYB.ATGL1    YKKGLWTVEEDNILMDYVLNHGTGQWNRIVRKTGLKRCGKSCRLRWMNYLSPN
MYB.AM305    VRKGPWTMEEDLILINYIANHGEGVWNSLARSAGLKRTGKSCRLRWLNYLRPD
MYB.AM308    TNKGAWTKEEDDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINYLRPD
MYB.AM306    VKKGPWTPEEDIILVSYIQEHGPGNWRAIPSNTGLLRCSKSCRLRWTNYLRPG
MYB.AM315    LKRGPWTEEEDQKLTSYVLKNGIQGWRVIPKLAGLSRCGKSCRLRWMNYLRPD
MYB.AM330    TNKGAWTKEEDQRLINYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINYLRPD
MYB.AM340    VRKGPWTMEEDLILINFISNHGEGVWNTIARSAGLKRTGKSCRLRWLNYLRPD
MYB.ATHMYB   VRKGPWTEEEDAILVNFVSIHGDARWNHIARSSGLKRTGKSCRLRWLNYLRPD
MYB.NT1      LKKGPWTAAEDAILMEYVKKNGEGNWNAVQRNSGLMRCGKSCRLRWANHLRPN
MYB.NT2      LKKGPWTATEDAILMEYVKKNGEGNWNAVQRNSGLMRCGKSCRLRWANHLRPN
MYB.PH3      LKKGPWTAAEDSILMEYVKKHGEGNWNAVQRNSGLMRCGKSCRLRWANHLRPN
MYB.PH2      LKKGPWTPEEDQILVSYIEKNGHGNWRALPKLAGLLRCGKSCRLRWTNYLRPD
MYB.REB1     EQRGKWTAEEEQELAKLCAEK-EGQWAEIGKT--LGRMPEDCRDRWRNYVKCG

MYB.PLANT    ..+G*WT.eED..L*.Y*..hG.G.W..**+.aGL.RcgKSCRLRW*NyLrp.

Repeat III
                   .                           .                          .      123
MYB.HSC      VKKTSWTEEEDRIIYQAHKRLG NRWAEIAKLLP GRTDNAIKNHWNSTMRRK
MYB.HSA      VKKSSWTEEEDRIIYEAHKRLG NRWAEIAKLLP GRTDNSIKNHWNSTMRRK
MYB.HsB      VKKSCWTEEEDRIICEAHKVLG NRWAEIAKMLP GRTDNAVKNHWNSTIKRK
MYB.Dm       IKKTAWTEKEDEIIYQAHLELG NQWAKIAKRLP GRTDNAIKNHWNSTMRRK
MYB.ZmC1     IRRGNISYDEEDLIIRLHRLLG NRWSLIAGRLP GRTDNEIKNYWNSTLGRR
MYB.ZmP1     VKRGNISKEEEDIIIKLHATLG NRWSLIASHLP GRTDNEIKNYWNSHLSRQ
MYB.Zm1      LKRGNFTDEEEEAIIRLHGLLG NKWSKIAACLP GRTDNEIKNVWNTHLKKK
MYB.Zm38     LKRGNFTADEDDLIVKLHSLLG NKWSLIAARLP GRTDNEIKNYWNTHVRRK
MYB.Hv1      LKRGNFSHEEDELIIKLHSLLG NKWSLIAGRLP GRTDNEIKNYWNTHIRRK
MYB.Gl1      VNKGNFTEQEEDLIIRLHKLLG NRWSLIAKRVP GRTDNQVKNYWNTHLSKK
MYB.Am305    VRRGNITPEEQLLIMELHAKWG NRWSKIAKTLP GRTDNEIKNYWRTRIQKH
MYB.Am308    LKRGNFTEEEDELIIKLHSLLG NKWSLIAGRLP GRTDNEIKNYWNTHIRRK
MYB.Am306    IKRGDFTEHEEKMIIHLQALLG NRWAAIASYLP HRTDNDIKNYWNTHLKKK
MYB.Am315    LKKGPLTEMEENQIIELHAHLG NRWSKIALHIP GRTDNEIKNYWNTHIKKK
MYB.Am330    LKRGNFTEEEDEIIIKLHSLLG NKWSLIAGALP GRTDNEIKNYWNTHIKRK
MYB.Am340    VRRGNITPEEQLLIMELHAKWG NRWSKIAKHLP GRTDNEIKNYWRTRIQKH
MYB.Athmyb1  LIRNSFTEVEDQAIIAAHAIHG NKWAVIAKLLP GRTDNAIKNHWNSALRRR
MYB.Athmyb2  VRRGNITLEEQFMILKLHSLWG NRWSKIAQYLP GRTDNEIKNYWRTRVQKQ
MYB.NT1      LKKGAFSLEEERFIVELHAKLG NKWARMAAQMP GRTDNEIKNYWNTRLKRR
MYB.NT2      LKKGAFSLEEERIIVELHAKLG NKWARMAAQMP GRTDNEIKNYWNTRLKRR
MYB.Ph3      LKKGAFTVEEERIIIELHAKLG NKWARMAAQLP GRTDNEIKNYWNTRLKRR
MYB.Ph2      IKRGNFTREEEDTIIQLHEMLG NRWSAIAARLP GRTDNEIKNVWHTHLKKR

MYB.PLANT    *+rG.*t..E-.*Ii.Lh..lG N+Ws.iA..lP gRTDNeiKNyWnt+*.++
                                                    turn  recognition helix
```
```
           ┌───────────┐   ┌───────────┐   ┌───────────┐
           │  helix 1  │───│  helix 2  │───│  helix 3  │
           └───────────┘   └───────────┘   └───────────┘
```
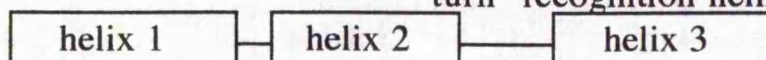
Figure 4.12

Dendrogram derived from percentage identity between peptide sequences of selected MYB-domains.

Performed using default parameters on the PILEUP program (Devereux *et al.*, 1984).

This comparison uses whole peptide sequences. References are given in the legend for Figure 4.11, except MYBST-1 from potato (Baranowski *et al.*, 1994) and ATHMYB2 from *A. thaliana* (Urao *et al.*, 1993). NTM-1 and NTM-2 represent MYB.NT1 and MYB.NT2 respectively.

PILEUP of: *.MYB   November 18, 1995 13:01

NTM-1
NTM-2
PH3
AM308
ZM38
HV-1
AM330
ZM1
ZMC1
ZMP1
AM305
AM340
ATHMYB2
GL-1
AM315
PH2
AM306
HSA-MYB
HSC-MYB
HSB-MYB
DM
ATHMYB1
REB-1
MYBST-1
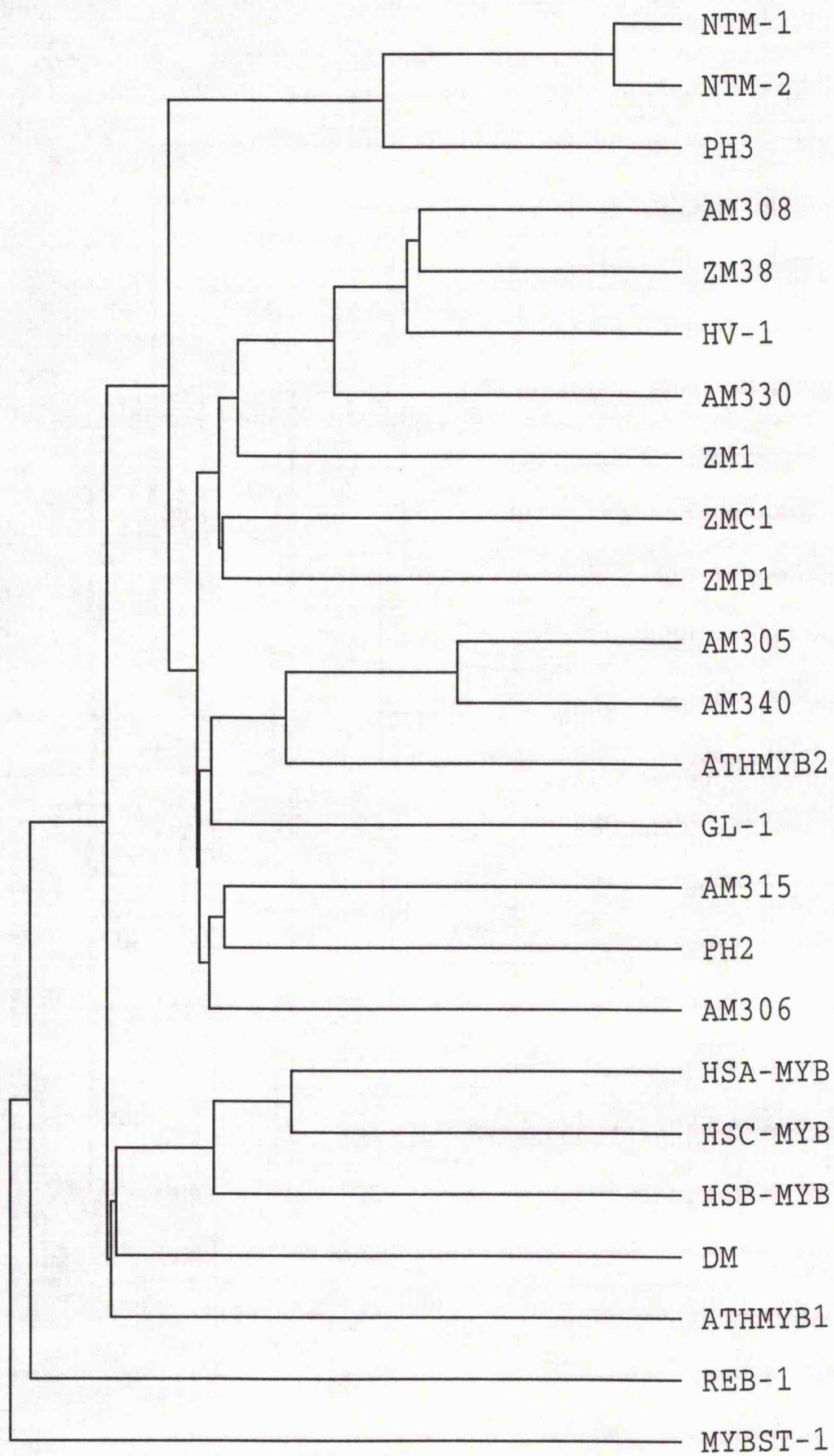
of the residues adjacent to the tryptophan residues. These basic residues form the surface of the recognition helix which contacts the DNA double helix.

An idea of the percentage identity among MYB protein sequences was obtained by constructing a dendrogram, using whole amino acid sequences (Figure 4.12). The sequence similarity of the plant MYB proteins, distinct from their animal counterparts can be seen in the grouping of the proteins. An exception is the MYB.ST1 protein from potato (Baranowski et al., 1994), which has a distinct sequence, possessing a novel single repeat MYB-domain, but retaining the highly conserved amino acids such as the tryptophan residues. The MYB.NT proteins form a tightly related group with the MYB.PH3 protein. The MYB.NT proteins can be seen to have had a relatively recent sequence divergence from each other, reflecting the high sequence similarity between the two proteins.

4.2.3.2 NTL1 and NTL2

The NTL1 and NTL2 predicted proteins have a basic character, with estimated pI values of 8.52 and 8.66 respectively. The lack of any significant hydrophobic signal peptide sequence can be seen in the predictions of Kyte and Doolittle (Kyte and Doolittle, 1982) (Figure 4.13) and Chou and Fasman (Chou and Fasman, 1978) (Figure 4.14). Hence the N. tabacum LIM proteins can be predicted to be intracellular proteins. Three putative N-glycosylation sites (NXS/T) on each of the proteins at positions 93, 125 and 172 are also noteworthy.

A self comparison of either of the NTL proteins using a matrix analysis program shows two regions of perfectly aligned homology (Figure 4.15). These regions span approximately 70 residues and are shifted by 100 residues with respect to one another. When the predicted amino acid sequences of the NTL1 and NTL2 are aligned (Figure 4.16), a 97 % sequence homology is seen over the length of the proteins. Further alignment of the N. tabacum LIM proteins with the sunflower SF3 protein (Figure 4.17a.), reveals a high homology over the length of the proteins with the notable exception that the acidic carboxy-terminal pentapeptide repeat present in SF3 is absent in the NTL proteins. Two potential zinc fingers per LIM domain can be defined in the N. tabacum proteins, spanning two 52 and 53 amino acid regions from 10 to 63 and 110 to 162 respectively. These regions have a characteristic pattern of cysteine residues. In addition a region of basic amino acid residues is present at the end of the second zinc finger of each LIM domain at positions 49 to 59 and 159 to 170 respectively. Also noteworthy are three potential protein kinase C phosphorylation

107

sites at positions 7, 43 and 81 and spaced by 33 amino acids which are conserved in all three proteins. Serine/threonine residues are present in all three proteins which are also potential sites of phosphorylation.

A comparison of the amino acid sequences between regions of the NTL1, NTL2 and SF3 proteins (Figure 4.17b.), shows that the highest homology occurs within the LIM domain regions. An alignment of the LIM repeats of the NTL proteins and SF3 with repeats from selected members of the animal LIM domain containing family (Figure 4.18), highlights the conservation of the regularly spaced cysteine and histidine residues in each of the repeats. These residues have been shown to be important in chelating zinc ions in the formation of fingers (Michelson *et al.*, 1993). The most frequently occurring metal chelating residues are cysteines and histidines, however aspartate is used as the last chelating residue in the second finger in the majority of animal LIM proteins. The plant LIM repeats also differ from their animal counterparts in the spacing of the second cysteine residue of the second zinc finger of the second LIM domain, from two to four residues.

A dendrogram was constructed using whole amino acid sequences of the proteins used in the LIM domain alignment, in order to gain an understanding of the percentage identity between the sequences (Figure 4.19). The three LIM protein classification groups can be seen clearly as clusters of related proteins, with the NTL and SF3 proteins forming a tightly related sub-group to muscle LIM protein from rat (Arber *et al.*, 1994) and cysteine rich protein from humans (Liebhaber *et al.*, 1990). These five proteins are classified as group 2 proteins based on their LIM domain sequence similarity (Taira *et al.*, 1995)), and on the basis of them possessing one or two LIM domains spaced 47 to 68 amino acid residues apart as opposed to the average distance in other LIM domain proteins of 10 residues.

4.3 Discussion
4.3.1 *Myb.Nt1* and *myb.Nt2*: pollen-expressed cDNAs encoding putative transcription factors

The deduced amino acid sequences of *myb.Nt1* and *myb.Nt2* isolated from the *N. tabacum* pollen cDNA library encode typical plant MYB proteins in that the MYB domain consists of two repeats (R2 and R3) corresponding to the second and third repeats of the MYB domain found in animal proteins (R1, R2 and R3; Figure 4.11). Regularly spaced tryptophan residues conserved in the MYB domains of plant and animal proteins are also present in the MYB.NT proteins. A characteristic exception of

Figure 4.13

Hydrophobicity plot of the deduced NTL1 and NTL2 peptides using the Kyte and Doolittle prediction.

Hydropathy plot according to Kyte and Doolittle (Kyte and Doolittle, 1982); window size: 9 residues. Bold arrows indicate LIM domains.

a. NTL1.
b. NTL2.

a.

b.
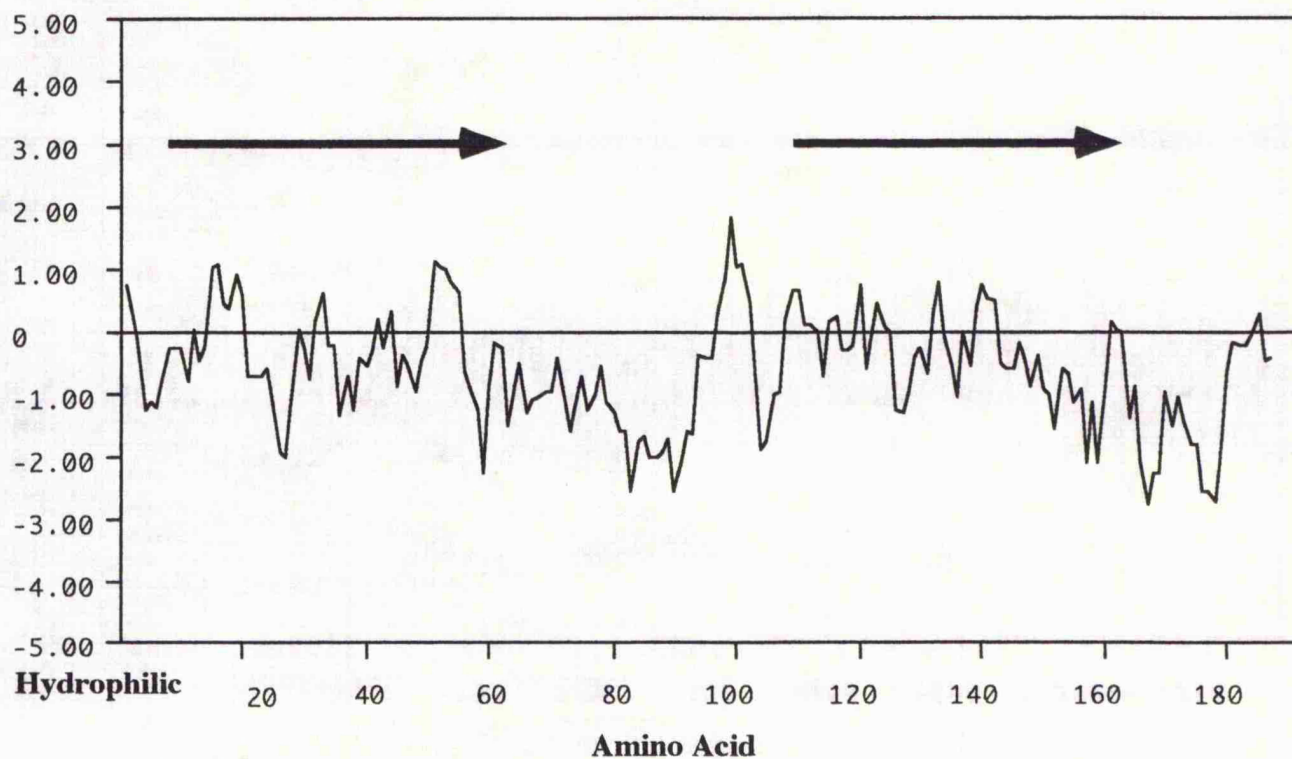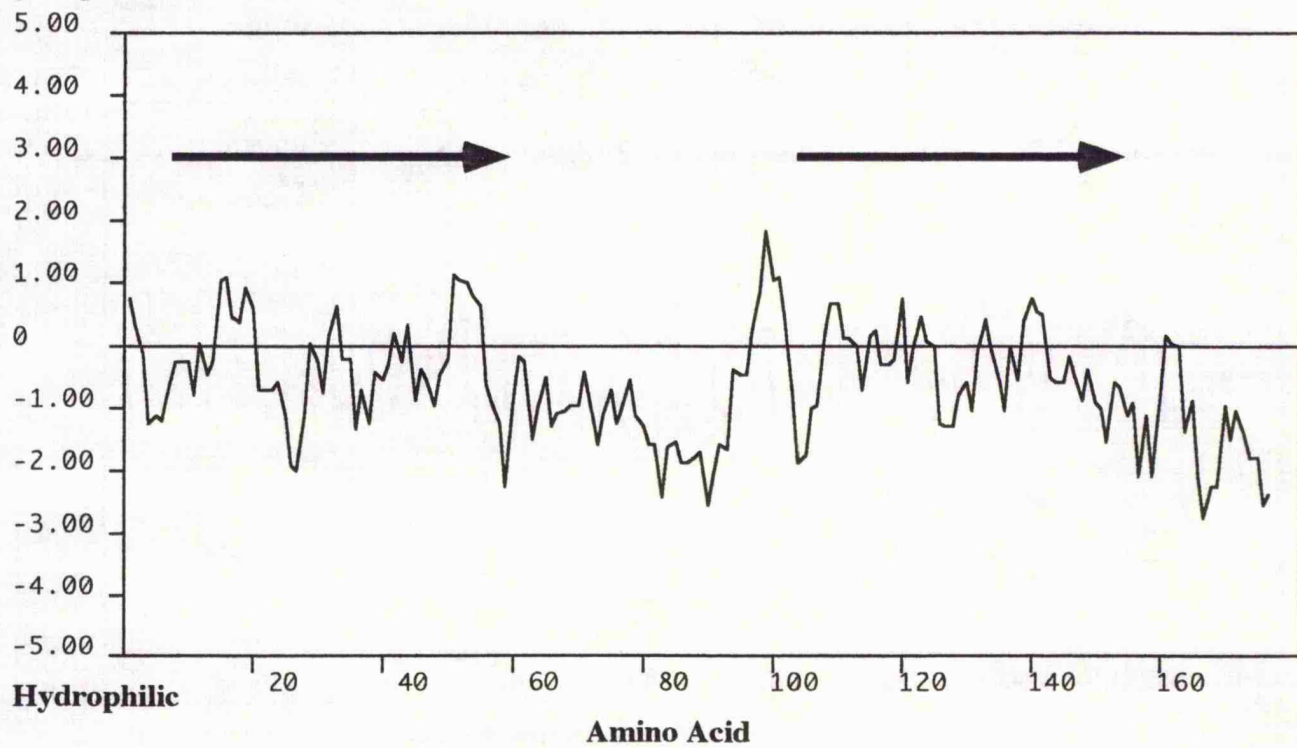
Figure 4.14

Secondary structure illustration of the deduced NTL1 peptide using the Chou-Fasman prediction.

Performed using the PLOTSTRUCTURE program from the GCG suite of programs (Devereux *et al.*, 1984).

Putative glycosylation sites are indicated with a 'lollipop'. Hydrophilic and hydrophobic regions are represented with oval and diamond symbols, respectively.

PLOTSTRUCTURE of: NTL1 ck: 2252

Chou-Fasman Prediction
November 25, 1995 14:20

NTL1 translated from 116 to 691

◯ KD Hydrophilicity >=1.3
◇ KD Hydrophobicity >=1.3

NH2

50

100

150

COOH
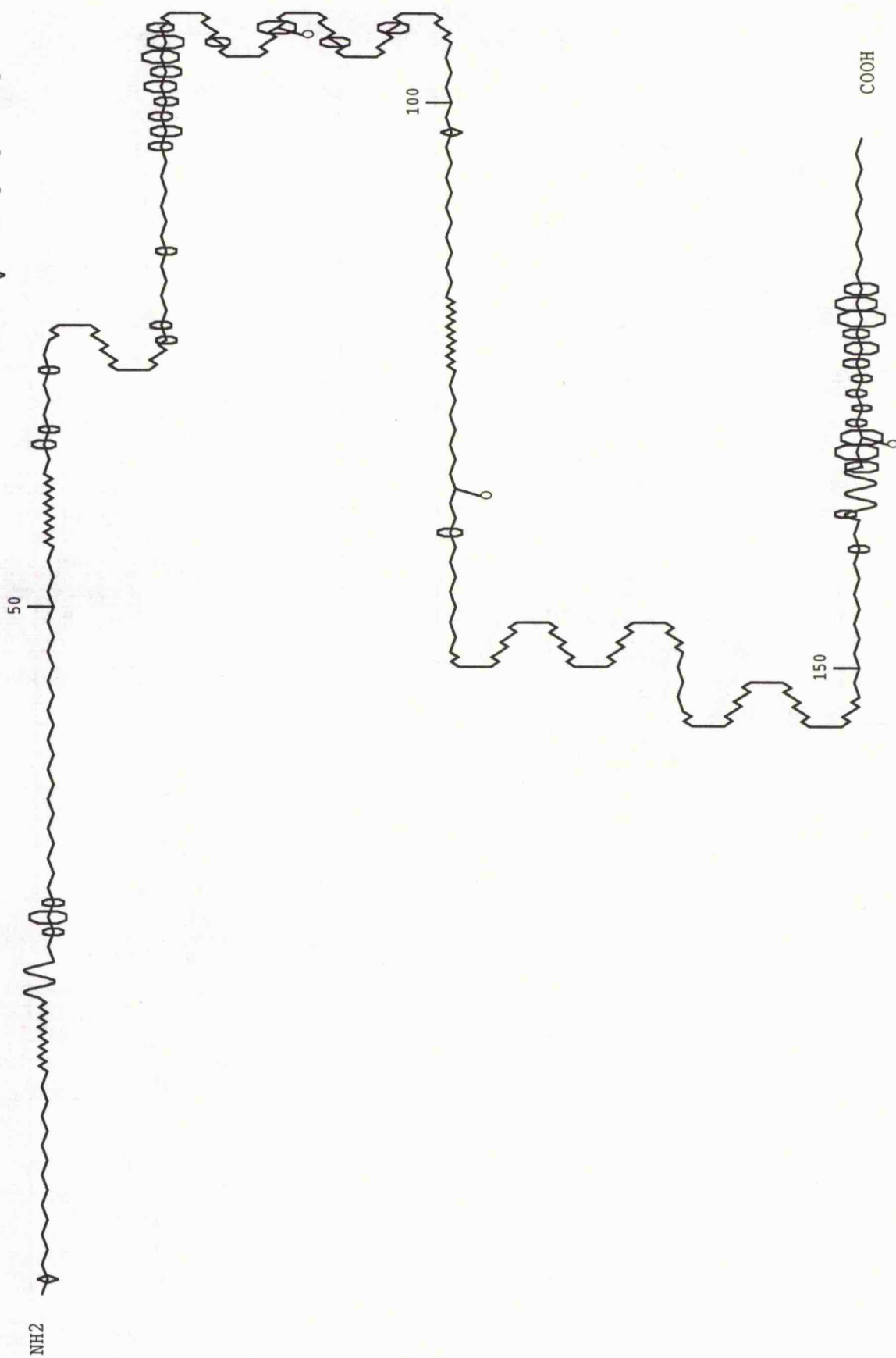
Figure 4.15
Self-comparison of the NTL1 and NTL2 peptides.

Performed using a matrix analysis program (GeneJockey™) using the following parameters: window size: 5; minimum score: 2.

a. NTL1.
b. NTL2.

a.
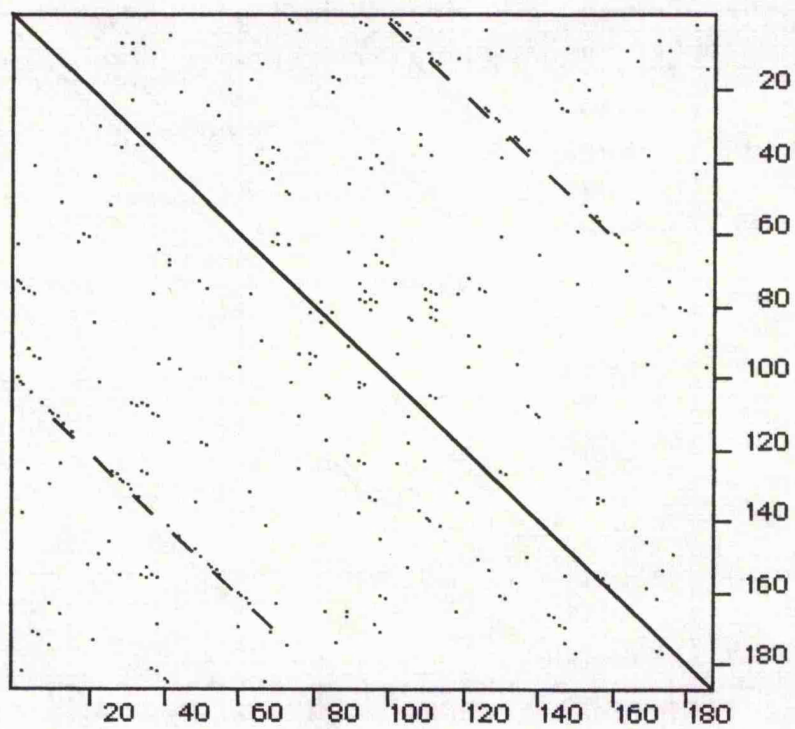
b.
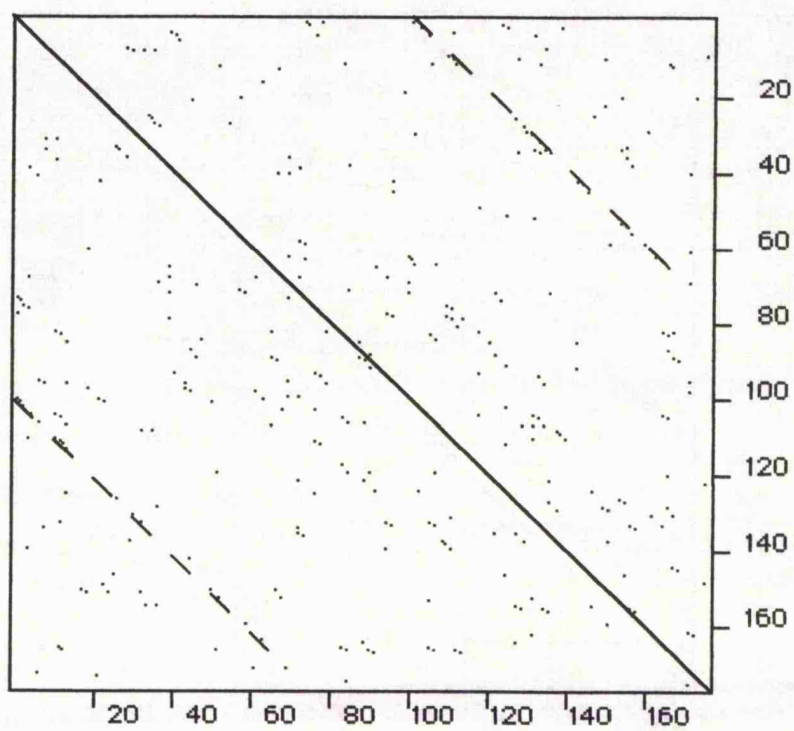
Figure 4.16

Sequence alignment of the deduced NTL1 and NTL2 proteins.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

```
MTFAGTTQKCSACEKTVYLVDRLAADNRIYHKACFRCYHCKSTLKLSNFNSFEGVIY 57
.........................................................
MTFAGTTQKCSACEKTVYLVDRLAADNRIYHKACFRCYHCKSTLKLSNFNSFEGVIY 57


CRPHFDQLFKRTGSLDKSFEGTPKVTKPEKPVDNENGSGTKVSSLFAGTREKCVGCT 114
.................. . ...... ................
CRPHFDQLFKRTGSLDKSFDGTPKVTKPEKSVENENGSGSKVSSLFAGTREKCVGCT 114


KTVYPIEKVSVNGTAYHKACFKCSHGGCTISPSNYIAHEGRLYCKHHHIQLFKEKGN 171
.................. ......................................
KTVYPIEKVSVNGTAYHKGCFKCSHGGCTISPSNYIAHEGRLYCKHHHIQLFKEKGN 171


YSQLESDHETDPALSTQSLT 191
..........
YSQLESDHET 181
```

Figure 4.17

Sequence alignment of the deduced NTL1, NTL2 and SF3 proteins.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

'o' signifies an identical amino acid;

'l' signifies a conservative amino acid substitution;

'.' mark every tenth amino acid residue.

a. The LIM domains are boxed. The penta-peptide repeat of SF3 is underlined '____'. Putative protein kinase C phosphorylation sites are underlined '▬▬'. Putative N-glycosylation sites are boxed and shaded. Basic LIM domain regions are marked under each residue by 'x'.

b. The Table illustrating the percentage identity at the nucleotide level between the *N. tabacum Ntl1* cDNA and the *Ntl2 and Sf3* cDNA clones (Baltz *et al.*, 1992a.).

a.

```
               .  | . | .    .    . . . . . . . . . . | .  . . | | . | . . . . . . .   . .  | . . . . . . . 
NTL1    M-TFAGTTQK|CSACEKTVYLVDRLAADNRIYHKACFRCYHCKSTLKLSN  48
NTL2    M-TFAGTTQK|CSACEKTVYLVDRLAADNRIYHKACFRCYHCKSTLKLSN  48
SF3     MKSFTGTTQK|CTVCEKTVYLVDKLVANQRVYHKACFRCHHCNSTLKLSN  49
                                                   ――              ――――

          . . . . . . . . . . . . . . . | . . | . . . . | . . .    .   .     |   . |    . . |
NTL1    FNSFEGVIYCRPHF|DQLFKRTGSLDKSFEGTPKVTKPEKSVDNE|NGS|GT  97
NTL2    FNSFEGVIYCRPHF|DQLFKRTGSLDKSFDGTPKVTKPEKPVENE|NGS|GS  97
SF3     FNSFDGVVYCRHHF|DQLFKRTGSLEKSFDGTPKF-KPERTFSQETQSAN  97
                       xxxxxxxxxx                ―――

          | | . . | .  . . . | . .     .    |   .      .   | . . . .    .  | |          .
NTL1    KVSSLFAGTREK|CVGCTKTVYPIEKVSV|NGT|AYHKACFKCSHGGCTISP  146
NTL2    KVSSLFAGTREK|CVGCTKTVYPIEKVSV|NGT|AYHKGCFKCSHGGCTISP  146
SF3     RLSSFFEGTRDK|CNACAKIVYPIERVKVDGTAYHRACFKCCHGGCTISP  146

          . . . . . . . . . . . . . . . . . | . . . . . . . .   | |          .
NTL1    SNYIAHEGRLYCKHHHI|QLFKEKG|NYS|QLESDH-----------------  179
NTL2    SNYIAHEGRLYCKHHHI|QLFKEKG|NYS|QLESDH-----------------  179
SF3     SNYIAHEGRLYCKHHHI|QLFKKKG|NYS|QLEVEETVAAPAESETQNTETQ  195
                         xxxxxxxxxx

          .                | . | | | | | | |
NTL1    --------------ETDPALSTQSLT  191
NTL2    --------------ET          181
SF3     NAETQNADTQNADTQNTETQNGSV  219
```

b.

NTL-1

LIM1   Inter-lim   LIM2

	N-terminus	LIM1	Inter-lim	LIM2	C-terminus	Overall
NTL-2	100	100	91	98	100	97
SF3	88	77	52	81	63	70
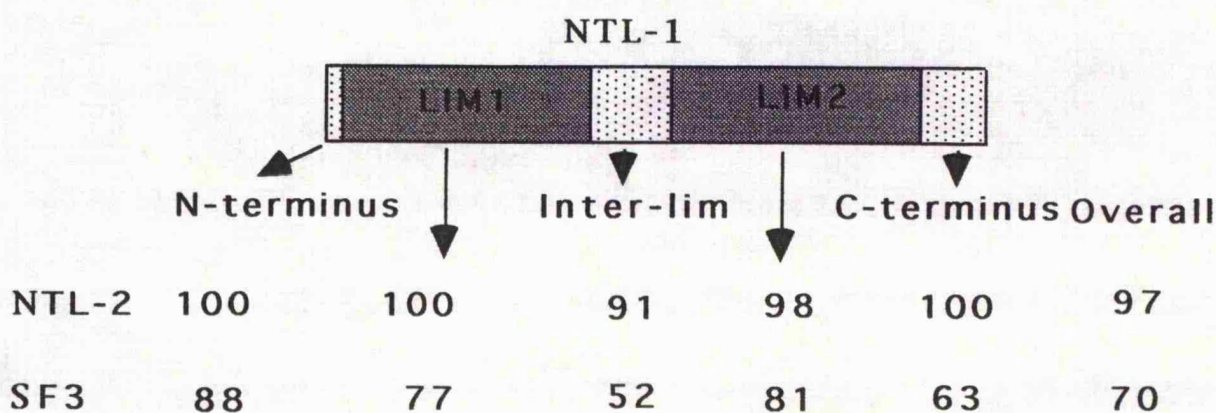
Figure 4.18

Sequence alignment of the deduced NTL1 LIM-domains with selected LIM domains.

Performed using the CLUSTAL programme (Higgins and Sharp, 1989) with default parameters.

LIM proteins are, APTEROUS from *D. melanogaster* ((Bourgouin *et al.*, 1992) and (Cohen *et al.*, 1992)), TTG-1 from humans (McGuire *et al.*, 1989), TTG-2 from humans (Royer-Pakora *et al.*, 1991), ISL-1 from rat (Karlsson *et al.*, 1990), XLIM-1 from *Xenopus leavis* (Taira *et al.*, 1992), MEC-3 from *Caenorhabitis elegans* (Xue *et al.*, 1992), LIMK from humans (Mizuno *et al.*, 1994), SF3 from sunflower (Baltz *et al.*, 1992a.), hCRP from humans (Liebhaber *et al.*, 1990), MLP from *D.melanogaster* (Arber *et al.*, 1994), RIL from rat (Kiess *et al.*, 1995), ENIGMA from humans (Wu and Gill, 1994), PAXILLIN from humans (Turner and Millar, 1994), PINCH from humans (Rearden, 1994) and ZYXIN from chicken (Sadler *et al.*, 1992).

The consensus LIM sequence is shown for each domain. Conserved zinc-binding residues are marked '·' and residues with greater than 80 % conservation in the LIM proteins compared are marked 'l'. An 'n' in the second LIM domain consensus motif denotes a 2 or 4 residue spacing.

LIM 1 Domains

```
Group1               .   .                .  |  .   .                       .  |
APTEROUS     CSGCGRQIQDRFYLS-AVEKRWHASCLQCYACRQPLER-ESSCYSRDGNIYCKNDY
TTG-1        CAGCNRKIKDRYLLKA-LDKYWHEDCLKCACCDCRLGEVGSTLYTKANLILCRRDY
TTG-2        CGGCQQNIGDRYFLKA-IDQYWHEDCLSCDLCGCRLGEVGRRLYYKLGRKLCRRDY
ISL-1        CVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDE-SCTCFVRDGKTYCKRDY
LIN-11       CAACAQPILDRYVFTV-LGKCWHQSCLRC--CDCR-APMSMTCFSRDGLILCKTDF
XLIM-1       CAGCERPILDRFLLNV-LDRAWHVKCVQC--CECK-CNLTEKCFSREGKLYCKNDF
MEC-3        CNCCNEQIYDRYIYRMD-NRSYHENCVKCTICESPLAE---KCFWKNGRIYCSQHY
LIMK         CASCGQRIYDGQYLQQALNADWHADCFRC--CDCS-ASLSHQYYEKDGQLFCKKDY


Group 2
NTL-1        CSACEKTVYLVDR-LAADNRIYHKACFRCYHCKSTLKL--SNFNSFEGVIYCRPHF
NTL-2        CSACEKTVYLVDR-LAADNRIYHKACFRCYHCKSTLKL--SNFNSFEGVIYCRPHF
SF3          CTVCEKTVYLVDK-LVANQRVYHKACFRCHHCNSTLKL--SNFNSFDGVVYCRHHF
hCRP         CGVCQKTVYFAEE-VQCEGNSFHKSCFLCMVCKKNLDS--TTVAVHGEEIYCKSCY
MLP          CGACDKTVYHAEE-IQCNGRSFHKTCFHCMACRKALDS--TTVAAHESEIYCKVCY


Group 3
RIL          CTRCGHGIV--GTIVKARDKLYHPECFMCSDCGLNLKQ--RGYFFLDERLYCENHA
ENIGMA       CHQCHKVIR--GRYLVALGHAYHPEEFVCSQCGKVLEE--GGFFEEKGAIFCPPCY
PAXILLIN     CGACKKPIA--GQVVTAMGKTWHPEHFVCTHCQEEIGS--RNFFERDGQPYCEKDY
PINCH        CERCKGGFAPAEKIVNSNGELYHEQCFVCAQCFQQFPE--GLFYEFEGRKYCEHDF
ZYXIN        CGFCRKPLSRTQPAVRALDCLFHVECFTCFKCEKQLQG--QQFYNVDEKPFCEDCY


                                                                       D
LIM motif    CX2C        X16-18        HX2CX2CX2C        X16-19    CX2H
                                                                       C
```

LIM 2 Domains

```
Group 1              .   .                .  |  .  |                          .  |
APTEROUS     CSRCLASISSNELVMRARN-LVFHVNCFCCTVCHTPLTKGDQYGIIDALIY--CRTHY
TTG-1        CAACSKLIPAFEMVMRARD-NVYHLDCFACQLCNQRFCVGDKFFLKNNMI-L-CQMDY
TTG-2        CASCDKRIRAYEMTMRVKD-KVYHLECFKCAACQKHFCVGDRYLLINSDI-V-CEQDI
ISL-1        CAKCSIGFSKNDFVMRARS-KVYHIECFRCVACSRQLIPGDEFALREDGLF--CRADH
LIN-11       CAGCDGKLEKEDLVRRARD-KVFHIRCFQCSVCQRLLDTGDQLYIMEGNRFV-CQSDF
XLIM-1       CAGCAQGISPSDLVRRARS-KVFHLNCFTCMMCNKQLSTGEELYIIDENKFV-CKEDY
MEC-3        CAGCKKGVSPTDMVYKLKAGLVFHVECHCCSLCGRHLSPGEQILVDDTMMTVSCMSHY
LIMK         CHGCSEQITK-GLVMVAGE-LKYHPECFICLTCGTFIGDGDTYTLVEHSKLY-CGHCY


Group 2
NTL1         CVGCTKTVYPIEKVSV--NGTAYHKGCFKCSHGGCTI-SPSNYIAHEGR-LY-CKHHH
NTL2         CVGCTKTVYPIEKVSV--NGTAYHKACFKCSHGGCTI-SPSNYIAHEGR-LY-CKHHH
SF3          CNACAKIVYPIERVKV--DGTAYHRACFKCCHGGCTI-SPSNYIAHEGR-LY-CKHHH
hCRP         CPRCSQAVYAAEKVIG--AGKSWHKACFRCAKCGKGL-ESTTLADKDGE-IY-CKGCY
MLP          CPRCGKSVYAAEKVMG--GGKPWHKTCFPCAICGKSL-ESTNVTDKDGE-LY-CKVCY


Group 3
ENIGMA       CAKCKKKI-TGEIMHA--LKMTWHVHCFTCAACKTPI-RNRAFYMEEGV-PY-CERDY
PAXILLIN     CYYCNGPI-LDKVVTA--LDRTWHPEHFFCAQCGVFF-GPEGFHEKDGK-AY-CRKDY
PINCH        CHQCGEFI-IGRVIKA--MNNSWHPECFRCDLCQEVL-ADIGFVKNAGRH-L-CRPCH
ZYXIN        CSVCKQTI-TDRMLKA--TGNSYHPQCFTCVMCHTPL-EGASFIVDQANQPH-CVDDY


                                                                       D
LIM motif    CX2C        X16-19        HX2CX2CXnC        X15-20    CX2H
                                                                       C
```

LIM 3 Domains

```
Group 3      ·   ·                           ·  |    ·   ·    |                    ·   |
ENIGMA     CHGCDFKID-----AGDRFLEALGFSWHDTCFVCAICQINL----EGKTFYSKKDRPLCKSHA
PAXILLIN   CGGCARAIL-------ENYISALNTLWHPECFVCRECFTPF----INGSFFEHDGQPYCEVHY
PINCH      CQKCH-AII-----DEQPLIFKND-PYHPDHFNCANCGKEL----TADAR-ELKGELYCLPCH
ZYXIN      CSVCSEPIMPEPGKDETVRVVALEKNFHMKCYKCEDCGRPLSIEADENGCFPLDGHVLCMKCH
```

LIM motif	CX_2C	X_{23}	$HX_2CX_2CX_2C$	X_{21}	CX_2H C

LIM 4 Domains

```
Group 3          ·    ·                       ·    ·    ·    ·                    ·  |
PAXILLIN       CSGCQKPITGRCITAMGKKFHPEHFVCAFCLKQLNKGTFKEQNDKPYCQNCF
PINCH          CGACRRPIEGRVVNAMGKQWHVEHFVCAKCEKPFLGHRHYERKGLAYCETHY
```

LIM motif	CX_2C	X_{16}	$HX_2HX_2CX_2C$	X_{17}	CX_2H C

LIM 5 Domain

```
Group 3          ·    ·                       ·    ·    ·    ·                    ·    ·
PINCH          CFHCNRVIEGDVVSALNKAWCVNCFACSTCNTKLTLKNKFVEFDMKPVCKKCY
```

LIM motif	CX_2C	X_{16}	$CX_2CX_2CX_2C$	X_{17}	CX_2C

Figure 4.19

Dendrogram derived from percentage identity between peptide sequences of selected LIM proteins.

Performed using default parameters on the PILEUP program (Devereux *et al.*, 1984).

This comparison uses whole peptide sequences. References are given in the legend to Figure 4.18.

a. Branch representing divergence of group 2 type LIM proteins.
b. Branch representing divergence of group 3 type LIM proteins.
c. Branch representing divergence of group 1 type LIM proteins.

115

PILEUP of: *.LIM November 18, 1995 12:52
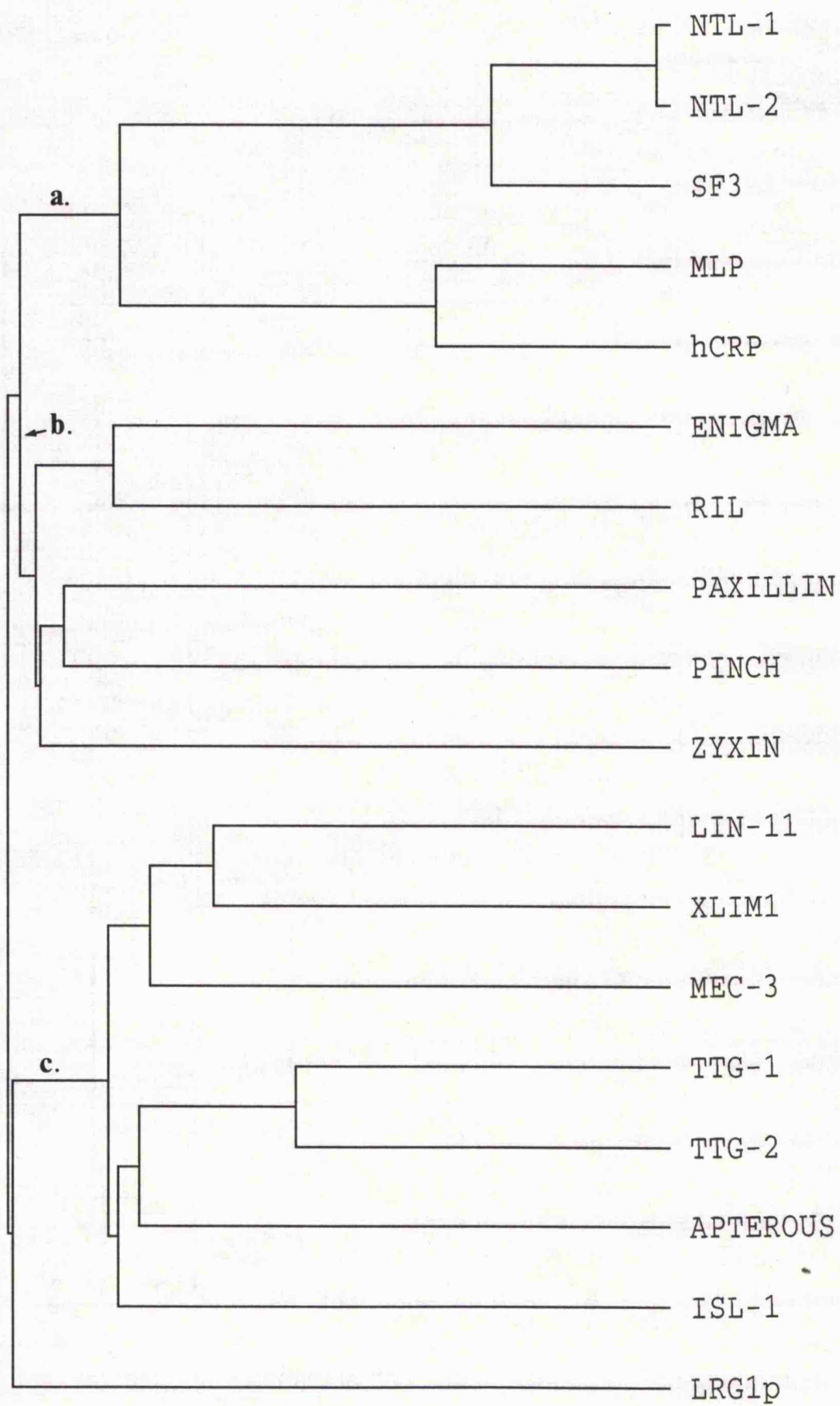
plant MYB domains and one that is seen in both the *N. tabacum* proteins is the conservative substitution of the first tryptophan in the second repeat with phenylalanine, another aromatic amino acid. Replacement of this residue with other aromatic amino acids did not appreciably affect DNA binding of c-MYB in gel retardation assays (Saikumar *et al.*, 1990).

Nuclear magnetic resonance imaging and molecular modelling of c-MYB (Ogata *et al.*, 1994) was used to show that the three tryptophan residues in R2 and R3 form hydrophobic cores around which are positioned three α-helices. The third helix in each of the R2 and R3 repeats forms a recognition helix. The two recognition helices from each repeat contact each other directly to bind to the consensus core sequence 'AACNG' cooperatively, neither repeat can bind to DNA alone. For this to be possible the R2 and R3 repeats are closely packed on the major groove of the DNA double helix. The absence of α-helical secondary structure means that the R1 repeat is not involved in the DNA binding process, although deletion of this repeat did result in a decrease in the stability of the protein-DNA interaction (Tanikawa *et al.*, 1993).

The core sequence of 'AAC' is recognised by an asparagine (N) and lysine (K) residue in the third MYB repeat (R3) and a lysine residue in the second repeat (R2). These residues are highly conserved in all animal and plant MYB repeats and are positioned at 114, 113 and 59 amino acid residues respectively in MYB.NT1 (Figure 4.11).

A highly conserved cysteine residue (C) in R2 of plant and animal MYB domains (position 61 in MYB.NT1) has been shown to have functional significance in that substitution for a serine (S) residue abolishes DNA binding (Grasser *et al.*, 1992). It was suggested that the redox state of this residue might induce a conformational change in R2 of the MYB domain, enabling DNA binding of MYB to its recognition site (Myrset *et al.*, 1993).

Other regions of possible structural significance in the MYB.NT proteins include acidic regions with predicted α-helical secondary structure. Such types of structures have been shown to function as transcriptional activation domains in c-MYB (Weston and Bishop, 1989) and a number of eukaryotic transcription factors.

It is likely, considering the presence of putative DNA binding and transcriptional activation domains that MYB.NT1 and MYB.NT2 exhibit sequence specific DNA binding and transcriptional activation properties. To this end the presence of conserved phosphorylation sites between the MYB.NT proteins and MYB.PH3 may indicate a mechanism of post-translational activation of these proteins by phosphorylation after DNA binding. Similar mechanisms have been suggested for *c-fos* and SP1 ((Lech *et*

116

al., 1988); (Jackson et al., 1990)). Such post-translational regulation enables a rapid response by transcription factors, to internal and external signals.

The 98 % amino acid identity of the derived sequence of the MYB.NT1 and MYB.NT2 MYB domains, or 100 % identity considering conservative substitutions, might mean the two proteins bind to the same target sequences and thus regulate the same target genes *in vivo*. The 92 % similarity of these sequences to the DNA-binding domain of MYB.PH3, 100 % considering conservative substitutions, again might suggest a similarity in the target genes regulated. All eight of the conservative substitutions made between the MYB domain of the three proteins (Figure 4.9) do not represent amino acid residues which are 100 % conserved in the MYB domains of plant MYB proteins and could be proposed as not likely to affect DNA binding. However, single amino acid changes in the DNA binding motifs of other regulatory proteins such as homeodomain proteins have been shown to completely alter their specificity for recognition sequences (Treisman *et al.*, 1989).

The MYB.PH3 protein, which was shown by Northern analysis to be expressed in flowers but not leaves, sepals or roots (Avila *et al.*, 1993), has been shown to recognise two distinct binding site sequences *in vitro* and induce similar DNA distortions on each site (Solano *et al.*, 1995b.). Such DNA distortions have been shown to be necessary to facilitate the protein-protein and DNA-protein interactions which are crucial for interaction with the RNA polymerase II complex in transcriptional activation. The first site, MSI (AAAAAAC(C/G)GTTA), resembles the consensus motif of c-MYB (AACNGTT) to which all known animal and two plant proteins, ATMYB1 and ATMYB2, from *Arabidopsis* have been shown to bind (Urao *et al.*, 1993a). The second site, MSII (AAAAGTTAGTTA) resembles the consensus binding site of the two other plant MYB proteins, AM305 from *Antirrinhum* (Sablowski *et al.*, 1994) and P from maize (Grotewold *et al.*, 1994) G(G/T)T(T/A)GGT(T/G) or GGT(T/A)GGT(T/G), which are closely related to ATMYB2 but recognise different sequences.

Further, MYB.PH3 could activate transcription of GUS in yeast from a promoter containing either of its two binding sites, and in plant protoplasts from the promoter of the *Petunia chsJ* gene (chalcone synthase), which contains two MYB.PH3 binding site sequences. Several other *chs* genes from different species also contain sequences related to the MSI and MSII binding sites. These include *chs* from *Antirrinhum* (ATTGGGTGGTTA; (Sommer and Saedler, 1986)), *chs* from *Hordem vulgare* (GGGAGTTAGTTG; (Rohde *et al.*, 1991)) and *chs1* from *Pisum sativum* (AAAAGTTAGTTA; (An *et al.*, 1993)). Two MYB proteins from *Aradidopsis*,

117

ATMYB6 and ATMYB7, were shown to bind preferentially to the maize P binding site but not c-MYB binding site *in vitro* (Li and Parish, 1995). The core motif in the potato MYBST1 binding site (GGATA) (Baranowski *et al.*, 1994) differs from the above sequences although there is some similarity to the AM305 and P binding sites. This may be a consequence of the unique MYBST1 DNA-binding domain which contains only one distantly related repeat (Figure 4.12).

A single conservative substitution of residue 117 in MYB.PH3 of aspartate (D) to glutamate (E) (position 99 in MYB.NT1), which is conserved in plant and animal MYB proteins, completely abolished binding to both the MSI and MSII binding sites in protoplasts and yeast. Also removal of the carboxy-terminal portion of the MYB.PH3 protein which contains the putative acidic activating domains greatly increased DNA binding. However the presence of possible negative regulatory domains in the MYB.PH3 protein may mean that the protein may repress transcription rather than activate under certain conditions. The presence of such a negative regulatory domain has been demonstrated in c-MYB (Sakura *et al.*, 1989).

These studies suggest that MYB.PH3 may be involved in the regulation of the synthesis of flavonoids and/or related compounds, which is a role proposed for several other plant MYB proteins (C1, Pl, P, Zm1, Zm38, Am305; section 1.4.4.1). Considering the high sequence similarity of the MYB domains of MYB.PH3 to MYB.NT1 and MYB.NT2, it is possible that a homologous function exists for the *N. tabacum* proteins. Anti-sense inhibition of the chalcone synthase gene, a highly expressed and key enzyme in flavonoid biosynthesis, in transgenic *P. hybrida* resulted in plants which were completely self-sterile, with germinating pollen showing rupturing of the emerging pollen tube tips due to a decrease in flavonols in the pollen and stile. Flavonols, a widely distributed and common class of flavonoids, are present at high levels in germinating pollen-tubes (Ylstra *et al.*, 1994). Addition of exogenous flavonol, complemented this phenotype. A vascular bundle-, tapetal- and pollen-expressed *BA42* cDNA clone encoding chalcone synthase, was isolated from *Brassica napus* (Shen and Hsu, 1992). *BA42* transcripts were present in microspores prior to PMI and a rapidly accumulated in pollen to maturity. In addition, the promoter of the *pall* gene in *Arabidopsis* was fused to the *gus* reporter gene and the activity analysed in transgenic plants (Ohl *et al.*, 1990). GUS activity was detectable prior to the accumulation of yellow pigment in spores, which occurs prior to PMI in *Arabidopsis thaliana* and accumulated rapidly to pollen maturity. *Pall* encodes a phenylalanine ammonium lyase enzyme which forms part of the hydroxycinnamic acid pathway which functions to supply 4-coumaroyl CoA precursors for flavonoid biosynthesis.

118

Another flavonoid biosynthetic enzyme which has been demonstrated to be expressed in pollen is the chalcone flavanone isomerase ((van Tunen *et al.*, 1989); (van Tunen *et al.*, 1990)). A *chiA* gene from *Petunia hybrida* is expressed from two independent promoters in corolla, anthers and pollen. A 440 bp region was shown to direct pollen specific gene expression of the *gus* reporter gene. MYB.NT1 and MYB.NT2 may be involved in the regulation of these flavonoid biosynthetic steps in *N. tabacum* pollen. A determination of the binding sites of the *N. tabacum* proteins and a demonstration of transactivation from promoters of pollen expressed flavonoid genes containing similar binding sites by transient assays and *in vivo* would be needed to test this hypothesis.

4.3.2 *Ntl1* and *Ntl2*: pollen-expressed LIM-only cDNAs encoding putative proteins with DNA-binding and protein-protein interaction potential

Two pollen expressed cDNAs, *Ntl1* and *Ntl2*, were isolated from an *N. tabacum* pollen cDNA library by heterologous hybridisation with a probe containing conserved zinc finger repeats from a pollen-specific cDNA *sf3*. The cDNA clones had related but unique nucleotide sequences. Data base searching with the derived amino acid sequences revealed the derived proteins had significant sequence similarity to the SF3 protein. Alignment of these sequences revealed the positions of the two imperfect LIM domains separated by 46 amino acid residues which were previously characterised in SF3. Each repeat has a characteristic pattern of cysteine and histidine residues which are involved with chelating two zinc ions, to form a double tetrahedral zinc finger motif (Figures 1.3 and 4.18). In addition, a short basic region at the end of the second zinc finger of each LIM domain is conserved between the NTL proteins and SF3. Despite the sequence similarity, both of the *N. tabacum* proteins lack a carboxy-terminal pentapeptide repeat present in SF3. This domain may represent a putative transactivation domain having an acidic character, with a net charge of -10 (10 acidic residues in the 44 residues of the domain, (Baltz *et al.*, 1992a.)), and 10 serine and threonine residues, which are targets for phosphorylation. In view of the structural similarity to the functional domains of eukaryotic transcription factors, it was suggested that SF3 could activate pollen transcription by binding to the promoters of pollen genes, via the zinc fingers and basic region of the LIM domain and interact with the RNA polymerase transcriptional complex via its carboxy-terminal repeat (Baltz *et al.*, 1992a.).

The sf3 cDNA was used as a probe to isolate a clone from a sunflower ovary cDNA library (OLIM) as well as from an *Arabidopsis thaliana* genomic library (ATL-1) (Dr. A. Steinmetz, personal communication). Both of these clones have strong sequence homology (90 %) to SF3. Interestingly, these clones also lack the acidic pentapeptide repeat.

The LIM domain is related to another cysteine-rich motif, the RING finger (Lovering *et al.*, 1993). Despite sequence similarities however, two important differences distinguish the motifs: the 'CCCH-CCCC' motif found in the RING finger is distinct from the 'CCHC-CCCC/H/D' motif of the LIM domain; the inter-repeat spacing is one residue in the RING finger and two residues in the LIM motif. The alignment of the LIM domains of different LIM domain proteins, identifies a number of conserved residues which distinguish LIM zinc fingers from other related zinc finger structures. These sequence differences together with the presence of associated domains were used as the basis of a classification of all LIM domain proteins (Figure 1.4; (Taira *et al.*, 1995)).

Nuclear magnetic resonance spectroscopy studies of the LIM domain of avian cysteine rich protein (CRP; (Perez-Alvarado *et al.*, 1994)), has shown that the structure of the carboxy-terminal 'CCCC' motif of the LIM domain is essentially identical to that observed for the DNA interactive 'CCCC' motifs of the GATA-1 and steroid hormone receptor binding proteins. Suggesting that this LIM motif may have a DNA binding function.

In vitro DNA binding assays using a variety of deletion mutants has shown that the LIM domains of SF3 are required for DNA binding (Baltz *et al.* 1996). Deletion of the first LIM domain and destabilisation of the second LIM domain by increasing or decreasing the number of amino acid residues separating the two zinc fingers, abolishes complex formation. Removal of the carboxy-terminal pentapeptide repeat and shortening the inter-LIM region has no effect on DNA binding, indicating that these regions are not involved in the interaction of the protein with DNA.

While the similarity to other classes of zinc finger domains and the demonstration of DNA binding by SF3 suggests that LIM domains may bind DNA and RNA *in vivo*, no published evidence for sequence specific DNA binding by LIM domains exists. However, there is strong evidence for the role of LIM domains in protein-protein interactions. Compared with animal LIM domain proteins highest homology to the plant LIM proteins was found in data-base homology searches to the cysteine rich protein from humans and chicken (hCRP (Liebhaber *et al.*, 1990); cCRP (Crawford *et al.*, 1994)) and muscle LIM protein from humans and *Drosophila* (MLP and DMPL1

(Arber *et al.*, 1994)). These LIM proteins, like the NTL and SF3 are classed as Group 2 LIM proteins by the classification system of Taira *et al.*. The sequence similarities of these Group 2 proteins may suggest similar functional roles.

A role in the promotion of myogenic differentiation has been assigned to MLP and DMLP1. These genes encode proteins with nuclear and cytoplasmic localisation that are expressed during muscle cell differentiation, and contain one or two LIM domain(s) respectively. Experiments involving overexpression and antisense mediated reduction of MLP in myogenic cell lines, provide evidence for a positive role of MLP in myogenesis. Interestingly, other LIM-only proteins not normally associated with muscle could substitute for MLP in the overexpression experiments, indicating that the LIM domains are the relevant functional groups involved in the myogenic activity of MLP. Specificity may exist under physiological conditions by a mechanism of regulated subcellular localisation rather than sequence-specificity.

Strong evidence for the involvement of LIM only proteins in protein-protein interactions came from the demonstration that hCRP can form homodimers with itself (Feuerstein *et al.*, 1994) and heterodimers with a Group 3 LIM domain protein, ZYXIN (Schmeichel and Beckerle, 1994). Like MLP, CRP contains two LIM domains only. Using *in vitro* filter binding assays and two-hybrid cell based interaction assays, it was shown that hCRP can efficiently homodimerise. Using deletions of the protein, the dimerisation domain was mapped to the LIM domains and the dimerisation ability was transferred to an unrelated protein by fusion of a single minimum LIM double zinc finger motif (70 % of the interaction with full length hCRP). Dimerisation was shown to be independent of DNA binding, but dependant on the coordination of zinc ions by introducing cysteine to serine substitutions.

ZYXIN, a protein with three LIM domains located at the carboxy-terminus to an extensive proline-rich region has been localised to cell adhesion plaques in the cytoskeleton in fibroblasts. Using *in vitro* blot overlay assays, a specific interaction of ZYXIN and cCRP was demonstrated. This interaction was dependant on zinc, illustrating that the LIM zinc fingers are needed for the protein-binding function. In this way, it was suggested ZYXIN may bind a protein at more than one LIM domain site and so may function as a cytoskeletal docking site for the assembly of macromolecular complexes. From the studies with MLP and CRP the function of LIM-only Group 2 proteins may be to regulate the function of other proteins via protein-protein interactions.

The presence of two conserved LIM domains in NTL1 and NTL2 suggests their evolutionary origin may be a result of gene duplication or gene conversion. The NTL

proteins may have evolved from an ancestral unit of 98 amino acids comprising the double zinc finger LIM domain. The near perfect alignment of the conserved amino acid residues suggests that tandem duplication has led to the functional NTL proteins. The high sequence conservation of the metal chelating residues (56 %) and basic residues (67 %) supports this hypothesis. Conservation of these residues might indicate functional importance. High sequence divergence in the non-LIM domain regions (65 %) suggests that the duplication event was ancient. A study of the intron positions in the *Ntl* genes would be needed to argue for or against such a duplication event.

4.4 Summary

Two classes of pollen expressed cDNAs encoding putative transcriptional regulators have been isolated from *N. tabacum* pollen. Amino acid sequence analysis of the predicted MYB.NT1 and MYB.NT2 proteins revealed high homology to the MYB.PH3 predicted protein from *Petunia*. Comparison of the MYB domain sequence suggested the three proteins could bind to similar recognition sites and regulate similar genes. A possible functional role for the *myb.Nt1* and *myb.Nt2* is therefore regulation of flavonoid biosynthesis in pollen.

The predicted proteins of two cDNA clones, *Ntl1* and *Ntl2*, were found to have high homology to the sunflower SF3 predicted protein. The structure and sequence conservation of the zinc finger LIM domains suggests a potential role in the post-translational regulation of the function of other proteins by protein-protein interactions. A role in transcriptional regulation by DNA binding is also possible.

In order to characterise these cDNA clones, their detailed expression patterns were investigated by spatial and temporal Northern blot analysis. In addition, the genomic copy number of the *Ntl* and *myb.Nt* genes and the complexity of the *lim* and *myb* gene families in *N. tabacum* and parental species were investigated by Southern analysis.

Chapter Five.

Expression and genomic organisation of sequences encoding MYB and LIM domain proteins.

5.1 Introduction and aims

This chapter describes the results of experiments designed to further characterise the *myb.Nt1*, *myb.Nt2*, *Ntl1* and *Ntl2* cDNAs. One route taken was to investigate the patterns of expression in *N. tabacum*, by Northern blot analysis, both spatially in plant tissues and temporally in the developing pollen grain. The development of pollen in *N. tabacum* is marked by defined stages (section 1.2), which correspond to specific bud lengths. Release of free microspores from tetrads occurs at bud lengths of 1 cm, the free uninucleate microspore stage extends to bud lengths of 1.5 cm when pollen mitosis I (PMI) results in the production of bicellular pollen. The pollen grains reach functional maturity at anthesis when bud sizes reach 5 cm in length. Harvesting of *N. tabacum* buds and sorting into specific size groups allows the collection of purified populations of microspores and pollen grains at defined stages of development. The homogeneity of cells collected in this way is estimated to be between 85 % and 93 % depending on the developmental stage and the number of buds collected ((Schrauwen *et al.*, 1990); (Bedinger and Edgerton, 1990b)). By performing Northern blot analysis with blots containing RNA extracted from each of the pure populations of microspores and pollen cells, it is possible to monitor steady-state mRNA transcript levels of a gene throughout the history of a single cell. In this way changes in transcript levels can be correlated with stages of development. Further, by correlating bud lengths to days before anthesis it is possible to relate changes in mRNA transcript levels to time and thus estimate actual rates of accumulation.

Previous characterisation of the temporal expression patterns of pollen-specific genes during pollen development had defined three broad groups of genes which differ with respect to the timing of appearance of steady-state transcript during pollen development (Stinson *et al.*, 1987). One group of genes are transcribed throughout pollen development, a second group, named 'early' genes, are transcribed during microspore development with transcript levels decreasing to pollen mitosis I (PMI). Transcripts of a third group of 'late' pollen-specific genes start accumulating after PMI and increase to anthesis. The majority of pollen-expressed genes isolated belong to this third class.

Previous studies of pollen gene expression have used whole anthers or flower buds at different developmental stages. By isolating relatively pure populations of microspores and pollen at defined developmental stages a more accurate and detailed analysis of pollen gene expression should be possible. This will define with more precision the three broad groups of pollen gene expression. For example, by defining pollen gene

expression just prior to and after the key developmental switch PMI, it will be possible to determine if there is a strong correlation between the first appearance of late pollen transcripts and PMI. If this is the case then this would indicate that late pollen transcripts are activated as a result of PMI.

A second route taken to characterise the pollen expressed genes isolated was an analysis of the *myb-* and *lim-*gene copy number in *N. tabacum*, in order to investigate whether other related genes to the *myb.Nt* and *Ntl* pairs of cDNA clones are present in the *N. tabacum* genome.

5.2 Results
5.2.1 An investigation into the relationship between RNA loading amount and scanning densitometry signal

As a preliminary study in the investigation of the spatial and temporal expression patterns of *myb.Nt* and *Ntl* gene expression by Northern blot analysis it was considered necessary to establish the relationship between amount of RNA immoblised onto a nylon membrane and quantification of the photo-autoradiographic signal after radioactive probing with a homologous DNA sequence probe.

The concentration of a sample of intact *N. tabacum* leaf total RNA was determined by spectroscopy and quantification on an agarose gel next to a known amount of commercial RNA ladder. A dilution series of the RNA was prepared and eight RNA samples loaded in adjacent lanes on an agarose gel and separated by electrophoresis (10, 5, 2.5, 1.125, 0.625, 0.312, 0.156 and 0.078 μg respectively). The agarose gel was blotted and the blot probed with an 18S rRNA ^{32}P-labelled DNA probe. Radioactive signals were visualised and quantified by scanning densitometry. Values were expressed relative to the 10 μg signal and plotted against μg of RNA loaded on the agarose gel (Figure 5.1). A linear relationship between signal and loading amount was observed. The assumption was made that transfer of RNA from the agarose gel to the nylon membrane was 100 % efficient and radiolabelled probe was in excess in the hybridisation reaction.

Figure 5.1

An investigation into the relationship between total RNA loading amount and scanning densitometry signal.

Graph representing relationship between scanning densitometry signal, obtained by phosphoimaging, and total RNA amount (μg).
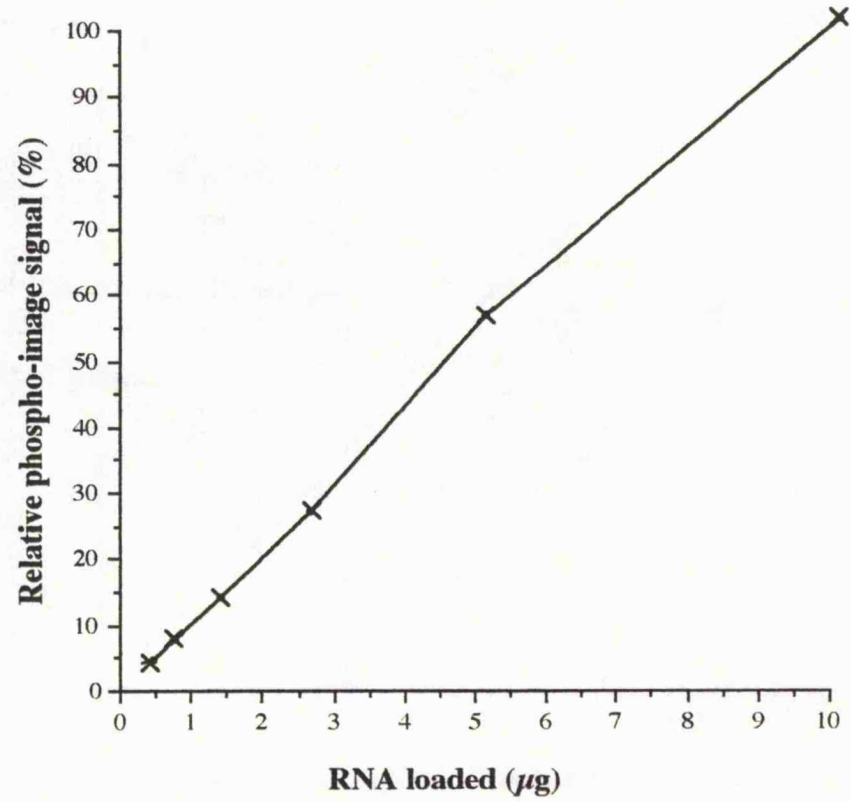
5.2.2 An investigation of the spatial expression of *myb.Nt*- and *Ntl*-specific transcripts in *N. tabacum*

To determine the overall organ-specific accumulation of *myb.Nt*- and *Ntl*- specific transcripts, total RNA was extracted from seedling (13 and 17 day), root, young leaf, sepal, petal, pistil, stamen and pollen (mature and 8 hours germinating) tissues. Total RNA from each tissue was separated by agarose gel electrophoresis and the gel blotted to a nylon membrane. Duplicate blots prepared in the same way were probed with *myb.Nt1* and *Ntl1* EcoRl/*Xho*l digested cDNA inserts separately (1.8 kb and 0.9 kb respectively) at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C). Quantification of the radioactive signals was achieved by scanning densitometry using a phosphoimager (Molecular Dynamics), and all values normalised against the 18S ribosomal RNA (rRNA) signal obtained by stripping and reprobing the same blot with an 18S rRNA DNA probe. Normalised signals from each of the samples were compared to that of mature pollen on the same blot, which was arbitrarily set at 100 %.

A 1.8 kb transcript was detected in RNA from stamens and mature and germinating pollen using the *myb.Nt1* 1.8 kb cDNA insert to probe a spatial RNA blot (Figure 5.2). A 58 % transcript abundance was seen in stamens compared to mature pollen. Although *myb.Nt*-specific transcript may be produced in anther and filament tissue at a low level it seems likely that this signal can be accounted for by transcript from the pollen grains. A 28 % relative transcript abundance was seen in germinating pollen. Upon overexposure of the blot a 1.8 kb transcript signal was seen at a constant level in the other tissue lanes. These signals either represent low level expression of *myb.Nt* genes (0.8 % average relative abundance), or transcript from distantly related *N. tabacum myb* genes. Alternatively, it could be due to trapping of the probe by the abundant 18S rRNA transcripts (approximately 1.9 kb in size). In summary, these results show that *myb.Nt* transcripts are present at high abundance in mature and germinating pollen and further suggest that *myb.Nt1* and *myb.Nt2* are pollen-specific genes.

A similar conclusion can be drawn from the signals obtained using the *Ntl1* EcoRl/*Xho*l digested 0.9 kb cDNA insert as a probe to a similar spatial RNA blot (Figure 5.3). A 0.9 kb transcript was observed in mature and germinating pollen and stamens at 100 %, 27 % and 32 % levels. Although *Ntl*-specific transcript may be produced in anther wall and filament tissue at a low level it is likely that this signal can

Figure 5.2

Spatial Northern analysis of *myb.Nt*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different organs of *N. tabacum*, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.

a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

17 Day seedling

13 Day seedling

Root

Young leaf

Sepal

Petal

Pistil

Stamen

Mature pollen

Germinating pollen

Figure 5.3
Spatial Northern analysis of *Ntl*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different organs of *N. tabacum*, hybridised with whole cDNA at high stringency ((20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.
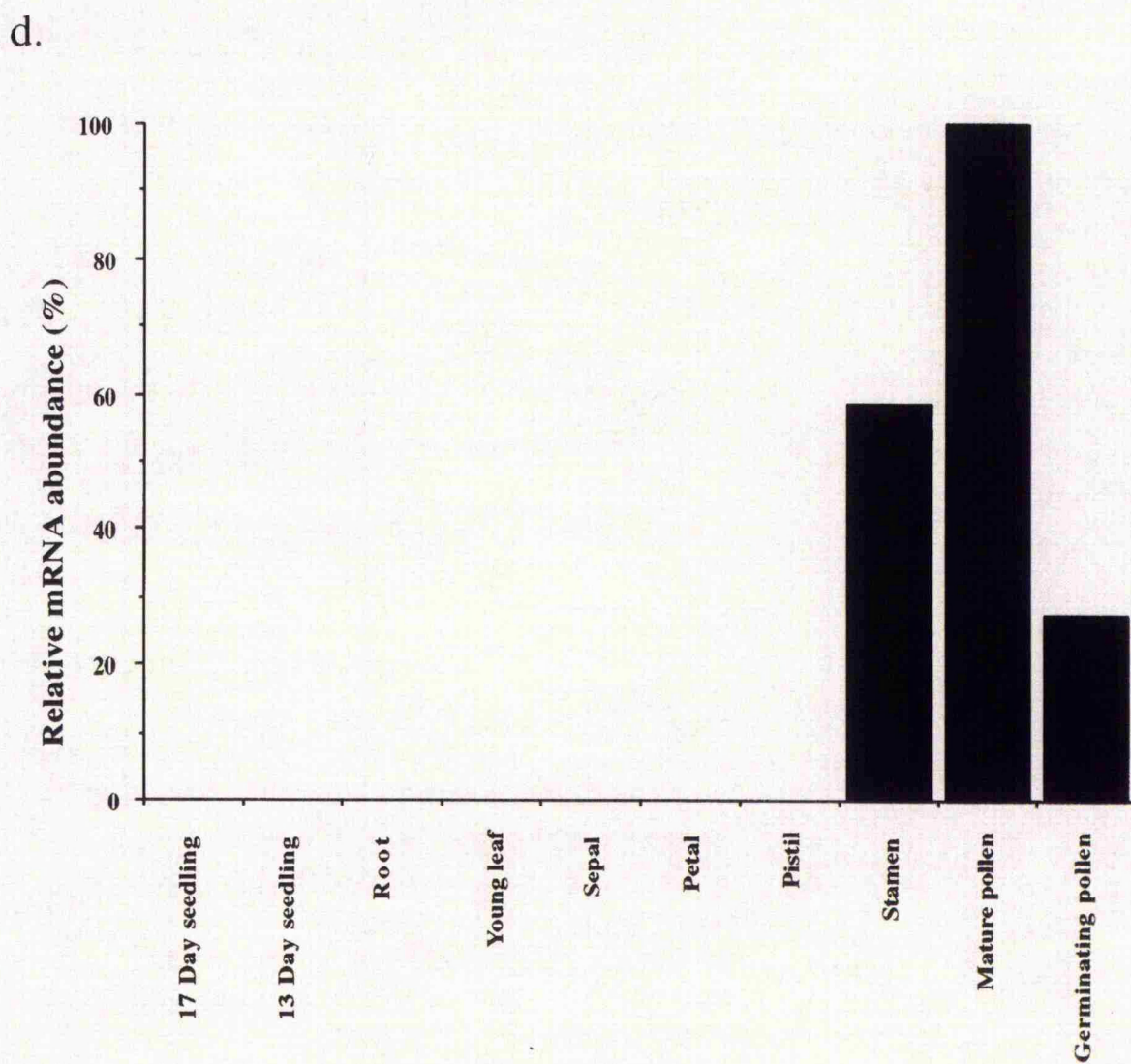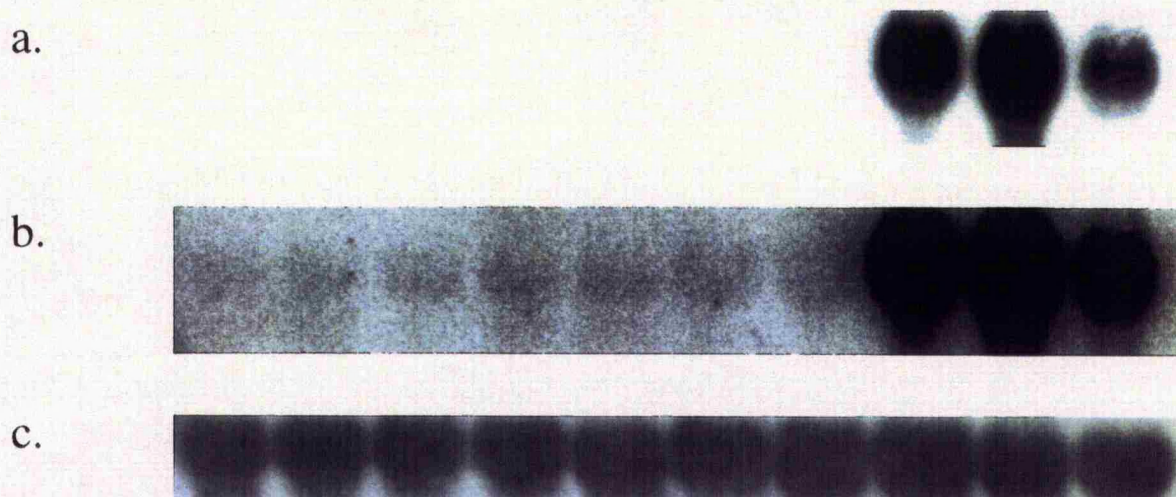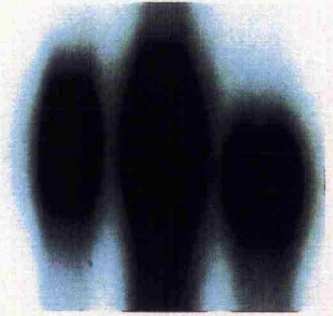
a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

17 Day seedling
13 Day seedling
Root
Young leaf
Sepal
Petal
Pistil
Stamen
Mature pollen
Germinating pollen

be accounted for by transcript from the pollen grains, considering the strong signal obtained from pollen. Overexposure of the blot revealed the presence of *Ntl*-specific transcripts in root and young leaf tissue. These signals either represent low level expression of *Ntl* genes (0.008 % and 0.046 % relative abundance in root and young leaf tissue respectively) or expression of more distantly related *N. tabacum lim* genes. In summary, these results show that *Ntl*-specific transcripts are present at high abundance in mature and germinating pollen and further suggest that *Ntl1* and *Ntl2* are pollen-specific genes.

5.2.3 An investigation of the temporal expression of *myb.Nt*-and *Ntl*- specific transcripts in *N. tabacum*

The temporal accumulation pattern of *myb.Nt*- and *Ntl*-specific transcripts was analysed. Buds were collected of field grown *N. tabacum* cv. Samsun and sorted into five size groups: 10-12 mm, 12-15 mm, 17-30 mm, 30-40 mm and 40-50 mm. The microspores and pollen cells which were present in these anthers were early uninucleate microspores released from tetrads, late uninucleate microspores, early binucleate pollen cells, mid binucleate pollen cells and late binucleate pollen cells respectively ((Koltunow *et al.*, 1990), (Schrauwen *et al.*, 1990)). Non-dehisced anthers were manually squeezed using thumb and forefinger from the buds and placed into an isotonic 0.3 M mannitol solution until all the buds had been processed. Microspores and pollen were isolated from the anther material by grinding in a pestle and mortar in a strongly denaturing buffer containing 4M guanidinium isothiocyanate (section 2.7.1), to protect the RNA from degradation, followed by filtration to separate the cells from anther debris and sedimentation of the cells to concentrate the material. In addition mature and germinating pollen was used as a source of material for total RNA extraction. Mature pollen was germinated on filter paper soaked in Tupy's germination medium (section 2.12.4) for 8 hours, by which time approximately 95 % of the pollen grains had produced long tubes. Total RNA was prepared from the isolated microspores and pollen, separated by agarose gel electrophoresis and the gel blotted to a nylon membrane. Two identical blots were probed at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C) with either of the *myb.Nt1* or *Ntl1* whole cDNA inserts.

Myb.Nt-specific transcripts (1.8 kb) were detectable at 2 % abundance relative to mature pollen in late microspores immediately prior to PMI, which occurs between

130

Figure 5.4

Temporal Northern blot analysis of *myb.Nt*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of *N. tabacum*, hybridised with whole cDNA at high stringency ((20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.
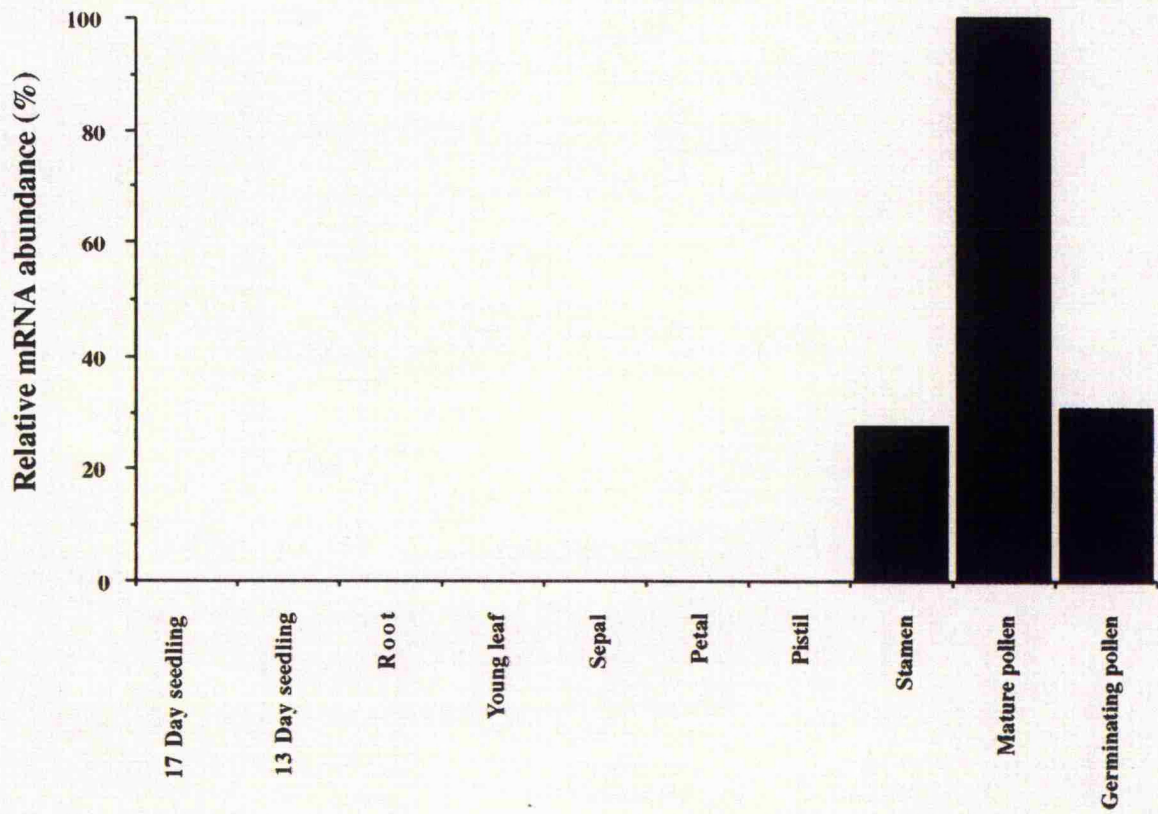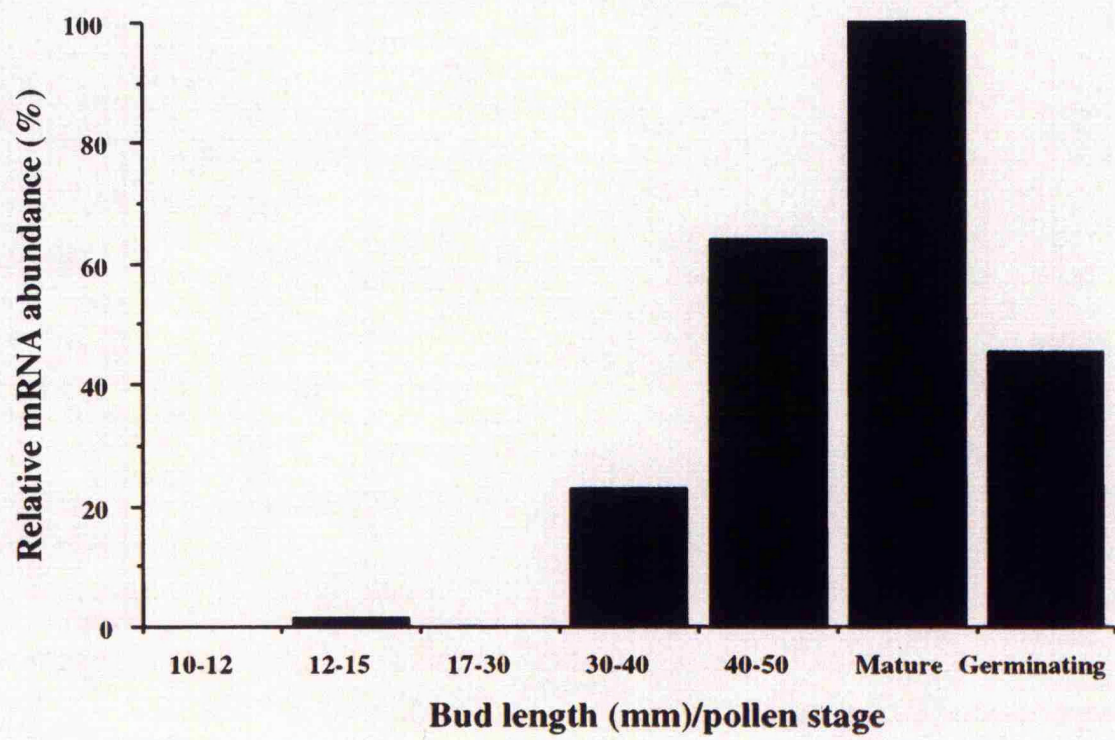
a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

10-12   12-15   17-30   30-40   40-50   Mature   Germinating

Bud length (mm)/pollen stage

stages 2 and 3 (Figure 5.4). Transcript levels decreased approximately nine fold to 0.16 % in early binucleate pollen. Transcripts then subsequently accumulated rapidly in mid binucleate pollen cells and increase in late binucleate and mature pollen (24 %, 65 % and 100 % respectively). Germinating pollen showed a two fold drop in transcript levels to 46 %.

To provide a valuable 'real-time' perspective of the rate of mRNA accumulation in microspores and pollen at each stage of development bud sizes were related to days before anthesis (Figure 5.5). An increase in transcript levels can be accounted for by an upregulation of transcription and/or a decrease in transcript turn over. The pulse of transcript present in late uninucleate microspores was seen at 6 days before anthesis, with PMI at 5 days. A dramatic increase in transcript levels was seen between 3 and 4 days, marking early to late bicellular pollen development, with levels increasing until anthesis. Transcript accumulation then increased four fold from mid to late binucleate pollen development until maturity.

A similar pattern of temporal expression was seen with *Ntl*-specific transcript accumulation (Figure 5.6). A 0.9 kb transcript signal was first seen in late uninucleate microspores at a relative abundance of 0.24 %. After PMI in early binucleate pollen transcript levels decreased to approximately 0.06 % relative abundance. Transcript production then increased dramatically in mid, late and mature bicellular pollen (10 %, 50 % and 100 % respectively). Transcript levels were approximately two fold lower in germinating pollen (48 %). Relating bud size to days before anthesis (Figure 5.5), the transcripts were detectable at 6 days before anthesis. Transcripts appeared at approximately 3 days and rose to 10 % steady state levels by approximately 2 days 18 hours in mid binucleate pollen . In the next 24 hours transcript levels rose to 48 % relative abundance during late binucleate pollen development reaching 100 % at anthesis.

A number of *N. tabacum* pollen-specific cDNAs were available which were used as control probes to the same temporal RNA blots to investigate the patterns of expression of characterised 'early', 'late' and constitutively expressed genes in pollen. By comparing these expression profiles to the patterns of expression of the *myb.Nt* and *Ntl* genes it is possible to classify with accuracy *myb-* and *lim-*gene expression in pollen.

Ntp303 which is thought to have a role during pollen germination (Weterings *et al.*, 1992) and *Npg1*, which encodes a polygalacturonase (Tebbutt *et al.*, 1994) are both 'late' genes, while *ntm19* (Dr. M. Oldenhoff, unpublished) is an 'early' gene with no

132

known function and *pma1* (Michelet *et al.*, 1994) a constitutively expressed gene encoding the β-subunit of the ATPase proton pump.

An *ntp303* cDNA (1.9 kb) was used as a probe to a temporal Northern blot (Dr. K.A.P. Weterings, Nijmegen) used previously with the *myb.Nt1* probe and stripped of signal to background levels (Figures 5.7 and 5.5). Signal was seen in all bud size groups. A signal equivalent to 0.28 % of that obtained with mature pollen was seen in buds containing early uninucleate microspores. An increase to 2.66 % relative abundance was seen in late uninucleate microspores with a slight decrease in early binucleate pollen cells to 1.58 %. A rapid increase was then seen in early to mid binucleate pollen cells (48 %) and a more gradual increase to pollen maturity and anthesis (72 % and 90 % respectively). Transcript abundance then increases in germinating pollen (100 %).

Previous characterisation of *ntp303* expression during microgametogenesis by Northern blotting had not detected any transcript in uninucleate microspores but a dramatic rise in transcript levels through early, mid and late binucleate pollen cells. A further increase in transcript levels had been seen during germination, with levels rising to a maximum of 270 % relative to mature pollen in pollen germinated for two hours. From this pattern of expression, NTP303 is thought to have a role in germination. Confocal laser scanning microscopy and *in situ* hybridisation detected *ntp303* transcript first in mid binucleate pollen cells and then in increasing levels to pollen maturity.

A cDNA encoding an *N. tabacum* polygalacturonase, *Npg1* (1.5 kb; Dr. D. Lonsdale, Norwich) was used as a probe to a temporal RNA blot used previously with the *Nt1* cDNA probe and stripped of signal (Figures 5.8 and 5.5). Signal was first seen in late uninucleate microspores with 0.89 % relative abundance. During development a transcript signal representing 1.14 % in early binucleate pollen cells was followed by a dramatic rise in levels to 47 %, 50 % and 100 % in mid, late and mature bicellular pollen. A six fold drop in transcript abundance was seen in germinating pollen.

Ntg1 is thought to function as an exopolygalacturonase for two reasons. First as a result of its high expression in mature pollen, a rich source of exopolygalacturonases, and second, due to its distinct sequence characteristics shared with other exopolygalacturonases. The original characterisation of *Npg1*-specific transcript levels in pollen development, had used RNA extracted from anthers in buds of five defined size groups, representing microspores and pollen at meiosis/tetrads, uninucleate microspores, PMI, early to mid binucleate pollen cells, mid to late binucleate pollen cells and germinating pollen (Tebbutt *et al.*, 1994). Low levels of transcript were

133

detected in all stages, with a dramatic rise in transcript levels seen in mid to late binucleate pollen cells. Transcript levels decreased in germinating pollen.

A different pattern of transcription was seen with an early *N. tabacum* microspore specific cDNA, *ntm19* (Figure 5.9; Dr. J.A.M. Schrauwen, Nijmegen). Using the whole cDNA as a probe (1 kb), to a new *N. tabacum* temporal Northern blot, transcript was first detected in early uninucleate microspores at 4 % relative abundance. A dramatic rise in transcript levels was seen in late uninucleate microspores and a decrease to 8 % levels in early binucleate pollen cells. No transcript is detectable at any other stage of development. The rapid peak in transcript levels was seen clearly in the relationship of *ntm19*-specific transcript abundance to days before anthesis (Figure 5.5). Here a 98 % increase in transcript levels was seen within 33 hours corresponding to late uninucleate microspore development. Such a specific pattern of transcription would indicate a role in microspore development.

A contrasting pattern of transcription was seen with the *pma1* cDNA encoding a *N. plumbaginifolia* pollen expressed β-subunit ATP synthase (ATPase) (Figure 5.10 and 5.5; Dr. D. Twell, Leicester). Transcript was detected in all stages of development. A transcript abundance of 4 %, 42 %, 16 %, 100 %, 86 %, 27% and 9 % was seen in early and late uninucleate, early, mid, late binucleate, mature and germinating pollen respectively.

Pma1 is a 'house keeping' gene being essential for basic cell function. This is reflected in its pattern of expression in all stages of development. The ATPase protein complex is present in the plasma membrane where it pumps protons out of the cell. The presence of GUS in transgenic *N. plumbaginifolia* plants containing a *pma1* promoter-GUS construct in mature pollen and pollen tubes was demonstrated, suggesting that these tissues are involved in active transport processes, leading to the uptake or delivery of ions or nutrients.

5.2.4 Southern analysis of *myb.Nt1* and *Nt11*

To estimate the number of genes in the *N. tabacum* genome that are closely related to the *myb.Nt* and *Nt1* cDNAs, a DNA gel blot was prepared from *N. tabacum* genomic DNA that had been digested with *Hind*III and *Eco*RI restriction endonucleases and probed with either a 1 kb *Eco*R1/*Hind*III fragment of *myb.Nt1* containing the MYB-domain sequence, or the whole *Nt11* 0.9 kb cDNA insert, separately at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C). The size of the respective gene families

Figure 5.5

Graph representing relative steady-state mRNA abundance of *myb.Nt-*, *Ntl-*, *ntp303-*, *npg1-*, *ntm19-*, and *pmal*-specific RNAs versus days of before anthesis.

Abbreviations: PMI, pollen mitosis I.
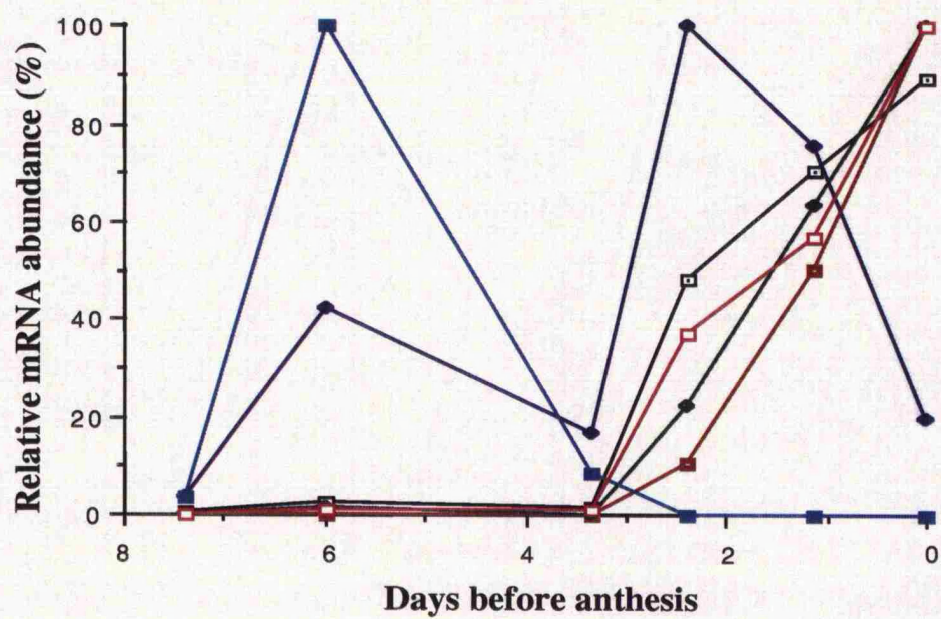
Figure 5.6
Temporal Northern blot analysis of *Ntl*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of *N. tabacum*, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure,

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.

a.

b.

c.

d.
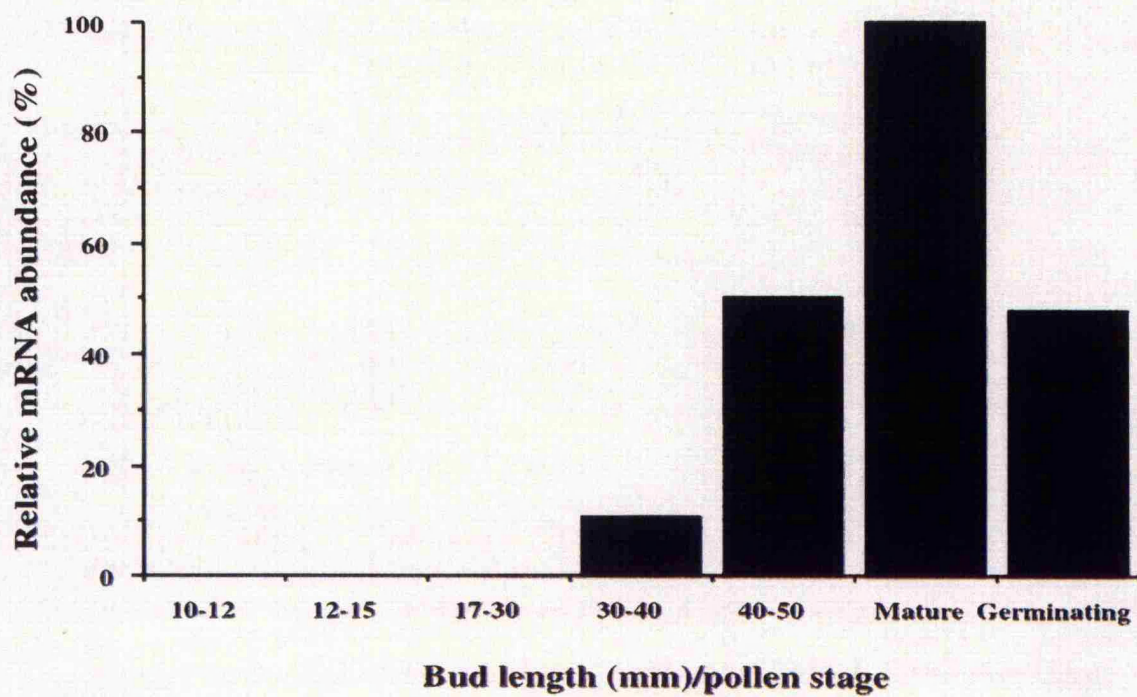
Figure 5.7
Temporal Northern blot analysis of _Ntp303_-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of _N. tabacum_, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.

a.

b.

c.

d.

Relative mRNA abundance (%)

Bud length (mm)/pollen stage
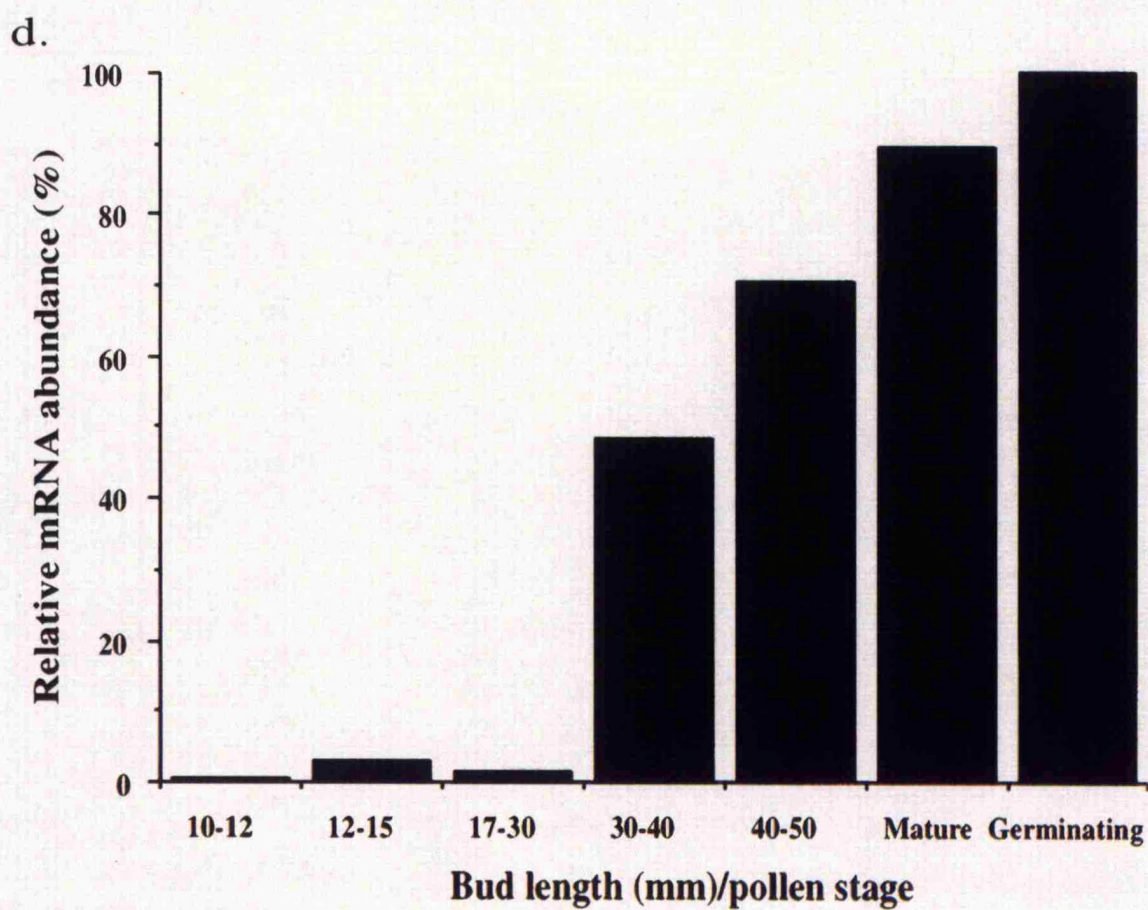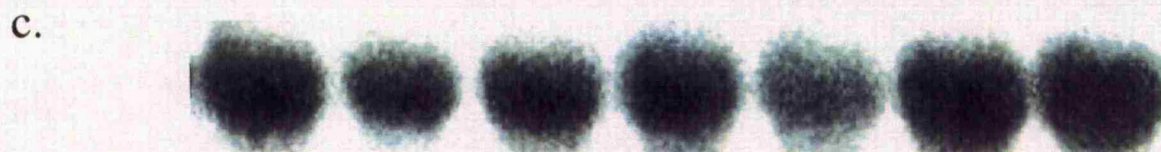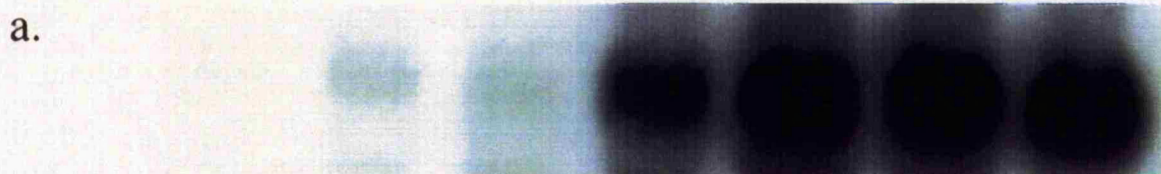
Figure 5.8
Temporal Northern blot analysis of *npg1*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of *N. tabacum*, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.
b. Over exposed image.
c. 18S rRNA loading control.
d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.

a.
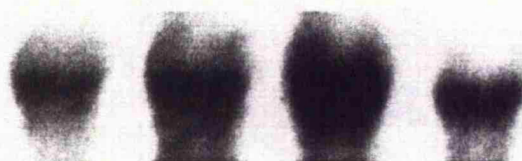
b.

c.

d.
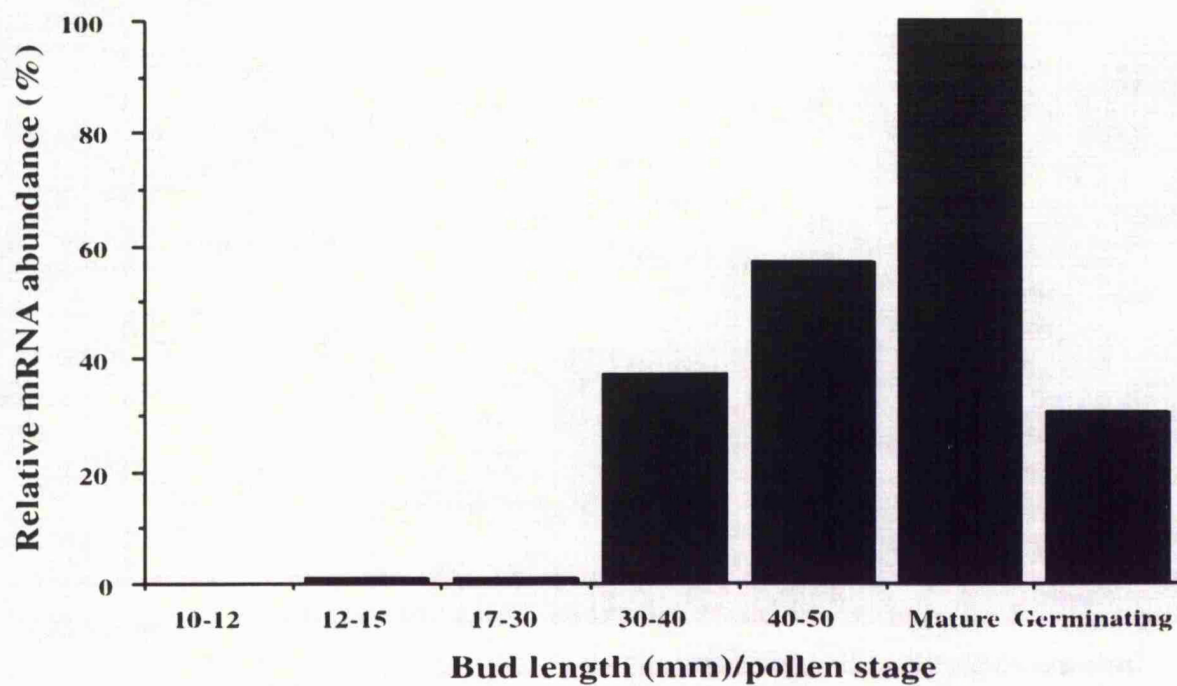
a.

b.

c.

d.
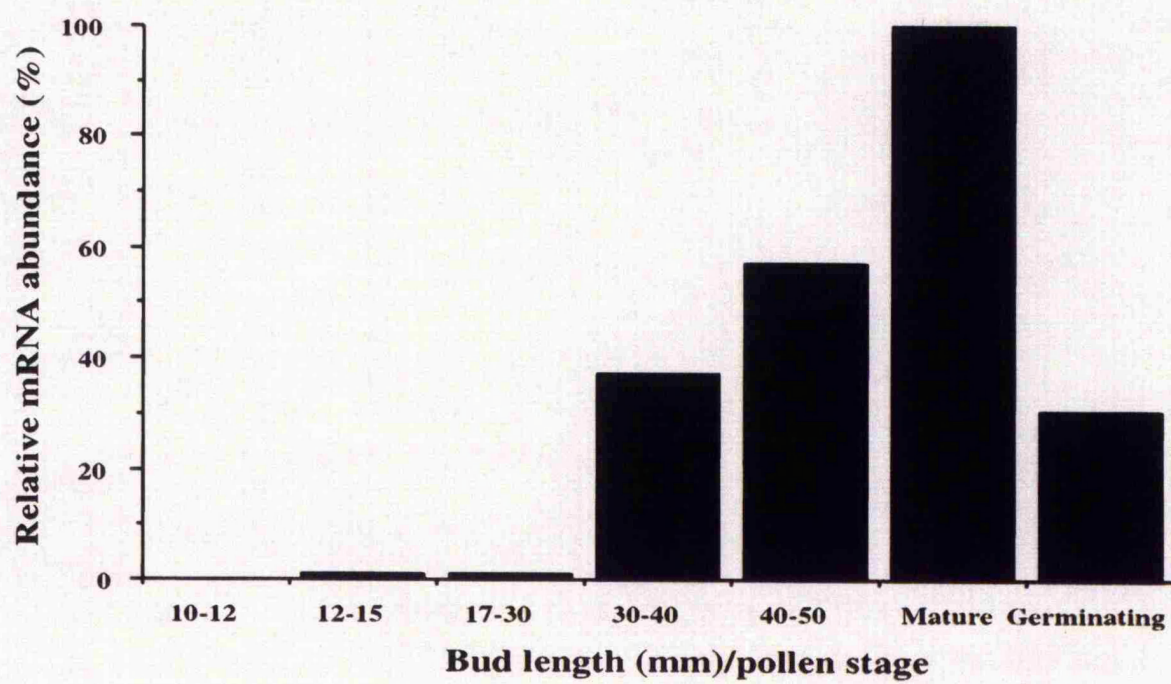
Figure 5.9

Temporal Northern blot analysis of *ntm19*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of *N. tabacum*, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.

a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

10-12   12-15   17-30   30-40   40-50   Mature

Bud length (mm)/pollen stage
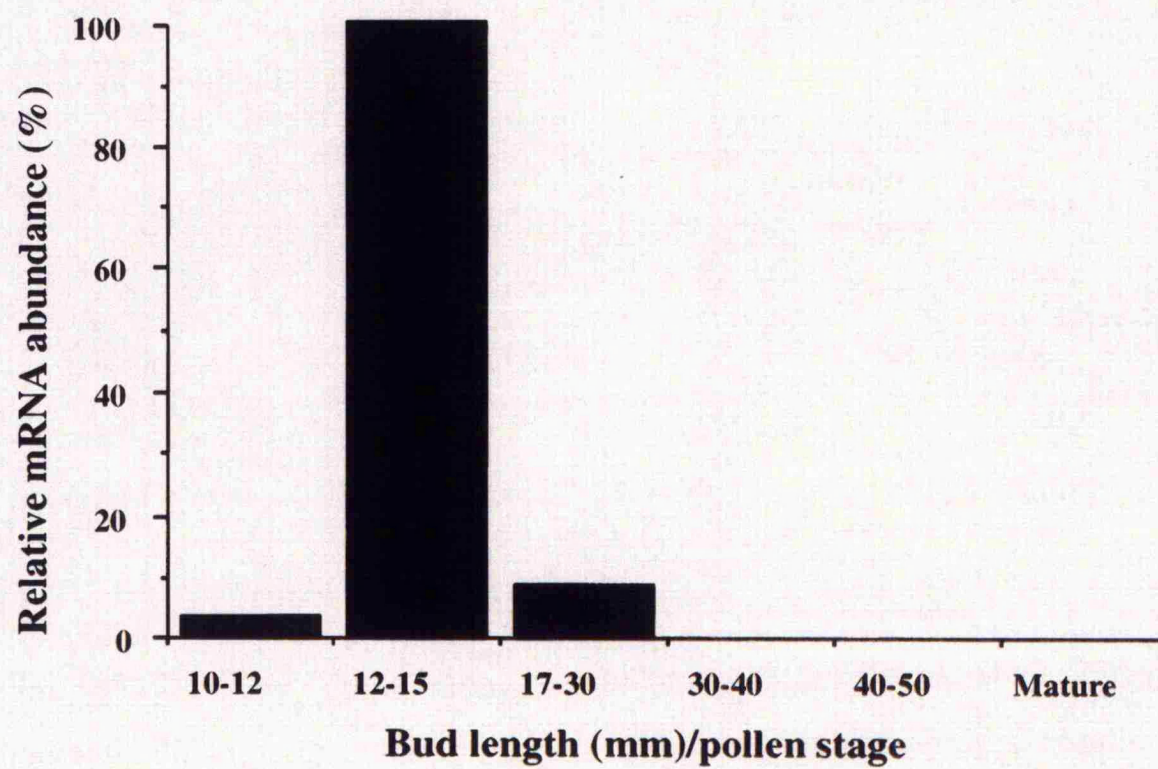
Figure 5.10
Temporal Northern blot analysis of *pma-1*-specific RNAs.

Image derived by scanning densitometry of total RNA (10 μg per lane) from different bud size classes of *N. tabacum*, hybridised with whole cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight photo-autoradiographic exposure.

b. Over exposed image.

c. 18S rRNA loading control.

d. Graph representing relative steady-state mRNA abundance. All values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.
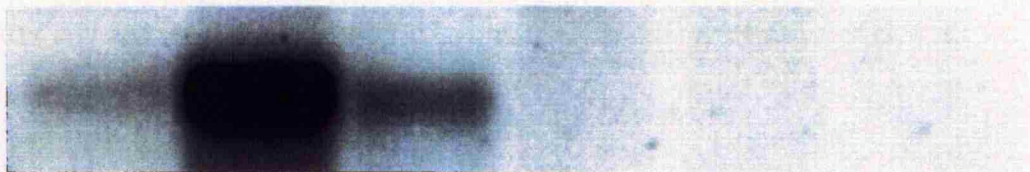
a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

10-12   12-15   17-30   30-40   40-50   Mature Germinating

Bud length (mm)/pollen stage
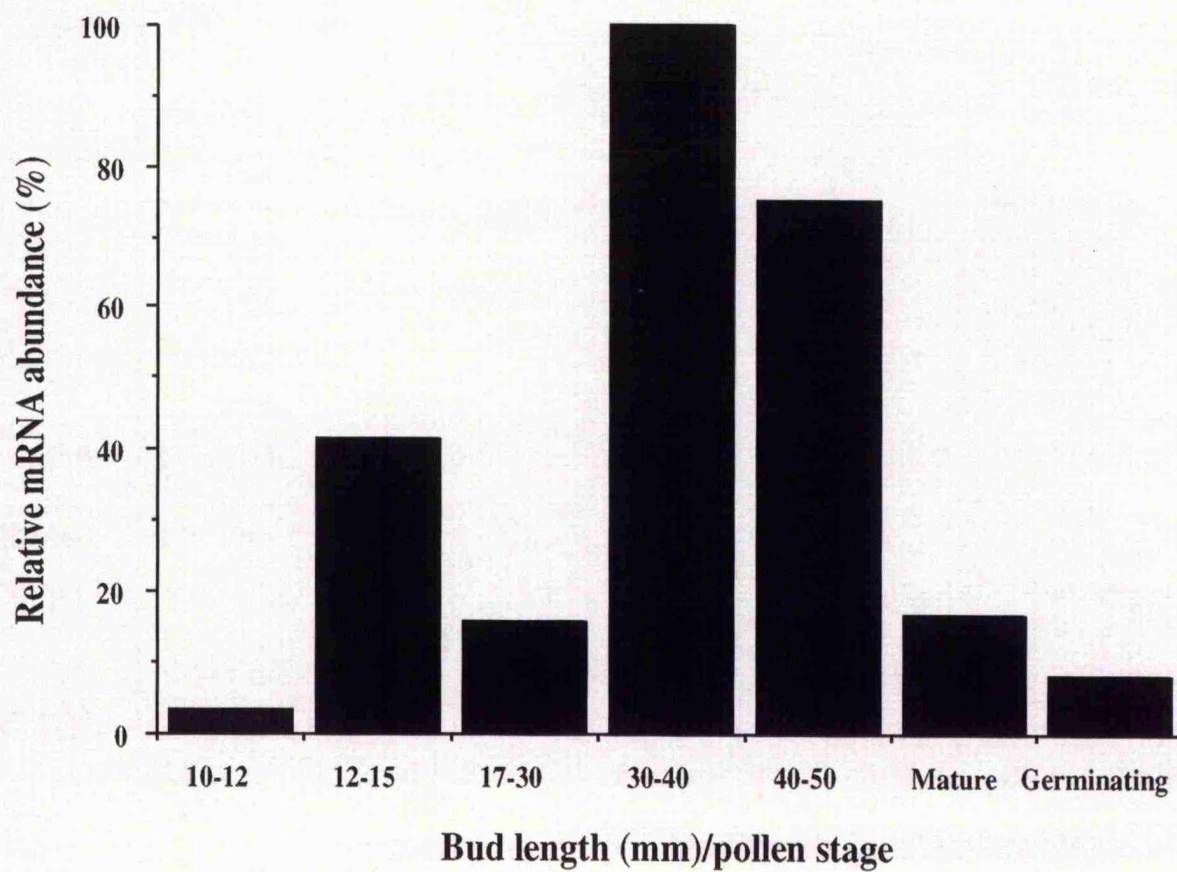
was approximated based on the number of bands in each lane and the intensity of the bands compared with copy number standards. *N. tabacum* has an amphidiploid genome, resulting from the fusion of the diploid genomes of *N. sylvestris* and *N. tomentosiformis* (Goodspeed, 1954). Genomic DNA was prepared and digested as with *N. tabacum*. DNA digested with the same restriction enzyme from *N. tabacum* and each of the parents was run in adjacent lanes on an agarose gel. Duplicate gels were blotted to a nylon membrane and each of the two blots probed with one of each of the pairs of cDNAs. By analysing the pattern of hybridising bands in each of the parents, it is possible to gain a clearer picture of the copy number of *myb.Nt* and *Ntl* related sequences in *N. tabacum*.

A similar pattern of hybridising fragments was obtained with each of the *myb.Nt1* and *myb.Nt2* cDNAs (Figure 5.11). After digestion with *Eco*RI and *Hind*III, a single hybridising fragment in each of the parental genomes was seen. Corresponding fragments of the same size were present in the *N. tabacum* genome. With *Eco*R1 restriction digestion, a fragment of approximately 4 kb was present in the *N. tabacum* genome and a corresponding fragment was present in the *N. tomentosiformis* genome. The 4 kb fragment in the *N.tabacum* genome was approximately half the intensity of the corresponding band in the *N. tomentosiformis* genome. This is expected as half the amount of target DNA was present for hybridisation in the *N. tabacum* lane as a result of eqivalent amounts of DNA from each genome having been loaded. The *N. tabacum* genome being amphidiploid has twice the amount of DNA in it's genome. Therefore loading equal amounts of DNA per track will result in half the signal for each gene. Two smaller hybridising fragments are seen at 1.5 kb and 0.7 kb. No similarly sized fragments were seen in either of the two parental lines. However, a single fragment of 6 kb was present in the genome of *N. sylvestris*. Sequence polymorphism between the *N. tabacum* and *N. sylvestris* genomes could account for the differently sized hybridising fragments. A more simple pattern of hybridising fragments was seen in *Hind*III digested DNA. Here two fragments of 3.8 kb and 3.5 kb correspond to similarly sized fragments in the two parental lines. Both of the *N. tabacum* fragments were approximately half the intensity of the corresponding fragments in the parental genomes.

From the numbers of hybridising fragments in each of the genomes and a comparison of the intensities of the bands to the reconstruction lanes, it can be deduced that the *myb.Nt1* and *myb.Nt2* genes are single copy in the genome of *N. tabacum*, *N. sylvestris* and *N. tomentosiformis* . Further each parent contributes one of the genes to the *N. tabacum* genome.

Using the *Ntl1* and *Ntl2* cDNA inserts as probes to the same pair of blots which had previously been stripped of radioactive signal, a similar pattern of hybridising fragments were obtained with each of the clones (Figure 5.12). With *Eco*R1 digestion, an 8.5 kb fragment was seen in the *N. tabacum* genome which corresponded to a similarly sized fragment, double the intensity, in the *N. tomentosiformis* genome. Three additional hybridising fragments of 5 kb, 4kb and 1 kb were present in the *N. tabacum* genome and similarly sized fragments were present in the *N. sylvestris* genome. However the intensity of bands in the *N. sylvestris* genome were not double those of those of the corresponding fragments in *N. tabacum*. A simpler pattern was seen with *Hind*III digestion. Two fragments of 6 kb and 4 kb were present in the *N. tabacum* genome which corresponded to a 6 kb fragment in the *N. tomentosiformis* genome and a 4 kb fragment in the *N. sylvestris* genome. However, while the intensity of the 6 kb fragment in *N. tabacum* was half that of the corresponding fragment in *N. tomentosiformis*, the intensities of the 4 kb fragments in *N. tabacum* and *N. sylvestris* are were the same.

From the numbers of hybridising fragments in each of the genomes and a comparison of the intensities of the bands to the reconstruction lanes, it can be deduced that the *Ntl1* and *Ntl2* genes are single copy in the genome of *N. tabacum, N. sylvestris* and *N. tomentosiformis*. Further each parent contributes one of the genes to the *N. tabacum* genome. The anomalies seen with the relative intensities of the hybridising fragments could be due to sequence divergence or to variations in blotting or hybridisation efficiencies. It has been shown previously that the parental genomes in *N. tabacum* have diverged considerably from corresponding *N. sylvestris* and *N. tomentosiformis* genomes ((Okamuro and Goldberg, 1985), (van Buuren *et al.*, 1992)).

Each of the *myb.Ntl* and *Ntl1* probes were reused to probe one of the pairs of blots which had previously been stripped of signal. Low stringency (20 mM sodium phosphate, 5 % SDS, 50 °C) hybridisation conditions were used in order to estimate the number of *myb* and *lim* genes in the *N. tabacum* genome. The size of the gene families was based on the number of hybridising bands in the *N. tabacum* lane.

The pattern of hybridising fragments obtained with the *myb.Ntl* 1 kb *Eco*RI/*Hind*III cDNA probe was complex (Figure 5.13). Approximately six hybridising fragments in addition to the two pollen-specific fragments were seen with *Eco*RI digestion. Of these two are major bands and minor bands. One of the major bands (4kb) is contributed from the *N. tomentosiformis* genome, while the other (6 kb) did not seem to be contributed from either of the parental genomes. Although an equivalent size band was

seen in the *N. sylvestris* genome this fragment contributes the 1.5 kb and 0.7 kb fragments in the *N. tabacum* genome. Two of the minor bands (3.2 kb and 3.0 kb) were contributed from the *N. tomentosiformis* genome. The two additional minor bands (8.5 kb and 8.0 kb) did not seem to be contributed from any of the parental genomes, although a band at 8.3 kb was present in the *N. tomentosiformis* genome. This suggests that gene amplification has occurred in the parental genomes after the original fusion had taken place which generated the *N. tabacum* genome. Alternatively, restriction fragment length polymorphism (RFLP) changes in the *N. tabacum* genome with time could account for the different banding patterns.

Approximately seven hybridising fragments in addition to the two pollen-specific fragments were seen with *Hind*III digestion. Of these one was major (1 kb) and six minor bands (8.5 kb, 8.0 kb, 5.5 kb, 5.0 kb, 3.2 kb and 0.5 kb). All the bands seem have been contributed from the *N. tomentosiformis* genome, although the major 1 kb band was probably a multiple band being composed of the contribution from the *N. sylvestris* genome in addition to the band from the *N. tomentosiformis* genome. The total number of hybridising fragments suggests an *N. tabacum myb* gene family size of approximately six to eight members. Weakly hybridising fragments suggest the presence of sequences distantly related to *myb.Nt1* and *myb.Nt2*.

A simpler pattern of hybridising fragments is seen with the 0.9 kb *Eco*RI/*Hind*III *Ntl1* cDNA probe (Figure 5.14). Approximately three bands in addition to the two pollen-specific fragments were seen with *Eco*RI digestion (6.0 kb, 2 kb and 0.75 kb). The 2 kb and 0.75 kb minor bands are contributed from the *N. tomentosiformis* genome. The 6 kb minor band did not seem to be contributed from either of the parental genomes, suggesting again that gene amplification has occurred in the parental genomes after the original fusion had taken place which generated the *N. tabacum* genome or RFLP changes has taken place in the *N. tabacum* genome, although background hybridisation is high in the adjacent *N. tomentosiformis* lane. Four additional minor bands were seen with *Hind*III digestion (3.5 kb, 3.0 kb, 1.7 kb and 1.4 kb). All four bands seem to be contributed from the *N. tomentosiformis* genome. A 2 kb major band, present in the *N. sylvestris* genome is not seen in the *N. tabacum* genome. The total number of hybridising fragments suggests an *N. tabacum lim* gene family size of approximately three to four members.

5.3 Discussion

5.3.1 *Myb.Nt1*, *myb.Nt2*, *Ntl1* and *Ntl2* are pollen-specific genes

143

Figure 5.11

High stringency Southern blot analysis of *myb.Nt* related sequences.

Image derived by scanning densitometry of total genomic DNA (7 μg per lane) from *N.tabacum (N. tob.)* and the parentals *N. sylvestris (N. syl.)* and *N. tomentosiformis (N. tom.)* digested with two restriction enzymes and hybridised at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1) with whole *myb.Nt1* cDNA.

The sizes indicated are in kilo bases. cDNA prepared from restiction digestion of a large scale plasmid preparation and serially diluted was used in the reconstruction lanes, accompanying labels indicate number of equivalent gene copies.

a. *Myb.Nt1*.

b. Reconstruction lane set.

c. *Myb.Nt2*.

d. Reconstruction lane set.

c.

6.0 —
4.0 /
3.8 —
3.5 \

1.5 —

0.7 —

d.   5 2 1 0.5

N.syl. N.tab. N.tom. N.syl. N.tab. N.tom.
¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯ ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯
      EcoRI   HindIII

a.

6.0 —
4.0 /
3.8 —
3.5 \

1.5 —

0.7 —

b.   5 2 1 0.5

N.syl. N.tab. N.tom. N.syl. N.tab. N.tom.
¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯ ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯
      EcoRI   HindIII

Figure 5.12

High stringency Southern blot analysis of *Ntl* related sequences.

Image derived by scanning densitometry of genomic DNA (7 μg per lane) from *N.tabacum (N. tob.)* and the parentals *N. sylvestris (N. syl.)* and *N. tomentosiformis (N. tom.)* digested with two restriction enzymes and hybridised at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1) with whole *Ntl1* cDNA..

The sizes indicated are in kilo bases. cDNA prepared from restriction digestion of a large scale plasmid preparation and serially diluted was used in the reconstruction lanes, accompanying labels indicate number of equivalent gene copies.

a. *Ntl1*.
b. Reconstruction lane set.
c. *Nlt2*.
d. Reconstruction lane set.

a.

8.5 —

6.0 —
5.0 —
4.0 —

b.

5   2   1   0.5

N.syl. N.tab. N.tom.
EcoRI

N.syl. N.tab. N.tom.
HindIII

c.

8.5 —

6.0 —
5.0 —
4.0 —

1.0 —

d.

5   2   1   0.5

N.syl. N.tab. N.tom.
EcoRI

N.syl. N.tab. N.to
HindIII

Figure 5.13

Low stringency Southern blot analysis of *myb.Nt* related sequences.

Image derived by scanning densitometry of genomic DNA (7 μg per lane) from *N.tabacum (N. tob.)* and the parentals *N. sylvestris (N. syl.)* and *N. tomentosiformis (N. tom.)* digested with two restriction enzymes and hybridised at low stringency (20 mM sodium phosphate, 5 % SDS, 50 °C; section 2.6.3.1) with whole *myb.Nt1* cDNA.

The sizes indicated are in kilo bases. Arrows indicate the presence of a hybridising band.

8.5 ——

6.0 —
5.0 —

4.0 —
3.8 ——
3.5 —
3.0 —

1.5 ——

0.7 ——

N.syl. N.tab. N.tom.
_____
EcoRI

N.syl. N.tab. N.tom.
_____
HindIII

Figure 5.14
Low stringency Southern blot analysis of *Ntl* related sequences.

Image derived by scanning densitometry of total genomic DNA (7 μg per lane) from
N.tabacum (N. tob.) and the parentals *N. sylvestris (N. syl.)* and *N. tomentosiformis*
(N. tom.) digested with two restriction enzymes and hybridised at low stringency (20
mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1) with whole *Ntl1* cDNA.

The sizes indicated are in kilo bases. Arrows indicate the presence of a hybridising
band.

8.5 —————

6.0 —————→

5.0 —————

4.0 —————

2.0 —————→

1.0 —————

0.75 —————→
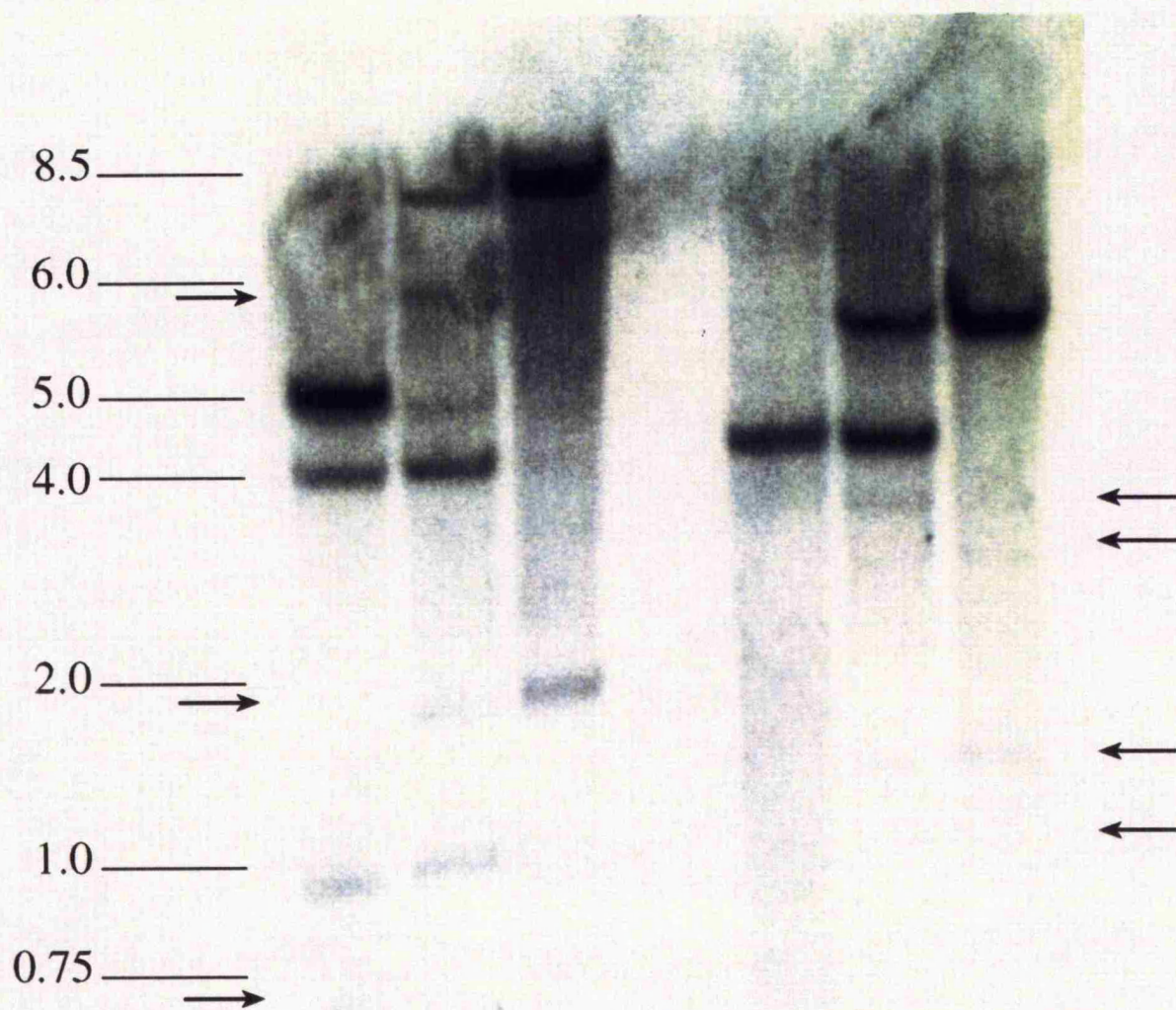
N.syl. _N.tab._ _N.tom._   _N.syl._ _N.tab._ _N.tom._
_Eco_RI   _Hind_III

The spatial pattern of expression of the *myb.Nt-* and *Ntl*-specific transcripts suggests that the *myb.Nt1*, *myb.Nt2*, *Ntl1* and *Ntl2* genes are transcribed in a pollen-specific manner. The *myb.Nt1* and *Ntl1* cDNA probes hybridise to transcripts 1800 and 900 nucleotides in length. These sizes are similar to the sizes of the cDNA inserts which suggests that the cDNAs are full length. Low level signal was detected in seedling (13 and 17 day), root, young leaf, sepal, petal and pistil tissues with the *myb.Nt1* probe. These signals may represent low level expression of the *myb.Nt* genes in these tissues or more possibly transcript from more distantly related *N. tabacum myb* genes. The presence of the highly conserved *myb* domain sequences in the probe fragment could potentially result in these low level signals being detected. The existence of a *myb* gene family in *N. tabacum* supports the later hypothesis (section 5.2.4), however no cross hybridisation to other *myb* genes was seen in the Southern blots at the same stringency. The uniformity of the signal in each lane suggests it may conceivably be due to the trapping of the probe in the abundant and similarly sized 18S rRNA.

Overexposure of the RNA blot probed with the *Ntl1* cDNA probe revealed low levels of signal in root and young leaf tissue. Again, this signal may be due to low level expression of the *Ntl* genes in these tissues or expression of distantly related *N. tabacum lim* genes. Such expression of root and young leaf *lim* genes may represent a developmentally early *lim* gene expression program in the young *N. tabacum* plant. The existence of a *lim* gene family in *N. tabacum* supports the later hypothesis (section 5.2.4).

Although the term 'pollen-specific' is applied when describing the spatial expression of the *myb.Nt* and *Ntl* genes, very low expression of these genes may be occurring in other tissues. Experiments using sensitive techniques such as *in situ* localisation of an antisense transcript specific probe may be used to further investigate the spatial accumulation of *myb.Nt1* and *myb.Nt2* transcripts in *N. tabacum* tissues. Such techniques have been used to localise transcripts of the *ntp303* gene (Reijen *et al.*, 1991).

More detailed analysis of the spatial expression patterns of previously characterised pollen-specific genes may reveal subtle differences to the established expression profiles. The difficulty of obtaining pure samples of anther tissue devoid of any contaminating pollen or microspores has led to problems in establishing anther- or pollen-specific expression by Northern analysis.

5.3.2 The *myb.Nt* and *Ntl* genes are expressed prior to the developmental switch pollen mitosis I

The expression profiles of *myb.Nt* and *Ntl* specific transcripts during pollen development is characterised by a low level pulse of transcript prior to pollen mitosis I (PMI) in late uninucleate microspores, a lag period immediately after PMI in early bicellular pollen when transcript levels decline almost to background levels followed by an abrupt rise in abundance at the mid binucleate pollen stage and accumulation of transcript to maximal levels in mature pollen.

Expression of the *myb.Nt* and *Ntl* genes initiates when the nursing tapetal cells of the anther are decaying (Koltunow *et al.*, 1990). This suggests that the expression of the *myb.Nt* and *Ntl* genes are regulated within the gametophyte itself and does not involve gene products from the surrounding sporophytic tissue (section 1.3.3). As the chromatin of the generative nucleus is highly condensed resulting in a low transcription level (LaFountain and Mascarenhas, 1972), the *myb.Nt* and *Ntl* genes will probably be expressed in the vegetative cell. The abundance increase of the *myb.Nt* and *Ntl* transcripts may reflect the massive cytoplasmic increase and nuclear activity of the vegetative cell (section 1.2).

Previous characterisation of the developmental expression patterns of other pollen-specific genes *ntp303* and *npg1* from *N. tabacum* ((Weterings *et al.*, 1992); (Tebbutt *et al.*, 1994)) had demonstrated a rapid rise in transcript levels after PMI with maximal levels in mature pollen, but had not detected transcript prior to PMI. These and other late pollen expressed genes are thought to be coordinately regulated (section 1.4.1). The expression patterns of these genes have been characterised in more detail revealing subtle expression prior to PMI, redefining the established late gene expression pattern in these cases. *Gus*-fusion analysis in transgenic plants using the promoters of the *Arabidopsis pal1* (Ohl *et al.*, 1990) and *tua1* (Carpenter *et al.*, 1992) genes have detected low levels of GUS activity prior to PMI in young and late microspores with much higher levels of activity accumulating after PMI until pollen maturity. A different pattern of pollen late gene expression again has been defined for the *p6* gene from *Oenothera* which does not accumulate until just before anthesis (Brown and Crouch, 1990). Further detailed studies of other late pollen expressed genes may also reveal differences to the established 'early' and 'late' expression profiles.

Expression of the *myb.Nt* and *Ntl* genes prior to PMI suggests that they might have a role in the regulation of late pollen expressed genes and thus the correct maturation of the pollen grain. A mechanism could be envisaged whereby accumulation of late gene transcription factors prior to PMI would result in activation and maintenance of late gene promoter activity. A number of late pollen expressed gene promoters have been sequenced which could be used to test their activation by *trans*-acting MYB.NT1 and

NTL factors in transient expression assays. Such putative regulatory factors might act in concert with other factors or as part of a regulatory cascade to activate late gene expression.

The activation of late pollen gene expression by the *myb.Nt* and *Ntl* genes may be connected with the asymmetric cell division PMI producing the generative cell and the larger transcriptionally more active vegetative cell within the confines of the original microspore cell wall. Production of transcription factors in the late uninucleate microspores could conceivably reach threshold levels at PMI. Exclusion of the factor from the generative cell pole would lead to activation of vegetative cell specific genes and induction of chromatin dispersion (Eady *et al.*, 1995).

The *lin-11* and *mec-3 lim* genes are expressed in one daughter cell as a result of asymmetric cell division in the nematode worm *Caenorhabitis elegans (C. elegans)* ((Ferguson *et al.*, 1987); (Way and Chalfie, 1988)). The *lin-11* gene is involved with vulval development in *C. elegans*. In *lin-11* mutants, the two daughter cells follow an identical fate and fail to form a functional vulva. Expression of the *mec-3* gene of *C. elegans* is required for the formation of mechanosensory neurons in one of two daughter cells of an asymmetric division. In the absence of functional MEC-3 the cells form but do not differentiate properly. Such asymmetric cell divisions require polarisation of proteins in the parent cell. Such evidence suggests the LIM domain might function in asymmetric cell division. Further, in the light of the LIM protein ZYXIN interacting with the actin cytoskeleton (Sadler *et al.*, 1992), the LIM domain might function by mediating protein segregation.

Based on the expression pattern of the *myb.Nt* and *Ntl* genes, a role in germination is possible, although transcript levels from both classes of genes decrease approximately four fold in germinating pollen. The use of specific inhibitors of transcription and translation has shown that germination and pollen tube growth of pollen occurs largely as a result of translation but independently of transcription (Hoekstra and Bruinsma, 1979). However, it is possible that the MYB.NT and NTL proteins are stored in the pollen grain for use in germination. Post-translational activation processes such as phosphorylation or glycosylation could then activate the factors once the germination process had begun.

A comparison of the relationship between transcript abundance and time with all six pollen expressed cDNAs, probably reflects the different patterns of transcript accumulation (Figure 5.5). The 'late' expressed *myb.Nt1*, *Ntl1*, *ntp303* and *ntg1* show the vast majority of transcript present in bicellular pollen cells, with maximal transcript levels in mature pollen. In contrast, an early pattern of transcription is

150

exemplified by *ntm19*, which shows a defined pattern of expression in microspores. A constitutive pattern of transcript accumulation is illustrated by *pma1*, with peaks of expression at late microspore and mid binucleate stages of development.

5.3.3 *Myb.Nt1*, *myb.Nt2*, *Ntl1* and *Ntl2* are single copy genes and form part of multi-gene families

Southern blot analysis was used to investigate the number of *myb.Nt* and *Ntl* related genes present in the *N. tabacum* genome. High stringency hybridisation produced patterns of hybridising fragments which suggested that both pairs of clones are single copy genes, with the *N.sylvestris* and *N. tomentosiformis* genomes contributing one gene each. The use of gene specific probes would identify which hybridising fragment corresponds to each of the pairs of genes and enable the exact determination of the origin of each of the genes.

Low stringency hybridisation was used to estimate the complexity of the *myb* and *lim* gene families in *N. tabacum*. The *myb.Nt1* and *Ntl1* genes belonged to families of approximately six to eight and three to four members respectively. Although most of the hybridising fragments in the *N. tabacum* genome corresponded to fragments from either the *N. sylvestris* or *N. tomentosiformis* genomes, the overall pattern of the fragments in each of the probings was too complex to enable accurate determination of the origin of the fragments in *N.tabacum*. In addition, most of the hybridising fragments were very weak suggesting these sequences are distantly related.

Previous estimates of *myb* gene family size in other plant species have been extremely variable. Thus while large gene families have been described in *Petunia hybrida* with 20-40 members and *Arabidopsis* with at least 100 members (Avila *et al.*, 1993) the moss *Physcomitrella patens* has an estimated gene family size of six or seven members (Leech *et al.*, 1993). An estimate of the *lim* gene family size in sunflower was approximately 10 members although clustering of hybridising bands made the estimation difficult (Baltz *et al.*, 1992a.).

5.4 Summary

Characterisation of two pairs of cDNAs encoding putative *myb* and *lim* transcription factor sequences revealed tight transcript accumulation in mature and germinating pollen. *Myb.Nt* and *Ntl* transcripts were first detectable prior to PMI in late uninucleate microspores, with a reduction in transcript levels immediately after PMI. A

dramatic rise in levels is seen in mid binucleate pollen cells with maximal levels in mature pollen grains. cDNA library screening allowed the estimation of a relatively high abundance for the *myb.Nt* and especially the *Ntl* putative transcription factor transcripts at maturity (0.05 % and 0.5 % respectively). These abundance estimates together with the spatial and temporal expression profiles suggest that the *myb.Nt* and *Ntl* specific genes perform an important function in pollen. Together with the *sf3* lim gene, the isolation the *myb.Nt1*, *myb.Nt2*, *ntl1* and *Ntl2* cDNAs represents the first examples of pollen-specific putative transcription factors in plants. The role of these *N. tabacum* pollen-specific genes was investigated *in planta* in transgenic *N. tabacum* plants by loss-of-function, overexpression and ectopic expression reverse genetics experiments presented in Chapter 6.

Chapter Six.

An investigation into the function of MYB.NT1
and NTL1 during pollen development.

6.1 Introduction and aims

Two classes of cDNAs encoding MYB and LIM proteins were isolated by heterologous hybridisation from an *N. tabacum* pollen cDNA library (section 4.2.1). The first clues as to the function of these clones came through sequence analysis, which revealed a high amino acid sequence homology between the MYB domains of the MYB.NT predicted proteins and the MYB domain of the MYB.PH3 protein from *P. hybrida* (section 4.2.3.1; Avila *et al.*, 1993). This homology suggested that the three proteins may bind similar DNA target sequences in the promoters of flavonoid biosynthetic genes. A high overall amino acid homology between the LIM domains of the NTL predicted proteins and those of a class LIM only proteins including SF3 from sunflower (section 4.2.3.2; Baltz *et al.*, 1992a.), MLP from rat (Arber *et al.*, 1994) and CRP from humans (Liebhaber *et al.*, 1990) suggested that the NTL proteins function through protein-protein interactions to regulate the function of other proteins. Investigation of the spatial and temporal expression of the *myb.Nt* and *Ntl* cDNA's in *N. tabacum* plants revealed a high expression in mature pollen (section 5.3). Low levels of transcripts were present in microspores prior to the asymmetric cell division pollen mitosis I, with an abrupt rise in transcript levels seen in mid bicellular pollen cell and levels increasing to pollen maturity. This expression profile suggested that the *myb.Nt* and *Ntl* genes are involved in the regulation of late pollen gene expression, either directly through DNA binding and transactivation from late pollen gene promoters, in the case of the MYB.NT proteins, or by a mechanism involving protein interactions such as protein segregation via the actin cytoskeleton, in the case of the LIM-domain containing NTL proteins.

It was decided to test these hypotheses and investigate the roles of the *myb.Nt* and *Ntl* genes directly by antisense and overexpression experiments *in planta*, using the cDNA sequences to alter the expression level and/or expression pattern of the regulatory proteins in transgenic *N. tabacum* plants. Analysis of any phenotypic effects would provide clues as to the function of the proteins. The use of transgene antisense suppression to decrease the steady-state level of a specific gene product has been used to assign functions to a number of proteins in plants ((Ecker and Davies, 1986); (Sheehy *et al.*, 1988); (van der Krol *et al.*, 1990a)). Although most of these studies have involved proteins with known function some have been used to elucidate the function of unknown proteins ((Hamilton *et al.*, 1990); (Bird *et al.*, 1991); (Stockhaus *et al.*, 1990)). The phenotypic effects seen in antisense studies range from a minimal decrease in activity to an almost complete inhibition of the target gene.

A second method which has been used to alter the level of steady-state RNA of a specific gene in plants is the addition of sense copies of an homologous transgene to increase the level of a particular RNA or restore a null phenotype. In a number of transgenic plant studies however the simultaneous sense suppression (cosuppression) of the endogenous and exogenous gene has also been observed ((van der Krol *et al.*, 1990b) and (Napoli *et al.*, 1990)).

Although the basis for antisense and sense gene suppression is not known, four mechanisms have been proposed. One consistent feature of antisense and sense gene suppression is the reduction of steady-state levels of the target gene poly A$^+$ RNA. It has been suggested that the mechanisms of antisense and sense mediated gene suppression are the same ((Grierson *et al.*, 1991), (Jorgenson, 1992)). Three models have been proposed to unite the mechanisms of sense and antisense gene-mediated suppression. The first proposes some sort of transient pairing of homologous DNA sequences which would result in an alteration of chromatin structure and the interruption of some aspect of gene transcription. A second involves antisense RNA-mediated gene suppression and relies on the specific base pairing of the transgene transcript with the target gene's transcript resulting in the degradation of the later. This could provide a mechanism for sense gene-mediated suppression if the sense transgene gives rise to unintended antisense transcripts from promoters within or external to the T-DNA insert. A third model involves RNA-DNA interactions in which either the sense or the antisense transgene transcript base pairs with the target gene. This could either directly or indirectly interfere with transcription.

In order to target antisense and sense *myb.Nt* and *Ntl* transcripts to pollen, the pollen-specific *lat52* gene promoter was used (Twell *et al.*, 1990). Analysis of the *lat52* temporal activity in transgenic *N. tabacum* plants detected low levels of GUS activity in anthers from 15 mm buds with levels dramatically increasing in buds upto 50 mm (anther dehiscence). This temporal expression profile matched that of the *myb.Nt* and *Ntl* genes (section 5.2.3) ensuring that antisense and sense transcripts will be produced at the same temporal time point to the endogenous transcript during pollen · development. The *lat52* promoter has been used to drive antisense expression of it's own transcripts in transgenic tomato plants which gave pollen with abnormal hydration and germination properties (Muschietti *et al.*, 1994). A recent study used the *lat52* promoter to drive antisense expression of the *Bcp1* gene, known to be required for male fertility in *Arabidopsis*, which showed 50 % mature pollen abortion (Huiling *et al.*, 1995).

If a pollen expressed gene plays an essential role during pollen development, an altered pollen phenotype should be found in the transgenic plants, from which a function may be deduced. As pollen genes are expressed gametophytically, only 50 % of the pollen of primary transformants will be expected to express the antisense/sense construct. Distortion of Mendelian segregation ratios observed from a process of self-crossing or back-crossing of the transgenic plants and kanamycin resistance gene segregation analysis would determine which transgenic plants are showing an effect on pollen viability and whether transmission is affected through the male or female gametes.

The ectopic overexpression of sense transcripts of the *myb.Nt* and *Ntl* genes in other plant tissues was also used in order to generate a phenotype which could give clues as to the function of the gene products. For this purpose the constitutive *cauliflower mosaic virus (CaMV) 35S* promoter was used, which has been shown to drive high level expression of foreign genes in most plant tissues (Odell *et al.*, 1985), although expression in pollen is low (Twell *et al.*, 1989b). The *CaMV 35S* promoter has been used to express a number of foreign genes in plants in order to study the protein function or to generate plants with novel characteristics for commercial applications ((Kang *et al.*, 1995); (Morelli *et al.*, 1985); (Bevan *et al.*, 1985)). A duplicated *CaMV 35S* promoter sequence was used to drive high level expression of *myb.Nt* and *Ntl* genes in *N. tabacum*. This duplication was shown to increase levels of expression of plant genes (Kay *et al.*, 1987).

N. tabacum has been used widely in studies of plant gene expression, primarily because leaf cells can be transformed easily with foreign DNA and subsequently regenerated into whole fertile plants. For this reason, *N. tabacum* was one of the first plants to be genetically engineered (Barton *et al.*, 1983). Although *N. tabacum* is a amphidipliod, its behaves as a diploid with respect to inheritance of foreign genes. A common method and one which was used for the introduction of the *myb.Nt1* and *Ntl1* genes into *N. tabacum* uses vectors based on the transferred T-DNA region of *Agrobacterium tumefaciens* (Ti) inducing plasmid in conjunction with the leaf-disc method (section 2.4.2; (Horsch *et al.*, 1985)). In this method, *N. tabacum* leaf pieces are incubated with a suspension of *Agrobacterium* cells containing a modified Ti plasmid that includes an antibiotic selection marker able to function in plant cells. Single leaf cells which have received the selectable marker gene and the test gene positioned between the T-DNA borders are able to divide and differentiate on a medium containing the appropriate antibiotic and a plant hormone to induce shoot regeneration.

In vivo transient assays were also used to investigate the function of the *myb.Nt* genes. Here, two constructs, an effector and a reporter construct were introduced into *N. tabacum* leaf cells. The effector construct comprised the *myb.Nt1* cDNA sequence linked to the constitutive *CaMV 35S* promoter. The reporter construct consisted of a reporter gene (β-glucuronidase or luciferase) linked to one of a number of pollen-specific promoters which were available for use. These constructs were introduced into leaf cells by microprojectile bombardment. In this way, transient expression of the MYB.NT1 protein factor might be expected to *trans*-activate from appropriate *cis*-element sequences present in the pollen-specific promoters in the reporter construct, whose activity can be monitored.

In summary, the main aim of the work described in this chapter is the deduction of a role for the *myb.Nt* and *Ntl* genes by analysing potential phenotypes generated in pollen and other tissues of *N. tabacum* by the use of antisense, sense and ectopic overexpression reverse genetics experiments.

6.2 Results

6.2.1 Construction of sense and antisense constructs for the modified expression of *myb.Nt* and *Ntl* genes in pollen

Full length *myb.Nt1* cDNA was cloned into the binary vector pMOG402 (Mogen; Figure 6.1a.). This vector has the multi-cloning site positioned at the right T-DNA border (RB), ensuring that the transgene cassette will be transferred into the plant genome before the kanamycin selectable marker gene (RB to LB insertion). Kanamycin resistant plants are therefore likely to contain an intact integrated transgene cassette. Full length *myb.Nt1* cDNA was isolated by *SstI/KpnI* restriction digestion of a pBluescript construct originally prepared from bacterial colonies resulting from the *in vivo* excision of cDNA from the λZAPII XR vector during cDNA library screening (section 4.2.1.1). The 1.8 kb fragment was cloned into the *SacI/KpnI* sites of the vector *pSL301*(Brosius, 1989). This subcloning step enabled an *XbaI* restriction site to be picked up at the 3' end of the cDNA. *XbaI* restriction digestion of the *pSL301-myb.Nt1* subclone generated a fragment which was cloned into the equivalent site between a *lat52* (-492/+110) pollen-specific promoter and the *CaMV 35S* 3' polyadenylation sequence. In this way the *myb.Nt1* cDNA was cloned in sense and

Figure 6.1

Construction of *lat52 myb.Nt1* and *Ntl1* sense and antisense constructs.

Schematic representation of the whole *Myb.Nt1* and *Ntl1* cDNA and the strategy used for cloning in sense and antisense orientations into pMOG402 (MOGEN). All constructs are represented as circular vector maps in Appendix1. cDNA orientations are indicated with arrows.

a. *Myb.Nt1* cloning strategy.
b. *Ntl1* cloning strategy.

Abbreviations: *CaMV, cauliflower mosiac virus*; LB and RB, left and right border T-DNA repeats; *NptII*, T-DNA kanamycin resistance gene; *nos*, nopaline synthase.

158

a.

*Hind*III
*Sst*1   *Xba*1   *Kpn*I

Myb.Nt 1
pBluescript SK+

*Sst*I/*Kpn*I fragment
into *Sac*I/*Kpn*I cut
pSL301

*Sst*I   *Xba*I   *Hind*III *Kpn*I   *Xba*I

*Xba*I fragment into
phosphatased *Xba*I
cut

*Hind*III
*Sst*I   *Xba*I   Sense   *Kpn*I *Xba*I
*Hind*III
lat52   Antisense   *CaMV*   pUC19
promoter   *35S 3'*

*Sst*I/*Hind*III partial
fragment into
*Sac*I/*Hind*III cut
pMOG402

*Hind*III
NptII   *Sst*I   Sense   *Hind*III

LB *Nos3'*   *Nos5'*   Antisense   RB

b.

*Sst*I   *Xba*I   *Kpn*I

Nt1
pBluescript SK+

*Sst*I/*Kpn*I fragment
into *Sac*I/*Kpn*I cut
pSL301

*Sst*I   *Xba*I   *Kpn*I   *Xba*I

*Xba*I fragment into
phosphatased *Xba*1
cut

*Xba*I   *Kpn*I   *Xba*I
*Sst*I   Sense   *Hind*III
lat52   Antisense   *CaMV*   pUC19
promoter   *35S 3'*

*Sst*I/*Hind*III frgment
into *Sac*I/*Hind*III cut
pMOG402

*Hind*III
NptII   *Sst*I   Sense

LB *Nos3'*   *Nos5'*   Antisense   RB

antisense orientations to the *lat52* promoter and ensure its pollen-specific expression in transgenic plants. The orientations of the constructs was determined by restriction digestion of plasmids recovered from transformed *E. coli*. Due the presence of an internal *Hind*III site (nucleotide position 803) in the *myb.Nt1* cDNA coding sequence, SstI/partial *Hind*III restriction digestion was performed to generate a 2.7 kb fragment, containing the lat52 promoter, *myb.Nt1* cDNA in sense and antisense orientations and the *CaMV 35S* polyadentlation signal, which was cloned into the SacI/*Hind*III sites in pMOG402. The cloning procedure used to generate the *lat52-Ntl1* cDNA sense and antisense T-DNA constructs was essentially the same as that used for *myb.Nt1* except that the last subcloning step into pMOG402 did not require partial *Hind*III digestion, as the *Ntl1* cDNA did not contain any internal sites (Figure 6.1b.).

The resulting pMOG402 constructs contained two cassettes from the right to the left border of the T-DNA; the cassette for sense and antisense expression of *myb.Nt1* and Ntl1 controlled by the *lat52* pollen-specific promoter sequence and *CaMV 35S* polyadenylation sequences and the *nptII* gene cassette for kanamycin resistance in plant tissues controlled by the nopaline synthase *(nos)* promoter and polyadenylation sequences. The binary vectors were transferred to *Agrobacterium tumefaciens* LBA4404 containing the pAL4404 disarmed Ti plasmid (section 2.3.7; (Hoekema *et al.*, 1985)). The integrity of the vectors were verified in *Agrobacterium* by restriction digestion and agarose gel electrophoresis.

6.2.2 Construction of sense constructs for ectopic expression of *myb.Nt1* and *Ntl1* in other plant tissues

In order to produce the T-DNA constructs used for ectopic overexpression of *myb.Nt1* and *Ntl1* transcripts in *N. tabacum* tissues, a different cloning strategy was adopted (Figure 6.2). A 0.942 kb SalI/*Bam*HI fragment obtained by restriction digestion of the pRTL2 vector (N. Bate, Leicester), and composed of the double *CaMV 35S* (-343/+9 and -343/-90 fragments; (Kay *et al.*, 1987)) and the *N. tabacum* etch virus 5' translational enhancer (144 bp; (Carrington and Freed, 1990)). This fragment produces a strong constitutive promoter for overexpression of protein in plant tissues. The SalI/*Bam*HI fragment was cloned into the respective sites of the *lat52-myb.Nt1/Ntl1* sense constructs. In this way, the *lat52* promoter was removed and replaced with the duplicated *CaMV 35S/tev* fragment. These cassettes were then removed by *Hind*III restriction digestion and cloned into the respective site on

159

Figure 6.2

Construction of *myb.Nt1* and *Nt11* ectopic overexpression sense constructs.

Schematic representation of the whole *Myb.Nt1* and *Nt11* cDNAs and the strategy used for cloning in sense orientation into pMOG402 (MOGEN). All constructs are represented as circular vector maps in Appendix 1.
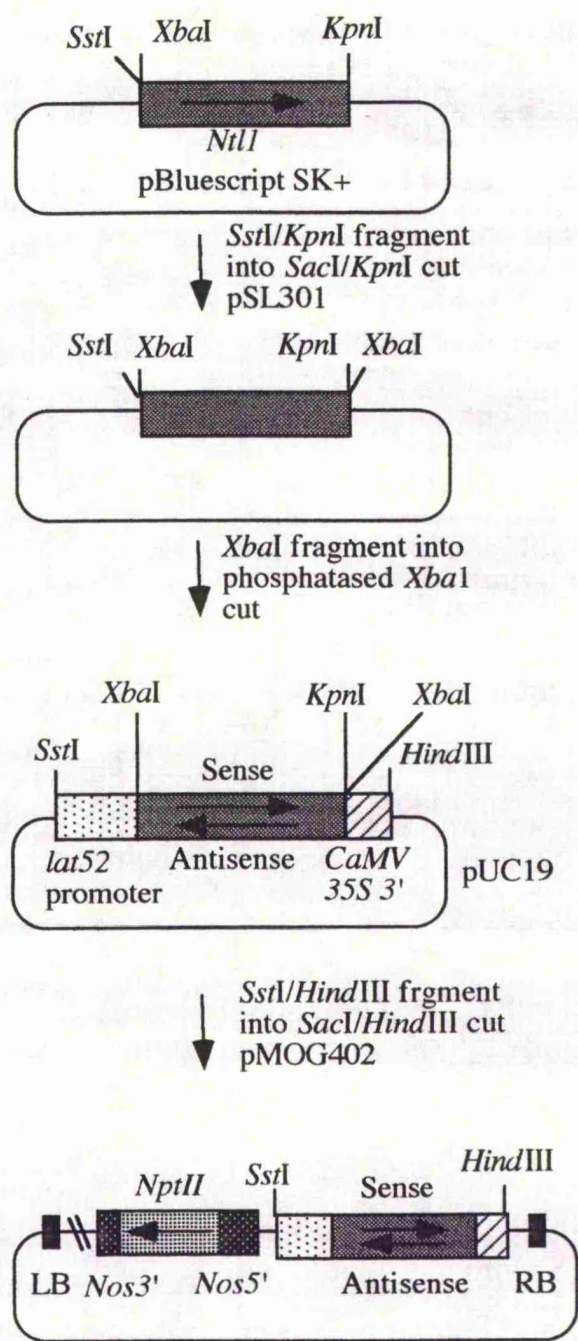
cDNA orientations are indicated with arrows.

Abbreviations: *CaMV, cauliflower mosiac virus*; LB and RB, left and right border T-DNA repeats; *NptII*, T-DNA kanamycin resistance gene; TEV 5' enhancer, *N. tabacum* etch virus 5' untranslated translational enhancer; *nos*, nopaline synthase.

SalI HindIII

BamHI

2x35S   TEV 5'
CaMV 5'   enhancer
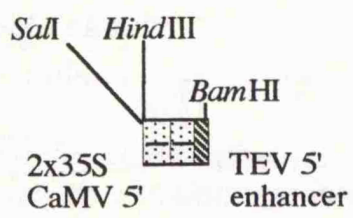
SalI/BamHI fragment
into SalI/BamHI cut
pMcNtl-1/pMcMyb.Nt1
sense constructs

SalI HindIII

BamHI   Sense   HindIII

myb.Nt1/Ntl1   CaMV
               35S 3'
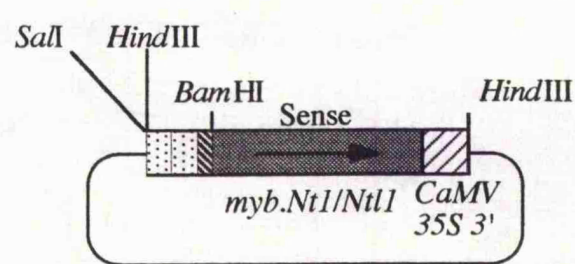
HindIII cut fragment
into phosphatased
pMOG402

HindIII   Sense   HindIII

LB   Nos3' NptII   Nos5'   RB
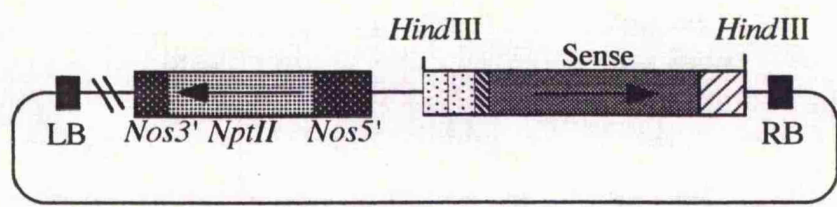
pMOG402. Partial *Hind*III digestion was required in the case of the *myb.Nt1* cassette due to the presence of the internal *Hind*III site in the *myb.Nt1* coding region.

Restriction digestion and gel electrophoresis was used to verify the integrity and orientation of the T-DNA vectors. The resulting pMOG402 constructs contained two cassettes from the right to the left border of the T-DNA; the cassette for sense expression of *myb.Nt1* and *Nt11* controlled by the *CaMV 35S* duplicated promoter sequence and *CaMV 35S* polyadenylation sequences and the nptII gene cassette for kanamycin resistance in plant tissues controlled by the nopaline synthase *(nos)* promoter and polyadenylation sequences. The binary vectors were transferred to *Agrobacterium tumefaciens* LBA4404 as before. The integrity of the vectors were verified in *Agrobacterium* by restriction digestion and agarose gel electrophoresis.

All six strains of *Agrobacterium* containing one of the T-DNA constructs described were used to transform *N. tabacum* (section 2.4.2). Transformed plantlets were transferred from MS0 media (section 2.4.1) to soil and grown under greenhouse conditions until flowering. The plants were allowed to self pollinate, transgenic pollen used to backcross to wild type plants and wild type pollen used to backcross to transgenic plants. F_1 seeds were harvested and selected on germination medium containing kanamycin (section 2.4.3). Leaf material was collected from plants containing the *CaMV 35S myb.Nt1/Nt11* constructs and stored at -80 °C until further use.

6.2.3 Characterisation of transgenic plants by kanamycin resistance segregation analysis and PCR

A total of 117 primary transformant (T_0) kanamycin resistant *N. tabacum* plants were generated in two co-cultivation experiments. In order to score for transmission of the *myb.Nt1* and *Nt11* sense and antisense genes, transmission of the linked kanamycin resistance gene was tested by germinating the F_0 seeds in medium containing kanamycin (Table 6.1). Inheritance of kanamycin resistance could be monitored by plating seeds from selfings and outcrosses on medium containing the antibiotic and scoring seedlings with dark green leaflets and long roots (resistant) and lighter green or bleached cotyledons and short roots (sensitive) (Potrykus *et al.*, 1985). Analysis of these crosses enabled classification of the plants into seven classes. Inclusion of transgenic plants into one of the inheritance classes I to VI was determined on the basis that at least two of the three segregation ratios fitted or deviated from the

161

Table 6.1

Numbers of transgenic *N. tabacum* plants containing one of six T-DNA constructs falling into six kanamycin resistance segregation classes (I-VI), and the percentage of second generation sensitive seedlings from selfcrosses and outcrosses through the transgenic male.

Kanamycin resistance segregation classes:

I Single dominant locus (1: 1, 3 : 1, 1 : 1).

II Two independent dominant loci (3 : 1, 15 : 1, 3 : 1).

III All resistant (1 : 0, 1 : 0, 1: 0).

IV All sensitive (0 : 1, 0 : 1, 0 : 1).

V Anomalous ratios (1 : 1, 3 : 1, 1 : 1).

VI Reduced transmission through the transgenic male gametophyte (RTM; 1 : 1, 1 : 1, 0 : 1).

For table of seedling counts see Appendix I.

The expected ratios are shown for each class;

Results that fitted the expected Mendelian ratios did so with at least a 95 % probability level ($\chi^2 >= 3.84$). Inclusion in the Mendelian classes of inheritance (I-VI) was on the basis that the three ratios fitted the expected ratios with 95 % probability.

All seeds germinated with a frequency of greater than 95 %.

T-DNA constructs	I	II	III	IV	V	VI	Second generation VI: % sensitive seedlings	
							selfcross	outcross
Lat52 myb.Nt1 sense	22	6	1	4	3	1	39	64
Lat52 myb.Nt1 antisense	9	1	1	2	2	3	36 38	65 57
CaMV 35S myb.Nt1 sense	7	1	-	-	1	-	-	-
Lat52 Ntl1 sense	11	5	-	-	2	2	36	56
Lat52 Ntl1 antisense	11	5	-	1	1	2	-	-
CaMV 35S Ntl1 sense	10	2	-	-	1	-	-	-

Classes

expected Mendelian ratio with greater than 95 % or 99 % probability (p < 0.05 or p < 0.01), as determined by chi-square tests.

The majority of plants containing one of the six T-DNA constructs (*lat52-myb.Nt1* sense, *lat52 myb.Nt1* antisense, *CaMV 35S myb.Nt1* sense, *lat52-Ntl1* sense, *lat52 Ntl1* antisense and *CaMV 35S Ntl1* sense) showed segregation ratios that indicated inheritance of the kanamycin resistance phenotype in a Mendelian manner. Class I plants showed the expected pattern for a single dominant gene (3 resistant : 1 sensitive selfcross, 1: 1 in both outcrosses). Class II plants showed evidence for two independent insertions of a kanamycin resistance gene locus (15 : 1 selfcross, 3 : 1 outcrosses). Two plants (*lat52-myb.Nt1* sense plant 32 and *lat52-myb.Nt1* antisense plant 18), showed entirely kanamycin resistant progeny in both selfcross and both outcrosses suggesting multiple independent insertions (class III).

All plants exhibiting non-Mendelian ratios were grouped into one of three classes. Class IV produced no kanamycin resistant progeny. Class V plants showed non-Mendelian ratios for all three crosses which could not be explained in terms of inheritance of independent loci. Lastly, class VI plants showed observed ratios that were similar to those expected of a reduced transmission through the male line (RMT) phenotype; 1 : 1 selfcross and backcross through the transgenic female and 0 : 1 backcross through the transgenic male. These ratios supported the idea that the kanamycin resistance gene was not transmitted normally through the male line. These plants consisted of one lat52-*myb.Nt1* plant (number 10) which showed a pollen lethal phenotype and three *lat52-myb.Nt1* antisense plants (numbers 6, 15 and 16), one *lat52-Ntl1* sense plant (number 19) and two *lat52-Ntl1* antisense plants (numbers 5 and 18) which showed ratios similar to the RMT phenotype, as determined by chi-square tests.

The presence of the expected transgene cassette in the kanamycin resistant F_0 plants was confirmed by using the polymerase chain reaction (PCR) on leaf tissue collected from the plants containing one of the six T-DNA constructs (section 2.10.3). *Lat52* and *tev 5'* primers were used in conjunction with one of the *myb.Nt1* and *Ntl1* sequencing primers (reverse or forward primers were used depending on the orientation of the cDNA insert). In most cases diagnostic sized bands were obtained for each of the transgenic plants, confirming the presence of the intact transgene and promoter. However the intensity of the PCR amplified bands did not correlate with the estimation of T-DNA loci number obtained from kanamycin resistance analysis.

6.2.4 Second generation plants expressing *myb.Nt1* and *Nt11* from the *lat52* promoter but not from the *CaMV 35S* promoter show a loss of viability of pollen in transgenic *N. tabacum*

Heterozygote kanamycin resistant seedlings of the six primary transformant plants exhibiting an RTM phenotype, obtained from outcrosses through the transgenic female of the primary transformant plant, were grown to maturity in soil. Self- and outcrosses through the transgenic male and female were made. Seed from one of the F_1 plants from each of the transgenic lines was collected and germinated on medium containing kanamycin. In this way, the F_1 plant was tested for the stable inheritance of the RTM phenotype (Table 6.1). Chi square tests on the F1 segregation data revealed that *lat52 myb.Nt1* sense plant line 10, *lat52 myb.Nt1* antisense plant lines 15 and 16 and *lat52 Nt11* sense plant line 7 inherited the RTM phenotype from the parents (ie. primary transformants showing a RTM phenotype again gave a RTM phenotype in the F_1 progeny). Selfcrosses and outcrosses through the male indicated significant deviation from the Mendelian ratios in at least one of the ratios with at least a 95 % probability (p < 0.05). However, the severity of the RMT phenotype seen in the primary transformant *lat52 myb.Nt1* sense plant number 10 was significantly reduced in the F_1 generation. Further, *lat52 myb.Nt1* sense plant lines 11 and 12, *lat52 myb.Nt1* antisense plant line 4, *lat52 Nt11* sense plant line 19 and *lat52 Nt11* antisense plant lines 5 and 18 did not show a statistically significant bias from the Mendelian ratios (p > 0.05) in the F_1 generation, despite showing such a bias in the primary transformant plants. Pollen from the F_1 plant lines exhibiting the stable RTM phenotype resulting from a self-cross were used for subsequent analysis.

6.2.5 *Myb.Nt*-related transcript was down-regulated in pollen of transgenic *N. tabacum* plants containing *Lat52-myb.Nt1* sense/antisense constructs

In order to test whether the aberrant genetic ratios could be correlated with downregulation of the *myb.Nt1* and *Nt11* genes, levels of *myb.Nt* and *Nt1* endogenous transcripts and transgene transcripts were analysed (Figure 6.3 Total RNA extracted from mature pollen of three heterozygote F_1 plants from each of the four transgenic plant lines exhibiting the inherited RTM phenotype and two wild type control plants

Figure 6.3

Northern blot showing downregulation of *myb.Nt* sense transcript in pollen from transgenic and wild type plants.

Image derived by scanning densitometry of pollen total RNA (10 μg per lane) extracted from transgenic *N. tabacum* plants hybridised at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. RNA from pollen from transgenic and wild type control plants probed with *CaMV 35S* 3' sequence, the image in lane 2 corresponds to pollen RNA from *Ntl1* S#7, *lat52 Ntl1* sense plant 7.

b. RNA from pollen from transgenic and wild type control plants probed with whole *myb.Nt1* and *Ntl1* cDNA (see d. for lane loadings).

c. 18S rRNA loading control.

d. Graph representing relative steady-state *myb.Nt1/Ntl1* mRNA abundance, all values were normalised against those of the loading control. Each bar represents normalised quantification of the lane image in b. directly above.

Abbreviations: *myb.Nt1* S#10, *lat52 myb.Nt1* sense plant 10; *myb.Nt1* AS#15, *lat52 myb.Nt1* antisense plant 15; *myb.Nt1* AS#16, *lat52 myb.Nt1* antisense plant 16; *Ntl1* S#7, *lat52 Ntl1* sense plant 7; wt, wild type control plant.
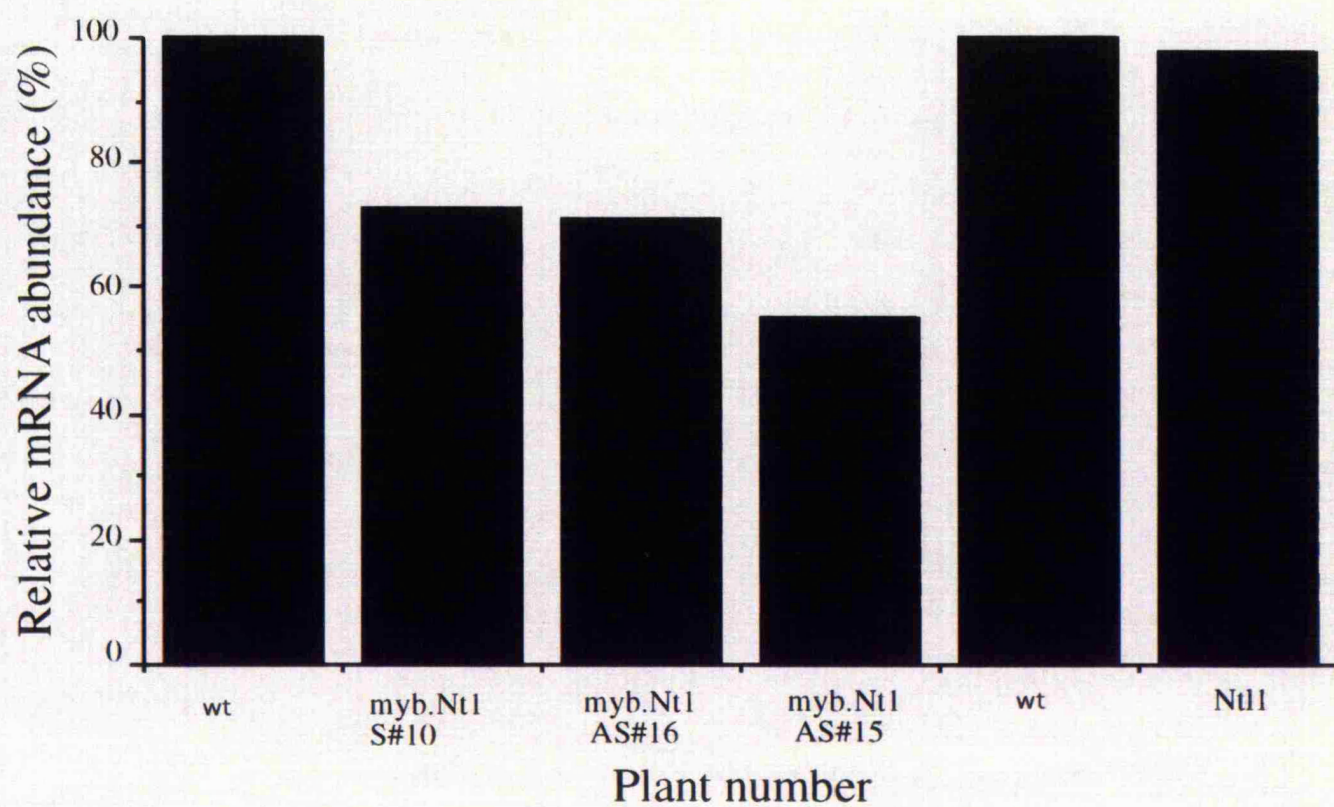
a.

b.

c.

d.

Relative mRNA abundance (%)

100

80

60

40

20

0

wt | myb.Nt1 S#10 | myb.Nt1 AS#16 | myb.Nt1 AS#15 | wt | Ntl1

Plant number

were analysed by Northern blot hybridisation. Whole *myb.Nt1* cDNA was used as a probe to RNA from wild type and *lat52 myb.Nt1* sense plant line 10, *lat52 myb.Nt1* antisense plant lines 15 and 16. Whole *Nt11* cDNA was used as a probe to RNA from wild type and the *lat52 Nt11* sense plant 7. Signals were quantified by scanning densitometry and values normalised against 18S rRNA signal.

The three *lat52 myb.Nt* plant lines showed a specific reduction in endogenous *myb.Nt1* transcript (1.8 kb) compared to a wild type control (72 %, 70 % and 56 % decrease in *Lat52 myb.nt1* sense plant line 10 and *lat52 myb.Nt1* antisense plant lines 15 and 16 respectively; panel b. Figure 6.3)), suggesting a link between downregulation of the *myb.Nt* message and the genetic data. In contrast, the *lat52 Nt11* sense plant line showed no increase in endogenous *Nt11* transcript (0.9 kb) compared to a wild type control. However, a larger and more abundant 1.2 kb transcript corresponding in size to the transgenic transcript of the *Nt11* cDNA and *CaMV 35S* 3' polyadenylation signal sequence was present.

A *CaMV 35S* 3' sequence was used as a probe to the same Northern blots (panel a. Figure 6.3). No transcript was detected in RNA from wild type controls or the *lat52 myb.Nt1* transformed plants. A relatively abundant 1.2 kb transcript corresponding to the *Nt11* transgene transcript was detected in RNA from the *lat52 Nt11* sense plant line 7.

6.2.6 An investigation into the phenotype of pollen from transgenic *N. tabacum* plants containing *Lat52-myb.Nt1* sense/antisense constructs

The pollen phenotype of the transformed plants showing the RTM phenotypes in the F_1 generation was examined. Mature pollen from the samples used in Northern analysis above (section 6.2.5) from heterozygote F_1 plants (*lat52 myb.Nt1* sense plant line 10, *lat52 myb.Nt1* antisense plant lines 15 and 16 and *lat52 Nt11* sense plant line 7) was hydrated in 0.3 M mannitol. Two phenotypic classes of pollen, large and round and small and collapsed, were noticed for each of the transgenic plants and wild type control plants with light microscopy. While the proportion of the small and collapsed pollen grains was higher in pollen from the *lat52 myb.Nt1* sense plant line 10 and *lat52 myb.Nt1* antisense plant lines 15 and 16 than wild type pollen, proportions of normal and collapsed were similar to wild type in *lat52 Nt11* sense plant 7 (Figure 6.4).

166

Figure 6.4

Bar chart showing proportions of large and round and small and collapsed pollen grains in R_1 pollen samples from four plant lines exhibiting the RMT phenotype.

Abbreviations: LR, large and round pollen grains; SC, small and collapsed pollen grains.

Pollen sample
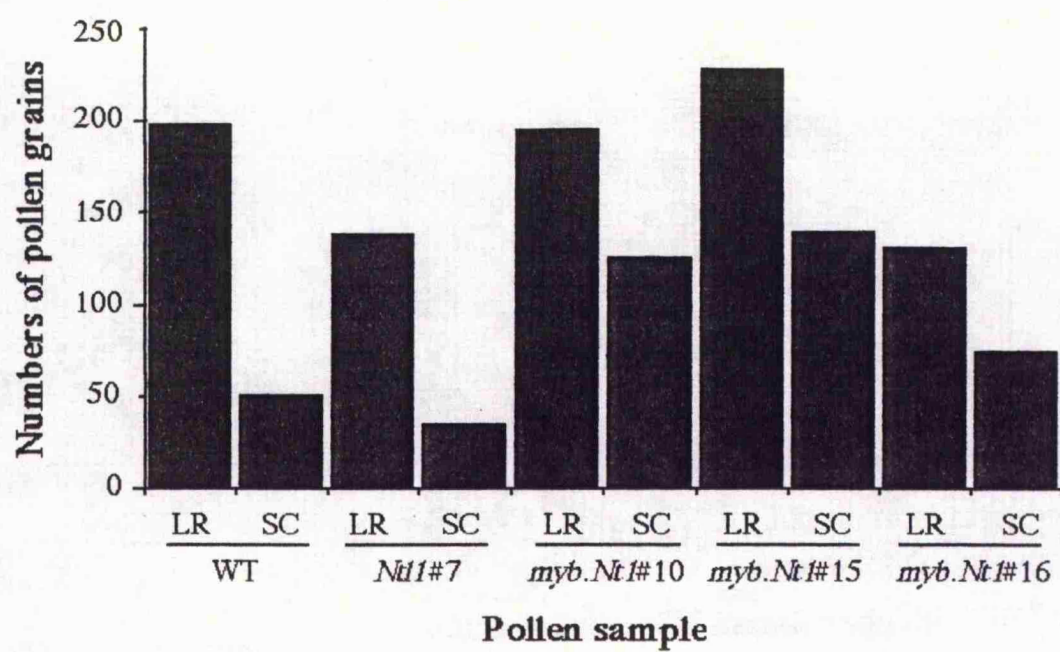
Cytochemical stains were used to investigate possible dysfunction of the small and collapsed class of pollen grains. DAPI binds to double-stranded DNA and fluoresces blue under ultra violet light (section 2.12.3; (Regan and Moffatt, 1990)). The normal looking pollen grains from each of the four transgenic plants and wild type control plants gave bright nuclear staining. In pollen samples from each of the plants vegetative nuclei were larger with more diffuse staining while generative nuclei were elongated in shape and stained more brightly. However the collapsed pollen grains from each of the plants gave very faint and diffuse staining or no staining at all indicating degradation or absence of DNA and therefore cell death.

Two methods of vital staining was used, Alexanders stain and fluorescein diacetate (FDA; sections 2.12.1 and 2.12.2 respectively). Alexanders stain (Alexander, 1969) is a stain composed of four dyes that stain nuclear material and cell walls. Therefore, living pollen grains stained purple and dead pollen a turquoise. The normal pollen grains from each of the four transgenic plants and wild type control plants stained purple and the collapsed grains in each of the plants turquoise indicating the two classes of pollen were alive and dead respectively.

FDA staining is the basis of the flurochromatic reaction involving esterase cleavage of non-fluorescent FDA to release fluorescein (Heslop-Harrison *et al.*, 1984). Levels of fluorescein in the normal pollen grains from the transgenic plants and wild type control plants were high, whereas levels was low in collapsed grains. The low level of fluorescence seen in the collapsed grains was probably due to exine auto-fluorescence as fragments of exine gave similar fluorescence levels to the collapsed grains.

In vitro germination of pollen from each of the transgenic plants and wild type control plants was carried out. Dessicated pollen grains were resuspended in pollen germination medium and germinated for 60 minutes (section 2.12.4). The round pollen grains germinated normally compared to wild type pollen. However, the collapsed pollen grains did not germinate, supporting the evidence from cytochemical staining that these grains were dead.

6.2.7 Expression of the *myb.Nt1* from the *CaMV 35S* promoter but not from the *lat52* promoter causes a dwarf phenotype

Figure 6.5
Primary transgenic *CaMV 35S myb.Nt1* sense *N.tabacum* plants with height at flowering.

a. Comparison of height of *CaMV 35S Myb.Nt1* sense transgenic *N. tabacum* plants at flowering. 200 mm is represented by the large black bar at the bottom of the scale.
b. Bar chart showing height in cm. Each bar represents the height of the plant directly above. Height was measured as the distance from the base of the stem to the tip of the first flower. Each plant was grown under similar greenhouse conditions. Wild type plants grown under the same conditions grew to an average of 109cm.

a.

b.

Height at flowering (cm)

CaMV 35S myb.Nt1 sense plant number
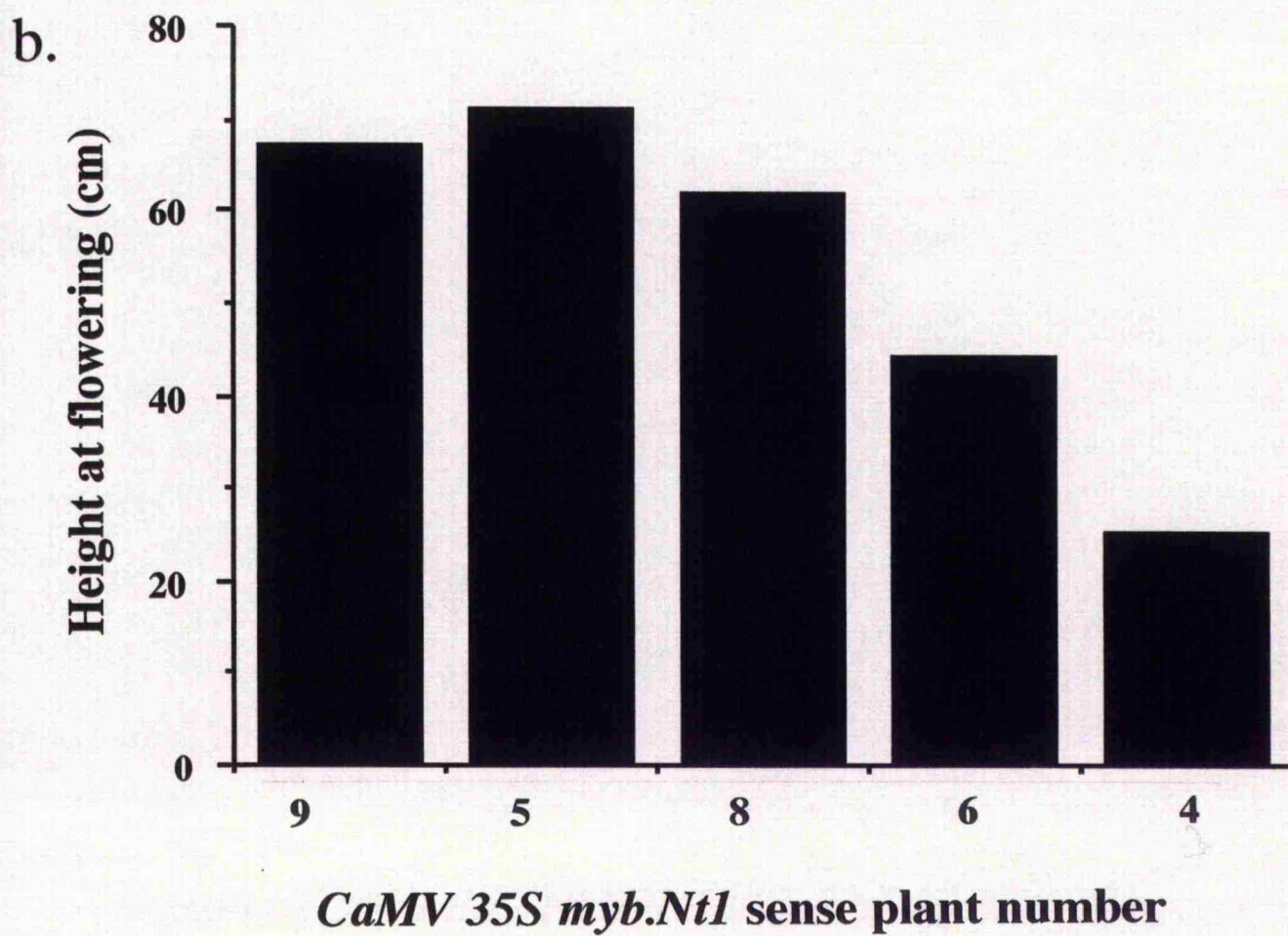
All of the regenerated transgenic *N. tabacum* plants had normal sporophytic growth habits and phenotypic appearance except for the majority (6/8) of the plants containing the *CaMV 35S-myb.Nt1* sense construct, which displayed varying degrees of dwarfism (plant numbers 3, 4, 5, 6, 8 and 9; Figure 6.5). Kanamycin resistance segregation analysis of the dwarf phenotype plants established that plants 4, 5, 6, 8 and 9 had a single T-DNA insertion and plant number 3 two independent T-DNA insertions. Microscopic examination of pollen from these plants showed a normal phenotype.

Plant numbers 4 and 6 were 26 cm and 45 cm at flowering. Plants 8, 9 and 5 reached heights of 62 cm, 68 cm and 72 cm respectively. In comparison, wild type *N. tabacum* plants of the same age and grown under the same conditions grew to 109 cm. The severe dwarf plants, 4 and 6, showed a significant reduction in internodal stem growth, while leaves had a more rounded shape in comparison to wild type leaves of approximately the same age and position. Onset of flowering was reduced in the severe dwarf plants 4 and 6. Flowers had a wild type morphology with normal levels of pollen at anthesis, but were reduced in size. In addition, the leaves of plant number 5 showed a distinct morphology from the rest of the dwarf plants, having an 'inverted spoon'-like appearance.

6.2.8 The *myb.Nt1* transcript is produced in leaves of plants exhibiting the dwarf phenotype

RNA blot analysis was conducted on five of the F$_1$ *CaMV 35S myb* sense plants (plants 4, 5, 6, 8 and 9). In addition two wild type control plants were used, one of which was *CaMV 35S myb* sense plant number 7, which had showed a kanamycin sensitive genotype, suggesting that this plant was untransformed (Figure 6.6). RNA was extracted from a young leaf of approximately the same age from each of the plants. The RNA was separated by agarose gel electrophoresis and blotted onto a membrane. The blot was probed under high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C) with the *myb.Nt1* cDNA. A transcript of approximately 2.0 kb was detected in all of the dwarf plants. This transcript size corresponded to the expected size of the *myb.Nt1* transcript (1.85 kb) and the *CaMV 35S* polyadenylation signal (0.22 kb). No transcript was detected in either of the wild type control lanes. Transcript levels were quantified by scanning densitometry and values normalised against the 18S rRNA band.

Figure 6.6

Northern blot analysis of transgenic *CaMV 35S myb.Nt1* **sense** *N.tabacum* **plants.**

Image derived by scanning densitometry of total RNA (10 μg per lane) from a young leaf of *CaMV 35S Myb.Nt1* sense transgenic *N.tabacum* plants, hybridised with whole *Myb.Nt1* cDNA at high stringency (20 mM sodium phosphate, 1 % SDS, 65 °C; section 2.6.3.1).

a. Image equivalent to overnight autoradiographic exposure.

b. 18S rRNA loading control.

c. Graph representing relative steady-state *myb.Nt1* mRNA abundance, all values were normalised against those of the loading control. Each bar represents quantification of the lane image in a. directly above.
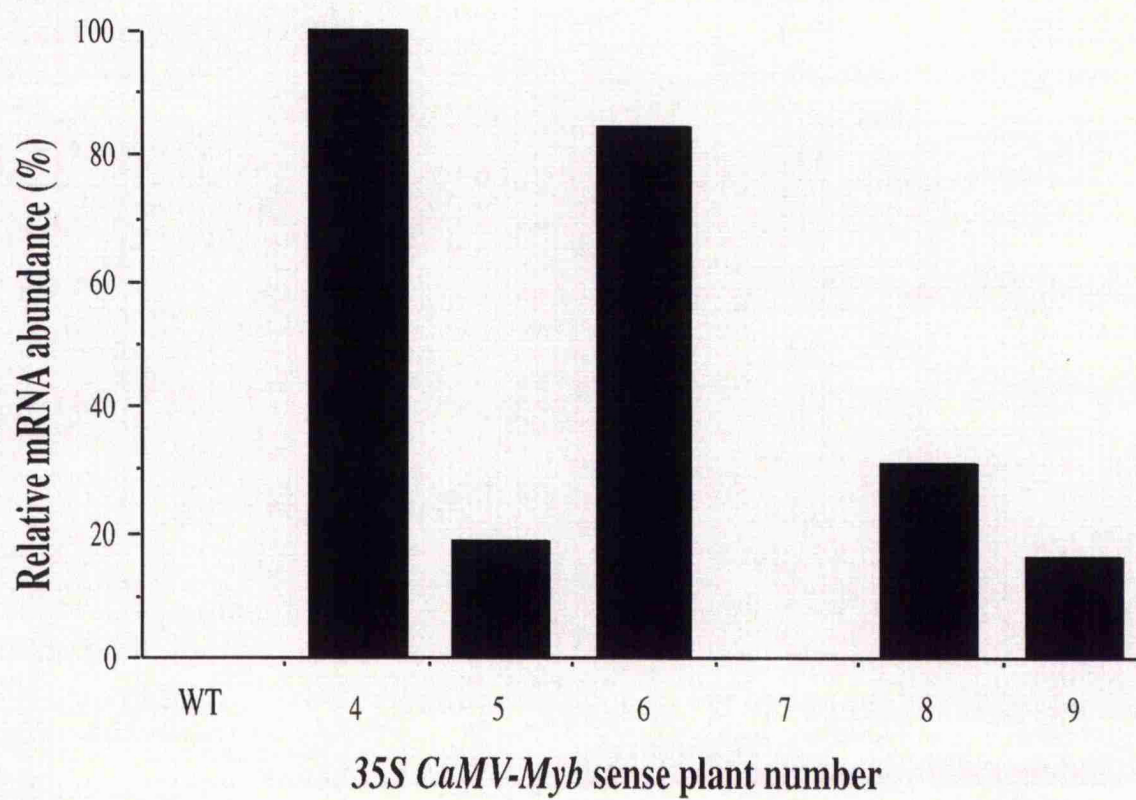
a.

b.

c.

Relative mRNA abundance (%)

35S CaMV-Myb sense plant number

There appeared to be a correlation between the degree of dwarfing seen in the *CaMV 35S* myb sense plants and the level of *myb.Nt1* transcript levels. Plants 4 and 6 showed the highest levels of transcript (100 % and 84 % respectively) and also showed the most severe phenotypes. Similarly, plants 5, 8 and 9 which showed a less severe phenotype also showed a lower transcript abundance (19 %, 31 % and 17 % respectively).

6.2.9 The dwarf phenotype analysed in second generation plants

Seeds collected from self crosses of the dwarf phenotype *CaMV 35S myb* sense plants were germinated in MS0 media (section 2.4.1) without kanamycin selection along with wild type control seeds. The dwarf phenotype could be seen in seedlings at 10 days post germination (cotyledon expansion and growth of first true leaves) (Figure 6.7). Further three types of seedling morphology could be distinguished and the numbers of seedlings showing each phenotype segregated in a Mendelian segregation ratio 1 : 2 : 1, which suggested that the plants had stably inherited a single T-DNA insertion and that the three phenotypes corresponded to the putative homozygous, hemizygous and null genotypic states. It has subsequently been shown by Southern analysis that these second generation plants do correspond to homozygote, hemizygote and null plants based on T-DNA copy number (Twell, personal communication). Homozygotes showed a severe phenotype, the cotyledons and first leaves were not expanded which with the minimal growth in height gave the seedlings the appearance of a small ball of green tissue on the surface of the medium. Hemizygotes had expanded coteyledons and first leaves but were reduced in size, to an intermediate state between homozygotes and null phenotype plants. The null plants were similar in size to the wild type control seedlings. Seedlings of the *CaMV 35S myb* sense plant numbers 3, 5, 8 and 9 also displayed the dwarf phenotype and displayed the three classes of seedling growth whose numbers segregated into the Mendelian ratios characterised with the original primary transformant plant (1 : 2 : 1 for plant lines 5, 8 and 9, indicating a single T-DNA insertion and 9 : 6 : 1 indicating two independent T-DNA insertions). However the distinction in phenotype between homozygotes and hemizygotes was not as clear in *CaMV 35S myb* sense plant lines 5, 8 and 9.

In order to investigate the affect of the dwarf phenotype on root growth and whether it was dependant on light, seeds of the *CaMV 35S myb* sense plant lines 4 and 6 and wild type controls were germinated horizontally on medium without kanamycin selection both and grown in continuous light and darkness (Figure 6.8). In the light

grown seedlings, root growth was inhibited to the same extent that shoot growth was. Homozygotes showed minimal root growth and putative null seedlings wild type root growth with hemizygotes intermediate growth between the two. The wild type controls showed more continuity in shoot growth although root growth was more variable. Dark grown seedlings still exhibited the dwarf phenotype in their etiolated stem elongation distinguishing the three types of growth.

Seeds collected from self crosses of the *CaMV 35S myb* sense dwarf plants were germinated in soil at the same time and grown in the greenhouse (Figure 6.9). Again, homozygous, hemizygous and null phenotype plants could be distinguished. Four plantlets exhibiting the homozygous, hemizygous and null phenotypes from each of the *CaMV 35S myb* sense dwarf plant lines were potted at the same time and grown under greenhouse conditions. The difference in internodal growth between the homozygous, hemizygous and null phenotypes could clearly be seen in each line (figure 6.11a.). Dwarf lines 3, 4, 6, 8 and 9 all showed a severe reduction in internodal growth in homozygous and hemizygous plants. Average heights of homozygous, hemizygous and null phenotype plants for each of the lines were similar with the exception of plants from line 5 which showed a less severe dwarf phenotype (Figure 6.10). Hemizygotes from this line were approximately half the size of the null phenotype plants and similarly, homozygotes were approximately half the size of the hemizygotes. The extent of the dwarfing phenotype was clearly seen by lining a homozygote from the severe dwarf line 6 next to an average internodal stem segment from the null plant of the same line which were both approximately 6.5 cm in height (Figure 6.11b.).

As well as a reduction in internodal growth there was an increase in axial leaf development relative to plant height in the homozygotes and hemizygotes in lines 4, 5 and 6 (Figure 6.12a.). This development was not present in the present in null plants from the same lines. The reduced stature and increase in axillary leaf development in lines 4 and 6 gave the homozygotes and hemizygotes the appearance of being bushy.

The dwarf phenotype was also characterised by alterations in leaf morphology in addition to their reduced size (Figure 6.12b). Homozygotes and hemizygotes of plants from lines 3, 4, 6, 8 and 9 showed a flat leaf surface with a more rounded

173

Figure 6.7
In vitro **grown second generation seedlings from three lines of transgenic** *CaMV*
35S myb.Nt1 **sense** *N.tabacum* **plants showing an early dwarf phenotype.**

All seedlings were grown for 14 days under constant light conditions. On the left of each photograph is a plate of wild type seedlings of the same age, grown under the same conditions for comparison. Seedlings were grown in the absence of kanamycin.

Abbreviations: Hom, homozygous seedling; Het, hemizygous seedling; null, null phenotype seedling.

a. Seedlings from line 9 (Table 6.1).
b. Seedlings from line 4.
c. Seedlings from line 6.
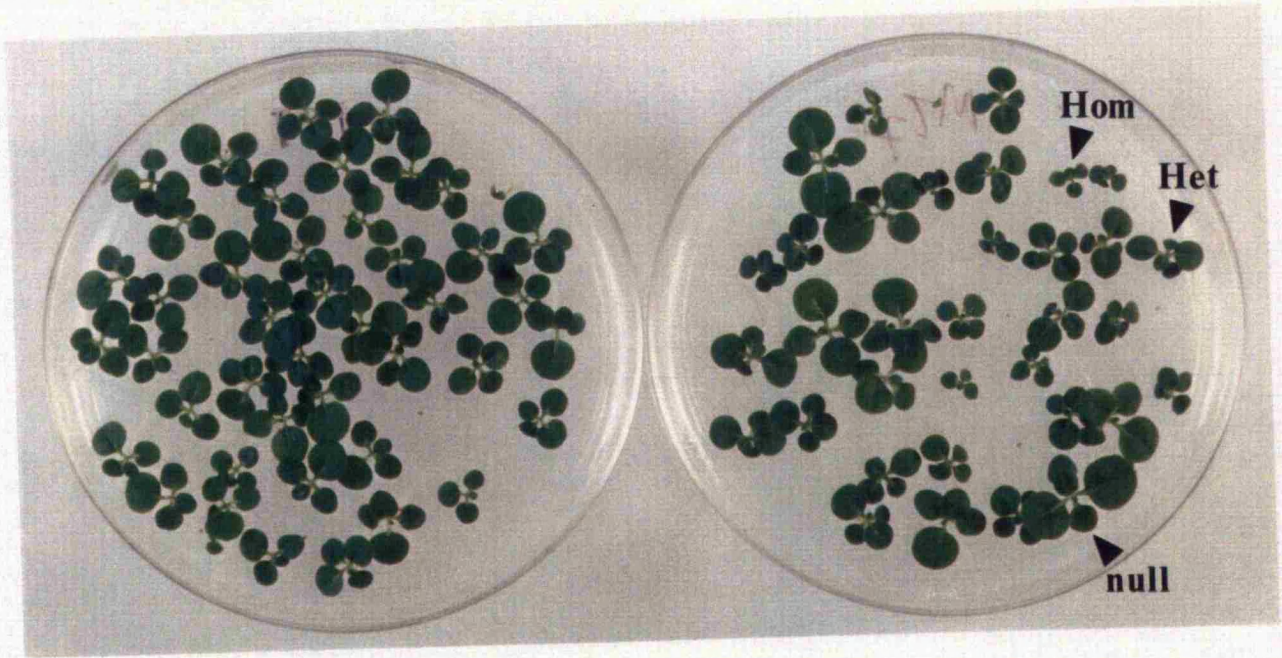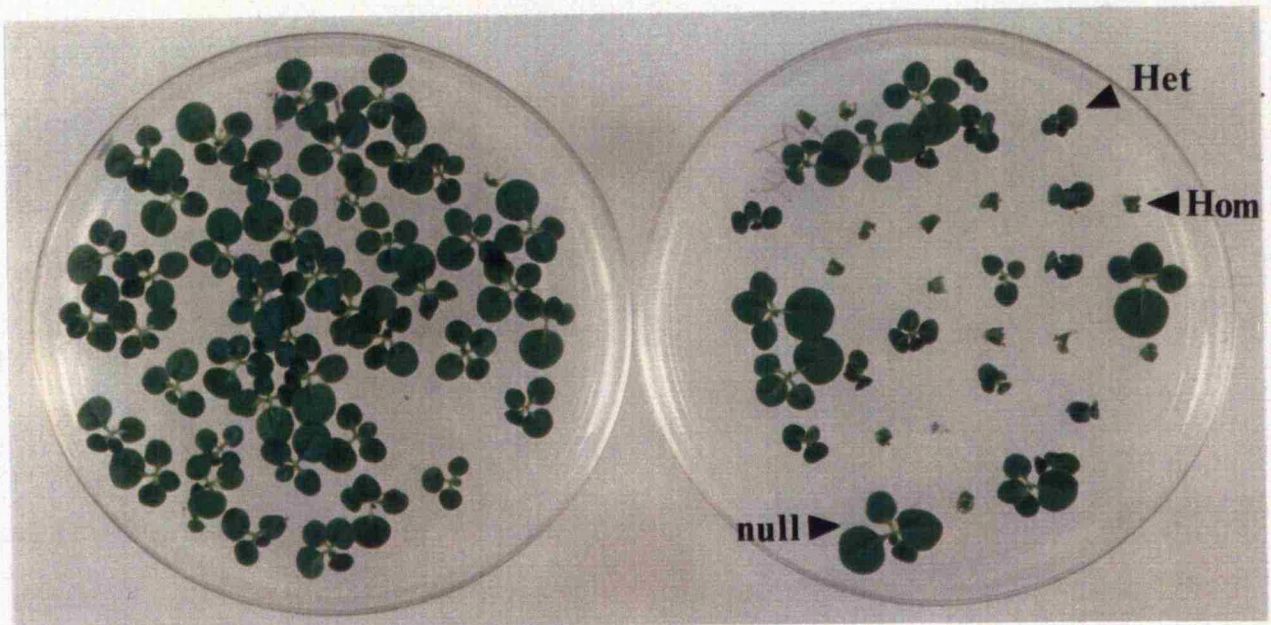
Figure 6.8

In vitro **grown second generation seedlings from transgenic** *CaMV 35S myb.Nt1* **sense** *N.tabacum* **plants grown horizontally in the light and dark.**

Seedlings from line 6 (Table 6.1) grown for 24 days under constant light or dark conditions. On the left of each photograph is a plate of wild type seedlings of the same age, grown under the same conditions for comparison. Seedlings were grown in the absence of kanamycin.

Abbreviations: Hom, homozygous seedling; Het, hemizygous seedling; null, null phenotype seedling.

a. Seedlings grown horizontally in the light.
b. Seedlings grown horizontally in the dark.

a.

Het   null Hom

b.

null

Hom   Het
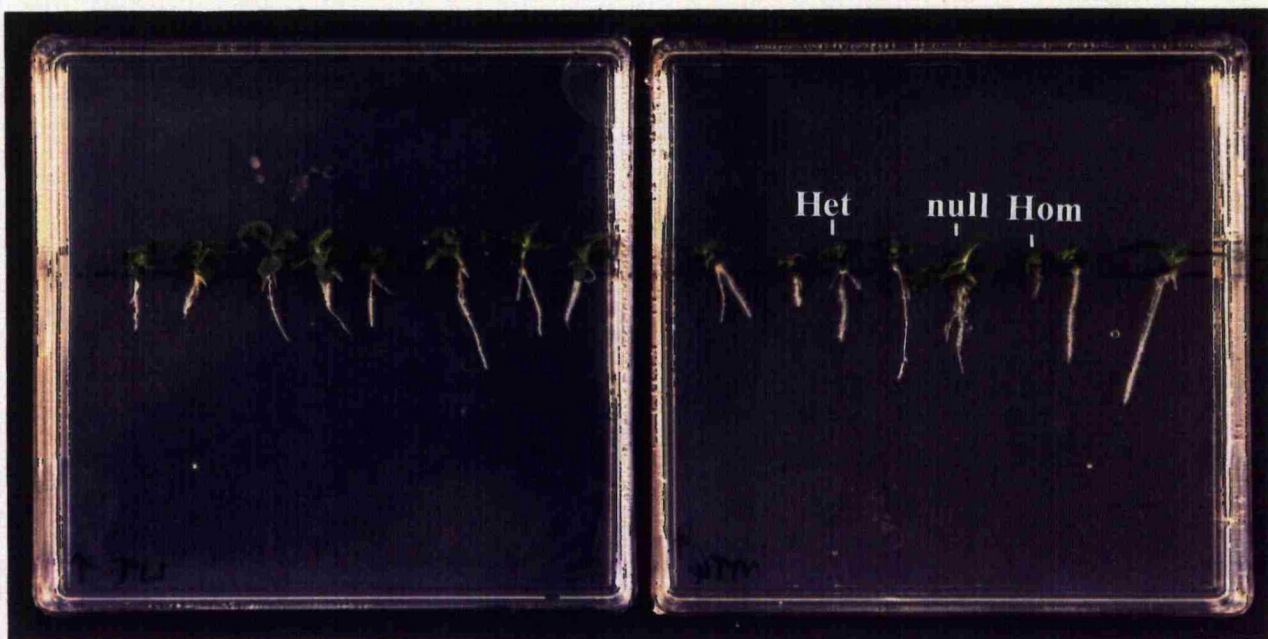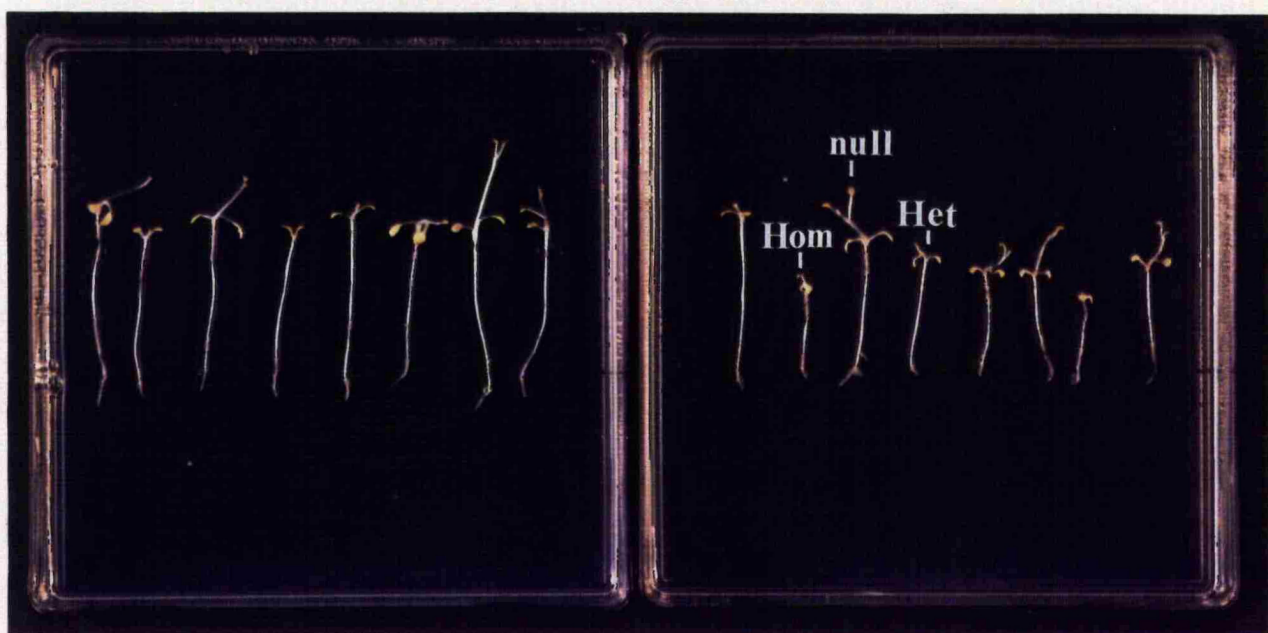
Figure 6.9

Second generation plantlets from two lines of transgenic *CaMV 35S myb.Nt1* sense *N.tabacum* plants.

Plantlets were grown for 30 days from germination under constant greenhouse conditions. On the left of each photograph is a tray of wild type plantlets of the same age grown under the same conditions for comparison.

Abbreviations: Hom, homozygous plantlet; Het, hemizygous plantlet; null, null phenotype plantlet.

a. Plantlets from line 4 (Table 6.1).
b. Plantlets from line 6.

a.

-Het

null

Hom

b.

null

Het

Hom
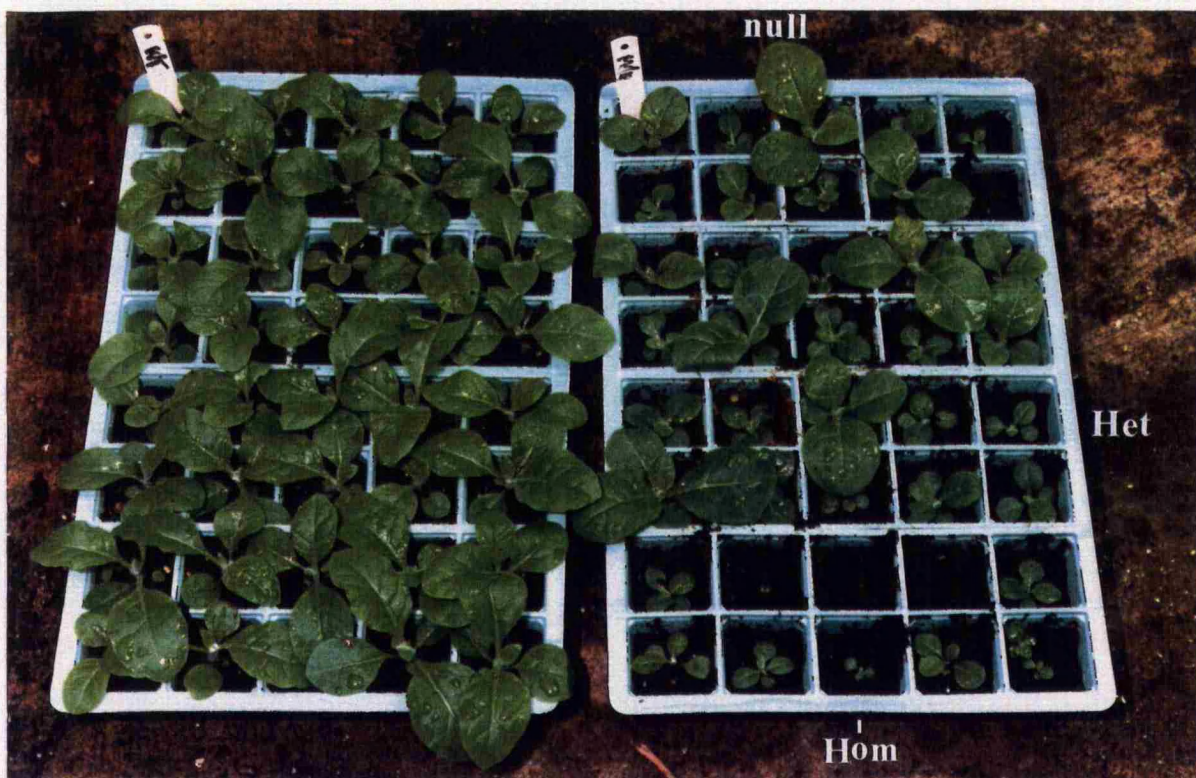
Figure 6.10

Heights of second generation *CaMV 35S myb.Nt1* sense *N.tabacum* plant lines at 62 days post germination.

Numbers on X-axis refer to plant line number.

Abbreviations: Null, null phenotype plantlet; Het, hemizygous plantlet; Hom, homozygous plantlet.

Bar chart showing height in cm. Each bar represents the average height of four individual plants. Error bars indicate standard deviation (n=4). Height was measured as the distance from the base of the stem to the tip of the first flower. Each plant was under similar greenhouse conditions.

177
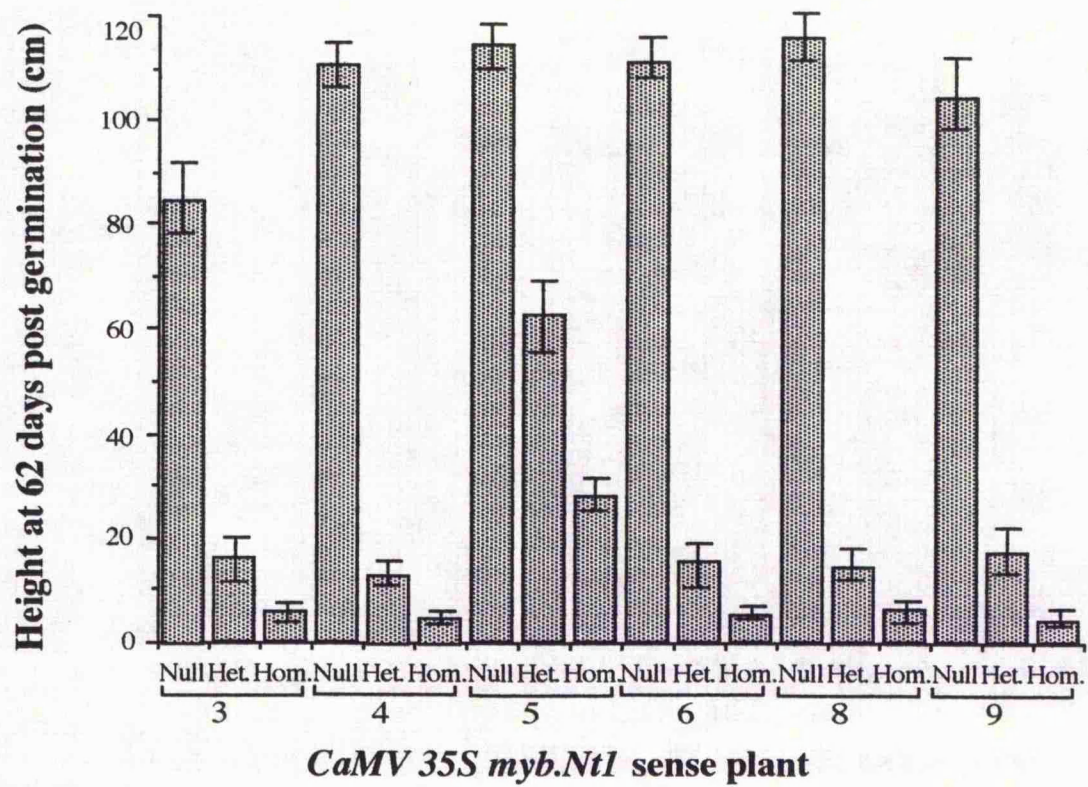
Figure 6.11

Second generation transgenic *CaMV 35S myb.Nt1* sense *N.tabacum* plants showing homozygous, hemizygous and null dwarf phenotypes.

a. Plants from line 4 (Table 6.1) showing the three classes of dwarf growth: left, null phenotype plant; middle, hemizygous plant; right, homozygous plant. All plants were grown for 83 days from germination under constant greenhouse conditions.

b. A homozygous plant from line 6 (right), which had the primary leaves removed to expose axial leaf growth, next to an internode segment from a null phenotype plant of the same line and age (left). All plants were grown for 62 days from germination under constant greenhouse conditions.

a.

b.

Figure 6.12

Second generation transgenic *CaMV 35S myb.Nt1* sense *N.tabacum* plants showing altered axial growth and leaf morphology

All plants were grown for 62 days from germination under constant greenhouse conditions.

a. Homozygous plants from line 6 (left) and line 5 (right) showing advanced axial growth and a difference in leaf morphology.

b. A comparison of leaf morphology from three lines of dwarf plants: leaf from null phenotype line 6 plant (left); leaf from hemizygous phenotype line 4 (middle); leaf from homozygous phenotype line 5 plant (right). Leaves from each plant were comparable in age and position.

a.

b.

shape compared to leaves of null plants of the same line and age, which have an extended shape and undulating surface. Leaves and stems in homozygotes and hemizygotes from each of these lines were also greener than those in the equivalent null plants. Young leaves from homozygote and hemizygote plants from lines 3, 5, and 8 showed the 'inverted spoon' morphology with a slender petiole seen previously in the original plant 3 primary transformant. Young leaf morphology was pear shaped compared to leaves of null phenotype plants of the same line. In addition the average number of leaf veins was reduced in homozygotes and hemizygotes compared to null plant leaves of the same size and age.

The onset of flowering was delayed by approximately 7 days in hemizygotes from lines 3, 4, 6, 8 and 9 compared to null plants from the same lines. Flowering of the homozygotes was delayed by approximately a further 14 days in lines 5, 8 and 9 and by a further 14 days in lines 3, 4 and 6. These observations suggest that there is a correlation between the height of the plant and flowering time.

Flower morphology was similar in null, hemizygote and homozygote plants from each line, except line 5 which showed oval shaped stigmas on each flower on hemizygote and homozygote plants compared to round stigmas on null plants. There was no obvious correlation between plant height and visible intensity of pigmentation in the petals of null, hemizygotes and homozygotes in each line. Flowers from hemizygotes and homozygotes from lines 4 and 9 showed slightly more intense pigmentation than flowers of null plants from the same lines. However, this effect was not seen with flowers from the other lines.

6.2.10 An investigation of the activation potential of MYB.NT1 in transient assays

6.2.10.1 Use of LUC and GUS as reporter genes in microprojectile bombardment in leaves to test for *trans*-activation from pollen-specific promoters

Trans-activation by MYB.NT1 from a number of pollen-specific promoters was tested using microprojectile bombardment in *N. tabacum* leaf tissue. Reporter constructs were based on a series *N. tabacum* pollen-specific promoters which were available for use in the laboratory. These consisted of the promoters of the *lat52* (Twell *et al.*, 1991), *lat59* (Twell *et al.*, 1991), *ntp303* (Weterings *et al.*, 1992), *npg-1* (polygalacturonase, (Tebbutt *et al.*, 1994)), *tp10* (pectate lyase, (Rogers *et al.*, 1992a)) and *tac25* (actin, (Thangavelu *et al.*, 1993)). These promoters were positioned

180

immediately upstream of the luciferase (LUC) reporter gene in the *lat52, lat59* and *ntp303* promoter constructs and the *β–glucoronidase* (GUS) reporter gene in the *npg-1, tp10* and *tac25* promoter constructs. The effector constructs was the *E. coli* plasmid constructed as part of the cloning strategy used to generate a T-DNA vector for the ectopic over-expression of *myb.Nt1* in *N. tabacum* tissues (Figure 6.2). This construct consisted of the duplicated *CaMV 35S* promoter placed upstream of a tev translational enhancer and the *myb.Nt1* cDNA. In addition, an internal control plasmid was used to measure plasmid delivery by bombardment in the leaf tissue. These constructs consisted of the duplicated *CaMV 35S* promoter placed upstream of either the *gus* or *luc* reporter genes.

Each bombardment consisted of delivery of effector, reporter and internal control plasmids together in greenhouse grown *N. tabacum* leaf material (section 2.13.4). In addition, a negative control bombardment consisted of delivery of reporter and internal control plasmids and non-specific DNA *(pUC19)*. Each bombardment experiment was repeated twice. Quantitative determinations of LUC and GUS levels were measured from crude protein extracts of the bombarded leaves (sections 2.13.5.2 and 2.13.5.3).

No activation of expression from the pollen-specific promoters was seen despite high levels of expression of the internal control plasmids.

6.2.10.2 The MYB.AM305 protein from *Antirrinhum* shows transactivation from a chalcone isomerase promoter

In order to test the sensitivity of the transient assay system, transactivation of a promoter sequence from the *Antirrinum* chalcone isomerase *(chi)* gene by a previously characterised MYB.AM305 protein was investigated (plasmids were a personal gift from Dr. C. Martin, Norwich). A previous study had shown that MYB.AM305, a flower specific MYB protein from *Antirrinum*, could transactivate from a *gPAL2* promoter fragment containing a MYB-binding site (AACCTAAC; (Sablowski *et al.*, 1994)). Similar sites were found to be present in a variety of promoters of phenyl-propanoid genes including *chi* (ACCTAAC; (Martin *et al.*, 1991)).

An effector plasmid was based on the *CaMV 35S* promoter upstream of the *myb.Am305* coding sequence. Reporter constructs contained the *chi* promoter fragment upstream of the *CaMV 35S -60* minimal promoter fragment containing a

Figure 6.13

Transient transactivation of a chalcone isomerase promoter by MYB.AM305 with controls ((Sablowski *et al.*, 1994); (Jackson *et al.*, 1991)).

Graph showing relative activity of each construct in microprojectile bombardment. Error bars indicate standard deviation (n=2).

Abbreviations: MYB, MYB.AM305; -60, -60 deletion minimal *cauliflower mosaic virus* promoter; CHI, chalcone isomerase promoter.
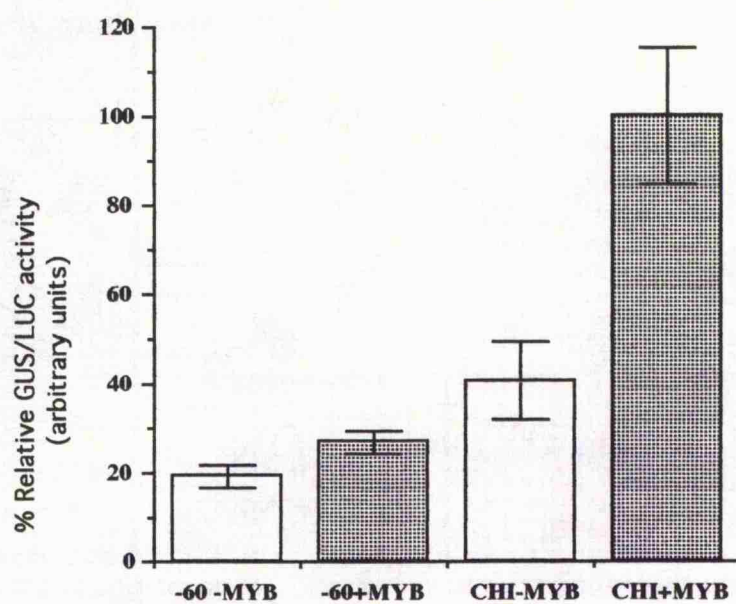
TATA sequence which was upstream of the GUS reporter gene. An CaMV-LUC internal control construct was also used to measure the efficiency of bombardment. Activated expression of the *chi* promoter fragment by MYB.AM305 was observed (Figure 6.13). Levels of expression for each bombardment was expressed as a ratio of GUS/LUC activity. Average values were taken for each set of plasmids and expressed as a relative activity with the value for *chi*+MYB set at 100 %. Activation from the *-60 CaMV 35S* construct by MYB.AM305 averaged 1.9 fold. This stimulation of expression may reflect the presence of a MYB-binding site in *pUC19* sequence upstream of the *CaMV 35S* promoter (Sablowski *et al.*, 1994). Both reporter plasmids gave levels of expression without MYB.AM305 (20 % with *-60 CaMV 35S* and 40 % with *chi*) suggesting that these fragments were targets for endogenous transcriptional activators in *N. tabacum* leaves. However expression of GUS was consistently activated 2.5 fold over that obtained with the *-60 CaMV 35S* promoter fragment alone by co-bombardment with the plasmid expressing MYB.AM305.

Demonstration of *trans*-activation in the transient assay system suggested that the system was sensitive enough to detect an activation from the *N. tabacum* pollen-specific *lat52, lat59* and *ntp303* promoters by *myb.Nt1* expression using the *luciferase* reporter gene, as the sensitivity of detection of *luciferase* is much greater than that of β-glucuronidase.

6.3 Discussion
6.3.1 Analysis of a population of transgenic *N. tabacum* plants

A sample of 117 transgenic *N. tabacum* plants were analysed for expression and inheritance of the kanamycin resistance gene. The expression test consisted of plating seeds from self- and outcrosses on medium containing kanamycin, which showed that approximately 79 % (93) of them gave Mendelian inheritance ratios for the kanamycin resistance phenotype (Classes I to III). Previous published studies of transmission ratios for a large sample of transgenic plants ((Budar *et al.*, 1986); (Deroles and Gardner, 1988)), found that Mendelian transmission of the kanamycin gene in 40 out of 44 *N. tabacum* plants and 59 out of 104 *P. hybrida* plants respectively.

Class I plants transmitted kanamycin resistance to their progeny as a single dominant gene. A further 20 plants from Class II and 2 plants from Class III gave inheritance data consistent with their having two independent copies of the kanamycin resistance gene. This proportion of multiple gene plants is slightly higher than that obtained in a study of *Agrobacterium* mediated gene transfer in *N. tabacum* transgenic plants (5 out

183

of 45 plants; (Budar *et al.*, 1986)). The difference in the figures of single gene versus multiple gene transformants is probably due to the co-cultivation conditions and selection protocols used.

A group of 26 plants showed a non-Mendelian inheritance of their kanamycin resistance gene, despite rooting on kanamycin containing medium. Class IV plants (8) showed no transfer of the kanamycin resistance gene. These plants were likely to be non-transformants, however inactivation of the kanamycin resistance gene by deletion or rearrangement of the T-DNA at meiosis could account for loss of resistance. Southern analysis would be needed to investigate the presence of a T-DNA locus. A group of 10 Class V plants transmitted their kanamycin resistance phenotype to self and backcross progeny with ratios that were not consistent with one of the Mendelian patterns described or with no expression. A further 7 plants produced patterns consistent with the reduced transmission of the kanamycin resistance gene through the transgenic male line (RTM phenotype; Class VI). One of these plants (*lat52 myb.Nt1* sense plant 10), showed a gamete lethal phenotype (1 : 1, 1 : 1, 0 : 1 for selfcross and outcrosses through the transgenic female and male respectively). F_1 plants were generated from seed of these primary transformants and the inheritance of the RTM phenotype investigated by kanamycin resistance gene segregation in seed from F_1 plants.

6.3.2 *Myb.Nt1* is an important gene required for pollen function

The stable inheritance of the RTM phenotype in the F_1 generation was demonstrated in four transgenic plant lines: *lat52 myb.Nt1* sense plant line 10, *lat52 myb.Nt1* antisense plant lines 15 and 16 and *lat52 Nt11* sense line 7. The primary transgenic plants had demonstrated a 1 : 1 segregation ratio of kanamycin resistant to sensitive seedlings with a backcross through the transgenic female indicating a single T-DNA insertion. A significant reduction in the level of the phenotype was observed in the plant line 10. This plant showed pollen cell death in approximately 50 % of grains from a self-cross indicating that presence of the transgenic transcript resulted in cell death. In addition, other primary transformants identified as having a RTM phenotype from their segregation ratios did not seem to transmit these ratios to the F_1 progeny. This apparent loss of transmission might have been due in fact to low population counts in some of the crosses of the primary transformants which resulted in statistically inaccurate ratios. Alternatively, non-Mendelian ratios in the primary

tansformants may have been due to unstable affects on pollen viability due to the tissue culture procedure which subsequently stabilised in the F_1 progeny.

A reduction of endogenous *myb.Nt* transcript in the pollen of the *lat52 myb.Nt1* sense line 10 and antisense lines 15 and 16 suggests there exists a link to the genetic ratios. Those plants carrying the *lat52 myb.Nt1* sense/antisense transgenes but exhibiting Mendelian kanamycin resistance ratios must not express enough of the transgene transcript to have a significant effect on the sense expression of the endogenous *myb.Nt* genes. Thus, partial antisense suppression of *myb.Nt* expression is operating in plant lines 15 and 16 and partial sense cosuppression of *myb.Nt* expression in plant line 10. The proportion of transgenic plants exhibiting transgene suppression of endogenous *myb.Nt* transcript levels is relatively low (3 % for sense suppression and 6 % for antisense suppression). Variable frequencies of between 2 % and 50 % of transgenic plants exhibiting sense/antisense suppression have been reported in other studies. The reasons for this variability of antisense and sense cosuppression are not understood, but it seems likely that features such as the transgenes themselves are involved such as expression level, integration sites and the structure of the integrated T-DNAs ((De Lange *et al.*, 1994); (van der Krol *et al.*, 1990c)). Generation of a larger number of primary transformants should reveal more plants exhibiting transgene-mediated gene suppression, including plants with complete inhibition of gene expression in 50 % of the pollen grains in a selfcross.

Microscopic analysis of pollen from each of the *myb.Nt* transgene lines 10, 15 and 16 showed an increase in the proportion of small and collapsed grains over that seen in wild type pollen samples (approximately 3 : 2 round to collapsed grains in *myb.Nt* transgenic lines as opposed to 4 : 1 in wild type). Cytochemical analysis and *in vitro* germination of the collapsed grains confirmed that the grains were dead. Evidence from the genetic, molecular and cytochemical studies suggest that the *myb.Nt* genes are crucial for pollen maturation and are not involved in pollen germination. By analysing the changes in the proportions of normal and collapsed pollen grains at each stage of development it should be possible to identify at which stage pollen grains show the collapsed phenotype, which could lend further evidence to the involvement of transgene suppression in pollen death.

The overexpression of the *Ntl1* transcript in *lat52 Ntl1* sense plant line 7 did not produce a pollen phenotype, based on a comparision of the proportion of normal and collapsed pollen grains to wild type plants, thus high level expression of *Ntl* genes may not be detrimental to developing pollen. Alternatively, low population counts in the crosses of the primary transformant plant may have resulted in statistically

185

inaccurate ratios. A third possible explanation for the lack of a pollen phenotype in line 7 may be that the *lat52* promoter may have conferred an incorrect pattern of *Ntl1* transgene expression in developing pollen for gene suppression to occur. By permitting low levels of endogenous NTL proteins to accumulate before transgene transcript is produced a normal phenotype may result. In this respect the *ntm19* promoter (section 5.2.3) which drives high level expression in late microspores prior to pollen mitosis I may be better for transgene mediated gene suppression by producing high levels of transgene transcript immediately before expression of endogenous transcript. However, this promoter was not available for use at the time the sense/antisense constructs were made.

6.3.3 Ectopic expression of *myb.Nt1* causes a dwarf phenotype

Ectopic expression of the *myb.Nt1* cDNA from the duplicated constitutive *CaMV 35S* promoter resulted in a dwarf phenotype to varying degrees in a group of six out of eight primary transformants. The combination of the low number of primary transformants and the high proportion of these transformants exhibiting a phenotype suggests high ectopic expression of the *myb.Nt1* cDNA results in cell death. Ectopic expression of the *Ntl1* cDNA did not produce any discernible phenotype. Northern blot analysis established a correlation between levels of *myb.Nt1* transcript and the severity of the dwarf phenotype in these plants. The variation in expression levels of *myb.Nt1* may be due to differences in the genomic integration site of the T-DNA locus. Insertion into a low expressing region of the genome due to chromatin secondary structure may result in a lower expression of the *myb.Nt1* gene. Alternatively, rearrangement of a portion of the integrated T-DNA during meiosis could be an alternative explanation of the lowered transcript levels (De Lange *et al.*, 1994).

Selfcrosses and outcrosses through the transgenic male line of the primary transformants produced seed which was germinated on medium without kanamycin selection. Time of germination was similar in all the dwarf lines and wild type controls. Two classes of resistant seedlings were identified, these were subsequently shown by Southern analysis to correspond to homozygote and hemizygote plants (Twell, personal communication). The numbers of resistant and sensitive seedlings segregated in Mendelian ratios indicating single dominant T-DNA insertions for five of the six lines and normal transmission through the transgenic male for all the dwarf lines. Germination of seeds of line 6 horizontally and in the dark indicated that the

dwarf phenotype was independent of light. Suggesting that the alteration of gene expression which produces the dwarf phenotype is separate from phytochrome signal transduction pathways. Horizontal growth in the light demonstrated that root elongated growth was also affected in dwarf lines.

The dwarf phenotype was analysed in mature plants grown in soil and was characterised by a general reduction in size of all plant tissues and organs. A severe reduction in internodal growth was seen in lines 3, 4, 6, 8 and 9 with average height reductions of hemizygotes and homozygotes compared to the putative null plants in these lines of 87 % and 94 % respectively. In contrast, internodal growth inhibition in plants from line 5 was less severe with height reductions of hemizygotes and homozygotes of 54 % and 25 % respectively at 62 days post germination..

A different leaf morphology was seen in lines 3, 5 and 8 compared to the other dwarf lines with an 'inverted spoon' phenotype in young leaves with a more pronounced petiole. This suggests that the rate of cell division and/or cell expansion at the leaf edge is slower than the rest of the leaf. Therefore, the dwarf phenotype in these leaves seems to affect cell-specific gene expression. In contrast young leaves from lines 4, 6 and 9 had a flat upper leaf surface.

While the null and hemizygote plants flower after different periods of time they produce approximately the same number of mature leaves and internodes at flowering. This observation indicates that the MYB.NT1 protein does not interact with the developmental process of floral determination in *N. tabacum*. A phenotypic effect seen only in flowers from hemizygotes and homozygotes line 5 was ovalised stigmas compared to round stigmas in null plants. Thus, the phenotype seen in line 5 is distinct from the phenotype seen in other lines.

The dwarf phenotype was seen in six independent lines. Co-inheritance of the dwarf phenotype with the kanamycin resistance gene as a dominant trait and a correlation of plant height and transgene expression level, suggests that the pollen-specific MYB.NT1 protein is modulating gene expression levels to produce the dwarf phenotype. An alternative explanation would be that the phenotype is the result of a random somaclonal or a T-DNA knockout mutation of an unrelated gene. Although the occurrence of this random event in six independent lines is extremely unlikely.

Transient expression assays have been used to show that MYB proteins in maize can bind to the promoters of flavonoid biosynthetic genes, and activate expression of these genes (section 1.4.4.1). MYB proteins bind either on their own, as in P binding to the promoter of the *A1* gene (Grotewold *et al.*, 1994), required for phlobaphene biosynthesis, or via heterodimerisation to a *myc* gene partner, as with C1 and and B

biosynthesis, or via heterodimerisation to a *myc* gene partner, as with C1 and and B which bind to composite sites on the promoter of the *Bz-1* gene ((Goff *et al.*, 1991); (Goff *et al.*, 1992)), or C1 and R, which bind to the promoter of the *Bz-2* gene (Bodeau and Walbot, 1992). The high level of expression of MYB.NT1 may lead to a squelching effect (Ptashne, 1988), whereby MYB.NT1 may bind to potential protein partners via heterodimerisation in solution and bind at composite *myb* sites. If the target factor is present at limiting concentrations, then it will be sequestered away from genes that require it for transcription, resulting in their inhibition. The target factor may be a protein binding partner, such as a MYC protein, or a component of the basal transcription apparatus, required for the transcription of a wide range of genes. Thus, MYB.NT1 may repress activation of *myb*-responsive genes indirectly by competing for the DNA-binding sites of other MYB factors and preventing the binding of these MYB factors to protein partners. Alternatively, in the light of possible low level expression of MYB.NT1 in most tissues of *N. tabacum* (Figure 5.2), the dwarf phenotype may be the result of overexpression of the pollen-specific protein rather than ectopic expression. An increase in levels of MYB.NT1 over a threshold value would activate gene expression.

MYB.NT1 seems to affect a wide variety of growth traits, therefore it either interacts with a common 'master' regulatory switch, or at multiple points in plant growth and development.

The simplest model to explain these observations of altered plant growth by MYB.NT1 in the dwarf phenotype is that MYB.NT1 affects gibberellin (GA) biosynthesis, either directly or via another factor, or that MYB.NT1 interacts with the GA signal transduction pathway. The gibberellins are a large and complex family of tetracyclic diterpenoid molecules, some of which act as endogenous regulators of various processes including internode extension growth, seed germination, fruit and flower formation. Gibberellin mutants exist in several plant species, and can be classified as GA biosynthesis mutants or GA response mutants, depending upon whether or not the mutation can be complemented by exogenous application of GAs (Ross, 1994). The mechanism of action of internode elongation by giberellins is not well understood. In lettuce, GA_3 had multiple affects on hypocotyl elongation, increasing cell wall extensibility and cell mitosis rate (Matraux, 1987). The minimal stem growth seen in the dwarf homozygous plants may be reflecting incomplete inhibition of gibberellin biosynthetic or gibberellin induced gene expression by MYB.NT1, allowing 'leaking' of low level gene expression.

GAmyb, cDNA clone recently isolated from barley aleurone layers (Gubler *et al.*, 1995), was shown to bind specifically to a 'TAACAAA' box motif present in the gibberellin responsive complex (GARC) in the promoter of the high-pI α-amylase gene. Transient expression assays were used to show that *GAmyb* could transactivate from a high-pI α-amylase gene promoter fused to a β-Glucuronidase reporter gene in the absence of GA. Other GA responsive genes such as the barley *EII (1-3, 1-4)-β-glucanase* (Wolf, 1992) and the wheat cathepsin B-like gene (Cajudo *et al.*, 1992) also contain 'TAACAAA'-like sequences in their promoters. Thus *GAmyb* may act as a master regulator of gibberellin responsive genes. Interestingly, GAMYB shares relatively high MYB domain sequence homology with MYB.NT1, MYB.NT2 and MYB.PH3 (81 %, 83 % and 86 % respectively). Further, the GAMYB binding site (TAACAAA) shows a one base pair mismatch to the MYB.PH3 MBSII binding site (TAACTAA; opposite strand shown). In the light of similarity of the MYB.NT and MYB.PH3 binding sites (section 4.2.3.1), MYB.NT1 may interact with *N. tabacum* homolog GAMYB binding sites to repress gene expression and thus produce a dwarf phenotype in *N. tabacum* by the mechanisms discussed above.

While no significant correlation could be defined between the degree of dwarfism and intensity of pigmentation by eye, extraction and quantitation of the anthocyanin pigments from plants grown under carefully controlled light conditions would be needed to determine whether MYB.NT1 interacts with the production of pigments via the flavonoid biosynthetic pathway. In *P. hybrida*, the expression of a number of key anthocyanin biosynthetic enzymes, chalcone synthase, chalcone flavanone isomerase and dehydroflavanol 4-reductase, which are coordinately expressed during flower development, are induced by GA3 (Weiss *et al.*, 1990). In the case of chalcone synthase, poly A$^+$ RNA accumulation was subsequently shown to be controlled by GA3 at the level of transcription, as it was inhibited by cycloheximide (Weiss *et al.*, 1992). A time lag of approximately 5 hours was seen between between application of GA3 to *P. hybrida* corollas and activation of *chs* expression. This time lag suggests that GA3 controls transcription in an indirect way perhaps by inducing the synthesis of regulatory proteins including MYB factors. A *Pmyb 92* cDNA clone, has been isolated from *P. hybrida* corolla tissue and is induced by gibberellic acid within 30 minutes after GA3 application. Promoter-GUS fusions revealed that *Pmyb 92* was also expressed in trichomes (hair cells), ovaries, stomium cells, stigmas and seedlings (Mur *et al.*, 1994).

MYB.PH3 has been demonstrated to *trans*-activate from the *P. hybrida* chalcone synthase J promoter in *N. tabacum* protoplasts presumably from two MYB.PH3

binding sites (Solano *et al.*, 1995a.). Thus, it has been proposed that MYB.PH3 may play a similar role to that proposed for GAMYB (Gubler *et al.*, 1995), which suggests that MYB.NT1 may interact at the chalcone synthase promoter and compete with MYB.PH3 for binding sites.

Considering that the pollen-specific MYB.NT proteins may represent a specific class of MYB proteins in *N. tabacum* it is interesting that they interact with other *myb*-gene regulated pathways. Southern analysis suggested that there are multiple *myb* genes in *N. tabacum* (section 5.2.4). Thus the opportunities that MYB.NT1 may have to alter GA-induced or other unrelated *myb* gene expression patterns in *N. tabacum* to produce the dwarf phenotype may be multiple.

6.4 Summary

Using a combination of genetic, molecular and histochemical analysis tentative evidence was collected which demonstrated that pollen grains which over-expressed antisense or sense *myb.Nt1* transcript's died. These results indicated an essential role for the *myb.Nt* genes in pollen development. Over-expression of the *Nt11* transcript in pollen did not produce a pollen phenotype. Ectopic expression of the *myb.Nt1* transcript resulted in a severe dwarf phenotype characterised by a significant reduction in internode growth, leaf shape changes and a delay in flowering. These phenotypic changes suggest an interaction of MYB.NT1 with MYB-regulated gene expression such as gibberellin induced gene activation or activation of genes encoding gibberellin biosynthetic enzymes.

Chapter Seven.

Discussion.

7.1 Characterisation of myb.Nt1, myb.Nt2, Ntl1 and Ntl2 cDNAs

7.1.1 cDNA sequence characterisation

A cDNA library was made from poly A$^+$ RNA from mature pollen of *N. tabacum* and screened with heterologous DNA sequence probes which resulted in the isolation of cDNAs corresponding to two *myb*- and two *lim*-related genes (*myb.Nt1*, *myb.Nt2* and *Ntl1*, *Ntl2* respectively; Chapter 4). These genes form part of a *myb*-gene family of approximately 6 to 8 gene members and a *lim*-gene family of 3 to 4 gene members. The products of the *myb.Nt1* and *myb.Nt2* genes have the structural characteristics of transcriptional activators, containing a MYB-domain composed of two repeats (R2 and R3), in which the residues known and/or predicted to be important for DNA-binding activity of *myb* proto-oncogenes have been conserved. In this respect, three tryptophan residues separated by 18 to 19 residues are present in each repeat except the first residue in the second repeat which is replaced with a phenylalanine. This MYB-domain structure occurs in other plant MYB proteins, and exchange of an aromatic or hydrophobic residue for the first tryptophan of the second repeat does not affect the DNA-binding potential of c-MYB (Saikumar *et al.*, 1990). In a comparison of MYB-domain sequences from animal and plant proteins, three α–helices are present in each of the R2 and R3 repeats, the last two of which on each repeat form a helix-turn-helix motif which contacts a specific DNA sequence ((Frampton *et al.*, 1991); (Ogata *et al.*, 1994)). The MYB-domain sequences of the MYB.NT proteins are identical, which predicts that the two proteins bind to the same DNA sequences. In a similar comparison of MYB-domains, the MYB.NT MYB-domains share most sequence identity with that of MYB.PH3 from *P. hybrida* and MYB.GA from Barley (93 % and 81 % respectively for MYB.NT1), which have been implicated in the activation of flavonoid biosynthetic gene expression and gibberellin-induced activation of α-amylase gene expression respectively ((Avila *et al.*, 1993); (Gubler *et al.*, 1995)).

The characterisation of the DNA-binding sites of the MYB.NT proteins could be investigated by gel retardation assay experiments ((Fried and Crothers, 1981); (Garner and Revzin, 1981)) using labelled DNA fragments containing the MYB binding sites identified for MYB.PH3 (Avila *et al.*, 1993); c-MYB from humans (Majello *et al.*, 1986); maize P (Grotewold *et al.*, 1994) and AM305 from *Antirrinhum* (Sablowski *et al.*, 1994). Alternatively, the binding site selection strategy based on the gel retardation assay (Blackwell and Weintrub, 1990) used for the determination of the MYB.PH3

binding sites (Solano *et al.*, 1995a.), could be used to identify the MYB.NT binding sites. The identification of the determined DNA-binding sites in the promoters of pollen-expressed genes, including the *myb.Nt* gene promoters, and the demonstration of activation from these promoters in transient assays would make an important contribution to the functional analysis of the pollen *myb* genes. Positive autoregulation of human *c-myb* has been demonstrated in transient assays and co-transfection studies (Nicolaides *et al.*, 1991) and could represent a mechanism for *myb* gene regulation in plants.

Outside the MYB-domain, two other regions of significant sequence similarity are shared between the MYB.NT and the MYB.PH3 protein. These regions have secondary structural characteristics of acidic activator domains. Several regions rich in serine/threonine residues which are potential sites for phosphorylation as well as several potential sites for glycosylation (NXS/T motifs) are present in the MYB.NT proteins which suggests their activity could be modulated post-transcriptionally like that of other transcriptional regulators (Mitchell and Tijan, 1989). Thus although the role of the MYB.NT proteins has not been demonstrated directly, the structural characteristics of the proteins suggests that they are transcription factors.

The products of the *Ntl1* and *Ntl2* genes are very similar to one another (97 % identity) and to the sequence of sf3 from sunflower (70 % identity; (Baltz *et al.*, 1992a.)), a fragment of which was used to isolate the *Ntl* genes. These genes are the only examples of *lim* genes to have been isolated from plants. All three encoded proteins consist mainly of two LIM-domains, separated by 47 amino acid residues. Each LIM repeat contains a conserved pattern of cysteine and histidine residues which form a secondary structure motif implicated in protein-protein interactions (Dawid *et al.*, 1995). However a significant difference between the NTL and SF3 proteins is the absence of a carboxy-terminal acidic pentapeptide repeat which was identified as a putative acidic activator domain in SF3. This region contains a number of serine and threonine residues which are potential sites for phosphorylation. Despite possessing zinc-finger and acidic activator domains sequence-specific DNA-binding has not been demonstrated for SF3 or any other LIM-only proteins.

Assuming the NTL LIM-domains represent a protein-binding motif, Far-Western blot analysis using phosphorylated protein could be used to investigate potential protein interactions with pollen proteins. Once cloned into an expression vector, the NTL proteins could be expressed as a recombinant protein in *E. coli* and *in vitro* phosphorylated by a cAMP dependant protein kinase. Using a ^{32}P-phosphorylated protein in a Far-Western blot, pollen proteins separated by polyacrylamide gel

electrophoresis and immobilised on a membrane could be tested for protein-protein interactions. Alternatively, a yeast genetic protein-protein interactive assay could be used to clone interacting proteins (Fields and Sternglanz, 1994). A 'bait' for the system would be the phosphorylated NTL LIM-domain, used to screen the mature pollen cDNA library. This later strategy has been used to clone LIM interacting proteins from mice (Goyal and Longmore, 1995) and the ENIGMA LIM-protein which interacts with the insulin receptor (Wu and Gill, 1994).

7.1.2 Further characterisation of *myb.Nt* and *Ntl* cDNAs

Northern blot analysis was used to investigate the expression of the *myb.Nt* and *Ntl* genes in *N. tabacum* (Chapter 5). Their spatial pattern of transcript accumulation suggested that they are highly expressed specifically in pollen. Recently, a group of monoclonal antibodies raised against NTL1 was used to demonstrate the presence of the protein in pollen. Identification of a protein band of approximately 30 kDa in *N. tabacum* pollen extract was absent in leaf extract (Dr. A. Steinmetz, personnel communication). The protein recognised by this protein in the pollen extract was larger than that predicted for the NTL1 protein (21 kDa). This discrepancy has been reported for a number of other regulatory proteins including *Drosophila* ZESTE, yeast GCN4 (Benson and Pirrota, 1987), (Hope and Struhl, 1986) and MYB.PH3 (Avila *et al.*, 1993), and appears to be the result of anomalous behaviour in SDS-PAGE (section 2.5.4). One antibody recognised the interlim domain and the other the second LIM-repeat.

Antibodies specific to the NTL and MYB.NT proteins could be used to localise the presence of the proteins in anther sections by *in situ* hybridisation experiments. Alternatively, direct immunoblot analysis of pollen nuclear and cytoplasmic fractions on a Western blot could be used to localise the proteins in the pollen grain. Nuclear localisation of the MYB.NT proteins would provide further evidence that these proteins function as transcription factors. A cytoplasmic localisation of the NTL proteins on the other hand would suggest a function in protein-protein interactions to regulate the activity of other proteins.

In situ hybridisation using riboprobes or antisense oligonucleotides has been used to provide direct evidence for the presence of transcripts within the vegetative cell cytoplasm of tomato (Hanson *et al.*, 1989), *Oenothera* (Brown and Crouch, 1990), *Brassica campestris* (Theerakulpisut *et al.*, 1991) and *N. tabacum* (Reijen *et al.*, 1991) in mature or germinated pollen. While these studies did not distinguish between the

194

presence of transcripts in the generative and vegetative cell cytoplasm, a novel approach has involved linking the *lat52* promoter to a nuclear targeting signal of the *N. tabacum* etch virus (Restrepo *et al.*, 1990) and the β-glucuronidase reporter gene (Twell, 1992). In this way the *lat52* promoter was shown to be specifically activated in the vegetative nucleus and not in the generative cell nucleus during pollen development or pollen tube growth. Considering the low transcriptional activity of the generative cell, it is likely that the *myb.Nt* and *Ntl* genes and their protein products are preferentially or specifically activated in the vegetative cell based on the high level of expression of these genes in pollen.

Pollen-specific *myb.Nt* and *Ntl* mRNA were first detectable at a low level in late microspores immediately before PMI and a lower level still in young bicellular pollen after PMI. Steady-state transcript levels then increased dramatically in mid-bicellular pollen at approximately 3.5 days before anthesis and continued to increase at a constant rate to anthesis but declined in germinating pollen. However the high abundance of the *myb.Nt* and *Ntl* transcripts in mature pollen and their presence in germinating pollen suggests that the translation products may play a role in germination or early tube growth. In comparison to the temporal expression profiles of previously characterised 'late' genes, such as *ntp303* (Weterings *et al.*, 1992) and *npg1* (Tebbutt *et al.*, 1994) the profiles of the *myb.Nt* and *Ntl* genes represented a largely late pattern of accumulation. However, the significance of this low level 'pulse' of transcript accumulation (1.4 % *myb.Nt* and 0.25 % *Ntl* transcript relative abundance) is not clear at this time. A similar pulse of transcript accumulation was seen for *ntp303* and *npg1*. However, in previous temporal Northern analysis this early pattern of transcript accumulation was not observed because the studies did not characterise separate early and late microspore stages. Furthermore, anther RNA was used which would decrease the sensitivity of detection of small changes in microspore and pollen transcript levels. Assuming that the low level expression seen in microspores is not due to cross contamination with late pollen stages, passage through PMI was not necessary for the initial transcriptional activation of *myb.Nt* and *Ntl* promoters in *N. tabacum*, but may be an important switch for high level activation and transcription of these genes.

Detailed analysis of transgenic *N. tabacum* and *Arabidopsis* plants containing the *lat52*-β-glucuronidase fusion allowed a direct comparison of the activation and developmental regulation of the *lat52* promoter in bicellular and tricellular pollen (Eady *et al.*, 1994). In *Arabidopsis*, the *lat52* promoter is activated immediately prior to PMI and in *N. tabacum* was activated immediately after PMI, but during pollen maturation

both species showed a very similar pattern of activity. A model could be proposed whereby accumulation of late gene transcription factors prior to PMI and accumulation to threshold levels after PMI activates late gene promoters (Eady *et al.*, 1994). This would indicate that the close correlation between observed late gene promoter activation and PMI may represent a coincidental relationship. Experimental support for the uncoupling of PMI and vegetative cell specific gene transcription involved the treatment of cultures of transgenic *N. tabacum* microspores containing nuclear targeted *lat52*-β-glucuronidase fusion with colchicine. Low levels of colchicine induced symmetrical division at PMI producing two similar sized daughter cells in which the generative cell chromatin condensation was absent (Eady *et al.*, 1995). Further, *lat52* promoter activation occurred in both symmetrical daughter cells. These results demonstrate that division asymmetry at PMI is essential for correct generative cell differentiation and that activation of vegetative cell specific transcription is independent from cell division at PMI. Two models were proposed to explain the relationship between activation or accumulation of transcription factor expression and late gene expression in the vegetative cell. The passive repression model proposes that a transcription factor is asymmetrically distributed or activated within the microspore such that unequal cell division partitions the active factor specifically to the vegetative cell. In this way, generative cell repression of vegetative transcription occurs passively. In the active repression model, vegetative cell-specific genes are not activated in the generative cell due to the asymmetrical distribution of a generative cell-specific repressor (GCR) in the microspore. One possible mode of action of the GCR would be to maintain chromatin condensation in the generative cell.

By investigating the accumulation of protein levels in the same developmental stages defined for temporal Northern analysis by using antibodies specific to the MYB.NT and NTL proteins in developmental Western blots, it would be possible to correlate accumulation of protein to transcript levels. A difference in the relative accumulation of transcript and protein levels may indicate translational or post-translational regulation. Further, a monoclonal antibody which recognises an epitope unique to each of the MYB.NT and NTL proteins could be used on these blots to investigate if the gene pairs are coordinately regulated during pollen development.

7.2 An investigation into the role of MYB.NT and NTL *in vivo*

7.2.1 Antisense and co-sense suppression *myb.Nt* constructs causes reduced male transmission

Three plant lines were identified which showed stable reduced transmission through the transgenic male line (RMT): *lat52 myb.Nt1* sense plant line 10, *lat52 myb.Nt1* antisense plant line 15 and *lat52 myb.Nt1* antisense plant line 16. Kanamycin resistance in seedlings was used as the assay for determining segregation ratios. Resistance was scored at 10-14 days post-germination. A plant from each of the lines showed reduced levels of endogenous *myb.Nt* transcript in pollen (approximately 50 % to 65 % of wild type levels). These observations suggest that reduced *myb.Nt* transcript levels in lines 15 and 16 occurs through antisense gene suppression and in line 10 by sense cosuppression. Detection of MYB.NT protein levels using Western blot analysis could be used to further establish the correlation between the downregulation of *myb.Nt* message and protein levels and the genetic data.

Transgene transcript could not be detected in the three RMT lines using a *CaMV 35S* 3' probe to a Northern blot. This maybe due to degradation of the transgene and endogenous transcript duplex. Alternatively, termination of the transgene transcript at the *myb.Nt1* polyadenylation site rather than that of the *CaMV 35S* site could have occurred.

Preliminary studies suggest that a correlation exists between the genetic ratios and the proportion of pollen grains which show a small and collapsed phenotype in the three RMT lines (approximately 55 % to 60 % in each line). *Myb.Nt1* has a gametophytic expression which in a heterozygote plant means that 50 % of the pollen grains carry the T-DNA construct. Therefore the presence of the antisense and sense *myb.Nt1* T-DNA construct probably causes the collapsed pollen phenotype. More extensive counting of kanamycin resistant and sensitive seedlings are needed here to characterise the genetic ratios more accurately. Unlike the *lat52* abnormal shape antisense pollen (Muschietti *et al.*, 1994), the *myb.Nt* small and collapsed pollen grains were shown by cytochemical staining and *in vitro* germination to be dead.

Previous studies with pollen-specific genes have used a variety of techniques to investigate their function. Tissue 'printing' was used to immunolocalise the polygalacturonase-like P2 protein to the pollen-tube which correlated with a possible role in pollen-tube wall synthesis (Brown and Crouch, 1990). *In situ* immunolocalisation was used to show that a maize exo-polygalacturonase had a

197

cytosolic location before germination, but was present in the pollen-tube wall after germination (Dubald *et al.*, 1993). On the other hand sporophytically determined genetic mutants, deficient in the particular protein under study, have also been used to investigate pollen function. The *Arabidopsis pop1* mutant, defective in wax formation in the pollen coat is male sterile under low humidity, which suggests that lipids in the pollen coat are important in pollen hydration (Preuss *et al.*, 1993). An *arpt⁻* mutant of *Arabidopsis* (adenine phosphoribosyl transferase) was shown to have defective pollen, suggesting a role in the nucleotide salvage pathway in pollen development (Regan and Moffatt, 1990).

Antisense gene suppression has also been used in the investigation of pollen function, to show that the LAT52 protein is essential for pollen hydration and germination (Muschietti *et al.*, 1994). Over 60 % of transgenic tomato plants containing an antisense transgene consisting of the *lat52* promoter driving expression of the lat52 cDNA showed a 1 : 1 segregation in selfcrosses. These segregation ratios were correlated with decrease in *lat52* mRNA and protein levels compared to transgenic plants exhibiting Mendelian ratios. *In vitro* germination showed 50 % of the pollen grains had abnormal hydration but were normal looking in a higher water potential medium. *In vivo* germination experiments further indicated that 50 % of the germinating pollen grains arrested in the style.

The lack of a phenotype in pollen from antisense and sense *Ntl1* expression (*lat52 Ntl1* sense plant line 7) suggests that the *Ntl* genes may be redundant in pollen. End sequencing of further cDNA clones isolated from the heterologous screen of the pollen cDNA library with the *sf3* LIM-domain probe could result in the isolation of further *lim* cDNAs sufficiently divergent from the *Ntl* sequences so that their expression is not affected by transgene transcript. A simpler explanation for the apparent lack of a *Ntl* pollen phenotype, was that a phenotype was not observed due to lack of an appropriate assay. Further studies using pollen samples from *Ntl* expressing transgenic plants could include *in vivo* germination on tobacco pistils and rehydration of pollen in media of different water potentials to investigate hydration properties.

Clearly more transgenic plants need to be generated and analysed to characterise more accurately the RMT phenotype and identify other pollen phenotypes.

7.2.2 Ectopic expression of *myb.Nt1* produces a dwarf phenotype

Section 6.2.7 described how ectopic expression of *myb.Nt1* in transgenic *N. tabacum* produced a dwarf phenotype in six out of eight primary transformant plants. The lack of any observed phenotype with ectopic expression of *Nt11* complemented the lack of any pollen phenotype (section 7.2.1). Kanamycin resistance segregation analysis of R_0 seed from *CaMV 35S myb.Nt1* sense plants revealed Mendelian ratios in selfcrosses and backcrosses indicating that transmission through the male and female germ lines was normal. Indeed, germination of R_0 seed from a selfcross of five of the primary transformants without kanamycin selection showed a 1 : 2 : 1 ratio of putative homozygotes, heterozygotes and null dwarf phenotype seedlings. The transmission of the dwarf phenotype in second generation plants and the demonstration of a direct correlation between increasing height reduction and increasing levels of *myb.Nt1* transcript in leaves of primary transformants strongly suggests that expression of *myb.Nt1* and not an unrelated T-DNA knockout mutation results in the dwarf phenotype. Recently the relationship between height reduction and steady state levels of *myb.Nt1* transcript in leaves was established in second generation dwarf plant lines (Spurr and Twell, unpublished results). While no transcript was detected in putative null plants, higher levels of transcript was detected in putative homozygote plants compared with heterozygotes. These results demonstrate the stable inheritance of *myb.Nt1* expression and support the hypothesis that the three groups of plants seen in each dwarf line correspond to null, heterozygote and homozygote plants. Southern analysis would be needed to confirm that putative null, heterozygote and homozygote plants possess 0, 1 and 2 genetic T-DNA copies respectively. Genetic crosses could confirm the genotype classification, such as that between homozygote and heterozygote plants would be expected to produce a 1 : 1 ratio of homozygotes and heterozygotes.

The growth of seedlings in darkness demonstrated that the dwarf phenotype was independent of light. This suggests that MYB.NT1 does not interact with phytochrome gene expression pathways. Observation of the dwarf phenotype in second generation plants grown in soil under greenhouse conditions revealed a number of distinct phenotypic traits compared to wild type plants and putative null phenotype plants from each line. The dramatic reduction in internode length and alterations of leaf morphology could suggest that the MYB.NT1 protein is interacting, probably by a squelching mechanism, with gibberellin biosynthesis or GA-induced

gene expression. In order to investigate whether the MYB.NT proteins interact with gibberellin biosynthesis exogenous application of GA to dwarf plants might complement the phenotype. While the molecular mechanisms of plant hormone action are not well understood, the involvement of *myb* genes in gibberellin induced gene expression in barley aleurone tissue (*Gamyb*; (Gubler *et al.*, 1995)) and gibberellin induced gene expression in trichomes, ovaries, stomium cells, stigmas and seedlings in *P. hybrida* (*Pmyb92*; (Mur *et al.*, 1994)), provides an insight into how MYB.NT1 could interact with *myb* gene expression to produce the dwarf phenotype. For instance, altered stigma morphology was seen in flowers of dwarf line 5, which could be reflecting an interaction with *Pmyb92* activity in stigmas. In this respect, an interesting observation is that the MYB-domain sequences of GAMYB and MYB.NT1 are similar (85 % amino acid identity), which suggests that the proteins might bind to similar DNA sites.

While no definite correlation between the dwarf phenotype and intensity of pigmentation was seen, plant *myb* genes have been shown to be involved with activation of flavonoid biosynthetic genes (*Cl, Pl, Zmc1, Zmc38* from maize, *Am305* from *Antirrinhum* and *myb.Ph3* from *P. hybrida*, (Cone *et al.*, 1986); (Cone *et al.*, 1993); (Franken *et al.*, 1994); (Grotewold *et al.*, 1991); (Paz-Ares *et al.*, 1987); (Sablowski *et al.*, 1994); (Avila *et al.*, 1993)). A link between activation of flavonoid enzymes and gibberellins was provided by the demonstration that flavonoid gene transcript levels such as chalcone synthase, are increased in corollas' by the application of gibberellins, which probably act by inducing transcription factor gene expression (Weiss *et al.*, 1992). With regard to the demonstration that *myb.Ph3* activates transcription from binding sites in the *P. hybrida* chalcone synthase gene (*chsJ*) promoter in *N. tabacum* protoplasts (Solano *et al.*, 1995a.) and the similarity of the MYB-domain to that of the MYB.NT proteins it would be interesting to investigate if these genes are induced by gibberellin application and if this constitutes a common mechanism of *myb* gene expression.

A closer analysis of phenotypic changes in homozygous dwarf plants from each line could involve an analysis of alterations in secondary metabolic pathways by extraction and quantification of phenylpropanoids such as anthocyanins, phlobaphenes and phenolics. In order to achieve this, dwarf and control plants would need to be grown under carefully controlled conditions for light. Such studies may reveal with more accuracy the target pathways and genes of MYB.NT1.

By using the homozygote plants as a genetic background, it would be possible to test for protein interactions with MYB.NT1 or test for competition for binding sites by

introducing genes into dwarf lines and screening for restoration of the null phenotype. In this way suppresser genes of the dwarf phenotype could be identified.

While phenotypic characteristics such as internode length and increased axial leaf growth was seen in most of the dwarf plant lines, the altered 'inverted spoon' leaf morphology was only seen in one of the primary transformants and in two other lines in second generation plants. This leaf shape presumably results from a greater division and/or expansion rate of cells in the middle of the leaves than at the edges, suggesting that myb.Nt1 may affect cell-specific gene expression. Thus this leaf phenotype seems to be an additional phenotype seen in some plant lines but not others. A situation where a threshold of myb.Nt1 expression has to be reached before a phenotype is seen in transgenic plants could be envisaged. However, plants from line 5 which display this 'inverted spoon' leaf morphology did not display a high myb.Nt1 transcript level. Addition of extra copies of the myb.Nt1 T-DNA cassette in homozygous plants exhibiting the 'inverted spoon' leaf morphology may lead to a greater degree of radially uneven lamina expansion.

The existence of a novel method for the production of dwarf N. tabacum plants could be of commercial interest to the plant biotechnology industry. Myb gene families have been characterised in other plants including Arabidopsis thaliana(>100 gene members; (Avila et al., 1993)) and P. hybrida (20-40; (Solano et al., 1995a.)). In addition considering the evolutionary conservation of the MYB-domain in plants (Figure 4.11) it seems possible that high level ectopic expression of a specialised myb gene such as the pollen myb.Nt1 in plants via a strong promoter such as CaMV 35S would produce a dwarf phenotype in other plants by the mechanisms discussed in section 6.3.3.

Dwarf plants are used commercially in flowering crop plants such as lilies, chrysanthenums and pointsettias. Here, restrictions in elongated growth can be achieved by applications of gibberellin synthesis inhibitors such as ancymidol (A-rest) or paclobutrazole (Bonzi). The same is true of cereal crops where plants of reduced height are useful as they are less susceptible to lodging (ie. falling over due to high winds). The use of such crops increases the proportion of the crop that can be harvested thus increasing production. Restriction of extension growth in roadside shrub plantings are also achieved with chemical agents.

The process of spraying gibberellin synthesis inhibitors can be expensive, time consuming and may pose the environmental problems of toxicity, to humans and wildlife, including inadvertent spraying of non-crop plants. The production of dwarf varieties through conventional breeding may not always produce plant lines carrying

all desirable features. The production of transgenic plants with stable high level, non-toxic MYB expression may therefore be of commercial and environmental benefit.

7.3 Summary

This thesis describes the investigation of the characteristics and function of genes encoding putative transcriptional regulatory proteins expressed specifically in pollen of *N. tabacum*. The development of the pollen grain is the result of the coordinated cooperation of gene products from both the gametophyte and the surrounding sporophytic tissue of the anther. A group of specifically expressed genes have been the focus of molecular biological studies into pollen development. An overview of the genes expressed during pollen development, their function and the regulatory mechanisms of these genes are described in Chapter 1.

The primary goal of the work presented in Chapters 3 to 6 was to investigate regulatory molecules controlling pollen gene expression. In order to isolate transcription factors regulating known pollen specific genes, a cDNA library was made from mature pollen of *N. tabacum*. Initially, expression screening of this library using probes containing defined *cis*-element sequences from the promoters of a number of pollen specific genes, failed to identify any sequence specific DNA-binding proteins (Chapter 3). Heterologous screening of the cDNA library with DNA probes to conserved MYB and LIM protein domains however, resulted in the isolation of two unique *myb*- and *lim*-related cDNAs: *myb.Nt1, myb.Nt2, Ntl1* and *Ntl2*. The sequence characterisation of these cDNA clones were described in Chapter 4.

Northern blot analysis (Chapter 5) was used to show that the corresponding *myb.Nt* and *Ntl* transcripts were detected specifically in pollen. Transcripts were detected immediately before the mitotic division of the microspore, with levels increasing dramatically in mid bicellular pollen grains with approximately 98 % of the *myb.Nt* and *Ntl* transcripts accumulating in the three days prior to anthesis.

The MYB.NT encoded proteins showed very high overall sequence identity to each other and the MYB domain sequences showed high sequence identity to the domains of two other *myb*-genes: *myb.Ph3* from *P. hybrida* (Avila *et al.*, 1993) and *Gamyb* from Barley (Gubler *et al.*, 1995). These proteins are involved in the expression of flavonoid genes in flowers and gibberellin responsive activation of expression of α-amylase genes in aleurone tissues. Further work needs to be done to establish whether the MYB.NT proteins activates gene expression from similar promoters in pollen.

The NTL encoded proteins showed high overall sequence identity to the SF3 pollen-specific LIM-domain protein from sunflower (Baltz *et al.*, 1992a.). The LIM-domain sequences of the NTL proteins showed relatively high similarity to those of a group of animal LIM-only proteins which have been implicated in the regulation of the function of other proteins by homo- and hetero-dimerisation via the LIM-domain.

The approach taken to investigate the function of the MYB.NT and NTL proteins involved the generation of transgenic *N. tabacum* plants containing pollen targeted antisense, pollen targeted sense and ectopic expression constructs (Chapter 6). Segregation analysis identified plants which showed a reduction in the viability of their pollen. Two *myb.Nt1* antisense plants and one *myb.Nt1* sense plant showed a reduction of endogenous *myb.Nt* transcript levels in pollen. Microscopic examination of the pollen from these plant lines showed that approximately 50 % of the pollen was dead. These results indicate that the *myb.Nt* genes are crucial for pollen development. Overexpression or antisense NTL1 constructs in pollen of transgenic plants did not result in an observable phenotype. One explanation for this observation was that the *Ntl* genes are redundant in pollen development.

Ectopic expression of *myb.Nt1* resulted in several transgenic plant lines exhibiting a dwarf phenotype characterised by a reduction in internodal growth, an alteration in leaf morphology and increased axial leaf growth. An increase in the levels of *myb.Nt1* transcript in the leaves of the dwarf plants increased the severity of the dwarf phenotype. One mechanistic explanation for this phenotype could be that MYB.NT1 is interfering with the expression of genes encoding precursors of plant hormones such as gibberellins, and/or the expression of hormone responsive genes in plants.

The work done in this thesis provides the groundwork for further characterisation and functional analysis of these *N. tabacum* pollen-specific transcriptional regulators (Chapter 7). Finally, the isolation and characterisation of the *myb.Nt* and *Ntl* cDNAs has provided new information and insight regarding regulatory molecules involved in pollen development.

Appendices.

Appendix I: Table of seedling counts: Analysis of transgenic *N.tabacum* plants as judged by kanamycin resistance segregation.

I. First generation kanamycin resistance segregation analysis.

Construct/ Class/Plant #	Male out-cross to untransformed female	Self-cross	Female out-cross to untransformed male

a.*Lat52-myb.Nt1* sense

CLASS I: Single dominant locus (1:1, 3:1, 1:1)

1	120 : 148	45 : 16	82 : 79
4	94 : 121	50 : 17	171 : 152
5	92 : 104	55 : 19	146 : 145
6	98 : 105	50 : 18	110 : 119
7	339 : 338	60 : 25	45 : 48
8	176 : 143	29 : 9	41 : 44
11	146 : 185	58 : 27	29 : 41
12	171 : 160	35 : 20	27 : 34
17	72 : 70	72 : 24	68 : 66
20	88 : 77	60 : 24	46 : 43
21	86 : 85	74 : 33	42 : 40
22	104 : 100	60 : 22	70 : 72
23	47 : 40	74 : 29	76 : 72
25	90 : 92	64 : 22	40 : 42
27	70 : 72	43 : 14	58 : 60
28	56 : 60	72 : 22	80 : 77
30	50 : 52	62 : 29	29 : 42
31	56 : 60	53 : 18	112 : 108
34	60 : 55	60 : 24	72 : 70
35	60 : 58	52 : 21	62 : 60
36	46 : 42	67 : 22	52 : 50
37	52 : 50	66 : 22	54 : 53

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

15	63 : 22	51 : 4	46 : 16
18	60 : 22	66 : 5	78 : 26
19	72 : 20	74 : 5	64 : 20
26	60 : 20	44 : 3	44 :18
29	49 : 14	86 : 8	70 : 30
33	74 : 26	67 : 5	54 : 18

CLASS III: All resistant (1:0, 1:0, 1:0)

32	43 : 0	50 : 0	42 : 0

CLASS IV: All sensitive (0:1, 0:1, 0:1)

3	0 : 129	0 : 105	0 : 85
13	0 : 155	0 : 50	0 : 95
14	0 : 181	0 : 150	0 : 125
16	0 : 50	0 : 49	0 : 46

CLASS V: Anomalous ratios (1:1, 3:1, 1:1)

2	154 : 26b	44 : 17	106 : 12b
9	187 : 15b	55 : 2b	79 : 0b
24	68 : 32b	42 : 14	41 : 20b

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

Construct/ Class/Plant #	Male out-cross to untransformed female	Self-cross	Female out-cross to untransformed male
10	118 : 116	45 : 42b	0 : 56b

b. *Lat52-myb.Nt1* **antisense**
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

1	66 : 65	45 : 15	59 : 60
2	323 : 315	184 : 60	121 : 120
4	235 : 286	56 : 24	44 : 57
5	84 : 89	66 : 30	72 : 88
8	72 : 78	63 : 30	62 : 62
10	50 : 50	76 : 28	70 : 70
11	70 : 78	88 : 29	50 : 45
12	56 : 58	50 : 18	60 : 70
17	62 : 68	78 : 24	46 : 45

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

13	60 : 20	62 : 5	72 : 23

CLASS III: All resistant (1:0, 1:0, 1:0)

18	56 : 0	42 : 0	40 : 0

CLASS IV: All sensitive (0:1, 0:1, 0:1)

9	0 : 50	0 : 47	0 : 62
14	0 : 46	0 : 50	0 : 46

CLASS V: Anomalous ratios (1:1, 3:1, 1:1)

3	240 : 43b	73 : 26	83 : 11b
7	63 : 12b	41 : 10	74 : 40b

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

6	178 : 171	73 : 29	27 : 43a
15	68 : 63	60 : 33a	26 : 88b
16	84 : 80	98 : 46	32 : 62b

c. *CaMV 35S-myb.Nt1* **sense**
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

1	166 : 148	55 : 19	34 : 32
2	155 : 165	71 : 19	50 : 55
4	57 : 62	128 : 42	ND
5	62 : 66	40 : 12	ND
6	138 : 145	113 : 34	ND
8	82 : 78	56 : 18	ND
9	63 : 58	46 : 16	ND

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

3	59 : 18	149 : 9	ND

CLASS V: All sensitive (0:1, 0:1, 0:1)

Construct/ Class/Plant #	Male out-cross to untransformed female	Self-cross	Female out-cross to untransformed male
7	0 : 48	0 : 46	ND

d. *Lat52-Ntl1* sense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

1	116 : 115	73 : 23	102 : 89
2	77 : 88	82 : 24	40 : 45
4	32 : 38	62 : 26	79 : 65
6	103 : 100	163 : 49	36 : 36
8	35 : 39	46 : 15	31 : 37
9	124 : 93	126 : 50	48 : 52
12	64 : 62	72 : 22	46 : 48
13	60 : 56	46 : 14	46 : 40
16	43 : 38	49 : 15	56 : 55
17	62 : 66	40 : 12	58 : 60
18	62 : 61	70 : 25	38 : 43

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

3	247 : 80	50 : 3	61 : 20
5	143 : 50	100 : 8	40 : 12
10	76 : 26	125 : 15	45 : 15
15	70 : 20	62 : 4	43 : 12
20	74 : 23	40 : 4	72 : 20

CLASS V: Anomalous ratios (1:1, 3:1, 1:1)

11	67 : 34b	51 : 6a	61 : 30b
14	60 : 8b	60 : 2b	71 : 7b

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

7	94 : 96	88 : 35	65 : 110a
19	54 : 60	56 : 31a	26 : 47a

e. *Lat52-Ntl1* antisense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

2	211 : 220	100 : 35	77 : 58
3	104 : 115	23 : 9	41 : 39
4	80 : 82	56 : 20	40 : 42
6	168 : 177	42 : 28	30 : 31
8	294 : 330	80 : 25	43 : 40
12	55 : 59	50 : 16	47 : 41
14	84 : 86	96 : 28	68 : 62
16	72 : 68	62 : 20	74 : 64
17	46 : 46	60 : 18	70 : 72
19	36 : 40	46 : 14	58 : 64
20	46 : 40	62 : 20	72 : 70

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

1	104 : 35	168 : 14	136 : 45
7	129 : 51	ND	76 : 27
9	42 : 13	46 : 3	61 : 18
13	58 : 20	56 : 4	62 : 24
15	74 : 24	50 : 4	90 : 30

208

Construct/ Class/Plant #	Male out-cross to untransformed female	Self-cross	Female out-cross to untransformed male

CLASS IV: All sensitive (0:1, 0:1, 0:1)

11	0 : 47	0 : 37	0 : 54

CLASS V: Anomalous ratios (1:1, 3:1, 1:1)

10	56 : 12b	72 : 3b	66 : 32b

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

5	78 : 74	57 : 33a	22 : 34
18	54 : 52	64 : 22	40 : 62a

f. CaMV 35S-Ntl1 sense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

2	331 : 340	68 : 20	43 : 42
3	38 : 48	96 : 31	78 : 71
4	146 : 147	65 : 25	126 : 118
5	210 : 228	160 : 60	89 : 93
6	173 : 136	34 : 18	42 : 48
7	209 : 220	205 : 72	48 : 35
8	72 : 45	24 : 7	43 : 25
9	90 : 111	45 : 17	35 : 33
10	331 : 340	232 : 80	163 : 170
11	64 : 60	46 : 14	62 : 60

CLASS II: Two independent dominant loci (3:1, 15:1, 3:1)

12	88 : 29	46 : 4	88 : 27
13	50 : 15	86 : 6	60 : 16

CLASS V: Anomalous ratios (1:1, 3:1, 1:1)

1	118 : 48b	33 : 4b	104 : 10b

II. Second generation kanamycin resistance segregation analysis.

a. Lat52-myb.Nt1 sense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

11	44 : 47	41 : 21	40 : 57
12	47 : 45	46 : 20	37 : 46

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

10	43 : 46	30 : 19a	58 : 104b

b. Lat52-myb.Nt1 antisense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

4	60 : 55	56 : 24	56 : 73

Construct/ Class/Plant #	Male out-cross to untransformed female	Self-cross	Female out-cross to untransformed male

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

15	48 : 52	56 : 31a	32 : 60b
16	36 : 32	64 : 40b	35 : 47

c. *Lat52-Ntl1* sense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

19	46 : 52	73 : 23	42 : 50

CLASS VI: Transmission through the male gametophyte affected (1:1, 1:1, 0:1)

7	42 : 40	57 : 32a	75 : 97

d. *Lat52-Ntl1* antisense
CLASS I: Single dominant locus (1:1, 3:1, 1:1)

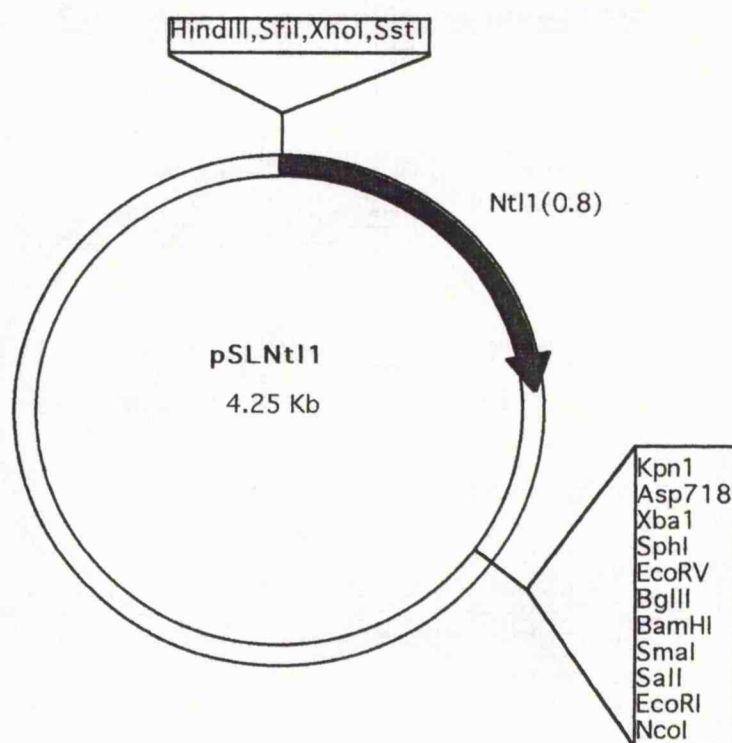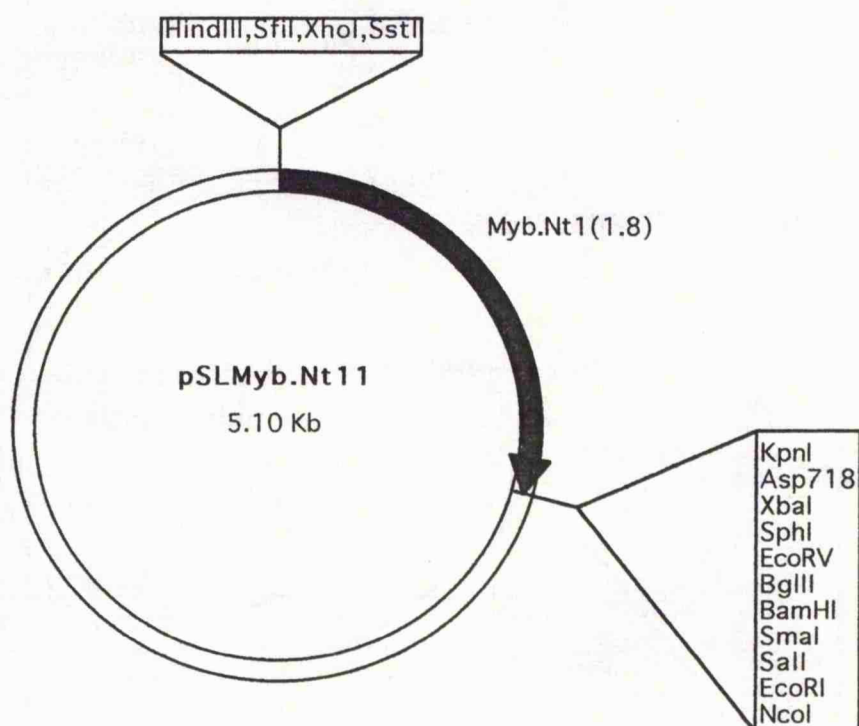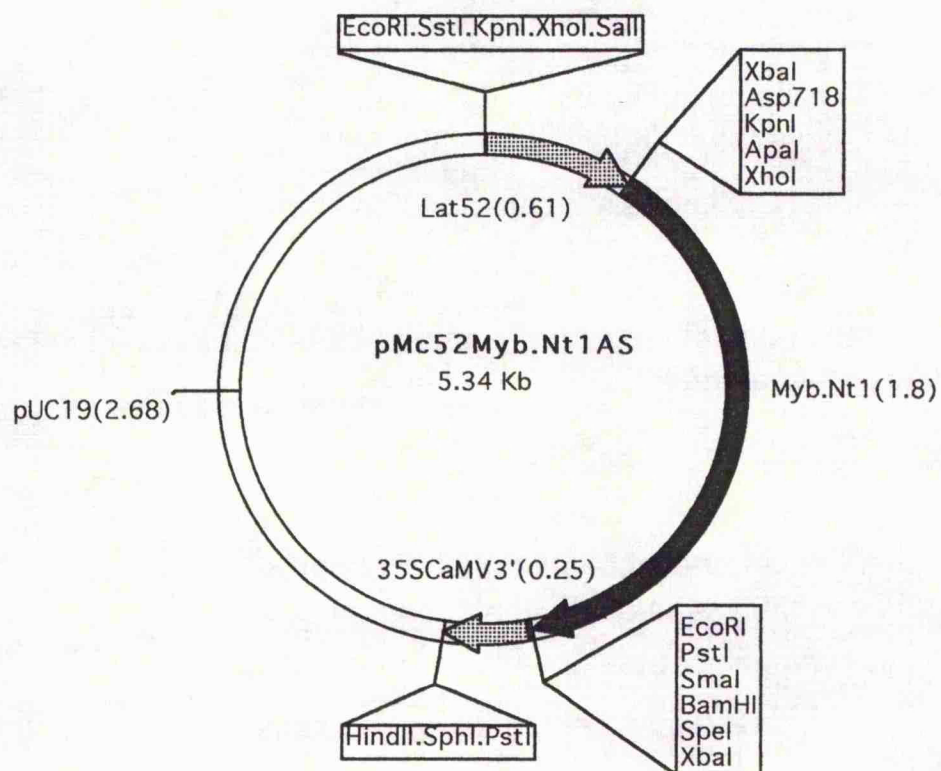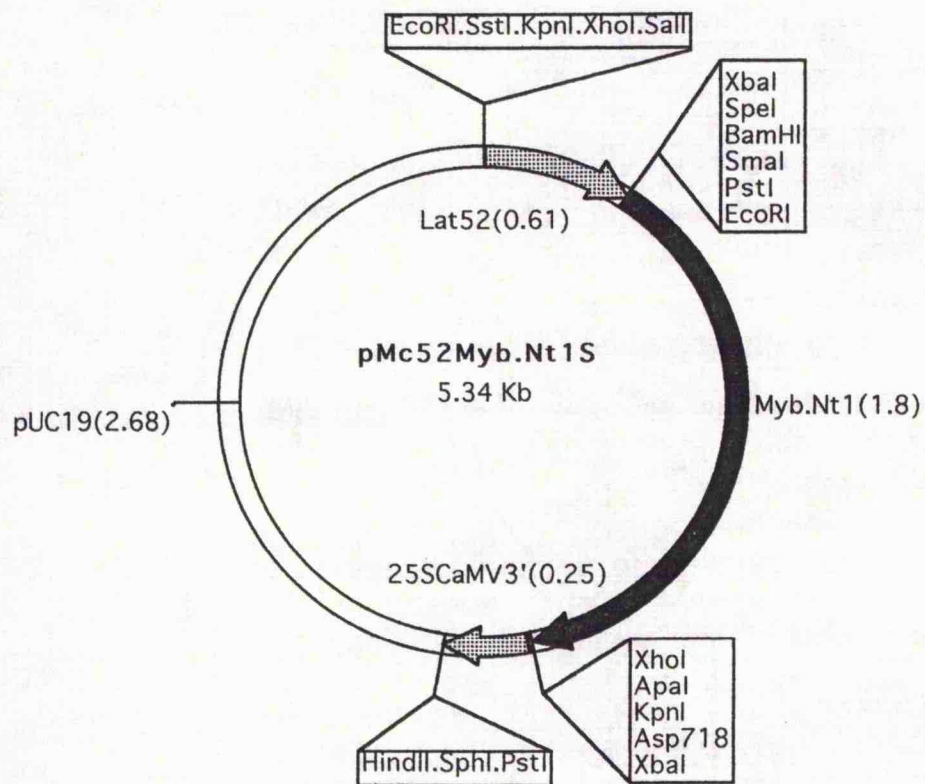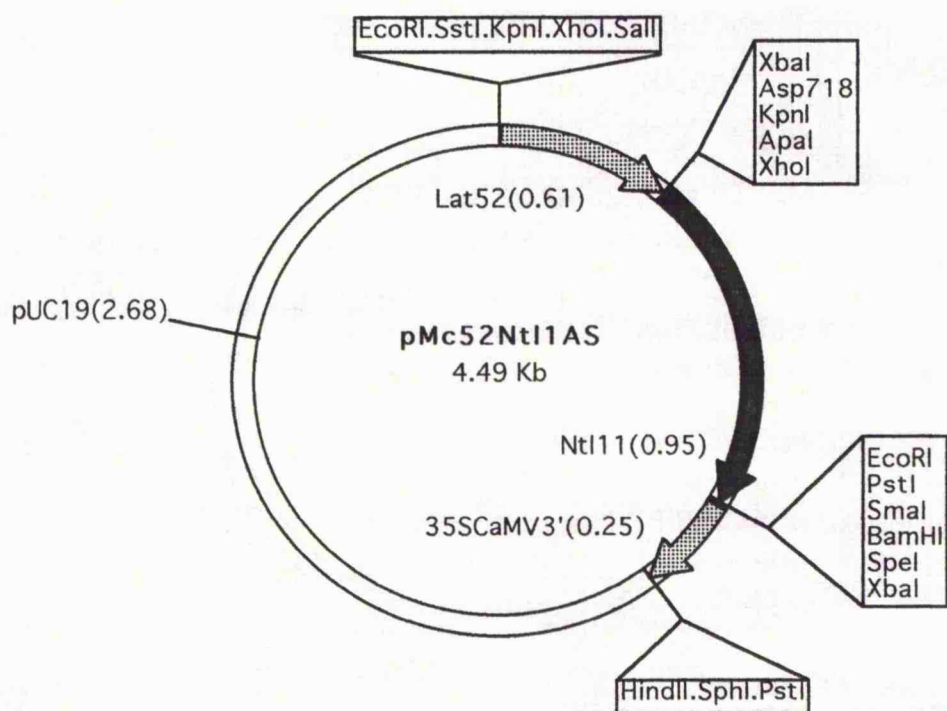|			
5	50 : 48	76 : 36	52 : 72
18	62 : 58	62 : 26	48 : 51

Legend.
a Chi-square > 95 % probability.
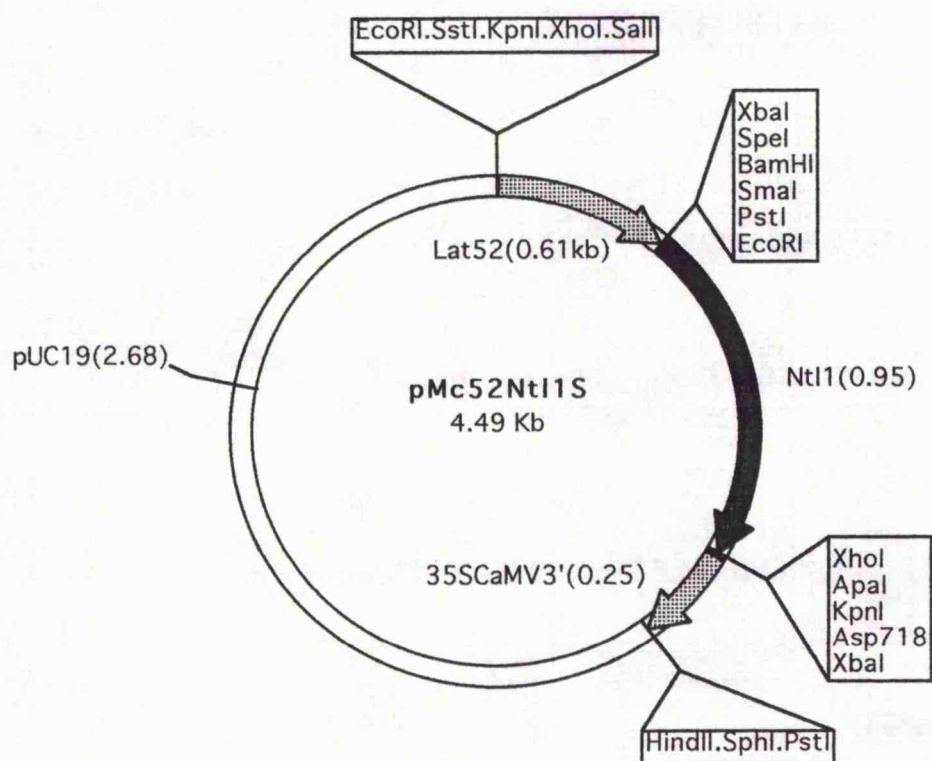b Chi-square > 99 % probability.

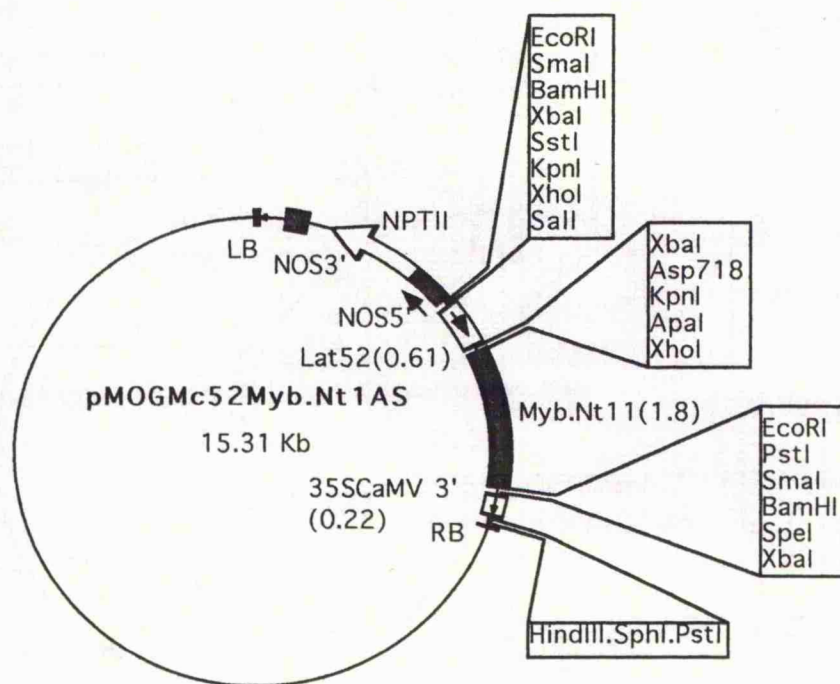# Appendix II: Circular maps of sub-clone and T-DNA vector constructs

(see sections 6.2.1 and 6.2.2).

SacI.BstXI.NotI.EagI.XbaI.SpeI.BamHI.SmaI.PstI.EcoRI

pBSMyb.Nt1
4.76 Kb

Myb.Nt1(1.8kb)

pBluescript SK

KpnI.ApaI.XhoI

SacI.BstXI.NotI.EagI.XbaI.SpeI.BamHI.SmaI.PstI.EcoRI

pBSNtl1
3.91 Kb

Ntl1(0.95kb)

pBluescriptSK

KpnI.ApaI.XhoI

212

EcoRI.SstI.KpnI.XhoI.SalI

Xbal
Spel
BamHI
Smal
PstI
EcoRI

Lat52(0.61)

pMc52Myb.Nt1S
5.34 Kb

pUC19(2.68)

Myb.Nt1(1.8)

25SCaMV3'(0.25)

Xhol
Apal
Kpnl
Asp718
Xbal

HindII.SphI.PstI

EcoRI.SstI.KpnI.XhoI.SalI

Xbal
Asp718
Kpnl
Apal
Xhol

Lat52(0.61)

pMc52Myb.Nt1AS
5.34 Kb

pUC19(2.68)

Myb.Nt1(1.8)

35SCaMV3'(0.25)

EcoRI
PstI
Smal
BamHI
Spel
Xbal

HindII.SphI.PstI

214

EcoRI.SstI.KpnI.XhoI.SalI

Xbal
Spel
BamHI
Smal
Pstl
EcoRI

Lat52(0.61kb)

pUC19(2.68)

pMc52Ntl1S
4.49 Kb

Ntl1(0.95)

35SCaMV3'(0.25)

Xhol
Apal
Kpnl
Asp718
Xbal

HindII.SphI.PstI

EcoRI.SstI.KpnI.XhoI.SalI

Xbal
Asp718
Kpnl
Apal
Xhol

Lat52(0.61)

pUC19(2.68)

pMc52Ntl1AS
4.49 Kb

Ntl11(0.95)

35SCaMV3'(0.25)

EcoRI
Pstl
Smal
BamHI
Spel
Xbal

HindII.SphI.PstI

215

pMOGMc52Ntl1S
14.46 Kb

EcoRI
SmaI
BamHI
XbaI
SstI
KpnI
XhoI
SalI

XbaI
SpeI
BamHI
SmaI
PstI
EcoRI

XhoI
ApaI
KpnI
Asp718
XbaI

NPTII
LB
NOS3'
NOS5'
Lat52(0.61)
Ntl11(0.95)
35SCaMV 3'(0.22)
RB
HindIII.SphI.PstI

pMOGMc52Ntl1AS
14.46 Kb

EcoRI
SmaI
BamHI
XbaI
SstI
KpnI
XhoI
SalI

XbaI
Asp718
KpnI
ApaI
XhoI

EcoRI
PstI
SmaI
BamHI
SpeI
XbaI

NPTII
LB
NOS3'
NOS5'
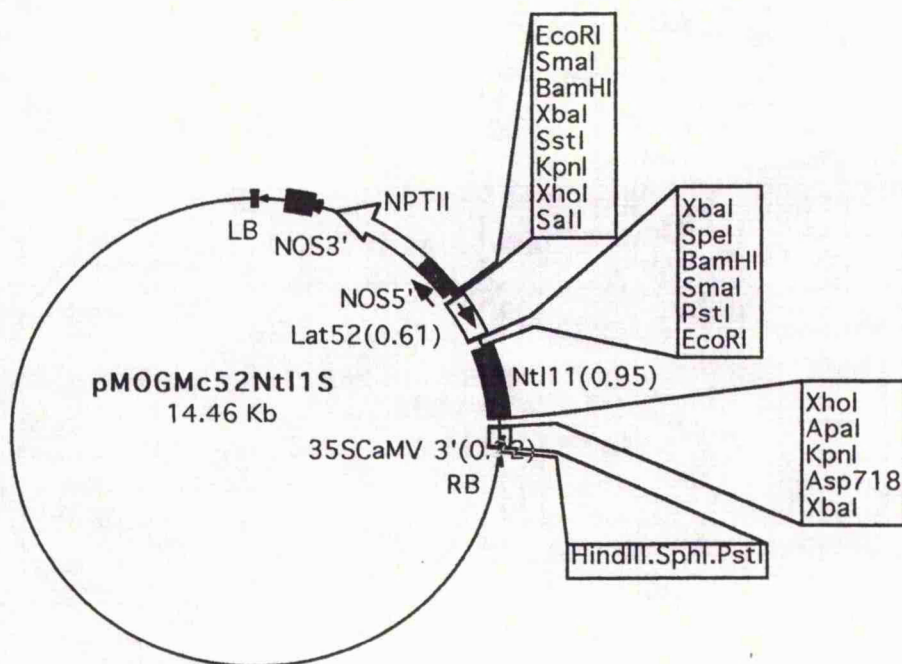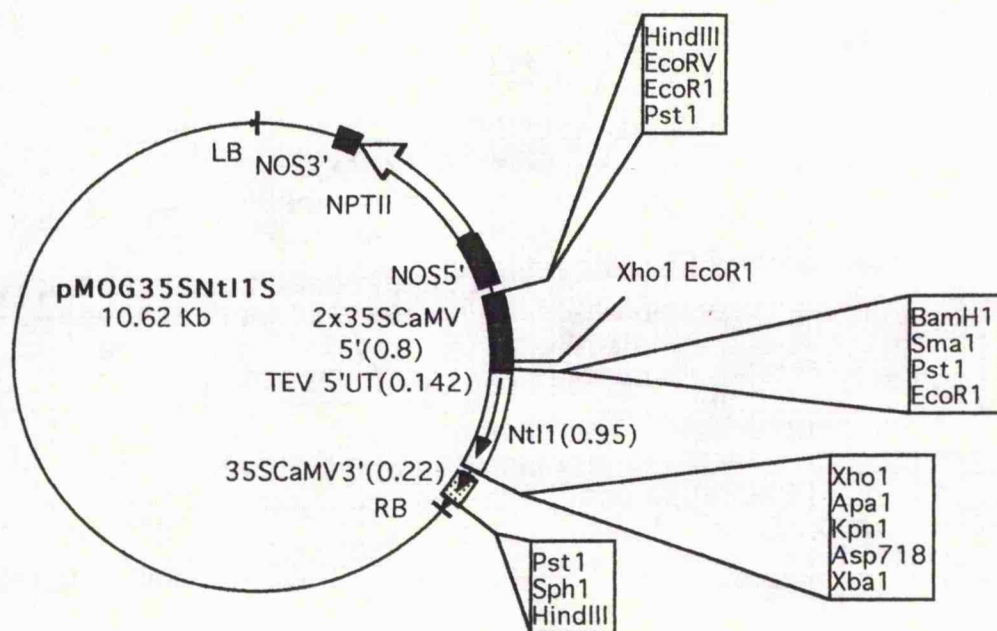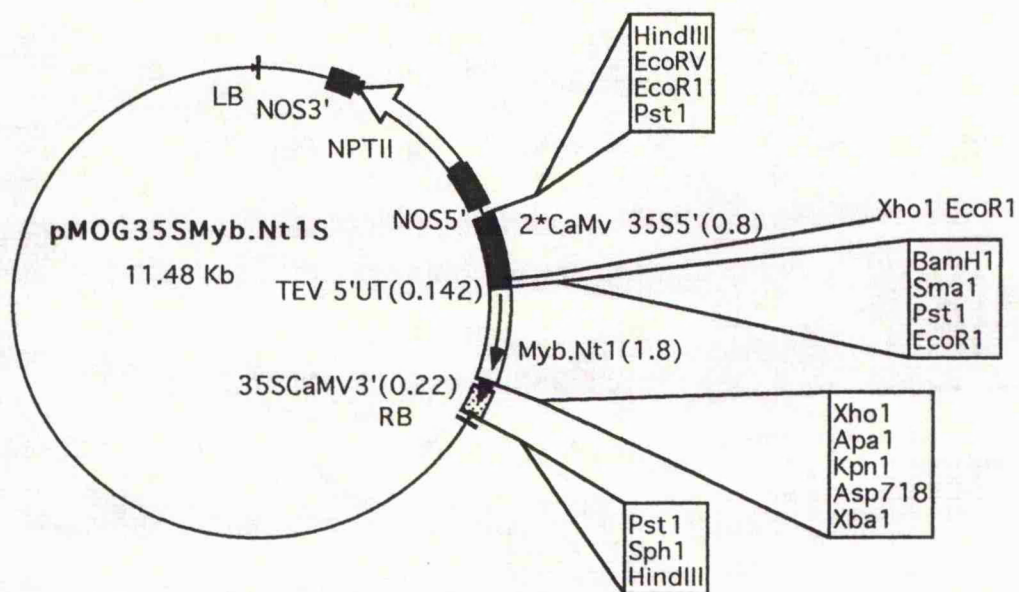Lat52(0.61)
Ntl1(0.95)
35SCaMV 3'
(0.22)
RB
HindIII.SphI.PstI

Bibliography.

Albani, D., Altosaar, I., Arnison, P.G. and Fabijanski, S.F. (1991). A gene showing sequence similarity to pectin esterase is specifically expressed in developing pollen of *Brassica napus*. Sequences in its 5' flanking region are conserved in other pollen-specific promoters. Plant Mol. Biol. 16, 501-513.

Albani, D., Robert, L.S., Donaldson, P.A., Altosarr, I., Arnison, P.G. and Fabijanski, S.F. (1990). Characterisation of a pollen-specific gene family from *Brassica napus* which is activated during early microspore development. Plant Mol. Biol. 15, 605-622.

Albani, D., Sardana, R., Robert, L.S., Altosaar, I., Arnison, P.G. and Fabijanski, S.F. (1992). A *Brassica napus* gene family which shows sequence similarity to ascorbate oxidase is expressed in developing pollen. Molecular characterisation and analysis of promoter activity in transgenic tobacco plants. Plant J. 2, 331-342.

Alexander, M.P. (1969). Differential staining of aborted and non-aborted pollen. Stain Tech. 44, 117-122.

An, C., Ichinose, Y., Yamada, T., Tanaka, Y., Shiraishi, T. and Oku, H. (1993). Organisation of the genes encoding CHS in Pisum sativum. Plant Mol. Biol. 21, 789-803.

Arber, S., Halder, G. and Caroni, P. (1994). Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. Cell 79, 221-231.

Archer, V.E., Breton, J., Sanchez-Garcia, I., Osada, H. and Forster, A. (1994). Cysteine-rich LIM domains of LIM-homeodomain and LIM-only proteins contain zinc and not iron. Proc. Natl. Acad. Sci. USA 91, 316-320.

Avila, J., Nieto, C., Canas, L., Benito, M.J. and Paz-Ares, J. (1993). *Petunia hybrida* genes related to the maize regulatory *C1* gene and to animal *myb* proto-oncogenes. Plant J. 3, 553-562.

Baltz, R., Domon, C., Pillay, D.T.N. and Steinmetz, A. (1992a.). Characterisation of a pollen-specific cDNA from sunflower encoding a zinc-finger protein. Plant J. 2, 713-721.

Baltz, R., Evrard, J.-L., Bourdon, V. and Steinmetz, A. (1996). The pollen-specific LIM protein PLIM-1 from sunflower binds nucleic acids in vitro. Sex Plant Reprod. 9, 264-268.

Baltz, R., Evrard, J.-L., Domon, C. and Steinmetz, A. (1992b.). A LIM motif is present in a pollen-specifc protein. Plant Cell 4, 1465-1466.

Baranowski, N., Frohberg, C., Prat, S. and Willmitzer, L. (1994). A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J. 13, 5383-5392.

Barnabus, B. and Fridvalszky, L. (1984). Adhesion and germination of differently treated maize pollen grains on the stigma surface. Acta Bot. Hung. 30, 329-332.

Barton, K., Binns, A., Matze, A.J.M. and Chilton, M.-D. (1983). Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. Cell 32, 1033-1043.

Bedinger, P. (1992). The remarkable biology of pollen. Plant Cell 4, 879-887.

Bedinger, P. and Edgerton, M.D. (1990a). Developmental staging of maize microspore proteins. Plant Physiol. 92, 474-479.

Bedinger, P. and Edgerton, M.D. (1990b). Developmental staging of maize microspores reveals a transition in developing microspore proteins. Plant Physiol. 92, 474-479.

Benson, M. and Pirrota, V. (1987). The product of the Drosophila zeste gene binds to specific DNA sequences in *white* and *Ubx*. EMBO J. 5, 1387-1392.

Bevan, M.W., Mason, S.E. and Goelet, P. (1985). Expression of tobacco mosaic virus coat protein by a cauliflower mosaic virus promoter in plants transformed with *Agrobacterium*. EMBO J. 4, 1921.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.-H. (1988). Viral myb oncogene encodes a sequence specific DNA-binding activity. Nature 335, 835-837.

Bird, C.R., Ray, J.A. and Fletcher, J.D. (1991). Using antisense RNA to study gene function: inhibition of carotenoid biosynthesis in transgenic tomatoes. Biotechnology 9, 635-639.

Blackwell, T.K. and Weintrub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250, 1104-1110.

Bodeau, J.P. and Walbot, V. (1992). Regulated transcription of the maize *Bronze-2* promoter in electroporated protoplasts requires the *C1* and *R* gene products. Mol. Gen. Genet. 233, 379-387.

Boehm, T., Greenberg, J.N., Buliwela, I., Lavenir, I., Forster, A. and Rabbitts, T.H. (1990). An unusual structure of a putative T-cell oncogene which allows production of similar proteins from distinct mRNAs. EMBO J. 9, 857-868.

Bourgouin, C., Lundgren, S.E. and Thomas, J.B. (1992). Apterus is a *Drosophila* LIM domain protein required for the development of a subset of embryonic muscles. Neuron 7, 881-889.

Brosius, J. (1989). Superpolylinkers in cloning and expression vectors. DNA 8, 759-778.

221

Brown, R.C. and Lemmon, B.E. (1991). Pollen development in orchids. 3. A novel generative pole microtubule system predicts unequal pollen mitosis. J. Cell Sci. **99**, 273-281.

Brown, S.M. and Crouch, M.L. (1990). Characterisation of a gene family abundantly expressed in *Oenothera organensis* pollen that shows sequence similarity to polygalacturonase. Plant Cell **2**, 263-274.

Bryce, W.H. and Nelson, O.E. (1979). Starch synthesizing enzymes in the endosperm and pollen of maize. Plant Physiol. **63**, 312-317.

Budar, F., Thia-toong, L., Van Montagu, M. and Hernalsteens, J.-P. (1986). Agrobacterium mediated gene transfer results mainly in transgenic plants transmitting T-DNA as a single Mendelian factor. Genetics **114**, 303-313.

Cajudo, F.J., Murphy, G., Chinoy, C. and Baulcombe, D.C. (1992). A gibberellin-regulated gene from wheat with sequence homology to cathepsin B of mammalian cells. Plant J. **2**, 937-948.

Campell, M.M., Whetton, R.W. and Sederoff, R.R. (1994). Transcription factors in Loblolly pine xylem. 4th International Congress of Plant Molecular Biology, 220.

Carpenter, J.L., Ploense, S.E., Snustad, D.P. and Silflow, C.D. (1992). Preferential expression of an alpha-tubulin gene of *Arabidopsis* in pollen. Plant Cell **4**, 557-571.

Carrington, J.C. and Freed, D.D. (1990). Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. J. Virol. **64**, 1590-1597.

Chou, P.Y. and Fasman, G.D. (1978). Secondary structure of proteins from their amino-acid sequence. Adv. Enzymol. **47**, 45-148.

Cohen, B., McGuffin, M.E., Pfeifle, C., Segal, D. and Cohen, S.M. (1992). *Apterous*, a gene required for imaginal disc development in Drosophila encodes a member of the LIM family of developmental regulatory proteins. Genes Dev. **6**, 715-729.

Coleman, A.W. and Goff, L.J. (1985). Applications of fluorochromes to pollen biology. Mithramycin and 4', 6-diamidino-2-phenylindole (DAPI) as vital stains and for quantitation of nuclear DNA. stain Technol. **60**, 145-154.

Cone, K.C., Burr, F.A. and Burr, B. (1986). Molecular analysis of the maize anthocyanin regulatory locus C1. Proc. Natl. Acad. Sci. USA **83**, 9631-9635.

Cone, K.C., Cocciolone, S.M., Burr, F.A. and Burr, B. (1993). Maize anthocyanin regulatory gene P1 is a duplicate of C1 that functions in the plant. Plant Cell **5**, 1795-1805.

Crawford, A.W., Michelson, J.W. and Becerle, M.C. (1992). An interaction between zyxin and alpha actinin. J. Cell Biol. **116**, 1381-1391.

Crawford, A.W., Pino, J.D. and Beckerle, M.C. (1994). Biochemical and molecular characterisation of the chicken cysteine-rich protein, a developmentally regulated LIM-domain protein that is associated with the actin cytoskeleton. J. Cell Biol. **124**, 117-127.

Dawid, I.B., Toyama, R. and Taira, M. (1995). LIM domain proteins. C. R. Acad. Sci. Paris. **318**, 295-306.

De Lange, P., Van Blockland, P., Kooter, J.M. and Mol, J.N.M. (1994). In P. Meyer (ed.) *Suppression of genes in higher plants by the introduction of antisense and sense genes*, Springer Verlag, Heidelberg, pp. 57-76.

Dehesh, K., Hung, H., Tepperman, J.M. and Quail, P.H. (1992). GT-2: a transcription factor with twin autonomous DNA-binding domains of closely related but different target sequence specificity. EMBO J. **11**, 4131-4144.

Deroles, S.C. and Gardner, R.C. (1988). Expression and inheritance of kanamycin resistance in a large number of transgenic petunias generated by *Agrobacterium*-mediated transformation. Plant Mol. Biol. **11**, 355-364.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nuc. Acids Res. **12**, 387-395.

Dickinson, H.G. (1993). The regulation of sexual development in plants. Phil. Trans. R. Soc. Lond. B. **339**, 147-149.

Dong, J., Asa, S.L. and Drucker, D.J. (1991). Islet cell and extrapancreatic expression of the LIM domain homeobox gene *Isl-1*. Mol. Endrocrin. **5**, 1633-1641.

Dooner, H.K., Robbins, T.P. and Jorgenson, R.A. (1991). Genetic and developmental control of Anthocyanin biosynthesis. Annu. Rev. Genet. **25**, 173-199.

Dubald, M., Barakate, A., Mandaron, P. and Mache, R. (1993). The ubiquitous presence of exopolygalacturonases in maize suggests a fundamental cellular function for this enzyme. Plant J. **4**, 781-791.

Eady, C., Lindsey, K. and Twell, D. (1994). Differential activation and conserved vegetative cell-specific activity of a late gene promoter in species with bi- and tri-cellular pollen. Plant J. **5**, 543-550.

Eady, C., Lindsey, K. and Twell, D. (1995). The significance of microspore division and division symmetry for vegetative cell-specific transcription and generative cell differentiation. Plant Cell **7**, 65-74.

Ecker, J.R. and Davies, R.W. (1986). Inhibition of gene expression in plant cells by expression of antisense RNA. Proc. Natl. Acad. Sci. USA **83**, 5372-5376.

Edwards, K., Johnstone, C. and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nuc. Acids Res. **19**, 1349.

Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. USA **81**, 140-144.

Feinberg, A. and Vogelstein, A.J. (1983). A technique for radiolabeling DNA restriction enzyme fragments to high specific activity. Anal. Biochem. **137**, 266-267.

Feldbrugge, H., Parniske, M., Halbrock, K. and Weisshar, B. (1994). A novel MYB-like transcription factor, putatively involved in the uv-light activation of the parsley chalcone synthase (chs) promoter. 4th International Congress of Plant Molecular Biology, 210.

Ferguson, E.L., Sternberg, P.W. and Horvitz, H.R. (1987). A genetic pathway for the specification of the vulval cell lineages of the nematode Caenorhabitis elegans. Genetics **110**, 17-72.

Feuerstein, R., Wang, X., DeCheng, S., Cooke, N.E. and Liebhaber, S.A. (1994). The LIM/double zinc-finger motif functions as a protein dimerisation damain. Proc. Natl. Acad. Sci. USA **91**, 10655-10659.

Fields, S. and Sternglanz, R. (1994). The two-hybrid assay system: an assay for protein-protein interactions. TIG **10**, 286-291.

Foley, R.C., Grossman, C., Ellis, J.G., Llewellyn, D.J., Dennis, E.S., Peacock, W.J. and Singh, K.B. (1993). Isolation of a maizw bZIP protein subfamily: candidates for the ocs-element transcription factor. Plant J. **3**, 669-679.

Foos, G., Natour, S. and Klempneuer, K.H. (1993). TATA-box dependant transactivation of the human HSP70 promoter by myb proteins. Oncogene **8**, 1775-1782.

Frampton, J., Gibson, T.J., Ness, S.A., Doderlein, G. and Graf, T. (1991). Proposed structure for the DNA-binding domain of the myb oncoprotein based on model building and mutational analysis. Protein Eng. **4**, 891-901.

Franken, P., Schrell, S., Peterson, P.A., Saedler, H. and Wienand, U. (1994). Molecular analysis of protein domain function encoded by the *myb*-homologous maize genes *C1*, *Zm1* and *Zm38*. Plant J. **6**, 21-30.

Freyd, G., Kim, S.K. and Horvitz, H.R. (1990). Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. Nature **344**, 876-879.

Fried, M. and Crothers, D.M. (1981). Equilibria and kinetics of the *lac* repressor-operator interactions by polyacrylamide electrophoresis. Nuc. Acids Res. **9**, 6505-6525.

Frova, C., Binelli, G. and Ottaviano, E. (1987). In M. C. Rattazzi and J. G. Scandalios (eds.), *Isozyme and hsp gene expression during male gametophyte development in maize*, Alan R. Liss, New York, Vol. 15, pp. 97-120.

Gaillard, A., Vergne, P. and Beckert, M. (1991). Optimisation of maize microspore isolation and culture conditions for reliable plant regeneration. Plant Cell Reprod. **10**, 55-58.

Garner, M.M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nuc. Acids Res. **9**, 3047-3060.

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting secondary structure of globular proteins. J. Mol. Biol. **120**, 97-120.

German, M.S., Juehu, W., Chadwick, R.B. and Rutter, W.J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes & Dev. **6**, 2165-2176.

Gerwirtz, A.M. and Calabretta, B. (1988). A c-myb antisense oligonucleotide inhibits normal human hematopoesis *in vitro*. Science **242**, 1303-1306.

Gewirtz, A.M., Anfossi, G., Venturelli, D., Valpreda, S., Sims, R. and Calabretta, B. (1989). G_1/S transistion in normal T-lymphocytes requires the nuclear protein encoded by c-myb. Science **242**, 1303-1306.

Gilmartin, P.M., Memelink, J., Hiratsuka, K., Kay, S.A. and Chua, N.-H. (1992). characterisation of a gene encoding a DNA binding protein with specificity for a light responsive element. Plant Cell **4**, 839-849.

Goff, S., Cone, K.C. and Chandler, V.L. (1992). Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional interaction between two classes of regulatory proteins. Genes Dev. **6**, 864-875.

Goff, S.A., Cone, K.C. and M.E., F. (1991). Identification of functional domains in the maize transcriptional activator C1: comparision of wild-type and dominant inhibitor proteins. Genes Dev. **5**, 298-309.

Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P. and Chandler, V.L. (1990). Transactivation of anthocyanin biosynthetic genes following transfer of *B* regulatory genes into maize tissues. EMBO J. **9**, 2517-2522.

225

Goodspeed, T.H. (1954). The genus *Nicotiana*: Origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics. Watham: Chronica Botanica.

Goyal, R.K. and Longmore, G. (1995). A novel mouse LIM domain protein expressed early during development. Workshop on LIM proteins and the LIM domain, 19.

Graf, T. (1992). Myb: a transcriptional activator linking proliferation and differentiation in hematopoietic cells. Curr. Opin. Genet. Dev. **2**, 249-255.

Grasser, F.A., LaMontagne, K., Whittaker, L., Stohr, S. and Lipsick, J.S. (1992). A highly conserved cysteine in the v-MYB DNA binding domain is essential for transformation and transcriptional transactivation. Oncogene **7**, 1005-1009.

Green, P.J., Yong, M.-H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P. and Chua, N.-H. (1988). Binding site requirements for pea nuclear factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. EMBO J. **7**, 4035-4044.

Grierson, D., Fray, R.G., Hamilton, A.J., Smith, C.J.S. and Watson, C. (1991). Does co-suppression of sense genes in transgenic plants involve antisense RNA. Trends Biotechnol. **9**, 122-123.

Grotewold, E., Athma, P. and Peterson, T. (1991). Alternatively spliced products of the maize *P* gene encode proteins with homology to the DNA-binding domain of *myb*-like transcription factors. Proc. Natl. Acad. Sci. USA **88**, 4587-4591.

Grotewold, E., Drummond, B.J., Bowen, B. and Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell **76**, 543-553.

Gubler, F., Kalla, R., Roberts, J.K. and Jacobson, J.V. (1995). Gibberellin-regulated expression of a myb gene in Barley aleurone cells: Evidence for myb tranactivation of a high-pI α-amylase gene promoter. Plant Cell **7**, 1879-1891.

Guerrero, F.D., Crossland, L., Smutzer, G.S., Hamilton, D.A. and Mascarenhas, J.P. (1990). Promoter sequences from a maize pollen-specific gene directs tissue-specfic transcription in tobacco. Mol. Gen. Genet. **224**, 161-168.

Hamilton, A.J., Lycett, G.W. and Grierson, D. (1990). Antisense gene that inhibits synthesis of the hormone ethylene in transgenic plants. Nature **346**, 284-287.

Hanson, D.D., Hamilton, D.A., Travis, J.I., Bashe, D.M. and Mascarenhas, J.P. (1989). Characterisation of a pollen-specific cDNA clone from *Zea mays* and its expression. Plant Cell **1**, 173-179.

He, X.Y., Antao, V.P., Basila, D., Marx, J.C. and Davies, B.R. (1992). Isolation and molecular characterisation of the human CD34 gene. Blood **79**, 2296-2302.

Heslop-Harrison, J. (1987). In J. K. Giles and J. Prakash (eds.), *Pollen germination and pollen tube growth*, Academic Press, London, Vol. 107, pp. 1-78.

Heslop-Harrison, J., Heslop-Harrison, Y. and Shivanna, K.R. (1984). The evaluation of pollen quaility, and further appraisal of the fluorochromatic (FCR) test. Theor. Appl. Genet. **67**, 367-375.

Higgins, D.G. and Sharp, P.M. (1989). CLUSTAL: a package for perfrominig multiple sequence alignments on a microcomputer. Gene **73**, 237-244.

Hodge, R., Paul, W., Draper, J. and Scott, R. (1991). Cold plaque screening: a simple technique for the isolation of low abundance, differentially expressed transcripts from conventional cDNA libraries. Plant J. **2**, 257-260.

Hoekema, A., van Haaren, M.J.J., Fellinger, A.J., Hooykaas, P.J.J. and Schilperoort, R.A. (1985). Non-oncogenic plant vectors for use in the Agrobacterium binary system. Plant Mol. Biol. **5**, 85-89.

Hoekstra, F.A. and Bruinsma, J. (1979). Protein synthesis of binucleate and trinucleate pollen and its relationship to tube emergence and growth. Planta **146**, 559-566.

Hope, I.A. and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein. Cell **46**, 885-894.

Horner, H.T. (1977). A comparative light and electron microscopic study of microsporogenesis in male-fertile and cytoplasmic male-sterile sunflower *(Helianthus annus)*. Can. J. Bot. **52**, 435-449.

Horsch, R.B., Fry, J.E., Hofman, N.L., Eicholtz, D., Rogers, S.G. and Fraley, F.T. (1985). A simple and general method for transferring genes into plants. Science **227**, 1229-1231.

Huiling, X., Knox, B., Taylor, P.E. and Singh, M.B. (1995). Bcp1, a gene required for male fertility in Arabidopsis. Proc. Natl. Acad. Sci. USA **92**, 2106-2110.

Ingraham, H.A., Chen, R., Mangalam, H.J., Eisholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfield, M.G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell **55**, 519-529.

Iturriaga, G., Leyns, L., Gharaibeh, R., Hernandez, F., Salamini, F. and Bartels, D. (1994). Isolation and characterisation of a MYB-related gene from the resurrection plant *Craterostigma plantagineum*. 4th International Congress of Plant Molecular Biology, 216.

Jackson, D., Culianez-Marcia, F., Prescott, A.G., Roberts, K. and Martin, C. (1991). Expression patterns of *myb* genes from *Antirrhinum* flowers. Plant Cell 3, 115-125.

Jackson, S.P., MacDonald, J.J., Lees-Miller, S. and Tijan, R. (1990). GC box binding induces phosphorylation of SP1 by a DNA-dependant protein kinase. Cell 63, 155-165.

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

Jorgenson, R. (1992). Silencing of plant genes by homologous transgenes. Agbiotech. News Inf. 4, 265-273.

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

Kamaley, J.C. and Goldberg, R.B. (1980). Regulation of structural gene expression in tobacco. Cell 19, 935-946.

Kanei-Ishii, C., sarai, A., Sawzaki, T., He, D.H., Ogata, K., Nishimura, Y. and Ishii, S. (1990). The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the *myb* proto-oncogene product. J. Biol. Chem. 265, 19990-19995.

Kang, H.-G., Noh, Y.-S., Chung, Y.-Y., Costa, M.A., An, K. and An, G. (1995). Phenotypic alterations of petal and sepal by ectopic expression of a rice MADS box gene in tobacco. Plant Mol. Biol. 29, 1-10.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cyc-His domain. Nature 344, 879-882.

Katagiri, F., Lam, E. and Chua, N.-H. (1989). Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature 340, 727-730.

Katzen, A.L., Kornberg, T.B. and Bishop, J.M. (1985). Isolation of the proto-oncogene *c-myb* from *D. melanogaster*. Cell 41, 449-456.

Kay, R., Chan, A., Daly, M. and McPhearson, J. (1987). Duplication of the CaMV 35S promoter sequences creates a strongh enhancer for plant genes. Science 236, 1299-1302.

Kiess, M., Scharm, B., Aguzzi, A., Hajnal, A., Klemenz, R., Schwarte-Waldhoff, I. and Schafer, R. (1995). Expression of *ril*, a novel LIM domain gene, is down-regulated in HRAS-transformed cells and restored in phenotypic revertants. Oncogene 10, 61-68.

Klempnauer, K.-H., Gonda, T.J. and Bishop, J.M. (1982). Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: the architecture of a transduced oncogene. Cell **31**, 453-453.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.B. (1990). Different temporal and spatial gene expression patterns occur during anther development. Plant Cell **2**, 1201-1224.

Korzh, V., Edlund, T. and Thor, S. (1993). Zebrafish primary neurons initiate expression of the LIM homeodomain protein Isl-1 at the end of gastrulation. Development **118**, 417-425.

Kosa, J.L., Michelson, J.W., Louis, H.A., Olsen, J.I. and Davies, D.R. (1994). Common metal ion coordination in LIM domain proteins. Biochemistry **33**, 468-477.

Kozak, M. (1986). Point mutations define a sequence flanKing the AUG intiation codon that modulates translation by eukaryotic ribosomes. Cell **44**, 283-292.

Ku, D.H.e.a. (1993). c-myb transactivates cdc2 expression via myb binding sites in the 5' flanking region of the human *cdc2* gene. J. Biol. Chem. **268**, 2255-2259.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathis character of a protein. J. Mol. Biol. **157**, 105-137.

LaFountain, J.R. and Mascarenhas, J.P. (1972). Isolation of vegetative and generative nuclei from pollen tubes. Exp. Cell Res. **72**, 233-236.

Larkin, J.C., Oppeinheimer, D.G., Pollock, S. and Marks, D.M. (1993). Arabidopsis *GLABROUS1* gene requires downstream sequences for function. Plant Cell **5**, 1739-1748.

Lech, K., Anderson, K. and Brent, R. (1988). DNA bound Fos proteins activate transcription in yeast. Cell **52**, 179-184.

Leech, M.J., Krammerer, W., Cove, D.J., Martin, C. and Wang, T.L. (1993). Expression of *myb*-related genes in the moss, *Physcomitrella patens*. Plant J. **3**, 51-61.

Li, S.F. and Parish, R.W. (1995). Isolation of two novel *myb*-like genes from *Arabidopsis* and studies on the DNA-binding properties of their products. Plant J. **8**, 963-972.

Lichtsteiner, S. and Tjian, R. (1995). Synergistic activation of transcription by UNC-86 and MEC-3 in *Caenorhabditis elegans* embryo extracts. EMBO J. **14**, 3937-3945.

Liebhaber, S.A., Emery, J.G., Urbanek, M., Wang, X. and Cooke, N.E. (1990). Characterisation of a human cDNA encoding a widely expressed and highly conserved cysteine-rich protein with an unusual zinc-finger motif. Nucl. Acids Res. **18**, 3871-3879.

Lin, Q., Hamilton, W.D.O. and Merrywaether, A. (1996). Cloning and initial characterisation of 14 *myb*-related cDNAs from tomato (*Lycopersicon esculentum* cv. Ailsa Craig). Plant Mol. Biol. **30**, 1009-1020.

Lovering, R., Hanson, I.M., Borden, K.L.B., Martin, S., O'Reilly, N.J., Evan, G.I., Rahman, D., Pappin, D.J.C., Trowsdale, J. and Fremont, P.S. (1993). Identification and preliminary characterisation of a protein motif related to the zinc finger. Proc. Natl. Acad. Sci. USA **90**, 2112-2116.

Lundgren, S.E., Callahan, C.A., Thor, S. and Thomas, J.B. (1995). Control of neuronal pathway selection by the *Drosophila* LIM homeodomain gene *apterous*. Development **121**, 1769-1773.

Luscher, B. and Eisenmann, R.N. (1990). New light on Myc and Myb. Part II. Myb. Genes Dev. **4**, 2235-2241.

Majello, B., Kenyon, L.C. and Dalla-Favera, R. (1986). Human *c-myb* oncogene: Nucleotide sequence of cDNA and organization of the genomic locus. Proc. Natl. Acad. Sci. USA **83**, 9636-9640.

Mandaron, P., Niogret, M.F., Mache, R. and Monegar, F. (1990). In vitro protein synthesis in isolated microspores of Zea mays at several stages of development. Theoret. Appl. Genet. **80**, 134-138.

Marocco, A., Wissenbach, M., Becker, D., Paz-Ares, J., Saedler, H., Salamini, F. and Rohde, W. (1989). Mutiple genes are transcribed in *Hordeum vulgare* and *Zea mays* that carry the DNA binding domain of the *myb* oncoproteins. Mol. Gen. Genet. **216**, 183-187.

Martin, C., Prescott, A., Bartlett, J. and Vrijlandt, E. (1991). Control of antocyanin biosynthesis in flowers of *Antirrhinum majus*. Plant J. **1**, 37-49.

Mascarenhas, J.P. (1975). The biochemistry of Angiosperm pollen development. Bot. Rev. **41**, 259-314.

Mascarenhas, J.P. (1988). In D. P. S. Verma and R. B. Goldberg (eds.), *Anther- and pollen expressed genes*, Springer Verlag, Berlin, pp. 97-115.

Mascarenhas, J.P. (1990). Gene activity during pollen development. Annu. Rev. Plant Physiol. Mol. Biol. **41**, 317-338.

Mascarenhas, N.T., Bashe, D., Eisenberg, A., Willing, R.P., Xiao, C.M. and Mascarenhas, J.P. (1984). Messenger RNAs in corn pollen and protein synthesis during germination and pollen tube growth. Theor. Appl. Genet. **68**, 323-326.

Matraux, J.-P. (1987). In P. J. Davies (ed.) *Gibberellins and plant cell elongation*, Kluwer, Boston, pp. 431-462.

McCormick, S. (1993). Male Gametophyte development. Plant Cell **5**, 1265-1275.

McGuire, E.A., Hackett, R.D., Pollock, K.M., Bartholdi, M.F., O'Brien, S.J. and Korsmeyer, S.J. (1989). The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including *Ttg-1*, a gene encoding a potential zinc finger protein. Mol. Cell Biol. **9**, 2124-2134.

Michelet, B., Lukaszewicz, M., Dupriez, V. and Boutry, M. (1994). A plant plasma membrane proton-ATPase gene is regulated by development and environment and shows signs of a translational regulation. Plant Cell **6**, 1375-1389.

Michelson, J.W., Schmeichel, K.L., Beckerle, M.C. and Winge, D.R. (1993). The LIM motif defines a specific zinc-binding protein domain. Proc. Natl. Acad. Sci. USA **90**, 4404-4408.

Mitchell, P.J. and Tijan, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**, 371-378.

Mizuno, K., Okano, I., Ohashi, K., Nunoue, K., Kuma, K.-i., Miyata, T. and Nakamura, T. (1994). Identification of a human cDNA encoding a novel protein kinase with two repeats of the LIM/double zinc finger motif. Oncogene **9**, 1605-1612.

Morelli, G., Nagy, F., Fraley, R.T., Rogers, S.G. and Chua, N.-H. (1985). A short conserved sequence is involved in the light inducibility of a gene encoding ribulose 1,5 bisphosphate carboxylase small subunit of pea. Nature **315**, 200.

Mucenski, M.I., McLain, K., Kier, A.B., Swerflow, S.H., Schreiner, C.M., Millar, T.A., Pietryga, D.W., Scott, W.J. and Potter, S.S. (1991). A functional *c-myb* gene is required for normal murine fetal hepatic heamatopoesis. Cell **65**, 677-689.

Mulcahy, D.L. (1979). The rise of the Angiosperms: A genecological factor. Science **206**, 20-23.

Mur, L.A., Weiss, D., Spelt, C., Mol, J.N.M. and Kooter, J.M. (1994). A member of the myb family of transcription factors from Petunia is induced by gibberellic acid, and shows a very cell type-specific expression pattern. 4th International Congress of Plant Molecular Biology, 214.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. **15**, 473-497.

Muschietti, J., Dircks, L., Vancanneyt, G. and McCormick, S. (1994). LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilisation. Plant J. **6**, 321-338.

Myrset, A.H., Bostad, A., Jamin, N., Lirsac, P.N., Toma, F. and Gabrielsen, O.S. (1993). DNA and redox state induced conformational changes in the DNA binding domain of the MYB oncoprotein. EMBO J. **12**, 4625-4633.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes in *trans*. Plant Cell 2, 279-289.

Nicolaides, N.C., Gualdi, R., Casadevall, C., Manzella, L. and Calabretta, B. (1991). Positive autoregulation of c-*myb* expression via Myb binding sites in the 5' flanking region of the human c-*myb* gene. Mol. Cell. Biol. 11, 6166-6176.

Noda, K.-I., Glover, B.J., Linstead, P. and Martin, C. (1994). Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature 369, 661-664.

Nomura, N., Takahashi, M., Matasui, M., Ishii, M., Date, T., Sasmoto, S. and Ishizaki, R. (1988). Isolation of human cDNA clones of *myb*-related genes. Nucl. Acids Res. 16, 11075-11089.

Nomura, T., Sakai, N., Sarai, A., Sudo, T., Kanei-Ishii, C., Ramsey, R.G., Favier, D., Gonda, T.J. and Ishii, S. (1993). Negative auto-regulation of c-Myb activity by homodimer formation through the leucine zipper. J. Biol. Chem. 268, 21914-21923.

Nunoue, I., Ohashi, K., Okano, I. and Mizuno, K. (1995). LIMK-1 and LIMK-2, two members of a LIM motif-containing protein kinase family. Oncogene 11, 701-710.

Odell, J.T., Nagy, F. and Chua, N.-H. (1985). Identification of DNA sequences required for activation of the cauliflower mosaic virus 35S promoter. Nature 313, 810-812.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994). Solution sructure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell 79, 639-648.

Ohl, S., Hedrick, S.A., Chory, J. and Lamb, C.J. (1990). Functional properties of a phenylalanine ammonium-lyase gene promoter from *Arabidopsis*. Plant Cell 2, 837-848.

Okamuro, J.K. and Goldberg, R.B. (1985). Tobacco single-copy DNA is highly homologous to sequencespresent in the genomes of its diploid progenitors. Mol. Gen. Genet. 198, 290-298.

Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J. and Marks, M.D. (1991). A *myb* gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell 67, 483-493.

Ow, D.W., Wood, M.D., DeLuca, M., De Wet, J.R., Helsinki, D.R. and Howell, S.H. (1986). Transcient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. Science 234, 856-859.

Pacini, E. (1990). In S. Blackmore and R. B. Knox (eds.), *Tapetum and microspore function*, Academic Press, London, pp. 213-237.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987). The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. 6, 3553-3558.

Pelham, H.R.B. (1982). A regulatory upstream promoter element in the *Drosophila hsp70* heat-shock gene. Cell 30, 517-528.

Perdue, T.D., Mondayan, V. and Parthasarathy, M.V. (1985). *In situ* localisation of F actin in pollen tubes. Eur. J. Cell Biol. 39, 13-20.

Perez-Alvarado, G.C., Miles, C., Michelson, J.W., Louis, H.A., Winge, D.R., Becerle, M.C. and Summers, M.F. (1994). Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. Struct. Biol. 1, 388-398.

Potrykus, I., Paskowski, J., Saul, M.W., Petruska, J. and Shillite, R.D. (1985). Molecular and general genetics of hybrid foreign gene introduced into tobacco by direct gene transfer. Mol. Gen Genet. 199, 169-177.

Preuss, D., Lemieux, B., Yen, G. and Davies, R.W. (1993). A conditional sterile mutation eliminates surface components from Arabidopsis pollen and disrupts cell signalling during fertilisation. Genes Dev. 7, 974-985.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature 335, 683-689.

Rafner, T., Griffith, I.J., Kuo, M.C., Bond, J.F., Rogers, B.L. and Klapper, D.G. (1991). Isolation of cDNA clones coding for the major short ragweed allergen Amb a I (antigen E.). J. Biol. Chem. 266, 1229-1236.

Ramachandran, S., Hiratsuka, K. and Chua, N.-H. (1994). Transcription factors in plant growth and development. Curr. Biol. 4, 642-646.

Rearden, A. (1994). a new LIM protein containing an autoepitope homologous to 'senescent cell antigen". Biochem Biophys. Res. 201, 1124-1131.

Regan, S.M. and Moffatt, B.A. (1990). Cytochemical analysis of pollen development in wild-type Arabidopsis and a male-sterile mutant. Plant Cell 2, 877-889.

Reijen, W.H., Vanherpen, M.M.A., Degroot, P.F.M., Olmedilla, A., Schrawen, J.A.M., Weterings, K.A.P. and Wullems, G.J. (1991). Cellular localization of a pollen-specific mRNA by *in situ* hybridization and confocal laser scanning microscopy. Sexual Plant Reprodroduction 4, 877-889.

Restrepo, M.A., Freed, D.D. and Carrington, J.C. (1990). Nuclear transport of plant potyviral proteins. Plant Cell 2, 987-988.

Roberts, M.R., Foster, G.D., Blundell, R.P., Robinson, S.W., Kumar, A., Draper, J. and Scott, R. (1993a). Gametophytic and sporophytic expression of an anther-specific *Aradidopsis thaliana* gene. Plant J. 3, 111-120.

Rogers, H.J., Allen, R.L., Hamilton, W.D.O. and Lonsdale, D.M. (1991). Pollen-specific cDNA clones from *zea mays*. Biochim. Biophys. Acta 1089, 411-413.

Rogers, H.J., Harvey, A. and Lonsdale, D.M. (1992a). Isolation and characterisation of a tobacco gene with homology to pectate lyase which is specifically expressed during microsporogenesis. Plant Mol. Biol. 20, 493-502.

Rohde, W., Dorr, S., Salamini, F. and Becker, D. (1991). Structure of a chalcone synthase gene from *Hordeum vulgare*. Plant Mol. Biol. 16, 1103-1106.

Ross, J. (1994). Recent advances in the study of gibberellin mutants. Plant growth regul. 15, 193-206.

Roth, B.A., Goff, S.A., Klein, T.M. and Fromm, M.E. (1991). *C1-* and *R*-dependant expression of the maize *Bz1* gene requires sequences with homology to mammalian *myb* and *myc* binding sites. Plant Cell 3, 317-325.

Royer-Pakora, B., Loos, U. and Ludwig, W.D. (1991). TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukemia with the t(11;14)(p13;q11). Oncogene 6, 1887-1893.

Sablowski, R.W.M., Moyano, E., Culianez-Macia, F.A., Schuch, W. and Martin, C. (1994). A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. EMBO J. 13, 128-137.

Sachs, M.M., Dennis, E.S., Gerlach, W.L. and Peacock, W.G. (1986). Two alleles of maize alcohol dehydrogenase 1 have 3' structural and additional polymorphisms. Genetics 113, 449-467.

Sadler, I., Crawford, A.W., Michelsen, J.W. and Beckerle, M.C. (1992). Zyxin and cCRP: Two interactive LIM domain proteins associated with the cytoskeleton. J. Cell Biol. 119, 1573-1587.

Saikumar, P., Murali, R. and Reddy, E.P. (1990). Role of tryptophan repeats and flanking amino-acids in Myb-DNA interactions. Proc. Natl. Acad. Sci. USA 87, 8452-8456.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T.J. and Ishii, S. (1989). Delination of three functional domains of the transcriptional activator encoded by the *c-myb* protooncogene. Proc. Natl. Acad. Sci. USA 86, 5758-5752.

Sanchez-Garcia, I. and Rabbitts, T.H. (1994). The LIM domain: a new structural motif found in zinc-finger like proteins. TIG 10, 315-320.

234

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 54-63.

Schmeichel, K.L. and Beckerle, M.C. (1994). The LIM domain is a modular protein-binding interface. Cell 79, 211-219.

Schrauwen, J.A.M., de Groot, P.F.M., van Herpen, M.M.A., van der Lee, T., Reynen, W.H., Weterings, K.A.P. and Wullems, G.J. (1990). Stage-related expression of mRNAs during pollen development in lily and tobacco. Planta 182, 298-304.

Schwartz, D. (1971). Genetic control of alcohol dehydrogenase-a competition model for regulation of gene action. Genetics 67, 411-425.

Sheehy, R.E., Kramer, M. and Hiatt, W.R. (1988). Reduction of polygalacturonase activity in tomato fruit by antisense RNA. Proc. Natl. Acad. Sci. USA 85, 8805-8809.

Shen, J.B. and Hsu, F.C. (1992). Brassica anther-specific genes: Characterisation and *in situ* localisation of expression. Mol. Gen. Genet. 234, 379-389.

Shen-Ong, G.L.C., Morse, H.C., Potter, M. and Mushinski, J.F. (1986). Two modes of c-myb activation in virus induced mouse myeloid tumours. Mol. Cell Biol. 6, 380-392.

Shinozaki, K., Yamaguchi-Shinozaki, K., Urao, T. and Koizumi, M. (1992). Nucleotide sequence of a gene from *Arabidopsis thaliana* encoding a myb homologue. Plant Mol. Biol. 19, 493-499.

Short, J.M., Fernandez, J.A., Sorge, J.A. and Huse, W.D. (1988). λZAP: A bacteriophage λ expression vector with in vitro excision properties. Nuc. Acids Res. 16, 7583-7600.

Singh, A. and Kao, T.-H. (1992). In S. D. Russell and C. Dumas (eds.), *Gametophytic self-incompatability: biochemical, molecular genetic and evolutionary aspects*, Academic Press, San Diego, Vol. 140, pp. 449-483.

Singh, H., Clerc, R.G. and LeBowitz, J.H. (1989). Molecular cloning of sequence-specific DNA binding proteins using recogonition site probes. BioTechniques 7, 252-261.

Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988). Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. Cell 52, 415-423.

Singh, M.B., Hough, T., Theerakulpisut, P., Avjioglu, A., Davies, S., Smith, P.M., Taira, M., Evrard, J.-L., Steinmetz, A. and Dawid, I.B. (1995). Classification of LIM proteins. TIG 11, 431-432.

235

Siu, G., Wurster, A.L., Lipsick, J.S. and Hedrick, S.M. (1992). Expression of the CD4 gene requires a myb transcription factor. Mol. Cell Biol. 12, .

Snow, A.A. and Spira, T.P. (1991). Pollen vigour and the potential for sexual selection in plants. Nature 352, 796-797.

Solano, R., Nieto, C., Avila, J., Canas, L., Diaz, I. and Paz-Ares, J. (1995a.). Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB-Ph3) from Petunia hybrida. EMBO J. 14, 1773-1784.

Solano, R., Nieto, C. and Pax-Ares, J. (1995b.). MYB.Ph3 transcription factor from Petunia hybrida induces similar DNA-bending/distortions on its two types of binding site. Plant J. 8, 673-682.

Sommer, H. and Saedler, H. (1986). Structure of the chalcone synthase gene of Antrrinhum majus. Mol. Gen. Genet. 202, 429-434.

Stauldt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A. and Baltimore, D. (1988). Cloning of a lymphoid specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science 241, 577-580.

Steiglitz, H. (1974) Somatic regulation of microspore release: An analysis of meiocyte wall breakdown, Ph.D. thesis, University of California.

Steiglitz, H. (1977). Role of beta-1,3-glucanase in postmeiotic microspore release. Dev. Biol. 57, 87-97.

Stinson, J.R., Eisenberg, A.J., Willing, R.P., Pe, P.E., Hanson, D.D. and Mascarenhas, J.P. (1987). Genes expressed in the male gametophyte of flowering plants and their isolation. Plant Physiol. 83, 442-447.

Stockhaus, J., Hofer, M., Renger, G., Westhof, P., Wydrzyski, T. and Willmitzer, L. (1990). Anti-sense RNA efficiently inhibits formation of the 10 kDa polypeptide of photosystem II in transgenic tomato plants: analysis of the role of the 10 kDa protein. EMBO J. 9, 3013-3021.

Sturm, L.M., Das, G. and Herr, W. (1988). The ubiquitous octamer binding protein Oct-1 contains a POU domain with an homeodomain subunit. Genes Dev. 2, 1582-1599.

Sudo, T., Miyazawa, H., Hanaoka, F. and Ishii, S. (1992). The c-myb proto-oncogene product binds to but does not activate the promoter of the DNA polymerase polymerase alpha expression. Oncogene 7, 1999-2006.

Taira, M., Evrard, J.-L., Steinmetz, A. and Dawid, I.B. (1995). Classification of LIM proteins. TIG 11, 431-432.

Taira, M., Jamrich, M., Good, P.J. and Dawid, I.B. (1992). The LIM-cotaining homeobox gene *Xlim-1* is expressed specifically in the organizer region of *Xenopus* gastrula embryos. Genes Dev. **6**, 356-366.

Taira, M., Otani, H., Saint-Jeannet, J.-P. and Dawid, I.B. (1994). Role of the LIM class homeodomain protein Xlim-1 in neural and muscle induction by the Spemann organizer in *Xenopus*. Nature **372**, 677-679.

Tang, X., Helper, P.K. and Scordilis, S.P. (1989). Immunochemical and immunocytochemical identification of a myosin heavy chain polypeptide in *Nicotiana* pollen tubes. J. Cell Sci. **92**, 569-574.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S. and Sarai, A. (1993). Recognition of specific DNA sequences by the c-MYB protooncogene product: role of three repeat units in the DNA-binding domain. Proc. Natl. Acad. Sci. USA **90**, 9320-9324.

Tebbutt, S.J., Rogers, H.J. and Lonsdale, D.M. (1994). Characterisation of a tobacco gene encoding a pollen-specific polygalacturonase. Plant Mol. Biol. **25**, 283-297.

Terasaka, O. and Niitsu, T. (1987). Unequal cell division and chromatin differentiation in pollen grain cells. I. centrifugal, cold and caffeine treatments. Bot. Mag. Tok. **100**, 205-216.

Thangavelu, M., Belostotsky, D., Bevan, M.W., Flavell, R.B., Rogers, H.J. and Lonsdale, D.M. (1993). Partial characterisation of the *Nicotiana tabacum* actin gene family: Evidence for pollen-specific expression of one of the gene family members. Mol. Gen. Genet. **240**, 290-295.

Theerakulpisut, P., Xu, H.L., Singh, M.B., Pettit, J.M. and Knox, R.B. (1991). Isolation and developmental expression of *Bcp1*, an anther-specific cDNA clone in *Brassica campestris*. Plant Cell **3**, 1073-1084.

Thompson, M.A. and Ramsey, R.G. (1995). Myb: an old oncoprotein with new roles. Bioessays **17**, 341-350.

Thor, S., Ericson, J., Brannstrom, T. and Edlund, T. (1991). The homeodomain LIM protein Isl-1 is expressed in subsets of neurons and endocrine cells in the adult rat. Neuron **7**, 881-889.

Tice-Baldwin, K., Fink, G.R. and Arndt, K.T. (1989). BAS1 has a Myb motif and activates *HIS4* transcription only in combination with BAS2. Science **246**, 931-935.

Treisman, J., Gonczy, P., Vashishtha, M., Esther, H. and Desplan, M. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell **59**, 553-562.

Tupy, J., Suss, J., Hrabetova, E. and Rihova, L. (1983). Developmental changes in gene expression during pollen differentiation and maturation in *Nicotiana tabacum* L. Biologia Plant. **25**, 231-237.

Turner, C.E. (1994). Paxillin: a cytoskeletal target for tyrosine kinases. BioEssays **16**, 47-52.

Turner, C.E. and Millar, J.T. (1994). Primary sequence of paxillin contains putative SH2 and SH3 domain binding motifs and multiple LIM domains: identification of a vincullin and pp125^{Fak}-binding protein. J. Cell Sci. **107**, 1583-1591.

Twell, D. (1992). Use of a nuclear-targeted β-glucuronidase fusion protein to demonstrate vegetative cell-specific gene expression in developing pollen. Plant J. **2**, 887-892.

Twell, D. (1994). In R. J. Scott and A. D. Stead (eds.), *The diversity and regulation of gene expression in the pathway of male gametophyte development*, Cambridge University Press, Vol. **55**, pp. 83-115.

Twell, D., Klein, T.M., Fromm, M.E. and McCormick, S. (1989b). Transcient expression of chimeric genes delivered into pollen by micro projectile bombardment. Plant Physiol. **91**, 1270-1274.

Twell, D., Wing, R., Yamaguchi, J. and McCormick, S. (1989a). Isolation and expression of an anther-specific gene from tomato. Mol. Gen. Genet. **217**, 240-245.

Twell, D., Yamaguchi, J. and McCormick, S. (1990). Pollen-specific gene expression in transgenic plants-coordinate regulation of 2-different tomato gene promoters during microsporogenesis. Development **109**, 705-712.

Twell, D., Yamaguichi, J., Wing, R.A., Ushiba, J. and McCormick, S. (1991). Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. Genes Dev. **5**, 496-507.

Urao, T., Yamaguchi-Shinozaki, K., Urao, S. and Shinozaki, K. (1993a). An Arabidopsis *myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5**, 1529-1539.

Urao, T., Yamaguchi-Shinozki, K., Urao, S. and Shinozaki, K. (1993b). An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5**, 1529-1539.

Ursin, V.M., Yamaguchi, J. and McCormick, S. (1989). Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. *Plant Cell* **1**, 727-736.

238

Valenta, R., Duchene, M., Pettenberger, K., Sillaber, C., Valent, P., Bettelheim, P., Breitenbach, M., m, Rumpold, H., Kraft, D. and Scheiner; O. (1991). Identification of profilin as a novel pollen antigen; IgE autoreactivitity in sensitized individuals. Science 253, 557-560.

van Buuren, M., Neuhas, J.-M., Shinshi, H., Ryals, J. and Meins, J.F. (1992). The structure and regulation of homeologous tobacco endochitinase genes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis* origin. Mol. Gen. Genet. 232, 460-469.

van der Krol, A., Mur, L.A., Beld, M., Mol, J.N.M. and Stuitje, A.S. (1990b). Flavonoid genes in Petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2, 291-299.

van der Krol, A.R., Lenting, P.E., Veenstra, J., van der Meer, I.M., Koes, R.E., Gerats, A.G.M., Mol, J.N.M. and Stuitje, A.R. (1990a). An antisense chalcone synthase gene in transgenic plants inhibits flower pigmentation. Nature 333, 866-869.

van der Krol, A.R., Mur, L.A., de Lange, P., Mol, J.N.M. and Stuitje, A.R. (1990c). Inhibition of flower pigmentation by antisense CHS genes: promoter and minimal sequence requirements for the antisense effect. Plant Mol. Biol. 14, 457-466.

van Tunen, A.J., Hartman, S.A., Mur, L.A. and Mol, J.N.M. (1989). Regulation of the chalcone flavanone isomerase (CHI) gene expression in *Petunia hybrida*: The use of alternative promoters in corolla, anthers and pollen. Plant Mol. Biol. 12, 539-551.

van Tunen, A.J., Mur, L.A., Brouns, G.S., Rienstra, J.-D., Koes, R.E. and Mol, J.N.M. (1990). Pollen- and anther-specific promoters from *Petunia*: tandem promoter regulation of the *chiA* gene. Plant Cell 2, 393-401.

Verwoerd, T.C., Dekker, B.M.M. and Hoekema, A. (1989). A small scale procedure for rapid isolation of plant RNAs. Nuc. Acids. Res. 17, 2362.

Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988). *In situ* detection of sequence-specific DNA binding activity specified by a recombinant phage. Genes Dev. 2, 801-806.

Wagner, V.T., Cresti, M., Salvatici, P. and Tiezzi, A. (1990). Changes in volume, surface area and frequency of nuclear pores on the vegetative nucleus of tobacco pollen in fresh, hydrated and activated conditions. Planta 181, 304-309.

Warren, A.J., Colledge, W.H., Carlton, M.B.L., Evans, M.J., Smith, A.J.H. and Rabbitts, T.H. (1994). The oncogenic cysteine-rich LIM domain protein Rbtn2 is essential for erythroid development. Cell 78, 45-57.

Way, J.C. and Chalfie, M. (1988). *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. Cell 54, 5-16.

Weber, B.L., Westin, E.H. and Clarke, M.F. (1990). Differentiation of mouse erythroleukemia cells enhanced by alternatively spliced c-myb mRNA. Science 249, 1291-1293.

Weiss, D., van Blokland, R., Kooter, J.M., Mol, J.N.M. and van Tunen, A.J. (1992). Gibberellic acid regulates chalcone synthase gene transcription in the corolla of Petunia hybrida. Plant Physiol. 98, 191-197.

Weiss, d., van Tunen, A.J., Halevy, A.H., Mol, J.N.M. and Gerats, A.G.M. (1990). Stamens and gibberellic acid in the regulation of flavanoid gene expression in the corolla of Petunia hybrida. Plant Physiol. 98, 191-197.

Weisshaar, B. (1991). Light inducible and constituitively expressed DNA-binding proteins recognise a plant promoter with functional relevance in light responsiveness. EMBO J. 10, 1777-1786.

Weston, K. and Bishop, J.M. (1989). Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb. Cell 58, 85-93.

Weterings, K. (1994). Pollen gene regulation. Ph.D. Thesis.

Weterings, K., Reijen, W., van Aarssen, R., Kortstee, A., Spijkers, J., Herpen, M., Schrauwen, J.A.M. and Wullems, G.J. (1992). Characterisation of a pollen-specific cDNA clone fron Nicotiana tabacum expressed during microgametogenesis and germination. Plant Mol. Biol. 18, 1101-1111.

Williams, J.A., Paddock, S.W., Vorwerk, K. and Carroll, S.B. (1994). Organisation of wing formation and induction of wing-patterning gene at the dorsal/ventral compartment boundry. Nature 368, 299-305.

Willing, R.P., Bashe, D. and Mascarenhas, J.P. (1988). An analysis of the quantity and diversity of mRNAs from pollen and shoots of Zea mays. Theor. Appl. Genet. 75, 751-753.

Willing, R.P. and Mascarenhas, J.P. (1984). Analysis of the complexity and diversity of mRNAs from pollen and shoots of Tradescantia. Plant Physiol. 75, 865-868.

Wing, R.A., Yamaguchi, J., Larabell, S.K., Ursin, V.M. and McCormick, S. (1989). Molecular and genetic characterisation of two pollen-expressed genes that have sequence similarities to pectate lyases of the plant pathogen Erwinia. Plant Mol. Biol. 14, 17-28.

Wolf, N. (1992). Structure of the genes encoding Hordeum vulgare (1-3, 1-4)-β-glucanase isoenzymes I and II and functional analysis of their promoters in barley aleurone protoplasts. Mol. Gen. Genet. 234, 33-42.

Wu, R.-Y. and Gill, G.N. (1994). LIM domain recognition of a tyrosine-containing tight turn. J. Biol. Chem. 269, 25085-25090.

Xue, D., Finney, M., Ruvkun, G. and Chalfie, M. (1992). Regulation of the mec-3 gene by the C. elegans homeoproteins UNC-86 and MEC-3. EMBO J. 11, 4969-4979.

Ylstra, B., Busscher, J., Franken, J., Hollman, P.C.H., Mol, J.N.M. and van Tunen, A.J. (1994). Flavonols and fertilisation in *Petunia hybrida:* localisation and mode of action during pollen tibe growth. Plant J. 6, 201-212.

Zaki, M.A.M. and Dickinson, H.G. (1991). Microspore derived embryos in *Brassica:* the significance of division asymmetry in pollen mitosis I to embryogenic development. Sex. Plant Reprod. 4, 48-55.

Zobel, A., Kalkbrenner, F., Guehmann, S., Nawrath, M., Vorbrueggen, G. and Moelling, K. (1991). Interaction of the v- and c-myb proteins with regulatory sequences of the human *c-myc* gene. Oncogene 6, 1397-1407.