The molecular basis of gene expression variability in transgenic tobacco plants

Laverty, Edward

The molecular basis of gene expression variability in transgenic tobacco plants.

Thesis submitted for the degree of Ph.D.

at

The University of Durham

by

Edward Laverty B.Sc. (Stirling)

May 1996

The molecular basis of gene expression variability in transgenic tobacco plants.

Abstract of a thesis submitted for the degree of Ph.D. at the Department of Biological Sciences, University of Durham by Edward J. Laverty B.Sc.

An extensive investigation into and characterisation of factors influencing transgene expression following introduction of the transgene into tobacco via *Agrobacterium*-mediated transformation was carried out. Characterisation of material supplied at the outset of this project revealed that this material was unacceptable for further analysis. It was thus deemed necessary to obtain large populations of transgenic tobacco heterogenous for levels of transgene expression. Characterisation of these populations (*CaMV-lecA* and *ssRubisco-lecA* plants) showed that all plants fell into one of four segregation classes based on segregation of the kanamycin-resistance selectable marker. Results showed that the majority of regenerants contained multiple *nptII*-containing inserts, while the presence of one or two such inserts was also found, albeit at a much lower frequency. Segregation analysis based on detection of the *lecA* transgene agreed, in the majority of cases, with these results. However, in a few cases it was found that data obtained from both segregation analyses did not agree, with the presence of a single *lecA*-containing transgene being detected in plants shown to contain two copies of the *nptII*-containing transgene. This result indicates the occurrence of T-DNA rearrangement either within the tobacco genome or during T-DNA transfer and integration. Southern blot analyses allowed a detailed characterisation of T-DNA structure, copy number and number of integration sites to be undertaken. Results from these analyses revealed a higher frequency of T-DNA rearrangement within plants containing multiple inserts. However, such rearrangements did not correlate with a significant reduction in levels of transgene expression since all detected rearrangements were found to occur at or towards the left hand border of the T-DNA, that border distant to the *lecA* transgene. Plants containing more than one T-DNA were also frequently found to contain these T-DNAs arranged as an inverted repeat at a single locus although no significant relationship between copy number and the presence of such structures was found. Correlating transgene expression levels, as determined by radioimmunoassay-based quantitation of lectin protein in tissues of transgenic plants, with T-DNA copy number, organisation and structure revealed no significant relationship. It is thus feasible to conclude that the major contributory factor influencing levels of transgene expression is the location of T-DNA integration within the plant genome. Subsequent work concerned with investigating the nature of those integration site-specific factors i.e. 'position effect' indicated a possible role for methylation-induced modulation of gene expression. Results presented in this thesis provide an insight into the fate of transgenes following introduction into the plant genome and clearly demonstrate the importance of further exploring the molecular mechanisms underlying transgene expression variability.

Acknowledgements

I wish to express my gratitude to the following for their help, support and interesting scientific discussion: Drs. R.R.D. Croy (my supervisor), J. Gatehouse, G. Edwards, C. Scutt, A. Skelton, A. Shirsat and B. Stanchev (University of Durham) and A. Gill (University of Glasgow).

This work was funded by a research studentship from the A.F.R.C.

Abbreviations

(w/v)	weight:volume ratio
(v/v)	volume:volume ratio
^{o}C	degrees Celsius
Abs	absorbance
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
CoA	coenzyme A
DABA	diaminobenzoic acid
DNA	deoxyribonucleic acid
ds	double stranded
EDTA	ethylenediaminetetraacetic acid (disodium salt)
g	gramme
g	relative centrifugal force
GUS	ß-glucuronidase, encoded by the *uidA* gene from *E. coli* (Jefferson *et al.*, 1987)
gus	*uidA* coding sequence
h	hour
Kb	kilobase pair
kDa	kilodalton
l	litre
p.s.i.	pounds per square inch
M	molar
mM	millimolar
min.	minute
MW	molecular weight

mRNA	messenger RNA
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
SDS	sodium dodecyl sulphate
vol.	volume

Table of Contents.

Chapter 2: Materials and Methods.

Chapter 3: Results and Discussion.

List of Figures.

List of Tables.

Chapter 1:

Introduction.

In higher plants recombinant DNA technology is being viewed with increasing importance for its potential in breeding allowing the direct modification of useful plants to improve their production as well as providing an important contribution in attempts to understand the molecular basis of the control of plant gene expression.

Many of the advances in plant molecular biology have been promoted by the development of efficient procedures for introducing *in-vitro* manipulated DNA into plant cells. The plant transformation system which has proved to be the most successful is that based on the Ti plasmid of *Agrobacterium tumefaciens*. The production of transgenic plants via this route is efficient and routine with some selected plant species and varieties (Gasser and Fraley, 1989). However, for the majority of agronomically important plant species and varieties, efficient and reproducible production of transgenic plants is still an unsolved problem.

1.1 Gene transfer techniques.

The numerous methods for gene transfer to plants may be divided into two general types. The first approach utilises the soil bacterium *A. tumefaciens* as a vector for introducing foreign DNA, whereas the second approach involves the application of naked DNA. Methods for introducing naked DNA include treatment of protoplasts with DNA in the presence of polyvalent cations and polymers such as PEG (Paszkowski *et al.*, 1984), electroporation (Fromm *et al.*, 1986) and microbombardment (Klein *et al.*, 1988). Each of these methods has its specific merits and drawbacks and there is no single method which has the potential to solve every practical problem or utility with every plant species. For example, frequent multimerisation and rearrangement of the foreign DNA sequences prior to integration are observed following transformation with

naked DNA. As a consequence, *Agrobacterium*-mediated DNA transfer is the most favoured method for producing stable transgenic tissue. On the other hand techniques involving the application of naked DNA have been used extensively in studying transient gene expression and in combination with the microprojectile gun may prove to be a sufficiently versatile alternative to *Agrobacterium* to be a universal method for plant transformation.

1.2 Transformation using *Agrobacterium.*

Agrobacterium tumefaciens and *Agrobacterium rhizogenes* cause crown gall tumours and hairy root disease, respectively, on a wide range of dicotyledonous plants and gymnosperms and on some monocotyledonous plants. Virulent strains of *Agrobacterium* harbour a large tumour-inducing (Ti) or root inducing (Ri) plasmid. During the process of infection at wound sites in susceptible plants, a specific segment of the plasmid, T-DNA (transferred DNA), is transferred from the bacterium to host cells and is inserted into the nuclear genome. The T-DNA contains genes for the synthesis of cytokinin and auxin, phytohormones that cause proliferation of plant cells to form a gall. The plant cells are also directed by T-DNA genes to produce unusual amino acids called opines, such as octopine and nopaline, which serve as particular nutrients for the *Agrobacterium* (Zambryski, 1992) and are often specific to type Ti plasmids. T-DNA on the Ti-plasmid is bordered by directly repeated 25-bp sequences that are required in *cis* for its efficient excision and transfer. No other part of the T-DNA is necessary for its transfer.

1.2.1 *vir* genes.

Excision and transfer functions of T-DNA are mediated by so called virulence functions that map on the Ti plasmid outside the T-DNA and are encoded by several *vir* genes and operons. Extensive biochemical and genetic studies of the *vir* genes have shown that the virulence region (35Kb) on the Ti plasmid consists of 6 loci (*vir*A,*vir*B,*vir*G,*vir*C,*vir*D and *vir*E) (Koukolikova-Nicola *et al.,* 1987, and Zambryski

et al., 1992). Stachel *et al.*, (1988) identified plant signal molecules that induced expression of most of the *vir* genes. This was demonstrated using ß-galactosidase as a reporter for inducible gene promoters to monitor *A.tumefaciens* specific gene expression. The coding sequences for ß-galactosidase were incorporated into a transposon and then inserted randomly throughout the Ti-plasmid. These fusions showed that most genes were not expressed during vegetative growth. However when *A.tumefaciens* was grown in the presence of plant cells the ß-galactosidase fusions in the *vir*-region were induced specifically to high levels of expression. Subsequently these ß-galactosidase fusions were used as a bioassay to purify and identify the plant cell factors responsible for the activation of *vir*-gene expression. Molecules isolated from wounded tobacco cells found to be most efficient in activating *vir* gene expression were identified as low molecular weight phenolic compounds acetosyringone and hydroxy-acetosyringone. Other phenolic compounds found to induce *vir* genes such as catechol, pyrogallic acid, syringyl alcohol, and certain chalcone derivatives have also been identified (Winans, 1992). These molecules share features with lignin (cell wall subunits, breakdown products, and precursors such as synapic acid. *Agrobacterium* may recognise these inducers as indicators of the wound environment mediated via gene products of the chromosomal virulence (*Chv*) genes. The expression of the *Chv* loci is constitutive and their genes may therefore play a general role in a mechanism allowing *Agrobacterium* to seek out plants.

The phenolics act through two component sensor/regulator systems that are highly conserved in a variety of prokaryotes. In *Agrobacterium* the sensor component is encoded by *vir*A and the regulator by *vir*G. The constitutively expressed *vir*A gene produces a protein located in the inner membrane that functions as a chemoreceptor to sense the presence of plant wound metabolites (Zambryski, 1992). The resulting autophosphorylation of VirA protein activates the intracellular signal transducing protein VirG by phosphorylation of an aspartic residue (Jin *et al.*, 1990). The activated VirG binds to *vir* gene promoters and acts as a transcriptional activator of virulence genes (Jin *et al.*, 1990; Pazour *et al.*, 1992).

1.2.2 T-DNA transfer.

The mobility of the T-DNA is largely determined by its flanking 25bp repeat sequences, the T-DNA borders. Genetic analysis has emphasised the importance of the right border sequence, whereas deletion of the left border sequence does not have a severe effect. The molecular mechanism of the process was shown by Miranda *et al.*, (1992), who reversed the orientation of the right border and showed that the entire 200Kb Ti plasmid can be transferred, proceeding clockwise around the plasmid with the actual T-DNA transferred last. This indicates that the T-DNA transfer normally occurs in a polar fashion from the right toward the left border sequence and also that large lengths of DNA can be transferred.

T-DNA transfer involves production of a single-stranded DNA molecule termed a T-strand (Stachel *et al.*, 1986). Recently strong support for a single stranded intermediate has come from two very different strategies used to assay the nature of the T-DNA copy upon arrival in the plant cell. Yusibov *et al.*, (1994) monitored the synthesis and disappearance of the T-DNA element within *Agrobacterium* and the appearance of T-DNA segment in tobacco protoplasts as detected by PCR. The disappearance of the T-DNA copy from *Agrobacterium* was correlated with the appearance in the plant cytoplasmic fraction of the T-DNA segment. The latter product was not amplified if the plant fraction was first treated with S1 nuclease absent from double stranded DNaseI activity therefore indicating the transfer of a single stranded T-DNA molecule. In a second study conducted by Tinland *et al.*, (1994) a sensitive extrachromosomal recombination assay was used. The strategy employed was to develop a T-DNA with a *gus* gene divided into two overlapping segments. *In planta* recombination is then required to yield a full length functional gene. Extra-chromosomal recombination that takes place early after the entry of the T-DNA into the nucleus can discriminate between the transfer of ssDNA or dsDNA. If the transfer intermediate is dsDNA, recombination should produce an intact *gus* gene regardless of whether the segments in the T-DNA are of the same or opposite polarity. If the transfer intermediate is ssDNA, however, then a

complete *gus* gene can be obtained through recombination only if the segments are of opposite polarity. GUS activity in infected protoplasts was an order of magnitude greater from the T-DNA bearing *gus* segments of opposite polarity relative to *gus* segments of the same polarity. Thus both the above studies provide strong evidence for a ssDNA transfer intermediate.

Essential to the production of a single stranded intermediate (T-strand) are the *vir* D1 and *vir*D2 gene products (Filichkin and Gelvin, 1993). Together, VirD1/VirD2 recognise the 25bp border sequence and produce a singe-stranded endonucleolytic cleavage in the bottom strand of each border. The 5' end of the T-strand is formed at the right T-DNA border. After nicking, VirD2 remains tightly associated with the 5' end of the T-strand and it is hypothesised to assist T-strand transfer out of the bacterial cell and into the plant nucleus. To preserve the integrity of the T-strand during its transfer from the bacterial cell to the plant nucleus it was hypothesised that the T-strand is transferred as a ssDNA-protein complex. Vir E2 is a wound inducible single stranded nucleic acid binding protein without sequence specificity encoded by the *vir* E locus. The VirE2 is thought to completely coat the T-strand preventing its degradation by nucleases and is also involved in the unfolding and extension of ssDNA to a narrow diameter of 2nm which may facilitate its transfer through membrane channels. The T-strand together with VirD2 and VirE2 are termed the T-complex. The exit of the T-complex from the bacterial cell and its entry into the plant cell are as yet not fully understood. However it has been hypothesised that the nuclear uptake of the T-complex is assisted by its associated proteins VirD2 and VirE2. Support for this hypothesis comes from deduced sequence analysis of *vir*D2 and *vir* E2, which shows in the last 30 amino acids of the C-terminus a sequence homologous to the bipartite type of NLS(Nuclear localisation signal) (Howard *et al.*, 1992). The nuclear localising function of this sequence was confirmed by fusing its coding region to the *gus* reporter gene and transiently expressing this construct in tobacco protoplasts. The nuclear localisation of the fusion protein was inferred from the accumulation of blue product in the nucleus. When the putative NLS was deleted from full length *Vir*D2 fused to *gus*,

the *gus* gene product was only in the cytoplasm. Nuclear accumulation of transiently expressed VirD2 has also been demonstrated by immunolocalisation (Tinland *et al.*, 1992). Sequence analysis has also revealed two potential bipartite NLS's in VirE2. As with VirD2 the nuclear localisation of VirE2 was tested by fusing its coding regions to the *gus* gene. Although mutants in which one or other of the NLS were deleted exhibited some nuclear transport, both NLS sequences were required for maximum accumulation in the nucleus (Citovsky *et al.*, 1992, Citovsky *et al.*, 1994) Therefore it can be envisaged that VirD2 initiates uptake into the nucleus ensuring that the 5' end of the T-strand enters the nucleus firsts and the multiple NLS of VirE2 may be important for uninterrupted nuclear import of the T-complex, e.g. keeping the cytoplasmic and nucleoplasmic sides of the nuclear pore open simultaneously.

1.2.3 Similarities with conjugative transfer systems in other prokaryotes.

The mechanism by which the T-DNA integrates into the host chromosome is not known and is currently one of the most intriguing questions in plant molecular biology. It has been proposed that the polarity of transfer as well as the occurrence of single stranded copies of the T-region are evolutionary related to other bacterial systems that produce ssDNA, during bacterial conjugation.

More recently functional and molecular similarities of T-DNA transfer and conjugation have been completely characterised with respect to the conjugative transfer system encoded by the broad host range plasmid RP4 (IncP). The plasmid RP4 (60Kb) was originally identified in a *Pseudomonas* clinical isolate, in which nearly half of the plasmid genome is devoted to conjugative transfer functions. These are encoded by two distinct regions of the plasmid, known as *Tra*1 and *Tra*2. *Tra*1 encodes the gene products essential for the initiation of conjugative DNA synthesis and contains the replicative start point or the so called origin of transfer (*ori* T). Functions of Tra2 together with one component of Tra1 are needed to establish physical contact between the donor and recipient cell. The RP4 transfer apparatus can mobilise DNA to a variety

of phylogenetically distinct micro-organisms, including Gram-negative, Gram-positive bacteria and yeasts. The detailed analyses of the RP4 conjugative machinery has provided data confirming the predicted analogies between most of the processes occurring in the donor cell (i.e. cell-cell contact, initiation of DNA transmission, and DNA transport through the membranes) for *Agrobacterium*-plant and conjugative DNA transfer systems. An essential prerequisite for DNA transfer is an intimate contact between donor and recipients cells. In conjugation with Gram-negative bacteria a plasmid-encoded pilus is necessary for the establishment of the initial contact between mating cells. Experiments with donor specific phages indicate that for RP4, most probably all functions of the Tra2 region in concert with TraF of Tra1 are responsible for donor-recipient contact (Lessl *et al.*, 1992). Electron microscopy studies confirm that these so-called mating pair formation (Mpf) determinants are involved in the production of an RP4 conjugative pilus. To date only chromosomally encoded products have been identified to be involved in the attachment of *Agrobacterium* to the plant cell with no evidence to confirm whether a Ti-plasmid encoded pilus like structure is also involved. However six of the *Tra*2 gene products show significant similarities to the proteins encoded by the Ti *vir*B region. (Lessl and Lanka, 1994), especially in features that suggest membrane localisation, e.g. hydrophobic domains. The genes for five of these Tra2 proteins are collinear with their *vir*B counter parts and show sequence similarities to the pilus subunit of TraA of the *Escherichia coli* F factor (Lessl and Lanka, 1994). It is also interesting to consider since most of the RP4 *Tra*2 and the Ti *vir*B gene product functions are thought to be membrane associated that each might form a transmembrane pore allowing the transferring DNA to be exported from the donor and imported into the recipient (for a review see Zupan and Zambryski, 1995). In addition to the similarities in the transport machinery used for conjugation and *Agrobacterium*-plant cell interaction, recent studies demonstrate that the initiation of the DNA transfer process follows similar mechanistic principles in both systems. According to the current model (Lessl and Lanka, 1994), components of the RP4 relaxases (TraI, TraJ) and the Ti *Vir*D (VirD1,VirD2) operon specifically interact with

their cognate targets (nick region of *ori*T or T-DNA border sequences respectively) to catalyse a site- and strand-specific cleavage reaction.

Comparison of the right border (T-DNA) and the nick region of *ori*T (RP4) sequences identified a 12-nucleotide consensus sequence (AC/TC/AT/CATCCTGC/TC/A) that has also been found among various rolling circle-type replicons (Waters and Guiney, 1993, Lessl and Lanka, 1994). Functional analogy between a conjugative transfer origin and a T-DNA border sequence was originally shown in experiments conducted by Buchanan-Wollaston *et al.*, (1987). They demonstrated that the origin of transfer oriT of the conjugative IncQ plasmid RSF1010 together with its cognate Mob functions can substitute for the T-DNA borders in directing DNA transfer to plant cells from *A. tumefaciens*. This hybrid transfer system required an intact Ti-plasmid *vir*-region and a region of RSF1010 encoding polypeptides involved in plasmid mobilisation. The ori-T of RSF1010 and its cognate mobilisation proteins presumably generate a conjugative DNA transfer intermediate which is then mobilised to plant cells by using Vir specific transfer machinery. Similarly VirD1, D2 form a complex on the right border(T-DNA) as does TraI, along with TraH assemble at the nick site on RP4. Site specific cleavage observed by VirD2 is also observed by Tra1 where both remain attached to the 5' end of the DNA to be transferred. *In vitro* Tra1 and VirD2 are each sufficient in the presence of Mg^{2+} to produce nicks in single stranded oligonucleotides bearing their respective cognate sites. Also in the presence of cleavage products, both Tra1 and virD2 can catalyse the opposite reaction joining two pieces of ssDNA. At the molecular level, Tra1 has three highly conserved domains at the N-terminus that it shares with VirD2 and relaxases of other conjugative systems (Pansegrau and Lanka, 1994). Critical amino acids for functionality of Tra1 have been identified in both motif I and III which are also predicted to be conserved in VirD2 based on sequence comparison. A role for VirD1 may also be deduced by comparison with its counterpart in the RP4 system. In RP4 TraJ first binds to the 3' site of the oriT and then recruits Tra1. By analogy, VirD1 may first recognise and bind the T-DNA border, which promotes VirD2 binding. The transfer intermediates in both systems are single

stranded, however unlike the T-DNA transfer system in which a T-complex is formed after the co-operative binding of the single stranded binding protein VirE2 no such ssDNA binding proteins have been identified in the RP4 system. T-DNA transfer and single stranded DNA conjugative transfer both occur directionally from 5' to 3'.

Therefore the *Agrobacterium* plant cell interaction appears to represent novel adaptations of two general prokaryotic processes, activation of gene expression in response to external stimuli and conjugable transfer of DNA from donor to recipient cell.

1.2.4 T-DNA integration.

The sequences of several T-DNA host plant DNA junctions have been determined. Gheysen *et al.*, (1987) pointed out that these junctions, in general appear more variable than the junctions created by insertions of transposons, retroviruses, or retrotransposons. Rather they are similar to the junctions created by insertions of non-retro-transforming viruses such as simian virus 40 or adenovirus.

The target sites of T-DNA integration before and after have been sequenced in several studies (Gheyson *et al.*, 1987, Matsumoto *et al.*, 1990, Mayerhofer *et al.*, 1991). Gheyson *et al.*, (1987) examined a single integration event of T-DNA into the tobacco genome. This insertion caused a duplication of 158bp of host DNA, a small deletion, a translocation, and several single base pair transitions. In a study conducted by Matsumoto *et al.*, (1990), seven different integration events into the genome of *Arabidopsis thaliana* were examined. As in earlier studies, the left junctions between T-DNA and plant T-DNA showed less precision than the right junctions. At the left junction, only two insertions contained all sequences found on the T-strand, including the 22 bases derived from the left border. The other insertions lacked either, 7, 16, 29, 31, or 100 bases from the left end of the T-strand. At the right junction, four of the T-DNA inserts contained all sequences found at the right end of the T-strands, including the three bases derived from the right border, while two inserts lacked two bases and one insert lacked 32 bases of T-strand DNA. In all cases, the target sequences suffered

small (29- to 73-base) deletions. In three cases, either 9, 15, or 19 bases of filler DNA separating T-DNA and host DNA were found. Perhaps most significantly, in four inserts a limited sequence identity (5 to 7 bases) was found between the extreme ends of the inserted T-DNA and the corresponding target sequences. Mayerhofer *et al.*, (1991) compared T-DNA junctions with target sequences, investigating a single insertion into the tobacco genome. In this study, the junctions between the T-DNA and host DNA occurred 350bp internal to the left border and 7.3kb internal to the right border. As observed by Matsumoto *et al.*, 1990, a small deletion (23 bases) occurred at the target site and the target site contained limited but significant sequence similarity to the boundaries of the inserted T-DNA sequences. It is therefore possible that the short stretches of homology (5-10 bases) between the T-DNA ends and plant target sites may play a role in integration. The characteristics of the integration events are classically analogous to illegitimate recombination events exemplified by those involving the integration of viral or transfected DNA in mammalian cells. It is suggested that the VirD2-bound 5' end of the T-DNA joins a nick in the plant DNA. The plant DNA may further unwind to form a gap, and the 3' end of the T-strand may pair with another region of plant DNA close by. Plant repair and recombination enzymes may then function to covalently join the 3' end to plant DNA. Theses reactions would result in the introduction of a T-strand into one strand of plant DNA, torsional strain would then result in the introduction of a nick into the opposite strand of plant DNA. Gap repair and DNA synthesis using the T-strand as template result in the final integration product. It may thus be seen that this system has the potential to be exploited in the introduction of specific genes into the plant genome, having targeted that gene for integration by introducing its sequence into the Ti plasmid.

1.3 *Agrobacterium* transformation vectors.

The cloning and genetic manipulation of genes to be transferred to the plant genome are first performed in *E.coli* and then the gene construct containing the gene of interest is transferred by conjugation into *Agrobacterium*. Many of the pBR322-like vectors from

E.coli cannot be directly conjugated into *Agrobacterium* since they do not normally encode the transfer and mobility functions. However they usually retain the bom site where conjugative transfer is initiated by mob functions. Therefore if the mob and tra functions are provided in *trans* by a helper plasmid (such as R64drd11, pGJ28, or pRK2013), mob- plasmids (pBR322) can be mobilised to *Agrobacterium* or any other Gram-negative bacteria. Conjugative transfer of pBR322-like vectors from *E.coli* to *Agrobacterium* can be achieved either as a two step conjugative process or via a triparental mating method. The two-step conjugative process involves conjugation of the helper plasmid into the cloning vector containing the strain. The strain containing both sets of plasmids is then isolated and used to mate with the *Agrobacterium* acceptor strain. A 'triparental' mating may be carried out where the strain containing the cloning vector, that containing the helper plasmid and the *Agrobacterium* strain are mixed together and *E.coli-E.coli-Agrobacterium* conjugations occur simultaneously. The transformation vectors used with *Agrobacterium* are divided into two types, binary or co-integrative vectors.

This system is based on the fact that the *vir* region and the T-DNA need not be linked on the same replicon for functional T-DNA transfer to occur. The binary vector system consists of two components. A cloning vector consisting of an origin of replication both active in *E.coli* and *Agrobacterium*, a selectable marker functional in both bacteria and a small T-DNA. The T-DNA is composed of the 25bp border sequences flanking multiple restriction sites for cloning desired sequences and a dominant selectable marker gene functional in plant cells. The acceptor *Agrobacterium* strain contains a Ti-plasmid depleted of its T-DNA, however provides the *vir* region which can act in *trans* mediating the transfer of the gene of interest and selectable gene marker in plants between the right and left border to the plant cell. Binary vectors contain origins of replication and selectable markers functional in *Escherichia coli* and *A tumefaciens* and hence can be maintained in both types of bacteria e.g. pBin19 (Bevan, 1984). These vectors can be transferred between bacteria by either conjugation or introduced directly as naked DNA by transformation (Ebert *et al.*, 1987). Moreover binary vectors contain

the 25bp repeats between which are located the appropriate cloning sites for insertion of foreign DNA and selectable markers functional in plant cells.

The disarmed Ti-plasmid vector pGV3850 (Zambryski *et al.*, 1983) is an example of a co-integrative vector in which all of the oncogenic phytohormone biosynthetic genes were removed from the T-region and replaced with the linearised plasmid pBR322. The pBR322 DNA provides homology for co-integration of the T-region of pGV3850 with any of the commonly used pBR322-type vectors carrying a cloned gene of interest. The pBR322 intermediate vector is introduced into the *Agrobacterium* harbouring the co-integrative vector by conjugation. As the pBR322 vector lacks an origin of replication which is functional in *Agrobacterium*, it is lost during growth and cell division. However, homologous recombination between the co-integrative and intermediate vectors results in the foreign DNA being integrated between the border repeats of the T-DNA and can be selected for by the appropriate genetic markers usually antibiotic resistance genes. Unlike the binary vector system, both the T-DNA and *vir*-region are present on the same Ti plasmid.

Both vector systems have advantages, but also limitations. Co-integrative vectors have the disadvantage that their efficiency relies on the frequencies of both conjugal transfer and homologous recombination, both of which are relatively low. The major disadvantages of binary vectors are that most of them are still very large and carry a limited number of cloning sites. Stability of theses plasmids is not always optimal in *Agrobacterium* and it has also been witnessed that the efficiency of transfer by binary vectors is significantly lower than that using co-integrative vectors for some plant species.

The use of *Agrobacterium*-mediated plant transformation systems are now firmly established. Tobacco remains the plant species of choice in gene transfer experiments, simply because it is so amenable to tissue culture and plant regeneration can be rapid with plants being regenerated in as short a period as 6-8 weeks. In time *Arabidopsis* may replace tobacco as the 'model' plant species because of the additional benefits it

offers by its small genome and stature, rapid life cycle and extensive genetics.

1.4 Chimaeric genes: Selectable marker and Reporter genes.

Chimaeric genes are usually constructed to contain a fragment from the gene of interest ligated to some heterologous DNA, which is not present in the host system. The chimaeric genes also provide efficient dominant selectable marker genes for use with disarmed Ti-plasmid vectors to allow the selection of cells that have acquired new genetic information as well as to provide convenient reporter genes to facilitate the functional analysis of gene regulatory mechanisms in plants.

The first of these chimaeric genes linked 5' (upstream) and 3'(downstream) sequences of opine synthase genes derived form the Ti-plasmid T-regions with the coding sequences of bacterial genes for enzymes such as chloramphenicol acetyltransferase (CAT), neomycin phosphotransferase (*nptII*), hygromycin phosphotransferase (*hpt*), mouse methotrexate resistant dihydrofolate reductase, and the bleomycin resistant gene from Tn5.

At present the most widely used selectable marker genes are based on the *nptII* gene from the bacterial transposon Tn5. This gene confers resistance in plant cells to aminoglycoside antibiotics, such as kanamycin and G418. The *nptII* gene has been shown to be an effective dominant marker conferring kanamycin resistance in tobacco protoplasts, callus, tissue explants, germinating seedlings and whole plants (Deblock, 1984). The hygromycin phophotransferase (*hpt*) gene from *E.coli* has also been used for the construction of selectable marker genes for plant cells and not as widely used as the *nptII* gene, has also been successfully used with several plant species. In contrast, while chloramphenicol acetyltransferase (CAT) can be used to confer resistance to chloramphenicol in germinating seedlings it is not effective at other developmental stages. Bleomycin is a glycopeptide which interacts with DNA producing single- and double strand- breaks. The bleomycin resistance gene was also found on Tn5 and has

been used successfully as a selectable marker gene in *Nicotiana plumbaginifolia*. Other selectable marker genes such as the dihydrofolate-reductase, have been used, but are not generally useful as those based on the *nptII* or *hpt* genes.

Reporter genes are those which are well characterised both genetically and biochemically, and have a coding region which can be easily fused to the regulatory sequences of other genes. The enzymatic activity of most reporter genes is not normally present in the host plant, or is easily distinguishable from other endogenous activities. This allows the study of a gene under different environmental conditions or in different organs of a plant.

Two types of particular attractive reporter enzymes commonly used are the luciferase genes from either firefly or bacteria (Ow *et al.*, 1987) and ß-glucoronidase(GUS) (Jefferson *et al.*, 1987). The enzymatic activity of GUS can be detected histochemically in plant tissues or accurately measured in fluorimetric assays on very small (milligram) amounts of transformed plant tissues. The activity of luciferin can also be assayed accurately in different plant tissues by simple luminometer measurements. However, it still remains difficult to quantify levels of expression of such markers due to difference in cell size, metabolic activity and penetration of substrate into the cell.

1.5 Promoters and enhancers.

In higher plants initial experimental studies using a combination of in-*vitro* mutagenesis and *Agrobacterium* -mediated or direct gene transfer identified two distinct types of *cis*-acting regulatory sequences: promoters which are located close to the transcription initiation site and act in a position-dependent fashion and enhancers which can be located far from the transcription initiation site and act in a position and orientation-independent fashion.

Promoters can be sub-divided into proximal *cis*-elements including the cap site itself, the TATA box located 25-30bp upstream from the transcription start site, and also a CCAAT box approximately 80bp upstream form the transcription initiation site, as well

as distal regulatory *cis*-elements which can spread over several hundred base pairs. Both the TATA box and CCAAT boxes perform vital functions in transcription. The TATA box functions to position the start site for RNA synthesis. However many promoters particularly those of the house keeping genes lack TATA boxes and are instead composed of GC rich elements that are often located within methylation free islands (Bird *et al.*, 1995). Transcription from this latter class of promoters often initiates at multiple sites.

The *cis*-acting elements have been defined as short DNA segments or motifs whose integrity is required for promoter activity. These motifs reside in functional domains or modules which by themselves have no or only weak enhancing activity, but which act synergistically to give high levels of activity. Enhancers are also organised much like promoters in that they too are composed of many individual modules. Unique promoter specificity and regulation can be conferred by the particular composition and organisation of *cis*-acting DNA sequence elements found in the promoter and enhancer regions. The *cis*-acting elements operate by interacting with specific trans-acting nuclear protein factors, many of which have been identified by gel retardation and foot printing assays and some of which have been purified to homogeneity. In some cases the corresponding genes have been cloned. The promoter-specific transcription factors function by binding the *cis*-acting elements and the overall rate of transcription initiation is determined by protein-DNA and protein-protein interactions.

1.5.1 Plant pathogen promoters.

Plant specific promoters derived from functioning in plant pathogen interactions i.e. T-DNA or *CaMV*, have been used most extensively to direct the expression of foreign genes in plants. Initially it was considered that these promoters were constitutive in their mode of expression i.e.. expressed at all times in all cells. However, with the use of sensitive reporter genes such as ß-glucoronidase or luciferase which are capable of detecting cell-specific expression it has become clear that these promoters can be affected by subtle environmental and developmental changes in the levels of growth

substances within the plant. For example, the nopaline synthase promoter of T-DNA has been found to be most active in the basal region of plants with its activity lower in the upper parts of the plants suggesting that it may respond to differing levels of auxin (An *et al.*, 1988). Following flowering, the activity of the promoter in vegetative organs declines whereas in flower tissues there is a generally higher level of activity. Other T-DNA promoters also appear to respond to endogenous plant growth substances. The gene 5 promoter is most active in callus, stems and roots while activity in leaf tissue is barely detectable whereas the 1'/2' dual promoter is the most active in the basal region of the plant suggesting that auxins control its expression (Langridge *et al.*, 1989). Promoters of T-DNA genes such as the nopaline synthase (*nos*) promoter, the octopine synthase (*ocs*) promoter and the mannopine synthase (*mas*) promoters have all been studied in transgenic plants using sensitive reporter genes. Although these promoters are active in undifferentiated cells and in various plant organs, a tissue or organ preference has been reported, with the highest activity often in the roots or wounded leaf tissue (An *et al.*, 1988, Teeri *et al.*, 1989).

Ellis *et al.*, (1987) have defined within the octopine synthase promoter a specific 16bp palindromic *cis*-regulatory sequence, known as the *ocs*-element. The *ocs*-element can function as an enhancer in dicot (tobacco) and monocot(maize) cells. The insertion of synthetic oligonucleotides containing the *ocs*-element upstream of the maize *Adh1* promoter, inactivated by deletion, reactivated the expression of this promoter in maize and tobacco protoplasts. The transcriptional enhancing activity of the *ocs*-element correlated with the *in-vitro* binding of a transcription factor OCSTF. The *ocs*-element has now been identified in six other opine synthase genes and also in promoters of 3 caulimoviruses. The 8bp direct repeats of the 35S promoter overlap the pentanucleotide direct repeats of the 35S promoter that make up the *as-1 cis*-element.

1.5.1.1 Cauliflower mosaic virus promoter.

Cauliflower mosaic virus (*CaMV*) is a DNA virus that has a specific host range infecting only members of the *Cruciferae*. The double-stranded genome is

approximately 8Kb long and its complete nucleotide sequence has been determined (Hohn *et al.*, 1982). Replication occurs via reverse transcriptase and does not involve integration into the host genome. Transcript mapping experiments have identified two viral promoters, designated 19S and 35S. During the virus life cycle the 35S promoter is transcribed from the viral DNA minus strand to produce an 8Kb transcript referred to as 35S RNA (Guilley *et al.*, 1982). The 5' and 3' termini of the RNA have been determined by S1 nuclease analysis and shown to have an overlapping sequence of about 200 nucleotides (Covey *et al.*, 1981, Guilley *et al.*, 1982). In addition to serving as template for translation, the 35S RNA with its direct terminal repeat sequence also functions as an intermediate for viral DNA synthesis through a reverse transcription process (Pfeiffer and Hohn 1983). Rapid viral replication in infected cells requires a copious supply of the 35SRNA. Therefore it is not suprising that this RNA is a major transcript in infected cells (Guilley *et al.*, 1982). The high levels of production of the 35S RNA is evidence of the strength of this promoter. The promoter is also nominally constitutive which is beneficial in allowing successful reproduction within most plant cells.

The *CaMV* 35S promoter has somehow evolved *cis*-elements regulated by cellular factors, not viral proteins. Information on the 35S transcriptional start site as well as the availability of cloned *CaMV* sequences have encouraged the use of the 35S promoter as a model system to characterise the modular organisation and interactions of enhancers which function in higher plant gene expression. *CaMV* fragments containing 400 to 1000 base pairs of 35S upstream sequences have been shown to be active in most organs of transgenic plants (Odell *et al.*, 1985, Jefferson *et al.*, 1987, Kay *et al*, 1987; Sanders *et al.*, 1987). In the early studies conducted by Odell *et al.*, (1985) various 5' deletions of the promoter were fused to human growth hormone as a reporter gene and integrated into the tobacco genome. This work demonstrated that sequences up to 46bp upstream from the transcriptional start site were able to produce a low level of transcript. The levels of transcript were increased about 20-fold if the region up to 168bp from the transcriptional start site were used. The promoter can also be expressed

transiently in protoplasts of several dicots and monocots (Onlee *et al.*, 1986, Ow *et al.*, 1987; Fromm *et al.*, 1986). Ow *et al.*, (1987) continued the 5' deletion analysis of the 35S promoter using the luciferase gene as a reporter in transient expression assays in carrot cells. Analysis of the expression levels led to several conclusions. At least three functional domains were identified: a proximal region (PR) composed of a TATA box (-46 to +9); a medial region (MR) containing a CCAAT-like box (-46 to -86); and a distal region (DR) with three elements similar to the SV40 core enhancer (-86 to -148). This however did not define the exact boundaries of the elements. Apart from the discovery of the three regions, these results also varied from those of Odell *et al.*, (1985): the Pr fragment was found to have more activity when integrated compared with those found by Ow *et al.*, (1987). This was suggested to be due to the effects of heterologous flanking sequences upon genome integration which may then activate the proximal element. The DR element was orientation independent when both MR and PR elements were present indicating enhancer-like activity. It was however position dependent, although possibly distant effects could be seen with more copies of DR. Multiple DR elements were effective and additive at close range. Further analysis using CAT as a reporter gene in transgenic plants showed that the majority of promoter strength was conferred by an upstream fragment between -343 and -46 (Fang *et al.*, 1989). Fang *et al.*, (1989) found that the addition of the -343 to -208 region to the promoter at -90 can activate transcription. A deletion between -343 and -105 led to a 75% decrease in promoter strength, and duplication of this region can enhance transcription 10-fold. It therefore seems likely that -343 to -208 contains an enhancer sequence. 5' and 3' internal deletion mutants indicated the presence of three regions between -343 and -46 which are needed for maximum promoter activity. The first lies between -343 and -208 and is responsible for 50% of total 35S activity. It also requires the presence of the -90 to -46 region to function. The second is between -208 and -90, and when deleted leads to a severe decrease in expression of approximately 40%. This therefore may contain an enhancer sequence as its function is orientation independent and a linear increase in transcription with multimerisation of up to 4 copies occurs. A chimaeric promoter

containing the enhancer element of the 35S promoter (-208 to -46 fragment) and a truncated promoter of wheat alpha amylase was shown to function in maize cells and tissues (Omirulleh *et al.*, 1993). The highest GUS specific activity was found in younger leaves with progressively less activity in older leaves, stem and root. The third region lies between -90 and -46 and plays an accessory role in increasing the transcriptional activity of upstream enhancers. The presence of CCAAT boxes in the region would facilitate the activity of enhancers, as observed in animal systems where some enhancers require the presence of these boxes Zenke *et al.*, (1986). These results indicate that maximal expression requires the co-operation of these multiple *cis*-elements.

Benfey *et al.*, (1990a) divided the promoter into two domains : domain A (-90 to +8) and domain B (-343 to -90) as illustrated in figure 1.1. Each domain is able to confer tissue specific and developmental regulated expression in transgenic plants. Domain A (which also contains the TATA box at -31 to -25) is able to confer expression principally in root tissue. A *cis* element located between -85 and -64 appears to primarily responsible for this expression. The element termed activation sequence (*as-1*) contains a tandem repeat of the sequence TGACG and binds a nuclear factor activation sequence factor (asF)-1 in tobacco nuclear extracts (Lam *et al.*, 1989). A cDNA clone for factor that binds to the *as-1* motif was isolated by (Katagiri *et al.*, 1989) and demonstrated to give increased levels of hybridisation to RNA from root than in leaf tissue. Domain B is almost entirely within an open reading frame(ORF) that codes for proteins implicated in host range specificity and disease symptoms. Domain B is able to confer expression in most cell types of leaf and stem as well as vascular tissue of the root (Benfey *et al.*, 1990b). Extensive detailed deletion analysis of this region has revealed 5 *cis*-elements residing within domain B which are responsible for this expression profile. Constructs containing various combinations of the 5 subdomains fused to domain A during the expression of the *E.coli* reporter gene GUS were used to compare the patterns of expression throughout the development of transgenic tobacco lines. Only three of these sub-domains were able to confer

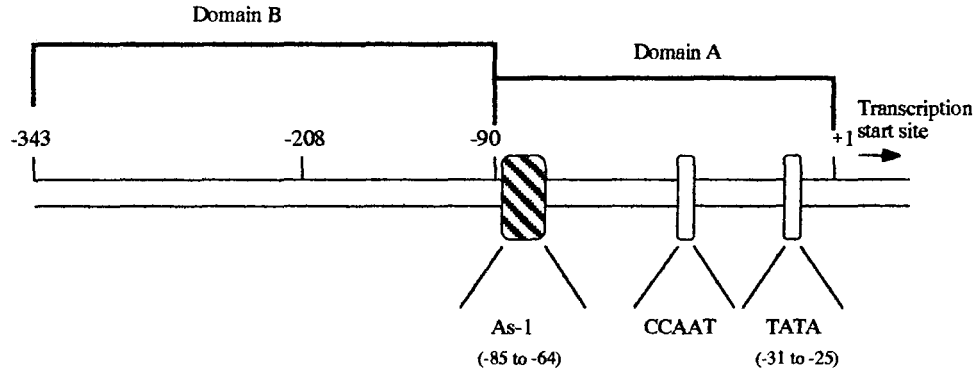
expression during the early stages of development when fused to a minimal promoter. Each sub-domain with only domain A conferred a different expression pattern. These findings indicate that three are active *cis*-elements in each sub-domain. The different expression patterns obtained from each sub-region suggest that interactions with different trans-factors is occurring, implying a modular organisation of domain B. Synergistic interactions were also found to occur between the subdomains as well as between domains A and B. Sub-domain B2 was found to be orientation independent suggesting that it has enhancer like activity. B3 produced different expression patterns in different seeds. This may have been due to overlapping *cis*-elements which are affected differently in different insertion sites in the host genome. An accompanying paper by the same authors analysed the expression patterns from the B subdomains in mature transgenic plants. Developmental regulation of expression was found, as low expression was apparent in sub domains giving none in earlier development. Specific cell type expression was also found, for example B4 only gives expression in immature vascular tissue and B2 expression was found specifically in phloem cells. B1 and B5 required fusion to domain A to confer expression in non-vascular tissue of the leaf and stem. B3 has a complicated expression pattern depending on the age of tissue, and the type of tissue. This may indicate that it contains many active *cis*-elements. Synergistic interactions were also described, for example a combination of B4 and B5 produces expression in root phloem tissue which is not obtained with either subdomain alone. Comparisons of expression patterns conferred by various combinations of 35S enhancer elements suggested that synergistic interactions among *cis*-elements may play a role in defining tissue specific expression.

Evidence has therefore been accrued for both additive and synergistic interactions among the B subdomains of the 35S promoter. The combinatorial properties of the *cis*-elements and the exact function of the trans-acting factors that control developmentally regulated and tissue-specific expression in plants, still need to be precisely defined.

Figure 1.1:

Schematic representation of regions of interest within the CaMV 35S promoter.

For a detailed interpretation see section 1.5.1.1.


Domain B
Domain A
-343
-208
-90
+1
Transcription start site
As-1
(-85 to -64)
CCAAT
TATA
(-31 to -25)

1.5.2 Light-inducible plant promoters.

The plant promoters that have been studied in most detail are those that respond to environmental stimuli such as light, anaerobic or heat stress, wounding, induction and or developmental cues such as leaf, seed, flower and fruit development. Moreover tissue specific expression has been investigated in green tissue, seeds, nodules, tubers and flowers.

Genes encoding proteins of the photosynthetic apparatus, such as Chlorophyll a/b binding (CAB) proteins and proteins of ribulose-1,5-bisphosphate carboxylase oxygenase small subunits (ssRUBISCO) exhibit light regulated expression. For *rbcS* genes this light regulation has been shown to be mediated primarily at the level of transcription and involves both phytochrome and the blue light photoreceptor (Tobin-Silverthorne 1985). Furthermore the expression of these photosynthetic genes is under developmental regulation and is commonly restricted to chloroplast containing photosynthetic tissues (Fluhr *et al.,* 1986). *Agrobacterium*-mediated gene transfer studies with reporter genes have demonstrated that 5' flanking DNA sequences of the *rbcS* genes from several plant species confer light inducible and tissue\organ-specific expression (Nagy *et al.*, 1985). There have been a number of reports demonstrating the existence of both positive and negative *cis*-acting elements within the 5' flanking sequences of light regulated genes including *rbcS* and chlorophyll a/b binding genes. The level and pattern of expression of members of the *rbcS* gene family has been found to be highly variable (Sugita and Gruissem 1987). When a 1.1KB fragment containing the tomato *rbcS*-3A promoter was fused to the CAT coding sequence and transferred to tobacco plants, high level CAT activity was measured in mature leaves (Ueda *et al.*, 1989). In contrast, no or very low expression was detected in roots, stems, flower buds, sepals, petals, stamens, ovaries or stigmas. The activity in young leaves was approximately 10% of that in mature leaves. When the region fused to the CAT gene was restricted to the 374bp sequence located upstream of the transcription start, the tissue-specific and light-inducible expression was maintained but the level of expression

was 5 times lower than that obtained with the full length promoter. During the development of the leaves the enhancement of expression level was only visible in mature leaves, suggesting that the modulation of expression level is established via the interaction of *trans*-acting factors with *cis*-elements present in this upstream region and that the activities of these *trans* acting factors are under developmental control. The levels of expression given by the full length tomato *rbcS*-3A-CAT fusion was estimated to be 50-70% of that of the 35S-CAT construct.

Extensive *cis*-element analysis of the 5' upstream region of the pea*rbcS*-3A gene (Green *et al.*, 1987) revealed multiple light responsive elements (LRE's). One such element located between -166 and -50 can confer light responsive expression upon the homologous pea promoter as well as heterologous promoters (35S and *nos*). Moreover, transcription of -166 pea *rbcS*-3A in transgenic etiolated seedlings is induced by red light and suppressed by far red light, indicating that it contains a phytochrome responsive element (Gilmartin *et al.*, 1990).

The -166 to -50 region contains two conserved sequences, boxes II and III that serve as binding sites for the nuclear factor GT-1. Analysis of site-specific mutants of pea *rbcS*-3A promoter demonstrated that GT-1 binding *in-vitro* is correlated with light responsive expression of the *rbcS* promoter in transgenic plants (Gilmartin *et al.*, 1990). Indeed a synthetic tetramer of box II inserted upstream of the -90 *CaMV* 35S promoter was sufficient to confer light responsive transcription in chloroplasts containing cells of transgenic plants. Recently Krapp *et al.*, (1993) have shown that the level of carbohydrates can regulate the expression of *rbcS* and other photosynthetic genes. It was reported that both the addition of carbohydrate, and the accumulation of carbohydrate due to inhibition of export lead to a rapid and reversible inhibition of the expression of several genes for proteins involved in photosynthesis in autotrophic cells and in detached leaves and whole plants. *De novo* ribulose-1,5-bisphosphate carboxylase oxygenase protein levels were found to decrease followed by a gradual decline in ribulose-1,5-bisphosphate carboxylase oxygenase activity, as well as the

activity of other photosynthetic enzymes.

1.6 Factors influencing transgene expression in stable transformants.

In the analysis of transgenic plants very little is currently known about the actual molecular processes acting upon the foreign DNA during uptake, intracellular transport and stable integration into the genome of plant cells. It is clear that genetic material newly introduced into plant cells must be subject to considerable modification and processing between its entry into the target cell, and its subsequent expression as part of the structurally organised chromatin of the resulting transgenic plant. Of specific importance to the genetic engineer is the question of how the process of genetic transformation affects both the character, level of expression and stability of expression of chromosomally integrated genes. One such widely reported effect in both animals and plants is a seemingly random clonal variability in the level of expression of newly introduced transgenes.

Expression level variability between different transgenic lines has been observed after introduction of many unrelated genes, both natural and chimaeric into numerous plant species (Jones *et al.*, 1985, Shirsat *et al.*, 1989). The observed variability has often been referred to as 'position effect' based on the assumption that expression levels of an introduced gene are directly influenced by host DNA sequence or chromosomal structure at or near the site of integration (Hobbs *et al.*, 1990, Prols and Meyer, 1992).

Much of the work has focused on the structure, organisation and number of plant T-DNA inserts in transgenic plants generated using disarmed *Agrobacterium* T-DNA vectors carrying foreign chimaeric genes.

1.6.1 T-DNA structure.

Southern hybridisation analysis has revealed that usually between one and three T-DNA copies are frequently found to be integrated into the genomes of various plant species,

including tobacco, petunia, tomato, sunflower, morning glory, and *Crepis capilaris* by the transformation system used. Although with one type of *Agrobacterium* strain C58, multiple copies of inserted DNA can be preferentially obtained. Multiple insertions of T-DNA of between 20 and 50 copies have been detected (Chyi *et al.*, 1986, Jorgensen *et al.*, 1987, Spielman and Simpson, 1986). These multiple insertions can occur at different loci on different chromosomes and may result from the *Agrobacterium* transferring more than one copy of the T-DNA prior to its integration. Multiple copies can also be either linked in tandem arrays or unlinked at different chromosomal locations. These results coupled with the ability of *A.tumefaciens* to transform a wide range of dicotyledenous plant species indicate that the T-DNA insertion is not specific to target sequence. This issue is both of fundamental interest and practical importance in the use of *A.tumefaciens* as a generalised vector for gene transfer to plants.

Tandem arrangement of the T-DNA can be composed of either direct or indirect repeats and both inverted and direct repeats may be present in a single T-DNA array. It has been demonstrated in mammalian cells that when two genes are placed in tandem, transcriptional read through from the first into the second reduces the expression of the second gene (Proudfoot 1986). Similar observations were made by Ingelbrecht *et al.*, (1991) using hybrid constructs consisting of the *nos* promoter driving the *nptII* gene in one direction and the 35S *CaMV* promoter located downstream in the opposite orientation to the *nptII* gene. It was demonstrated that transcriptional activity interfered with the expression of a downstream gene in the opposite orientation both in transient expression systems and also in transgenic plants. One possible explanation is that the RNA polymerase transcribing from convergent promoters interfere with one another or alternatively the production of antisense *nptII* RNA and subsequent base pairing with sense *nptII* RNA could be responsible for the reduced expression levels. Although the above possibility cannot be ruled out, the majority of reports have shown both single copy and repeated T-DNA structures to be equally capable of expressing internal T-DNA sequences. The variability in expression is most often due to the insertion into different chromosomal sites i.e. position effect. There is, however, some evidence that

multiple T-DNA copies, especially inverted copies are unstable and possible targets for rearrangements within the plant genome (Jones *et al.*, 1987). Therefore in minimising transgene expression level variability, it would be best to use transformants with single or few directly repeated T-DNAs.

1.6.2 T-DNA integration site specificity.

The T-DNA is thought to integrate at random positions in the plant chromosome. DNA regions undergoing transcription are presumably more accessible to invading exogenous DNA owing to requisite unravelling of nucleosomes and other chromatin structures. Eckdahl *et al.*, (1989) have shown that the T-DNA borders possess segments with the structural and sequence features for autonomous replication in yeast. The function of these segments in the plant cell is not known however, it is possible that initiation within the integrated T-DNA sequences results in their synchronous replication during the s-phase of the cell cycle. For example replication early in s-phase could ensure that each integrated T-DNA is transcribed, since early replicating sequences are packaged into a potentially active chromatin conformation. Evidence put forward by Koncz *et al.*, (1989) suggests that DNA introduced into plant cells using *Agrobacterium* vector systems is preferentially integrated into actively transcribed regions. In this work the T-DNA was engineered so that a promoter-less *aph*(3') II (amino glycoside phosphotransferase II) reporter gene was linked to the right end of the T-DNA and transformed into plants along with a plasmid origin of replication and a selectable hygromycin resistance gene. Transcriptional and translational reporter gene fusions were identified by screening for *aph*(3')II enzyme in transgenic tobacco and *Arabidopsis*.

The results indicated that the T-DNA insertions in *Arabidopsis* and tobacco induced transcriptional and translational gene fusions at similar sequences. This is intriguing because the genome organisation of these species is strikingly different. *Arabidopsis* has a small genome (1*C* = 0.07pg, Bennet and Smith, 1991) with a low content of repeated DNA, whereas the tobacco genome is large (1*C* = 3.3pg, Bennet and Smith,

1991) and consists of 60% repeated and 40% single copy DNA . This suggests that the T-DNA preferentially inserts into potentially active regions of the plant genome, however given the close proximity of the selectable marker gene in the T-DNA vector to the reporter gene, selection pressure may have been biased in favour those transformants with T-DNA inserts in regions of active chromatin and thus eliminating many transformants with low expression levels.

A desirable goal in gene transfer is to direct DNA into defined sites in the genome. Site-specific integration can potentially control the variability caused by 'position effects' by placing introduced DNA into the same genomic locus. It might also be useful for clustering transgenes and thereby ease the introgression of multiple traits into elite field varieties. Moreover, the placement of DNA into a known locus will remove the possibility of random insertions behaving as insertional mutagens. In principle, site selectivity of DNA insertion into eukaryotic cells can be achieved through homologous recombination as in some fungal and animal systems. Recombination via homology also permits the replacement of genomic sequences with introduced versions. Unfortunately, gene integration through homologous recombination has yet to be practical with plant nuclear DNA (Lichtenstein and Barrena, 1993). Alternative strategies for gene targeting in plants include recombinase-mediated systems, such as the Cre-*lox* system derived from bacteriophage P1 (Sauer, 1987) and the FLP-*frt* system of the 2-µm plasmid of *Saccharomyces cerevisiae* (Broach, 1981). These site-specific recombination systems are shown to be induced in various heterologous organisms including plants (Kilby *et al.*, 1993, Qin *et al.*, 1994, Lloyd and Davis, 1994). The frequency of recombination with the above systems is found to be remarkably high (Sauer and Henderson, 1988, Broach, 1981), unlike that of homologous recombination which usually occurs only rarely in the somatic cells of higher eukaryotes (Paszowski *et al.*, 1988, Swoboda *et al.*, 1993). Qin *et al.*, (1994) conducted an investigation into the Cre-*lox* system for site specific integration of DNA into the tobacco genome. The bacteriophage P1 encodes a site specific recombination system where the 38Kd Cre protein encoded by the Cre gene mediates recombination

between 34bp *lox* sites. Thus recombination between a *lox* site on a circular molecule and a *lox* site on a chromosome results in the insertion of the circular molecule into the chromosome. To test for site specific insertion into the plant genome, Qin *et al.*, (1994) first placed a 35S-*lox*-Cre construct into the tobacco genome by *Agrobacterium*-mediated gene transfer. The resultant transgenic lines were used to produce protoplasts harbouring the 35S-*lox* Cre for PEG mediated transformation with an insertion plasmid containing a promoter-less *lox-hpt* transgene, where *hpt* encodes hygromycin phosphotransferase. The integration of the plasmid via the *lox*-site recombination would produce a 35S-*lox-hpt-lox*-Cre linkage, thus terminating Cre transcription by displacement of the 35S from Cre and selection of plants with this specific integration via their hygromycin-resistant phenotype. To assess the efficiency of random integration a plasmid that confers hygromycin resistance under the control of the 35S promoter was used. Tens to hundreds of hygromycin resistant calli per million protoplasts recovered were found to be 10-100% the frequencies obtained with the 35S-*hpt* reference plasmid. It was therefore concluded that DNA can delivered by Cre-*lox* recombination into genomic *lox* sites at a workable frequency approaching that of random integration. It was also noted that DNA was found as a single copy inserted at the target locus and additional insertions in the genome were infrequent. When mutated *lox* sites were used no rearrangements of the integrant DNA were observed in a large majority of cases. Thus it would seem that Cre-*lox* recombination may offer a method for precise single copy insertion of the DNA into genomic targets.

1.6.3 Chromatin structure.

The regulation of gene expression is also directly related to structure of the chromosome; for the most part the DNA is compacted too tightly to be available to the transcription apparatus. Therefore the chromatin structure of genes which are transcribed (approximately 10-20% of the total DNA) has to be different from that of the bulk DNA. Both the 'housekeeping' genes, which are expressed in all cell types in the organism, and the genes which distinguish one cell type from another, exist in the

active conformation.

In the first level of chromatin organisation, DNA is wrapped in $1^{3}/_{4}$ turns around the histone octamer (2 x H2A, 2 x H2B, 2 x H3, and 2 x H4) to form nucleosomes and the 11-nm nucleosome fibre. A fifth histone H1 is associated with the DNA linking each nucleosome and is thought to be involved in the second level of organisation resulting in the folding of nucleosome fibers into a higher order structure, the 30-nm chromatin fibre (solenoid model). In the next higher level of organisation, loop domains are postulated to form by attachment of the DNA sequences, scaffold or matrix attachment regions (SARs/MARs) to the proteinaceous network of filaments that runs throughout the nucleus, known as the nuclear scaffold or nuclear matrix. The different levels of chromatin organisation are illustrated in figure 1.2.

The importance of studies of plant nuclear architecture and the three dimensional structure of chromosomes has been recognised as studies on gene expression have progressed. Not surprisingly plant nuclei have been shown to have many features in common with those of other eukaryotes. Extraction of plant nuclei with non-ionic detergents, high molarity salt solutions and treatment with nucleases, has resulted in preparations of the nuclear matrix, a proteinaceous network within the nucleus. The nuclear matrix together with the associated nuclear lamina underlying the nuclear envelope, is likely to be of considerable importance for interphase chromosome structure. Field emission scanning electron microscopy has recently revealed a plant chromosome structure that appears consistent with the organisation of DNA into 10-15nm beads on a string fibres, which are then wound into 30-nm solenoids and probably into loops. Such an organisation is known to exist in animal cells. Further condensation of these loops is believed to lead to structures visible in the light microscope.

Figure 1.2:

The structure of chromatin.

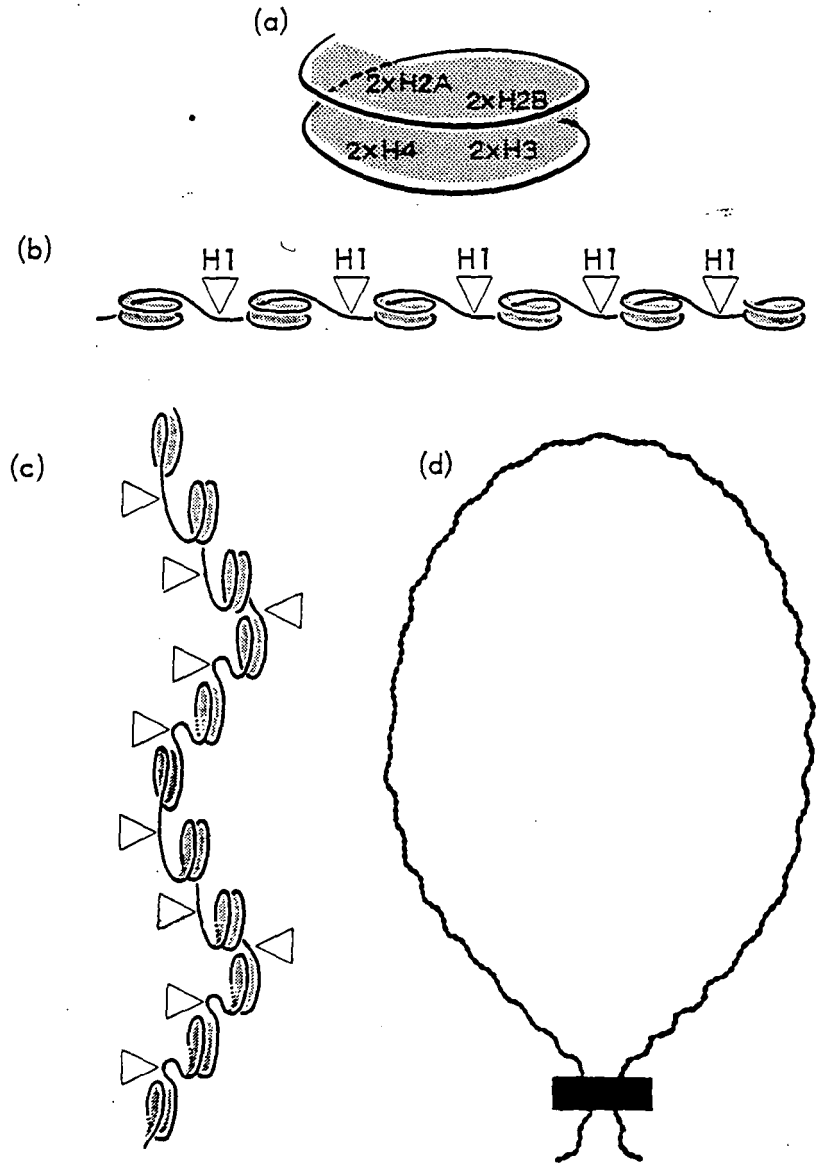
(a): A diagrammatic structure of a single nucleosome with eight histone molecules in the centre and 1.75 turns of DNA on the outside.

(b): Several nucleosomes in an oligonucleosome showing the internucleosomal position of histone 1.

(c): The coiling of an H1-containing polynucleosome into a helix.

(d): A polynucleosome loop restrained so that the removal of the histone protein H1 leads to a loop of superhelical DNA.

From : Adams, et al., 1981.


(a)
2xH2A
2xH2B
2xH4
2xH3
(b)
H1
H1
H1
H1
H1
(c)
(d)

1.6.3.1 The role of chromatin structure in gene expression.

The control of gene expression at the chromatin structure level involves access of RNA polymerase and transcription factors to their binding sites on DNA. These binding sites can be inaccessible because of the highly compact structure of chromatin fibres. A partial unfolding to a less compact, more 'open' structure can render the binding sites accessible. The transcriptional potential of large genomic regions is thought to be controlled by the extent to which the 30-nm fibre (inactive chromatin) is unwound to the level of the 11-nm nucleosome fibre thus permitting access to the DNA by RNA polymerases and transcriptionally regulatory proteins. One of the hallmarks of unfolded, transcriptionally competent chromatin is heightened general sensitivity to endonuclease DNaseI (Weintraub and Groudine , 1974).

In contrast to the abundant information on changes in chromatin organisation of animal genes very few reports have been published investigating the effect of chromatin organisation on the control of gene expression in plants. Murray and Kennard(1984) demonstrated that in nuclei isolated from broad bean cotyledons the genes encoding phaseolin exhibit a higher degree of sensitivity to DNaseI than in nuclei of leaves in which the phaseolin genes are not expressed. However no hypersensitive sites were detected. Similar experiments were conducted on the chromatin structure of legumin genes in nuclei isolated from leaves and cotyledons of pea (Sawyer *et al.*, 1987). It was reported that the transcriptional activity of the legumin genes in pea cotyledons correlated with greater DNaseI sensitivity than that of leaf tissue where the legumin genes are inactive. Similar positive correlations between transcription and enhanced DNaseI sensitivity were demonstrated in maize for the endogenous histone H3 promoter. (Brignon *et al.*, 1993). Both general DNaseI sensitivity and DNaseI hypersensitive sites were observed in nuclei isolated from 4-day germinated maize embryos, the point at which expression of the histone genes reaches a maximum. The hypersensitive sites were found to be located within consensus sequences of the promoter known to act as *cis*-elements. Steinmuller and Apel (1986) conducted an

investigation into the DNaseI activity of three plant genes; the NADPH-protochlorophyllide oxidoreductase (an actively transcribed gene in etiolated barley leaf tissue) ; hordein seed storage protein and 26S ribosomal DNA (known to be inactive in leaves). The hordein and rDNA genes were found to be resistant to low levels of DNaseI, while the gene for the NADPH-protochlorophyllide oxidoreductase was highly sensitive to DNaseI. It was found during the course of digestion of the NADPH-protochlorophyllide oxidoreductase gene, discrete cleavage products were generated, indicating the presence of DNaseI hypersensitive sites in the vicinity of the NADPH-protochlorophyllide oxidoreductase gene in etiolated leaves. Coates *et al.,* (1987) investigated the chromatin structure of the T-DNA when integrated in tumour lines. This work set out to correlate transcriptional activity of the T-DNA genes with an altered chromatin configuration using non-specific endonucleases, micrococcal nuclease and deoxyribonuclease I (DNaseI). In active or potentially active sequences, the chromatin was found to exhibit an enhanced sensitivity to digestion with DNaseI. Coates *et al.,* (1987) concluded that the integrated T-DNA sequences were found in 'open' (active) nucleosome confirmations displaying enhanced sensitivity to DNaseI. This indicated that the majority of the T-DNA integrated sequences were transcriptionally competent.

Gorz *et al.*, (1990) detected increased DNaseI sensitivity and the appearance of DNaseI hypersensitive sites in the 5' upstream regulatory sequences and coding regions of transcriptionally active pea *rbcS* genes (light grown pea plants) when compared with transcriptionally inactive pea *rbcS* genes (dark grown pea plants). It was demonstrated that within the same organ individual members of the pea *rbcS* multigene family are differentially sensitive to DNaseI suggesting differential chromatin structure in the nuclei. During light activation general DNaseI sensitivity increases in some genes, especially their 5' upstream regulatory sequences. DNaseI-hypersensitive sites were constitutively present in 5' upstream regulatory sequences around positions -335, -465, -650 and -945(5' constitutive domain) and in the coding region around position +340, +450, +530, +640, and +810 (3' constitutive domain). One additional hypersensitive

site appears after light induction, this region is centred around position -90 and flanked by light responsive elements (LREs). In spite of the changes in chromatin structure of the pea *rbcS* genes during their transition from an inactive to active state, their cytosine methylation in the promoter region remained uniform. However it was not certain that the methylation sites investigated were those involved in the transcriptional control of the *rbcS* genes. Similar observations to those of Coates *et al.*, (1987) and Gorz *et al.*, (1990) have also been noted in animal systems where the level of DNA methylation does not correlate well with either transcriptional activity or specific chromatin conformations. It is however thought that if repression of transcriptional activity were mediated via nucleosomes, the combined effects of both systems would be weakened. Therefore the multiplication of effectiveness is only possible if the two systems achieve their effects through independent mediators (Bird, 1995). Methylation specific DNA binding proteins have been identified in nuclear extracts of peas (DPB-m) (Zhang *et al.*, 1989) and mammalian cells (MeCP-2 and MDBP-2) (Meehan *et al.*, 1989; Pawlak *et al.*, 1991). MeCP-2 has been found to be associated with heterochromatin (Lewis *et al.*, 1992) whereas MDBP-2 shares sequence homology to histone H1 (Jost and Hofsteenge, 1992), implying a function for both proteins in heterochromatin formation.

1.6.3.2 Covalent modification of histones.

A fundamental manner in which chromatin conformation can be varied is by covalent modification of its histones. Post-synthetic histone modification by phosphorylation, acetylation, methylation and poly-ADP ribosylation are well characterised features of chromatin. Attempts have been made to link these modifications, particularly acetylation, to transcriptional activation Csordas (1985). However even though there are numerous correlations between an increase in histone acetylation and increased RNA synthesis, and considerable evidence for the enrichment of acetylated histones in transcriptional chromatin, it seems unlikely that acetylation alone is sufficient either to produce an active chromatin conformation or to increase transcription rate. Studies of *in vitro* assembled simian virus 40 complexes have shown that acetylation does not

destabilise or open up the nucleosome structure, and the kinetics of transcription of these complexes by calf thymus RNA polymerases A or B is not affected by acetylation. Moreover, no difference is observed on comparing the DNAseI digestion rates of the nucleosome particles from control cells and those treated with butyrate which inhibits cellular deacetylase, implying that acetylation of core histones is not necessarily correlated with DNaseI sensitivity. So although histone hyperacetylation is a characteristic of active genes its function in transcription, is not fully understood.

1.6.3.3 DNA methylation.

Some of the work on T-DNA integration has been concerned with determining which aspects of the plant DNA structure, such as DNA methylation and chromatin arrangement contribute to the integration and transcription of the T-DNA. Recently there has been much interest in the possible role DNA methylation may play in the regulation of gene expression in plants.

Methylation of DNA is used by a wide range of prokaryotes and eukaryotes for the marking or tagging of DNA sequences as part of a range of epigenetic processes, including restriction-modification systems, DNA replication and repair, and regulation of gene expression. Methylation occurs subsequent to DNA synthesis by the activity of a DNA methyl-transferase catalysing the transfer of a methyl group from S-adenosyl-L-methionine to the C5 position of cytosine. The mechanism of methyl group transfer has been studied in detail for the human and murine DNA methyl-transferases (Bestor, 1992, Smith *et al.*, 1992) and the prokaryotic methylase *HhaI* (Wu and Santi, 1987). In addition, an enzyme-S-adenosyl methionine (AdoMet) complex has recently been crystallized (Cheng *et al.*, 1993). A single mammalian methyl transferase enzyme which has been characterised and the gene cloned (Bestor, 1992), displays two methylating activities namely maintenance methylation of the symmetrical cytosine in a hemi-methylated CG doublet and *de novo* methylation of unmethylated CG dinucleotides. *In vitro* and in mammalian cell lines, *de novo* methylation occurs at a lower rate than maintenance methylation, in fact the N-terminal domain of the enzyme

acts to inhibit *de novo* methylation (Bestor, 1992). Although it is not known how tissue-specific methylation patterns are established in mammalian cells, it appears that extensive *de novo* methylation occurs predominantly in the embryo, the methylation patterns are then subsequently inherited via maintenance methylation which occurs throughout development (Razin and Cedar, 1991). Plant cells are also found to exhibit a tissue specific methylation pattern which is maintained for many generations in tissue cultured cell lines (Kovarik *et al.*, 1994, Philips *et al.*, 1994). In plant genomes it is well established that 5-methyl cytosine is present in both CG dinucleotides and in CNG trinucleotides, where N can be any base. Besides symmetrical methylation at CG doublets and CNG triplets a non-symmetrical cytosine methylation has been recently detected in a plant transgene (Meyer *et al.*, 1994). Recently cDNA sequences for *Aribidopsis thaliana* DNA methyl-transferases have been cloned (Finnegan and Dennis, 1993, Schiedt *et al.*, 1994). These have been shown to contain regions of homology with DNAs coding for prokaryotic DNA methylases. As for methylation of CG doublets and CNG triplets in higher plant genomes, it is still not certain whether it is accomplished by a single multi-specific DNA methylase or by multiple enzymes. Southern studies using probes for cloned DNA methylases have revealed multiple homologous sequences in the genome of *Arabidopsis thaliana* which could reflect the existence of a family of DNA methylase genes (Finnegan and Dennis, 1993, Scheidt *et al.*, 1994). Investigation of DNA substrate specificity of pea methylases indicates that more than one enzyme or its several isoforms may exist (Houlston *et al.*, 1993). These data suggest the existence of a complex methylation system in plants.

Information about the distribution of methylated cytosines in plants has come almost exclusively from studies with a pair of isoschizomeric restriction endonucleases, *MspI* cleaves CCGG sequences whether or not the internal cytosine is methylated, *HpaII* will act only on unmethylated DNA. Using each enzyme on the same genomic DNA will allow conclusions to be made about the methylation status of the DNA containing a gene of interest. Cleavage of genomic DNA with the restriction enzyme, followed by electrophoresis, Southern transfer and probing with the appropriate fragment of the

cloned gene, gives a pattern of bands from which the presence of methyl groups in particular CCGG sequences can be determined. The disadvantage of this is that only about 4% of methyl cytosines occur in the sequence and so the analysis is at best patchy. The alternative and a more superior method involves genomic sequencing and a recent modification of this method (Clark *et al.*, 1994) allows the high sensitivity mapping of methylated cytosines on both strands of the target sequence. In this method sodium bisulphite is used to convert cytosine residues to uracil residues in single stranded DNA, under conditions whereby the 5-methyl cytosine remains non-reactive. The converted DNA is amplified with two sets of strand specific primers to yield a pair of fragments, one derived from each strand, in which all uracil and thymine residues have been amplified as thymine and only 5-methyl cytosine residues have been amplified as cytosine. The position of each methyl cytosine is given by a distinct band on a sequencing gel. This approach is able to give information as to the location *in vivo* of all methylcytosines in a gene and its flanking sequences.

1.6.4 DNA methylation and the regulation of gene expression.

It is generally assumed that once a sequence becomes methylated it will remain so through subsequent rounds of replication via the action of a maintenance methylase, however the signals for the initial *de novo* methylation of the unmodified DNA are not completely understood. Accumulating experimental support for methylation of foreign DNA sequences in plants suggest that methylation may occur in response to foreign DNA (Meyer and Heidmann, 1994), via DNA-DNA (Flavell, 1994) pairing and as an RNA -directed process (Wassenegger *et al.*, 1994).

DNA methylation is correlated with decreased expression of many eukaryotic genes, a process which has been studied extensively in animal systems (reviewed by Adams, 1990). In animals, where methylation occurs at CG sites, DNA methylation has been implicated in the control of gene expression (Bestor, 1990), epigenesis (Holliday, 1987), carcinogenesis (Counts and Goodman, 1995) and genomic imprinting Monk, *et al.*, 1987). Methylation also appears to be involved in the regulation of plant gene

expression. An association between methylation and gene expression was observed for genes in crown gall tumours (Hepburn *et al.*, 1983), maize transposable elements (Dennis and Brettell, 1990, Schlappi *et al.*, 1994), endogenous storage protein genes (Bianchi and Viotti, 1988), parental imprinting (Lund *et al.*, 1995a) and ribosomal RNA genes (Torres-Ruiz and Hemleben, 1994). Using a construct containing the *CaMV* 35S promoter and the *gus* reporter gene methylated at every cytosine using methyl-dCTP, Weber *et al.*, (1990) showed that both transient and long term gene expression in transformed tobacco cells was inhibited and that the integrated DNA from non-expressing plants was methylated at both CG and CNG sequences. It has become clear from several studies that transgene inactivation has been associated with an increased level of DNA methylation (Hobbs *et al.*, 1990, Linn *et al.*, 1990, Matzke *et al.*, 1989, Prols and Meyer, 1992) which in two cases analyses led to promoter inactivation (Meyer *et al.*, 1993, Neuhuber *et al.*, 1994). It has been reported that methylation towards the 3' end of the gene can also be as effective as methylation at the 5' end (Keshet *et al.*, 1986, John and Amasino, 1989).

Inactivation associated with methylation occurs to DNA introduced into plant cells using *Agrobacterium tumefaciens* or by direct DNA transfer methods.(Amasino *et al.*, 1984, Peerbolte *et al.*, 1986, Matzke *et al.*, 1989). Gelvin *et al.*, (1983) investigated the methylation of T-DNA in several crown gall tumours using the isoschizomeric restriction endonucleases *MspI* and *HpaII*. Using Southern blot hybridisation they determined the T-DNA was unmethylated at this sequence, however in tumour lines containing multiple T-DNA insertions, T-DNA methylation occurs. In a petunia line carrying three copies of a tms2 gene construct as a negative selectable marker the expression of the transgene was also efficiently shut down and the sequence became hypermethylated, while plants containing one transgene copy did not lose its activity during vegetative culture (Renckens *et al.*, 1992). Multiple transgene copies integrated into the plant genome have frequently been observed to undergo inactivation accompanied with the methylation of cytosine residues in the promoter regions of the transgenes. Linn *et al.*, (1992) employed the A1 marker gene to analyse certain

parameters that influence DNA methylation and to define the molecular mechanism by which DNA methylation influences gene transcription in plants. The A1 gene of *Zea mays* encodes a dihydroflavanol reductase (DFR) an enzyme of the anthocyanin pathway. Due to a block in the anthocyanin pigmentation pathway petunia line R101 produces very low amounts of cyanidin and delphinidin, the two pigments that determine floral pigmentation in petunia. The white flowers of R101 accumulate dihydrokaempferol, a substrate of maize DFR. Introduction of a chimaeric maize A1 gene from *Zea mays* under the control of the 35S *CaMV* promoter into a flowering mutant of *Petunia hybrida* R101, resulted in the conversion of dihydrokaemfrol into leucopelragonidin for further processing by endogenous enzymes into a red pelargonidin pigment. Primary transformants generated could be classified into three groups, plants with whitish flowers without any indication of A1 activity, transformants with salmon red flowers exhibiting continuous A1 activity and plants carrying variegated flower with A1 activity present in some petal cells. The uniformity of petal pigmentation was inversely correlated with the number of integrated A1 copies Most plants with white flowers had multiple copies of the transgene integrated at one or more sites, while flowers with uniform pigmentation generally occurred in plants which carried a single intact copy of the introduced gene. Inactivation of the chimaeric gene was associated with increased methylation of sites within the 35S *CaMV* promoter. Transgene inactivation and the existence of more than one T-DNA copy was also observed to correlate with DNA methylation in studies conducted by Matzke *et al.*, (1989, 1993). In this study doubly transformed tobacco plants were created following sequential transformation steps using two T-DNAs encoding different selection and screening markers. In 15% of the double transformants the phenotype of the gene conferring kanamycin resistance (introduced first) was lost after introduction of the second gene conferring hygromycin resistance. The loss of kanamycin resistance was accompanied by methylation of a specific CpG residues in its nopaline synthase promoter sequence. Segregation of the hygromycin gene away from the kanamycin gene resulted in the restoration of kanamycin activity corresponding with the partial or

complete demethylation of the promoter. It was also observed that different transgene combinations produced no, partial (unstable) or complete (stable) *trans*-inactivation of kanamycin resistance. The results imply that the ability to *trans*-inactivate or to be *trans*-inactivated is defined by the state of the gene loci. The results seemed to indicate that silencing *H* loci (encoding resistance to hygromycin) contained multiple, methylated copies of the H transgene construct and that the degree of methylation of a silencing locus could determine the extent to which it is inactivated and methylated the *K* locus (encoding resistance to kanamycin). The two loci are homologous in 300bp of their nopaline synthase promoter regions. To account for the apparent imposition of methylation from the silencing *H* loci to the target *K* locus, Matzke *et al.*, (1994) proposed a DNA-pairing-dependent process that takes advantage of the fact that maintenance methylase prefers hemimethylated DNA as a substrate. It was also interesting to note that methylation occurred predominantly in the promoter regions and was associated with inactivation at the level of transcription.

Reversible methylation and activation of genes can be induced using the substrate azacytidine (5-azaC), a chemical analogue of, which is incorporated into the nucleic acids in place of cytosine. When DNA methylase interacts with azacytidine containing DNA it forms a dead-end complex in which the enzyme is covalently linked to the azacytidine (Santi *et al.*, 1983). This depletes the cell of active enzyme (Tanaka *et al.*, 1980) and DNA subsequently replicated fails to become methylated, i.e. it is hemi-methylated after one phase. In plants, 5-azaC treatment had already been shown to result in demethylation and subsequent expression of T-DNA genes. (Klaas *et al.*, 1989, Weber *et al.*, 1990). Azacytidine was used by Renkens *et al.*, (1992) to reactivate the T-DNA 2 gene of *A.tumefaciens* in transgenic petunia after it had been silenced by methylation resulting in the reactivation of gene 2 expression. Further evidence for reversible methylation was presented by Weber *et al.*, (1990). A hemi-methylated chimaeric gene, containing the cauliflower mosaic virus 35S promoter, the ß-glucuronidase coding region and the poly-adenylation signal of nopaline synthase was introduced into tobacco protoplasts by polyethylene glycol mediated transfection.

Hemimethylation led to complete inhibition of transient gene expression. In regenerated transgenic plants the integrated gene was constitutively hypermethylated at the sequences CG and CNG and this was correlated with an inactivation of ß-glucuronidase in 12 out of 18 plants analysed. Shoots regenerated from 5-azaC treated calli revealed stable enzyme restoration and demethylation of the integrated transgene.

1.6.4.1 Homology-dependent gene silencing and methylation.

In work carried out by Linn *et al.*, (1990) and Matzke and Matzke, (1994) both have related transgene inactivation in petunia and tobacco to involve transcriptional inactivation and methylation of repeated genes. Both linked and unlinked repeated genes are observed become methylated and silenced (*cis*- and *trans*-inactivated respectively). The possibility that a methylated transgene allele or locus can impose by means of a DNA-DNA pairing a hypermethylated state on a homologous, undermethylated allele or an unlinked locus have also been shown (Meyer *et al.*, 1993, Matzke *et al.*, 1993). Analysis of such homology dependent or repeat induced gene silencing (RIGS) (Assad *et al.*, 1993) in plants has revealed intriguing parallels to related phenomena in some fungi. The first discovered instance of this phenomenon, repeat induced point mutation (RIP), occurs in the filamentous fungus *Neurospora crassa* (Selker, 1990). RIP is a DNA pairing dependent process in which duplicated sequences become riddled with point mutations in the form of C-T transitions and also become methylated. RIP presumably exists to increase sequence divergence between repeated sequences. This reduces the possibility of ectopic recombination, an event that could lead to deleterious genome rearrangements in the *Neurospora* genome which is otherwise deficient in repetitive DNA. Selker *et al.*, (1993) described the initial events in RIP as cytosine methylation followed by conversion of 5-methyl cytosine to thymine, however methylation minus mutants are still subject to RIP, such that point mutations appear in the absence of methylation. It should be noted that certain other fungi inactivate duplicated sequences by a reversible methylation process that is not associated with point mutations; the so called MIP (Methylation induced pre-

meiotically) phenomenon has been described in *Ascobulus*, another filamentous fungi (Rossignol and Faugeron, 1994), and in a basidomycete, *Coprinus cinereus* (Freedman and Pukkila, 1993).

The extent to which gene silencing in plants is mechanistically related to RIP is not known. In that RIP occurs pre-meiotically in a few haploid cells during the sexual phase of *Neurospora* whereas homology dependant silencing events in plants take place in diploid (or polyploid) somatic cells. Furthermore unlike RIP mutations, point mutations do not seem to occur to any great degree with methylated plant genes, as demonstrated by the reversibility of silencing (Matzke *et al.*, 1994). The plant mechanism may therefore more closely resemble MIP than RIP. It is suggested that such mechanisms in plants may have evolved to help silence the many copies of transposable elements thus reducing their detrimental and mutagenic effects on the genome.

1.6.4.2 Paramutation.

Accumulating evidence to date suggests that the behaviour of transgenes in which one gene induces an epigenetic change in another is strongly reminiscent of paramutation phenomenon where differences in DNA methylation are also found to be associated. In plants the term paramutation was coined by Brink, (1973) who discovered and characterised the phenomenon at the maize R locus, which encodes a transcription factor that is involved in the anthocyanin biosynthetic pathway. Paramutation refers to an allelic interaction in which the activity of a sensitive allele [the 'paramutatable' allele] is heritably weakened following an interaction with an inducing [or 'paramutagenic'] allele in a heterozygote. Paramutation violates Mendel's first law, which states that alleles exert no effect on one another once they have segregated in progeny, because the weakened, or paramutated, allele remains weakened for several generations after crossing out the inducing allele. The possibility that one allele could influence and bring about a genetic change of the other allele in heterozygotes was recognised as early as 1930 by Winkler, and is known to occur in a large number of plant species, including

Lycopersicon esculentum, Zea mays, Antirrhinum majus, Pisum sativum, Malva parviflora, Celosia crista, and certain varieties of *Prunus, Fragaria* and *Oenothera.* (Meyer *et al.*, 1993). However, the molecular mechanisms responsible for paramutation phenomena remains unclear. Paramutation has been most extensively studied for the R locus (Brink, 1973) and the B locus (Coe, 1966) of *Zea mays*. The reduced pigmentation of the *r* gene following its interaction with the paramutatable alleles appears to involve methylation and a paramutagenic allele that contains multiple copies of the *r* gene (Patterson and Chandler, 1995). However no changes in DNA methylation have been detected in paramutated alleles of the maize *b* gene (Patterson *et al.*, 1993). Other differences between paramutation of *b* and *r* relate to penetrance and stability of the paramutated allele. These genes, although homologous, differ in their structure with the *r* gene having a complicated pattern of duplicated sequences not seen at the B locus. It has been proposed (Patterson *et al.*, 1993) that paramutation at the B locus results from an allelic interaction which promotes a heritable change in expression state similar to heritable transcription states which occur in the absence of DNA methylation in Drosophila (Paro, 1990) and yeast (Laurenson and Rine, 1992).

Matzke and co-workers (1993) showed that methylation and inactivation of the aforementioned target locus *K* could be reversed after crossing out the silencing locus *H*, however was observed to occur gradually over several generations. It was therefore suggested that the silencing locus *H* causes a heritable alteration in the expression of the target locus *K*. These interlocus effects are similar to the directed heritable (epigenetic) changes that occur when certain alleles interact, in paramutation. Similarly Meyer and Heidmann (1994) have discovered paramutagenic-like versions of the aforementioned maize A1 gene inserted into petunia under the control of the *CaMV* promoter. When the hypermethylated A1-allele of a single copy transformant was crossed with a homologous, but hypomethylated allele, the hypomethylated allele became methylated and silenced in a semi-dominant way. The F1 plants which received a hypomethylated A1 allele from one parent and a hypermethylated allele from the other showed a significant reduction in A1 expression correlating with hypermethylation of the

promoter region. It was suggested that the hypermethylated allele has paramutagenic potential, inducing methylation in the paramutatable A1 allele that had been previously hypomethylated. After both alleles had segregated form each other, the paramutated A1 allele remained hypermethylated over the next two generations with occasional weak reactivation in a few cells. Homology dependent silencing in plants does not always appear to be associated with DNA methylation of the suppressed genes (Mol *et al.*, 1991, Hart *et al.*, 1992, Goring *et al.*, 1991). Jorgensen, (1995) described co-ordinate silencing (co-suppression) of chalcone synthase (*chs*) transgenes and endogenous genes in petunia. This was a post-transcriptional process involving RNA turnover (Van Blokland *et al* ., 1994) that appears not to be associated with extensive methylation. The *chs* system in petunia involves fascinating changes in corolla pigmentation patterns that must involve meristem-wide changes, such as alterations in the turnover of transcript. The different patterns have a epigenetic basis because they are both meiotically heritable yet also reversible. The *chs* system is complicated in that not only are the *chs* endogenous genes affected by expression of the transgene, but also the transgene itself can undergo paramutation- like changes. The meiotically heritable state imposed on a paramutatable allele is an example of an epiallele or an epimutation. The questions remaining to be answered is how does one gene impose an inactive meiotically heritable state on another and what modifications in DNA and/or chromatin structure does it entail.

Models for a mechanism underlying allelic *trans*-inactivation in *D.melanogaster* suggest the inactivation of a gene by direct contact with heterochromatic proteins (Dreeson *et al.*, 1991) and a direct transmission of the resulting chromatin structure (heterochromatinisation) between paired homologues (Henikoff and Dreeson 1989). However it is difficult to envisage the inheritance of chromatin structures as this would require the dismantling and subsequent reassembly of chromatin states during DNA replication. Because plants do not have a separate germ line, specific somatic methylation of genes could be inherited through meiosis as the maintenance methylase will ensure the inheritance of methylation on symmetrical cytosines during DNA

replication. Lund *et al.*, (1995b), described an epimutation of the P-locus in maize that changes during plant development. Changes in the *p* gene are marked both by demethylation of specific sites and by decondensation of chromatin in regions that are important for gene expression. This provides the first direct connection between chromatin structure at a plant locus and its DNA methylation state.

1.6.5 Stabilising transgene expression.

There is considerable evidence from many non-plant sources including Drosophila (Gasser *et al.*, 1989), chicken, mouse, HeLa cells and yeast that the organisation of DNA into loops is maintained by the periodic attachment of the DNA to a proteinaceous chromosome scaffold. How these scaffolds relate to the nuclear matrix is as yet unclear. In many eukaryotes, nuclear scaffold or matrix attachment regions (SARs/MARs) of DNA have been isolated and analysed for common features. These have often been found to lie in A-T rich regions and to contain sequences related to the Drosophila topoisomerase II consensus sequence. They may also contain sequences related to the T-box and A-box sequences identified by Gasser and Laemmli (1986) and to autonomously replicating sequences (ARSs). Several of the ARS-like regions have been shown to function as SARs/MARs *in vivo* . In some cases, co-mapping of SARs/MARs with transcriptional enhancers have been noted. All of these characteristics have led to the speculation that enzymes regulating torsional stress, proteins enhancing or silencing transcription, and even replication complexes may bind at the base of loops to exert their effects on each loop and therefore on the genes within the loop (Jackson, 1991). Some credence has been given to this speculation by another property of SARs/MARs, namely their ability when placed upstream and downstream of a reporter gene in transgenic organisms to alleviate commonly observed position effects on the expression of the reporter gene. Studies in animal systems have provided strong evidence that SARs/MARs can insulate transgenes from surrounding chromatin (Stief *et al.*, 1989, Phi-Van *et al.*, 1990). When SARs/MARs are included on both sides of the transgene the expression level becomes proportional to transgene copy

number, indicating that gene activity is independent of position in the chromosome. In a plant study Breyne *et al.,* (1992) characterised a plant SAR from the soya bean genome which flanks a seed specific lectin gene and four other genes with different tissue specificities. The p1-SAR was used to flank a GUS reporter gene at either end and was introduced into tobacco via *A. tumefaciens* mediated gene transfer. The results seemed to contradict the published animal data, in that only slight decreases in variability were noted and more surprisingly overall expression levels were unaffected or slightly decreased. However it is difficult to compare this work with others because no gene copy numbers were determined. Van der Geest and co-workers (1994) found that MARs flanking the ß-phaseolin gene caused a modest increase in overall expression and modest decrease in variability in expression of a reporter gene in tobacco plants transformed using *Agrobacterium.* The work of Schoffl and co-workers (1993) comes the closest to that obtained in the animal work. In this work a MARs isolated from a clone containing a soybean heat-shock gene was used to generate transgenic tobacco plants by *Agrobacterium* mediated transformation. The MARs resulted in 5- to 9- fold increase in expression of the GUS reporter gene, but no effect on the variability in transgene expression. There was however, at least partial copy-number dependence of expression. Mlynárová *et al.*, (1994, 1995) used an animal MAR chick lysozyme and a *gus* reporter gene in *Agrobacterium* mediated transformation of tobacco plants. They found a moderate increases in average transgene expression (4-fold) and a substantial decrease in transformant to transformant variability in transgene expression (8-fold decrease in variance of logarithmically transformed data). In contrast to the work of Schoffl *et al.*, (1993) no-copy number dependence in transgene expression was found. In work carried out by Allen *et al.*, 1993 and Spiker *et al.*, 1995, both a heterologous yeast and tobacco MARs were used to flank a *gus* reporter gene. Microprojectile bombardment was used to co-transform tobacco cells with the selectable marker and reporter gene on separate plasmids. The objective of this approach was to avoid the bias in favour of expression that may result from the close physical linkage of selectable marker and reporter genes in used in most *Agrobacterium* vectors. It was noted with

the yeast MARs an overall stimulation in transgene expression of 12-fold(24-fold on a per copy basis). Spiker *et al.*, 1995 using a tobacco MAR observed greater than 60-fold stimulation (nearly 140-fold on a per-copy basis). In both cases, the variability of transgene expression was reduced only slightly with no evidence for copy number dependence. Spiker *et al.*, (1995) also noted that transgene expression was inhibited by higher copy numbers. It was suggested that these results may indicate the involvement of "homology-dependent gene silencing".

1.7 Scope of this thesis.

It may thus be seen that expression of a foreign gene introduced into plants by the process of *Agrobacterium*-mediated transformation is not solely dependent on a successful integration event but rather depends to a greater or lesser extent on factors associated with or dependent on the number and location of integration sites within the genome. As a direct consequence of this extensive variation in levels of transgene expression are commonly observed. In basic research such variation between independent transformants is a major obstacle to precise comparisons between different gene constructs. In plant breeding programs inter-line variability necessitates the screening of large numbers of transformants such that the desired trait may be selected, and thus makes such breeding programs less cost efficient. The situation may also be further complicated should the introduced gene prove to be developmentally and/or genetically unstable. Despite a recent survey report (Finnegan and McElroy, 1994) revealing that, of the 30 companies engaged in the commercialisation of transgenic plants questioned, almost all reported some level of transgene inactivation few reports of such problems reach the literature.

This thesis is concerned with an investigation into the reasons underlying the observed variability described above. To this end populations of transgenic tobacco plants harbouring a known transgene will be analysed in order characterise the structure and number of gene copies integrated into the tobacco genome, the organisation of integration sites within this genetic background and the influence of integration site-

specific factors. These data will then be correlated with observed levels of transgene expression and an attempt made to determine the extent to which each of the above factors exerts an effect over transgene expression. While many workers seem willing to accept that transgene expression level variability is a mere 'fact of life' and compensate by undertaking analyses of increasingly large populations of transformants, attempts to optimise plant transformation techniques such that stable, determinable levels of expression are reproducibly obtained will continue to be hampered by a lack of understanding of the extent to which expression levels are dependent on the above factors, independently and in concert. It is hoped that this study will increase our awareness of these issues and serve to enhance our knowledge of the little explored phenomena underlying transgene expression in plants.

Chapter 2:
Materials and Methods.

2.1 Glassware and plasticware.

All glassware and plasticware used in experiments involving nucleic acids were siliconised and autoclaved at 121^{o}C, 15 p.s.i., for 20 minutes.

2.2 Chemicals and biological reagents.

Unless otherwise indicated all chemicals and reagents used were purchased from BDH Chemicals Ltd., Poole, Dorset, UK and were of analytical grade (AnaLar). Other materials were obtained from the designated sources listed below.

Bovine serum albumin (BSA), polyvinyl pyrolidone (PVP), herring sperm DNA, sodium iodide, xylene cyanol, bromophenol blue, ethidium bromide (EtBr), dithiothreitol (DTT), adenosine triphosphate (ATP), RNase A, pronase, proteinase K, lysozyme, sodium dodecyl sulphate (SDS), sodium lauryl sarcosinate, ampicillin, carbenicillin, gentamycin, rifampicin, isoamyl alcohol, dimethyl formamide (DMF), dimethyl sulfoxide (DMSO), micrococcal nuclease and deoxyribonuclease I were from Sigma Chemical Company, Poole, Dorset, UK.

Restriction endonucleases and DNA modifying enzymes were purchased from Northumbrian Biologicals Ltd., Cramlington, UK, Boehringer Mannheim GmbH, Mannheim, W.Germany and New England Biolabs, Beverly, Maryland, USA. dNTP's, random hexanucleotide primers and the digoxigenin DNA labelling and detection kit were also purchased from Boehringer and Mannheim.

3MM paper, and filter papers were from Whatman Ltd., Maidstone, Kent, UK. Nitrocellulose filters were from Schleicher and Schuell GmbH, Dassel, W.Germany. 'Hybond-N' nylon filters, radioactive ^{32}P-dCTP and ^{125}I conjugated donkey anti-

rabbit IgG antibodies were from Amersham International plc., Amersham, Oxon, UK. X-ray films (Fuji RX100) were from Fujimex, Swindon, Wiltshire UK.

Sephadex G-50, G-100, Ficoll-400, glycogen and percoll were purchased from Pharmacia Fine Chemicals, Milton Keynes, Bucks., UK.

Yeast extract was from Biolife s.r.l., Milan, Italy. Bacto-agar was from DIFCO, Detroit, Michigan, USA. Bactotryptone and beef extract were from Oxoid Ltd., Basingstoke Hertfordshire, UK.

Agarose was from BRL Ltd., Uxbridge, Middlesex, UK. Augmentin was from Beecham's Pharmaceuticals, Brentford, Middlesex UK.

Rabbit anti-pea lectin antibodies were obtained from Glyn Edwards. Present address: Shell Research Bioresearch Centre, Sittingbourne, Kent.

Rabbit anti-acetylated, and anti-unacetylated Histone H4 antibodies were received as a gift from D.C Allis. Present address: Dept. of Biology, Syracuse University, New York, USA.

2.3 Bacterial strains and plasmids.

All bacterial strains and plasmids used are described in table (2.1)

2.4 Bacterial culture media and antibiotics.

The following media were used for the growth of bacterial cultures:

- **LB-medium** (Luria-Bertani):- 10g/l bactotryptone; 5g/l bacto-yeast extract; 10g/l NaCl; and adjust to pH to 7.0 with NaOH
- **YEB-medium**:- 5g/l bactotryptone; 1g/l bacto-yeast extract; 5g/l bacto-beef extract; $MgSO_4$ to 2mM adjust to pH to 7.2 with NaOH; add 10ml 50%(w/v) sterile sucrose solution added after autoclaving.

For solid media, 15g/l agar was added before autoclaving. After autoclaving (121°C, 15 p.s.i.), and the media cooled to 55°C before antibiotics were added (see table 2.2).

2.5 Storage of bacteria.

Bacterial cultures were stored at 4°C for up to six weeks on inverted agar plates sealed with Nesco-film. For long term storage, bacterial lawns grown from single colonies on selective agar plates were transferred to sterile 2ml aliquots of a solution containing 60% L-broth and 40% glycerol mixed thoroughly by vortexing and stored at -80°C.

2.6 Plant material.

Nicotiana tabacum SR1 cv. Petit Havana (Maliga *et al.*, 1973). *N.tabacum* seeds were surface sterilised by immersing in 20% (v/v) sodium perchlorate for one hour with gentle agitation every fifteen minutes. Five washings with sterile distilled water were performed before transferring the seeds to MS salts agar in beatson jars to germinate and produce axenic plants for transformation.

2.7 Production of transgenic plants.

The method employed was based on the leaf disc method of Horsch *et al* (1985) with modifications.

A.tumefaciens (GV3101) harbouring either pGV3850::pDUB130 or pGV3850::pDUB133 co-integrates was grown overnight at 28°C in 50ml YEB-broth containing appropriate antibiotics with vigorous shaking. The cells were harvested by centrifugation and washed three times in 2mM $MgSO_4$, the cells were resuspended, reharvested and the supernatant discarded in each case. After the final wash the cells were resuspended in 50ml co-cultivation media(details of the constituents of all the media used are given in table (2.3), to give a cell density of approximately a 10^{10} cells/ml.

All plant material used in the transformation was obtained from tissue culture to obviate

Antibiotic	Abbreviation	*A.tumefaciens* µg/ml	*E.coli* µg/ml
Ampicillin	Ap	-	50
Carbenicillin	Cb	100	-
Gentamycin	Gm	10	2.5
Kanamycin	Km	-	20
Rifampicin	Rif	100	-

Table 2.2:

Antibiotic concentrations used in bacterial culture media. Concentrations used in the selection of bacterial strains harbouring recombinant plasmids are shown.

Ampicillin, carbenicillin, gentamycin and kanamycin were each dissolved in sterile, distilled water to the appropriate concentration and filter-sterilised prior to use.

Stocks of rifampicin were made by dissolving the antibiotic to the desired concentrationn in 70% methanol.

All antibiotic stock solutions were stored at -20^{o}C and defrosted thoroughly before use.

Bacterial Strains *E.coli*	Genotypes/Characteristics	Reference/Source
DH5α	F-,recA1,endA1,supE44,thi hsdR17	R.Croy[2]
Agrobacterium		
GV3101	(C58C),Ti-cured,Rif^{r}C58 derivative	Glyn Edwards(Ph.D)[1]
Plasmids		
pBR322	Ap^r,Tc^r,ColE1 replicon bom $^+$.	R.Croy[2]
pUC18/19	Ap^r, lacZ cloning region from M13mp18/19	R.Croy[2]
pNOS-NPT	*nos* promoter/*nptII* gene fusion in pUC9	Glyn Edwards(Ph.D)[1]
pDUB80	pUC18 containing pea *lecA* gene	R.Croy[2]
pCaMV	800 bp 35S promoter from ROK2(a derivative of ROK1) in pUC18	Glyn Edwards(Ph.D)[1]
pDUB133	1020 bp Tobacco *ssRubisco* promoter in pUC18	R.Croy[2]
pDUB1104	pBR322 containing RB & LB *HindIII* fragments 10& 23	C. Shaw[2]
pGV3850	pGV3100,pBR322 replacing T-DNA between *HindIII* fragments 23-10	Zambryski *et al* (1983)
pGV3850::pDUB130	*npt(II)*,*CaMV*35S-*lecA*, replacing T-DNA between *HindIII* fragments 23-10	Glyn Edwards(Ph.D)[1]
pGV3850::pDUB133	*npt(II)*,Tobacco *ssRubisco-lecA*, replacing T-DNA between *HindIII* fragments 23-10	Glyn Edwards(Ph.D)[1]

Table 2.1:

Escherichia coli & *Agrobacterium tumefaciens* strains & plasmids

Key: Ap^r: ampicillin resistance; Tc^r: tetracycline resistance; *RecA*: recombinant deficient; *hsdR*: endonuclease R host restriction activity; *supE*: supressor of amber mutations; bom^+: mobilisation site; LB & RB: left border and right hand border sequence; [1]: Present address - Shell Research, Bioresearch Centre, Sittingbourne, Kent; [2]: Present address - Dept. of Biological Sciences, University of Durham.

the need for surface sterilisation. Leaf squares ($1cm^2$) were floated on the surface of the bacterial culture for 10 minutes and then incubated on co-cultivation agar for 48h at 25°C in tissue culture with an 18h photoperiod (6,000 Lux). The leaf squares were then washed overnight in explant medium with gentle shaking, four squares were washed in 30ml of media. The squares were blotted on sterile filter paper, and placed on selective agar plates. The plates were sealed and incubated under the conditions previously described.

Shoots obtained were excised after four to six weeks at the second or third internode. The base of the stems were cut at an angle to maximise the surface area available for rooting and placed a few mm in rooting agar (40ml) in 250ml beatson jars.

2.8 Growth of transgenic plants.

Rooted plants which survived the kanamycin selection after four weeks were potted in clean four inch pots, two-thirds filled with autoclaved (2h,121°C) 50:50 compost and perlite mixture, together with some agar around the root system. The pots were then sealed in thin clear polythene bags, secured round the rim by an elastic band, and transferred to the containment growth room. All plants were grown at 25°C with an 18h photoperiod (6,000 Lux). Plants were watered every two days with MS salts until established and the plants repotted every four weeks.

2.9 Biochemical isolation and purification of nucleic acids.

2.9.1 Deproteinisation of nucleic acid samples using phenol.

Phenol for deproteinising nucleic acids was obtained by the distillation of phenol crystals under nitrogen, collection and storage under TE (pH 7.2) at -20°C.

Solutions of DNA were deproteinised by two successive extractions with phenol-chloroform-isoamyl alcohol, henceforth referred to as "phenol". 1.5 volumes of phenol were added to the DNA sample and mixed by vortexing. The aqueous phase and

phenolic phase were separated by a brief centrifugation in a microcentrifuge. The upper phase was transferred to a fresh tube and the phenol extraction repeated. When extracting small amounts of DNA, the phenolic phase was back extracted with an equal volume of TE buffer, and the resultant aqueous phase was combined with the original aqueous phase. Phenol extractions were followed by two extractions with an equal volume of chloroform-isoamyl alcohol (24:1 v/v) to remove the remaining traces of phenol. In some cases, when the DNA sample was very viscous, the remaining traces of phenol were removed by three to four extractions with an equal volume of diethyl ether. Traces of diethyl ether were then removed by aspirating the sample with nitrogen for 5 min. After deproteinisation, DNA was recovered by alcohol precipitation.

2.9.2 DNA precipitation using alcohols

0.1 volumes of 3M sodium acetate pH 5.2 and two volumes of ethanol were added to the DNA solution and kept at -70°C for 30 min, or at -20°C overnight. For precipitating DNA samples of concentrations lower than 200ng/ml, 1 to 2µg of glycogen per 10ml total volume was added as a carrier. The precipitated DNA was pelleted by centrifugation at 12,000 x g for 10min (MSE Microcentaur microcentrifuge) for small samples, or at 25,000 x g (Beckman JS-21 high speed centrifuge) for larger samples. The pellet was washed twice in 70%(v/v) ethanol, dried briefly under vacuum and then redissolved in a small volume of sterile water or TE (10mM Tris-HCl, pH7.4, 1mM EDTA). To minimize the volume of solution to be centrifuged, isopropanol was sometimes used instead of ethanol. In these cases, 0.6-1.0 volume of isopropanol was added to the DNA solution and the mixture was kept at -20°C for 30 minutes.

2.9.3 Analysis of nucleic acids by UV absorbance.

In order to determine the concentration of nucleic acids in DNA, a small sample (1µl to 5µl) was added to a 1cm quartz cuvette containing 1ml of sterile water or TE buffer. The solution was thoroughly mixed and the absorbance (A) was measured at a wavelength of 260nm against a blank of sterile water or TE buffer. An A_{260} of 0.02

Co-cultivation media	1xMS medium (4.71g/l) Murashige & Skooge (1962),2%(w/v) sucrose, 1mg/ml BAP, 0.1mg/ml NAA, pH to 5.8 with HCl hereafter called MS2SBN.
Washing medium	as MS2SBN with added anti-*A.tumefaciens* antibiotics (400µg/ml) Augmentin, 200(µg/ml) Kanamycin.
Selection medium	as MS2SBN, 0.8%(w/v) agar, (200µg/ml) Kanamycin
Rooting medium	0.5x MS, 0.8%(w/v) agar, (100µg/ml) Kanamycin
Seed germination media	0.5x MS, 0.8%(w/v) agar

Table 2.3:

Plant tissue culture media constituents.

All constituents, excluding antibiotics, were dissolved in approximately 800 ml distilled water and the pH adjusted appropriately before making the solution up to a final volume of 1 l with distilled water. Media was dispensed into aliquots and autoclaved at 121^{o}C, 15 p.s.i. Antibiotics, if required, were added immediately before use.

corresponds to a DNA concentration of 1µg/ml. An A_{260} of 0.024 corresponds to an RNA concentration of 1µg/ml.

Analysis of the purity of the DNA was estimated from a spectrophotomeric scan from 200nm to 300nm by calculating the A_{260}/A_{280} ratio, whereby ratios for pure DNA were 1.8.

2.9.4 Fluorimetric analysis of DNA using diaminobenzoic acid (DABA).

Estimation of DNA concentration was determined by performing fluorimetric analysis using diaminobenzoic acid (DABA) as described by Thomas and Farquar (1978). 1 to 5µl samples of DNA solution to be analysed were placed in 1.5ml eppendorf tubes. Samples greater than 5µl were first dried under vacuum. 20µl DABA solution (400mg/ml in H_20) was added to each tube, mixed, centrifuged briefly, and incubated at 60°C for 30 minutes. 1.6ml of 1M HCl was added to each tube, mixed thoroughly, and the relative fluorescence was measured on a Baird-Atomic Fluoripoint fluorimeter at 405nm excitation wavelength and 505nm emission wavelength. A standard curve was constructed using lambda DNA standards containing 20ng to 1µg of DNA. The concentration of DNA was then estimated using the standard curve.

2.9.5 Preparation of competent *E.coli* cells.

For cloning and maintaining stocks of plasmids competent *E.coli* cells were prepared for transformation experiments by one of the two methods described below:

2.9.5.1 RbCl/$CaCl_2$ Method

5ml of LB medium were inoculated with a single colony of DH5α or other desired bacterial strain and shaken at 37°C for 2-3h. This starter culture was then used to inoculate 100ml of pre-warmed LB medium in a 1L conical flask which was incubated at 37°C with vigorous shaking until the OD_{550} reached 0.35. The culture was transferred to sterile 50ml capped centrifuge tubes and chilled on ice for 15min. The

cells were then centrifuged for 5min at 5,000 x g at 4°C. The supernatant was discarded and the cells resuspended in 10.5ml of ice cold TFB1 (100mM RbCl, 50mM $MnCl_2$, 30mM KOAc, 10mM $CaCl_2$, 15% (v/v) glycerol). Both suspensions were pooled, incubated on ice for 90 minutes and the cells pelleted as before. The supernatant was decanted and the cells gently resuspended in 2.8ml of ice cold TFB2 (10mM MOPS pH7.0, 10mM RbCl, 75mM $CaCl_2$,15% (v/v) glycerol). 100-200µl of competent cells were used per transformation reaction. For long term storage 200µl aliquots in microfuge tubes were flash frozen in liquid nitrogen and kept at -80°C.

2.9.5.2 Polyethylene/dimethylsulfoxide method.

The method described was essentially that of Chung *et al.*, (1989). 5ml of LB medium were inoculated with a single colony of DH5α or other required strain and incubated at 37°C for 2-3h with vigorous shaking. This starter culture was then used to inoculate 50ml of prewarmed LB medium and the bacteria allowed to grow at 37°C with vigorous shaking until an OD of 0.4 was reached at 650nm. the cells were pelleted by centrifugation 5,000 x g for 5min. The supernatant was discarded and the pellet resuspended in 1/10 th of the original volume of the culture using ice cold sterile transformation and storage solution buffer pH6.5 (LB-medium containing 10% (w/v) PEG8,000, 50mM $MgCl_2$ and 5% (v/v) DMSO; pH6.5). 200µl of competent cells was used for every transformation reaction. For long term storage, 200µl aliquots of competent cells in microfuge tubes were frozen on dry ice and stored at -80°C.

2.9.6 Small-scale preparation of plasmid DNA.

The modified method of Birnboim and Doly (1979) was used in the preparation of small quantities of plasmid DNA. 10ml aliquots of LB medium in universal bottles with appropriate antibiotics added were each inoculated with bacteria from single colonies using sterile tooth picks, and grown overnight at 37°C on a rotary shaker. The cells were pelleted by centrifugation at 4,200rpm (5,000 x g) for 10 minutes and resuspended in 200µl of lysis buffer (50mM glucose, 25mM Tris-HCl pH8.0, and

10mM EDTA pH8.0) containing 4mg/ml lysozyme (freshly added to the buffer). The cell suspension was transferred to a 1.5ml Eppendorf tube and placed on ice for approximately 30 minutes. 400µl of freshly prepared alkaline-SDS (1% (w/v) SDS,0.2M NaOH) was added to the tube, mixed gently by inverting the tube several times and kept on ice for a further 15 minutes with occasional mixing. 300µl of cold 3M sodium acetate, pH4.6 was added, mixed thoroughly and incubated on ice for 10 minutes. The sample was centrifuged in a microcentrifuge at 12,500rpm (12,000 x g) for 10 minutes. The supernatant was phenol extracted once, chloroform extracted twice and precipitated with two volumes of cold ethanol. The pellet was washed with 70% ethanol, vacuum dried and resuspended in T.E buffer (pH7.2). The plasmid DNA was stored at -80°C. 1 to 2µl RNaseA (10mg/ml), previously heat treated by boiling for 20 minutes to inactivate DNases, was normally added during restriction digestions of the plasmid DNA isolated by the above method.

2.9.7 Large-scale preparation of plasmid DNA.

The method used was adapted from the Promega catalogue of "Protocols and Applications" (1990), allowing for the rapid isolation of plasmid DNA without the need for banding in caesium chloride gradients. The plasmid containing strain of *E.coli* was grown overnight in 250ml of L-broth containing the appropriate antibiotics, at 37°C with vigorous shaking. The cells were harvested by centrifugation at 5,000 x g for 15 minutes at 4°C, the pellet resuspended in 6ml of freshly prepared lysozyme buffer (25mM Tris-HCl, pH7.5, 10mM EDTA, 15%(w/v) sucrose, 2mg/ml lysozyme) and placed on ice for 20 minutes. To this 12ml of 0.2M NaOH, 1%(w/v) SDS was added, mixed carefully by inversion and incubated on ice for 10 minutes. 7.5ml of ice cold 3M sodium acetate, pH 4.6 was added to the lysate, mixed thoroughly and the suspension placed on ice for 10 minutes, followed by centrifugation at 10,000 x g for 15 minutes. The supernatant was recovered, to which 50µl of RNase A (1mg/ml stock) was added and incubated for 20 minutes at 37°C. After incubation the supernatant was extracted with one volume of phenol/chloroform by vortexing for 5 minutes and centrifugation at

10,000 x g for 10 minutes. The upper aqueous phase was saved and the phenol extraction repeated. The upper phase was then extracted with chloroform/isoamyl alcohol (24:1) by vortexing for 5 minutes and centrifugation at 10,000 x g for 10 minutes.The upper aqueous phase was then transferred to a fresh tube and the DNA precipitated with two volumes of ethanol at -20°C for 30 minutes. The precipitate was recovered by centrifugation at 10,000 x g for 20 minutes. The pellet obtained was dissolved in 1.6ml of sterile water, mixed with 0.4ml of 4M NaCl, 2ml of 13% (w/v) polyethylene glycol (PEG 6000) and incubated on ice for 60 minutes. The PEG precipitated plasmid DNA was pelleted by centrifugation at 10,000 x g for 10 minutes, washed in 70%(w/v) ethanol, dried under vacuum and finally dissolved in sterile water or TE buffer.

2.9.8 *A.tumefaciens* total DNA preparation.

The method used was essentially that of Herrera-Estrella and Simpson, (1988). *A. tumefaciens* strains were grown in 5ml YEB-broth containing the appropriate antibiotics for 24-48h at 28°C with vigorous shaking. 1.5ml of culture was centrifuged for 30s at 12,000 x g in an Eppendorf tube and the cells resuspended in 380µl pronase buffer (50mM Tris-HCl pH8.0, 20mM EDTA, 0.8% Na-lauroyl sarcosinate). 20µl of a 20mg/ml solution of pronase was added and the suspension incubated at 42°C for 1 hour. The resulting lysate was sheared by passing it through a Pasteur pipette 4-5 times. The sheared lysate was deproteinised twice by extraction with phenol followed by four extractions with diethyl ether. The DNA was precipitated with alcohol, redissolved in 200µl TE, reprecipitated, and washed twice in 70% (v/v) ethanol. The DNA pellet was dissolved in 100µl TE and stored at -70°C.

2.9.9 Extraction of high molecular weight DNA from leaf tissues.

Total genomic DNAs were isolated from plant tissues according to the procedure described by Shirsat *et al.*, (1989). Briefly, 5g of leaf tissue material was frozen in liquid air and ground in a cold mortar and pestle to a fine powder. 5ml homogenising

buffer(0.1M NaCl, 0.025M EDTA, 2% (w/v) SDS), 2.5ml 5M sodium perchlorate, 5ml redistilled phenol, and 5ml 1%(v/v) octanol/chloroform mixture was added sequentially with thorough mixing. The resulting suspension was transferred to a sterile 100ml conical flask and was shaken on a rotary shaker (100rpm) for 1.5h at 4°C. The suspension was then centrifuged in corex tubes for 10 minutes at 10,000rpm at 4°C to separate the phases. The aqueous phase was removed and extracted once with an equal volume of octanol/chloroform mixture. The DNA was then recovered by spooling with a glass hook using two to three volumes of cold (-20°C) 100% ethanol gently layered over the supernatant. The spooled DNA was gently resuspended in 1ml of TE buffer overnight at 4°C. 25µl of stock pronase (25mg/ml; previously self-digested for 2h at 37°C) was added and the mixture incubated at 37°C for 2h. The DNA was then purified by centrifuging twice, if necessary, on caesium chloride/ethidium bromide gradients using a Sorvall OTD65B Ultracentrifuge. The band containing the DNA was collected, extracted with caesium chloride-saturated isoamyl alcohol to remove the ethidium bromide and dialysed against TE buffer for 48h at 4°C with several buffer changes. Dialysed DNA was then concentrated by ethanol precipitation (overnight at -20°C), pelleted, redissolved in TE buffer, and the concentration and purity determined using a spectrophotometer as described in section 2.9.2. All extracted DNAs were tested for restrictability using various restriction enzymes followed by analysis on agarose gels. Samples of unrestricted DNAs were also analysed in the experiment to check the size and quality of the isolated DNA.

2.9.10 DNA restriction.

DNA saamples were digested with type I restriction endonucleases according to the manufacturers recommendations. Buffers for the restriction enzymes were either supplied by the manufacturers or prepared as described in table 2.4.

Generally, the enzymes were used at a concentration of 2.5u/µg DNA and incubated at temperatures recommended by the manufacturers for 1-3h. Many of the enzymes have been shown to work adequately at different NaCl concentrations (Sambrook *et al.*,

Table 2.4:

Buffers for restriction endonucleases.

All components were added from 1M stock solutions and combined in sterile, distilled water. Aliquots of 1 ml were stored at -20°C and thawed completely before use.

Buffer (10x)	Components (mM)			
	Tris-HCl pH 7.5	$MgCl_2$	DTT	NaCl
Low Salt	10	10	1.0	-
Medium Salt	10	10	1.0	50
High Salt	50	10	1.0	100

1989), and hence multiple digestions were usually performed simultaneously in the same buffer.

2.9.11 Agarose gel electrophoresis of DNA.

The methods used for the preparation of gels and subsequent electrophoresis of DNA were as described by Maniatis *et al.*, (1982). Briefly, the required amount of agarose [0.5-1.5%(w/v] depending on the size of DNA to be resolved, was added to tris-acetate EDTA (TAE) buffer (40mM tris acetate pH7.2, 2mM EDTA) and boiled until the agarose dissolved. The solution was cooled to 50-60°C and EtBr added to a final concentration of 0.5µg/ml. The solution was allowed to set in a perspex mould (190x 150 x 6mm) adhered to a glass plate using silicone grease, and containing a suitable well former. The formed gel was transferred to an electrophoresis tank and TAE buffer added to a level 1-2mm above the surface of the gel. The DNA samples containing 20% (v/v) loading buffer (0.25%(w/v) bromophenol blue, 0.25%(w/v) xylene cyanol, 30%(v.v) glycerol, 10mM EDTA) were loaded into the wells and the electrophoresis performed, usually at 50mA (30V) overnight. The DNA was visualised under 300nm UV light and photographed with a polaroid MP-4 land camera through a Kodak 23A Wratten filter, using Polaroid type 667 film.

Minigel electrophoresis was used to monitor the progress of restriction digests and to estimate DNA concentrations, due to its rapidity and ability to detect small amounts of DNA. These were essentially the same as the gels described except that a 100x80x5mm gel could be cast directly into the tank and tris-borate EDTA (TBE) buffer (0.089 tris-borate, 0.089 boric acid, 2mM EDTA) was used. Minigels were run for 30 min to 1h at 50V (40mA).

2.9.12 Southern transfer of DNA fragments onto filters.

Southern transfers were carried out essentially according to the procedure of Southern (1975). After photography, the gel was treated for blotting by gentle agitation in denaturing solution (0.4M NaOH, 0.8M NaCl), for 1h, followed by neutralising

solution (0.5M Tris-HCl, 1.5M NaCl, pH7.5) for 1h, and 20xSSC (0.3M sodium citrate, 3M NaCl) for 30 minutes or until blotted. For genomic digests, the gel was first depurinated with 0.25M HCl for 15 to 30 minutes prior to denaturation to cleave large fragments of DNA to facilitate their transfer to nitrocellulose or nylon filters. The DNA fragments were transferred in 20 x SSC onto nylon filters "Hybond-N" (Amersham) or nitrocellulose filters (0.45µm pore size, Schleicher and Schuell) either by capillary blotting overnight or by vacuum blotting for 2-3h using a Hybaid "Vacu Aid" vacuum blotter (Hybaid Ltd., Teddington, Middlesex, UK.), both at room temperature. Capillary blotting was performed by placing the gel on an apparatus consisting of a tray containing a glass plate on a platform overlaid with 2 layers of oversized 3MM paper acting as wicks on all sides of the glass plate and filled up to a level 1cm below the top of the glass plate with 20x SSC. A nylon/nitrocellulose filter presoaked in 20x SSC was carefully placed over the gel and several layers of presoaked 3MM paper were overlaid on top of the filter. Finally four layers of disposable nappies cut to an appropriate size were placed on the 3MM paper and a 1Kg weight on a glass plate. After the transfer, the position of the wells and the orientation of the gel were marked on the filter before lifting it off the gel and placed to dry on 3MM paper.

Transferred DNA was fixed to nylon or nitrocellulose filters by baking for 1.5h at 80°C in a vacuum oven sandwiched between two glass plates.

2.9.13 Recovery of DNA from agarose gels.

Two methods were used for isolating DNA fragments from agarose gels.

2.9.13.1 "Freeze-squeeze" method.

The restricted DNA sample was electrophoresed on an agarose gel and the required fragment was cut out using a clean scalpel blade and placed into a 1.5ml Eppendorf. The gel piece was equilibrated in 1ml of 1x TEA (40mM Tris-HCl pH8.0, 1mM EDTA, 20mM sodium acetate) for 15 minutes. The gel piece was then transferred to a small Eppendorf tube (0.5ml) containing a plug of siliconised glass wool and a small hole in

the base. The small Eppendorf tube was placed inside a large Eppendorf tube (1.5ml) and the whole assembly frozen at -80°C. Once frozen the assembly was spun in a haematocrit microcentrifuge for 15 minutes. The expressed solution containing the DNA was collected in the large Eppendorf tube. The solution was extracted with an equal volume of phenol followed by four extractions with chloroform. Each time the aqueous layer was back extracted with TE buffer (pH8.0) and added to the total. The aqueous layer was mixed with 5µl of 1M $MgCl_2$ and 20µl of 3M sodium acetate pH4.8. Two to three volumes of cold ethanol (-20°C) was added and the DNA precipitated at -80°C for 1h. The DNA pellet was washed twice in 80% ethanol, vacuum dried and resuspended in sterile water or TE buffer.

2.9.13.2 Electroelution method.

The DNA was digested with the appropriate enzymes, electrophoresed and a small piece of gel containing the required fragment of DNA visualised on a UV-transilluminator was excised. The gel piece was put into a dialysis tubing secured at one end with a dialysis clip. 0.5ml of sterile TBE electrophoresis buffer was added, and after securing the open end of the tubing with another clip, the DNA fragments were electroeluted in TBE buffer using a minigel tank by electrophoresing at 60V (50mA) for 20 min or until the DNA was visible as a thin line on the side of the tubing under UV-illumination. The polarity of the current was then reversed for about 1 minute to detach the eluted DNA form the dialysis tubing. The gel piece was carefully removed from the tubing and the buffer containing the eluted DNA fragment was transferred to a 1.5ml Eppendorf tube. The DNA was extracted with phenol and chloroform and finally precipitated with ethanol.

2.9.14 Radioactive and non-radioactive labelling of DNA probes by random priming.

DNA fragments isolated by one of the methods listed above were labelled for use as hybridisation probes using the following protocols.

2.9.14.1 Non-radioactive labelling of probes using DIG-dUTP.

Digoxigenin-labelled DNA probes were generated with *E.coli* polymerase I (holoenzyme, Klenow fragment) by the random priming method of , Feinberg and Volstein (1984). The following was added to an Eppendorf tube (1.5ml) on ice; 10ng-300ng of linearised purified DNA previously denatured for 10 minutes at 95°C and quickly chilled on ice for 3min; 2µl of 10x hexanucleotide mix (hexanucleotides 90 OD units/ml in TE buffer); 2µl of 10x dNTP labelling mix (1mM each dATP, dCTP, dGTP, 0.65mM dTTP; 0.35mM DIG-dUTP, pH6.5). The volume was then made up to 19µl with sterile water. The random priming reaction was initiated using 1µl(2u) of Klenow enzyme producing a final reaction volume of 20µl. The reaction was centrifuged briefly and incubated overnight at 37°C to increase the yield of labelled DNA. The reaction was terminated by the addition of 2µl 0.2M EDTA, pH8.0. The labelled DNA was precipitated with 2.5µl 4M LiCl, 1µl glycogen as a carrier and 75µl of prechilled ethanol (-20°C). The tube was mixed well and placed at -70°C for 30 minutes. The DNA precipitate was pelleted by centrifugation in a microfuge (12,000 x g) for 10 minutes. The resultant pellet was washed with 50µl of cold 70%(v/v) ethanol, centrifuged as before, dried under vacuum and dissolved in 50µl of TE buffer, pH8.0. for up to 30 min at 37°C.

2.9.14.2 Radioactive labelling of DNA probes using ^{32}P-dCTP.

The method used was that of Feinberg and Vogelstein (1984) which allowed DNA fragments <10ng to be labelled to high specific activities of 10^9 cpm/µg or more. Double stranded DNA was first denatured by boiling for 3 minutes and stored briefly on ice for one minute. The labelling reaction was carried out by mixing the following components in the order given:

(i) Denatured DNA xµl (10ng)

(ii) Oligo-labelling buffer(OLB) 10µl

(iii) Nuclease free BSA(10mg/ml) 2µl

(iv) Sterile water to 50µl

(v) ^{32}P-dCTP, 50µCi at 410 Ci(mmol)

(vi) Klenow polymerase (2U)

The mixture was incubated at 37°C for 3h (30ng DNA), 5h (10-30ng DNA), or overnight (<10ng DNA). The reaction was stopped by adding 200µl of 20mM NaCl, 20mM Tris-HCl, pH7.5, 2mM EDTA, 0.25% SDS, 1mM dCTP, labelled DNA was separated from unincorporated ^{32}P-dCTP by gel filtration using a sephadex G-50 column with column buffer consisting of 150mM NaCl, 50mM Tris-HCl pH7.5, 10mM EDTA and 0.1%(w/v) SDS.

Oligo-labelling buffer(OLB) was prepared by mixing the following solutions listed below A,B and C in the ratio 100:250:150 and stored at -20°C.

- Solution O: 1.25M Tris-HCl pH8.0, 125mM $MgCl_2$, stored at 4°C.
- Solution A: 1ml solution O + 18µl ß-mercaptethanol + 4µl each of dATP, dGTP, and dTTP. Each nucleotide is made as a 0.1M stock in 3mM Tris-Hcl pH7.0, 0.2mM EDTA, and stored at -20°C.
- Solution B: 2M HEPES-NaOH,pH6.6,stored at 4°C.
- Solution C: Hexadeoxyribonucleotides (P-L Biochem N 2166) at 90 OD units/ml in 10 mM Tris-HCl, 1mM EDTA, pH7.5 and stored at -30°C.

2.9.15 Hybridisation with Digoxigenin-11-dUTP DNA probes.

Southern blots on nitrocellulose filters were prepared for hybridisation according to the following protocol of Boehringer Mannheim catalogue (1990). Filters were prehybridised in a siliconised glass hybridisation vessel Techne with 20ml of hybridisation solution (5xSSC), 1%(w/v) N-lauroylsarcosine-Na salt, 0.02%(w/v) SDS, 1%(w/v) blocking solution vial 11 Boehringer and Mannheim kit) per 100cm^2 of filter at 68°C overnight. After overnight incubation the isolated digoxigenin-DNA probe was denatured for 10 minutes, cooled in an ice bath for 2 minutes and then added to the prehybridisation solution and incubated overnight for hybridisation at 68 °C.

After hybridisation the probe was recovered for later use and stored at -20°C. The filters were washed twice for 5 minutes each at room temperature with 50ml 2xSSC, 0.1%(w/v) SDS, once for 10 minutes at 68°C with 1xSSC, 0.1%(w/v) SDS and for high stringency twice for fifteen minutes each with 0.1xSSC, 0.1%(w/v) SDS at 68°C. Filters were then used directly for detection.

2.9.16 Detection of DIG-DNA probe target DNA hybrids.

After hybridisation, nitrocellulose filters were processed for detection of DIG bound probe by washing briefly in buffer 1,(100mM Tris-HCl, 150mM NaCl, pH7.5) for 1 minute. The filters were then equilibrated in 100ml buffer 2, (1%(w/v) blocking reagent; vial 6 Boehringer and Mannheim kit) prepared in buffer 1, for 30 minutes. 150u of anti-digoxigenin alkaline phosphatase F_{ab} fragments in 50 mM triethanolamine buffer (3mM NaCl, 1mM $MgCl_2$, 0.1mM $ZnCl_2$, 0.1%(w/v) BSA, pH7.6) was diluted to 10ml in buffer 2 (1:5,000). After blocking the filters for 30 minutes buffer 2 was replaced with 10ml of the diluted antibody-conjugate solution for a further 30 minutes. After incubation, unbound antibody-conjugate was removed by washing twice for 15 minutes each time with 100ml of buffer 1. The filter was then equilibrated in alkaline buffer 3 (100mM Tris-HCl; 100mM NaCl; 50mM $MgCl_2$, pH9.5) for two minutes and then incubated in the dark with about 10ml of freshly prepared substrate solution (45µl NBT solution vial 4, 35µl x-phosphate solution vial 5) in 10 ml of buffer 3 sealed in a plastic bag. The development of the colour precipitate was complete usually after 4h. The reaction was stopped by washing the filter for 5 minutes in TE buffer pH 8.0, dried and stored at room temperature.

2.9.17 Hybridisation of ^{32}P-labelled probes to filter-bound DNA.

The technique was used to detect homologous DNA sequences transferred to nylon or nitrocellulose filters by Southern transfer. Filters were equilibrated in prehybridisation solution containing 3x SSC (0.45M NaCl, 0.045M Na-citrate, pH7.0), 5x Denhardt's solution (0.1% (w/v) each of Ficoll 400, BSA, and PVP), 0.1%(w/v) SDS and

100mg/ml sheared and denatured herring sperm DNA at 65°C with rotation overnight in a techne hybridisation vessel.) 0.5-1.0 ml of this solution was used per cm^2 of filter. After equilibration, ^{32}P-labelled DNA previously denatured by boiling for 5 minutes, was added and the incubation continued as previously described for 16-20h. After incubation, the hybridisation solution was removed and the filter washed twice for 10 minutes at 65°C in solutions containing 0.1% SDS and varying concentration of SSC, depending on the stringency required. In most cases where probe was fully homologous to sequences to be detected, two minute washes in each of 3xSSC, 1xSSC and 0.1xSSC were performed. After washing, the filters were air dried and the presence and position of hybridised radioactive probes detected by autoradiography.

2.9.18 Autoradiography.

Autoradiography for ^{32}P was carried out using preflashed film (Fuji RX, Fujimex, Swindon, Wilts, UK) and an intensifying screen (Dupont, Wilmington, Delaware, USA), exposing the autoradiograph at -80°C for appropriate times.

2.9.19 Isolation of tobacco (*Nicotiana tabacum* cv.SR1) nuclei.

Tobacco nuclei were prepared as described by Spiker *et al.*,(1983) with some modifications. All glassware was siliconised autoclaved, and chilled at -20°C and operations conducted at 4°C. To maintain histone protein acetylation sodium butyrate (inhibitor of de-acetylase activity) was incorporated into all buffers to a final concentration of 20mM. Approximately 20g of fresh tobacco leaf tissue was ground to a fine powder with a mortar and pestle under liquid nitrogen. The liquid nitrogen was allowed to evaporate and the powder transferred to a 200ml beaker on ice. 25ml of ice cold Honda buffer (5% (w/v) Dextran T-40, 25mM Tris-HCl, pH8.5, 0.5% Triton-x-100, 0.44M sucrose, after autoclaving 2.5mM DTT, 1mM PMSF and 2mM spermine were added) was added to the powder with gentle stirring until a smooth slurry was obtained. The homogenate was then filtered through 3 layers of muslin and an 80 micron mesh. A further 25ml of Honda buffer was used to wash the filters through.

Table 2.5:

Percoll gradients.

Gradient constituents were prepared as follows:

Percoll (Pharmacia): Sterile stocks were stored at 4°C.

3M sucrose: The solution was heated to 75°C to ensure that the sucrose was completely dissolved prior to autoclaving.

10x Grad. Buffer (250 mM Tris-HCl pH 8.5, 100 mM $MgCl_2$): This 10x stock was made up in distilled water and autoclaved prior to use.

Solutions corresponding to 0, 40, 60 and 76% Percoll were made up by combining the stock solutions as shown. Gradients were formed in 30 ml Corex tubes by layering 4.5 ml of each solution in the order 76%/60%/40%/0% Percoll and allowed to stand at 4°C for at least 30 min. before use.

% Percoll	Percoll (ml)	3M Sucrose (ml)	10x Grad. buffer (ml)	H_20 (ml)	Volume (ml)
40	8	3	2	7	20
60	12	3	2	3	20
76	16	3	2	0	21
0	0	13	2	6	21

The filtrate was then centrifuged in 30ml corex tubes at 2,500 x g in a Beckman JS-21 centrifuge for 10 minutes at 4°C. The supernatant was discarded and the pellets washed in a total of 20ml of Honda buffer. Centrifugation was repeated as before. The crude nuclear pellets were resuspended in 5ml of Honda buffer and layered onto two four step gradients. Percoll solutions were prepared as described in table 2.5

Gradients were formed in 30ml corex tubes with each step layered from the bottom up. Gradients were left to stand for thirty minutes on ice before use. Centrifugation was performed at 4,000 x g for 30 minutes at 4°C in a swing-out rotor. The nuclei were observed to band as a white layer between the 40% and 60% percoll layers in the gradient. The fraction containing the nuclei was removed and washed with 5ml of nuclei wash buffer (50mM Tris-HCl, pH8.5, 5mM $MgCl_2$, 20%(w/v) glycerol). The nuclei were pelleted by centrifugation at 3,000 x g for 10 minutes at 4°C. This was repeated twice. Finally the nuclei were resuspended in 0.5ml of nuclei resuspension buffer (50mM Tris-HCl, pH8.5, 5mM $MgCl_2$, 50%(w/v) glycerol), frozen in dry ice and stored in a prechilled box at -80°C. Nuclei were stained with Carmine acetic acid and counted using a haemocytometer. Average yields were between $7x10^7$-$8x10^7$ nuclei per 10g of tissue. Nuclei were shown to be intact by the generation of a micrococcal nuclease ladder.

2.9.20 Nuclease digestion and DNA purification.

Micrococcal nuclease and DNaseI digestion experiments were carried out on 100µl aliquots of suspended nuclei. The suspended nuclei were pelleted in a microfuge for 10s to remove glycerol and resuspended in the appropriate digestion buffer for digestion with micrococcal nuclease or DNaseI. For micrococcal nuclease digestions the nuclear pellet was resuspended in 50mM Tris-HCl pH8.0, 0.3M sucrose, 5mM $MgCl_2$, 1.5mM NaCl, 0.005mM ß-mercaptoethanol, 0.1mM $CaCl_2$. Digests with several dilutions of micrococcal nuclease were performed ranging from 0.0(control digest) to 2.0 units/ml. Reactions were incubated for 1 minute at 37°C and halted by placing on ice and quickly adding 100µl of 0.1M EDTA pH7.0. The nuclei were

pelleted by centrifugation in a microfuge for 10s. To lyse the nuclei, the pellet was resuspended in 300µl extraction buffer (100mM Tris-HCl, pH8.0, 50mM EDTA, 500mM NaCl, 10mM ß-mercaptoethanol) with 5µl of 20mg/ml proteinase K and 50µl of 10%(w/v) SDS and incubated at 65°C for 40-60 min. After incubation, 100µl 5M potassium acetate (pH4.5) was added, mixed well and then set on ice for 30 min. The precipitated material was then pelleted at 13,000 x g in a microcentrifuge for 10 minutes and the supernatant transferred to another tube. The DNA was precipitated from the supernatant by adding 600µl of isopropanol, chilling for at least 15 minutes (-80°C) and centrifuging as before. The DNA pellet was vacuum dried and resuspended in 150µl of TE pH8.0. To this 50µl of 7.5M ammonium acetate was added followed by 500µl 95%(w/v) ethanol and incubated at -20°C for at least 15 minutes and centrifugation as before. The DNA pellets were washed with 80%(w/v) ethanol, vacuum dried and resuspended in 40µl of TE pH 8.0. The concentration of DNA was measured by fluorescence with DABA.(see section 2.9.4).

2.10 Biochemical techniques: Proteins.

2.10.1 Isolation of pea lectin.

The method used was essentially that of Trowbridge (1974). Pea lectin was prepared from mature pea seed; 200g was ground to a fine meal in a water cooled mill and extracted with stirring at 4°C with 800ml 25mM sodium acetate pH 4.8, for at least 3h. Pea meal residue was discarded after centrifugation at 1,500 x g for 30min. Solid $(NH_4)SO_4$ was added slowly to the supernatant to give 60% relative saturation. Precipitated protein was recovered by centrifugation at 3,000 x g for 30min and dissolved in 80ml of distilled water and dialysed at 4°C against 1litre of 0.15M sodium chloride, 0.01M Phosphate buffer pH7.2 for 3h each. The crude lectin extract was loaded onto a sephadex G-100 column (500ml). The column was then washed with 0.15M sodium chloride, 0.01 phosphate buffer pH7.2 until the A_{280} was less than 0.05 and the pea lectin was eluted with the same buffer containing 0.2M glucose. Absorbance was monitored at 280nm and 10ml fractions were collected using a Frac-

100 fraction collector (Pharmacia). Fractions correlating to the peak of released lectin were pooled and dialysed exhaustively against distilled water, followed by storage as a sterile solution. The lectin concentration was determined spectrophotometrically using lectin $E^{280}_{1\%}$=15.0 (Trowbridge, 1974).

2.10.2 Total plant protein extraction from tobacco.

Plant tissue protein extracts were prepared by freezing tissue in liquid nitrogen, grinding to a fine powder and freeze drying. A suitable volume of phosphate buffered saline (PBS: 8g/l NaCl, 0.02g/l KCl, 1.15g/l Na_2HPO_4, 0.02g/l KH_2PO_4, pH7.2) was added and the homogenate shaken on a rotary shaker for 1h, sonicated and shaking continued for 1-2h at 4°C. The homogenate was centrifuged at 12,000 x g for 5min and the protein content of the supernatant determined (see section 2.10.3).

2.10.3 Determination of protein content of plant extracts.

The method used to estimate soluble proteins was the Coomassie brilliant blue (G250) binding assay of Bradford (1976). The method was used directly on extracts (or diluted extracts) prepared for use in radioimmunological assays. The Bradford reagent was obtained in concentrated form from Bio-Rad. The reagent was diluted 1:5 with distilled water, 1ml was removed and mixed with 20µl of plant extract and the absorbance at 595nm measured after 2 minutes in a 1cm cuvette. The blank consisted of 20µl of the appropriate extraction buffer and 1ml of the reagent. Protein concentrations were determined by extrapolation from a standard curve constructed using BSA standards and pea lectin standards in the concentration range (0-100µg/ml) diluted in the appropriate buffer.

2.10.4 Slot-blot radioimmunoassay.

The method used was adapted from that of Jahn *et al,* (1984). Nitrocellulose filters (SA85) were pre-equilibrated in phosphate buffered saline (PBS:8g/l NaCl, 0.02g/l KCl, 1.15g/l Na_2HPO_4, 0.02g/l KH_2PO_4, pH7.2) for a few minutes prior to blotter

assembly. Proteins were bound to the nitrocellulose by passing 200µl of leaf extracts containing a known amount of protein (20µg/ml) through a filter using a Schleicher and Schuell SRC 072/0 vacuum slot blot apparatus connected to a constant vacuum pump. Controls such as pea-lectin, were diluted two-fold steps from 90ng/ml-0.3ng/ml in microtitre plates previously blocked with 1%(w/v) BSA and then applied to the same filter. After binding, the filters were dried under vacuum for 1h at room temperature. The non-specific protein binding sites were blocked by incubating the filter with 5%(w/v) skimmed milk powder in PBS (blocking buffer) for 2h at 37°C with gentle shaking. Immunological detection of bound pea lectin was conducted using rabbit anti-pea lectin primary antibodies(2mg/ml) diluted 1:500 in blocking solution for 2h at 37°C. The filter was washed 3 x 15min in blocking buffer to remove unbound pea lectin antibody. The ^{125}I-conjugated donkey anti-rabbit secondary antibodies (5µCi) were then added and incubated in blocking solution for 2h at 37°C. Before exposure the filter was washed 3 x 15min at 40°C incorporating 0.1%(v/v) Triton-X-100 into the final wash. When the background was sufficiently low as judged by Geiger coumter, the filters were air dried, autoradiographed and signal intensities of the slots on the autoradiograph determined using an LKB Ultroscan XL laser densitometer.

2.10.5 Squash-blot radioimmunolocalisation.

A nitrocellulose filter (20 x 20cm, Scleicher and Schuell) and a sheet of Whatman 3MM paper were equilibrated in Tris-buffered saline (TBS: 20mM Tris-HCl, pH7.2, 0.9%(w/v) NaCl) and placed, with the nitrocellulose uppermost on the porous sheet plate of a vacuum gel drier (Pharmacia). Tissue-culture-grown plants or leaf tissue taken from individual tansformants were arranged on the filter and a gentle vacuum was applied. The plant tissues were then carefully squashed onto the nitrocellulose using a flat ended spatula and the released cell contents sucked into the filter medium. After squashing, the plant tissues were removed with forceps and the filter air-dried. Filters were stained with Ponceau red to assess the efficiency of protein transfer and subsequently processed as described for the slot blot radioimmunoassay to determine

the distribution of expressed pea lectin.

2.10.6 Extraction of tobacco histone proteins.

The method used was essentially that of Allis *et al.*, (1979). Tobacco nuclei prepared as described in section 2.9.19 were extracted once with 0.2M H_2S0_4 for 12-18 hr at 4°C. The sulphuric acid solution was added directly to nuclei pelleted by centrifugation in plastic Eppendorf tubes which were then shaken gently overnight. The supernatant was recovered fter centrifugation (16,000 x g) for 15min) and the acid soluble material was precipitated by the addition of 100% trichloroacetic acid to a final concentration of 20% (w/v) and chilling for 15min at 0°C. The precipitate was collected, washed once with 0.024M HCl in acetone and once with acetone and dried under reduced pressure. Throughout this procedure only plastic or siliconised glassware was used.

2.10.7 SDS-Polyacrylamide gel electrophoresis of proteins (SDS-PAGE).

Sodium dodecyl sulphate (SDS) polyacryamide gels were prepared according to Laemmli (1970). Recipes for the preparation of the resolving gel and stacking gels are given in table 2.6.

- Separating Gel Buffer: 1.5M Tris
  0.4% (w/v) SDS
  pH to 8.7 with HCl

- Stacking Gel Buffer: 0.5M Tris
  0.4%(w/v) SDS
  pH to 6.8 with HCl

- Acrylamide stock: 30%(w/v) acrylamide
  0.8%(w/v) Bis-acrylamide

The reservoir buffer consisted of 192mM glycine, 25mM Tris base, 0.1% (w/v) SDS,

Table 2.6:

Recipes for gel solutions for 1.5mm SDS-polyacrylamide gels, capacity 40ml.

Separating gels of the appropriate acrylamide concentration were cast and overlain with a layer of distilled water before allowing to polymerise. Once set the water was completely removed and the stacking gel cast immediately on top of the separating gel. The comb was inserted into the stacking gel and polymerisation allowed to proceed. Before loading the gel wells were flushed out with running buffer to completely remove residual non-polymerised acrylamide.

Separating Gel:

Final acrylamide concentration (%)	**7.5**	**12**	**15**	**20**
Separating gel buffer (ml)	10	10	10	14.7
Distilled water (ml)	20	14	10	0
Acrylamide stock (ml)	10	16	20	28
TEMED (µl)	30	30	30	30
10% Ammonium persulphate (µl)	140	140	140	140

Stacking Gel:

Final acrylamide concentration (%)	**4**
Stacking gel buffer (ml)	2.5
Distilled water (ml)	6.1
Acrylamide stock (ml)	1.34
TEMED (µl)	10
10% Ammonium persulphate (µl)	50

pH8.3). The stacking gel (at least 2cm high) was 4% acrylamide with a few drops of tracking dye added (1%(w/v) Bromophenol blue in distilled H_2O) and the resolving gel (9.0cm) was 22% acrylamide. Preparations of histone proteins purified as outlined in section (2.10.6) were mixed 1:1 with 2x Sample Buffer (30ml 10% (w/v) SDS, 12.5ml stacking gel buffer, 10ml glycerol, pH6.8),and stacked in 1.5mm gels of 4% (w/v) acrylamide and separated with 22% (w/v) acrylamide. For basic histone proteins gels were electrophoresed at 130-140v until the samples were in the separating gel where the voltage was lowered to 120v and the run continued for approximately 4h. The histone preparations were run in parallel with a prepared mixture of molecular weight markers (Dalton VII-SDS) consisting of BSA 66,000; ovalbumin 45,000; G-3-P-dehydrogenase 36,000; carbonic anhydrase 29,000; trypsinogen 25,000; soya bean trypsin inhibitor 21,000; and lactalbumin 14,000. On completion the gels were stained in 0.07% (w/v) Coomasie brilliant blue R 250 (Sigma) in 50% (v/v) methanol 5% (v/v) acetic acid for at least 1h, and destained in 5% (v/v) methanol, 7.5% (v/v) acetic acid.

2.10.8 Western blotting: Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose filters.

The method used was that of Towbin *et al.*, (1979) for wet electroblotting. Transfer was conducted in a Bio-Rad electroblotter according to the instructions supplied. Briefly the gel was first equilibrated for 15 minutes in transfer buffer (125mM glycine, 25mM Tris-base, 0.1%(w/v) SDS,20%(v/v) methanol) and then placed in a cassette with holes to allow passage of current. A sheet of nitrocellulose pre-soaked in transfer buffer was carefully positioned on top of the gel and the cassette closed to sandwich both together. The cassette was immersed in the electroblotter tank containing 3l of transfer buffer. The apparatus was usually run at 200v overnight for a good transfer . After transfer the filters were processed for immunological detection of acetylated and unacetylated histone 4.

2.10.9 Immunological detection of acetylated and unacetylated Histone 4 immobilised on nitrocellulose filters.

To evaluate the efficiency of transfer, blots were briefly stained with Ponceau red, quickly destained in water, and photographed before processing for immunoblotting to provide a photographic record of each blot. All incubations described were at room temperature with shaking and all washing steps conducted under the same conditions for 15min each. Filters were incubated and washed once in blocking solution (5%(w/v) non-fat dried milk powder in 20mM Tris-HCl pH7.2, 0.9%(w/v),NaCl. Filters were then incubated with rabbit antibodies against acetylated or unacetylated histone 4 (25µl/25ml blocking solution) for 2h, washed three times in blocking solution, and incubated with ^{125}I-conjugated anti-rabbit IgG antibodies (10µl/25ml in 25ml of blocking solution). Filters were then washed three times in blocking solution, followed by autoradiography.

2.10.10 Immunofractionation of chromatin.

Chromatin was immunofractionated using the method of Hebbes *et al.*, (1988). Glycerol was removed from stored nuclei by centrifugation and the pelleted nuclei washed by centrifugation in 5-10 vol. of micrococcal nuclease digestion buffer (50mM Tris-HCl, pH8.0, 0.3M sucrose, 5mM $MgCl_2$, 1.5mM NaCl, 0.005mM ß-mercaptoethanol, 0.1mM $CaCl_2$, 10mM sodium butyrate). Native chromatin was prepared from tobacco nuclei by resuspending nuclei at a DNA concentration of 0.5mg/ml as estimated by fluorescence using DABA (see section (2.9.4) in micrococcal nuclease buffer and digesting with micrococcal nuclease at 20U/ml for 30s at 37°C. Digestion was terminated by the addition of EDTA to a final concentration of 5mM and the nuclei lysed by sonication. Insoluble material was removed by centrifugation. The chromatin-containing supernatant approximately 200µg of DNA as estimated by the diaminobenzoic acid method was incubated with 30µg of anti-acetylated histone 4 antibody in 500µl of incubation buffer (PBS: 8g/l NaCl, 0.02g/l LiCl, 1.15g/l Na_4HPO_4, 0.02g/l KH_2PO_4 pH7.4), with 20mM sodium butyrate, 5mM EDTA,

EDTA, 0.1mM PMSF) for 3h at 37°C under constant agitation. The chromatin/antibody mixture was added to 20µl of PROSEP^{TM}-A (Protein A immobilised on controlled pore glass) and incubated at room temperature for 1h. with constant agitation. Complexes were collected by centrifugation and the unbound fraction in the supernatant retained. Bound material was eluted by resuspending the pellet in 250µl of elution buffer (0.1M glycine/HCl pH3.0) and incubating at room temperature for 15min. Prosep-A was removed by centrifugation and the supernatant containing the released chromatin and antibody (the"bound" fraction) collected and dialysed against TE pH7.5. DNA from all chromatin fractions was obtained by two phenol/chloroform and one chloroform extraction. The DNA was ethanol precipitated, redissolved and treated with 50µg/ml RNaseA and subsequently with 50µg/ml proteinase K before final precipitation with ethanol. DNAs from both antibody-bound and -unbound fractions were analysed in 1% (w/v) agarose gels, blotted and processed for hybridisation with ^{32}P-labelled pea lectin sequence.

Chapter 3:
Results and Discussion.

3.1 Introduction.

In the analysis of transgenic plants many workers have commonly observed that the expression level of a particular construct can vary between independent transformants, independent regenerants from the same transformed cell, and progeny plants, a phenomenon referred to as 'position effect' and of great importance to both research and commercial interests. The reasons for position effect remain unclear; possible contributory factors include number, location and distribution of integration sites. Moreover, methylation of the foreign DNA following stable integration into the genome can affect its expression (Meyer and Heidmann, 1994). This project is concerned with exploring the phenomenon of position effect, examining the extent to which the observed variable phenotypes are influenced by such factors.

3.2 Source material.

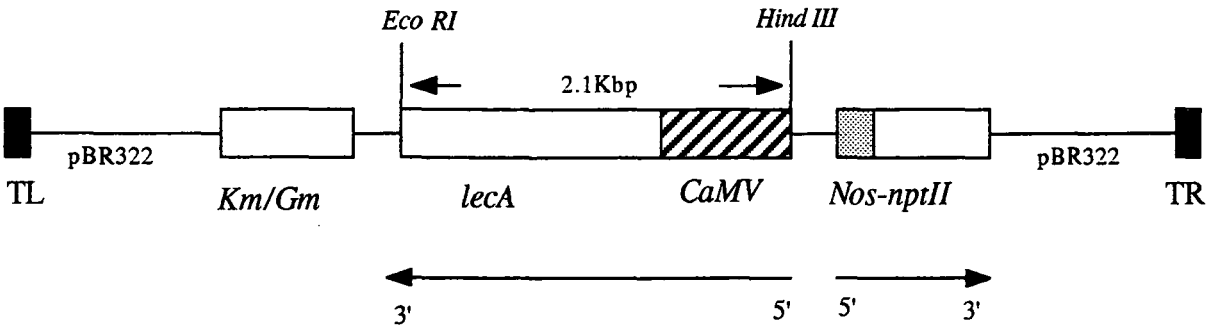
In order to examine the phenomenon of 'position effect' suitable transgenic material was supplied by Dr. G. Edwards (Ph.D., University of Durham, 1987). The material consisted of transgenic tobacco (*Nicotiana tabacum* cv., Petit Havana SR1) seed produced from the self-crossing of three parental lines designated L1, L4, and L12. These *Nicotiana tabacum* lines had been generated via *Agrobacterium tumefaciens* (strain GV3101) mediated transformation with the co-integrate pGV3850::pDUB130 (*CaMV-lecA*). The resultant T-DNA integrated within the tobacco genome contained between its left and right T-DNA borders the pea lectin gene (*lecA*) controlled by the full 35S cauliflower mosaic virus (*CaMV*) promoter, the neomycin phosphotransferase gene II (*nptII*) driven by the nopaline synthase promoter (*nos*), and the plasmid sequence of pBR322 (figure 3.1; for more detail on *Agrobacterium* strain and plant transformation see section 3.2.1). The use of the lectin reporter gene afforded a

Figure 3.1:

The arrangement of the *CaMV-lecA* chimaeric gene within the T-DNA expression cassette of the plant transformation vector pGV3850.

Key: *CaMV*: cauliflower mosaic virus 35S promoter; *lecA*: pea lectin coding region; *nos*: nopaline synthase promoter; *nptII*: neomycin phosphotranspherase II gene, conferring kanamycin resistance in plants; Km^r/Gm^r: kanamycin/gentamycin resistance in *E.coli/Agrobacterium*; pBR322: plasmid sequence of pBR322 containing the ampicillin resistant gene; TL & TR : left and right hand T-DNA border sequences.

pGV3850::pDUB130 Co-integrate



number of important advantages in this study upon comparison with other reporter systems available. The lectin protein is plant-derived and therefore DNA sequence encoding this protein will possess the correct codon bias and processing sequences necessary for optimal expression in a plant-based system. Thus produced protein is easily quantified using lectin-specific antibodies and activity values readily determined. The antibody assay itself is directly quantitative for the actual amount of protein present, and thus indirect quantitation methods such as determination of potentially variable enzyme activities, as with the reporter genes *uidA* and *luc* are redundant. The lectin protein itself is a stable protein, being resistant to proteolysis, and not subject to modification by glycosylation. These considerations illustrate the distinct advantages of the *lecA* reporter gene and, coupled with the capacity to handle large numbers of samples in the assay system employed were factors taken into account in initial experimental design.

The integrated T-DNA for all *CaMV-lecA* lines (L1, L4 and L12) had been shown by G. Edwards (Ph.D., 1987) to be structurally stable during vegetative proliferation and sexual transmission under the applied kanamycin selection conditions. As determined by Southern analysis no rearrangements of the integrated T-DNA were evident. For all three lines only one copy of the *lecA* gene was present as judged by comparison with gene copy reconstructions. Finally, the expression levels of pea lectin were found to vary between the independent *lecA* lines and progeny of the plants, thus making the material ideal for an investigation into the molecular reasons underlying transgene expression variability.

3.2.1 Selection of transgenic material.

Tobacco seed collected from the self crossing of the parental *CaMV-lecA* lines L1, L4, and L12 were tested for kanamycin resistance by germination on agar plates of half strength MS salts (5mg/ml) supplemented with kanamycin sulphate (100µg/ml).

In table 3.1 germinated seeds from only one parent, *CaMV-lecA* plant L4 segregated

Table 3.1:

Frequency of kanamycin resistance in populations of T1 *CaMV-lecA* lines L1, L4 and L12.

Seed was germinated on kanamycin-containing (100 μg/ml) medium and sensitivity to kanamycin determined as bleaching of cotyledons at 21 d post-germination. Observed segregation ratios were then statistically compared to predicted ratios and the probability that observed and predicted ratios were the same expressed as a p value.

Parent plant *CaMV-lecA*	Number of seeds germinated	Resistant to kanamycin (green)	Sensitive to kanamycin (white)	Fit to ratio Kan^r:Kan^S	χ^2	p
L1	0	0	0	-	-	-
L4	200	141	59	3:1	1.33	<0.05
L12	14	11	3	3:1	0.896	<0.05
SR1	200	0	200	-	-	-

into kanamycin resistant and sensitive phenotypes. The seedlings were scored for resistance to kanamycin after 1 week by counting the number of green resistant seedlings to white sensitive seedlings. The counts were then compared to the expected segregation ratio for one independent locus for kanamycin resistance using the chi-square test:

$$X^2_{[1]} = \frac{([Ob_{Green} - Ex_{Green}])^2}{Ex_{Green}} + \frac{([Ob_{White} - Ex_{White}])^2}{Ex_{White}}$$

where 'Ob' is the observed frequency for each of the classes and 'Ex' is the corresponding expected value.

For *CaMV-lecA* line 4 progeny a computed value of $X^2_{[1]} = 1.33$ was obtained which, with reference to a chi-squared table, is less than $X^2_{[1]}$ value at the 5%(3.84) significant level. Therefore the null hypothesis (that the data do not differ from those predicted by Mendel's first law) is not rejected; in other words, a segregation ratio of approximately 3:1 Kan^r : Kan^s is the phenotype frequency as predicted by Mendel's first law, for the *nptII* gene segregating as a single dominant locus. The expected segregation ratio of 3:1 Kan^r:Kan^s was also displayed by L12, however the result was unreliable in that greater than 90% of the L12 seed failed to germinate in the above and repeat experiments.

One explanation for the reduction in germination of L12 seed may reflect inappropriate harvesting and storage of seed. No germination of L1 was observed as a result of fungal contamination of the seed. Attempts to stem the growth of the fungal contaminant were carried out by prolonging the incubation of L1 seed with 20%(w/v) sodium hypochlorite during surface sterilisation and by the incorporation of an anti-fungal agent nyastatin (50µg/ml) into the germination agar; both however proved unsuccessful.

From the above data it is clear that, of the material supplied, only one line (L4), which appears to contain the transgene at a single locus, was suitable for further analysis.

Thus seven kanamycin resistant tobacco plants representing the T1 population of*CaMV-lecA* L4 were selected and further characterised by Southern blotting analysis.

3.2.2 Southern analysis of selected plants.

Genomic DNA was isolated from the leaves (5-10cm) of seven kanamycin-resistant individual tobacco plants representing a sample of the T1 population of L4. The DNA (10 µg) for each plant was analysed on 0.7%(w/v) agarose gels after complete digestion with *EcoRI* and *HindIII* together with linearised pDUB80 (spiked into *EcoRI* and *HindIII* restricted SR1 genomic DNA) representing gene copy reconstructions of one, three, and five gene copies of the *lecA* gene (figure 3.3(A)). The amount of pDUB80 DNA estimated to contain the equivalent of one gene copy per 10µg of genomic DNA was calculated using a haploid *N. tabacum* genomic content of 1*C* = 3.3pg DNA (Bennet and Smith, 1991). The required concentrations of linearised pDUB80 were determined fluorimetrically using diaminobenzoic acid (DABA) and a standard curve constructed from a two-fold serial dilution of commercial pUC18 of known concentration (1µg/ml) (figure 3.2). In Southern hybridisation analysis an expected 2.1 Kb band (figure 3.3(B), indicated by arrow) for the *CaMV-lecA* chimaeric gene was detected with varying intensities in the seven plants when probed with a 1.2 Kb *SphI/ EcoRI* fragment containing the *lecA* coding region. Gene copy reconstructions were shown to be linear after densometric scanning (LKB laser densitometer) and plotting the areas under the curve against pDUB80 concentrations (figure 3.3(C)). For samples corresponding to lanes 1, 2, and 3 plants were estimated to contain one copy of the *CaMV-lecA* gene per genome with no rearrangements. In lane 4 (plant 4) two bands were observed, one the expected size of 2.1Kb equivalent to one gene copy, the other slightly larger and greater in intensity. This larger band may have been the result of incomplete genomic DNA digestion or possibly rearrangements within the T-DNA and/or tobacco genome. An identical pattern of three bands of homology were observed in lanes 5 and 7, two of which were much larger than the predicted size (2.1Kb) and may again have resulted from incomplete digestion or, perhaps, gene

Figure 3.2:

Calibration curve with diaminobenzoic acid (DABA), using commercial pUC18 (2.5µg-0.5µg) as the plasmid DNA standard.

Estimations of pDUB80 plasmid concentrations (purified on a caesium chloride gradient) were determined using DABA and extrapolating from the standard curve for dilutions representing one, three, and five gene copy reconstructions.

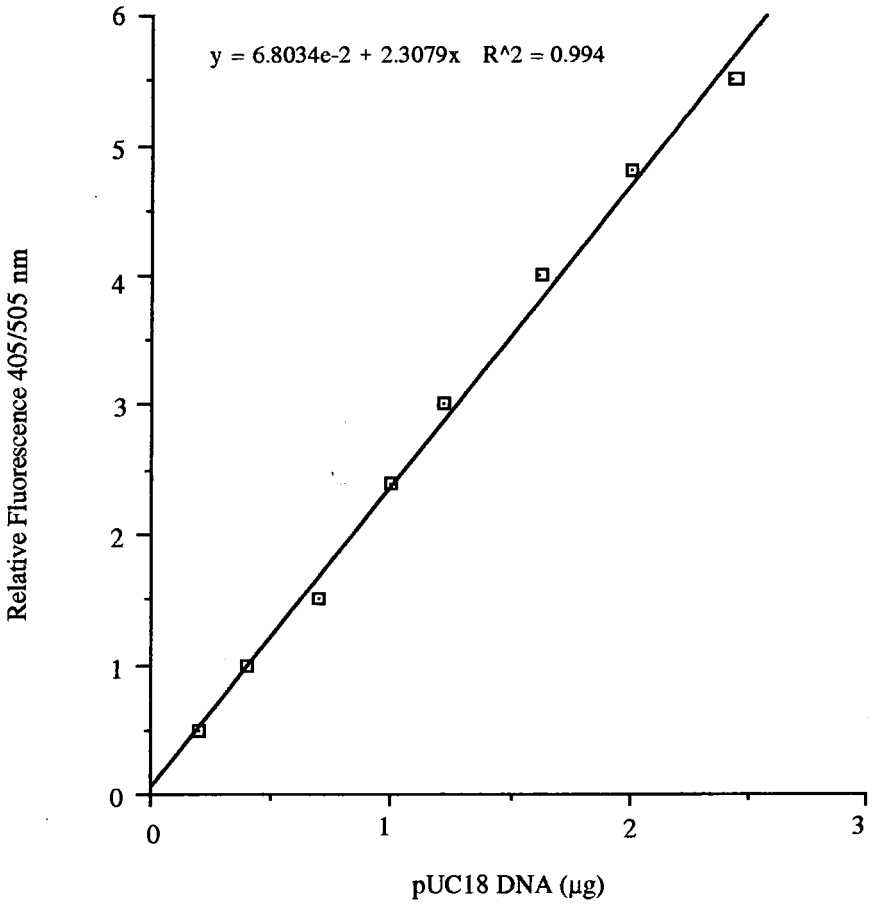

y = 6.8034e-2 + 2.3079x R^2 = 0.994
Relative Fluorescence 405/505 nm
pUC18 DNA (µg)
0
1
2
3
4
5
6

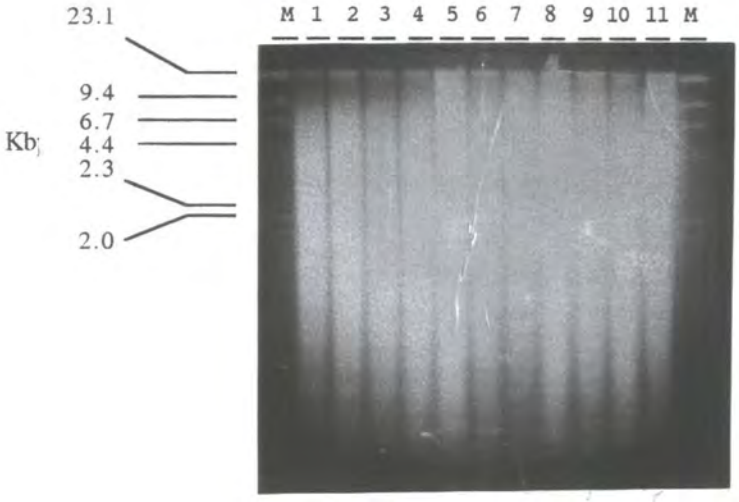
Figure 3.3:

(A) Agarose gel (0.7% w/v) analysis on seven kanamycin resistant progeny for parental *CaMV-lecA* line L4. 10µg aliquots of genomic DNA were used per lane. Lanes 1-7: *EcoRI* and *HindIII* digested genomic DNA for *CaMV-lecA* line L4 T1 individuals, plant numbers 1-7 respectively. Lane 8: *EcoR I* and *HindIII* digested SR1 genomic DNA. Lanes 9-11: *EcoR I* and *HindIII* digested SR1 genomic DNA, spiked with *EcoRI* linearised pDUB80 representing *CaMV-lecA* gene copy reconstructions one, three, and five respectively. Lane M: Lambda *HindIII* molecular weight DNA markers

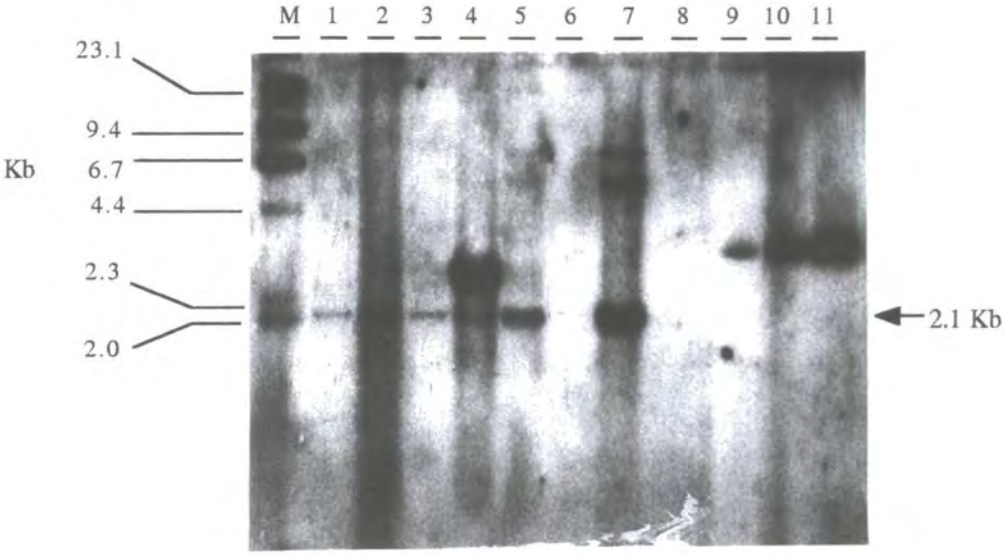
(B) Southern blot analysis of *EcoRI* and *HindIII* digested genomic DNA (10µg) from leaves (5-10cm long), representing seven *CaMV-lecA* T1 plants, probed with a ^{32}P-labelled 1.2kb *Sph I/EcoRI* fragment of pDUB80 containing the *lecA* coding region. Lanes 1-7: plant numbers 1-7 respectively for *CaMV-lecA* T1 individuals, arrow indicates the predicted size (2.1Kb) of an unrearranged chimaeric *CaMV-lecA* gene; Lane 8: *EcoRI* and *HindIII* digested SR1 genomic DNA (untransformed control); Lanes 9-11: *EcoRI* linearised pDUB80 (3.8kb) representing, one, three, and five gene copies per haploid *Nicotiana tabacum* genome, in the order given.

(C) Gene copy reconstructions. A plot of the relative area percentages under the peaks after densometric scanning (L.K.B. densitometer) for *CaMV-lecA* gene copy equivalents of one, three, and five with pDUB80 plasmid DNA concentrations. Plot indicates linear relationship between gene copy reconstructions. X-ray film exposed under described conditions (section 2.9.18).

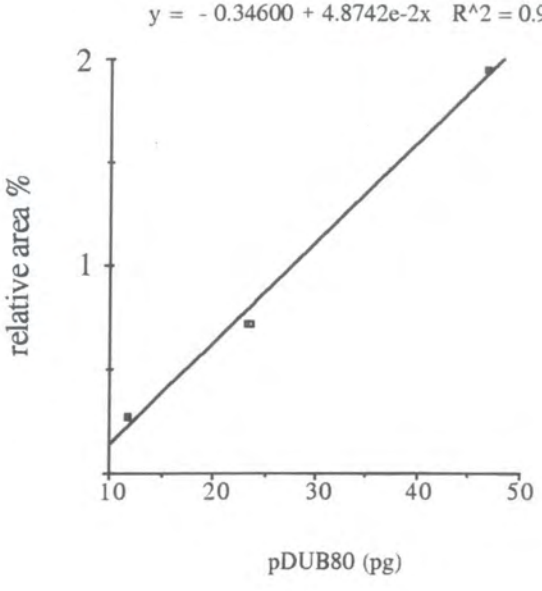
A.


M 1 2 3 4 5 6 7 8 9 10 11 M
23.1
9.4
6.7
4.4
2.3
2.0
Kb

B.


M 1 2 3 4 5 6 7 8 9 10 11
23.1
9.4
6.7
4.4
2.3
2.0
Kb
2.1 Kb

C.


y = - 0.34600 + 4.8742e-2x R^2 = 0.994
relative area %
2
1
10
20
30
40
50
pDUB80 (pg)

rearrangements. The strongly intense third band of predicted size 2.1kb was equivalent in intensity to 2 gene copies. The absence of bands with homology to the *lecA* probe in lane 6 (plant 6) suggested that the transgene may have been deleted from the genome. Thus Southern results in conjunction with the genetic analysis suggests the integration of at least one copy of the *lecA* gene at a single locus with plants 1, 2, 3, and 4 being hemizygous and plants 5 and 7 homozygous for the transgene.

3.2.3 Detection of pea lectin in the T1 population of transgenic line L4.

Having extensively characterised seven L4 plants and determined the presence of a single insertion locus containing one copy of the T-DNA in either a hemizygous or homozygous state, it was desired to investigate the influence of copy number on levels of transgene expression within a T1 population. The amount of pea lectin protein in the leaves of transgenic tobacco plants was estimated using a radio-immunoassay (RAI) (Jahn *et al.*, 1984) and used as a measure of levels of transgene expression. Such an assumption is possible given that previous studies (Edwards, 1987) have shown an absolute correlation between levels of lectin protein and corresponding transcript. Prior to analysis of transgenic material lectin isolation and RAI procedures were further optimised.

3.2.3.1 Isolation of pea lectin from material.

The purification of pea lectin involved affinity chromatography on a sephadex G-100 column by way of utilising the lectin carbohydrate binding properties. As shown in figure 3.4 the pea lectin in a buffer extract from pea meal (section 2.10.1) was allowed to bind to the column while all the weakly unbound and associated proteins were washed off in elution buffer. Pure, bound pea lectin was eluted using a pulse of 0.2 M glucose (figure 3.4, indicated by arrow) and monitored as a single peak. Aliquots of the eluant were analysed on SDS-PAGE gels to confirm purity. The pure fractions corresponding to the peak in figure 3.4 were pooled together to yield 54mg of lectin per

200g of pea seeds. The pea lectin was stored as a sterile solution 0.5mg-1mg/ml in 0.15M sodium chloride, 0.01M sodium phosphate pH 7.2. This protein was found to agglutinate rabbit erythrocytes at a minimum concentration of 0.8µg/ml under the conditions of the agglutination assay (Trowbridge *et al.*, 1974). Examination of the affinity purified pea lectin by SDS-gel electrophoresis on a 15% (v/v) acrylamide gel revealed only two bands (7,000 and 17,000 daltons) with no other contaminating proteins. This is consistent with the data of Trowbridge *et al.*, (1974) where pea lectin has been demonstrated to be a tetramer consisting of two α-subunits (MW. 7,000) and two ß-subunits (MW. 17,000). The anti-lectin antiserum was obtained from Dr. G. Edwards (Shell Research Ltd.). The specificity of this antibody had previously been determined, showing an immunoreaction with the ß-subunits (figure 3.5; Edwards, 1987).

3.2.3.2 Optimisation of detection assay.

To optimise the RAI assay and to ensure quantitative estimates of lectin the maximum protein (lectin) binding capacity of the nitrocellulose matrix within the slots of a slot blotter (Schleicher and Schuell) was determined. A dilution series of SR1 tobacco leaf total protein extracts, from 20µg down to 0.05µg in a 200µl volume of PBS was applied in each slot of the slot blotter containing two sheets of nitrocellulose. From figure 3.6(A), it was observed from staining both nitrocellulose sheets (upper and lower) with Ponceau red (0.2% (w/v) in 3% (w/v) TCA) that protein concentrations greater than 6.0µg per slot resulted in saturation of all protein binding sites within the upper nitrocellulose sheet. At higher concentrations unbound protein passes through onto the lower second nitrocellulose sheet. In row 1 the tobacco leaf protein extract was spiked with pea lectin two-fold serially diluted from a starting concentration of 3µg. After probing with a rabbit anti-pea lectin antibody and detection with donkey anti-rabbit IgG ^{125}I conjugate in figure 3.6(B) (row a), it was evident that the optimum concentration range over which to detect pea lectin was between 90ng and 0.3ng producing discrete slot bands which yielded a linear response when scanned by a

densitometer. It was also observed that no cross reactivity was detected with SR1 tobacco protein extracts below 20μg (figure 3.6 row (b)) and therefore this could be used as a suitable negative control.

3.2.3.3 Estimation of lectin levels in transgenic plants.

Total soluble protein was extracted from the first upper leaf (3-5cm long) of each of the seven previously characterised transgenic *CaMV-lecA* tobacco plants at the six leaf stage in phosphate buffered saline (PBS: 8g/l NaCl, 0.02g/l KCl, 1.15g/l Na_2HPO_4, 0.02g/l KH_2PO_4, pH7.2) and their protein content determined by the Coomasie brilliant blue binding assay of Bradford (1976). Extracts were diluted to a protein concentration of 20μg/ml and 200μl of each extract (approx. 4μg of protein) was used per slot, alongside a row containing a two-fold serial dilution of purified pea lectin (90ng-0.3ng) as shown in figure 3.7(A) (row a). Figure 3.8(A) displays a densometric trace, over 8 slots containing the standard dilutions of pea lectin in figure 3.7 (A) (row a). A standard curve was constructed by calculating the area under the peaks (integrated values from densitometer) and plotting these against the logarithm of the pea lectin concentration (figure 3.8 (B)). Peak areas were also calculated for densometric scans of the seven transgenic *CaMV-lecA* plants and quantitative lectin levels determined by extrapolating from the curve in figure 3.8 (B). All plants analysed (figure 3.7 (B)) with the exception of plant 6 were observed to contain detectable amounts of pea lectin protein greater than 0.05% of total soluble leaf protein. Variation in pea lectin expression levels within the T1 population was observed to be low with greater than 50% of the plants centred around the mean value of 0.11% of total soluble leaf protein. Plant 5 which was estimated to contain 2 *lecA* gene copies produced pea lectin protein levels comparable with plant 3 estimated to contain 1 copy of the *lecA* gene. Similar pea lectin levels were also observed in plant 4 which possibly contains rearranged copies of the *CaMV-lecA* gene and at least one intact copy. However plant 1 which was estimated to have 1 copy of the *lecA* gene differentiated from the rest of the T1 plants having increased levels of detectable pea lectin protein than expected for a

Figure 3.4:

Affinity purification of pea lectin on sephadex G-100 from a crude pea seed protein preparation, after fractionation with solid ammonium sulphate and dialysis (section 2.10.1). Unbound pea seed proteins, corresponding to fractions 30-110 were washed through with 2.5l of 0.15M NaCl, 0.01M sodium phosphate buffer (pH7.2). Fractions 150-210 containing pea lectin were eluted with 1l of 0.2M glucose made up in 0.15M NaCl, 0.01M sodium phosphate buffer (pH7.2) as indicated by the arrow.

* refers to A_{280} monitored after elution with 0.2M glucose.

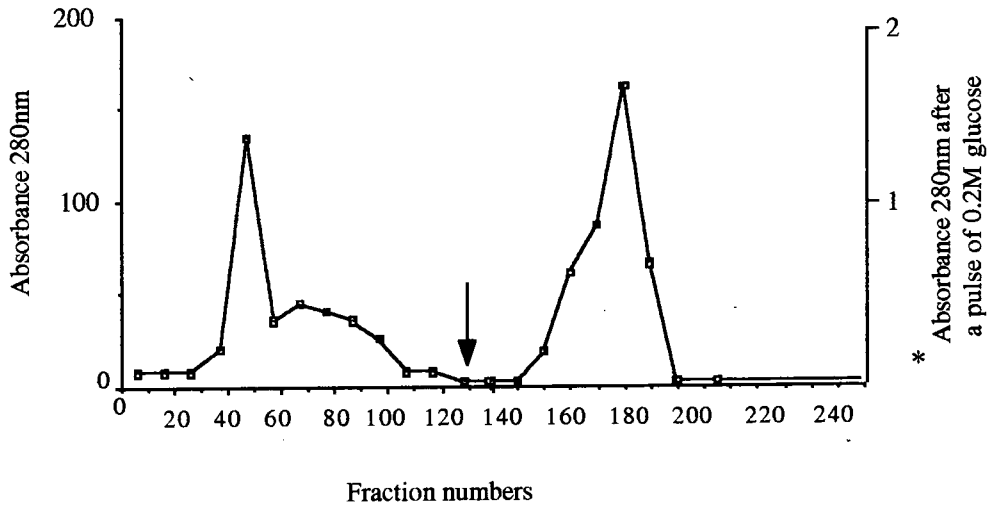

200
100
0
0
20 40 60 80 100 120 140 160 180 200 220 240
Fraction numbers
Absorbance 280nm
2
1
* Absorbance 280nm after a pulse of 0.2M glucose

Figure 3.5:

Left: SDS-PAGE of extracts of transgenic plants. Lane 1: 5 μg size markers (pea vicillin); Lane 2: 500 ng pea cotyledon lectin. Lanes 3: 100μg leaf extract from control plant. Lanes 4, 5, and 6: 100 μg leaf extract of transgenic plants containing the *CaMV-lecA* transgene.

Right: Filter of above following immunological detection of lectin. Lanes 7-11 as lanes 2-6 above respectively.

From: Edwards, 1987.

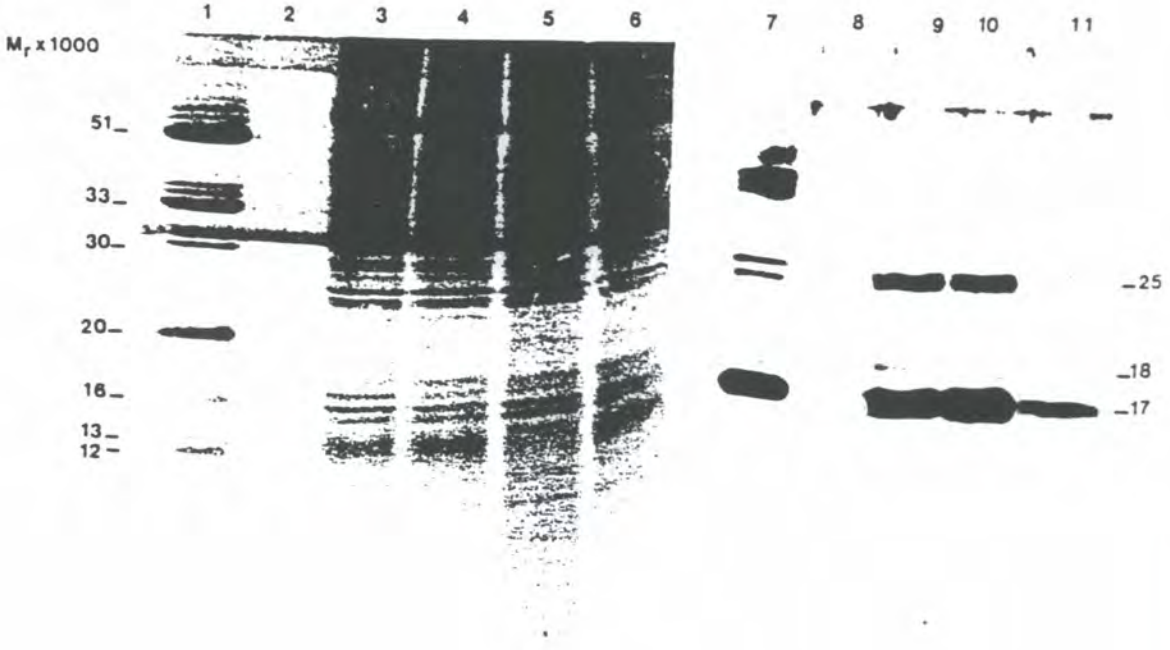

1
2
3
4
5
6
7
8
9
10
11
M_r x 1000
51
33
30
20
16
13
12
25
18
17

Figure 3.6:

(A) Nitrocellulose filters stained with Ponceau red (0.2%(w/v) in 3%(w/v) TCA after binding a range(20µg-0.05µg) of SR1 total soluble leaf protein in PBS. Top filter sheet, row 4: SR1 total soluble leaf protein from 20µg down to 0.05µg, stained with Ponceau red. Row 3: SR1 total soluble leaf protein from 20µg down to 0.05µg, spiked with a two-fold serial dilution of purified pea lectin (3µg-0.1ng) and stained with Ponceau red. Rows 1 and 2: represent the bottom nitrocellulose sheet stained with ponceau red for unbound SR1 tobacco leaf protein passing through the top nitrocellulose sheet, arrow indicates the threshold at which most or all of the protein binding sites within the nitrocellulose matrix are saturated.

(B) Autoradiograph demonstrating a western slot blot after reacting with a rabbit anti-pea lectin antibody (1:500) and followed by an ^{125} I labelled donkey anti-rabbit IgG of the top nitrocellulose sheet (rows 3 and 4) in figure 3.6 (A) for SR1 leaf protein alone and SR1 leaf protein spiked with dilutions of pea lectin (rows a and b in figure 3.6B). Arrow indicates the concentration range (90ng-0.3ng), overwhich the pea lectin protein could be used, producing clearly defined slots without distortion, for densometric scanning.

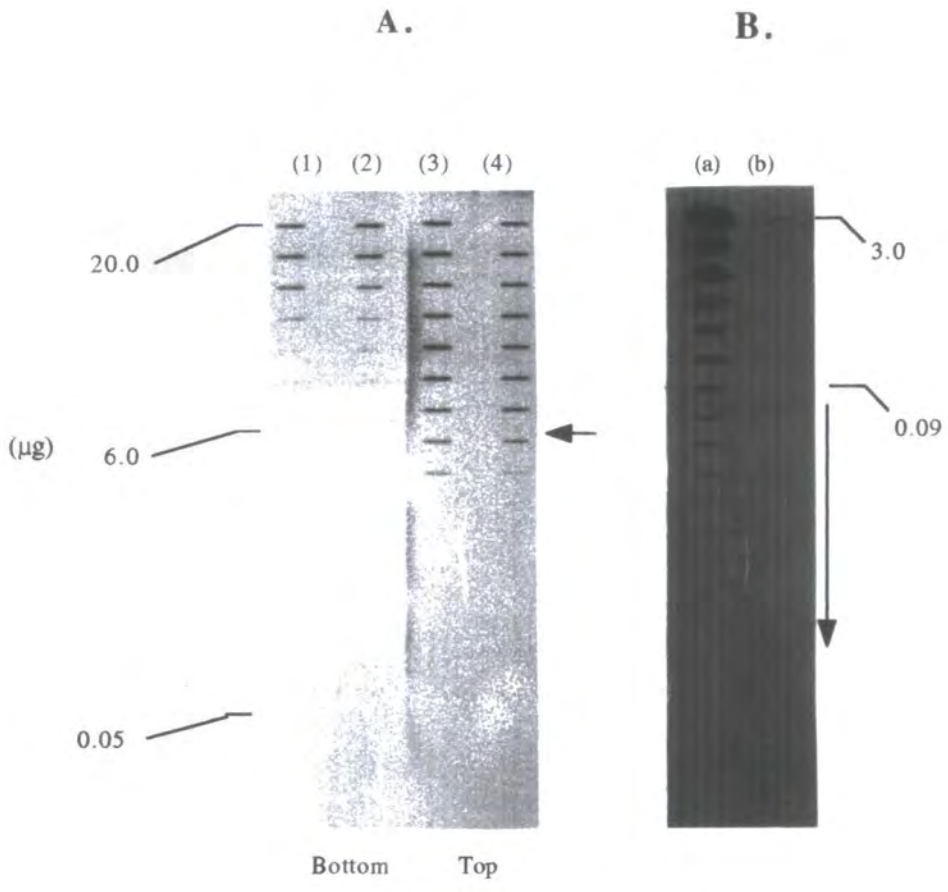

A.
B.
(1)
(2)
(3)
(4)
(a)
(b)
20.0
(μg)
6.0
0.05
3.0
0.09
Bottom
Top

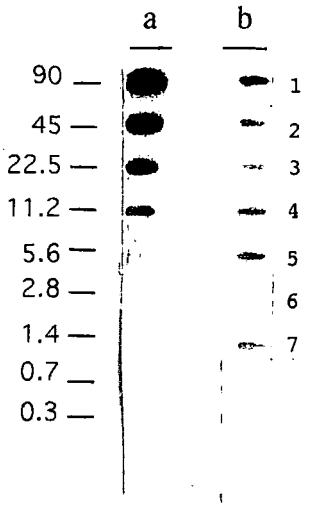
Figure 3.7:

(A) Autoradiograph of a western slot blot after probing with a pea lectin antibody and a ^{125}I labelled IgG of total soluble leaf protein extracted from seven *CaMV-lecA* line 4 individuals, aligned with pea lectin protein standards and SR1(untransformed plant) leaf soluble protein.

Row a: Two-fold dilution series of purified pea lectin protein (90ng-0.3ng), arrow indicates 4µg of SR1 (untransformed plant) total leaf soluble protein. Row b: Total soluble protein extracts (4µg per slot in PBS) from the first upper leaf of *CaMV-lecA* T1 individuals numbered 1-7, at the six leaf stage.

(B) Pea lectin levels in seven *CaMV-lecA* line L4 T1 tobacco plants for the first upper leaf of tobacco plants at the six leaf stage. Pea lectin levels in the seven T1 individuals (figure 3.7A) were determined as a percentage of total soluble leaf protein for three separate assays, the mean value calculated and the lectin levels for each plant arranged in ascending order. Levels of pea lectin protein were compared with gene copy estimations in figure 3.3B (see section 3.2.2 for more detail).

A.



(ng)

T1 population of
Camv-lecA L4

B.

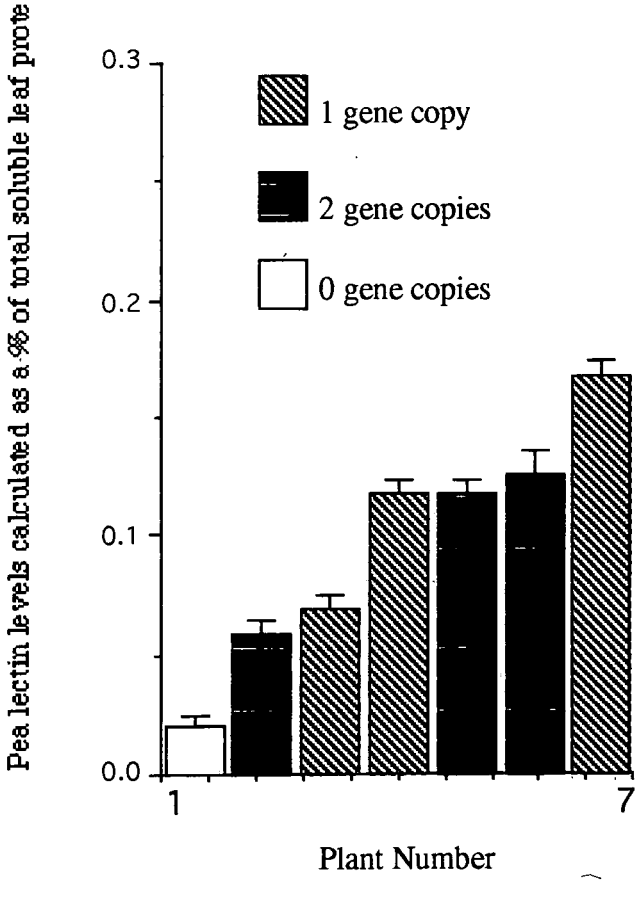
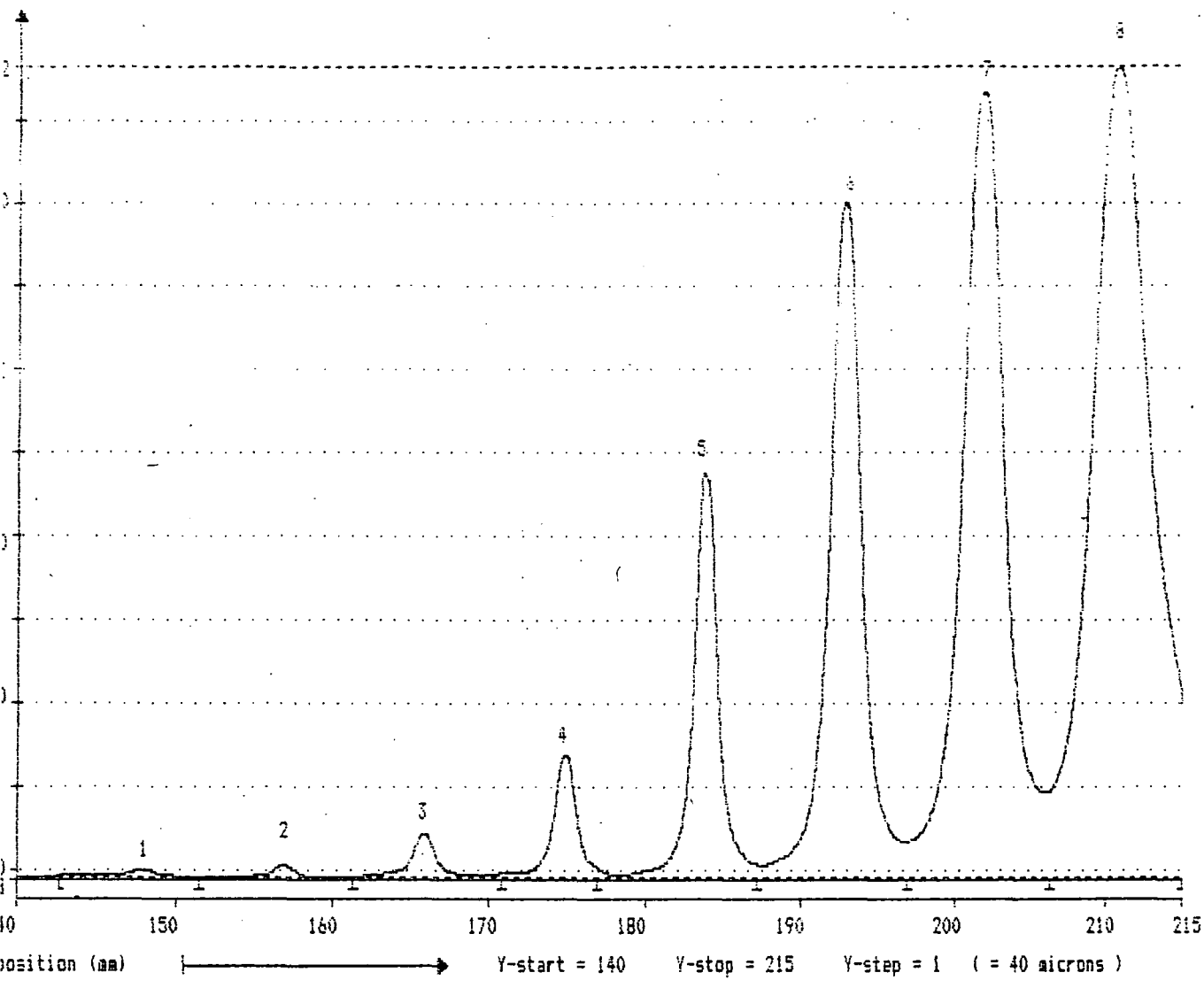


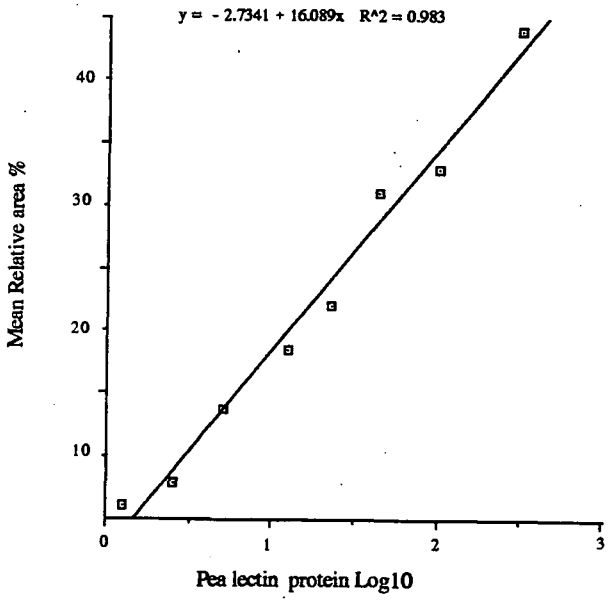
Figure 3.8:

(A) Densitometric trace of the serial two-fold dilutions of purified pea lectin in figure 3.7(A). The autoradiogram of the dilution series of purified pea lectin in (row a figure 3.7(A)) was scanned with a laser densitometer (LKB Ultroscan, LKB Biotechnologies). The peaks numbered from the right 8-1 correspond to pea lectin two-fold serial dilutions from 90ng down to 0.3ng. The mean relative area percentages under the peak were integrated and the data used to construct a standard curve.

(B) A plot of the mean relative area under the peaks (8-1) with the $\log_{10}$ of pea lectin concentrations (90ng-0.3ng). Areas under the peaks were determined for all 10 *CaMV-lecA* T1 individuals and amounts of detectable pea lectin protein estimated by extrapolating from the standard curve.


1
2
3
4
5
150
160
170
180
190
200
210
215
position (mm)
Y-start = 140
Y-stop = 215
Y-step = 1
(= 40 microns)

B.


y = - 2.7341 + 16.089x R^2 = 0.983
Mean Relative area %
10
20
30
40
0
1
2
3
Pea lectin protein Log10

homozygous individual. In the analysis it was postulated that variation of pea lectin protein levels within the T1 population would reflect the presence of *lecA* hemizygous and homozygous individuals. In most cases, however, pea lectin protein levels were discovered to vary only slightly between individuals of the T1 population, with estimates of pea lectin levels in hemizygous individuals similar to those of homozygous individuals. It may thus be seen that variabilty in levels of pea lectin production in the T1 population did not necessarily correlate with differences in copy number. For example, plant 1 containing one *lecA* copy displayed greater amounts of pea lectin protein on comparison with plant 7, found to contain 2 *lecA* copies. A linear correlation coefficient r= -0.101 was obtained for the relationship of copy number and pea lectin levels within the L4 T1 population. Comparison of the calculated r value (-0.101) with that of the tabulated r value (0.7545) indicated no significant correlation at the 5% confidence level. In other words no significant correlation could be found between the number of *lecA* copies and levels of pea lectin protein. Therefore based upon the limited number of T1 plants analysed it can be concluded that differences in copy number do not fully account for the differences in pea lectin levels between individuals of a T1 population.

3.2.4 Concluding Remarks.

It was evident from the germination results that only one independent transgenic line could be used to examine position effect, *CaMV-lecA* line L4. Genetic and Southern analysis identified L4 T1 plants containing one copy of the *CaMV-lecA* gene integrated at a single locus given the assumption that both the *CaMV-lecA* and *nos-nptII* gene were carried on the same T-DNA. Correlation of pea lectin levels in leaves of the T1 population with *lecA* gene copy number concluded that an increase in expression through doubling of gene copy number did not necessarily result, as in most cases expression levels of homozygous individuals were identical or similar to those of hemizygous progeny. The reasons for the observed similarity may include the silencing of other copies of the transgene via methylation (Matzke *et al.*, 1993) or, as may be the

case in plant 4, only intact copies of the *CaMV-lecA* gene being transcriptionally active and other copies inactive due to rearrangement.

It was concluded from the above assessment that the observed frequency of apparent rearrangements within the *lecA* gene coupled with the lack of a sufficient number of independent transgenic lines, and consequentially the lack of material containing T-DNA integrated at different positions within the genome necessitated the need to generate more transgenic lines in order to adequately carry out this study.

3.3 Production of a large population of independent transgenic *lecA* tobacco plants.

3.3.1 Plant transformation vectors.

The production of a large population of independent transgenic *lecA* tobacco plants was performed via leaf disc *Agrobacterium tumefaciens* transformation of *Nicotiana tabacum* (SR1). The *Agrobacterium* transformation vectors used in this study were based on the 'disarmed' Ti plasmid vector pGV3850 (Zambryski *et al.*, 1983). This vector comprises a Ti-plasmid containing a T-DNA from which all the oncogenic genes have been deleted and replaced by the cloning vehicle pBR322. Any sequence cloned in pBR322 derivatives can then be conjugated into *Agrobacterium* and inserted between the T-DNA borders by a single homologous recombination event between the two plasmid sequences. Since pBR322 vectors cannot replicate in *Agrobacterium*, exconjugants harbouring the co-integrates can be easily isolated by selecting for an antibiotic resistant marker present in the pBR322 derivative which is absent from the Ti-plasmid.

The 35S cauliflower mosaic virus promoter (*CaMV*)-pea (*Pisum sativum*) lectin gene construct (pDUB129) and the tobacco ribulose bisphosphate carboxylase small subunit (*ssRubisco*) promoter-pea lectin gene construct (pDUB132) were both made by Edwards *et al.*, (1991), as shown in figure 3.9 (A) and (B). Both vectors consisted of either the 810bp 35S *CaMV* promoter derived from pROK2 (Bevan, 1984) or the

1020bp tobacco *ssRubisco* promoter (Bevan, 1984) as a transcriptional fusion to 1.25Kb of coding and 403bp of 3' untranslated sequence (including the polyadenylation site) of the pea lectin gene (Gatehouse *et al.*, 1987). This fusion replaced the native *lecA* gene transcription start site and 5' controlling sequences that confer seed specificity with the 35S *CaMV* promoter conferring constitutive expression in all tissues, or the photosynthesis-specific promoter from the *ssRubisco,* allowing expression in green tissue of transgenic tobacco. Sequence analysis of the pea lectin gene identified an extra ATG intitiation codon, within the *SphI* site used for both constructs. This extra upstream ATG is out of frame with the *lecA* coding sequence in that translation initiation from the upstream ATG would result in 6bp encoding two amino acids followed by a termination codon. Based on its efficiency to direct translational initiation a strong ATG is defined by the presence of a purine at position -3 and a guanine in position +1 (Kozak, 1987; Guerineau *et al.*, 1992). Thus when the sequences surrounding both ATG's are compared the downstream ATG falls into the optimum context.

The intermediate vectors pDUB130 and pDUB133 (figure 3.10) were generated by cloning the *HindIII/EcoRI* fragment of pDUB129 and pDUB132 into the multiple cloning site of pDUB126A, a pBR322 based cloning vector (Edwards *et al.*, 1991). These plasmids contain the following properties: pBR322 homology (including the ampicillin resistance gene), *nos/nptII* gene, *bom* (basis of mobolisation) site, and a gene encoding kanamycin/gentamycin resistance in bacteria. These intermediate vectors were then transferred to *A.tumefaciens* harbouring the Ti-plasmid pGV3850 by a two-step mating technique with the aid of *E.coli* strain GJ23. This *E.coli* strain contains two plasmids: R64 drd11, which has a tetracycline resistance (Tc^r) gene and is fully auto-transferable and pGJ28 which has a Km^r/Nm^r gene and is non-conjugable but can be mobolised in the presence of a conjugable plasmid. It does this by means of a mob protein which acts on a *bom* site to initiate transfer. If another plasmid is present containing a *bom* site, the mob protein will act in *trans* and this plasmid (i.e. pDUB130 or pDUB133) will also be mobilised to the acceptor strain of *A.tumefaciens*.. Once in

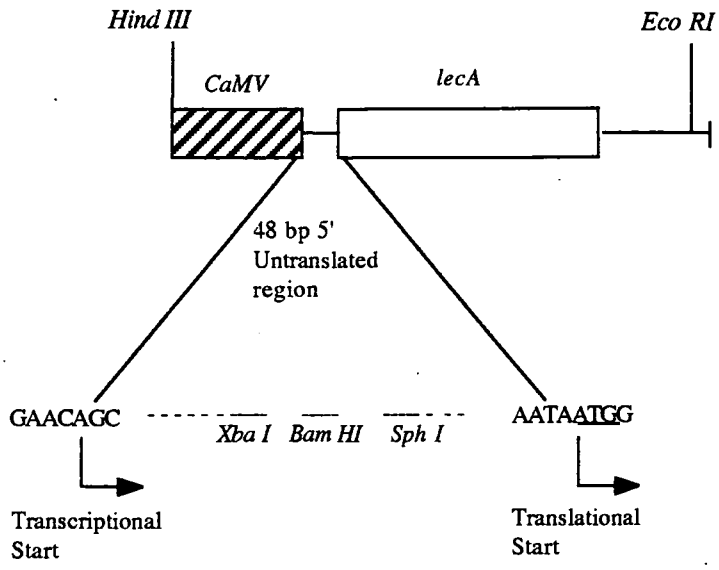
Figure 3.9:

(A) Detail of the plasmid pDUB129 containing the *CaMV-lecA* transcriptional fusion with the pea *lecA* gene, cloned as a *HindIII/ EcoRI* fragment into the *lacZ* cloning region of pUC18. *CaMV*=810bp cauliflower mosaic virus 35S promoter, *lecA*=25Kb and 403bp of 3' untranslated sequence of the pea lectin gene.

(B) Detail of the plasmid pDUB132 containing the *ssRubisco-lecA* transcriptional fusion with the pea *lecA* gene cloned as a *HindIII/EcoRI* fragment into the *lacZ* cloning region of pUC18. *ssRubisco*=1020bp tobacco ribulose bis-phosphate carboxylase oxygenase small subunit promoter, *lecA*=1.25Kb and 403bp of 3' untranslated sequence of the pea lectin gene.

pDUB129

A.



pDUB132

B.

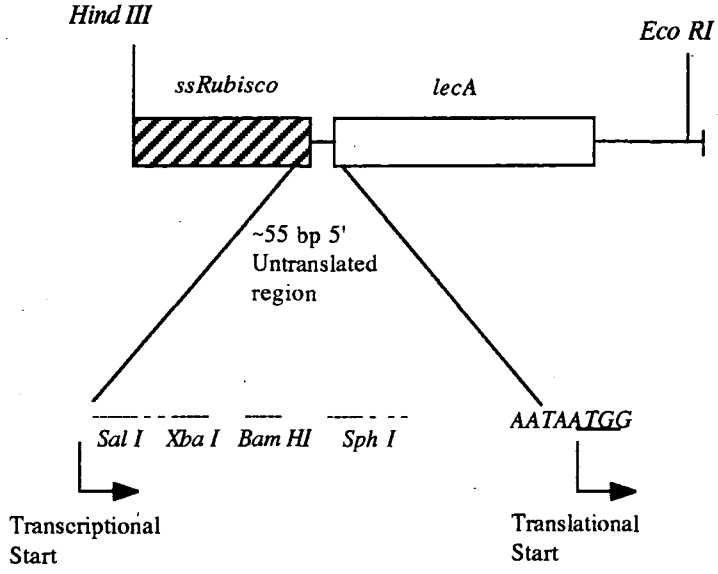


Figure 3.10:

(A) Isolated *HindIII/EcoRI CaMV-lecA* and *ssRubisco-lecA* chimaeric genes for cloning into the *HindIII/ EcoRI* sites of pDUB126A.

(B) Detail of cloning vector pDUB126A. *nptII*=neomycin phosphotransferase II gene, conferring kanamycin resistance in plants, *Km/Gm*=kanamycin/gentamycin resistance in *E.coli/Agrobacterium*, Amp=ampicillin resistance in *E.coli*.

(C) Resultant intermediate vector pDUB130 containing the chimaeric *CaMV-lecA* gene inserted into the *HindIII/EcoRI* sites of pDUB126A. *nptII*= neomycin phosphotransferase II gene, conferring kanamycin resistance in plants, *Km/Gm*= kanamycin/gentamycin resistance in *E.coli/ Agrobacterium*, *Amp*= ampicillin resistance in *E.coli*.

(D) Resultant intermediate vector pDUB133 containing the chimaeric *ssRubisco-lecA* gene inserted into the *HindIII/EcoRI* sites of pDUB126A. *nptII*=neomycin phosphotransferase II gene, conferring kanamycin resistance in plants, *Km/Gm*=kanamycin/gentamycin resistance in *E.coli/Agrobacterium, Amp*=ampicillin resistance in *E.coli*.

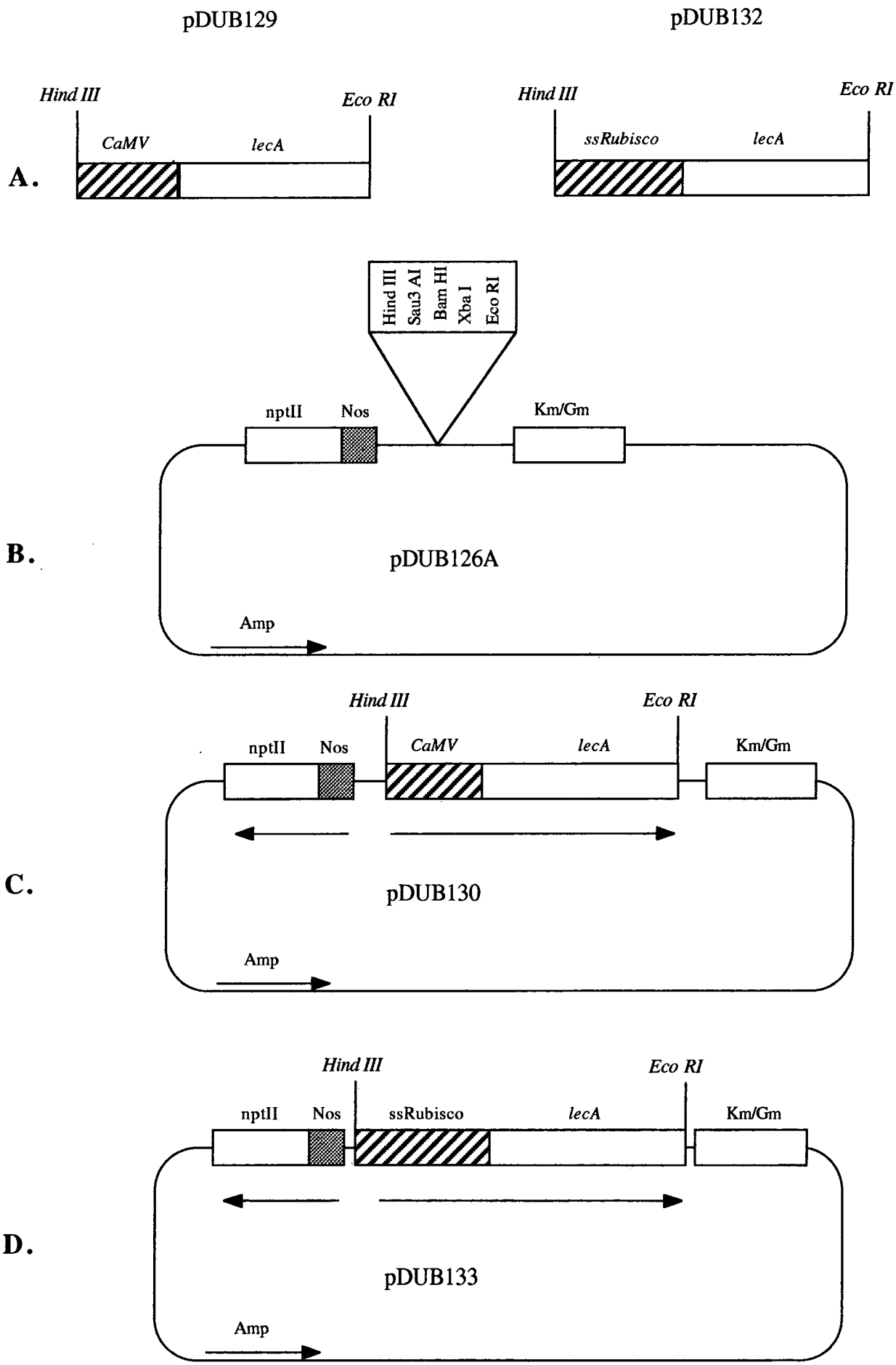

pDUB129
pDUB132
A.
Hind III
Eco RI
CaMV
lecA
Hind III
Eco RI
ssRubisco
lecA
B.
Hind III
Sau3 AI
Bam HI
Xba I
Eco RI
nptII
Nos
Km/Gm
pDUB126A
Amp
C.
Hind III
Eco RI
nptII
Nos
CaMV
lecA
Km/Gm
pDUB130
Amp
D.
Hind III
Eco RI
nptII
Nos
ssRubisco
lecA
Km/Gm
pDUB133
Amp

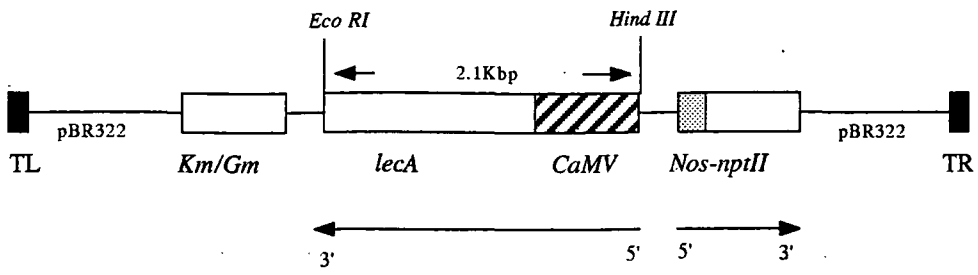
Figure 3.11:

(A) The arrangement of the *CaMV-lecA* chimaeric gene within the T-DNA expression cassette of the plant transformation vector pGV3850. *CaMV*=cauliflower mosaic virus 35S promoter, *lecA*=pea lectin coding region, *nos*=nopaline synthase promoter, *nptII*=neomycin phosphotranspherase II gene conferring kanamycin resistance in plants, Km^r/Gm^r=kanamycin/gentamycin resistance in *E.coli/Agrobacterium*, pBR322=plasmid sequence of pBR322 containing the ampicillin resistant gene, TL & TR=left and right hand T-DNA border sequences.

(B) The arrangement of the *ssRubisco-lecA* chimaeric gene within the T-DNA expression cassette of the plant transformation vector pGV3850. *ssRubisco*=tobacco small subunit ribulose bisphosphate carboxylase promoter, *lecA*=pea lectin coding region, *nos*=nopaline synthase promoter, *nptII*=neomycin phosphotranspherase II gene conferring kanamycin resistance in plants, Km^r/Gm^r=kanamycin/gentamycin resistance in *E.coli/Agrobacterium*, pBR322=plasmid sequence of pBR322 containing the ampicillin resistant gene, TL & TR=left and right hand T-DNA border sequences.

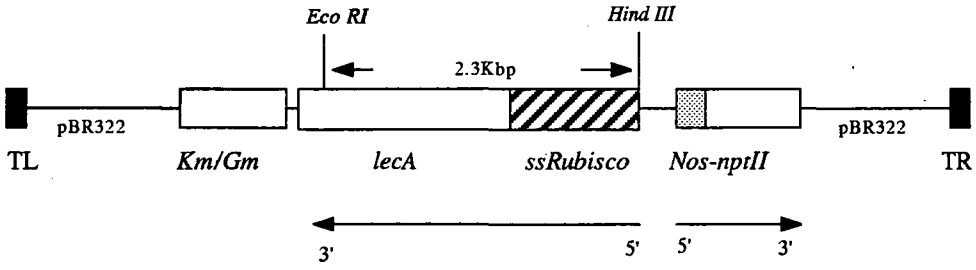
pGV3850::pDUB130 Co-integrate

A.



pGV3850::pDUB133 Co-integrate

B.



the acceptor strain of *A.tumefaciens* the intermediate vectors cannot replicate and the only way to obtain expression of the gentamycin resistance (Gm^r) gene is by homologous recombination of the intermediate vectors with a region of DNA within *A.tumefaciens* (i.e. marker rescue). The disarmed Ti-plasmid pGV3850 as mentioned previously, contains between its T-DNA borders pBR322 homology and can therefore form a co-integrate with the intermediate vectors (Zambyrski *et al*, 1983).

Thus two co-integrates were generated, one (pDUB130) containing the *CaMV* promoter fused to the *lecA* coding sequence, and the other (pDUB133) having the *ssRubisco* promoter fused to the *lecA* coding sequence (figure 3.11). Both were provided by Dr. G. Edwards (Shell Research Ltd.).

3.3.2 Characterisation of *Agrobacterium* transformation vectors pGV3850::pDUB130 and pGV3850::pDUB133.

Prior to employing the above constructs in a plant transformation programme it was deemed prudent to undertake a detailed characterisation of the supplied material.

The co-integrates pGV3850::130 (*CaMV-lecA*) and pGV3850::133 (*ssRubisco-lecA*) in *Agrobacterium* strain GV3101, were selected on YEB-agar containing 100µg/ml of both rifampicin and carbenicillin, and 5µg/ml gentamycin. Total DNA was prepared from cultures of single colonies and the presence of an intact, non-rearranged T-DNA confirmed by restriction analyses on 0.7%(w/v) agarose gels and Southern hybridization with homologous probes. The co-integrates pGV3850::126A (control i.e. containing only the *lacZ* cloning region without an insert), pGV3850::130, and pGV3850::133 were digested with *KpnI, XbaI, XhoI* and hybridized to a ^{32}P-labeled linearized, pBR322 probe (figure 3.12(A) and (B)). The *KpnI* digests (Lanes 1-3) revealed a common 6.1Kb fragment for all three co-integrates and larger fragments of 12.4Kb and 12.6Kb for pGV3850::130 and pGV2850::133 (lanes 2 and 3) when compared with the smaller 10.3Kb fragment of pGV3850::126A (lane 1). Likewise the *XbaI* digestion for all three co-integrates (lanes 4-6) revealed a common 24.9Kb

fragment and larger fragments of 9.5Kb for pGV3850::130 and pGV3850::133 (lanes 5 and 6) when compared with the smaller 8.6Kb fragment of pGV3850::126A (lane 4). The *XhoI* digest on the other hand produced only one fragment with homology to the pBR322 probe suggesting that there was no *XhoI* restriction site within the T-DNA delimited by the left and right hand border sequences. The Southern analysis in figure 3.12 (B) therefore confirmed the presence of both co-integrates pGV3850::130 and pGV3850::133 in *Agrobacterium.*

To obtain detailed restriction maps of the co-integrates pGV3850::130 and pGV3850::133, further restriction analyses were carried out on the same DNA samples with the enzymes *KpnI, BamHI, EcoRI, HindIII*, and *EcoRI/HindIII* followed by Southern hybridization with digoxigenin labelled pBR322 (figure 3.13 (A) and (B)), *CaMV* (figure 3.14 (A) and (B)), *ssRubisco* (figure 3.15 (A)and (B)) or *lecA* (figure 3.14 (B) and 3.15 (B)) probes. The presence of the *CaMV-lecA* gene construct is indicated as a 2.1Kb fragment after hybridzation with the probes to *CaMV* (lane 10) and *lecA* (lane 5) figure 3.14 (B). Similarily the presence of a 2.3Kb *ssRubisco-lecA* gene is confirmed in figure 3.15 (B) after hybridzation with probes to *ssRubisco* (lane 10) and *lecA* (lane 5). The sizes of the major bands of hybridization were determined and are presented in figures 3.16 and 3.17 as fully detailed restriction maps for both co-integrates.

The detailed restriction maps were essential for choosing enzymes and probes for the mapping of the T-DNA sequences found in subsequent transgenic plants. For example digestion with *HindIII* and probing with pBR322 sequences (figure 3.16 (A) and figure 3.17 (A)) will determine whether or not the T-DNA expression cassette has been transferred.

Figure 3.12:

(A) Restriction analysis of total DNA (6μg per lane) isolated from *Agrobacterium* (GV3101), harbouring the co-integrates pGV3850::pDUB126A, pGV3850::pDUB130, and pGV3850::pDUB133 T-DNA. Lanes 1-3: pGV3850::pDUB126A, pDUB130, and pDUB133 digested with *KpnI*. Lanes 4-6: digested with *XbaI*. Lanes 7-9: digested with *XhoI*.

(B) Autoradiogram of a Southern blot of the gel in figure 3.12(A) hybridised to a ^{32} P-labelled linearised 4.3Kb pBR322 probe. Lanes 1-3: pGV3850::pDUB126A, pDUB130, and pDUB133 digested with *KpnI*. Lanes 4-6: digested with *XbaI*. Lanes 7-9: digested with *XhoI*.

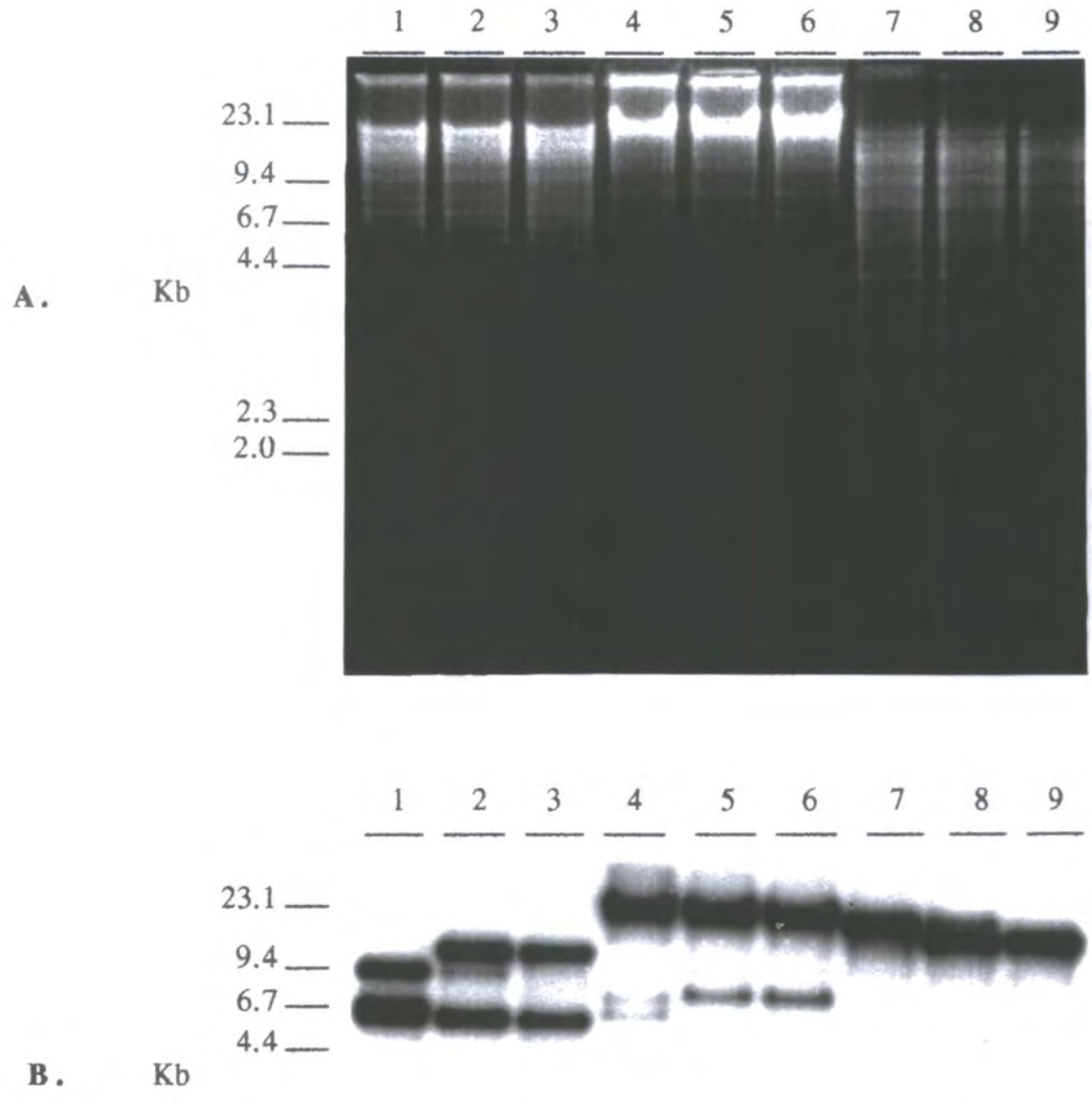
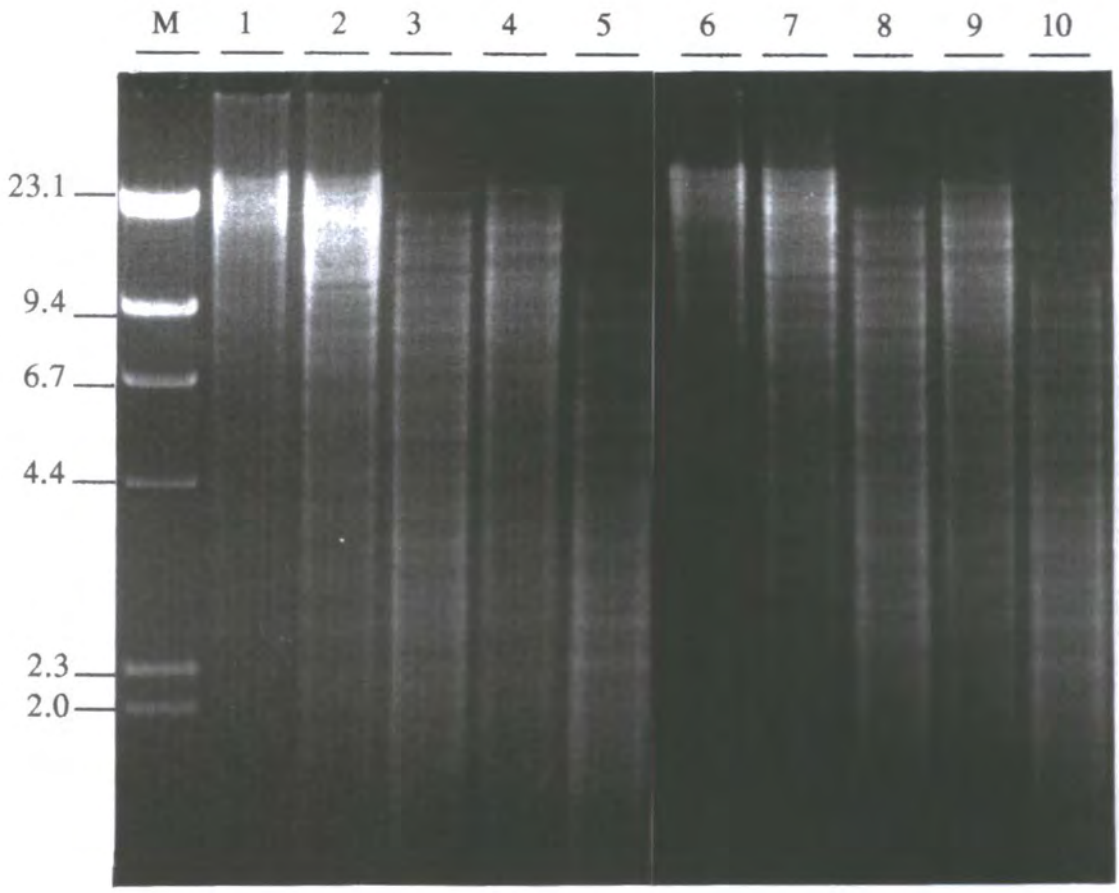

1 2 3 4 5 6 7 8 9
23.1
9.4
6.7
4.4
Kb
2.3
2.0
A.
1 2 3 4 5 6 7 8 9
23.1
9.4
6.7
4.4
Kb
2.3
2.0
B.

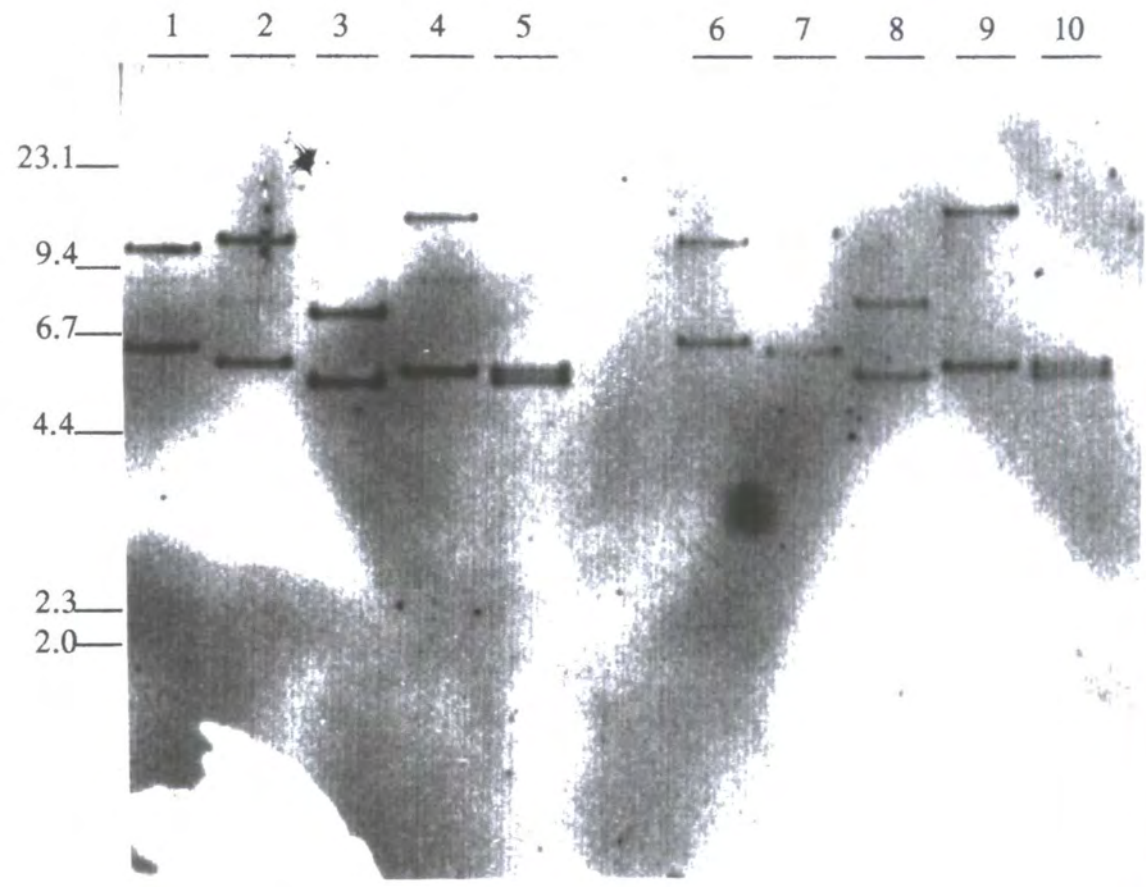
Figure 3.13:

(A) Restriction analysis of Total DNA (6µg per lane) isolated from *Agrobacterium* (GV3101) harbouring the co-integrates pGV3850::pDUB130, and pGV3850::pDUB133. Lanes 1-5: pGV3850::pDUB130A, digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.* Lanes 6-7: pGV3850::pDUB133 digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.*

(B) Stained nitrocellulose filter of the gel shown in figure 3.13(A) after southern transfer, and hybridisation to a digoxigenin labelled, linearised 4.3Kb pBR322 probe. Lanes 1-5: pGV3850::pDUB130A, digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.* Lanes 6-7: pGV3850::pDUB133 digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.* Detection of fragments homologous to the probe was achieved with colorimetric substrates nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) after hybridisation with an anti-digoxigenin alkaline phosphatase conjugate antibody (1:5,000).


M
1
2
3
4
5
6
7
8
9
10
23.1
9.4
6.7
4.4
2.3
2.0
Kb

A.


1
2
3
4
5
6
7
8
9
10
23.1
9.4
6.7
4.4
2.3
2.0
Kb

B.

Figure 3.18:

(A) Leaf squashes of primary regenerants in tissue culture after co-cultivation with *Agrobacterium* (GV3101) harbouring the co-integrate pGV3850::130.

(B) Autoradiograph of the immunoblot in figure 3.18(A) after probing with affinity purified rabbit anti-pea lectin antibodies (1:2000) and detection after reaction with ^{125}I labelled IgG antibodies (1:5000). 200ng of purified pea lectin and a SR1 untransformed leaf squash blot are indicated by arrows.

(C) Autoradiograph of an immunoblot of further primary regenerant leaf squashes from transformations with the co-integrate pGV3850::130 after reaction with primary rabbit anti-pea lectin antibodies (1:2000), secondary donkey anti-rabbit IgG ^{125}I conjugated antibodies 1:5000) and detection via autoradiography. Numbers represent individual transgenic lines (plants 55 -114 as shown)

3.3.3 Production of *Nicotiana tabacum* primary regenerants containing the *CaMV-lecA* and *ssRubisco-lecA* chimaeric genes.

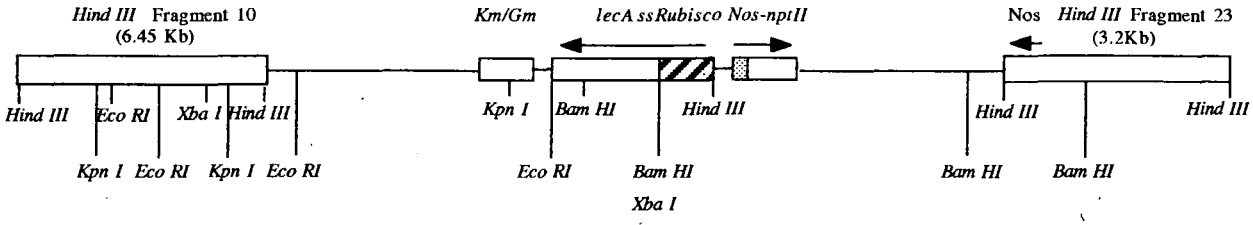
Having confirmed the integrity of the co-integrates pDUB130 and pDUB133 in *Agrobacterium* these chimaeric gene constructs were transferred to the genome of *Nicotiana tabacum* SR1 plants via leaf disc transformation (Horsch *et al.*, 1985). Leaf discs were infected with *Agrobacterium* cultures harbouring either co-integrate, the chimaeric 2.1Kb *CaMV-lecA* or 2.3Kb *ssRubisco-lecA* constructs. Shoot initiation took place within 2-3 weeks on Murashige and Skoog (MS) agar (Murashige and Skoog, 1962) supplemented with 2% (w/v) sucrose, 0.1 mg/l α-napthalene acetic acid (NAA), 0.5 mg/l benzylaminopurine (BAP) under kanamycin selection of 100mg/l. To ensure selection of independent transformation events, not more than two shoots normal in appearance were taken per leaf disk and these from opposite sides of the disc, avoiding associated callus. Rooting of shoots and further development into small plantlets was achieved by transferring the kanamycin resistant shoots onto 1/2 MS agar supplemented with 1%(w/v) sucrose containing 100mg/l kanamycin but lacking phytohormones.

The transgenic nature of regenerated plantlets was confirmed using detection of pea lectin with affinity purified rabbit anti-pea lectin antibodies performed on leaf tissue squashes of shoots with three to four leaves (figures 3.18, 3.19). Figure 3.18 (A) displays 54 squashes of whole leaves or leaf discs on nitrocellulose, taken from the leaves of shoots arising from independent transformation events with co-integrate pGV3850::130, together with SR1 and pea lectin protein (200ng) controls. After incubation with lectin antibodies and ^{125}I labelled anti rabbit IgG antibodies five shoots were identified by autoradiography as accumulating pea lectin protein in figure 3.18 (B). Likewise in figure 3.19 (A), 24 leaf squashes are shown for individual shoots after transformation with pGV3850::133. Subsequent immunoblotting with rabbit anti pea lectin antibodies and ^{125}I conjugated donkey anti rabbit IgG antibodies (figure 3.19 (B)) showed the majority of the shoots accummulating pea lectin protein when

compared with the relative controls (SR1 and purified pea lectin). Transformation was somewhat variable as judged by the variable number of positive pea lectin shoots obtained and therefore it was necessary to continually pass SR1 material through the transformation procedure to generate a sufficiently large population of primary regenerants for both chimaeric genes. Figures 3.18 (C) and 3.19 (C) represent leaf squashes from shoots selected from further rounds of transformation with pGV3850::130 and pGV3850::133 co-integrates.

In figures 3.20 and 3.21 the localisation of pea lectin accumulation in whole transgenic plantlets is demonstrated. The plantlets representing primary regenerants were first squashed onto nitrocellulose and stained with Ponceau red (figures 3.20 (A) and 3.21 (A)) before reaction with antibodies and autoradiography. This allowed useful comparisons between the Ponceau red stained images and autoradiograms, demonstrating that the absence of a reaction signal is not the result of poor or ineffective squashing of the plantlets. The pattern of pea lectin accumulation under control of the *CaMV* promoter (figure 3.20 (B)) was detected in most tissues of the primary regenerant including the leaves, stem and roots. On the other hand pea lectin accumulation under the *ssRubisco* promoter was only ever detected only in leaves and not in the roots (figure 3.21 (B)). Quantitative studies conducted by Edwards *et al.*, (1991) for the *CaMV-lecA* chimaeric gene in transgenic potato plants revealed similar patterns of expression, in leaves, stems, and roots. In *ssRubisco-lecA* potato transformants, Edwards *et al.*, (1991) also observed high levels of pea lectin in the shoot apex and expanded leaves; however no lectin protein was detected in roots or tubers, consistent with the results displayed in figure 3.21 (B).

Rooting shoots were maintained in tissue culture for 1 month before being transferred to soil to ensure their continued resistance to kanamycin. From observations of primary regenerants in tissue culture between 16% and 18% of all primary regenerants did not develop roots, their leaves were usually curled and elongated and proliferating to high numbers. Such abnormal phenotypes may be a consequence of somaclonal variation



A.

Kpn I	6.1 kb	12.6 Kb
Xho I	14.4 Kb	
Xba I	9.5 Kb	24.9 Kb
Eco RI / Hind III	6.2 Kb	6.8 Kb
Hind III	8.5 Kb	6.8 Kb
Eco RI	6.2 Kb	17.6 Kb
Bam HI	15.3 Kb	6.7 Kb

B.

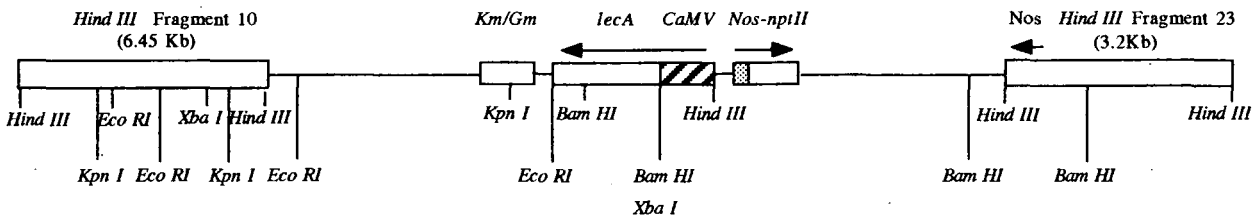
Kpn I	12.6 Kb	
Bam HI	0.7Kb	
Eco RI	16.5 Kb	
Hind III	8.6 Kb	
Eco RI / Hind III	2.3 Kb	

C.

Kpn I	12.6 Kb	
Bam HI	6.7 Kb	
Eco RI	16.5 Kb	
Hind III	8.6 Kb	
Eco RI / Hind III	2.3 Kb	

Figure 3.17:

Detailed restriction map and homologous probes to the *Agrobacterium* co-integrate pGV3850::133. (A) T-DNA fragments homologous to a ^{32}P-labelled linearised 4.3KB pBR322 probe, (B) T-DNA fragments homologous to a digoxigenin labelled 700bp *lecA* coding sequence, (C) T-DNA fragments homologous to a digoxigenin labelled 1020bp *ssRubisco* promoter fragment. Arrows above genes represent direction of transcription.


Hind III Fragment 10
(6.45 Kb)
Km/Gm
lecA
CaMV
Nos-nptII
Nos
Hind III Fragment 23
(3.2Kb)
Hind III
Eco RI
Xba I
Hind III
Kpn I
Eco RI
Kpn I
Eco RI
Kpn I
Bam HI
Hind III
Eco RI
Bam HI
Xba I
Hind III
Hind III
Bam HI
Bam HI

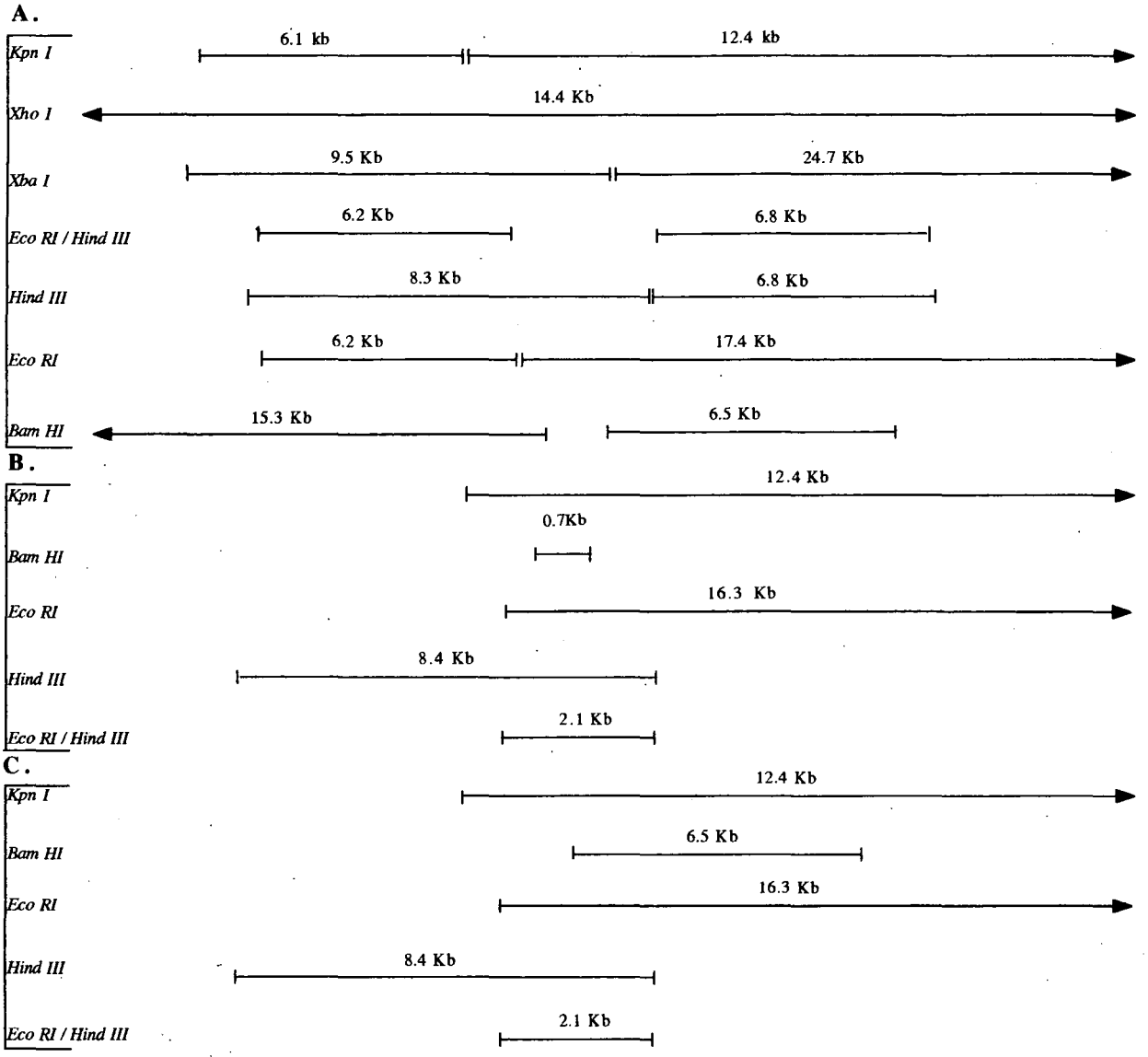
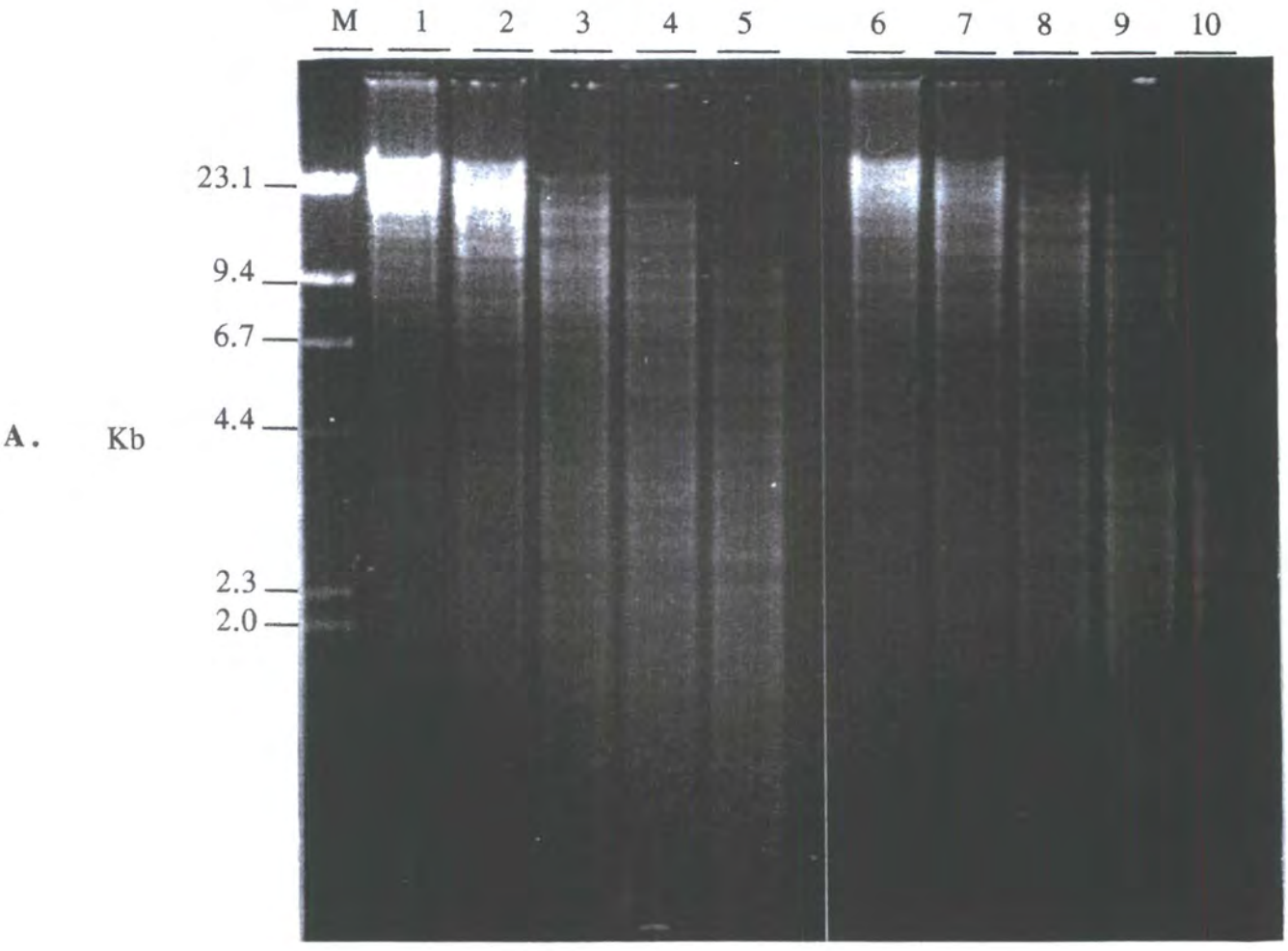
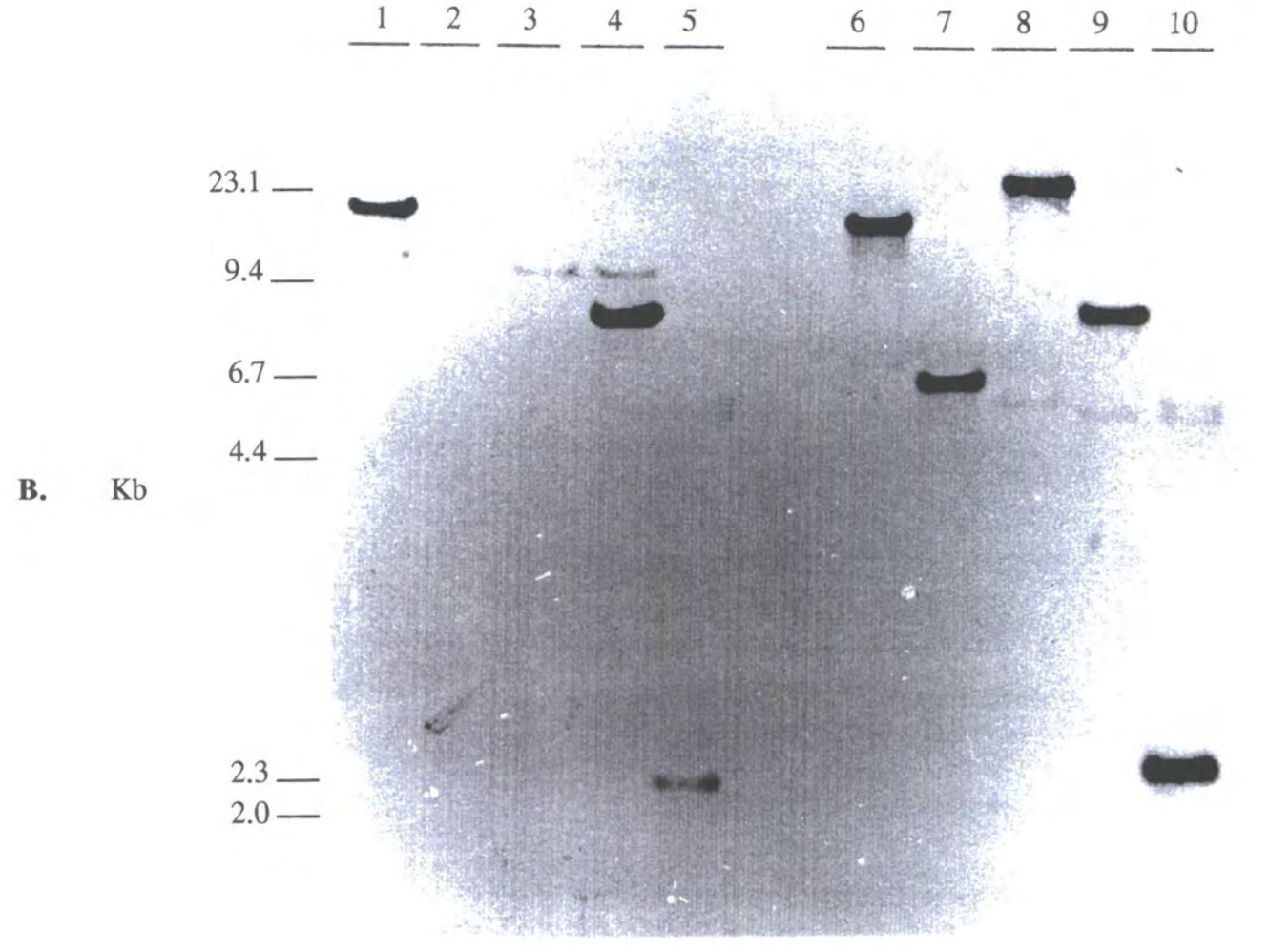

A.
Kpn I
6.1 kb
12.4 kb
Xho I
14.4 Kb
Xba I
9.5 Kb
24.7 Kb
Eco RI / Hind III
6.2 Kb
6.8 Kb
Hind III
8.3 Kb
6.8 Kb
Eco RI
6.2 Kb
17.4 Kb
Bam HI
15.3 Kb
6.5 Kb
B.
Kpn I
12.4 Kb
Bam HI
0.7Kb
Eco RI
16.3 Kb
Hind III
8.4 Kb
Eco RI / Hind III
2.1 Kb
C.
Kpn I
12.4 Kb
Bam HI
6.5 Kb
Eco RI
16.3 Kb
Hind III
8.4 Kb
Eco RI / Hind III
2.1 Kb

Figure 3.16:

Detailed restriction map and homologous probes to the *Agrobacterium* co-integrate pGV3850::130. (A) T-DNA fragments homologous to a ^{32}P-labelled linearised 4.3KB pBR322 probe, (B) T-DNA fragments homologous to a digoxigenin labelled 700bp *lecA* coding sequence, (C) T-DNA fragments homologous to a digoxigenin labelled 810bp *CaMV* promoter fragment. Arrows above genes represent direction of transcription.


M 1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.7
4.4
2.3
2.0
Kb

A.


1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.7
4.4
2.3
2.0
Kb

B.

Figure 3.15:

(A) Restriction analysis of total DNA (6µg per lane) isolated from *Agrobacterium* (GV3101) harbouring the co-integrate pGV3850::pDUB133. Lanes 1-5: pGV3850::pDUB133, digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.* Lanes 6-7: pGV3850::pDUB133 digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.*

(B) Stained nitrocellulose filter of the gel shown in figure 3.15(A) after southern transfer, and hybridisation to a digoxigenin labelled 700bp *BamHI* fragment of pDUB80 containing part of the pea lectin coding region (lanes 1-5) and a digoxigenin labelled 1020bp *HindIII/BamHI* fragment from pDUB132 containing the tobacco *ssRubisco* promoter region (lanes 6-10). Detection of fragments homologous to the probes was achieved with colourmetric substrates nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) after hybridisation with an anti-digoxigenin alkaline phosphatase conjugate antibody (1:5,000).

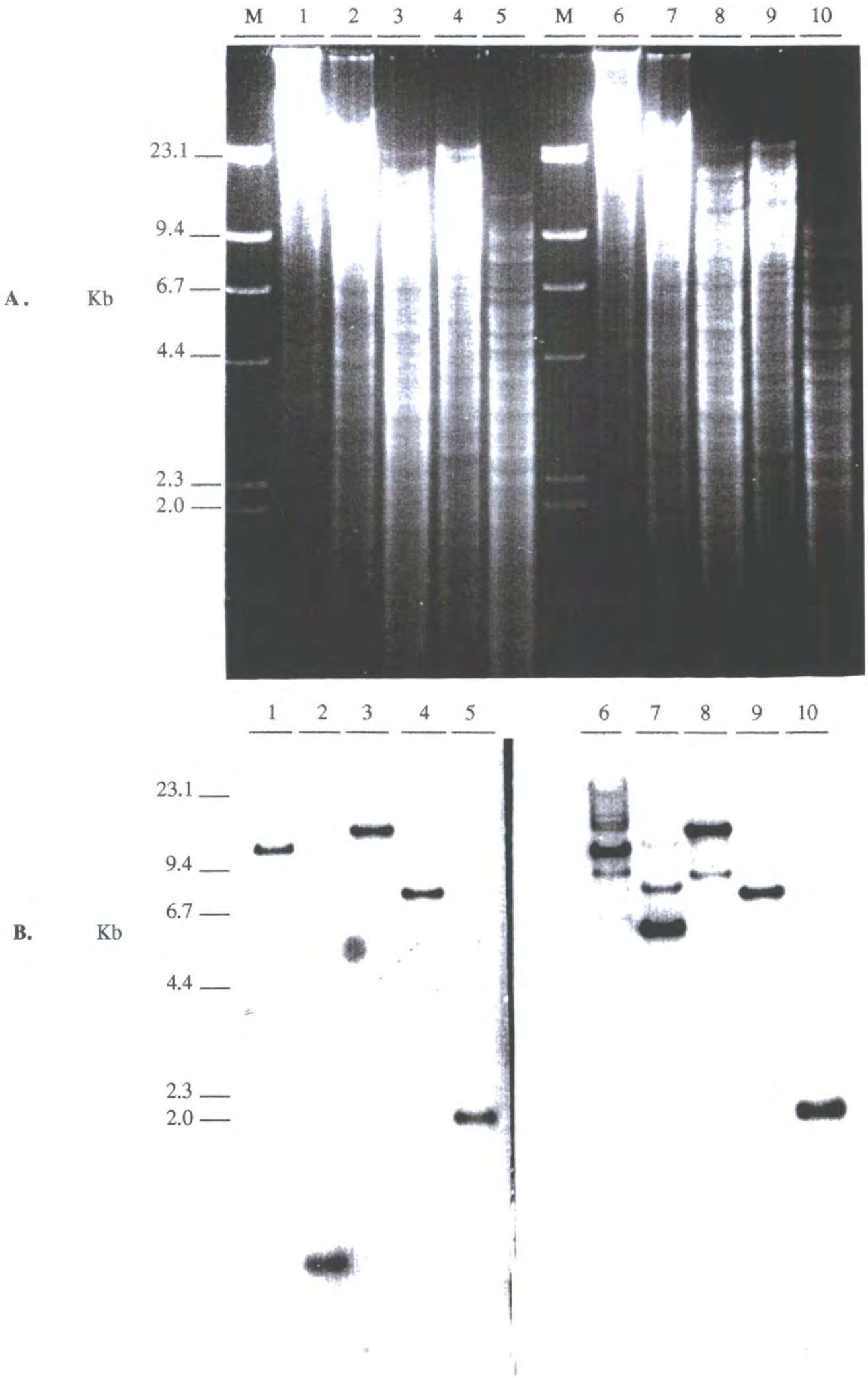

M 1 2 3 4 5 M 6 7 8 9 10
23.1
9.4
6.7
4.4
2.3
2.0
Kb
A.
1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.7
4.4
2.3
2.0
Kb
B.

Figure 3.14:

(A) Restriction analysis of total DNA (6μg per lane) isolated from *Agrobacterium* (GV3101) harbouring the co-integrate pGV3850::pDUB130. Lanes 1-5: pGV3850::pDUB130, digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.* Lanes 6-7: pGV3850::pDUB130 digested with *KpnI, BamHI, EcoRI, HindIII* and *EcoRI/HindIII.*

(B) Stained nitrocellulose filter of the gel shown in figure 3.14(A) after southern transfer, and hybridisation to a digoxigenin labelled 700bp *BamHI* fragment of pDUB80 containing part of the pea lectin coding region (lanes 1-5) and a digoxigenin labelled 810bp *HindIII/BamHI* fragment from pDUB129 containing the 35S *CaMV* promoter region (lanes 6-10). Detection of fragments homologous to the probes was achieved with colorimetric substrates nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) after hybridisation with an anti-digoxigenin alkaline phosphatase conjugate antibody (1:5,000).

A.

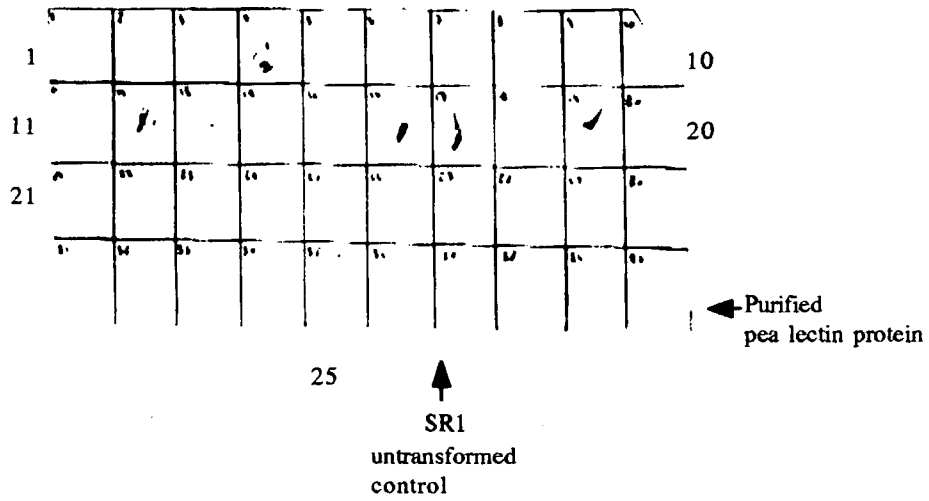
B.

Figure 3.20:

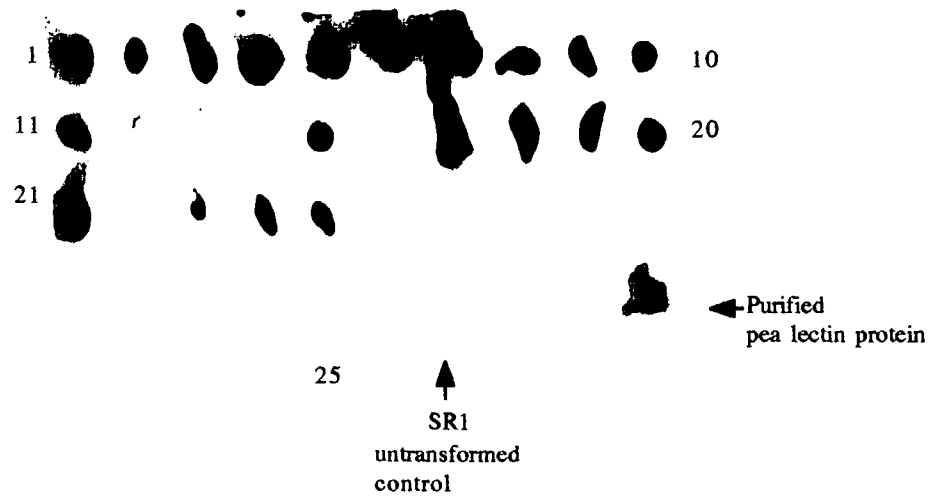
(A) Nitrocellulose filter of a whole plantlet squash, stained with Ponceau red (0.2%(w/v) in 3%(w/v) TCA), for one example of a kanamycin resistant primary regenerant after transformation with *CaMV-lecA* construct

(B) Immunoblot of a whole plantlet squash in figure 3.20(A) after reaction with pea lectin antibodies(1:2,000), and ^{125}I labelled IgG antibodies(1:5000) detected via autoradiography.

A.


1
10
11
20
21
Purified
pea lectin protein
25
SR1
untransformed
control

B.


1
10
11
20
21
Purified
pea lectin protein
25
SR1
untransformed
control

C.

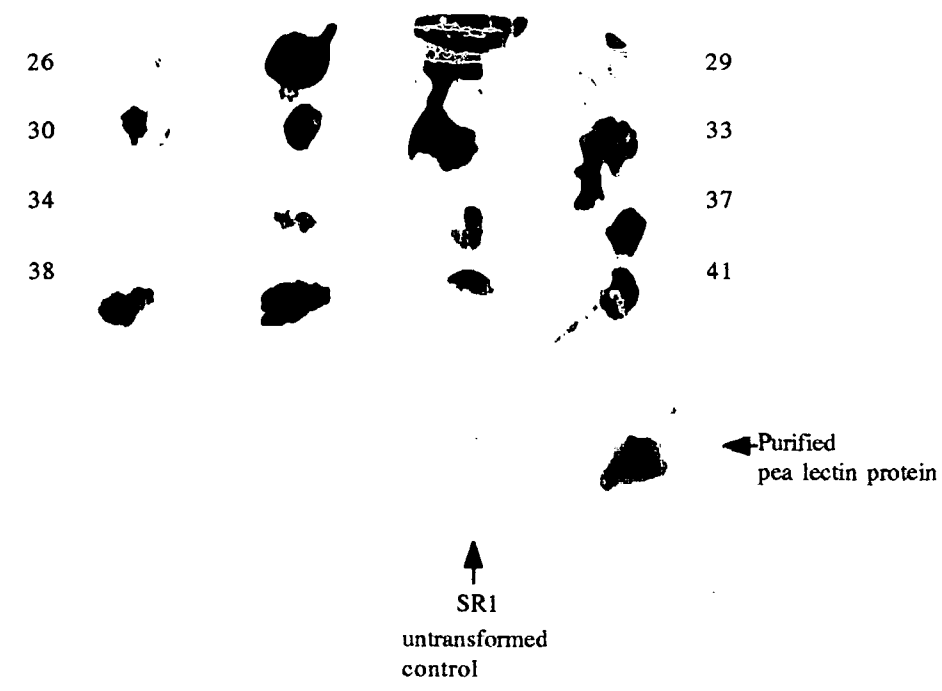

26
29
30
33
34
37
38
41
Purified
pea lectin protein
SR1
untransformed
control

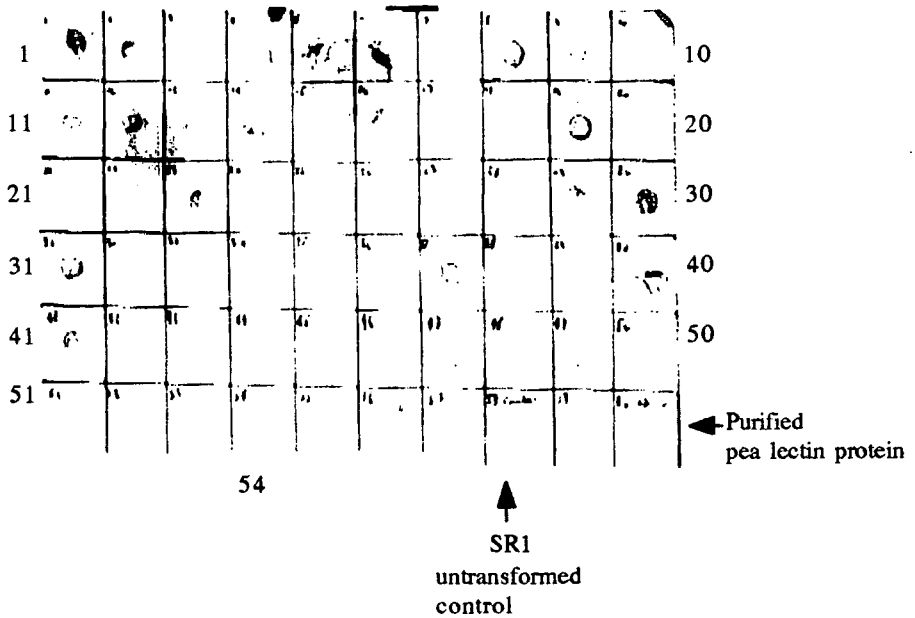
Figure 3.19:

(A) Leaf squashes of primary regenerants in tissue culture after co-cultivation with *Agrobacterium* (GV3101) harbouring the co-integrate pGV3850::133.

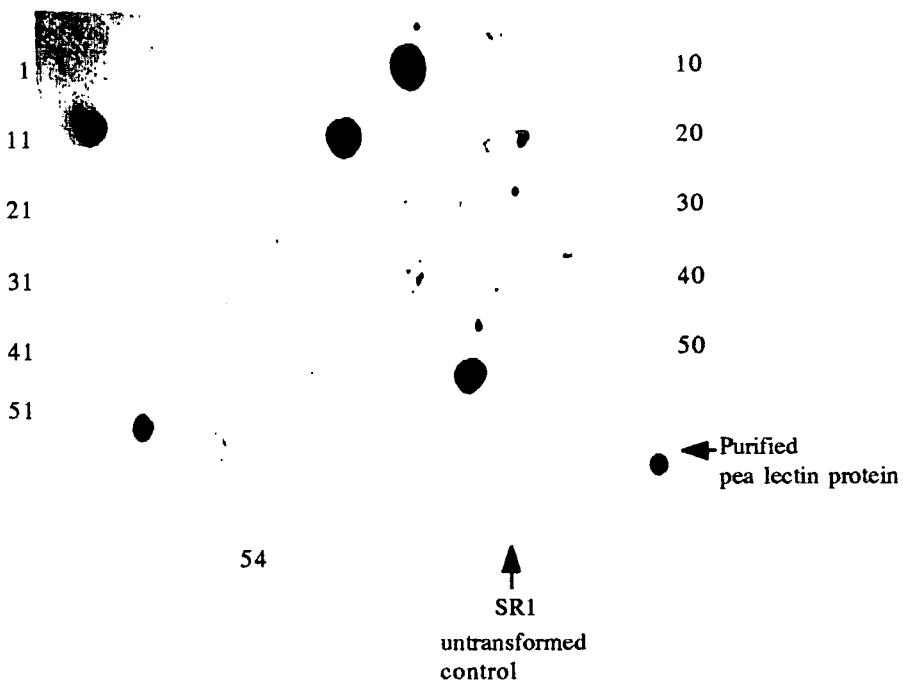
(B) Autoradiograph of the immunoblot in figure 3.19(A) after probing with affinity purified rabbit anti-pea lectin antibodies (1:2000) and after reaction with labelled IgG antibodies (1:5000). 200ng of purified pea lectin and a SR1 untransformed leaf squash blot are indicated by arrows.

(C) Autoradiograph of an immunoblot of further primary regenerant leaf squashes from transformations with *Agrobacterium* GV3101) harbouring the co-integrate pGV3850::133 after reaction with primary rabbit anti-pea lectin antibodies (1:2000), secondary donkey anti-rabbit IgG ^{125}I conjugated antibodies (1:5000) and detection via autoradiography. Numbers represent individual transgenic lines (1-41).

A.


1
10
11
20
21
30
31
40
41
50
51
Purified
pea lectin protein
54
SR1
untransformed
control

B.


1
10
11
20
21
30
31
40
41
50
51
Purified
pea lectin protein
54
SR1
untransformed
control

C.

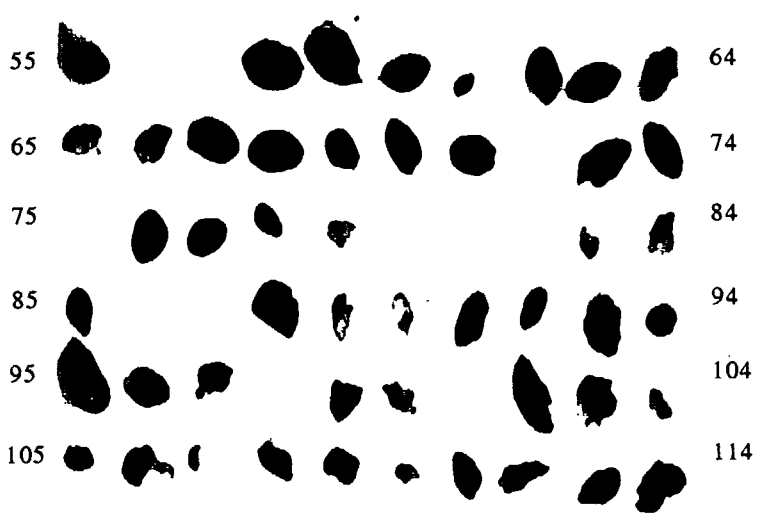

55
64
65
74
75
84
85
94
95
104
105
114

Figure 3.21:

(A) Nitrocellulose filter of a whole plantlet squash, stained with Ponceau red (0.2%(w/v) in 3%(w/v) TCA), for a one example of a kanamycin resistant primary regenerant after transformation with *ssRubisco-lecA* construct

(B) Immunoblot of a whole plantlet squash in figure 3.21(A) after reaction with pea lectin antibodies(1:2,000), and ^{125}I labelled IgG antibodies (1:5000) detected via autoradiography.

A.

B.

Figure 3.22:

T0 individuals regenerated after co-cultivation with *Agrobacterium* (GV3101) harbouring pGV3850::pDUB130 co-integrate. The plants are shown flowering 2 months after potting out. The arrows indicate plants which displayed dwarf characteristics (2) when compared with phenotypically normal plants (1) at the same physiological stage.

(1) (2)

(Larkin and Scowcroft, 1981). This variation is commonly detected where cell proliferation has gone through a disorganized phase. Frequently the variation is due to polyploidy (extra sets of chromosomes 3N, 4N, 6N etc.) in the resultant cells. Possible mechanisms suggested for somaclonal variation are the activation of transposable elements (McClintock, 1984; Scowcroft, 1985) or mutations induced by retrotransposons during tissue culture (Hirochika, 1993). 4% of the plantlets displayed small white leaves with no root formation and these were considered to be kanamycin sensitive. Those plants which rooted and appeared phenotypically normal were transferred to soil. A total of 64 *CaMV-lecA* plants and 37 *ssRubisco-lecA* plants were potted out in sterile compost, grown to full maturity in a containment growth room and allowed to self pollinate on flowering. The primary regenerants were designated T0 and the progeny obtained from selfing T1. Ten weeks after potting out in soil 6% of the plants appeared to regenerate abnormally as characterised by a reduction in height when compared with other T0 individuals at the same physiological stage (figure 3.22). The dwarfed plants obtained were unable to produce seed and suggests that their abnormal phenotype may be as a result of somaclonal variation which commonly produces sterile plants. Such plants were not taken on for further analysis.

3.3.4 Analysis of pea lectin accumulation in T0 individuals.

Pea lectin accumulation was estimated by a quantitative slot blot radioimmunoassay (RAI) for 60 *CaMV-lecA* T0 individuals and 34 *ssRubisco-lecA* T0 individuals (figures 3.23 (A) and 3.24 (A)). Total soluble protein was extracted in PBS from the first upper leaf (3-5cm long) of each individual plant at the six leaf stage and the protein content of the extract determined by the coomassie brilliant blue binding assay of Bradford (1976). This was a rapid, convenient and quantitative method for comparison of closely similar protein extracts. Exactly 4µg of total protein was loaded per slot, aligned with a row of a serial dilution of purified pea lectin protein (90ng-0.3ng). This amount of protein

was selected based on results of earlier experiments (section 3.2.3.2) to ensure that blots were not overloaded with protein, leading to a loss of protein and underestimation of lectin levels. The resulting autoradiographs are shown in figures 3.23 (A) and 3.24 (A). The autoradiograph slots were scanned with an L.K.B. densitometer and a standard curve constructed by integration of the peak areas and plotting these against the logarithm of the pea lectin concentration (figures 3.23 (B) and 3.24 (B)). Peak areas calculated from the densitometric scans of the T0 individuals were converted to ng of pea lectin protein by extrapolating from the standard curve and values expressed as a percentage of total soluble leaf protein. The results are represented in figures 3.23 (C) and 3.24 (C). As expected the levels of pea lectin accumulated showed a wide variation in T0 individuals. The large number of T0 individuals assayed permitted statistical analysis, collated into tables 3.2, 3.3 and 3.4. Table 3.2 gives the mean, median, standard deviation, variance, and variation coefficient of pea lectin levels for both *CaMV-lecA* and *ssRubisco-lecA* populations. Since no detailed Southern analysis was undertaken on the T0 individuals it was assumed that the *lecA* gene was integrated together with the *nos/nptII* gene and thus kanamycin-resistant T0 individuals giving undetectable amounts of pea lectin were not omitted from the statistcal analysis as they may have undergone transgene silencing during regeneration or as a consequence of some environmental stimulus. This possibility must be considered given that all plants potted out were previously shown to be lectin-positive on squash blots (section 3.3.3; figures 3.18, 3.19). All statistical analyses was carried out on pea lectin protein values expressed as a percentage of total soluble leaf protein.

The estimates of pea lectin protein levels were arranged in ascending order for both *CaMV-lecA* and *ssRubisco-lecA* T0 populations (figures 3.25 (A) and (B)) to illustrate the overall range of pea lectin levels observed. No apparent differences between promoter strengths at the stages of leaf development selected were observed as reflected by similar ranges in the levels of pea lectin accumulation for both populations (figure 3.25 (A) and (B)). The frequency distribution curves for *CaMV-lecA* and *ssRubisco-lecA* (figure 3.26) are skewed, indicating that T0 individuals from both populations

Table 3.2:

Main statistical data of pea lectin accumulation in both *CaMV-lecA* and *ssRubisco-lecA* T0 populations.

Table 3.3:

Computation of Bartlett's test of homogeneity of variance for *CaMV-lecA*[a] and *ssRubisco-lecA*[b] T0 populations differing in size.

Table 3.4:

An approximate method of testing the hypothesis whether the sample means of the two different series were significantly different.

Data	*CaMV-lecA* T0	*ssRubisco-lecA* T0
Nº of Plants (n)	60	34
Mean (x)	0.086 +/- 0.0110	0.054 +/- 0.0124
Median	0.067	0.025
Variance (vx)	0.00728	0.00526
Standard deviation (sx)	+/- 0.0853	+/- 0.0724
Coefficient of variation (C)	100.1%	74.7%

Population	$\sum x^2$	D.F.(n-1)	1/(n-1)	vx	log(vx)	(n-1)log(vx)
1[a]	0.873	59	0.0169	0.00728	-2.138	-126.142
2[b]	0.273	33	0.0303	0.00526	-2.279	-75.207
n=2	1.146	92	0.0472			-201.349

Population	Number of plants(n)	D.F.	Mean	$\sum x^2$	vx	$\sum x^2$=vx/n	$t_{.05}$
CaMV-lecA	60	59	0.086	0.873	0.00728	1.21×10^{-4}	2.000
ssRubisco-lecA	34	33	0.054	0.273	0.00526	1.55×10^{-4}	2.030
	94	92	0.032			2.76×10^{-4}	

tend to accumulate low levels of pea lectin protein. In fact 58% of the analysed *CaMV-lecA* T0 and 68% of the *ssRubisco-lecA* T0 individuals contain levels which are lower than the mean. Similar observations have been noted by Peach and velten, (1991), Breyne *et al.*, (1992), Allen *et al.*, (1993), De Neve *et al.*, (1993), and Mlynárová *et al.*, (1994) when analyzing populations of primary regenerants. The extent of variation within the *CaMV-lecA* and *ssRubisco-lecA* populations was observed to differ as deduced fron a higher coefficient of variation obtained with the *CaMV-lecA* population. The Bartletts test was used to determine whether the variances of both populations were homogenous (table 3.3).

Mean variance = $\sum x^2/\sum(n\text{-}1) = 1.146/92 = 0.0125$

(log mean variance) $\sum(n\text{-}1) = (-1.90460)\ (92) = -175.2235$

Chi-squared, $X^2 = 2.3026[(\text{log mean variance}) \sum(n\text{-}1) - \sum(n\text{-}1)\ (\log vx)$

$= 2.3026[-175.2235 - (-201.349)]$

$= 60.16$

It was revealed that $X^2(60.16)$ was greater than the tabular X^2 value for one degree of freedom at the 5%(3.841) and 1%(6.635) significance level rejecting the null hypothesis, suggesting the variances were heterogenous. As the variances were found to be heterogenous a Cochran and Cox method (table 3.4) was used to test whether the population means of two different series were significantly different with no hypothesis about the population variances.

$\sum x = \sqrt{2.76 \times 10^{-4}} = 0.0166$

$x/\sum x = 0.032/0.0166 = 1.93$

$5\%\ \text{level} = \dfrac{(1.21\text{x}10\text{-}4)(2.000) + (1.55\text{x}10\text{-}4)(2.030)}{1.21\text{x}10\text{-}4 + 1.55\text{x}10\text{-}4}$

$t_{0.5} = 2.02$

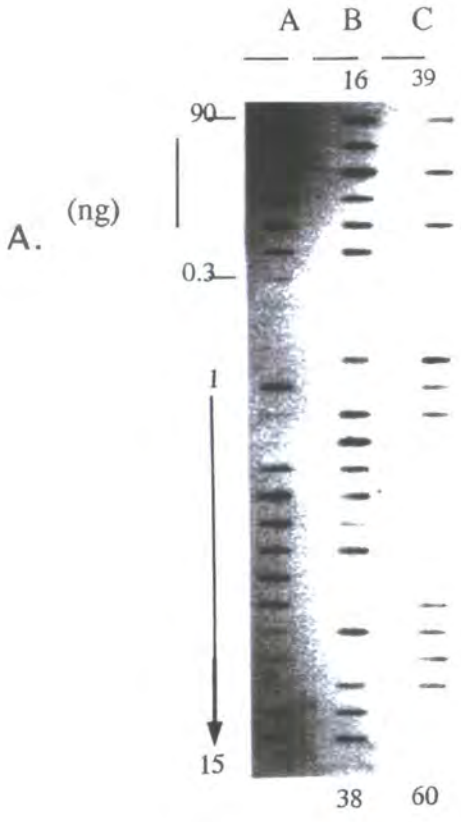
Since the ratio 1.93 was smaller than the 5% level(2.02), the average level of pea lectin accumulation estimated for both populations is not significantly different however the

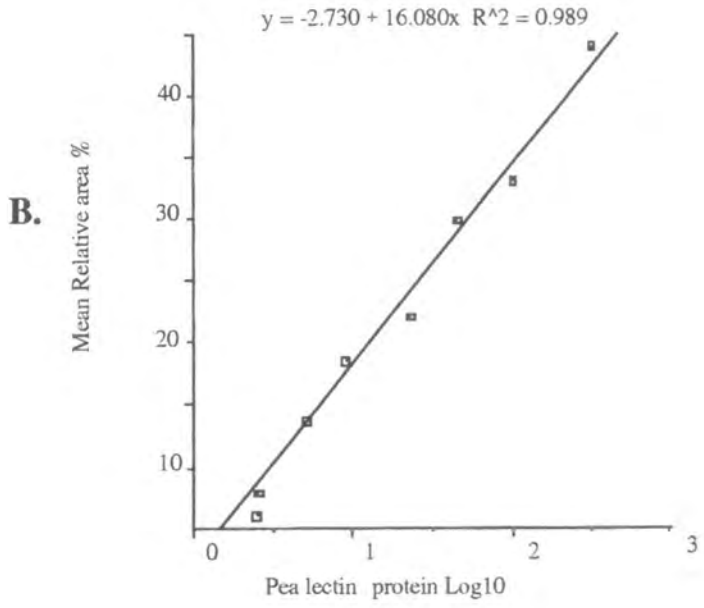
Figure 3.23:

(A) Autoradiograph of a Western slot blot after reacting with rabbit ant-pea lectin antibodies and ^{125}I labelled IgG antibodies. Total soluble protein extracted from the first upper leaf of each of 60 *CaMV-lecA* T0 individuals (6 leaf stage) were blotted along with pea lectin standards (90ng-0.3ng) 4µg/slot and an SR1(untransformed plant) leaf extract. Arrow indicates plants numbered 1-15 (row A), numbers at the top and bottom of row B indicate plant numbers 16-38 and numbers at the top and bottom of row C indicate plant numbers 39-60.

(B) A plot of the mean relative area percentages under the peaks against $\log_{10}$ lectin concentration. Areas under the peaks were determined for all 60 T0 *CaMV-lecA* individuals and the amounts of pea lectin estimated by extrapolating from the standard curve.

(C) Quantitative results from 3.23(A) for pea lectin levels in *CaMV-lecA* T0 individuals. Pea lectin levels in T0 individuals were determined as a percentage of total soluble protein loaded per slot (4µg) for three separate assays and the mean value calculated. Error bars represent the spread of three repeat experiments.


A
A B C
16 39
90
(ng)
0.3
1
15
38 60


B.
y = -2.730 + 16.080x R^2 = 0.989
Mean Relative area %
40
30
20
10
0
1
2
3
Pea lectin protein Log10

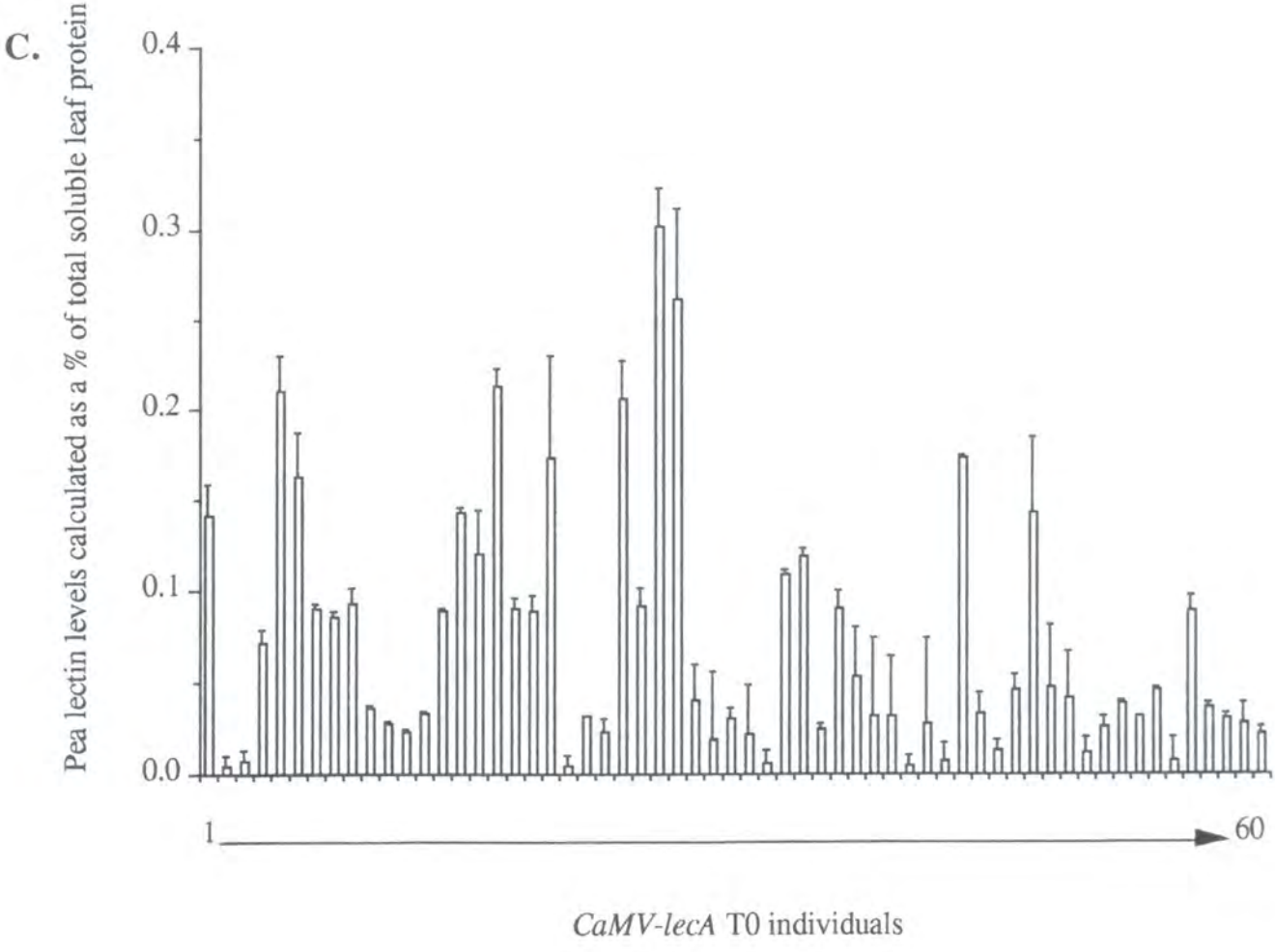
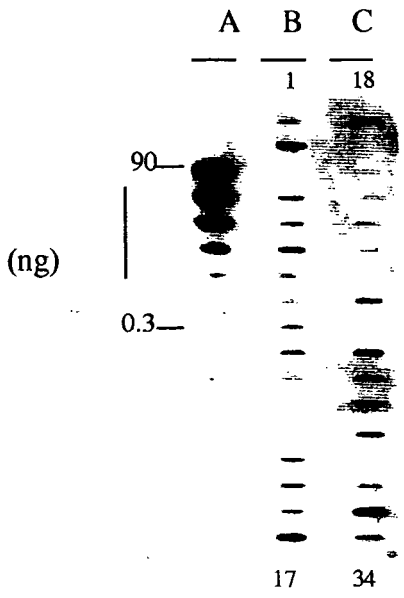

C.
Pea lectin levels calculated as a % of total soluble leaf protein
0.4
0.3
0.2
0.1
0.0
1
60
CaMV-lecA T0 individuals

Figure 3.24:

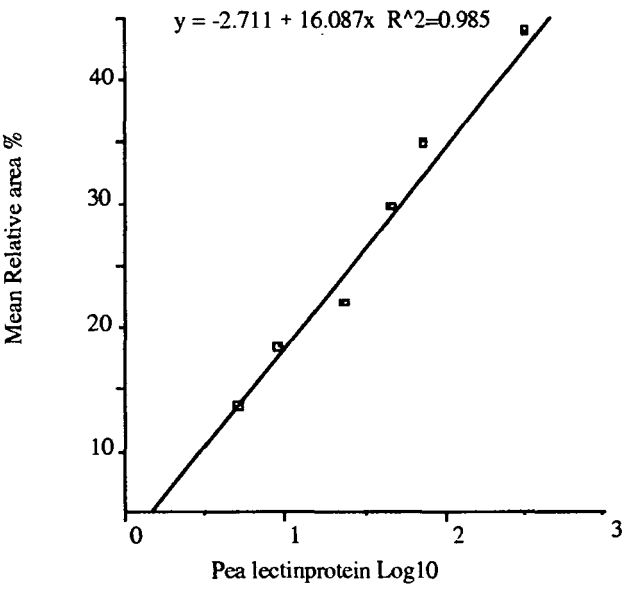
(A) Autoradiograph of a Western slot blot after probing with rabbit ant-pea lectin antibodies and detection via autoradiography with a donkey anti-rabbit IgG ^{125}I conjugate for total soluble protein extracted from the first upper leaf for 34 *ssRubisco-lecA* T0 individuals at the six leaf stage aligned with pea lectin standards(90ng-0.3ng) and SR1 (untransformed plant) leaf soluble protein.

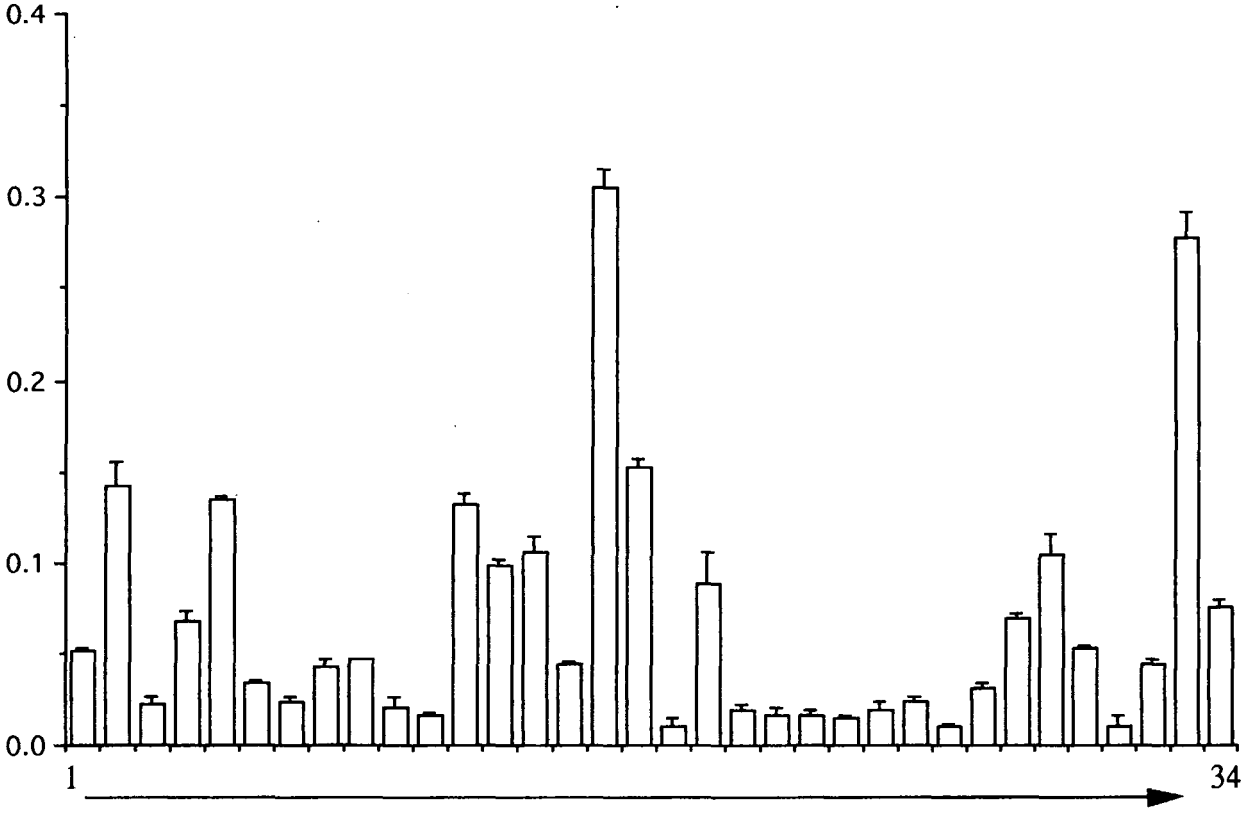
(B) A plot of the mean relative area percentages under the peaks with the $\log_{10}$ of standard two-fold pea lectin dilutions (90ng-0.3ng). Areas under the peaks were determined for all 34 T0 *ssRubisco-lecA* individuals and the amounts of pea, lectin protein estimated by extrapolating from the standard curve.

(C) Graphical representation of 3.24(A) for pea lectin estimated levels in *ssRubisco-lecA* T0 individuals. Pea lectin levels in T0 individuals were determined as a percentage of total soluble protein loaded per slot (4µg) for three separate assays and the mean value calculated. Error bars represent the spread of three repeat experiments.



B.





ssRubisco-lecA T0 levels

Table 3.5:

Kanamycin segregation data for progeny derived from a number of *CaMV-lecA* T0 individuals.

Seed was germinated on kanamycin-containing (100 µg/ml) medium and sensitivity to kanamycin determined as bleaching of cotyledons at 21 d post-germination. Observed segregation ratios were then statistically compared to predicted ratios and the probability that observed and predicted ratios were the same expressed as a p value.

variances of the populations as judged by the Bartletts test evidently are. Therefore the statistical data indicates that the accumulation of pea lectin protein is in the same range in both populations; however alterations in the distribution of pea lectin levels within both populations may have arisen as a result of possible different T-DNA integration events and the effects of surrounding sequences or chromatin structure on transgene expresion (i.e. position effect).

3.3.5 Concluding Remarks.

A large population of T0 individuals for both *CaMV-lecA* and *ssRubisco-lecA* constructs were generated, selected on kanamycin and comparable stages of developing leaf tissue assayed for pea lectin protein levels. Both populations exhibit great variability in the levels of pea lectin protein. These data are in agreement with other studies of variability in gene expression in transgenic *Nicotiana* populations (Peach and Velten, 1991, Breyne *et al.*, 1992, and De Neve*et al.*, 1993). Thus such material is ideally suited to further analysis, allowing a detailed study into the stable inheritence of the chimaeric genes, T-DNA structure, T-DNA copy number and other factors such as methylation and site of genome integration which may contribute to the observed transgene expression variability .

3.4 Characterisation of integrated T-DNAs within T1 *CaMV-lecA* and *ssRubisco-lecA* tobacco plants.

It was important from the outset that a thorough knowledge of the structure of the integrated T-DNA was available for the plants selected for the studies on lectin gene expression. The aim of this analysis was to investigate the frequency with which the T0 plants transmitted their T-DNA genes normally and stably to their T1 progeny. It was also important to determine the frequency with which single copies of the non-rearranged T-DNA were introduced. Once this was established independent transgenic lines were selected with a single T-DNA integration to investigate various factors contributing to transgene expression variability.

3.4.1 Inheritance of the kanamycin-resistant and lectin-positive phenotype within T1 plants.

Inheritance of the kanamycin-resistant phenotype was monitored as before (section 3.1.2) by germinating seedlings on media containing the appropriate antibiotic and scoring green (resistant) and white (sensitive) progeny. Seed collected from 31 selfed *CaMV-lecA* and 16 *ssRubisco-lecA* T0 parental plants were spread onto 1/2MS plates containing kanamycin to determine the segregation ratios of the *nos-nptII* transgene (which confers resistance to kanamycin). The data for the ratio of kanamycin resistant to kanamycin sensitive *CaMV-lecA* and *ssRubisco-lecA* T1 seedlings are summarised in tables 3.5 and 3.6 respectively. The inheritance patterns of the kanamycin resistance gene in the combined T1 seedling population could be placed into four classes as summarised in table 3.7. Out of a total of 47 parental lines for both constructs tested, class I plants (6/47) showed the expected Mendelian inheritance pattern for the *nos-nptII* gene carried at a single locus (Kan^r:Kan^s; 3:1; $p<0.05$). For class II plants (9/47) the ratio approximated to 15:1, $p<0.05$ (Kan^r:Kan^s) indicating at least two independent insertion sites of the kanamycin resistance gene. In class III plants (30/47) the germinated seedlings were entirely kanamycin resistant suggesting the segregation of more than two independent insertion sites for the *nos-nptII* gene. Class IV plants were observed to produce lower than expected ratios of kanamycin resistance in their seeds. In this instance segregation data suggests the inheritance of less than one *nos-nptII* gene and may perhaps be explained by silencing of transgenes inserted at other loci, a phenomenon observed by others (Meyer *et al.*, 1993, Matzke *et al.*, 1989, Ingelbrecht *et al.*, 1994). Ten *CaMV-lecA* and five *ssRubisco-lecA* T0 parental lines, having been selfed and the progeny analysed genetically, were selected with functional *nos-nptII* genes carried at one or more loci.

Seed collected from the self-fertilization of these *lecA* T0 lines were squashed onto nitrocellulose, releasing the stored proteins for probing with anti lectin antibody and ^{125}I conjugated antibody. Figure 3.27 illustrates examples of results obtained. The

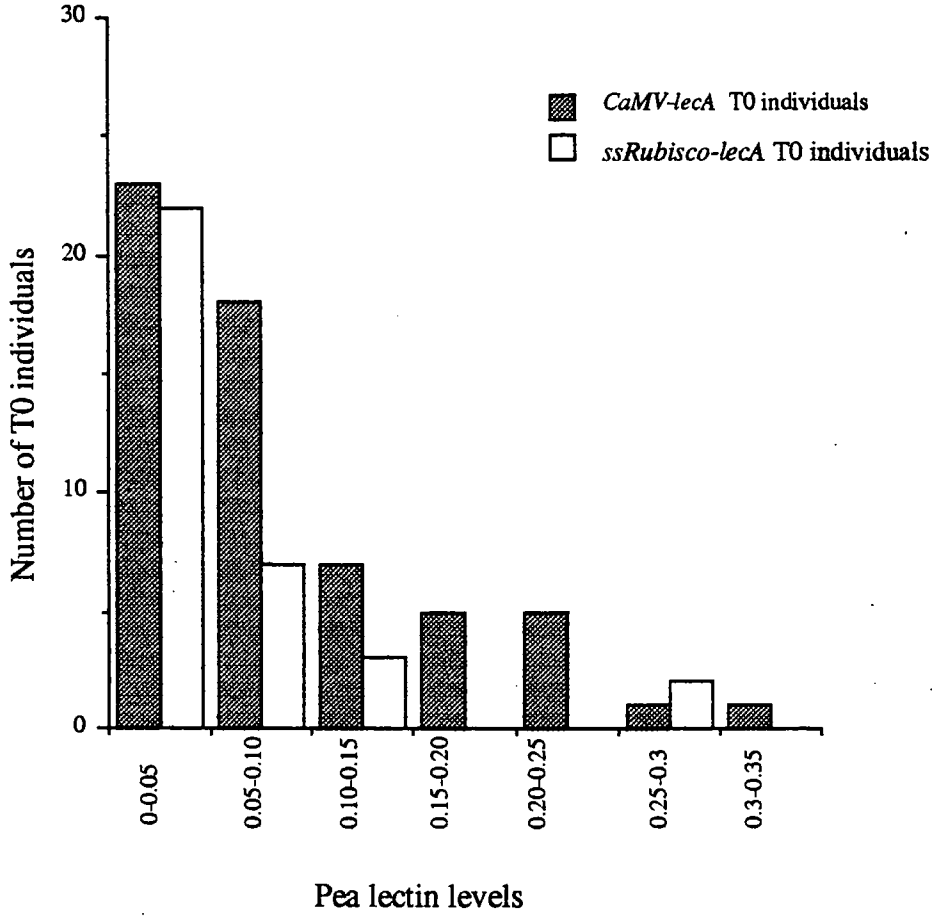
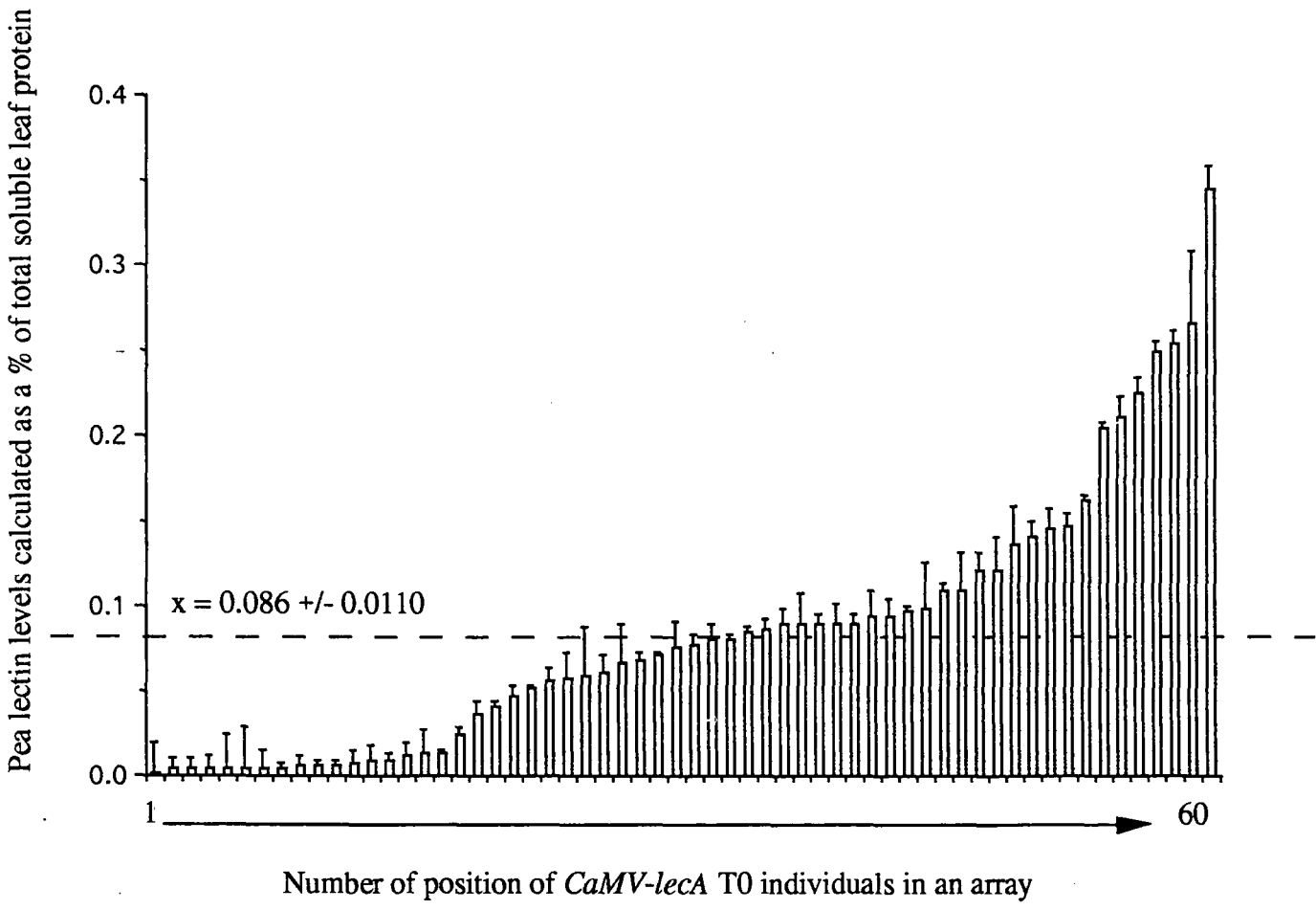

30
20
10
0
Number of T0 individuals
CaMV-lecA T0 individuals
ssRubisco-lecA T0 individuals
0-0.05
0.05-0.10
0.10-0.15
0.15-0.20
0.20-0.25
0.25-0.3
0.3-0.35
Pea lectin levels

Figure 3.26:

Frequency distribution of pea lectin accumulation within the *CaMV-lecA* and *ssRubisco-lecA* T0 populations. Numbers of T0 individuals are plotted against the corresponding level of pea lectin protein (given as a %(w/w) of total soluble leaf protein), these have been grouped into classes of 0.05%(w/w) of total soluble protein.

A.


Pea lectin levels calculated as a % of total soluble leaf protein
0.4
0.3
0.2
0.1
0.0
x = 0.086 +/- 0.0110
1
60
Number of position of *CaMV-lecA* T0 individuals in an array

B.

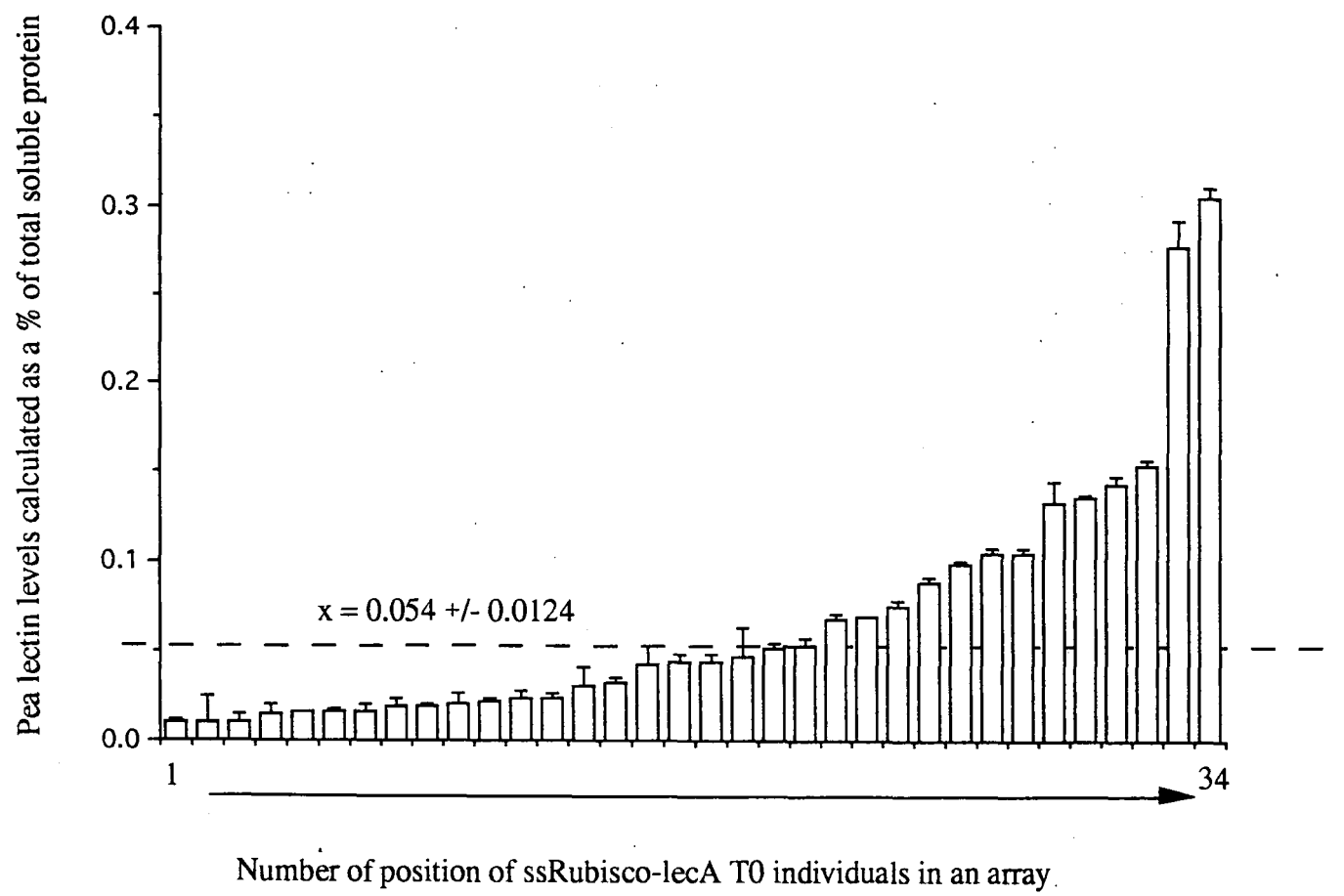

Pea lectin levels calculated as a % of total soluble protein
0.4
0.3
0.2
0.1
0.0
x = 0.054 +/- 0.0124
1
34
Number of position of ssRubisco-lecA T0 individuals in an array

Figure 3.25:

(A) Results from 3.23(C) rearranged according to ascending order for *CaMV-lecA* T0 individuals. The mean levels of pea lectin accumulation are indicated by the dashed line. Error bars represent the spread of three repeat experiments.

(B) Results from 3.24(C) rearranged according to ascending order for *ssRubisco-lecA* T0 individuals. The mean levels of pea lectin accumulation are indicated by the dashed line. Error bars represent the spread of three repeat experiments.

Parent plant *CaMV-lecA*	Number of seeds germinated	Resistant to kanamycin (green)	Sensitive to Kanamycin (white)	Fit to ratio $Kan^R:Kan^S$	χ^2	p
3	100	89	11	15:1	2.97	<0.05
4	'	68	32	3:1	0.08	<0.05
5	'	100	0	1:0	-	-
7	'	61	39	3:1	0.54	<0.05
10	'	91	9	15:1	0.85	<0.05
11	'	100	0	1:0	-	-
12	'	90	10	15:1	1.63	<0.05
13	'	100	0	1:0	-	-
14	'	100	0	1:0	-	-
18	'	59	41	3:1	2.67	<0.05
19	'	100	0	1:0	-	-
20	'	71	29	3:1	0.83	<0.05
22	'	100	0	1:0	-	-
23	'	54	46	3:1	7.26	>0.05 >0.01
26	'	66	34	3:1	0.02	<0.05
28	'	90	10	15:1	1.63	<0.05
31	'	100	0	1:0	-	-
33	'	89	11	15:1	2.97	<0.05
35	'	53	47	3:1	8.45	>0.05 >0.01
36	'	100	0	1:0	-	-
37	'	90	10	15:1	1.63	<0.05
38	'	100	0	1:0	-	-
40	'	100	0	1:0	-	-
42	'	100	0	1:0	-	-
45	'	99	1	1:0	-	-
50	'	100	0	1:0	-	-
51	'	91	9	15:1	0.85	<0.05
55	'	100	0	1:0	-	-
56	'	100	0	1:0	-	-
57	'	98	2	1:0	-	-
60	'	100	0	1:0	-	-

Table 3.6:

Kanamycin segregation data for progeny derived from a number of *ssRubisco-lecA* T0 individuals.

Seed was germinated on kanamycin-containing (100 μg/ml) medium and sensitivity to kanamycin determined as bleaching of cotyledons at 21 d post-germination. Observed segregation ratios were then statistically compared to predicted ratios and the probability that observed and predicted ratios were the same expressed as a p value.

Parent plant *ssRubisco-lecA*	Number of seeds germinated	Resistant to Kanamycin	Sensitive to Kanamycin	Fit to ratio Kan^r:Kan^S	χ^2	p
2	100	91	9	15:1	0.85	<0.05
8	"	92	8	15:1	0.27	<0.05
9	"	100	0	1:0	-	-
13	"	100	0	1:0	-	-
15	"	100	0	1:0	-	-
16	"	100	0	1:0	-	-
17	"	100	0	1:0	-	-
19	"	100	0	1:0	-	-
20	"	100	0	1:0	-	-
23	"	100	0	1:0	-	-
24	"	100	0	1:0	-	-
25	"	100	0	1:0	-	-
28	"	100	0	1:0	-	-
30	"	67	33	3:1	0.004	<0.05
31	"	100	0	1:0	-	-
34	"	100	0	1:0	-	-

Table 3.7:

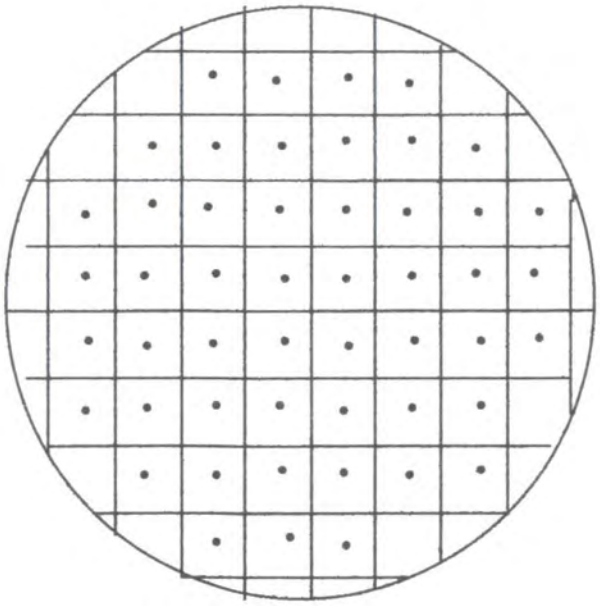
Classification of segregating T1 *ssRubisco-lecA* and *CaMV-lecA* T1 plants. Segregating ratios were determined as the ratio of kanamycin-resistant to -sensitive seedlings when germinated on medium containing 100 µg/ml kanamycin.

Inheritance class	Ratio (green/white) Self	Number of plants
I Single dominant gene	3:1	6
II Two dominant genes	15:1	9
III All resistant	1:0	30
IV Low ratios	<3:1	2
	Total number of plants	47

Figure 3.27:

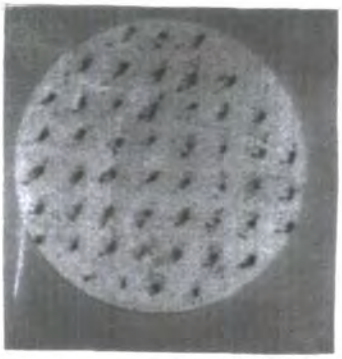
Segregation ratio of lectin gene detected by squash blots of seed collected from the self-crossing of *CaMV-lecA* lines (33, 26, 20, and 4), *ssRubisco-lecA* line (16) and SR1 as an untransformed control. Seed squash blots were probed with a pea lectin antibody and ^{125}I-labelled IgG.

A: Schematic representation of grid used for 50 seed squashes on to nitrocellulose; **B:** Seed squashes of progeny seed collected from T0 *ssRubisco-lecA* parental line 16 displaying all (1:0) to be lectin positive; **C:** Seed squashes of progeny seed collected from the selfing of T0 *CaMV-lecA* parental line 33 displaying a phenotypic ratio of 15:1 lectin positive; **D:** Squashes of SR1 seed (untransformed control); **E, F** and **G:** Seed squashes of progeny seed collected from the selfing of T0 *CaMV-lecA* parental lines 26, 20, and 4, respectively. These lines display a phenotypic ratio of 3:1 lectin positive.



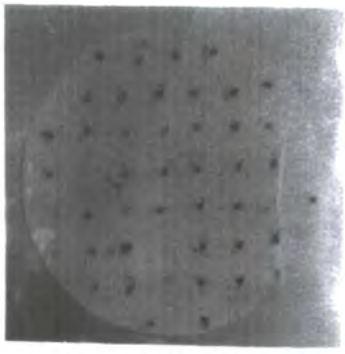
Schematic representation of tobacco seed squashes

A.



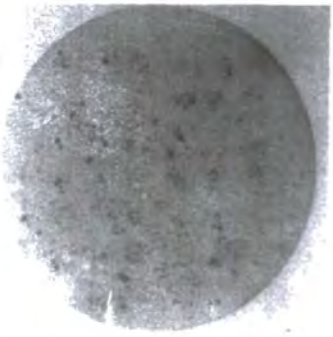
ssRubisco-lecA line 16 T1 seed

B.



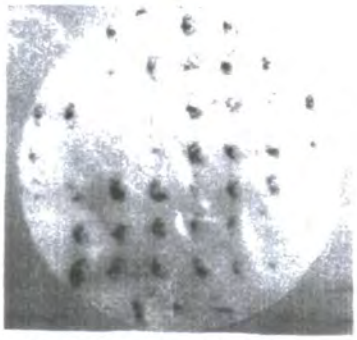
CaMV-lecA line 33 T1 seed

C.



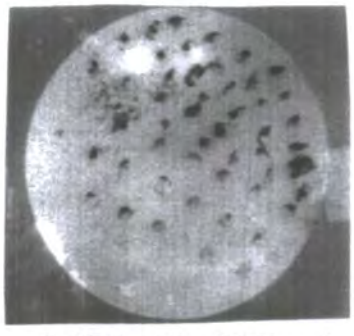
SR1 (untransformed control)

D.



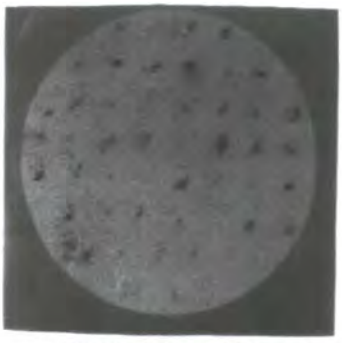
CaMV-lecA line 26 T1 seed

E.



CaMV-lecA line 20 T1 seed

F.



CaMV-lecA line 4 T1 seed

G.

majority of lines displayed the inheritance of the *lecA* gene with ratios identical to that observed for kanamycin resistance. However, two populations of *CaMV-lecA* lines (28 and 12) showed a 15:1 kan^r:kan^s segregation ratio and a 3:1 *lecA*$^+$:*lecA*$^-$ ratio, indicating that a single locus locus carries a *lecA* gene, whereas more than one locus carries a functional *nos-nptII* gene.

T1 plants arising from the selfed ten T0 *CaMV-lecA* and five *ssRubisco-lecA* T0 parental lines were further investigated by Southern analyses (section 3.3.3) to determine the arrangement and number of T-DNA insertion events within the genome.

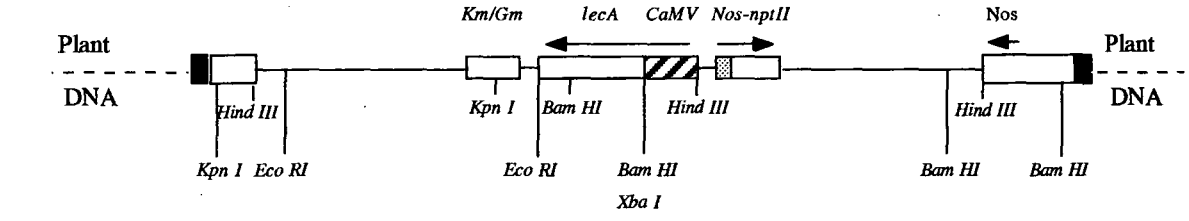
3.4.2 Southern analyses of selected *CaMV-lecA* and *ssRubisco-lecA* T1 individuals.

Southern hybridisations were conducted on selected T1 plants (section 3.3.2) in order to analyse in detail the structure and number of integrated T-DNAs. The probes and restriction enzymes used in the analyses of the structure of the integrated T-DNAs for each construct are shown in figures 3.28 and 3.29 along with the predicted sizes of bands of homology in non-rearranged, integrated T-DNA. The probes used were: (i) 4.37Kb linearised pBR322, (ii) 600bp *Sph I/Pst I* fragment from pNOS-NPT(see table 2.1, section 2.3) homologous to the neomycin phosphotransferase II gene of Tn5, (iii) 3.0Kb *EcoR I / HindIII* fragment from pDUB1105 homologous to T-DNA near the left border, (iv) 1.2Kb *Sph I/ EcoRI lecA* coding region from pDUB80 homologous to the pea lectin gene and (v) 2.1Kb *HindIII/SacII* fragment from pDUB1105 homologous to the T-DNA near the right hand border. The complete analysis consisted of running 0.7% (w/v) agarose gels containing 10µg of digested DNA per lane from plants transformed using pGV3850:130 or 133. The probes used in the analysis were labeled with ^{32}P to high specific activity $>10^6$ c.p.m./µg by the procedure of random priming (Feinberg and Vogelstein, 1984). In all cases hybridisations were carried out with positive and negative controls. The negative control was DNA from untransformed tobacco and the positive controls were one, three and five-copy *lecA* reconstructions

Figure 3.28:

Details of the strategy used for the Southern analysis of T-DNA in transgenic tobacco plants transformed with the *CaMV-lecA* construct.

A: represents the predicted sizes(kb) of DNA fragments with homology to a 4.3Kb pBR322 probe after a *HindIII* genomic digest. **B**: represents the predicted size of T-DNA with homology to a 600bp *Sph I/Pst I* fragment from pNOS-NPT (containing the *nos-nptII* gene) after a *HindIII* plant genomic DNA digest (see section 3.4.2 in text for more detail). **C**: represents the predicted junction fragment with homology to a 3Kb *EcoRI/HindIII* fragment (containing the left hand border sequence) after a *HindIII* genomic digest. **D**: represents the predicted sizes (kb) of DNA fragments with homology to a 1.2Kb *EcoRI/Sph I* fragment (containing the *lecA* gene) after an *EcoRI/HindIII* genomic digest. **E**: represents the predicted sizes (Kb) of DNA fragments with homology to a 2.1Kb *HindIII/SacI* fragment(containing part of the right hand border sequence with homology to the *nos* promoter) after a *BamHI* digest. All probes used were ^{32}P-labelled



A.

Hind III — 8.3 Kb — 6.8 Kb

B.

Hind III — 6.8 Kb

C.

Hind III — Junction Fragment

D.

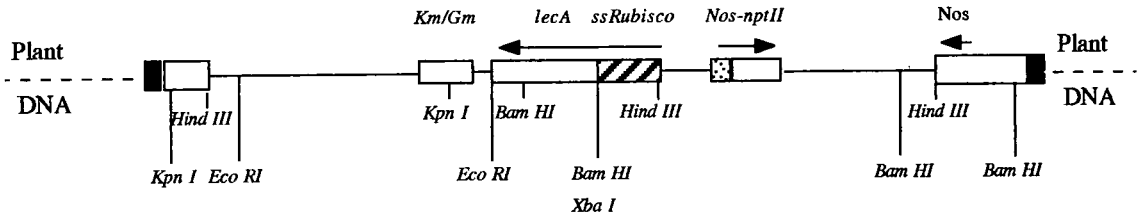
Hind III/Eco RI — 2.1 Kb

E.

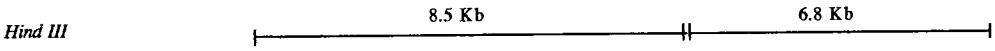
Bam HI — 6.5 Kb — 1.4 Kb — Junction Fragment

Figure 3.29:

Details of the strategy used for the Southern analysis of T-DNA in transgenic tobacco plants transformed with the *ssRubisco-lecA* construct. **A**: represents the predicted sizes(kb) of DNA fragments with homology to a 4.3Kb pBR322 probe after a *HindIII* genomic digest. **B**: represents the predicted sizes of T-DNA with homology to a 600bp *Sph I/Pst I* fragment (containing the *nos-nptII* gene) after a *HindIII* genomic digest.. **C**: represents the predicted junction fragment with homology to a 3Kb *EcoRI/HindIII* fragment (containing the left hand border sequence) after a *HindIII* genomic digest.


Km/Gm
lecA
ssRubisco
Nos-nptII
Nos
Plant
DNA
Plant
DNA
Hind III
Kpn I
Eco RI
Kpn I
Bam HI
Hind III
Eco RI
Bam HI
Xba I
Hind III
Bam HI
Bam HI

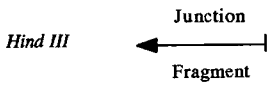
A.


Hind III
8.5 Kb
6.8 Kb

B.


6.8 Kb
Hind III

C.


Junction
Hind III
Fragment

with purified pDUB80 plasmid DNA and DNA from the *Agrobacterium* strains with co-integrates used in the tobacco transformations.

3.4.2.1 Blot hybridisation analysis with a pBR322 probe.

10µg of genomic DNA from 10 *CaMV-lecA* individual T1 plants (33, 28, 26, 20,18, 12, 10, 7, 4, and 3) and 5 *ssRubisco-lecA* T1 individuals (30, 28, 16, 8, and 2) was digested to completion with the restriction endonuclease *HindIII* and the fragments electrophoretically separated on 0.7% (w/v) agarose gels (figure 3.30 (A) and 3.36 (A)). The DNA was blotted onto nylon and hybridised to ^{32}P-labeled linearised plasmid pBR322 in order to detect whether the 15.1Kb (*CaMV-lecA* plants) or 15.3Kb (*ssRubisco-lecA* plants) central region of the T-DNA had been transferred. Out of a total of 10 *CaMV-lecA* T1 individuals analysed 7 *CaMV-lecA* plants 33, 26, 20, 18, 10, 7, and 4 (lanes 2, 4, 5, 6, 8, 9 and 10, figure 3.30 (B)) were found to contain both the predicted 8.3Kb and 6.8Kb internal *HindIII* fragments. Three *CaMV-lecA* T1 plants 28, 12, and 3 (lanes 3, 7, and 11 figure 3.30 (B)) were observed to have homology only to the 6.8Kb region of the T-DNA, the predicted size of the left hand pBR322 fragment (see figure 3.28) known to contain the neomycin phosphotransferase II gene (*nptII*) driven by the nopaline synthase (*nos*) promoter conferring kanamycin resistance. Hybridisations of *ssRubisco-lecA* T1 plants 30, 16, 8, and 2 (lanes 1, 3, 4, and 5, figure 3.36 (B)) revealed two T-DNA internal fragments of 8.7Kb and 6.8Kb with homology to the pBR322 probe. Rearrangements were evident in *ssRubisco-lecA* T1 plant 30 (lane 1), which generated bands of the predicted sizes of 6.8Kb and 8.7Kb and one additional band of 5.4Kb. Extreme forms of rearrangement were observed in the analysis pattern from *ssRubisco-lecA* T1 plant 28 (lane 2). This line contained one band of homology to the predicted 6.8Kb pBR322 region of the T-DNA, but also had additional hybridising bands which did not correspond to any predicted fragment sizes.

The majority of rearangements observed in the T1 plants (4/15) was the absence of fragments with homology to the left hand pBR322 sequence containing the chimaeric *lecA* genes. The rearrangements suggest the T-DNAs in these plants not to be full

length and may have resulted from incomplete transfer to the plant genome or, possibly, rearrangements or deletions within the plant genome. A more detailed discusion will be presented in section 3.4.3 where the relationship between T-DNA structure, number of loci, copy number and T-DNA organisation will be discussed.

In all T1 plants analysed with the pBR322 probe the 6.8Kb fragment containing the kanamycin resistance gene was observed. To confirm the presence of this gene directly the same blots were stripped and reprobed with the *nos-nptII* gene.

3.4.2.2 Blot hybridisation analysis with the *nos-nptII* probe.

Figures 3.30 (C) and 3.36 (C) show *HindIII* digests of the same DNA samples as above hybridised with a *nos-nptII* probe for *CaMV-lecA* and *ssRubisco-lecA* T1 plants. The expected 6.8Kb *HindIII* fragment containing the *nos-nptII* gene was observed in 14 out of the 15 *CaMV-lecA* and *ssRubisco-lecA* plants tested. The presence of such a band was obscured in lane 2 (figure 3.30 (C)) by high background labeling but was expected to be present since it had previously been detected with the pBR322 probe in figure 3.30 (B) (lane 2) and only kanamycin resistant T1 seedlings had been selected. The intensities of bands were observed to vary between the T1 plants indicating possible differences in copy number. No attempt was made to determine the copy number of the *nos-nptII* gene due to time constraints; however gene copy determinations were made for the lectin gene allowing conclusions to be drawn as to the number of *lecA* copies at a given locus.

3.4.2.3 Blot hybridisation analysis with the *lecA* probe.

To determine the number of integrated *CaMV-lecA* gene copies within the tobacco genome, 10µg of DNA from each plant was analysed on a 0.7% (w/v) agarose gel (figure 3.31 (B)) after complete digestion with *EcoRI* and *HindIII*. Linearised pDUB80 was spiked into *EcoRI* and *HindIII* restricted SR1 genomic DNA as gene copy reconstructions of one, three, and five gene copies. Gene copy reconstructions were calculated using a haploid *N. tabacum* genomic content of $1C$ = 3.3pg DNA

(Bennet and Smith 1991). After Southern hybridisation with a ^{32}P-labelled 1.2Kb *Sph I/EcoR I* fragment containing the *lecA* coding region, the expected 2.1 Kb band corresponding to the *CaMV-lecA* chimaeric gene was detected with varying intensities in all nine plants used in the analysis. The presence, in all plants tested, of a single hybridising fragment of a size corresponding to that of the non-rearranged *lecA* coding sequence indicates that, in all cases, the *lecA* gene retained its integrity within T1 plants.

Based on hybridisation intensities above, *CaMV-lecA* T1 individuals 28, 26, 20, and 12 (lanes 3, 4, 5, and 7) were estimated to contain 1 copy of the transgene per tetraploid genome. While *CaMV-lecA* T1 individuals 18, and 4 (lanes 6, and 9) were estimated to contain at least two gene copies. *CaMV-lecA* T1 individuals 33, 7, and 3 (lanes 2, 8, and 10) contained an estimated 3-5 gene copies and were designated multicopy. The number and organisation of T-DNA integration events within the genome were investigated in section 3.4.3. Conclusions to the number of *lecA* gene copies per locus are also discussed in section 3.4.3.

3.4.2.4 Blot hybridisation analysis with the left hand border probe.

The left boundary probe is defined by a *HindIII* site at its right end which lies about 3.0Kb to the right of the left repeat and an *EcoRI* site at its right end which lies within the T-DNA left border a short distance from the left direct repeat (figures 3.28 and 3.29). The left border probe and associated *HindIII* digest provide criteria for recognising the presence of inverted repeats (IR) at the ends of T-DNA. In principle, an IR about the left T-DNA end will give border fragments of a size twice (6.0Kb) the distance from the border to the next *HindIII* site. Also bands corresponding to such fragments should hybridise with twice, or an integral multiple of twice, the signal of a band from a non-inverted repeat border fragment.

Figures 3.32 (B) and 3.36 (D) present the *HindIII* Southern blot of 8 *CaMV-lecA* T1 plants (33, 28, 26, 20, 18, 12, 10, and 4) and 5 *ssRubisco-lecA* (30, 28, 16, 8 and 2) plants probed with the left hand border fragment. In figure 3.32 (A) *CaMV-lecA* T1

plants 26, 20, 12, and 4 (lanes 3, 4, 6, and 8) give rise to simple patterns consistent with only one left border fragment. On the other hand *CaMV-lecA* T1 plants 33, 18, and 10 (lanes 1, 5, and 7) show patterns with two bands and *CaMV-lecA* T1 plant 28 (lane 2) shows three bands with homology to the probe. Fragments approximating 6.0Kb were observed with T1 plants 33, and 28 (lanes 1 and 2) consistent with the existence of inverted repeats using the criteria defined above. The greater intensity of bands depicted by arrows for inverted repeats (T1 plants 28 and 18) when compared with single left hand borders (e.g. T1 plant 26) further substantiates this type of arrangement about the left border.

In figure 3.36 (D) *ssRubisco-lecA* T1 plants displayed fragment patterns of zero (plant 30; lane 1), two (plants 8 and 2; lanes 4 and 5), three (plant 16; lane 3), and four bands (plant 28; lane 2) with homology to the left hand border probe. Bands of homology with estimated sizes of 6.0Kb were found in T1 plants 28 and 16 (lanes 2 and 3) and were the predicted size for that fragment between the *HindIII* sites of the left hand borders assuming the T-DNA had integrated as an inverted repeat. The intensity of the bands for plant 16 (lane 3) suggest the arrangement of multicopy repeat T-DNA structures about the left border. Bands of homology larger than 6.0Kb common in four *ssRubisco-lecA* T1 individuals in lanes 2, 3, 4, and 5 (figure 3.36 (D)) did not correspond to sizes predicted for inverted or possible direct repeat like structures and were interpreted as complex arranged T-DNA arrays about the left border. The absence of bands homologous to the left border probe for plant 30 (lane 1) suggests that it was completely rearranged or possibly not transferred to the plant genome.

3.4.2.5 Blot hybridisation analysis with the right hand border probe.

Southern analysis was performed on T1 *CaMV-lecA* plants for lines 26, 20 and 4 to determine the types of arrangement centred around the right hand border. These three lines were selected for this analysis given that they had failed to show any apparent T-DNA rearrangements in previous Southern analyses. DNA was extracted, cut with

BamHI and hybridised with a probe corresponding to the right hand border. The probe was a 2.1Kb *HindIII/SacII* fragment from pDUB1105 homologous to the T-DNA near the right border repeat. The probe spans the right hand border region from the *HindIII* site to a *SacII* site which lies 176bp left of the right border repeat. A *BamHI* digest generates two internal fragments of 1.4Kb and 6.5 Kb with homology to the probe. The 6.5Kb band is observed since the probe has 300bp homology with the *nos* promoter which drives the *nptII* gene. The *BamHI* digest also produces fragments of variable size spanning the T-DNA right border and plant DNA junctions with homology to the probe. The size of these fragments can also be used to predict the presence of inverted repeats around the right hand border. In theory an inverted repeat around the right hand border would produce fragments of 2.3Kb.

Genomic DNA extracted from T1 plants for each line was digested with *BamHI* (figures 3.33 (A), 3.34 (A), and 3.35 (A)) and hybridised to a ^{32}P-labelled 2.1 Kb *HindIII/SacII* (pDUB1105) fragment containing part of the right hand border. A control of total *Agrobacterium* (pGV3850::130) DNA digested together with SR1 genomic DNA was used to evaluate the extent of digestion and to determine the precise size of non-rearranged internal T-DNA fragments with homology to the probe. In figures 3.33 (B), 3.34 (B), and 3.35 (B) two internal fragments of 6.5Kb and 1.4 Kb were produced for the majority of T1 plants analysed. A number of different sized fragments spanning the T-DNA and plant DNA junctions can also be seen for most of T1 plants analysed. In figure 3.33 (B) one junction fragment greater than 23.1Kb was observed in all T1 plants for line 26. In the case of line 20 all T1 plants produced the predicted internal fragments and either three (lane 2) or two other hybridising fragments (lanes 3-8) (figure 3.34 (B)). The size of one of the fragments in lanes (2-8) corresponds to the size of a predicted *BamHI* fragment if the T-DNA is arranged as a inverted repeat. Also, a 5.8Kb hybridising fraggment is found in all plants of this line with an intensity similar to that of the inverted repeat fragment, suggesting the possibility that some form of rearrangement may have occurred. In lane 1 (plant 1) of figure 3.34 (B) (line 20) there is a homologous band greater than 6.5 Kb suggesting the

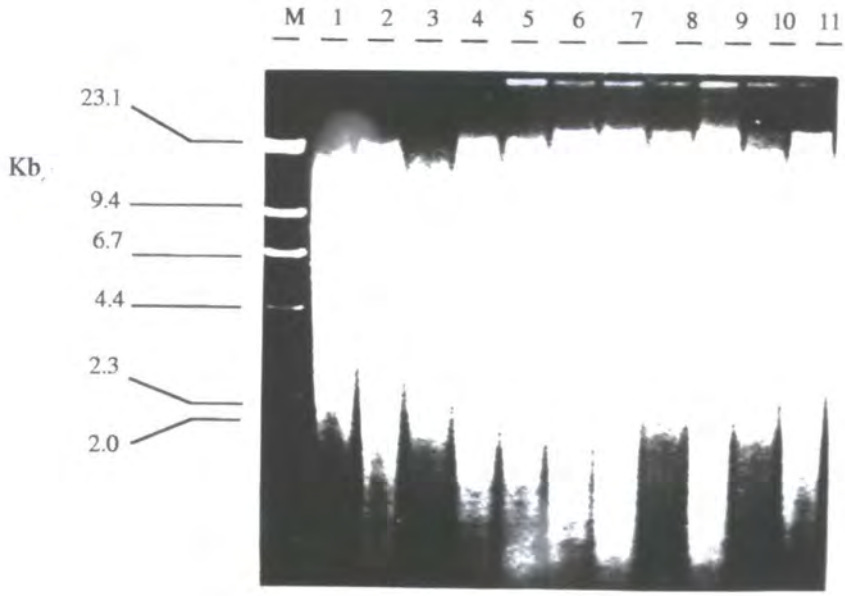
Figure 3.30:

(A) Restriction analysis of ten selected *CaMV-lecA* T1 individuals separated on 0.7%(w/v) agarose gel. Lane 1: Control of *HindIII* digested SR1 genomic DNA. Lanes 2-11: represent *HindIII* digested genomic DNA for *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, 7, 4, & 3). Lane M: Size markers Lambda *HindIII* digest.

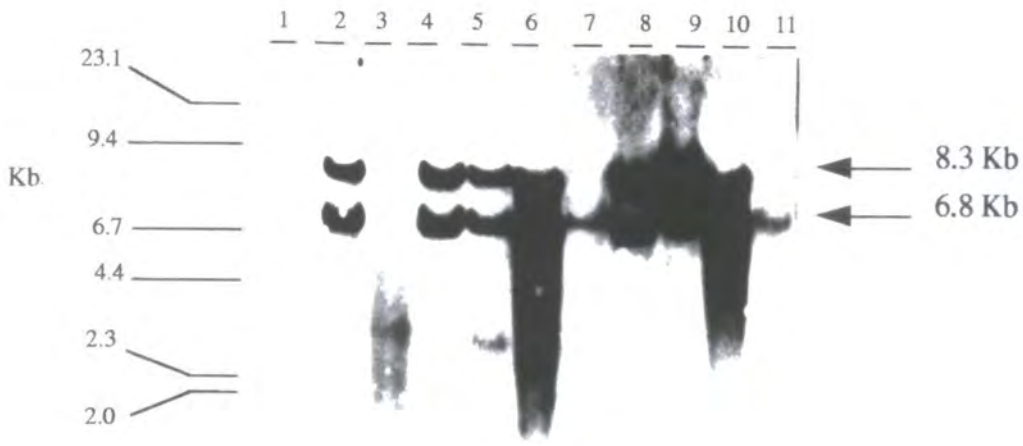
(B) Southern blot analysis of the *HindIII* digested genomic DNA from the 10 *CaMV-lecA* T1 individuals as shown in figure 3.30(A) and probed with a ^{32}P-labelled linearised pBR322 plasmid DNA 4.3Kb. Lane 1: control, SR1 genomic DNA. Lanes 2-11: *CaMV-lecA* T1 individuals (33, 28, 26, 20,18,12, 10, 7, 4, & 3). Arrows indicate the predicted sizes of *HindIII* fragments (8.3Kb & 6.8Kb) showing the presence of the 15.1Kb central T-DNA region.

(C) Southern blot analysis of *HindIII* digested genomic DNA for 10 *CaMV-lecA* T1 individuals (figure 3.30(A)) probed with a ^{32}P-labelled 600bp *Sph I/Pst I* fragment, containing the *nos-nptII* gene (see figure 3.28 B and section 3.4.2). Lane 1: control, SR1 genomic DNA. Lanes 2-11: *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, 7, 4, & 3). Arrow indicates the predicted size(6.8Kb) of a T-DNA fragment with homology to the probe.

A.


M 1 2 3 4 5 6 7 8 9 10 11
Kb
23.1
9.4
6.7
4.4
2.3
2.0

B.


1 2 3 4 5 6 7 8 9 10 11
Kb
23.1
9.4
6.7
4.4
2.3
2.0
8.3 Kb
6.8 Kb

C.

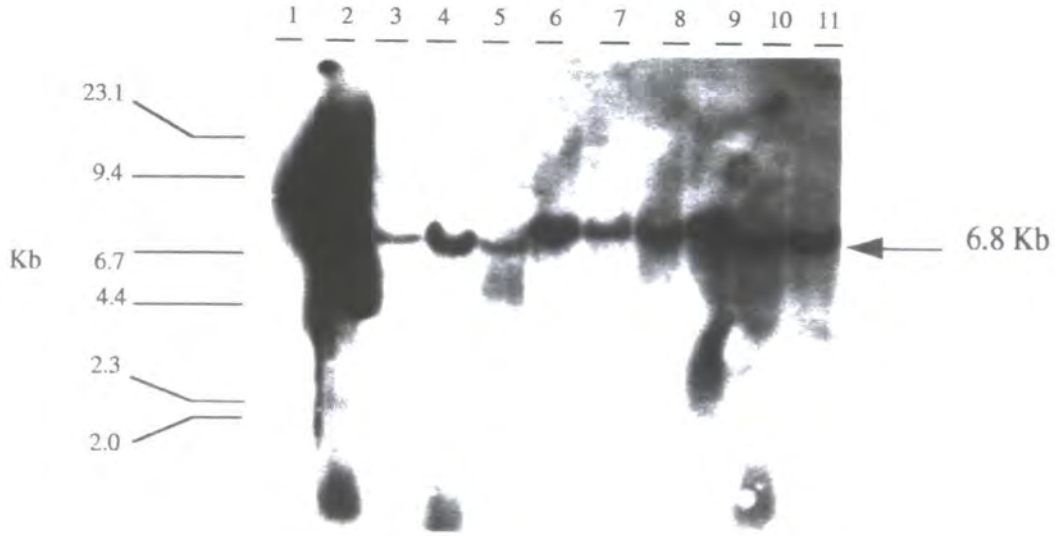
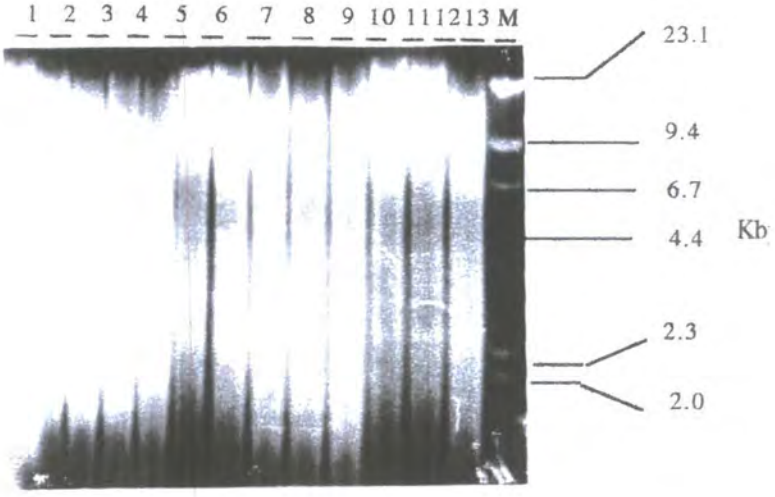

1 2 3 4 5 6 7 8 9 10 11
Kb
23.1
9.4
6.7
4.4
2.3
2.0
6.8 Kb

Figure 3.31:

(A) Agarose gel analysis on nine selected *CaMV-lecA* T1 individuals. lane 1: *EcoRI/ HindIII* digested SR1 genomic DNA. Lanes 2-10: *EcoRI/HindIII* digested genomic DNA for *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, 4, & 3). Lanes 11-13: *EcoRI/HindIII* digested SR1 genomic DNA spiked with *EcoRI* linearised pDUB80 representing *CaMV-lecA* gene copy reconstructions one three and five respectively. Lane M: Lambda *HindIII* digest.

(B) Southern blot analysis of *EcoRI/HindIII* digested genomic DNA on nine selected *CaMV-lecA* T1 individuals, probed with a ^{32}P-labelled 1.2Kb *Sph I/EcoRI* fragment of pDUB80 containing the *lecA* coding region. Lane 1: *EcoRI/HindIII* digested SR1 genomic DNA (untransformed control). Lanes 2-10: *EcoRI/HindIII* digested *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, 4, & 3). Lanes 11-13: *EcoRI* linearised pDUB80 (3.8Kb) representing one, three, and five *CaMV-lecA* gene copies per haploid *N.tabacum* genome.

A.


1 2 3 4 5 6 7 8 9 10 11 12 13 M
23.1
9.4
6.7
4.4
Kb
2.3
2.0

B.

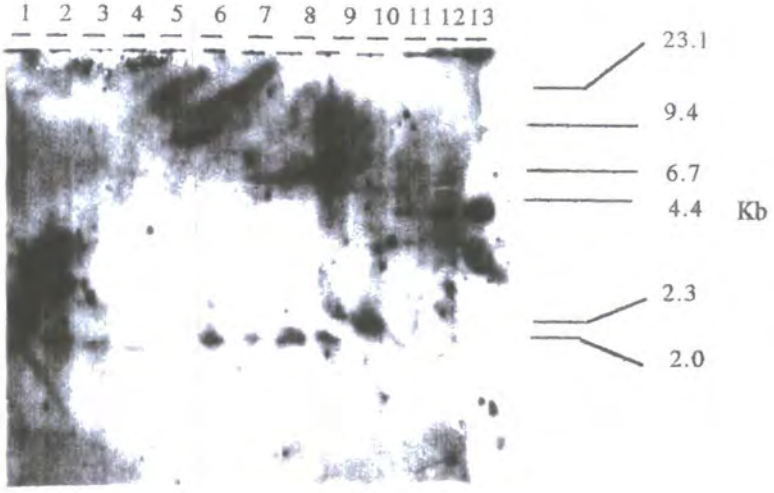
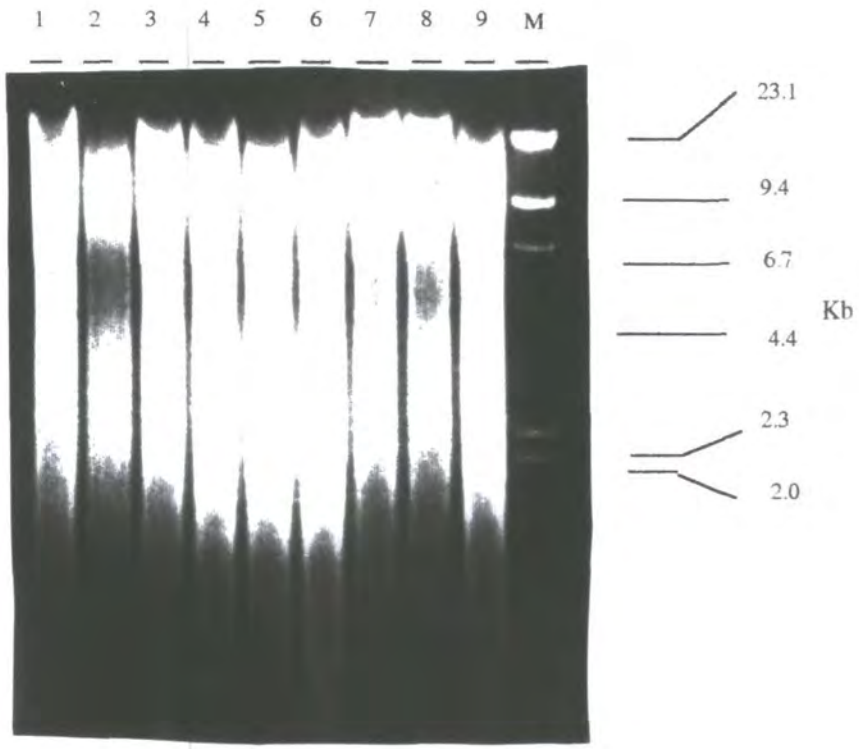

1 2 3 4 5 6 7 8 9 10 11 12 13
23.1
9.4
6.7
4.4
Kb
2.3
2.0

Figure 3.32:

(A) Restriction analysis of eight selected *CaMV-lecA* T1 individuals separated on 0.7%(w/v) agarose gel. Lanes 1-8: represent *HindIII* digested genomic DNA for *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, and 4). Lane 9: control of *HindIII* digested SR1 genomic DNA. Lane M: Size markers Lambda *HindIII* digest.

(B) Southern blot analysis of *HindIII* digested genomic DNA for eight selected *CaMV-lecA* T1 individuals probed with a ^{32}P-labelled left hand border fragment. Lanes 1-8: *CaMV-lecA* T1 individuals (33, 28, 26, 20, 18, 12, 10, and 4). Arrows indicate the presence of inverted repeats.

A.



B.

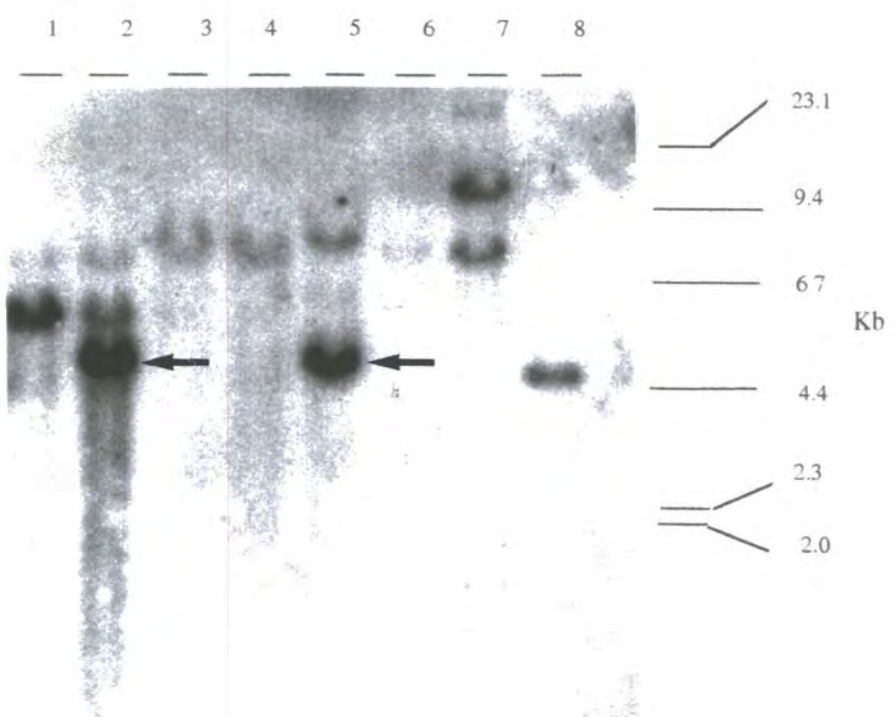
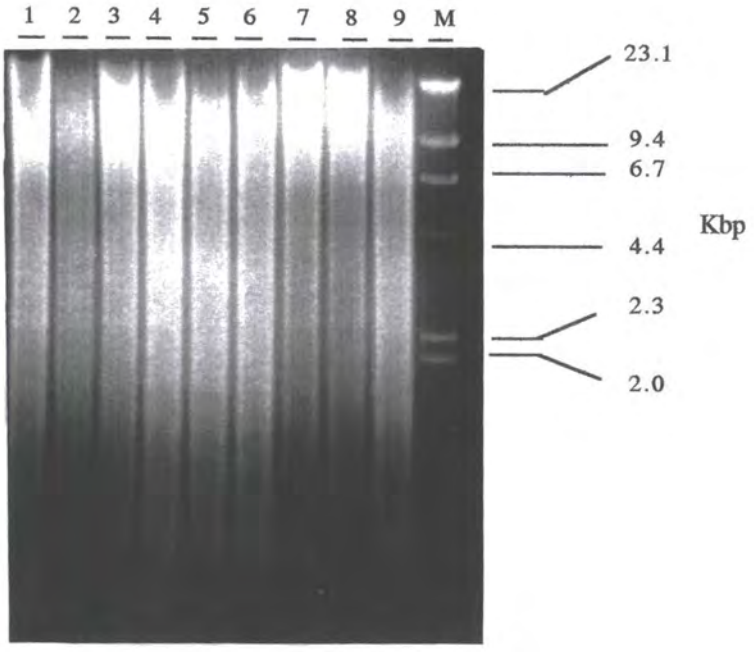


Figure 3.35:

(A) Agarose gel(0.7%w/v) analysis of DNA(10µg) extracted from seven individual T1 plants from the *CaMV-lecA* T1 4 family after digestion with *BamHI*. Lanes 1-9: T1 plants numbers 1-9, respectively. Lane 10: SR1 genomic DNA(10µg) and 2µg of total *Agrobacterium* (pGV3850:130) DNA. Lane M: Lambda *HindIII* digest.

(B) Southern blot analysis of figure 3.35(A) after hybridisation with a 2.1Kb *HindIII/SacI* fragment (containing part of the right hand border sequence with homology to the *nos* promoter). Lanes 1-9: T1 plants numbers 1-9, respectively. Lane 10: SR1 genomic DNA and 2µg of total *Agrobacterium* (pGV3850:130) DNA. Arrows indicate the predicted sizes for internal T-DNA fragments 6.5Kb and 1.4Kb.

A.


1 2 3 4 5 6 7 8 9 M
23.1
9.4
6.7
4.4
2.3
2.0
Kbp

B.

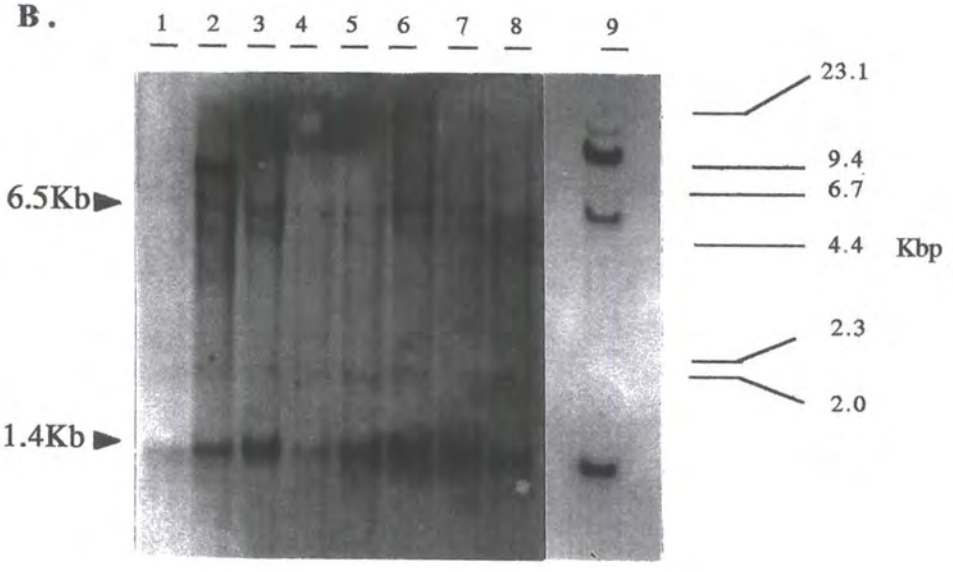

1 2 3 4 5 6 7 8 9
6.5Kb
1.4Kb
23.1
9.4
6.7
4.4
2.3
2.0
Kbp

Figure 3.34:

(A) Agarose gel(0.7%w/v) analysis of DNA(10μg) extracted from eight individual T1 plants from the *CaMV-lecA* T1 20 family after digestion with *BamHI*. Lanes 1-8: T1 plants numbers 1-8, respectively. Lane 9: SR1 genomic DNA(10μg) and 2μg of total *Agrobacterium* (pGV3850:130) DNA.

(B) Southern blot analysis of figure 3.34(A) after hybridisation with a 2.1Kb *HindIII/SacI* fragment (containing part of the right hand border sequence with homology to the *nos* promoter). Lanes 1-8: genomic DNA from T1 plants numbers 1-8, respectively. Lane 9: SR1 genomic DNA(10μg) and 2μg of total *Agrobacterium* (pGV3850:130) DNA. Arrows indicate the predicted sizes for internal T-DNA fragments 6.5Kb and 1.4Kb.

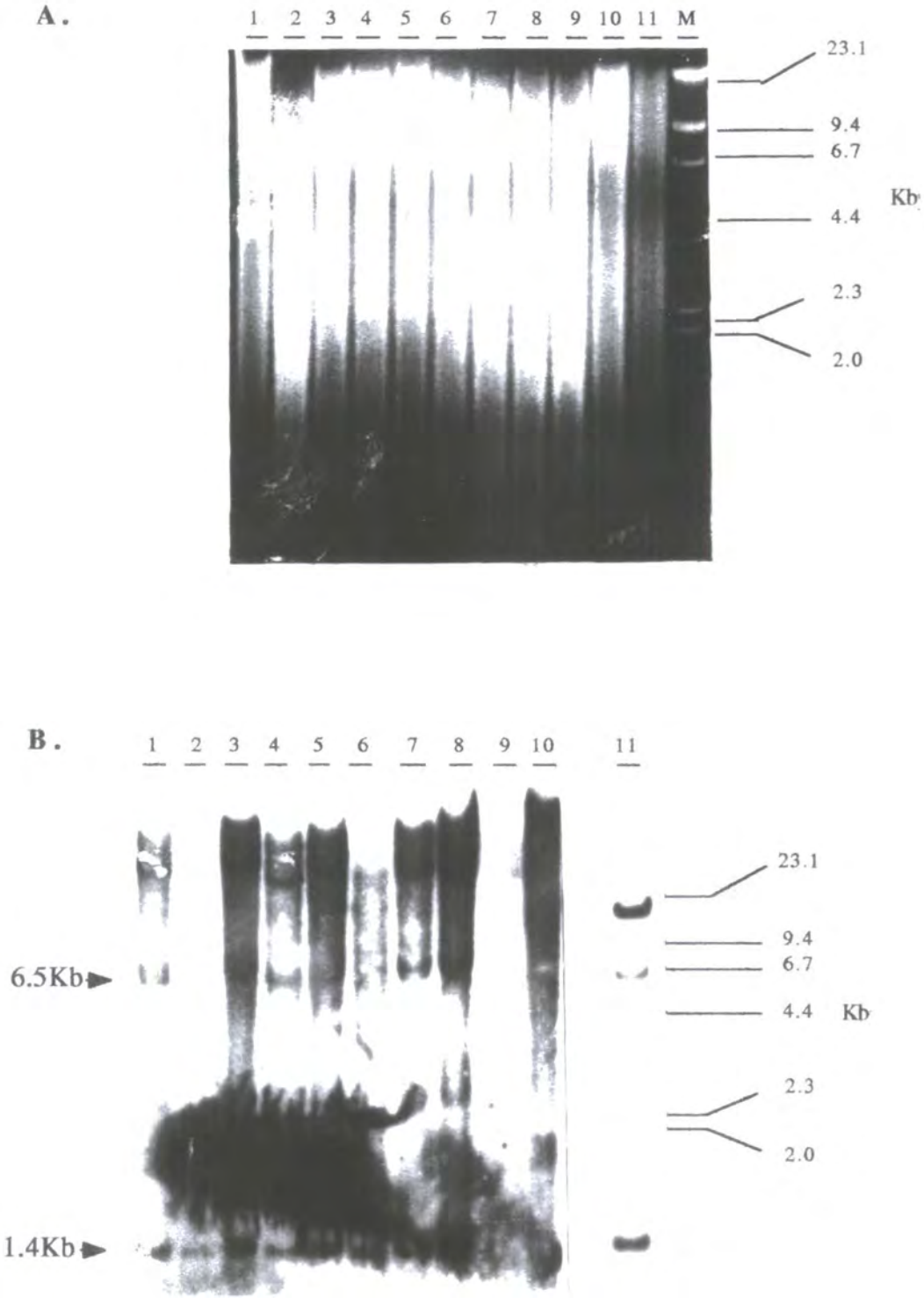
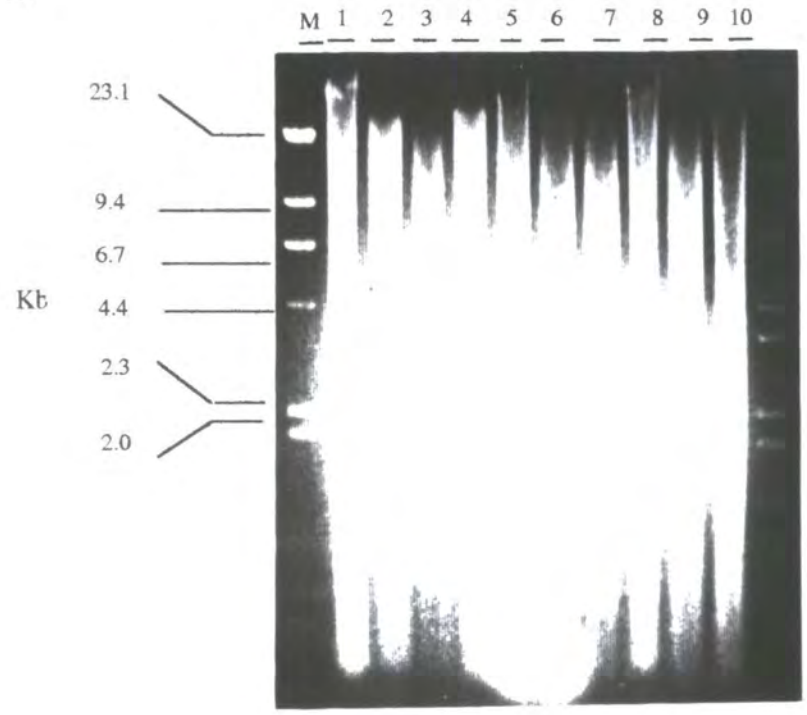

A.
1 2 3 4 5 6 7 8 9 10 11 M
23.1
9.4
6.7
4.4
2.3
2.0
Kb
B.
1 2 3 4 5 6 7 8 9 10 11
6.5Kb
1.4Kb
23.1
9.4
6.7
4.4
2.3
2.0
Kb

Figure 3.33:

(A) Agarose gel(0.7% w/v) analysis of DNA(10µg) extracted from ten individual T1 plants from the *CaMV-lecA* T1 26 family after digestion with *BamHI*. Lanes 1-10: T1 plants numbers 1-10, respectively. Lane 11: SR1 genomic DNA and 2µg of total *Agrobacterium* (pGV3850:130) DNA. Lane M: Lambda *HindIII* digest.

(B) Southern blot analysis of gel shown in figure 3.33(A) after hybridisation with a 2.1Kb *HindIII/SacI* fragment (containing part of the right hand border sequence with homology to the *nos* promoter). Lanes 1-10: genomic DNA(10µg) from T1 plants numbers 1-10, respectively. Lane 11: SR1 genomic DNA(10µg) and 2µg of total *Agrobacterium* (pGV3850:130) DNA. Arrows indicate the predicted sizes for internal T-DNA fragments 6.5Kb and 1.4Kb.

A.


M 1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.7
Kb
4.4
2.3
2.0

B.

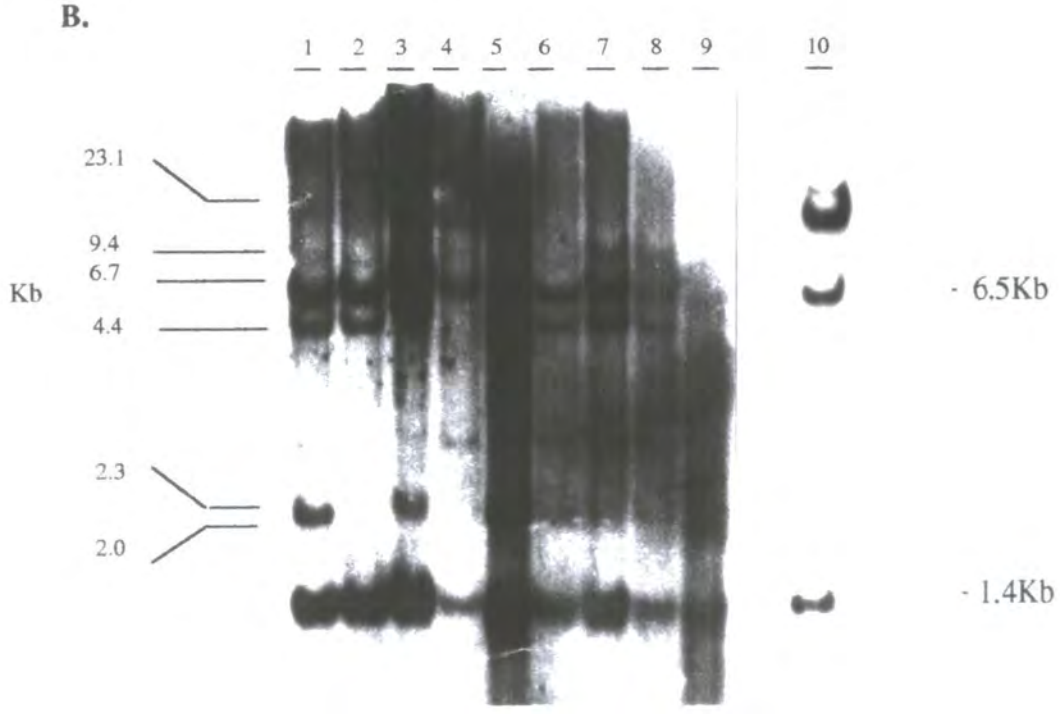

1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.7
Kb
4.4
2.3
2.0
- 6.5Kb
- 1.4Kb

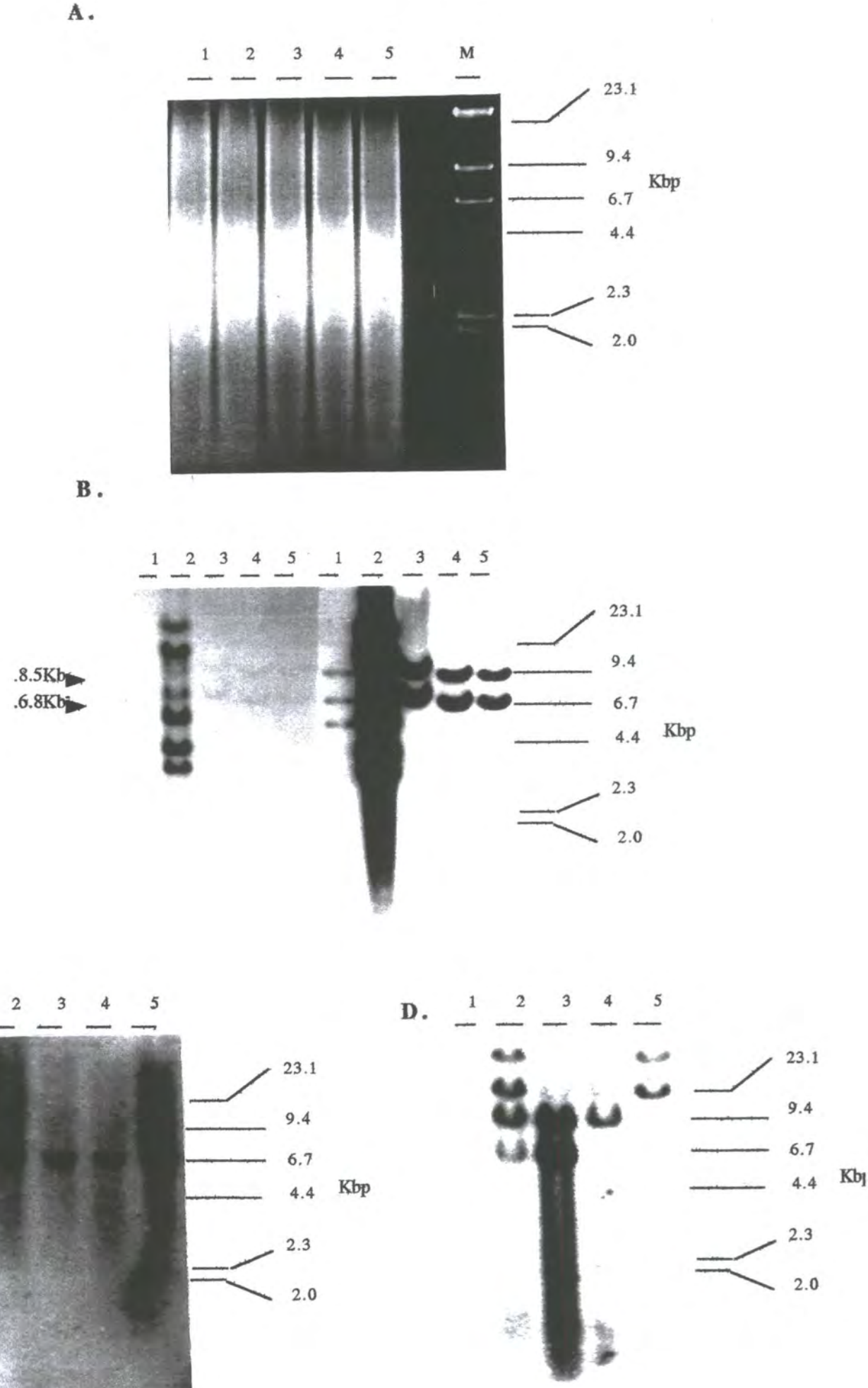
Figure 3.36:

(A) Restriction analysis of 5 selected *ssRubisco-lecA* T1 plants analysed on agarose gel(0.7% w/v). Lanes 1-5: represent *HindIII* digested genomic DNA for 5 *ssRubisco-lecA* T1 individuals (30, 28, 16, 8, & 2). Lane M: Lambda *HindIII* size markers.

(B) Southern blot analysis of the *HindIII* digested genomic DNA from the *ssRubisco-lecA* T1 plants shown in figure 3.36(A) probed with a 4.3Kb linearised pBR322 plasmid. Lanes 1-5: *HindIII* digested genomic DNA for the *ssRubisco-lecA* T1 plants 30, 28, 16, 8, and 2. The blot was underexposed to distinguish the number of bands hybridising in lane 2. Arrows indicate the predicted sizes of 8.5Kb and 6.8Kb for the presence of a 15.3Kb integrated T-DNA.

(C) Southern blot analysis of the *HindIII* digested genomic DNA from the *ssRubisco-lecA* T1 plants shown in figure 3.36(A) probed with a 600bp *Sph I/Pst I* fragment containing the *nos-nptII* gene. Lanes 1-5: *HindIII* digested genomic DNA for the *ssRubisco-lecA* T1 plants 30, 28, 16, 8, and 2. Arrow indicates the predicted size (6.8kb) T-DNA fragment with homology to the probe.

(D) Southern blot analysis of the *HindIII* digested genomic DNA from the *ssRubisco-lecA* T1 plants shown in figure 3.36(A) probed with a 3Kb *EcoRI/HindIII* fragment (containing the left hand border sequence). Lanes 1-5: *HindIII* digested genomic DNA for the *ssRubisco-lecA* T1 plants 30, 28, 16, 8, and 2.


A.
1 2 3 4 5 M
23.1
9.4
6.7
4.4
2.3
2.0
Kbp
B.
1 2 3 4 5 1 2 3 4 5
.8.5Kb
.6.8Kb
23.1
9.4
6.7
4.4
2.3
2.0
Kbp
C.
1 2 3 4 5
8Kb
23.1
9.4
6.7
4.4
2.3
2.0
Kbp
D.
1 2 3 4 5
23.1
9.4
6.7
4.4
2.3
2.0
Kbp

possibility of incomplete restriction digest or the rearrangement of T-DNA structure. In figure 3.35 (B) (line 4) both internal fragments of 6.5Kb and 1.4Kb were observed with homology to the probe in all cases. As with line 20, bands of 5.8Kb and 2.3Kb with homology to the probe were also observed (lanes 1, 3, 5, 6, 7, and 8). The smaller fragment of 2.3Kb could be attributed to the existence of an inverted repeat as observed by Jorgensen *et al.*, (1987). The similar restriction profiles for line 20 and line 4 suggests the same type T-DNA arrangement may be present. The number and organisation of T-DNA integrations within the genome are discused in relation to the genetic anlaysis (section 3.4.3).

3.4.3 Interpretation of T-DNA structure and copy number.

The results of the Southern analyses are summarised in tables 3.8 and 3.9 for *CaMV-lecA* and *ssRubisco-lecA* plants respectively from which conclusions can be drawn on the copy number, organisation and integrity of fragments hybridising to each probe. An attempt will also be made to relate the results of the genetic analyses and Southern hybridisations to the number and arrangement of loci.

The Southern analyses were conducted on both *CaMV-lecA* and *ssRubisco-lecA* T1 plants for three of the inheritance classes (I, II, and III) (section 3.4.1). For class I plants identical ratios for the inheritance of the kanamycin resistance and the lectin phenotype were observed indicating that both transgenes were carried at a single locus. In table 3.8 *CaMV-lecA* T1 plants (26, 20, 18, 7, and 4) belonging to class I show the presence of an intact T-DNA with homology to *lecA*, *nos-nptII* and pBR322 probes. Hybridisation of the left hand border probe to *CaMV-lecA* plants 26, 20, and 4 reveals the presence of a single junction fragment. Plant 18 shows two hybridising fragments, and plant 7 was not included in the analysis. The number of T-DNA locus insertions can be estimated from the number of fragments that represent junctions between the plant DNA and T-DNA. Therefore the presence of one junction fragment for *CaMV-lecA* plants 26, 20, and 4 suggests only one T-DNA integration event in good agreement with the genetic analysis. The difference in intensities observed for the two

junction fragments with plant 18 suggests that the T-DNA has integrated as a repetitive structure carried at a single locus.

In table 3.9 *ssRubisco-lecA* plant 30 belonging to class I displays homology to both *nos-nptII* and pBR322 probes. However, *ssRubisco-lecA* plant 30 shows no homology to the left hand border sequence suggesting rearrangements up to the *HindIII* site as the left hand pBR322 fragment for the T1 plant was observed to be intact. Another rearrangement evident in plant 30 is the presence of an additional fragment of 5.4Kb with homology to the pBR322 probe. It can therefore be concluded that this plant does not contain an intact T-DNA with rearrangements at the left hand border consistent with reports that the junction between the plant DNA and the T-DNA is much more variable at the left border than the right (Matsumoto *et al.*, 1990).

Seven T1 plants were analysed from populations with class II patterns of kanamycin resistance inheritance for the segregation of two independent loci. Five out of the seven plants came from populations where both the kanamycin resistance and lectin phenotype segregated with identical ratios. Two exceptions within this class were *CaMV-lecA* T1 plants 28 and 12 from populations which showed a 15:1 segregation ratio for kanamycin resistance and a 3:1 ratio for lectin expression indicating that one locus carries functional *nos-nptII* and *CaMV-lecA* genes whereas another locus carries only a functional *nos-nptII* transgene, data similar to recent observations by Palaqui and Vaucheret (1995). Both plants showed homology to the *lecA* and *nos-nptII* genes; however both only displayed one fragment of 6.8Kb with homology to the pBR322 probe. The presence of a single junction fragment with homology to the left hand border probe in plant 12 does not agree with the genetic analysis for kanamycin resistance which suggests the presence of two insertions, but does match that for the lectin phenotype. It would therefore seem that plant 12 contains rearranged T-DNAs about the left hand border and within the left hand pBR322 sequence. *CaMV-lecA* plant 28 on the other hand produced three fragments hybridising to the left hand border probe, one of which was the predicted size of an inverted repeat. This would lead one

Table 3.8:

Summary of segregation data and Southern analyses for *CaMV-lecA* plants.

Kanamycin segregation ratios were determined as the ratio of resistant to sensitive seedlings when germinated on medium containing 100 µg/ml kanamycin. Segregation of the *lecA* phenotype was determined by seed squash analysis. Number of loci, T-DNA structure and organisation were all determined from Southern analyses. Copy number was estimated from comparisons with gene copy equivalents.

Key: +: positive for the specified charcteristic.

-: negative for the specified characteristc.

n.d: not determined.

Line	Kan^r:Kan^S ratio	Number of loci	*lecA*$^+$:*lecA*$^-$ ratio	Number of loci	Observed bands with homology to the probe							Copy number	Rearrangement at sequences internal to the T-DNA	T-DNA organised as an inverted repeat
					lecA	*Nos*	pBR322		LB					
					2.1Kb	6.8Kb	6.8Kb	8.3Kb	>3.0Kb	<3.0Kb	6.0Kb			
33	15:1	2	15:1	2	+	+	+	+	+ (2)	-	+	3-5	-	+
28	15:1	2	3:1	1	+	+	+	-	+ (3)	-	+	1	+	+
26	3:1	1	3:1	1	+	+	+	+	+ (1)	-	-	1	-	-
20	3:1	1	3:1	1	+	+	+	+	+ (1)	-	-	1	-	-
18	3:1	1	3:1	1	+	+	+	+	+ (2)	-	-	2	-	-
12	15:1	2	3:1	1	+	+	+	-	+ (1)	-	-	1	+	-
10	15:1	2	15:1	2	n.d	+	+	+	+ (2)	-	-	n.d.	-	-
7	3:1	1	3:1	1	+	+	+	+	n.d.	n.d	n.d.	3-5	-	n.d.
4	3:1	1	3:1	1	+	+	+	+	+ (1)	-	-	2	-	-
3	15:1	2	15:1	2	+	+	+	-	n.d.	n.d	n.d.	3-5	+	n.d.

Table 3.9:

Summary of segregation data and Southern analyses for *ssRubisco-lecA* plants.

Kanamycin segregation ratios were determined as the ratio of kanamycin-resistant to -sensitive seedlings when germinated on medium containing 100 µg/ml kanamycin. Segregation of the *lecA* phenotype was determined from seed squash analysis. Number of loci, T-DNA structure and organisation were determined from Southern analyses.

Key: +: positive for the specified charcteristic.

-: negative for the specified characteristc.

n.d: not determined.

Line	Kan^r:Kan^s ratio	Number of loci	*lecA*$^+$:*lecA*$^-$ ratio	Number of loci	Observed bands with homology to the probe						Rearrangement at sequences internal to the T-DNA	T-DNA organised as an inverted repeat
					Nos	pBR322			LB			
					6.8Kb	6.8Kb	8.5Kb	>3.0Kb	<3.0Kb	6.0Kb		
30	3:1	1	3:1	1	+	+	+	-	-	-	+	-
28	1:0	>2	1:0	>2	+	+	-	+ (4)	-	+	+	+
16	1:0	>2	1:0	>2	+	+	+	+ (3)	-	+	-	+
8	15:1	2	15:1	2	+	+	+	+ (2)	-	-	-	-
2	15:1	2	15:1	2	+	+	+	+ (2)	-	-	-	-

to the conclusion that plant 28 contains four independent insertions in which two copies of the T-DNA were possibly arranged as an inverted repeats. Copy number estimates do not, however, substantiate this, concurring instead with *lecA* segregation data which indicates the presence of a single copy (section 3.4.1) The absence of the left hand pBR322 fragment was also observed for *CaMV-lecA* plant 3 belonging to class II. Plant 3 also showed homology with the *nos-nptII* and *lecA* probes, although no conclusion could be made to the number and arrangement of loci about the left hand border since this plant was not tested. However in the T1 population both the lectin phenoytpe and the kanamycin resistance segregated with identical ratios indicating that both were linked at two independent loci. In the case of *CaMV-lecA* plant 33 all fragments were observed with homology to the probes. The number of junction fragments hybridising to the left hand border was 2 with one of the T-DNAs being arranged as an inverted repeat. This was in agreement with the genetic analysis and suggests that the T-DNA was carried as an inverted repeat at one locus and a single T-DNA element at another. Class II *CaMV-lecA* plant 10 and *ssRubisco-lecA* plants 8 and 2 showed all the expected fragments with homology to the probes indicating the presence of intact T-DNAs, and the production of two junction fragments is in agreement with the genetic analysis for the T-DNA segregating as two independent loci.

Class III plants analysed were *ssRubisco-lecA* plants 28 and 16. Both these plants displayed diferent types of T-DNA structure. Plant 28 showed no homology to the left hand pBR322 probe, as previously observed for *CaMV-lecA* plants 28, 12, and 3. Gross rearrangements were also evident in *ssRubisco-lecA* plant 28 (lane 2, figure 3.36 (B)) where many of the fragments hybridising to labelled pBR322 did not correspond to partial digestion products, suggesting that they represent further altered forms of the T-DNA present in the genome. The number of junction fragments hybridising was four for which one of the fragments was the predicted size of an inverted repeat. The genetic analysis therefore agreed wih the Southern analysis for the segregation of more than two independent loci at which the T-DNA was grossly reaaranged. *ssRubisco-lecA* plant 16 was shown to be intact with bands of homology to all probes used and

displayed three 3 junction fragments with homology to the probe. One of the junction fragments was predicted for the T-DNA arranged as an inverted repeat. The arrangement and number of loci were in agreement with the genetic analysis for more than two independent integration events at which the T-DNA was arranged as an inverted repeat and intact T-DNA molecules at individual genomic locations.

The relationship between T-DNA structure and the number of integrations based on the genetic data for the inheritance of the kanamycin gene is shown in table 3.10. Table 3.10 shows that the frequency of rearrangements is apparently greater with plants containing two or more T-DNA integration events. The majority of the rearrangements were observed at the left hand border and it would therefore seem that there is an apparent increase in this type of rearrangement with increasing number of T-DNA integrations. The organisation of the T-DNA as an inverted repeat and T-DNA structure was also investigated in table 3.11. However no direct relationship was apparent, sugesting that inverted repeat structures are not necessarily associated with aberrant T-DNA structures. It was also evident that inverted repeats occured in greater than 50% of the plants analysed.

Copy number was determined for nine *CaMV-lecA* T1 individuals allowing comparisons with the genetic analyses for the segregation of the lectin phenotype (figure 3.27). Table 3.12 summarises the relationship between *lecA* copy number and the number of *lecA* integrations from which several concusions can be drawn. The results were as expected in that the majority of T1 plants with single gene insertions contain copy numbers of 1 and 2 indicative of hemizygous and homozygous individuals. Four *CaMV-lecA* plants (28, 26, 20, and 12) were found to contain a single copy of the *lecA* gene in agreement with lectin segregation analysis suggesting that these plants were hemizygous for the *lecA* gene. Two plants *CaMV-lecA* 18 and 4 were estimated to contain 2 copies of the *lecA* gene upon comparison with gene copy equivalents and to contain a single insertion locus, thereby indicating that the T1 plants were homozygous for the *lecA* gene at a single locus. Plant 7 on the otherhand with an

Table 3.10:

Analysis of T-DNA structure at the right hand border as determined by Southern analysis for a number of *CaMV-lecA* T1 plants of each of three lines - 26, 20, and 4.

Key: +: positive for the specified charcteristic.

-: negative for the specified characteristc.

n.d: not determined.

Line	Observed bands with homolgy to the RB probe					T-DNA organised as an inverted repeat
	1.4Kb	6.5Kb	>1.15Kb	<1.15Kb	2.3Kb	
4-1	+	+	+ (2)	-	+	+
4-2	+	+	+ (2)	-	+	+
4-3	+	+	+ (2)	-	+	+
4-4	+	+	+ (1)	-	-	-
4-5	+	+	+ (2)	-	+	+
4-6	+	+	+ (2)	-	+	+
4-7	+	+	+ (2)	-	+	+
4-8	+	+	+ (2)	-	+	+
4-9	+	-	+ (1)	-	+	+
4-10	n.d.	n.d.	n.d	n.d	n.d.	n.d.
	1.4Kb	6.5Kb	>1.15Kb	<1.15Kb	2.3Kb	
20-1	+	-	-	-	-	-
20-2	+	+	+ (3)	-	+	+
20-3	+	+	+ (2)	-	+	+
20-4	+	+	+ (2)	-	+	+
20-5	+	+	+ (2)	-	+	+
20-6	+	+	+ (2)	-	+	+
20-7	+	+	+ (2)	-	+	+
20-8	+	+	+ (2)	-	+	+
20-9	n.d.	n.d.	n.d	n.d	n.d.	n.d.
20-10	n.d.	n.d.	n.d	n.d	n.d.	n.d.
	1.4Kb	6.5Kb	>1.15Kb	<1.15Kb	2.3Kb	
26-1	+	+	+ (1)	-	-	-
26-2	+	+	+ (1)	-	-	-
26-3	+	+	+ (1)	-	-	-
26-4	+	+	+ (1)	-	-	-
26-5	+	+	+ (1)	-	-	-
26-6	+	+	+ (1)	-	-	-
26-7	+	+	+ (1)	-	-	-
26-8	+	+	+ (1)	-	-	-
26-9	-	-	-	-	-	-
26-10	+	+	+ (1)	-	-	-

Table 3.11:

Tabulated representation of the relationship between T-DNA structure and number of integration loci, as determined from genetic and Southern analyses, for 15 T1 transgenic lines representing both *ssRubisco-lecA* and *CaMV-lecA* plants.

Table 3.12:

Tabulated representation of the relationship between T-DNA structure and organisation, as determined from genetic and Southern analyses, for 13 T1 transgenic lines representing both *ssRubisco-lecA* and *CaMV-lecA* plants.

Key: IR=inverted repeat.

Number of integrations	T-DNA structure		Total
	Intact	Rearranged	
1	5	1	6
2	4	3	7
>2	1	1	2
Total	10	5	15

T-DNA organisation	T-DNA structure		Total
	Intact	Rearranged	
IR	5	2	7
None	4	2	6
Total	9	4	13

estimated 3-5 *lecA* copies presumably contained several copies at one locus. *CaMV-lecA* plants 33 and 3 from segregating populations for two independent loci were estimated to have 3-5 copies of the *lecA* gene suggesting these plants were homozygous for more than one T-DNA insertion.

The relationship between copy number and T-DNA structure was also investigated; however there was no apparent relationship (table 3.13). This is because most of the rearrangements were at the left hand border and not within the *lecA* gene itself. A comparison of copy number with T-DNA organisation in table 3.14 indicated that there was no clear relationship between copy number and inverted repeats.

CaMV-lecA lines 26, 20 and 4 were chosen to complete the molecular analysis at the right hand border because they already contained an intact T-DNA at a single locus insertion as determined by genetic and Southern analysis. This characterisation involved the analysis of a number of T1 plants for each chosen line representing both hemizygotes and homozygotes. The results of the analysis are summarised in table 3.15 from which conclusions can be drawn concerning the complete arrangement of the T-DNA at a single locus. The majority of T1 plants for each line showed homology with all the internal fragments up to the *Bam H1* site within the right border and junction fragments greater than 1.15Kb suggesting no rearrangements of the T-DNA extending from the *BamHI* site within the *CaMV-lecA* gene to the *SacII* site which lies close to the right repeat of the right hand border. The production of one junction fragment for most of line 26 T1 plants, not consistent with the predicted sizes for repeat formation, was in agreement with previous results for the single integration of an intact T-DNA. On the other hand the majority of line 4 and line 20 T1 plants displayed two junction fragments of 5.8Kb and 2.3Kb of equal intensities. The 2.3kb band is the predicted size between *BamHI* sites when the T-DNA is arranged as an inverted repeat around the right hand border. The 5.8Kb band may be predicted to represent a left hand pBR322 fragment containing *nos-npt II* gene up to the *HindIII* site of the *CaMV* promoter. The envisaged inverted repeat structure for *CaMV-lecA* plants 4 and 20 is

Figure 3.38:

(A) Autoradiograph of a western slot blot after probing with rabbit anti-pea lectin antibodies and detection via autoradiography with a donkey anti-rabbit IgG ^{125}I conjugate for total soluble protein extracted from 32 *CaMV-lecA* T1 families.

(B) A plot of the mean relative peak area percentages with the log $_{10}$ of standard pea lectin dilutions(90ng-0.3ng). Peak areas were determined for all 32 *CaMV-lecA* T1 families and the amounts of pea lectin protein estimated by extrapolating from the standard curve.

(C) Graphical representation of figure 3.38(A) for pea lectin levels estimated for 32 *CaMV-lecA* T1 families given as a percentage of total soluble protein loaded per slot (4µg). Error bars represent the spread of three repeat experiments.

shown in figure 3.37 (B). In conjunction with the previous Southern analysis it can be concluded that T1 plants 26, 20, and 4 contain an intact T-DNA between the right and left hand borders carried at a single locus. On the other hand in respect of T-DNA organisation at a given locus, two out of the three lines (20 and 4) had the T-DNA organised as an inverted repeat at their right hand border.

3.4.4 Concluding Remarks.

From the above investigations it may be observed that in the majority of cases kanamycin segregation data was found to be consistent with the *lecA* segregation data, indicating that both genes linked on the T-DNA generally remain so following integration into the plant genome.

A high frequency of multiple itegration events was observed, as determined from the kanamycin segregation data. This is a phenomenon frequently observed with the type of *Agrobacterium* plant transformation system used here. It was also found that those plants containing two or more T-DNAs integrated at two or more sites showed a higher frequency of T-DNA rearrangements. All rearrangements noted were observed to occur at or near the left hand border.

The insertion of more than one T-DNA into a single locus was found to result, in a number of plants, in the generation of inverted repeat structures (determined by Southern analysis). However, no significant correlation was observed between the presence of these structures and the number of T-DNAs integrated into the plant genome.

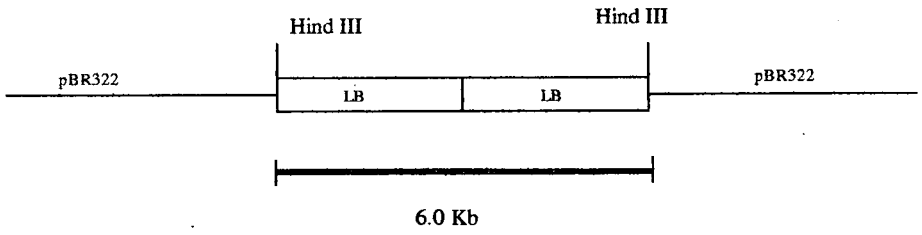
The T-DNA integrity and organisation at individual loci for both *CaMV-lecA* and *ssRubisco-lecA* lines had now been well established by both genetic and molecular techniques. Subsequent analysis of pea lectin levels in the T1 plants would allow conclusions to be reached as to the effect of copy number, number of integration loci, rearrangements and T-DNA organisation on levels of transgene expression.

3.5 Measurement of lectin levels in the T1 *CaMV-lecA* and *ssRubisco-lecA* plants.

Seed collected from the selfing of *CaMV-lecA* and *ssRubisco-lecA* T0 independent lines were germinated simultaneously. T1 plants were transferred to soil at the same development age allowing pea lectin level measurements to be made on plants grown synchronously (i.e. at the same time and under exactly the same conditions of temperature, humidity and lighting).

Leaf discs were removed from the first upper leaf at the six leaf stage and pooled together for 10 T1 plants derived from a single T0 parental line. Pea lectin accumulation was estimated in T1 families for each of the selfed 31*CaMV-lecA* T0 and 16 *ssRubisco-lecA* T0 lines using quantitative slot blot radioimmunoassays (RAI). The term T1 family will be used henceforth to refer to any number of T1 individuals derived from a single T0 parental line. The slot blots were scanned using an L.K.B. laser densitometer and a standard curve constructed from a pea lectin dilution series (two-fold dilutions) from 90ng down to 0.3ng (figures 3.38 (B) and 3.39 (B)). Pea lectin levels were expressed as a %(w/w) of total leaf soluble protein and the results representing pea lectin levels within *CaMV-lecA* and *ssRubisco-lecA* T1 families displayed in figures 3.38 (C) and 3.39 (C). Pea lectin levels within both T1 populations were observed to be variable. Statistical analysis was used to compare the accumulation of pea lectin protein between the *CaMV-lecA* and *ssRubisco-lecA* T0 and T1 populations (figure 3.40). Results were termed significant when $P<0.05$. Table 3.16 gives the mean, median, standard deviation and variation coefficient of the accumulation levels. The Bartletts test was used to determine whether the variances between the T0 and T1 *CaMV-lecA* populations were homogenous (see table 3.17).

A.


Hind III
Hind III
pBR322
LB
LB
pBR322
6.0 Kb

B.

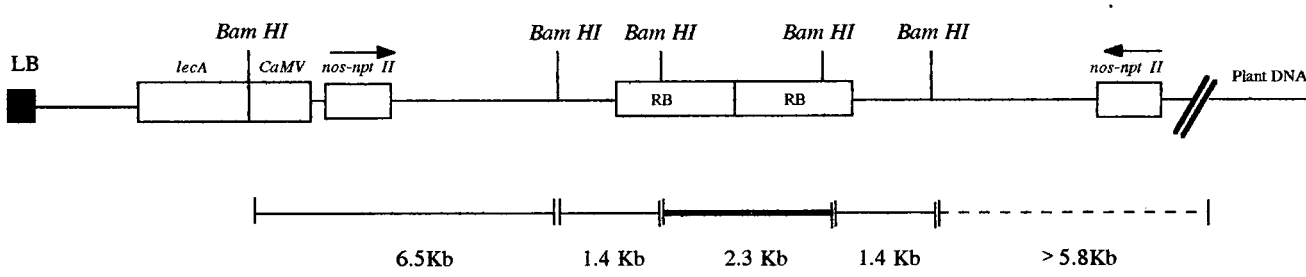

LB
Bam HI
lecA
CaMV
nos-npt II
Bam HI
Bam HI
Bam HI
Bam HI
RB
RB
nos-npt II
Plant DNA
6.5Kb
1.4 Kb
2.3 Kb
1.4 Kb
> 5.8Kb

Figure 3.37

Schematic representation of predicted T-DNA organisation following integration as inverted repeats at (A) the left hand border and (B) the right hand border. Predicted fragment sizes are shown for (A) *HindIII* -digested DNA and (B) *BamHI*-digested DNA.

Copy number	Number of integrations 1	2	Total
1	4	0	4
2	2	0	2
3-5	1	2	3
Total	7	2	9

Copy number	T-DNA organisation IR	none	Total
1	3	1	4
2	1	1	2
3-5	1	0	1
Total	5	2	7

Copy number	T-DNA structure Intact	Rearranged	Total
1	2	2	4
2	2	0	2
3-5	2	1	3
Total	6	3	9

Table 3.13:

Illustration of the relationship between *lecA* copy number and number of integration loci.

Estimates of *lecA* copy number were obtained from comparisons with gene copy equivalents. The number of integration loci was determined from genetic and Southern analyses.

Table 3.14:

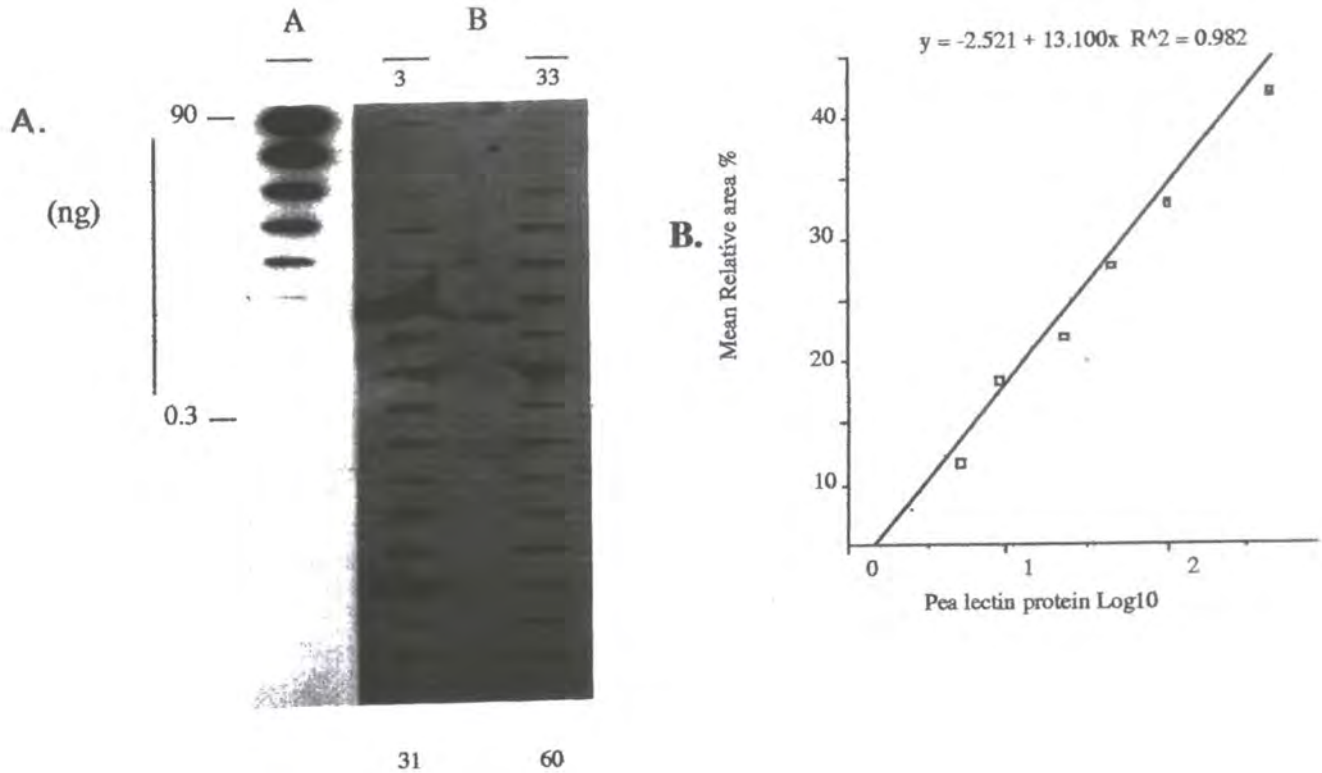
Illustration of the relationship between *lecA* copy number and T-DNA organisation.

Estimates of *lecA* copy number and T-DNA organisation were determined as above.
Key: IR=inverted repeat.

Table 3.15:

Illustration of the relationship between *lecA* copy number and T-DNA structure.

Estimates of *lecA* copy number and T-DNA structure were determined as above.


A.
A
B
3
33
90
(ng)
0.3
31
60
B.
y = -2.521 + 13.100x R^2 = 0.982
Mean Relative area %
40
30
20
10
0
1
2
Pea lectin protein Log10

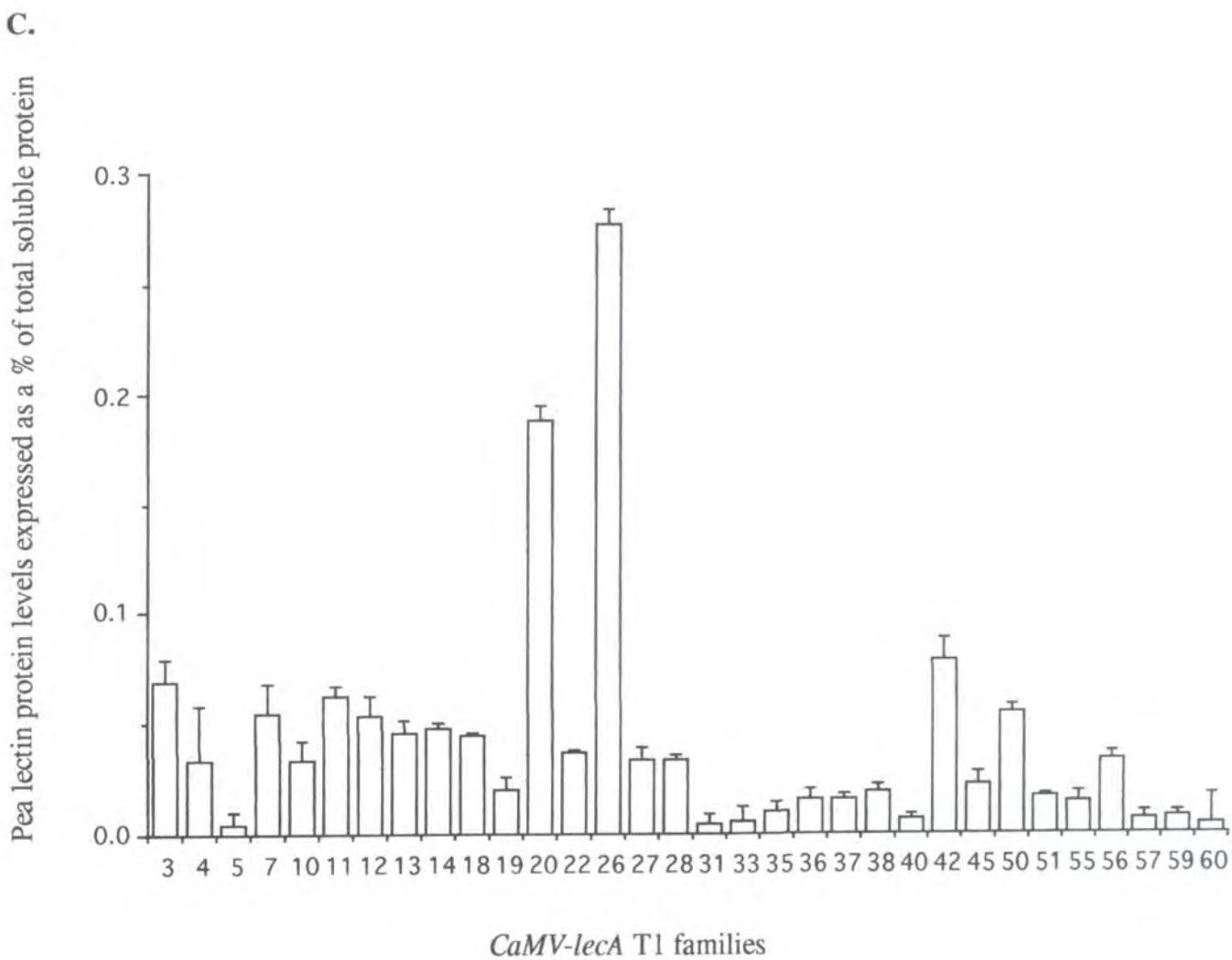

C.
Pea lectin protein levels expressed as a % of total soluble protein
0.3
0.2
0.1
0.0
3 4 5 7 10 11 12 13 14 18 19 20 22 26 27 28 31 33 35 36 37 38 40 42 45 50 51 55 56 57 59 60
CaMV-lecA T1 families

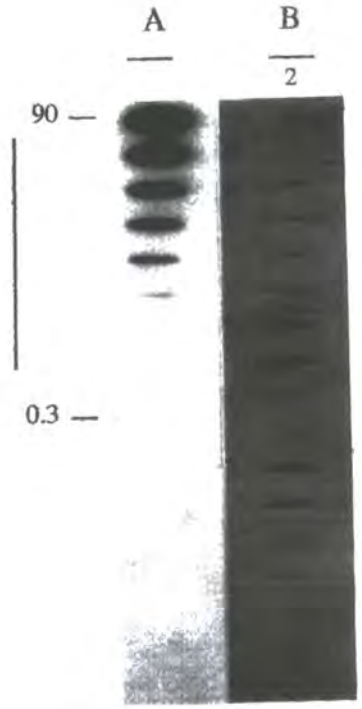
Figure 3.39:

(A) Autoradiograph of a western slot blot after probing with rabbit anti-pea lectin antibodies and detection via autoradiography with a donkey anti-rabbit IgG ^{125}I conjugate for total soluble protein extracted from 16 *ssRubisco-lecA* T1 families.

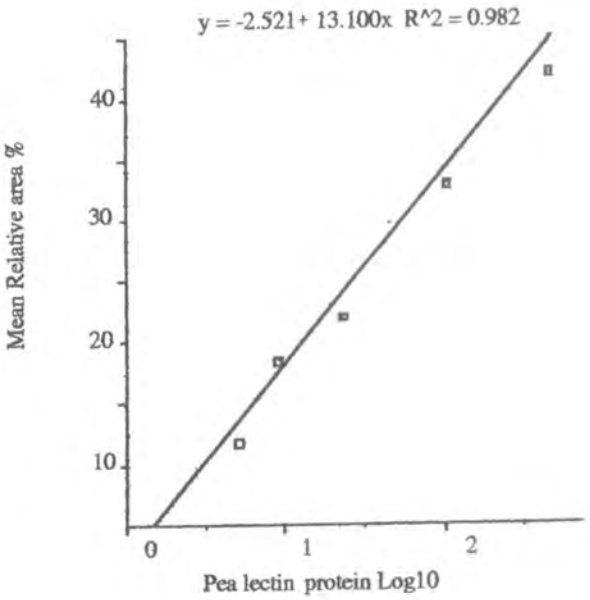
(B) A plot of the mean relative area percentages under the peaks with the log $_{10}$ of standard pea lectin dilutions(90ng-0.3ng). Peak areas were determined for all 16 *ssRubisco-lecA* T1 families and the amounts of pea lectin protein estimated by extrapolating from the standard curve.

(C) Graphical representation of figure 3.39(A) for pea lectin levels estimated for 16 *ssRubisco-lecA* T1 families given as a percentage of total soluble protein loaded per slot (4µg). Error bars represent the spread of three repeat experiments.

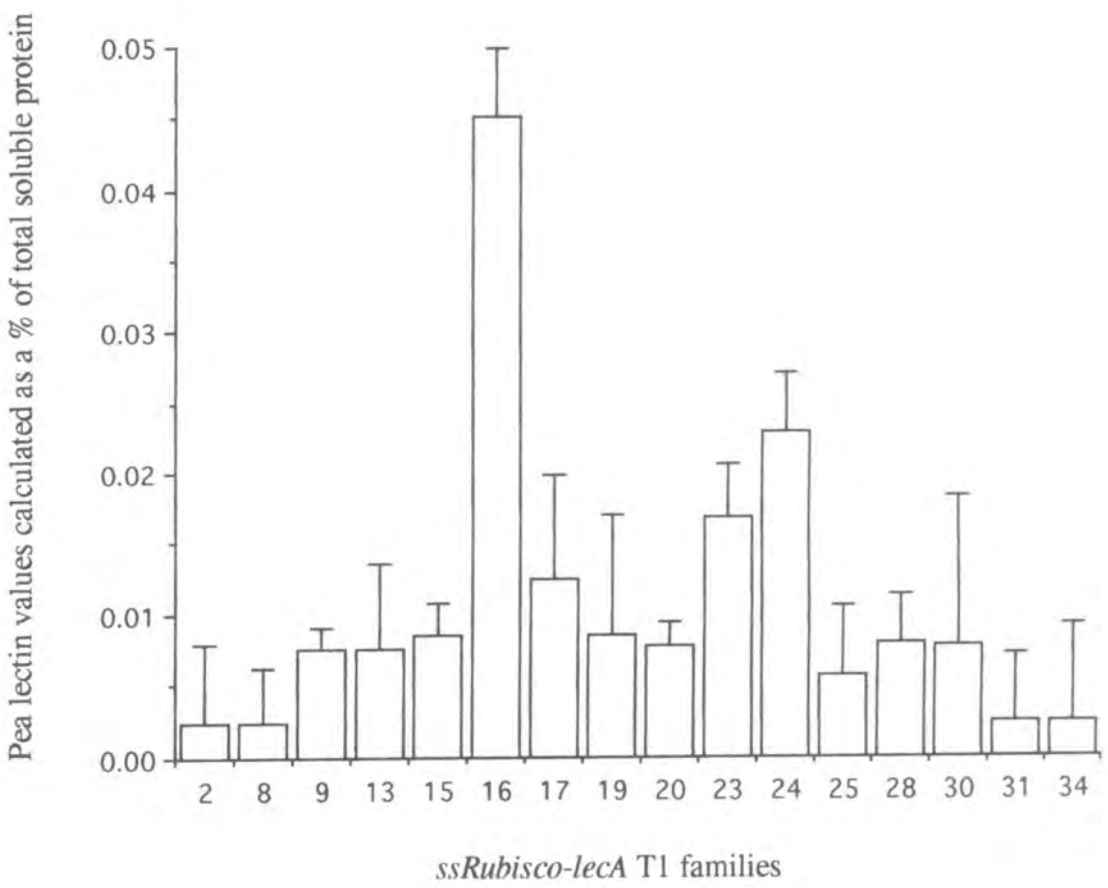
A.


A
B
2
90
(ng)
0.3
34

B.


y = -2.521+ 13.100x R^2 = 0.982
Mean Relative area %
40
30
20
10
0
1
2
Pea lectin protein Log10

C.


Pea lectin values calculated as a % of total soluble protein
0.05
0.04
0.03
0.02
0.01
0.00
2 8 9 13 15 16 17 19 20 23 24 25 28 30 31 34
ssRubisco-lecA T1 families

Mean variance = $\sum(vx)/\sum(n) = 4.39x10^{-3}/2 = 2.195x10^{-3}$

n(log mean variance) = 2(-2.6586) = -5.3171

Chi-squared, X^2 = 2.3026[(n-1)(nlog mean variance - $\sum$ (log vx)

= 2.3026[-5.3171-(-5.3245)]

= 0.017

This analysis revealed that X^2(0.017) was less than the tabular X^2 value for one degree of freedom at the 5%(3.841) and 1%(6.635) significance levels suggesting that the variances between the *CaMV-lecA* T0 and T1 populations were homogenous.

Similarly the Bartletts test was applied to the population data for both the *ssRubisco-lecA* T0 and T1 populations (see table 3.18).

Mean variance = $\sum(vx)/\sum(n) = 5.63x10^{-3}/2 = 2.815x10^{-3}$

n(log mean variance) = 2(-2.5505) = -5.1010

Chi-squared, X^2 = 2.3026[(n-1)(nlog mean variance - $\sum$ (log vx)

= 2.3026[-5.3171-(-6.3398))]

= 2.85

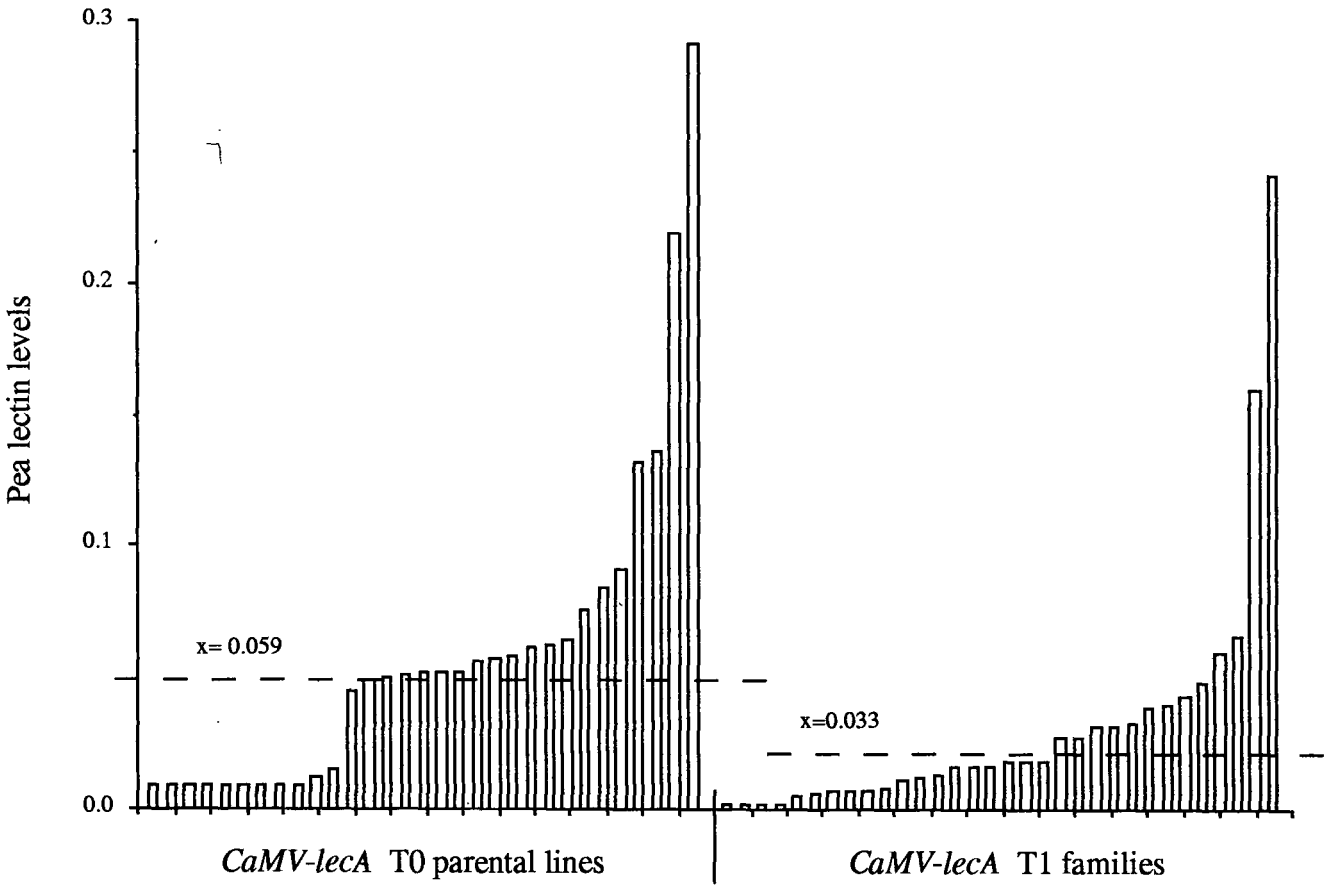
This analysis revealed that X^2(2.85) was less than the tabular X^2 value for one degree of freedom at the 5%(3.841) and 1%(6.635) significance level suggesting that the variances between the *ssRubisco-lecA* T1 and T0 populations were also homogenous. As the variances for both the *CaMV-lecA* (T0 and T1) and *ssRubisco-lecA* (T0 and T1) populations were found to be homogenous, a t-test could be applied to test the hypothesis whether the sample means were significantly different for *CaMV-lecA* (T0 and T1) and *ssRubisco-lecA* (T0 and T1) populations.

$t(n-1) = \frac{d}{Sd} = \frac{\text{variation between the series}}{\text{Variation within the series}}$

t = 0.024/0.0190

t= 66.33

A.


0.3
0.2
0.1
0.0
Pea lectin levels
x= 0.059
x=0.033
CaMV-lecA T0 parental lines
CaMV-lecA T1 families

B.

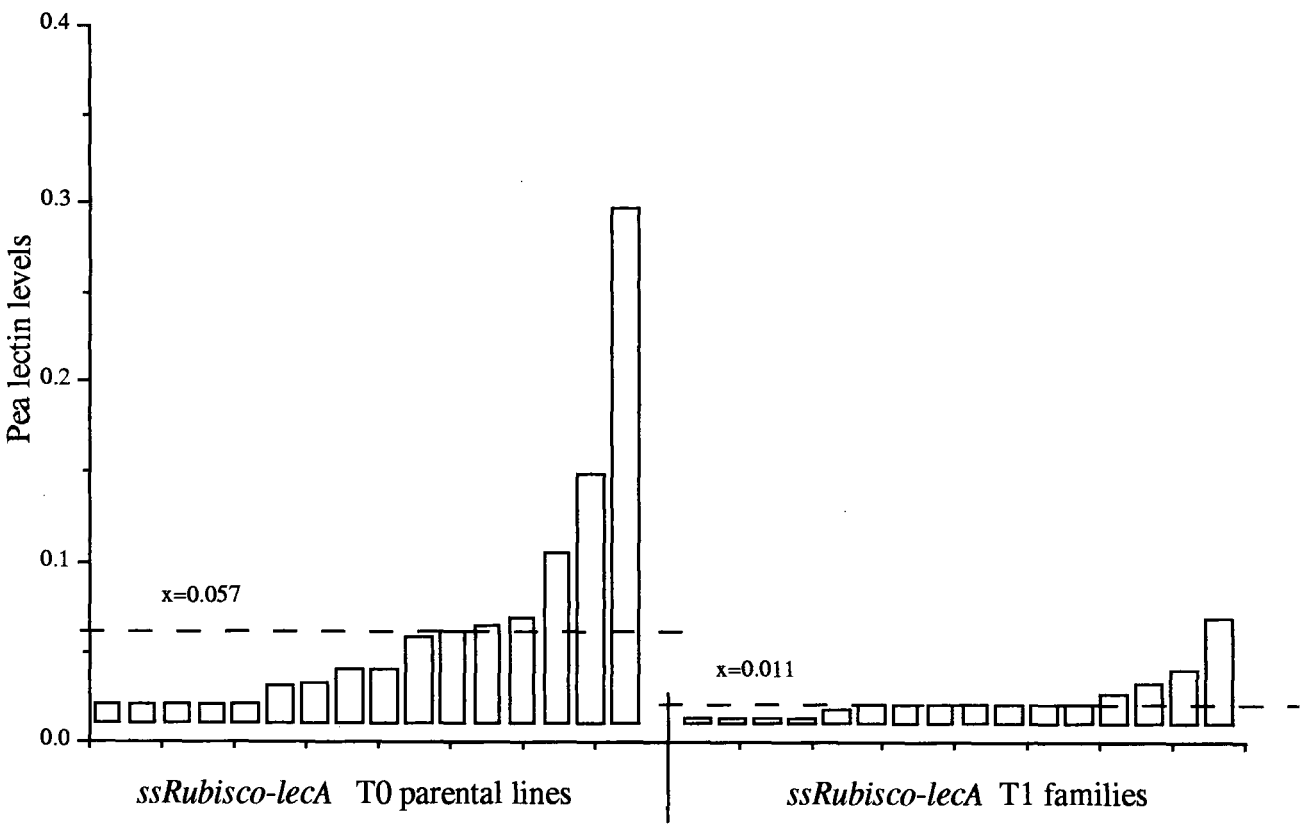

0.4
0.3
0.2
0.1
0.0
Pea lectin levels
x=0.057
x=0.011
ssRubisco-lecA T0 parental lines
ssRubisco-lecA T1 families

Figure 3.40:

(A) A graphical representation of pea lectin protein levels within T0 *CaMV-lecA* parental lines and respective progeny *CaMV-lecA* T1 families. Samples have been ranked in order of increasing lectin level and do not nescessarily correspond directly.

(B) A graphical representation of pea lectin protein levels within T0 *ssRubisco-lecA* parental lines and respective progeny *ssRubisco-lecA* T1 families. Samples have been ranked in order of increasing lectin level and do not nescessarily correspond directly. Samples have been ranked in order of increasing lectin level and do not nescessarily correspond directly.

The t value obtained for both *CaMV-lecA* T0 and T1 populations was greater than the 5% level ($P<0.05$) suggesting that the mean level of pea lectin accumulation is significantly different.

$t(n-1) = \underline{d} = \underline{\text{variation between the series}}$

Sd Variation within the series

$t = 0.046/0.0187$

$t = 2.45$

The t value obtained for both *ssRubisco-lecA* T0 and T1 populations was greater than the 5%($P<0.05$) level suggesting that the mean level of pea lectin accumulation is also significantly different.

Since a Bartletts test had shown the variances between the *ssRubisco-LecA* and *CaMV-lecA* T0 populations to be heterogeneous and the difference in average pea lectin accumulation to be similar (see section 3.2.4), the variances between the T1 populations were also tested (see table 3.19).

Mean variance $= \sum(vx)/\sum(n-1) = 9.1824 \times 10^{-3}/46 = 1.996 \times 10^{-4}$

(log mean variance) $\sum(n-1) = (-3.6998)\ (46) = -170.1908$

Chi-squared, $X^2 = 2.3026[(\text{log mean variance}) \sum(n-1) - \sum(n-1)\ (\log vx)$

$= 2.3026[-170.1908 - (-311.88)]$

$= 141.68$

This analysis revealed that X^2 (141.68) was greater than the tabular X^2 (3.841) value for one degree of reedom at the 5% significance level, suggesting they were heterogenous. As the variances were found to be heterogenous a Cochran and Cox test was used to test whether the population means of the two different series were significantly different with no hypothesis about the population variances (table 3.20).

Table 3.16:

Main statistical data of pea lectin accumulation in both *CaMV-lecA* and *ssRubisco-lecA* T0 and T1 populations.

Table 3.17:

Computation of Bartlett's test of homogeneity of variance for *CaMV-lecA* T0 and T1 populations similar in size.

Table 3.18:

Computation of Bartlett's test of homogeneity of variance for *ssRubisco-lecA* T0 and T1 populations similar in size.

Data	*CaMV-lecA* T0	*CaMV-lecA* T1	*ssRubisco-lecA* T0	*ssRubisco-lecA* T1
N^{o} of Plants (n)	32	32	16	16
Mean (x)	0.057	0.033	0.057	0.011
Median	0.057	0.018	0.031	0.008
Variance (vx)	0.00248	0.00191	0.00555	0.0000824
Standard deviation (sx)	+/- 0.0498	+/- 0.0437	+/- 0.0745	+/- 0.00227
Coefficient of variation (C)	114%	75%	76.5%	121%

Population	$\sum x^2$	D.F.(n-1)	vx	log(vx)
T0	0.181	31	0.00248	-2.606
T1	0.094	31	0.00191	-2.719
n=2	0.275			

Population	$\sum x^2$	D.F.(n-1)	vx	log(vx)
T0	0.137	15	0.00555	-2.255
T1	0.094	15	0.000082	-4.084
n=2	0.003			

Table 3.19:

Computation of Bartlett's test of homogeneity of variance for *CaMV-lecA*[a] and *ssRubisco-lecA*[b] T1 populations differing in size.

Table 3.20:

An approximate method of testing the hypothesis whether the sample means of the two different series were significantly different.

Population	$\sum x^2$	D.F.(n-1)	1/(n-1)	vx	log(vx)	(n-1)log(vx)
1[a]	0.094	31	0.0322	0.00191	-2.718	-84.258
2[b]	0.003	15	0.0667	0.00008	-4.084	-61.26
n=2	0.097	46	0.0989			-145.518

Population	Number of plants(n)	D.F.	Mean	$\sum x^2$	vx	$\sum x^2$=vx/n	$t_{.05}$
CaMV-lecA	32	31	0.033	0.094	0.00191	2.84×10^{-4}	2.000
ssRubisco-lecA	16	15	0.011	0.003	0.00008	5.15×10^{-6}	2.030
	48	46				5.78×10^{-4}	

$\sum x = \sqrt{6.48 \times 10^{-6}} = 8.05 \times 10^{-3}$

$x/\sum x = 0.022/8.05 \times 10^{-3} = 2.73$

$$5\% \text{ level} = \frac{(5.96 \times 10^{-5})(2.042) + (5.15 \times 10^{-6})(2.131)}{5.96 \times 10^{-5} + 5.15 \times 10^{-6}}$$

$t_{0.5} = 2.04$

Since the ratio 2.73 was greater than the 5% level (t=2.04), the difference in the average pea lectin accumulation was significant($P<0.05$) between the T1 populations of both constructs.

The aim of the statistical analysis was to compare the T0 and T1 populations for both constructs on the basis of their mean level accumulation of pea lectin protein. Similar statisitical treatment of data has been carried out by (De Neve *et al.*, 1993, Breyne *et al.*, 1992) in analysis of transgenic *Nicotiana* populations. Table 3.16 shows that the mean level of accumulation of pea lectin protein is 1.7 fold lower in the T1 *CaMV-lecA* population and 5 fold lower in the T1 *ssRubisco-lecA* population, a difference which was significant ($P<0.05$) when compared with their respective T0 parental populations. The variation within both populations was the same as might be expected since the site of T-DNA integration did not change. However the diminished levels of pea lectin accumulation within the T1 populations suggests some degree of gene silencing. Statistical comparison of the mean accumulation levels between T1 populations for each construct showed a 3-fold lower level of lectin protein within the *ssRubiso-lecA* T1 population than in the *CaMV-lecA* T1 population, a difference which was significant ($P<0.05$). This was suprising since there was no difference between both T0 populations and indicated that there may be a greater occurrence of gene silencing within the *ssRubisco-lecA* T1 population. This observation illustrates the possibility that promoter-mediated co-suppression is occurring, a phenomenon described by (Brusslan and Tobin, 1995). This possibility, coupled with the low frequency with which single integration events appeared to occur within that population and the high frequency with which T-DNA rearrangements were detected led to the decision not to

include this material in future analysis.

To assess the relative impact of all possible T-DNA-related factors influencing levels of transgene expression (number of T-DNAs integrated (copy number), number of integration sites, structure and organisation of integrated T-DNAs, and location of integration sites within the genome) a series of further studies were carried out.

3.6 T-DNA structure and pea lectin expression.

To investigate the influence of T-DNA copy number and structure on levels of transgene expression the data from plants analysed by Southern hybridisation were coupled together with their respective pea lectin levels in table 3.21 from which some conclusions may be drawn.

3.6.1 The influence of copy number on transgene expression.

A representation of the relationship between copy number and levels of *lecA* gene expression is shown in figure 3.41 (B). Although a relationship does appear to exist between both parameters suggesting that an increase in copy number leads to a reduction in expression levels this hypothesis is not borne out by statistical analysis. A correlation coefficient (r) of -0.55 was obtained which, upon comparison with tabulated values, was found to be not significant at the 5% confidence limit. Therefore one must conclude that the number of T-DNAs integrated does not correlate with measured levels of transgene expression, and hence cannot fully account for the variability in expression levels observed between independent transformants. A possible contributory effect however cannot be fully discounted.

3.6.2 The influence of T-DNA structure and organisation on transgene expression.

Correlating the number of integration sites (locus number) with levels of transgene expression results in a correlation coefficient (r) of -0.304, a value which is not significant ($p<0.05$) indicating no significant relationship between these variables. This

Table 3.21:

Tabulation of levels of transgene expression[a], number of integrations[b] and *lecA* copy number[c] for a number of *CaMV-lecA* T1 plants.

Lectin levels were determined as a percentage of total soluble protein; number of loci were determined from segregation data and *lecA* copy number from Southern analyses. All data was collated for individual *CaMV-lecA* T1 lines from selected T1 families shown in figure 3.38.

CaMV-lecA T1 individual	Pea lectin levels[a]	loci number[b]	copy number[c]
33	0.048	2	3-5
28	0.047	1	1
26	0.31	1	1
20	0.26	1	1
18	0.051	1	2
12	0.049	1	1
10	0.043	2	n.t.
7	0.041	1	3-5
4	0.038	1	2
3	0.022	2	3-5

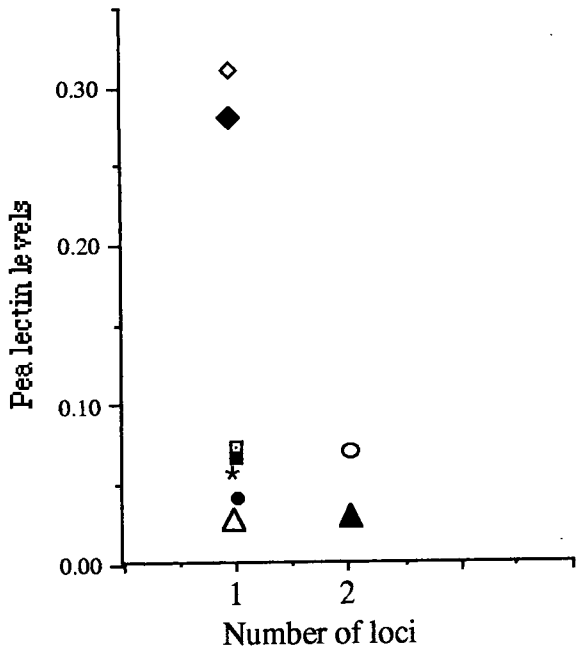
Figure 3.41:

(A) Graph showing the relationship of locus number and pea lectin levels for nine *CaMV-lecA* T1 plants.

(B) Graph showing the relationship of copy number and pea lectin levels for nine *CaMV-lecA* T1 plants

Key:
CaMV-lecA plants: 33, (○), 28 (●), 26 (◇), 20 (◆), 18 (▫), 12 (⋆), 7 (△), 4 (■), and 3 (▲).

A.


Pea lectin levels
0.30
0.20
0.10
0.00
1
2
Number of loci

B.

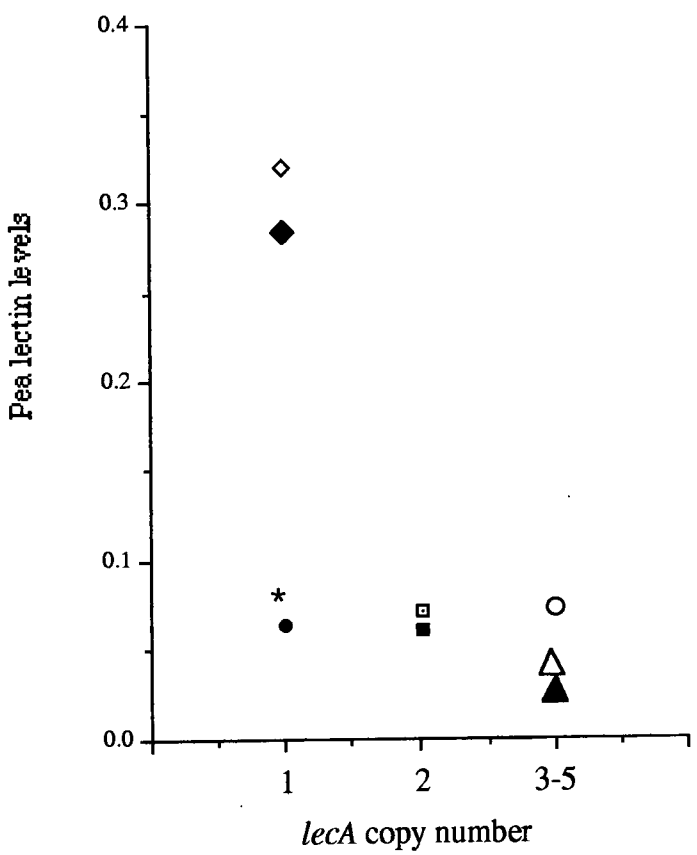

Pea lectin levels
0.4
0.3
0.2
0.1
0.0
1
2
3-5
lecA copy number

Table 3.22:

Tabular representation of the relationship between levels of transgene expression and T-DNA organisation.

Lectin levels were determined by radioimmunoassay, T-DNA organisation by Southern analysis.

Key: IR=inverted repeat.

Table 3.23:

Tabular representation of the relationship between levels of transgene expression and T-DNA structure.

Lectin levels and T-DNA structure were determined as above.

T-DNA organisation	Pea lectin levels		Total
	Low	High	
IR	2	2	4
None	5	0	5
Total	7	2	9

T-DNA structure	Pea lectin levels		Total
	Low	High	
intact	5	2	7
rearranged	3	0	3
Total	8	2	10

data is illustrated in figure 3.41 (A). Thus, as in the case of copy number effects, the number of integration events is not the major factor in determining levels of transgene expression.

Tables 3.22 and 3.23 illustrates the relationship between T-DNA organisation and structure on levels of *lecA* gene expression. It may be seen that the presence of inverted repeats does not appear to exert an influential effect; indeed it can be noted that a higher percentage of plants with no repeats tend towards low levels of *lecA* gene expression than do those with inverted repeats. These data are in agreement with previously published reports (Jorgensen *et al.*, 1987).

As shown in table 3.23 there does not appear to be a positive relationship between low levels of gene expression (*lecA*) and the presence of T-DNA rearrangements. This result may be expected since most rearrangements detected occurred at the left hand border of T-DNA and therefore did not affect the expression of the lectin gene.

These data suggest that levels of transgene expression are predominantly influenced by the location of the integration site within the genome, although a possible contributory role for any one, or indeed all, of the above factors cannot be conclusively discounted. To analyse in more detail the effect position (i.e. the phenomenon of 'position effect') a number of T1 families from T0 parental lines hemizygous for a single non-rearranged copy of the *CaMV-lecA* transgene were selected (lines 28, 26, 20, 18, 12, and 4).

3.7 Variability in pea lectin levels between tobacco plants of selected *CaMV-lecA* T1 families.

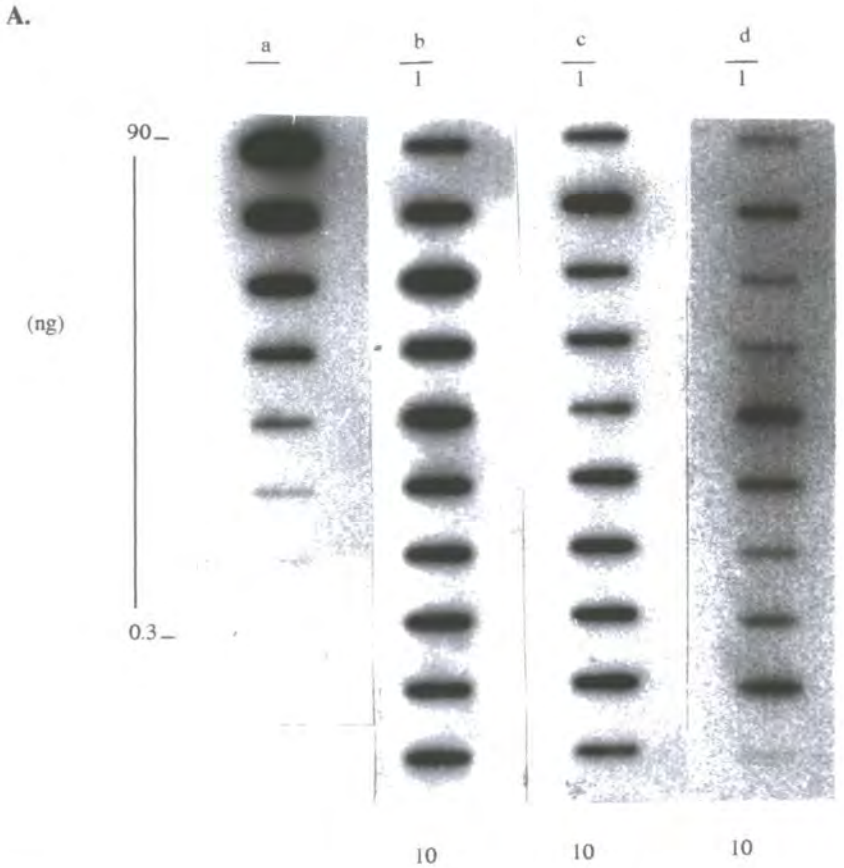
In this study leaf discs taken from 10 kanamycin resistant progeny tobacco plants for selected *CaMV-lecA* lines were analysed separately for the accumulation of pea lectin protein. Pea lectin levels within of each T1 family were estimated by slot blot radioimmunoassay (figure 3.42 (A), 3.43 (A)) using lectin standards (figure 3.42 (A) (blot a), 3.43 (A) (blot a)). The lectin levels in the plants would reflect both homozygous and hemizygous individuals. It would therefore be expected for the plants

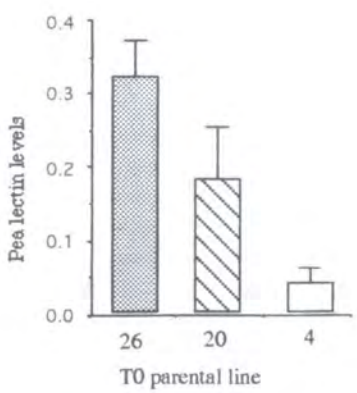
Figure 3.42:

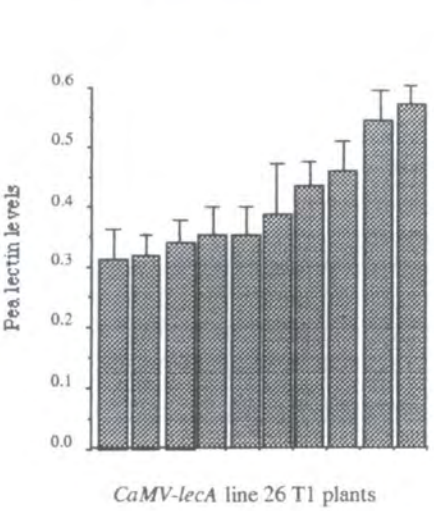
(A) Autoradiograph of a western slot blot for total soluble protein extracted from 10 individual members(kanamycin resistant) of each selected *CaMV-lecA* T1 family. Blot a: standard two-fold dilution series of purified pea lectin from 90ng-0.3ng per slot. Blot b: T1 plants of the *CaMV-lecA* T1 26 family. Blot c: T1 plants of the *CaMV-lecA* T1 20 family. Blot d: T1 plants of the *CaMV-lecA* T1 4 family.

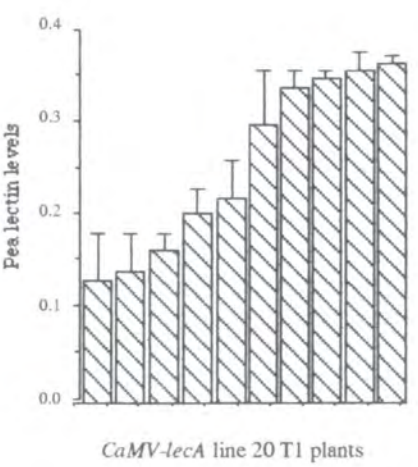
(B) Graphical representation of pea lectin protein levels in selected T1 plants for CaMV-*lecA* T1 families 26, 20, and 4. Error bars represent the spread of three repeat experiments.

(C) Graphical representation of pea lectin protein levels in *CaMV-lecA* T0 parental lines 26, 20, and 4.


A.
a
b
c
d
1
1
1
90
(ng)
0.3
10
10
10


B.
Pea lectin levels
0.4
0.3
0.2
0.1
0.0
26
20
4
T0 parental line


Pea lectin levels
0.6
0.5
0.4
0.3
0.2
0.1
0.0
CaMV-lecA line 26 T1 plants


Pea lectin levels
0.4
0.3
0.2
0.1
0.0
CaMV-lecA line 20 T1 plants

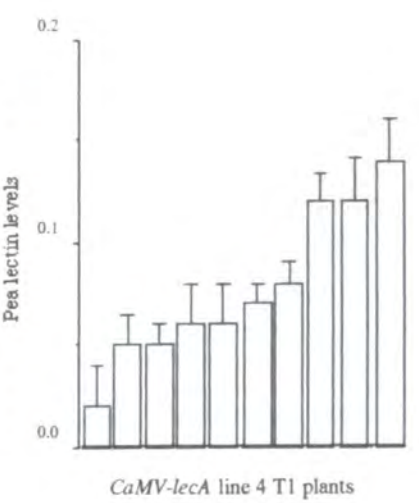
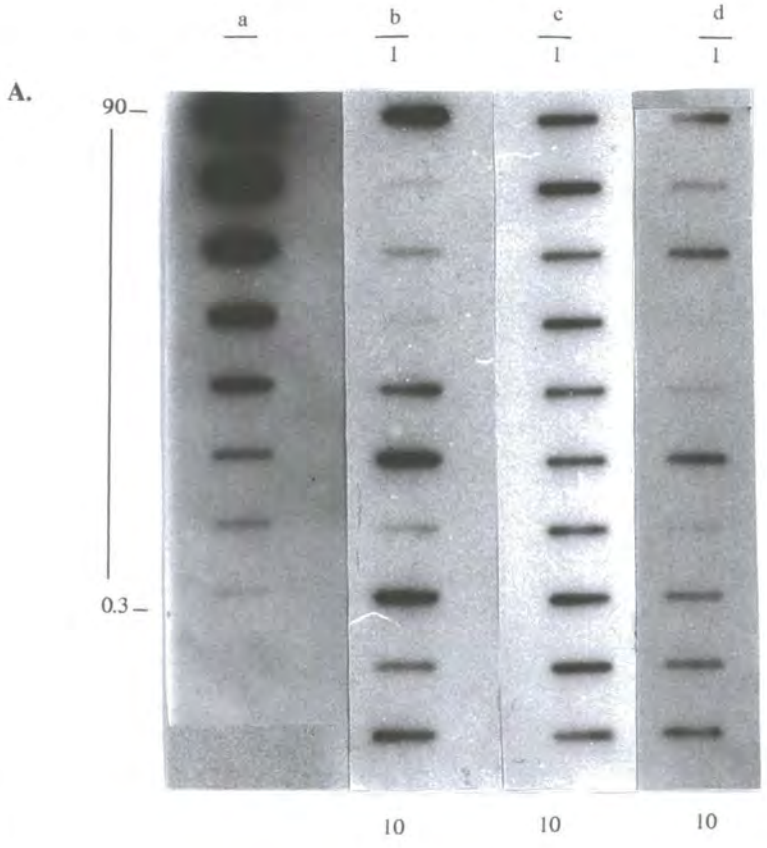

Pea lectin levels
0.2
0.1
0.0
CaMV-lecA line 4 T1 plants

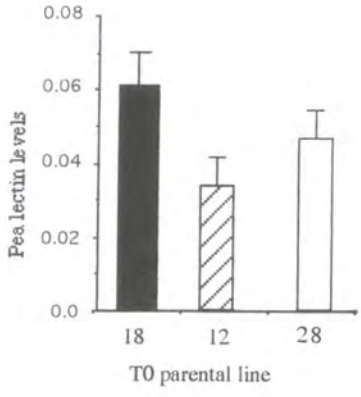
Figure 3.43:

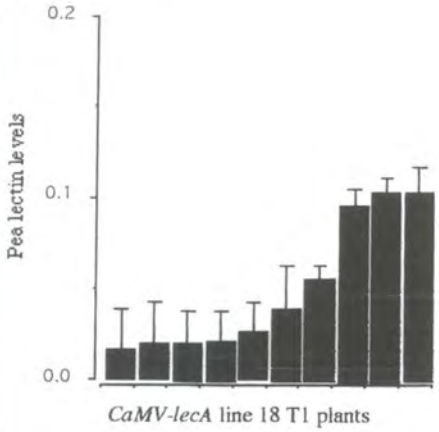
(A) Autoradiograph of a western slot blot for total soluble protein extracted from 10 individual members (kanamycin resistant) of each selected *CaMV-lecA* T1 family. Blot a: standard two-fold dilution series of purified pea lectin from 90ng-0.3ng per slot. Blot b: T1 plants of the *CaMV-lecA* T1 18 family. Blot c: T1 plants of the *CaMV-lecA* T1 12 family. Blot d: T1 plants of the *CaMV-lecA* T1 28 family.

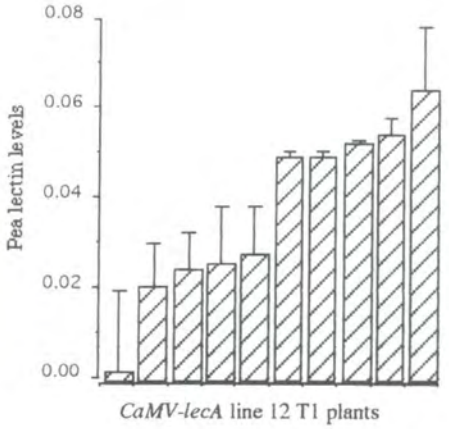
(B) Graphical representation of pea lectin protein levels in selected T1 plants for *CaMV-lecA* T1 families 18, 12, and 28. Error bars represent the spread of three repeat experiments.

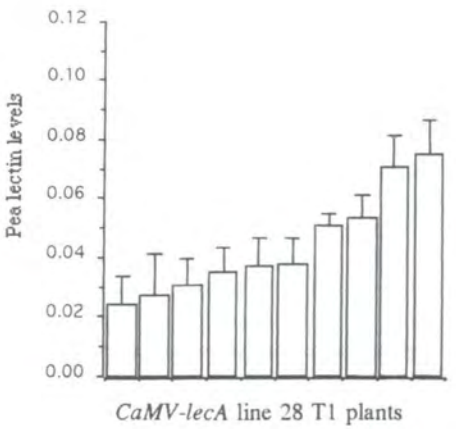
(C) Graphical representation of pea lectin protein levels in *CaMV-lecA* T0 parental lines 18, 12, and 28.


A.
a
b
c
d
1
1
1
90
0.3
10
10
10


0.08
0.06
0.04
0.02
0.0
Pea lectin levels
18
12
28
T0 parental line


0.2
0.1
0.0
Pea lectin levels
CaMV-lecA line 18 T1 plants


0.08
0.06
0.04
0.02
0.00
Pea lectin levels
CaMV-lecA line 12 T1 plants


0.12
0.10
0.08
0.06
0.04
0.02
0.00
Pea lectin levels
CaMV-lecA line 28 T1 plants

that levels should either equal those of the hemizygous parental line or be approximately two-fold greater for homozygous individuals. The results were as expected for most T1 plants, with lectin levels similar to the parental line or approximately double repereseting both hemizygous and homozygous plants. It was therefore concluded that the variability between T1 plants of similar genetic background was low with the majority of T1 plants exhibiting lectin levels similar to the parental line representing hemizygous individuals and a minority of plants within T1 populations displaying lectin levels approximately double representing homozygous lines. However the inter-transformant variability was estimated to be much greater with comparisons of expression levels between families indicating a similar result to that obtained upon comparison of expression levels between parental lines. Thus it may be concluded that it is the location of integration of the T-DNA within the genome that is the major influencing factor determining levels of expression of the transgene, since within each of the six lines analysed the genetic background into which the T-DNA was transformed is identical (SR1), the nature (*CaMV-lecA*) and number of integrated transgenes is the same, and hence so the number of insertion loci is the same. The number of homozygous and hemizygous individuals analysed is also the same for each population. The only apparent difference between the six populations is the site into which the T-DNA has integrated.

A series of investigations was then undertaken in order to determine the nature of those position-specific factors underlying the observed variation in transgene expression levels.

3.8 An investigation into chromatin structure of the integrated T-DNA for *CaMV-lecA* T1 plants.

The role of chromatin organisation is known to regulate eukaryotic gene expression at several levels (Parajape *et al.*, 1994) and hence chromatin structure at and proximal to the site of T-DNA integration may play a role in modulation of transgene expression.

A more open configuration of regulatory sequences seems to be a cause or effect of gene activation. Such open configurations occur preferentially in the 5' upstream region of a gene and can then be associated with regulatory proteins and other transcription factors. Generally chromatin of active genes is about 10 times more sensitive towards the endonuclease DNaseI than is bulk chromatin. This general sensitivity is again 10 fold higher for distinct stretches of a gene (DNaseI hypersensitive sites). These areas, though referred to as sites may extend over a length of 50-200bp. In plant genomes, increased DNaseI sensitivity of transcribed genes (Spiker *et al.*, 1983, Murray and kennard 1984, Coates *et al.*, 1987, Sawyer *et al.*, 1987) as well as the existence of hypersensitive sites have been established in only a limited number of genes (Steinmuller and Apel 1986, Gorz *et al.*, 1990, Brignon *et al.*, 1993). One reason for this lack of information on the architecture of plant genes results from the difficulty in isolating suitable amounts of purified intact nuclei from tissues with specific patterns of gene expression. The plants chosen for use in this analysis were hemizygous individuals from *CaMV-lecA* lines 26 and 4.

3.8.1 Nuclei isolation and assessment of integrity.

Tobacco leaf nuclei (figure 3.44 (A)) were isolated from *CaMV-lecA* T1 plants and subsequently purified on percoll gradients to remove cell debris and contaminating organelles (figure 3.44 (B)). The morphology of the nuclei was also checked using the fluorescent DNA specific dye, DAPI (figure 3.44 (C)). The nuclei were observed to be spheres of 7-10µm in diameter. Nuclei were incubated with various concentrations of micrococcal nuclease and the DNA isolated as described previously (section 2.9.19). The purified DNA was subjected to electrophoresis through 1%(w/v) agarose gel (figure 3.45 (A)). Included on the gel were two sets of size markers, lambda *HindIII* and pBR322 *BstEII* to provide a size scale from 0.4Kb to 23.1KB. For all isolated nuclei a distinct nucleosome ladder could be seen on the gel. The sizes of the trimer, tetramer and pentamer repeats were determined by reference to the standards (figure 3.45 (B)). The nucleosome repeat length was calculated by subtracting the size of the

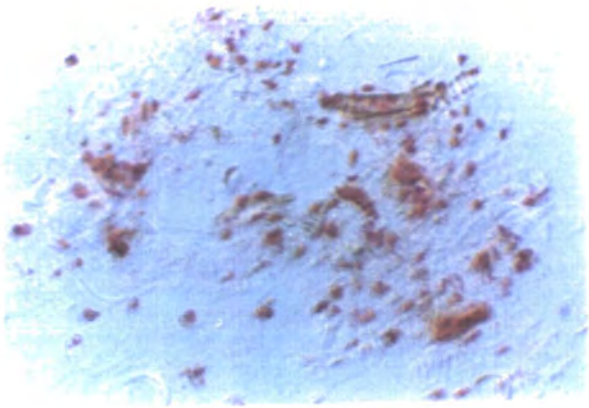
Figure 3.44:

(A) An aliquot of crude tobacco leaf nuclei, before purification on percoll gradients, stained with acetocarmine.

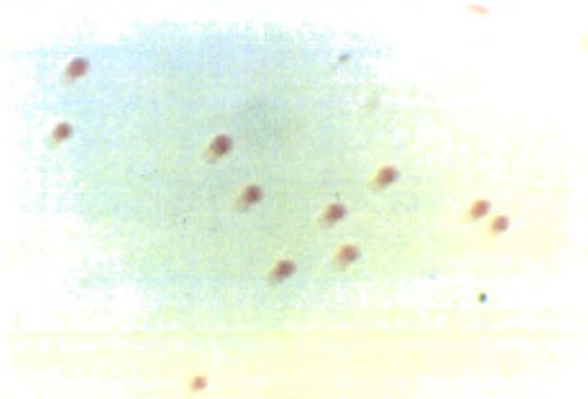
(B) An aliquot of tobacco leaf nuclei after purification on percoll gradients and stained with acetocarmine.

(C) Morphology of tobacco nuclei determined using the fluorescent stain DAPI.

A.



B.



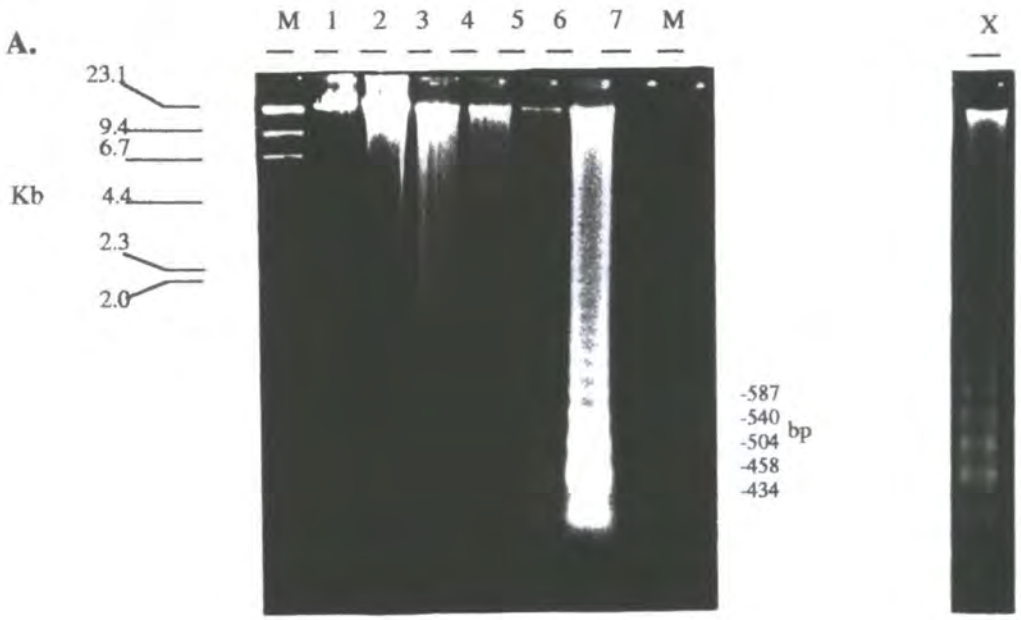
C.

Figure 3.45:

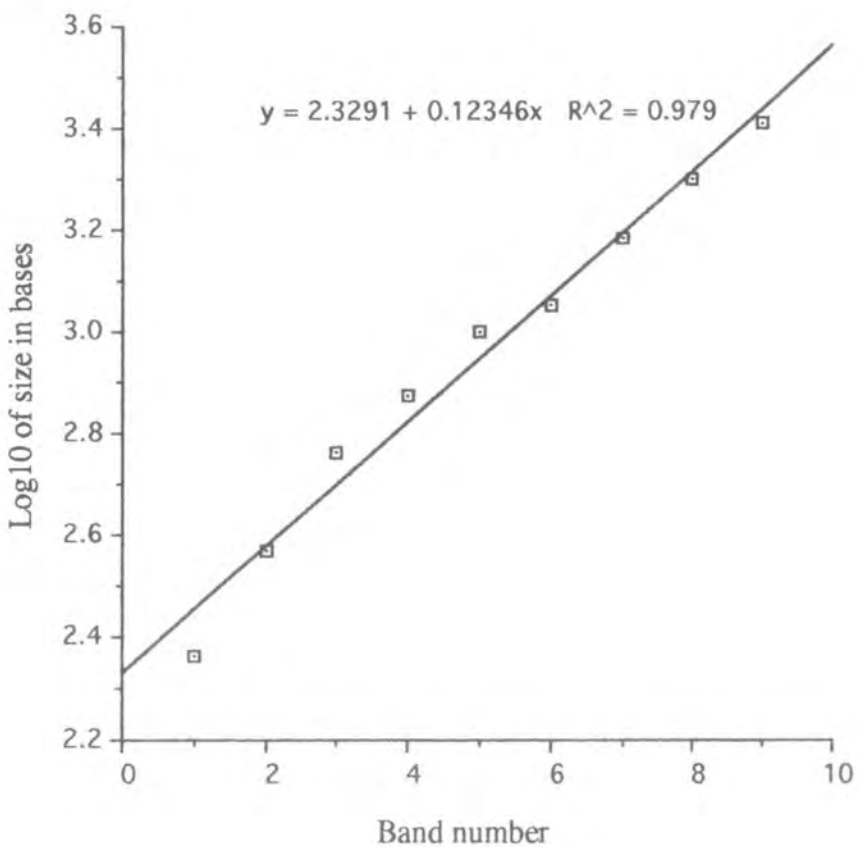
(A) Micrococcal nuclease digestion of tobacco leaf nuclei. 10µg of DNA in intact nuclei were incubated at 37^0C with micrococcal nuclease at 0.4 (Lane 2), 0.8 (Lane 3), 1.2 (Lane 4), 1.6 (Lane 5), and 2.0 (Lane 6) units/ml. The control (Lane 1) contained buffer instead of exogenous nuclease. Lane M to the left: Lambda *HindIII* size markers. Lane M to the right: pBR322 *HaeIII* size markers. Lane X: same as lane 7, only destained further to allow more accurate size mapping of nucleosome bands.

(B) Nucleosome band number with the $\log_{10}$ of size in base pairs.

A.


M 1 2 3 4 5 6 7 M
X
23.1
9.4
6.7
Kb
4.4
2.3
2.0
-587
-540
-504 bp
-458
-434

B.


y = 2.3291 + 0.12346x R^2 = 0.979
Log10 of size in bases
Band number
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
0
2
4
6
8
10

multimer from the size of the next highest multimer and averaging the results for each T1 individual investigated. In all cases the repeat length was 192±4bp. This estimate is consistent with other data published on plant nucleosome repeat lengths (McGhee and Engel, 1975). The results of the micrococcal nuclease digestion of tobacco nuclei showed that the nucleosome structure of tobacco chromatin was retained during the isolation procedure.

3.8.2 Southern blot analysis of the T-DNA from DNaseI-treated nuclei for *CaMV-lecA* T1 plants.

In previous work with chromatin from animal cells, it was demonstrated that the detection of DNaseI sensitive regions of actively transcribed genes was optimal when the overall extent of digestion reached 10 to 15% of the total DNA (Weintraub and Groudine, 1974). Nuclei isolated from upper leaves of *CaMV-lecA* lines 26 and 4 individuals were subjected to digestion with DNaseI, aliquots removed and reactions terminated after incubation times of between 0-45 minutes at room temperature. Optimal digestion (10-15%) was achieved over an incubation period of between 4 and 32 minutes as determined by agarose gel electrophoresis. These fractions were then used in subsequent analysis. After purification, the DNA was digested to completion with *EcoRI* analysed by agarose gel electrophoresis (figures 3.46 (A) and 3.47 (A)) and, following Southern blottting, hybridised succesively to probes specific for (i) the *lecA* coding region and (ii) the pea ribosomal clone pHA1 (obtained from Dr. R. Croy). The genomes of higher eukaryotes contain nuclear encoded ribosomal genes (rDNA) in multiple repeats. The repeats consist of the 18S, 5.8S, and 25S rDNA structural regions which are separated by intergenic spacers (IGS). Large tandem clusters of rDNA repeats are located at one or a few chromosomal loci, cytologically located in the heterochromatin of nucleolar organisation regions (NOR). The expression of only a limited number of the ribosomal RNA genes seems to be sufficient for the observed level of ribosome synthesis (Flavell *et al.*, 1985). Thus the pea ribosomal clone (pHA1) was used as a reference for untranscribed genes, since the bulk of the

ribosomal genes should remain unexpressed in leaf tissue. The resulting patterns of DNA fragments hybridising to both probes for *CaMV-lecA* T1 individuals (lines 26 and 4) are displayed in figures 3.46 and 3.47.

In *CaMV-lecA* line 26 the *lecA* specific probe hybridised to a fragment of >14.0Kb which contained some of the plant flanking DNA; suggesting that the next *EcoRI* site is located well into the flanking plant DNA (figure 3.46 (B)). *CaMV-lecA* line 4 on the other hand produced a strong hybridising junction fragment of <14.0Kb (figure 3.47 (B)). These differences in junction fragment size reflect differences in location of integration sites. On comparison of hybridisation patterns both individuals display a strong intense hybridising band throughout the DNaseI incubation although additional lower molecular weight bands, indicative of DNaseI hypersensitive sites, were detected in *CaMV-lecA* line 4. Following autoradiography the blots were stripped and reprobed with a pea ribosomal gene clone which was demonstrated to cross hybridise with DNA extracted from *Nicotiana tabacum* cv. SR1, Lotus, *Brassica napus*, and with the endogenous ribosomal genes in *Pisum sativum* L.(figure 3.47 (D); lanes 1, 2, 3, and 4). On hybridisation to DNA for *CaMV-lecA* line 4 two bands were detected, one at 14Kb and one band of 8.5Kb. A similar pattern was observed with *CaMV-lecA* line 26. However the lower molecular weight band of 8.5Kb was not detected after until 32 minutes of digestion and suggests that the nuclei for *CaMV-lecA* line 26 were more intact than those isolated from line 4.

Although these results appear to indicate that, in both cases, the T-DNAs are located within DNaseI-resistant regions of the genome, difficulties exist in drawing absolute conclusions from data obtained using this approach. Thus it was decided to take a more indirect approach in determining chromatin structure at the site of integration, using analysis of methylation and acetylation status of the T-DNA in assessing the likliehood that a particular T-DNA will adopt a heterochromatin structure.

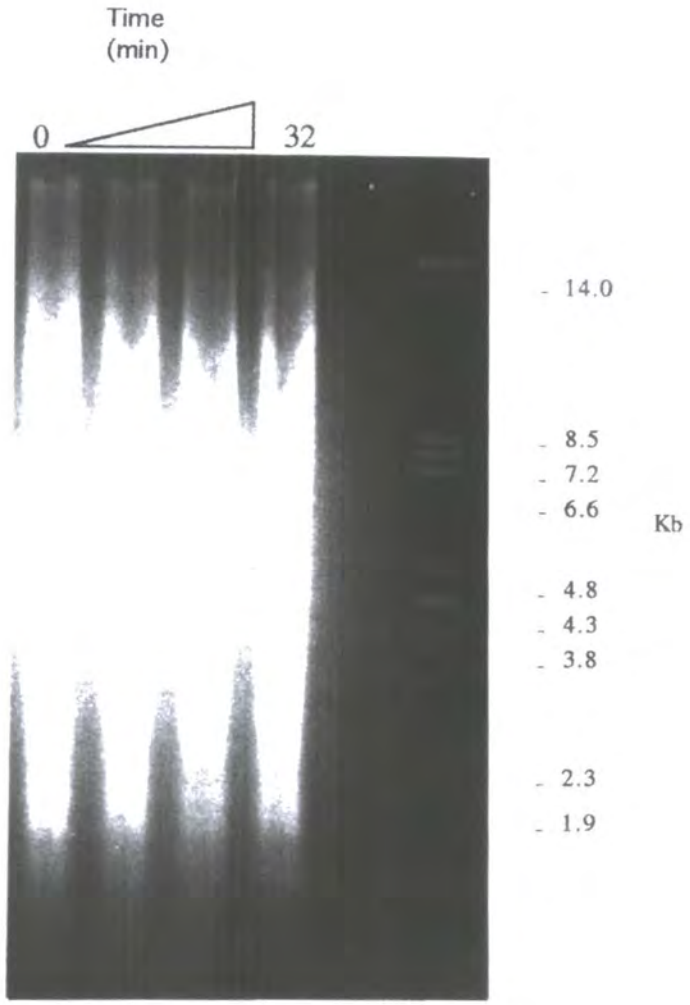
Figure 3.46:

(A) Analysis of DNA (10µg per lane) isolated from nuclei for *CaMV-lecA* line 26 T1 individual after digestion with DNaseI (0.2 units/ml) for various lengths of time and subsequent restriction with *EcoRI.*

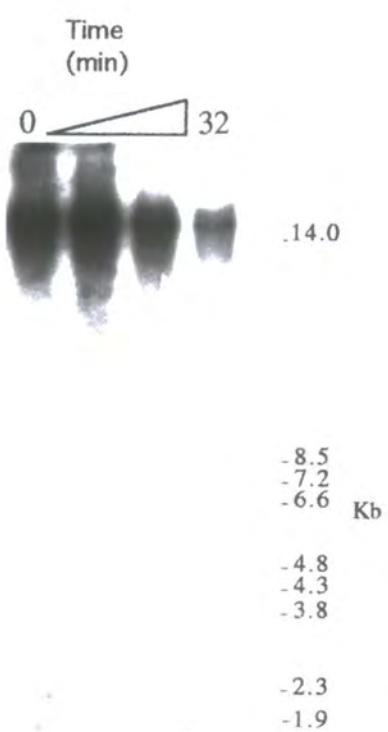
(B) Autoradiograph of a blot of the gel in figure 3.46(A) hybridised to a ^{32}P-labelled *Sph I/EcoRI* fragment containing the *lecA* coding region.

(C) Autoradiograph of a blot of the gel in figure 3.46(A) hybridised to a ^{32}P-labelled pea ribosomal clone (pHA1).

A.


Time
(min)
0
32
14.0
8.5
7.2
6.6
4.8
4.3
3.8
2.3
1.9
Kb

B.


Time
(min)
0
32
14.0
8.5
7.2
6.6
4.8
4.3
3.8
2.3
1.9
Kb

C.

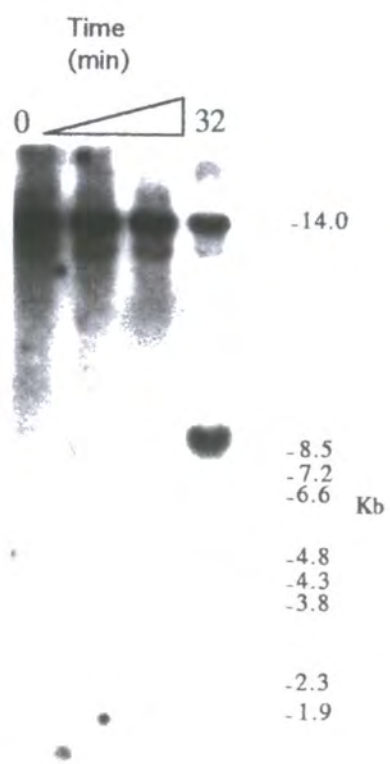

Time
(min)
0
32
14.0
8.5
7.2
6.6
4.8
4.3
3.8
2.3
1.9
Kb

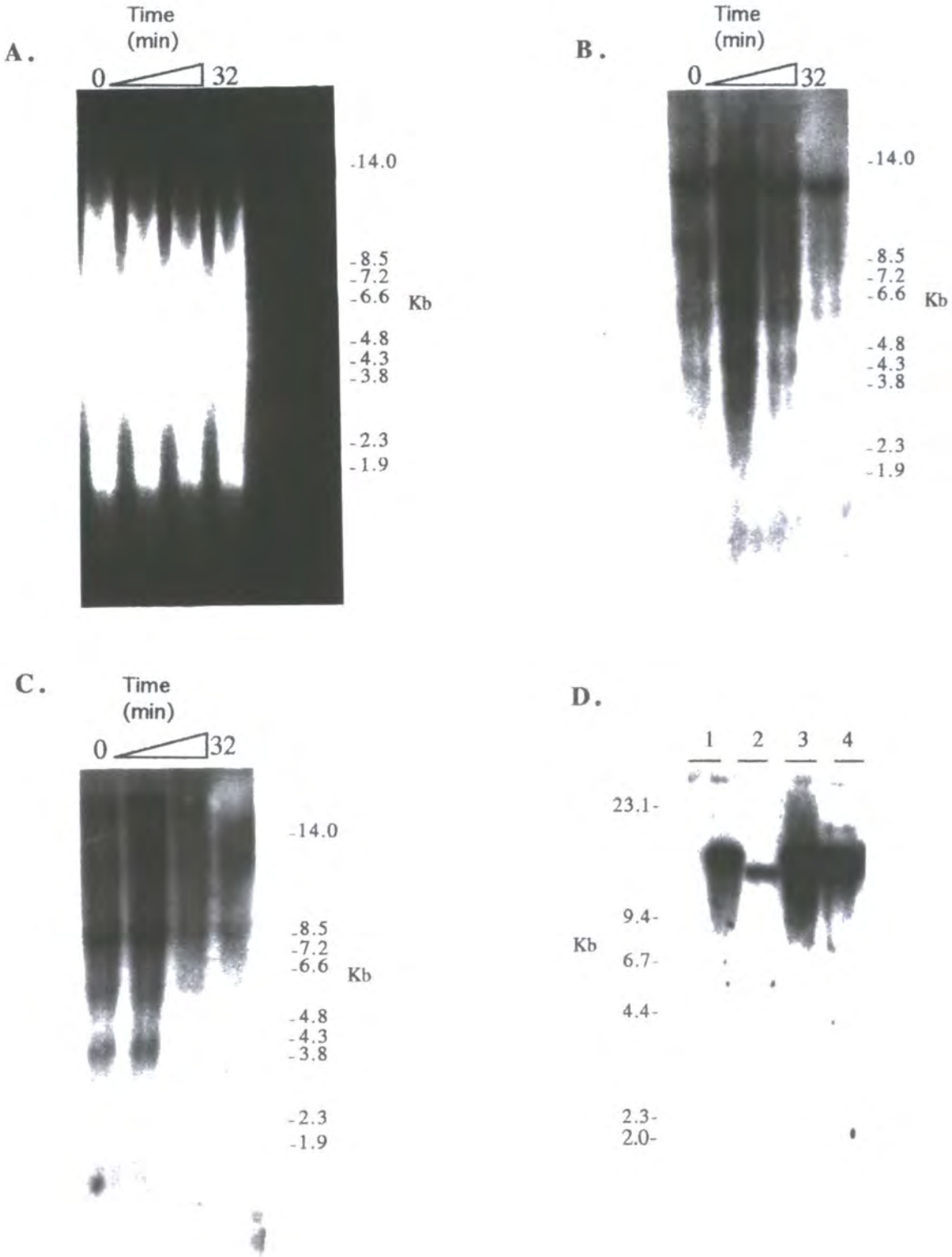
Figure 3.47:

(A) Analysis of DNA(10µg per lane) isolated from nuclei for *CaMV-lecA* line 4 T1 individual after digestion with DNaseI(0.2 units/ml) for various lengths of time and subsequent restriction with *EcoRI.*

(B) Autoradiograph of a blot of the gel in figure 3.47(A) hybridised to a ^{32}P-labelled *Sph I/EcoRI* fragment containing the *lecA* coding region.

(C) Autoradiograph of a blot of the gel in figure 3.47(A) hybridised to a ^{32}P-labelled pea ribosomal clone (pHA1).

(D) Southern blot of genomic DNA extracted from tobacco, lotus, rape and pea, restricted with *HindIII* and hybridised to a ^{32}P-labelled pea ribosomal gene. Lane 1: SR1 genomic DNA. Lane 2: lotus genomic DNA. Lane 3: rape genomic DNA. Lane 4: pea genomic DNA.


A.
Time
(min)
0
32
14.0
8.5
7.2
6.6
Kb
4.8
4.3
3.8
2.3
1.9
B.
Time
(min)
0
32
14.0
8.5
7.2
6.6
Kb
4.8
4.3
3.8
2.3
1.9
C.
Time
(min)
0
32
14.0
8.5
7.2
6.6
Kb
4.8
4.3
3.8
2.3
1.9
D.
1
2
3
4
23.1
9.4
Kb
6.7
4.4
2.3
2.0

3.8.3 A study of the methylation status of integrated T-DNA for selected *CaMV-lecA* T1 plants.

The DNA of higher eukaryotes is methylated at carbon 5 of some cytosine residues. In vertebrates, 3-8% of cytosine residues are methylated (Shapiro, 1975), whereas in plants as many as 30% of the total cytosines are methylated (Adams and Burdon, 1985). The higher content of methylated cytosine in some plants could be partly attributed to the large genome which contains many repetitive DNA sequences. However, in the vertebrate genome 5-methyl-cytosine (5mC) is largely confined to CG dinucleotides, whereas in higher plants both CG dinucleotides and CNG trinucleotides are methylated (Gruenbaum *et al.*, 1981). Methylation of DNA in plants, as in vertebrates, is implicated in the regulation of gene expression (Antequera and Bird, 1988; Finnegan *et al.*, 1993); an effect that may be direct through DNA-transcription factor interaction, or indirect via an alteration in chromatin structure (Adams, 1990; Lewis and Bird, 1991; Razin and cedar, 1991) with increased methylation being frequently correlated with heterochromatinisation. Indeed it has recently been shown that hypermethylation can directly stimulate heterochromatinistion in plants (Lund *et al.*, 1995b). Methylation of bases in the DNA at or near gene promoters has been shown to be associated with the modulation of gene expression. An investigation was undertaken to examine whether the variability in pea lectin levels between *CaMV-lecA* T1 plants could be explained by different extents of DNA methylation in relevant portions of the transgene.

T1 plants from *CaMV-lecA* lines 26, 20 and 4 with levels of lectin accumulation comparable to those of their parental line (and hence hemizygous for the transgene) were selected for methylation analysis. The methylation status of the integrated transgene was analysed by restriction enzyme treatment having first cut the DNA with *EcoRI/HindIII* or *EcoRI* (figure 3.48). Enzymes used in the methylation analysis were *HpaII-MspI*, *HhaI*, *HaeIII*, *MboI-Sau3A*, and *EcoRII-BstNI*. Their mode of action on the transgene is schematically shown in figure 3.49. The isoschizomers *HpaII-MspI*,

Figure 3.48:

Strategy used for the methylation analysis of T-DNA in transgenic plants transformed with the *CaMV-lecA* construct after digestion with *EcoRI/HindIII* and methylation sensitive enzymes *HpaII* or *MspI*. **A**: represents the predicted size (Kb) of T-DNA with homology to a ^{32}P-labelled 1.2Kb *EcoR I/SphI* fragment containing the *lecA* coding region when sites 1 and 2 are not methylated (*Agrobacterium* control). **B**: represents the predicted sizes (Kb) with homology to the *lecA* probe when both sites 1 and 2 are methylated. **C**: represents the predicted sizes (Kb) with homology to the *lecA* probe when sites 1, 2, and 3 are methylated.

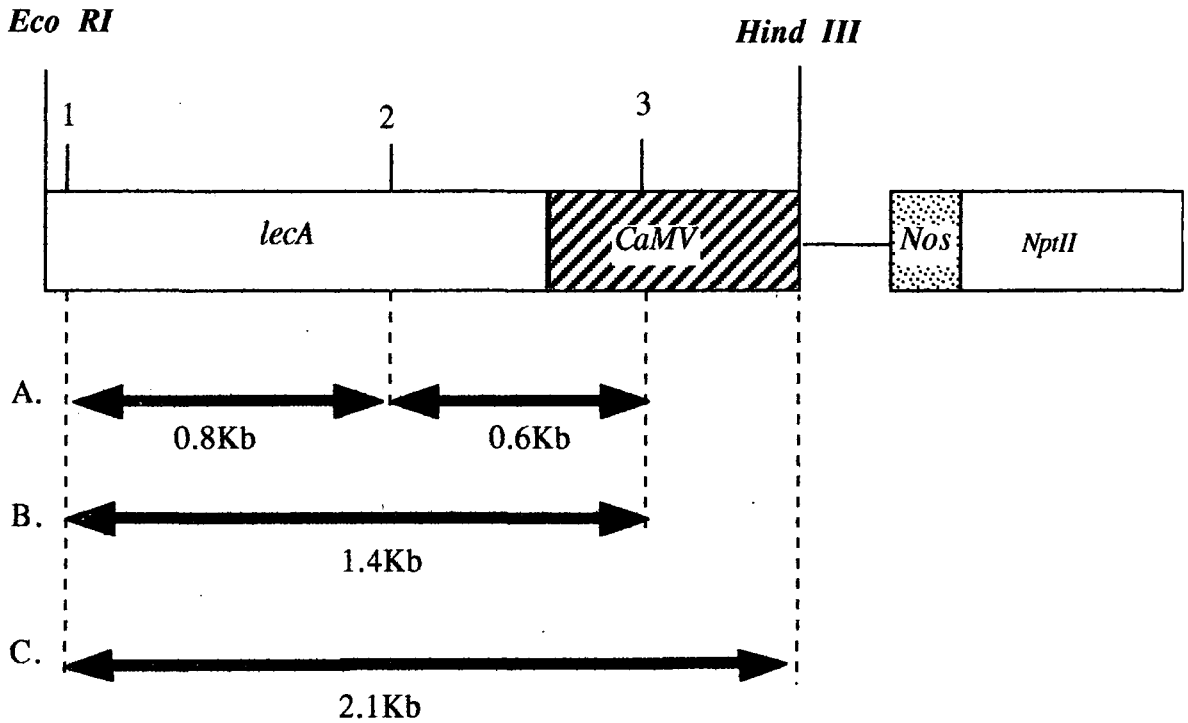

Eco RI
Hind III
1
2
3
lecA
CaMV
Nos
NptII
A.
0.8Kb
0.6Kb
B.
1.4Kb
C.
2.1Kb

Figure 3.50:

(A) Agarose gel analysis of *CaMV-lecA* T1 plants 26, 20, and 4. Lanes 1-3: *EcoRI/HindIII* digested genomic DNA (10µg) for *CaMV-lecA* T1 plants 26, 20, and 4 respectively. Lane 4: *EcoRI/HindIII* digested SR1 genomic DNA (10µg). Lanes 5-7: *EcoRI/HindIII* digested SR1 genomic DNA (10µg) spiked with *EcoRI* linearised pDUB80 representing *lecA* gene copy reconstructions one, three and five respectively. Lane M: Lambda *HindIII* digest.

(B) Autoradiograph of the gel in figure (3.50A) after Southern transfer and hybridisation to the *lecA* coding sequence. (see text for further details).

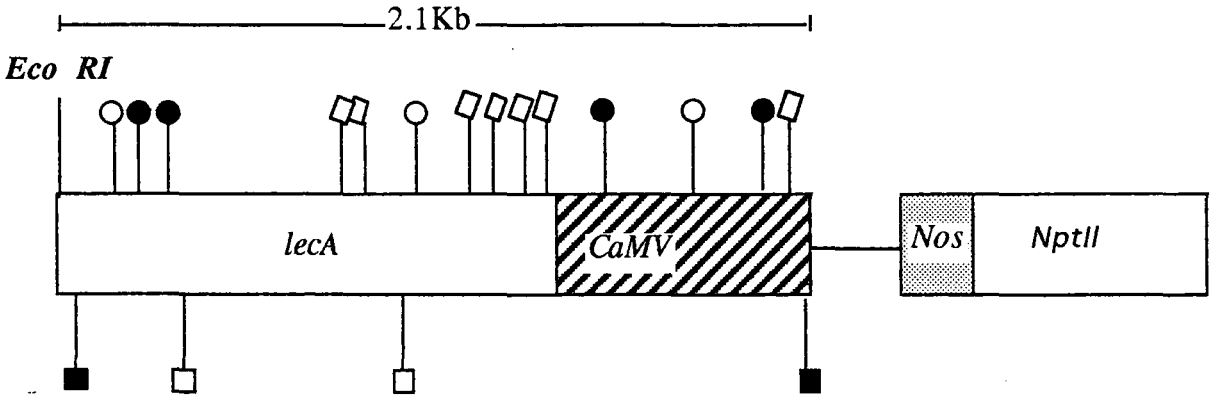
MboI-Sau3A, and *EcoRII/BstNI* differ in their sensitivity to methylation in their respective recognotion sequences and are useful in determining the methylation at CG or CNG sites. Restriction endonucleases *HpaII* and *MspI* were used to monitor DNA methylation at CCGG sites. While *HpaII* only cleaves target sites containing unmethylated C residues, *MspI* is blocked by CCGG residues that have the first residue methylated. Hence if, in a DNA digestion experiment, no cleavage is seen with either enzyme, the CNG sequence is methylated. If only CG is methylated the recognition site is cleaved by *MspI*, but not by *HpaII*. The sites indicated for *MboI-Sau3A* form putative methylatable sites either CG or CNG, therefore resistance to cleavage by *Sau3A* (*Sau3A* does not cleave the methylated GAT^{m}C site) would indicate methylation at a CG or CNG recognition site. Isoschizomers *EcoRII* and *BstNI* both recognise the sequence CC$_{A/T}$GG. *BstNI* cleaves the DNA when methylated at both C residues whereas *EcoRII* is blocked when the internal C is methylated. Thus if no cleavage is observed with *EcoRII*, the CNG sequence is methylated. *HhaI* is sensitive to methylation of the CG dinucleotide in its recognition site CGCG. *HaeIII* however does not contain methylatable C residues at CG or CNG sites within its GGCC sequence and such a site is not generated by G or NG nucleotides adjacent to the *HaeIII* sites as deduced from the sequence of the transgene. Therefore if no cleavage is observed with *HaeIII* it may indicate methylation of cytosines in non-CG and non-CNG sequences.

3.8.3.1 Methylation status of the integrated *CaMV-lecA* gene in leaf tissue using *MspI-HpaII*.

The probe used in this analysis was a 1.2Kb *Sph I/EcoRI lecA* coding region from pDUB80 homologous to the pea lectin gene. The strategy for determining the methylation status of the *CaMV-lecA* gene using isoschizomers *HpaII-MspI* is shown in figure 3.48. The recognition sequence CCGG (site 3) is unique at position -315 within the 35S *CaMV* promoter. 3' to this site are two CCGG (1 and 2) sequences present within the pea lectin coding region. Non-methylation of the CCGG sites (1 and 2) located within the *lecA* gene results in 2 predicted fragments on hybridisation to the

probe of 0.6Kb and 0.8Kb whereas methylation of any site will generate fragments much larger in size.

DNA (10µg) restricted to completion with *EcoRI/HindIII* (figure 3.50) was digested with *MspI* or *HpaII* (figure 3.51 (A)). Southern blot analysis of *CaMV-lecA* (26 and 20) T1 individuals (figure 3.51 (B); lanes 3-6) showed non-methylation of CCGG sites 1 and 2 (see figure 3.48) within the *lecA* gene by having homology to the predicted fragments 0.8Kb and 0.6Kb on digestion with both *MspI* and *HpaII* as observed with the control of unmethylated *Agrobacterium* DNA (figure 3.51 (B); lane 9). In addition to the predicted non-methylated fragments a much larger fragment of 1.2Kb was observed for both *CaMV-lecA* (26 and 20) T1 individuals (denoted by an asterisk in figure 3.51 (B); lanes 3-6). This band can be interpreted if the *Hpa II-MspI* -315 CCGG site 3 within the *CaMV* promoter is methylated (see figure 3.48 (B)). T-DNA in the octopine type Ti-plasmid pTiB6806 and the nopaline type Ti-plasmid pTiT37 were reported by Gelvin *et al.*, (1983) to be non-methylated at the sequence 5'CCGG 3' in *Agrobacterium,* indicating that the observed methylation has arisen as a consequence of an *in planta* event. In contrast *CaMV-lec A* line 4 displayed only one band of homology to the non-methylated predicted size of 0.8Kb (figure 3.51(B); lanes 7 and 8) and two larger bands, one of 1.2Kb as observed with *CaMV-lecA* (26 and 20) T1 individuals and an additional band of 1.5Kb. The presence of 1.5Kb band can be allocated on the assumption that site 2 (see figure 3.48) within the *lecA* gene was methylated. It may thus be seen that the methylation profiles observed for the three lines analysed do not fit with any of the predicted profiles in that sites are found both methylated and unmethylated in the same leaf. These apparent discrepencies may perhaps be explained on the basis of the fact that the tissue type analysed i.e. leaf is comprised of a number of different cell types, each of which may exhibit a methylation profile specific to that cell type and therefore give rise to anomalous hybridising fragments under the conditions used here.

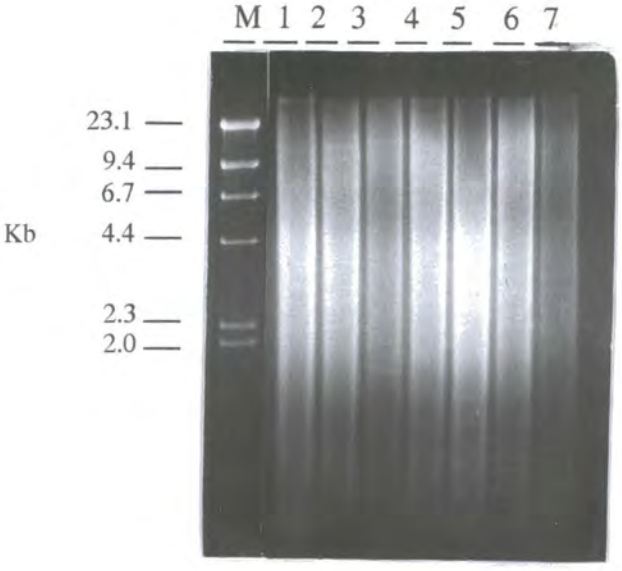


	Cut	Not Cut	Number of sites
Hpa II *Msp I*	CCGG C*CGG	C*CGG CCGG*	3
Hae III	GG*CC	GG*CC	4
Mbo I *Sau 3A*	GA*TC GATC	GATC*	7
Hha I	GCGC	G*CGC GCG*C	2
Bst NI *Eco RII*	C*C*C(A/T)GG CC(A/T)GG*	CC(A/T)GG	2

Figure 3.49:

Strategy used for the methylation analysis of T-DNA in transgenic tobacco plants transformed with the *CaMV-lecA* construct. Enzymes used for the methylation assay by Southern analysis are listed with regard to their methylation sensitivity. 5mdCTP is indicated as C^{*}.

A.



B.

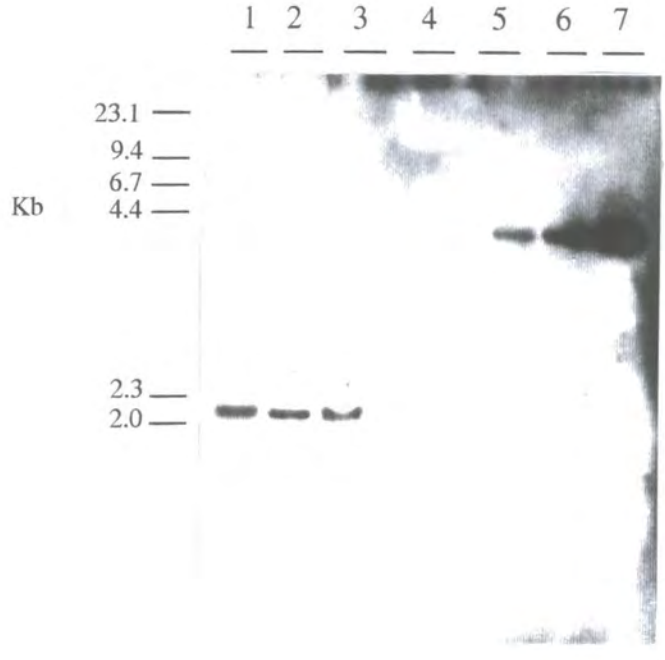
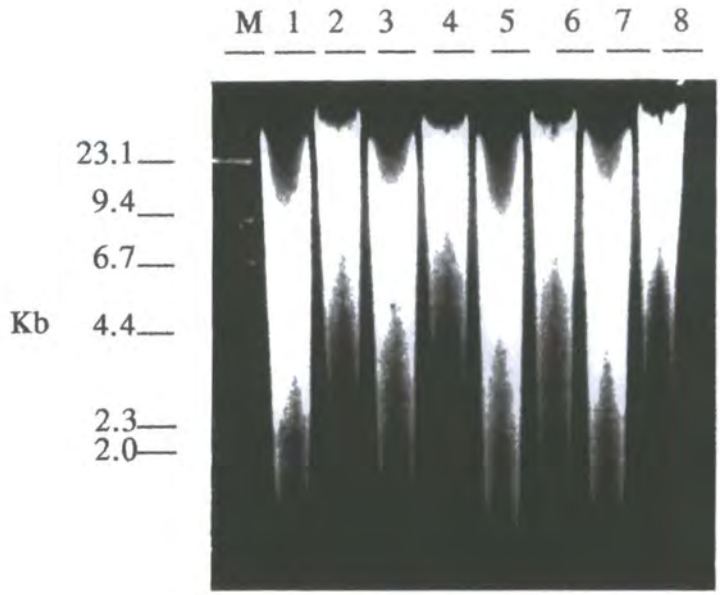


Figure 3.51:

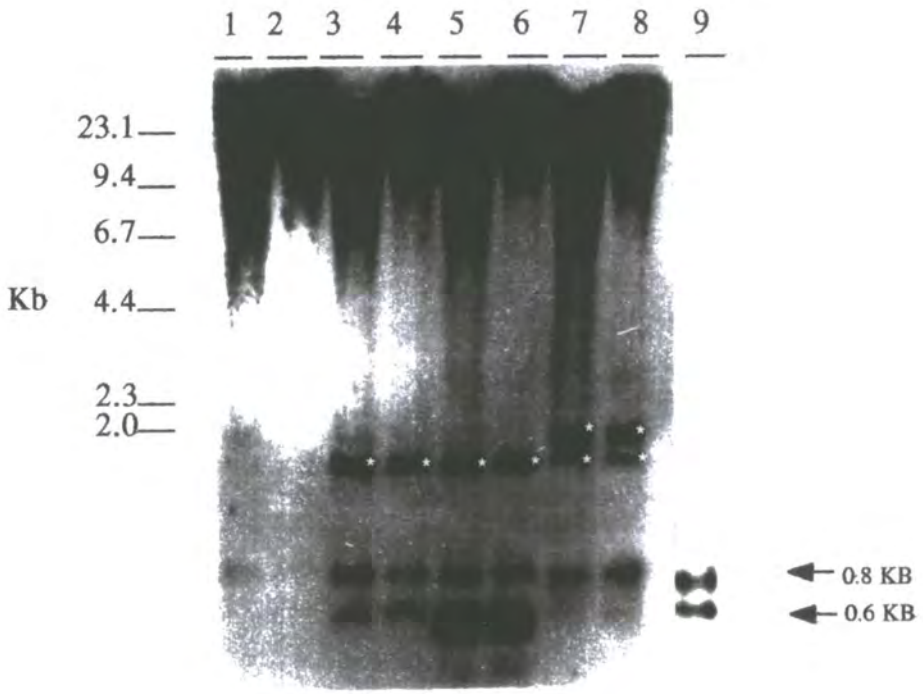
(A) Agarose gel analysis of genomic DNA extracted from *CaMV-lecA* (26, 20, & 4) T1 individuals after digestion with *EcoRI/HindIII* and methylation sensitive restriction endonucleases *HpaII* and *MspI*. Lanes 1 and 2: *SR1* genomic DNA restricted with *HpaII* and *MspI* Lanes 3 and 4: *CaMV-lecA* 26 T1 individual restricted with *HpaII* and*MspI* . Lanes 5 and 6: *CaMV-lecA* T1 20 individual restricted with *HpaII* and*MspI*. Lanes 7 and 8: *CaMV-lecA* T1 4 individual restricted with *HpaII* and *MspI* .

(B) Autoradiograph of the gel in figure 3.51(A) after Southern transfer and hybridisation to a ^{32}P-labelled 1.2Kb *Sph I/EcoRI* fragment containing the *lecA* coding region. Lane 9: *Agrobacterium* (pGV3850::130) DNA restricted with *MspI*

A.


M 1 2 3 4 5 6 7 8
23.1
9.4
6.7
Kb 4.4
2.3
2.0

B.


1 2 3 4 5 6 7 8 9
23.1
9.4
6.7
Kb 4.4
2.3
2.0
0.8 KB
0.6 KB

3.8.3.2 Methylation status of the integrated *CaMV-lecA* gene in leaf tissue using *MspI-HpaII, HaeIII, MboI-Sau 3a, HhaI,* and *BstNI-EcoRII.*

A more extensive analysis was conducted using restriction enzymes *HpaII- MspI, HaeIII, MboI-Sau3A, HhaI,* and *BstNI-EcoRII.* The number of sites together with the recognition sequences of the restriction enzymes are detailed in figure 3.49. DNA (90ug) from *CaMV-lecA* T1 individuals (26, 20, and 4) was digested first with *EcoRI* and 10µg removed for restriction with methylation sensitive/insensitive restriction endonucleases (figures 3.52 (A); figure 3.53 (A); and figure 3.54 (A)). Parallel digestions of *Agrobacterium* (pGV3850::130) DNA were performed for all enzymes to evaluate the extents of digestion and to determine the precise size of non-methylated T-DNA fragments with homology to the probe (figure 3.49). Southern blot analysis of *CaMV-lecA* line 26 after hybridisation to a ^{32}P-labelled *lecA* gene displayed resistance to methylation sensitive sites *HpaII, MspI, Sau3A, HhaI* and *EcoRII* (figure 3.52 (B); lanes 2, 3, 6, 7, and 9) when compared with the *Agrobacterium* control (figure 3.55; lanes 7, 8, 4, 3, and 1). Hybridisation patterns of *CaMV-lecA* line 26 after digestion with *HpaII-MspI* showed a predicted fragment of 0.8Kb and larger molecular weight fragments of 2.1Kb and >2.1Kb. After digestion with *MboI* and *Sau3A* resistance to *Sau3A* cleavage was observed (*Sau3A* does not cleave the methylated GAT^{m}C site). *HhaI* was expected to release a 0.4Kb fragment with homology to the probe, however much larger fragments were observed greater than 2.1Kb. The Southern profile of *CaMV-lecA* line 20 (figure 3.53 (B)) was similar to that of *CaMV-lecA* 26 T1 individual in that *HpaII-MspI* cleavage yielded identical fragments of 0.8Kb, 2.1Kb, and >2.1Kb (figure 3.53 (B); lanes 2 and 3). The profile of *HaeIII* digestion produced the predicted fragment of 1Kb identical to the *Agrobacterium* control (figure 3.55; lane 6). As observed with *CaMV-lecA* line 26 resistance to cleavage by *Sau3A* was also found (lane 6). An extra band of hybridisation was observed after an *EcoRII* digestion >2.1Kb (lane 9). The hybridisation pattern obtained with *CaMV-lecA* line 4 on the other hand did not resemble that of the other two T1 individuals and displayed greater

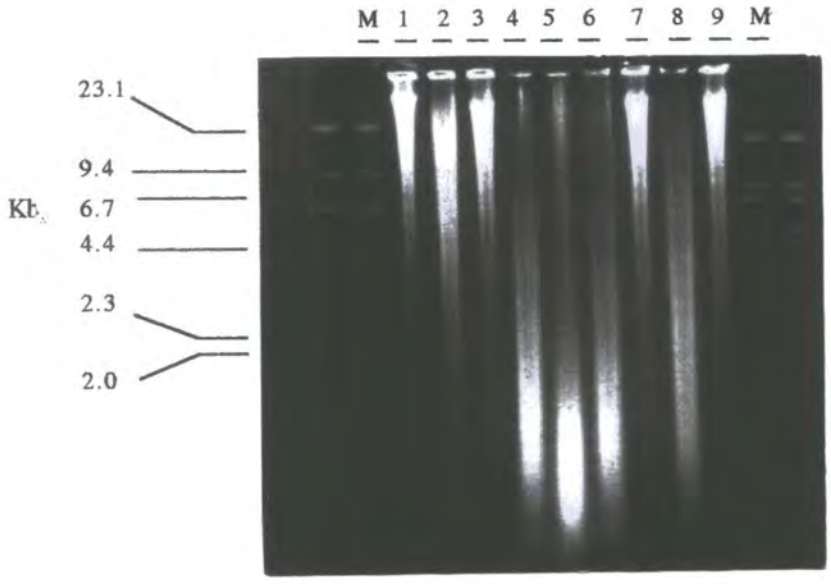
Figure 3.52:

(A) Agarose gel analysis of DNA extracted from *CaMV-lecA* 26 T1 individual restricted with the following restriction enzymes. Lane 1: *EcoRI*. Lane 2: *HpaII*. Lane 3: *MspI*. Lane 4: *HaeIII*. Lane 5: *MboI*. Lane 6: *Sau3A*. Lane 7: *HhaI*. Lane 8: *BstNI*. Lane 9: *EcoRII*. Lane M: Lambda *HindIII*.

(B) Southern blot analysis of DNA extracted from *CaMV-lecA* 26 T1 individual, restricted with *EcoRI* (Lane 1), followed by cleavage with *HpaII, MspI, HaeIII, MboI, Sau3A, HhaI, BstNI, EcoRII* (Lane 2-9), and hybridised to a ^{32}P-labelled 1.2Kb *SphI/EcoRI* fragment containing the *lecA* coding region.

Asterisks denote hybridising fragments showing altered size on comparison with similar hybridising fragments in figure 3.55. The observed size alterations have arisen as a consequence of resistance to endonuclease cleavage caused by methylation at specific sites within the transgene.

A.


M 1 2 3 4 5 6 7 8 9 M
23.1
9.4
Kb 6.7
4.4
2.3
2.0

B.

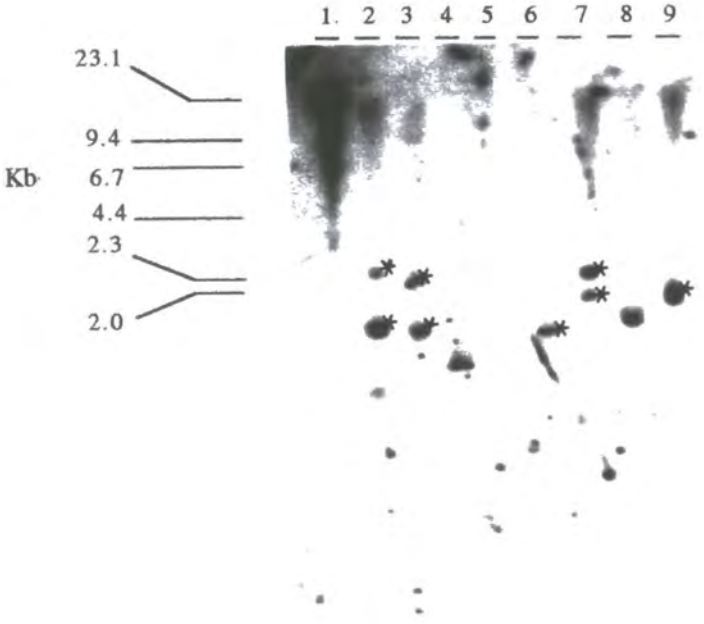
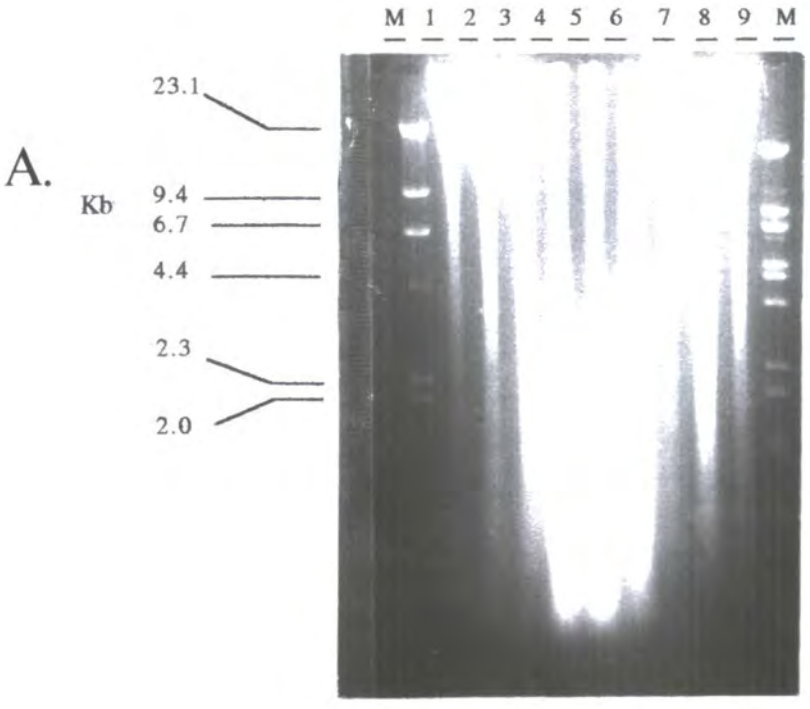

1. 2 3 4 5 6 7 8 9
23.1
9.4
Kb 6.7
4.4
2.3
2.0

Figure 3.53:

(A) Agarose gel analysis of DNA extracted from *CaMV-lecA* 20 T1 individual restricted with the following restriction enzymes. Lane 1: *EcoRI*. Lane 2: *HpaII*. Lane 3: *MspI*. Lane 4: *HaeIII*. Lane 5: *MboI*. Lane 6: *Sau3A*. Lane 7: *HhaI*. Lane 8: *BstNI*. Lane 9: *EcoRII*. Lane M: Lambda *HindIII*.

(B) Southern blot analysis of DNA extracted from *CaMV-lecA* 20 T1 individual restricted with *EcoRI* (Lane 1), followed by cleavage with *HpaII, MspI, HaeIII, MboI, Sau3A, HhaI, BstNI, EcoRII* (Lane 2-9), and hybridised to a ^{32}P-labelled 1.2Kb *SphI/EcoRI* fragment containing the *lecA* coding region.

Asterisks denote hybridising fragments showing altered size on comparison with similar hybridising fragments in figure 3.55. The observed size alterations have arisen as a consequence of resistance to endonuclease cleavage caused by methylation at specific sites within the transgene.


A.
Kb
23.1
9.4
6.7
4.4
2.3
2.0
M 1 2 3 4 5 6 7 8 9 M

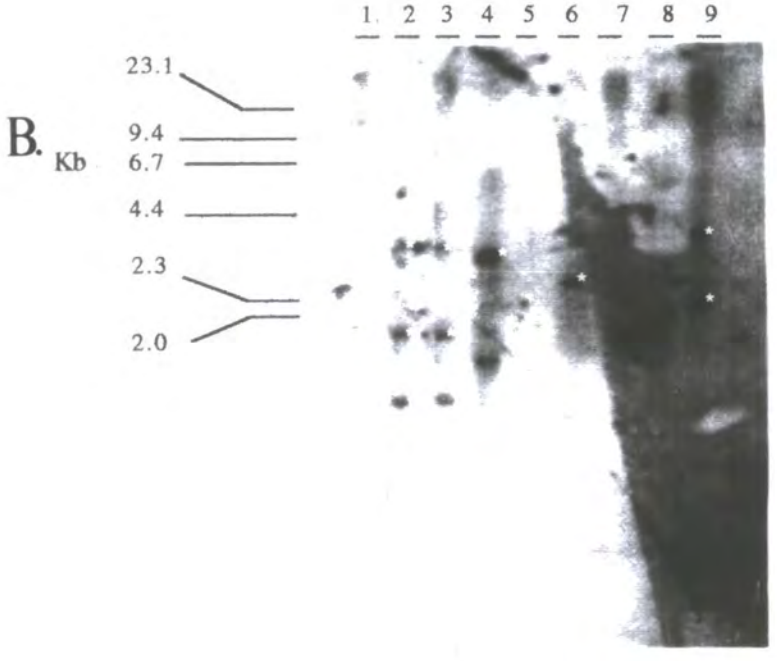

B.
Kb
23.1
9.4
6.7
4.4
2.3
2.0
1. 2 3 4 5 6 7 8 9

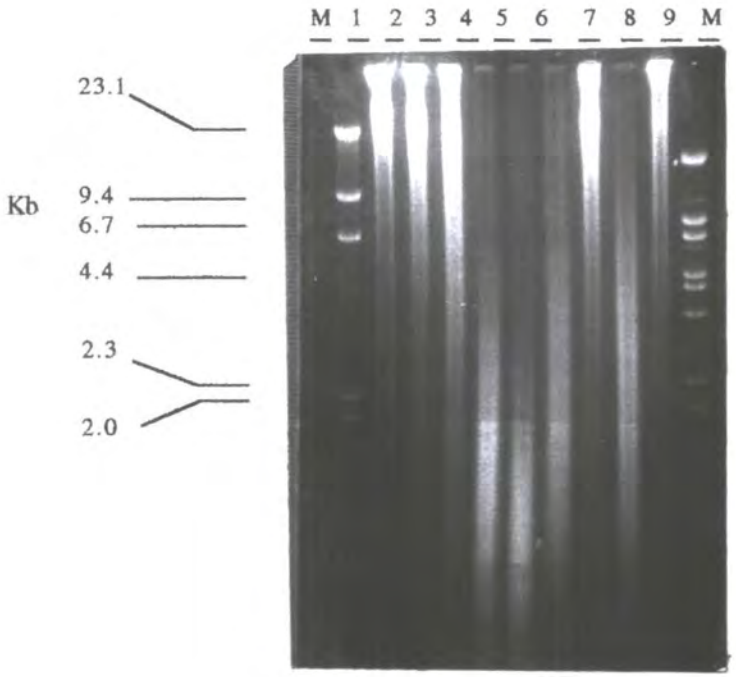
Figure 3.54:

(A) Agarose gel analysis of DNA extracted from *CaMV-lecA* 4 T1 individual restricted with the following restriction enzymes. Lane 1: *EcoRI*. Lane 2: *HpaII*. Lane 3: *MspI*. Lane 4: *HaeIII*. Lane 5: *MboI*. Lane 6: *Sau3A*. Lane 7: *HhaI*. Lane 8: *BstNI*. Lane 9: *EcoRII*. Lane M: Lambda *HindIII*.

(B) Southern blot analysis of DNA extracted from *CaMV-lecA* 4 T1 individual (plant), restricted with *EcoRI* (Lane 1), followed by cleavage with *HpaII, MspI, HaeIII, MboI, Sau3A, HhaI, BstNI, EcoRII* (Lane 2-9), and hybridised to a ^{32}P-labelled 1.2Kb *SphI/EcoRI* fragment containing the *lecA* coding region.

Asterisks denote hybridising fragments showing altered size on comparison with similar hybridising fragments in figure 3.55. The observed size alterations have arisen as a consequence of resistance to endonuclease cleavage caused by methylation at specific sites within the transgene.

A.


M 1 2 3 4 5 6 7 8 9 M
23.1
Kb
9.4
6.7
4.4
2.3
2.0

B.

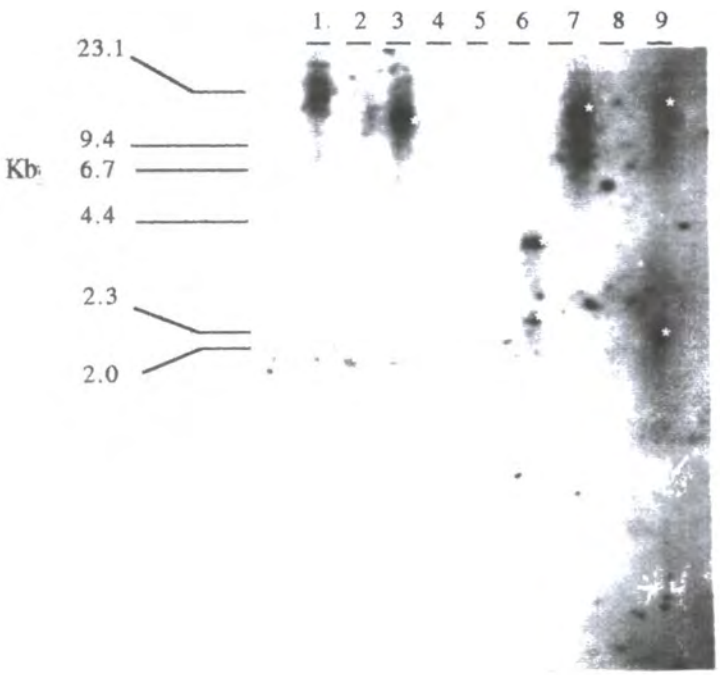

1. 2 3 4 5 6 7 8 9
23.1
Kb
9.4
6.7
4.4
2.3
2.0

Figure 3.55:

Southern blot of total DNA isolated from *Agrobacterium* (GV3101) harbouring the co-integrate pGV3850::pDUB130, hybridised to a digoxigenin labelled 1.2Kb *Sph I/EcoRI* fragment containing the *lecA* coding region. Lanes 1-8: pGV3850::pDUB130 DNA digested first with *EcoRI* then with *EcoRII, BstNI, HhaI, Sau3A, MboI, HaeIII, MspI, HpaII.* Detection of fragments homologous to the probe was achieved with colourmetric substrates nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indoyl phosphate (BCIP) after hybridisation with an alkaline phosphatase conjugated anti-digoxigenin antibody(1:5,000)

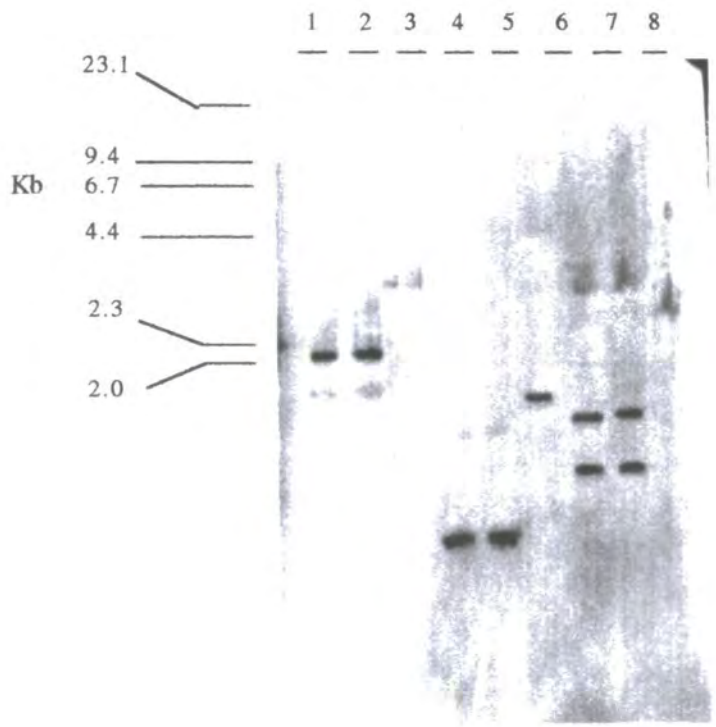

1 2 3 4 5 6 7 8
23.1
9.4
Kb 6.7
4.4
2.3
2.0

Table 3.24:

Summary of data obtained from methylation analysis of transgenic *CaMV-lecA* lines 26, 20 and 4. None, partial and complete refer to the level of methylation observed.

Key: +: positive for the specified charcteristic.

-: negative for the specified characteristc.

n.d: not determined.

Table 3.25:

Interpretation of data obtained from methylation analysis of transgenic *CaMV-lecA* lines 26, 20 and 4.

Key: H: High levels of pea lectin; L: Low levels of pea lectin.

Restriction site	Line 26			Line 20			Line 4		
	none	partial	complete	none	partial	complete	none	partial	complete
Hpa II	+	+	-	+	+	-	-	+	+
Msp I	+	+	-	+	+	-	-	+	+
Sau 3A	-	-	+	-	-	+	-	-	+
Hha I	-	-	+	-	-	+	-	-	+
Eco RII	-	-	+	-	-	+	-	-	+

Methylation level	Pea lectin levels		
	Line 26 (H)	Line 20 (H)	Line 4 (L)
none	2	2	0
partial	2	2	2
complete	3	3	5

resistance to cleavage by *HpaII, MspI , Sau3A, HhaI*, and *EcoRII* (figure 3.54 (B); lanes 2, 3, 6, 7, and 9). Many of the fragments homologous to the probe were observed to be greater than 9.4Kb.

3.8.3.3 Interpretation of methylation data.

The results of the methylation analysis were compiled together into a tables 3.24 and 3.25 based on whether the sites were non-methylated, partially methylated or completely methylated within the *lecA* sequence. This allows one to speculate as to the relationship between the extent of methylation occuring over regions of the *lecA* portion of the transgene and detectable pea lectin levels.

From table 3.23 one may observe that in all plants analysed methylation was found to occur at both CG and CNG sites (*HpaII-MspI, HhaI, EcoR II*, and *Sau3A* digests), sites known to be targets of plant methyl transferases (Adams and Pradhan, 1995). At non-CG and CNG sites methylation was not detected (*HaeIII* digest). Therefore methylation does occur within leaf tissue. However what is apparent from the data is that within the leaf different cell types may potentially exhibit differential methylation profiles. Because it appears to happen in this cell specific manner it is impossible to draw definite conclusions regarding the influence of methylation on transgene expression, within the leaf tissue.

3.8.4 Detection of histone acetylation in *CaMV-lecA* plants.

Histone acetylation, the acetylation of specific lysines within the amino terminal domains of each core histone is an active process whose precise role(s) remains controversial. Because the acetylation of lysine residues neutralises their positive charge, it has long been suggested that this modification modulates the interaction of amino termini with the negatively charged DNA backbone (Alfrey, 1977), leading to chromatin decondensation. Thus determination of the acetylation status of histones associated with the T-DNA may allow one to speculate as to the chromatin configuration at the site of integration. Recently the generation of antibodies which

recognise the hyperacetylated amino termini of core histones have become useful new probes to investigate the acetylation status of nucleosomes in transcriptionally active/inactive chromatin. In studies conducted by Jeppesen and Turner (1993) antibodies specific for acetylated isoforms of histone H4 were used to immunolabel human and mouse metaphase chromosomes. It was discovered that all chromosomes were labelled in regions corresponding to conventional R-bands (regions enriched in coding DNA) except for a single chromosome in female cells which was largely unlabelled and identified as the inactive X-chromosome. Many aspects of the inactive X-chromosome resemble non-transcribed constitutive heterochromatin, including the methylation of CpG islands at the 5' ends of several X-linked genes (Grant and Chapman, 1988). Tazi and Bird, 1990 have shown that nucleosomes from non-methylated CpG islands contain highly acetylated histones H3 and H4. A direct link between core histone acetylation and active genes was established using an antibody against the N^{ε}-acetyl-lysine for the isolation of acetylated nucleosomes from embryonic chicken chromatin Hebbes *et al.*, (1988). Immunofractionation was performed by incubating chromatin with antibody, and the complexes were isolated by centrifugation after binding to formalin fixed *Staphylococcus aureus* cells that had protein A in the outer membrane. On probing with gene sequences, antibody bound chromatin was found to be enriched in the actively transcribed α-globin gene and not enriched in the inactive ovalbumin gene. An investigation into the chromatin acetylation status of *CaMV-lecA* (26 and 4) T1 individuals was conducted similar to the approach described by Hebbes *et al.*, (1988), using antibody probes specific to the tetra-acetylated N-terminal domain of histone H4 (Lin *et al.*, 1989).

3.8.4.1 Investigation into the relationship between histone acetylation and transgene expression levels.

Nuclei from *CaMV-lecA* T1 individuals of lines 26 and 4 were isolated in the presence of 5mM sodium butyrate a known inhibitor of histone deactetylases which does not interfere with the activity of acetyl transferases (Hebbes *et al.*, 1988), thus preserving

core histone acetylation. The histones extracted with 0.4M sulphuric acid, were subjected to SDS-PAGE, blotted onto nitrocellulose and probed with rabbit polyclonal antibodies (1:500) specific to the acetylated terminal domain of histone H4 (figure 3.56) and detected via autoradiography after reacting with a secondary conjugated ^{125}I anti-rabbit IgG antibody. The results indicated banding patterns for the most conserved histones (H4, H3, H2A, and H2B) with less similarity in mobilities for proteins which could correspond to the least conserved histone H1 (figure 3.56; lane 1). On probing with a rabbit anti-acetylated histone H4 antibody, a band with similar mobility to histone H4 was detected (figure 3.56; lane A), suggesting that histone H4 in plant cells as in animal cells can be modified by acetylation. Since the rabbit anti-acetylated histone H4 antibody was observed to be functional it was used to fractionate chromatin fragments for both T1 individuals (see material and methods). Nuclei were first treated with micrococcal nuclease to generate a nucleosome ladder for both *CaMV-lecA* (26 and 4) T1 individuals (figure 3.57 (A); lanes 1, and 2). The nucleosome preparations were then fractionated by immune precipitation using the antibody, and DNA isolated from the antibody bound (acetylated fraction) and unbound (unacetylated fraction) nucleosomes. Analysis of DNA on 0.8%(w/v) agarose gels in figure 3.57 (B) revealed some faint nucleosomal banding; however it appeared that most of chromatin structure had been lost after the fractionation procedure. The gel was blotted onto nylon and probed with a ^{32}P-labelled 1.2Kb *Sph I/EcoRI* fragment containing the *lecA* coding region, however no enrichment was found for pea lectin sequences in either fraction. This may reflect the protocol used by which the DNA was analysed by agarose gel electrophoresis whereby concentrating the DNA and performing dot-blot hybridisations could have proved more sensitive. Due to time constraints it was not possible to continue this work further.

3.9 Conclusions, discussion and suggested further research.

This research programme has been concerned with detailing the major underlying factors exerting a determinatory effect on levels of transgene expression following *Agrobacterium*-mediated transformation of tobacco. A detailed investigation and characterisation of these factors should, it was hoped, allow the determination of the significance of the role played by factors such as transgene copy number and number and location of integration sites in determining the levels of transgene expression observed in stable transgenics.

Initial work centered on the characterisation of supplied material. Of the three lines provided only one, L4, was deemed suitable for detailed analysis. Having determined both transgene expression levels and copy number for this line it was noted that expression levels appeared to be independent of the number of integrated transgenes. However the obvious lack of suitable material i.e. of sufficient numbers of independent transgenic lines carrying the transgene at a number of different loci necessitated the generation of further transgenic populations in order to carry out this study in sufficient depth.

Thus, a large population of transgenic plants were generated harbouring both the *CaMV-lecA* and *ssRubisco-lecA* transgenes. Analysis of lectin levels (directly correlated with transgene expression levels) revealed extensive variation in expression levels within both populations. It was observed that, whille the range over which this variation was distributed was the same for both populations, the distribution of variation was significantly different between the two. Investigating the segregation of kanamycin resistance within these populations showed that transformants could be placed into one of four apparent inheritance classes: Class I plants had a segregation ratio of 3:1 kan^r:kan^s indicating the presence of a single integration locus; Class II plants segregated with a resistant:sensitive ratio of 15:1, a ratio associated with the

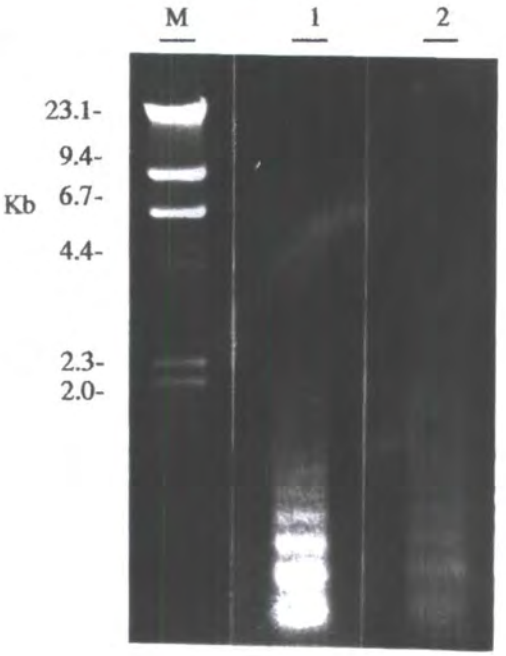
presence of two integration loci; plants of Class III possessed more than two integrations as shown by a ratio of 1:0, while Class IV plants segregated with a ratio of <3:1 which would appear to indicate the presence of less than one integration site, possibly reflecting the occurrence of some form of transgene silencing. A number of plants of classes I, II and III were then selected for further investigation.

Southern blot analysis allowed a detailed characterisation of T-DNA structure, copy number and number of integration sites to be undertaken. Results confirmed, in the majority of cases, the data previously obtained using kanamycin segregation analysis. A high frequency of multiple integrations was observed, with those plants containing >2 T-DNAs integrated at >2 sites showing a higher frequency of T-DNA rearrangements, all of which were observed to occur at or towards the left hand border. Plants containing more than one T-DNA were also frequently found to contain these T-DNAs arranged as an inverted repeat at a single locus although no significant relationship between copy number and the presence of such structures was found.

Having thus characterised in detail both T-DNA integrity and organisation further analysis was undertaken to determine the impact of T-DNA-related factors on transgene expression. Analysing levels of lectin accumulation in T1 populations of *CaMV-lecA* and *ssRubisco-lecA* revealed a significant difference in average levels of lectin accumulated. Within population variation was not statistically significant; this was expected since all plants within each T1 population were derived from a single transformant (T0). The variability in expression levels determined for T0 individuals was found to be inherited in the progeny, in agreement with previously published reports (Dean *et al.*, 1988).

Correlating transgene expression levels with T-DNA copy number and structure revealed no significant relationship to be evident. Thus one may conclude that the major factor influencing levels of transgene expression is the location of T-DNA integration within the genome. Subsequent work was thus concerned with investigating the nature of those integration site-specific factors i.e. 'position effect'.

A.



B.

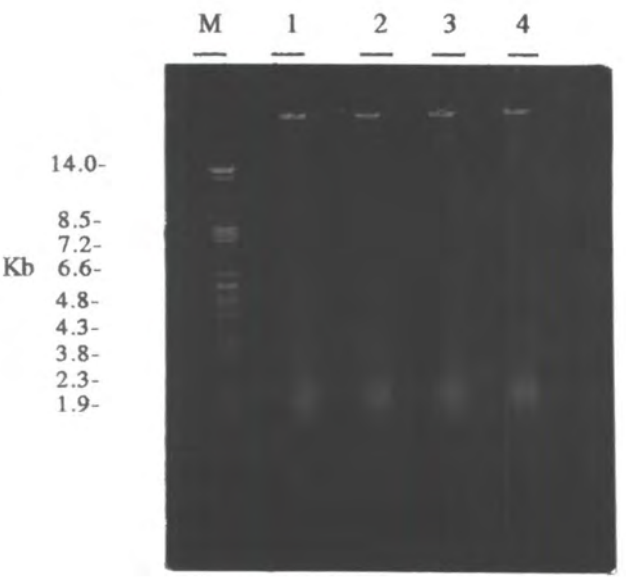


Figure 3.57:

(A) Agarose gel analysis of nuclei from *CaMV-lecA* 26 and 4 T1 individuals digested with micrococcal nuclease. Lane 1: Micrococcal nuclease digestion of nuclei from *CaMV-lecA* 26 T1 individual. Lane 2: Micrococcal nuclease digestion of nuclei from *CaMV-lecA* 4 T1 individual. Lane M: Lambda *HindIII* DNA size markers.

(B) Agarose gel analysis of nucleosomes fractionated using an anti-acetylated antibody. Lanes 1 and 2: *CaMV-lecA* 26 T1 individual unacetylated and acetylated fractions respectively. Lanes 3 and 4: *CaMV-lecA* 4 T1 individual unacetylated and acetylated fractions respectively. Lane M: Lambda *BstEII* size markers

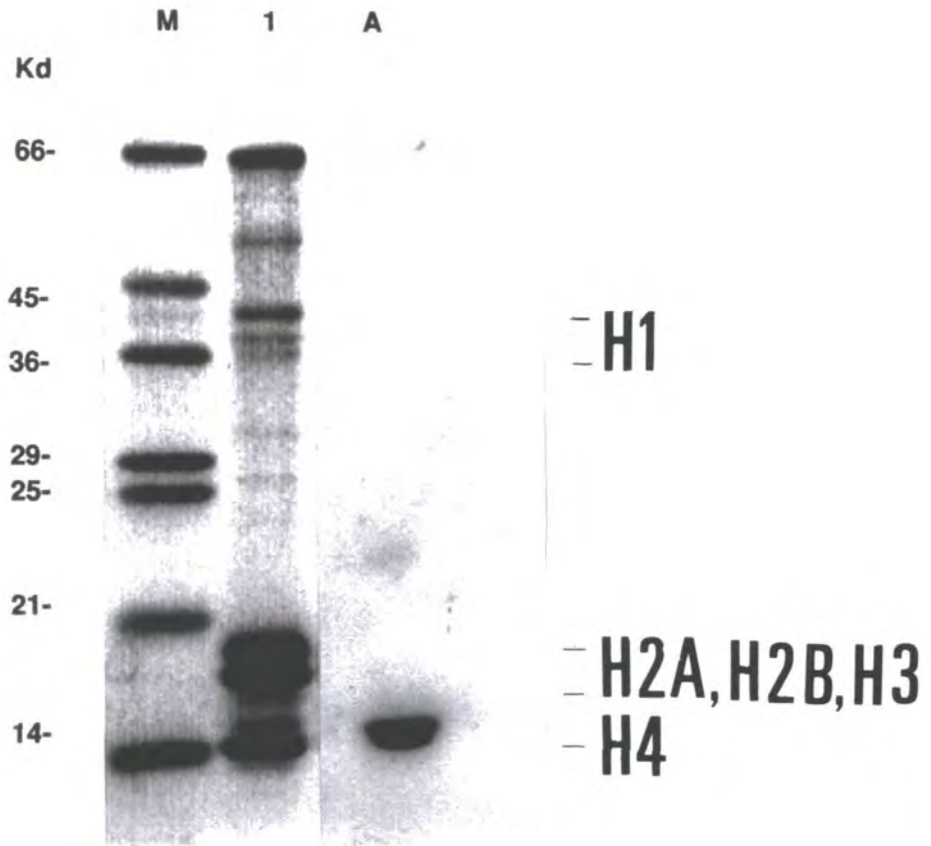

M
1
A
Kd
66-
45-
36-
29-
25-
21-
14-
H1
H2A, H2B, H3
H4

Figure 3.56:

SDS-PAGE analysis of isolated tobacco histone proteins and immunoblot detection of acetylated Histone 4. Lane M: Dalton VII SDS-PAGE molecular weight size marker. Lane 1: SDS-PAGE analysis of 0.2M H_2SO_4 extracted tobacco histone proteins for *CaMV-lecA* T1 individual 26, stained with coomassie blue. Lane A: Immunoblot of gel after probing with a rabbit anti-acetylated histone 4 antibody and detection with a ^{125}I conjugated goat anti-rabbit antibody via autoradiography.

In order to assess the chromatin conformation of the integration site a number of analyses were undertaken. Results obtained following DNaseI assay of selected material proved inconclusive, although would appear to indicate that, in all plants analysed, the integrated T-DNAs are located within DNaseI-resistant regions of the genome. Resistance to DNaseI digestion is taken to reflect a condensed chromatin conformation. Methylation analysis revealed that in all cases methylation was found to occur at both CG and CNG sites, sites known to be targets of plant methyl transferases. However, it was difficult to draw absolute conclusions from these analyses given the apparent cell type-specificity of methylation observed. Characterisation of histone acetylation within the integration site proved unsuccessful.

It thus appears that in optimising transgene expression in stably transformed plants a 'hierarchy' of contributory factors must be considered. Results obtained in this thesis show that, having eliminated the contribution of T-DNA copy number by selecting for single copy integrations, and as a direct consequence of this the elimination of factors associated with the presence of aberrant T-DNA structures, extreme levels of expression level variability are still apparent. Thus one can conclude that, having shown that these factors exert no significant effect on transgene expression levels, it is the site of T-DNA integration within the genome and factors associated with this that exert the major determinatory effect, giving rise to the observed levels of inter-transformant variability.

It must be accepted, therefore, that in designing a plant transformation strategy either for commercial exploitation or fundamental research purposes one must either accomodate transgene variability through the generation and analysis of large, and in many cases excessively so, transgenic populations or strive to minimise the extent to which such variability is manifest. Producing, maintaining and analysing large populations of transformed plants is clearly dependent on the provision of adequate facilities and large areas of containment growth space, and is a labour-intensive, costly process. Although such expense may be easily absorbed by larger commercial projects,

less so by smaller research groups, the cost-efficiency of such programmes would be greatly increased were it possible to reduce the extent of expression level variability. To this end only two approaches have met with substantial interest to date: the use of homologous recombination-based systems and nuclear scaffold/matrix attachment regions (SARs/MARs).

In principle, DNA integration into the plant genome via homologous recombination should be more precise than the random DNA insertion achieved by *Agrobacterium*-mediated transformation. In practice however, there has been a lack of success with homologous recombination reflecting the low frequencies with which recombination events occur within the genome. For example Swoboda *et al.*, (1994) set up a non-selective assay system to visualise intrachromosomal homologous recombination events throughout the whole life-cycle of *Arabidopsis thialana* transformed with two overlapping, non-functional deletion mutants of a chimaeric ß-glucoronidase (*uidA*) genes. Upon recombination, a functional *gus* gene was restored as determined by Southern blotting and detection by histochemical staining for GUS enzyme activity. Recombination was observed in most plant organs analysed throughout plant development. Recombination frequencies were typically low in the range 10^{-6}-10^{-7} events per genome. It was also demonstrated that recombination frequencies differed between organs within the same plant line. Thus it can be seen that homologous recombination within plant somatic cells occurs at extremely low frequencies and therefore at this time is not a viable consideration for large scale transformation programs. In contrast with homologous recombination, the application of site-specific recombination systems has been progressing steadily in recent years. Qin *et al.*, (1994) described the use of the Cre-*lox* system to integrate DNA into a *lox* site previously placed in the tobacco genome by transformation of protoplasts with a plasmid carrying a *lox* site. The efficiency of obtaining transformants with this system due to site specific integration approaches that of random integration. Most Cre-*lox*-mediated DNA integration events were found to be the insertion of a non-rearranged single copy of the introduced molecule at the chromosomal *lox* site. Because it allows some of the

variables common to current gene transformation to be controlled, this system would appear to be the likely candidate for future exploitation in the analysis of transgene behaviour. Although one limitation is that the system is dependent on a primary transformation programme which necessitates extensive characterisation of integration site and selection of different level expressors.

An alternative to site specific recombination is the modification of the T-DNA itself such that integration in some way is targeted. The primary interest in this area is in the use of nuclear scaffold or matrix attachment regions (SARs/MARs). In general, SARs/MARs appear to increase overall levels of transgene expression and decrease transformant-to-transformant variability in transgene expression (position effects). However there exists within the literature conflicting data with some workers reporting an increase in expression levels with either slight or no effect on variability within the population (Schöffl *et al.*, 1993, Spiker *et al.*, 1995) and others reporting a reduction in variability together with an overall increase in expression levels (Mlynárová *et al.*, 1994, 1995). The role of SARs/MARs in overcomimg position effects remains ambiguous and is currently the subject of much research.

Since the site of transgene integration within the genome appears to be the major contributory factor to transgene expression variability future research interests are to look at the influence of the different integration regions on gene activity. This would involve the isolation of the genomic insertion sites either by constructing genomic libraries in bacteriophage lambda or recovering the plant flanking sequences by inverse PCR (polymerase chain reaction). This analysis would be useful for a detailed characterisation of the T-DNA/plant DNA junctions to elucidate if sequence or structure-specific characteristics of the integration sites can be related to transgene expression level variability between independent transformants.

It may now be seen that the expression of a transgene as a consequence of its integration into the plant genome via *Agrobacterium*-mediated transformation is a complex series of events, influenced by a number of factors acting both independently

and together. This thesis has attempted to illuminate the worker as to the influence exerted by each of these factors and to address the phenomenon of transgene expression variability. It is hoped that future research here at Durham and elsewhere will allow us to undertake 'designer' transformation programmes, furthering the usefulness of transgenic plant research in the commercial and everyday environments.

References

Adams, R.L.P. and Burdon, R.H. (1985) eds. *Molecular Biology of DNA Methylation.* Springer-Verlag, NY.

Adams, R.L.P. (1990) DNA methylation: The effect of minor bases on DNA-protein interactions. Biochem. J. **265**: 309-320

Adams, R.L.P., and Pradhan S. (1995) Distinct CG and CNG DNA methylatransferases in *Pisum sativum.* Plant J. **7**:471-481.

Adams, R.L.P., Burdon, R.H., Campbell, A.M., Leader, D.P., and Smellie, R.M.S. (eds.) (1981) *The biochemistry of the nucleic acids. 9th edn.* Chapman and Hall, London pp. 517.

Alfrey, V.G. (1977) In: *Chromatin and Chromosome Structure.* H.J. Li and R.A. Eckhardt (eds.) Academic Press, NY. pp. 167-191.

Allen, G.C., Hall, G.E., Childs, L.C., Weissinger, A.K., Spiker, S. and Thompson, W.F. (1993) Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. Plant Cell **5**:603-613.

Allis, C.D., Glover, C.V.V. and Gorovsky, M.A. (1979) Micronuclei of *Tetrahymena* contain two types of histone H3. Proc. Natl. Acad. Sci. U.S.A. **76**:4857-4861.

Amasino, R.M., Powell, A.L.T. and Gordon, M.P. (1984) Changes in T-DNA methylation and expression are associated with phenotypic variation and plant regeneration in a crown gall tumour line. Mol. Gen. Genet. **197**:437-446.

An, G., Ebert, P.R., Mitra, A. and Ha, J.B. (1988) Binary vectors. In: *Plant Molecular Biology Manual.* S.B. Gelvin and R.A. Schileport (eds.) Kluwer Academic Publishers, Dordecht. pp. A3/1-A3/19.

Antequera, F. and Bird, A.P. (1988) Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J. **7**(8):2295-2299.

Assad, F.F., Tucker, K.L., and Signer, E.R. (1993) Epigenetic repeat-induced silencing (RIGS) in *Arabidopsis*. Plant. Mol. Biol. **22**:1067-1085.

Bevan, M. (1984) Binary *Agrobacterium* vectors for plant transformation. Nuc. Acids Res. **12**:8711-8721.

Benfey, P.N., Ren, L. and Chua, N.H. (1990a) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. EMBO J. **9**:1677-1684.

Benfey, P.N., Ren, L. and Chua, N.H. (1990b) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. EMBO J. **9**:1685-1696.

Bennet, M.D., and Smith, J.B. (1991) Nuclear DNA amounts in angiosperms. Phil. Trans. R. Soc. London **B334**:309-345.

Bestor, T.H., (1990) DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Phil. Trans. R. Soc. London **B236**:179-187.

Bestor , T.H., (1992) Activation of mammalian DNA methyltransferase by cleavage of a Zn binding domain. EMBO J. **11**:2611-2617.

Bianchi, M.W. and Viotti, A. (1988) DNA methylation and tissue-specific transcription of the storage protein genes of maize. Plant. Mol. Biol. **11**:203-214.

Bird, A.P. (1995) Gene number, noise reduction and biological complexity. Trends. Genet. **11**(13):94-100.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nuc. Acids Res. **7**:1513-1523.

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

Breyne, P. Gheyson, G. Jacobs, A., van Montagu, M. and Depicker, A. (1992a) Effects of T-DNA configuration on transgene expression. Mol. Gen. Genet. **235**:389-396.

Breyne, P., van Montagu, M.V., Depicker, A. and Gheyson, G. (1992b) Characterisation of a plant scaffold attachment region in a DNA fragment that normalises transgene expression in tobacco. Plant Cell **4**:463-471.

Brignon, P. Lepetit, M., Gigot, C. and Chaubert, N. (1993) Nuclease sensitivity and functional analysis of a maize histone H3 gene promoter. Plant Mol. Biol. **22**:1007-1015.

Brink, R.A., (1973). Paramutation. Ann. Rev. Genet. **7**:129-152

Broach, J.R., (1981). In: *The Molecular Biology of the Yeast Saccharomyces: Life cycle and Inheritance.* Strathern J.N., Jones, E.W., Broach, J.R. (eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor. pp. 445-470.

Brusslan J.A., and Tobin, E.M. (1995) Isolation of new promoter mediated co-suppressed lines in *Arabidopsis thaliana*. Plant Mol. Biol. **27**:809-813.

Buchanan-Wallaston, V., Passiatore, J.E. and Cannon, F. (1987) The mobilisation functions (mob and ori-T) of a bacterial plasmid promoter its transfer to plants. Nature **328**:172-175.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., and Roberts, R.J (1993) Crystal structure of *HhaI* methyl-transferase complexed with S-Adenosyl-L-Methionine. Cell **74**:299-307.

Chung, C.T., Niemla, S.L., and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA. **86**:2172-2175.

Chyi, Y.S., Jorgansen, R.A., Goldstein, D., Tanksley, S.D. and Loaiza-Figueroa, F. (1986) Location and stability of *Agrobacterium*-mediated T-DNA insertions in the *Lycopersicum* genome. J. Bacteriol. **170**:2659-2667.

Citovsky, V., Warnick, D., and Zambryski, P. (1992) Nuclear localisation of Agrobacterium VirE2 protein in plant cells. Science **256**:1802-1805.

Citovsky, V., Zupan, J., Warnick, D., and Zambryski, P. (1994) Nuclear import of Agrobacterium tumefaciens VirD2 and VirE2 proteins in maize and tobacco. Proc.Natl.Acad.Sci. USA **91**:3210-3214.

Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. Nuc. Acids Res. **22**(15):2990-2997.

Coates, D., Talierco, E.W. and Gelvin, S.B. (1987) Chromatin structure of integrated T-DNA in crown gall tumours. Plant Mol. Biol. **8**:159-168.

Coe, E.H.J. (1966) The properties, origin and mechanism of conversion-type inheritance at the B locus in maize. Genetics **53**:1035-1063.

Csordas, A. (1990) On the biological role of histone acetylation. Biochem. J. **265**:23-28.

Counts J.L., and Goodman, J.I. (1995) Alterations in DNA methylation may play a variety of roles in carcinogenesis. Cell **83**:13-15.

Covey, S.N., Lomonossoff, G.P., and Hull, R. (1981) Characterisation of cauliflower mosaic virus DNA sequences which encode major polyadenylated transcripts. Nuc. Acids Res. **24**:6735-6747.

Dean, C., Jones, J., Favreau, M., Dumsmuir, P. and Bedbook, J. (1988) Influence of flanking sequences on variability in expression levels of an introduced gene in transgenic tobacco plants. Nuc. Acids Res. **16**:9267-0282.

DeBlock, 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.

DeNeve, M., Deloose, M., Jacobs, A., Van Houdt, H., Kaluza, B., Weidle, U., van Montagu, M. and Depicker, A. (1993) Assembly of an antibody and its derived antibody fragment in *Nicotiana* and *Arabidopsis*. Transgenic Res. **2**:227-237.

Dennis, E.S. and Brettel, R.I.S. (1990) DNA methylation of transposable elements is correlated with activity. Phil. Trans. Royal Soc. Lond. (Biol.) **326**:217-229.

Dreeson, T.D., Henikoff, S., and Loughney, K. (1991) A pairing-sensitive element that mediates trans-inactivation is associated with the Drosophila brown gene. Genes Develop. **5**:331-340.

Ebert, P.R., Ha, S.B. and An, G. (1987) Identification of an essential upstream element in the nopaline synthase promoter by stable and transient assays. Proc. Natl. Acad. Sci. USA **84**:5745-5749.

Eckdahl, T.T., Bennetzen, J.L. and Anderson, J.N. (1989) DNA structures associated with autonomously replicating sequences from plants. Plant Mol. Biol. **12**:507-516.

Edwards, G.A., Hepher, A., Clerk, S.P. and Boulter, D. (1991) Pea lectin is correctly processed, stable and active in leaves of transgenic potato plants. Plant Mol. Biol. **17**:89-100.

Edwards, G.A. (1987) Plant transformation using an *Agrobacterium tumefaciens* Ti plasmid vector system. Ph.D. thesis, University of Durham.

Ellis, J.G., Llewellyn, D.J., Dennis, E.S., and Peacock, W.J. (1987) ADH-1 promoter sequences control anaerobic regulation: addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. EMBO J. **6**:11-16.

Fang, R.-X., Nagy, F., Sivasubramanian, S. and Chua, N.-H. (1989) Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell **1**:141-150.

Feinberg, A.P. and Vogelstein, E. (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266-267.

Filichkin S.A., and Gelvin S.B. (1993) Formation of a putative relaxation intermediate during T-DNA processing directed by the *Agrobacterium tumefaciens* VirD2, D1 endonuclease. Mol. Microbiol. **8**:915-926.

Finnegan, J.E. and Dennis, E.S. (1993) Isolation and identification by sequence homology of putative cytosine methyltransferase from *Arabidopsis thaliana.* Nuc. Acids Res. **21**(10):2383-2388.

Finnegan, J.E. and McElroy, D. (1994) Transgene inactivation: Plants fight back! Bio/Technology **12**:883-888.

Finnegan, E.J., Brettell, R.I.S., and Dennis, E.S. (1993) The role of DNA methylation in the regulation of plant gene expression. In: *DNA methylation: Molecular Biology and Biological Signifigance.* J.P. Jost and H.P. Saluz (eds). BaseI: Birkhauser, pp 218-261.

Flavell, R.B., O'Dell, M., Smith, D.B., and Thompson W.F. (1985) Chromosome architecture: the distribution of recombination sites, the structure of ribosomal DNA loci and the multiplicity of sequences containing inverted repeats. In: *Molecular Form and Function of the Plant Genome.* L. van Vloten-Doting, G.S.P. Groot, and T.C. Hall (eds.) NATO ASI series A(83). Life sciences Plenum Press, New York. pp1-14.

Flavell, R.B. (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc. Natl. Acad. Sci. USA **91**:3490-3496.

Fluhr, R., Kuhlemeier, C., Nagy, F. and Chua, N.-H. (1986) Organ-specific and light-induced expression of plant genes. Science **232**:1106-1112.

Freedman, T., and Pukkila, P.J. (1993). *De novo* methylation of repeated sequences in *Coprinus cinereus.* Genetics **135**:357-366.

Fromm, M.E., Taylor, L.P., and Walbot, V. (1986) Stable transformation of maize after gene transfer by electroporation. Nature **319**:791-793.

Gasser, S.M., Amati, B.B., Cardenas, M.E., and Hofmann, J.F.X. (1989) Studies on scaffold attachment sites and their relation to genome function. Internat. Rev. Cytol. **119**:57-96.

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

Gasser, S.M., and Laemmli, U.K. (1986) The organization of chromatin loops: Characterization of a scaffold attachment site. EMBO J. **5**:511-518

Gatehouse, J.A., Brown, D., Evans, I.M., Gatehouse, L.N., Jobes, D., Preston, P. and Croy, R.R.D. (1987) Sequence of the seed lectin gene from pea (*Pisum sativum* L.) Nuc. Acids. Res. **15**:7642.

Gelvin, S.B., Karcher, S.J. and DiRita, V.J. (1983) Methylation of the T-DNA in *Agrobacterium tumefaciens* and in several crown gall tumours. Nuc. Acids Res. **11**(1):159-174.

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

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

Goring, D.R., Thomson, L., and Rothstein, S.J. (1991) Transformation of a partial nopaline synthase gene into tobacco suppresses the expression of a resident wild-type gene. Proc. Natl. Acad. Sci. USA **88**:1770-1774.

Gorz, A., Schafer, W., Hirasawa, E. and Kahl, G. (1990) Constitutive and light induced DNaseI hypersensitive sites in the *rbcS* genes of pea (*Pisum sativum*). Plant Mol. Biol. **11**:561-573.

Grant, S.G., and Chapman, V.M., (1988) Mechanisms of X-chromosome regulation. Annu. Rev. Genet. **22**:199-233.

Green, P.J., Kay, S.K., and Chua, N-H. (1987) Sequence specific interactions of a pea nuclear factor with light responsive elements upstream of the *rbcS*-3A gene. EMBO J. **6**:2543-2544.

Gruenbaum, Y., Naveh-Mavy, T., Cedar, H. and Razin, A. (1981) Sequence specificity of methylation in higher plant DNA. Nature **292**:860-862.

Guerineau, F., Lucy, A. and Mullineaux, P. (1992) Effect of two consensus sequences preceding the translational initiator codon on gene expression in plant protoplasts. Plant Mol. Biol. **18**:815-818.

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

Hart, C.M., Fischer, B., Neuhaus, J-M., and Meins, F.J. (1992) Regulated inactivation of homologous gene expression in transgenic *Nicotiana sylvestris* plants containing a defence-related tobacco chitinase gene. Mol. Gen. Genet. **235**:179-188.

Hebbes, T.R., Thorne, A.W. and Crane-Robinson, C. (1988) A direct link between core acetylation and transcriptionally active chromatin. EMBO J. **7**:1395-1402.

Henikoff, S., and Dreeson, T.D. (1989) Trans-inactivation of the Drosophila brown gene: Evidence for transcriptional repression and somatic pairing dependence. Proc. Natl. Acad. Sci. USA **86**:6704-6708.

Hepburn, A.G., Clarke, L.E., Pearson, L., and White, J. (1983) The role of cytosine methylation in the control of nopaline synthase gene expression in a plant tumour. J. Mol. Appl. Genet. **2**:315-329.

Herrera-Estrella, L., and Simpson, J. (1988) Foreign gene expression in plants. In: *Plant molecular biology-a practical approach.* C.H. Shaw (ed.) IRL press, Oxford. pp. 131-160.

Herrman, L., Jacobs, A., van Montagu, M. and Depicker, A. (1990) Plant chromosome-marker gene fusion assay to study normal and truncated T-DNA integration events. Mol. Gen. Genet. **224**:248-256.

Hirochika, H. (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J. **12**(6):2521-2528.

Hobbs, S.L.A., Kpodar, P., and Delong, C.M.O. (1990) The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. Plant Mol. Biol. **15**:851-864.

Hohn, T., Richards, K. and Lebeuvier, G. (1982) Cauliflower mosaic virus on its way to becoming a useful plant vector. Curr. Top. Microbiol. Immunol. **96**:194-236.

Holliday, R. (1987) The inheritance of epigenetic defects. Science **238**:163-170.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eicholtz, D., Rogers, S.G. and Fraley, R.T. (1985) A simple and general method for transferring genes into plants. Science **227**:1229-1231.

Houlston, C.E., Lindsay, H., Pradhan, S. and Adams, R.L.P. (1993) DNA substrate specificity of pea DNA methylase. Biochem. J. **293**:617-624.

Howard, A.E., Zupan, J.R., and Zambryski, P. (1992) The VirD2 protein of Agrobacterium tumefaciens contains a C-terminal bipartite nuclear localisation signal: implications for nuclear uptake of DNA in plant cells. Cell **68**:109-118.

Ingelbrecht, I.L.W., Breyne, P., Vancomperolle, K., Jacons, A., van Montagu, M. and Depicker, A. (1991) Transcriptional interference in transgenic plants. Gene **109**:239-242.

Ingelbrecht, I., Van Houdt, H., Van Montagu M., and Depicker A. (1994) Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. Proc. Natl. Acad. Sci. USA **91**:10502-10506.

Jackson, D.A. (1991) Structure-function relationships in eukaryotic nuclei. BioEssays **13**:1-10

Jahn, R., Scheibler, W. and Greengard, P. (1984) A quantitative dot-blot immunobinding assay for proteins using nitrocellulose. Proc. Natl. Acad. Sci. USA **81**:1684-1687.

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

Jeppesen, P. and Turner, B.M.(1993) The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell **74**:291-289.

Jin, S., Prusti, T., Roitsch, R.G., Ankenbauer, R.G., and Nester, E.W. (1990) Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG. J. Bacteriol. **72**:4945-4950.

John, M.C. and Amasino, R.M. (1989) Extensive changes in DNA methylation patterns accompanying activation of a silent T-DNA *nptII* gene in *Agrobacterium tumefaciens* -transformed plant cells. Mol. Cell Biol. **9**:4298-4303.

Jones, J.D.G., Dunsmuir, P. and Bedbrook, J. (1985) High level expression of introduced chimaeric genes in regenerated transformed plants. EMBO J. **4**:2411-2418.

Jones, J.D.G., Gilbert, D.E., Grady, K.L. and Jorgensen, R.A. (1987) T-DNA structure and gene expression in petunia plants transformed by *Agrobacterium tumefaciens* C58 derivatives. Mol. Gen. Genet. **207**:478-485.

Jorgensen, R.A., Snyder, C. and Jones, J.D.G. (1987) T-DNA is organised predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. Mol. Gen. Genet. **207**:471-477.

Jorgensen, R.A. (1995) Cosuppression, flower color patterns, and metastable gene expression states. Science **268**:868-691.

Jost, J-P., and Hofsteenge, J. (1992) The repressor MDBP-2 is a member of the histone H1 family that binds preferentially *in-vitro* and *in-vivo* to methylated non-specific DNA sequences. Proc. Natl. Acad. Sci. USA **89**:9499-9503.

Katagiri, F., Lam, E., and Chua, N-H. (1989) Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature **340**:727-730.

Koukolikova-Nicola, Z., Albright, L., and Hohn B., (1987) The mechanism of T-DNA transfer from Agrobacterium tumefaciens to the plant cell. In:*Plant DNA infectious agents.* Hohn, T., and Schell, J. (eds.) Springer Verlag, Vienna.

Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. Science **236**:1299-1302.

Keshet, I., Liemen-Hurwitz, J. and Cedar, H. (1986) DNA methylation affects the formation of active chromatin. Cell **44**:535-543.

Kilby, N.J., Snaith, M.R., and Murray, J.H. (1993) Site-specific recombinases: tools for genome engineering. Trends. Genet. **9**:413-421.

Klaas, M., John, M.C., Crowell, D.N. and Amasino, R.M. (1989) Rapid induction of genomic demethylation and T-DNA gene expression in plant cells by 5-azacytidine derivatives. Plant Mol. Biol. **12**:413-423.

Klein, T.M., Harper, E.C., Svab, Z., Sanford, J.C., Fromm, M.E., and Maliga, P. (1988) Stable genetic transformation of intact Nicotiana cells by the particle bombardment process. Proc. Natl. Acad. Sci. USA **85**:8502-8505.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Redei, G.P. and Schell, J. (1989) High frequency T-DNA-mediated gene tagging in plants. Proc. Natl. Acad. Sci. USA **86**:8467-8471.

Kovarik, A., Koukalova, B., Holy, A., and Bezdek, M. (1994) Sequence-specific hypomethylation of the tobacco genome induced with dihydroxypropyladenine, ethionine, and 5-azacytidine. FEBS Lett. **353**:309-311

Kozak, M. (1987) An analysis of 5' non-coding sequences from 699 vertebrate messenger RNAs. Nuc. Acids. Res. **15**:8125-8148.

Krapp, A., Hofmann, B., Scaefer, C. and Sitt, M. (1993) Regulation of the expression of the *rbcS* and other photosynthetic genes by carbohydrates: a mechanism for the 'sink regulation' of photosynthesis? Plant J. **38**:221-257.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of Bacteriophage T4. Nature **227**:680-685.

Lam, E., Benfey, I.E., Gilmartin, P.N., Fang, P.M. and Chua, N.-H. (1989) Site specific mutations alter *in vitro* factor binding and charge promoter expression pattern in transgenic plants. Proc. Natl. Acad. Sci. USA **86**:7890-7894.

Langridge, W.H.R., Fitzgerald, K.J., Koncz, C., Schell, J. and Szalay, A.A. (1989) Dual promoter of *Agrobacterium tumefaciens* mannopine synthase genes is regulated by plant growth hormones. Proc. Natl. Acad. Sci. USA **86**:3219-3223.

Larkin, P.J., and Scowcroft, W.R. (1981) Somoclonal variation: a novel source of variability from cell cultures for plant improvement. Theor. Appl. Genet. **60**:197-214.

Laurenson, P., and Rine, J. (1992) Silencers, silencing and heritable transcriptional states. Micorobiol. Rev. **56**:543-560.

Lessl, M. and Lanka, E. (1994) Common mechanisms in bacterial conjugation and Ti-mediated transfer to plant cells. Cell **77**:321-324.

Lessl, M., Balzer, D, Pansegrau, W., and Lanka, E. (1992) Sequence similarities between the RP4 Tra2 and the Ti virB region strongly support the conjugation model for T-DNA transfer. J. Biol. Chem. **267**:20471-20480.

Lewis, J. and Bird, A. (1991) Methylation and chromatin structure. FEBS Lett. **285**:155-159.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) Purification, sequence and cellular localisation of a novel chromosomal protein that binds to methylated DNA. Cell **69**:905-914.

Lichtenstein, C., and Barrena, E. (1993) Prospects for reverse genetics in plants using recombination. Plant Mol. Biol. **21**:5-12.

Lin, R., Leone, J.W., Cook, R.G. and Allis, D.C. (1989) Antibodies specific to acetylated histones document the existence of deposition- and transcription-related histone acetylation in *Tetrahymena*. J. Cell. Biol. **108**:1577-1588.

Linn, F., Heidmann, I., Saedler, H. and Meyer, P. (1990) Epigenetic changes in the expression of the maize A1 gene in *Petunia hybrida*: role of numbers of integrated copies and the state of methylation. Mol. Gen. Genet. **222**:329-336.

Lloyd, A.M., and Davis, R.W. (1994) Functional expression of the yeast FLP/*frt* site-specific recombination system in *Nicotiana tabacum*. Mol. Gen. Genet. **242**:653-657.

Lund, G., Ciceri, P., and Viotti, A. (1995a) Maternal-specific demethylation and expression of specific allels of zein genes in the endosperm of *Zea mays* L. Plant J. **8**:571-581.

Lund, G., Das, O.P., Messing, J. (1995b) Tissue-specific DNaseI-sensitive sites of the maize *P* gene and their changes upon epimutation. Plant J. **7**:797-807

Maliga, P., Sz-Brenznovits, A. and Marton, L. (1973) Streptomycin resistant plants from callus culture of haploid tobacco. Nature **244**:28-30.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A laboratory manual*. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY.

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. EMBO J. **8**:643-649.

Matzke, A.J.M., Neuhuber, F., Park, Y-D., Ambros, P., and Matzke, M.A. (1994) Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. Mol. Gen. Genet. **244**:219-229.

Matzke, M.A., Neuhuber, F., Matzke, A.J.M. (1993) A variety of epistatic interactions can occur between partially homologous transgene loci brought together by sexual crossing. Mol. Gen. Genet. **236**:379-386.

Matsumoto, S., Ito, Y., Hosoi, Y., 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.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, 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.

McClintock, B. (1984) The significance of responses of the genome to the challenge. Science **226**:792-801.

McGhee, J.D., and Engel, J.D. (1975) Subunit structure of chromatin is the same in plants and animals. Nature **254**:449-450.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell **58**:499-507.

Meyer, P., and Heidmann, I. (1994) Epigenetic variants of a transgenic petunia line show hypermethylation in transgene DNA: An indication for specific recognition of foreign DNA in transgenic plants. Mol. Gen. Genet. **243**:390-399.

Meyer, P., Heidmann, I., and Niedenhof, I. (1993) Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. Plant J. **4**:89-100.

Meyer, P., Niedenhof, I., and Ten Lohuis, M. (1994) Evidence for cytosine methylation of non-symmetrical sequences in transgenic Petunia hybrida. EMBO J. **13**:2084-2088.

Miranda, A., Janssen, G., Hodges, L., Peralta, E.G., and Ream, W. (1992) *Agrobacterium tumefaciens* transfers extremely long T-DNAs by a unidirectional mechanism. J. Bacteriol. **174**:2288-2297.

Mlynárová, L., Loonen, A., Heldens, J., Jansen, R.D., Keizer, P., Stiekema, W.J. and Nap, J.-P. (1994) Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. Plant Cell **6**:417-426.

Mlynávorá, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., and Nap, J-P. (1995) The MAR-mediated reduction in position effect can be uncoupled from copy-number expression in transgenic plants. Plant Cell **6**:417-426.

Mol, J., van Blockland, R., and Kooter, J. (1991) More about co-suppression. Trend. Biotech. **9**:182-183.

Monk, M., Boubelik, M., and Lehnert, S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Devel. **99**:371-382.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant. **15**:473-497.

Murray, M.G., and Kennard, W.C. (1984) Altered chromatin conformation in the higher plant gene phaseolin.. Biochemistry **23**:4225-4230.

Nagy, F., Morelli, G., Fraley, R.T., Rogers, S.G. and Chua, N.-H. (1985) Photoregulated expression of a pea *rbcS* gene in leaves of transgenic plants. EMBO J. **12**:3063-3068.

Neuhuber, F., Park, Y-D., Matzke, A.J.M., and Matzke, M.A. (1994) Susceptibility of transgene loci to homology dependent gene silencing. Mol. Gen. Genet. **244**:230-241.

Omirulleh, S., Abraham, M., Golovkin, M., Stefanov, I., Karabaev, M.K., Mastardy, L., Morocz, S. and Dudits, D. (1993) Activity of a chimaeric promoter with the doubled CaMV 35S promoter element in protoplast-derived cells and transgenic plants and maize. Plant Mol. Biol. **21**:415-428.

Onlee, T.M., Turgeon, R. and Wu, R. (1986) Expression of a foreign gene linked to either a plant virus or a *Drosophila* promoter after electroporation of protoplasts of rice, wheat and sorghum. Proc. Natl. Acad. Sci. USA **83**:6815-6819.

Odell, J.T., Nagy, F. and Chua, N.-H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature **313**:810-812.

Ow, D.W., Jacobs, J.D. and Howell, S.H. (1987) Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by the firefly luciferase gene as a reporter of promoter activity. Proc. Natl. Acad. Sci. USA **84**:4870-4874.

Palaqui, J-C., and Vaucheret, H. (1995) Field trial analysis of nitrate reductase co-supression: a comparative study of 38 combinations of transgene loci. Plant Mol. Biol. **29**:149-159.

Pansegrau, W., and Lanka, E. (1991) Common sequence motifs in DNA relaxases and nick regions from a variety of DNA transfer systems. Nuc. Acids. Res. **19**:3455

Paranjape, S.M., Kamakaka, R.T., and Kadonaga, J.T. (1994) Role of chromatin structure in the regulation of transcription by RNA polymerase II. Annu. Rev. Biochem. **63**:265-297.

Paro, R. (1990) Imprinting a determined state into the chromatin of Drosophila. Curr. Opin. Cell Biol. **5**:999-1005.

Paszkowski , J., Shillito, R.D., Saul, M., Mandak, V., Hohn, T.B., Hohn, B., and Potrykus, I. (1984) Direct gene transfer to plants. EMBO J. **3**:2717-2722.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. (1988) Gene targeting in plants. EMBO J. **7**:4021-4026.

Patterson, G.I., and Chandler, V.L. (1995) Paramutation in maize and related allelic interactions. In: *Gene silencing in Higher plants and Related phenomena in Other Eukaryotes*, and edition of *Current Topics in Microbiology and Immunology* Meyer P (ed.) Springer, Berlin: **197** pp121-141.

Patterson, G.I., Thorpe, C.J., and Chandler, V.L. (1993) Paramutation, an allelic interaction, is associated with a stable and heritable reduction of transcription of the maize-B regulatory gene. Genetics **135**:881-894.

Pawlak, A., Bryans, M., and Jost, J-P. (1991) An avian 40Kda nucleoprotein binds preferentially to a promoter sequence containing one single pair of methylated CpG. Nuc. Acids. Res. **19**:1029-1034.

Pazour, G.J., Ta, C.N., and Das, A. (1992) Constitutive mutations of *Agrobacterium tumefaciens* transcriptional activator virG. J. Bacteriol. **174**:4169-4174.

Peach, C. and Velten, J. (1991) Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. Plant Mol. Biol. **17**:49-60.

Peerbolte, R., Leenhouts, K., Hooykas-van Slogteren, G.M.S., Wullems, G.J. and Schilperoot, R.A. (1986) Clones from a shooty tobacco crown gall tumour II: Irregular T-DNA structures and organisation, T-DNA methylation and conditional expression of opine genes. Plant Mol. Biol. **7**:285-299.

Pfeiffer, P. and Hohn, T. (1983) Involvement of reverse transcription in the replication of cauliflower mosaic virus: A detailed model and test of some aspects. Cell **33**:7781-789.

Philips, R.L., Kaeppler, S.M., and Olhoft, P. (1994) Genetic instability of plant tissue cultures: Breakdown of normal controls. Proc. Natl. Acad. Sci. USA **91**:5222-5226.

Phi-Van, L., von Kries, J.P., Ostertag, W., and Stratling, W.H. (1990) The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol. Cell. Biol. **10**:2302-2307.

Prols, F. and Meyer, P. (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in *Petunia hybrida*. Plant J. **2**:465-475.

Proudfoot, N.J. (1986) Transcriptional interference and termination between duplicated α-globin gene constructs suggests a novel mechanism for gene regulation. Nature **322**:562-565.

Qin, M., Bayley, C., Stockton, T. and Ow, D.W. (1994) Cre recombinase-mediated site-specific recombination between plant chromosomes. Proc. Natl. Acad. Sci. USA **91**:1706-1710.

Razin, A., and Cedar, H. (1991) DNA methylation and gene expression. Microbiol. Rev. **55**:451-458.

Renckens, S., De Greve, H., van Montagu, M. and Hernalsteens, J.-P. (1992) Petunia plants escape from negative selection against a transgene by silencing the foreign DNA via methylation. Mol. Gen. Genet. **233**:53-64.

Rossignol, J.-L., Faugeron, G. (1994) Gene inactivation triggered by recognition between DNA repeats. Experientia **50**:307-317.

Sambrook, J., Fritsch, E.F., and Maniatis (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanders, P.R., Winter, J.A., Zarnason, A.R., Rogers, S.G. and Fraley, R.T. (1987) Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. Nuc. Acids Res. **15**:1543-1558.

Santi, D.V., Garret, C.E. and Barr, P.J. (1983) On the mechanism of inhibition of DNA cytosine methyltransferases by cytosine analogues. Cell **33**:9-10.

Sauer, B., and Henderson, N. (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc. Natl. Acad. Sci. USA **85**:5166-5170.

Sauer, B (1987) Functional expression of the Cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisae.* Mol. Cell. Biol. **7**:2087-2096.

Sawyer, R.M., Boulter, D. and Gatehouse, J.A. (1987) Nuclease sensitivity of storage protein genes in isolated nuclei of pea seeds. Planta **171**:254-258.

Schiedt, G., Weber, H., Graessmann, M. and Graessmann, A. (1994) Are there two DNA methyltransferase gene families in plant cells? A new potential methyltransferase gene isolated from an *Arabidopsis thaliana* genomic library. Nuc. Acids Res. **22**(6):953-958.

Schlappi, M., Raina, R. and Federoff, N. (1994) Epigenetic regulation of the maize *Spm* transposable element: novel avtivation of a methylated promoter by TnpA. Cell **77**:427-437.

Schoffl, F., Schroeder, G., Kleim, M., and Rieping, M. (1993) An SAR sequence containing 395bp DNA fragment mediates enhanced, gene-dosage-correlated expression of a chimaeric heat shock gene in transgenic tobacco plants. Transgenic Res. **2**:93-100.

Scowcroft, W.R. (1985) Somoclonal variation: the myth of clonal uniformity. In: *Genetic flux in plants*. Hohn B., and Dennis, E.S. (eds.) Springer -Verlag, Wien, pp. 217-245.

Selker, E.U., Fritz, D.Y., and Singer, M.J. (1993) Dense non-symmetrical DNA methylation resulting from repeat induce point mutation (RIP) in *Neurospora*. Science **262**:1724-1728.

Selker, E.U. (1990) Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. **24**:579-613.

Shapiro, H.S. (1975) Distribution of purines and pyrimidines in DNA. In: *CRC Handbook of Biochemistry and Molecular Biology* (Sober, H.A., ed.). Boca Raton, FL:CRC Press, pp. H30-H51.

Shirsat, A.H., Wilford, N. and Croy, R.R.D. (1989) Gene copy number and levels of expression in transgenic plants of a seed-specific promoter. Plant Sci. **61**:75-80.

Smith, S.S., Kaplan, B.E., Sowers, L.C., and Newman, E.M. (1992) Mechanism of human methyl-directed DNA methyl-transferase and the fidelity of cytosine methylation. Proc. Natl. Acad. Sci. USA **89**:4744-4748

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503-517.

Spielman, A. and Simpson, R.B. (1986) T-DNA structure in transgenic tobacco plants with multiple independent integration sites. Mol. Gen. Genet. **205**:34-41.

Spiker, S., Murray, M.G. and Thompson, W.F. (1983) DNaseI sensitivity of transcriptionally active genes in intact nuclei and isolated chromatin of plants. Proc. Natl. Acad. Sci. USA **80**:815-819.

Spiker, S., Allen., G.C., Hall, G.E. Jr., Michalowski, S., Newman, W., Thompson, W., and Weissinger, A.K. (1995). Nuclear matrix attachment regions (MARs) in plants: affinity for the nuclear matrix and effect on transient and stable gene expression. J. Cell Bio-chem **21B**:167.

Stachel, S.E., Timmerman, B. and Zambryski, P. (1986) Generation of single stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cells. Nature **322**:706-712.

Steinmuller, K. and Apel, K. (1986) A simple and efficient procedure for isolating plant chromatin which is suitable for studies of DNaseI sensitive domains and hypersensitive sites. Plant Mol. Biol. **7**:87-94.

Stief, A., Winter, D.M., Stratling, W.H. and Sippel, A.E. (1989) A nuclear DNA attachment element mediates elevated and position independent gene activity. Nature **341**:343-345.

Sugita, M. and Gruissem, W. (1987) Developmental, organ-specific and light-dependent expression of the tomato ribulose 1,5-bisphosphate carboxylase small subunit gene family. Proc. Natl. Acad. Sci. USA **84**:7104-7108.

Swoboda, P., Hohn, B., and Gal, S. (1993) Somatic homologous recombination in planta: The recombination frequency is dependent on the allelic state of the recombining sequences and may be influenced by genomic position effects. Mol. Gen. Genet. **237**:33-40.

Swoboda, P., Gal, S., Hohn, B., and Puchta, H. (1994) Intrachromosomal homologous recombination in whole plants. EMBO J. **13**:484-489.

Tanaka, A., Mita, S., Ohta, S., Kyazyka, J., Shimamoto, K. and Nakamura, K. (1990) Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron. Nuc. Acids Res. **18**:6767-6770.

Tazi, J., and Bird, A. (1990) Alternative chromatin structure at CpG islands. Cell **60**:909-920.

Teeri, T.H., Lehvaslaiho, H., Frank, M., Uotila, J., Heino, P., Palva, E.T., van Montagu, M. and Herrera-Estrella, L. (1989) Gene fusions to *lacZ* reveal new expression patterns of chimaeric genes in transgenic plants. EMBO J. **8**:343-350.

Thomas, P.S., and Farquar, M.N (1978) Specific measurement of DNA in nuclei and nucleic acids using diaminobenzoic acid. Anal. Biochem **89**:35-44.

Tinland, B., Koukolikova, N.Z., Hall, M.N., and Hohn, B. (1992) The T-DNA-linked VirD2 protein contains two distinct functional nuclear localisation signals. Proc Natl. Acad. Sci. USA **89**:1123-1130.

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

Tobin, E.M. and Silverthorne, J. (1985) Light regulation of gene expression in higher plants. Annu. Rev. Plant Physiol. **36**:569-593.

Torres-Ruiz, R.A., and Hemleben, V. (1994) Pattern and degree of methylation in ribosomal RNA genes of *Cucurbita pepo* L. Plant Mol. Biol. **26**:1167-1179.

Towbin, J., Staehelin, T., Gordon, J., (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc.Natl.Acad.Sci. **76**:4350-4354.

Trowbridge, L.S. (1974) Isolation and characterisation of a mitogenic lectin from *Pisum sativum.* J. Biol. Chem. **249**:6004-6012.

Ueda, T., Pichersky, E., Malik, V.S. and Cashmore, A.R. (1989) Level of expression of the tomato *rbcS*-3A gene is modulated by a far upstream promoter element in a developmentally regulated manner. Plant Cell **1**:217-227.

Van der Geest, A.H.M., Hall, G.E. Jr., Spiker, S., and Hall, T.C. (1994) The ß-phaseolin gene is flanked by matrix attachment regions. Plant J. **6**:413-423.

VanBlockland, R., Van der Geest, N., Mol, J.N.M., and Kooter, J.M. (1994) Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. Plant J. **6**:861-877.

Waters, V.L., and Guiney , D.G., (1993) Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. Mol. Microbiol. **9**:1123-1130.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H. (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell **76**:567-576.

Weber, H., Ziechmann, C. and Graessman, A. (1990) *In vitro* DNA methylation inhibits gene expression in transgenic tobacco. EMBO J. **9**:4409-4415.

Weintraub, H., and Groudine, M. (1974) Chromosomal subunits in active genes have an altered conformation. Science **193**:848-853.

Winans, S.C. (1992) Two-way chemical signalling in *Agrobacterium*-plant interactions. Microbiol. Rev. **56**(1):12-31.

Wu, J.C., and Santi, D.V. (1987) Kinetic and catalytic mechanism of *HhaI* methyl-transferase. J. Biol. Chem. **262**:4778-4786.

Yusibov, V.M., Steck, T.R., Gupta, V., and Gelvin S.B. (1994) Association of single-stranded transferred DNA from *Agrobacterium tumefaciens* with tobacco cells. Proc. Natl. Acad. Sci. USA **91**:2994-2998.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M., and Schell, J. (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. EMBO J. **2**:2143-2150.

Zambryski, P.C. (1992) Chronicles from the *Agrobacterium*-plant cell DNA transfer story. Ann. Rev. Plant Physiol. Plant Mol. Biol. **43**:465-495.

Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A., and Chambon, P. (1986) Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5**:387-397.

Zhang, D., Ehrlich, K.C., Supakar, P.C., and Ehrlich, M. (1989) A plant DNA binding protein that recognizes 5-methylcytosine residues. Mol. Cell. Biol. **9**:1351-1356

Zupan, J.R. and Zambryski, P. (1995) Transfer of T-DNA from *Agrobacterium* to the plant cell. Plant Physiol. **107**:1041-1047.