1

PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

50-108-1-01/0000

0000-108-50-0000

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Isolation and sequence analysis of a *Nicotiana
tabacum* seed coat-specific genomic region
identified by T-DNA tagging.

by
John Cosmopoulos, B.Sc.

thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment
of the requirements for the degree of

Master of Science
Department of Biology
Ottawa-Carleton Institute of Biology

Carleton University
Ottawa, Ontario
January 15, 1993
© John Cosmopoulos

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Author's Note

Author's Note

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-84175-3

Canada

Carleton University
Ottawa, Canada K1S 5B6

Thesis contains black & white
photographs &/or explanatory tables
which when microfilmed may lose their
significance. The hardcopy of the
thesis is available upon request from
Carleton University Library.

The undersigned recommend to the Faculty of Graduate
Studies and Research acceptance of the thesis

"Isolation and sequence analysis of a *Nicotiana
tabacum* seed coat-specific genomic region
identified by T-DNA tagging."

submitted by John Cosmopoulos, B.Sc.
in partial fulfilment of the requirements for
the degree of Master of Science

Thesis Co-Supervisor

Thesis Co-Supervisor

Chair, Department of Biology

Ottawa-Carleton Institute of Biology
Carleton University

ABSTRACT

A *Nicotiana tabacum* SR1 genomic library was constructed in a Lambda-Dash II vector by Strategene and screened to isolate the wild-type tobacco sequences corresponding to a seed coat-specific genomic region previously identified through T-DNA tagging (Fobert, 1992). A Lambda clone (L3) was isolated and characterized by restriction and Southern blot analyses. This established that the L3 contained the wild-type *N. tabacum* genomic region that corresponded to the seed coat-specific genomic region. A 3.0 kb *Eco* RI L3 fragment containing the region of interest was subcloned in pGEM-4Z (pJC-1) and characterized by sequence analysis.

The primary objective of the sequence analysis was to determine whether the integration of T-DNA in the plant genome and caused rearrangents and whether the seed coat-specific pattern of expression was a result of rearrangements in DNA at the right border junction. The sequencing data indicated that only minor rearrangements had resulted from the T-DNA integration. Only small deletions, base substitutions and insertions, as well as a 12 bp insert at the site of integration were detected. This indicated that the seed coat-specific pattern of expression was not the result of gross rearrangements caused by T-DNA integration.

A secondary objective of the sequence analysis was to use the pJC-1 sequencing data to determine whether the T-DNA tag had inserted near or into an ORF in the transgenic plant line T218. The sequencing data was insufficient to identify an ORF.

ACKNOWLEDGEMENTS

I would first like to thank my supervisors, Dr. V.N. Iyer and Dr. Brian Miki for their help, advice and support during all stages of this project. I am also grateful to Dr. Brian Miki and the Plant Research Centre, Agriculture Canada for allowing me to use their research facilities and Dr. Bob Iyer for financial assistance from his NSERC grants. The advice of Dr. V. Seligy is also gratefully acknowledged.

The help, advice, encouragement and friendship of people from Dr. Brian Miki's lab, past and present is well appreciated. Special thanks to Dr. Pierre Fobert, Kim Boutilier, Hélène Labbé, Dr. Sylvia McHugh, Dr. Thérèse Ouellet, Dr. Janice DeMoor and Dr. Jiro Hattori.

The financial support of Carleton University is gratefully acknowledged.

TABLE OF CONTENTS

Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Abbreviations	ix
 CHAPTER 1	
INTRODUCTION	1
 CHAPTER 2	
LITERATURE REVIEW	5
2.A. Organization of the Plant Genome	5
2.B. Regulation of Plant Gene Expression	8
2.B.1. Overview	8
2.B.2. <i>cis</i> -acting factors	9
2.B.3. <i>trans</i> -acting factors	11
2.B.4. <i>cis</i> -elements in post-transcriptional regulation	12
2.C. The Biology of <i>Agrobacterium</i> T-DNA	13
2.D. Gene-Tagging in Plants	21
2.D.1. Transposon tagging	23
2.D.2. T-DNA tagging	24
 CHAPTER 3	
METHODS AND MATERIALS	28
3.A. Bacterial Strains	28
3.A.1. Growth and storage conditions	28
3.B. Preparation of Competent Cells and Bacterial Transformation	28
3.B.1. Preparation of competent cells	30
3.B.2. Transformation of competent cells	31
3.C. DNA Extraction Protocols	32
3.C.1. Rapid extraction of plasmid DNA with Holmes and Quigley's boiling method	32
3.C.2. Large scale plasmid DNA extraction	32
3.C.3. Isolation of lambda phage DNA	34
3.C.3a. Elution of lambda particles from solid media	34
3.C.3b. Growth of lambda particles	35
3.C.3c. Isolation of lambda DNA	35

3.C.4. Isolation of total plant genomic DNA	36
3.C.5. Isolation of DNA fragments	37
3.C.5a. Isolation of DNA with GENECLEAN	37
3.C.5b. Isolation of DNA with low-melting agarose	38
3.D. Ethanol Precipitation of DNA	38
3.E. DNA Restriction Digestion	39
3.E.1. Restriction digestion of plasmid DNA	40
3.E.2. Restriction digestion of plant genomic DNA	40
3.E.3. Restriction digestion of lambda phage DNA	40
3.F. DNA Hybridization Techniques	41
3.F.1. Southern Hybridizations	41
3.F.1a. Agarose gel electrophoresis and Southern blotting	41
3.F.1b. Preparation of radioactive probes	43
3.F.1c. Hybridization and autoradiography	44
3.F.2. Plaque Hybridization	45
3.F.2.1. Preparation of lysed bacterial plates	45
3.F.2.2. Screening procedure	45
3.G. DNA Sequencing Techniques	46
 CHAPTER 4	
RESULTS	48
4.1. Molecular Analysis of the seed coat-specific genomic region in untransformed <i>Nicotiana tabacum</i> SR1.	48
4.2. Isolation of the Lambda Clones.	53
4.2.1. Cloning of the genomic region corresponding to pT218-1.	53
4.2.2. Characterization of Lambda clones.	55
4.2.3. Mapping and cloning of L3.	63
4.2.4. Southern analysis of the 3.0 kb L3 Eco RI fragment.	64
4.3. Restriction pattern analysis of the 3.0 kb pJC-1 clone.	65
4.4. Sequence analysis of clone pJC-1.	70
 CHAPTER 5	
DISCUSSION AND CONCLUSIONS	82
5.1. Background and objectives of Present Study.	82
5.2. Rearrangements in the target sequence.	83
5.3. T-DNA Recombination Mechanism.	86
5.4. Is there an ORF associated with this genomic region?	86
5.5. Concluding remarks	87
 REFERENCES	90

LIST OF TABLES

Table 3.1.	Bacterial strains used.	29
Table 4.1.	Screening the genomic libraries L1, L2, L3	54

LIST OF FIGURES

FIGURE 4.1.	Southern blot of <i>Nicotiana tabacum</i> SR1 genomic DNA with probe #2.	49
FIGURE 4.2.	A partial map of pT218-1 and of the genomic region of interest.	51
FIGURE 4.3.	Ethidium bromide and Southern blot analysis of L3, pT218-1 and L2 digests.	56
FIGURE 4.4.	Southern blot analysis of Lambda L3 against probe #1.	59
FIGURE 4.5.	A partial map of the Lambda L3 clone	61
FIGURE 4.6.	Southern blot analysis of four L3 <i>Eco</i> RI fragments with probe #2.	66
FIGURE 4.7.	Ethidium bromide and Southern blot analysis of seven L3 <i>Eco</i> RI subclones.	68
FIGURE 4.8.	Ethidium bromide and Southern blot analysis of pJC-1 and pT218-1 digests.	71
FIGURE 4.9.	Sequencing strategy followed for clone pJC-1.	74
FIGURE 4.10.	Nucleotide sequence of clone pJC-1.	77
FIGURE 4.11.	A schematic diagram of the sequencing data.	80

LIST OF ABBREVIATIONS

A	adenine
bp	base pair(s)
C	cytosine
dATP	deoxyadenosine triphosphate
DNA	deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetra acetic acid
g	gram or centrifugal force
G	guanine
GUS	gene coding for β-glucuronidase
hr	hour(s)
kb	kilo base pair(s)
L	Litre
LB	Luria-Bertani broth
<i>luc</i>	gene encoding firefly luciferase
<i>lux</i>	gene encoding <i>Vibrio harveyi</i> luciferase
mRNA	messenger RNA
mM	millimolar
M	molar
min	minute(s)
ml	millilitre(s)
mg	milligram(s)
MgCl₂	magnesium chloride
NaCl	sodium chloride

nptII	gene coding for neomycin phosphotransferase
pfu	plaque forming units
rpm	revolutions per minute
RNA	ribonucleic acid
rRNA	ribosomal RNA
sec	second(s)
SDS	sodium dodecyl sulfate
T	thymine
TA	Tris acetate restriction buffer
TAE	Tris acetate EDTA electrophoresis buffer
TBE	Tris borate EDTA electrophoresis buffer
T-DNA	transferred DNA
TE	Tris EDTA
Ti	tumour-inducing
tRNA	transfer RNA
vir	<i>Agrobacterium</i> virulence genes
v/v	volume per volume
w/v	weight per volume
μ Ci	microcurrie(s)
μ g	microgram(s)
μ l	microlitre(s)
μ M	micromolar

The natural characteristics for plant transformation exhibited by the bacterial species *Agrobacterium tumefaciens* have been the objects of in depth studies during the past ten to twenty years (Zambryski, 1992). These studies have led to an understanding of the bacterial processes involved (see Section 2.C.). In particular, it has now become clear that the transformation of the plant cell is a result of the transfer and stable incorporation in the plant genome of a particular DNA segment (referred to as the Transferred or T-DNA) of a large *Agrobacterium* plasmid. This plasmid has been named the Tumour-inducing (Ti) plasmid, since it induces the neoplastic transformation of plant cells. The T-DNA does not itself code the functions that mediate its transfer from the bacterial to the plant cell. All of the wild-type (tumour-inducing) T-DNA genes may be deleted and replaced with any other gene constructs, without affecting the efficiency of the T-DNA transfer. The only wild-type T-DNA sequences required for efficient transfer are the regions bordering the T-DNA. These are referred to as the left and right border regions and are approximately 25 bp long (see Section 2.C.). This has permitted the utilization of the transfer mechanism for the development of efficient genetic transformation systems for many dicotyledonous plant species (Klee et al., 1987; Zambryski, 1992).

These same bacterial processes are now being employed for purposes that are beyond those of gene transfer (Section 2.D.2.).

The ability of the T-DNA to insert at random in the plant genome is being employed to detect novel plant developmental genes by insertion mutagenesis (Feldmann, 1991). With the development of appropriate reporter genes for plant systems, the T-DNA system is also being used to characterize plant cis-acting regulatory elements and their neighbouring Open Reading Frames (ORFs). Widely used reporter genes are GUS (β -glucuronidase), *nptII* (neomycin phosphotransferase), *luc* (firefly luciferase) and *lux* (bacterial *-Vibrio harveyi-* luciferase). Reporter genes are used in conjunction with a selectable antibiotic marker gene to screen for transformants. This approach screens for random T-DNA insertional events that place the reporter gene under the control of neighbouring plant regulatory elements in the chromosomal region into which the T-DNA has inserted, thus initiating or enhancing their expression. These events are referred to as gene fusions and reporter gene constructs have the ability to form transcriptional or translational gene fusions with the neighbouring chromosomal region (Walden et al., 1991). Translational gene fusions differ from transcriptional gene fusions because the reporter gene construct has inserted in frame with a plant ORF and, as a result, additional amino acids are to be found at the N' end of the protein, perhaps altering its expression patterns.

Previous studies (Fobert, 1992) were undertaken to characterize the chromosomal regions into which the T-DNA integrates based on their expression properties in *Nicotiana tabacum*. This was accomplished by generating, upon insertion of

the T-DNA into the plant chromosome, *in vivo* gene fusions between a promoterless reporter gene (GUS), and endogenous promoters residing in the plant chromosome.

One result of this study (Fobert, 1992) was the identification and isolation, by T-DNA tagging (Section 2.D.), of a 2.6 kb chromosomal region responsible for seed coat-specific pattern of expression in transgenic plant line T218. This *cis*-regulatory chromosomal region was partially sequenced (ca. 1.0 kb) and reintroduced into non-transformed tobacco plants for functional testing (Fobert, 1992). The results obtained indicated that the sequences controlling the seed-coat pattern of expression were confined to a 200 bp *Xba* I/*Sna* BI region situated 700 bp upstream of the insertion site.

The objective of this thesis was to study the wild-type *N. tabacum* genomic region that corresponds to the region into which the T-DNA inserted in plant line T218. A chromosomal region that spans the 2.6 kb seed coat-specific region isolated previously (Fobert, 1992) was identified, subcloned and analyzed by sequence analysis. The unsequenced portion of the T218 clone was also sequenced. Sequence alignments were made between the wild-type sequence and the one originating from transgenic plant line T218.

The sequencing data was used to address the two objectives of this study. The first objective was to determine whether any rearrangements (deletions, duplications or base substitutions) were caused by the T-DNA integration in plant T218 and whether the seed coat-specific pattern of expression of this genomic region may be

the result of gross rearrangements.

Sequence alignments were also performed between this region and the T-DNA vector used to transform plant line T218, to determine which recombination mechanism may have played a role in the T-DNA integration (section 2.C.2.). Homologous recombination is a rare event, but has been proposed for a few T-DNA insertional events studied (Matsumoto et al., 1990) and in the integration mechanisms of foreign DNA in the plant genome by other transformation procedures, such as direct gene transfer (Paszkowski, 1989). Illegitimate or non-homologous recombination is the insertional mechanism that best explains most of the T-DNA integrations studied to date (Gheysen et al., 1987; Gheysen et al., 1991; Mayerhofer et al., 1991).

The second objective of this thesis was to analyze the sequencing data, especially the sequences situated downstream of the site of T-DNA insertion for the presence of an ORF. Any detected ORF is very likely, by virtue of its close proximity, to be under the control of the *cis*-acting regulatory elements tagged in plant line T218.

2. LITERATURE REVIEW

2.A. Organization of the Plant Genome: An Overview

Biochemical and molecular studies have shown that there is no such thing as a typical plant genome. Plant genomes from different species are usually heterologous in size and evolutionarily dynamic (Okamuro and Goldberg, 1989). Analyses performed by Bernatzky and Tanksley (1986) indicate that the majority of plant structural genes are present in 1-2 copies per haploid genome. However, certain genes that code for highly abundant mRNA species may be represented many times. For example, Histone 3 (H3) and Histone 4 (H4) genes may be present in 60-80 and 100-120 copies, respectively, in maize (Chaubert, 1986). The genes that code for the light harvesting chlorophyll a/b binding protein may range between 3-50 copies per haploid genome. This occurrence of gene redundancy may provide the organism with the opportunity to evolve new modified genetic patterns and ultimately new physiological functions, without risking the disruption of existing and essential functions. An indication of this ongoing event may be the qualitative, as well as quantitative differences in gene expression that have been described for several plant gene families. Some recent examples include glutamine synthase (Gebhardt et al., 1986; Tingey et al., 1987), the small subunit of ribulose-biphosphate carboxylase (Coruzzi et al., 1984; Dean et al., 1985; Fluhr et al., 1986) or acetohydroxyacid synthase (Rutledge et al., 1991). Among

higher plants there is one important exception to this redundancy: *Arabidopsis thaliana*. In *A. thaliana* the size of any given multigene family, the number of introns in a gene, as well as the size of these introns, are generally smaller than most plant species. This is a reflection of its much smaller genome size which is itself very likely the result of evolutionary selective pressure (Chang and Meyerowitz, 1986; Meyerowitz, 1987).

Irrespective of the differentiated states of a plant tissue or organ, the component cells have genomes that remain constant in size and structure. Evidence that cells in distinct differentiated state have identical DNA sequence content has been provided by the ability of many plant cells or tissues to be totipotent under the right culturing conditions.

Although plant species have fewer organs and cell types than most animals, they possess genomes that are equal to and sometimes even greater than those of many higher animals, including humans, in both size and complexity (Okamuro and Goldberg, 1989). Most plant genomes are between 10^5 and 10^7 kb in size (Okamuro and Goldberg, 1989). *Nicotiana tabacum*, the species used in this study, has a genome size of 2.4×10^6 kb (Okamuro and Goldberg, 1989), whereas the size of the haploid human genome is 3×10^6 . Conversely, *A. thaliana* has a genome size of only 7.0×10^4 , comparable to the genome size of most species of algae and molds (Okamuro and Goldberg, 1989).

This variation in genome size between animals and plants is largely due to dramatic differences in repetitive DNA sequence

content. However, this explanation by itself is insufficient, since even when repeated DNA sequences are subtracted, plants still have a genome complexity that is comparable to the complexity of most higher animal genomes (Okamuro and Goldberg, 1989).

Gene activity during plant development is complex and highly regulated. Each stage of plant development is associated with the expression of unique sets of developmentally regulated genes. Over the course of sporophytic and gametophytic development a plant expresses a minimum of 60,000 structural genes or 7×10^4 kb of DNA. Each organ of the sporophyte is programmed by a unique set of structural genes that encode primarily rare class mRNAs. With the exception of leaf and petal whose mRNA sequence sets overlap almost completely, 25-50% of the structural genes expressed in each organ are organ-specific. Approximately 30% of the genes expressed in each organ encode a set of ubiquitous or "housekeeping" mRNAs. Furthermore, quantitative, as well as qualitative, changes in gene expression play an important role in the establishment and maintenance of the unique structure and function of each plant organ type (Goldberg et al., 1989; Okamuro and Goldberg, 1989).

2.B. Regulation of Plant Gene Expression

2.B.1. Overview

In general, the processes that govern plant gene regulation and the structure and organization of plant genes are similar to animals (Kuhlmeier, 1992). Plant genomes use the same genetic code, are split by introns, and use regulatory mechanisms that are very similar in principle. However, this does not imply that genes or promoters could be shuffled at will between animals and plants. Although plant and animal promoters are functionally very similar, the subtle differences that do exist are sufficient to minimize their expression capability if transferred to the other system. Similarly, although introns are removed in plants by a similar mechanism as in animals, the sequence requirements for splice sites may be very different (Kuhlmeier, 1992).

Plant nuclear genes are regulated primarily at the transcriptional level (Kuhlmeier, 1992). A good example is the genes coding for the small subunit of Rubisco. In this system, it has been demonstrated that modulations in the level of mRNA, after exposure to specific light conditions, are a result of an increased rate of transcription, rather than post-transcriptional processes, such as differential mRNA stability (Thompson and White, 1991).

2.B.2. *cis*-acting sequences

The coding region is preceded, in the 5' upstream region, by *cis*-regulatory sequences that are important in initiating and modulating plant gene expression (Kuhlmeier, 1992). The *cis*-regulatory elements closest to the transcription start site are referred to as the proximal promoter. Although not absolute requirements, common features of many proximal promoter regions is the TATA box at around -30 bp (where 0 bp is the adenine of the ATG start codon) and an upstream element at around -60 bp (White and Jackson, 1992). The upstream element binds to gene-specific factors, while the TATA box is recognized by the general transcription factor TFIID. Two well known upstream elements are the CAAT or AGGA boxes (Elliston and Messing, 1988). The TATA box determines at what site transcription will start, usually around 30 bp downstream. Mutations in the TATA box have been demonstrated to interfere with proper transcription initiation. However, reporter gene constructs that contain only the TATA box elements of a promoter usually give extremely low levels of expression. The proximal promoter may contain elements that confer organ-specific (Gilmartin et al., 1990) or light-inducible (Sheen, 1991) regulation. So, different proximal promoter elements are not functionally equivalent. Recent studies have demonstrated that the proximal promoter regions of some eukaryotes are not A/T rich and that they do not possess TATA boxes. In these systems, a GC-rich motif seems to serve as the binding site for the RNA transcription

complex (Sehgal, 1991).

The best known plant promoter is the cauliflower mosaic virus (CaMV) 35S genomic RNA promoter. This is a very strong constitutive promoter. Deletion analysis studies of this promoter (Benfey et al., 1990) have identified the promoter regions sufficient for high expression. However, these studies also determined that certain portions of this promoter, when separated from the whole, may confer organ-specific expression. This is a good example of the complex interactions that take place between the various domains of a *cis*-regulatory region and DNA-binding protein factors. These observations have raised the question of whether a constitutive promoter is only a complex array of *cis*-regulatory elements that as a whole confer constitutive expression, but separately confer a more specific pattern of gene expression (Kuhlmeier, 1992).

The description given above for the proximal promoter region corresponds to most class II genes (i.e. the ones that are transcribed by RNA polymerase II). However, most class III genes (i.e. transcribed by RNA polymerase III), have intragenic promoters in which the essential *cis*-acting elements comprise highly conserved sequence blocks downstream of the transcription start site, both of which are recognized by the general class III factor TFIIC. Examples of class III genes are the tRNA genes (White and Jackson, 1992).

The other *cis*-acting elements that modulate (enhance) transcription are referred to as enhancers. Enhancers may be

located at a significant distance from the proximal promoter region (1-2 kb), may function independently of their position and orientation (Dyner, 1989) and by themselves are unable to initiate transcription. It has also been demonstrated that animal enhancers have the ability to regulate transcription even when situated at the 3' end of the coding region (Dyner, 1989). This class of regulatory DNA elements may constitute binding sites for proteins that interact with the RNA polymerase complex (Kuhlmeier, 1992).

2.B.3. *trans*-acting factors

Although, transcription is usually very specific within a cell type, a purified RNA polymerase transcribes DNA essentially randomly. Transcription will once again become specific when a cell-free protein extract is added because such extracts would contain transcription initiation factors, which serve to recruit RNA polymerases to the appropriate sets of genes. Other *cis*-regulatory sequences bind to a whole array of protein factors that interact with the RNA polymerase complex either separately or through other proteins. Some upstream binding proteins are probably general transcription factors, present and active in a large array of cell types, whereas other factors are more specialized (Kuhlmeier, 1992). *Trans*-acting factors interact with each other and certain DNA motifs to modulate gene regulation. Since there are no *cis*-acting DNA elements or motifs that are responsible for gene activation or for a specific pattern of gene

expression, ultimately it is the availability of trans-acting factors at a specific cellular location at the right moment in development that control gene regulation. A large number of plant transcription (trans-acting) factors have been identified (Katagiri and Chua, 1992) by mutant analysis (Sommer et al., 1990), cDNA expression libraries (Katagiri and Chua, 1992) and homology to known factors (Ma et al., 1991). Most plant transcription factors characterized to date contain DNA-binding domains previously described in animal systems (Johnson and McKnight, 1989).

2.B.4. Post-transcriptional regulation

Post-transcriptional cis-regulatory elements also exist downstream of the ATG start codon that may affect the qualitative expression of a gene. Examples of post-transcriptional regulatory elements are the following: 5'cap site (Gallie et al., 1991), non-translated leader sequences (Marcotte et al., 1989; Gallie et al., 1991), initiation codons (Kuhlmeier, 1992), introns (Callis et al., 1987; Luehrsen and Walbot, 1991), 3' untranslated regions (Gallie et al., 1991; Dietrich et al., 1992) and polyadenylation sites (Kuhlmeier, 1992). The 5' non-translated leader sequence may enhance reporter gene expression ten-fold (Marcotte et al., 1989), probably by influencing stability or translation of the mRNA.

The expression of some genes appears to have a strong and absolute requirement for the presence of an intron. The first intron of the maize ADH gene is important to achieve a high

transcription rates of reporter genes, even those that are of bacterial origin and, as a result, would not normally contain introns (Callis et al., 1987; Luehrsen and Walbot, 1991).

A recent study (Dietrich et al., 1992) examined the developmental regulation of the AX92 gene that is expressed at an early stage of cortex differentiation in the root apex of oilseed rape seedlings. It was observed that sequences that are situated 3' in relation to the protein coding region are necessary for the activation of the AX92 gene in both embryos and seedlings.

2.C. The Biology of *Agrobacterium* T-DNA

Agrobacterium tumefaciens, a phytopathogenic soil bacterium, is the causative agent of the plant neoplastic disease known as crown gall (Kado, 1991; Ream, 1990; Zambryski, 1992). The natural characteristics of *Agrobacterium*, described below, have been adapted to create an efficient DNA transfer system that is widely used for molecular and genetic engineering experiments (Zambryski, 1992).

When *Agrobacterium* infects a plant, a specific segment of its large, 200 kb, tumour-inducing (Ti) plasmid, the transferred or T-DNA, is transferred to the plant cell and is eventually integrated into the plant genome. The integrated T-DNA is stable during the proliferation of the transformed cells. The tumorous

growth of the transformed cells is a direct result of the expression of several T-DNA genes that code for the synthesis of the phytohormones auxin and cytokinin (Stachel et al., 1986). Transfer of the T-DNA requires the functions of the virulence (*vir*) region on the Ti-plasmid. The *vir* region is transcribed following induction of the bacterium by phenolic compounds from wounded plant cells (Stachel et al., 1985; Boulton et al., 1986). The *vir* region consists of seven loci (*vir* A, B, G, C, D, E and H) and extensive biochemical and genetic studies of their functions have been carried out (Stachel et al., 1986; Veluthambi et al., 1988; Winans et al., 1988). The Vir proteins mediate the appearance within the bacterium of a single-stranded DNA, the T-strand, which represents the T-DNA. They also mediate the transfer of the T-strand to the plant cell by the formation of the T-complex, which is comprised of the T-strand and various Vir proteins. The VirE2 protein exhibits a single-stranded DNA-binding activity and probably plays a role in protecting the T-DNA during transfer from the bacterium to the plant genome (Christie et al., 1988; Das, 1988; Dürrenberger et al., 1989). The VirD2 protein is covalently linked to the 5' end of the T-strand and may play a role in its translocation (Young and Nester, 1988; Herrera et al., 1990). There is also speculation that the VirD2 protein may mediate the integration of the right end of the T-DNA strand (Zambryski, 1992). While single copies of the T-DNA are commonly inserted, tandem and multicopy inserts can be obtained (Jorgensen et al., 1987).

The T-DNA region on the Ti plasmid is flanked by imperfect direct repeats of 25 bp (Van Haaren et al., 1988). Genetic analysis has shown the importance of the right terminal sequence; it appears to direct T-DNA synthesis. If the right border is deleted, the transfer frequency is severely reduced (Wang et al., 1984), as much as 500-fold (Herman et al., 1990). Deletion of the left border does not have a severe effect (Shaw et al., 1984; Peralta et al., 1985), probably because it acts as a termination signal beyond which DNA sequences are not normally transferred. The differential functions of the two border sequences is an indication of the polarity of T-DNA transfer (Wang et al., 1984). These border sequences have been shown to serve as sites recognized by the VirD1/D2 endonuclease (Albright et al., 1987; Wang et al., 1987).

Recently, it was reported that there also exists the possibility for a low frequency transfer of variable amounts of DNA from the *Agrobacterium* Ti plasmid to the plant genome in the absence of both the left and right border regions (Herman et al., 1990).

The ends of the integrated T-DNAs are not always at the same position within the border repeats but may lie in the region internal to these repeats as well (Simpson et al., 1982; Yadav et al., 1982; Zambryski et al., 1982; Holsters et al., 1983; Kwok et al., 1985). Although the breakpoints vary, no right breakpoint has been found beyond the third base of the right (25 bp) border and almost all left breakpoints are not beyond the middle of the left

border. In only two cases have they been mapped outside of the third base: one was mapped between the first and second base of the left border (Simpson et al., 1982) and the other outside of the left border entirely (Ooms et al., 1982). Since the VirD1/D2 nuclease cleaves exactly between the third and fourth bases of the 25 bp repeat to generate the processed T-DNA molecules in bacteria (Albright et al., 1987; Wang et al., 1987; Dürrenberger et al., 1989), the variations observed in the T-DNA ends are probably generated in the plant cells after the T-DNA transfer. A protein/T-strand complex is transferred to the plant cell by a process thought to be analogous to plant conjugation and is mediated by further products of the vir genes (Zambryski, 1992).

The T-DNA fragments integrate at random positions in the plant genome and appears to have no requirements for sequence specificity (Wallroth et al., 1986). This was demonstrated by *in situ* hybridization (Ambros et al., 1986) and genetic mapping (Chyi et al., 1986). The integrated DNA is genetically stable and is transmitted in a Mendelian fashion (DeBlock et al., 1984). It is possible for the integrated DNA to become methylated, resulting in a reduction in gene expression and phenotypic instability (Matzke et al., 1989).

Notwithstanding the randomness of T-DNA integration (Ambros et al., 1986), T-DNA appears to integrate preferentially into transcriptionally active regions (Koncz et al., 1989; Herman et al., 1990; Fobert, 1992). The open chromatin structure that characterizes transcriptionally active regions might facilitate

recombination and, as a result, serve as a preferential target for the integration of the T-DNA (Herman et al., 1990). Koncz et al. (1989) determined that gene fusions were obtained at a remarkably high frequency (25%). Furthermore this frequency was the same regardless of whether *Nicotiana tabacum* or *Arabidopsis thaliana* was used. When considering the vast difference in genome size, transcript complexity and relative amount of repetitive DNA between these two species, these results strongly support preferential integration. Herman et al. (1990) observed that as many as 70% to 100% of the transformed calli contained one of their T-DNA insertions in a transcriptionally active region with a transcript running from the left or right. Although they admit that their frequency may be overestimated, since they do not take into account the frequency of multiple insertions in the same callus, their observations give additional support to the idea of preferential T-DNA integration.

Fobert (1992) observed that the frequency of leaf-specific gene fusions was higher than that of any other organ-specific fusion. Given that this study used the leaf transformation method, this seems to support the proposition of preferential integration. However, in most studies where regulatory elements in the plant genome are tagged (Koncz et al., 1989; Topping et al., 1991; Fobert, 1992), gene fusions are also expressed in tissues (e.g. the seed coat in Fobert, 1992 or the floral organs in Topping et al., 1991) that were not transcriptionally active at the time of the T-DNA integration. This would seem to indicate that preferential

integration of the T-DNA does occur, but in genomic regions that are capable of being transcribed in addition to those that are actively being transcribed at the time of transformation. This observation may be correlated with the different structure of euchromatin (van Holde, 1989). At any given time in development, euchromatin may be differentiated between "active", which is actively being transcribed in that plant component, and "competent", which has the capability to be transcribed. On the basis of nuclease sensitivity studies (van Holde, 1989), both "active" and "competent" euchromatin conditions are characterized by a relatively "open" conformation, contrary to non-transcribed (non-competent) heterochromatin.

These observations are similar to others made with retroviral integration sites in transgenic Mov mice. These integration sites have also been shown to map in transcribed DNA regions (Mooslehner et al., 1990). It has also been suggested that retroviral insertion in animals (Jaenisch, 1988) and P-element insertion in *Drosophila* (Brownes, 1990) occurs in regions of the genome that are potentially or actually transcriptionally active.

To understand the mechanisms behind T-DNA integration into the plant genome requires analysis of the plant target DNA before and after integration. The findings of two such studies of T-DNA insertion events are discussed here (Gheysen et al., 1991; Mayerhofer et al., 1991). In these studies, the nucleotide sequences overlapping the T-DNA ends and the plant target DNA before and after integration were studied.

The common observations were that T-DNA insertions do not usually create large rearrangements of plant DNA. Most insertions induced a small deletion, usually between 13-28 bp and sometimes up to 79 bp, at the target site (Gheysen et al., 1991). The deleted sequences show small patches of homology with the left and/or right T-DNA ends. Additional homologies are found between the plant DNA and T-DNA outside of the breakpoints and which is deleted at the final junctions. At a much lower frequency, larger chromosomal rearrangements may also occur, including large deletions or inversions of plant target DNA. These can happen when free plant ends are fused to other plant DNA nicks during repair of the junctions. A duplication of the target sequence is possible when integration occurs at breakpoints between sister chromatids within a replication bubble.

Another consistent observation was the presence of a short stretch of "filler" DNA at the T-DNA-plant junctions (Gheysen et al., 1991). Such filler DNA is composed of sequences homologous to plant DNA close to the junctions and may be explained by polymerase slippage or template switching during repair at the junctions.

One final observation was that T-DNA integration has no significant sequence requirement at the plant target site, although sequences that flank the T-DNA have been found to be relatively A/T rich (Gheysen et al., 1991, Mayerhofer et al., 1991). However, very short stretches of homology (5-10 bp) between the T-DNA ends and the plant target sites may have a role in integration (Matsumoto et al., 1990).

Illegitimate recombination is probably a two step process: DNA ends are first generated and then joined. There is no conclusive evidence whether the invading T-DNA is single- or double-stranded. However, other observations indicate that ssDNA generates a higher transformation frequency. The following mechanism for T-DNA integration was proposed by Gheysen et al. (1991). It assumes that the T-DNA ends, if not the entire T-DNA, is ssDNA. The invading T-DNA takes advantage of nicks or breaks, perhaps near the replication fork, and the invading (single-stranded) ends of the T-DNA form a heteroduplex with single-stranded target DNA in the gap. The small size of the nick in the target sequence indicates that the ends of the T-DNA are situated near each other at this point. The overhangs of the T-DNA ends are nicked and the T-DNA is ligated to the target ends. The finding that the right border is truncated less often could be due to the protection by the covalently attached VirD2 protein. However, this result may also be linked to a VirD2-supported integration mechanism. Truncated T-DNA borders, especially left borders (Deroles and Gardner, 1988; Gheysen et al., 1991) could be generated by breaks during or after transfer or by the very mechanism of integration (Gheysen et al., 1991).

Illegitimate recombination events in mammalian cells share many of the same characteristics (Roth et al., 1988). For example, SV40 (Simian Virus 40) integration sites are characterized by the absence of a well-defined target sequence requirement. SV40 also appears to require only patches of homology for integration; 1-6 complementary bp are sufficient. SV40 integration sites may

contain 1-40 bp of filler DNA. Furthermore, SV40 integration may also generate microdeletions and larger target rearrangements.

The preferential T-DNA integration into transcriptionally active regions (Koncz et al., 1989; Herman et al., 1990; Fobert, 1992) is readily explainable by this integration model. First of all, these genomic regions are more accessible to invading DNA due to unravelling of the nucleosomes of transcribed DNA, as compared to tightly coiled, transcriptionally silent sequences (Patient and Allen, 1989). According to this integration model, the integration of the T-strand into the plant genome occurs at nicks within the DNA sequence. Integration into these regions will predominate since they are characterized by a higher frequency of random, often transcriptionally related, nicks.

2.D. Gene Tagging

There are a number of approaches for isolating genes from plant genomes. One approach utilizes standard recombinant DNA technology by using cDNAs. This approach has proven very successful, especially for the isolation of genes that code for abundant mRNAs.

Another approach, gene tagging, utilizes the natural characteristics of certain DNA elements (T-DNA and transposable elements) that have the ability to insert into the plant genome (Balcells et al., 1991; Walden et al., 1991; Gierl and Saedler,

1992). With gene tagging, the insertion of a defined DNA segment in the vicinity of a DNA sequence or within the target gene occurs. The insertion of this foreign DNA (the tag) may alter the expression of the neighbouring gene, or, conversely, the surrounding sequences may alter the expression of the sequences contained within the tag. For the tagging to be successful, the insertion must result in the gain or loss of a function which then produces a screenable phenotype. Thereafter, it is relatively easy to use hybridization probes based on the sequence of the gene tag to isolate the sequence. So, the essential requirements for successful gene tagging are: a well characterized element that can insert randomly into the genome, a means to carry out the insertion and a scorable-detectable phenotype resulting from the insertion (Walden et al., 1991).

T-DNA and transposon tags are utilized in plants in an analogous manner to that of insertional elements in yeast (Roeder et al., 1980), P elements in *Drosophila* (Bingham et al., 1981) and retroviruses in animals (Gridley et al., 1987).

T-DNA tagging is also the approach used in a previous study (Fobert, 1992) that detected a genomic region in a transgenic plant line that exhibited seed-specific gene expression. It is this genomic region in the wild type and transgenic line that is the object of this study.

2.D.1. Transposon tagging

There are two approaches to gene tagging. The first utilizes plant transposable elements (Gierl and Saedler, 1992). Plant transposable elements were first described and characterized in *Zea mays* by Barabara McClintock. She named them "controlling elements" because they seemed capable of influencing the expression of a given locus when integrated into it or nearby. By the criteria of inheritance instability, transposable elements have been characterized in 35 different mono- and dicotyledonous plant species (Nevers et al., 1986).

Transposable elements have been very useful in identifying and isolating plant genes, especially from *Z. mays* and *Antirrhinum majus*. Some recent examples include a gene coding for chalcone synthase (Wienand et al., 1986) and other genes involved in photosynthetic membrane organization (Martienssen et al., 1989), flower morphogenesis (Sommer et al., 1990), as well as the biochemical processes in the anthocyanin pathway (Paz-Arez et al., 1986; McLaughlin and Walbot, 1987; Theres et al., 1987; Luo et al., 1991).

The maize elements Ac (Activator) and En/Spm (Enhancer/Suppressor-mutator) are currently being used (Balcells et al., 1991; Gierl and Saedler, 1992) to extend transposon tagging to plant species in which no suitable elements have been identified or exist. One advantage of gene tagging in heterologous species is that the transposon introduced is present in a low copy number (1-5

copies). This should make the identification of mutants much easier.

2.D.2. T-DNA tagging

A second approach to gene tagging employs the natural transformation process carried out by *A. tumefaciens* (Walden et al., 1991). The mechanisms of T-DNA transfer have been sufficiently elucidated (Zambryski, 1992) to permit the design of disarmed vectors used in T-DNA tagging. Apart from the border repeats, the T-DNA may be modified at will without interfering with the process of its transfer to the plant cell and integration into the genome.

T-DNA tagging may be carried out with specialized or non-specialized transformation vectors. Non-specialized T-DNA vectors are used to create insertional mutations. Conceptually, this approach is very similar to transposon tagging because both strategies capitalize on random insertion events that create visible/detectable mutations. The wild-type gene may then be easily isolated using the tag as a probe. However, this approach has the disadvantage that unless a visible mutation is created, it is ineffective. This has restricted this approach primarily to *Arabidopsis thaliana* for its small genome size, relatively low amount of repetitive DNA and low ploidy number. Additionally, *A. thaliana* presents the benefits of an extensive genetic map and a low generation time (Feldmann, 1991). There are several recent

reports that detail the isolation of novel plant genes by way of T-DNA insertional mutagenesis. These mutant phenotypes include dwarfs (Feldman et al., 1989), embryo lethals (Errampoli et al., 1991), plants whose stems lacked trichomes (Herman and Marks, 1989; Marks and Feldmann, 1989) and a homeotic flower mutation, agamous (Yanofsky et al., 1990), which eventually led to the identification of an *Arabidopsis* gene family of transcription factors which may play an important role in flower development (Ma et al., 1991). These examples highlight the strengths of this approach; however, they also indicate its weakness; besides the few plant species that may be used, only visible mutations will be detected. For example, tobacco, being tetraploid and with much repetitive DNA, is not a good candidate for insertional mutation research. In one recent study (Fobert, 1992), over a thousand transgenic plants were generated and visible mutations were not detected, even though the frequency of gene fusions (i.e. likely insertions into or directly before ORFs) was well over 5%.

T-DNA tags have been developed that may also have more specialized functions (Walden et al., 1991). Some specialized vectors can only react passively to the expression patterns of the chromosomal environment into which they have inserted. These transformation vectors will generally carry a fusion marker gene, such as *nptII*, *lux*, *luc* or GUS that will either lack a promoter sequence or be linked to a minimal promoter sequence. After the tag randomly inserts into the plant genome, if the plant sequences that flank the tag are able to direct gene expression, then this

will lead to the expression of the marker gene. The expression of the marker gene can be detected by sensitive assay systems. Otherwise the tag will not be expressed. These vectors must also contain a second marker gene, to select for transformants at the callus level. This approach has been used to functionally tag plant genomes for the identification of *cis*-regulatory elements (Koncz et al., 1989; Herman et al., 1990; Fobert et al., 1991) and for T-DNA insertional mutagenesis (Koncz et al., 1990). Fobert (1992) reports the isolation of a seed-specific *cis*-regulatory chromosomal region from tobacco. Koncz et al. (1991) report the isolation of a nuclear gene from *Arabidopsis thaliana* that codes for a chloroplast protein, using a combination of the specialized vector and insertional mutagenesis approaches.

However, other specialized vectors have been developed that are capable of modifying the expression pattern into which they have inserted (Walden et al., 1991). The published information on this topic is currently quite sparse. These vectors should permit the direct selection of a specific phenotype at the level of the primary transformant. This may be achieved with vectors that no longer respond passively to the expression patterns of the chromosomal environment into which they have inserted but, instead, actively modify it in a positive, or negative, way.

T-DNA tags could carry a transcriptional enhancer that would affect the expression of neighbouring genes by altering their normal expression patterns. This event could then be detected and screened for with the appropriate selection system. Depending on

the target gene, such an enhancer/tag would either give a constitutive or a developmental mode of expression.

Walden et al. (1991) reports experiments that use T-DNA tags with the cauliflower mosaic virus 35S RNA promoter located at the right border sequences, to transform tobacco protoplasts. The transformed protoplasts were selected for their ability to grow in the absence of auxin in the culture media. From these calli, a population of plants has been regenerated that displays differing phenotypes indicative of disturbances in endogenous auxin perception and/or action that is genetically linked to the T-DNA tag. The major limitation of such systems would be the limited selection systems.

3. MATERIALS AND METHODS

3.A. Bacterial strains

The bacterial strains that were used in this thesis are listed in Table 3.1.

3.A.1. Growth conditions

All *Escherichia coli* strains were grown in Luria Broth (LB; 10 g/L Bactotryptone, 5g/L Yeast Extract, 5g/L NaCl). All media were sterilized by autoclaving at 121°C for 17-20 minutes. If solid LB medium was required, then 1.5% Bactoagar was added prior to sterilization. *E. coli* strains were grown at 37°C for 5-7 hours and at 30°C overnight with moderate agitation (approximately 200rpm). The only antibiotic used in this study was carbenicillin at a concentration of 100µg/ml for liquid media and 200µg/ml for solid media. The carbenicillin was dissolved in sterile, double distilled water and stored at -20°C until needed. The antibiotic was added to the media following the sterilization procedure and after they had cooled to approximately 50°C.

3.B. Preparation of Competent Cells and Bacterial Transformation

Competent cells were prepared according to the method outlined by Dr. D. Johnson (University of Ottawa, Canada) and communicated

Table 3.1 List of bacterial strains used in this thesis.

<u>Bacterial strain</u>	<u>Genotype/Phenotype</u>	<u>Reference/Source</u>
<i>Escherichia coli</i> K12		
LE392	<i>hsdE514</i> (rk ⁻ , mk ⁻) <i>supE44</i> Promega <i>supF58 lacY1</i> or (<i>lacIZY</i>) ⁶ <i>galK2 galT22 metB1 trpR55</i>	(Madison) (WI, USA)
WA802	<i>lacY1</i> or (<i>lac</i>) ⁶ <i>supE44</i> <i>galK2 galT22-rfbD1</i> <i>metB1 hsdR2 McrA- McrB-</i>	Raleigh et al (1988)
DH5 α	<i>endA1 hsdR17</i> (rk ⁻ , mk ⁻) <i>supE44 thi-1 recA1 gyrA96</i> <i>relA1 ϕ80dlacZ M15</i>	BRL (Burlington) (Ontario)
(P2) SURE-RB:	<i>recJ, sbcC201, uvrC,</i> <i>umuC::Tn5(Km^r), mcrA, mcrB,</i> <i>mrr, lac, (hsdRMS), endA1,</i> <i>pro, gyrA96, thi-1, relA1,</i> <i>supE44 {F', proAB,</i> <i>lac19Z M15}</i>	Greener et al (1990)

to me by Dr. S. McHugh (Agriculture Canada). All manipulations were performed at 4°C, under sterile conditions and all solutions were cooled to 0°C in ice-water before use.

3.B.1. Preparation of competent cells

A frozen stock of DH5 α cells was used to streak an SOB (20 g/L Bactotryptone, 5 g/L Yeast Extract, 0.57g/L NaCl; 4.84g/L MgSO₄, pH 7.0 with KOH with 1.5% Bactoagar for solid media) plate. A single, large colony was used to inoculate 20 ml of SOC (SOB with 20 mM glucose) in a 250 mL Erlenmeyer flask prewarmed to 37°C. The culture was grown at 37°C while shaking at 200 rpm for about 2-3 hr until the absorption at 600 nm was approximately 0.2. The culture was chilled on ice for 10 min before being transferred to a sterile 50 mL polycarbonate tube and being spun in an HB4 rotor for 5 min at 4,000 rpm. The rotor was cooled to less than 0°C. Afterwards, the supernatant was discarded completely, the pellet resuspended by gentle shaking in 12 mL of 10 mM MgSO₄, left on ice for 20 min and spun as above. Then the MgSO₄ was discarded, the pellet was resuspended in 6 mL of 50 mM CaCl₂, left on ice for 20 min and spun as before. Finally, the CaCl₂ was also discarded and the pellet was gently dissolved in 1 ml of 50 mM CaCl₂. The cells were ready for transformation at this point. The competent cells prepared according to this method were used only when fresh.

3.B.2. Transformation of competent cells

Approximately 50-100 ng of plasmid DNA was transformed into competent cells. DNA was added to 200 μ L of competent cells in a 1.5 mL microfuge tube. The mixture was incubated on ice for at least 40 minutes and heat treated at 37°C for 5 min. The mixture was then returned to ice and 1.0 mL of room temperature SOB was added. After an incubation at 37°C for 30-60 min with moderate aeration (150 rpm), 50-200 μ L aliquots were plated on selective media together with 40 μ L of 20 mM X-Gal (5-Brom-4-chlor-3-indolyl- β -D-galactopyranoside) and incubated at 37°C for 12-18 hr. An additional 2-4 hr at room temperature was required for the blue colour of the background-cells transformed with a plasmid lacking the insert-to become clearly visible.

The following controls were used:

- (1) Competent cells that received DNA which had been treated exactly as the experimental DNA, except that (i) the ligase enzyme had been omitted from the ligation reaction, background control, (ii) the insert DNA had been omitted, so the plasmid vector was being re-ligated, ligation efficiency control (iii) the plasmid cloning vector had been omitted from the ligation mix, ligation negative control.
- (2) Competent cells were treated exactly as above, except that they did not receive transforming DNA, transformation negative control.
- (3) The efficiency of the competent cells was assessed by transforming the cells with a known quantity (10-20 ng) of CsCl

purified plasmid DNA.

3.C. DNA Extraction Protocols

3.C.1. Rapid extraction of plasmid DNA (Mini-Prep)

The boiling method of Holmes and Quigley (1981) was used for the rapid extraction of plasmids. Two ml of overnight bacterial culture was centrifuged for 60 seconds in a microfuge. The pellet was resuspended in 200 μ l of STET buffer (8% w/v sucrose, 5% v/v Triton-X-100, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA, pH 8.0) by vortexing. Twenty μ l of a lysozyme solution (10 mg/ml) was added (whether the lysozyme solution was freshly prepared or stored at -20°C made no difference) and the mixture was immediately incubated in a boiling water bath for 60 seconds. After centrifugation at 4°C for 30 min, the gelatinous pellet was discarded and to the supernatant an equal volume of cold (-20°C) isopropanol was added to precipitate the plasmid DNA. The mixture was incubated at -20°C for 30 min and centrifuged for 30 min. The DNA pellet was washed once with 500 μ l ice-cold 70% ethanol and once with 95% ethanol and dried at 37°C . The DNA was resuspended in 100 μ L TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0).

3.C.2. Large scale plasmid extraction

Two LB plates were flooded with 200 μ L of a fresh culture of

the appropriate bacterial strain and incubated overnight (ca 12-18 hr) at 37°C. Three mL of saline solution (0.9% NaCl) was added to each plate and the cells were scraped off with a sterile glass spreader. Alternatively, a 50 ml liquid bacterial culture was grown overnight in a 500 mL Erlenmeyer flask. The cells were then centrifuged at 5,000 rpm for 5 min and redissolved in 1 mL sucrose buffer (50 mM Tris-HCl, pH 8.0, and 25% w/v sucrose). Three hundred μ l of 0.5 M EDTA, pH 8.0, 125 μ l of 10 mg/ml lysozyme solution and 2.5 ml of Triton-X-lytic solution (50 mM Tris-HCl, pH 8.0, 60 mM EDTA, pH 8.0, 2% v/v Triton-X-100) were added and the solution was mixed by gentle inversion. After a 15 min incubation period on ice, lysed cells were centrifuged at 12,000 rpm for 25 min at 4°C. Another 4 g CsCl and 0.8 ml of a 4 mg/ml ethidium bromide (EtBr) were added to each 4.2 ml of cell lysate. The mixture was then placed in a Beckman Quick-seal ultracentrifuge tube. The remainder of the tube was filled with light paraffin oil, the tube was sealed with a Beckman tube heat-sealer and spun at 60,000 rpm for 16 hr or at 48,000 rpm for 48 hr. After centrifugation, the plasmid DNA (present in the lower of the two expected bands) was collected by needle and syringe under U.V. light and extracted with 10 mM Tris-HCl (pH 7.5)-saturated n-butanol to remove the EtBr. This was repeated 3-4 times. The DNA was precipitated by the addition of 2 volumes of sterile, distilled water followed by 2 volumes of ice-cold 95% ethanol. After an incubation period of at least 30 min at -20°C, the DNA was recovered by centrifugation for 15-30 min at 10,000 rpm at 4°C.

The DNA pellet was washed in 70% and 95% ethanol, dried at 37°C, dissolved in 200-400 μ l TE buffer and then quantified by U.V. absorption at 260 nm (an absorption of 1.0 reflects a DNA concentration of 45 μ g/ml).

2.C.3. Isolation of Lambda DNA

The plate lysate method of Sambrook et al (1989) with the modifications added by Dr. L. Robert and S. Bird (Agriculture Canada) was used. When larger DNA quantities were required, this method was scaled up to 10 or 20 plates.

2.C.3a. Elution of phage particles from solid media

Using a pasteur pipette, a single, well-isolated bacteriophage plaque was picked and placed in 1 ml of SM [phage buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM MgSO₄) and 2% gelatin] buffer containing a drop of chloroform in a small, sterile tube. The suspension is stored at room temperature for ca 2 hr, or at 4°C for 4-6 hr, to allow the bacteriophage particles to diffuse out of the agarose (SeaKem agarose, FMC BioProducts, ME, USA). The phage suspension is then transferred to a fresh tube and with a series of dilutions the concentration of plaque-forming-units/ μ l (pfu/ μ l) is estimated.

2.C.3b. Growth of Lambda Particles

Approximately, 105 pfu are added to 100 μ l of plating bacteria, *E. coli* strains LE 392 (Promega, Madison, WI) or WA 802 (Raleigh et al, 1988), and incubated at 37°C for 20-30 min. The above mix was added to 3 mL of molten (47-50°C) top agarose (10 g/L Bactotryptone, 5 g/L NaCl, 6 g/L agarose) and to 10 mM MgSO₄ (added after sterilization and just prior to use) and spread over the surface of a freshly poured plate of LB solid medium. In this protocol agarose was used instead of bactoagar because bactoagar contains potent inhibitors of restriction enzymes. Additionally, using freshly prepared media improved the Lambda DNA yield.

3.C.3c. Isolation of Lambda DNA

After the plates have been incubated at 37°C for at least 12 hr, 6 mL of lambda diluent (10 mM Tris-HCl, pH 7.5 and 10 mM MgSO₄) are added directly onto the top agarose. This is followed by 2 hr of storage at room temperature with constant, gentle shaking, to allow the bacteriophage particles to elute into the buffer. The lambda diluent was then centrifuged at 8,000 rpm for 10 min at 4°C and the supernatant was removed and treated with DNase I and RNase A, each to a final concentration of 1 μ g/ml for 30-45 min at 37°C. An equal volume of an ice-cold PEG/NaCl [20% w/v polyethylene glycol (PEG 8000) and 2 M NaCl in Lambda diluent) solution was then added, the tube was mixed by gentle vortexing and incubated for 1

hr in an ice water bath (ca 0°C). The precipitated bacteriophage particles were collected by centrifugation at 20,000 rpm for 10 min at 4°C. Special care was given to removing all of the supernatant because PEG remnants could complicate the following steps. The pellet was resuspended in 500 µl of TE (pH 8.0) and briefly centrifuged to removed any remaining bacterial debris. Five µL of 10% SDS were added and the mixture was incubated for 5 min at 68°C. Ten µl of 5 M NaCl were then added and the DNA was purified by extracting once with 1 volume phenol, once with 1 volume phenol:chloroform:isoamylalcohol (25:24:1) and once with one volume chloroform:isoamylalcohol (24:1). After thoroughly mixing the aqueous and organic phases, the tubes were centrifuged for 5 min following each extraction. An equal volume of isopropanol was added to the tubes, which were then incubated at -80°C for 15 min and centrifuged at 12,000 g for 30 min at 4°C. The pellet was washed in ice-cold 70% and 95% ethanol and then dried at room temperature. Finally, the recovered phage DNA was dissolved in 100-200 µl and quantified.

3.C.4. Isolation of plant genomic DNA

The method outlined by Sanders et al (1987) was used to isolate plant genomic DNA with the modifications added by P.R. Fobert (John Innes Institute, UK) and S. McHugh (Agriculture Canada). Using a mortar and pestle, 2.5-3.0 g of freeze-dried material was pulverized to a fine powder in the presence of 0.5 g

alumina/g tissue, liquid nitrogen and 15 ml extraction buffer (50 mM Tris-HCl, pH 8.0, EDTA, pH 8.0, 50 mM NaCl, 1% N-Laurosarkosyl). The mixture is then placed in a 50 ml polypropylene tube and spun for 10 min at 10,000 rpm. At the end of this, the supernatant is purified with phenol extraction; an equal volume of phenol is added, the solution is mixed well and spun as above. The DNA is recovered from the aqueous by ammonium acetate precipitation. The DNA pellet is washed with 70% and 95% ethanol and redissolved in 9 ml of TE buffer with the addition of 9 g (1g/ml of mixture) CsCl (the density of the mixture at this point should be 1.5 g/ml), together with 1 ml of 400 µg/ml EtBr. This mixture is then spun in a Beckman ultracentrifuge for 16 hr (at 60,000 rpm) or 48 hr (at 45,000 rpm). Only one DNA band should be visible under U.V. light. This band is extracted and purified (see 3.C.2.).

3.C.5. Isolation of DNA fragments

Restriction fragments were isolated from agarose gels using the GENECLAN kit and the manufacturer's protocol (BIO 101 Inc, LaJolla, CA). When a restriction fragment smaller than 0.4 kb was needed as a radioactive probe, this was isolated from low-melting-agarose according to Sambrook et al (1989).

3.C.5a. Isolation of DNA fragments with GENECLAN

DNA fragments were isolated from TAE (40 mM Tris-HCl, pH 8.0;

20 mM sodium acetate, 2 mM EDTA, pH 8.0) agarose gels. At the end of the GENE CLEAN procedure, the concentration of the purified fragment(s) was estimated by running an aliquot on an agarose gel next to a known quantity of lambda DNA digested with Hind III.

3.C.5b. DNA fragment isolation with Low-Melting-Temperature

Agarose

DNA digested with the appropriate enzymes was separated on a 1% TAE low-melting-temperature agarose (SeaPlaque, Low-Gelling-Temperature agarose, FMC Co, ME, USA) gel already containing 0.5 µg/ml EtBr. The gel was poured and run at 4°C at a low voltage (ca 30 volts). The DNA was visualised on a long wavelength U.V. transilluminator. Desired bands were cut out from the gel, transferred to a microfuge tube and heated at 70°C for 15 min to melt the agarose. A sufficient quantity was then transferred immediately to 37°C (the concentration was roughly estimated on the basis of the original quantity digested and run) and used directly for random priming. DNA in low-melting-temperature agarose was stored for short periods of time at 4°C and reheated at 70°C each time just prior to use.

2.D. Ethanol precipitation of DNA

DNA was precipitated by bringing the NaCl concentration of the mixture to 200-250 mM, by the addition of 2.2-2.5 volumes of

ice-cold 95% ethanol and the incubation at 4°C for at least 30 min. The DNA was then collected by centrifugation at 4°C for 30 min. The pellet was washed with 70% and 95% ice-cold ethanol and then dried at 37°C or in a SpeedVac Concentrator (Savant Instruments, Farmingdale, NY, USA).

Alternatively, in some instances, DNA was precipitated as described above except 1/10 volume 10 M ammonium acetate and 2 volumes ice-cold 95% ethanol were added.

3.E. DNA Restriction Digestion

Plant genomic and plasmid DNA were digested using the buffers provided by the manufacturer. All digestions were performed at 37°C. If the DNA was to be digested with more than one enzyme, this was done simultaneously, if they had the same buffer requirements. If they did not, then the DNA was digested with one enzyme first, ethanol precipitated, redissolved in water and then digested with the next enzyme. With some enzymes used widely in this study, the buffers differed only with respect to the NaCl concentration needed. It was therefore possible to digest the enzyme first with the enzyme that required the least amount of NaCl, then add the appropriate amount of NaCl and continue the digestion with the second enzyme. It was important that the first enzyme be inactivated at this point (which occurs after a digestion of ca 3 hr at 37°C). Enzymes that were capable of non-specific or "Star" activity were always used last.

3.E.1. Restriction digestion of plasmid DNA

Plasmid DNA was digested in small volumes (20-30 μ l) and for short periods of time (0.5-2 hr). Three units of enzyme per μ g of plasmid DNA for a single restriction site were used.

3.E.2. Restriction digestion of plant genomic DNA

Plant DNA digests were supplemented with a 2.5 mM concentration of spermidine trihydrochloride. Genomic DNA was digested for several hours (ca 3 hr) for every enzyme used separately; the digestion process was aided by intermittent and gentle mixing. The digested DNA was precipitated by ethanol precipitation and redissolved in water, prior to gel electrophoresis. This reduced the total volume and removed any salts from the buffer that might distort the electrophoresis running pattern.

Five units of enzyme were added per μ g of DNA.

3.E.3. Restriction digestion of Lambda DNA

Lambda DNA, isolated according the procedure described in 2.C.3., could not be digested by restriction enzymes using the manufacturer's buffer. So, the restriction buffer used in this instance was TA (33 mM Tris-Acetate, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DL-dithiothreitol, 100 μ g/ml gelatine)

which was described by O'Farrell et al (1980).

The ratio of enzyme per μg of DNA for one restriction site was 1:1:1.

3.F. DNA Hybridization Techniques

3.F.1. Southern hybridizations

Southern hybridizations were performed in the manner described by Rutledge et al (1991) with the modifications added by P.R. Fobert (1992).

3.F.1a. Agarose gel electrophoresis and Southern blotting

Five to 10 μg of genomic DNA and 0.5-1.0 μg of plasmid or lambda DNA was digested as described in Section 3.E. and fractionated on a 0.8% TAE agarose gel at moderate voltage (20-50 volts depending on the duration of the run). End-labelled *Hind* III and *Hind* III/*EcoR* I lambda DNA (ca 5,000 cpm Cerenkov for an overnight exposure) were used as molecular weight markers.

When the DNA had migrated 3/4 of the gel, the gel was stained in EtBr and photographed. The DNA was then (i) denatured and (ii) neutralized by soaking (with moderate agitation at room temperature) in 2-3 volumes of the following solutions. For genomic blots, where the Nytran membrane (Schleicher and Schuell, Keene, NH, USA) was used, these were: (i) 0.5 M NaOH, 1.5 M NaCl (45 min)

and (ii) 1 M Tris-HCl, pH 7.0, 1.5 M NaCl, 1 mM EDTA (twice for 25 min). For all other blots, where the Hybond membrane (Amersham, Arlington Heights, IL, USA) was used, the solutions were the same except that (ii) had the addition of 1 mM EDTA and the incubation times were 30 min each.

The DNA was transferred to a piece of membrane, cut to the same size as the gel, by being placed above 4 layers of Whatman 3MM paper soaked in 20 X SSC (3 M NaCl, 0.3 M ammonium citrate). Directly above the gel was the membrane (Nytran was soaked in 5 X SSC) and three layers of Whatman 3MM paper the bottom one soaked in 20 X SSC. On top of this was 2 inches of paper towels, a light weight and it was covered by Saran Wrap, to keep the assemblage from drying out. DNA was transferred to the membrane overnight and was then immobilized onto the membrane by U.V. irradiation for 2-5 min on a transilluminator and rinsed in 5 X SSC.

3.F.1b. Preparation of radioactive probes

Twenty-five to 50 ng of an isolated restriction fragment was used as a template for random priming (Feinberg and Vogelstein, 1983) using kits purchased from BRL (Burlington, Ontario) and the manufacture's specifications. [α^{32} -P]-dCTP (3,000 cpm/mmol) (Dupont Corp, DE, USA) was used as a radioactive label.

Radioactive molecular weight markers were prepared according to the protocol provided by T. Ouellet (Agriculture Canada). This procedure used the Klenow fragment to end-fill with radioactive

nucleotides the sticky ends of *Hind* III or *Hind* III/*Eco*R I digested lambda DNA. To 1 μ g of digested lambda DNA was added 1 mM of each dATP, dGTP, dTTP, 1/10 volume restriction buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 50 mM NaCl), 40 μ Ci [α 32-P]-dCTP (3,000 cpm/mmol) and 1 unit Klenow fragment (BRL, Burlington, Ontario). The mixture was then incubated at 25°C for 10 min and the reaction stopped by the addition of 1 μ l of 0.5 M EDTA.

In both procedures, the unincorporated nucleotides were removed by chromatography through a Sephadex G-50 Nick Column (Pharmacia, Montreal, PQ).

The radioactive probes were denatured for 10 min in boiling water and placed on ice for at least 5 min, prior to use. If the nucleotide incorporation of the radioactive weight marker was been over 50-60%, then 5,000 cpm per lane was sufficient for overnight exposure.

3.F.1c. Hybridization and autoradiograph

The membranes were prehybridized from at least 0.5 (Hybond) or 1 (Nytran) to several hours at 45°C in: (Nytran) 1 M NaCl, 1% N-laurosarkosyl, 50 mM Tris-HCl (pH 8), 1 mM EDTA and 50 μ g/ml yeast t-RNA or (Hybond) 5 X Denhardt's solution (2% w/v for each of: BSA (bovine serum albumin), PVP (polyvinyl pyrrolidone) and ficoll), 5 X SSPE (120 mM NaCl, 10 mM NaPhosphate and 1 mM EDTA), 0.2% SDS and 500 μ g/ml yeast t-RNA, in heat-sealable bags with ca 5 ml of solution per 100 cm². They were hybridized in the same

buffer, with the addition of the radioactive probe, in the same conditions overnight (at least 12 hr). For genomic blots 5×10^5 to 1×10^6 cpm/ml were added, whereas lambda or plasmid blots required less cpm/ml, between 1×10^4 and 2×10^5 .

Membranes were then washed with moderate agitation, to remove non-specifically bound probe. Nytran membranes were washed firstly with 1 X SSC, 0.1 M NaPhosphate, pH 7.0, 0.1% w/v N-Laurosarkosyl (twice for 15 min each time at room temperature), secondly with 0.2 M NaCl, 0.5% w/v N-laurosarkosyl, 1 x TE (under the same conditions) and thirdly with 0.5% w/v N-laurosarkosyl and 1 X TE (twice for 30 min at 65°C). Hybond membranes were washed at 65°C firstly in 2 X SSC (15 min), secondly in 2 X SSC and 0.1% w/v SDS (30 min) and thirdly in 0.1 X SSC (10 min).

Membranes were exposed to Kodak X-Omatic AR film at -80°C in Kodak X-Omatic cassettes with regular intensifying screens. If the membrane was to be reused, then the probe was eluted from the membrane in the following ways. Nytran membranes were boiled for 20 min in: 0.01 X SSC, 0.1 mM NaPhosphate, pH 7.0, 0.01 mM EDTA (pH 8) and 0.5% w/v SDS, then rinsed in 2 X SSC at room temperature. Hybond membranes were incubated for 30 min at 45°C in: (i) 0.1 M NaOH and then (ii) 0.1 X SSC, 0.1 w/v SDS and 0.2 M Tris-HCl, pH 7.5. Thereafter, the hybridization procedure was exactly as described above.

3.F.2. Plaque Hybridization

3.F.2a. Preparation of lysed bacterial plates

For the primary screening of the lambda gene banks, from 3×10^6 to 6×10^6 (see Section 4.B.) were mixed with 1 ml of an overnight culture of WA 802 (Raleigh et al, 1988) cells and incubated at 37°C for 30 min. The suspension was divided into parts (so as to keep the pfu per plate at 1.5×10^6), each part was added to 30 ml molten (47-50°C) top agar and spread onto a 200 X 300 mm Pyrex plate containing 350 ml LB agar. The dish was then incubated in an inverted position at 37°C for 12 hr.

For the secondary and tertiary level screenings, small, square petri dishes were used and between 100-200 pfu were plated per dish for each plaque to be well isolated from each other.

3.F.2b. Screening procedure

After the incubation period, the plates were transferred to 4°C for 1 hr and then directly lifted (in duplicate) onto a Hybond membrane. The membranes were given characteristic labels and these labels were transferred as precisely as possible onto the dish.

The membranes were denatured in the appropriate Hybond solutions for 7 min and neutralized twice for 3 min each time. Thereafter, they were transferred to 2 X SSC for 5 min, the DNA immobilized by U.V. irradiation and they were washed (by gently

rubbing with a gloved hand) of any agar particles in 2 X SSC. The same procedure for Southern blots with Hybond membranes was followed here, as well.

The characteristic markings of the membranes were transferred to the autoradiograms and these permitted correlation between putative, positive signals on the autoradiograms with a specific area on the plate or a plaque (in further levels of screening). These areas on the dish were removed using the large end of a sterile Pasteur pipette, transferred to a microfuge tube containing SM buffer and the phage particles were eluted (as described in Section 3.C.3a.), replated and screened as for the primary screening.

3.G. DNA Sequencing

In this study, DNA was sequenced according to the dideoxy chain termination method of Sanger et al (1977). Sequencing was performed on double-stranded plasmid DNA templates. The plasmid template was either CsCl-purified or, alternatively, mini-prep DNA was phenol-purified. The kits were purchased from Pharmacia (Montréal, PQ) using [α^{35} S]-dATP (1,000 μ Ci/mmol) as radioactive label.

All templates originating from clone pJC-1 were inserted into the pGEM4Z vector. The primers T7 and SP6 specific to pGEM 4Z plasmid were used. Several other internal primers, specific to this genomic region were also used. Eight of these (PF3, PF4,

PF5-6, PF8-11) originated from a different study (P.R. Fobert, 1992). The primer JC-1 was synthesized especially for this study. All primers were synthesized by Dr. J. Hattori (Agriculture Canada, Ottawa

The manufacturer's protocol was followed precisely for the sample preparation. The nucleotides were resolved on a 6% denaturing polyacrylamide (ultrapure grade; BRL, Burlington, Ontario) gel. The electrophoresis buffer (TBE; 1 X TBE is 89 mM Trisma base, 2.5 mM EDTA, pH 8.0, 89 mM boric acid) was alternatively 1 X or 0.5 X concentration. After the end of the electrophoresis, the gel was placed on Whatman 3MM paper dried under a vacuum in a gel drier at 80°C for 3 hr. Dried gels were exposed to Kodak X-Omatic film at room temperature for 24-48 hr.

4. RESULTS

4.1. Molecular Analysis of the seed coat-specific genomic region in untransformed *Nicotiana tabacum* SR1.

The genomic region in *N. tabacum* SR1 that corresponds to the seed coat-specific region identified by T-DNA tagging in transgenic plant line T218 (see Introduction) was analyzed for its copy number in the tobacco genome and its restriction pattern. The restriction enzymes *Eco* RI, *Hind* III and *Xba* I were selected because these had also been used for the analysis of the seed coat-specific genomic region in plant line T218 (Fobert, 1992).

A Southern blot with tobacco genomic DNA was probed with the 0.35 kb *Xba* I-*Kpn* I fragment that is situated directly upstream of the T-DNA insert in line T218 (probe #2, see Figure 4.2). The results obtained (see Figure 4.1.) suggest that probe #2 is specific to only one *N. tabacum* SR1 genomic region because the probe hybridizes to only one band in genomic DNA digested with *Eco* RI and *Hind* III. For genomic DNA digested with *Xba* I (lane 3), a second band (2.6 kb) also appears in addition to the one stronger band (1.4 kb). This band is most likely the result of a partial digest.

This same Southern blot analysis indicates that the sizes of the *Eco* RI, *Hind* III and *Xba* I fragments are: 3.0 kb, 3.3 kb and 1.4 kb. Assuming that rearrangements in the pT218 genomes have not occurred, the pT218-1 partial map (Fobert, 1992) and the data from

FIGURE 4.1. A Southern blot hybridization of *Nicotiana tabacum* SR1 genomic DNA with probe #2 (see Figure 4.2.). Lanes 1-3 contain *N. tabacum* DNA digested with *Eco* RI, *Hind* III and *Xba* I, respectively. Lanes 4 and 5 contain probe #2 DNA as a positive control. The sizes of the restriction fragments that hybridize to the probe are the following: *Eco* RI (3.0 kb), *Hind* III (3.3 kb) and *Xba* I (1.4 kb). A second band can also be seen in lane 3 (2.6 kb) which is possibly the result of a partial digest.

1 2 3 4 5

FIGURE 4.2. (A) A partial map of the pT218-1 clone (adapted from Fobert, 1992) and (B) a partial map of the genomic region of interest. The three pT218-1 fragments that were used as probes are indicated above the pT218-1 map. These are probe #1: 2.6 kb fragment, probe #2: 0.35 kb fragment and probe #3: GUS gene. The partial map of L3 was constructed based on restriction fragment sizes from the Southern blot analyses with probe #2 (see Figure 4.1.) and probe #1 (see Figure 4.4.). (R1: *Eco* RI, Hc: *Hinc* II, H3: *Hind* III, Kp: *Kpn* I, Sp: *Sph* I, St: *Sst* I, X: *Xba* I).

this Southern blot analysis were used to construct a partial map for the genomic region. According to the pT218-1 map (Figure 4.2.A.), *Eco* RI, *Hind* III and *Xba* I restriction sites exist 2.4 kb, 2.0 kb and 0.35 kb, respectively, upstream of the site of T-DNA insertion. Given *Eco* RI, *Hind* III and *Xba* I fragments that are 3.0 kb, 3.3 kb and 1.4 kb in size, respectively, it was estimated that *Eco* RI, *Hind* III and *Xba* I sites may exist 0.6 kb, 1.3 kb and 1.05 kb, respectively, downstream from the site of T-DNA insertion (Figure 4.2.B.).

4.2. Isolation of Lambda Clones

4.2.1. Cloning of the Genomic Region Corresponding to pT218-1.

A Lambda-Dash II (Stratagene) genomic library was screened to isolate Lambda clones carrying the *N. tabacum* SR1 genomic region corresponding to the seed coat-specific region. This library was custom-made by Stratagene with *N. tabacum* SR1 genomic DNA. The genomic DNA was partially digested with the restriction enzyme *Mbo* I and cloned into *Bam* HI sites.

To screen the genomic library, three aliquots of the library termed Library 1, 2, and 3 were screened (see Table 4.1.). Two different probes were used, probe #1 (2.4 kb) and probe #2 (0.35 kb). Both correspond to different regions of the cloned seed coat-specific T218 genomic region, referred to as pT218-1 (see Figure 4.2). Probe #2 had been used previously in the Southern blot

Table 4.1. Screening the genomic libraries L1, L2, L3**Total number of positive signals^a**

	L1 ^b	L2 ^c	L3 ^d
Primary screening:			
Probe #1 ^e	12	26	>40
Probe #2 ^f	NA ^g	NA ^g	4
Secondary Screening:			
Probe #2	NA ^g	1	1
Tertiary Screening:			
Probe #2	NA ^g	1	1

^aPositive signals were identified on duplicate autoradiograms.

^bIn Library #1 3×10^5 pfu were screened.

^cIn Library #2 4×10^5 pfu were screened.

^dIn Library #3 6×10^5 pfu were screened.

^eRefers to the 2.6 kb probe (Figure 4.2.A.)

^fRefers to the 0.35 kb probe (Figure 4.2.A.)

^gNA: Not Applicable. This level of screening for the particular Library was not performed with the probe in question.

analysis (Figure 4.1) and had been shown to hybridize to only one genomic region in *N. tabacum*.

Library #1 (3×10^5 pfu) was screened only with probe #1. None of the twelve putative positives passed the secondary screening step (see Table 4.1.).

Library #2 (4×10^5 pfu) gave 26 putative positive signals at the primary level with probe #1. These primary level filters were not rescreened, but the secondary level screening was done only with probe #2. Only one of these twenty-six positives passed both the secondary and tertiary screening (see Table 4.1.).

Library #3 (6×10^5) gave over forty putative positive signals at the primary screening level with probe #1. The primary level filters were rescreened with probe #2 and four signals passed this screening. Of these four, one passed the following two levels of screening (see Table 4.1.).

The two clones that passed the secondary and tertiary level of screening were named L2 (originating from Library #2) and L3 (originating from Library #3).

4.2.2. Characterization of the Lambda clones.

The restriction patterns of the lambda clones L2 and L3 were compared to that of *N. tabacum* SR1 genomic DNA. A Southern blot analysis with four different restriction digests (*Eco* RI, *Hind* III, *Xba* I and *Eco* RI/*Xba* I) were performed on L2, L3 and pT218-1 DNA (see Figure 4.3.). This Southern blot was analyzed with both

FIGURE 4.3. (A) An ethidium bromide gel of L3, pT218-1 and L2 DNA digests. Lanes 1-4 contain L3 DNA, lanes 5-8 pT218-1 DNA and lanes 9-12 L2 DNA with various combinations of restriction enzymes. These combinations are: (1,5,9) *Eco* RI, (2,6,10) *Hind* III, (3,7,11) *Xba* I and (4,8,12) *Eco* RI/*Xba* I. Southern hybridizations of the same gel follow with probe #2 (B) and probe #1 (C). The sizes of the molecular weight marker are given in kb on the left side. On the right side, the size of various L3 and pT218-1 fragments are given. These Southern analyses demonstrated that L2 was a false positive and that the L3 clone seems to contain the region of interest. This conclusion was based on the fact that various L3 DNA fragments with the expected sizes (see Figure 4.2.) hybridize to probe #1 (Part A: a 3.0 kb *Eco* RI, a 1.4 kb *Xba* I and a faint 0.8 kb *Eco* RI/*Xba* I fragment) and to probe #2 (Part B: 3.0 kb *Eco* RI, 1.4 kb *Xba* I and 1.0 kb *Xba* I fragment).

08

1 2 3 4 5 6 7 8 9 10 11 12

06
0.5

probes #1 and #2. Neither probe hybridized to L2 DNA (lanes 9-12), under conditions specific for the pT218-1 region, so the L2 clone was determined to be a false positive. In contrast, L3 DNA yielded restriction patterns on Southern blots that were consistent with those observed with tobacco genomic DNA (Figure 4.1.) for the restriction digests with *Eco* RI, *Xba* I and *Xba* I/*Eco* RI with probe #2 (see Figure 4.3.B.). However, L3 was later found to be contaminated with an unidentified second Lambda phage that resulted in the complex patterns of bands detected by ethidium bromide staining (Figure 4.3.A.). In two Southern blots, 3.0 kb *Eco* RI and 1.4 kb *Xba* I fragments (see Figures 4.1. and 4.4.) hybridized to the probe.

Probe #1 which hybridizes to the various small pT218-1 fragments (*Xba* I, *Hind* III and *Eco* RI/*Xba* I, see Figure 4.2.) was used to better compare the Lambda L3 restriction patterns to those of clone pT218-1. The Southern hybridization with probe #1 shown in Figure 4.3.C. was weak did not give clear results. Therefore, the Southern hybridization with another filter in Figure 4.4. are presented. This Southern analysis permits a closer comparison between the L3 and pT218-1 restriction patterns (see Figure 4.4.). The L3 fragments with similar sizes to pT218-1 fragments (see Figures 4.2. and 4.3.C) are the following: the *Xba* I (1.0 and 0.6 kb) fragments, the *Eco* RI/*Xba* I (0.5 kb) fragment and the *Hind* III (0.35 kb) fragment (see Figure 4.4.). This is a strong indication that L3 clone contains the genomic region that corresponds to the seed coat-specific region from T218. Neither probe hybridizes to

FIGURE 4.4. A Southern blot of L3 DNA with probe #1. Lanes 1 and 2 contain the Lambda molecular weight marker. Lanes 3-9 contain L3 DNA digested with (3) *Eco* RI, (4) *Hind* III, (5) *Xba* I, (6) *Eco* RI/*Hind* III, (7) *Eco* RI/*Xba* I, (8) *Hind* III/*Xba* I and (9) *Eco* RI/*Hind* III/*Xba* I. The sizes are given in kb. The objective of this Southern analysis was to complement the Southern analysis in Figure 4.3.C. by comparing the L3 restriction patterns in the region of interest with the pattern expected from pT218-1. These two patterns match very closely. Fragments with similar sizes between pT218-1 and L3 are the 1.0 kb and 0.6 kb *Xba* I (lanes 5, 7-9), 0.5 kb *Eco* RI/*Xba* I (lane 6) and 0.35 kb *Hind* III (lanes 4,6,8-9) fragments (see Figure 4.2.A.). The 3.0 kb *Eco* RI (lane 3) and 1.4 kb (lane 5) *Xba* I fragments identified in the genomic Southern (see Figure 4.1.) may also be seen.

1 2 3 4 5 6 7 8 9

FIGURE 4.5. A partial map of the Lambda L3 clone positioning the L3 *Eco* RI fragments on the L3 insert. The insert is 13.4 kb in size. (R1: *Eco* RI, H3: *Hind* III, X: *Xba* I).

(Constructed by Dr. Jiro Hattori)

a 3.3 kb *Hind* III fragment because the L3 sequences that correspond to the this fragment remain attached to the Lambda arm.

The following conclusions were drawn from this Southern blot analysis (Figure 4.4.).

(1) The sizes of L3 *Eco* RI and *Xba* I fragments correspond to the sizes of the respective fragments from a genomic blot (Figure 4.1.). This indicates that the L3 clone contains the complete pT218-1 seed coat-specific sequences and that these sequences are not truncated at the edge of the Lambda clone.

(2) The 3.0 kb L3 *Eco* RI fragment may be a good candidate for further subcloning since it appears to contain the entire 2.6 kb seed coat-specific region, together with 0.4 kb of downstream sequences.

These restriction patterns were used to construct a partial map of the L3 clone in the region of interest for the enzymes *Eco* RI, *Hind* III and *Xba* I (see Figure 4.2).

4.2.3. Mapping and cloning of L3.

The size of the entire Lambda L3 insert was estimated to be 13.4 kb. A partial map of the entire L3 insert was constructed (see Figure 4.5.). (This information was provided by Dr. Jiro Hattori (Agriculture Canada, Ottawa) who detected the Lambda contaminant in L3 and corrected the data).

The L3 *Eco* RI fragments were subcloned simultaneously in a pGEM-42 cloning vector in DH5 α cells (see Figure 4.7.). In Figure

4.7., as a result of the Lambda contaminant in L3, only the subcloned fragments in lanes 2, 4 and 6 are actually from the L3 insert. The 3.0 kb subclone was tentatively identified as the region of interest by the size of the insert.

4.2.4. Southern analysis of the L3 3.0 kb *Eco* RI fragment

The next step was to verify that the 3.0 kb *Eco* RI fragment and no other L3 fragment (i.e. the 3.3 kb *Eco* RI fragment) contains the genomic region of interest. A Southern blot analysis was performed (Figure 4.6.) with the following DNA fragments (see Figure 4.7.): the 5.1 kb (Lane 2), 4.8 kb (Lane 3), 3.3 kb (Lane 4) and 3.0 kb (Lane 5) *Eco* RI fragments. When the contaminant was detected, it was realized that the first two fragments (the 4.8 kb and 5.1 kb fragments that had been added as negative controls) were actually from the contaminant. The Southern blot was probed with probe #2. The results indicated that it is indeed the 3.0 kb L3 *Eco* RI fragment that corresponds to the seed coat-specific region pT218-1.

The final step in the identification of the pJC-1 clone was a Southern blot analysis with probe #2 for all seven different L3 *Eco* RI subclones (Figure 4.7.). L3 DNA digested with *Eco* RI was added for comparison (lanes 1 and 7). The clone of interest (pJC-1) was identified in lane 6 (Figure 4.7.A.). This was evident from the size of the insert which is 3.0 kb-just larger than the pGEM-42 cloning vector at 2.75 kb, from the intense band detected

by the Southern blot analysis (Figure 4.7.B.) and the similar size of this insert and the 3.0 kb *Eco* RI L3 fragments running alongside it (lanes 1 and 7).

It is apparent from the Southern blot analysis (Figure 4.7.B.) that the pGEM-4Z vector has cross-hybridized to our probe. This may be the result of contamination of the probe DNA with vector sequences (pUC 19) during its excision. However, neither the other L3 clones (lanes 2 and 5-6) nor the two negative controls (lanes 10-11), which are unrelated tobacco genomic DNA from L2 subclones, nor the Lambda contaminant DNA (lanes 3-4 and 8-9) hybridize to the probe. These results indicate that subclone pJC-1 (lane 6) contains the region of interest.

4.3. Restriction pattern analysis of 3.0 kb pJC-1 clone.

The identity of the 3.0 kb *Eco* RI insert in pJC-1 was further established by comparing Southern blot analyses of certain restriction digests of pJC-1 and pT218-1 (see Figure 4.8.).

The restriction patterns produced by these two clones match the map shown in Figure 4.2. for the 3.0 kb *Eco* RI fragment and pT218-1. In lanes 4 and 6 (see Figure 4.8.B.) a novel fragment 0.8 kb *Eco* RI/*Xba* I, is unique to pJC-1 (see map in Figure 4.2.B.). This fragment contains sequences that correspond to the 0.35 kb *Xba* I/*Kpn* I fragment used as probe #2 (Lanes 12 and 13, Figure 4.8.B.) from line T218 and additional downstream tobacco sequences.

FIGURE 4.6. Southern blot hybridization with probe #2 of four L3 *Eco* RI fragments (see Figure 4.7., lanes 1 and 7) to verify which one contains the genomic region of interest. The marker in lane 1 is Lambda *Hind* III. Lanes 2 and 3 contain the two negative controls: the 5.1 kb (lane 2) and 4.8 kb (lane 3) *Eco* RI fragments. Lane 4 contains the 3.3 kb and Lane 5 the 3.0 kb *Eco* RI fragments. The 3.0 kb fragment in lane 5 hybridizes to probe #2, so it must contain the region of interest. This was anticipated, since the size of the *Eco* RI fragments that hybridize in the Southern in Figures 4.3. and 4.4. are also 3.0 kb. All sizes are given in kb.

FIGURE 4.7. Ethidium bromide and Southern blot analyses of pJC-1 and pT218-1 digests to compare the restriction patterns of the two genomic regions. (A) Lanes 1 and 7 contain DNA from L3 (with the contaminant DNA) restricted with *Eco* R I. Lanes 2 and 5-6 contain three L3 subclones digested with *Eco* RI. Clone pJC-1 which contains the 3.0 kb L3 *Eco* RI fragment is in lane 6. Lanes 3-4 and 8-9 contain DNA that has now been identified as originating from the Lambda contaminant in L3. Lanes 10 and 11 contain tobacco DNA as negative controls. The tobacco DNA used as negative controls are two L2 subclones in pGEM-4Z. The common fragment among all the clones is the cloning vector pGEM-4Z. (B) A Southern hybridization of this gel with probe #2. The only inserts to hybridize to the probe are the 3.0 kb fragments in lanes 1, 6 and 7, demonstrating that the pJC-1 insert corresponds to the L3 3.0 kb *Eco* RI fragment.

Some large fragments also hybridize to probe #2 (lanes 3, 5, 7-8, 9-11 and 14-15) because the 0.8 kb region and the 0.35 kb Xba I/Kpn I (probe #2) region remain attached to the vector DNA, due to the restriction enzyme combinations used in those particular lanes.

A similar Southern blot analysis with probe #1 was used to determine whether the various small restriction fragments with similar sizes in pJC-1 and pT218-1 restriction digests are homologous. This was demonstrated when they all hybridized to the large probe #1 (Figure 4.8.C).

One final question that was raised concerning the identity of the pJC-1 clone was that it might be some sort of pT218-1 derivative or contaminant. If this were true, then pJC-1 would have to contain the GUS gene (see Figure 4.2.A.). To determine whether this was the case, the same Southern blot was reprobed with the GUS gene (probe #3). It was observed (see Figure 4.8.D.) that only sequences belonging pT218-1 (lanes 9-15) and not pJC-1 sequences (lanes 2-8) hybridized to the probe, thus eliminating this concern.

4.4. Sequence analysis of clone pJC-1

The nucleotide sequence of clone pJC-1 and of the 1.4 kb unsequenced upstream portion of clone pT218-1 was analyzed by the dideoxy chain-termination reaction (Sanger et al., 1977). The templates used were the entire pJC-1 and pT218-1 clones, as well as a number of smaller fragments that were subcloned for this purpose

FIGURE 4.8. (A) An ethidium bromide gel of pGEM-4Z, pJC-1 and pT218-1 digests. Lane 1 contains pGEM-4Z digested with *Eco* RI. Lanes 2-8 contain clone pJC-1 DNA and lanes 9-15 contain clone pT218-1 DNA digested with various combinations of restriction enzymes. These combinations are: (2,9) *Eco* RI, (3,10) *Xba* I, (4,11) *Eco* RI/*Xba* I, (5,12) *Kpn* I/*Xba* I, (6,13) *Eco* RI/*Kpn* I/*Xba* I, (7,14) *Hind* III and (8,15) *Eco* RI/*Hind* III. This is followed by a Southern hybridization of the same gel with (B) probe #2, (C) probe #1 and (D) probe #3. Figure 4.8.D. is a negative control performed to discredit the possibility that clone pJC-1 might actually be a pT218-1 derivative. The fact that pJC-1 is not a pT218-1 derivative is demonstrated because only pT218-1 fragments (lanes 9-15) hybridize to the probe (the GUS gene, see Figure 4.2.).

in pGEM-4Z (see Figure 4.9.). For pJC-1, these were the 0.5 kb *Eco* RI/*Xba* I (pF4), 1.0 kb *Xba* I (pF5), 0.6 kb *Xba* I (pF6) and 0.8 kb *Eco* RI/*Xba* I (pF7) fragments. For pT218-1, these were the 0.5 kb *Eco* RI/*Xba* I (pT218-F4) and the 1.0 kb *Xba* I (pT218-F5) fragments.

The double-stranded plasmid DNA templates were sequenced from either ends of these clones using the T7 and SP6 primers which run off the pGEM-4Z plasmid (see Figure 4.9.). A number of additional primers, that are internal to these clones, were also used to complete the analysis. Most of these primers (PF3, PF4, PF5, PF6, PF8, PF9, PF10, PF11) originated from a previous study (Fobert, 1992), while one was constructed for this study (JC1).

The pJC-1 sequence is compared to pT218-1 in Figure 4.10. The pT218-1 sequence includes the 1.0 kb previously analyzed (Fobert, 1992) and 1.4 kb first sequenced in this study. The pJC-1 is just below 3.0 kb in size and includes 0.8 kb of sequence that is situated downstream of the site of T-DNA integration.

Overall, the two sequences are identical to each other with two exceptions.

(1) The pT218-1 clone contains a small DNA insert at the site of integration (see Figure 4.10.). This is a common feature for most T-DNA integration events and is referred to in the literature as "filler" DNA (Cheysen et al., 1991). The "filler" DNA in pT218-1 is 14 bp long.

(2) The pT218-1 clone also has some base substitutions and additions (between 1535-2523, see Figure 4.10.). Most of these are very close to the site of T-DNA integration. A comparison

FIGURE 4.9. Sequencing strategy: Both the 3.0 kb *Eco* RI pJC-1 clone and the 1.6 kb *Eco* RI-*Xba* I upstream region of pT218-1 were sequenced using the following sequencing strategy. The remaining 1.0 kb of clone pT218-1 was sequenced previously (Fobert, 1992). The arrows indicate the direction of the primers. A primer with an arrow located above a line was used to sequence one DNA strand, whereas a primer with an arrow below a line was used to sequence the opposite strand. The positioning of the primer on the main clone or on a smaller subclone indicates which was used as a template for that primer. Primers# 1: SP6. 2: T7, 3: PF3, 4: PF4, 5: PF5, 6: PF6, 7: PF8, 8: PF9, 9: PF10, 10: PF11, 11: JC1.

between T-DNA vector right border region and sequence in Figure 4.10. indicates that the end point of the T218 T-DNA is 8 bp from the right border.

In transgenic plant line T218, the promoter sequences situated within the 2.4 kb cloned region have been shown to regulate seed coat-specific mode of GUS gene expression (Fobert, 1992). One of the objectives of this study was to determine whether there is an ORF associated with the T218 seed coat-specific regulatory element. However, the sequence analysis results have not permitted the identification of any such ORF. Three putative ORF's have been identified within the 3.0 kb pJC-1 clone; however, all three are very small in size and far from the site of T-DNA integration. The first two are 105 bp and 85 bp long (sites 2825 and 2855, respectively, see Figure 4.10.) and are situated at the downstream end of pJC-1. In the opposite direction, there is a third small ORF, 136 bp long (site 136, see Figure 4.12.), that is situated at the upstream edge of pJC-1. The ATG codons of the first two are situated over 250 bp from the site of integration. All three ATG codons are preceded by TATA-like and CAAT-like elements; however, none of the ATG codons sequences is characterized by the eukaryotic ATG consensus sequence (Joshi, 1987). The best match is 4-5 nucleotides out of 9 (data not shown).

The sequencing data was analyzed in IBI genebank and did not show significant homology with any of the eukaryotic genes currently present in this computer data base.

FIGURE 4.10. Nucleotide sequence of clone pJC-1. Nucleotides are numbered on the left hand side. Nucleotide 1 represents the first nucleotide of the *Eco* RI site. Underlined and bold characters represent restriction sites. 1: *Eco* RI, 2: *Hinc* II, 3: *Hind* III, 4: *Sna* BI, 5: *Sph* I and 6: *Xba* I. The ATG (or CAT) codons of the truncated ORF's are represented in bold and underlined characters. The pJC-1 sequence is compared to the pT218-1 sequence. The top sequence represents pJC-1 and the bottom sequence represents pT218-1. The filler DNA (beginning at 2523) is in bold characters and the T-DNA polylinker is underlined. The pT218-1 *Xba* I- *Sna* BI region that has been shown to be to be responsible for GUS expression in T218 (Unpublished data) corresponds to the pJC-1 genomic region that ranges from bp 1790-1993.

1 **1**
 1 **GAATTC**GGGTCGAAGTCTTCTCCTCGACCTTGTCGGACCTAGGTGAAGCAGAATCAGTCC
3.....
 61 CCACCGGTCCCA**AGCTT**TACAAGTATCAGTACCGTCCCGTGCACGCCACCAACCTAGAG

 121 TTATCTTCTTCTT**CAT**CATCCCTTAGACGTAGGACCGATTCCATAGTCAATGGGATTGTG

 181 GTTCCCTTGGCTTACAGACTGTTCTCTTCTTAGGTTTTTGACCCCTCGGAATCCAAGGTCC

 241 TTTTCTCTTACCTTCCTTCAACGACTTTGAGACAGGCGGTAAA**ACT**TCTTACCACCAG

 301 ACGGGGCCCTCATGACTGCATGTTCCCGAGACCTGCGAAGGAAGAAGATAAGTAA**ACT**C
3.....
 361 ATGCTTCGAATAAACATGTAATCTCGGATGAAAAACGAATCGGCAGACCAAGATGA**AGC**

 421 **TT**ACCATGAGTACGAGCTTCCCCATCGACCCGGCGACAAATCGCCATGAGCGCTCGACAA
5.....
 481 ATGTTGATGT**CTAG**ACCAGGTTTTAGCCAGTTCTTCAAGGTAAGGAACCGGTACCCGG

 541 CATCCAAGCAACGACAGCATCGAGAAAGGCACTAATAAGAAAGGATAAAGAAGTAAGACA
4.....
 601 AACAGAGATTAAACGGGATCGTCATTACGTTTCATATTCATTCTCAGAAAATGGC**ATG**

 661 **CTCTC**AGCCGGGATTAGGTTCGAGGTCTTCACTTCGAACAAACCGGCCCATCCAGCCCCC

 721 GATCCTTATCTTCATCTATGCTCGATGTGA**ACT**GTGCACAGATATTGAAGCCTTATGAGC

 781 CCGCCACGTAATGCAGAGCTATATAATCTTATAAGTGGTCGAGAGTGAATAAGGAAAGTC

 841 CATCGATTTTACTCACAAAGAAACGAATCAAGATCACGATCCTCCAGAAAGAAGGATGAA

 901 TTTGACCGAGGGATCACCTGGTATTTCTTGCAAAAATCGATAATGACTGCATCGAGGGGA

 961 CCCATCGTGAAAGGGTAAGTGTAACACTCAGATACTCTTCCACATGGGTAGTGATCGAT

 1021 TCCTTCGGTGTGCGAGCCTAGAGACAAAAGAGACGTTATAGCCACTATGCATTTTAA**TTA**

 1081 TATCCAAAGTAGGGCTCCTCCATTATATTAGGGTGGCTCTCATT**TGT**AAAAGATTGATT

 1141 GGAATACTCATTATACACTCTCACATTCAGAGATTGAGATTACTATAACAAATATT**GT**

 1201 CATTCTTTGGCTCTTGGCATTGTTTCATATGTTGCTCTAGCAATCATTTTATACTCAATT

 1261 TGGGTTTGTATTCA**TTTCTTTT**TACAGTCAATATTTGATATATTTT**TACTT**ATCCATCCA

 1321 ATTTGTACCAATTTATACCACGTATCCTTAGGATCACGTATAAATTCAACTCTATCCGTT

 1381 TTTCCGGGTAAACAGATGGTGTA**AAAATT**ATTGCAATTATCAATGAATTTTAATATATAAT

 1441 TGTTGAGCAGTTATTATTTTATCATATTCAATGTTTATCAAGAGGCTGGCCGTAATTAT

 1501 CAAATAAATGACCTATACTTAAACCAATAAATAT-GACTTCTCTTTCTTTACATAATCCT

.....C.A.....
 1561 CTTCTTCTTTTTTTTGTAGTTTCTTCTGTTTATCCAAAAACGAATTATTGATTAAGA

 1621 AATACACCAGACAAGTTTTTACTTCTTTTTCTTTTTTTTTTGTGGTAAAAAATTACAC

 1681 CTGGACAAGTTTATCACGAAAATGAAAATTGCTATTTAAGGGATGTAGTTCCGGACTATT
5.....
 1741 TGGAAAGATAAGTGTTAACAAAATAAATAAAAAAGTTTATACAGTTAGATCTCTCTA

 1801 TAACAGTCATCCTTATTTATAACAATACTTTACTATAACCGTCAAATTTATTTTGAAACA
-.....
 1861 AAATTTTCATGTTATGTTACTATAACAGTATTTTATTATAGCAACCAAAAAATATCGAAA

 1921 CAGATACGATTGTTATAGAGCGATTTGATTGTTATCA-TATCCACATATTTTCGTAAGCCC
-.....T.....
 1981 AATTACTCC----TCCTACGATGAAAGTAAACCAATTTAAAGTTGCAAAAATCCAATAGATT
TACG.....
 2039 TCAATACTTCTTCAACTGGCGTTATGTTAGGTAATGACTCCTTTTAACTTTTCATCTTTAA
5.....
 2101 TTTGAAGTTTCTTTCATTAAAGAAAGTTTCTAGAAGAGAAGTGTTTAACTTCTAGC
5.....
 2161 TCTACTATTATCTGTGTTTCTAGAAG-AAAATAGAAAATA'GTGTCCACCTCAAAAACAAC
A.....
 2221 TAAAGGTGGGCAAATCTCCACCTATTTATTTTATTTTGGATTAATTAAGATATAGTAAAG

 2281 ATCAGTTATAAACGGAGTTTT-GAGTTGATACAGTGAATTTTAAGATGTGTACCGATTAA
 ...G.....G.....
 2341 CTTTATTTACATTTATGTTTCGCATATAAGAAGTCCGATTGGAATACTAGATTTTGTC
-----.....
 2401 AATCAGGCAATTCATGTGGTTGAAGAATTTAAGTTATATACAATGATGATATAAGAATT

 2461 TTTATACTATTAGTGCAAATTAATCGATTACTAAAAATTATTATTCTATTAATTTATGCT

 2521 ATCATCAAATTTAGTATTCGGTTTATCCTTATTTTATGTGGCACATGTTTATTTAGATT
 ...GTGCCTCCCCAACCCGTCGACCGGTACCGCGGTACCGGGTGGTTCAGTCCCTTATG
 2581 ATCCTAAAAACAATTATATTTTTTTTATATTATAAAATAATTTAACTTTAATCTTTTTT

 2641 TAATTTTTTCTTACTAGATTAATTTATTGTTATTCAAGTGTATTGTTTGTTCATAACAC

 2701 GAATTTCAAAAATCTTTATTTCTTAAATTCTATATCAGTGAAGTAGTGTCACATAAACAG

 2761 AGAAGGAAAAAAGTATGAAATTACCTATTATTATAGATCCAATATTGAAAGCAGAGGGTG

 2821 CCCTACATTACGTTAATTATTTGAGGGGGAAATATGGTACTAGTTGTTTGGCCAGTTCTA

 2881 TGTTGCAAACCAACATAGAAGTTGGGATTCAATTCAGTACACGACATAATCAGAGGAT

 2941 TGATCGACT

FIGURE 4.11. A schematic diagram indicating the differences between the pJC-1 and pT218-1 sequences.

pJC-1

4I 1S
1I

6D

1 kb

5. DISCUSSION AND CONCLUSION

5.1. Background and Objectives of Present Study

A previous study (Fobert, 1992) using T-DNA tagging identified the transgenic plant line T218 with a seed coat-specific mode of GUS expression. A 2.6 kb genomic region that was situated directly upstream of the T-DNA tag was subcloned and reintroduced in *N. tabacum* SR1 by T-DNA-mediated transformation. The results indicated (Fobert, 1992) that this 2.6 kb clone contained all the sequences responsible for the spatial and temporal seed coat-specific pattern of GUS expression observed in plant T218. Other analyses of this genomic region have since localized the cis-regulatory sequences to the pT218-1 region contained between the *Xba* I and *Sna* BI restriction sites (Unpublished data).

This thesis had three objectives that have addressed issues concerning the T218 seed coat-specific genomic region isolated in the previous study (Fobert, 1992). The first was to isolate the wild-type *N. tabacum* genomic region that corresponds to the seed coat-specific region identified in plant line T218 (Fobert, 1992) and to characterize this region by sequence analyses. The sequencing data was used for the second and principal objective of this thesis which was to determine whether the T-DNA insertion caused any rearrangements in the target sequences.

This issue of rearrangements in the target sequence is interesting for two reasons. First, the presence of rearrangements

may suggest whether the T-DNA integration itself had an effect on the expression pattern of this region. If this were the case, then the seed coat-specific pattern of expression would be an artifact, the result of rearrangements in the regulatory element. Second, the type of rearrangements present in the target sequence, together with a comparison of the vector and target sequences, may also help to understand which recombination mechanism may have mediated the integration of the T-DNA into the plant chromosome.

The third objective of this study was to use the sequencing data to determine whether there is a gene associated with the T218 seed coat-specific *cis*-regulatory element. If there was such a gene, then it would be the first seed coat-specific gene identified to date. On the other hand, if such a gene didn't exist, then the T218 genomic region would be an example of a *cis*-regulatory element that is not in the vicinity of the gene it may regulate or a remote element.

5.2. Rearrangements in target sequence

The type and extent of rearrangements in the target sequence, as well as the existence of homology between the T-DNA vector and the target sequence give an indication of the recombination mechanism by which the T-DNA integration event occurred. Minor deletions and insertions, as well as significant homology between the vector and target site are indications of homologous recombination. One study (Matsumoto et al., 1990) of T-DNA

integration events in an established tobacco cell line from crown gall tumours reported strong evidence that homologous recombination had mediated these T-DNA insertions. The homology between the T-DNA and the plant target site was reported to be as high as 70%.

Two comprehensive studies of T-DNA integration events in *N. tabacum* (Gheysen et al., 1991) and *Arabidopsis thaliana* (Mayerhofer et al., 1991) reported finding no significant overall homology between the T-DNA and the target sequence for 15 different T-DNA insertions. However, they did observe very short stretches of homology between the T-DNA ends and the target sites and the presence of short inserts at the T-DNA junctions, referred to as filler DNA. On the basis of these observations, illegitimate recombination was proposed to mediate T-DNA integration (Gheysen et al., 1991; Mayerhofer et al., 1991). These studies also reported that the integration usually causes target site deletions: the range observed was between 13-28 bp (Gheysen et al., 1991) and 29-73 bp (Mayerhofer et al., 1991). Other rearrangements that were also observed in a smaller number of target sites are duplications, inversions, as well as larger target rearrangements (usually between 1-3 kb but sometimes over 3 kb).

A comparison between the *N. tabacum* target sequence before (pJC-1) and after (pT218-1) integration indicates that the T218 T-DNA integration event did not result in any gross rearrangements (see Figure 4.10.). The rearrangements that were observed fall into two categories. First, the pT218-1 clone contains a short insert at the site of integration, filler DNA (see Figure 4.10.).

2 OF 2

PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

The filler DNA in pT218-1 is 14 bp long and is homologous to two other regions in pT218-1. The first is situated within the GUS gene and the other is in a region further upstream from the site of integration. Second, close to the site of T-DNA integration, the pT218-1 clone has a few base substitutions, insertions and deletions. Since no gross rearrangements were not observed, the conclusion drawn was that the seed coat-specific pattern of expression is not a result of gross rearrangements created in transgenic plant T218 by the T-DNA integration. Modification of the potential *cis*-elements by the less extensive alterations cannot be ruled out. In the region responsible for seed coat-specific GUS expression (between bp 1790-1892), two bp deletions and the insertion of one bp and four bp have been identified.

The small number of target rearrangements in pT218-1 may be a result of the particular T-DNA vector used. With the T-DNA vector used to transform plant T218, there was generally one T-DNA insert per genome (Fobert, 1992), whereas other T-DNA vectors usually insert in multiple copies (Jorgensen et al., 1987). In the transgenic plant T218 there was only one T-DNA copy (Fobert, 1992). Compared to other examples of T-DNA integration events (Gheysen et al., 1991; Mayerhofer et al., 1991) few rearrangements were observed at the right border junctions. The rearrangements that were observed were the filler DNA (14 bp), base substitutions (2 bp), deletions (10 bp) and insertions (3 bp) between 1771 bp and the site of integration at 2636 bp.

Alternatively, the small number of rearrangements observed may

be due to the fact that the left border junction was not examined. Since the endpoints of the left T-DNA border are much more irregular and more rearrangements have been reported in the left border T-DNA junctions (Walden et al., 1991), this study may not have been able to observe rearrangements that did occur upon integration but at the left border junction.

5.3. T-DNA Recombination mechanism

The sequence data from the tobacco target sequence (pJC-1) was compared to the T-DNA vector sequence. This comparison revealed that the T-DNA vector and the target sequence in the tobacco genome have no significant homology. This indicates T-DNA insertion by illegitimate recombination. The conclusion is reinforced by the presence of filler DNA, a common feature in most studied examples of illegitimate recombination (Gheysen et al, 1991; Mayerhofer et al, 1991).

5.4. Is there an ORF associated with this genomic region?

Three putative ORFs were identified within the 3.0 kb pJC-1 clone. However, all three ORFs are far from the site of T-DNA integration. The first two are 105 bp and 85 bp long and are located at the downstream end of pJC-1. Their ATG codons are situated over 300 bp from the site of integration. In the opposite direction, there is a third ORF, 136 bp long (site 136, see Figure

4.10.), at the upstream edge of pJC-1. All three ATG codons are preceded by TATA-like and CAAT-like elements; however, none of the sequences are characterized by the eukaryotic ATG consensus sequence (Joshi, 1987). The best match is 4-5 nucleotides out of 9 (data not shown).

All three of the putative ORFs are incompletely studied. The sequencing data presented in this thesis is insufficient to conclude whether there is an ORF associated with this cis-regulatory seed coat-specific region.

5.5. Concluding Remarks

There are four possible explanations for the seed coat-specific expression pattern of the T218 gene fusion. The first is that there is a seed coat-specific gene associated with the genomic region into which the T-DNA has inserted and the expression of the GUS reporter gene in plant line T218 mirrors the expression of this endogenous tobacco gene. The ATG start codon of this gene may be one of the three referred to in this study or a different one, even further away from the site of integration. If this were true, then more sequencing data would be required to confirm it.

The second possibility is that there is no gene associated with this genomic region. This implies that this genomic region has promoter-like features and can confer seed coat-specific expression on neighbouring sequences. If this were true, the *N. tabacum* seed coat-specific genomic region would be the first

example of a tissue-specific plant regulatory element not associated with a gene.

If there is an ORF associated with this genomic region, another question is the *in vivo* expression patterns of this ORF. The T-DNA may have inserted into a region that is not itself seed coat-specific. Some form of rearrangement of the T-DNA within the promoter results in an altered (i.e. seed coat-specific) expression of the reporter gene (GUS) in plant T218. However, few rearrangements were detected at the right border junction, so this possibility is unlikely. Another possibility is that the expression of the T-DNA is a result of a positioning effect due to the position in the plant promoter into which it has inserted. One recent study (Benfe^v and Chua, 1990), reported that the constitutive CaMV promoter is actually divided into various subdomains. Specific subdomains confer tissue-specific expression to reporter genes, whereas as a whole they confer constitutive expression. A similar question could be asked about the T218 seed coat-specific region. Are the expression patterns of the region that is associated with seed coat-specificity different from those observed for the GUS gene in plant line T218? Given the distance of any ORF from the seed coat-specific sequences, such a position effect (i.e. different expression pattern for the reporter gene due to the positioning of the T-DNA within the regulatory sequence) is not unlikely.

Related to the apparent absence of any ORF is which endogenous sequences acted as a promoter-proper (i.e. to initiate the

transcription of the GUS gene fusions) in plant line T218. The explanation may lie in the fact that this genomic region is relatively AT-rich (70%) and, as result, one of the many TATA-like and CAAT-like sequences may be exhibiting promoter-proper behaviour. On the other hand, GC-rich motifs rather than TATA box sequences have been detected in the promoter regions of several higher eukaryote genes (Sehgal et al., 1988). Since the filler DNA is very GC-rich, then it may be possible that the filler DNA is initiating transcription for the T218 GUS gene fusion.

To determine which scenario best explains the seed coat-specific expression pattern, further data is required. Mapping of the 5' termini of the GUS mRNA in T218 is needed to determine which sequences are acting as the proximal promoter. This study presents sequencing data from a region that is up to 400 bp downstream from the site of T-DNA integration. One more possibility for future work would be to use DNA from the wild-type region where the T-DNA inserted to as a probe for sensitive RNA assays such as poly A+ or PCR Northern, RNase protection assays. This would tell us whether this region is expressed. If further sequencing analyses uncovers an ORF downstream of the *EcoR* I site (see Figure 4.2.), then to determine whether this endogenous gene also has a similar seed coat-specific expression pattern, we would have to create a pJC-1/GUS gene fusion and functionally test it in transgenic plants.

REFERENCES

- Albright, L.M., Yanofsky, M.F., Leroux, B., Ma, D., and Nester E.W. (1987) Processing of the T-DNA of *Agrobacterium tumefaciens* generates border nicks and linear, single-stranded T-DNA. *J. Bacteriol.* **169**: 1046-1055.
- Ambros, P., Matzke, A.J.M., and Matzke. M.A. (1986) Localization of the *Agrobacterium rhizogenes* T-DNA in plant chromosomes by in situ hybridization. *EMBO J.* **5**: 2073-2077.
- André, D., Colau, D., Schell, J., Van Montagu, M., and Hernalsteens, J.P. (1986) Gene tagging in plants by a T-DNA insertion that generates APH(3')II plant gene fusions. *Mol. Gen. Genet.* **204**: 512-518.
- Balcells, L., Swinburne, J., and Coupland, G. (1991) Transposons as tools for the isolation of plant genes. *TIBTECH* **9**: 31-37.
- Benfey, P.N., and Chua, N.H. (1990) The cauliflower mosaic virus 35S promoter: Combination regulation of transcription in plants. *Science* **250**: 959-966.
- Bernatzky, R., and Tanksley, S.D. (1986) Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol. Gen. Genet.* **203**: 8-14.
- Bingham, P., Lewis, R., and Rubin, G.M. (1981) Cloning of DNA sequences from the white locus of *Drosophila melanogaster* by a novel and general method. *Cell* **20**: 579-595.
- Boulton, G.W., Nester, E.W, and Gordon, M.P. (1986) Plant phenolic compounds induce expression of *Agrobacterium tumefaciens* loci needed for virulence. *Science* **232**: 983-984.
- Brownes, M. (1990) Preferential insertion of P elements into genes expressed in the germ-line of *Drosophila melanogaster*. *Mol. Gen. Genet.* **222**: 457-460.
- Callis, J., Fromm, M., and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. *Genes Dev.* **4**: 1183-1200.
- Chang, C., and Meyerowitz, E.M. (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* **83**: 1408-1412.
- Chaubert, N., Philipps, G., Chaboute, M.-E., Ehling, M, and Gigot, C. (1986) Nucleotide sequences of two corn histone H3 genes. Genomic organization of the corn histone H3 and H4 genes. *Plant*

Mol. Biol. 6: 253-263.

Christie, P.J., Ward, J.E., Winans, S.C., and Nester E.W. (1988) the *Agrobacterium tumefaciens* virE2 gene product is a single-stranded DNA-binding protein that associates with T-DNA. J. Bacteriol. 170: 2659-2667.

Chyi, Y.-S., Jorgensen, R.A., Goldstein, D., Tanksley, S.D., and Loaiza-Figueroa, F. (1986) Location and stability of *Agrobacterium*-mediated insertions in the *Lycopersicon* genome. Mol. Gen. Genet. 204: 64-68.

Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA 69: 2110-2114.

Coruzzi, Gloria, Broglie, R., Edwards, C., and Chua, N-H. (1984) Tissue-specific and light-regulated expression of pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO J., 3: 1671-1679.

Das, A. (1988) *Agrobacterium tumefaciens* virE operon encodes a single-stranded DNA-binding protein. Proc. Natl. Acad. Sci. USA 85: 2909-2913.

De Block, M., Herrera-Estrella, L., Van Montagu, M., Schell, J., and Zambryski, P. (1984) Expression of foreign genes in regenerated plants and their progeny. EMBO J. 3: 1681-1689.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P., and Bedrock, J. (1985) Linkage and homology analysis divides the eight genes for the small subunit of petunia into three gene families. Proc. Natl. Acad. Sci. USA 82: 4964-4968.

Derolles, S.C., and Gardner, R.C. (1988) Analysis of the T-DNA structure in a large number of transgenic petunias generated by *Agrobacterium*-mediated transformation. Plant Mol. Biol. 11: 365-377.

Dietrich, R.A., Radke, S.E., and Harada, J.J. (1992) Downstream DNA sequences are required to activate a gene expressed in the root cortex of embryos and seedlings. The Plant Cell, 4: 1371-1382.

Dürrenberger, F., Cramer, A., Hohn, B., and Koukoloková-Nikola, Z. (1989) Covalently bound VirD2 protein of *Arabidopsis thaliana* protects the T-DNA from

Dynan, W.S. (1989) Modularity in promoters and enhancers. Cell 58: 1-4.

Elliston, K., and Messing, J. (1988) The molecular architecture of

plant genes: A phylogenetic perspective. In: *Architecture of Eukaryotic Genes*, G. Kahl (ed), pp. 21-56, VCH Publisher, NY

Errampalli, D., Patton, D., Castle, L., Mickleson, L., Kansen, K., Schell, J., Feldmann, K., and Marks, D. (1991) Embryonic lethals and T-DNA insertion mutagenesis in *Arabidopsis*. *Plant Cell* 3: 149-157.

Feinberg, A.P., and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* 132: 6-13.

Feldmann K.A., Marks, M.D., Christianson, M.L., and Quatrano, R.S. (1989) A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science* 243: 1351-1354.

Feldmann, K.A. (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* 1: 71-82.

Fluhr, R., Kuhlmeier, C, Nagy, F., and Chua, N.-H. (1986) Oragn-specific and light-induced expression of plant genes. *Science* 232: 1106-1111.

Fobert, P.R. (1992) PhD Thesis, Carleton University, Ottawa, Ontario, Canada.

Fobert, P.R., Miki, B.L., and Iyer, V.N. (1991) Detection of gene regulatory signals in plants revealed by T-DNA-mediated fusions. *Plant Molecular Biology* 17: 837-851.

Gallie, D.R., Gordon, K., Sanfacon, H., Bonneville, J-M, and Hohn, T. (1991) Post-transcriptional regulation in higher eukaryotes: The role of the reporter gene in controlling expression. *Mol. Gen. Genet.* 228: 258-264.

Gasser, C.S., and Fraley, R.T. (1989) Genetically engineering plants for crop improvement. *Science* 244: 1293-1299.

Gebhardt, C., Oliver, J.E., Forde, B.G., and Miflin, B.J. (1986) Primary structure and differntial expression of glutamine synthetase genes in nodules, roots and leaves of *Phaseolus vulgaris*. *EMBO J.* 5: 1429-1435.

Gheysen, G., Van Montagu, M., and Zambryski, P. (1987) Integration of *Agrobacterium tumefaciens* T-DNA involves rearrangements of target plant DNA sequences. *Proc. Natl. Acad. Sci. USA* 84: 6169-6173.

Gheysen, G., Herman, L., Breyen, P., Gielen, J., Van Montagu, M., and Depicker, A. (1990) Cloning and sequence analysis of truncated T-DNA inserts from *Nicotiana tabacum*. *Gene* 94: 155-163.

Gheysen, G., Villaroel, R., and Van Montagu, M. (1991) Illegitimate recombination in plants: a model for T-DNA insertion. *Genes Develop.* 8: 87-97.

Gierl, A., and Saedler, H. (1992) Plant transposable elements and gene tagging. *Plant Mol. Biol.* 19: 39-49.

Gilmartin, P.M., Sarokin, L., Memelink, and Chua, N-H (1990) Molecular light switches for plant genes. *The Plant Cell* 2: 369-378.

Goldberg, R.G., Barker, S.J., and Perez-Grau, L. (1989) Regulation of Gene Expression during Plant Embryogenesis *Cell* 56: 149-160.

Greener, A., Filutowicz, M.S., Meachern, M.J., and Helinski, D.R. (1990). N-Terminal truncated forms of the bifunctional P_i-initiation protein express negative activity on plasmid R6K replication. *Mol. Gen. Genet.* 224: 24-32.

Gridley, T., Soriano, P., and Jaenisch, R. (1987) Insertional mutagenesis in mice. *Trends Genet.* 3: 162-166.

Herman, P., and Marks, M.D. (1989) Trichome development in *Arabidopsis thaliana* II. Isolation and complementation of the GLABROUS I gene. *Plant Cell* 1: 1051-1055.

Herman, P., Jacobs, A., Van Montagu, M., and Depicker, A. (1990) Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. *Mol. Gen. Genet.* 224: 248-256.

Herrera-Estrella, A., Van Montagu, M., and Wang, K. (1990) A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs a beta-galactosidase fusion protein into tobacco nuclei. *Proc. Natl. Acad. Sci. USA* 87: 9534-9537.

Holmes, D.S., and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114: 193-197.

Holsters, M., Villaroel, R., Gielen, J., Seurinck, J., DeGreve, H., Van Montagu, M., and Schell, J. (1983) An analysis of the boundaries of the octopine TL-DNA in tumours induced by *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* 190: 35-41.

Jaenisch, R. (1988) Transgenic animals. *Science* 240: 1468-1474.

Johnson, P.F., and McKnight, S.L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* 58: 799-839.

Jorgensen, R., Snyder, C., and Jones, J.D.G. (1987) T-DNA is organized predominantly in repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. *Mol. Gen. Genet.* 297: 471-477.

- Joshi, C.P. (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucl. Acids Res.* 15: 6643-6653.
- Kado, C.I. (1991) Molecular mechanisms of crown gall tumorigenesis. *Crit. Rev. Plant Sci.* 10: 1-32.
- Katagiri, F., and Chua, N-H (1992) Plant transcription factors: present knowledge and future challenges. *Trends in Genetics* 8: 1, 22-27.
- Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M., and Hernalsteens, J.-P. (1991) In vivo random β -glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 88: 5212-5216.
- Klee, H., Horsch, R., and Rogers, S. (1987) *Agrobacterium*-mediated plant transformation and its further applications to plant biology. *Ann. Rev. Plant Biol.* 38: 467-486.
- Koncz, C., Martini, N., Mayerhofer, R. Koncz-Kalman, Zs., Körber, H. Redei, G.P., and Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl. Acad. Sci. (USA)* 86: 8467-8471.
- Koncz, C., Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Reiss, B., Redei, G.P., and Schell, J. (1990) Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. *EMBO J.* 9: 1337-1346.
- Koukoliková-Nikola, Z., Albright, L., and Hohn, B. (1987) The mechanism of T-DNA transfer from the *Agrobacterium tumefaciens* to plant cell. In Hohn, T, Schell, J (eds) *Plant Gene Research*, Vol. 4, Springer, Wien, pp. 109-148.
- Kwok, W.W., Nester, E.W., and Gordon, M.P. (1985) Unusual plasmid organization in an octopine crown gall tumour. *Nucleic Acids Res* 13: 459-471.
- Luehrsen, K.R., and Walbot, V. (1991) Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol. Gen. Genet.* 225: 81-93.
- Luo, D., Coen, F.S., Doyle, S., and Carpenter, R. (1991) Pigmentation mutants produced by the transposon mutagenesis in *Antirrhinum majus*. *Plant J.* 1: 59-69.
- Ma, H., Yanofsky, M.F., and Meyerowitch, E.M. (1991) AGL1-AGL6 an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Develop.* 5: 484-495.
- Mandel, M., and Higa, A. (1970) Calcium-dependant bacteriophage DNA infection. *J. Mol. Biol.* 53: 159-162.

- Marcotte, W.R., Russell, S., and Quatrano, R.S. (1989) Abscissic Acid-Responsive Sequences from the Em Gene of Wheat. *Plant Cell* 1: 969-976.
- Marks, M.D., and Feldmann, K.A. (1989) Trichome development in *Arabidopsis thaliana* I. T-DNA tagging of the GLABROUS I gene. *Plant Cell* 1: 1043-1050.
- Martienssen, R.A., Barkan, A., Freeling, M., and Taylor, W.C. (1989) Molecular cloning of a maize gene involved in photosynthetic membrane organization that is regulated by Robertson's Mutator. *EMBO J.* 8: 1633-1639.
- Matsumoto, S., Yukihiro, I., Tsuyoshi, H., Takahashi, Y. and Machida, Y. (1990) Integration of *Agrobacterium* T-DNA into a tobacco chromosome: Possible involvement of DNA homology between T-DNA and plant DNA. *Mol. Gen. Genet.* 224: 309-316.
- Matzke, M.A., Primig, M., Trnovsky, J., and Matzke, A.J.M. (1989) Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *Dev. Genet.* 11: 214-223.
- Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Cramen, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B., and Koncz, C. (1991) T-DNA integration: A mode of illegitimate recombination in plants. *EMBO J.* 10: 697-704.
- McLaughlin, M., and Walbot, V. (1987) Cloning of a mutable bz2 allele of maize by transposon tagging and differential hybridization. *Genetics* 117: 771-776.
- Meyerowitz, E.M. (1987) *Arabidopsis thaliana*, *Annu. Rev. Genet.* 21: 93-111.
- Mooslehner, K., Karls, U., and Harbers, K. (1990) Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J. Virol.* 64: 3056-3058.
- Nester, E.W., Gordon, M.P., Amasino, R.M. and Yanofsky, M.F. (1984) Crown gall: a molecular and physiological analysis. *Annu. Rev. Plant Physiol.* 35: 387-413.
- Nevers, P., Shepherd, N.A., and Saedler, H. (1986) Plant transposable elements. *Adv. Bot. Res.* 12: 102-203.
- O'Farrell, P.H., Kutter, E., and Nakanishi, M. (1980) A restriction map of the bacteriophage T4 genome. *Mol. Gen. Genet.* 179: 421-435.
- Okamuro, J.K., and Goldberg, R.B. (1986) Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. *Mol. Gen. Genet.* 198: 290-298.

Okamuro, J.K., and Goldberg, R.B. (1989) Regulation of Plant Gene Expression: General Principals. In "The Biochemistry of Plants", A. Marcus, ed. (San Diego: Academic Press), Vol. 15, pp. 1-82.

Ooms, G., Bakker, A., Molenddijk, L., Wullems, G.J., Gordon, M.P., Nester E.W., and Schilperoort, R.A. (1982) T-DNA oraganization in homogenous octopine-type crown gall tissues of *Nicotiana tabacum*. Cell 30: 589-597.

Patient, R.K., and Allen, J. (1989) Active chromatin. Curr. Opin. Cell Biol. 1: 454-459.

Paszkowski, J. (1989) Fate of Foreign DNA introduced to Plant Cells. In: Genetic Engineering: Principles and Methods, Vol. 11, J.K. Setlow (ed), pp.145-157.

Paz-Arez, J., Wienand, U., Peterson, P.A., and Saedler H. (1986) Molecular cloning of the *c* locus of *Zea mays*: a locus regulating the anthocyanin pathway. EMBO J. 5: 829-833.

Peralta, E.G., Helmiss, R., and Ream, W. (1986) Overdrive, a T-DNA transmission enhancer on the *Agrobacterium tumefaciens* tumour-inducing plasmid. EMBO J. 5: 1137-1141.

Promega Catalogue, Fisher Scientific, Madison, WI, USA (1991).

Raleigh, E.A., Hanahan, D., Rigby, P.W.J., Reith, A.D., Revel, H, Murray, N.E., Blumenthal, R.M., Elhai, J., and Westaway, D. (1988) Mcra and Mcrb restriction phenotypes of some *Escherichia coli* strains and implications for gene cloning. Nucl. Acids. Res. 16: 1563-1575.

Ream, W. (1990) *Agrobacterium tumefaciens* and interkingdom genetic exchange. Annu. Rev. Phytopathol. 27: 583-618.

Roeder, G.S., Farabough, P.J., Chaleff, D.F., and Fink, G.R. (1980) The origins of gene instabilty in yeast. Science 209: 1375-1380.

Roth, D.B., Chang, X-B, and Wilson, J.H. (1989) Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol. and Cellular Biology 9: 3049-3957.

Rutledge, R.G., Ouellet, T., Hattori, J., and Miki, B.L. (1991) Molecular characterization and genetic origin of the *Brassica napus* acetohydroxyacid synthase multigene family. Mol. Gen. Genet. 229: 31-40.

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

- Sanders, P.R., Winter, J.A., Barnason, A.R., Rogers, S.G. and Fraley, R.T (1987) Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucl. Acid Res.* 15: 1543-1558.
- Sandmeyer, S.B., Hansen, L.J., and Chalker, D.L. (1990) Integration specificity of retrotransposons and retroviruses. *Annu. Rev. Genet.* 24: 491-518.
- Sanger, F., Nicklin, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
- Schmid, C.W., and Deininger, P.L. (1975) Sequence organization of the human genome. *Cell* 6: 345-358.
- Sehgal, A., Patil, N., and Chao, M. (1991) A constitutive directs expression of the nerve growth factor receptor gene. *Mol. Cell Biol.* 8: 3160-3167.
- Shaw, C.H., Watson, M.D., Carter, G.H., and Shaw, C.H. (1984) The right hand copy of the nopaline Ti plasmid 25 bp repeat is required for tumour transformation. *Nucl. Acids Res.* 12: 6031-6041.
- Sheen, J. (1991) Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. *Plant Cell* 3: 225-245.
- Simpson, R.B., O'Hara, P.J., Kwok, W. W., Montoya, A.L. Lichtenstein, C., Gordon M.P. and Nester E.W. (1982) DNA from the A6/2 crown gall tumour contains scrambled Ti plasmid sequences near its junctions with the plant DNA. *Cell* 29: 1005-1014.
- Sommer H., Beltran, J-P, Huijser, P., Pape, H., Lönning, W-E, Saedler, H., and Schwarz-Sommer, Zs. (1990) Deficiens, a homeotic gene invoved in the control of flower morpphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* 9: 605-613.
- Stachel, S.E., Messens, E., Van Montagu, M., and Zambryski, P. (1985) Identification of the plant molecule produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. *Nature* 318: 624-629.
- Stachel, S.E., Nester, E.W., and Zambryski, P.C. (1986) A plant cell factor induces *Agrobacterium tumefaciens* vir gene expression. *Proc. Natl Acad. Sci. USA* 83: 379-383.
- Tingey, S.V., Walker, E.L., and Coruzzi, G. (1987) Glutamine synthase genes of pea encode distinct polypeptides which are differently expressed in leaves, roots and nodules. *EMBO J.* 6: 1-9.

- Theres, K., Scheele, T., and Starlinger, P. (1987) Cloning of the Bz2 locus of *Zea mays* using the transposable element Ds as a gene tag. *Mol. Gen. Genet.* 209: 193-197.
- Thompson, W.F. and White, M.J. (1991) Physiological and molecular studies of light-regulated nuclear genes in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42: 423-466.
- Topping, J., Wei, W., and Lindsey, K. (1991) Functional tagging of regulatory elements in the plant genome. *Development* 112: 1009-1119.
- Van Haaren, M.J.J., Sedee, N.J.A., Krul, R.A., Schilperoort, R.A., and Hooykaas, P.J.J. (1988) Function of heterologous and pseudo border repeats in the T region transfer via the octopine virulence system of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* 11: 773-781.
- van Holde, K.E. (1989) *Chromatin*. Springer-Verlag, New York.
- Veluthambi, K., Jayaswal, R.K., and Gelvin S.B. (1988) Virulence genes A, G and D mediate the double-stranded border cleavage of T-DNA from the *Agrobacterium Ti* plasmid. *Proc. Natl. Acad. Sci. USA* 84: 1881-1885.
- Walden, R., Hayashi H., and Schell J. (1991) T-DNA as a gene tag. *The Plant Journal* 1: 281-288.
- Jallroth, M., Gerats, A.G.M., Rogers, S.G., Fraley, R.T., and Horsch, R.B. (1986) Chromosomal localization of foreign genes in *Petunia hybrida*. *Mol. Gen. Genet.* 202: 6-15.
- Wang, K., Herrera-Estrella, L., van Montagu, M., and Zambryski, P. (1984) Right 25 bp terminus sequences of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. *Cell* 38: 455-462.
- Wang, K., Stachel, S.E., Timmerman, B., van Montagu, M., and Zambryski, P. (1987) Site-specific nick in the T-DNA border repeat element determines its relative activity during T-DNA transfer to plant cells. *Mol. Gen. Genet.* 210: 338-346.
- Wienand, U., Weydemann, U., Niesbach-Lösgen, U., Ptersson, P.A., and Saedler, H. (1986) Molecular cloning of the c2 locus of *Zea mays*, the gene coding for chalcone synthase. *Mol. Gen. Genet.* 203: 202-207.
- Winans, S.C., Kerstetter, R.A., and Nester E.W. (1988) Transcriptional regulation of the *virA* and *virG* genes of *Agrobacterium tumefaciens*. *J. Bacteriol.* 170: 4047-4054.
- White, R.J. and Jackson, S.P. (1992) The TATA-binding protein: a central role in transcription by RNA polymerases I, II, III.

Trends in Genetics 8: 8. 284-288.

Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M., and Chilton, M.-D. (1982) Short direct repeats flank the T-DNA on a nopaline Ti plasmid. Proc. Natl. Acad. Sci. USA 79: 6322-6326.

Yamamoto, A., Iwahishi, M., Yanofsky, M.F., Nester, E.W., Takebe, I., and Machida, Y. (1987) The promoter proximal region in the virD locus of *Agrobacterium tumefaciens* is necessary for the plant-inducible circularization of T-DNA. Mol. Gen. Genet. 206: 174-177.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990) The protein coded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35-39.

Yanofsky, M.F., Poter, S.G., Young, C., Albright, L.M., Gordon, M.P., and Nester, E.W. (1986) The virD operon of *Agrobacterium tumefaciens* encodes a site specific endonuclease. Cell 47: 471-477.

Young, C., and Nester, E.W. (1988) Association of the VirD2 protein with the 5' end of T-strands in *Agrobacterium tumefaciens*. J. Bacteriol. 170: 3367-3374.

Zambryski, P., Depicker, A., Kruger, K., and Goodman, H.M. (1982) Tumour induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. J. Mol. Appl. Genet. 1: 361-370.

Zambryski, P.C. (1992) Chronicles from the *Agrobacterium*-plant cell DNA transfer story. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43: 465-490.

Zambryski, P.C., Tempe, J., and Schell, J. (1989) Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. Cell 56: 193-201.

END

1 6 - 1 2 - 9 3

FIN